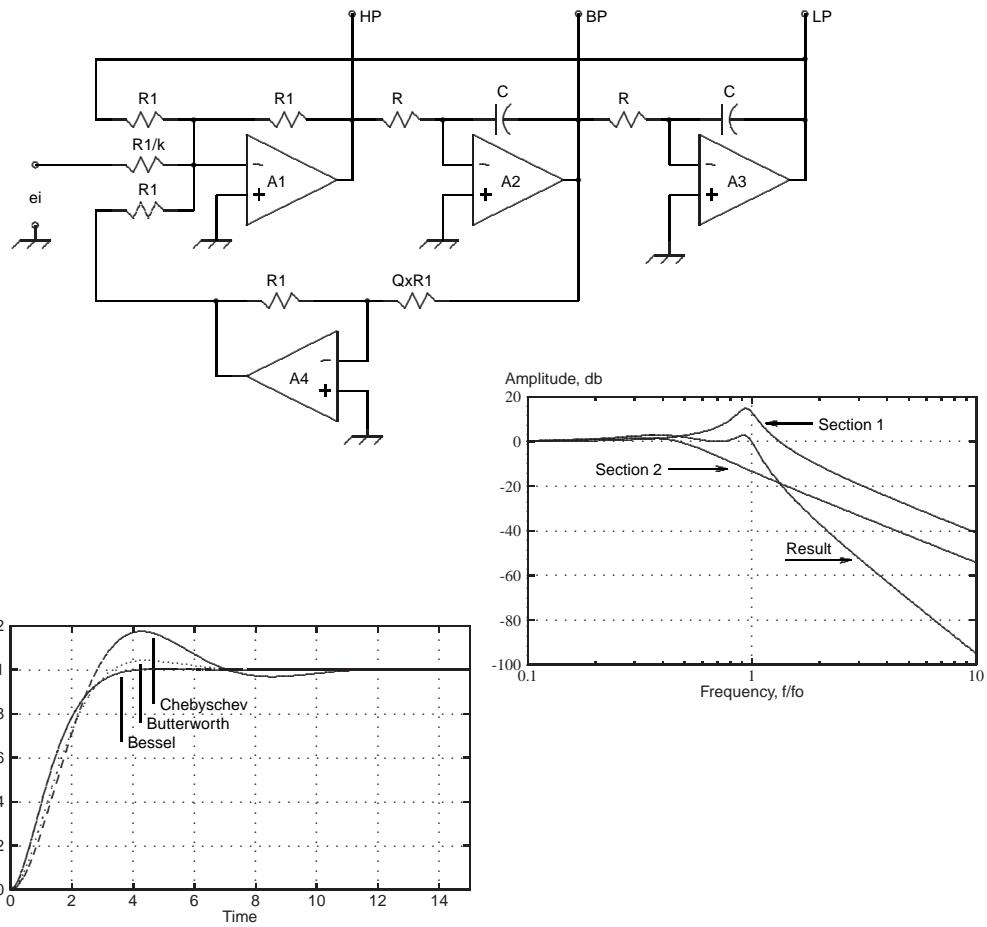


Analog Circuit Design



Second Edition
Peter D. Hiscock

Analog Circuit Design

Peter D. Hiscocks

Professor Emeritus
Department of Electrical and Computer Engineering
Ryerson University

CEO
Syscomp Electronic Design Limited

Email: phiscock@ee.ryerson.ca
First Edition 14 September, 2010
Second Edition 1 March, 2011
©Peter D. Hiscocks, 2010

ISBN 978-0-9867235-0-6

Understanding (an analog design) is like understanding a language. It doesn't take long to look at a schematic and know what it is all about if you know the language.

Bob Dobkin, Vice President, Engineering, Maxim Integrated Circuits
Designer of the LM118 Op Amp
Quoted in [1]

Contents

Table of Contents	iii
List of Figures	xx
1 Introduction	43
1.1 Outline of the Material	46
1.2 Converting Units of Measurement: A Systematic Method	48
2 Basic DC Circuits	50
2.1 Voltage, Current and Resistance	50
2.2 Electron Flow and Conventional Current: The Unlucky Guess of Benjamin Franklin	50
2.3 Electronic Charge	51
2.4 Voltage	51
2.4.1 Example Voltages	52
2.4.2 Voltmeter	53
2.4.3 Voltage Standards	53
2.5 Current	53
2.5.1 Current Meter (Ammeter)	54
2.5.2 Current Standards	54
2.5.3 Example Currents	54
2.5.4 Electron Conduction Physics	54
2.5.5 Velocity of Electron in a Vacuum Tube	55
2.6 Resistance and Ohm's Law	56
2.6.1 Example Resistances	57
2.6.2 Resistance Standards	58
2.6.3 Standard Values	58
2.6.4 Conductivity	58
2.6.5 Designing Resistors	59
2.6.6 European Style Notation	61
2.7 Ideal Voltage Source	61
2.8 Ideal Current Source	62
2.9 Kirchhoff's Voltage Law	63
2.10 Kirchhoff's Current Law	64
2.11 KVL, KCL and Circuit Analysis	65
2.12 <i>Grounded</i> and <i>Floating</i> Components	65
2.13 Resistors in Series	66
2.14 Resistors in Parallel	67
2.15 Resistor Networks	67
2.16 Meters for Voltage, Current and Resistance	68
2.16.1 Moving Coil Meter Movement	68
2.16.2 Characterising an Unknown Meter Movement	68
2.16.3 Moving Coil Voltmeter	69
2.16.4 Moving Coil Ammeter	70
2.16.5 Moving Coil Ohmmeter	71
2.17 Practical Voltage Source	72
2.18 Internal Resistance: Case Histories	74

2.19	Practical Current Source	77
2.20	Measuring Internal Resistance	77
2.21	The <i>Load Line</i>	79
2.22	Voltage and Current Source, Compared	81
2.23	Power Dissipation	83
2.24	Battery Tester	84
2.25	The Capacitor	85
2.25.1	Capacitor Construction	86
2.25.2	Energy Storage in a Capacitor	87
2.25.3	Energy Storage: Battery and Capacitor	87
2.25.4	Common Misconception: Kirchhoff's Current Law and The Capacitor	88
2.26	The Capacitor: Time Domain	88
2.26.1	Defective Circuit	90
2.26.2	Energy Storage in a Capacitor: Reprise	90
2.26.3	Application: Ramp Generator	91
2.27	Capacitances in Series and Parallel	92
2.28	The Inductor	95
2.28.1	Voltage Induction	95
2.28.2	Moving Magnet Induction	95
2.28.3	Faraday's Law and Lenz's Law	96
2.28.4	Inductance	98
2.28.5	Energy Storage in an Inductor	100
2.29	Inductance and Magnetism	101
2.29.1	The Electric Circuit Analogy	101
2.29.2	Reluctance	102
2.29.3	Example: Variable Reluctance Sensor	102
2.29.4	Deriving the Inductance	103
2.29.5	Saturation	104
2.29.6	Summary	105
2.29.7	Solenoid Inductance on Ferrite Rod	106
2.30	The <i>Amplifier</i> Concept	106
2.31	Switches and Relays	108
2.32	Marx Generator	110
2.33	Exercises	112
3	Tools for Circuit Analysis: Part 1	119
3.1	Thevenin's Theorem	119
3.2	Norton's Theorem	121
3.3	The Voltage Divider	121
3.4	Voltage Divider Design	125
3.4.1	Exact Method	125
3.4.2	Simple Method	127
3.4.3	Graphical Method: Divider Design with Standard 5% Resistors	130
3.4.4	Large Ratio Divider Shortcut	131
3.4.5	Derivation of the Nomogram	132
3.5	The Current Divider	133
3.5.1	Current Divider Application: Meter Characterization	133
3.6	The Superposition Theorem	134

3.6.1	Applications	134
3.6.2	Superposition and Power	136
3.7	The Passive Adder	137
3.8	The Wheatstone Bridge	138
3.9	<i>Load Line Application: Digital Logic Interface</i>	139
3.10	Component Tolerances	145
3.11	Cumulative Tolerance	145
3.12	Complex Cases: Spreadsheet to the Rescue	147
3.13	Larger Problems	149
3.14	Error Budget Analysis	149
3.15	Exercises	150
4	Tools for Circuit Analysis: Part 2	154
4.1	Alternating Current	154
4.1.1	Waveforms in General	154
4.1.2	Sinusoidal Waveforms	155
4.2	Vector Notation and Sine Waves	157
4.3	Reactance	157
4.4	Complex Numbers	158
4.5	Capacitive Reactance	161
4.5.1	Capacitive Reactance in Vector Notation	162
4.5.2	Example: Capacitive Reactance	162
4.5.3	Example: Current in a Capacitor	163
4.6	Inductive Reactance	164
4.7	Reactance Calculators	166
4.7.1	Reactance Chart	166
4.7.2	Reactance Slide Rule	168
4.7.3	Gnuplot Routine for Capacitive Reactance Chart	168
4.8	RC Lowpass Network	169
4.8.1	Example: Charging Capacitor	170
4.9	RC Highpass Network	172
4.10	Resonance, an Introduction	173
4.11	Series Resonance	174
4.11.1	Frequency Response	175
4.11.2	Voltages and Currents in the Series Resonant Circuit	176
4.12	Parallel Resonance	176
4.12.1	Frequency Response	177
4.13	Boing vs Thud: the Q Factor	178
4.13.1	Defining Q Factor	178
4.13.2	Q Factor and Energy Dissipation	179
4.13.3	Q Factor and Voltage Magnification	180
4.13.4	Q Factor, Bandwidth and Filter Shape	180
4.14	Bandwidth of a Resonant Circuit	181
4.14.1	Bandwidth and Q Factor	182
4.15	The Compensated Attenuator	184
4.16	The Ideal Transformer	186
4.17	Impedance	190
4.17.1	Impedance	191

4.17.2 Using Impedance	193
4.17.3 Impedance Matching	193
4.18 Exercises	193
5 The Laplace Transform	195
5.1 Overview	195
5.1.1 Laplace Notation as Shorthand	197
5.1.2 A Simple Example: Capacitor Charging Equation	197
5.1.3 Inductor Differential Equation	198
5.1.4 Capacitive Reactance	199
5.1.5 Summary	199
5.2 Transfer Function and Frequency Response	199
5.2.1 Transfer Function of Low Pass RC Filter	199
5.2.2 Magnitude and Phase of The Lowpass RC Filter	201
5.2.3 Measurement	202
5.2.4 Transfer Function of Low Pass LR Filter	203
5.2.5 Measurement	203
5.2.6 Transfer Function of a Second Order RLC Lowpass Filter	204
5.2.7 Amplitude and Phase of the Lowpass RLC Filter	206
5.2.8 Measurement	208
5.2.9 Transfer Function of the Two Stage (Second Order) RC Lowpass Filter	210
5.3 Time Domain Response	211
5.3.1 Computer Algebra: Using Maxima	212
5.3.2 Table of Transforms	214
5.3.3 Table of Transforms, Quadratic	214
5.3.4 Forcing Function: Unit Impulse	215
5.3.5 Forcing Function: Unit Step	216
5.3.6 Forcing Function: Unit Ramp	216
5.3.7 Unit Step and RC Lowpass	220
5.3.8 Unit Ramp and RC Lowpass	223
5.3.9 Capacitor with Initial Charge	226
5.3.10 LC Resonant Circuit	228
5.3.11 RLC Resonant Circuit	232
5.3.12 Two Stage (Second Order) Lowpass, Impulse Response	236
5.3.13 Comparing the Filters	237
5.4 Arbitrary Waveforms	238
5.4.1 Linearity Property	238
5.4.2 Time Shift Property	239
5.4.3 Example: Square Pulse	239
5.4.4 Example: Triangular Pulse	240
5.5 Appendices	240
5.5.1 Laplace Transform of Unit Impulse	240
5.5.2 Laplace Transform of the Unit Step	242
5.5.3 Gnuplot Magnitude and Phase Plotting Routines	242
5.5.4 Further Reading	244
5.6 Exercises	245

6 The Semiconductor Diode	246
6.1 Semiconductor Diode	246
6.2 Diode Logic OR Gate	247
6.3 Diode Logic AND Gate	248
6.4 Half Wave Rectifier	249
6.5 Full Wave Rectifier	249
6.6 Diode Bridge Rectifier	251
6.7 Diode Clamp	252
6.8 Diode Clipping Circuit	252
6.9 Bi-Directional Clipping	253
6.10 The Flyback Diode	254
6.11 Diode Types	254
6.12 Exercises	255
7 The Zener Diode	259
7.1 The Zener Diode	259
7.2 Zener Regulator	259
7.3 Zener Impedance	261
7.4 Zener Regulator Amplifier	262
7.5 Zener Current Drive	263
7.6 The Integrated Circuit Shunt Regulator (IC Zener)	263
7.7 Exercises	265
8 Active Devices	266
8.1 The BJT: Bipolar Junction Transistor	266
8.2 The JFET: Junction Field Effect Transistor	268
8.3 The MOSFET: Metal Oxide Semiconductor Field Effect Transistor	269
8.4 Application Circuit	271
8.5 The MOSFET Switch	272
8.6 The BJT Switch	274
8.7 BJT Logic Devices	276
8.8 The BJT Flip Flop	278
8.9 Exercises	279
9 Motors and Generators	280
9.1 DC Motor and Generator	280
9.1.1 The Ideal DC Motor	281
9.1.2 Real DC Generator Behaviour	282
9.1.3 Speed Control	282
9.1.4 Torque Control	283
9.1.5 Alternatives	283
10 Negative Feedback	284
10.1 Negative Feedback and Amplifiers	286
10.2 Negative Feedback and the Operational Amplifier	287
10.3 Minimizing Disturbances	288
10.4 Negative Feedback Goes Bad	290
10.5 Application: Temperature Controller	291
10.6 Exercises	294

11 Reactance and Bode Plots	300
11.1 Reactance Plot	301
11.2 The Bode Plot	304
11.3 Constant, Integrator and Differentiator Factors	305
11.4 First Order Lead and Lag Factors	307
11.5 First-order Lead	307
11.6 First-order Lag	308
11.7 Example: PID (Proportional, Integral, Differential) Controller	309
11.8 Quadratic Factor	312
11.9 Exercises	313
12 The Operational Amplifier: Basics	315
12.1 Introduction	315
12.2 The Op-Amp Comparator	317
12.3 The Buffer	318
12.4 The Non-Inverting Amplifier	319
12.5 The Inverting Amplifier	320
12.6 Formal Analysis of the Inverting Amplifier	323
12.7 The Rules of Op-Amp Behaviour	325
12.8 Single Supply Operation	326
12.9 Exercises	328
13 The Operational Amplifier: Applications	333
13.1 Miller Effect	333
13.2 The Op-Amp Adder	336
13.3 Non-Inverting Adder	337
13.4 Sign Changer (Inverter)	337
13.5 Switchable Sign Changer	338
13.6 The Subtractor	338
13.7 The Instrumentation Amplifier	341
13.8 Universal Amplifier	343
13.9 Current to Voltage Converter	345
13.10 Integrator, Inverting	346
13.11 Integrator, Non Inverting	349
13.12 Differentiator	351
13.13 Application: Spring-Mass Simulation	354
13.14 Booster Stage	356
13.15 Current Generator, Floating Load	358
13.16 Current Generator, Grounded Load	359
13.17 Charge Amplifier	360
13.18 AC-Coupled Inverting Amplifier	364
13.19 AC-Coupled Inverting Amplifier, Frequency Response	367
13.20 AC-Coupled Non-Inverting Amplifier	370
13.21 AC-Coupled Non-Inverting Amplifier, Frequency Response	372
13.22 AC-Coupled Amplifier, Miscellaneous Issues (Advanced)	375
13.23 Input and Output Resistance	377
13.24 Summary of Amplifier Circuits	383
13.25 Exercises	388

14 The Schmitt Trigger	401
14.1 Introduction	401
14.2 Schmitt Trigger, Inverting	403
14.3 Schmitt Trigger, Non-Inverting	405
14.4 Schmitt Trigger Design Equations	406
14.5 Inverting Schmitt Trigger, Design Equations	407
14.6 Non-Inverting Schmitt Trigger, Design Equations	408
14.7 Summary of Design Equations	410
14.8 Fan Thermostat, Design Example	410
14.9 Appendix: Confirming the Schmitt Trigger Lever Diagram	413
14.10 Exercises	414
15 Power Supplies	416
15.1 AC Power Distribution	416
15.2 Power Supply Configuration	418
15.3 Rectifier-Filter Configurations	421
15.4 Series and Shunt Voltage Regulators	428
15.5 Feedback Voltage Regulator	431
15.6 Three-terminal Linear Regulator	433
15.7 The AC Adaptor	435
15.8 Switching Regulator	437
15.9 Switching Regulator, Design Example	449
15.10 Charge Pump	455
15.11 Voltage Multiplier	459
15.12 Transformer Design	460
15.13 Transformer Design Parameters	464
15.14 Application: Transformer Isolated Interface	467
15.15 Example: Transformer Design	468
15.16 Summary of Power Supply Design	474
15.17 Appendix: The Transformer Voltage Equation	476
15.18 Appendix: Transformer Power Equation	478
15.19 Wire Table	482
15.20 Appendix: Calculating the Effective (RMS) Value of a Waveform	483
15.21 Exercises	488
16 Precision Rectifiers	495
16.1 Precision Diode	497
16.2 Precision Half-Wave Rectifier	499
16.3 Precision Full-Wave Rectifier	500
16.4 Full-Wave Meter Driver	502
16.5 Phase-Sensitive Rectifier	503
16.6 RMS Conversion	505
16.6.1 Obvious Implementation	506
16.6.2 The Implicit Method	506
16.7 Thermal Method	507
16.8 Peak Detector	507
16.8.1 Improved Peak Detector	508
16.9 Sample-Hold Circuit	509

16.10 Exercises	511
17 Active Filters	513
17.1 Introduction	513
17.1.1 Applications	513
17.1.2 Function	514
17.1.3 Symbols	514
17.1.4 Order	515
17.1.5 Shape	515
17.1.6 Sensitivity	516
17.1.7 Implementation	517
17.1.8 Roadmap to the Section	517
17.2 First-Order Filters	518
17.2.1 First-Order Passive Filters	518
17.2.2 First Order Active Filters	520
17.3 Example: A Simple Bandpass Filter	521
17.4 Second-Order Low-Pass Filter	524
17.4.1 Transfer Function of the First Order Cascade	524
17.5 Second-Order Band-Pass Filter	529
17.6 Second-Order High-Pass Filter	531
17.7 Second-Order Band-Reject (Notch) Filter	533
17.8 Summary, Second-Order Filters	536
17.9 The <i>Bronzite</i> Circuit	537
17.10 The <i>Sallen-Key</i> Lowpass Circuit	539
17.11 The <i>Sallen-Key</i> Highpass Circuit	545
17.12 The <i>State Variable</i> Circuit	546
17.13 Notch Filter Circuits	550
17.14 State Variable Notch	550
17.15 The Twin-Tee Network	551
17.15.1 Tuneable Passive Notch Filter	553
17.16 Higher Order Filters	553
17.17 The <i>Butterworth</i> Characteristic	554
17.18 Designing A Butterworth Filter	554
17.19 The <i>Chebyschev</i> Characteristic	557
17.20 The <i>Bessel</i> Characteristic	564
17.21 Plotting High-Order Filter Characteristics	565
17.22 First-Order Transient Response	567
17.23 Second-Order Transient Response	567
17.24 Filter Tables	570
17.25 Design Method Overview	572
17.26 The Complex Frequency Plane	573
17.26.1 Poles and Zeros	573
17.26.2 The Second-Order Lowpass Filter, Normalized	574
17.26.3 Diagram of the Complex Frequency Plane	576
17.26.4 Zeros at Infinity	576
17.26.5 The <i>Membrane</i> Model	577
17.26.6 Amplitude and Phase Response from the Complex Frequency Plane	578
17.26.7 Generalizing the Amplitude and Phase Response	580

17.26.8 Example: The Highpass Filter	580
17.27 Summary of Common Pole-Zero Configurations	582
17.28 Exercises	585
18 The Zoo: Unusual Circuits	593
18.1 NIC: Negative Impedance Converter	593
18.2 Capacitance Multiplier	595
18.3 Howland Current Generator	597
18.4 Simple Gyrator	599
18.4.1 Transient Analysis: A Reality Check	600
18.4.2 Assessment	602
18.5 Improved Gyrator	603
18.6 FDNR: Frequency Dependent Negative Resistance	605
18.6.1 Application	605
18.6.2 Circuit	606
18.7 Circulator	607
18.8 Exercises	609
19 Phase Shift Networks	613
19.1 Passive Network Phase Shifter	613
19.1.1 Improved Phase Shifter	613
19.1.2 Analysis of the All-Pass Network	613
19.1.3 Bode Plots	614
19.2 Op-Amp Phase Shifter	615
19.3 Wideband Phase Shifter	616
19.4 Exercises	618
20 Oscillators	619
20.1 Phase Shift Oscillator	620
20.1.1 A Design Example	620
20.2 Wien Bridge Oscillator	622
20.2.1 References	624
20.3 Wien Bridge Oscillator, Amplitude Stabilization	624
20.4 Quadrature Oscillator	626
20.5 Three-Phase Oscillator	629
20.5.1 Analysis	629
20.6 Schmitt Trigger Astable	631
20.7 Function Generator	633
20.8 Exercises	635
21 The Non-Ideal Operational Amplifier	639
21.1 Finite Gain	640
21.2 Bias Current	641
21.3 Offset Voltage	643
21.4 Power Supply Rejection Ratio	646
21.5 Common Mode Rejection Ratio	649
21.6 Output Voltage and Current Limit	652
21.7 Output Resistance	653
21.8 Slew Rate	655

21.9	Noise and Bandwidth	659
21.10	Exercises	659
22	Operational Amplifier Noise	667
22.1	Resistor Thermal Noise	670
22.2	Noise Spectrum	671
22.3	Effective Noise Bandwidth	673
22.4	Op-amp Noise	674
22.5	Noise Gain	678
22.6	Noise Example: Voltage Follower	679
22.7	Noise Example: Inverter	681
22.8	Noise Formulae	684
22.9	Exercises	686
23	Operational Amplifier Frequency Response	691
23.1	Introduction	691
23.2	Gain-Bandwidth	692
23.3	Clarifying Gain	694
23.4	Maximum Useful Frequency	696
23.5	Single Pole Response	698
23.6	Quadratic Response	701
23.7	Exercises	704
24	Stability	706
24.1	Introduction	706
24.2	Decoupling	707
24.3	Phase Margin and Stability	708
24.4	The <i>Intercept</i> Rule	711
24.5	Enhanced Noise Gain	713
24.6	Output Buffer Stage	716
24.7	Input Capacitance	718
24.8	Frequency Compensation: The Lag-Lead Network	719
24.9	Capacitive Loading	722
24.10	Case History: Oscillating Follower	727
24.11	Exercises	729
25	Digital to Analog Conversion	732
25.1	D/A Specifications	733
25.2	Weighted Resistor D/A	735
25.3	R-2R D/A: Current Steering Mode	736
25.4	R-2R D/A: Voltage Switching Mode	737
25.5	The MDAC: Multiplying Digital to Analog Converter	739
25.6	MDAC Applications	741
25.7	Pulse Width Modulation DAC	744
25.8	Exercises	747

26 Analog to Digital Conversion	749
26.1 Quantization	749
26.2 Signal to Noise Ratio	750
26.3 Aliasing	752
26.4 Feedback Converters	754
26.5 Single Slope Technique	756
26.6 Dual Slope Technique	758
26.7 Sigma-Delta Oversampling Converter	761
26.8 Voltage to Frequency Converter	764
26.9 Exercises	768
27 Opto-Electronic Devices	772
27.1 Introduction	772
27.2 Light Emitting Diode	773
27.3 Photodiode	774
27.4 Photoresistive Cell	776
27.5 Phototransistor	777
27.6 Optocoupler	778
27.7 Exercises	779
28 Cooling	783
28.1 Thermal Circuits	784
28.2 The Derating Curve	785
28.3 Designing a Heatsink: Calculating Thermal Resistance	786
28.4 Heatsink Area	788
28.5 Thermal Capacitance	789
28.6 Fan Cooling	791
28.7 Power Dissipation in an Operational Amplifier	793
28.7.1 Example	794
28.7.2 Solution	794
28.8 Exercises	796
29 The Single Stage BJT Amplifier	798
29.1 Introduction	798
29.2 The Diode Model	799
29.3 BJT Nomenclature	800
29.4 The Current Generator Model	801
29.5 Base Current Control	802
29.6 Base Voltage Control	805
29.6.1 The <i>Ebers-Moll</i> Equation	807
29.7 The Ramp Generator	808
29.8 Emitter Current Control	810
29.9 The Ramp Generator: Improved	811
29.10 Reflecting Resistance	815
29.11 Collector Resistance	817
29.12 Temperature and Base-Emitter Voltage	822
29.13 Temperature and Leakage Current	825
29.14 DC BJT Circuits and Equations	828

29.15 Bias Stability and Beta	833
29.16 Emitter Incremental Resistance and BJT Transconductance	834
29.17 The <i>T</i> Model for the Transistor	837
29.18 Transconductance and Distortion	838
29.19 The Common Emitter Amplifier	839
29.20 The Common Emitter Amplifier, Analysis	841
29.21 Common Emitter Amplifier with Emitter Resistance	847
29.22 Common Emitter Amplifier with Enhancements	851
29.23 The Common Emitter Amplifier, Base Current Drive	858
29.24 Common-Emitter Amplifier with Collector-Base Feedback	861
29.25 A Catalogue of Common Emitter Configurations	866
29.26 Summary: Common-Emitter Equations	869
29.27 The Common Collector (Emitter Follower) Amplifier	870
29.28 Emitter Follower Oscillation	875
29.29 Summary: Emitter Follower Equations	876
29.30 The Emitter Follower as a Negative Feedback System	877
29.31 Emitter Follower Design Example	880
29.32 The Common Base Amplifier	886
29.33 Summary: Common-Base Equations	890
29.34 The Inverted Transistor	891
29.35 Noise in the BJT	892
29.36 BJT Noise Current and Other Considerations	893
29.37 Problems	894
29.38 Appendix: Determining the Transistor Parameters	894
29.39 Appendix: A discussion of Models	900
29.40 Appendix: Base-Emitter Voltage Function of Temperature	900
29.41 Appendix: BJT Datasheet	907
30 BJT-Compounds	911
30.1 The Darlington Pair	911
30.2 The Sziklai Pair	913
30.3 The Current Source	915
30.4 Discrete Component Sink	915
30.5 The Current Mirror	917
30.6 Discrete Component Current Mirror	919
30.7 Multiple-Output Current Mirror	923
30.8 The Wilson Current Mirror	925
30.9 Emitter Resistor and Output Resistance	927
30.10 Appendix: Mirror Emitter Degeneration Resistors	928
30.11 Appendix: Wilson Mirror Output Resistance	929
30.12 The Cascode Pair	931
30.13 Folded Cascode Pair	935
30.14 The Differential Pair	935
30.14.1 BJT Differential Pair, Transfer Characteristic	936
30.15 Enhanced Differential Pair	943
30.15.1 Darlington-Differential	947
30.15.2 Cascode-Differential	947
30.15.3 Differential Stage Coupling	948

30.15.4 Appendix: Differential Amplifier Transfer Characteristic	948
30.15.5 The <i>tanh</i> Function	950
30.15.6 Appendix: Dual Stage Differential Amplifier Bias and Gain	951
30.16 Series-Shunt Feedback Pair	953
30.17 An Operational Amplifier Design	959
30.18 Exercises	966
31 JFET Characteristics and Applications	976
31.1 Constant Current Diode	976
31.2 Transfer Characteristic	978
31.3 The JFET Switch	981
31.3.1 Opto-Isolated JFET Switch	983
31.4 The JFET Amplifier	983
31.4.1 Playing to the Strengths of the JFET	983
31.4.2 BJT and JFET Op-Amps	984
31.4.3 Transconductance of the JFET	984
31.4.4 Transresistance of the JFET	986
31.5 The Source Follower	986
31.5.1 Notes on the Source Follower Circuit	986
31.5.2 The JFET Small Signal Model	987
31.5.3 Voltage Gain of the Source Follower	988
31.6 JFET Follower Oscillation	988
31.6.1 Summary: Source Follower Equations	990
31.7 JFET Common Source Amplifier and Common Gate Amplifier	991
31.8 JFET Compound Amplifiers	991
31.8.1 JFET Differential Amplifier	991
31.8.2 Oscilloscope Buffer Amplifier	991
31.9 Appendix: Interpreting the JFET Datasheet	993
31.10 Exercises	995
32 The MOSFET Family	997
32.1 Linear Applications of the MOSFET Transistor	997
32.2 MOSFET Transfer Characteristic	997
32.3 The Enhancement Mode MOSFET	998
32.4 The Depletion Mode MOSFET	999
32.5 DC Coupled Source Follower	999
32.5.1 Biasing the Follower	999
32.5.2 Voltage Gain of the Follower	1000
32.5.3 Output Resistance of the Follower	1001
32.6 Biasing the Enhancement Mode MOSFET: AC Coupled Follower	1003
32.7 Applying MOSFETs in Linear Applications	1004
32.8 Vertical and Lateral Format MOSFETs	1004
32.9 MOSFET Power Amplifier Output Stage	1004
32.9.1 Drive Current and Voltage	1005
32.9.2 Thermal Stability and Crossover Distortion	1006
32.9.3 Output Voltage Swing	1007
32.9.4 Bandwidth	1007
32.10 Exercises	1007

33 CMOS Analog Switch and Applications	1009
33.1 Applications	1010
33.1.1 Mixer Input Select	1010
33.1.2 Switched Gain Amplifier	1011
33.1.3 Logic Circuits	1012
34 Interference	1013
34.1 Scope of the Problem	1013
34.2 The Mechanism of Interference	1015
34.3 The Source	1015
34.4 The Coupling Mechanism	1015
34.5 Conductive Coupling	1016
34.5.1 Implications for PC Board Layout	1017
34.6 Electric Field Coupling	1017
34.6.1 Electric-Field Detection	1018
34.6.2 Electric-Field Shielding	1018
34.6.3 Demonstration: Electric Field Pickup	1019
34.6.4 Demonstration: Reducing Electric Field Pickup by Lowering Impedance	1020
34.6.5 Demonstration: Reducing Electric Field Pickup by Shielding	1020
34.7 Magnetic Field Coupling	1020
34.7.1 Magnetic Field Identification	1022
34.7.2 Voltage in the Conductor	1022
34.7.3 Demonstration: Magnetic Field Pickup	1024
34.7.4 Magnetic Field Shielding	1024
34.8 Spectrum of Pulse Waveform	1026
34.8.1 Square Wave Spectrum	1026
34.8.2 Trapezoidal Wave Spectrum	1027
34.8.3 The Bode Approximation	1027
34.8.4 Derivation: Spectrum of Trapezoidal Waveform	1028
34.9 Components of The Hidden Schematic	1029
34.9.1 Wiring Resistance	1029
34.9.2 Resistance of Printed Wiring Traces	1031
34.9.3 Wiring Inductance	1032
34.9.4 Inductance of a Printed Wiring Trace	1033
34.9.5 Power Supply Inductance and The Bypass Capacitor	1034
34.9.6 Parasitic Inductance in Capacitors	1036
34.9.7 Capacitor Resonance	1037
34.9.8 Broadbanding the Bypass Capacitor	1037
34.10 Common Mode and Differential Signals	1038
34.10.1 Rejection of Ground Conductor Noise	1038
34.10.2 Differential Amplifier Enhancements	1039
34.10.3 Guarding	1040
34.10.4 Bifilar Choke	1040
34.11 EMC: Electromagnetic Compliance	1041
34.11.1 EMI from Differential and Common Mode Currents	1042
34.11.2 Common Mode Current, EMI, and the Bifilar Choke	1043
34.11.3 Differential Drive and Common-Mode EMI Current	1044
34.11.4 Evaluating EMC	1044

34.12 Electromagnetic Field Coupling	1045
34.13 Measurement Technique	1046
34.13.1 Single Lead Connection	1047
34.13.2 Twin Lead Connection	1047
34.13.3 Coaxial Cable Connection	1048
34.13.4 x10 Probe Connection	1049
34.13.5 Active Probe Connection	1049
34.14 Printed Wiring Board Layout	1049
34.14.1 A/D Converter Layout and Grounding	1051
34.14.2 Ground Plane and the A/D Converter	1054
34.15 Case Histories: Solved	1054
34.16 Exercises	1055
35 Beyond the Op-Amp	1058
35.1 Introduction	1058
35.1.1 Voltage-Mode and Current Mode Circuits	1058
35.1.2 Why is the Op-Amp so Popular?	1058
35.1.3 Why Consider Other Amplifier Architectures?	1059
35.1.4 Other Types of Amplifiers	1059
35.2 The Current Conveyor	1060
35.2.1 Example	1060
35.2.2 Internal Circuit Operation	1061
35.2.3 Application	1061
35.3 The <i>Current Feedback Amplifier</i>	1062
35.3.1 Basic Operation	1063
35.3.2 Gain Analysis	1064
35.3.3 Loop Gain	1064
35.3.4 Maximum Frequency of Operation	1065
35.3.5 Transresistance Equivalent Circuit	1065
35.3.6 Slew Rate	1066
35.3.7 References	1066
35.4 The <i>Norton</i> Amplifier	1067
35.4.1 Norton Internals	1067
35.4.2 Equivalent Circuit	1067
35.4.3 Bias for AC Amplification	1068
35.4.4 AC Amplifier	1068
35.4.5 Discussion	1068
35.5 Operational Transconductance Amplifier	1069
35.5.1 Internal Circuitry	1069
35.5.2 OTA Application: Voltage Controlled Amplifier	1070
35.6 The Comparator Device	1071
35.7 Instrumentation Amplifier	1072
35.8 Fully Differential Amplifier	1074
35.8.1 Differential and Common-Mode Signals	1074
35.8.2 Differential Signalling	1075
35.8.3 Internal Circuit	1076
35.8.4 Example Design: Transient Capture	1076
35.8.5 Rules of Ideal Behaviour	1077

35.8.6 Differential Gain Analysis	1078
35.8.7 Common-Mode Analysis	1079
35.8.8 Design Example: Single-Ended to Differential Converter	1080
35.8.9 Selecting Resistor Value	1082
35.9 Autozero Amplifier	1083
35.10 Chopper Amplifier	1085
35.11 Composite Amplifier	1088
35.12 Exercises	1089
36 Multipliers	1094
36.1 Introduction	1094
36.2 MDAC: Multiplying Digital to Analog Converter	1096
36.3 Duty Cycle (Pulse Width-Height) Multiplier	1096
36.4 Transconductance Multiplier	1098
36.5 Exercises	1103
37 Complete Systems	1105
37.1 Equation Solver	1105
37.2 DC Motor Controller	1107
37.2.1 Real DC Motor Behaviour	1107
37.2.2 The Control System	1108
37.2.3 Control System Equations	1108
37.2.4 Electronic Circuit	1110
37.2.5 Results	1111
37.3 Sumo Competition Robot	1112
37.3.1 Robot Behaviour	1112
37.3.2 Robot Power Supply	1113
37.3.3 Motor Driver Circuit	1114
37.3.4 Edge Detection Subsystem	1115
37.3.5 Object Detector and Speed Control	1116
37.3.6 Object Detector Circuit	1117
37.4 The Synchro Resolver	1119
37.4.1 Resolver Interface, Design	1121
37.4.2 Resolver Interface, Circuit Diagram	1122
37.5 The Linear Variable Differential Transformer	1124
37.6 Analog Correlation Techniques	1125
37.6.1 Correlation Overview	1125
37.6.2 The Correlation Time Delay	1125
37.6.3 Correlation and Signal Processing	1126
37.6.4 Correlation and the Phase Sensitive Rectifier	1128
37.7 Barometer Interface	1129
37.7.1 The Pressure Sensor	1129
37.7.2 Sensor Specifications	1130
37.7.3 Pressure Sensor Equivalent Circuit	1130
37.7.4 Barometer Resolution and Dynamic Range	1131
37.7.5 Transfer Function	1132
37.7.6 Designing the Circuit	1133
37.7.7 Complete Circuit	1135

37.7.8 Reliability of the Design	1136
37.7.9 Circuit Analysis for Component Tolerance and Temperature Drift	1137
37.7.10 Calibration	1138
37.8 Exercises	1140
37.9 Design Exercises	1142
37.10 Extended Design Exercise: Telescope Heater	1146
Bibliography and References	1153
Index	1175

List of Figures

Basic DC Circuits	50
1 Hydraulic Analogy	50
2 Ohm's Law	56
3 Incremental Resistance	56
4 1% Standard Values	58
5 European Style Notation	61
6 Ideal Voltage Source	61
7 Identical Voltage Sources	62
8 Ideal Current Source	62
9 Kirchhoff's Voltage Law	63
10 Kirchhoff's Voltage Law	63
11 Kirchhoff's Current Law	64
12 Floating and Grounded Components	66
13 Series Connections	66
14 Parallel Connections	67
15 Resistor Network	67
16 Moving Coil Meter	68
17 Characterizing a Moving Coil Meter	68
18 Voltmeter, 10V FSD	69
19 Ammeter, 1A FSD	70
20 Constant Current Ohmmeter	71
21 Constant Voltage Ohmmeter	72
22 Practical Voltage Source	72
23 Practical Source with Load	73
24 Jumpering a Car Battery	76
25 Practical Current Source	77
26 Output Resistance	77
27 Test Generator Method	78
28 Blocking the DC current	78
29 The <i>Load Line</i>	79
30 The <i>Load Line</i> with Load Resistance	79
31 Load Line with AC Voltage	79
32 Source with $R_L \gg$ Internal Resistance	81
33 Source with $R_L \ll$ Internal Resistance	82
34 Power Dissipation	83
35 Battery Tester	84
36 Capacitor Symbol	85
37 Capacitor Operation	85
38 Charging a Capacitor	87
39 Energy Density	88
40 Capacitor Integrator	89
41 Capacitor Behaviour	90
42 Defective Circuit	90
43 Ramp Generator	91
44 Practical Ramp Generator	91

45	Capacitors in Parallel	92
46	Capacitors in Series	92
47	High Voltage Capacitor Array	94
48	Inductor Symbol	95
49	Magnetic Induction Waveform	95
50	Self Induction	96
51	Self Induction Waveforms	97
52	Inductor Snubber Network	98
53	Inductor Transient	99
54	Inductor Energy Storage	100
55	Magnetic Units	101
56	Magnetic Circuit Analogy	101
57	Rotation Detector	102
58	BH Curve	104
59	Flux and Voltage Waveforms	104
60	Ferrite Rod Permeability [2]	106
61	Amplifier	107
62	Unidirectional Mechanical Amplifier	107
63	Pneumatic Amplifier	108
64	Switch Types	109
65	SPDT Relay	110
66	Marx Generator	111
67	Capacitor Currents	117
	Tools for Circuit Analysis: Part 1	119
68	Thevenin's Theorem	119
69	Calculating R_{int}	119
70	Thevenizing a Voltage Divider	120
71	Norton's Theorem	121
72	Voltage Divider Evolution	121
73	Voltage Divider	122
74	Voltage Divider with n Resistors	122
75	Voltage Divider Exercise	123
76	Voltage Divider	124
77	Voltage Divider Equivalent Circuit	124
78	Voltage Divider Example	124
79	Voltage Divider Design	125
80	Voltage Divider Design	127
81	Voltage Design Rule	128
82	Voltage Divider, Approximate Method of Design	128
83	Voltage Divider Nomograph	130
84	Large Ratio Divider	132
85	Nomograph Derivation	132
86	Current Divider	133
87	Meter Characterization	133
88	Superposition Theorem Example	135
89	Superposition Example, cont'd	136
90	Superposition and Power	136

91	The Passive Adder	137
92	The Wheatstone Bridge	138
93	Measuring R_x	138
94	Logic Device Equivalent Circuit	140
95	4000 Series CMOS Gate	141
96	The Effect of a Load Resistor	141
97	LSTTL Output Characteristic	142
98	Interfacing Options: Source or Sink	144
99	Binary Sequence	147
100	Tolerance Example	148
Tools for Circuit Analysis: Part 2		154
101	Alternating Voltage Source	154
102	Waveforms	154
103	AC Generator	155
104	Sine Wave and Vector	157
105	The Number Line	158
106	Cartesian Graph	158
107	Rectangular Notation	159
108	Imaginary Operator j	159
109	Polar Notation	160
110	AC Voltage and Current in a Capacitor	161
111	Impedance Vector Diagram	162
112	Capacitive Reactance vs Frequency	164
113	AC Voltage and Current in an Inductor	164
114	Impedance Vector Diagram	165
115	Inductive Reactance vs Frequency	165
116	Reactance Chart for Capacitance	167
117	Reactance Slide Rule	168
118	RC Lowpass Filter	169
119	RC Lowpass Charging Curve	170
120	RC Lowpass Amplitude and Phase Response	171
121	RC Highpass Filter	172
122	RC Highpass Step Response	172
123	RC Highpass Amplitude and Phase Response	173
124	Tank Circuit and Waveforms	174
125	Series Resonant Circuit	174
126	Series RLC Circuit Response	175
127	Parallel Resonant Circuit	176
128	Parallel RLC Circuit Response	177
129	Damped Oscillations	178
130	Bandwidth	181
131	The Compensated Divider	185
132	Compensation Adjustment	185
133	Ideal Transformer	186
134	Resistance and Source Reflection	187
135	Transformer with Multiple Windings	189
136	Winding Connections	189

137	Impedance Summary	191
138	Impedance, Example	192
The Laplace Transform		195
139	Laplace Transform Method	195
140	Impulse Noise Filter	196
141	Noise Filter Candidates	196
142	Differentiator and Integrator	197
143	RC Lowpass Filter	199
144	Amplitude and Phase of RC Lowpass Network	201
145	RC Lowpass Measurement	202
146	LR Lowpass Filter	203
147	LR Lowpass Measurement	204
148	Second-Order Lowpass Filter	204
149	Amplitude and Phase of RLC Lowpass Network	206
150	RLC Lowpass Measurement	209
151	2nd Order RC Lowpass Filter	210
152	Maxima Screen Shot	213
153	Impulse Signal	215
154	Unit Step	216
155	Unit Ramp	217
156	Impulse with RC Lowpass	217
157	Impulse and RC Lowpass	219
158	Step with RC Lowpass	220
159	Step and RC Lowpass	223
160	Ramp with RC Lowpass	223
161	RC Lowpass Output vs Time with Ramp Kt Input	224
162	Ramp and RC Lowpass	226
163	Capacitor with Initial Charge in RC Circuit	227
164	LC Resonant Circuit and Waveforms	228
165	Current and Voltage in LC Resonant Circuit	230
166	RLC Resonant Circuit	233
167	Damped Sine Wave	235
168	2nd Order RC Lowpass Filter	236
169	Impulse Response of 2nd Order Lowpass Filter	238
170	Shift Property	239
171	Pulse Waveform	239
172	Triangular Pulse	240
The Semiconductor Diode		246
173	The Semiconductor Diode	246
174	Diode Equation	247
175	Diode OR Gate	247
176	Diode AND Gate	248
177	Half-Wave Rectifier	250
178	Full-Wave Centre-tapped Rectifier	250
179	Full-Wave Centre-Tapped Supply	250
180	Bridge Rectifier	251

181 Bridge Rectifier Power Supply	251
182 Clamp	252
183 Clip Circuit	253
184 Bidirectional Clip Circuit	253
185 Inductor Flyback Diode	254
186 Diodes in Current Use	254
187 Obsolete Diodes	255
The Zener Diode	259
188 The Zener Diode	259
189 Zener Diode Reference Circuit	260
190 Zener Regulator with Load	261
191 The Zener Incremental Resistance	261
192 The Zener Incremental Resistance	261
193 Zener with Current Amplifier	262
194 Current-Driven Zener Regulator	263
195 The IC Zener	264
Active Devices	266
196 Controlled Sources	266
197 Bipolar Junction Transistor	266
198 BJT Saturation	267
199 Junction Field Effect Transistor	268
200 JFET Saturation	269
201 Alternative MOSFET Symbol	269
202 MOSFET	270
203 MOSFET Motor Drive Circuit	271
204 MOSFET Switch	272
205 Transfer Characteristic	273
206 MOSFET Switch Example	273
207 NPN BJT Switch	274
208 Transfer Characteristic	276
209 BJT Logic Gates	277
210 BJT Flip Flop	278
Motors and Generators	280
211 Generator	280
212 Ideal Motor, Speed-Torque Characteristic	281
213 Real Motor, Equivalent Circuit	281
214 Real Motor, Speed-Torque Characteristic	282
215 Real Motor with Fan Load	282
Negative Feedback	284
216 Block Diagram Elements	284
217 Driver and Vehicle	284
218 Negative Feedback System	286
219 Negative Feedback using the Operational Amplifier	287
220 Modelling a Disturbance	288
221 Temperature Controller	292

222 Thermistor Characteristic	292
223 Oven with Diode Sensor	298
224 Thermistor Oven Controller	299
Reactance and Bode Plots	300
225 Reactance Types	301
226 Reactance Plots, Series and Parallel	302
227 Reactance Phase Plot	302
228 Example Network	303
229 Reactance Plots	303
230 Constant Factor	305
231 Differentiator	306
232 Integrator	306
233 First Order Lead	307
234 First-order Lead	308
235 First Order Lag	308
236 First-order Lag	309
237 Industrial Control System	309
238 PID Controller	309
239 PID Controller Bode Plot	310
240 Quadratic Factor, Magnitude	312
241 Quadratic Factor, Phase	312
The Operational Amplifier: Basics	315
242 Negative Feedback using the Operational Amplifier	315
243 Op Amp Power and Input Voltages	315
244 Op Amp Potentiometer Analogy	316
245 Common Mode and Differential Inputs	316
246 Op-Amp Comparator	317
247 Zero Crossing Detector	318
248 Operational Amplifier Buffer	318
249 Capacitively Coupled Buffer	319
250 Non-Inverting Amplifier	319
251 Geometric Construction	320
252 Non-Inverting Amplifier Lever Diagram	320
253 Amplifier Configurations	321
254 Inverting Amplifier	321
255 Inverting Amplifier Lever Diagram	322
256 Pan Pot	322
257 Inverting Amplifier Circuit Model	323
258 Inverting Amplifier	325
259 Split Supply Operation	326
260 Single Supply Operation	326
261 Inverting Amplifier, Split Supply	327
262 Inverting Amplifier, Single Supply	327
263 Pseudo Ground	328
264 Operational Amplifier Circuits	329

The Operational Amplifier: Applications	333
265 Miller Effect	333
266 Applying the Miller Effect	334
267 Miller Effect with Capacitance	334
268 Miller Capacitance and High Frequency Response	335
269 Op-Amp Adder	336
270 Offset-generator	336
271 Non-Inverting Adder	337
272 Op-Amp Sign Changer	337
273 Op-Amp Switchable Sign Changer	338
274 Op-Amp Subtractor #1	338
275 Difference Amplifier	339
276 High Impedance Input Diff Amp	340
277 Instrumentation Amplifier	341
278 Generalized Inverting/Non-Inverting Amplifier	343
279 Lever Diagrams	343
280 Voltage Distribution Diagram, Generalized Amplifier	344
281 Current to Voltage Converter	345
282 Equivalent Circuit	346
283 Integrator Circuit	346
284 Waveforms and their Integrals	347
285 Integrator, Frequency Domain	348
286 Op-Amp Integrator Frequency Response	349
287 Non-Inverting Integrator	349
288 Howland Non-Inverting Integrator	350
289 Differentiator	351
290 Waveforms and their Differentials	351
291 Differentiator, Frequency Domain	352
292 Op-Amp Differentiator with Gain	352
293 Improved Differentiator	353
294 Mass-Spring System	354
295 Mass-Spring Simulation	355
296 Basic Booster	356
297 Improved Booster Stages	357
298 Current Generator, Floating Load	358
299 Current Generator, Increased Output	358
300 Current Generator, Grounded Load	359
301 Current Generator, Grounded Load	359
302 Charge-Amplifier	360
303 Piezo-Electric Accelerometer	360
304 Electret Microphone	360
305 Charge Sensor Interface	361
306 Charge Sensor Interface, Improved	361
307 Basic Charge Amplifier	362
308 Complete Charge Amplifier	362
309 Charge Amplifier Example	363
310 Charge Amplifier Frequency Response	364
311 Defective Amplifier	364

312 Better Amplifier	365
313 Working Amplifier	365
314 AC Coupled Inverting Amplifier Signals	366
315 Inverting Amplifier, Single Supply, Equivalent Circuits	366
316 AC Coupled Inverting Amplifier	367
317 Bode Amplitude Plot for G_{13}	367
318 Bode Amplitude Plot for G_{34}	368
319 Completed Bode Amplitude Plot	369
320 Evolution of the Non-Inverting AC Coupled Amplifier	370
321 Non-Inverting AC Coupled Amplifier Equivalent Circuits	371
322 AC Coupled Non-Inverting Amplifier Signals	371
323 AC Coupled Amplifier Sub-Circuits	372
324 Bode Plot for G_{23}	373
325 Design Example	374
326 Example Design Completed	374
327 Offset Voltage in AC Coupled Amplifier	375
328 Inverting and Non-Inverting AC Coupled Amplifiers	376
329 Split Supply with Limited Frequency Response	376
330 Measuring R_x	377
331 Non-Inverting Amplifier: Input Resistance	378
332 Input Resistance: Analysis	379
333 Output Resistance	380
334 Current-Voltage Converter: Input Resistance	381
335 Current Generator, Output Resistance	381
336 Output Resistance: Example	383
337 Generalized Amplifier	383
338 Transimpedance Amplifier Application	384
339 Ideal Voltage Amplifier	384
340 Transconductance Amplifier Application	385
341 Relating Amplifier Parameters	386
342 Temperature Regulator Circuit	392
343 Pressure Measurement Circuit	392
344 Differential Amplifier	393
The Schmitt Trigger	401
345 Noisy Input to a Comparator	401
346 Mechanical Hysteresis	401
347 Comparator and Schmitt Transfer Function	402
348 Inverting Schmitt Trigger	403
349 Lever Diagram, Inverting Schmitt Trigger	403
350 Schmitt Trigger Design Requirements	404
351 Inverting Schmitt Trigger	404
352 Non-Inverting Schmitt Trigger	405
353 Non-Inverting Schmitt Trigger	406
354 Schmitt Voltage Divider	406
355 Inverting Schmitt Trigger	407
356 Schmitt Trigger Design Requirements	408
357 Non-Inverting Schmitt Trigger	409

358	Fan Thermostat	411
359	Inverting Schmitt Trigger Lever Diagram	413
	Power Supplies	416
360	AC Power Distribution Wiring	417
361	AC Power Wiring Safety	418
362	A DC Power Supply	418
363	Configuration Set #1	421
364	Rectifier Configurations #2	422
365	Half-Wave Rectifier, Waveforms	423
366	Transformer Resistance	425
367	Rectifier Simulation	426
368	Rectifier-Inductor Filters	427
369	Series Regulator	428
370	Shunt Regulator	429
371	Zener Regulator	430
372	Amplified Zener Regulator	430
373	Op Amp Regulator	431
374	Regulator with Series Pass Transistor	431
375	Regulator with Current Limit	432
376	Regulator with Remote Sensing	432
377	3 Terminal Voltage Regulators	433
378	Regulator with NPN Output	434
379	PNP LDO Regulator	435
380	AC Adaptor Output	436
381	Switching Regulator Concept	437
382	Switching Regulator Example	439
383	Switching Regulator Concept, With Diode	439
384	Switching Regulator Waveforms	440
385	Capacitor Waveforms	443
386	Capacitor Equivalent Circuit	444
387	Vector Diagram	445
388	Switching Regulator, Discontinuous Mode Waveforms	446
389	Output Voltage during Discontinuous Operation	447
390	Switching Regulator, Basic Circuit	449
391	Switching Regulator, Internal Circuit	450
392	Inductor Voltage and Current	452
393	Switching Regulator Layout	455
394	Charge Pump Inverter	456
395	Charge Pump Doubler	457
396	Charge Pump Efficiency	457
397	Voltage Doubler	459
398	Voltage Multiplier (N=6)	459
399	Voltage Multiplier Redrawn	460
400	Power Transformers	461
401	Power Supply Block Diagrams	461
402	Simple Transformer Drive	462
403	Bridged Transformer Drive	463

404 Push-Pull Transformer Drive	464
405 Transformer Saturation	464
406 Toroid Core	466
407 Transformer Isolated Interface	467
408 Toroid Core Dimensions	469
409 Magnetizing Inductance	470
410 Transformer Equivalent Circuit	473
411 Transformer Symbols	476
412 Candidate Toroid Core	479
413 RMS Calculation, Graphical Method	484
414 Sawtooth Waveform	485
415 Pulse Waveform	486
Precision Rectifiers	495
416 Sine Wave Amplitude Measurements	495
417 Half-Wave Rectifier	497
418 Precision Diode	498
419 Op Amp Output Waveform	498
420 Precision Half Wave Rectifier	499
421 Op Amp Output Waveform	499
422 Precision Full Wave Rectifier, Block Diagram	500
423 Precision Full Wave Rectifier, Circuit	500
424 Full-Wave Rectifier #3	500
425 Alternative Full Wave Rectifier	501
426 Analysis	502
427 Full Wave Meter Driver	503
428 Phase-Sensitive Rectifier	503
429 Phase Sensitive Rectifier, Waveforms	505
430 RMS Conversion	505
431 RMS Conversion Circuit	506
432 RMS Conversion, Implicit Method	506
433 RMS Conversion, Thermal Method	507
434 Simple Peak Detector	507
435 Improved Peak Detector	508
436 Basic Sample Hold	509
437 Sample Hold Circuits	510
Active Filters	513
438 A Lowpass Filter	513
439 Filter Functions	514
440 First and Second Order Lowpass Filters	515
441 Chebyshev and Butterworth Characteristics	516
442 Filter Implementations	517
443 First Order Passive Filter Characteristics	519
444 First Order Active Filter Characteristics	520
445 Bandpass Filter	521
446 Filter Bode Plot	522
447 Cascade of First-Order Lowpass Sections	524

448	Response of Cascaded RC Sections	524
449	Lowpass Filter, Magnitude Plot	525
450	Second-Order Lowpass Filter	525
451	Lowpass Filter, Phase Plot	528
452	Bandpass Filter	529
453	Bandpass Filter Amplitude	530
454	Bandpass Filter Phase	531
455	Highpass Filter	531
456	Highpass Filter, Magnitude Plot	532
457	Highpass Filter, Phase Plot	533
458	Band-Reject Filter	533
459	Notch Filter, Magnitude Plot	535
460	The <i>Bronzite</i> Lowpass Filter	537
461	Bronzite Filter Analysis	537
462	The <i>Bronzite</i> Highpass Filter	538
463	The <i>Bronzite</i> Lowpass Filter, DC Current Path	539
464	Sallen-Key Lowpass Filter	539
465	Sallen-Key Lowpass Filter, Unity Gain	541
466	Sallen-Key Filter Drives Coaxial Cable	542
467	Sallen-Key Lowpass Filter Designs	544
468	Sallen-Key Highpass Filter Designs	545
469	The Series Resonant Circuit	546
470	Series Resonant Circuit Simulation	548
471	State Variable Filter	548
472	State Variable Notch Filter	550
473	Twin-Tee Notch Filter	552
474	Twin-Tee Notch Filter, Analysis	552
475	Tunable Twin-Tee Notch Filter	553
476	Graphic Equalizer Block Diagram	553
477	Butterworth Characteristic	554
478	Fifth Order Lowpass Filter	556
479	Butterworth Highpass Characteristic	557
480	Chebyschev and Butterworth Characteristics	557
481	Fifth Order Highpass Filter	562
482	Bessel Magnitude Characteristic	564
483	Bessel Phase Characteristic	565
484	Fourth Order Chebyschev Filter	565
485	Chebyschev Components and Result	566
486	Second-Order Lowpass, Transient Response	567
487	Chebyschev, Butterworth and Bessel Step Response	569
488	High-Order Step Responses	570
489	Pole Migration	576
490	Membrane Model, Lowpass	577
491	Membrane Model, Highpass	578
492	Response Calculation	579
493	Pole-Zero Response	580
494	Highpass Filter Response from Pole-Zero Pattern	581
495	Filter Circuit	585

496 Octave Filter Response	587
497 Multiple Feedback Bandpass Filter	587
498 Multiple Feedback Bandpass Filter, Modified	588
499 State Variable Filter Variant	589
500 Biquad Filter	590
The Zoo: Unusual Circuits	593
501 Negative Impedance Converter	593
502 NIC Analysis	594
503 Negative Incremental Resistance	595
504 Capacitance Multiplier	595
505 Howland Current Generator	597
506 Howland Circuit Equivalent	597
507 Alternate Input	598
508 Gyrator	599
509 Gyrator Transient Response	600
510 Gyrator Responses	602
511 Gyrator Example	602
512 Gyrator	603
513 Gyrator Analysis	604
514 Response of the D Element	605
515 D Element Example	606
516 Frequency Dependent Negative Resistance	606
517 Three Port Circulator	607
518 Circulator Applications	607
519 Op-Amp Circulator	608
Phase Shift Networks	613
520 Phase Shifter, Passive	613
521 All-Pass Network, Magnitude Response	614
522 All-Pass Network, Phase Response	614
523 Op-Amp Phase Shifter	615
524 Single Sideband Generation, Phasing Method	616
525 Wideband Phase Shifter 1	616
526 Wideband Phase Shifter 2	617
527 Wideband Phase Shifter 2, Phase Response	618
Oscillators	619
528 Phase Shift Network	620
529 Phase Shift Oscillator	620
530 1 KHz Phase Shift Oscillator	622
531 Wien Bridge Oscillator	622
532 Wien Bridge Analysis	623
533 Wien Bridge Magnitude and Phase	623
534 Lamp Stabilization	624
535 Diode Limiter	625
536 Integrator-Comparator Limiter	625
537 Quadrature Oscillator	627

538 Quadrature Oscillator, Howland Type	628
539 Quadrature Oscillator, State Variable Type	628
540 Three Phase Oscillator	629
541 Section Magnitude and Phase	630
542 Schmitt Trigger Astable	631
543 Astable Waveforms	632
544 Function Generator	633
545 Function Generator Waveforms	633
546 Level Clamping Circuit	634
547 Function Generator with Controlled Inverter	635
The Non-Ideal Operational Amplifier	639
548 Bias Current Model	641
549 Bias Current Example	642
550 Bias Current Correction	642
551 Offset Voltage Model	643
552 Offset Voltage Example	644
553 Offset Voltage Correction	645
554 Power Supply Rejection Ratio, Model	647
555 PSRR vs Frequency	648
556 Common-Mode Rejection Ratio, Model 1	649
557 Common-Mode Rejection Ratio, Model 2	649
558 Common-Mode Rejection Ratio, Model 3	650
559 Difference Amplifier	650
560 Inverting and Non-Inverting Amplifiers	651
561 CMRR vs Frequency	651
562 Positive Output Characteristic	652
563 Output Resistance Model	653
564 Output Resistance Analysis	654
565 Slew Rate Limiting	655
566 Large Signal Frequency Response	658
Noise	667
567 Noise Waveform	667
568 Probability Density Function	667
569 Resistor Noise Measurement	670
570 Resistor Noise Equivalents	671
571 Octave Filter Band Edges	672
572 Noise Bandwidth	673
573 Noise Bandwidth of a Bandpass Filter	674
574 Op-Amp Noise Model	674
575 NE5534 Voltage Noise Spectrum	676
576 Noise Gain	678
577 Operational Amplifier Voltage Follower	679
578 Voltage Follower Noise Model	680
579 Operational Amplifier Inverter	681
580 Inverter Noise Model	682
581 Op-Amp Noise Spectrum	684

582 Flicker Noise Power	685
583 Magnetometer Circuit	689
584 Alternative Magnetometer Circuit	689
Frequency Response	691
585 Open-Loop Gain vs Frequency	691
586 Op Amp Open Loop Response	692
587 Dominant Pole Response	692
588 Gain-Bandwidth	693
589 741 Gain-Bandwidth	693
590 Open and Closed Loop Gain	695
591 Open and Closed Loop Response	696
592 741 Gain-Bandwidth	697
593 Single Pole Response	698
594 Single Pole Transient Response	699
595 Two Pole Op-Amp Frequency Response	701
596 Lowpass Filter, Magnitude Plot	702
597 Two Pole System Example	703
598 Closed Loop Responses	704
Stability	706
599 Single Supply Decoupling	707
600 Capacitive Reactance vs Frequency	707
601 Op Amp Open Loop Response	709
602 Loop Gain and Phase	709
603 Q Factor, Phase Margin and Frequency Response	710
604 Q Factor and Step Response	710
605 Increased Closed Loop Gain	711
606 Differentiator	712
607 Improved Differentiator	712
608 Reduced Noise Gain	713
609 Stability Analysis Circuit	713
610 Improved Circuit	714
611 Noise Gain Network	714
612 Op-Amp with Buffer Stage	716
613 Amplifier and Buffer Bode Plots	717
614 Stabilized Amplifier+Buffer	717
615 Input Capacitance	718
616 Lag-Lead Network Bode Plot	720
617 Lag-Lead Compensation	722
618 Capacitive Loading	723
619 Loop Gain	723
620 Compensation for Load Capacitance, 1	724
621 Lag-Lead Network	724
622 Enclosed Buildout Resistance	725
623 Figure 622(b) Redrawn	725
624 Compensated Follower	726
625 Compensation using Compensated Attenuator	727

626 Oscillating Follower	728
627 Oscillating Follower, Fixed	728
628 Stabilizing the Voltage Follower	729
629 Amplifier with Booster	730
630 Step Response	730
Digital to Analog Conversion	732
631 3 Bit D/A Converter	733
632 Monotonicity	734
633 Settling Time	734
634 Output Glitch	735
635 Weighted Resistor D/A	735
636 R-2R Network	736
637 R-2R Current Steering D/A	737
638 R-2R Voltage Switching D/A	737
639 MSB Analysis	738
640 NSB Analysis	738
641 R-2R Voltage Switching D/A, Inverting Buffer	739
642 CMOS Driven Voltage Switching D/A	739
643 Basic MDAC, 8 bits	739
644 Gain Control with Single Quadrant MDAC	740
645 Two-Quadrant MDAC	740
646 Four-Quadrant MDAC Example	741
647 MDAC Gain Control	741
648 MDAC Controlled Function Generator	742
649 MDAC Frequency Control	743
650 D/A Conversion using Pulse Width Modulation	744
651 The Ripple Waveform	745
652 PWM Spectrum	746
Analog to Digital Conversion	749
653 Signal-Noise Ratio	749
654 Quantization Noise	749
655 Normal Sampling	752
656 Undersampling	753
657 Anti-alias Filter	753
658 A/D Feedback Converter Concept	754
659 Up-Counter Converter Waveforms	754
660 Up-Down Counter Converter Waveforms	754
661 Successive Approximation Converter Waveforms	755
662 Single Slope Technique	756
663 Single Slope A/D, Simplified	757
664 Single Slope A/D, Alternate Configuration	758
665 Dual Slope Technique	759
666 Sigma-Delta Converter, Block Diagram	761
667 Quantization Noise, Normal Sampling	762
668 Quantization Noise, Oversampling	762
669 Quantization Noise, Oversampled and Decimated	763

670 Delta-Sigma Circuit, Noise Analog	763
671 Voltage to Frequency Conversion	765
672 V to F Application	768
673 Dual Slope A/D Converter	769
674 Voltage-Frequency Converter	770
Opto-Electronic Devices	772
675 The Visible Spectrum and Neighbourhood	772
676 Symbol, Light Emitting Diode	773
677 Photodiode	774
678 Voltage-Current Characteristic	774
679 Equivalent Circuit	775
680 Photodiode Amplifiers	775
681 Photoresistor	776
682 CdS Cell Spectral Response	776
683 Photosensitive Relay	777
684 Phototransistor	777
685 Device Comparison	777
686 Photodarlington	777
687 Phototransistor Spectral Response	778
688 Optocoupler	778
Cooling	783
689 Heat Sink	783
690 Thermal Circuit Example	784
691 Effect of Insulating Washer	785
692 Derating Curve	786
693 Device to Heatsink Assembly	787
694 Heatsink Equation	788
695 Thermal Circuit with Capacitance	790
696 Power Dissipation in Op Amp	793
The Single Stage BJT Amplifier	798
697 Diode Model of the BJT	799
698 BJT Invention Test Circuit	800
699 Example Circuit	801
700 Current Gain vs Collector Current	801
701 Base Current Control of the BJT	802
702 Collector Characteristic	802
703 Practical Circuits	803
704 Current Generator Model	804
705 Current Generator Model, Simplified	804
706 Current Gain vs Collector Current	805
707 Voltage-Controlled Current Generator Model	806
708 Collector Current vs Base-Emitter Voltage	807
709 Ramp Generator	808
710 Ramp Generator	809
711 Emitter Current Control of Collector Current	810

712	Emitter Current Control, Practical Circuits	810
713	Evolution of the Current Generator	811
714	Evolution of the Current Generator	811
715	Evolution of the Current Generator	812
716	Evolution of the Current Generator	812
717	Evolution of the Current Generator	812
718	Completed Current Generator	813
719	Improved Generator Waveform	814
720	Improved Generator Circuit	814
721	Reflecting Resistance	815
722	Reflecting Resistance	817
723	Collector Incremental Resistance	818
724	BJT models, including collector incremental resistance	818
725	Ramp Generator Circuit	819
726	Ramp Generator Equivalent Circuits	819
727	Ramp Generator Charging Curve	820
728	<i>Early</i> Voltage V_A	820
729	Base vs Emitter Current Control	821
730	The Effect of V_{be}	823
731	Ramp Generator with V_{be} Compensation	824
732	Leakage Current I_{cbo} in the BJT	826
733	Final Circuit	832
734	Base Voltage and Collector Current	834
735	Evolution of the T Model of the BJT	837
736	Base Voltage and Collector Current	838
737	A Basic Common Emitter Amplifier	839
738	Common Emitter Amplifier, Model	842
739	Common Emitter Amplifier, Output Resistance	843
740	Common Emitter Amplifier Parameters	844
741	Example Amplifier	845
742	Common Emitter Amplifier	846
743	Amplifier with Emitter Resistance	847
744	Signal Capability	848
745	CE Amplifier with R_e , Parameters	849
746	Final Circuit	850
747	Input Bias Network	851
748	Final Circuit	852
749	Output Network	853
750	Emitter Bypass Capacitor	854
751	Emitter Capacitor Calculation	855
752	Example Circuit Complete	856
753	Common Emitter Amplifier, Dual Supply	857
754	Common Emitter Amplifier, Base Current Drive	858
755	Collector-Base Feedback Bias	861
756	Collector Voltage Variation	864
757	Common Emitter Circuits	866
758	Common Emitter Circuits, cont'd	867
759	Common Emitter Circuits, cont'd	868

760	Common Collector Concept	870
761	Bias Current	871
762	Bidirectional Emitter Follower	871
763	Basic Emitter Follower	872
764	Dual Supply Emitter Follower	872
765	Single Supply Emitter Follower	872
766	Emitter Follower Model	873
767	Voltage Gain of Emitter Follower	874
768	Emitter Follower as NFB System	877
769	Buffer	878
770	The <i>Collector Follower</i>	879
771	Emitter Follower Design Problem	880
772	Circuit Configuration	880
773	Base Bias Network	881
774	Thevenising the Input Network	881
775	Calculating C_i	882
776	Calculating C_o	883
777	Current Mirror Current Sink	883
778	Long Tailed Current Sink	884
779	Resistor Current Sink	884
780	Completed Emitter Follower	885
781	Basic Common Base Amplifier	886
782	Common Base Single Supply AC Amplifier	887
783	Modelling the Common Base Amplifier	887
784	Common Base Amp with Source Resistance	888
785	Noisy Amplifier	892
786	Noisy Amplifier Example	893
787	V_{be} vs I_C for the 2N4401	894
788	H Parameter Black Box Model	896
789	Tee Model	897
790	Base Resistance Function	898
791	Current Generator Model, Alternate Versions	901
792	Effect of Temperature on Collector Current	904
793	Temperature Change, Effect on V_{be} and I_c	906
794	Datasheet Cover Page	907
795	Datasheet (continued)	908
796	Datasheet (continued)	909
797	Datasheet (continued)	910
BJT-Compounds		911
798	Darlington Pair	911
799	Optocouplers	911
800	Darlington Leakage	912
801	Darlington Follower	912
802	Sziklai Pair	913
803	Complementary Pair	913
804	Sziklai Follower	914
805	Voltage Shift Circuit	915

806	Discrete Current Sink	915
807	Current Mirror	917
808	Current Mirror, Analysis	919
809	Current Mirror, Discrete Version	919
810	Power Dissipation	920
811	Design Example	922
812	Design Complete	923
813	Multiple Output Mirror	923
814	Multiple Output Mirror Analysis	924
815	Wilson Current Mirror	925
816	Current Analysis	926
817	Output Resistance Analysis	927
818	Emitter Degeneration Analysis	928
819	Wilson Mirror, Output Resistance	930
820	Cascode Pair	931
821	Cascode Output Resistance	932
822	Common-Emitter Amplifier	933
823	Cascode Frequency Response	934
824	Folded Cascode	935
825	Differential Pair	936
826	Diff Amp Transfer Characteristic	936
827	Example Differential Amplifier	937
828	Differential Pair DC Analysis	938
829	Differential Amplifier AC Analysis	940
830	Common-Mode AC Analysis	942
831	Transconductance Stage	943
832	Constant Tail Current	945
833	Emitter Degeneration Resistors R_E	946
834	Single Gain-Setting Resistor	946
835	LM124 Input Stage	947
836	Cascode Differential	947
837	Two Stage Diff Amp	948
838	V_{be} Loop	949
839	Two Stage Diff Amp: Bias	951
840	Two Stage Diff Amp: Gain	952
841	Two Stage Diff Amp: Final	953
842	Series-Shunt Pair	953
843	Series-Shunt Voltage Gain	954
844	Series-Shunt Gain Analysis	956
845	Closed-Loop Gain	958
846	Operational Amplifier	959
847	Block Diagram	960
848	Emitter Resistors	960
849	Compensated Response	961
850	Compensation Technique	962
851	Short Circuit Protection	963
852	Enhanced Output Stage	964
853	Offset Adjustment	965

854 P-Channel JFET Inputs	965
855 Logic Level Translator	966
856 Cascode Current Sink	967
857 Current Sink	967
858 Complementary Buffer Amplifier	968
859 Amplifier Design	969
860 Attempt 3	970
861 Operational Amplifier	971
862 Op-Amp Short Circuit Protection	972
863 μ A733 Differential Amplifier	973
864 μ A723 Regulator	974
865 AC Coupled Amplifier	974
866 Audio Power Amplifier	975
JFET Characteristics and Applications	976
867 JFET Constant Current Diode	976
868 Adjustable Current Source	977
869 Long-Tailed Current Source	978
870 Low Voltage Region	979
871 Compressor	980
872 Compressor Details	981
873 JFET Analog Switch	982
874 Shunt Switch with Drive	982
875 Opto-Isolated JFET	983
876 Transconductance of the JFET	984
877 Measuring JFET Transconductance	985
878 JFET Source Follower	986
879 JFET Small Signal Models	987
880 JFET Source Follower Model	988
881 JFET Differential Amplifier	991
882 JFET Buffer Amplifier	992
883 MPF102 Datasheet	994
884 Modern JFET Voltmeter	995
885 Square Law Function	997
886 MOSFET Transfer Curve	997
887 Threshold Voltage V_t	998
888 N and P Channel Enhancement Mode MOSFETs	998
889 Depletion Mode MOSFET Transfer Curve	999
890 N Channel, Enhancement Mode, Source Follower, DC Coupled	999
891 Bias Condition for DC Follower	1000
892 Small Signal AC model of Source Follower	1000
893 Simplified Model of Source Follower	1001
894 Output Resistance Model of Source Follower	1001
895 Graphical Determination of Transconductance g_m	1002
896 N Channel, Enhancement Mode, Source Follower	1003
897 Audio Amp, MOSFET Outputs	1005
898 Source Follower with Parasitic Capacitances	1005
899 Source Follower Equivalent Circuit	1006

CMOS Analog Switch and Applications	1009
900 Relay and Analog Switch	1009
901 Rotary Switch/Multiplexer	1010
902 Switcheable Mixer	1010
903 Switched-Gain Amplifier	1011
904 Analog Switch Logic	1012
Interference	1013
905 Interference in General	1015
906 Example, Common Impedance	1016
907 Common Point Grounding	1017
908 Electric Field Coupling	1018
909 Electric Field Shield	1018
910 Electric Field Pickup	1019
911 Equivalent Circuit	1020
912 Magnetic Coupling	1020
913 Coil Cancellation	1021
914 Measuring Voltage in a Conductor	1022
915 Magnetic Field Probes	1023
916 Magnetic Field Pickup	1024
917 Square Wave Spectrum	1026
918 Trapezoid Wave Spectra	1027
919 Sinc Function	1028
920 Power Supply Capacitive Filter	1029
921 Reducing Trace Resistance	1032
922 Wiring Inductance	1032
923 Analog/Digital Power Supply System	1035
924 Capacitor Parasitic Inductance	1036
925 Bypass Capacitor Resonance	1037
926 Broadbanding the Bypass Response	1037
927 Differential Amplifier Application	1038
928 Differential Amplifier Enhancements	1039
929 Guarding	1040
930 Bifilar Choke	1040
931 FCC Part 15 Class B Emission Limits	1041
932 Common-Mode Output Signal	1043
933 Unbalanced Differential Drive	1044
934 Shielded Enclosure	1045
935 Measurement Using AC Ground	1047
936 Twin Unshielded Leads	1047
937 Measurement Using Coaxial Cable	1048
938 Measurement Using $\times 10$ Probe	1049
939 PCB Power Trace Layout	1051
940 A/D Converter Internal Wiring	1052
941 A/D Converter System Wiring	1053
942 Segmented Ground	1054
943 Unbalanced Differential Drive	1056

Beyond the Op-Amp	1058
944 Amplifier Classification	1059
945 Current Conveyor	1060
946 Current Conveyor Differentiator	1061
947 CFA Non Inverting Amplifier	1062
948 CFA Equivalent Circuit	1063
949 CFA Transresistance Gain	1065
950 CFA Equivalent	1065
951 Op-amp slew rate model	1066
952 CFA slew rate model	1066
953 Norton Amplifier	1067
954 Norton Amplifier Application	1068
955 OTA Function	1069
956 OTA Internal Circuit	1069
957 Voltage-Controlled Amplifier	1070
958 Time Delay	1071
959 Instrumentation Amplifier	1072
960 Fully Differential Amplifier	1074
961 Internal Circuit	1076
962 Fully-Differential Amplifier Application	1077
963 Differential Gain Analysis	1078
964 Differential Gain Analysis	1079
965 Single to Differential Signal Converter	1080
966 Lever Diagram, Single to Differential Signal Converter	1080
967 Adding Common Mode Output	1081
968 Feedback Resistor Value and Bandwidth	1082
969 Autozero Operation	1084
970 Chopper Amplifier Concept	1085
971 Single-Ended Chopper Amplifier	1086
972 Chopper Amplifier Generalized	1087
973 Composite Amplifier	1089
974 Difference Amplifiers	1093
975 Duty Cycle Multiplier	1096
976 Predistortion Diodes	1099
977 Generating Drive Currents	1100
978 Gilbert Quad	1101
979 Complete 4-Quadrant Multiplier	1102
980 Alternate Duty Cycle Multiplier	1103
981 Photo-Resistive Multiplier	1104
Complete Systems	1105
982 Analog and Microprocessor System Cost	1105
983 Equation Solver Block Diagram	1105
984 Equation Solver Circuit	1106
985 Real Motor, Equivalent Circuit	1107
986 Motor Power Circuit	1108
987 Motor Speed Controller Block Diagram	1108
988 Motor Speed Controller Circuit Diagram	1110

989 Mechbot	1112
990 Mechbot	1113
991 Battery Capacity	1113
992 Power Supply Wiring	1113
993 Motor Driver Circuit	1114
994 Edge Detection, Back and Turn Concept	1115
995 Edge Detection Circuit	1116
996 Object Detector Concept	1117
997 Object Detector Circuit	1118
998 Resolver	1119
999 Resolver SINE Output	1119
1000 Resolver COSINE Output	1120
1001 Resolver Stator Voltage Waveform vs Shaft Angle	1120
1002 Resolver Interface Block Diagram	1121
1003 Phase Sensitive Rectifier	1121
1004 Phase-Sensitive Rectifier Output	1122
1005 Resolver Interface Schematic	1123
1006 LVDT Interface	1124
1007 Correlator	1125
1008 Correlator With Delay	1125
1009 Correlation Application	1126
1010 Signal Processing with Correlation	1126
1011 Experiment with Bandpass Filter	1127
1012 Correlator using Phase Sensitive Rectifier	1128
1013 Pressure Sensor	1129
1014 Resistor Bridge	1129
1015 Sensor Equivalent Circuit	1130
1016 Barometer Interface Transfer Function	1131
1017 Barometer Interface Block Diagram	1132
1018 Splitting the Gain	1133
1019 Subtractors	1134
1020 Barometer Interface Schematic	1135
1021 Circuit Equations	1137
1022 Amplifier Spreadsheet Model and Results	1138
1023 Water Manometer	1138
1024 Telescope Dew Heater, Block Diagram	1147
1025 Comparator Detail	1149
1026 Enclosure Concept	1152

1 Introduction

Electrical technology has two large subject areas: power systems and signals.

Power Systems concerns the generation and transmission of electrical energy. Electricity is a convenient way of moving and distributing energy at a central location (the power plant) to the end user, where it powers your lights, refrigerator and computer.

Signals concerns the use of electricity to convey information. For example, the video image of a television display and the sounds of a music player are transmitted and processed in electrical form. The information in the video image and sound track are represented by small-scale voltages and currents¹.

This representation of the signal can be *analog* or *digital*. In analog form² the magnitude of some voltage or current represents the original signal. For example, the magnitude of a voltage varies in the same fashion as a sound wave: the voltage is said to be an analog representation of the sound wave. Signals almost always begin in analog format. In many cases, the signal is subsequently converted to digital format, in which the original signal is represented by a stream of numbers. In some cases, it is subsequently reconverted back to analog form.

This text focussed on electronic *signals* in the *analog* format.

Why study analog circuit design?

- Signals usually begin in analog form and it is important to understand the various tradeoffs of analog and digital signal representations. A versatile designer will be able to work in both analog and digital domains.
- In some cases, a small analog circuit can replace a larger, more complicated analog-digital design, with consequent savings in parts cost and power consumption.
- Ultimately, all circuitry functions according to analog principles, even if the circuitry is managing a stream of numbers. Issues that are critical to the correct functioning of digital circuits – stray capacitance, termination of transmission lines, supply of power, interference and shielding – are all analog in nature.
- The interface between measuring instrument and circuit is analog in nature. Measurement affects a circuit. It's important to understand these effects and mitigate them when they are a problem.
- Some circuits – audio power amplifiers and power supply designs – are inherently analog from start to finish.

Getting to a Working Design

There are various paths to a final design, and the one taken will depend on circumstances. Here are two representative situations:

- **We need one functional circuit for a single application.** In this case, there is only one instance of the circuit. So, if the circuit works, that's the end of the matter. Getting there can be accomplished by some calculations on a beer coaster, more-or-less-inspired tinkering at the workbench and some cut and try circuits on a protoboard. If the circuit has to work over a range of conditions (varying temperature, changing line voltage) then it would be a good idea to check circuit operation under those circumstances. Last-minute tweaking of the circuit is fine and will probably be required.

¹At the university where I taught for many years, engineers worked in power or electronics, and only very rarely in both. One power engineer derisively referred to electronics as *the hobby course*. Power engineers worked with large voltages and currents, which he thought a more serious pursuit. Tribalism is everywhere.

²In general, we use the American spelling, *analog*. The British spelling is *analogue*.

This circumstance often arises in scientific exploration, where one circuit is needed for some measurement. With the results in hand, the circuit is scrap. Furthermore, because there is only one, we can use parts from a junk bin. If it saves time, it's worthwhile to spend money on components or assemblies. For example, it's probably best to buy a pre-assembled power supply.

- **This is a consumer-grade product that will be built in large quantities.** In this case, cost is critical. The cost has two principal components: the parts and the labour to assemble them, and both must be kept under control. Product recalls, warranty repairs and field failures are the stuff of engineering nightmares, so considerable effort must be made to avoid these.

This design must be *engineered* to meet requirements of cost, function and reliability. Engineering implies *planning*: careful thought in advance of building. Then, when a prototype is finally constructed, there should be no mysteries and no surprises.

A prototype of working hardware is, to borrow a term from mathematics, a *necessary but not sufficient condition*. If the prototype works it means that one instance of the circuit design functions correctly. It doesn't mean that all 10,000 instances will function correctly. That's the job of the circuit analysis, which took place weeks or months before. On the other hand, if the prototype does not work, then one instance of the circuit has failed and the circuit design is flawed.

Documentation

The product of engineering work is not the thing itself. The product is the *information* to build the thing. The circuit concept, detailed drawings, printed circuit board layout, mechanical drawing, bill of material, assembly procedure: those are the product of creative engineering work. Those are the items that must be carefully archived and protected.

In these days of computer tools, there is no excuse for shoddy documentation. There are word processors for manuals, spreadsheets for lists, and drawing programs for schematics. And until someone finds the definitive solution for long-term storage? Keep a paper copy. Just in case.

Simulation and Analysis

In the pursuit of electronic design, we must determine the properties and behaviour of an electrical component or network. We could do this by calculation or by simulation with a computer program. Which is better?

For a circuit design where component values are known, the simulation has the advantage – it may be faster to do and it can take into account the imperfections of components like an operational amplifier. However, that is a *specific* solution and it doesn't give full insight into the circuit behaviour.

You can change the simulation circuit values and determine their effect on the behaviour of the circuit. However, analysis gives a deeper understanding of the circuit operation. For example, an analysis can tell you that a certain cutoff frequency is proportional to the square root of the ratio of two capacitor values. Knowing that, you might be able to select two capacitors that track with temperature and therefore result in a more stable filter design.

On the other hand, some circuits are particularly nasty to analyse. For example, the transformer-driven half-wave power supply with a capacitor filter is often analysed in textbooks as if the transformer has no resistance, that is, the rectifier-filter is driven by a pure voltage source. This is completely unrealistic in practice, but it's difficult to take source resistance into account. However, using simulation one can play with the values of the components and get a feel for how the circuit functions, so simulation is very useful in that situation.

Whether you use analysis or simulation, there is no substitute for understanding the operation of the circuit and doing a sanity check on the results. Run some cases where you know the results and see if the method

generates something reasonable. In the preface to [3], Pease gives 10 examples of simulation failures. As he says, sometimes the simulator gives *lies*. It's critical to be able to detect those cases.

To summarize, a simulation gives *circuit specific results* when numeric values are known. Analysis provides *generalized results*. They both have their uses.

A philosophical note

The reader could be excused for finding a certain amount of magic in the mathematical expositions. For example, in the section on the Laplace Transform, there are a mere dozen or so shown out of the vast number of possible Laplace Transforms³. That requires a-priori knowledge of what will be useful. We also know that a term like LC or R/L has special significance. That requires experience. Finally, every author has his/her preference for method. What may be clear to the writer is not necessarily clear to the reader. Throughout, I have tried to indicate not only *what we do next to progress to the solution*, but *why we choose to do this and not something else*.

That said, after doing a few examples, patterns do emerge and hopefully the reader will be able to extrapolate to other circuits.

Words of Encouragement

For a hobbyist or engineer, this is a wonderful time to be practising analog circuit design.

Analog circuit design is not at the point where one can simply patch together the necessary integrated circuits. Analog circuit design still requires skill and ingenuity. However, analog integrated circuits such as the operational amplifier and voltage regulator have made circuit design much easier.

- Wonderful computer tools are available for design and documentation.
 - Symbolic math programs such as Mathematica and Maxima for solving equations,
 - Spreadsheets such as Excel and Calc for tolerancing circuits and listing components.
 - Circuit simulation programs such as LTSpice and Multisim
 - Access to datasheets and applications information via the Web.
 - Printed circuit board layout programs that result in reliable circuits that are easy replicate.
 - Mechanical design programs to build and check a virtual 3D design.
- Excellent parts are available at modest price. For example, operational amplifiers have evolved from low-frequency units packaged as expensive modules to gigahertz bandwidth units available for a few cents⁴.
- Small quantities are available for quick delivery and reasonable price from distributors such as Digikey, Mouser and Jameco.
- Small-quantity printed circuit boards are available at reasonable price. Email a description of the board to the company and receive a printed circuit board later the same week.
- Excellent surplus and new electronic instruments are available at modest cost for measuring and testing circuits⁵.

It's still possible and an enjoyable pastime to experiment with electronic circuits. For those who wish to practice professional-level analog circuit design, the components and tools have never been better.

³Reference [4] lists 168 of them.

⁴Be prepared to invest in a microscope to identify some of these components: they are *tiny*.

⁵Syscomp has contributed to this trend with its signal generators and oscilloscopes [5].

Acknowledgement

Special thanks to my family, friends and colleagues for their help and encouragement.

I particularly wish to thank John Foster for his masterful work in the demanding task of proofreading the manuscript. In addition to a careful review of the entire manuscript, John created software tools that were tremendously effective in finding transcription and typographical errors and inconsistencies.

1.1 Outline of the Material

Here is an overview of the book material.

Sections 2, 3 and 4: a review of the basic electrical concepts.

When I was teaching electronic circuit design, I found that students were often insufficiently familiar with the tools of the trade. Their original coverage of these topics had a classical emphasis with topics like Kirchhoff's laws, Thevenin's theorem and the reciprocity theorem getting equal emphasis. These subjects are not all of equal usefulness. As is not uncommon in academia, some of these topics have very little practical application at all. Others are critically important.

This treatment is not intended to replace more detailed study. What this material does indicate is *what is important* out of the vast subject that is electrical circuit analysis. For example, circuit analysis using Kirchhoff's laws can be applied manually in a simple one or two loop circuit. In more complicated circuits analysis can be done by hand with great effort, but it really requires a computer. It's not a useful tool for manual analysis of a complicated electronic circuit. It's much more effective and useful to be able to recognize and analyse subsystems such as the voltage divider and Thevenin source. It's absolutely essential to understand the superposition theorem.

Section 2.30 introduces the concept of *amplifier*, giving some mechanical and pneumatic analogs. Section 3.10 shows how component tolerances affect a real-world design.

Section 4 covers concepts that are specific to AC circuit analysis and useful in electronic circuit design. Many topics are omitted. For example, there is nothing on three-phase circuits, which is of great importance in power systems.

Section 5, the Laplace Transform is optional material. For most of the text, it is sufficient to see the Laplace variable s simply as a shorthand for $j\omega$. Although not essential, the Laplace Transform is a useful tool and the essentials are not difficult to understand. Furthermore, a computer algebra program (such as the open-source *Maxima*) can handle the heavy lifting. This makes it much more accessible for the non-specialist to use in electronic circuit design.

Sections 6 and 7 introduce two-terminal semiconductor devices (the diode and zener diode) and their circuit models.

Section 8 introduces three-terminal devices: the MOSFET, JFET and BJT (transistors), explaining basic operation and their application as a switch.

Section 9 provides simple electronic models for the DC motor. For an ideal motor, the speed is proportional to the terminal voltage. The torque is proportional to the current. From these simple concepts it is straightforward to model a real world motor. This is useful in the many applications that consist of electronics with a DC motor actuator – such as the focussing control of a camera. An application is in section 37.3.

Section 10 provides an intuitive introduction and develops simple mathematical models for *negative feedback*, a concept that pervades analog circuit design.

Section 11 introduces the *reactance plot* and *Bode plot*. These tools equip the circuit designer for intuitive and *paper and pencil* design for frequency dependent networks.

Section 12 introduces the operational amplifier, with applications in section 13. The advent of the low cost operational amplifier revolutionized the design of analog circuits. Many discrete component circuits could be replaced with smaller, simpler, lower cost circuits. Moreover, the simplicity of designing with op-amps opened up the field of analog circuit design to the non-specialist.

Section 14 introduces the best known example of *positive* feedback, the *Schmitt Trigger*. The *lever diagram*, introduced in section 12 with basic op-amp circuits, is used again in this section for the design of the Schmitt trigger circuit.

Section 15 discusses the design of power supplies for electronic systems. The power supply is an important component of the system and it is critical that it be designed properly, not added as an afterthought.

Section 16 The precision rectifier and variants are useful circuits in their own right, and illustrative of non-linear op-amp operation in general.

Section 17 is a detailed exposition on the theory and practice of active filters, covering the circuits that are most likely to be useful in practice. The derivations are included, but the emphasis is on obtaining a functional design to meet engineering requirements.

Section 18 describes unusual circuits, with a focus on those that manipulate impedance in some fashion.

Section 19 describes active filters that affect the phase of a signal, without appreciably changing the amplitude. Phase shifters are useful in measurement and communications systems.

Section 20 describes several low-frequency oscillators. (Oscillators for radio frequencies are considered to be a specialized topic outside the purview of this text.)

Section 21 further explores the behaviour of the operational amplifier. In many applications, the op-amp may be treated as an ideal device, but there are applications where their limitations surface. This section explores those properties.

Sections 25 and 26 describe the D/A and A/D converters that transition signals between the world of analog signal processing into the digital realm, and vice versa.

Section 22 shows how intrinsic noise arises in an electronic circuit. Prediction and control of noise levels is important in low-level instrumentation circuits and audio signal processing.

Section 23 explores the frequency response behaviour of op-amps, and shows how to predict the maximum usable frequency of a given device and circuit.

Section 24 explores the issue of stability in op-amp circuits. Dominant pole compensation in op-amps has greatly reduced the problem of oscillation in op-amp circuits, but it is still important to understand the source of and cure for marginal stability or outright oscillation.

Section 27 catalogues some useful opto-electronic devices, along with their characteristics and applications information.

Section 28 discusses an oft-neglected topic: managing the heat produced by an electronic system. The engineering of heatsinks is quite straightforward and should be considered at an early stage in a system design.

Section 29 is a large section on the operation and use of the Bipolar Junction Transistor (BJT). It is now uncommon to design complete systems using discrete BJTs. Much of the work can be done by analog integrated circuits. However, BJTs do continue to have an important place in analog circuit design, and it's important to understand them thoroughly.

Section 30 shows how pairs of BJTs are used as building blocks to create functional circuits.

Section 31 describes the operation of the junction field effect transistor (JFET) and its application in linear circuits. The JFET is somewhat deprecated in practice. However, it does find some applications, its function is similar to the MOSFET (Metal Oxide Field Effect Transistor), and engineers may encounter it in existing circuitry, so we include it here.

Section 32: The MOSFET is most commonly found either as a switch in power control applications or as one of thousands in an integrated circuit. However, there are some linear applications of the MOSFET – specifically in power audio amplifiers.

Section 33: The analog switch, described in this section, is a tremendously useful component in analog system design. For an example application, see the duty-cycle multiplier of section 36.3.

Section 34: Electro-magnetic interference (EMI) can prevent a circuit from operating or may cause it to fail regulatory testing. In the absence of a complete understanding of the problem, it can be very difficult to remediate. This section introduces the principles and tools related to EMI so that problems may be investigated and cured.

Section 35 describes variants of the standard operational amplifier that are useful in particular application domains. For example, the current-feedback amplifier (CFA) is well suited for high-frequency and slew rate applications. The instrumentation amplifier is particularly useful when dealing with small signals in the presence of common-mode noise.

Section 36: Analog multiplication is a useful technique when two signals must be multiplied and conversion to the digital domain is not worthwhile. This section describes two analog multiplier techniques: the pulse-width-height technique for low speed, high accuracy, the translinear technique for high speed, moderate accuracy.

Section 37 brings together the ideas of the book to show examples of complete analog systems. The *exercises* section suggests a number of open-ended analog circuit design exercises that can be tackled by students. The end of the section follows through an *Extended Design Exercise* to illustrate the thought process of an analog circuit design engineer.

A note on the references

Many of the bibliographic references point to web pages. At release time all the addresses were valid, but that may change. The World Wide Web is a wonderful resource but it's dynamic. Web pages move and disappear.

If you find that one of the addresses doesn't work, copy and paste the document title and author name into a search engine such as Google, and see if that turns it up. Failing that, choose some approximation to the description, and try that.

1.2 Converting Units of Measurement: A Systematic Method

The conversion of units can cause great confusion, so here is a systematic method.

In the example on page 59, we are given the diameter of the wire in millimetres and the resistance formula in square metres. To be consistent, we need to work in the MKS (Metre, Kilogram, Second) system of units, so millimetres must be converted to metres.

A millimetre is one thousandth of a metre, so it must be divided by 1000 to convert to metres. This is fairly clear. However, there are cases where the conversion process is not so clear.

A systematic method of unit conversion is the following:

1. Express the conversion factor as a fraction that is numerically equal to unity. We'll call this the *converter fraction*. In this case, the converter fraction is obtained by the following reasoning:

$$1000 \text{ mm} = 1 \text{ metre}$$

Divide both sides by the RHS (Right Hand Side of the equation), and we can write:

$$1 = \frac{1 \text{ metre}}{1000 \text{ mm}}$$

Because this fraction is equal to unity, we can multiply anything by it without affecting the result.

This equation could equally well be written:

$$1 = \frac{1000 \text{ mm}}{1 \text{ metre}}$$

How do we select between these two? That brings us to the second rule:

2. Arrange the numerator and denominator units of the converter fraction so that there is a cancellation of the old units (millimetres in this case) in the equation to be converted. Then the result is expressed in the units of the numerator.

As a very simple example, to convert 1 millimetre to metres, we would write:

$$\begin{aligned}1 \text{ mm} &= \frac{1 \text{ mm}}{1} \times \frac{1 \text{ metre}}{1000 \text{ mm}} \\&= \frac{1}{1000} \text{ metre}\end{aligned}$$

This technique is especially useful when there are multiple steps to the conversion. There are multiple converter fractions, one for each conversion step. Ultimately, the units must cancel correctly to give the desired final units. This helps avoid mistakes of *multiplying by 6.3 instead of dividing by 6.3*.

Examples of unit conversion using this technique are in section 2.6.5 on page 59, and in section 28.6 on page 792.

2 Basic DC Circuits

2.1 Voltage, Current and Resistance

The *hydraulic analogy* of figure 1 helps explain the concepts of voltage, current and resistance.

Figure 1(a) is a *hydraulic circuit*. A pump (driven by a motor for example) drives water through a pipe to a valve. Adjusting the valve changes resistance to the flow. As the valve is closed, the flow of water decreases.

Figure 1(b) is the electrical equivalent. A battery provides electrical pressure to drive an electrical current through the variable resistance. If the resistance is increased, the flow of electrical current decreases.

An amount of water in liters is analogous to an amount of electrical charge in *coulombs*. The *pressure* in the hydraulic system is similar to the *voltage* in the electrical system. The quantity of water flowing (in liters per second, for example) is similar to the electrical current in coulombs per second (amperes). The resistance of the hydraulic valve is similar to the electrical resistance.

In both cases, the power used in the system is the product of the pressure and current.

In both cases, the pressure rise is equal to the pressure drop. That is, the algebraic sum of the pressures (voltages) around this *circuit loop* is zero. We'll meet this concept again as Kirchhoff's Voltage Law in section 2.9.

In both cases, the quantity of water and current are constant. We'll meet this concept again as Kirchhoff's Current Law in section 2.10.

The hydraulic analogy is not perfect. For example, if there is a break a pipe of the hydraulic system, water flows out onto the floor. That's not the case in an electrical system: a break causes the current to stop. In effect, creating a break (an *open circuit*) inserts a large resistance.

Furthermore, in the hydraulic system, the flowing quantity is water. In the electrical circuit, the flow is composed negatively charged particles, electrons.

Now we are in a position for more formal definitions of voltage and current.

2.2 Electron Flow and Conventional Current: The Unlucky Guess of Benjamin Franklin

Early electricity experimenter Benjamin Franklin [6] guessed that an electrical current was created by the movement of a positively charged particle which moved from a positive terminal to the negative terminal⁶. Many years later, J.J.Thompson used cathode-ray tube experiments to discover the electron. These experiments determined that Ben Franklin had guessed incorrectly: the electron carries a negative charge [8].

⁶Franklin suggested that electricity could be collected by flying a kite in a thunderstorm. The image of Franklin and his kite subsequently assumed mythic status in the history of electricity. He gave the impression that he had actually performed the experiment, but Tucker [7] makes a strong case that the experiment never took place. Subsequently, St. Petersburg electrical experimenter Georg Vil'gel'm Rikhman did something similar and was electrocuted.

Among Franklin's many accomplishments, he invented the lightning rod and furthered the case that electricity was caused by the flow of one charged particle rather than two dissimilar fluids – a competing theory of the time.

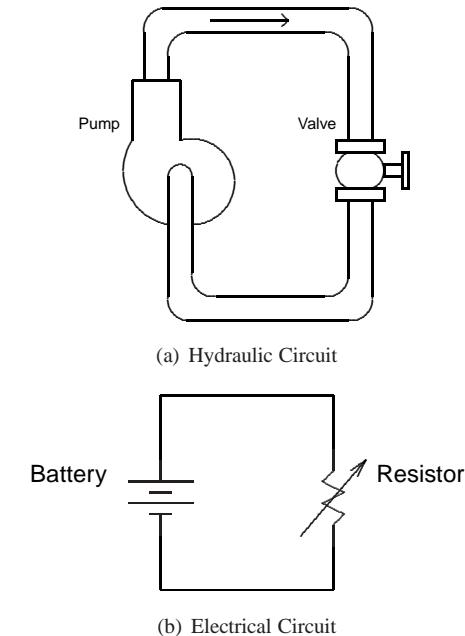


Figure 1: Hydraulic Analogy

At that point, electrical conventions were well established. Electrical circuit design calculations are largely unaffected whether an electrical current is caused by a positive particle flowing from positive to negative (*conventional current flow*, as it is now known) or a negative particle flowing from negative to positive (*electron flow*).

For example, suppose we describe the effect of rush hour traffic as a flow of positive cars into a city. Cars accumulate in the city. We could equally well describe the same flow as negative cars (car spaces) flowing out of the city. Car spaces leaving the city must be equivalent to cars accumulating in the city.

Conventional current has the psychological advantage that it flows from positive to negative, where the positive terminal may be regarded as having a surplus of positive charge. The symbols for semiconductor devices indicate with an arrow the direction of conventional current flow (See for example the diode symbol in figure 173 on page 246).

Electron flow is essential in understanding the behaviour of electron tubes and semiconductor device physics. However, in most situations it is not essential and we will use conventional current flow in this text.

2.3 Electronic Charge

An *electronic charge* may be thought of as a collection of electrons. The unit of electronic charge is the *coulomb* (symbol C) , equal to the total charge on 6.242×10^{18} electrons⁷.

$$1 \text{ coulomb} = 6.242 \times 10^{18} q_e \quad (1)$$

where q_e is the charge on an individual electron.

As indicated above, a charge created by a surplus of electrons is negative. Materials are normally electrically neutral, that is, the positive and negative charges balance out. A charge created by a *deficiency* of electrons is then positive.

A surplus of charge may be created by a number of mechanisms. The classic example from high school physics is an effect called *triboelectricity*. Rubbing certain combinations of substances such as cat fur on a glass surface causes a charge accumulation⁸. The charge is known as *static electricity* because the charge is stationary – until it is discharged, possibly causing a visible spark.

The charge q_e on an electron can be measured directly in the *Millikan Oil-Drop experiment* [10].

2.4 Voltage

There is another analogy for the electronic pressure that is known as voltage: it is similar to the difference in height between two locations in a gravitational field. Objects in a gravitational field tend to move downhill – that is, to fall from a large to a lesser height. As well, it takes work to move a mass from a lesser to a greater height.

Similarly, charge tends to move from a location of greater voltage to one of lesser voltage. As well, it takes work to move a charge *uphill* from a lesser to a larger voltage.

This is a useful analogy. Just as height must be measured between two locations, so must voltage. A 9 volt battery is one in which the voltage (also known as the *potential difference*) between the two terminals is equal to 9 volts.

Voltage is defined in terms of the work required to move a charge between two points at different potential.

One volt is the potential difference between two points when the work required to move one coulomb of charge between them is one joule.

⁷This number representation is *scientific notation*, which has a number of advantages in engineering and science. See [9] for a detailed explanation.

⁸It helps but is not essential that the fur be unpopulated by a live cat.

In equation form:

$$W = QV \quad (2)$$

where V is the voltage difference between two points, Q is the charge being moved in coulombs, and W is the work in joules.

2.4.1 Example Voltages

Typical units of voltage:

nanovolt	nV	10^{-9}V
microvolt	μV	10^{-6}V
millivolt	mV	10^{-3}V
kilovolt	kV	10^{+3}V

The following table gives the approximate magnitude of various voltages.

$1\mu\text{V}$	Antenna output voltage due to radio signal
$500\mu\text{V}$	Output of moving-coil phonograph cartridge [11]
10 to $100\mu\text{V}$	Electroencephalogram (EEG) signal (activity of brain)
1mV	Electrocardiogram (ECG) signal (activity of heart)
2mV	Electromyograph (EMG) signal (activity of muscles)
5mV	Output of moving-magnet phonograph cartridge [11]
10mV	Dynamic microphone in moderate sound field
0.1 to 10mV	Electret microphone in moderate sound field [12]
70mV	Nerve membrane potential
1.5V	Nominal voltage dry cell, which is <i>stacked</i> to form higher voltage batteries.
6V	Lantern battery, some vehicle electrical systems
9V	Nominal voltage of electronic battery
12V	Nominal voltage of car electrical system
28V	Nominal voltage of aircraft electrical system
0 to 30V	Electronics lab power supply
60V	Threshold of dangerous voltage [13]
48 to 90V	Telephone loop voltage
117V	Household voltage (North America)
230V	Household voltage (UK)
300 to 750V	Electric tram and railway overhead voltage
350V	Typical voltage for vacuum tube supply
3MV/m	Breakdown voltage of dry air
20kV	Accelerating voltage of cathode ray tube (television tube)
1GV (10^9V)	A 300 m (1000 ft) lightning bolt [14]

The magnitude of a voltage may or may not correlate with its danger. For example, very large voltages are generated by frictional electricity. These voltages are sufficient to create visible sparks, which requires kilovolts of electromotive force, but they do not cause electrocution⁹.

⁹A spark due to a static discharge can ignite an explosive gas or dust cloud. Static electricity must be strictly controlled in such an environment. As one example, an aircraft can pick up a static charge while flying. The tires are insulating, so before refueling the aircraft must be electrically connected to earth by a conducting cable to dissipate the charge.

2.4.2 Voltmeter

A *voltmeter* is a device to measure the potential difference (the voltage) between two points in a circuit. An ideal voltmeter conducts no current, that is, it appears to be an electrical open circuit.

When a voltmeter is not ideal, it affects the circuit and the voltage at that point in the circuit changes when the meter is attached. This is known as the *loading effect*.

Early voltmeters were far from the ideal and the loading effect could be a significant problem. However, modern voltmeters have improved to the point where loading of the circuit is no longer an issue.

2.4.3 Voltage Standards

The definition of the volt given by equation 2 above, ties the volt to fundamental units in the MKS system of measurement but it does not lead to a particularly convenient working lab standard. For over 100 years the primary voltage standard was an electrochemical battery called the *Weston Cell*, which produces a voltage of precisely 1.0183 volts [15]. However, the voltage of a Weston Cell changes with temperature (albeit in a predictable manner) and it cannot supply current.

At this time, the voltage primary standard is the Josephson Voltage Standard [16]. This semiconductor device produces a voltage that is related to the electron charge q_e , Planck's constant h , and the applied frequency. These constants are known to considerable accuracy and frequency may be measured very precisely, so the Josephson Voltage Standard is a reliable primary standard.

Primary standards are expensive and inconvenient for use in an industrial lab. (The Josephson voltage standard requires cooling the semiconductor unit in liquid helium.) The usual practice is to calibrate a secondary standard against the primary and then use that.

Fortunately, excellent voltage standards based on semiconductor technology are now available. Voltage drift is measured in *ppm, parts per million*. The Linear Technology LTZ1000 is an ultra-stable reference which produces an output between 7.0 and 7.5V with temperature drifts of $0.05\text{ppm}/^\circ\text{C}$ (ie, $\approx 7 \times 0.05 \times 10^{-6} = 0.35\mu\text{V}/^\circ\text{C}$). The long-term stability is $2\mu\text{V}/\sqrt{\text{kHr}}$ (ie, less than 2×10^{-6} volts drift every 31 hours).

The National Semiconductor LM4132 produces a 2.5V output with 0.05% initial accuracy and stability of 10ppm/degree C. For a 2.5 volt output at the reference temperature, the LM4132 is then within 1.25mV of the exact value 2.500 volts. This makes it suitable for routine calibration of many lab voltmeters.

2.5 Current

An electrical current is measured in *amperes*. One ampere of current is equivalent to the flow of one coulomb of charge per second. That is:

$$I = \frac{Q}{t} \quad (3)$$

where I is the current in amperes, Q is the charge in coulombs and t the elapsed time in seconds. Turning this around, we can write:

$$Q = I \times t \quad (4)$$

If the current is a function of time, we can write:

$$Q = \int_0^t i(t)dt \quad (5)$$

This turns out to be useful in some devices that are *charge balancing*.

2.5.1 Current Meter (Ammeter)

An ammeter is connected so that the current to be measured flows through the meter. An ideal current meter appears as a short circuit, so it creates no voltage drop and has no effect on the behaviour of the circuit.

Real ammeters depart from this ideal. Section 2.16.4 (page 70) describes one example of a practical (ie, non-ideal) ammeter.

2.5.2 Current Standards

An electrical current passing through an electrolyte can be made to deposit a substance on one of the electrodes. This process of *electroplating* can be used to create a standard for current. The current is passed through a solution of silver nitrate and the current measured by the rate of deposition of silver in grams per second [17].

A current-carrying conductor in a magnetic field creates a linear force or torque. This is the basis for more recent determinations of the ampere standard [18].

As in the case of the voltage standard, this primary standard is not suitable for engineering practice. Modern practice is to use voltage and resistance industrial standards. Then current can be determined using Ohm's law, as described in section 2.6 on page 56.

For example, if the voltage is known to a precision of $\pm 0.05\%$ and the resistance to $\pm 0.01\%$, then a current measurement can be known to an accuracy of $\pm 0.06\%$.

2.5.3 Example Currents

Typical units of current:

picoamp	pA	10^{-12} A
nanoamp	nA	10^{-9} A
microamp	μA	10^{-6} A
milliampere	mA	10^{-3} A
kiloampere	kA	10^3 A

Examples of current:

5mA	Ground fault interrupter trip current (USA) [19]
10mA	Typical current in light emitting diode
3 to 30mA	Current per square cm, anodizing electroplate process
500mA	Current in 60 watt household lamp (117VAC supply)
1A	Typical current limit for electronic bench power supply
10A	Current in toaster, teakettle (117V appliances)
20A	Current in dryer (220V appliance)
30A	Current limit in modern household circuits
100A	Starting current for automobile
150A	Total current capability of household electrical service (North America)
1300A	750V, two-car light rapid transit vehicle
46kA	Fusion reactor solenoid current
120kA	Lightning bolt [14]

2.5.4 Electron Conduction Physics

It's not necessary in electronic design to consider the microscopic behaviour of a conductor. However, the behaviour and numbers are interesting [20].

As we have indicated, a coulomb contains 6.242×10^{18} electrons. One ampere of current at some point in a circuit is equivalent to an electric charge of one coulomb flowing past that point every second.

A piece of wire one metre long has in the order of 10^{22} free electrons available for conduction. As a result, to set up a current of one ampere the average electron velocity is very small: a fraction of a millimetre per second.

This is the *average* or *drift* velocity of an electron. The actual instantaneous movement of each electron is at a high speed known as the *Fermi Velocity* – about 1.5×10^6 metres/sec – changing direction each time it collides with an atom in the conductor.

Consequently, a current-carrying metallic conductor contains a cloud of electrons moving at high speed in random directions. The current is evident in a small average speed of this cloud.

When a current is started (by closing a switch in a circuit like figure 1(b), for example), the movement is propagated very quickly throughout the entire conductor. This propagation velocity is some fraction, known as the *velocity factor*, of the speed of light. The velocity factor is typically between 0.5 and 0.9 depending on the wiring configuration.

All of this description applies to a so-called *DC circuit*, where the current flows in one direction. Later we will explain the functioning of an *AC circuit* such as the power distribution system.

2.5.5 Velocity of Electron in a Vacuum Tube

In a vacuum tube, electrons are emitted from a heated metal surface (the *cathode*) and then accelerated through the vacuum by an electric field toward another electric surface, the *anode*. A typical voltage between the cathode and anode terminals (electrodes) might be 350 volts. What is the velocity of the electron when it strikes the anode?

This is one of many problems in physics that can be solved by calculating the exchange of potential energy for kinetic energy.

The potential energy is equal to the work done by one electron of charge q_e moving through the voltage E between the two electrodes. The kinetic energy is given by $\frac{1}{2}m_e v^2$, where m_e is the mass of an electron and v is its velocity. These must balance: the loss in potential energy is equal to the gain in kinetic energy. In equation form:

$$q_e E = \frac{1}{2} m_e v^2 \quad (6)$$

where the algebraic quantities are:

q_e charge on the electron, in coulombs

E voltage between the two electrodes, in volts

m_e mass of the electron, kg

v velocity of the electron, metres/second

From the definition of Coulomb above, equation 1, the electron charge is:

$$\begin{aligned} q_e &= \frac{1}{6.242 \times 10^{18}} \\ &= 1.6 \times 10^{-19} \text{ Coulombs} \end{aligned}$$

The mass of an electron is 9.109×10^{-31} kg. Rearrange equation 6 to solve for velocity v , and then substitute values for q_e and m_e :

$$\begin{aligned} v &= \sqrt{\frac{2q_e E}{m_e}} \\ &= \sqrt{\frac{2 \times (1.6 \times 10^{-19}) \times 350}{9.109 \times 10^{-31}}} \\ &= 11.1 \times 10^6 \text{ metres/sec} \end{aligned}$$

The velocity of light is 3×10^8 metres per second, so the electron velocity is about 3.7% of the speed of light.

Notice that the spacing of the electrodes has no effect on the final velocity of the electron. Because of the square root, the accelerating voltage must be quadrupled to double the final velocity. At higher accelerating voltages, the velocity becomes sufficiently close to the speed of light that relativistic effects apply and the electron mass increases. Relativistic effects are not accounted for in these equations.

2.6 Resistance and Ohm's Law

Current is proportional to (increases with) voltage V . The constant of proportionality is the *conductance*, symbol G , measured in *siemens*.

$$I = GV \quad (7)$$

Conductance is somewhat useful, but its inverse, *resistance* R measured in *ohms* (greek symbol Ω) is far more common¹⁰.

$$R = \frac{1}{G} \Omega \quad (8)$$

Then the current:voltage relationship using resistance is

$$I = \frac{V}{R} \quad (9)$$

This is *Ohm's law*¹¹, usually written as:

$$V = I \times R \quad (10)$$

where the algebraic quantities are

- V Voltage across the resistor, volts
- I Current through the resistor, amps
- R Resistance, ohms.

Notice that the same equation applies if current is in milliamperes and resistance is in kilohms, or current is in microamperes and resistance is in megohms. (Both these are common in practice).

As the graph of figure 10 shows, a *resistance* is a constant of proportionality between voltage and current that extends into the first and third quadrants. If the current reverses, the voltage will also reverse, so a resistor is a *bilateral* device. Not all devices are - a diode has low resistance in one direction and high resistance in the other.

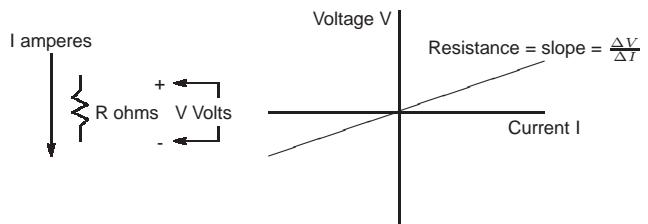


Figure 2: Ohm's Law

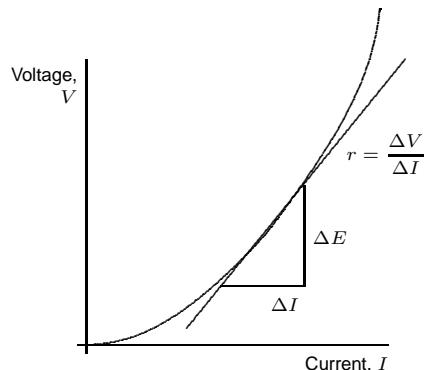


Figure 3: Incremental Resistance

¹⁰In an early example of geek humour, the unit of conductance was originally known as the *mho*.

¹¹Georg Ohm published his original paper in 1827 [21]. To determine his law of electrical conduction, he used a thermocouple [22] as a voltage source, a galvanometer (section 2.16 on page 68) as a current detector, and various conductive materials as resistances [23].

The capitalized quantities represent DC current and voltage, but the same equation applies to alternating current and voltage. In that case, the equation would be written as

$$v = i \times R \quad (11)$$

where v and i are the AC quantities.

It's often useful to think in terms of *incremental* quantities, ie *the change of* some quantity, represented by Δ . So the change of voltage ΔV is equal to the change of current ΔI times resistance:

$$\Delta V = \Delta I \times r \quad (12)$$

This is illustrated in figure 3, where the relationship between voltage and current is non-linear. A tangent to the curve defines the *incremental resistance* at that particular point. As the tangent point moves along the curve, the slope of the tangent changes and the incremental resistance changes.

Discovering Ohm's Law

Ohm's law is easy enough to verify if you have a modern voltmeter and ammeter. Apply various voltages to a fixed resistor. Measure the corresponding currents. Plot the voltage and current values. Place the best straight line through the results, and measure the slope. This is the resistance.

However, there is a fallacy in doing this. Modern digital voltmeters and ammeters require an internal *voltage reference*. The implicit assumption is that there is a known voltage available to compare to the unknown. Georg Simon Ohm, who discovered the law that bears his name, did not have access to such a reference. This makes his discovery quite remarkable.

2.6.1 Example Resistances

Typical units of resistance in electronics:

millionohm	$m\Omega$	$10^{-3}\Omega$
ohm	Ω	
kilohm	$k\Omega$	$10^3\Omega$
megohm	$M\Omega$	$10^6\Omega$

There is some possibility of confusion between millionohm ($m\Omega$) and megohm ($M\Omega$). Fortunately, millionohm value resistors are quite uncommon. Megohm resistors are much more usual.

Notice that $10^3\Omega$ is written as *kilohms*, not *kiloohms*. Similarly, $10^6\Omega$ is written as *megohms*, not *megaohms*.

Consistent Units

Ohm's law can be expressed in volts, amps and ohms. It can also be expressed in volts, millamps and kilohms or volts, microamps and megohms. This is very convenient when analysing a circuit.

Example What is the current through a $2k\Omega$ resistor if there is 10 volts across it?

Solution $I = V/R = 10/2 = 5$ mA. Because resistance is in $k\Omega$, the current is in millamps.

Example What is the voltage across a $5M\Omega$ resistor if there are $3\mu A$ through it?

Solution $V = I \cdot R = 5 \times 3 = 15$ V. Because resistance is in $M\Omega$ and the current is in μA , the voltage is in volts.

2.6.2 Resistance Standards

Like the voltage standard, there is a resistance standard that can be related to fundamental constants: the *quantum hall effect* [24]. Unfortunately, it requires very sophisticated equipment including a cryostat that can cool the standard to within a degree or so of absolute zero. However, this device can act as a primary standard to calibrate secondary standards that are practical for lab use.

High precision resistors are available at reasonable cost. Reference [25] shows resistors with 0.005% initial accuracy and 1ppm/K temperature stability. For example, a 1000 ohm resistor of this type would be expected to vary by $\pm 0.05\Omega$ initial value and $1m\Omega$ per K (or $^{\circ}\text{C}$).

A voltage standard (section 2.4.3) and a resistance standard can then be used to produce a standard current, according to Ohm's Law. For example, a voltage standard with 0.05% accuracy combined with a resistance of 0.01% accuracy would produce a current known to 0.06% accuracy (section 3.11)¹².

2.6.3 Standard Values

Component manufacturing inevitably produces components with a certain amount of variation, known as a *tolerance* in value around the *nominal* value.

Since a 619Ω nominal value resistor with 1% tolerance can have any value between 612.8Ω and 625.2Ω , there is no point in manufacturing, say, a 620Ω resistor. Because a resistor of any given nominal value may have a range of values, manufacturers make resistor values in certain *standard values*. In the case of 1% resistors, the next higher value is 634Ω , which may have any value between 627.7Ω and 640.3Ω .

The standard values for 1% resistors between 100 and 1000Ω are shown in figure 4. These values repeat at decade intervals. For example, to obtain the values between 10 and 100Ω , divide the table entries by ten.¹³

If you need a 620Ω , 1% resistor, then your options are to choose a nominal value of 619Ω or 634Ω .

100	102	105	107	110	113	115	118
121	124	127	130	133	137	140	143
147	150	154	158	162	165	169	174
178	182	187	191	196	200	205	210
215	221	226	232	237	243	249	255
261	267	274	280	287	294	301	309
316	324	332	340	348	357	365	374
383	392	402	412	422	432	442	453
464	475	487	499	511	523	536	549
562	576	590	604	619	634	649	665
681	698	715	732	750	768	787	806
825	845	866	887	909	931	953	976

Figure 4: 1% Standard Values

2.6.4 Conductivity

For a conductor, the resistance is proportional to its length and inversely proportional to its cross-sectional area. That is, making a conductor longer increases its resistance. Making it thicker decreases the resistance. The constant of proportionality is a property of the material, known as the *resistivity*.

In formula form, this is:

$$R = \rho \frac{L}{A} \quad (13)$$

where the quantities are:

¹²Considerable care must be taken to ensure that the device measuring the current has no effect on the resistance. Common current-measuring devices have a non-zero resistance themselves. As well, passing current through a resistor causes it to heat and thereby to change resistance slightly. This effect must be taken into account.

¹³The table of figure 4 is from [26]. That reference shows standard values for other common tolerances.

R Resistance in ohms

ρ The resistivity of the material, measured in ohm-metres

L Length of the sample, in metres

A Area of the sample (such as the cross-sectional area of a wire), in square metres

A table of resistivities for different materials is shown in [27].

2.6.5 Designing Resistors

A knowledge of the conductivity equation equips one to custom design resistors.

Example Nichrome wire is commonly used for electrical heating and as the conductor in resistors that must dissipate considerable power [28]. Assuming #20AWG (American Wire Gauge) nichrome wire is to be used to construct a power resistor, how many metres are required to produce a resistor of 1000Ω ? According to [27], the resistivity of nichrome is 1.10×10^{-6} ohm-metres.

Solution According to the wire table (section 15.19 on page 482) the area of #20AWG wire is 8.023×10^{-4} inches 2 . To use this in equation 13, the area needs to be converted into square metres. (See section 1.2 on page 48 for an explanation of the method).

First, let's convert from square inches to square centimetres. The conversion factor is 1 inch = 2.54 cm. Rearrange this so that the constant 1 is on the left side of the equation. Then we have:

$$1 = \frac{2.54 \text{ cm}}{\text{inch}}$$

If we multiply the area in square inches by this (squared), the square inches will cancel, leaving square centimetres as the unit of measurement:

$$\begin{aligned} \text{Area inches}^2 &= \text{Area inches}^2 \times 1 \\ &= \text{Area inches}^2 \times \left[\frac{2.54 \text{ cm}}{\text{inch}} \right]^2 \\ &= 8.023 \times 10^{-4} \times 2.54^2 \\ &= 5.176 \times 10^{-3} \text{ cm}^2 \end{aligned}$$

Now we repeat the process, this time converting from square centimetres to square metres. The conversion factor is 100 cm = 1 metre. Then we have:

$$1 = \frac{\text{metre}}{100 \text{ cm}}$$

Multiply that by the value for area in cm^2 :

$$\begin{aligned} \text{Area cm}^2 &= \text{Area cm}^2 \times 1 \\ &= \text{Area cm}^2 \times \left[\frac{\text{metres}}{100 \text{ cm}} \right]^2 \\ &= 5.176 \times 10^{-3} \times \frac{1}{10000} \text{ metres}^2 \\ &= 5.176 \times 10^{-7} \text{ metres}^2 \end{aligned}$$

Now rearrange equation 13 to solve for length L and substitute numerical values:

$$L = \frac{RA}{\rho}$$

$$\begin{aligned}
 &= \frac{1000 \times (5.176 \times 10^{-7})}{1.10 \times 10^{-6}} \\
 &= 470 \text{ metres}
 \end{aligned}$$

In most applications it would be hopelessly impractical to coil up 470 metres of nichrome wire to make a 1000Ω resistor. To reduce the length, one would choose a smaller diameter (higher gauge number) wire. On the other hand, the wire has to be large enough to carry whatever current flows through the resistor. The absolute maximum current is shown in the Wire Table as the *Fusing Current*.

Example

A designer plans to use the resistance of a trace on a printed circuit board as a current sensing resistor. The resistor is to generate 100mV at 1 amp current, so the resistance is 0.1Ω . Assume that the thickness of the copper is 1.35 mils (1.35×10^{-3} inches), which is typical for 1 oz copper plating [29]. (The measure ‘1 oz copper’ is the weight of the copper per square foot.)¹⁴ The trace is 20 mils wide. How long should the trace be, in centimetres?

Solution

We can apply equation 13 to solve for length, once we have the cross-section of the conductor and the resistivity of copper.

According to [27], the resistivity ρ of copper is 1.72×10^{-8} ohm-metres.

Determining the area of the wire is an exercise in converting units (section 1.2, page 48).

$$\begin{aligned}
 A &= w \times h \\
 &= \left[\frac{20}{1000} \text{ inches} \right] \times (1.35 \times 10^{-3} \text{ inches}) \\
 &= 27 \times 10^{-6} \text{ inches}^2
 \end{aligned} \tag{14}$$

where w and h are the width and height of the copper conductor. We need to convert this to square metres. Ready to hand is the conversion factor of $2.54 \text{ cm} = 1 \text{ inch}$, so we'll start there. We'll also use 100 centimetres (cm) = 1 metre.

$$\begin{aligned}
 27 \times 10^{-6} \text{ inches}^2 &= 27 \times 10^{-6} \text{ inches}^2 \times \left[\frac{2.54 \text{ cm}}{1 \text{ inch}} \right]^2 \\
 &= 174 \times 10^{-6} \text{ cm}^2 \times \left[\frac{1 \text{ metre}}{100 \text{ cm}} \right]^2 \\
 &= 17.4 \times 10^{-9} \text{ metres}^2
 \end{aligned}$$

Rearrange equation 13 to solve for length L and substitute numerical values:

$$\begin{aligned}
 L &= \frac{RA}{\rho} \\
 &= \frac{0.1 \times (17.4 \times 10^{-9})}{1.72 \times 10^{-8}} \\
 &= 0.101 \text{ metres} \\
 &= 0.101 \text{ metres} \times \left[\frac{100 \text{ centimetres}}{1 \text{ metre}} \right] \\
 &= 10.1 \text{ centimetres}
 \end{aligned}$$

¹⁴We're not being perverse in using inches for this problem. North American practice is dictated by the USA, which still uses English units of measurement. However, we're in Canada so we'll state the result in a metric unit.

Using a PCB (printed circuit board) trace for the resistor saves the cost of a resistor component and the cost of assembly. However, the design should be checked for the heating effect of the current, which may cause the resistance to vary or shorten the life of the trace. Vandersleen in [30] relates current and temperature rise in a copper trace as

$$I = 0.025\Delta T^{0.45}W^{0.79}Th^{0.53} \quad (15)$$

where the variables are:

I current in amperes

ΔT temperature rise above ambient, °C

W width of the PCB trace, mils

Th thickness of the PCB trace, mils

Anything more than a 10°C temperature rise should be viewed with caution.

2.6.6 European Style Notation

The decimal point in component values can easily be lost or confused, especially on a schematic diagram that has been reproduced over several generations. The european style of component value notation deals with this by replacing the decimal point with the multiplier symbol. For example, $5.6k\Omega$ would be written as $5k6$. In using the first form, it is possible that the decimal point may get lost. With the second form, the multiplier symbol is less likely to get lost. Furthermore, it should be clear from the position of the notation on the diagram that this refers to a resistor, so the Ω symbol is not necessary.

When the multiplier is unity, the convention is to put a component symbol such as R, C, L, A, V, or W in the decimal place position. Examples are shown in figure 5.

This is most commonly seen in the designation of resistors, but it can be used for other values as well. For example, a voltage supply could be labelled $2V6$ for 2.6 volts.

2.7 Ideal Voltage Source

An *ideal voltage source* is produces a constant electro-motive force (output voltage) regardless of the output current. As shown in the figure, the output characteristic is a horizontal straight line, reaching out into the dangerous world of infinite currents.

Notice that the current extends into negative as well as positive regions. Here we have defined positive current as flowing out of the voltage source. If the ideal voltage source were a battery, then output current would be regarded as battery discharge current.

Negative current flows into the voltage source, and would charge a battery.

The ideal voltage source is useful as a concept but does not exist in practice. The terminal voltage of any real-world source decreases as the output current increases. (section 2.17, page 72).

However, there are many situations where a real-world voltage source may be treated as ideal. For that reason, the symbol for the ideal voltage source, figure 6, is also frequently used to represent a real battery or other real-world voltage source.

Value	Notation
0.1Ω	R1
5.3Ω	5R3
28Ω	28R
$33\mu F$	33C
$4.7k\Omega$	4k7
$0.33\mu H$	$\mu 33$

Figure 5: European Style Notation

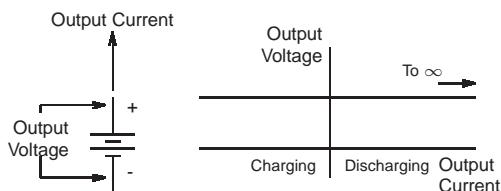


Figure 6: Ideal Voltage Source

Polarity

The positive and negative signs next to the voltage source indicate the *reference polarity* of the source. This reference polarity works together with the sign of the voltage to characterize the voltage source. (The positive and negative symbols are redundant: the long bar of the voltage source graphic is positive.)

A voltage measuring instrument (a voltmeter) has a reference polarity as well. *When the reference polarities of the voltage source and voltmeter match, then the voltmeter will read the nameplate value of the voltage source.*

Consider the source of figure 7(a). This is a +9 volt source. When the reference terminals of the voltmeter are connected to the reference terminals of the source, plus to plus, minus to minus, the voltmeter will read +9 volts.

This is the usual situation. The battery manufacturer marks the terminal on the battery that sources current as the positive terminal *and* what is equally important, shows the battery voltage as a positive 9 volts.

However perverse this may seem, there is another possibility. As shown in figure 7(b) this same source could be marketed as a -9 volt source, providing that its reference terminals are interchanged. The terminal that sources current is now the negative terminal. When a voltmeter is connected to the reference terminals on this battery, it will read -9 volts. This is correct: interchanging the reference terminals is equivalent to changing the sign on the voltage value.

Put another way, a -9 volt source will produce current out of the negative reference terminal. The sources of figure 7(a) and figure 7(b) are identical.

Why is this important? In the process of doing circuit calculations, it is sometimes necessary to begin with an arbitrary choice for the reference polarity of a voltage source. If the result of the calculations is a negative value for this voltage source (eg, -9 volts), we could change this to a positive value (+9 volts) by reversing the reference polarity of the voltage source.

2.8 Ideal Current Source

An *ideal current source* is a device that produces a constant output current regardless of the output voltage. As shown in the figure, the output characteristic is a vertical straight line.

The voltage source is easier to understand, because a battery is an approximate voltage source and batteries can be purchased at the Dollar Store. Current sources are less familiar: they aren't available over the counter. Their utility is in modelling the behaviour of other devices, such as the output of a junction transistor.

In the real world, current sources usually produce small currents and are rather limited in the allowable voltage that they can sustain, known as the *voltage compliance*. The voltage compliance is the current-source analog to the short-circuit current limit for a voltage source.

The symbol and characteristic for a DC current source is shown in figure 8.

Just as the current direction in the ideal voltage source was unspecified, the polarity of the voltage across the ideal current source is unspecified. In other words, the voltage across the ideal current source can be of either polarity. This has no effect on the output current.

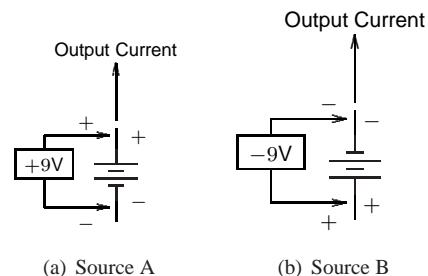


Figure 7: Identical Voltage Sources

2.9 Kirchhoff's Voltage Law

The algebraic sum of voltages around any circuit loop, is zero.

This is best done by example, so here we go:

Example

What's the voltage V_{ab} ?

The current out of the generator I_o is fixed at 3mA and the voltage across its terminals is defined by the external circuitry. In this case, the external circuitry is a $1\text{k}\Omega$ resistor with 3mA through it, so V_{ab} is 3 volts. Going clockwise around the circuit loop, there is a 3 volt rise across the current generator and a 3 volt drop across the resistor, for an algebraic sum around the loop of zero.

Example

Now consider the example shown in figure 10. What's the current in the loop?

- Choose a direction around the loop. Anything will work if you are consistent, but it's easier to choose a direction that is the same as the direction of flow of current around the loop – if it is known. In this case, it's apparent that the large voltage source E_1 will drive the current, so the current will run clockwise around the loop.
- Label the resistors with their voltage drops. You can arbitrarily choose which end to label positive. If the result of the calculations is negative voltage, that indicates that the original choice was incorrect.
- Do the algebraic sum of voltages around the loop. The **second** sign of a voltage defines its polarity, so V_2 is negative and V_3 is positive.

$$+E_1 - V_1 - E_2 - V_2 + V_3 = 0$$

(Notice the sign of V_3 .)

- Use Ohm's law to substitute for the voltages across the resistors.

$$+E_1 - IR_1 - E_2 - IR_2 - IR_3 = 0$$

- Collect terms, substitute the known quantities and solve the equation for the current I in the loop.

$$\begin{aligned} +20 - 50I - 5 - 50I - 50I &= 0 \\ 150I &= 15 \\ I &= 0.1 \text{ amps} \end{aligned}$$

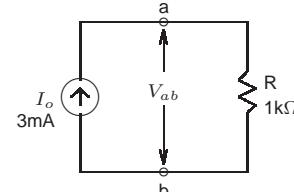


Figure 9: Kirchhoff's Voltage Law

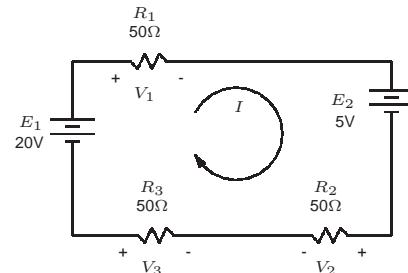


Figure 10: Kirchhoff's Voltage Law

Here are some questions to test your understanding:

Question

A 0.1mA current source is placed in series with the loop, facing in the clockwise direction. What's the voltage across the current source? (No math should be necessary, reason it out.)

Question

A 0.2mA current source is placed in series with the loop, facing in the clockwise direction. What's the voltage across the current source?

2.10 Kirchhoff's Current Law

The algebraic sum of currents into any circuit node, is zero.

With KVL under our belt from the previous section, we can tackle something a bit more complex:

Example

What's the voltage V_{ab} ? (The currents are in mA, the resistances in k Ω , and the voltages in volts.)

The current out of the generator I_o is fixed at 1mA and the voltage V_{ab} across its terminals is defined by the external circuitry. Calculate V_{ab} .

Solution

The key is to get as many equations as there are unknowns and then collapse them into each other until there is a solution. Notice that there are only 2 independent circuit loops and 1 independent circuit node in this circuit. The 3rd loop and the 2nd node are restatements of the others.

The currents at the upper-middle node X, by Kirchhoff's current law:

$$I_1 + I_3 - I_2 = 0 \quad (16)$$

The current I_3 is known to be 1mA:

$$I_3 = 1 \quad (17)$$

Expressing the voltages in terms of the currents through them:

$$V_1 = 2I_1 \quad (18)$$

$$V_2 = 4I_2 \quad (19)$$

$$V_3 = 3I_3 \quad (20)$$

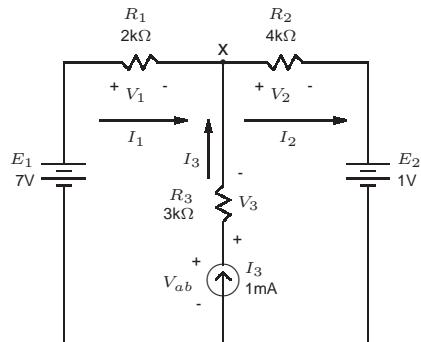


Figure 11: Kirchhoff's Current Law

Now the loop equations by KVL: first the left loop

$$+ 7 - V_1 + V_3 - V_{ab} = 0 \quad (21)$$

and then the right loop

$$- 1 + V_{ab} - V_3 - V_2 = 0 \quad (22)$$

where I have substituted 7 for E_1 and 1 for E_2 .

In total, we have the 7 unknowns $I_1, I_2, I_3, V_1, V_2, V_3$ and V_{ab} and 7 equations to find them.

A certain amount of algebraic manipulation eventually leads to the result: $V_{ab} = 9.3V$

This is a fine method for computer solution, because there are algorithms that can automate the process. However, this brute-force solution method is tedious and, more important, it doesn't yield much understanding of the circuit. There are easier and more revealing ways to solve the circuit, rather than solve 7 equations in 7 unknowns. In subsequent sections these will be revealed.

2.11 KVL, KCL and Circuit Analysis

Kirchhoff's Voltage and Current Law (KVL and KCL, as they are known) are important and should be understood by anyone who practises this trade. We use them implicitly in many cases: by KVL, if the voltage rise is *plus whatever* between these points, it must be *minus whatever* between those points.

Similarly, we understand from KCL that current cannot pile up at a circuit node: the total of what flows into a circuit node must flow out.

Those are fundamental. However, grinding through lines of simultaneous equations is *not* the way practising designers create or reverse-engineer circuits. There are tools that are much easier and faster to use: *Thevenin's Theorem*, the *Superposition Theorem* and *combining resistors in series and parallel*, for example. On the other hand, very simple applications of KVL and KCL are required to develop these useful tools.

The simultaneous solution of KVL and KCL equations *is* also essential for the solution of electrical networks by computers. It is the basis of programs like *Spice* that simulate and solve electronic circuits. Software to monitor the flow of electrical currents in a power grid would make heavy use of KVL and KCL.

So:

- KVL and KCL are fundamental circuit analysis tools.
- KVL, KCL and simultaneous equations are the basis for computer analysis of circuits.
- Solving equations by hand is not a practical technique for humans.
- There are better tools for human-based circuit design.

2.12 Grounded and Floating Components

Figure 12 shows a variety of possible *grounded* and *floating* component arrangements.

It is usual to designate some point in the circuit as the reference terminal for voltage measurements. Since this terminal is sometimes connected to earth ground, it is colloquially referred to as the *ground* terminal. When the enclosure or frame of the equipment is conductive, it is also usual practice for this reference terminal to be connected to the chassis.

Although this is not always done, it is good practice to differentiate between these two ground connections, using the symbol  for an earth ground and  for a chassis ground. (A third symbol  is sometimes used to indicate a special ground for noise-sensitive analog circuits.)

The current flowing through resistor R is the same in all the circuits of figure 12.

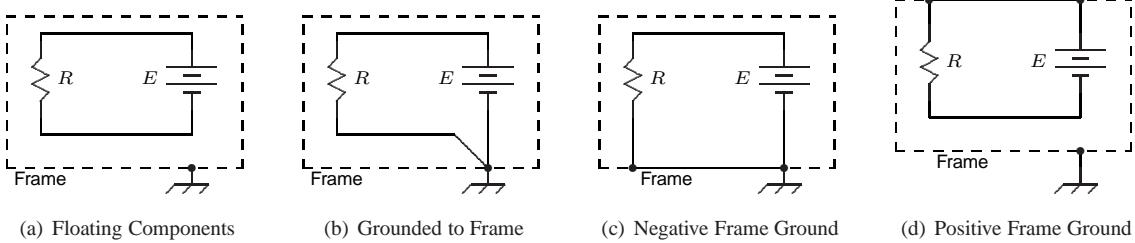


Figure 12: Floating and Grounded Components

- In figure 12(a), both components are completely *floating*, that is, not connected to the frame. No ground reference point is specified.
- In figure 12(b), the negative terminal of the power supply E is connected to the frame, which is considered to be a reference point for the circuit. This circuit consists of a *grounded source* and a *floating load*.
- In figure 12(c), the return current from the load resistance flows through the frame. This saves wire, but requires that the frame – and the connections to it – be a reliable path for current.
- In figure 12(d), the *supply* current flows through the frame and the reference (ground) terminal is now the positive terminal of the power supply.

At one point in history, automobiles used the schemes in figures 12(c) and (d). The steel chassis of the automobile acted as one of the conducting paths for electrical current. American vehicles generally used a negative ground and British vehicles a positive ground.

A device may be grounded to the frame by an explicit wiring connection or implicitly via its own frame. For example, the mounting frame of a DC generator is the negative electrical terminal, and this must attach to a metal support structure. In that case, the generator is implicitly a grounded device and cannot easily be floated off ground. Furthermore, for safety reasons, it may be *required* that the frame be attached to an earth ground. Consequently, any circuitry that attaches to the generator must do so at its positive terminal. The negative terminal is not accessible.

2.13 Resistors in Series

Devices are said to be connected in *series* when they conduct the same current.

The resistors in figure 13 are electrically in series¹⁵ even though they are drawn parallel to each other.

When resistors are connected in series, the total effective resistance is the sum of the individual resistances. For three resistors,

$$R_T = R_1 + R_2 + R_3 \quad (23)$$

This may seem obvious, but bear with me:

When one resistor is much larger than the other, the largest resistance dominates: the total resistance is approximately equal to the larger resistance.

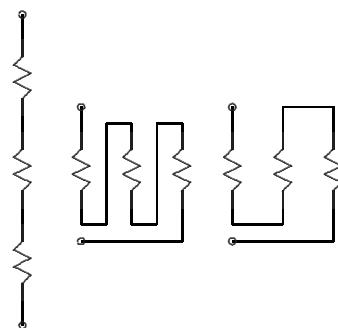


Figure 13: Series Connections

¹⁵Note: *series*, not *serious*, as more than one neophyte has mistaken.

2.14 Resistors in Parallel

Devices are said to be connected in *parallel* when they receive the same voltage across their terminals. The resistors in figure 14 are electrically in parallel.

When resistances are connected in parallel, the total conductance (reciprocal of resistance) is equal to the sum of the individual conductances. In formula form for three resistances:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \quad (24)$$

To find the parallel equivalent of some resistors, sum up the reciprocals of the individual resistances and then take the reciprocal of the result, equation 25.

$$R_T = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)^{-1} \quad (25)$$

For the special case of two resistors, we have:

$$R_T = \frac{R_1 R_2}{R_1 + R_2} \quad (26)$$

Equation 26 does not scale easily to more than two resistors: it's better to deal with conductances as in equation 25.

When one resistor is much smaller than the other, the smaller resistor dominates: total parallel resistance is approximately equal to the smaller resistance.

2.15 Resistor Networks

More complex networks of resistors can often be reduced to one equivalent resistance by combining resistors in series and parallel. For example, in figure 15 suppose each resistor is $1k\Omega$. Then the total resistance is $1k5\Omega$ (1500Ω).

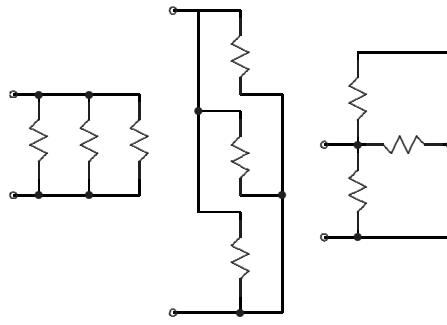


Figure 14: Parallel Connections

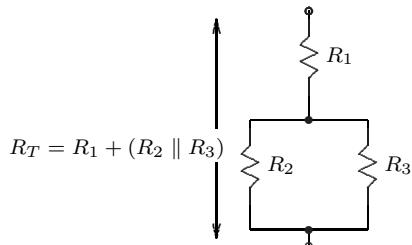


Figure 15: Resistor Network

2.16 Meters for Voltage, Current and Resistance

2.16.1 Moving Coil Meter Movement

Analog meters are based on the Moving Coil Meter Movement¹⁶ shown in figure 16. A meter movement consists of a rotating coil of wire (the *rotor*) in a magnetic field (the *stator*). The rotation of the coil is opposed by a delicate coil spring. When a current passes through the coil, it sets up a magnetic field that creates a torque on the coil. The coil then rotates an amount that is proportional to the current. An indicator needle is attached to the coil. This is calibrated to read out the current, or some quantity proportional to the current. Based on this, the moving coil meter is inherently a current measuring device, but it can be used to measure voltage and resistance.

The two parameters of interest in a moving coil meter are the full-scale current (FSD, full scale deflection) and resistance. The full-scale current is typically 1mA or 100 μ A. The resistance is typically a few hundred ohms.

In these days of multi-digit meters, we tend to regard moving-coil instruments as antiques. However, they did have one major advantage – it was easy to determine a reading *trend* (is the value increasing, decreasing or staying the same?) by watching the movement of the meter needle. It's much more difficult to detect a trend from a digital readout, in which the displayed value is updated at intervals. For that reason, some digital readouts include a bar display which indicates the reading trend.

2.16.2 Characterising an Unknown Meter Movement

The full scale current I_{FSD} is determined by driving current through the meter and noting the current at full scale deflection. An adjustable current source would be ideal, but they are not a common item of lab equipment.

An adjustable voltage supply cannot be used directly since moving coil meters are very low resistance. The resulting current – even at very low supply voltage – will exceed the meter capability and it's likely to be destroyed. However, an adjustable voltage supply can be used if a fixed resistance is added in series with the meter movement.

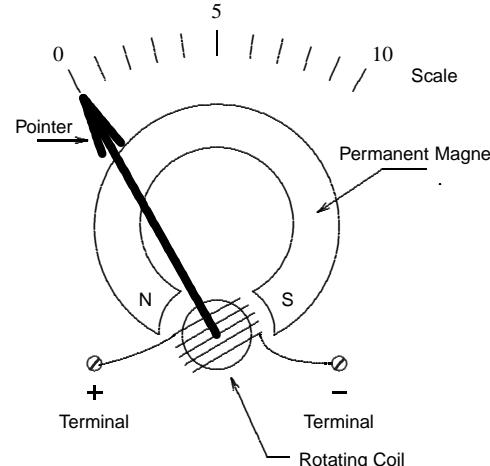


Figure 16: Moving Coil Meter

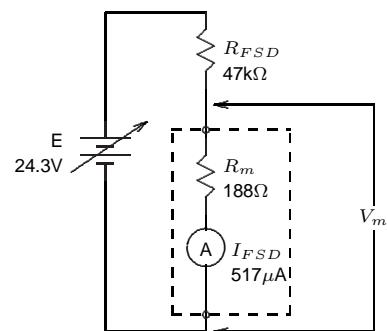


Figure 17: Characterizing a Moving Coil Meter

Example

An unknown moving coil meter is to be characterized¹⁷. The test setup is shown in figure 17. Full scale deflection in a meter movement is typically in the range 100 μ A to 1mA. With a 25 volt adjustable supply, a resistor in

¹⁶Also known as a *galvanometer* [31] or *d'Arsonval meter movement*.

¹⁷It can be difficult to determine what is inside an unknown meter movement. Sometimes there is a large series resistance, to convert the movement to measure voltage. Sometimes there are rectifier diodes to convert AC to DC. So if this technique does not work, you may have to disassemble the meter.

the vicinity of $50\text{k}\Omega$ will allow up to $500\mu\text{A}$, so that seems like a suitable value to put in series with the meter movement. This assumes that the resistance of the meter movement is negligible by comparison, which is usually the case.

We have a $47\text{k}\Omega$ resistor so we connect that in series with the meter and adjust the output voltage of the supply until that the meter reads full scale. The voltage across the resistor is 24.3 volts, so the full scale current is $517\mu\text{A}$. This is probably a $500\mu\text{A}$ meter movement.

The voltage V_m across the meter movement is 94mV when the meter current I_m is $517\mu\text{A}$, so the resistance of the meter is:

$$\begin{aligned} R_m &= \frac{V_m}{I_m} \\ &= \frac{94 \times 10^{-3}}{517 \times 10^{-6}} \\ &= 188\Omega \end{aligned}$$

As expected, the meter movement resistance is much less than the external resistance¹⁸.

Now this meter can be designed into a voltage, current or resistance measurement circuit.

2.16.3 Moving Coil Voltmeter

To begin with, suppose we need a 10V (FSD) voltmeter. We intend to use a meter movement with full-scale current I_{FSD} of $100\mu\text{A}$ and meter resistance R_m of 200Ω . Notice that the circuit will need to supply this measuring instrument with $100\mu\text{A}$. It must be able to do so without a significant effect on the operation of the circuit.

The design is very simple: a resistor is in series with the meter movement. The value is chosen so that the current through it is $100\mu\text{A}$ when the voltage across it is 10 volts. The value of the resistance is an application of Ohm's law:

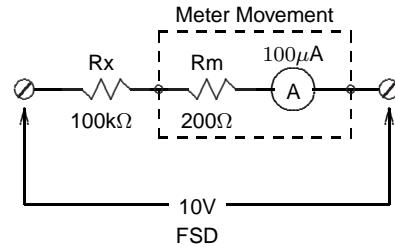


Figure 18: Voltmeter, 10V FSD

$$\begin{aligned} R &= \frac{V}{I} \\ &= \frac{10}{100 \times 10^{-6}} \\ &= 100,000 \Omega \end{aligned}$$

The resistance of the meter movement is included in the total, so the external resistance in series with the meter is equal to $100\text{k}\Omega$ minus the meter resistance (figure 16(b)). In this case, the meter resistance is less than 1% of the series resistance, so we'll ignore it. The accuracy of the meter depends on the meter movement and the precision of that resistor. Alternatively, part of that series resistance would be a variable resistor that is adjusted to make the meter movement read full scale when the applied voltage is 10 volts.

¹⁸There is an alternative technique to determine meter resistance, if one does not have access to a millivolt meter like the one used here. See section 3.5, page 133

Voltmeter Sensitivity

The sensitivity of the basic voltmeter can be increased by driving it with an amplifier. The amplifier can have a very large input resistance, so the loading effect is negligible¹⁹.

The current drawn by a voltmeter is sometimes specified in *ohms per volt*, which we will call the *sensitivity*. The sensitivity is the reciprocal of the full scale meter current, so a smaller current gives a larger value of sensitivity, which makes intuitive sense. The resistance presented to the measurement circuit is the product of the full scale deflection in volts times the sensitivity.

Example

In the case of the previous voltmeter, it has a 10 volt full scale deflection and its sensitivity S is:

$$\begin{aligned} S &= \frac{1}{I_{FSD}} \\ &= \frac{1}{100 \times 10^{-6}} \\ &= 10\text{k ohms/volt} \end{aligned} \quad (27)$$

Then the voltmeter appears as a resistance of:

$$\begin{aligned} R &= V_{FSD} \times S \\ &= 10 \text{ volts} \times (10 \times 10^3) \text{ ohms/volt} \\ &= 100\text{k}\Omega \end{aligned} \quad (28)$$

The value of $100\text{k}\Omega$ makes sense, since we previously determined that the total resistance in series with the meter movement should be $100\text{k}\Omega$.

2.16.4 Moving Coil Ammeter

The ideal current meter appears as a short circuit, zero resistance. It is connected so the current to be measured flows through the meter. Since the meter has no resistance, it creates no ohmic voltage drop, and it has no effect on the circuit.

In general, a moving coil meter cannot be inserted directly in a circuit to measure current. For example, the meter movement characterised in section 2.16.2 above has a series resistance of 188Ω and full scale current of $517\mu\text{A}$. The series resistance is too large and the full scale current is unlikely to match the full scale current to be measured. Instead, the current to be measured is caused to flow through a *current sensing resistor*. The voltage across that resistor then creates a proportional current in the meter movement.

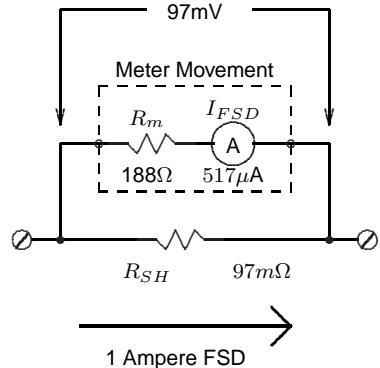


Figure 19: Ammeter, 1A FSD

¹⁹Early voltmeter amplifiers used vacuum tubes and were known as VTVM: Vacuum Tube Volt Meter. An example circuit with two pages of description is in reference [32]. The amplifier requires a power supply, which makes the meter less portable. In the case of the VTVM, the meter required 117V line power. A solid-state version of this circuit, using JFETs (section 8.2 on page 268), is in [33]. The circuit uses a single 9 volt battery. A modern version would use an operational amplifier, such as shown in [34].

For example, a 1 amp measuring circuit, using the moving coil meter characterised in section 2.16.2, is shown in figure 19. The *shunt resistance* R_{SH} conducts most of the current and creates a voltage that drives the meter movement to its maximum deflection, $517\mu\text{A}$.

Under these conditions, the voltage across the meter movement is:

$$\begin{aligned} V_m &= I_{FSD} \times R_m \\ &= (517 \times 10^{-6}) \times 188 \\ &= 97 \times 10^{-3} \text{ volts} \end{aligned} \quad (29)$$

The current through the shunt resistor is 1.0 amperes minus the meter current of $517\mu\text{A}$, which is small enough to be ignored. So the shunt resistance value is

$$\begin{aligned} R_{SH} &= \frac{V_m}{I} \\ &= \frac{97 \times 10^{-3}}{1.0} \\ &= 97 \times 10^{-3} \Omega \end{aligned} \quad (30)$$

This is not the sort of resistor one buys off the shelf, but there are several ways to construct it:

- Connect a number of small-value resistors in parallel. For example, a quantity of ten, 1Ω resistors in parallel create a resistance of 0.1Ω , close to the required value.
- Use the appropriate length of small-gauge wire. From section 15.19 on page 482, #30AWG wire has a resistance of 103Ω per thousand feet. A resistance of $97\text{m}\Omega$ would therefore be about 11 inches (28cm) in length. (This could be coiled up into a small area.)
- Use a small-width trace on a printed circuit board. An example is given in section 2.6.5 on page 59.
- For very large currents, there are commercially available *shunt* resistors [35]. For example, a $100\text{A}/75\text{mV}$ shunt would be capable of handling 100 amperes of current. At that current, the voltage across it is 75mV . The display device would be a voltmeter with 75mV full scale deflection.

2.16.5 Moving Coil Ohmmeter

Modern instruments force a known value of current I_{test} through the unknown resistance R_x . Then by Ohm's law the voltage across the resistance is the product of these two, as shown in figure 20. For example, if the test current is 1mA , then the voltage across R_x is numerically equal to the resistance in kilohms. If the test current is $1\mu\text{A}$, then the voltage is numerically equal to the resistance in megohms.

The current must be generated by a constant current source²⁰.

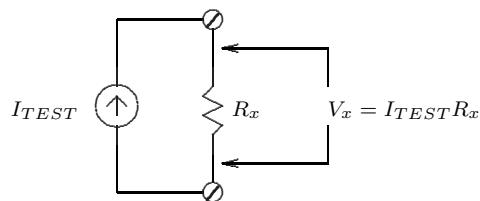


Figure 20: Constant Current Ohmmeter

²⁰A circuit for generating constant current is shown in section 13.15, page 358.

The voltmeter reading is directly proportional to resistance, so the readout scale is linear. This technique is ubiquitous in digital resistance measuring instruments such as the *digital multimeter*, DMM.

The constant-test-current method could be used in conjunction with a moving-coil voltmeter of the type discussed previously. However, the tradition in moving-coil instruments is to measure the current in a circuit of known voltage and unknown resistance, the constant-test-voltage method. The circuit for this technique is shown in figure 21. The current, which is the displayed value on the meter, is inversely proportional to the resistance. This results in a logarithmic scale, with zero resistance (short circuit) corresponding to maximum deflection of the meter and infinite resistance (open circuit) corresponding to zero deflection. Resistor R_{scale} sets the resistance value that appears for mid-scale (half full-scale current).

The adjustable resistor R_{adj} is required to compensate for changes in the battery voltage. Before taking a resistance reading, the operator shorts the meter terminals together and then adjusts R_{adj} so the meter reads full scale (zero ohms)²¹.

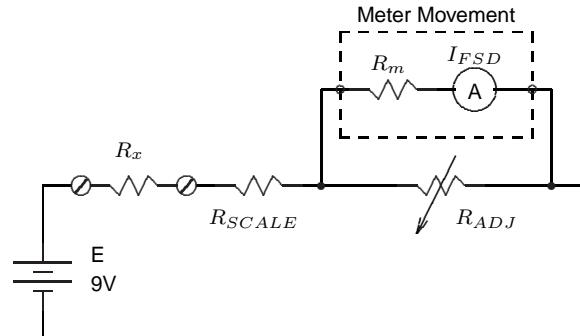
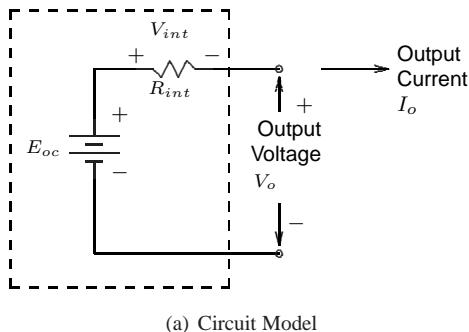


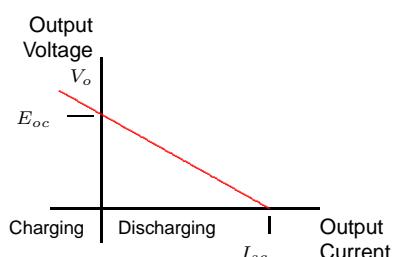
Figure 21: Constant Voltage Ohmmeter

2.17 Practical Voltage Source

Picture the following real-life scenarios:



- You are in your car at night. You turn on the dome light and then start the car. While the engine is cranking, you notice that the dome light dims significantly.
- In your kitchen, when the refrigerator starts you notice that the kitchen light dims.
- In your office on the third floor of an old house, you notice that the screen on your computer contracts slightly when your Lexmark printer runs.



(b) Output Voltage and Current

These are all indicators of the effect of *internal resistance*, R_{int} or R_{int} . These voltage sources – the car battery and the household electrical system – decrease in voltage when supplying current. We need some way of modelling this effect, and that is internal resistance.

Any practical²² voltage source has an *internal resistance* R_{int} . The circuit model is shown in figure 22. It includes the following quantities:

²¹The design equations and an example design for this type of ohmmeter are in [36].

²²I use the term *practical* as an alternative to the usual term, *non-ideal*.

Figure 22: Practical Voltage Source

V_o	Output voltage of the source at some output current, volts
E_{oc}	<i>Open Circuit</i> output voltage of the source
R_{int}	<i>Internal Resistance</i> of the source, ohms.
I_o	Output current of the source.
I_{sc}	<i>Short Circuit</i> output current of the source

The terminal voltage can be obtained using KVL. Summing voltages in a clockwise direction around the loop in figure 22, we have:

$$+ E_{oc} - V_{int} - V_o = 0 \quad (31)$$

The voltage V_{int} across the internal resistance R_{int} is given by Ohm's Law:

$$V_{int} = I_o R_{int} \quad (32)$$

Substitute for V_{int} from equation 32 into equation 31, and solve for V_o :

$$V_o = E_{oc} - I_o \times R_{int} \quad (33)$$

Equation 33 is the expression for the terminal voltage of a source (such as a battery) given the other three quantities: open circuit voltage, output current and internal resistance. This is a very useful expression and circuit model, not only for real world voltage sources such as batteries, but also to model sub-sections of electronic circuitry – as we'll see with Thevenin's Theorem (section 3.1, page 119).

Points to notice:

- The output voltage V_o is equal to the open circuit voltage E_{oc} when the source is open circuited and the output current I_o is zero. This provides a convenient way to measure E_{oc} .
- The internal resistance determines the output slope. A larger resistance causes a steeper dropoff of voltage as the current increases, ie, a larger negative slope.
- If the source is short-circuited, it produces an output current I_{sc} that is equal to the open-circuit voltage E_{oc} divided by the internal resistance R_{int} . This might suggest we could measure R_{int} by measuring the short-circuit current, but in practice many sources cannot be safely short-circuited – the output current is simply too large. So, in practice, we first measure E_{oc} , and then measure the output voltage V_o at some output current I_o . This gives us the required two points to define the output characteristic.
- The output characteristic extends into the region for negative currents. This implies that the terminal voltage rises above the open circuit value when a battery is being charged. This is in fact what happens in practice.

When a practical voltage source is connected to a resistance, the operating point (output voltage and current) of the source may be determined by superimposing the characteristic of the resistive load on the characteristic of the source, as shown in figure 23. As the load resistance decreases, its slope decreases, and the Q point moves down and to the right.

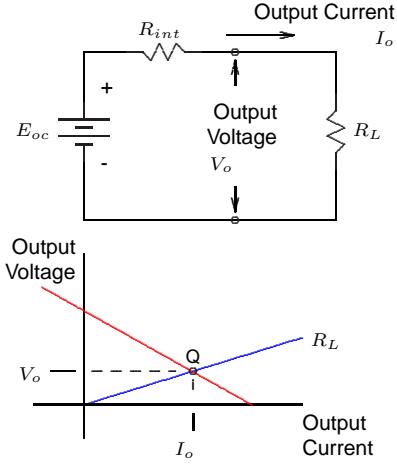


Figure 23: Practical Source with Load

When the internal resistance is *small* (whatever that is ²³) with respect to the load resistance, the internal resistance may be neglected. Consequently, we often ignore the effect of internal resistance if we know it's small compared to the load resistance.

Also, in terms of designing a voltage source, we must keep the internal resistance much smaller than the smallest load resistance.

2.18 Internal Resistance: Case Histories

Now we can revisit some common scenarios as explained by the concepts *ideal voltage source* and *internal resistance*.

Case 1: The Short Circuit

The term *open circuit* refers to a break in the conducting path. In effect, the break is a very large resistance. The term *short circuit* (or simply, a *short*) refers to section of a circuit where there is a low resistance path for current, a kind of electrical *short cut* – the current can bypass the circuit resistance.

An ideal voltage source is quite happy when connected to an open circuit. The output current drops to zero.

However, an ideal voltage source connected to a short circuit (a zero resistance load²⁴) will produce a very large current. In general, this is a *Bad Thing*, because the large current will cause large heat in the wiring resistance. In practice, a short circuit may melt the insulation off wires and lead to a fire.

Consequently, a voltage source may require either a *fuse*²⁵, *circuit breaker*²⁶, or *current limit*²⁷ to keep the output current below a safe maximum value

For an ideal current source, a short circuit load is a Good Thing; an open circuit is Bad because the terminal voltage would go to infinity.

Think of $E = IR$ where I is some constant value and R becomes very large. Then E , the terminal voltage of the current supply, must become very large.

For most practical current sources, the current simply stops when the terminal voltage exceeds the voltage limit for the supply. This voltage limit is referred to as the *voltage compliance* of the current supply and is analogous to the current limit in a voltage supply.

To summarize, for a voltage source an open circuit load is fine, a short circuit load is a problem. For a current source, it's the opposite: an open circuit load is a problem, a short circuit load is fine.

Case 2: House Wiring

A household electrical power distribution system is a *constant voltage* system²⁸. In effect, there is a 115 volt battery available at each outlet in the house. The internal resistance is very low, in the order of a fraction of an ohm.

We could describe this as a constant voltage system in which the current varies with demand.

The short circuit current is limited only by the internal resistance and so it can be extremely large. That's potentially dangerous, but there is a safety mechanism: a fuse or circuit breaker that disables (open circuits) that outlet if the current exceeds a certain value. The maximum current is typically 15 to 30 amperes.

²³A factor of 10 is a common choice.

²⁴To be precise, no conductor (except for a superconductor cooled to near absolute zero) has zero resistance. So a short circuit has a very small albeit non-zero resistance.

²⁵A fuse is a metallic strip with a low melting point. When the current exceeds some maximum value, the fuse melts, opening the circuit.

²⁶A circuit breaker is a switch that is actuated by a current sensing circuit. When the current exceeds a preset maximum the switch opens. Unlike a fuse, a circuit breaker can be reset to re-energize the circuit.

²⁷Electronic current limit circuitry is standard in DC constant-voltage lab power supplies. It can be adjusted to set the current limit and recovers automatically after an overload.

²⁸It's an alternating current system, but for our purposes, we can regard it as battery powered.

A short circuit is still potentially dangerous. This very large current can flow for several milliseconds until the fuse blows or the circuit breaker opens²⁹.

Case 3: Car Battery Internal Resistance

Some idea of the internal resistance of a 12 volt car battery may be obtained from its *cold cranking amps* figure. From [37]:

Cold Cranking Amps is a rating used in the battery industry to define a battery's ability to start an engine in cold temperatures. The rating is the number of amps a new, fully charged battery can deliver at 0° Fahrenheit for 30 seconds, while maintaining a voltage of at least 7.2 volts, for a 12 volt battery.

The higher the CCA rating, the greater the starting power of the battery.

A moderate sized 12 volt car battery might have an open circuit voltage of 14 volts and a CCA rating of 400 amperes. Then the internal resistance is:

$$\begin{aligned} R_{int} &= \frac{E_{oc} - V_L}{I_L} \\ &= \frac{14 - 7.2}{400} \\ &= 0.017 \Omega \end{aligned}$$

To any load resistance greater than ten times this, the battery will appear (approximately speaking) as a constant voltage source.

Case 4: Jump-Starting a Car

If a car battery is inadvertently discharged (by leaving the car lights on after shutting down the engine, for example), it must be *jump started*³⁰. One needs to find another car and jumper from that helper battery to the dead battery.

Figure 24 shows why this must be done with *great care*.

In figure 24(a) the connection is correct: positive to positive, negative to negative. The helper battery voltage is higher than the 'dead' battery, so current will flow from the helper into the dead battery, charging it. As the dead battery charges, the voltage difference decreases and the flow of current decreases. Assuming the helper battery is being charged by a running engine, then the dead battery will charge up to the helper battery voltage and the current flow then decreases to zero.

Figure 24(b) shows the **incorrect** connection. This is a disaster for several reasons. First, the helper and the dead battery push current in the same direction, *out* of the dead battery terminal. This has the effect of further discharging the dead battery and possibly discharging the helper as well. Second, because both batteries drive current in the same direction, the current is created by the sum of the battery voltages divided by the total of the internal resistances. The result is an extremely large current that can melt the insulation on the jumper cables, damage the electrical system of either car or overheat the batteries.

We can gain some insight by applying KVL to this circuit³¹. In the correct wiring of figure 24(a), summing voltages clockwise around the loop we have:

$$+ E_1 - V_1 - V_2 - E_2 = 0 \quad (34)$$

²⁹When a circuit breaker is activated in this fashion, it is said to be *tripped*. A tripped breaker can be recognized by the position of the handle, which is mid-way between the ON and OFF positions.

³⁰If the car uses a standard transmission and if the car is on a hill, it may be possible to get up sufficient speed to crank the engine and start it.

³¹For simplicity, we assume than the two internal resistances are equal. In practice, that's unlikely, but it doesn't affect the conclusions.

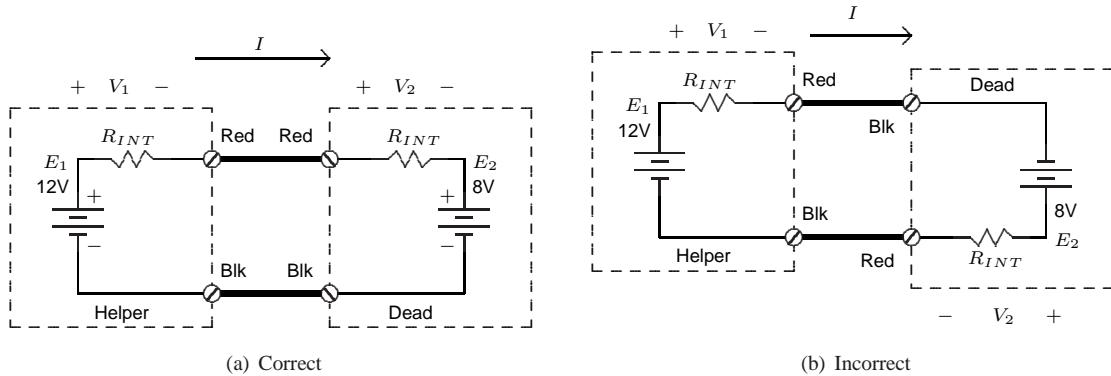


Figure 24: Jumpering a Car Battery

Notice that both resistors conduct the same current I (an implicit application of KCL). So we can apply Ohm's Law to the resistor voltage drops:

$$\begin{aligned} V_1 &= IR_{int} \\ V_2 &= IR_{int} \end{aligned} \quad (35)$$

Substitute for the voltages in equation 34 and solve for the current I :

$$+ E_1 - (IR_{int}) - (IR_{int}) - E2 = 0 \\ I = \frac{E_1 - E_2}{2R_{int}} \quad (36)$$

A similar analysis of the incorrect wiring of figure 24(b) gives:

$$I = \frac{E_1 + E_2}{2R_{int}} \quad (37)$$

Comparing equations 36 and 37, the current in the correct wiring is dependent on the *difference* in voltage between the two battery voltages. The current in the incorrect wiring is dependent on the *sum* of the two battery voltages.

In the first case, the current will be much smaller and decrease as the dead battery charges and its voltage increases. In the second case, there is a much larger voltage driving current around the circuit (a factor of 5 for the numbers shown) and it doesn't decrease.

Case 5: Identifying a Discharged Battery

When a battery discharges, we speak of it as being *run down*. And it is the case that the output voltage decreases. However, as anyone who has done it has noted, in a mix of fresh and discharged batteries, it's difficult to tell the difference with a voltmeter. A voltmeter appears to the battery as an open circuit, so it measures the open circuit voltage E_{oc} of the battery. That doesn't change much when a battery discharges. What does change is the internal resistance R_{int} , which increases significantly as the battery approaches a discharged condition. In that case the battery terminal voltage V_o decreases if there is any load current.

Consequently, the reliable way to test a battery is to place it under load (ie, connect it to a load resistor so the battery must supply current) and measure its terminal voltage. If this is much lower than the open-circuit voltage, then the battery is near end of charge.

2.19 Practical Current Source

As we saw in section 2.8, an ideal current source is a device that produces a constant output current regardless of the output voltage. The output characteristic is a vertical straight line. Real-world current sources have internal resistance, and this vertical graph develops a tilt upward to the left. The terminal current of any real-world current source decreases as the output voltage increases, figure 25.

This has exactly the same shape as figure 22, and in fact the practical voltage source and practical current source are equivalent, where

$$E_{oc} = I_{sc} \times R_{int} \quad (38)$$

Consequently, whether one considers a real-world source as primarily a voltage source or primarily a current source, depends on the size of the internal resistances compared to the load resistance. A *voltage source* has a low internal resistance compared to the load resistance. A *current source* has a high internal resistance compared to the load resistance.

2.20 Measuring Internal Resistance

In the previous section, we established that any practical source has some amount of resistance in series with the output terminal. In the case of a battery or DC power supply, this is known as the *internal resistance*, R_{int} . The upper-case symbol indicates that the internal resistance is relatively constant.

In the case of a simple battery, we established that the output voltage was given by:

$$V_o = E_{oc} - I_o \times R_{int} \quad (39)$$

where the algebraic quantities are

- V_o Output voltage of the source at some output current, volts
- E_{oc} Open Circuit output voltage of the source
- R_{int} Internal Resistance of the source, ohms.
- I_o Output current of the source.

If the source is short-circuited the output current will be I_{sc} and the output voltage V_o will be zero. Then we can write:

$$0 = E_{oc} - I_{sc}R_{int} \quad (40)$$

Consequently, as we indicated in section 2.17, if the open circuit voltage and short circuit current are both known, we can calculate the internal resistance:

$$R_{int} = \frac{E_{oc}}{I_{sc}} \quad (41)$$

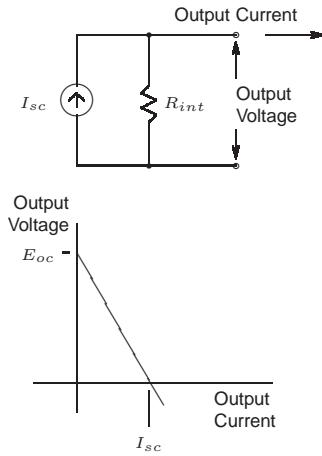


Figure 25: Practical Current Source

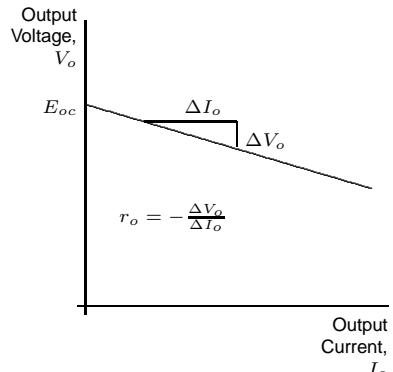


Figure 26: Output Resistance

The slope method

The open-circuit-voltage, short-circuit-current method is a simple way of measuring internal resistance. Unfortunately, it's not always possible to short circuit the source. In some cases (a 12 volt car battery, for example) the short circuit current is simply too large to be safe. In other cases, shorting the source causes it to stop operating.

As another example, the output terminal of certain electronic circuits such as the operational amplifier of section 12.1, may be modelled as a voltage source with internal resistance. Such sources cannot be short circuited, so the output resistance cannot be determined using the E_{oc}/I_{sc} method. The alternative is indicated in figure 26: to measure the slope of the output characteristic. Switching the name from internal resistance R_{int} to output resistance r_o , and taking the differential of equation 39, we have:

$$\Delta V_o = \Delta E_{oc} - \Delta I_o R_{int} \quad (42)$$

(we assume that R_{int} is a constant.) When the open-circuit voltage E_{oc} is a constant, $\Delta E_{oc} = 0$. Then:

$$\Delta V_o = -\Delta I_o R_{int} \quad (43)$$

and

$$R_{int} = -\frac{\Delta V_o}{\Delta I_o} \quad (44)$$

Consequently, *the internal resistance is the negative of the slope of the output voltage-current characteristic.*

The Test Voltage Method

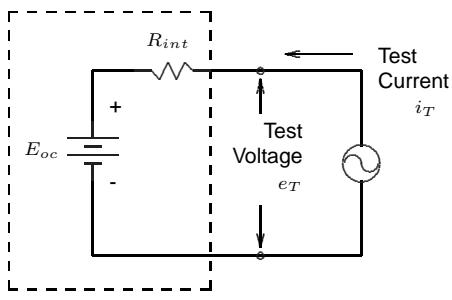


Figure 27: Test Generator Method

This method requires some advanced concepts – superposition and capacitive reactance. We mention it here for completeness, but you might come back to this at a later stage.

Consider the circuit shown in figure 27. An AC voltage generator e_T is attached to the output terminals of the source, and we measure the resultant AC current i_T . By the superposition theorem, (section 3.6) the open circuit voltage E_{oc} becomes a short circuit for AC, and the value of the internal resistance is simply the ratio of test voltage to test current.

$$R_{int} = \frac{e_T}{i_T} \quad (45)$$

As it is shown, the DC voltage source will drive DC current through the test generator. In fact, the DC current would be equal to the DC short-circuit current of the source. Were there not a solution to this problem, it wouldn't be a practical method.

However, this can be made to be a practical method if the AC test voltage is coupled into the supply terminals via a capacitor, figure 28. The capacitor will appear as an open circuit for the DC current, effectively blocking it from flowing through the test generator.

If the reactance of the capacitor is small compared to the internal resistance, then the capacitor will appear as a short circuit for the AC test signal, and the ratio of AC test voltage to AC test current will be an accurate measurement of the internal resistance.

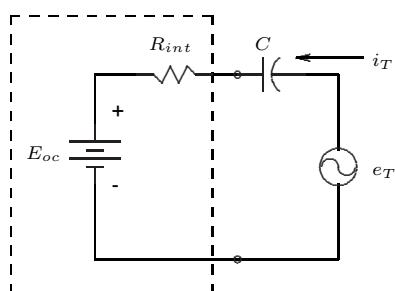
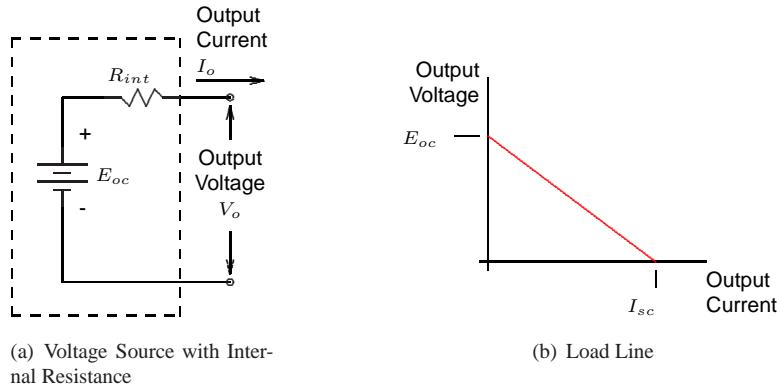
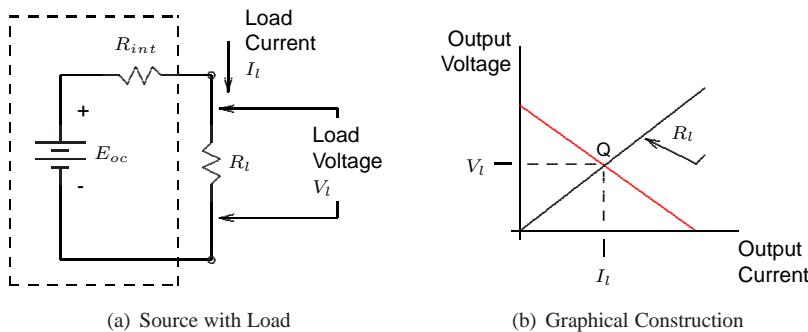


Figure 28: Blocking the DC current

2.21 The Load Line

Figure 29: The *Load Line*Figure 30: The *Load Line* with Load Resistance

Many circuits can be reduced to the circuit shown in figure 29(a). The output voltage-current characteristic for this electrical source is shown in figure 29(b).

Now consider that a resistor R_l is attached to the terminals of the source as shown in figure 30(a). The voltage-current characteristic of this resistor can be plotted on the same graph, with the result shown in figure 30(b). The intersection of the load-line with the resistance is the operating point or Q (for *Quiescent*) point of the circuit. The current through the resistance I_l and voltage across the resistance V_l can be read off the graph.

As is, this is not particularly earthshaking. The Q point can be found quite easily by an algebraic solution of two linear equations. However, this becomes a very useful technique when the load is a non-linear device such as the diode of figure 174 on page 247. Again, the Q point of the circuit can

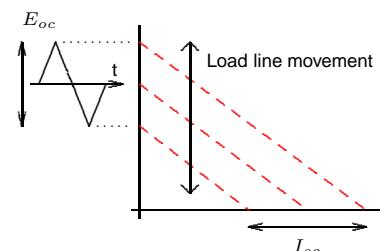


Figure 31: Load Line with AC Voltage

be found by superimposing the load characteristic on the load line, and determining the point of intersection.

Variations

Depending on the circuit and device, the axes of the load line and device characteristic may be interchanged, with current on the vertical axis and voltage horizontal. This does not change the basic idea.

The load line can be used with an AC source, as shown in figure 31. In that case, the value of E_{oc} changes over time which varies the position of the load line. The slope, which is dependent on the internal resistance, stays constant.

Each load line corresponds to one particular value of input voltage. If the corresponding Q points are plotted, then one can determine the shape of the voltage-time waveform across the load device.

2.22 Voltage and Current Source, Compared

Whether a source can be treated as a voltage source or a current source depends, not on the absolute value of the internal resistance, but on the relative value of the internal resistance and the load resistance. That is, the same source can behave as a voltage source or a current source, depending on the value of the load resistance. We illustrate this with two examples.

Generalized Source as Voltage Source

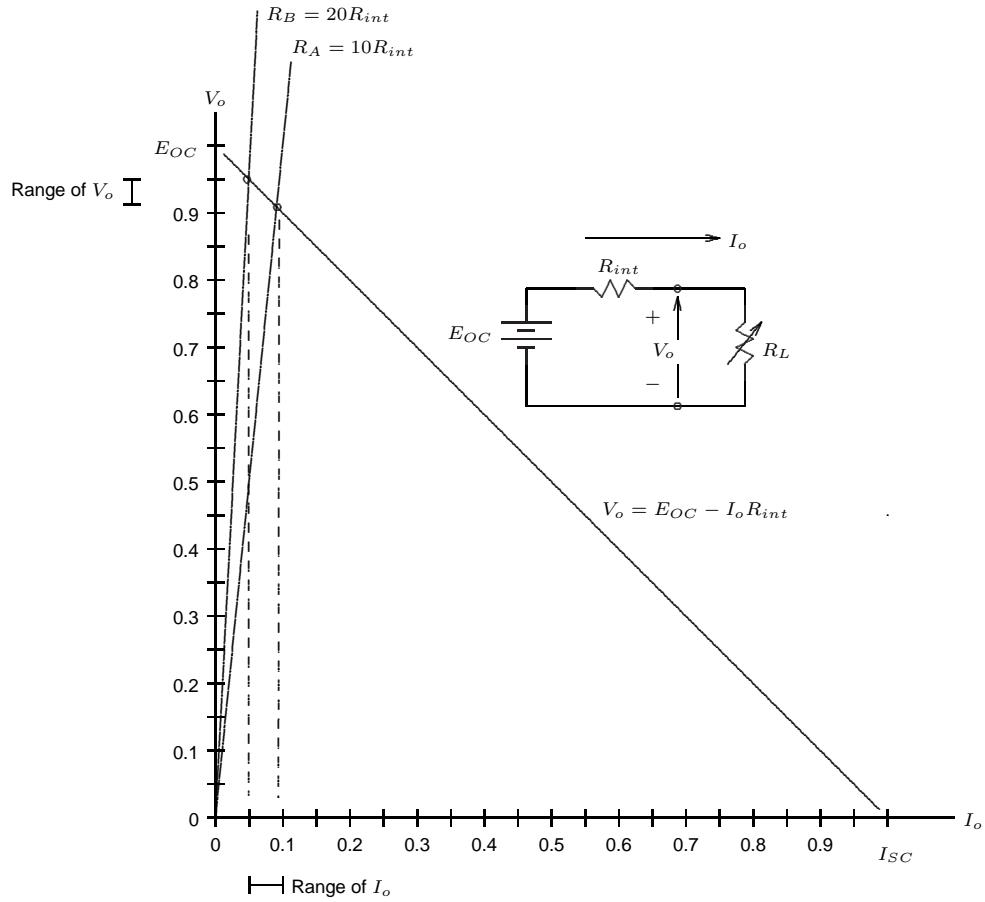


Figure 32: Source with $R_L \gg R_{int}$

The behaviour of some source is illustrated in figure 32. The source has an open circuit voltage E_{oc} , a short-circuit current I_{sc} and an internal resistance R_{int} . As usual, the output voltage is given by

$$V_o = E_{oc} - I_o R_{int} \quad (46)$$

The output characteristic appears as a line running between E_{oc} and I_{sc} .

If the characteristic of the load resistance R_L is plotted on this same chart, then the intersection of source output characteristic and the load resistance occurs at the terminal voltage V_o and terminal current I_o .

Figure 32 shows two load resistances, R_A and R_B . Load resistance R_A is equal to $10R_{int}$, R_B is equal to $20R_{int}$. Consequently, since the load resistance is much larger than the internal resistance, we can consider that the source behaves approximately as a voltage source for these loads. Let's see if that's the case.

For these load resistances, the output voltage V_o varies between $0.95 E_{oc}$ and about $0.91E_{oc}$, that is, approximately equal to E_{oc} , an approximately constant voltage source. The output current is much farther from constant, varying over approximately a 2:1 range for these two resistances. Put another way, a 100% variation in current is accompanied by a 5% variation in voltage.

Generalized Source as Current Source

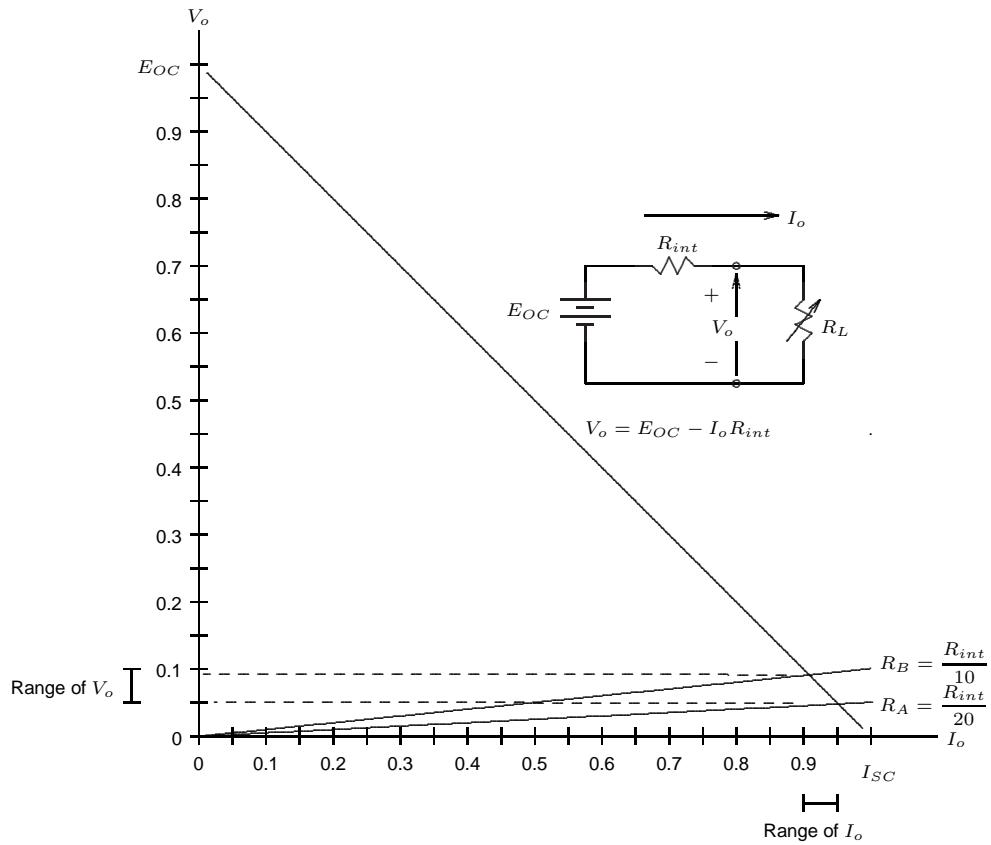


Figure 33: Source with $R_L \ll$ Internal Resistance

The same source approximates a current source when the load resistance is much lower than the internal resistance, ie, no larger than one-tenth the internal resistance. The situation is shown in figure 33. In this case, the output current varies over the range $0.95I_{sc}$ to $0.91I_{sc}$ while the terminal voltage varies over a 2:1 range. That is, a 100% variation in voltage is accompanied by a 5% variation in current.

2.23 Power Dissipation

The end result of many electrical systems is to output some form of work or energy. The rate at which this energy is produced is the system output power.

One watt of power is equivalent to one joule of energy being produced per second. Conversely, one joule of energy is one watt of power produced for a duration of one second (one *watt-second*). The flow of an electrical current through a voltage drop requires electrical power, in watts. Over a period of time, this electrical power represents some amount of energy in *watt-seconds* or *joules*³².

Figure 34 shows a very simple example. Voltage source E drives current I through resistor R . As a result, the resistor dissipates power P .

In this case, electrical energy is transformed into heat energy. However, with a different type of electrical device than a resistor the electrical energy could have been transformed into some other type of energy – light, sound, or mechanical motion, for example.

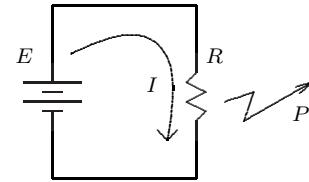


Figure 34: Power Dissipation

Development of the Power Equation by Analogy

In this section, we'll develop an equation for the power of an electrical current and voltage by a further extension of the hydraulic analogy previously discussed in section 2.1 on page 50.

The hydraulic analogy is a hydro-electric generating plant, where a quantity of water falls through a vertical height (under the influence of gravity) and the resultant power is used to operate generating turbines.

Somewhat more formally, recall that work or energy is accomplished by applying a force to some object and moving it over a distance.

$$W = F \times S \text{ joules} \quad (47)$$

where F is the force in newtons and S is the distance in metres.

Power is the *rate of doing work*, that is, work divided by time.

$$\begin{aligned} P &= \frac{W}{t} \\ &= \frac{F \times S}{t} \text{ watts} \end{aligned} \quad (48)$$

But S/t is velocity v , so

$$P = F \times v \text{ watts} \quad (49)$$

where v is the velocity in metres/second.

Now, let us think of the pressure in the pipe. (I'll call the pressure E to avoid bringing another P into the discussion.) The total force on water in a pipe is equal to the pressure times the pipe area.

$$F = E \times A \text{ Newtons} \quad (50)$$

where A is the area of the pipe in square metres. Substitute for F from equation 50 into equation 49, and we have:

³²Electrical energy is billed to a household in units of *kilowatt-hours*. We speak colloquially of the *power* bill, but strictly speaking it's an *energy* bill.

$$\begin{aligned}
 P &= F \times v \\
 &= E \times A \times v \text{ watts}
 \end{aligned} \tag{51}$$

But the product of area A and velocity v is the quantity of water flowing in cubic metres per second. Let's call that flow rate the *current* with symbol I . Then equation 51 becomes:

$$P = E \times I \text{ Watts} \tag{52}$$

So the power is equal to the pressure E times the flow rate I . By analogy with electricity, the pressure is voltage and the flow rate is the electrical current. In fact, the same equation applies. Electrical power is the product of voltage and current.

This is *far* from a rigorous proof of the power equation, but it does lend it some credence.

2.24 Battery Tester

To give some context to the ideas in preceding sections, we now design a simple but useful device: a *battery tester*.

At first blush, this can be a simple device. As a battery discharges, the terminal voltage drops, correct? Well, yes and no. Depending on the chemistry of the battery, the open circuit voltage may not change very much between a fully charged battery and one that is discharged. The internal resistance is often a better indicator of battery state: it increases as the battery discharges.

In this design exercise, we need to test the state of a 9 volt battery. When new and fully charged, the internal resistance is 1.7Ω . When discharged, the internal resistance is equal to or greater than 170Ω . We'll assume that the open-circuit voltage remains at or near 9 volts.

The internal resistance may be determined by measuring the open-circuit voltage E_{oc} and then the short-circuit current I_{sc} . Then

$$R_{int} = \frac{E_{oc}}{I_{sc}} \tag{53}$$

Unfortunately, applying a short circuit to a battery is a bad idea. Even a few seconds of short circuit will discharge the battery significantly. Many batteries are damaged³³ by discharging into a short circuit. However, it is possible to connect a resistive load to the battery. The terminal voltage will decrease when the load is connected, and the amount of that decrease is indicative of the internal resistance.

One possible circuit for the battery tester is shown in figure 35. When the pushbutton switch is open, the voltmeter reads the open-circuit voltage of the battery. When the pushbutton switch is closed, it loads the battery with an output current. For an internal resistance of 1.7Ω , the load resistor R_{load} is much larger than the internal resistance and the indicated voltage will drop only slightly. For an internal resistance of 170Ω , the load resistor R_{load} is much *smaller* than the internal resistance, and the indicated voltage will collapse to near zero. The load resistor must be sized to handle the worst case power dissipation, that is, when the internal resistance is low and the operator – fixated on the results, perhaps – holds their finger on the *TEST* button indefinitely.

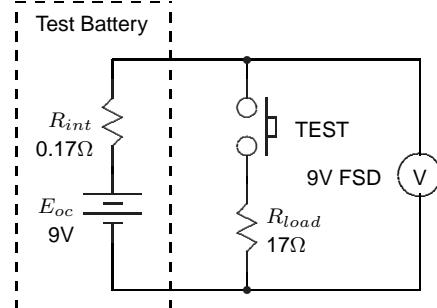


Figure 35: Battery Tester

³³Some show their distress by catching fire or exploding.

2.25 The Capacitor

A *capacitor*, symbol in figure 36 is an electronic device for storing energy. The circuit of figure 37 provides an introductory demonstration³⁴.

The LEDs in this circuit are *light emitting diodes*. They are explained in section 27.2 on page 773. For the purpose of this demonstration, you need only know that a current passing through the LED in the direction of the arrow causes it to illuminate. The LED is preferable to an incandescent lamp because the LED can respond much more quickly to a short duration pulse of current.

1. The demonstration starts with both switches open.
2. If we measure the voltage across the terminals of the capacitor, it is zero.
3. We close switch S_1 . The LED_1 flashes briefly, indicating that current passed through it.
4. If we open and close switch S_1 again, nothing happens.
5. If we now measure the voltage across the capacitor, we find it has charged to the same voltage as the battery V .
6. Ensure switch S_1 is open. Close switch S_2 . LED_2 flashes briefly, indicating that current passed through it.
7. If we open and close switch S_2 again, nothing happens.
8. If we now measure the voltage across the capacitor, we find it has discharged back to zero volts.

This demonstration shows charge migrating from the voltage source into the capacitor. The initial capacitor voltage is zero. When switch S_1 is closed, current flows into the capacitor. The capacitor voltage rises until it is equal to the voltage source. At that point, the charging current ceases. It can be shown that the stored charge is proportional to the applied voltage, with a constant of proportionality called *capacitance*.

$$Q = CV \quad (54)$$

where the algebraic quantities are

Q charge stored in the capacitor, in coulombs

C the *capacitance*, in farads

V the voltage between the capacitor plates, in volts

A one farad capacitor is very large (although not entirely out of the question) and most capacitors are measured in microfarads (μF , 10^{-6}F) or smaller units.

When we close switch S_2 , the capacitor discharges through the second LED.

In both cases, charging and discharging, there is very little resistance in the circuit: just the internal resistance of the voltage source and the wiring resistance. As a result, the charging and discharging voltages appear across

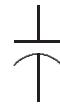


Figure 36: Capacitor Symbol

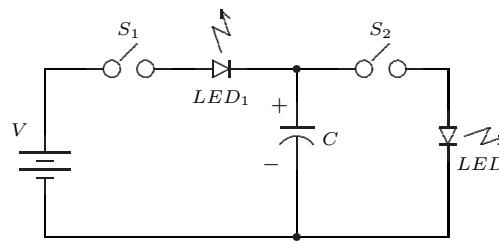


Figure 37: Capacitor Operation

³⁴If you want to try this yourself, use a *small* 9 or 12 volt battery for V and a thousand microfarad capacitor for C . Ensure the capacitor polarity correct (figure 37) or the capacitor may explode.

small resistances and the currents can be quite large. Because the currents are large, they need only be of short duration to move a quantity of charge.

This is one application of capacitors: to store charge and then discharge it in a burst of current.

The Hydraulic Analog of the Capacitor

In the hydraulic analogy, a capacitor is analogous to a water storage tank. The stored electric energy is the amount of water in the tank. The pressure at the bottom of the tank, which is proportional to the height of the water, is analogous to the capacitor voltage. The capacitance is analogous to the diameter of the tank.

Then the stored energy in a capacitor can be increased by increasing the size of the capacitor or increasing the charging voltage, just as the amount of water in a tank can be increased by increasing the size of the tank or the height of the water in the tank.

2.25.1 Capacitor Construction

The original capacitor was the *Leyden Jar* [38]. A Leyden Jar consists of a glass bottle with separate metallic layers on the inside and outside of the jar. Connections are made to the two metal layers. This capacitor has modest capacitance but can store extremely high voltages, and it is still used for charge storage by electrostatic generators such as the Wimshurst Machine [39].

In general, an electronic capacitor consists of two parallel metallic conductors separated by an insulating material, which is known as a *dielectric*.

The capacitance is given by

$$C = \frac{\epsilon A}{s} \quad (55)$$

where the algebraic quantities are

C the *capacitance*, in farads

ϵ the *dielectric constant* of the insulating material

A area of the conductors

s spacing between the conductors

Capacitance can be made to vary by changing any of these quantities. For example, in a mechanically adjustable variable capacitor, the overlapping area of the plates is changed to adjust the capacitance.

Capacitors are constructed with a wide variety of dielectric materials, and these determine both the size and properties of the capacitor.

The capacitor may be regarded as the analog of a mechanical spring, in which the applied voltage is analogous to the applied force on the spring and the stored charge is the resultant compression distance of the spring.

Application: Transducer

Piezoelectric materials have the property that they generate a charge on their surface when flexed or compressed. This makes them useful as a force-to-charge transducer. In electronic terms, a piezoelectric transducer can be regarded as a capacitor where the charge varies according to stress. Then if the capacitance is fixed, the output voltage will vary. From 54,

$$\Delta V = \frac{\Delta Q}{C} \quad (56)$$

The *electret* microphone uses a different approach. A fixed charge is imprisoned into the dielectric material of a capacitor, and one of the plates caused to vibrate by sound waves. The variation of the plate spacing causes the capacitance to vary, which then causes the output voltage to change.

$$\Delta V = \frac{Q}{\Delta C} \quad (57)$$

Both piezoelectric and electret devices have very high internal resistances – they can't supply significant current. So they must be buffered by an impedance converter such as a JFET device (section 8.2, page 268).

2.25.2 Energy Storage in a Capacitor

Picture pumping water into a tank, via an entry point at the bottom of the tank. As the height of the water rises, the back pressure at the entry point increases, requiring more work to push water into the tank. The water in the tank is a form of stored energy. Similarly, it requires more work to push charge into a capacitor as its voltage increases.

In figure 38(a), a current source charges a capacitor³⁵. According to equation 54 (page 85) the voltage increases linearly with the stored charge, as plotted in figure 38(b)³⁶.

$$\frac{V_c}{Q_c} = \frac{1}{C} \quad (58)$$

In general, the work done by moving charge Q through voltage V is given by equation 2, repeated here:

$$W = QV \quad (59)$$

where V is the voltage difference between two points, Q is the charge being moved in coulombs, and W is the work in joules.

In the case of the charging capacitor, the *average* voltage affecting the movement of charge is half the final capacitor voltage V_c . Substitute for V in equation 59:

$$W = Q_c \times \frac{1}{2}V_c \quad (60)$$

Now get rid of the charge Q_c , substituting CV_c from equation 58:

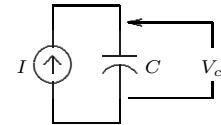
$$W = \frac{1}{2}CV_c^2 \quad (61)$$

The energy is stored in the *electric field* of the capacitor. (This is in contrast with an inductor, which stores energy in a magnetic field.)

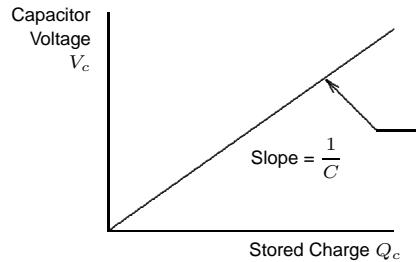
2.25.3 Energy Storage: Battery and Capacitor

³⁵An alternative development of the equation for stored energy in a capacitor is given in section 2.26.2 on page 90.

³⁶If the capacitor is charged by a *constant* current source, it turns out that the capacitor charges at a linear rate: both the voltage and the stored charge increase as a linear function of time. However, this equation development for stored energy is valid regardless of the rate at which the capacitor is charged. The voltage and charge are a linear function of each other as shown in figure 38(b), regardless of the rate of charge. The capacitor could equally well be charged from a voltage source via a resistor.



(a) Circuit



(b) Voltage vs Charge

Figure 38: Charging a Capacitor

Both a rechargeable battery and a capacitor are capable of storing and producing electrical energy. What's the difference?

A battery stores and produces its energy by means of an electrochemical reaction. This appears as internal resistance, which is heated by a charging or discharging current. As a result, there is a limit to the maximum charge or discharge current.

In contrast, a capacitor can be charged and discharged *very* rapidly without damage. That makes the capacitor ideal for situations where a large pulse of current is required – to fire a strobe lamp or in spot welding two metals together.

On the other hand, the energy density is much higher for a battery, at 128 watt-hours per kilogram for a lithium-ion battery³⁷ versus 30 Wh/kg for the so-called *ultracapacitor* [40], [41].

Another source [42] shows the values in figure 39 for energy storage in megajoules vs weight in kilograms and size in litres. On the basis of these figures a lithium-ion battery can store 30 times the energy of an ultracapacitor per unit weight and 28 times the energy per unit volume.

2.25.4 Common Misconception: Kirchhoff's Current Law and The Capacitor

The equation $Q = CV$ (equation 54 above) states that the charge Q stored in a capacitor is proportional to the capacitance C and the charging voltage V . This implies that charge flows into a capacitor as water flows into a tank. It turns out that this is incorrect: Kirchhoff's Current Law is always observed for a capacitor. During charging and discharging, the current flowing into the positive terminal is equal to the current flowing out of the negative terminal.

If one terminal of the capacitor is connected to ground, then we see the other terminal changing voltage as charge flows into or out of the other terminal, and it's easy to picture charge accumulating in the capacitor. However, charge is flowing out of the grounded terminal at the same time.

How can equation 54 and KCL both be true at the same time? What is happening is that a positive charge is accumulating on one plate of the capacitor and *an equal and opposite charge on the other plate of the capacitor*. So equation 54 is correct *when the charge referred to is the charge on one plate of the capacitor*.

Overall, the positive and negative charges balance out, and there is no net charge in the capacitor.

If the overall charge in a capacitor does not change, then what is being stored? Energy.

It's correct to speak of a capacitor as *a charge storage device*, as long as you are referring to either one of the plates of the capacitor. However, the entire capacitor is storing energy in the electric field between the two plates, much as energy is stored in a spring by compressing the spring with a force.

2.26 The Capacitor: Time Domain

In the previous section, we studied the behaviour of the capacitor under more-or-less static conditions. Here we study its dynamic behaviour – with time-varying voltages and currents.

When equation 54 is differentiated with respect to time, we obtain

$$\frac{dQ_c}{dt} = C \frac{dV_c}{dt} \quad (62)$$

Since dQ_c/dt is equal to current i_c ,

$$i_c = C \frac{dV_c}{dt} \quad (63)$$

³⁷At this writing, of various battery technologies lithium-ion has the greatest energy density.

This is a very important relationship and worth committing to memory or adding to a cheat sheet. When the current is a constant quantity I_c , then the rate of change of voltage with time is constant, and

$$I_c = C \frac{\Delta V_c}{\Delta t} \quad (64)$$

This equation is relevant to the implementation of a *ramp generator*, a circuit that generates a steadily changing (increasing or decreasing) voltage over a period of time. A constant current flowing into a capacitor will generate a voltage that shows a constant rate of change (see section 2.19).

Integral Form

It is also possible to rewrite equation 63 as

$$dV_c = \frac{1}{C} i_c dt \quad (65)$$

Integrating both sides, we have

$$V_c = \frac{1}{C} \int i_c dt \quad (66)$$

This is another useful description of capacitor behaviour. It is sometimes encapsulated in the rule of thumb:

The average current into a capacitor must be zero.

Were this not so, over a long period of time the voltage on the capacitor would integrate toward infinity. (As a practical matter, the capacitor voltage rating would eventually be exceeded and the capacitor destroyed.) Consequently, there must always be some mechanism to ensure that the capacitor charging current is balanced by an equal discharging current.

Application

An example application of equation 66 is shown in figure 40.

In this application, a current generator $i(t)$ pumps rectangular pulses of current into a capacitor. At the end of the interval, the voltage on the capacitor is proportional to the time-integral (the area under the curve) of the current waveform. During a current pulse, the capacitor voltage ramps upward. During the quiescent interval between pulses, the capacitor voltage stays the same.

If the current pulses are proportional to some variable, then the capacitor voltage is proportional to the integral of that variable.

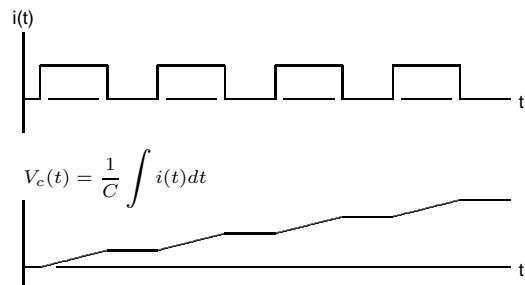


Figure 40: Capacitor Integrator

Rule of Thumb

The rate of change of capacitor voltage is proportional to the capacitor current (equation 63 above). Since capacitor current is usually finite, we can say that

The voltage across a capacitor cannot change instantaneously.

Example

An example of this rule of thumb is shown in figure 41.

In this circuit, the switch starts in position A. The capacitor discharges through the switch-resistor loop until its voltage is zero.

Then the switch is moved to position B. This causes a step input voltage E volts to be applied to the left plate of the capacitor. Since *the voltage across a capacitor cannot change instantaneously*, the capacitor voltage must still be zero volts immediately after the step. Then to satisfy Kirchhoff's voltage law around the E-C-R loop, the right plate of the capacitor must step upward by exactly the same E volts as the left plate.

At this point the resistor has E volts across it and current flows clockwise around the capacitor-resistor loop. This current charges the capacitor and after some time the current runs down to zero, as shown in the waveform for $e_o(t)$.

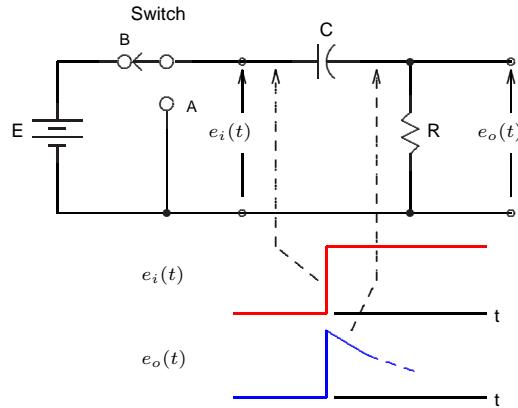


Figure 41: Capacitor Behaviour

Question

Now the capacitor is charged so that the left plate is at E volts and the right plate is at zero volts. What happens when the switch is moved back to position A?

2.26.1 Defective Circuit

Consider the circuit in figure 42. Assume the capacitor is initially uncharged. When the switch is closed, the voltage source will dump charge into the capacitor until the capacitor voltage V_c is equal to the supply voltage E .

However, there is a problem: the voltage source has no internal resistance, so the capacitor voltage will step up to the supply voltage in zero time. The rate of change dV_c/dt is equal to infinity, so the switch current is infinite. Well, perhaps not quite infinite, but very large, and for a very short interval of time.

The net effect is that the switch contacts may be welded together by this *inrush* current into the capacitor. When this situation arises, there must always be a small charging resistance to limit the current.

2.26.2 Energy Storage in a Capacitor: Reprise

Now that we have differential equation 63 relating voltage and current in a capacitor, we can develop the energy storage equation by a different method.

Suppose some current $i(t)$ flows into a capacitor, creating a voltage $v(t)$ on the capacitor. The instantaneous power is the product of the voltage and current. The total work (or energy) is the time-integral of power expended over the same interval.

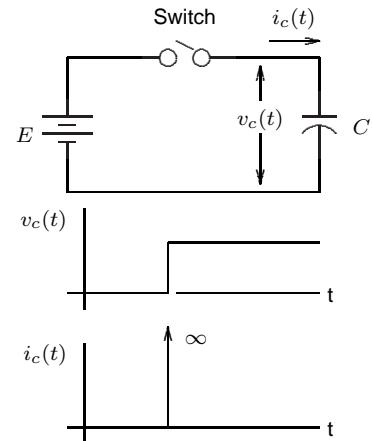


Figure 42: Defective Circuit

$$W = \int v(t)i(t)dt \quad (67)$$

where W is the stored energy in joules.

Substitute for $i(t)$ from equation 63, and we have

$$\begin{aligned} W &= \int v(t) \left\{ C \frac{dv}{dt} \right\} dt \\ &= C \int v(t) dv \\ &= \frac{1}{2} CV_c^2 \end{aligned} \quad (68)$$

where V_c is the voltage on the capacitor at the end of the interval.

2.26.3 Application: Ramp Generator

To create some confidence that the somewhat alien concept of *current generator* has some real-world validity, consider the ramp generator circuit of figure 43.

This circuit may be implemented with a half-dozen discrete components, and the current generator itself can be provided by a single junction transistor.

When the switch is closed, the voltage across the capacitor is zero. When the switch is opened, the current source begins to charge up the capacitor at a constant rate of I/C volts per second. For an ideal current source, this will go on forever. For a practical current source, when the terminal voltage reaches the compliance limit of the current source, the output current will cease or trend away from the constant rate.

Non-Ideal Ramp Generator

When the ramp generator circuit of figure 43 is implemented with a real-world current generator, the effect is to cause the linear ramp to deviate slightly downwards as the capacitor charges.

As the capacitor charges, the voltage across it increases. This voltage drives increasing current through R_{int} , so that some of the current that should be charging the capacitor is lost through R_{int} .

This effect is usually undesired. Consequently, in this and other applications, the internal resistance of a current source should be as large as possible. When the current source drives a load resistance, the internal resistance should be much *larger* than the load resistance.

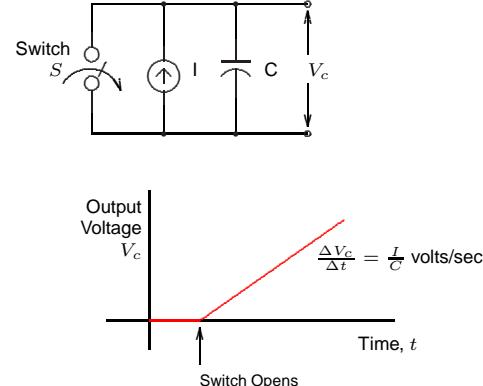


Figure 43: Ramp Generator

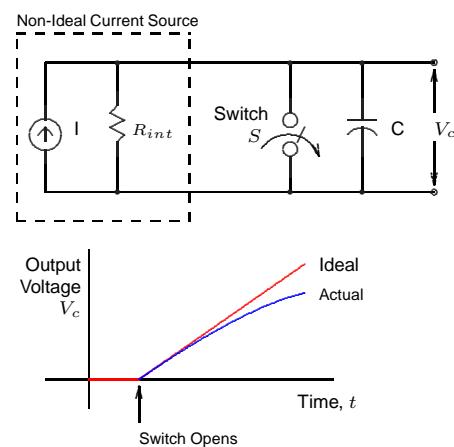


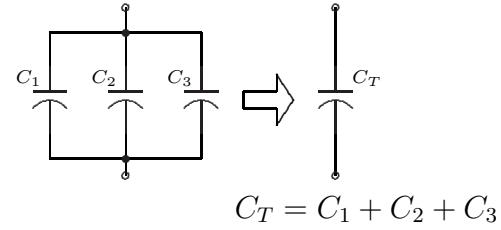
Figure 44: Practical Ramp Generator

2.27 Capacitances in Series and Parallel

Just as resistors in series and parallel can be combined to make an equivalent resistance, reactances can be combined as well. Capacitors in parallel is the easiest to understand, so we'll look at that first.

Capacitances in Parallel

When charged with a voltage V , a capacitor of C farads stores charge Q coulombs. (equation 54, page 54).



$$Q = C \times V \quad \text{Figure 45: Capacitors in Parallel} \quad (69)$$

Suppose we have three capacitors, each charged to a voltage V . Then the individual charges will be:

$$Q_1 = C_1V, Q_2 = C_2V, Q_3 = C_3V,$$

We connect those three capacitors together as shown in figure 45. This is a parallel connection because each capacitor must have the same voltage across it. No current flows, since there is no voltage difference between the capacitors. Therefore there is no movement of charge, and the total charge remains the same. The total charge Q_T is then equal to the sum of the individual charges.

$$\begin{aligned} Q_T &= Q_1 + Q_2 + Q_3 \\ &= C_1V + C_2V + C_3V \\ &= (C_1 + C_2 + C_3)V \\ &= C_TV \end{aligned} \quad (70)$$

where C_T is the total capacitance, equal to the sum of the individual capacitances. That is, capacitances in parallel add together.

This makes sense because the capacitance is proportional to the area of the plates of the capacitor. When the capacitors are in parallel each capacitor adds its plate area to the total plate area.

Application: Capacitors in Parallel

Capacitors are occasionally connected in parallel to make one large capacitor. However, a more common arrangement is to parallel a large-capacitance electrolytic capacitor with a relatively low capacitance ceramic capacitor. The total capacitance is not materially changed. However, as detailed in section 24.2, a large electrolytic capacitor effectively contains an internal inductance, which reduces its effectiveness at high frequencies. The smaller capacitor contains a smaller internal inductance, so it is more effective at high frequencies.

Capacitors in Series

Consider that we put the same amount of charge Q coulombs on three different capacitors, C_1 , C_2 and C_3 . Then according to equation 69, their individual voltages will be:

$$Q = C_1V_1, Q = C_2V_2, Q = C_3V_3,$$

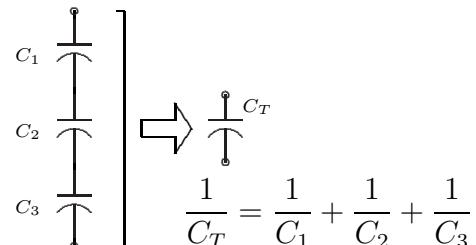


Figure 46: Capacitors in Series

Now we connect the capacitors in series, as shown in figure 46. The voltages on the capacitors are in series, so they add together. The total voltage is given by:

$$\begin{aligned}
 V_T &= V_1 + V_2 + V_3 \\
 &= \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} \\
 &= Q \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \right) \\
 &= \frac{Q}{C_T}
 \end{aligned} \tag{71}$$

where C_T is the equivalent total capacitance.

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \tag{72}$$

That is, series capacitors add in a similar fashion to parallel resistors: *the reciprocal of the total capacitance is equal to the sum of the reciprocals of the individual capacitances*.

This also makes sense, if you think of equation 55 (page 86) that relates capacitance and the spacing between the plates. For simplicity, let's assume the three capacitors have the same dielectric constant ϵ and the same plate area A with different plate spacings S . Then:

$$C_1 = \frac{\epsilon A}{S_1}, C_2 = \frac{\epsilon A}{S_2}, C_3 = \frac{\epsilon A}{S_3}$$

Putting three capacitors in series has the effect of increasing the spacing between the top plate of the upper capacitor and the bottom plate of the lower capacitor. Now it is equal to the sum of the individual plate spacings. Then the total capacitance C_T is given by:

$$C_T = \frac{\epsilon A}{S_1 + S_2 + S_3} \tag{73}$$

Invert this equation:

$$\begin{aligned}
 \frac{1}{C_T} &= \frac{S_1 + S_2 + S_3}{\epsilon A} \\
 &= \frac{S_1}{\epsilon A} + \frac{S_2}{\epsilon A} + \frac{S_3}{\epsilon A} \\
 &= \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}
 \end{aligned} \tag{74}$$

which is the same result as before.

Example

What is the combined capacitance of a $0.5\mu\text{F}$ and $0.2\mu\text{F}$ capacitor when they are connected in series?

Solution

Using equation 74, we have

$$\begin{aligned}
 \frac{1}{C_T} &= \frac{1}{C_1} + \frac{1}{C_2} \\
 &= \frac{1}{0.5} + \frac{1}{0.2} \\
 &= \frac{1}{0.14}
 \end{aligned}$$

Then

$$C_T = 0.14\mu\text{F}$$

Notice that the series combination of two capacitors is less than either of the individual capacitors.

Application: High Voltage Capacitor

In a high-voltage circuit, the required voltage rating may exceed that of any individual capacitor. The internal construction of a capacitor is such that each capacitor is effectively paralleled by a very large resistance which is known as the *leakage resistance* of the capacitor.

When N capacitors are placed in series, the leakage resistances form a voltage divider which distributes the applied voltage across the capacitors. If the leakage resistances are equal, then each capacitor sees $1/N$ of the total voltage. If the leakage resistances are not equal the voltage distribution will be uneven. To ensure an even distribution of voltage, the capacitive divider is usually paralleled by an external voltage divider of high-resistance resistors, figure 47.

For example, if the leakage resistance is in the region of $100\text{M}\Omega$, then the divider resistances might be chosen to be $10\text{M}\Omega$ each.

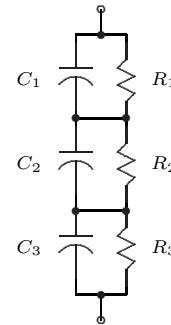


Figure 47: High Voltage Capacitor Array

2.28 The Inductor

The existence of the magnetic field may be seen by repeating the work of Hans Christian Ørsted, its discoverer. Place a magnetic compass below a wire conductor. With the current off, align the wire to point in the direction of the compass needle. Turn the current on, and the needle swings to point a right angles to the conductor.

When a current flows, it creates a magnetic field around the conductor. The direction of the field is indicated by the *right hand rule*. Grasp the conductor³⁸ by the right hand, with the thumb pointing in the direction of the current³⁹. Then the fingers point in the direction of the magnetic field.

The magnetic field contains stored energy. It absorbs electrical energy when it is increasing and produces electrical energy when it is collapsing. The energy is concentrated by winding the wire into a coil, and further concentrated by winding this coil on a magnetic permeable material, such as iron⁴⁰. This device is known as an *inductor*, symbol in figure 48.



Figure 48: Inductor Symbol

2.28.1 Voltage Induction

Michael Faraday discovered the law of magnetic induction: a time-varying magnetic field (flux) passing through a conductor creates a voltage between the two ends of the conductor. The effect of the flux variation can be magnified by placing a number of turns of wire in the magnetic field.

This varying field can be created by either of two methods:

1. physically move a magnetic field (from a magnet, for example) in the vicinity of the wire. We'll call this *moving magnet induction*⁴¹.
2. vary the current through the conductor: so-called *self-induction*.

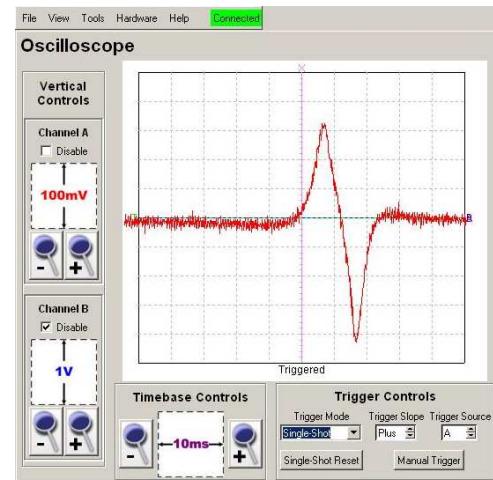


Figure 49: Magnetic Induction Waveform, Stationary Coil, Moving Magnet. Syscomp CircuitGear oscilloscope, 100mV/division vertical, 10mSec/division horizontal.

2.28.2 Moving Magnet Induction

The first case is illustrated in figure 49. A magnet was swept past a stationary coil and the resultant coil voltage recorded as an oscilloscope trace. The changing magnetic field creates a pulse of voltage in the coil, which displays on the oscilloscope screen.

³⁸Should there be any question of the safety of doing this, it should be done as a thought experiment only.

³⁹There, another reason to use conventional current flow rather than electron flow. For electron flow, this would become the *left hand rule*.

⁴⁰The wire conductor must be insulated from adjacent conductors by some sort of insulating coating, a layer of enamel or a plastic jacket.

⁴¹The same effect is obtained when a coil of wire is moved past a stationary magnet.

2.28.3 Faraday's Law and Lenz's Law

Faraday discovered that the voltage generated by a changing magnetic field in a coil of wire is given by:

$$e(t) = N \frac{d\phi}{dt} \quad (75)$$

where e is the so-called *induced* voltage, N is the number of turns on the coil, and $d\phi/dt$ is the rate of change of the magnetic field flux in webers per second.

This equation describes the magnitude of the induced voltage⁴².

Figure 49 on page 95 illustrates equation 75. A positive voltage is induced in a coil as the magnet approaches the coil and the flux is increasing. The polarity reverses as the magnet recedes from the coil and the flux is decreasing. If we substitute a coil with more turns and move the magnet at the same speed, or if move the magnet more quickly, the magnitude of the voltage pulse increases.

Notice that the total area under the two pulses is equal to zero. (We count the area under the positive pulse as a positive and the area under the negative pulse as a negative.) This isn't a coincidence.

Write equation 75 as:

$$d\phi = \frac{1}{N} e(t) dt \quad (76)$$

Integrating:

$$\phi = \frac{1}{N} \int e(t) dt \quad (77)$$

In words, the resultant flux is proportional to the integral of the generated voltage in volt-seconds. In figure 49 the final flux is zero so the integral of the voltage-time curve (the area under the voltage-time curve) must be zero. That is, the positive area must equal the negative area.

Self Induction

A demonstration of self induction is shown in figure 50. An iron-core inductor connects to a battery via a flying lead (the probe). Connect the probe briefly to the battery. Current builds up in the inductor, creating a magnetic field. Disconnect the probe. As the probe is withdrawn, an electrical spark occurs across the air gap⁴³.

When the probe breaks the circuit, the current stops. The stored magnetic field then collapses, rapidly. In doing so, it creates a very large voltage across the air gap, initiating a spark.

Faraday's Law predicts the magnitude of the voltage. For the polarity, we will need Lenz's Law: *the polarity of the induced voltage is such as to maintain the flow of current in the original direction*⁴⁴.

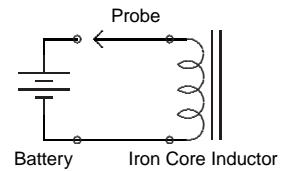


Figure 50: Self Induction

⁴² The usual statement of Faraday's law precedes the right-hand side of equation 75 with a minus sign. Then, if the field is collapsing, $d\phi/dt$ is negative. The preceding minus sign results in a positive value for the voltage. This gives the correct magnitude *and* sign for the induced voltage. However, rather than rely on the sign of the generated voltage it is often easier to determine the polarity by Lenz's law, given below.

⁴³This was confirmed to work with a 6 volt battery as the supply and the primary of a small power transformer as the inductor. It is best not to use a regulated power supply for this demonstration because the spark voltage may damage the internal circuitry of the supply.

⁴⁴Lenz's Law is usually stated somewhat differently: *the polarity of the induced voltage is such as to oppose the collapse of the magnetic field*. In practice, both formulations generate the same results. The reader can choose whichever is easier to understand.

Figure 51 shows the circuit waveforms. (The gap is replaced by a switch.)

When the switch is closed, during time interval t_1 to t_2 , there is a fixed voltage across the inductor. Then magnetic flux builds up in the coil at a constant rate, shown in the upper of the two graphs of figure 51. The equation of flux vs time is given by equation 75:

$$\frac{d\phi}{dt} = \frac{E}{N} \text{ webers per second} \quad (78)$$

Voltage E is a constant, number of turns in the inductor is a constant N , so the rate of increase of magnetic flux is a constant.⁴⁵

At time t_2 the switch opens. The current drops to zero and the magnetic field collapses, abruptly. Consequently the value of $d\phi/dt$ becomes $-\infty$. According to equation 75, this creates a very large voltage at the point of the open circuit, which initiates a spark.

The polarity of the induced voltage is such that it tries to continue the original current. In this case, the original current was clockwise around the circuit, through the battery and inductor. Consequently, the bottom end of the inductor becomes positive with respect to the top end. If the bottom end of the inductor is the reference voltage (zero volts), then the top end becomes negative at the probe.

This can be difficult to get right, so here are three additional ideas that may help.

1. The inductor appears as a *constant current* device. The current through it tries to remain constant, while the voltage across the inductor adjusts to make that so. Consequently, the voltage across an inductor can and often does change very suddenly.
2. When a gap appears in the circuit, think of it as a very large resistor. The inductor continues to produce the initial current. This current sets up a very large voltage across the resistor-gap. The current is flowing clockwise around the loop in figure 50, so the left end of the resistor is positive, the right end is negative. Applying KVL to the circuit, the top end of the inductor must be at a very negative voltage.
3. The voltage across the inductor abruptly switches polarity. When the inductor behaves as a *load*, it is absorbing energy. Then current flows into the positive terminal of the inductor. When a gap appears in the circuit, the inductor turns into a *source*, in which case current flows out of the positive terminal of the inductor. Since the inductor current does not change direction, then the change from load to source must be accompanied by a change in inductor polarity. That is, the bottom terminal must change from negative to positive polarity.

When a switch is used to open an inductive circuit, the spark created by the induced voltage will eventually corrode and damage the switch contacts. If the switch is a semiconductor device such as a transistor, it may be destroyed by this transient.

⁴⁵This can't go on forever. At some point, the magnetic core *saturates* and the flux cannot increase. We assume here that the flux level does not reach saturation.

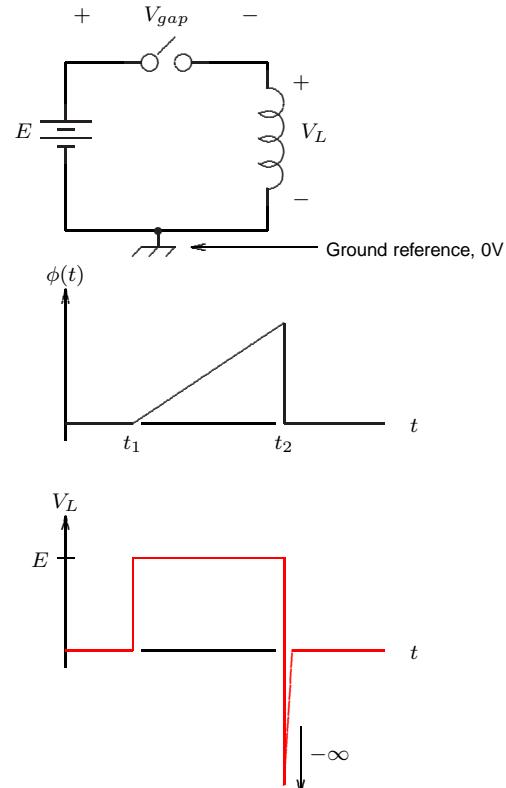


Figure 51: Self Induction Waveforms

There are several possible solutions to this problem, all of which essentially provide a path for the inductor current so that it is allowed to die away in a gradual fashion. The circuit that prevents the inductive formation of a high voltage is known as a *snubber*.

Snubber Network and the Kettering Ignition System

An example of a snubber circuit is shown in figure 52. This circuit is of historical interest – it shows the basic circuitry of the *Kettering ignition system* used in automobiles until the late 60's [43].

The inductor is an *ignition coil*, which consists of two windings, a *primary winding* and a *secondary winding*. The primary winding has relatively few turns, the secondary has many. When the switch closes, current ramps upward in the primary winding, creating a magnetic flux in the core. The switch then opens, stopping the primary current. The magnetic flux collapses, generating a large voltage and a miniature lightning-bolt at the spark gap. This ignites gasoline vapour in the engine cylinder.

The collapsing flux also generates a high voltage across the primary winding of the ignition coil, and that will create a spark across the switch contacts unless Something Is Done. That something is usually a capacitor (known at the time as a *condenser*). Since voltage cannot change instantaneously across a capacitor, the primary voltage drives current into the capacitor, and this restricts the capacitor (and switch) voltage to a safe level. In effect, the stored energy in the magnetic field of the primary is transferred to the stored electric field of the capacitor⁴⁶.

When the condenser became defective, the switch (known in this context as the *points*) would begin to spark (arc) with resultant damage to the switch contact surfaces.

With a single capacitor snubber as shown in this circuit, there is always the possibility that the capacitor and inductor will set up a resonant circuit, and the stored energy will slosh back and forth between the inductor and capacitor, creating a lightly-damped sine wave. If that is a problem, a resistor is often added in series with the capacitor to dissipate the energy as quickly as possible.

Why are there two windings, and not just one? After all, we've established that one winding can create a spark.

Having two windings allows them to be optimized separately. The primary winding has relatively few turns and therefore low resistance, allowing the battery to drive significant current through it. (This was especially important when car batteries were only 6 volts.) This significant current generates a substantial magnetic field. The resistance of the secondary is less important, so it has many turns to create a large induced voltage from the collapsing field.

2.28.4 Inductance

Faraday's law shows that an induced voltage is proportional to the rate of change of the magnetic flux, as we saw in equation 75, repeated here.

$$|e(t)| = N \frac{d\phi}{dt} \quad (79)$$

⁴⁶Section 6.10 shows an alternative snubber technique, using a semiconductor diode. Diodes were not available at the time the Kettering ignition system was invented.

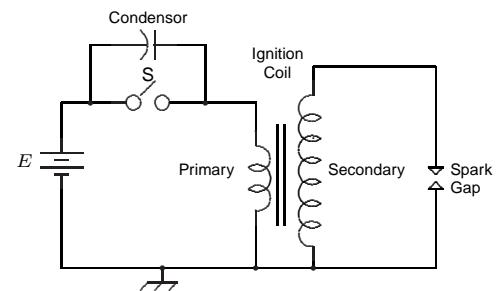


Figure 52: Inductor Snubber Network

In the case of self-induction the changing magnetic field is caused by a changing current. Then, with a different constant of proportionality⁴⁷ inductance can be defined in terms of its current-voltage differential equation:

$$e = L \frac{di}{dt} \quad (80)$$

where e is the voltage across the inductor, di/dt is the rate of change of current through it, and L is the constant of proportionality: the inductance in *Henries*⁴⁸.

In words,

The voltage across an inductor is proportional to the rate of change of the current through it.

An even more useful rule of thumb is

Current in an inductor cannot change instantaneously.

Notice the beautiful symmetry of the inductor equation 80 with the differential equation 63 for a capacitor (page 88):

Capacitor
 $i = C \frac{dv}{dt}$

Inductor
 $v = L \frac{di}{dt}$

We can now redescribe the previous example of figure 51 (page 97) in terms of inductance, figure 53. Since current and flux are proportional, we simply relabel the original *flux axis* $\phi(t)$ as *current* $i(t)$. This time, it's the current that increases at a linear rate, given by equation 81.

$$\frac{di}{dt} = \frac{E}{L} \text{ amps/second} \quad (81)$$

At time t_2 , the switch opens, the current stops abruptly and a very large voltage appears at the terminals of the inductor⁴⁹.

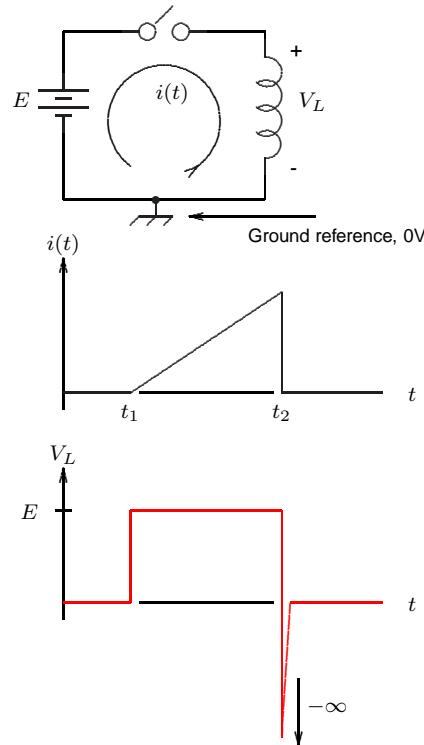


Figure 53: Inductor Transient

⁴⁷The details are in the appendix of section 2.29, Inductance and Magnetism.

⁴⁸Named after Joseph Henry, who discovered the property of *self inductance*.

⁴⁹With respect to magnetic core saturation, there is one very important difference between the flux-time curve of figure 51 and current-time waveform of 53. If the core saturates, the rate of change of flux flattens out (approaches a horizontal straight line vs time). However, on saturation, *the inductance drops and the current increases*.

2.28.5 Energy Storage in an Inductor

The magnetic field of an inductor stores energy. In an ideal inductor (no wiring resistance), the same amount of energy put into the inductor can be returned to the outside world. In practice, electromagnets are not much used for deliberate energy storage in the way that capacitors and batteries are used to store energy for, say, an electric car. However, they do have an important role in the short-term storage of energy, as in switching power supplies. Further, it may be important to plan for the energy released when a magnetic field collapses by accident.

Referring to figure 54, the voltage across the inductor is given by

$$v(t) = L \frac{di}{dt} \quad (82)$$

The instantaneous power flowing into the inductor is the product of inductor voltage and current:

$$p(t) = v(t)i(t) \text{ watts (ie, joules/sec)} \quad (83)$$

Assume the switch has been closed over the interval 0 to T seconds. Then the total energy W stored in the inductor is the time integral of the power:

$$W = \int_0^T p(t)dt \text{ joules} \quad (84)$$

Substitute for $v(t)$ in equation 83 from equation 82, and then substitute for $p(t)$ from equation 83 into equation 84:

$$\begin{aligned} W &= \int_0^T L \frac{di}{dt} i(t) dt \\ &= L \int_0^T i(t) di \\ &= \frac{1}{2} L I_L^2 \text{ joules} \end{aligned} \quad (85)$$

where I_L is the current at time T . The current and the inductance define the stored energy [44].

Notice the similarity to the equation for stored energy in a capacitor: (equation 68 on page 91):

$$W = \frac{1}{2} C V_C^2 \text{ joules} \quad (86)$$

where V_C is the voltage on the capacitor.

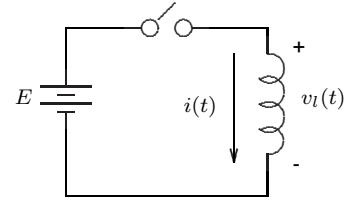


Figure 54: Inductor Energy Storage

2.29 Inductance and Magnetism

It's not difficult to construct a low-value capacitance, using equation 55 (page 86) as a guide. However, this is unusual: capacitors are usually purchased off-the-shelf.

Inductors are also available from suppliers, but it is not uncommon to have to construct an inductor to meet a particular requirement – a switching power supply filter, for example.

In this section we'll relate inductance to the magnetic properties of a material. We'll develop a formula for the inductance in terms of μ , the permeability of the magnetic material, the number of turns on the coil N , and the dimensions of the core (cross sectional area A and core length l).

Rather than repeat definitions throughout this section, magnetic units are grouped together in figure 55.

Symbol	Unit	Description
ϕ	Webers	Magnetic flux
B	Teslas	Magnetic flux density, webers/m ²
B	Gauss	Magnetic flux density, 10^{-4} Tesla
μ	Henries/metre	Magnetic Permeability
H	Amp-turns/metre	Magnetizing Force (Magneto-Motive Force, MMF)
\mathcal{R}	Amp-turns/weber	Reluctance
L	Henries	Inductance
l	Metres	Core length
A	Metres ²	Core cross-sectional area
N		Number of windings
i	Amperes	Current in the winding

Figure 55: Magnetic Units

2.29.1 The Electric Circuit Analogy

It is useful to think in terms of a *magnetic circuit*, which is analogous to an electric circuit. A *magnetomotive force MMF amp-turns/metre*, symbol H , is analogous to the circuit voltage. It creates a magnetic flux ϕ webers in a core material, where the flux is analogous to current. The ratio of the MMF to the flux is the *reluctance* \mathcal{R} . Then the magnetic analogy to Ohm's law is:

$$\phi = H\mathcal{R} \quad (87)$$

Figure 56 illustrates the relation between a magnetic circuit and its electrical analog.

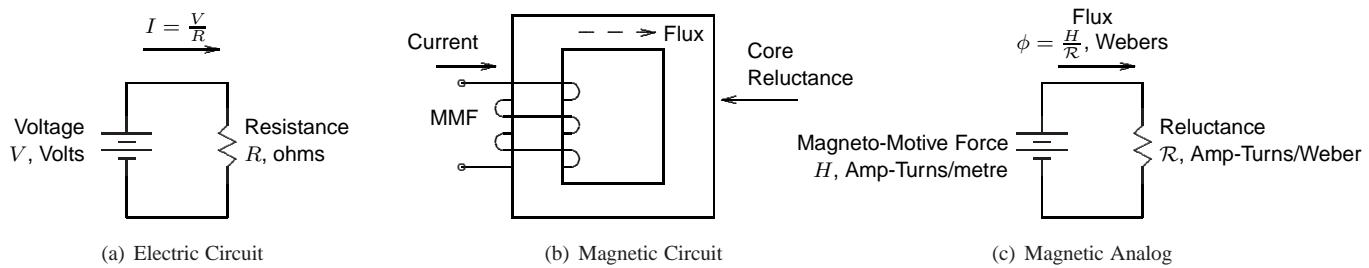


Figure 56: Magnetic Circuit Analogy

In the magnetic circuit of figure 56(b), a current-carrying coil of wire creates a *magneto-motive force, MMF* which drives flux through the core. The resistance to this flux is the *reluctance*.

The equivalent electrical circuit is in figure 56(c).

2.29.2 Reluctance

It turns out that the reluctance of a magnetic core is proportional to its length, inversely proportional to its area, and inversely proportional to its *magnetic permeability* μ . Magnetic permeability is a property of the material which is which analogous to the resistivity of a conductor. Putting this together we have

$$\mathcal{R} = \frac{l}{\mu A} \quad (88)$$

The magnetic permeability is in turn made up of two parts:

$$\mu = \mu_0 \mu_r \quad (89)$$

where

μ_0 is the *permeability of free space*, $4\pi \times 10^{-7}$.

μ_r is the *relative permeability* of some material

For example, one type of steel is 400 times as permeable as free space, that is, $\mu_r = 400$.

2.29.3 Example: Variable Reluctance Sensor

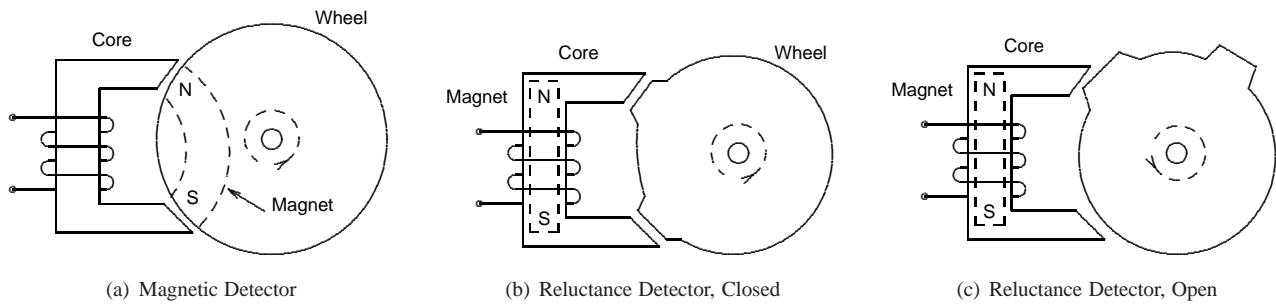


Figure 57: Rotation Detector

There are many industrial applications that require the detection of a rotating mechanic part. Figure 57(a) shows a sensor based on Faraday's law. A portion of the wheel is magnetized. As that magnetic field approaches and recedes from the core, flux from the magnet flows through the core and induces a voltage in the coil, similar to the pulse shown in figure 49 on page 95. The magnetic section in the wheel and the core form a *magnetic path* for the flux. They are made of materials such as steel that have low reluctance compared to the surrounding space, so the flux travels through them easily. We say that the core provides a *low reluctance path* for the magnetic flux.

As a practical matter, it can be difficult to magnetize a portion of a rotating part, especially if the part must rotate at high speed and remain balanced. Furthermore, the rotating part could be part of an engine that runs hot: magnets deteriorate when heated. The *variable reluctance* technique shown in figures 57(b) and (c) allows the magnet to be fixed, and it's especially attractive for rotating parts that have protrusions, such as gears.

This time, the magnet is fixed and embedded in the coil. In the case shown, the wheel has two protrusions⁵⁰. When the wheel is in the position of figure 57(b), the air gap is small, the magnetic path reluctance is low, and the flux is large.

As the wheel rotates into the configuration of figure 57(c), large gaps appear in the magnetic path. This increases the reluctance and decreases the flux through the core.

⁵⁰Mechanical engineers will object that the rotating part in figure 57(b) is badly unbalanced. However, this is a book about electronic design, and you can't expect two books for the price of one.

The net result is similar to the moving magnet case: changing flux in the core induces a voltage pulse that can be used as a signal.

Variable reluctance sensors of this type are sometimes used in an internal-combustion engine to sense crankshaft rotation and synchronize the ignition.

2.29.4 Deriving the Inductance

The standard expressions relating the flux, permeability and magneto-motive force in a magnetic material may be used to relate flux to the current in the coil:

The total flux ϕ is equal to the flux density B times the core area A :

$$\phi = BA \quad (90)$$

The flux density B depends on the magneto-motive force H times the permeability of the core, which can be read off a spec sheet or a BH curve like figure 58 on page 104.

$$B = \mu H \quad (91)$$

The magnetomotive force (MMF) is proportional to the current i in the winding and the number of turns N in the winding. It is inversely proportional to the length l of the magnetic path through the core.

$$H = \frac{Ni}{l} \quad (92)$$

Combining 90, 91 and 92 we have

$$\phi = \left[\frac{N\mu A}{l} \right] i \quad (93)$$

Now differentiate this with respect to time. The quantities in the square brackets are constants, so the only differential quantity is current:

$$\frac{d\phi}{dt} = \left[\frac{N\mu A}{l} \right] \frac{di}{dt} \quad (94)$$

Now, by Lenz's law

$$e = N \frac{d\phi}{dt} \quad (95)$$

Substitute into equation 95 from equation 94 for $\frac{d\phi}{dt}$, and we have:

$$e = \left\{ \frac{N^2 \mu A}{l} \right\} \frac{di}{dt} \quad (96)$$

The quantity inside the braces is a physical property of the inductor, known as the *inductance* and given the symbol L .

Then equation 96 can be rewritten as

$$e = L \frac{di}{dt} \quad (97)$$

which is the standard differential equation relating the behaviour of current and voltage in an inductance.

From equation 96, the inductance itself is given by:

$$L = \frac{N^2 \mu A}{l} \quad (98)$$

Thus the inductance depends on cross-sectional area, permeability of the core, and the number of turns squared.

For an inductor that has an identifiable core length and area (a toroid, for example) the results are reasonably predictable. But for a solenoid winding on a cylindrical core, part of the magnetic path is through the air and the values of length and area are much more difficult to predict.

For example, reference [45] gives the inductance of an air-core coil as

$$L = \frac{N^2 d^2}{40l + 18d} \text{ microhenries} \quad (99)$$

where d and l are the coil diameter and length in inches.

2.29.5 Saturation

There is an upper limit B_{max} to the flux density in a magnetic core, figure 58. In the figure, the value of B_{max} is about $430 \times 10^{-3} T$. The permeability, which is equal to the slope of the BH curve, decreases to a low value, and consequently the inductance decreases as well. The effect is that the inductance essentially evaporates at high flux densities, and this leads to large spikes in current.

Saturation is to be avoided, so an inductor is always operated at a flux density below saturation.

For a given core and input voltage waveform, the maximum flux density may be determined as follows:

By Lenz's law

$$e = N \frac{d\phi}{dt} \quad (100)$$

As well, the flux density B is equal to

$$\phi = BA \quad (101)$$

Substituting for ϕ in 100, we have

$$e = NA \frac{dB}{dt} \quad (102)$$

so

$$dB = \frac{1}{NA} edt \quad (103)$$

and

$$B = \frac{1}{NA} \int edt \quad (104)$$

Now we can substitute a waveform equation for e in equation 104 and determine the resultant flux density.

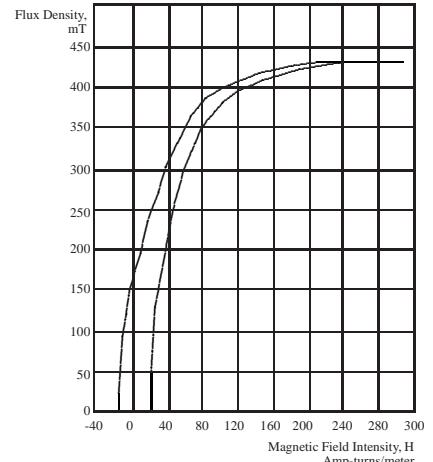


Figure 58: BH Curve

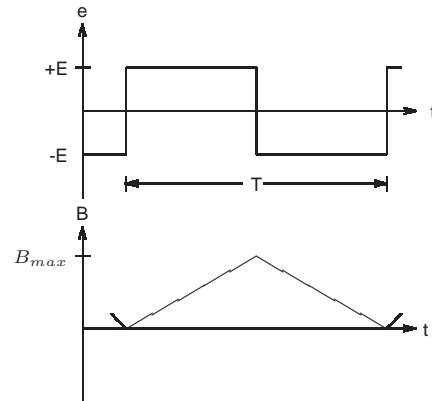


Figure 59: Flux and Voltage Waveforms

For example, if the input waveform is a square wave of magnitude E and period T (figure 59), then the maximum flux density is

$$\begin{aligned} B_{max} &= \frac{1}{NA} \int_0^{T/2} Edt \\ &= \frac{ET}{2NA} \end{aligned} \quad (105)$$

The maximum flux density may be kept below B_{max} by

- reducing the maximum input voltage E (usually not an option)
- reducing T by increasing the operating frequency
- Increasing the number of turns on the winding
- increasing the core cross-sectional area A

In electronic circuits increasing the operating frequency is an attractive option because, for a given B_{max} , the core size and number of turns can be reduced, leading to a smaller and less expensive component.

2.29.6 Summary

The important equations in this section are:

$e = L \frac{di}{dt}$	Differential equation for inductance
$\phi = BA$	Magnetic flux and flux density
$B = \mu H$	BH Curve relationship, permeability
$H = \frac{Ni}{l}$	Magnetizing Force
$e = N \frac{d\phi}{dt}$	Lenz's Law

All the other concepts can be derived from these.

2.29.7 Solenoid Inductance on Ferrite Rod

Inductors are often wound in solenoid fashion on a rod of magnetic material. When a magnetic circuit consists of a high permeability section (such as ferrite) with an air gap, the reluctance of the air gap dominates the total reluctance. Consequently, one would expect for a coil wound on a magnetic material rod, that the reluctance would be very high and the inductance very low. This suggests that the magnetic material serves no useful purpose at all.

On the other hand, we do know that a magnetic field flows by preference through a low reluctance path rather than a high reluctance path. Then magnetic lines of force will flow by preference through a ferrite rod rather than air, and a ferrite rod in space will therefore concentrate magnetic lines of force. This concentration of the lines of force is indicative of a lower reluctance and higher permeability. So there is an effect on the magnetic field, but the magnetic circuit model fails to explain it properly⁵¹.

Figure 60 explains this mystery. The effective permeability of the rod μ_{rod} is plotted against the length/diameter ratio, with the intrinsic permeability of the magnetic material μ_i as a running parameter. Consider for example the lowest trace, for a material with $\mu_i = 10$. As the length/diameter ratio increases, the effective permeability of the rod increases. At L/D values greater than 20, they are essentially equal. In effect, the magnetic field behaves as if it is entirely contained in the high-permeability rod material, and the rod is infinitely long.

A solenoid-wound coil on a ferrite rod will also vary in inductance, based on the placement of the coil (in the centre or near an end of the rod), the wire diameter, and the spacing of the turns on the winding. The design obtained by calculation must be verified by constructing a prototype.

The shape of the core is very important in interacting with an external magnetic field. For example, a toroid-wound coil has no interaction with an external field, nor does the field in the core extend outside the core. On the other hand, a solenoid coil on a rod couples very well with any external field, and is often used as an antenna for this reason. A solenoid coil on a rod also radiates its internal magnetic field into space, and so it is a prodigious generator of magnetic field interference when used in something like a switching power supply.

2.30 The Amplifier Concept

An *amplifier* is a device that strengthens a signal. Ideally, the amplifier does not distort the signal, that is, it maintains the relative amplitude and phase of the signal components while the output signal is larger than the input.

⁵¹To be fair to the model, it applies when the magnetic field is contained within a well-defined path. In the case of the solenoid-on-ferrite-rod, the magnetic field flows out into space, and there is a large amount of *fringing*. The path behaviour is not amenable to a simple analysis.

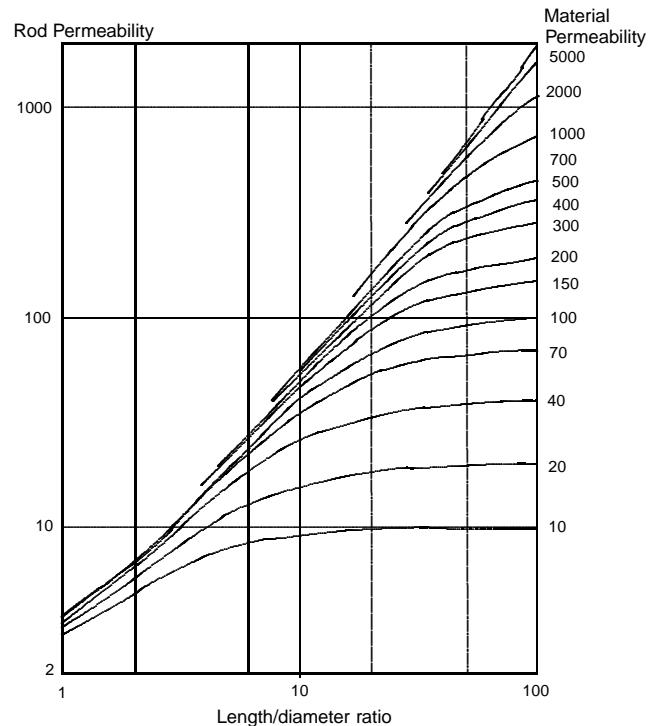


Figure 60: Ferrite Rod Permeability [2]

When we say that the signal is *larger*, this implies that the power in the output signal is greater than the power in the input signal. A transformer can step up voltage or current, but the power remains the same, so a transformer is not an amplifier.

This is not magic: an amplifier requires a *power supply* to provide the energy that is in the output signal. *In an amplifier, the input signal controls*⁵² *power from the power supply to produce a replica signal with greater power.*

Interestingly, the Canadian Oxford Dictionary defines amplifier as *an electronic device for increasing the strength of electrical signals*. Indeed, many amplifiers are electronic, but there are also amplifiers that are mechanical, pneumatic, hydraulic or biological.

The symbol for an amplifier is the triangle shown in figure 61. The pointy end indicates the direction of signal flow: from input to output⁵³.

Mechanical Amplifier

It is useful and entertaining to examine some mechanical amplifiers. The system shown in figure 62 is based on the idea of a *capstan*. The capstan is a shaft that is connected to a motor of some sort, so that the shaft is constantly rotating.

A cable is wound around the capstan. The input signal is some kind of tension force that is applied to the free end of the cable. When the cable is pulled, it tightens around the capstan. The friction between the cable and the capstan causes the cable to move to the right, driving the output linkage.

When the cable is released, the friction is reduced between the cable and the capstan. Then the spring pulls the cable to the left, and the output linkage moves to the left.

Ideally, the movement of the output should mimic the movement of the input, but be able to apply much more power. So this device is a genuine amplifier.

A device similar to this, known as a *torque amplifier* was used in a mechanical analog computer developed by Vannevar Bush, reference [46]. An accuracy in the order of 1% was achieved with the computer, so the torque amplifiers must have been very precise⁵⁴.

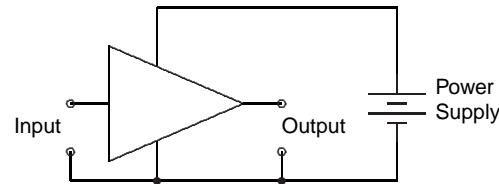


Figure 61: Amplifier

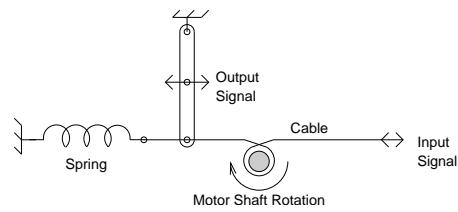


Figure 62: Unidirectional Mechanical Amplifier

⁵²The technical term for *controls* is *modulates*.

⁵³There was a brief but futile attempt at one point to reverse the amplifier signal on the basis that the signal got larger from input to output and so the wide end of the triangle symbol should be at the output.

⁵⁴There has been some recent interest in duplicating the torque amplifier in *Meccano*. (*Meccano* is a wonderful British mechanical construction set with which one can construct a wide variety of mechanisms and models of machines. Its American counterpart was called *Erector*). A search of the Internet using the terms *torque amplifier* and *Meccano* will identify the relevant sites.

Pneumatic Amplifier

A *pneumatic amplifier* uses compressed air as an energy source⁵⁵. In figure 63, compressed air is directed into the two bellows *B*. Two smaller tubes vent the bellows and straddle a *flapper*, a flat sheet of metal that can move to the left or right, blocking or opening the vents.

When the flapper is in the centre position, the rate of discharge flow through the two small pipes will be equal. The pressure in the two bellows will be equal, and the output will also be centred.

If the flapper is moved to the left side, for example, then the pressure will increase in the left bellows and decrease in the right bellows. The left bellows will expand and the right bellows contract, moving the output linkage to the right. So this is an inverting amplifier: leftward movement of the input causes rightward movement of the output.

The amplification in this device is based on the fact that pressure is everywhere the same in a given enclosure. Force is pressure times area, so a force may be scaled by scaling the area. The flapper works against a small area, so the forces against the input signal are relatively small. The bellows is a much larger area, so it generates larger forces in driving the output signal. Consequently, the device has power gain.

The restrictors are necessary to ensure that there is a pressure drop between the air supply and the input orifice when air is flowing. Otherwise, the pressure is everywhere the same.

2.31 Switches and Relays

The electromechanical switch is one of the most simple devices in the extended universe of electronic components. However, there are some points worth mentioning.

An electrical switch is characterised by some of the following properties.

- the amount of current it can carry and safely disconnect
- the maximum allowable voltage across the switch when it's open-circuit
- the number of independent circuits that are simultaneously connected when the switch is closed: the switch *throws*.
- the number of positions of the switch: the switch *poles*.
- the *action* of the switch: momentary (ie, spring release, has to be held down), toggle (two stable positions), centre-off (three possible positions), or rotary (many possible positions).

The schematic symbols for a variety of switches are shown in figure 64.

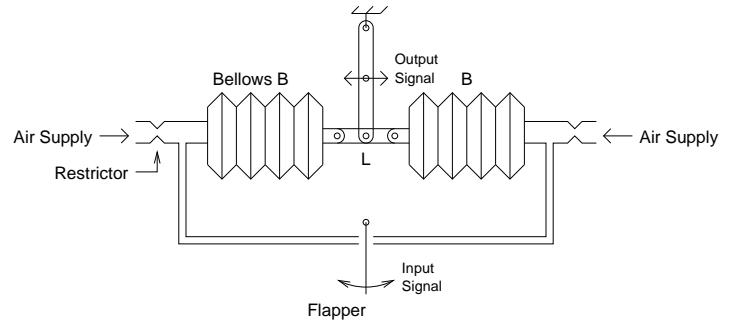


Figure 63: Pneumatic Amplifier

⁵⁵Figure 63 is based on my recollection of a pneumatic amplifier described in *The Amateur Scientist* column of *Scientific American* magazine. It was used to drive a pen recorder in a seismograph, a device for measuring and recording earthquakes.

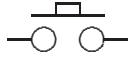
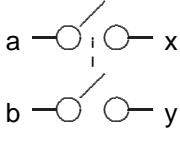
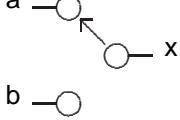
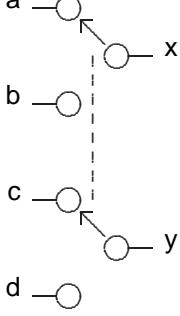
Symbol	Description	Notes
	SPST, NO, Toggle	Single pole, single throw, normally open, toggle action
	SPST, NO, PB	Single pole, single throw, normally open, pushbutton action
	DPST, NO	Double pole, single throw, normally open, toggle or pushbutton action. Points a and b are simultaneously connected and disconnected to x and y respectively.
	SPDT	Single pole, double throw, toggle or pushbutton action Point a or point b is connected to x.
	DPDT	Double pole, Double throw, toggle or pushbutton action. Point a is connected to x and c to y, or b to x and d to y.

Figure 64: Switch Types

Relays

A *relay* is a combination of electromagnet and a two-position switch mechanism. When current is passed through the electromagnet, it activates the switch, which then moves to its energized state. When the current is removed, the switch moves to its de-energized state.

An example of a relay, in schematic notation, is shown in figure 65. The connections to the coil supply current to activate the relay. The switch contact designated as *normally open*(NC) is connected to the pole contact when the relay is de-energized. When the relay is energized, the pole moves across to the *normally open* (NO) contact.

Relays can have a much more sophisticated contact structure, such as a 4PDT arrangement.

Relays were used as logic elements in early digital computers. However, they are relatively slow and power-hungry compared to solid-state logic. The four-bit counter shown in [47] requires more than 2 amperes at 12 volts to operate.

Relays are available in a wide variety of physical configurations, ranging from devices the size of a sugar cube to enormous *contactors* used to start industrial motors.

2.32 Marx Generator

The *Marx Generator* is a generator circuit that uses a cascade of capacitors and inductors to generate extremely high voltage pulses. Among other applications, it's used to test transformers for insulation breakdown [48] and as a rock disaggregator [49].

A Marx Generator is often built to a very large scale, but the circuit and operating principle are quite simple. Figure 66(a) shows the schematic.

The circle terminals reached by a diagonal path are the *stage spark gaps*. The arrow terminals comprise the final spark gap. In effect, the capacitors are charged in parallel from the DC supply and then discharge in series into the output spark gap. The result is an output which is N times the input voltage, where N is the number of stages (5 in the diagram shown).

Now the operation in more detail. When the input DC supply is first energized, the spark gaps appear as an open circuit and the inductors appear as a short circuit. Then the charging circuit is as in figure 66(b).

At some point, the voltage on the first capacitor becomes large enough to initiate a spark across the first spark gap. This causes an abrupt rise in voltage at the other side of the spark gap. Current cannot change instantaneously in an inductor, so the inductors in the first stage appear as open circuits. The second capacitor is now in series with the first, so the voltage at its spark gap fires. In short order, all the intermediate spark gaps fire. The discharging circuit then appears as in figure 66(c), creating sufficient voltage to discharge through the final spark gap.

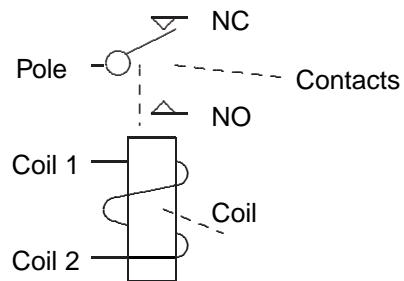


Figure 65: SPDT Relay

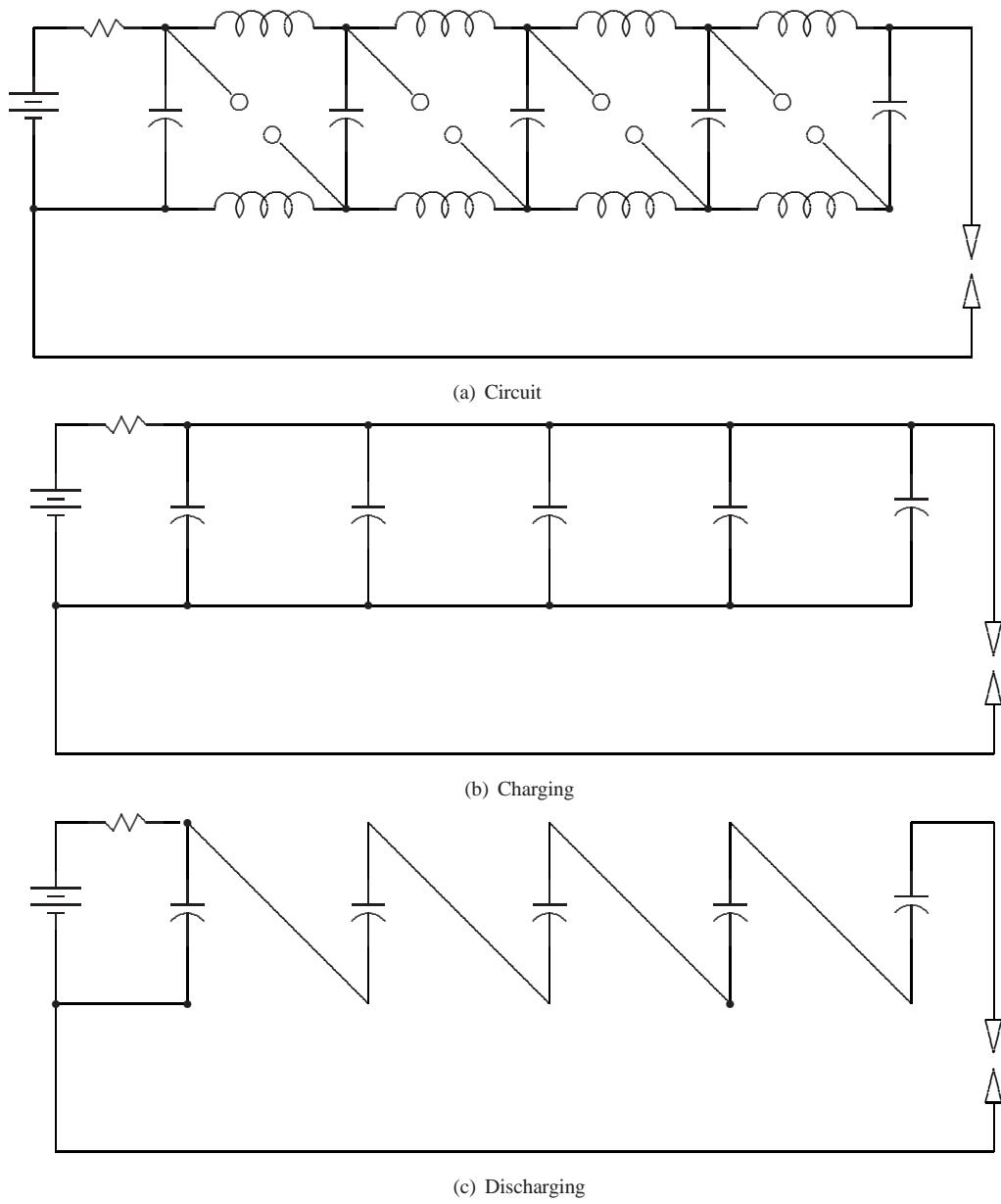
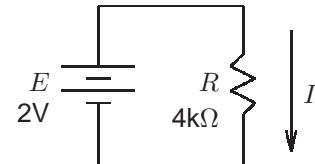
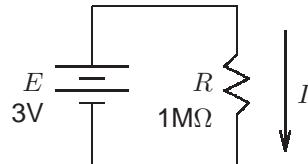
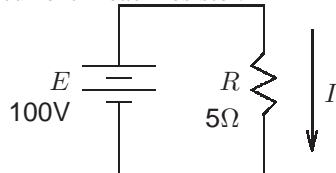


Figure 66: Marx Generator

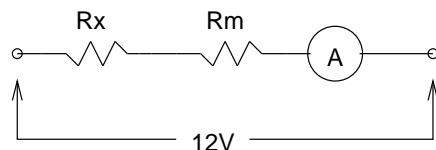
2.33 Exercises



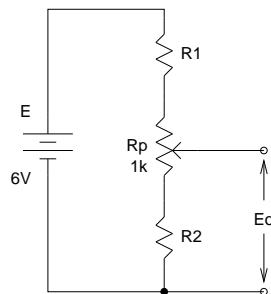
1. Calculate the current in each resistor.



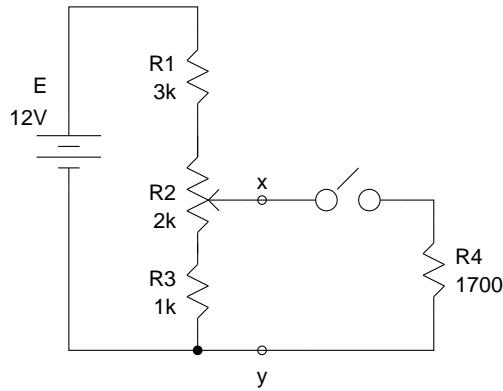
2. Sketch the VI characteristic of a 1000Ω resistor.
3. A certain car battery is measured with a voltmeter and found to measure 14 volts. When it is connected to a load resistor of 1Ω , the terminal voltage drops to 13 volts.
- What is the open circuit voltage E_{oc} of the battery?
 - What is the internal resistance R_{int} of the battery?
 - Draw the equivalent circuit for the battery.
 - What is the short circuit current I_{sc} of the battery?
4. A $100\mu\text{A}$ meter movement has an internal resistance R_m of 875Ω . What is the value of the resistance R_x that must be added in series with the meter if a 12V input drives the meter to full scale?



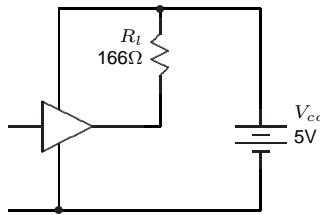
5. For the circuit shown, calculate values for R_1 and R_2 so that the voltage at the wiper of the potentiometer can be adjusted between 1.5 volts and 3.5 volts.



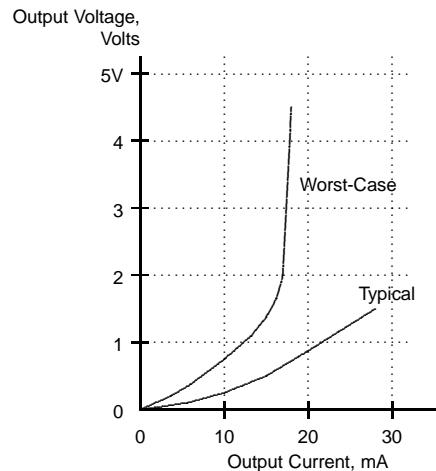
6. Determine the voltage V_{xy} for (a) switch open and (b) switch closed.



7. In the figure, a digital logic gate from the High Speed CMOS family is used to drive a relay coil, which may be characterised as a 166Ω resistor.



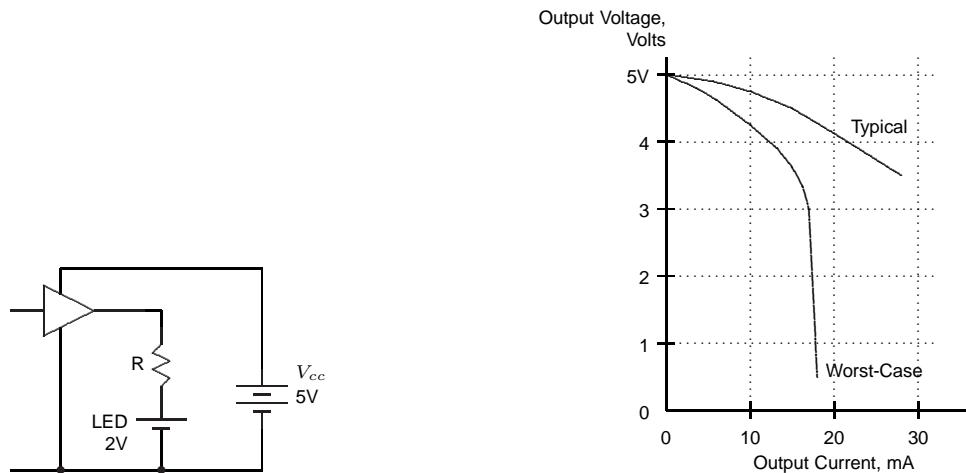
The output voltage-current characteristics of the HCMOS gate when sinking current is shown below. Characteristics are shown for a typical case and worst-case.



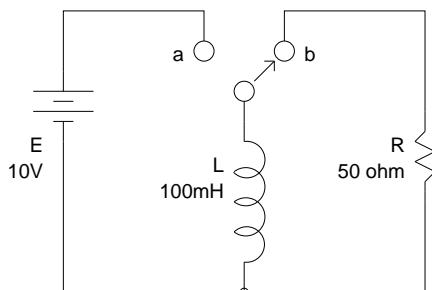
This is a 5 volt relay, so its coil current is designed to be $5V/166\Omega = 30mA$. The relay will operate reliably as long as its current exceeds 75% of this value.

Using load-line analysis, determine whether the circuit will function reliably.

8. In the figure shown below, a digital logic gate from the High Speed CMOS family is used to drive a light-emitting diode, which may be characterised as a 2 volt constant-voltage device. The output voltage-current characteristics of the HCMOS gate when sourcing current are also shown. Characteristics are shown for a typical case and worst-case.

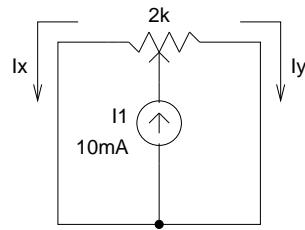


- (a) Choose a suitable value for R such that the LED current will be 10mA in the typical case.
 - (b) What is the LED current in the worst case?
 - (c) What is the output voltage of the logic gate in the worst case?
9. In the circuit shown, the switch starts in position (b). It is moved to position (a) for 2 seconds and then back to (b).



- (a) Switches can be obtained in one of two varieties: *make before break*, in which the new contact is closed before the old one is opened and *break before make*, in which the old contact is opened before the new one is closed. Which type of switch should be used here, and why?
 - (b) When the switch is moved back to position (b) show the direction of current through the resistor, and the polarity of the voltage across it. (Mark these on the diagram).
 - (c) Sketch a diagram of the inductor current vs time for 6 seconds from the time the switch is first moved.
10. For the circuit, what are I_x and I_y for the following potentiometer wiper positions (fill in the boxes):

Wiper Position	I_x	I_y
Fully left		
Centred		
Fully right		

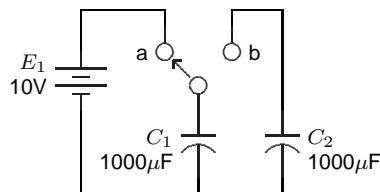


11. A nicad battery open-circuit voltage is 9.6 volts when fully charged. The open-circuit voltage decreases to 7 volts when the battery is nearly discharged.

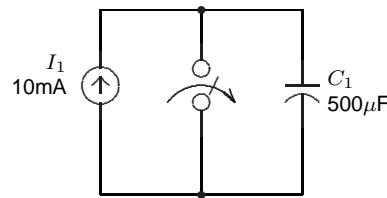
Suppose a nicad battery of this type, in discharged condition, is connected to one-amp constant current charger. The terminal voltage of the battery is measured as 9 volts while being refueled by this charger.

- (a) Calculate the internal resistance of this battery.
- (b) The battery is charged for 6 hours on this charger. At the end of the charging process, the open-circuit voltage is 10.2 volts. Assuming that the complete battery charge is available to drive a load, how long in hours will the battery drive a load of 100mA?

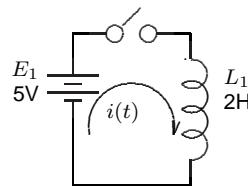
12. Assume that capacitor C_2 is uncharged to begin with. The switch starts off in position *a* and moves to position *b*.



- (a) What is the voltage across the two capacitors after the switch is moved to position *b*? (Hint: Charge is conserved.)
 - (b) Now consider the original situation with the switch in position *a*, and the second capacitor uncharged. What is the total energy stored in the two capacitors?
 - (c) When the switch has been moved to position *b*, what is the total energy stored in the two capacitors?
 - (d) Advanced question: Where did the rest of the energy go?
13. Assume the capacitor is uncharged to begin with. Draw the waveform of voltage vs time, with amplitude and time scales, for the two seconds after the switch is opened.



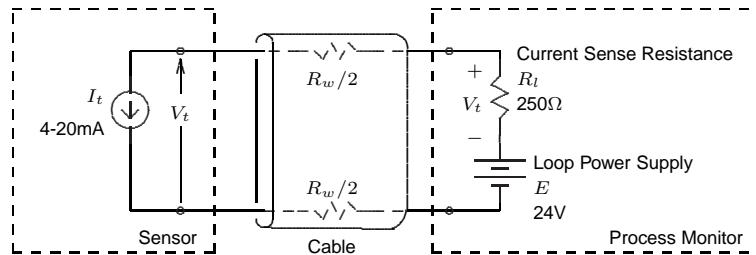
14. The switch is closed at time $T = 0$ seconds.



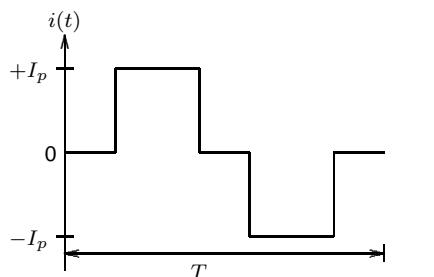
- (a) Sketch the waveform of $i(t)$ for the first 5 seconds after the switch is closed.
 (b) Sketch the voltage across the inductor between the intervals $T = 0$ seconds and $T = 6$ seconds if switch is opened at $T = 5$ seconds.
15. Two SPST toggle switches A and B are to be wired so that both of them must be closed to turn on a 12 volt lamp. Draw the circuit.
16. Two SPST toggle switches A and B are to be wired so that either of them turns on a 12 volt lamp. Draw the circuit.
17. Draw the circuit diagram for a *two-way* switch arrangement. There are two SPDT toggle switches A and B, a lamp and a battery. Either switch can turn on the lamp, and either can turn it off.
18. A 12 volt relay is to be used in a *lockup* circuit. There are two momentary-action pushbutton switches that control the circuit.
- Pushing switch A causes the relay to close. When switch A is released, the relay stays closed.
 - Pushing switch B causes the relay to open. When switch B is released, the relay stays open.
 - The relay has SPST contacts and a 12 volt coil.
 - Switch A is normally open.
 - Switch B is normally closed.
 - A 12 volt lamp illuminates when the relay is energized.
 - The circuit is supplied power by a 12 volt battery.

Draw the circuit.

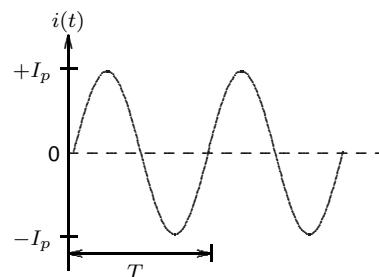
19. The 4-20mA current loop is widely used in industry to transmit information from a sensor to a process monitor, or from a process controller to a receiver. Referring to the figure, a 24 volt *loop power supply* operates the loop. The current generator I_t passes a current between 4 and 20mA, proportional to the sensor reading. The *current sense resistor* translates this measurement current into a voltage V_t which is read by the measurement system.



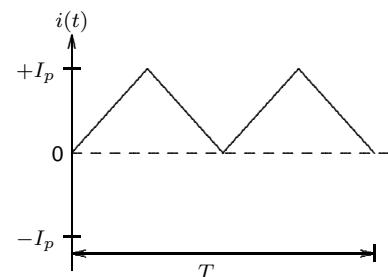
- (a) What effect does the wiring resistance have on the sensor voltage reading V_t ?
- (b) The loop current values are chosen so that an open circuit in the wiring is easily detectable from the sensor voltage reading V_t . Explain how the sensor equipment can detect a wiring open circuit.
- (c) The 24 volt supply is distributed around the loop. If the current generator requires a minimum of 8 volts across it to function correctly, what is the maximum allowable voltage drop across the wiring resistance?
- (d) It is usual to power the sensor electronics from the voltage V_t across the current generator. (This eliminates the requirement for a power supply at the sensor.) If the wiring resistance can vary from zero up to the maximum calculated above, what is the range of voltages to expect across the current generator?
20. For each of the following waveforms of capacitor current, the peak value is I_p . Sketch the corresponding waveform of capacitor voltage, indicating the magnitude in terms of the waveform period T and capacitance C .



(a) Pulse



(b) Sine



(c) Ramp

Figure 67: Capacitor Currents

21. Show that equation 99 on page 104 is dimensionally correct.
22. Repeat the analysis on page 104 to determine flux density B when the input waveform is a sine wave of frequency ω .
23. Redesign the mechanical torque amplifier of figure 62 so that it is bi-directional. That is, the spring is eliminated and the output signal is driven in both directions by the motor shaft.
24. With reference to the Marx Generator of section 2.32 and page 110:
- (a) One variation on the design is to use resistors instead of inductors [48]. What effect would this have on the operation of the generator?

- (b) The following questions refer to the Marx generator shown in [49]. That design uses a *pole pig* hydro transformer that is normally used as a step down transformer, ‘120kV to 240V’. In this application the transformer is connected to a 220V AC supply as step up transformer. A half-wave rectifier rectifies the AC and charges up the first capacitor to the peak value of the AC. What is the expected DC voltage on the first capacitor? What is the expected output voltage at the final spark gap?
- (c) The capacitors are $1\mu\text{F}$ units. At 100kV, what is the energy stored in one capacitor, in joules?
- (d) The inductors used in this design are about 1.25 inches in diameter by 12 inches long. Assuming that the wire used was #24 AWG and the coil was wound as a single layer, determine the approximate number of turns on the inductor. Using the equation for the inductance of an air-core coil (equation 99 on page 99) estimate the inductance of these coils.

3 Tools for Circuit Analysis: Part 1

3.1 Thevenin's Theorem

Thevenin's theorem is a very useful method of simplifying an electronic circuit. It states that

any network of sources and resistances can be reduced to one voltage source in series with one resistance. The reduced circuit is the Thevenin equivalent circuit.

The basic concept is illustrated in figure 68. Both of these circuits, the complicated one and the simple one, produce the same voltage at terminals AB. Both would produce the same voltage and current in a device attached at AB.

This process of simplifying a network to its Thevenin equivalent is referred to as *Thevenizing* a circuit. The simplified circuit is often referred to as the *Thevenin voltage* E_{th} in series with the *Thevenin resistance* R_{th} . We'll stick with E_{oc} and R_{int} .

The Thevenin voltage source is equal to the open-circuit voltage E_{oc} of the original network. The internal resistance R_{int} can be obtained by either of two methods:

- Determine the short circuit current I_{sc} , that is, the current out of the network when the terminals AB are short-circuited. Then

$$R_{int} = \frac{E_{oc}}{I_{sc}} \quad (106)$$

- Replace all the voltage and current sources by their internal resistances: voltage sources are short-circuited, current sources are open-circuited.

Now connect to the terminals of the network an imaginary ohm-meter and measure the resistance. Whatever is in the network must reduce to a group of resistors in parallel and series, which can be reduced to one resistance. This is R_{int} . Figure 69 shows how the circuit of figure 68 would be modified to determine the internal resistance, where the voltage sources have been short-circuited and the current source open-circuited.

The Voltage Divider

The voltage divider is often used to generate a low-current voltage source. (The voltage divider is explored more fully in section 3.3). It's important to be able to Thevenize a voltage divider, that is, convert it into an equivalent E_{oc} and R_{int} .

An example of this process of Thevenizing a voltage divider is shown in figure 70.

- The original voltage divider is shown in figure 70(a).

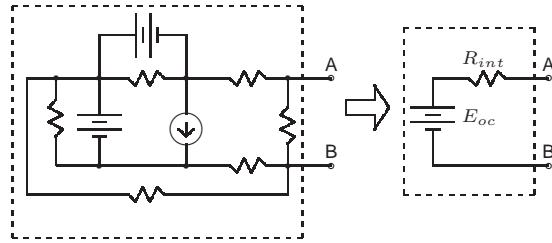


Figure 68: Thevenin's Theorem

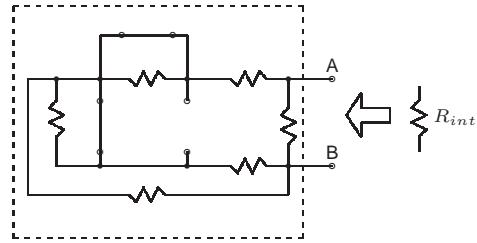


Figure 69: Calculating R_{int}

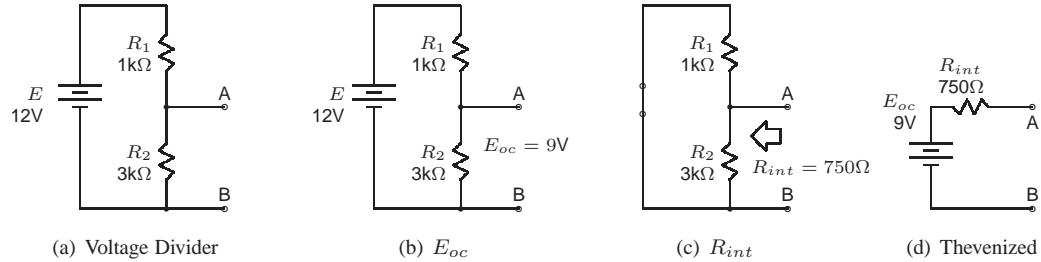


Figure 70: Thevenizing a Voltage Divider

- The calculation of E_{oc} is shown in figure 70(b), according to

$$\begin{aligned} E_{oc} &= E \left(\frac{R_2}{R_1 + R_2} \right) \\ &= 12 \left(\frac{3000}{1000 + 3000} \right) \\ &= 9 \text{ V} \end{aligned}$$

- As shown in figure 70(c), the value of R_{int} is determined by shorting E and measuring the resistance between the terminals A-B. Resistors R_1 and R_2 are in parallel, so

$$\begin{aligned} R_{int} &= R_1 \parallel R_2 \\ &= 3000 \parallel 1000 \\ &= 750\Omega \end{aligned}$$

- The final thevenized circuit consisting of E_{oc} and R_{int} is shown in figure 70(d).

Thevenin and Alternating Currents

The previous material shows the application of Thevenin's Theorem to direct current circuits, but it is equally applicable to AC circuits or some mix of AC and DC, and circuits that contain some combination of capacitors, inductors and resistors. In that case, the Thevenin source will be AC or some mix of AC and DC, and the internal resistance will be an internal *impedance* (section 4.17).

3.2 Norton's Theorem

If you know Thevenin's theorem (section 3.1), this is easy. Norton's theorem says that **any network of sources and resistances can be reduced to one current source in parallel with one resistance**. The basic concept is illustrated in figure 71. Both of these circuits, the complicated one and the simple one, produce the same voltage at terminals AB. Both would produce the same voltage and current in a device attached at AB.

This process of simplifying a network to its Norton equivalent is referred to as *Nortonizing* a circuit. The simplified circuit is often referred to as the *Norton current* I_{sc} in parallel with the *Norton resistance* R_{int} .

The Norton current source is equal to the short-circuit voltage I_{sc} of the original network. The internal resistance R_{int} is obtained by the same methods as it was in the Thevenin case.

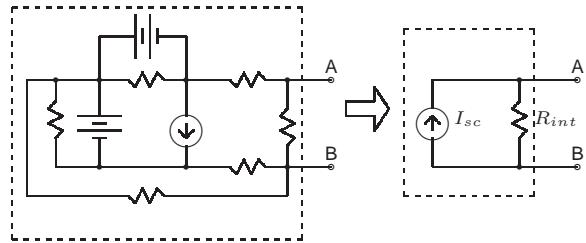


Figure 71: Norton's Theorem

Thevenin and Norton

These two equivalent circuits are themselves equivalent, according to the relationship in equation 107.

$$R_{int} = \frac{E_{oc}}{I_{sc}} \quad (107)$$

Depending on the particular circuit being analysed, the Thevenin or Norton equivalent circuit may be more convenient, and can be selected as needed.

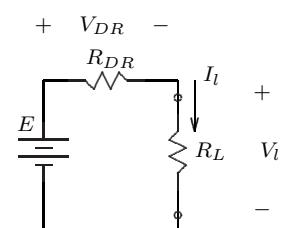
For example, a 1 volt, 10 ohm Thevenin source, seen from the terminals A-B, is exactly equivalent to a 1 amp, 10 ohm Norton source.

3.3 The Voltage Divider

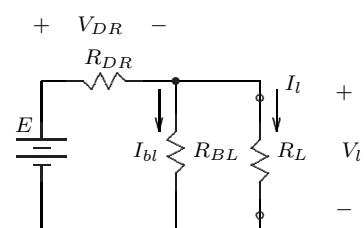
Evolution

To provide a voltage below the supply voltage, the simplest scheme is a *dropping resistor* R_{DR} as shown in figure 72(a). The voltage V_L across the load resistor R_L is equal to the supply voltage minus the voltage across the dropping resistor. The voltage drop is proportional to the load current. Consequently, this scheme requires that **the load current must be completely constant**. Any variation in load resistance will cause a variation in load voltage.

There are many situations where the load current is not constant and this circuit is not satisfactory. How could we make the load voltage less dependent on changes in load current?



(a) Dropper Resistance



(b) Added Bleeder Resistance

Figure 72: Voltage Divider Evolution

If we can somehow lower the resistance of the dropper R_{DR} , that will have the desired effect. However, just lowering R_{DR} while keeping the voltage drop V_{DR} constant requires that the current through R_{DR} increase. That extra current cannot be allowed to flow through the load resistor or it will change the load voltage. The answer is to add a *bleeder resistor* R_{BL} to provide a path for the extra current, as shown in figure 72(b).

Now the current through the dropper resistance can be much larger than the load current. For example, the dropper current might be made 10 times the load current, with 9 times the load current flowing through the bleeder resistance. When the load current changes by some amount ΔI_l , then this change in current is a small part of the total current through the dropper resistance, and the change in voltage drop (and output voltage) is correspondingly small.

The price to pay for this is the bleeder current. To be effective, the bleeder current must be much larger than the load current and, because it does not flow through the load resistance, it is entirely wasted. Consequently, this circuit is inherently an inefficient way of generating a required voltage. It is not a practical circuit for a large load current. However, if the load current is small, the wasted current is of minor importance and can be tolerated. This is in fact the situation in many circuits and the circuit is widely used.

This circuit is known as a *voltage divider*.

The Voltage Divider

A *voltage divider* consists of a string of two or more resistors that divide up the input voltage into smaller chunks of output voltage.

Notice that this circuit is a *ratioing* device. For example, in figure 73, we have

$$\frac{V_1}{V_2} = \frac{R_1}{R_2} \quad (108)$$

That is, the voltage across each resistance is proportional to its resistance.

The individual voltages sum to the supply voltage:

$$E = V_1 + V_2 \quad (109)$$

and the individual resistances sum to the total resistance:

$$R_t = R_1 + R_2 \quad (110)$$

Then we can write

$$\frac{V_1}{E} = \frac{R_1}{R_t} \quad (111)$$

and

$$\frac{V_2}{E} = \frac{R_2}{R_t} \quad (112)$$

In words, **the proportion of voltage across a given resistor to the total voltage is the same as the proportion of the given resistance to the total resistance**.

The same concept may be extended to an n resistor divider, as shown in figure 74. The output voltage proportion V_n/E is equal to R_n/R_T , where R_T is the total resistance of the voltage divider.

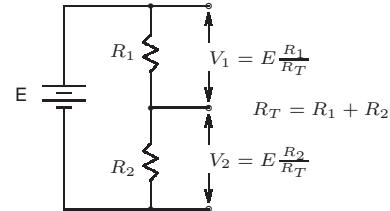


Figure 73: Voltage Divider
(109)

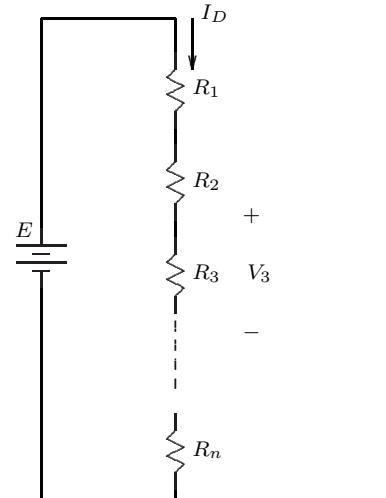


Figure 74: Voltage Divider with n Resistors

The absolute values of the resistances do not affect the output voltage. (They do have an important effect, however, and we'll get to that in a moment.)

For example, in figure 74,

$$\frac{V_3}{E} = \frac{R_3}{R_t} \quad (113)$$

where R_t is the total resistance of the divider string of resistors.

For example, in a two-resistor divider if the two resistors are equal, $V_1 = V_2 = E/2$, independent of the values of the resistances R_1 and R_2 . In the n resistor divider of figure 74, if all the resistors are equal the voltage across each one is E/n .

Example

For the voltage divider shown in figure 75, determine the voltage across each resistor *without calculating the divider current*.

Solution

The voltages across the resistors are proportional to their resistance. If x is the unknown quantity, then we can write

$$V_1 = 1000x \quad (114)$$

$$V_2 = 2000x \quad (115)$$

$$V_3 = 5000x \quad (116)$$

As well, these three voltages add up to the supply voltage:

$$E = V_1 + V_2 + V_3 \quad (117)$$

Substitute 16V for E and the expressions for V_1 , V_2 and V_3 from equations 114 through 116 in equation 117 and we have

$$16 = 1000x + 2000x + 5000x$$

Solving for x :

$$x = 2 \times 10^{-3}$$

Back-substituting for x in equations 114 through 116,

$$V_1 = 2V$$

$$V_2 = 4V$$

$$V_3 = 10V$$

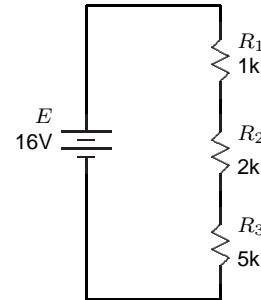


Figure 75: Voltage Divider Exercise

(117)

The Potentiometer

The voltage divider is so useful that it is provided by a special mechanical device: the *potentiometer* (pot), figure 76. The potentiometer consists of a resistive track which corresponds to R_T in figure 73, and a wiper, which connects to some point on the resistive track. By turning a mechanical shaft, the position of the wiper is changed, and this changes the output voltage.

Notice that a potentiometer is not a variable resistor. A variable resistor is a two-terminal device and this has three terminals. However, the pot can be used as a variable resistor by attaching to the wiper and one of the track terminals.

- When the wiper is at the top of the track, the output voltage is E .
- When the wiper is at the bottom of the track, the output voltage is 0.
- When the wiper is at the mid-point of the track, the output voltage is $E/2$.

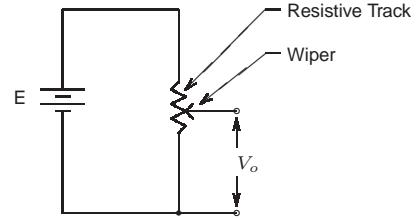


Figure 76: Voltage Divider

Voltage Divider Thevenin Equivalent

(See also section 3.1).

The voltage divider is most useful as a source of voltage that is some fraction of the input voltage. To use it effectively, we have to look at its equivalent circuit, shown in figure 77.

This says that a voltage divider behaves as a voltage source with an internal resistance. The open circuit voltage E_{oc} is proportional to the ratio of the two resistances (figure 73). The internal resistance is equal to the parallel combination of the two divider resistances.

The values of the resistors are important when the voltage divider is used as a voltage source, which is its usual application. From section 2.17, the internal resistance of a source must be much less than the load resistance if the source is to appear as a constant voltage.

Common sense tells us that there has to be a difference between a voltage divider constructed with two 1Ω resistors, and one constructed with two $1M\Omega$ resistors. Their values of open-circuit output voltage would be the same, but their internal resistances would be vastly different. The 1Ω resistor divider would have a much lower internal resistance than the $1M\Omega$ resistor divider.

Furthermore, notice that there is current down the divider that does not flow into the load resistance. This divider current must be much larger than the output (load) current, which is another way of saying that the internal resistance must be much less than the load resistance. This divider current is wasted, so the voltage divider is a very inefficient source and consequently practical only for very small output currents.

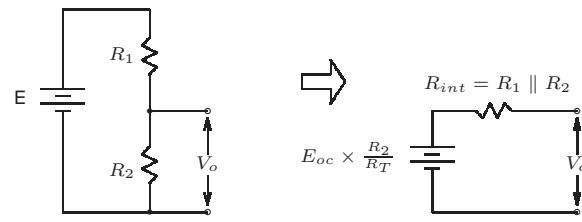


Figure 77: Voltage Divider Equivalent Circuit

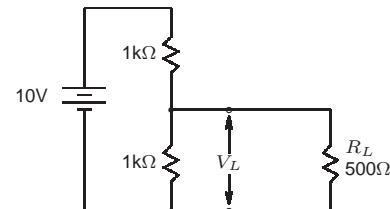


Figure 78: Voltage Divider Example

Whenever you see a voltage divider, **Think Thevenin**: mentally convert the voltage divider into its thevenin equivalent circuit. Using this approach, consider figure 78. You should be able to determine V_L by inspection. (Answer: 2.5 volts).

3.4 Voltage Divider Design

The voltage divider is a required component of many electronic designs. Because it is required so frequently, we'll spend some time developing design methods.

This section shows three design methods for the voltage divider:

- An *Exact Method* yields precise results but is time-consuming to apply
- A *Simplified Method* that quickly yields an approximate answer
- A *Graphical Method*, using a nomograph, which helps to choose standard 5% resistors for a voltage divider.
- A *Short-Cut for Large Division Ratios*. This method applies to dividers where the output voltage is very small compared to the input voltage.

In practice, the *Simplified Method* is usually adequate.

3.4.1 Exact Method

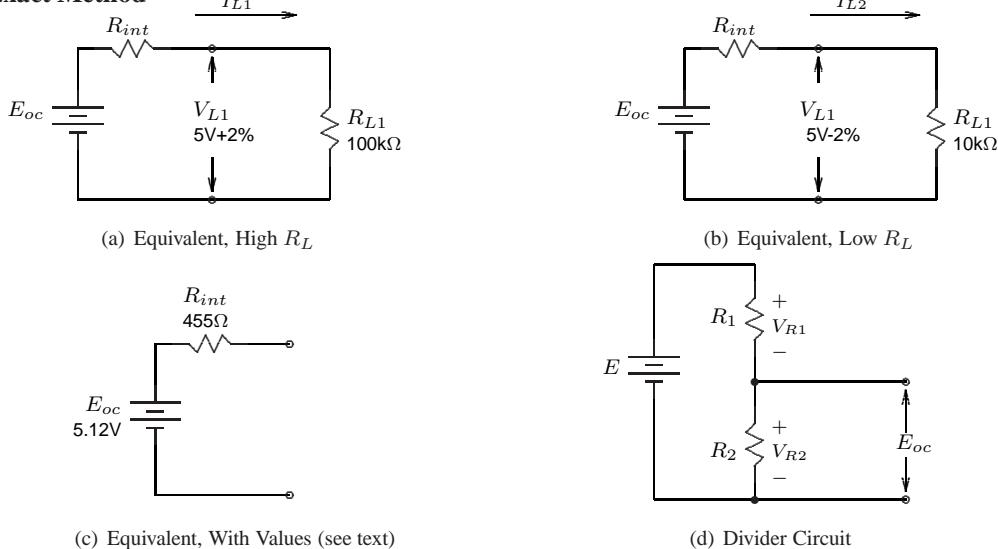


Figure 79: Voltage Divider Design

This *exact method* of designing a voltage divider generates a design which exactly meets the requirements. We'll illustrate the exact method with an example, but bear in mind that most voltage dividers can be designed by a much simpler process than this.

To apply the exact method, determine the requirements and then use the internal resistance equation:

$$V_o = E_{oc} - I_o R_{int} \quad (118)$$

to calculate the open-circuit voltage E_{oc} and internal resistance R_{int} for the divider. Then use those values for E_{oc} and R_{int} to determine the voltage divider resistors, using the voltage divider equations:

$$\frac{V_{R2}}{V_{R1}} = \frac{R_2}{R_1} \quad (119)$$

$$R_{int} = R_1 \parallel R_2 \quad (120)$$

Example

Design a voltage divider to generate 5 volts from a 12 volt supply. The output voltage should be within $\pm 2\%$ of 5 volts as the load resistance varies from $10k\Omega$ to $100k\Omega$.

Solution

Using the equivalent circuit, we will determine the open-circuit voltage E_{oc} and internal resistance R_{int} . Then we will translate that equivalent circuit into a voltage divider.

The equivalent circuit for the larger output voltage (and larger load resistance) is shown in figure 79(a).

$$V_{L1} = 5V + 2\% = 5.1V \text{ at } R_{L1} = 100k\Omega$$

$$I_{L1} = \frac{V_{L1}}{R_{L1}} = \frac{5.1}{100 \times 10^3} = 51\mu A$$

Similarly, the equivalent circuit for the smaller output voltage and smaller load resistance is shown in figure 79(b).

$$V_{L2} = 5V - 2\% = 4.9V \text{ at } R_{L2} = 10k\Omega$$

$$I_{L2} = \frac{V_{L2}}{R_{L2}} = \frac{4.9}{10 \times 10^3} = 490\mu A$$

Now we can apply the internal resistance equation twice, once for each of these situations:

$$\begin{aligned} V_{L1} &= E_{oc} - I_{L1}R_{int} \\ V_{L2} &= E_{oc} - I_{L2}R_{int} \end{aligned}$$

Subtracting these two equations, we have

$$\begin{aligned} V_{L1} - V_{L2} &= -I_{L1}R_{int} + I_{L2}R_{int} \\ &= R_{int}(I_{L2} - I_{L1}) \end{aligned}$$

Substituting numeric values, we can solve for the internal resistance.

$$\begin{aligned} R_{int} &= \frac{V_{L1} - V_{L2}}{I_{L2} - I_{L1}} \\ &= \frac{5.1 - 4.9}{490\mu A - 51\mu A} \\ &= 455\Omega \end{aligned}$$

Now we can back-substitute for R_{int} in one of the previous equations to solve for the open-circuit voltage E_{oc} .

$$\begin{aligned} E_{oc} &= V_{L1} + I_{L1}R_{int} \\ &= 5.1V + 51\mu A \times 455\Omega \\ &= 5.12V \end{aligned}$$

The voltage-divider equivalent circuit with its values is shown in figure 79(c).

Now we can convert this into a voltage divider with equations 119 and 120.

Starting with equation 119, it's convenient to introduce a constant m equal to the ratio of the voltages across the two resistances:

$$\begin{aligned} m &= \frac{V_{R2}}{V_{R1}} \\ &= \frac{5.12}{12 - 5.12} \\ &= \frac{5.12}{6.88} \\ &= 0.744 \\ &= \frac{R_2}{R_1} \end{aligned} \tag{121}$$

Consequently,

$$\begin{aligned} R_2 &= mR_1 \\ &= 0.744R_1 \end{aligned} \tag{122}$$

Recalling the formula for parallel resistors, it is convenient to rewrite equation 120 as:

$$\frac{1}{R_{int}} = \frac{1}{R_1} + \frac{1}{R_2} \tag{123}$$

Substitute for R_2 from equation 122 and we have:

$$\frac{1}{R_{int}} = \frac{1}{R_1} + \frac{1}{mR_1} \tag{124}$$

Solve for R_1 and substitute the value of k :

$$\begin{aligned} R_1 &= \left(\frac{1+m}{m} \right) R_{int} \\ &= \left(\frac{1+0.744}{0.744} \right) \times 455 \\ &= 1066\Omega \end{aligned} \tag{125}$$

Finally, use equation 122 to find R_2 :

$$\begin{aligned} R_2 &= mR_1 \\ &= 0.744 \times 1066 \\ &= 793\Omega \end{aligned} \tag{126}$$

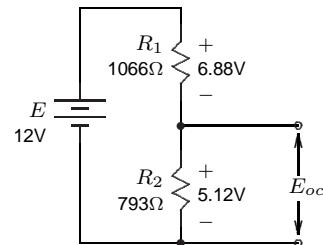


Figure 80: Voltage Divider Design

The final design is shown in figure 124.

3.4.2 Simple Method

In designing a voltage divider, the internal resistance should be low enough compared to the load resistance that load current does materially affect the open-circuit voltage. This is illustrated in figure 81(a). As R_L varies over its range, the load voltage should remain more-or-less constant.

The low internal resistance of the voltage divider is achieved at the expense of the bleeder current, which is much larger than the load current and is wasted. A lowering the internal resistance requires increasing the bleeder current, and so in the interest of current economy the internal resistance should be made no lower than necessary.

The *Exact Method* of the previous section generates a voltage divider which is optimum – it wastes exactly as much current as required to create the necessary internal resistance. However, the Exact Method is cumbersome and time-consuming to apply. It would be convenient to have a rule of thumb that could be applied quickly to generate a voltage divider design.

A suitable target for this quicky design method is that the internal resistance should be $1/20$ of the load resistance. Then the divider output voltage, when attached to a load resistance, will be within about 5% of the divider open-circuit value. Non-critical designs use 5% tolerance resistors, so this is in the same neighbourhood.

The rule of thumb:

- If the two resistors are approximately equal, make them approximately $1/10$ of the minimum load resistance.
- If the two resistors are substantially different, make the smaller one $1/20$ of the minimum load resistance.

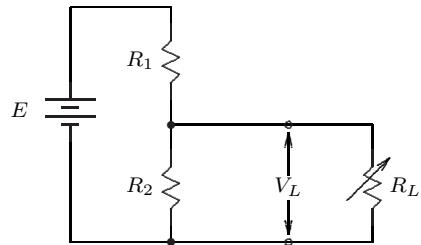
Example, Simple Method

Redesign the previous voltage divider, using the rules-of-thumb. Determine the load voltage in the completed design for minimum and maximum values of the load resistance, 10k to 100k.

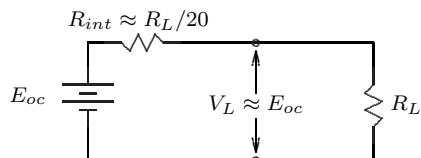
Solution, Design

- The minimum load resistance is $10\text{k}\Omega$. For an output voltage of 5 volts from a 12 volt supply, the voltages across the two resistors are 7 and 5 volts.
- The resistors are in the ratio of their voltage drops, so the resistors will be more the same than different. Consequently it is reasonable to treat them as *approximately equal*.
- Applying the rule that the one of the resistances should be approximately $R_L/10$, choose

$$R_2 = R_L/10 = 1\text{k}\Omega.$$

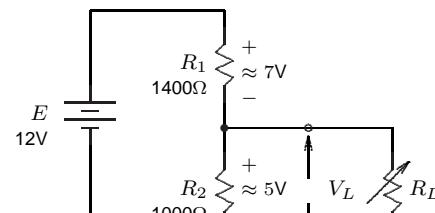


(a) Divider Circuit

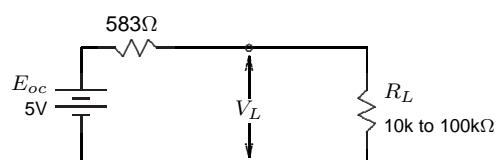


(b) Equivalent

Figure 81: Voltage Design Rule
a rule of thumb that could be applied quickly to generate a voltage divider design.



(a) Divider Circuit



(b) Equivalent

Figure 82: Voltage Divider, Approximate Method of Design

- The resistors are in the ratio of their voltage drops, so

$$\begin{aligned} R_1 &= \frac{V_{R1}}{V_{R2}} R_2 \\ &= \frac{7}{5} \times 1000 \\ &= 1k4\Omega \end{aligned}$$

The completed design is shown in figure 82(a). As you can see, the procedure is a lot simpler than the exact method.

Solution, Analysis

Now we will determine the performance of this voltage divider design. First we determine its Thevenin equivalent circuit. Then we determine the effect of attaching the two load resistances.

- The open circuit voltage is given by

$$\begin{aligned} E_{oc} &= E \frac{R_2}{R_1 + R_2} \\ &= 12 \times \left(\frac{1000}{1400 + 1000} \right) \\ &= 5 \text{ volts} \end{aligned}$$

- The internal resistance is given by

$$\begin{aligned} R_{int} &= R_1 \parallel R_2 \\ &= 1400 \parallel 1000 \\ &= 583\Omega \end{aligned}$$

The equivalent circuit is shown in figure 82(b). Now we will determine the effect of the load resistances.

With a 10k load resistance,

$$\begin{aligned} V_L &= E_{oc} \left(\frac{R_L}{R_L + R_{int}} \right) \\ &= 5 \left(\frac{10000}{10000 + 583} \right) \\ &= 4.72\text{V} \end{aligned}$$

This is an error of about 5%, as predicted.

With a 100k load resistance,

$$\begin{aligned} V_L &= E_{oc} \left(\frac{R_L}{R_L + R_{int}} \right) \\ &= 5 \left(\frac{100000}{100000 + 583} \right) \\ &= 4.97\text{V} \end{aligned}$$

This error is much less than 5%.

We can expect the load voltage to vary between 4.72 and 4.97 volts, depending on the load resistance. In many designs, that variation would be completely acceptable.

3.4.3 Graphical Method: Divider Design with Standard 5% Resistors

When using 5% tolerance resistors to construct a voltage divider a certain amount of trial-and-error choice of the resistor values is required, in order to get as close as possible to the required voltage ratio. The *Voltage Divider Nomograph* in figure 83 speeds up this process.

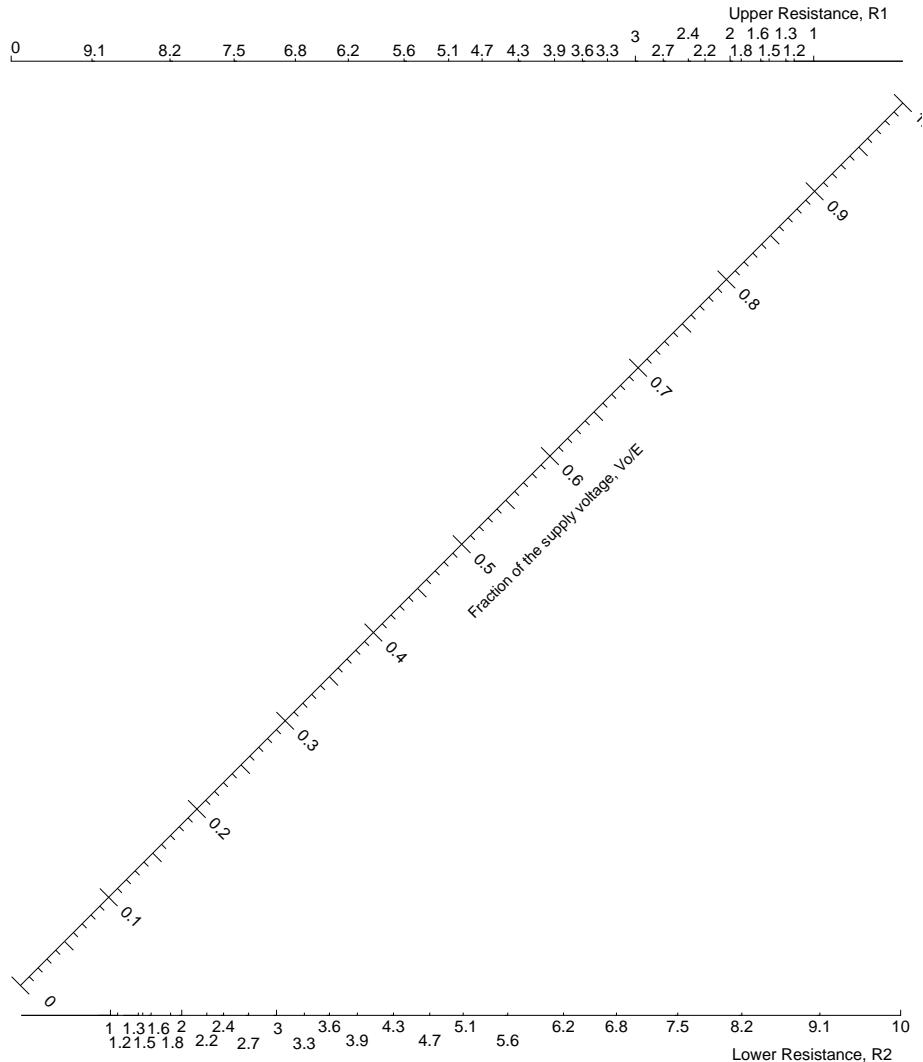


Figure 83: Voltage Divider Nomograph

To use the nomograph, determine the required voltage division ratio V_o/E and identify that point on the diagonal scale. Then choose one of the resistor values from the upper or lower horizontal scale. Draw a straight line through these two points and extend it until it intersects with the opposite scale. The intersection points on the two horizontal scales are the divider resistor values.

For example, if you choose equal resistor values on the upper and lower resistance scales and connect them by a line, the line will pass through $k = 0.5$ on the voltage division ratio scale.

The standard 5% values are marked on the resistance scales. It's a simple matter to experiment with different standard resistance values to obtain a particular value of voltage division ratio, and one can experiment with different pairs of values to see the effect on the division ratio.

Example

A 12 volt supply must be divided down to a voltage of 2.1 volts. The internal resistance should be in the region of $10\text{k}\Omega$. Use the nomograph to identify suitable resistor pairs.

Solution

The voltage division ratio k is:

$$\frac{V_o}{E} = \frac{2.1}{12} = 0.175$$

Using the nomograph, some possible pairs of resistors are

R_1	R_2	Voltage Division Ratio V_o/E	Output Voltage V_o for $E = 12\text{V}$
6.2	1.3	0.173	2.08
7.5	1.6	0.176	2.10
10	2.2	0.18	2.16

The pair 7.5, 1.6 yields the closest result to the required 2.1 volts. If the tolerance is not critical and one of these other values is used elsewhere in the design, that pair might be a suitable choice. Choosing $75\text{k}\Omega$ and $16\text{k}\Omega$ gives an internal resistance of $13\text{k}\Omega$.

Limitations

The nomograph of figure 83 is usable when the voltage ratio is between 0.05 and 0.95, which is the most common range of values. Outside that range, you'll still need to do the calculation.

3.4.4 Large Ratio Divider Shortcut

When the output voltage is very small compared to the input voltage⁵⁶, the upper resistor R_1 will be much larger than the lower resistor R_2 . In that case the exact value of R_2 will have very little effect on the divider current. Then the divider current may be treated as a constant value equal to E/R_1 . Then the output voltage may be calculated by multiplying the divider current by the value of the lower resistor R_2 .

Example

A voltage divider is required that will produce 10mV from a supply of 12V. The internal resistance should be about 100Ω . You may assume the load current is negligible compared to the divider current. Design the divider.

Solution

⁵⁶This design shortcut would also apply where the divider output voltage is very nearly equal to the supply voltage, but that's not a common design scenario.

Since the lower resistance is very small compared to the upper resistance, the lower resistance will define the divider internal resistance. Therefore choose the lower resistance R_2 as 100Ω . Assuming that the load current is negligible compared to the divider current, the divider current is:

$$\begin{aligned} I_D &= \frac{V_o}{R_2} \\ &= \frac{10 \times 10^{-3}}{100} \\ &= 100\mu\text{A} \end{aligned}$$

To create this divider current, the upper resistance should be approximately

$$\begin{aligned} R_2 &\approx \frac{E}{I_d} \\ &= \frac{12}{100 \times 10^{-6}} \\ &= 120\text{k}\Omega \end{aligned}$$

The completed design is shown in figure 84.

3.4.5 Derivation of the Nomogram

The nomograph is a useful tool for exploring various design alternatives, and so it is helpful to understand the derivation of the voltage-divider nomograph of figure 83.

Referring to figure 85(a), for the voltage divider we have:

$$\frac{V_{R1}}{V_{R2}} = \frac{R_1}{R_2} \quad (127)$$

The basic nomograph is pair of similar right-angle triangles, as shown in figure 85(a). If we make the horizontal components of the triangles proportional to R_1 and R_2 , then

$$\frac{A}{B} = \frac{R_1}{R_2} \quad (128)$$

Combining equations 127 and 128, we have that

$$\frac{A}{B} = \frac{V_{R1}}{V_{R2}} \quad (129)$$

For the power supply voltage E we can write

$$E = V_{R1} + V_{R2} \quad (130)$$

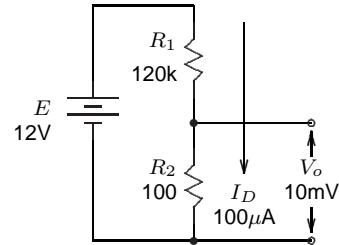
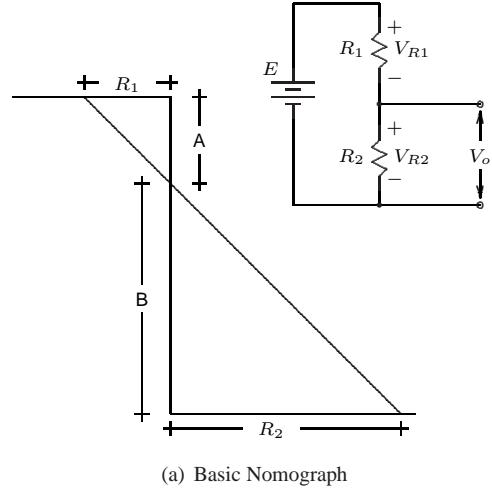
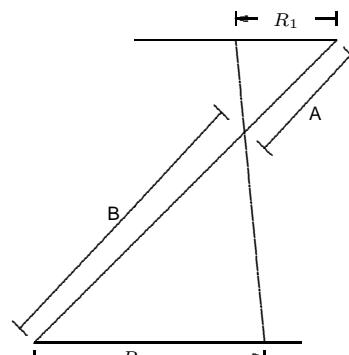


Figure 84: Large Ratio Divider



(a) Basic Nomograph



(b) Folded Version

Figure 85: Nomograph Derivation

Combining equations 129 and 130, we can obtain

$$\frac{V_{R2}}{E} = \frac{V_o}{E} = \frac{B}{A + B} \quad (131)$$

In words, the vertical line represents the fraction of the output voltage compared to the supply voltage.

This is a functional nomograph, but it can be made more compact by folding 85(a) as shown in figure 85(b). It's not as obvious, but the folded version consists of two similar triangles, so the same relationships apply. This is the version that is used in figure 83.

3.5 The Current Divider

A *current divider* consists of two or more parallel resistors that divide the current into proportional streams. A two-resistor circuit is shown in figure 86.

Referring to figure 86

$$I_1 = I \times \frac{R_2}{R_T} \quad (132)$$

$$I_2 = I \times \frac{R_1}{R_T} \quad (133)$$

where

$$R_T = R_1 + R_2 \quad (134)$$

The output current in R_2 is proportional to R_1/R_T , and vice versa. As in the case of the voltage divider, the absolute values of the resistances do not affect the output currents, it is their ratio which is important.

For example, if the two resistors are equal, $I_1 = I_2 = I/2$, independent of the values of the resistances R_1 and R_2 .

In practice, the current divider is much less common than the voltage divider.

3.5.1 Current Divider Application: Meter Characterization

In section 2.16.2 (page 68) we showed a technique for measuring the full scale current I_{FSD} and resistance R_m of a moving coil meter. That technique obtained the meter resistance by operating the meter at its full scale current while measuring the voltage across it. As mentioned then, the meter current and meter resistance are both small, so the voltage across the meter movement is very small. The measurement requires a millivolt-scale voltmeter.

Such a voltmeter may not be available, in which case a *current divider* technique can be used. The procedure is illustrated in figure 87.

First, as shown in figure 87(a), we connect the meter movement in series with an adjustable resistor R_{FSD} and voltage supply E . We adjust resistor R_{FSD} until the

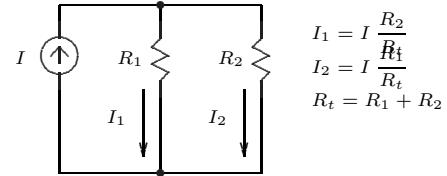


Figure 86: Current Divider

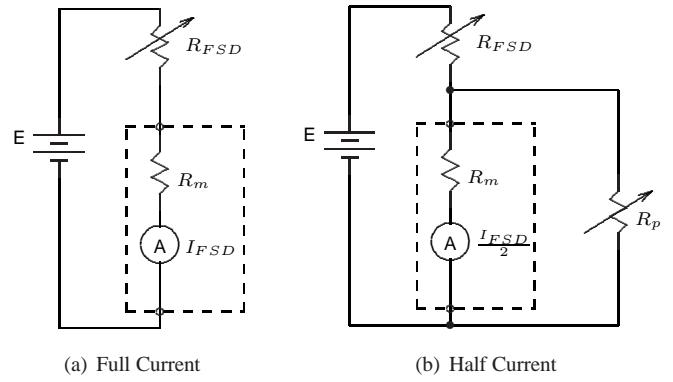


Figure 87: Meter Characterization

meter deflects full scale. Then, ignoring the meter resistance (because it is small compared to the adjustable resistance), we use Ohm's law to calculate the full scale current. It is equal to the supply voltage E divided by the setting of the adjustable resistance R_{FSD} . (We could equally well use a fixed resistor for R_{FSD} and an adjustable supply for E .)

Next, as shown in figure 87(b), we connect adjustable resistor R_p in parallel with the meter movement. Without altering the supply E or series resistor R_{FSD} , we adjust the parallel resistor R_p until the meter deflects *half* its full scale value. The value of the parallel resistor R_p is now equal to the meter resistance R_m . This works because the meter resistance R_m and parallel resistance R_p are much smaller than the series resistor R_{FSD} . Then voltage supply E and series resistor R_{FSD} act as a constant-current source. When the meter resistance R_m and the parallel resistance R_p are equal, they split the full scale current and the meter movement reads half its full scale value.

This method assumes that one has two calibrated variable resistances. Failing that, one could use fixed resistors, trial and error, and a lot of patience.

Other Examples

The one ampere current measuring circuit of figure 19 (page 70) is a current divider. The current in each of the two paths – one through the meter movement and one through the shunt resistance R_{SH} – is inversely proportional to their resistances.

The *pan-pot* shown in section 12.5 on page 322 is another application of the current divider.

3.6 The Superposition Theorem

The superposition theorem could be known as the *divide-and-conquer* theorem. It says that the electrical sources in a circuit may be treated independently. The effect of each source may be calculated separately and then the individual effects added to get the total effect of all sources.

When calculating the effect of an individual source, all other voltage sources and current sources should be replaced by their internal resistance. An ideal voltage source is replaced by a short circuit. An ideal current source is replaced by an open circuit.

The superposition theorem only applies to *linear* circuits. If there are non-linear elements in a circuit (diode, overdriven amplifier), then all bets are off.

3.6.1 Applications

An obvious application of the superposition theorem is to analyse a complex circuit by treating each source independently. An example is given below.

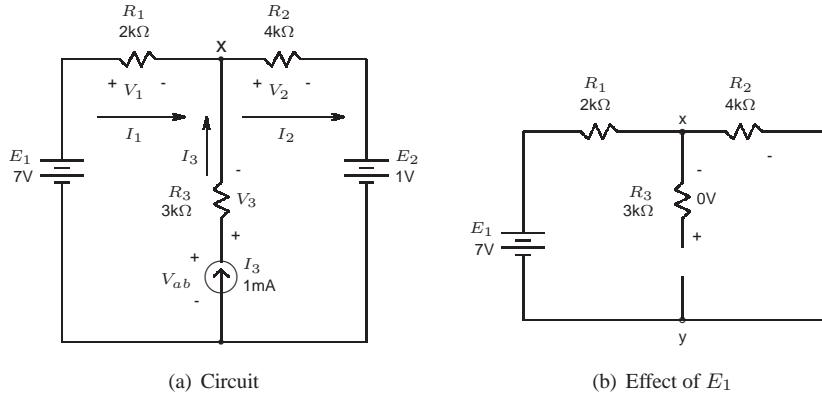
However, it is more important in its application to *biasing* in electronic circuits. Many electronic devices require a certain level of DC current or voltage in order to function correctly. This voltage or current is said to *bias the device*, or that it *places the device into its operating region*.

Many signals of interest to us are alternating: that is, they have a positive and negative excursion around some average value. To amplify this signal, it must be added to the bias signal, passed through the device, and then have the bias removed. For this to work, the effect of the DC bias and the AC signal must be superimposed and then separated. Using the superposition theorem, the effect of the DC bias and the AC signal may be determined separately and then added together.

This is often treated as two separate investigations. The *DC analysis* establishes the bias point and the *small-signal AC analysis* determines the gain (of an amplifier, for example.)

Example

We will now redo the example of figure 11 (section 2.10, page 64) using superposition. This involves drawing many diagrams, but the calculations are simple. The diagram is reproduced in figure 88(a):



(a) Circuit

(b) Effect of E_1

Figure 88: Superposition Theorem Example

Solution

The effect of E_1 is shown in figure 88(b): Voltage source E_2 is short-circuited and the current source I_3 is open-circuited. Now there is no current through R_3 , so its voltage drop is zero. Then R_1 and R_2 function as a voltage divider so

$$\begin{aligned} V_{xy} &= V_{ab} \\ &= E_1 \left(\frac{R_2}{R_1 + R_2} \right) \\ &= 4.66 \text{ volts} \end{aligned}$$

The effect of E_2 is shown in figure 89(a): This time, voltage source E_1 is short-circuited and the current source I_3 is again open-circuited. R_1 and R_2 again function as a voltage divider so

$$\begin{aligned} V_{xy} &= V_{ab} \\ &= E_1 \left(\frac{R_1}{R_1 + R_2} \right) \\ &= 0.33 \text{ volts} \end{aligned}$$

The effect of the current source I_1 is shown in figure 89(b): This time, both voltage sources are short-circuited. R_1 and R_2 appear in parallel and the combination is in series with R_3 .

$$\begin{aligned} V_{ab} &= I_1(R_1 \parallel R_2 + R_3) \\ &= 4.33 \text{ volts} \end{aligned}$$

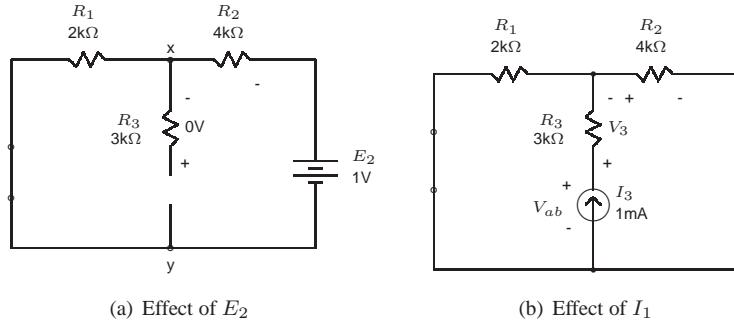


Figure 89: Superposition Example, cont'd

Finally, we add up these effects to get the resultant value of V_{ab} :

$$\begin{aligned} V_{ab} &= 4.66 + 0.33 + 4.33 \\ &= 9.3 \text{ volts} \end{aligned}$$

3.6.2 Superposition and Power

The superposition theorem applies to voltages and currents, but not power. You cannot calculate the power due to each voltage and current source, and then add these individual powers together to get the total.

Figure 90 illustrates this. The circuit consists of two equal voltage sources, e_1 and e_2 . Using superposition, the clockwise current due to e_1 is

$$I_{cw} = \frac{E}{R}$$

Similarly, the counter-clockwise current due to e_2 is

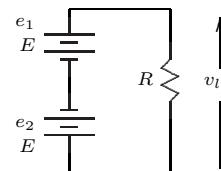


Figure 90: Superposition and Power

$$I_{ccw} = \frac{E}{R}$$

These two currents are in opposite directions and consequently cancel, so the net current in the circuit is zero. Then the power in the load resistor is

$$\begin{aligned} P_l &= I^2 R \\ &= 0 \times R \\ &= 0 \text{ watts} \end{aligned}$$

One might reason that the current due to e_1 causes power in the resistor, as does the current due to e_2 , and the total power in the resistor is the sum of these two. Clearly this is incorrect.

In general, for a circuit of the type shown in figure 90, we could write

$$\begin{aligned}
 P_l &= \frac{v_l^2}{R} \\
 &= \frac{(e_1 + e_2)^2}{R} \\
 &= \frac{(e_1^2 + 2e_1e_2 + e_2^2)}{R} \text{ watts}
 \end{aligned} \tag{135}$$

If we substitute $+E$ for e_1 and $-E$ for e_2 in equation 135, then the resultant power in the load is zero, which is the correct answer. The crossproduct term $2e_1e_2$ in equation 135 makes it incorrect to sum up the individual powers to obtain the total.

3.7 The Passive Adder

A *passive adder* circuit is a useful application of the superposition theorem.

Consider the following problem: We wish to add together two voltage sources so that the sum is some function of the two of them. This is to be done with passive components only (no transistors or op-amps, for example).

When both sources are AC signals e_1 and e_2 , then the circuit is referred to by audio designers as a *mixer*⁵⁷, although *adder* would be more accurate.

The same circuit may be used to add an AC signal e_1 to a DC signal E_1 so that the result is an AC signal riding on a DC offset of some magnitude.

Solution

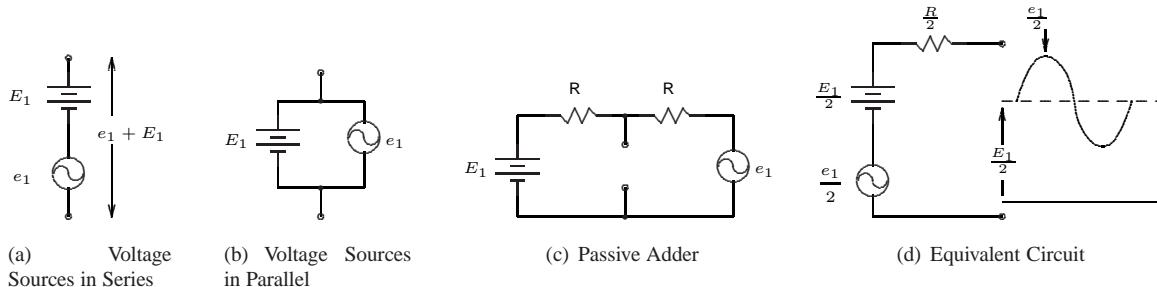


Figure 91: The Passive Adder

A simple solution is to wire the two sources in series, figure 91(a). If it's possible this will work well. However, one of the sources must have both terminals disconnected from ground (a *floating* source). Unfortunately, both sources usually have one terminal connected to the ground, so neither source can be floated in this fashion.

Notice that the sources cannot be connected in parallel as in figure 91(b), because each acts as a short circuit for the other, and the result is infinite current out of each source⁵⁸.

⁵⁷The term *mixer* is also used in Radio Frequency engineering, where it refers to a circuit that multiplies two signals together. The result is quite different from adding signals.

⁵⁸This raises the interesting question of whether it is allowable to connect batteries in parallel to increase the available current. This is common practice with certain types of batteries. However, it works because the batteries have internal resistance which limits the flow of current out of one battery into the other. It also assumes that both batteries are approximately equal open-circuit voltage. Otherwise, the transfer current between batteries can be very large. In the case of nicad batteries, which have very low internal resistance, it is not recommended to connect them in parallel, precisely because this transfer current is excessive.

A working passive mixer circuit, which adds the two voltages, is shown in figure 91(c).

The two resistors are often made equal, as shown in the diagram. Then the net effect is to create a source which has an open-circuit voltage half the sum of the sources, with an internal resistance equal to half the resistance value, figure 91(d).

The resistor value is dictated by the desired value of internal resistance of this source. A larger value reduces the wasted current that each source sends through the other. On the other hand, if the circuit must supply a finite value of load resistance R_L , then the internal resistance $R/2$ must be chosen to be much less than R_L .

3.8 The Wheatstone Bridge

The *Wheatstone Bridge* illustrates some techniques of circuit analysis and is a useful circuit in its own right. The circuit is shown in figure 92.

The bridge may be viewed as two voltage dividers, R_1, R_2 , and R_3, R_4 , both powered from the same source E . A galvanometer G bridges the output of the two dividers. The galvanometer is a sensitive centre-zero ammeter⁵⁹.

The bridge is said to be *balanced* when the current through the galvanometer is nulled (zero). We'll determine the relationship of the resistances for that to be true.

To analyse the circuit, we'll convert both the voltage dividers into their Thevenin equivalents. The values of R_{int} are irrelevant for the moment. But when the bridge is nulled, the open-circuit voltages of the two voltage dividers must be equal. That is,

$$E \frac{R_2}{R_1 + R_2} = E \frac{R_4}{R_3 + R_4} \quad (136)$$

The E 's cancel out, which is nice, because that implies that the bridge balance condition is *independent of the power supply voltage*. In the early days of electronics, when battery power was unreliable, this was an important consideration. With some algebra we can put this equation into the form

$$\frac{R_2}{R_1} = \frac{R_4}{R_3} \quad (137)$$

Now suppose you had to make an accurate resistance measuring device. You have one standard resistor R_{std} which is somehow known to a high accuracy. You have an unknown resistor R_x . And you have some resistance wire, which will be used to construct R_1 and R_2 .

The resistance measuring setup is shown in figure 93.

Since the resistor R_2, R_1 is composed of a length of resistance wire, the ratio $\frac{R_2}{R_1}$ is the ratio of the two lengths of wire, one below the wiper (R_2) and the other above the wiper (R_1).

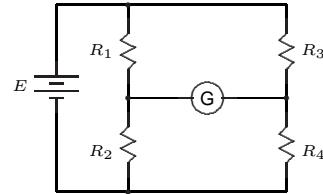


Figure 92: The Wheatstone Bridge

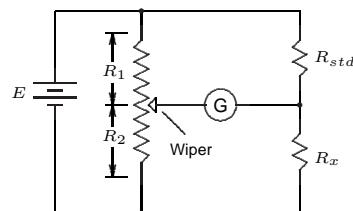


Figure 93: Measuring R_x

⁵⁹Before the days of inexpensive electronics, the output of the galvanometer was often equipped with a small mirror that rotated in response to the current. The rotation was effectively amplified with a light beam that reflected off the mirror and then onto a scale several meters away.

More precisely, the resistances are related to their length by equation 13 (page 58). The resistivity of the wire is ρ and cross-sectional area A .

$$R_1 = \rho \frac{L_1}{A} \quad (138)$$

$$R_2 = \rho \frac{L_2}{A} \quad (139)$$

Then

$$\begin{aligned} \frac{R_2}{R_1} &= \frac{\rho \frac{L_2}{A}}{\rho \frac{L_1}{A}} \\ &= \frac{L_2}{L_1} \end{aligned} \quad (140)$$

Substituting R_x for R_4 and R_{std} for R_3 in equation 137, we can find the expression for R_x is:

$$R_x = R_{std} \frac{R_2}{R_1} \quad (141)$$

Since R_{std} is known, an accurate measurement of R_x can be performed by measuring the ratio of two physical lengths. And the value of E is irrelevant.

In practice, a resistance would be placed in series with the galvanometer to limit the current through it when the bridge is unbalanced. As the bridge is moved closer to balance, this resistance is reduced to increase the sensitivity of the null measurement.

3.9 Load Line Application: Digital Logic Interface

It is a common requirement in electronics that a logic signal from a digital gate or microprocessor must drive some external device, such as a light-emitting diode (section 27.1), relay (section 2.31) or motor (section 9)⁶⁰.

This application can be analysed with tools from the previous section: voltage source, internal resistance, load-line, voltage divider. The analysis tells us something of importance in digital circuit design: will the circuit operate correctly when connected to some load resistance?

For our purposes here, we need understand the following:

- A logic gate produces a two-level signal representing digital signals zero and one. Any voltages will do, as long as they are sufficiently different to be distinguished. A common representation is +5V and 0V, referred to as *logic high* and *logic low* respectively..
- In order to work reliably, the output voltages of a logic gate are made slightly larger than the detection thresholds at the input of a logic gate. For example, in a 5V/0V system, the input detection thresholds might be set to recognize anything above 4.5V as logic high and anything below 0.5V as logic low.
- Suppose the output of gate A drives the input of gate B. We wish to attach some circuit to the output of gate A. In order that gate B continue to recognize logic high and logic low correctly, the added circuit must not reduce the magnitude of the signal below the detection thresholds of gate B.

⁶⁰Here we have an example of a digital circuit being analysed using analog techniques. At some level, *all* circuits are analog and need to be understood as such. The analysis and design of digital circuits can often proceed at a level of abstraction that hides the analog behaviour, and that's fine. But there are times – such as this example – when an analog understanding of the circuit is required.

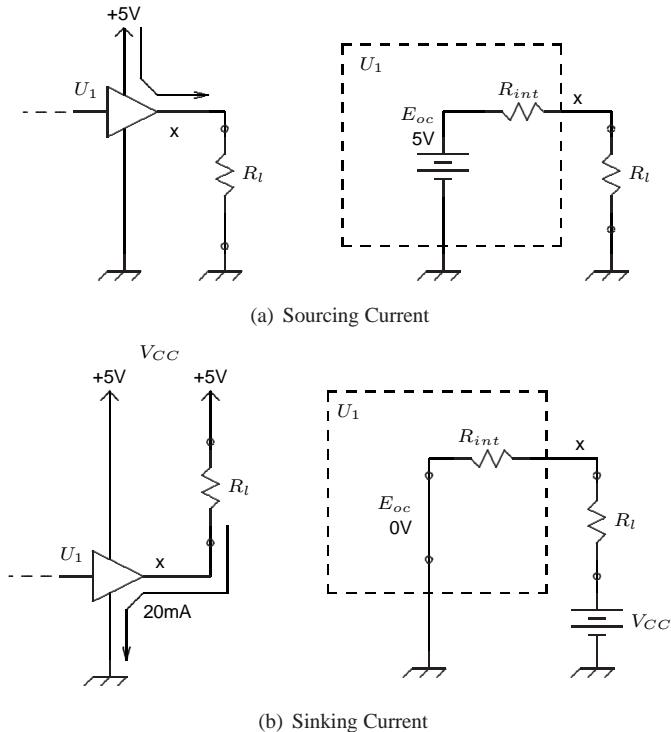


Figure 94: Logic Device Equivalent Circuit

The operation of a logic gate is illustrated in figure 94. (The logic gate is standing in for a wide variety of possible digital circuits – this could be the output of a microprocessor).

The logic gate has two possible output states:

- In the *output high* state, figure 94(a), the logic device behaves as a source with an open-circuit voltage E_{oc} equal to the supply voltage (in this case, 5 volts). If an external load R_l is connected to ground, current from the power supply flows through the device internal resistance and out of terminal x.
- In the *output low* state, figure 94(b), the open-circuit voltage E_{oc} becomes zero. If an external supply is connected as shown, current flows into terminal x, through the internal resistance R_{int} and to ground.

Modelling the Logic Device

The model for a logic device is determined from the device datasheet. It may be provided in tabular format or in the form of a graph. In general, all the devices of a particular logic family (such as CMOS or TTL logic) will have the same characteristic⁶¹. Here is an example.

⁶¹CMOS: Complementary Metal Oxide Semiconductor. TTL: Transistor-Transistor Logic. These terms and the details of the corresponding logic family do not concern us here since we are treating the logic device as a circuit model. All we need to know is in the behavioural characteristics.

Logic Output Characteristic: 4000 Series CMOS

The information for this logic family is provided in the form of a table of electrical characteristics, figure 95 (adapted and edited from [50]). (By convention, negative current indicates current flowing out of the device, positive current into the device.)

In this case, the figures indicate that the gate open circuit voltage switches between 0 and +5 volts. The internal resistance is such that a source or sink current results in a voltage drop of 0.4 volts at a current of 0.8mA, for an internal resistance of 500Ω . The internal resistance is symmetrical: it is the same value whether the gate is sourcing or sinking current.

The equivalent circuit of a 4000 series CMOS gate is therefore as shown in figure 95.

Resistance Load

When the load resistance is much larger than the internal resistance of the logic device, then the situation is very simple: the output voltage is essentially equal to the open-circuit voltage.

Consider figure 96(a), where the output of one logic gate drives the input of another logic gate. When the output of one CMOS logic gate actuates the input of another one, the load may be regarded as *high resistance*, and the output voltage of the first device will swing between zero volts and the supply voltage. In figure 96(a), the output voltage at point x would switch between 0 and V_{CC} volts. This signal would correctly operate the second gate U_2 .

Now consider the case shown in figure 96(b), where a load resistance R_l has been added at the output of U_1 . (This might be done to add an indicator lamp at that point in the circuit.)

In this case, the load resistance R_l forms a voltage divider with the internal resistance R_{int} of the first logic gate. At point x the low state voltage is unaffected but the high state voltage will be some fraction of the gate open-circuit voltage. For example, if the load resistance is equal to the internal resistance, then the high state output voltage of the first gate is *half* of E_{oc} . It is likely that this will prevent the second gate from working correctly.

Condition	Open Circuit Voltage	Output Current
Source	5V	0mA
	4.6V	-0.8mA
Sink	0V	0mA
	0.4V	+0.8mA

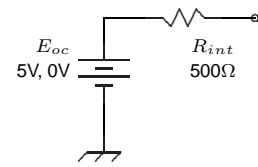
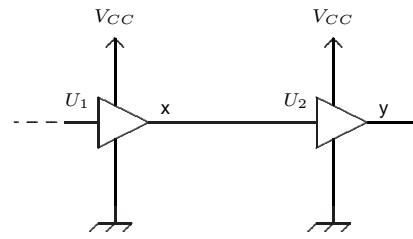
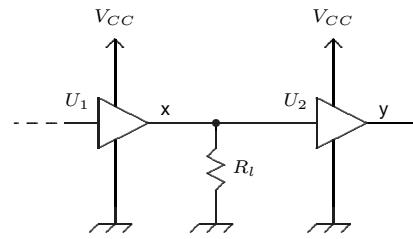


Figure 95: 4000 Series CMOS Gate



(a) Logic Gates



(b) Added Resistor

Figure 96: The Effect of a Load Resistor

Asymmetrical Drive: LSTTL Logic

Digital logic devices have certain functions such as the *OR* gate, *inverter* or *flip-flop*. These functions can be implemented in various ways – using relays or transistors, for example. A specific implementation of logic gates is known as a logic *family*. The logic devices in a particular logic family will have similar speed, power consumption and input-output characteristics.

When a logic device is used to drive some external device, we need to determine whether it can produce sufficient voltage and current to work reliably. That can be done by treating the logic device as an equivalent circuit, with an open-circuit voltage in series with an internal resistance.

Certain logic families are asymmetrical in their output characteristics, and this affects how a circuit is interfaced. For example, the output characteristics of the LSTTL (Low Power Schottky Transistor-Transistor Logic) family are shown in graphical form in figure 97.

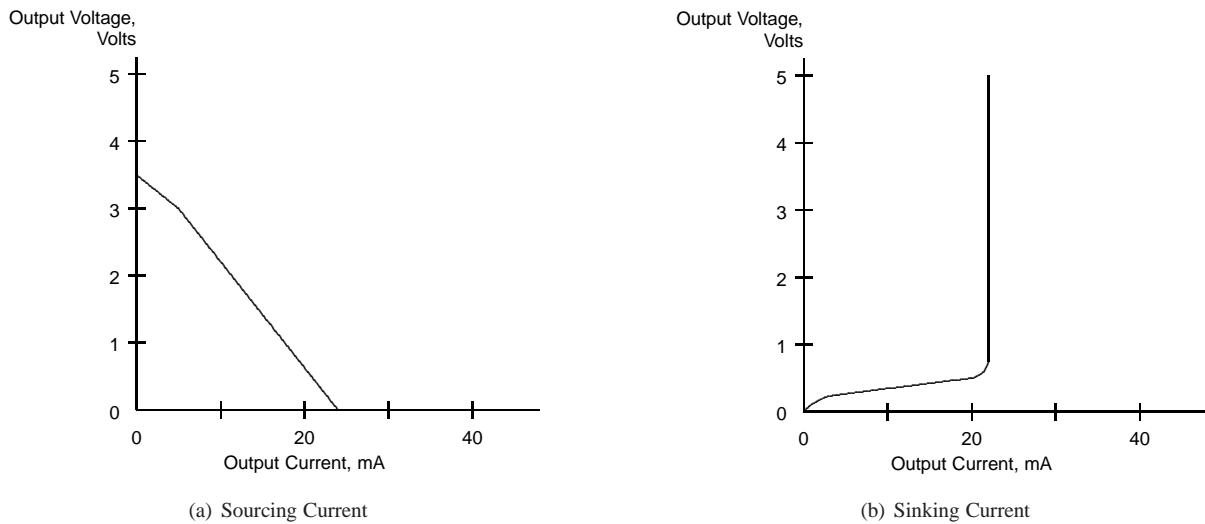


Figure 97: LSTTL Output Characteristic

The *sourcing* characteristic shows an open-circuit voltage of 3.5 volts and a short-circuit current of 22mA. The device could be modelled by an open circuit voltage of 3.5 volts in series with an internal resistance of 140Ω . The *sinking* characteristic approximates a constant-current sink of 22mA.

Assymetrical Drive Interface Example

Now let us consider that an LSTTL logic gate of the type shown in figure 97 is to be used to drive 20mA through a resistor. As shown in figure 98, there are two possible circuits. In figure 98(a), the logic device sources 20mA into the load resistor when its open-circuit voltage is high (3.5V). In figure 98(a), the logic device sinks 20mA through the load resistor, when its open-circuit voltage is low (0V).

Either of these circuits will work in some fashion, and put 20mA through the load resistance, as required.

However, assuming that the logic circuit can be arranged to generate either an output-high or output-low state when pumping current through the load resistor, is one of these circuits to be preferred? If so, which one and why?

For this particular family, the arrangement of figure 98 is likely to be preferable. There are two reasons, both with respect to the output voltage of the logic gate.

- In the case of the sourcing circuit, figure 98(a), the open circuit voltage is 3.5 volts and the output voltage V_x is about 0.6 volts. This is well below the amount necessary to run a subsequent logic device. (For the LSTTL logic family, at least 2.5 volts is required to drive another logic device.) So this logic signal is too small to drive additional logic devices.

Furthermore, a common requirement of this output voltage is that it exceed the threshold voltage of some semiconductor device like a diode, base-emitter junction of a transistor, or light-emitting diode. These threshold voltages range from about 0.6 volts to 2.5 volts. The output voltage is insufficient to actuate these devices.

- In the case of the sinking circuit, figure 98(b), the output voltage V_x is again about 0.6 volts. However, in this case, it is supposed to be recognized as a logic low signal. The LSTTL family requires a logic low output be no larger than 0.6 volts, so this output voltage is marginally satisfactory. Consequently, the gate could drive a subsequent logic device.

As well, the voltage across the load resistance is 4.4 volts, and this substantially exceeds the threshold voltage of many semiconductor devices. For example, a light-emitting-diode requires a threshold voltage of about 2 volts. It could be placed in series with the load resistor and would then operate correctly. (The load resistance would be adjusted to suit.)

Summary

- A logic device may be treated as an electrical source. Then the effect of connecting a load resistance or other device may be determined.
- In some cases, the logic gate may be modelled by a source with an open circuit voltage and fixed internal resistance.
- In other cases, the output characteristic of the logic gate may be non-linear and/or asymmetrical in the source and sink state. In that case, the load-line technique may be used to analyse the circuit and design the load.
- A logic gate can be operated in the sourcing condition, in which case the load current flows out of the logic gate, through the load, and to ground.

It may also be used in the sinking condition, in which case the logic current flows from some positive supply, into the logic gate, and then to ground.

The best arrangement depends on the output voltage-current characteristics of the logic gate.

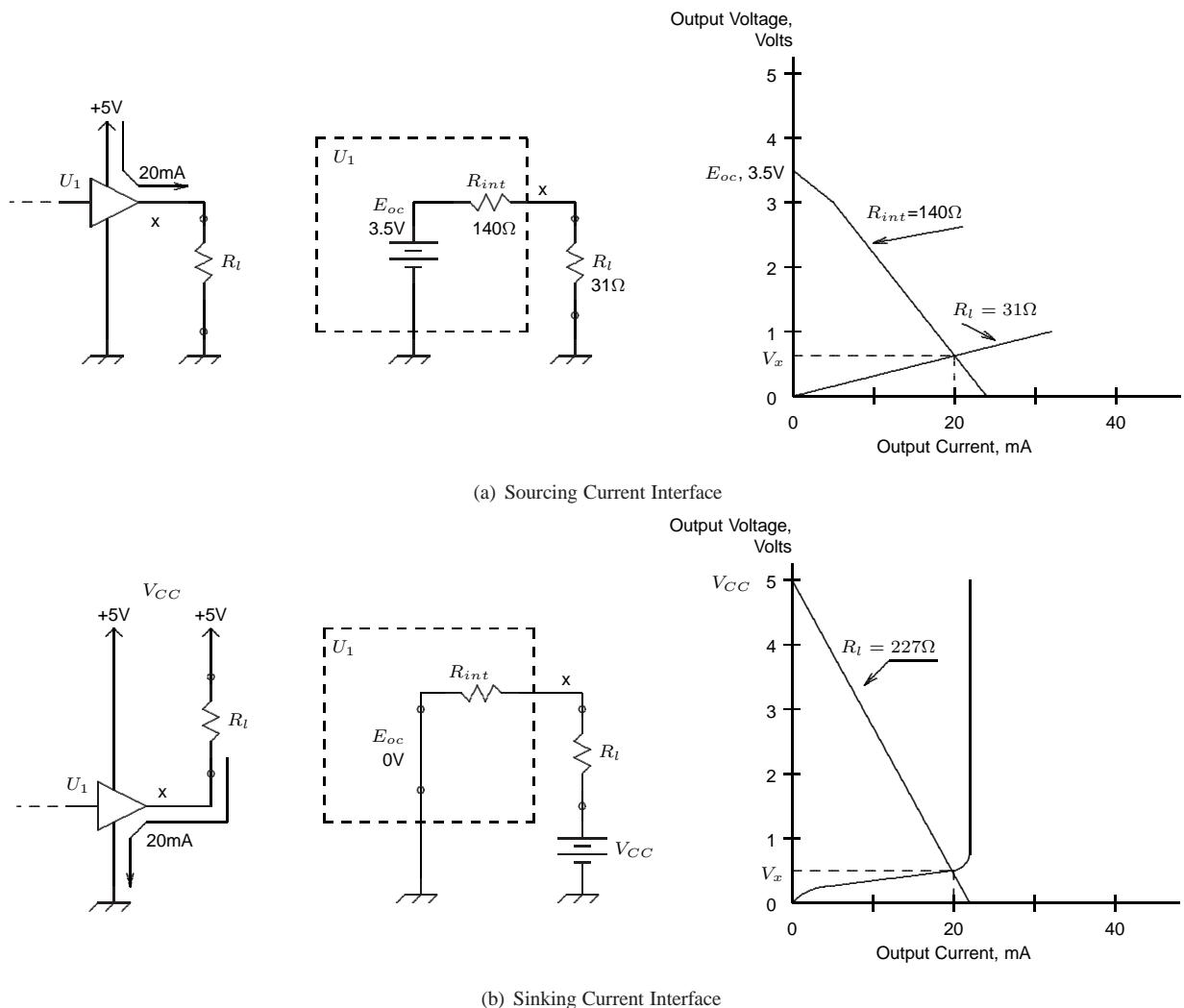


Figure 98: Interfacing Options: Source or Sink

3.10 Component Tolerances

As a result of unavoidable process variations in the manufacturing of electronic components, all components have a *tolerance* around their *nominal value*. For example, a 612Ω 1% resistor has a nominal value of 612Ω but could have any value in the range $612 - 1\% = 605.9\Omega$ to $612 + 1\% = 618.1\Omega$. Transistors have huge tolerances on their value of current gain. A range of 3:1 is not unusual. Some electrolytic capacitor values vary by hundreds of percent.

One of the tasks of the circuit designer is to ensure that the circuit will work with components that exhibit the full range of tolerances. For example, suppose a timing circuit relies on an electrolytic capacitor to establish the time interval. The capacitance value may vary over a 2:1 range. Then the time interval can be expected to vary over a similar range. If this is not acceptable (and it probably isn't), then the designer has to specify a timing capacitor with a tighter tolerance or provide another adjustment for the timing interval. Adjustments of any kind should be avoided in production equipment because they require the intervention of a human being or expensive robot, and because adjustable components are more expensive and less reliable than fixed components.

Unfortunately, one cannot assume any particular probability distribution of the tolerances. The only safe assumption is that all possible values within the tolerance range are equally likely⁶².

Furthermore, a circuit can suffer from *tolerance buildup*, where a particular combination of tolerance values can cause the circuit to malfunction. There are two common approaches to this:

- **Works under all component values.** In this scenario, the circuit operation is tested under all possible combinations of circuit tolerances and designed in such a way that no combination of tolerances can make it misbehave.

This can be expensive if the circuit must cope with a worst-case condition that is unlikely in practice.

One possible solution is to provide one adjustment that compensates for the combined tolerances of a group of components.

Another solution is to design the circuit so that it ignores certain tolerances. For example, a bridge circuit (section 3.8, page 138) tends to ignore variations in the power supply voltage, so tolerances in the power supply circuit are not critical.

- **Works under most component values.** In this case, the circuit designer simulates the circuit with many combinations of circuit tolerances, all chosen at random. If, say, all but one circuit in 1000 are likely to function properly then it may be most economical to scrap or rework the offending circuit.

3.11 Cumulative Tolerance

For circuits that are not too complex, it's useful to estimate the tolerance on the outcome. Here we look at some rules of thumb.

Multiplication of Variables

Suppose we wish to create a certain voltage as the result of a current passing through a resistor. The voltage is equal to the product of the current and resistance. If the current has a tolerance of $\pm a\%$ and the resistance has a tolerance of $\pm b\%$, what is the expected tolerance on the voltage? We'll now show that it's the *sum* of the current and voltage tolerances.

⁶²As an interesting example of this, suppose you are presented with a batch of resistors of 5% tolerance. Could you go through them and pick out the units that are within 1% of the nominal value? Possibly not. The manufacturer may have already done that, planning to sell the 1% units for a higher price. So the probability distribution would show a gap in the range of plus or minus 1% of the nominal value.

We wish to determine the worst possible case, which occurs when the individual tolerances are maximum and have the same sign. (If they have the opposite sign, they partially cancel.)

To begin with, we have

$$V = IR \quad (142)$$

where V, I and R are the nominal voltage, current and resistance.

Now assume that the tolerances are at their maximum positive values. Then we can write

$$V(1 + k) = I(1 + a) \times R(1 + b) \quad (143)$$

where k , a and b are the tolerances in decimal form. For example, if the worst case tolerance on current is 2%, then $a = 0.02$. Then

$$V(1 + k) = IR(1 + a)(1 + b) \quad (144)$$

Substitute V for IR on the RHS and then cancel V from both sides. We are left with:

$$\begin{aligned} 1 + k &= (1 + a)(1 + b) \\ &= 1 + b + a + ab \end{aligned}$$

and

$$k = b + a + ab \quad (145)$$

Since a and b are both small numbers, their product ab is even smaller and may be ignored. Then

$$k \approx a + b \quad (146)$$

That is, the resultant tolerance k is equal to the sum of the individual tolerances.

The other worst case occurs when both tolerances are negative. By a similar analysis,

$$\begin{aligned} k &= a + b - ab \\ &\approx a + b \end{aligned} \quad (147)$$

Consequently, we can say that a tolerance in current of $\pm a\%$ and a tolerance in resistance of $\pm b\%$ will create a tolerance in voltage of $\pm(a + b)\%$. For example, if the current tolerance is $\pm 3\%$ and the resistor tolerance $\pm 1\%$, then the voltage tolerance is $\pm 4\%$.

More generally, the tolerance of a product is equal to the sum of the multiplier and multiplicand tolerances.

Division of Variables

This may also be shown to be the case for a tolerance that is the result of dividing two toleranced quantities. For example, consider a current that is the result of a voltage divided by a resistance:

$$I = \frac{V}{R} \quad (148)$$

where again I, V, and R are the nominal current, voltage and resistance.

In taking tolerance into account, the worst case maximum current occurs when the voltage is at a maximum and the resistance at a minimum. In equation form:

$$I(1 + k) = \frac{V(1 + a)}{R(1 - b)} \quad (149)$$

Solving for k , we obtain:

$$\begin{aligned} k &= \frac{a+b}{1-b} \\ &\approx a+b \end{aligned} \quad (150)$$

providing $b \ll 1$. Once again, the tolerances add. Similarly, the worst case minimum current is approximately equal to the sum of the voltage and resistance tolerances.

Addition of Variables

Suppose we connect two resistors in series. Given the tolerances of the individual resistors, what is the tolerance on the combination?

Let's call the two resistors R_a and R_b , and their tolerances a and b as before. Then the total resistance R_t is the sum:

$$\begin{aligned} R_t(1+k) &= R_a(1+a) + R_b(1+b) \\ (R_a + R_b)(1+k) &= R_a + R_b + aR_a + bR_b \\ k &= \frac{aR_a + bR_b}{R_a + R_b} \end{aligned} \quad (151)$$

This is not particularly enlightening. However, assuming that the tolerances are equal (as they likely would be for two resistors in series) then the final tolerance is just equal to the tolerance of one resistor:

$$k = a \quad (152)$$

3.12 Complex Cases: Spreadsheet to the Rescue

The preceding rules are useful for a quick look at tolerances. However, most circuits have far too many variables and associated tolerances for a quick assessment of overall behaviour.

A more rigorous approach is a spreadsheet-based brute force calculation that determines all possible combinations of the component values. The possible combinations follow the same sequence as a binary counter. A three-variable example is shown in figure 99. An N variable table will have 2^N rows in the table. In figure 99, there are three variables V_0 , V_1 and V_2 , so there are $2^3 = 8$ rows. Now picture that the zero in this table represents some variable at the low extreme of its tolerance. A one represents that variable at its high extreme. Then each row is a possible combination of extreme tolerances.

A fourth column to this table would show the resultant circuit behaviour – some voltage, for example – that results from each combination of variables. Consequently the entries in this fourth column are an exhaustive list of all possible voltages. A scan down that column shows the minimum and maximum possible values of the voltage in a production of this circuit.

Once that is complete, it's time for some engineering judgement:

V_2	V_1	V_0
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Figure 99: Binary Sequence

- **Output is within tolerance.** Providing that these components do not exceed their allowed tolerances, the voltage of interest *cannot* exceed the limits specified on this spreadsheet. If those limits are acceptable, then the circuit has been toleranceed for production. We should never see a unit that fails because of tolerance variation of these components.
- **Output is slightly out of tolerance with some component tolerance combinations.** If the range of voltages is slightly outside the acceptable limits, the circuit may still be acceptable. It may be unlikely that all components will be simultaneously at this particular combination tolerance limits. Whether this is an acceptable decision depends on the application. For example, if an out-of-tolerance voltage endangers human life, then it's not acceptable.
- **Output is significantly out of tolerance for some component tolerances.** The circuit must be modified to use tighter-tolerance components or some manual method of adjusting the output.

Regardless of the outcome of this analysis, we are in control of the situation because we know exactly how the circuit will behave in response to component tolerances.

Example: Spreadsheet Tolerance Calculation

In this example, we'll show the use of a spreadsheet to determine the tolerance variation of the output voltage V_o in the circuit of figure 100(a), a voltage divider using 1% resistors. For this exercise the voltage supply E can be assumed to be exact.

Analysis

The voltage supply E and voltage divider resistors R_1 and R_2 of figure 100(a) can be collapsed into a Thevenin equivalent circuit of figure 100(b). The Thevenin open circuit voltage E_{oc} is given by:

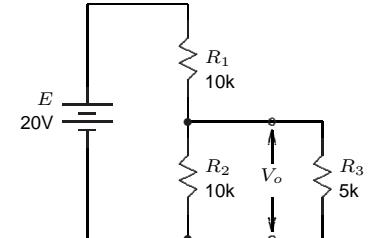
$$E_{oc} = E \times \frac{R_2}{R_1 + R_2} \quad (153)$$

If the supply E is 20 volts and the resistors are at their nominal values, then $E_{oc} = 10V$.

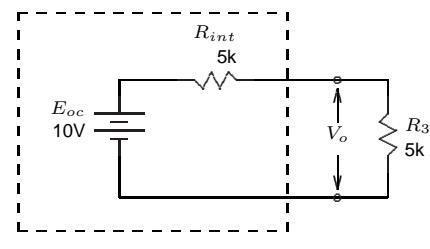
The Thevenin internal resistance R_{int} is equal to $R_1 \parallel R_2$. For nominal resistor values, $R_{int} = 5k\Omega$.

The Thevenized circuit of 100(b) is another voltage divider with equal resistors. So the output voltage V_o is nominally half of E_{oc} , 5 volts.

Now we are ready to throw this information into the spread-



(a) Original Circuit



(b) Thevenin Equivalent

R1	R2	R3	Eoc	Rint	Vo
10100	10100	5050	10	5050	5
10100	10100	4950	10	5050	4.95
10100	9900	5050	9.9	4999.5	4.97
10100	9900	4950	9.9	4999.5	4.93
9900	10100	5050	10.1	4999.5	5.08
9900	10100	4950	10.1	4999.5	5.02
9900	9900	5050	10	4950	5.05
9900	9900	4950	10	4950	5

(c) Spreadsheet

Figure 100: Tolerance Example

sheet, as shown in figure 100(c).

- The first three columns are resistors at their +1% and -1% tolerance values. They follow the pattern of figure 99 so that all possible combinations are represented.
- The value of E_{oc} is determined by plugging the values of resistors R_1 and R_2 , and voltage supply E , into equation 153. Once the first entry for the formula is determined, it can be copied and pasted into the rest of the column, and the formula entries will be adjusted to apply to each particular row.
- The value of R_{int} is determined by applying the parallel resistor formula to the values for R_1 and R_2 .
- Finally, a formula in the last column calculates output voltage V_o using a voltage divider equation with the entries for E_{oc} , R_{int} and R_3 .

A scan down the last column indicates that the output voltage could vary between a minimum of 4.93 volts and a maximum of 5.08 volts, which is a variation of -1.4% to +1.6%. This is greater than the individual component variation (1%) and it indicates why this type of analysis is important.

3.13 Larger Problems

In an exhaustive analysis like this the table size doubles with each additional variable, so it blows up with a large number of variables. However, some thought and understanding of the circuit operation can simplify the analysis.

In the above circuit, for example, we might reason as follows:

The maximum output voltage will occur under the following circumstances: R_1 minimum, R_2 maximum and R_3 maximum.

Similarly, the minimum output voltage will occur for the opposite condition: R_1 maximum, R_2 minimum and R_3 minimum.

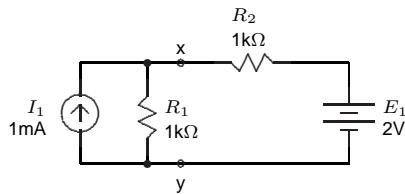
This reasoning process reduces the number of table entries from 8 to 2. In a large, complex circuit that type of thinking may be helpful.

3.14 Error Budget Analysis

The tolerances of the components may be regarded as *errors* that alter the output from its correct value. The output of the circuit may be further corrupted by noise or drift in amplifiers, leakage currents, variation in supply voltages and so forth. There is a *budget* for the total of the errors. The analysis of these effects is then referred to as an *error budget* analysis.

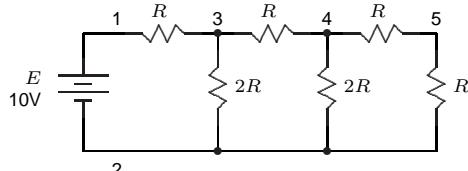
3.15 Exercises

1. Use Norton's theorem to calculate V_{xy} .



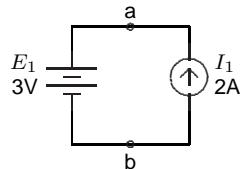
2. The circuit shown below is known as an *R-2R ladder network*. Determine each of the voltages in the circuit:

Voltage	Value
V_{12}	
V_{32}	
V_{42}	
V_{52}	



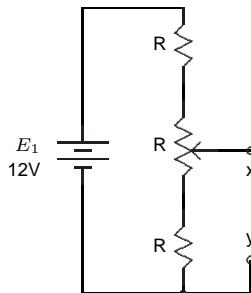
Hint: Start at the right end of the network and collapse it back toward the voltage supply.

3. For the sources in the diagram,



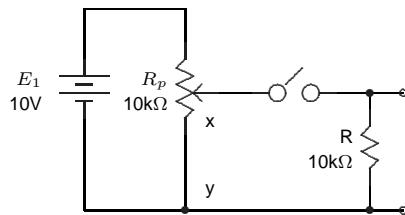
- (a) What is V_{ab} ?
- (b) How much power is dissipated in E_1 ?
- (c) How much power is dissipated in I_1 ?
- (d) What is the magnitude and direction of the current in E_1 ?

4. For the circuit shown, what are the maximum and minimum values of the potentiometer wiper voltage, with respect to point y , as the potentiometer is adjusted?



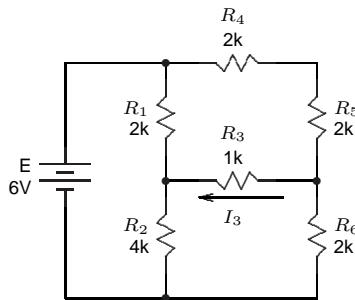
5. What is the value of the voltage V_{xy} when the wiper of the pot is at mid-position:

- (a) with the switch open?
- (b) with the switch closed?

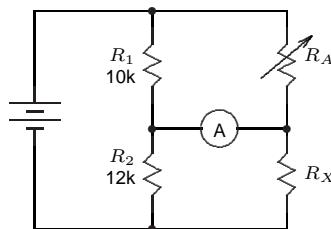


6. Determine the current I_3 through R_3 . The indicated direction of I_3 is assumed and may or may not be correct.

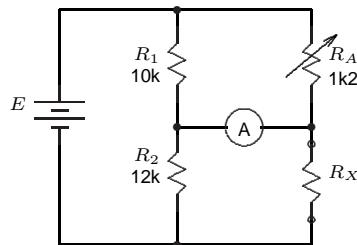
Hint: It should be possible to do this problem by redrawing the circuit, followed by perhaps 2 lines of math. It is not necessary to set up and solve simultaneous equations.



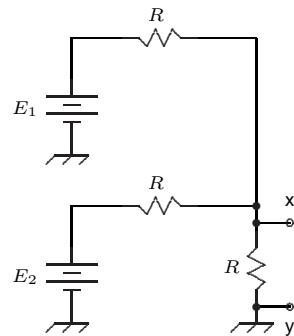
7. For the Wheatstone bridge shown below, the meter is a centre-zero type which can indicate the direction and magnitude of current flow. If the meter indicates current flow from left to right, how should R_A be adjusted: increased or decreased? Justify your answer.



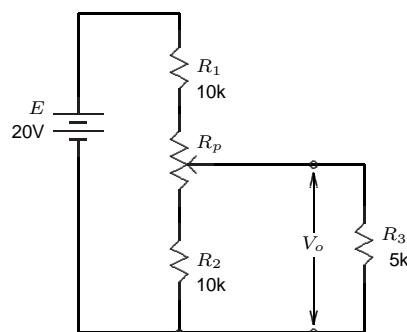
8. The Wheatstone bridge shown below is in balance. R_A is at its maximum setting. Determine R_X .



9. Use the superposition theorem to develop a mathematical expression for V_{xy} in terms of E_1 and E_2 .



10. Section 3.10 on page 145 showed a circuit where the output varied due to the tolerances of the components. The circuit diagram shows how the output could be adjusted using a potentiometer R_p .



Assuming that the supply E is exactly 20 volts and the resistor values vary by 1%, calculate a suitable value for the track resistance of potentiometer R_p if the output voltage V_o must be adjusted to exactly 5 volts.

Check your answer with a spreadsheet.

11. A Puzzle

You are given two black boxes, each with terminals A-B. One contains a Thevenin equivalent circuit of 10

volts in series with 1 ohm. The other contains a Norton equivalent circuit, 10 amperes in parallel with 1 ohm⁶³.

How could you tell the difference between them?

⁶³This puzzle was originally posed in Electronics Design or EDN magazine.

4 Tools for Circuit Analysis: Part 2

4.1 Alternating Current

The source shown in figure 6 (page 61) is a *direct-current* (DC) source⁶⁴. There are also voltage sources for which the output voltage changes with time. In general, these are referred to as *AC* for *alternating current* voltage sources. An AC voltage source is shown in figure 101.

In designing electronic circuits, we need some method of predicting the various voltages and currents in the circuit. For a DC circuit, this is relatively straightforward – Ohm’s law and Kirchhoff’s laws tell us everything we need to know⁶⁵.

However, it’s more complicated to calculate current and voltage in a circuit containing capacitors and inductors. The relationship of voltage and current is described by differential equations, such as

$$i_c(t) = C \frac{dV_c(t)}{dt} \quad (154)$$

for the capacitor. We’ll explain this in detail later. For the moment, it’s enough to observe that the current in a capacitor is a function of the time differential of the voltage across its terminals.

Differential equations are not easy to apply for circuit analysis. Fortunately, there are techniques that get around this difficulty, as we now explain..

4.1.1 Waveforms in General

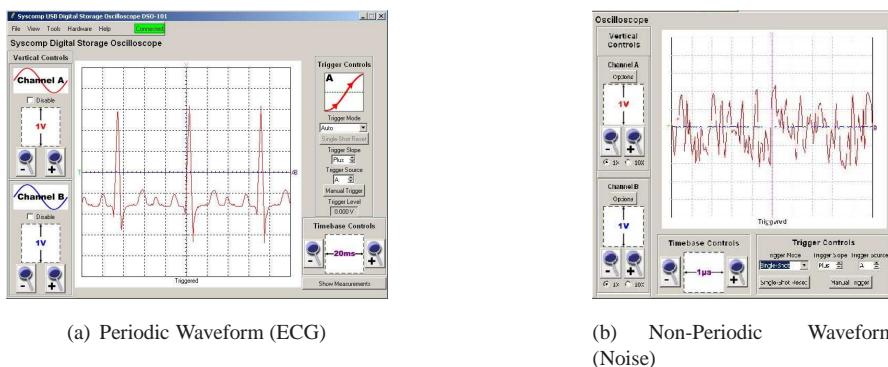


Figure 102: Waveforms

A *waveform* is the plot of some quantity against time. For example, the ECG (electrocardiogram) waveform shown in figure 102(a) is a plot of the electrical activity of a human heart, in volts, against time. The waveform is shown plotted on the screen of the *oscilloscope*, an electronic instrument for displaying such plots.

⁶⁴Sometimes referred to as a *zero frequency* (ZF) source.

⁶⁵In practice, we use various circuit analysis short cuts such as Thevenin’s theorem and The Superposition theorem, but these are based on Ohm and Kirchhoff.

A *periodic* waveform is one like the ECG that repeats on a regular basis. One complete repetition of the waveform is a *cycle*. The time between repetitions is the *period* in seconds per cycle (usually expressed in seconds). The *frequency* is the inverse of the period, in cycles per second, or Hertz (Hz)⁶⁶.

A non-periodic waveform does not repeat, as for example the noise waveform of figure 102(b)⁶⁷.

4.1.2 Sinusoidal Waveforms

For a variety of reasons, sinusoidal waveforms (*sine waves*) are popular in electrical engineering.

- When a sine wave is differentiated or integrated, the result is another sine wave – at a different amplitude and advanced or delayed in time, but at the same frequency and with the same shape.
- Sine waves are naturally generated and accepted by AC machinery. In North America the power line voltage is a sine wave with an amplitude of 165 volts and a frequency of 60Hz.
- A resonance, such as the vibration of mass-spring system, produces a sine wave of motion. Similarly, an electrical resonance produces a sine wave of voltage and current. As a consequence, many electronic oscillators naturally produce a sine waveform.
- Any periodic waveform which is not in a sinusoidal shape may be treated as the sum of various sinusoids, a process known as *Fourier Synthesis*.

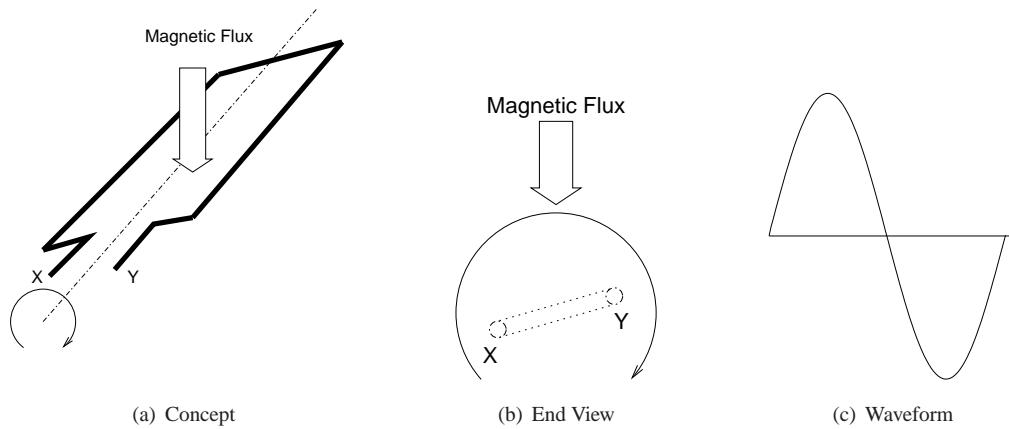


Figure 103: AC Generator

One method of producing a sine wave is shown in figure 103. A loop of wire turns in a stationary magnetic field. As the loop rotates a voltage is induced in each conductor, proportional (by Faraday's Law) to the rate of change of the magnetic flux swept out by that conductor.

The rotating coil is known as the *rotor*. The magnet that produces the stationary field is known as the *stator*.

When the coil is vertical, the conductors are moving horizontally and sweep out a significant flux area and the generated voltage is at a maximum. When the conductors are moving vertically, they sweep out no new flux area and the generated voltage is at a minimum. In between, the generated voltage varies with the angle of rotation. The polarity of the voltage reverses with each half-rotation.

⁶⁶Named after the German scientist Heinrich Hertz who in 1887 demonstrated the existence of radio waves predicted by Maxwell.

⁶⁷Waveforms captured from the Syscomp DSO-101 and CGR-101 oscilloscopes.

If the rotor is taken to be at zero degrees when it is horizontal then the output voltage is proportional to the sine of the rotation angle:

$$e(\theta) = E \sin \theta \quad (155)$$

where E is the maximum value of the sine wave and θ is the rotation angle in degrees or radians.

It turns out that angle measurement in degrees encumbers formulae with a factor of 2π . To avoid that, the angle can be measured in radians. Similarly, frequency can be measured in *radians per second*, in which case it is given the symbol ω .

This is convenient, because the angle in radians at any given instant is equal to the product of the radian frequency ω radians/second and the time t in seconds.

$$\begin{aligned} \theta &= \omega \frac{\text{radians}}{\text{second}} \times t \text{ seconds} \\ &= \omega t \text{ radians} \end{aligned} \quad (156)$$

The frequency in cycles per second (Hz) can be converted to radians per second, adapting the technique of section 1.2 on page 48. Recall that 2π radians equals 360 degrees (one cycle). Then we can write:

$$2\pi \text{ radians} = 1 \text{ cycle} \quad (157)$$

Divide both sides by *second*:

$$2\pi \frac{\text{radians}}{\text{second}} = 1 \frac{\text{cycle}}{\text{second}} \quad (158)$$

Multiply both sides of this equation by frequency of f , and interchange the LHS with the RHS:

$$f \frac{\text{cycle}}{\text{second}} = f \times 2\pi \frac{\text{radians}}{\text{second}} \quad (159)$$

That is, multiply frequency in Hz by 2π to convert it to radians per second.

Example

The frequency of the power system in North America is 60Hz.⁶⁸ What is the frequency of the power system in radians/second?

Solution

From equation 159 we have:

$$\begin{aligned} f \frac{\text{cycles}}{\text{second}} &= f \times 2\pi \frac{\text{radians}}{\text{second}} \\ &= 60 \times 2\pi \frac{\text{radians}}{\text{second}} \\ &= 377 \text{ radians/sec} \end{aligned}$$

⁶⁸The author can remember as a child when electrical power in Toronto was converted from 25 to 60Hz. An older engineering colleague of the author managed that conversion. A higher operating frequency reduces the required mass of steel in motors and transformers.

4.2 Vector Notation and Sine Waves

It is useful to represent a sine wave as the vertical component of a rotating vector, as shown in figure 104. Here we assume the waveform is a voltage and the length of the vector is E , the maximum value of the sine wave. In figure 104(a), the rotational angle of the vector is a measurement of the *phase angle* of the sine wave – in either radians or degrees.

In figure 104(b), the horizontal axis of the sine wave plot is time in seconds. The vector is rotating counterclockwise at a rate of f cycles per second or ω radians per second. The angle of the vector in radians is ωt , product of the radian frequency and elapsed time.

Figure 104(c) shows an example of vector representation. A sinusoidal voltage and current are shown in the upper half of the figure. The vector representation of the voltage and current is in the lower half of the figure. The current is *phase shifted* from the voltage and would be said to have a *lagging phase angle*. The phase angle between the voltage and current vectors provides an unambiguous description of their relationship.

4.3 Reactance

In a circuit with AC voltages and currents, we find that resistors behave in a simple manner: the current is proportional to the voltage and Ohm's law applies.

Capacitors and inductors behave differently. As in a resistor, the current is proportional to voltage. However, there are significant differences, and so these devices have the property of *reactance* – sometimes referred to as *AC resistance*.

- *Resistance* is constant with frequency and the current is in phase with the voltage. That is, the current and voltage vary in lock-step: as the voltage increases, so does the current.
- *Reactance* changes with frequency. In the case of capacitive reactance, we will see that it decreases at higher frequencies. Furthermore, the current through a capacitor is not in phase with the applied voltage. When we treat the voltage and current as rotating vectors, *the current leads the voltage by an angle of 90 degrees*.
- In an amazing coincidence, an inductor has precisely the opposite behaviour. Inductive reactance increases at high frequencies: an inductor is a short circuit at low frequencies and an open circuit at high frequencies. The current lags the voltage by 90 degrees.

Charles Steinmetz showed how reactance relates the magnitude and phase of the voltage and current in a capacitor or inductor: the voltages and currents are treated as phasor quantities that are affected by the reactance

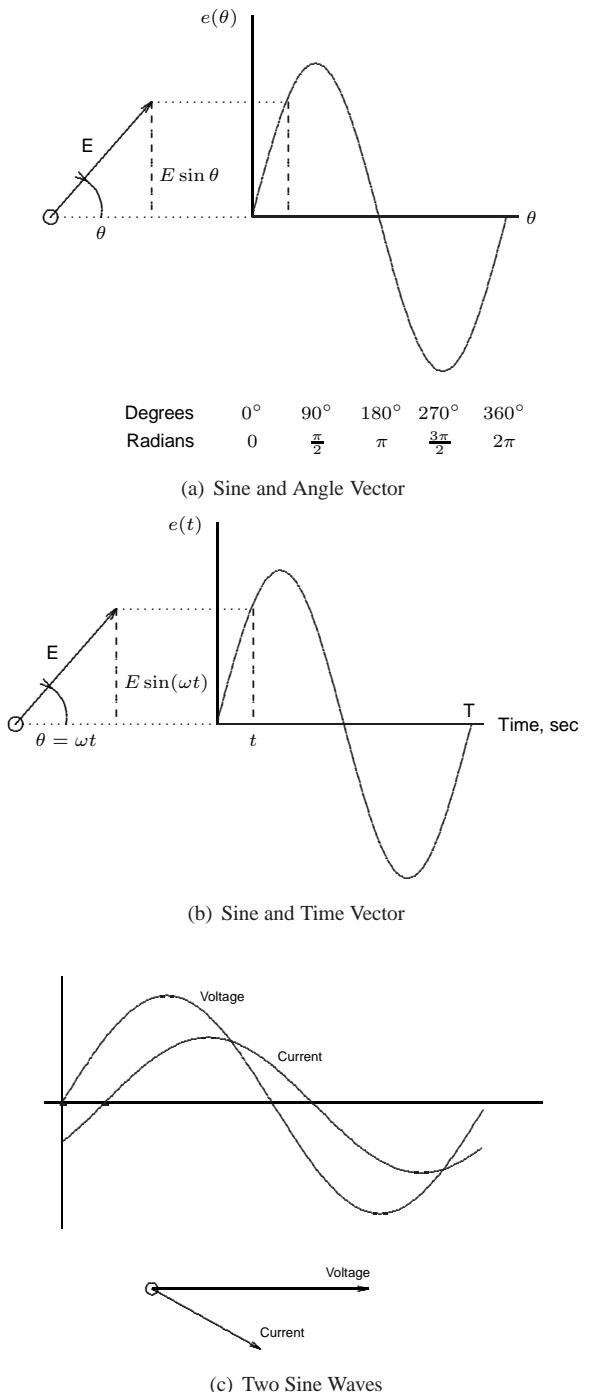


Figure 104: Sine Wave and Vector

of a capacitance or inductance. For this to work, reactance has both a magnitude (that depends on frequency) and a phase angle⁶⁹. Then the calculation of voltages and currents becomes a relatively simple exercise in vector arithmetic.

This is particularly convenient in circuits like power systems that operate at one particular frequency. In many cases, the phase-shifting property of a reactance may be ignored and as a useful approximation the reactance may be treated as a frequency dependent resistance.

Consequently, in the frequency domain, we totally avoid solving differential equations and can manipulate the behaviour of the reactances in much the same fashion as resistance.

When we use these methods of analysing a circuit – phasors, reactance, impedance – then we say that we are analysing the circuit *in the frequency domain*. When we are using the differential equations directly, then we say that we are analysing the circuit *in the time domain*. They are both useful techniques.

4.4 Complex Numbers

An ordinary number can be thought of as identifying a point on the *number line*, figure 105. A line has one dimension, and all points are on this number line. For example, the number 2.3 is 2.3 units to the right of zero on the number line. Negative numbers are to the left of zero on the same number line.

Now consider a Cartesian graph, figure 106. Any point on the graph is characterized by two numbers, which represent its x-position and y-position. The zero position on the graph is known as the *origin*, and has the position 0, 0. The point A is located at 4, 3, that is, 4 units to the right of the origin, and 3 units vertically up from the origin.

In this context, a *vector* is a line segment on a cartesian graph. It turns out that a vector is a very useful representation of an alternating current signal. The vector rotates at the frequency of the AC signal, and the length of the vector is equal to the peak magnitude of the AC signal.

A rotating vector is difficult to deal with. Fortunately it turns out to be sufficient for the purpose of AC circuit analysis to take a snapshot of the vector at some instant of time. The angle of the vector at that instant is the *phase angle* of the signal.

Furthermore, it turns out that the resistors, capacitors and inductors may themselves be treated as vector quantities. This sounds complicated, but it allows the application of familiar circuit laws (Ohm's law, KVL, KCL, etc).

Consequently, we need a process for mathematical manipulation of vectors: addition, subtraction, multiplication and division. This is accomplished by treating each vector as a *complex number*⁷⁰.

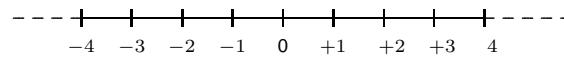


Figure 105: The Number Line

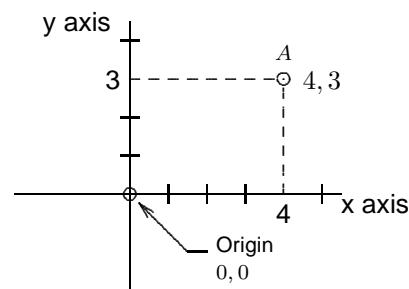


Figure 106: Cartesian Graph

⁶⁹Reactance is more general than resistance. We could say that resistance is a reactance for which the magnitude is frequency invariant and the phase angle is zero.

⁷⁰The name *complex* is an unfortunate choice, reflecting the suspicion visited on these numbers when they were first proposed. Complex numbers follow straightforward rules, and are not difficult to understand.

Complex numbers [51], [52], [53] and *imaginary* numbers are routinely used in many fields of engineering. This is especially true in electrical engineering.

Complex Number: Rectangular Notation

A two-dimensional vector Z may be represented by its horizontal and vertical components as a complex number (figure 107). When Z is a vector we could write:

$$Z = x + jy$$

where x is the horizontal component, said to be the *real* part. y is the vertical component, referred to as the *imaginary* part⁷¹. j is the *imaginary operator*, $\sqrt{-1}$. (Electrical engineers use j for the imaginary operator, mathematicians use i).

When a complex number is expressed like this as the sum of a real part and imaginary part, then it is said to be in *rectangular* notation. (The x component and the y component lie along the sides of an imaginary rectangle.)

For our purposes, a complex number might be better called a *vector number*. The imaginary operator, which poses such conceptual difficulty on first encounter, can be treated as a device that rotates a vector by 90° . It could be called a *quadrature rotation operator*, the term *quadrature* implying *right angle*.

Why use j as an operator on the vertical component? Because it leads to a consistent system of mathematical rules regarding operations such as addition and multiplication on the vectors. These rules are surprisingly consistent with familiar mathematical operations, as illustrated by the following example:

Refer to figure 108 and consider the vector 5 pointing along the positive X axis. This may be regarded as $5j^0 = 5 \times 1 = 5$, a vector 5 units long with no rotations applied.

Then $5j^1 = 5j$ is the same vector rotated CCW 90° to lie along the positive Y axis.

Now consider $5j^2$. This time vector is further rotated CCW by 90° so it points along the negative X axis. Then the original vector has two rotations and in complex notation becomes $5j^2$. But $j^2 = (\sqrt{-1})^2 = -1$ so $5j^2 = -5$, which is the cartesian graph notation for lying along the negative X axis. In other words, treating $j = \sqrt{-1}$ works out correctly with the standard rules of arithmetic.

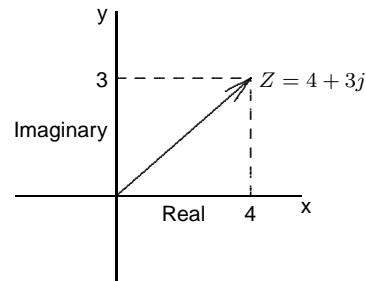


Figure 107: Rectangular Notation

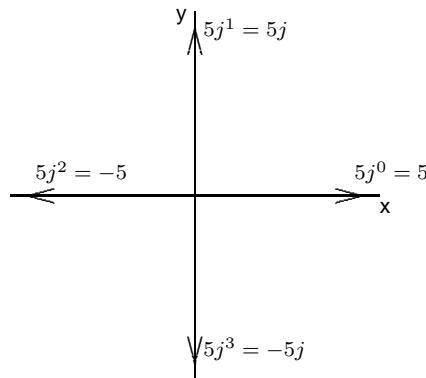


Figure 108: Imaginary Operator j

Complex Number Arithmetic: Addition and Subtraction

The addition and subtraction of complex numbers is quite straightforward when the numbers are expressed in rectangular notation. One simply adds the corresponding parts: real+real, imaginary+imaginary. For example:

⁷¹Another unfortunate name.

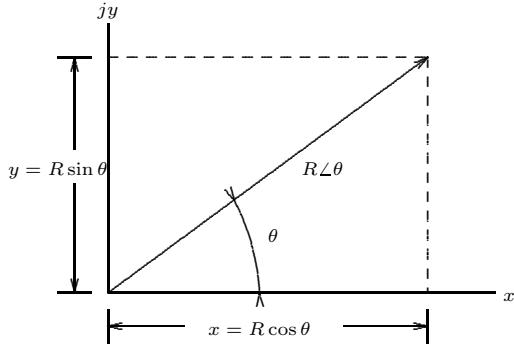
$$(x + jy) + (a + jb) = (x + a) + j(y + b)$$

Multiplication and division are not so simple. It is possible to perform these operations in rectangular notation, but an easier approach is *polar* notation.

Complex Number: Polar Notation

The rectangular form of the complex number may be written as follows (figure 109):

$$\begin{aligned} x + jy &= R \cos \theta + jR \sin \theta \\ &= R(\cos \theta + j \sin \theta) \end{aligned} \quad (160)$$



where:

R is the length of the vector: $R = \sqrt{x^2 + y^2}$

θ is the angle to the vector measured from the horizontal axis, $\tan^{-1}(y/x)$

Figure 109: Polar Notation

However, by Euler's Formula [54]

$$e^{j\theta} = \cos \theta + j \sin \theta \quad (161)$$

Then equation 160 can be written as:

$$x + jy = Re^{j\theta} \quad (162)$$

The right side of this equation is the *polar* form of complex notation. It is so useful in electrical engineering that many scientific calculators have built-in keystrokes to convert from rectangular to polar notation and vice-versa.

Complex Number Arithmetic: Multiplication and Division

The multiplication and division of complex numbers is straightforward when the numbers are expressed in polar notation. One simply multiplies the magnitude and adds the angles. For example:

$$Ze^{j\theta} \times We^{j\alpha} = Z \cdot W e^{j(\theta+\alpha)} \quad (163)$$

$$\frac{Ze^{j\theta}}{We^{j\alpha}} = \frac{Z}{W} e^{j(\theta-\alpha)} \quad (164)$$

$$\frac{1}{We^{j\alpha}} = \frac{1}{W} e^{j(0-\alpha)} = \frac{1}{W} e^{-j\alpha} \quad (165)$$

These operations are so common in electrical engineering that they have a shorthand form, where $e^{j(\text{something})}$ is replaced by the angle symbol \angle (something). Then examples 163 to 165 would be written as:

$$Z \angle \theta \times W \angle \alpha = Z \cdot W \angle (\theta + \alpha) \quad (166)$$

$$\frac{Z \angle \theta}{W \angle \alpha} = \frac{Z}{W} \angle (\theta - \alpha) \quad (167)$$

$$\frac{1}{W \angle \alpha} = \frac{1}{W} \angle -\alpha \quad (168)$$

4.5 Capacitive Reactance

To determine capacitive reactance, we start with the differential equation describing the behaviour of a capacitor in the time domain, equation 169:

$$i = C \frac{dV_c}{dt} \quad (169)$$

Now consider that a sinusoidal voltage $e(t)$ is applied to the capacitor, where

$$e(t) = E \sin \omega t \quad (170)$$

This defines $e(t)$ as a sine wave of peak magnitude E volts and frequency ω radians per second, where (from equation 159, page 156):

$$\omega = 2\pi f \quad (171)$$

and f is the frequency of the AC waveform in Hz.

Example: Voltage Equation for Line Voltage

The North American line voltage has a peak value of 165 volts and frequency of 60Hz or 377 radians/sec.

The equation for line voltage is given by equation 170, where 165 is substituted for E and 377 for ω :

$$\begin{aligned} e(t) &= E \sin \omega t \\ &= 165 \sin 377t \end{aligned}$$

A plot of that equation shows the waveform of line voltage: zero volts at 0 seconds, 165 volts at 4.16 milliseconds, zero again at 8.33 milliseconds, -165 volts at 12.5 milliseconds and so on.

The equation of current is obtained by substituting $e(t)$ for V_c in equation 169:

$$\begin{aligned} i(t) &= C \frac{d}{dt} E \sin \omega t \\ &= \omega C E \cos(\omega t) \\ &= \omega C E \sin(\omega t + 90^\circ) \end{aligned} \quad (172)$$

In words, the current leads the voltage in a capacitor, as shown in figure 110. That is, current must flow into a capacitor in order to create a stored charge, and hence a voltage.

Sanity Check

Equations and graphs imply certain real results. Notice in figure 110 that there is no change in the capacitor voltage when the current is zero (ie, the where the current trace crosses the X axis). This is consistent with our physical understanding: current must flow into a capacitor to change the voltage. It's also consistent with equation 169. When the current drops to zero then the rate of change of voltage dV_c/dt must become zero, ie, the slope of the voltage-time curve is zero (it's horizontal).

Conversely, in figure 110 when the current is at a maximum the rate of change of the voltage is at a maximum. The voltage curve has its largest positive or negative slope.

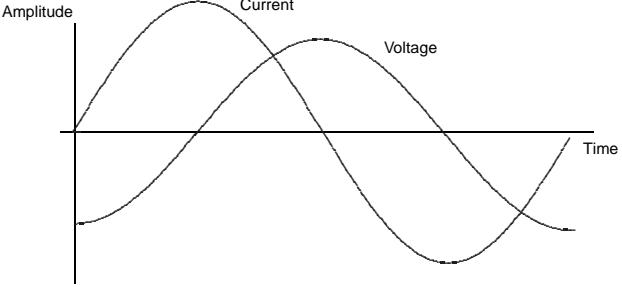


Figure 110: AC Voltage and Current in a Capacitor

4.5.1 Capacitive Reactance in Vector Notation

Now let us switch to vector notation and consider that the voltage is a reference vector of magnitude E and angle zero degrees.

$$\begin{aligned} e(t) &= E \sin \omega t \\ &= E \angle 0^\circ \end{aligned}$$

Then according to equation 172 the current has a magnitude of ωCE and leads by 90 degrees.

$$\begin{aligned} i(t) &= \omega CE \sin(\omega t + 90^\circ) \\ &= \omega CE \angle +90^\circ \end{aligned}$$

The capacitive reactance X_c is the ratio of current to voltage,

$$\begin{aligned} X_c &= \frac{E \angle 0}{\omega CE \angle 90} \\ &= \frac{1}{\omega C} \angle -90^\circ \end{aligned} \tag{173}$$

where the angles are given in degrees.

Now we can apply Ohm's law to a circuit containing a capacitor:

$$e(t) = i(t)X_c \tag{174}$$

The capacitive reactance may be drawn on an impedance vector diagram as shown in figure 111. A resistor lies along the X axis because it does not cause any phase shift to the current. In vector terms, it is $R \angle 0$.

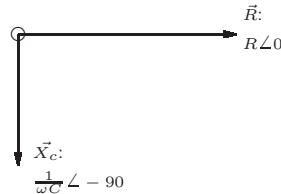


Figure 111: Impedance Vector Diagram

4.5.2 Example: Capacitive Reactance

What is the capacitive reactance of a 1uF capacitor at a frequency of 1000Hz, expressed in vector notation?

Solution

The radian frequency is given by equation 170

$$\begin{aligned} 1000 \text{ Hz} &= 1000 \times 2\pi \text{ radians/sec} \\ &= 6280 \text{ radians/sec} \end{aligned}$$

The capacitive reactance is given by equation 173 above:

$$\begin{aligned} X_c &= \frac{1}{\omega C} \angle -90^\circ \\ &= \frac{1}{6280 \times (1 \times 10^{-6})} \angle -90^\circ \\ &= 159.2 \Omega \angle -90^\circ \end{aligned}$$

4.5.3 Example: Current in a Capacitor

What is the current in this capacitor when the applied voltage is 5 volts peak at a frequency of 1000Hz?

Solution

We use Ohm's law, but taking into account vector angles. The applied voltage e in vector notation is considered to be a vector at zero degrees.

$$e = 5\angle 0^\circ \quad (175)$$

Then the current is

$$\begin{aligned} i &= \frac{e}{X_c} \\ &= \frac{5\angle 0^\circ}{159.2\angle -90^\circ} \\ &= 31.4 \times 10^{-3}\angle +90^\circ \end{aligned}$$

The current and voltage would appear as in figure 110 above, but with a peak value of 5 volts for the voltage and a peak value of 31.4mA for the current.

Capacitive Reactance and Frequency

Quite frequently, phase angle is not important, and we are only interested in the magnitude of the capacitive reactance. Then:

$$X_c = \frac{1}{\omega C} \quad (176)$$

Substituting $2\pi f$ for the radian frequency ω , we have:

$$X_c = \frac{1}{2\pi f C} \quad (177)$$

where X_c is in ohms, ω is in radians/sec, C is in farads and f is in Hz.

This is the most common form of the equation for calculating reactance.

Capacitive Reactance Summary

- Capacitive reactance decreases as frequency increases. So a capacitor appears as an open-circuit at low frequencies and a short-circuit at high frequencies. The precise values of *low frequency* and *high frequency* depend on the size of the capacitance.
- The voltage and current are 90° out of phase in a capacitor, the current leading the voltage.
- A capacitor has a reactance which may be viewed as a vector of magnitude $1/\omega C$ at an angle of -90° .
- The reactance may also be characterized in rectangular notation as $X_c = \frac{1}{j\omega C}$ or Laplace notation as $X_c = \frac{1}{sC}$ (see section 5).

When capacitive reactance is plotted against frequency on a log-log scale, the result is a straight line, as shown in figure 112 for a $1\mu\text{F}$ capacitor. This straight line shifts up or down with different capacitances, but maintains the same slope (see figure 116 on page 167). This image should come to mind wherever the capacitor symbol is seen in a circuit schematic diagram.

Practical methods of calculating capacitive reactance are in section 4.7, page 166.

4.6 Inductive Reactance

Everything that has been said about capacitive reactance has an equivalent in inductive reactance. Again, the AC current is proportional to voltage, and the constant of proportionality is known as the *inductive reactance*.

To determine inductive reactance, we start with the differential equation describing the behaviour of an inductor in the time domain, equation 178:

$$e = L \frac{di}{dt} \quad (178)$$

Now consider that a sinusoidal current $i(t)$ is applied to the inductor, where

$$i(t) = I \sin \omega t \quad (179)$$

This defines $i(t)$ as a sine wave of peak magnitude I amps and frequency ω radians per second, where

$$\omega = 2\pi f \quad (180)$$

and f is the frequency of the AC waveform in Hz.

The voltage is obtained by substituting $i(t)$ in equation 178:

$$\begin{aligned} e(t) &= L \frac{d}{dt} I \sin \omega t \\ &= \omega L I \cos(\omega t) \\ &= \omega L I \sin(\omega t + 90^\circ) \end{aligned} \quad (181)$$

The quantity ωL is known as the *inductive reactance*, in ohms, and given the symbol X_L .

In words, the voltage leads the current in an inductor, as shown in figure 113. That is, voltage must be applied to the terminals of an inductance for some time in order to create a current in the inductor. Also notice in figure 113 that there is no change in the inductor current when the voltage is crossing through the zero axis. It follows from equation 178 that when the voltage drops to zero then $\frac{di}{dt}$ must become zero, ie, the inductor current-time curve is horizontal (zero slope).

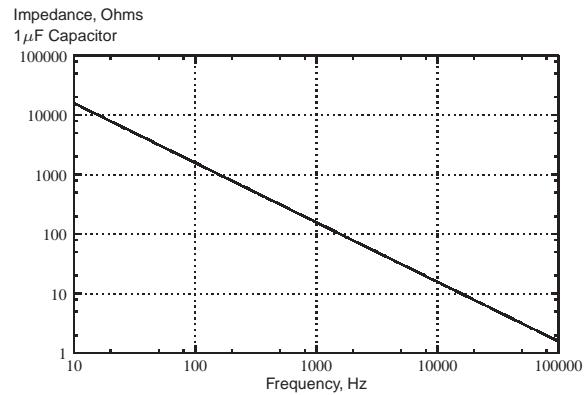


Figure 112: Capacitive Reactance vs Frequency

Figure 113 shows the AC voltage and current in an inductor. The vertical axis is labeled 'Amplitude' and the horizontal axis is labeled 'Time'. The 'Voltage' curve is a sine wave leading the 'Current' curve by 90 degrees. The current curve is a sine wave lagging the voltage by 90 degrees.

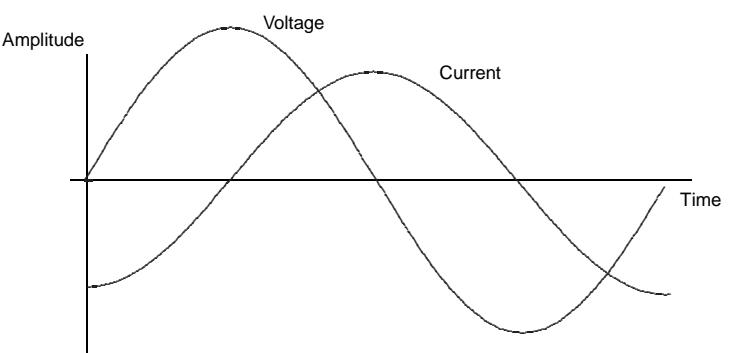


Figure 113: AC Voltage and Current in an Inductor

Vector Notation

Now let us switch to vector notation and consider that the voltage is a reference vector of magnitude E and angle zero degrees. Then the current has a magnitude of $E/\omega L$ and lags by 90 degrees. The inductive reactance X_l is the ratio of voltage to current,

$$\begin{aligned} X_l &= \frac{E\angle 0}{\frac{E}{\omega L}\angle -90} \\ &= \omega L\angle 90 \end{aligned} \quad (182)$$

where the angles are given in degrees.

Now we can apply Ohm's law to a circuit containing an inductor:

$$e(t) = i(t)X_l \quad (183)$$

The inductive reactance may be drawn on an impedance vector diagram as shown in figure 114.

A resistor lies along the X axis because it does not cause any phase shift to the current. In vector terms, it is $R\angle 0$.

Inductive Reactance and Frequency

Like capacitive reactance, inductive reactance also plots as a straight line, as shown in figure 115 for a 1mH inductor. This straight line shifts up or down with different inductances, but maintains the same slope. This image should come to mind wherever the inductor symbol is seen in a circuit schematic diagram.

Inductive Reactance Summary

- Inductive reactance increases as frequency increases. So an inductor appears as a short-circuit at low frequencies and an open-circuit at high frequencies. The precise values of *low frequency* and *high frequency* depend on the size of the inductance.
- The voltage and current are 90° out of phase in a inductor, the current lagging the voltage.
- An inductor has a reactance which may be viewed as a vector of magnitude ωL at an angle of $+90^\circ$.
- The reactance may also be characterized in rectangular notation as $X_l = j\omega L$ or Laplace notation (see section 5) as $X_l = sL$

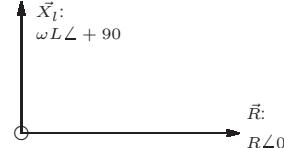


Figure 114: Impedance Vector Diagram

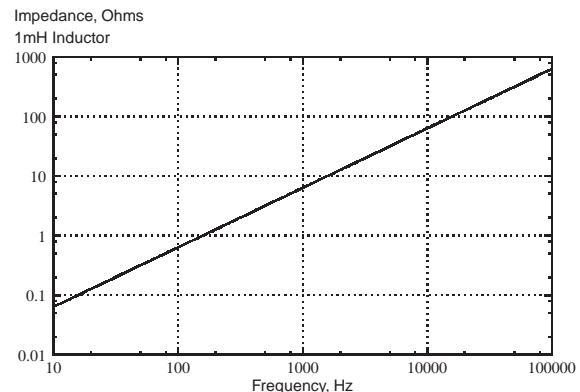


Figure 115: Inductive Reactance vs Frequency

4.7 Reactance Calculators

The calculation of capacitive reactance is a routine chore in analog circuit design. Three examples are shown below, using various forms of equation 184:

$$X_c = \frac{1}{2\pi f C} \quad (184)$$

where X_c is in ohms, ω is in radians/sec, C is in farads and f is in Hz.

1. What is the reactance of a 100nF capacitor at 1MHz?

$$\begin{aligned} X_c &= \frac{1}{2\pi f C} \\ &= \frac{1}{2\pi(1 \times 10^6) \times (100 \times 10^{-9})} \\ &= 1.592\Omega \end{aligned}$$

2. Choose a capacitor that has a reactance of 10Ω at 100Hz.

$$\begin{aligned} C &= \frac{1}{2\pi f X_c} \\ &= \frac{1}{2\pi \times 100 \times 10} \\ &= 159\mu F \end{aligned}$$

3. Find the frequency at which a 10nF capacitor has a reactance of 100Ω .

$$\begin{aligned} f &= \frac{1}{2\pi C X_c} \\ &= \frac{1}{2\pi \times (10 \times 10^{-9}) \times 100} \\ &= 159 \text{ kHz} \end{aligned}$$

This is all quite straightforward, but tedious. It requires that one rearrange equation 184 to solve for the desired quantity and then type the known values into a calculator. It does not lend itself easily to the exploration of various design scenarios.

A spreadsheet or purpose-built program is better, but that requires calling up the program and typing in a range of values.

There are two simple design aids⁷² that can speed up the process: the *reactance chart* and the *reactance slide rule*.

4.7.1 Reactance Chart

The simplest design aid is the *reactance chart*. The chart is an elaboration of figure 112 with multiple traces, one for each value of capacitor.

Reactance charts have a long history. They have appeared in the annual version of the ARRL handbook [45] for many years. Traditionally, the reactance chart shows capacitive reactance and inductive reactance on the same

⁷²Battery not required.

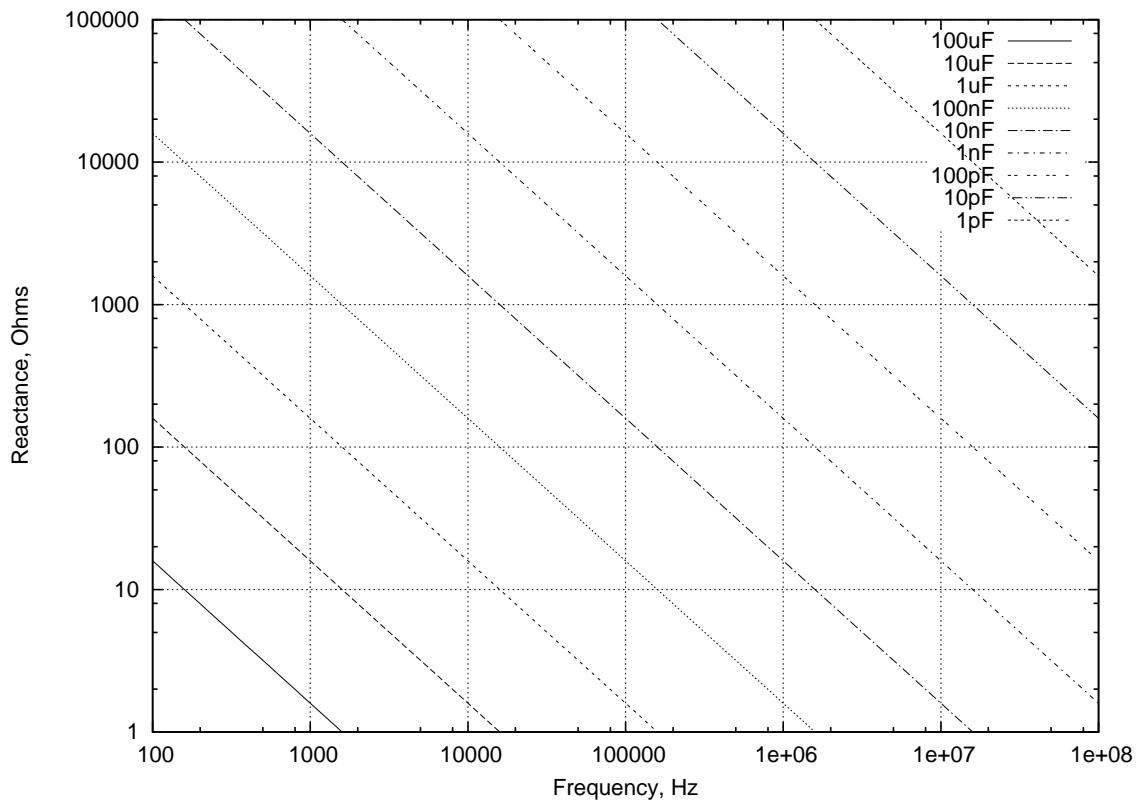


Figure 116: Reactance Chart for Capacitance

chart. Since there is less requirement to know inductive reactance, capacitive reactance is plotted by itself on this chart.

As an example, consider the first of the reactance calculations above: *What is the reactance of a 100nF capacitor at 1MHz?* Find the diagonal trace representing 100nF. Follow that down to the right to its intersection with 1MHz ($1\text{e}06$) on the frequency axis. Now carry that point over to the vertical axis and read off resistance in the order of 1.5Ω . Clearly this is far more approximate than a calculation, but that's often not important: we would choose the next standard value for the capacitor in any case.

The plot of figure 116 was done with a Gnuplot routine shown in section 4.7.3. A modern-day practitioner could use this code to generate their own reactance chart. For example, a designer of audio circuits could modify the Gnuplot code to show capacitive reactance between 1 Hz and 100 kHz. As a practical matter, the capacitance traces are widely spaced. A more usable chart would show capacitance in a 1:2:5:10 series.

It's somewhat more difficult to read but there is one advantage to putting capacitive and inductive reactance on the same graph: the point of intersection of an inductive and capacitive trace is a *resonant frequency* for those values. (In the case of resonance, the combined reactance of the inductance and capacitance is wildly different than the chart value.) Again, that's easily done by modifying the plotting code.

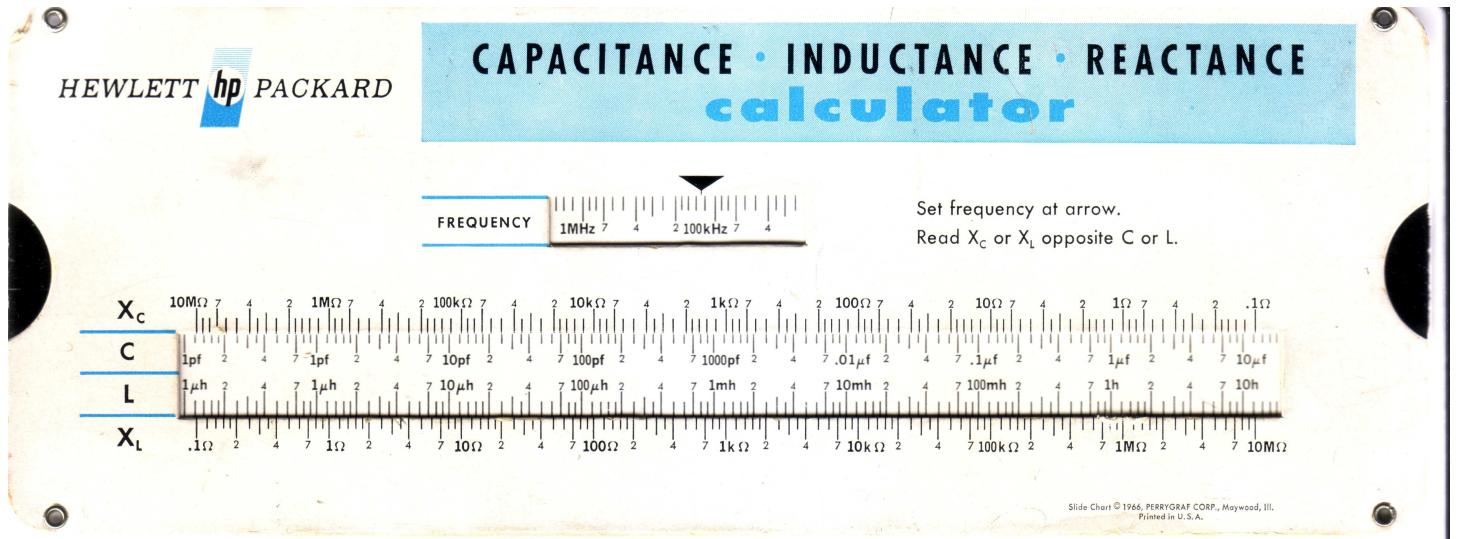


Figure 117: Reactance Slide Rule

4.7.2 Reactance Slide Rule

The *reactance slide rule* is shown in figure 117. These slide rules were produced by Perrygraf and given out in the 60's as a promotion from Hewlett-Packard. They immediately found a receptive audience among electrical engineers. Operation is simplicity itself: slide the inner part until the operating frequency is aligned with the arrow in the upper window. In the lower window upper scale, read the capacitive reactance opposite the value of capacitance. In the lower window lower scale, read the inductive reactance opposite the value of inductance.

Alternatively, line up some value of capacitance next to a desired value of capacitive reactance, and the frequency at which that occurs is in the upper window.

The accuracy is sufficient for much design work. It is much more versatile and convenient than a calculator.

A very nice computer simulation of the calculator in figure 117 is on the web [55]. Printouts of another reactance slide rule, by the Shure audio electronics company (circa 1943!), are also on the web [56].

Antique reactance slide rules are available through the surplus market. It would not be difficult to produce a modern version, using the sources cited above.

4.7.3 Gnuplot Routine for Capacitive Reactance Chart

```
# List of commands to plot the impedance of a capacitor vs frequency
set terminal postscript
set output "cap-reactance-chart.ps"
set ylabel "Reactance, Ohms"
set xlabel "Frequency, Hz"
set grid
set logscale y
set logscale x
plot [x=100:100000000] [1:100000] \
1/(2*3.1415*x*(100*10**-6)) title '100uF' with lines,\
```

```

1/(2*3.1415*x*(10*10**-6)) title '10uF' with lines,\n
1/(2*3.1415*x*(1*10**-6)) title '1uF' with lines,\n
1/(2*3.1415*x*(100*10**-9)) title '100nF' with lines,\n
1/(2*3.1415*x*(10*10**-9)) title '10nF' with lines,\n
1/(2*3.1415*x*(1*10**-9)) title '1nF' with lines,\n
1/(2*3.1415*x*(100*10**(-12.0))) title '100pF' with lines,\n
1/(2*3.1415*x*(10*10**-12.0)) title '10pF' with lines,\n
1/(2*3.1415*x*(1*10**-12.0)) title '1pF' with lines\n
set terminal x11
replot
pause -1

```

To change this to a plot of inductive reactance, invert each of the capacitance formulae, changing $1/(something)$ to $(something)/1$. Following that pattern, it is also possible to include inductive reactance on this chart.

This routine generates a postscript file `cap-reactance-chart.ps`, which can be printed.

4.8 RC Lowpass Network

In this section, we deal with resistor-capacitor networks configured as voltage dividers. These turn out to be very common, and so it is important to recognise their time-domain and frequency-domain behaviours.

Time Domain Response

The circuit for an *rc-lowpass filter* is shown in figure 118.

To understand the time-domain response, consider that a step voltage E is applied to the input at $e_i(t)$. The capacitor is initially uncharged, and voltage cannot change instantaneously across a capacitor. So the capacitor and output voltage $e_o(t)$ are both zero to start with. Then by KVL, the full E volts is across the resistor and the charging current is E/R .

The capacitor starts to charge up at a rate of

$$\frac{dV_c}{dt} = \frac{E}{R} \text{ volts per second} \quad (185)$$

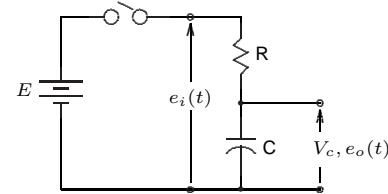


Figure 118: RC Lowpass Filter

As the capacitor charges, the voltage across the capacitor increases and the voltage across the resistor must decrease. This causes the flow of charging current to decrease, which slows down the rate of charging the capacitor. Consequently, the time charging curve flattens out, as shown in figure 119.

The rate of charging is determined by the *time constant* τ , where

$$\tau = RC \quad (186)$$

Figure 119 indicates that the capacitor is charged to about 63% of its final value in one time constant and essentially fully charged in 5 time constants.

The complete charging equation ⁷³ is:

$$V_c = E(1 - e^{-\frac{t}{RC}}) \quad (187)$$

Substituting τ for RC in equation 187, we have the alternative form:

$$V_c = E(1 - e^{-\frac{t}{\tau}}) \quad (188)$$

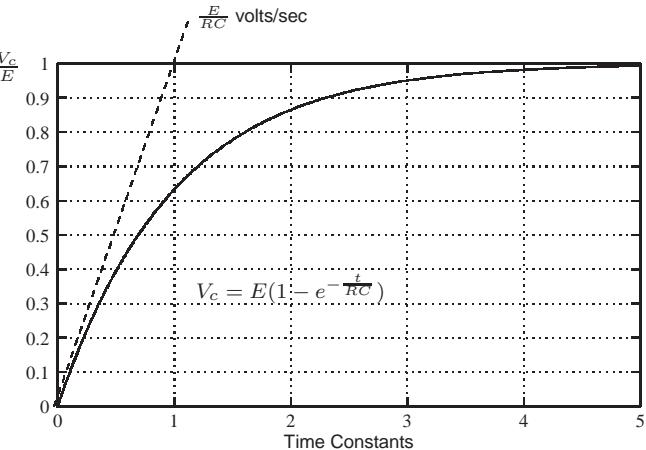


Figure 119: RC Lowpass Charging Curve

4.8.1 Example: Charging Capacitor

The voltage supply, resistor and capacitor values in the circuit of figure 119 are: $E = 10V$, $R = 500k\Omega$, $C = 0.2\mu F$. What is the voltage on the capacitor 0.3 seconds after the switch closes?

Solution

Substitute for R and C in equation 186 to calculate the time constant τ .

$$\begin{aligned} \tau &= RC \\ &= (500 \times 10^3) \times (0.2 \times 10^{-6}) \\ &= 0.1 \text{ seconds} \end{aligned}$$

Substitute time t , time constant τ and supply voltage E into equation 188:

$$\begin{aligned} V_c &= E(1 - e^{-\frac{t}{\tau}}) \\ &= 10 \times \left[1 - e^{-\frac{0.3}{0.1}} \right] \\ &= 9.5 \text{ volts} \end{aligned}$$

A time interval of 0.3 seconds represents 3 time constants, so as indicated in figure 119 the capacitor voltage is approaching 'fully charged', 10 volts.

Frequency Response

Now we will consider the behaviour of the lowpass when it is driven from an AC source.

The lowpass network behaves as a voltage divider with a frequency-dependant divide ratio. At low frequencies, the capacitor behaves like an open circuit, and the divider ratio (the *gain* of the network, the ratio of output to input voltage) is close to unity. At high frequencies, the divide ratio approaches zero.

⁷³The most straightforward mathematical development of this formula uses the Laplace Transform, section 5.3.7 on page 220.

The transition between the two behaviours occurs at the *cutoff frequency* ω_o , (also known as the *corner frequency*), where the capacitive reactance is equal to the resistance.

$$\begin{aligned} X_C &= R \\ \frac{1}{\omega_o C} &= R \\ \omega_o &= \frac{1}{RC} \text{ radians/sec} \end{aligned} \quad (189)$$

Notice that the cutoff frequency is equal to the reciprocal of the network time constant:

$$\omega_o = \frac{1}{\tau} \quad (190)$$

The frequency and phase responses of the lowpass network ⁷⁴ are shown in figure 120(a) and (b).

The magnitude response is expressed in decibels, where:

$$G_{db} = 20 \log_{10} \left(\frac{e_o}{e_i} \right) \text{ decibels} \quad (191)$$

and

$$\angle G = \angle e_o - \angle e_i \text{ degrees} \quad (192)$$

Frequency is expressed in logarithmic terms relative to the cutoff frequency ω_o . Notice how both these curves can be approximated by straight lines. This is only true when both axes are logarithmic.

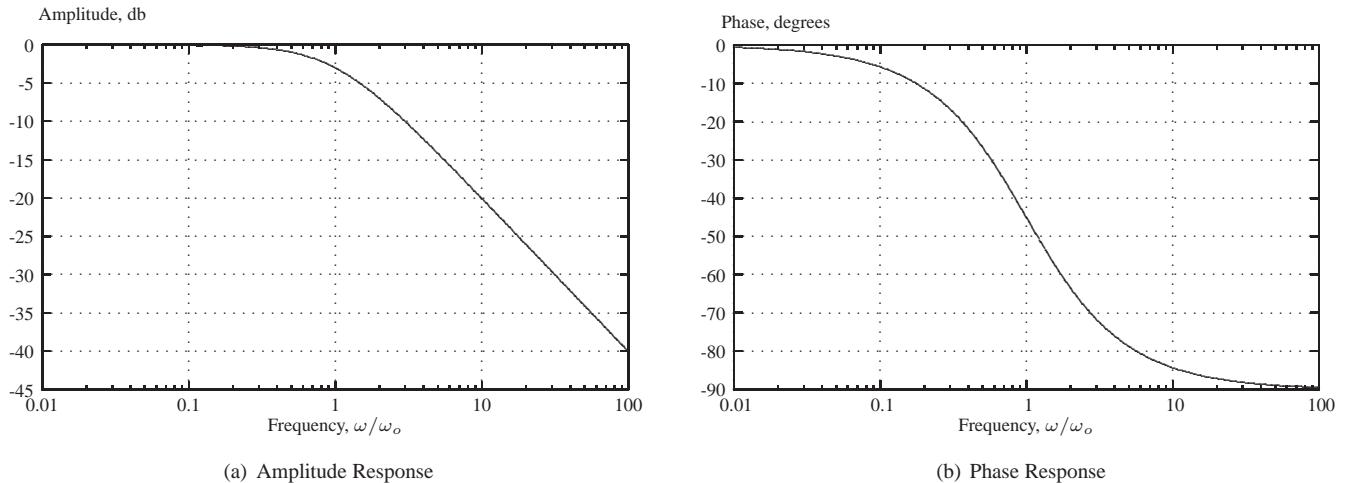


Figure 120: RC Lowpass Amplitude and Phase Response

Notice how the amplitude curve rolls off at a rate of -20db/decade above the cutoff frequency. (A decade is a factor of 10 in frequency.)

⁷⁴The development of these plots is given in detail section 5.2.2 on page 201.

The phase changes from zero at low frequencies to -45° at the corner frequency, to -90° at high frequencies. The transition takes place over two decades (a factor of 100).

These are important curves and should be committed to memory. Lowpass RC networks are useful in themselves for deliberately removing high-frequency components of a signal. They also occur unintentionally when a stray capacitance in the circuit interacts with some source resistance. This is the mechanism that restricts the maximum operating frequency of amplifiers and other circuits, so it's important to know how to predict and manage its effects.

4.9 RC Highpass Network

Time Domain Response

The circuit for an *rc-highpass filter* is shown in figure 118.

Consider that a step voltage E is applied to the input at $e_i(t)$. The capacitor is initially uncharged, and voltage cannot change instantaneously across a capacitor. So when the top plate of the capacitor steps up by E volts, the bottom plate must follow. At that point, there is E volts across the resistor, so current begins to flow through it, charging the capacitor. The capacitor starts to charge at a rate of

$$\frac{dV_c}{dt} = \frac{E}{R} \text{ volts per second} \quad (193)$$

As the capacitor voltage increases, the voltage across the resistor decreases until it has dropped completely to zero. Consequently, the time charging curve flattens out, as shown in figure 119.

As in the case of the lowpass filter, the rate of charging is determined by the *time constant* τ , where

$$\tau = RC \quad (194)$$

The complete equation of the output waveform (obtained by the Laplace Transform methods of section 5.3) is:

$$V_c = E(e^{-\frac{t}{RC}}) \quad (195)$$

Frequency Response

At low frequencies, the capacitor behaves like an open circuit, and the divide ratio is close to zero. At high frequencies, the capacitor approaches a short circuit and the divide ratio approaches unity.

The transition between the two behaviours occurs at the *cutoff frequency* ω_o , where

$$\omega_o = \frac{1}{RC} \text{ radians/sec} \quad (196)$$

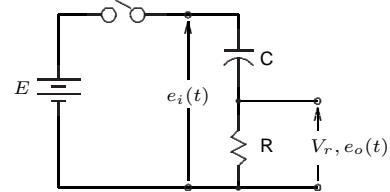


Figure 121: RC Highpass Filter

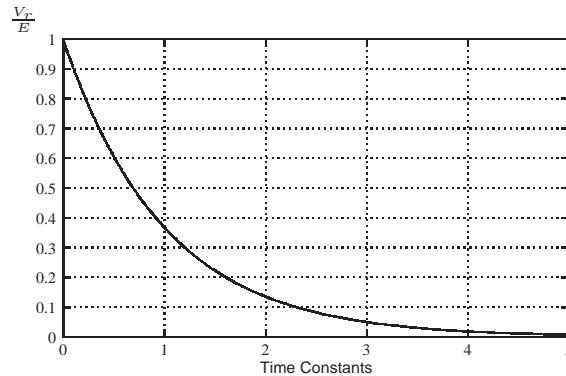


Figure 122: RC Highpass Step Response

as in the lowpass case. Also,

$$\omega_o = \frac{1}{\tau} \quad (197)$$

The frequency and phase responses of the highpass network are shown in figure 123(a) and (b).

The magnitude response is expressed in decibels, where:

$$G_{db} = 20 \log_{10} \frac{e_o}{e_i} \text{ decibels} \quad (198)$$

and

$$\angle G = \angle e_o - \angle e_i \text{ degrees} \quad (199)$$

Frequency is expressed in logarithmic terms relative to the cutoff frequency ω_o . Again, both these curves can be approximated by straight lines.

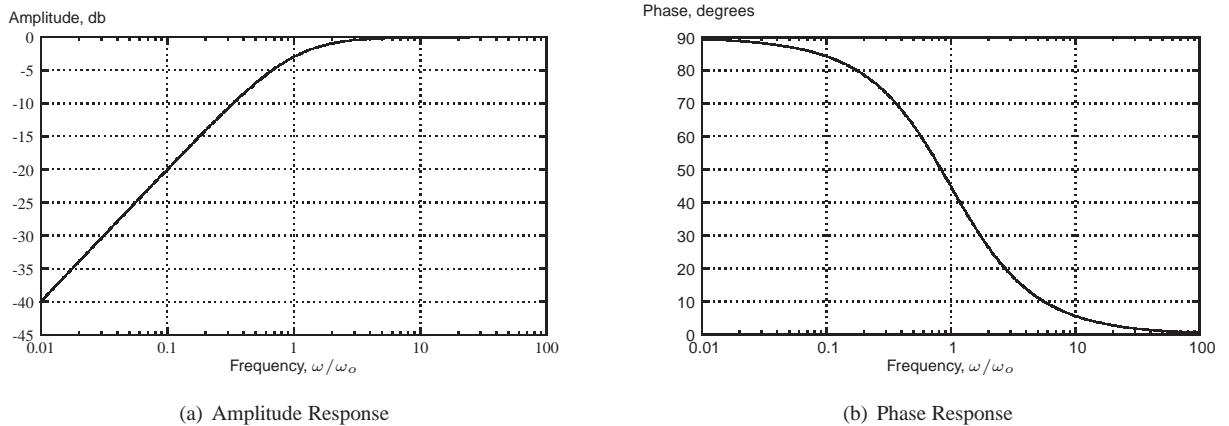


Figure 123: RC Highpass Amplitude and Phase Response

Below the cutoff frequency, the amplitude curve rolls off at a rate of -20db/decade. The phase changes from +90° at low frequencies to +45° at the corner frequency, to 0° at high frequencies. The transition takes place over two decades (a factor of 100).

Like the lowpass curves, these curves are important and should be committed to memory.

4.10 Resonance, an Introduction

The circuit shown in figure 124(a) illustrates the phenomenon of resonance and, with a few simple rules, is simple to analyse.

The parallel combination of the capacitance and inductance is known as a *tank* circuit [57], [58]. Consider that the switch is in position *a*, so the capacitor charges up to *E* volts⁷⁵. The switch is then moved to position *b*. Assume the circuit is lossless (no resistance). What happens to the voltage and current in the tank circuit?

- At the instant the switch is moved over to position *b*, (time t_1) there is a certain quantity of stored energy in the capacitor, given by equation 61. The voltage in the capacitor will cause a current in the inductor, so the capacitor starts to discharge and its voltage will begin to decrease as the inductor current increases.

⁷⁵It's bad practice to charge a capacitor directly from a voltage supply because the charging current is infinite for an infinitesimal instant, and this can be hard on the switch. The circuit should really have a resistance to limit the charging current.

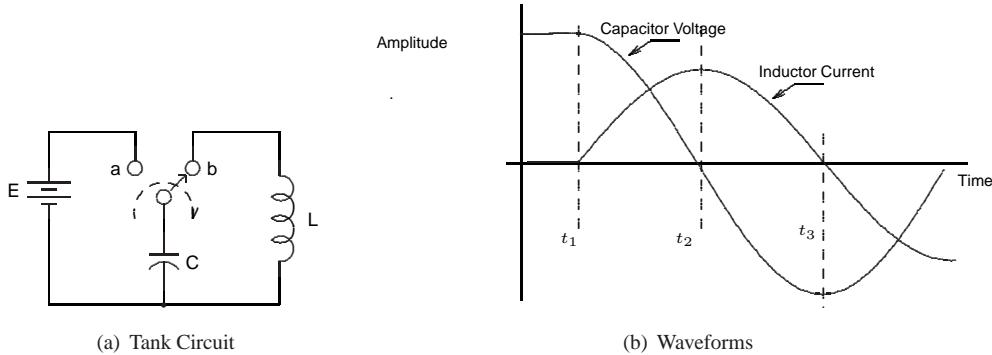


Figure 124: Tank Circuit and Waveforms

- Once the capacitor has completely discharged, (time t_2) all the energy previously stored in the electric field of the capacitor has been transferred to the magnetic field of the inductor, the current in the circuit is at a peak, and the stored energy in the system is given by equation 85 on page 100
 - Now the situation reverses. The inductor current cannot change rapidly, so it continues and recharges the capacitor to a negative voltage. As it does so, it decreases until it stops completely At this point the capacitor is charged again to the same peak voltage, but to the opposite polarity (time t_3).
 - Since this is a lossless circuit, the oscillation will continue forever. In practice, there is some resistance in the inductor and wiring and the sine waves of voltage and current will gradually die away.

A detailed analysis of this circuit is in section 5.3.10 on page 228. In that analysis, we find that the waveforms are sinusoidal in shape, and the oscillation occurs at the resonant frequency:

$$f_o = \frac{1}{2\pi\sqrt{LC}} \text{ Hz} \quad (200)$$

The complete waveforms are shown in figure 124(b).

4.11 Series Resonance

When an inductor and capacitor appear in series or in parallel in an electronic circuit, *resonance* will occur. The behaviour is analogous to a spring-mass system in mechanics: energy transfers back and forth between the two elements at some frequency dependent on the properties of the two elements.

There are two possible resonant circuit configurations: series and parallel. It is possible to analyse these circuits in some detail, but here we will stick to the main points. (For a more detailed development, see for example [59]).

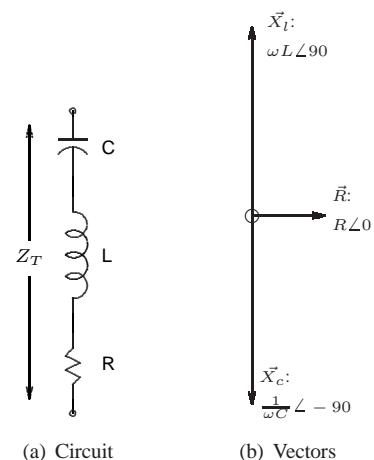


Figure 125: Series Resonant Circuit

Resonant circuits are most useful because they are *frequency selective*. Their impedance changes with frequency and shows a maximum or minimum at some particular frequency.

A series resonant circuit is shown in figure 125(a).

First, let's try to get a feel for the variation of the total impedance Z_T of this RLC circuit as the frequency changes.

- At low frequencies, the capacitor will look like an open circuit (high reactance), so Z_T will be high.
- At high frequencies, the inductor will look like an open circuit (high reactance), so Z_T will be high.
- As the frequency is swept from a low value to a high value, the capacitive reactance decreases and the inductive reactance increases (see figures 112 and 115).
- At some intermediate frequency, the capacitive and inductive reactances will be equal. Since they are of opposite sign on the vector diagram (figure 125(b)), they will cancel, and the net impedance will be the resistance R . This is the lowest value of the impedance of the network and occurs at the resonant frequency.
- In summary, the series resonant circuit is generally large but drops to a low value at the resonant frequency.

Based on this simple reasoning, the resonant frequency occurs when the magnitude of the capacitive reactance is equal to the magnitude of the inductive reactance, so that

$$X_L = X_C \quad (201)$$

$$\omega_o L = \frac{1}{\omega_o C} \quad (202)$$

$$\omega_o = \sqrt{\frac{1}{LC}} \quad (203)$$

where ω_o is the resonant frequency in radians/sec.

4.11.1 Frequency Response

The frequency response of the impedance is shown in figure 126.

Notice how the impedance decreases in a linear fashion at low frequencies (where the network is mainly capacitive) and increases in a linear fashion at high frequencies (where it is mainly inductive). At resonance, the curve departs from these asymptotes.

At resonance, the impedance drops to the value of the series resistance R . As R is made smaller, the resonant valley becomes narrower and extends to a lower minimum value. The narrowness of the resonant valley around the resonant frequency is defined by the *Q factor* (section 4.13, page 178).

The Q factor can be shown to be:

$$\begin{aligned} Q &= \frac{\omega_o}{\Delta\omega} \\ &= \frac{X_L}{R} \end{aligned} \quad (204)$$

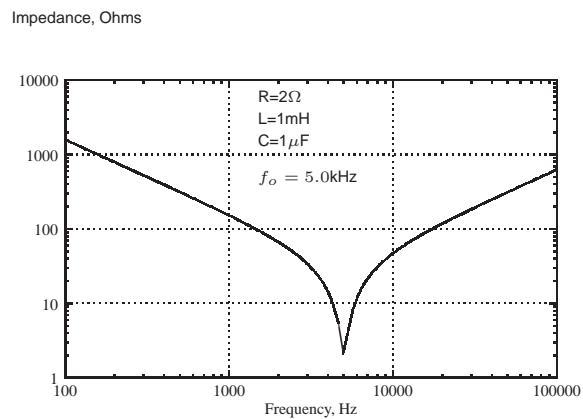


Figure 126: Series RLC Circuit Response

where $\Delta\omega$ is the *bandwidth*, the range in frequency where the response is different from the peak value by a factor of $\sqrt{2}$.

In this example, the resonant frequency ω_o is 31.6×10^3 radians per second, X_L at resonance is 31.6Ω , and the resistance is 2Ω . From equation 204, the Q factor is 15.8 and the bandwidth $\Delta\omega$ is 2000 radians per second (318Hz).

The circuit resistance is often determined by the series resistance of the inductance. A low-resistance inductor is capable of higher Q factors and narrower bandwidth, and consequently acts as a more selective frequency filter.

4.11.2 Voltages and Currents in the Series Resonant Circuit

An interesting phenomenon, voltage amplification, occurs at resonance. The voltages across the inductor and capacitor can be much larger than the input voltage. What causes this?

Referring to the vector diagram, figure 125(b), the impedance vectors cancel out because they are equal and opposite. So the current in the circuit is controlled by the series resistance. Consequently, there is a large AC current flowing through the circuit. If e is the voltage across the series-resonant circuit, then the current i through the circuit is:

$$i = \frac{e}{R} \quad (205)$$

This same AC current flows through the inductive and capacitive reactances, setting up voltages across them. For example, the voltage e_L across the inductor is:

$$\begin{aligned} e_L &= iX_L \\ &= \frac{e}{R} X_L \\ &= Qe \end{aligned} \quad (206)$$

If the Q factor is large, then the voltage across the inductor and across the capacitor is much larger than the input voltage. This can be a consideration in RF transmitter applications, where the voltages caused by this *Q multiplication effect* can cause components to break down. In the example given above, Q is 15.7. If the input voltage e is 10 volts, then the voltage across the inductor and across the capacitor is 157 volts.

An interesting example of Q amplification occurs in nature with the tides in the Minas Basin (part of the Bay of Fundy on the east coast of Canada). The tidal flow of the Minas Basin is a resonant system with a Q of 5 and a period of 13 hours. As a result, the tidal maximums are some of the highest in the world, up to 16 metres [60].

4.12 Parallel Resonance

A parallel resonant circuit is shown in figure 127.

Considering the impedance Z_T of this circuit as the frequency changes:

- At low frequencies, the inductor will look like a short circuit (low reactance), so Z_T will be low.
- At high frequencies, the capacitor will look like a short circuit (low reactance), so Z_T will again be low.
- As the frequency is swept from a low value to a high value, the inductive reactance increases and the capacitive reactance decreases.

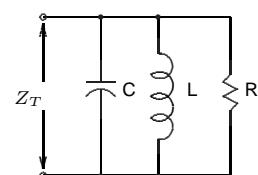


Figure 127: Parallel Resonant Circuit

- At some intermediate frequency, the capacitive and inductive reactances will be of equal magnitude and of opposite sign: $X_L = -X_C$. Then the circuit is said to be resonant, and the reactive component of the circuit X_T is given by the parallel combination of the capacitive and inductive reactances:

$$\begin{aligned}\frac{1}{X_T} &= \frac{1}{X_L} + \frac{1}{X_C} \\ &= \frac{1}{X_L} + \frac{1}{-X_L} \\ &= 0\end{aligned}$$

Then

$$X_T = \infty \quad (207)$$

i.e., the parallel combination of reactances appears as an open circuit. The total impedance Z_T is equal to the parallel combination of the resistance plus an open circuit, so it's simply equal to the resistance R .

In summary, the impedance of the parallel resonant circuit is generally low but increases to the value of R at the resonant frequency.

4.12.1 Frequency Response

The frequency response of the impedance is shown in figure 128.

Notice how the impedance increases in a linear fashion at low frequencies, where the network is mainly inductive. The impedance decreases in a linear fashion at high frequencies, where it is mainly capacitive.

At resonance, the impedance increases to the value of the parallel resistance R . As R is made large, the resonant valley becomes narrower and extends to a higher maximum value. The narrowness of the resonant valley around the resonant frequency is defined by the *Q factor*.

The value for resonant frequency is the same as the series resonant case.

The Q factor formulae are the reciprocals of the series case:

$$Q = \frac{\Delta\omega}{\omega_o} = \frac{R}{\omega_o L} = \frac{R}{X_L} \quad (208)$$

(As a sanity check on this, notice that the Q factor is proportional to the length of time that the circuit will ring, as explained in section 4.13 following. For a very large value of resistance, approaching an open circuit, the circuit has less dissipation and therefore will ring longer, and therefore has a larger Q factor.)

In practice, the parallel resonant circuit is not quite as simple as this example would imply. Usually, the parallel resistance is extremely large and there is a small resistance in series with the inductor, caused by the wiring resistance of the inductor. The Big Picture remains unchanged, but there are minor changes in detail. If you have to design a parallel resonant circuit these days, the best approach would be to understand the basic operation of the circuit, as given above, and then simulate the circuit to nail down the detailed response.

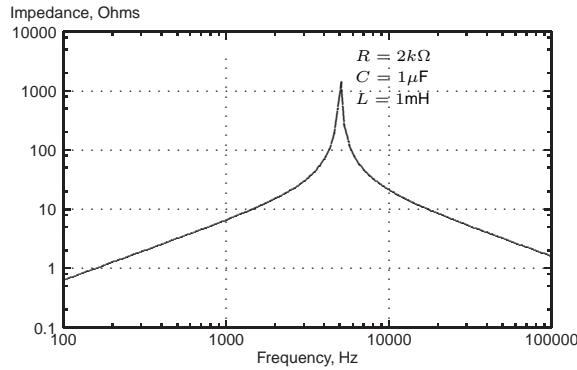


Figure 128: Parallel RLC Circuit Response

4.13 Boing vs Thud: the Q Factor

The *Q Factor* finds use in a number of related situations in electronic design.

To begin with, let us describe the way an oscillation decays in terms of *damping*, because that is easiest to understand. A long ringing oscillation (as in the example of figure 129(a)) is said to be *lightly damped* or *underdamped*. Increasing the damping has the effect shown in figure 129(b): the oscillation dies away more quickly. An oscillation that dies away very quickly is described as *heavily damped* or *over-damped*. The mid-point, which has some mathematical significance, is called *critical damping*. The term *damping* comes to us from mechanical engineering. Picture a mass-spring system suspended in open air. It bounces for some time after it's put in motion. Now picture that the mass is immersed in a bucket of water. The movement of the water is dissipative, that is, it removes energy from the bouncing system and the bouncing decays more quickly. Thus *damping* the mass-spring system increases the rate of decay of the oscillation.

The Q factor is an inverse of the damping factor. If the sound of a struck bell decays slowly, then it is said to be a *high Q* oscillation. If it decays quickly, then it is a *low Q* oscillation.

The term Q factor arises from radio engineering. A high Q inductor is one that has a high ratio of inductance to resistance. When used in a selective circuit, it has a narrow passband, (better selectivity). Consequently a high Q inductor is *high quality*.

4.13.1 Defining Q Factor

It turns out that Q factor is related to the damping factor δ by:

$$Q = \frac{1}{2\delta} \quad (209)$$

Q factor is commonly used in analog electronics and radio frequency circuits. Damping factor is commonly used in servomechanism design and negative feedback systems. They are two faces of the same phenomenon.

(To confuse matters further, there is a *damping coefficient d* which is defined as:

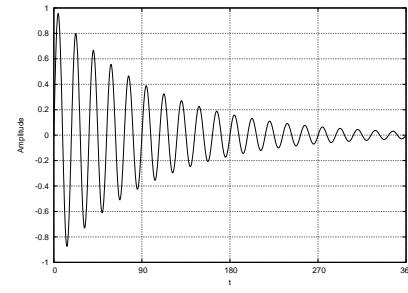
$$Q = \frac{1}{d} \quad (210)$$

This is sometimes used in connection with the specification of active filters, but is not common.)

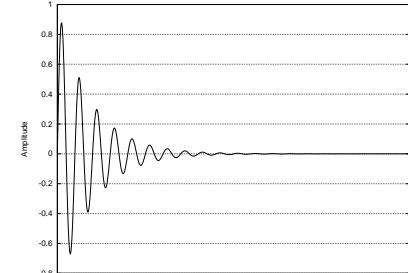
For an electrical engineer, a simple definition of Q factor relates the reactance of an inductor to its resistance.

$$Q = \frac{X_l}{R} \quad (211)$$

$$= \frac{\omega L}{R} \quad (212)$$



(a) Lightly Damped



(b) Heavily Damped

Figure 129: Damped Oscillations

4.13.2 Q Factor and Energy Dissipation

Now we'll show that there is a relationship between Q factor, energy dissipation in the resistor, and energy storage in the inductor.

Consider that an inductor L and resistor R are in series. They are conducting a sinusoidal current $i(t) = I_m \sin(\omega t)$. That is, the sine wave has a peak current value of I_m amperes and a radian frequency of ω radians per second.

When the current is at its peak value I_m , then the energy stored in the inductor is also at its peak (section 2.28.5, page 100, equation 85):

$$W_s = \frac{1}{2} L I_m^2 \text{ joules} \quad (213)$$

The inductor and resistor are in series so the peak value of the current in the resistor is the same as the inductor, I_m . Then the effective value of the current that causes heating in the resistor (the RMS value of the current, I_{eff}) is $1/\sqrt{2}$ times the peak value.

$$I_{eff} = \frac{I_m}{\sqrt{2}} \quad (214)$$

The power being dissipated in the resistor P_d is

$$\begin{aligned} P_d &= I_{eff}^2 R \\ &= \left(\frac{I_m}{\sqrt{2}} \right)^2 R \\ &= \frac{I_m^2}{2} R \text{ watts} \end{aligned} \quad (215)$$

The energy dissipated in the resistor is this power times the period T of one cycle.

$$W_d = \frac{I_m^2}{2} R T \text{ joules} \quad (216)$$

The period T can be written in terms of the radian frequency ω :

$$\omega T = 2\pi \text{ radians} \quad (217)$$

Substitute $\omega/2\pi$ from equation 217 for T into equation 216, and the energy dissipated in the resistor is:

$$W_d = \frac{I_m^2}{2} R \frac{2\pi}{\omega} \text{ joules} \quad (218)$$

Finally, the ratio of energy stored to energy dissipated is:

$$\begin{aligned} \frac{W_s}{W_d} &= \frac{\frac{1}{2} L I_m^2}{\frac{I_m^2}{2} R \frac{2\pi}{\omega}} \\ &= \frac{1}{2\pi} \frac{\omega L}{R} \end{aligned} \quad (219)$$

From equation 212, $\omega L/R = Q$, so

$$Q = 2\pi \frac{W_s}{W_d} \quad (220)$$

That is, the Q factor is proportional to the ratio of peak energy stored to average energy dissipated in one cycle. This definition is more general than equation 212 since it can apply to systems other than electrical, such as a mechanical vibration.

4.13.3 Q Factor and Voltage Magnification

For illustration, let us return to the spring-mass system. Without damping (the mass bouncing in air), it has a very large Q factor. Then the effect of resonance is quite dramatic. A very small input vibration at the resonant frequency causes *much* larger movements of the spring-mass system. For this reason, resonance is usually something to be avoided in mechanical or civil engineering structures. Technically, this called *exciting the structure at resonance*, and the results can be very exciting indeed – such as massive failure or collapse⁷⁶.

In an electrical resonant circuit, such as a series inductor and capacitor, the voltages across the resistor and capacitor are Q times the input voltage, as shown in section 4.11.2 on page 176.

4.13.4 Q Factor, Bandwidth and Filter Shape

In a high Q resonant system, the range of frequencies over which the system will resonate is quite narrow. That is, the frequency input vibration (from, say, a human shaking the upper end of the spring) must be very close to the resonant frequency in order to excite resonant behaviour in the system.

An example is shown in the resonant bandpass filter of figure 453 on page 530. The filter response narrows as the Q value increases. The same effect is not only evident in the LC resonant circuits we focus on here, but in other electronic circuits that exhibit similar behaviour. For example, the same effect can be seen in the lowpass filter of figure 449 on page 525 and the highpass filter of figure 456 on page 532. In all cases, the resonant peak becomes narrower as the Q factor increases.

In general, the Q factor is one of two parameters that define the frequency response curve of a filter. (The other is the resonant or cutoff frequency ω_o .) This is further explored in section 17.24 on page 570.

⁷⁶A dramatic example of resonance from civil engineering is the collapse of the Tacoma Narrows Bridge [61]. To prevent resonance, some machines and buildings are equipped with a damper that absorbs energy at the resonant frequency [62].

4.14 Bandwidth of a Resonant Circuit

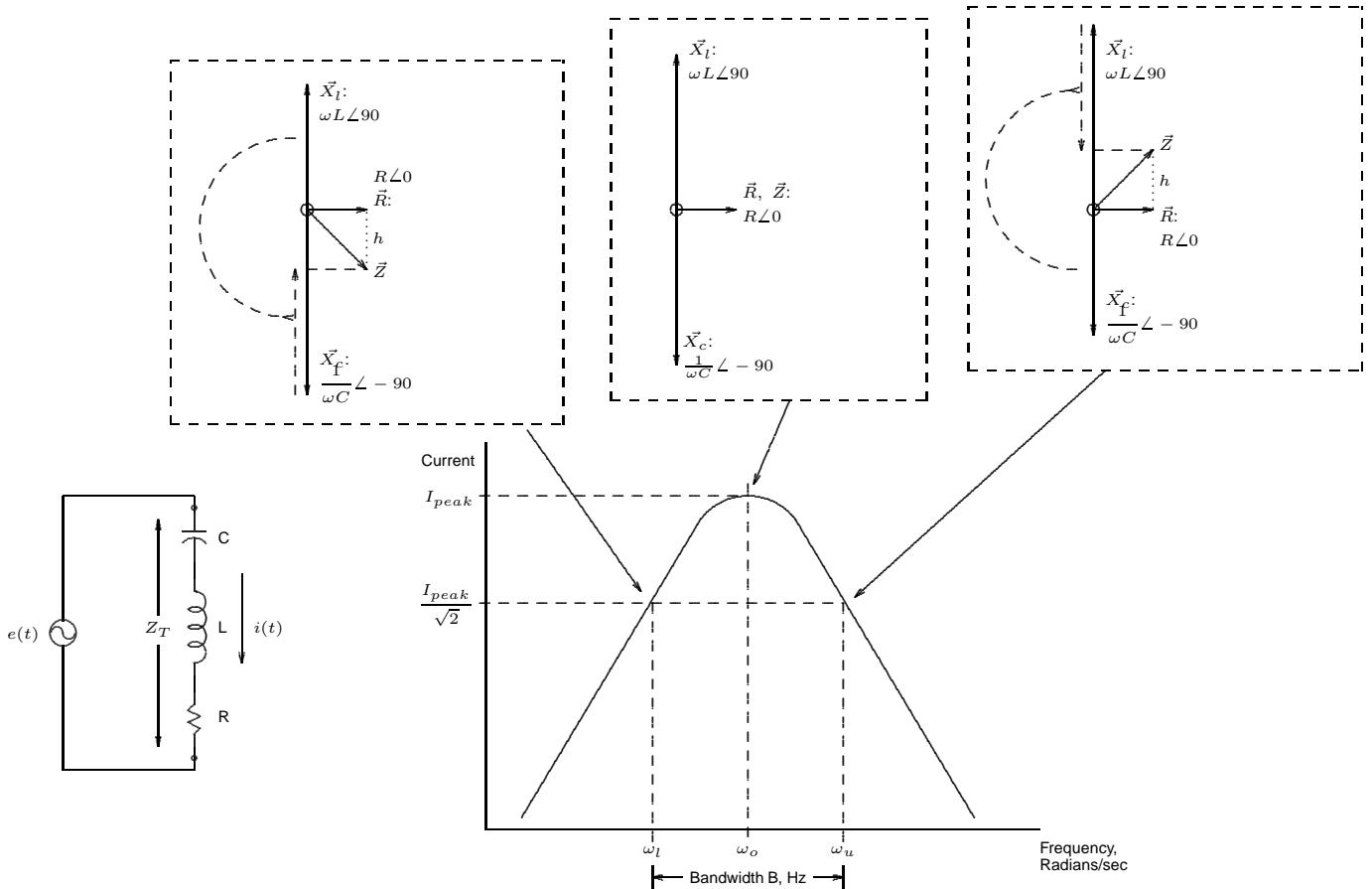


Figure 130: Bandwidth

In this section, we consider the *centre frequency* ω_o and bandwidth B of a resonant circuit. We'll use the series resonant circuit of figure 130 above for our example.

As the frequency approaches resonance, the circuit decreases in impedance as shown in figure 126 on page 126. If the circuit is driven by a constant-voltage AC source, then the current has the inverse shape. As shown in the plot of current vs frequency in figure 130 (the *frequency response curve*) the current peaks at the resonant frequency ω_o and decreases for frequencies above or below ω_o .

Three reactance vectors, corresponding to different points on the frequency response curve, are also shown on figure 130.

At resonance, as we showed in section 4.11 on page 174, the inductive and capacitive reactances cancel. The total impedance of the RLC circuit is simply the resistance. The current is at its largest value, $i_{peak} = e/R$, where e is the source voltage.

The *bandwidth B* is defined as the frequency span (in Hz) between the lower cutoff frequency ω_l and the upper cutoff frequency ω_u . A cutoff frequency is also known as a *half power point*, that is, where the power developed in the resistance is at half its peak value. Power is proportional to the square of current, so the current is equal to

$1/\sqrt{2}$ times the peak current, which occurs at resonance.

Notice the behaviour of the reactance vectors. For example, when the frequency is below resonance, the capacitive reactance vector X_c of the circuit in figure 130 increases and the inductive reactance vector X_l decreases. Above resonance, it's the reverse.

Now we can develop a relationship between the two cutoff frequencies ω_l and ω_u , and the resonant frequency ω_o . The key observation is to note that the curve is symmetrical. The current is the same value at the lower and upper half-power frequencies. Then the impedance vector \vec{Z} is the same in both cases. The resistance R is the same in both cases, so the value of h on the vector diagrams must be the same in both cases. That is, *the difference between the capacitive and inductive reactance at ω_l is the same as the difference between the inductive and capacitive reactance at ω_u* . In formula form:

$$\frac{1}{\omega_l C} - \omega_l L = \omega_u L - \frac{1}{\omega_u C} \quad (221)$$

We also have an equation for resonance:

$$\omega_o = \frac{1}{LC} \quad (222)$$

from which

$$L = \frac{1}{\omega_o^2 C} \quad (223)$$

Substitute for L from equation 223 into equation 221:

$$\frac{1}{\omega_l C} - \frac{\omega_l}{\omega_o^2 C} = \frac{\omega_u}{\omega_o^2 C} - \frac{1}{\omega_u C} \quad (224)$$

Capacitance C cancels out. Solving for ω_o results in:

$$\omega_o^2 = \omega_l \omega_u \quad (225)$$

$$\omega_o = \sqrt{\omega_l \omega_u} \quad (226)$$

In words, *the resonant frequency ω_o is the geometric mean of the lower and upper half-power frequencies ω_l and ω_u* .

4.14.1 Bandwidth and Q Factor

By a similar argument, we can determine a relationship between the bandwidth B , the resonant frequency ω_o , and the Q factor. Our strategy is to obtain an equation for each of the half power frequencies (ω_l, ω_u) in terms of the resonant frequency ω_o . Then the difference between these two frequencies is the bandwidth B .

First, consider the value of h in equation 130. The hypotenuse of the right-angle triangle is \vec{Z} . At the upper and lower half-power frequencies, the current is reduced by a factor of $\sqrt{2}$, so the impedance must be increased by $\sqrt{2}$. That is:

$$|Z| = \sqrt{2}R \quad (227)$$

That being the case, because this is a right-angle triangle, the sides of the triangle must both be equal to R , that is, $h = R$.

Now we can create two equations: in both cases, the difference between the capacitive and inductive reactance is equal to the circuit resistance R .

At the lower frequency ω_l , the equation is:

$$\frac{1}{\omega_l C} - \omega_l L = R \quad (228)$$

At the upper frequency ω_u , the equation is reversed:

$$\omega_u L - \frac{1}{\omega_u C} = R \quad (229)$$

We'll work with equation 228 and then extrapolate to equation 229. Equation 228 can be rearranged as:

$$1 - \omega_l^2 LC = \omega_l RC \quad (230)$$

Time for some substitutions:

From

$$\omega_o = \frac{1}{\sqrt{LC}} \quad (231)$$

we have

$$LC = \frac{1}{\omega_o^2} \quad (232)$$

Also, from

$$Q = \frac{1}{\omega_o RC} \quad (233)$$

we have

$$RC = \frac{1}{\omega_o Q} \quad (234)$$

Substitute for LC from equation 232 and for RC from equation 234, into equation 230. Then,

$$1 - \frac{\omega_l^2}{\omega_o^2} = \frac{\omega_l}{\omega_o Q} \quad (235)$$

Now solve for ω_l . After a certain amount of algebraic crank-turning, we can get the equation into the quadratic form:

$$\omega_l^2 + \omega_l \left(\frac{\omega_o}{Q} \right) - \omega_o^2 = 0 \quad (236)$$

This is similar to the quadratic equation

$$ax^2 + bx + c = 0 \quad (237)$$

where $x = \omega_l$, $a = 1$, $b = \omega_o/Q$ and $c = -\omega_o^2$. The standard solution to the quadratic equation 237 is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (238)$$

Then the solution to equation 236 is:

$$\omega_l = \frac{1}{2} \left[-\frac{\omega_o}{Q} \pm \sqrt{\frac{\omega_o^2}{Q^2} + 4\omega_o^2} \right] \quad (239)$$

Applying a similar process to equation 229 for the upper half-power frequency, we obtain:

$$\omega_u = \frac{1}{2} \left[\frac{\omega_o}{Q} \pm \sqrt{\frac{\omega_o^2}{Q^2} + 4\omega_o^2} \right] \quad (240)$$

We now apply the requirement that *frequency must have a positive value* to restrict these solutions⁷⁷. The quantity under the square root sign is larger than the first term. Consequently, in equation 239, the positive root must apply. In equation 240, the positive square root must also apply. Then our two equations are:

$$\omega_l = \frac{1}{2} \left[-\frac{\omega_o}{Q} + \sqrt{\frac{\omega_o^2}{Q^2} + 4\omega_o^2} \right] \quad (241)$$

Applying a similar process to equation 229 for the upper half-power frequency, we obtain:

$$\omega_u = \frac{1}{2} \left[\frac{\omega_o}{Q} + \sqrt{\frac{\omega_o^2}{Q^2} + 4\omega_o^2} \right] \quad (242)$$

The bandwidth B is the difference between these two frequencies. Subtracting equation 241 from equation 242, we have:

$$\begin{aligned} B &= \omega_u - \omega_l \\ &= \frac{1}{2} \left[2 \frac{\omega_o}{Q} \right] \\ &= \frac{\omega_o}{Q} \end{aligned} \quad (243)$$

Or, in a form that is easier to remember:

$$Q = \frac{\omega_o}{B} \quad (244)$$

That is, a high Q resonance has a smaller bandwidth.

4.15 The Compensated Attenuator

A voltage divider is commonly used at the front end of a voltmeter or oscilloscope to attenuate the input signal to a suitable range for the measuring circuitry. It's referred to as the *voltage attenuator* stage. When a voltage divider is used at high frequencies, the stray capacitance of the load may be a problem. The schematic is shown in figure 131(a). The load capacitance might be the junction capacitance of an active device such as a BJT or FET, or it might just be wiring capacitance. It's typically about 20pF.

We'll assume that the load resistance R_L is large enough to be ignored. Then the equivalent circuit is as shown in figure 131(b). The internal resistance R_{int} of the voltage divider and the load capacitance C_L form a lowpass filter with cutoff frequency $\omega_o = 1/R_{int}C_L$. The frequency response will decline at 20db per decade above that frequency.

In effect, the voltage division ratio has become a function of frequency, when we would prefer it to be frequency independent. To accomplish this, we need a *compensation* capacitance C_c across the upper resistor in the divider, figure 131(c).

If the capacitance C_c is adjusted such that $R_1C_c = R_2C_L$, then it turns out that the frequency response is frequency-independent.

The capacitor C_c is often a small screwdriver adjustment at the base of the oscilloscope probe⁷⁸. A square wave signal is applied to the input of an oscilloscope channel via a $\times 10$ probe. Figure 132(a) shows too large C_c , figure 132(b) shows correct adjustment, and figure 132(c) shows too small C_c . The probe screwdriver adjustment is changed until the scope trace resembles figure 132(b).⁷⁹

⁷⁷A tip of the analog circuit hat to John Foster, who clarified this point.

⁷⁸Sometimes there is a locking ring on the probe barrel. Unlock the ring and then adjust the barrel.

⁷⁹Images from Syscomp CGR-101 oscilloscope and generator.

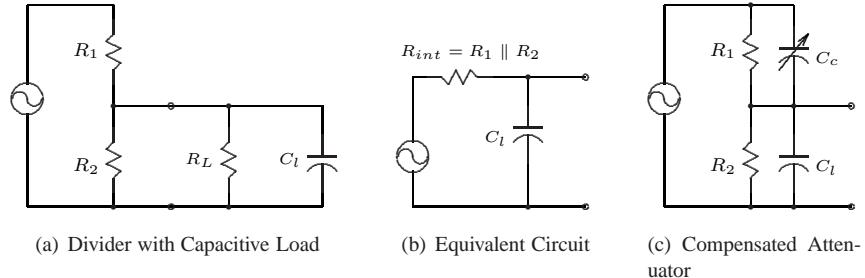


Figure 131: The Compensated Divider

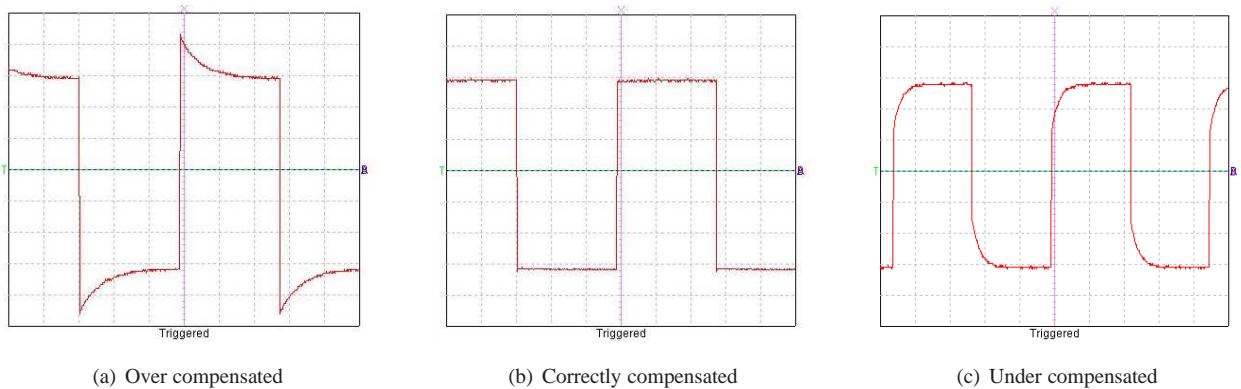


Figure 132: Compensation Adjustment

Analysis

Assume that the voltage divider consists of Z_1 and Z_2 . The upper part, Z_1 , is composed of $R_1 \parallel C_1$. The lower part, Z_2 , is composed of $R_2 \parallel C_2$.

Then the voltage division ratio K (the ratio of output to input signal) is equal to

$$K = \frac{Z_2}{Z_1 + Z_2} \quad (245)$$

where

$$\begin{aligned} Z_1 &= R_1 \parallel 1/sC_1 \\ &= \frac{R_1}{1 + sR_1C_1} \end{aligned} \quad (246)$$

$$(247)$$

and

$$\begin{aligned} Z_2 &= R_2 \parallel 1/sC_2 \\ &= \frac{R_2}{1 + sR_2C_2} \end{aligned} \quad (248)$$

(We are using the Laplace Notation for impedance, where $s = j\omega$ as a form of shorthand, section 5.1.1, page 197.)

Substitute for Z_1 and Z_2 in equation 245:

$$\begin{aligned} K &= \frac{\frac{R_2}{1 + sR_2C_2}}{\frac{R_1}{1 + sR_1C_1} + \frac{R_2}{1 + sR_2C_2}} \\ &= \frac{R_2(1 + sR_1C_1)}{R_1(1 + sR_2C_2) + R_2(1 + sR_1C_1)} \end{aligned} \quad (249)$$

If we make

$$R_1C_1 = R_2C_2 \quad (250)$$

then equation 249 simplifies to

$$K = \frac{R_2}{R_1 + R_2} \quad (251)$$

which is independent of frequency, since it contains no s terms.

4.16 The Ideal Transformer

Small transformers are useful in signal coupling between amplifier stages, for DC isolation of signals, and in generating different power supply voltages from a single supply.

The transformer consists of two or more coils of wire, known as *windings*, wound around a magnetic *core*. The core is usually a closed magnetic path, without any gaps, so that the magnetic flux level in the core can be as large as possible.

The schematic of a two-winding transformer is shown in figure 133.

Voltage Ratio

When a source of alternating voltage is connected to the primary of the transformer, it creates a time-varying magnetic field in the core according to Lenz's Law:

$$e_p = N_p \frac{d\phi}{dt} \quad (252)$$

where e_p is the primary voltage, N_p is the number of turns on the primary winding, and ϕ is the magnetic flux.

This same flux generates a voltage in the secondary winding:

$$e_s = N_s \frac{d\phi}{dt} \quad (253)$$

Substituting for $\frac{d\phi}{dt}$ from equation 253 in equation 252, we have

$$\frac{e_p}{e_s} = \frac{N_p}{N_s} \quad (254)$$

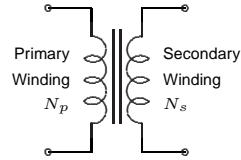


Figure 133: Ideal Transformer

The voltage is stepped up or down from primary to secondary in the same ratio as the number of turns.

Current Ratio

The ideal transformer is a lossless device⁸⁰, that is

$$P_{in} = P_{out} \quad (255)$$

so

$$\begin{aligned} e_p i_p &= e_s i_s \\ \frac{e_p}{e_s} &= \frac{i_s}{i_p} \end{aligned} \quad (256)$$

Substitute $\frac{N_p}{N_s}$ for $\frac{e_p}{e_s}$ from equation 254, and we have

$$\frac{i_p}{i_s} = \frac{n_s}{n_p} \quad (257)$$

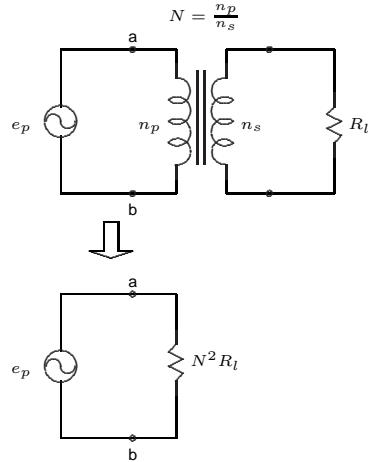
The current is stepped up or down from primary to secondary in the inverse of the turns ratio. For example, a transformer that steps the voltage down by a factor of 10 will step the current up by a factor of 10.

Resistance Ratio

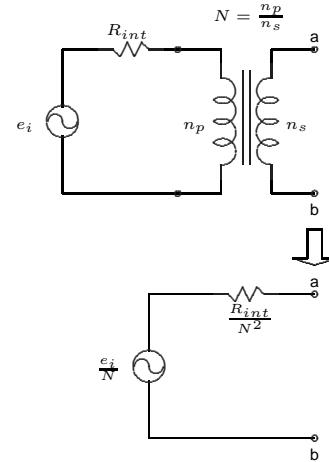
Consider that a transformer has a turns ratio

$$N = \frac{N_p}{N_s} \quad (258)$$

The secondary of the transformer is connected to a load resistance R_l and the primary is driven by a voltage source e_p (figure 134(a)).



(a) Resistance Reflection



(b) Source Reflection

Figure 134: Resistance and Source Reflection

⁸⁰Practical transformers have efficiencies less than but close to 100%.

The secondary current is

$$i_s = \frac{e_s}{R_l} \quad (259)$$

But

$$\begin{aligned} e_s &= e_p/N \\ i_s &= N i_p \end{aligned} \quad (260)$$

Substitute these in equation 259, and

$$\begin{aligned} N i_p &= \frac{e_p}{N R_l} \\ \frac{e_p}{i_p} &= N^2 R_L \end{aligned} \quad (261)$$

But e_p/i_p is the effective resistance seen at the primary of the transformer, so

The load resistance is multiplied by N^2 to reflect it into the primary.

Based on this concept, transformers are often specified in terms of their resistance ratio, which is equal to the square of the turns ratio. For example, a transformer identified as '10kΩ:600Ω' has a turns ratio such that a 600Ω resistor across the secondary will appear as a 10kΩ resistor across the primary. Then the turns ratio is:

$$\begin{aligned} N &= \sqrt{\frac{10k\Omega}{600\Omega}} \\ &= 4.08 \end{aligned}$$

Reflecting a Source into the Secondary

The same concept of *reflecting* a device through the transformer can be applied when moving a source into the secondary.

A source of e_i volts and internal resistance R_{int} , connected to the primary winding, is equivalent to a source e_i/N with an internal resistance of R_{int}/N^2 , figure 134(b).

A Floating Source

There are situations where a *floating* source is required, that is, neither end of the source voltage is connected to ground. The secondary winding of a transformer is such a source. Either end of the secondary may be connected to ground to select the phase of the secondary voltage with respect to the primary. Or one end of the secondary may be connected to some other voltage (see section 3.7 for an example).

Transformer Polarity

The transformer is an alternating current device and the commonly used term *transformer polarity* refers more precisely to the issue of *transformer winding phase*.

In some circumstances, the relative phase of different windings is unimportant. However, there are other circumstances where the relative phase of different windings is critical. The phase relationships could be between primary and secondary, or between multiple windings on the primary or secondary.

When phase is important, the transformer windings are each marked with a *dot*. At any given instant in time, all the dotted terminals have the same voltage polarity. For example, suppose a dotted terminal on the primary is at some positive voltage. Then the dotted terminal of the secondary is also at some positive voltage, possibly of a different magnitude. Notice that current is flowing into the primary winding and out of the secondary winding.

Figure 135 shows an example of a transformer with multiple windings, Signal A41-80-230. Each winding is rated at 117VAC, 40VA (volt-amps). This is a very versatile device. For example, primary P_1 could be powered from 117VAC. The other three windings then become isolated (floating) sources of 117VAC.

Two common applications are shown in figure 136.

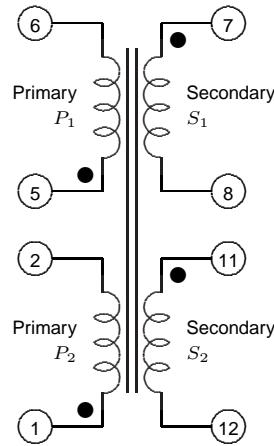
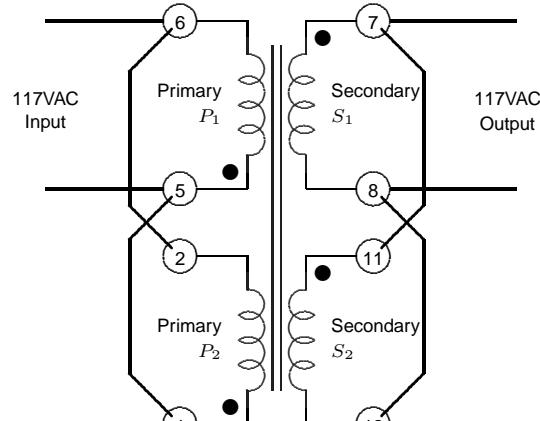
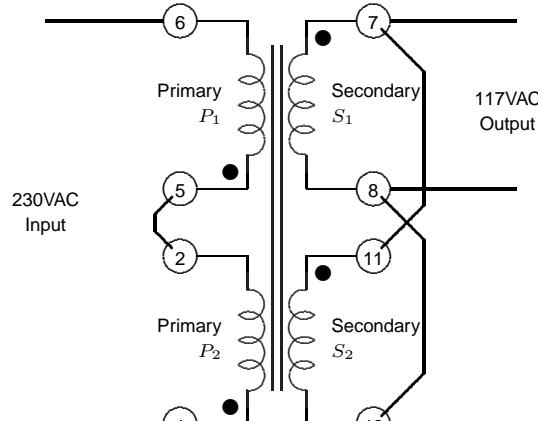


Figure 135: Transformer with Multiple Windings



(a) 117VAC Isolation



(b) 230VAC Step Down

Figure 136: Winding Connections

Figure 136(a) shows the transformer configured to operate as a 117VAC isolation transformer. Both primaries and secondaries are wired in parallel so that the total output is 80VA, double the individual winding output. In this configuration, it is critical to connect the dotted terminals together and the undotted terminals together, as shown in the figure. If the connection is reversed on the primary so that dot is connected to undot the polarities of the core flux due to the windings will oppose each other and the core fluxes will cancel. The result will be extremely

large currents in the primary windings which will destroy the transformer.

Figure 136(b) shows the transformer configured to operate as a 230VAC to 117VAC step-down transformer⁸¹. In this configuration the primary windings are operated in series and the dot connections ensure that the core fluxes add.

4.17 Impedance

First, let's review the behaviour of a capacitor or inductor in an AC circuit.

Previously in this section we showed how resistors, capacitors and inductors can be regarded as vector quantities. If a resistance has a vector angle of zero degrees, capacitive reactance has an angle of -90° and inductive reactance has an angle of $+90^\circ$. Then we can do calculation of, say, current using Ohm's law (and other circuit rules) and the rules of vector arithmetic (aka complex math, section 4.4).

Example

An AC voltage of frequency 60Hz and 10V peak is applied to a $2\mu\text{F}$ capacitor. What is the current in the circuit?

Solution

- Determine the frequency in radians per second.

$$\begin{aligned}\omega &= 2\pi f \\ &= 2 \times 3.14 \times 60 \\ &= 377 \text{ radians/sec}\end{aligned}$$

- Determine the magnitude of the capacitive reactance.

$$\begin{aligned}X_c &= \frac{1}{\omega C} \\ &= \frac{1}{377 \times (2 \times 10^{-6})} \\ &= 1326\Omega\end{aligned}$$

(If two significant figures are sufficient, the reactance chart or slide rule give the capacitive reactance without calculation, as shown in section 4.7.)

- Determine the current, using Ohm's law. We indicate that the current and voltage are AC quantities by using lower case letters: e and i . The resistance R is replaced by the capacitive reactance X_c .

We take the voltage phase angle as the reference of zero degrees, so in vector form it is $10\angle 0$. The capacitive reactance is the magnitude we previously calculated, at an angle of -90 degrees.

$$\begin{aligned}i &= \frac{e}{X_c} \\ &= \frac{40\angle 0}{1326\angle -90} \\ &= 0.030\angle +90\end{aligned}$$

⁸¹This might be useful in Europe to operate a 117VAC North-American device from 230VAC power. However, check that the AC frequencies are similar. Some supply frequencies are lower (50 vs 60 Hz, for example) and that may require decreasing the VA rating of the transformer.

In words, the current has a peak value of 30mA and a phase angle that leads the voltage by 90 degrees.

This analysis of an AC circuit totally avoids the differential equation of capacitor behaviour and allows us to use the tools of circuit analysis that apply to a DC circuit.

4.17.1 Impedance

To this point, we have been treating the capacitor and inductor as pure reactances. In practice, we frequently find a circuit element that contains both resistance and reactance. This is known as an *impedance*.

Like reactance, impedance is a vector quantity. A pure reactance vector has some magnitude at an angle of $+90^\circ$ or -90° . An impedance vector has some magnitude at some arbitrary angle. These properties are summarized in the table of figure 137.

Name	Symbol	Polar Notation	Rectangular Notation
Resistance		$R\angle 0^\circ$	$R + j0$
Capacitance		$X_c \angle -90^\circ$	$0 - jX_c$
Inductance		$X_l \angle +90^\circ$	$0 + jX_l$
Impedance	or 	$Z\angle\theta^\circ$	$R \pm jX$

Figure 137: Impedance Summary

Referring to this table:

- A resistance (row 1) may be regarded as an impedance without any reactive component, so its angle is zero and its quadrature component (the j factor) is zero.
- A capacitance (row 2) may be regarded as an impedance without any resistive component, and a fixed angle of -90° . The magnitude of the impedance is the capacitive reactance X_c , where $X_c = \frac{1}{\omega C}$, $\omega = 2\pi f$, where C is the capacitance in farads and f is the frequency of the AC signal in Hz.
- An inductance (row 3) may be regarded as an impedance without any resistive component, and a fixed angle of $+90^\circ$. The magnitude of the impedance is the inductive reactance X_l , where $X_l = \omega L$, $\omega = 2\pi f$, where L is the inductance in henries and f is the frequency of the AC signal in Hz.
- An impedance (row 4) consists of a resistance and inductive or capacitive reactance⁸².

In polar notation, we describe the impedance triangle by its hypotenuse and angle (figure 109 on page 160). The magnitude of the impedance Z (the hypotenuse of the impedance triangle) is given by

⁸²The table shows these elements in series, but they can equally well be in parallel.

$$|Z| = \sqrt{R^2 + X^2}, \text{ where } X \text{ is either } X_c \text{ or } X_l.$$

The angle θ is given by

$$\angle Z = \theta = \tan^{-1}(X/R) \text{ where } X \text{ is either } X_c \text{ or } X_l.$$

In rectangular notation, we describe the impedance triangle by its horizontal and vertical components, the resistance R and the reactance X .

Strictly speaking, we should distinguish in notation between the impedance vector \vec{Z} (notice the arrow to indicate vector quantity) and the magnitude of the impedance, $|Z|$. In practice this is not done, and the distinction is (hopefully) evident from the context.

It's not necessary to consider an impedance consisting of a resistance in series with a capacitance and inductance. The capacitive and inductive reactances partially cancel, resulting in a net single reactance that is capacitive or inductive.

Example

Determine the magnitude and phase (with respect to the voltage) of the current in the circuit of figure 138.

Solution

We'll use Ohm's law to solve for the current.

- Convert the frequency in Hz to radians per second.

$$\begin{aligned}\omega &= 2\pi f \\ &= 2 \times 3.14 \times 4000 \\ &= 25120 \text{ radians/sec}\end{aligned}$$

- Determine the reactance of the inductor.

$$X_L = \omega L = 25120 \times (90 \times 10^{-3}) = 2260\Omega$$

In rectangular form, the impedance is then:

$$Z = R + jX_L = 1300 + j2260\Omega$$

- In order to use this in the Ohm's law equation, we'll need the impedance in polar form:

$$Z \angle \theta = \sqrt{R^2 + X_L^2} \angle \tan^{-1} \left(\frac{X_L}{R} \right) = \sqrt{1300^2 + 2260^2} \angle \tan^{-1} \left(\frac{2260}{1300} \right) = 2607 \angle 60^\circ \Omega$$

- Now we can execute Ohm's law:

$$i = \frac{e}{Z} = \frac{6 \angle 0}{2607 \angle 60^\circ} = 2.3 \times 10^{-3} \angle -60^\circ$$

Notice that the voltage phase angle is zero degrees because it is the reference vector. The current is 2.3mA peak-peak, lagging by 60° behind the voltage.

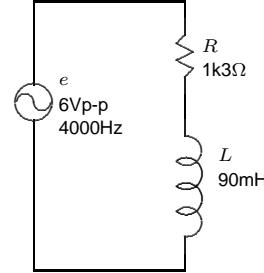


Figure 138: Impedance, Example

4.17.2 Using Impedance

In analog circuit design, the concept of impedance is important, but it is rare to do the sort of calculation shown on page 192. In general, analog circuits deal with a range of frequencies and so an exact calculation of impedance, which is valid at only one particular frequency, is not all that useful.

Impedance is a *much* more useful concept in power systems, which operate at a single frequency (60Hz in North America). So, for example, it is quite feasible to add reactance to a power system to change the phase relationship between current and voltage.

However, although precise calculations involving impedance are not all that common, *impedance* is a very common term in analog circuit design⁸³. That's because impedance is a general catch-all term for resistance, reactance and some combination of the two. For example, it's very common to speak of the *output impedance of a signal generator* where the impedance is inevitably a resistance (typically 50 ohms). More usefully, one speaks of *loudspeaker impedance*. This is reasonable, since the impedance of a moving coil loudspeaker varies dramatically over the audio frequency range – typically a combination of resistance and inductance, although there can be capacitive reactance as well. Consequently, referring to a circuit element as an impedance is safe but not all that informative.

4.17.3 Impedance Matching

There are situations, particularly in radio frequency design, where a particular source impedance must be matched by a specific load impedance. For example, the input of an amplifier stage may appear as a complex impedance, some mix of resistance and reactance. If we require maximum power transfer from the driver stage into the amplifier, then the impedances must be matched in some fashion.

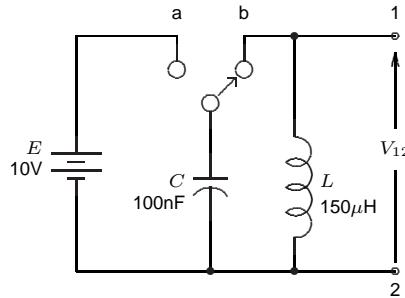
Impedance matching is also a requirement where a fast risetime pulse drives a transmission line such as a coaxial cable. A coaxial cable has a *characteristic impedance*, typically 50 or 75 ohms resistive. To avoid reflections, the source driving the cable should have an internal impedance equal to the cable impedance. As well, the receiver at the other end of the cable should present the cable with the same impedance. As the clock frequency of digital circuits rises into the stratosphere, this becomes an important issue in the design of printed circuit boards, where a PCB trace has a certain characteristic impedance [29].

Apart from those two situations, impedance matching is not a requirement in most analog circuit design. Typically, the source impedance is made much smaller than the load impedance so that there is a maximum transfer of a voltage signal.

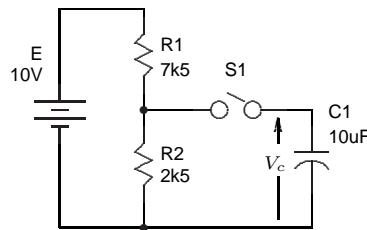
4.18 Exercises

1. A $1\mu\text{F}$ capacitor is placed in parallel with an ideal $100\mu\text{H}$ inductor.
 - (a) What is the resonant frequency in Hz?
 - (b) Does the parallel (tank) circuit look like a high impedance or low impedance at resonance?
 - (c) If the inductance has a resistance of 0.1Ω , what is the Q factor of the tank circuit?
2. For the circuit shown below, both components are ideal (lossless). The switch is initially in the *a* position and then moved to *b* at $t = 0$. Sketch the waveform of V_{12} versus time for the interval just after the switch is moved to *b* until it begins to repeat itself.

⁸³In this text, it's used some 340 times.



3. For the circuit shown, the capacitor is initially uncharged.



- (a) After the switch is closed, to what final voltage does the capacitor C_1 ultimately charge?
 (b) What is the charging time-constant τ ?
4. A certain transformer, obtained in the basement of Active Surplus, is labelled as $117V:12.6V, 1\text{ Amp}$. With our trusty digital voltmeter, we measure the winding resistances. The primary resistance R_p is 44Ω , the secondary resistance R_s is 1.3Ω . The transformer is to be connected to a load resistance R_L of 15Ω . The transformer primary is connected to a source of 117VAC line voltage, and the secondary is connected to the load resistance.
- (a) Determine the transformer turns ratio.
 (b) Draw an equivalent circuit of the transformer and its load resistance, showing the line voltage, primary resistance, the ideal transformer with its turns ratio, the secondary resistance and the load resistance.
 (c) Calculate the RMS voltage V_L across the load resistance.
5. A transformer is labelled as $117V:12.6V, 1\text{ Amp}$. If the transformer primary is connected to a source of 117VAC power, and the secondary is connected to a 100Ω load, determine:
- (a) The secondary voltage.
 (b) The secondary current.
 (c) The primary current.
6. An inductor of 100mH is in parallel with a resistance of $1\text{k}\Omega$. The circuit is operated at a frequency of 1600Hz .
- (a) Determine the impedance of this network.
 (b) The original network is to be transformed into a series network containing a resistor and inductor. Determine the values for the resistance and inductance.
 (c) If the frequency changes, do the parallel and series networks continue to be equivalent? Explain.

5 The Laplace Transform

Time is valuable in the practice of Engineering, and Engineers value tools which facilitate getting to an answer and a final design. The slide rule and nomograph are early examples of general-purpose engineering tools. The spreadsheet, simulation program (such as *Spice*) and the *Computer Algebra Program* are more recent additions to the toolkit. The Laplace Transform is a mathematical tool for solving differential equations.

In this section, we'll show a *cookbook* approach to the Laplace transform, that is, with sufficient background for applications in engineering, but without the rigor required by pure mathematics⁸⁴.

5.1 Overview

The Laplace transform is a technique for circuit analysis that facilitates the calculation of a network time response. An electrical network exists. A certain input signal is applied to the network. What is the output from the network as a function of time? This type of calculation can be done entirely in the time domain, but it requires solving differential equations, which is challenging and time-consuming.

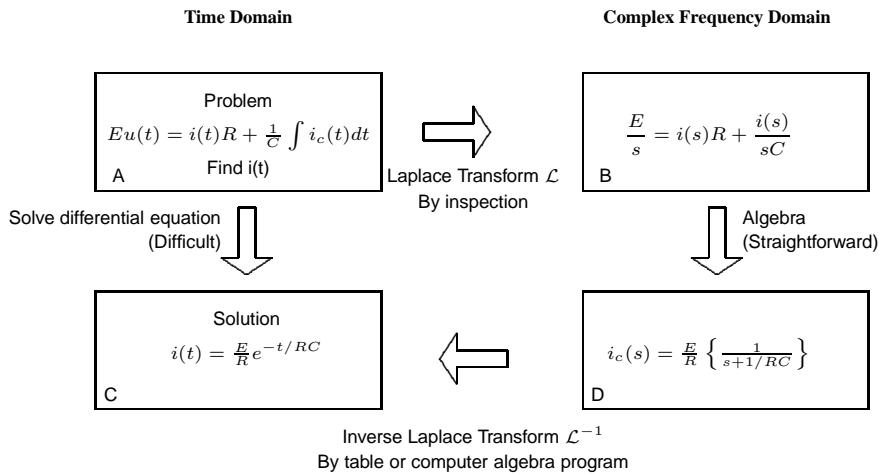


Figure 139: Laplace Transform Method

The process is represented in figure 139.

The problem equation is in block A in time-domain format. (The details of this equation are unimportant at this point.) Moving directly to a solution (block C) requires solving a differential equation, which is difficult. Alternatively, we use the Laplace transform and its inverse to outflank the problem, moving through blocks B and D and then to the solution in block C.

1. First, we move the equation from block A to block B. We thereby convert the time-domain equation into an equation in what is known as *the complex-frequency domain*. It turns out that this is relatively simple and can be done by inspection. The table shows common examples:

⁸⁴ This section was originally written as the *Laplace Transform Cookbook* available at <http://www.syscompdesign.com/AppNote.html>. My colleagues Glen Martinson and John Foster contributed significantly to this section. Glen showed me how to use the Laplace transform, and John did a very thorough review of the document. Of course, any errors remain my responsibility.

Component	Time Domain	Laplace Domain
Step voltage of E volts	$Eu(t)$	E/s
Resistor	R	R
Capacitor	C	C/s
Inductor	L	sL

2. Next, we solve for the quantity of interest. In figure 139, it's the current in the circuit. We also manipulate the equation into a form that is listed in a table of Laplace transforms. This may require *completing the square* or *partial fraction expansion* of the equation.

This puts us in block D of figure 139.

3. We find some comparable entry in the transform table, which determines the time-domain equivalent. This moves us to the solution in block C.

Alternatively, a computer algebra program such as Maxima saves the labour of manipulating the equation into the form of a table entry. It does the grunt work of algebraic manipulation and finds the inverse Laplace transform directly, effectively moving the process directly from block B to C.

The Laplace transform technique is a huge improvement over working directly with differential equations.

Case History

To set the stage for this next circuit and analysis, suppose that we wish to design a filter for noise pulses that occur on a power supply line, figure 140. We would like to attenuate this noise pulse sufficiently with as low cost a circuit as feasible⁸⁵. The noise pulse can be modelled as an *impulse*, a pulse that is short compared to the time-constants in the circuit.

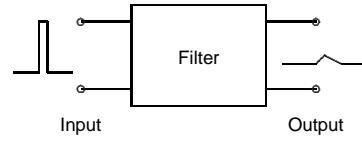


Figure 140: Impulse Noise Filter

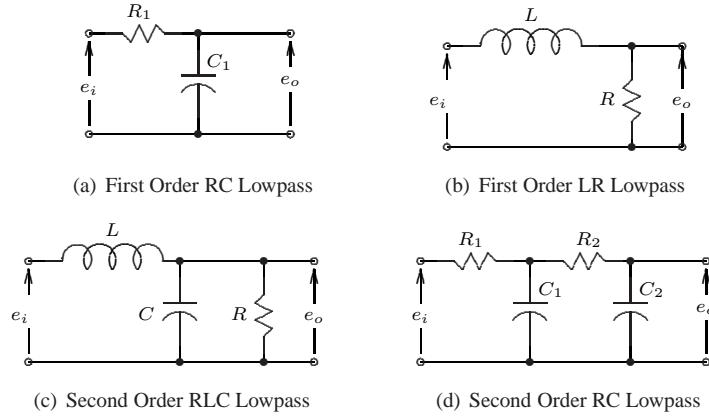


Figure 141: Noise Filter Candidates

The filter candidates are shown in figure 141.

⁸⁵In a final design, this vague description must be tightened up to specify the actual noise reduction and the resultant cost.

Each of these filters can be considered to be a voltage divider, in which the output voltage is some fraction of the input. The impedance of one or more components is variable with frequency. The net effect is to increase the voltage division ratio at high frequencies, that is, reduce the output signal at high frequencies. For example, in figure 141(a), the bottom arm of the voltage divider is a capacitor. At high frequencies, the impedance of the capacitor decreases and so also the output signal.

The single section RC filter is least expensive. The dual section RC filter requires 4 inexpensive components, but is a more effective filter. The RLC filter uses 3 components, one of which is a more expensive inductor. However, if the RLC filter has a significant advantage in performance, it might be preferable.

To do a meaningful comparison, we need to know the frequency response and the impulse response of each of these filters. First, some explanation of tools that can help get us there.

5.1.1 Laplace Notation as Shorthand

Very simply, the Laplace transform substitutes s , the Laplace transform operator for the differential operator d/dt . Then the s term may be manipulated like any other variable.

Thus one will see s in a control system block to indicate *differentiator* and $1/s$ to indicate *integrator*.

The substitution of s for d/dt leads to another one, s for $j\omega$. This is useful in determining the transfer function of an electrical network, and then its magnitude and phase responses.

This approach can be used without any real understanding of Laplace transform theory. It can be regarded simply as a kind of shorthand. Nonetheless, it's a useful technique when the requirement is to determine the *frequency response* of some network. In section 5.3 (page 211) we'll look at obtaining the *time-domain response*, which does require more in-depth understanding.

5.1.2 A Simple Example: Capacitor Charging Equation

As a first introduction to the Laplace operator s as differentiation and integration, consider the operation of a capacitor. The basic differential equation relating voltage and current in a capacitor is⁸⁶:

$$i(t) = C \frac{d}{dt} v(t) \quad (262)$$

where:

- $i(t)$ Current in the capacitor, amps, as a function of time
- $v(t)$ Voltage across the capacitor, volts, as a function of time
- C Capacitance, farads

In words, the time-varying current into a capacitor is proportional to the rate of change of the voltage across its terminals. The constant of proportionality is the capacitance C .

To apply the Laplace transform to this equation, we replace the differential operator d/dt by s and the voltage and current by their transformed versions:

⁸⁶A Note on Notation: It is common to indicate time domain quantities as lower case such as v or $v(t)$. We'll use the latter. Common practice is also to use upper case for the Laplace domain (aka the *complex frequency domain*), such as V . However, sometimes upper case is used for DC quantities and lower case for AC quantities, so we will use $v(s)$ for Laplace domain to make it very clear. British practice is to use p for s .

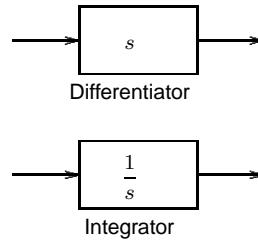


Figure 142: Differentiator and Integrator

$$\begin{aligned} i(s) &= C s v(s) \\ &= sC v(s) \end{aligned} \quad (263)$$

where:

- $i(s)$ Current in the capacitor, amps, in the Laplace domain
- $v(s)$ Voltage across the capacitor, volts, in the Laplace domain
- C Capacitance, farads

Let's reverse this and solve for the capacitor voltage:

$$v(s) = \frac{1}{C} \frac{1}{s} i(s) \quad (264)$$

Now, back to the time domain: voltage and current transform to their time-dependent values and $1/s$ becomes an integral:

$$v(t) = \frac{1}{C} \int i(t) dt \quad (265)$$

This is the long way round. Equation 265 could be obtained directly by inspection of equation 262. However, it shows a very simple application of the Laplace transform: we transformed the original equation into the Laplace domain, manipulated it, and then transformed the result back into the time domain⁸⁷.

5.1.3 Inductor Differential Equation

A similar reasoning process can be applied to the inductor. The basic differential equation relating voltage and current in an inductor is:

$$v(t) = L \frac{di(t)}{dt} \quad (266)$$

where:

- $v(t)$ Voltage across the inductor, volts
- $i(t)$ Current in the inductor, amps
- L Inductance, Henries

Proceeding as we did in the case of capacitance, we replace the differential operator d/dt by s and the voltage and current by their transformed versions:

$$\begin{aligned} v(s) &= L s i(s) \\ &= sL i(s) \end{aligned} \quad (267)$$

The reactance of the inductor (analogous to resistance, but affecting AC current only) is the ratio of voltage to current:

$$\frac{v(s)}{i(s)} = sL \quad (268)$$

Inductors may then be represented by inductive reactance as sL_1, sL_2 and so on.

⁸⁷The result of equation 265 assumes that there is no initial charge on the capacitor. If there is charge on the capacitor at the beginning of the interval of integration, it must be accounted for by an additional term which we might call v_o .

5.1.4 Capacitive Reactance

The reactance of the capacitor is the ratio of voltage to current:

$$\frac{v(s)}{i(s)} = \frac{1}{sC} \quad (269)$$

It's common to represent the capacitive reactance directly on a circuit diagram, so one sees capacitors labelled as $1/sC_1$, $1/sC_2$ and so on. For the purpose of circuit analysis these reactances may be treated as resistances.

The magnitude and phase of capacitive reactance are also represented as

$$X_c = \frac{1}{j\omega C} \quad (270)$$

where the variables are:

- X_c Capacitive reactance, ohms
- j Imaginary operator, $\sqrt{-1}$
- ω Circular frequency, radians/sec

Comparing equations 269 and 270, it can be seen that the Laplace transform operator s can be treated simply as shorthand notation for $j\omega$.

5.1.5 Summary

- The differential operator d/dt in a differential equation can be replaced by s . ⁸⁸
- Integration is the inverse: $1/s$. ⁸⁹
- The Laplace operator s may also be regarded as shorthand for the expression $j\omega$.
- Capacitive reactance is $1/sC$.
- Inductive reactance is sL .

5.2 Transfer Function and Frequency Response

Next, we'll develop the magnitude and phase response curves for the lowpass networks of figure 141.

5.2.1 Transfer Function of Low Pass RC Filter

A simple RC lowpass filter is shown in figure 143, where the capacitor is indicated by its reactance $1/sC$. Let us determine the frequency response of this filter.

The frequency response shows the relationship between output voltage and input voltage as a function of frequency.

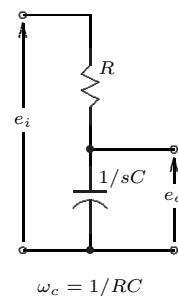


Figure 143: RC Lowpass Filter

⁸⁸This is straightforward to extend: double differentiation d^2/dt^2 becomes s^2 , and so on.

⁸⁹Similarly, double integration becomes $1/s^2$, and so on.

Consequently, a first step is to determine the relationship between e_o and e_i . The resistor and the reactance of the capacitor form a voltage divider, so we can write:

$$\frac{e_o}{e_i} = \frac{Z_2}{Z_1 + Z_2} \quad (271)$$

where:

- e_o AC Output voltage from the circuit
- e_i AC Input voltage to the circuit
- Z_1 Impedance of the upper half of the voltage divider (the resistor)
- Z_2 Impedance of the bottom half of the voltage divider (the capacitor)

Strictly speaking, e_o should be written as $e_o(\omega)$ or $e_o(f)$ to indicate that the value is a function of frequency, but we'll take that as understood.

Now substitute R for Z_1 , $1/sC$ for Z_2 and do some algebra:

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{1/sC}{R + 1/sC} \\ &= \frac{1}{1 + sRC} \end{aligned} \quad (272)$$

Now we need to introduce some new labels. The quantity RC is important in these circuits: it is known as the *time constant* τ and will turn up again when we look at the time-domain response of the filter.

$$\tau = RC \quad (273)$$

Then we could rewrite equation 272 this way:

$$\frac{e_o}{e_i} = \frac{1}{1 + s\tau} \quad (274)$$

We can do even better than this. In the frequency domain RC is related to the *corner* or *cutoff frequency* of the filter, which is referred to as ω_o in radians/sec notation or f_o in Hertz (cycles/second) .

$$\begin{aligned} \omega_o &= \frac{1}{\tau} \\ &= \frac{1}{RC} \end{aligned} \quad (275)$$

So equation 274 could be written as:

$$\frac{e_o}{e_i} = \frac{1}{1 + s/\omega_o} \quad (276)$$

This turns out to be a useful form of the equation as we'll see in a second. It is known as the *transfer function* or *characteristic equation* of the RC lowpass network.

Of course, frequency in radians/second and Hertz are related

$$\omega_o = 2\pi f_o \quad (277)$$

So if you prefer your frequencies in Hz, you could rewrite equation 276 as:

$$\frac{e_o}{e_i} = \frac{1}{1 + s/2\pi f_o} \quad (278)$$

In other words, we have three ways to define the *cutoff frequency*: by time-constant τ , by cutoff frequency ω_o in radians per second, and by cutoff frequency f_o in Hertz. All three are in use. Control system people like time constants. Electrical engineers use radians/second. Audio folks like Hertz.

We'll stick with the form of equation 276.

5.2.2 Magnitude and Phase of The Lowpass RC Filter

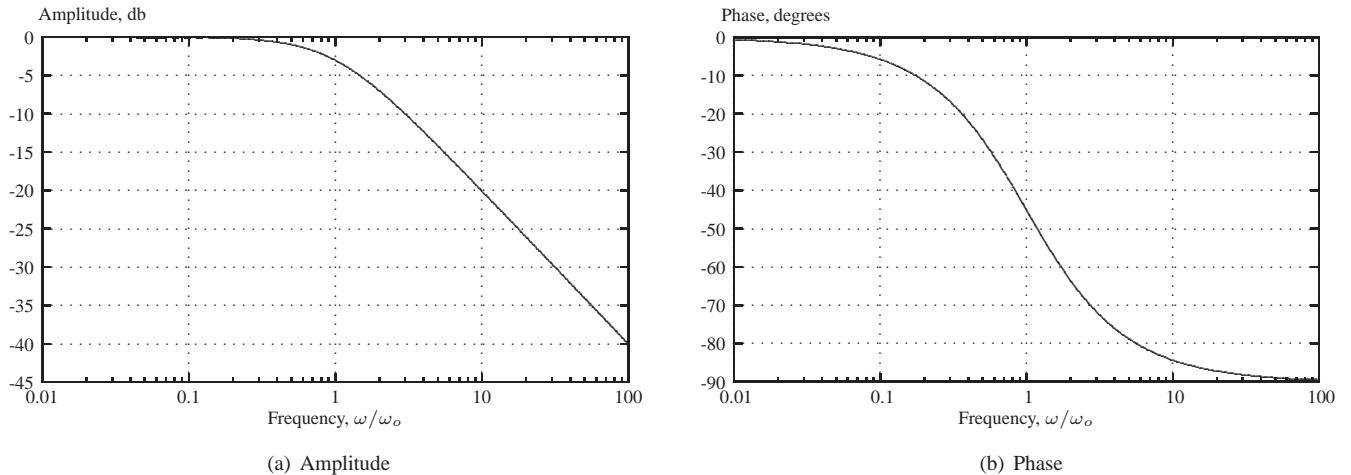


Figure 144: Amplitude and Phase of RC Lowpass Network

Now we'll put the transfer function of equation 276 into a suitable form for plotting the magnitude and phase response. To do that, we substitute $j\omega$ for s in the transfer function:

$$\frac{e_o}{e_i} = \frac{1}{1 + j\omega/\omega_o} \quad (279)$$

Now, the denominator is a complex number with a real component 1 pointing along the X axis and a quadrature part ω/ω_o pointing up the Y axis. (See section 4.4 on page 158 for a review of complex numbers.)

We need to convert this complex number from its current rectangular form to polar form. In polar form, the magnitude is the hypotenuse of the two rectangular form vectors:

$$\begin{aligned} \left| \frac{e_o}{e_i} \right| &= \frac{1}{\sqrt{1^2 + (\omega/\omega_o)^2}} \\ &= \frac{1}{\sqrt{1 + (\omega/\omega_o)^2}} \end{aligned} \quad (280)$$

The tangent of the phase angle is the ratio of the two rectangular components:

$$\begin{aligned} \angle \frac{e_o}{e_i} &= \frac{1}{\tan^{-1} \left(\frac{\omega/\omega_o}{1} \right)} \\ &= -\tan^{-1}(\omega/\omega_o) \end{aligned} \quad (281)$$

Plots of the magnitude and phase are shown in figure 144.

It is customary to plot magnitude in *decibels*, that is, $20 \log_{10}(e_o/e_i)$ against the logarithm of frequency. Then the magnitude plots may be approximated by straight lines.

Phase is simply plotted versus the logarithm of frequency. Notice that the frequency axis is plotted as the ratio of frequency to the cutoff frequency in radians per second: ω/ω_o . Without change, the frequency axes could equally well be the ratio of frequency to cutoff frequency in Hertz, f/f_o .

There is much that can be said about these plots. See for example [63].

The plotting routines for the Gnuplot program, are in section 5.5.3 on page 242.

5.2.3 Measurement

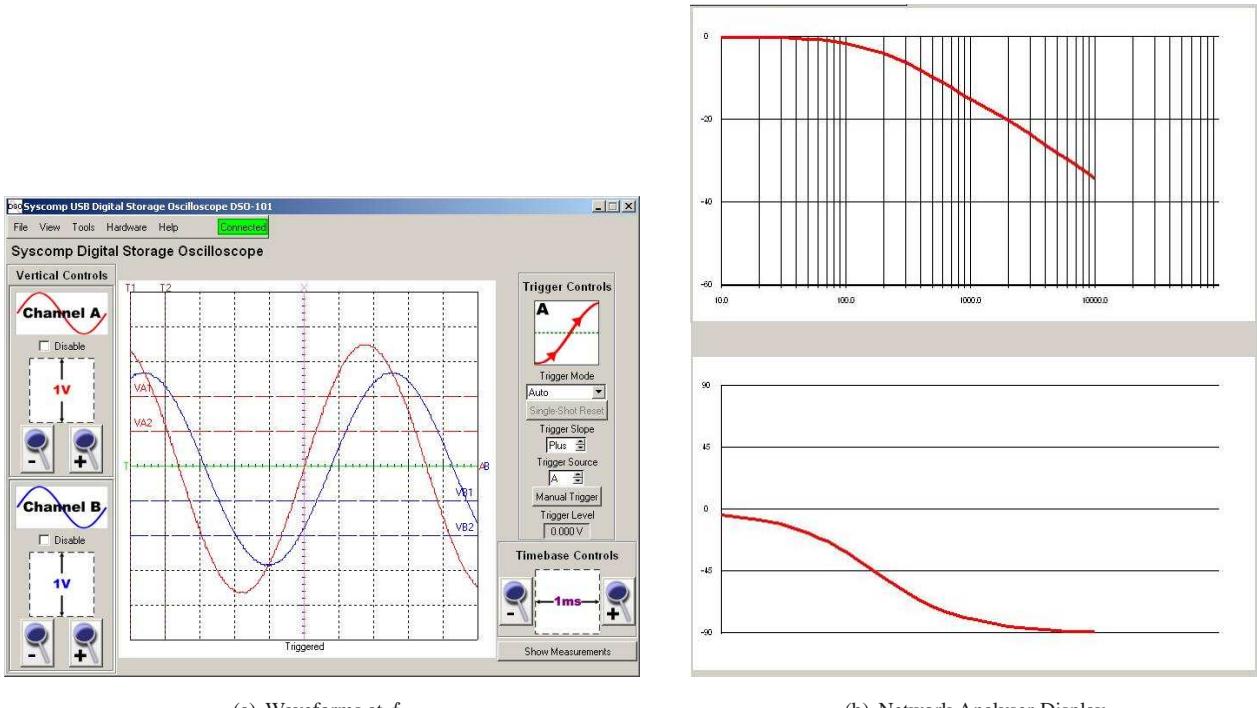


Figure 145: RC Lowpass Measurement

Figure 145 shows measurements on an RC lowpass filter with $R = 10k\Omega$ and $C = 100nF$. The cutoff frequency f_o is then 160Hz.

In figure 145(a), the generator was set to 160Hz. The output waveform (smaller of the two sine waves) lags the input by 45 degrees and is smaller by 3db.

Figure 145(b) shows the results of a swept-frequency measurement with the Vector Network Analyser (VNA)⁹⁰ software operating on a Syscomp WGM-101 waveform generator and DSO-101 oscilloscope [64].

The magnitude and phase are as predicted by the graphs of figure 144.

5.2.4 Transfer Function of Low Pass LR Filter

The LR lowpass filter is shown in figure 146. Now we'll determine the frequency response of this filter. As we'll see, this is very similar to the RC lowpass filter of the previous section.

The inductor and resistor form a voltage divider, so we can write:

$$\frac{e_o}{e_i} = \frac{Z_2}{Z_1 + Z_2} \quad (282)$$

where

e_o AC Output voltage from the circuit

e_i AC Input voltage to the circuit

Z_1 Impedance of the upper half of the voltage divider (the inductor)

Z_2 Impedance of the bottom half of the voltage divider (the resistor)

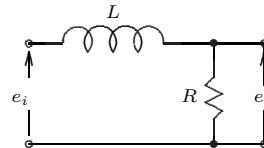


Figure 146: LR Lowpass Filter

Now substitute sL for Z_1 , R for Z_2 and do some algebra:

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{R}{R + sL} \\ &= \frac{1}{1 + s\frac{L}{R}} \end{aligned} \quad (283)$$

This time, we'll see that the quantity L/R is the time-constant of the network and the inverse of the lowpass cutoff frequency. Consequently, $\tau = L/R$ for the LR lowpass.

Then we could rewrite equation 283 as:

$$\frac{e_o}{e_i} = \frac{1}{1 + s\tau} \quad (284)$$

A Wonderful Thing happens at this point: this equation 284 for the LR lowpass filter has exactly the same form as equation 274 on page 200, which we previously found for the RC lowpass filter. Consequently, for equal values of time constant *they will have exactly the same magnitude and phase response*⁹¹.

Consequently, we can simply recycle the remainder of the information for the RC lowpass filter, recognizing that the time constant is $\tau = RC$ for the RC lowpass and $\tau = L/R$ for the LR lowpass.

5.2.5 Measurement

For these measurements, the inductor is 30mH and the resistor 2kΩ, for a cutoff frequency f_o of 10kHz. Measurements are shown in figure 147 on page 204. Based on the results of the previous section, the measurements

⁹⁰A Vector Network Analyser is sometimes referred to as a *Bode Plotter*.

⁹¹This assumes ideal behaviour of the components. In practice, the components have various forms of non-ideal behaviour, and that can be important.

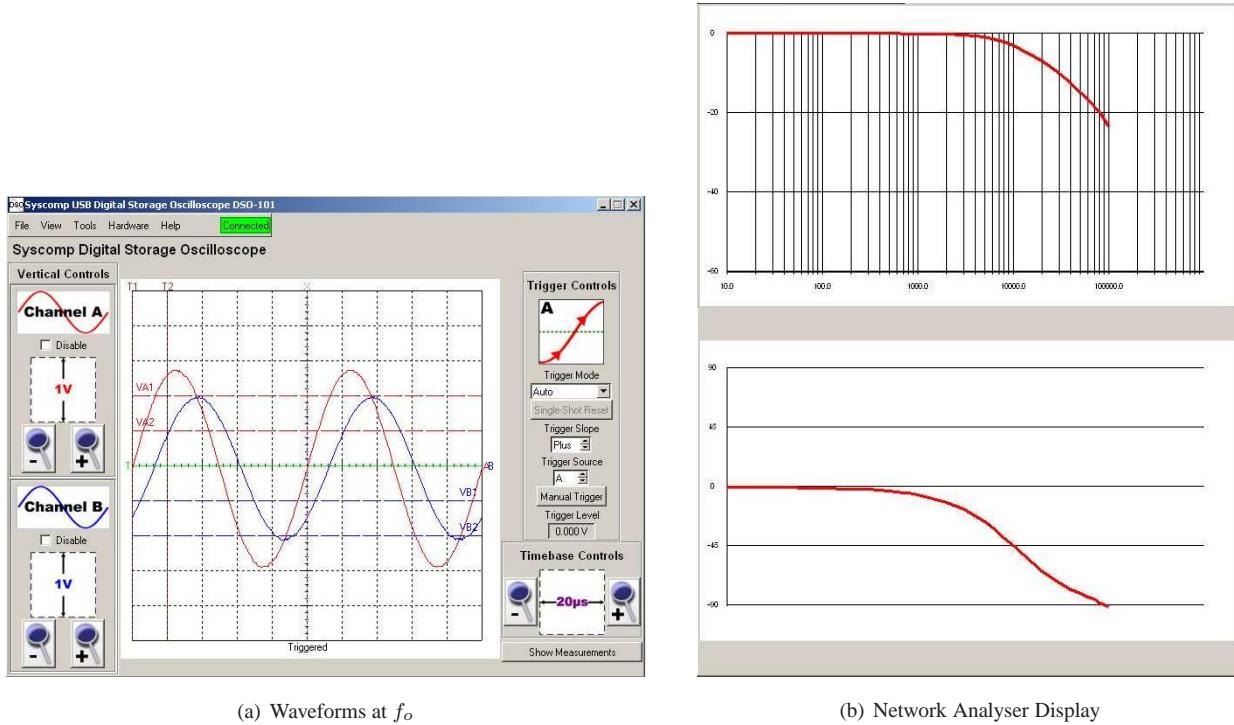


Figure 147: LR Lowpass Measurement

should be similar to those of the RC lowpass filter, adjusted for the different cutoff frequency.

The sine wave display of figure 147(a), taken at 10kHz, is essentially as expected. The output waveform lags the input by 45 degrees.

The VNA plot also shows the expected behaviour at low and mid frequencies. However, a careful examination at high frequencies shows something unexpected: the rolloff is slightly more than the expected 20db/decade. As well, the phase is headed for a larger angle than 90° . Inductance is often accompanied by significant series resistance and parallel stray capacitance (so-called *parasitic* components) and these are affecting the measurement.

5.2.6 Transfer Function of a Second Order RLC Lowpass Filter

Like the RC filter of figure 143, figure 148 is a lowpass filter: it passes low frequencies without attenuation and it attenuates high frequencies. However, the attenuation above the cutoff frequency occurs at *twice* the rate of the RC lowpass filter and there is less attenuation at the cutoff frequency.

Furthermore, the RC lowpass filter is a *first order* filter, because the largest power of s is unity. The RLC lowpass filter is second order, since it contains an s^2 term.

As we did in the previous section, we'll develop the transfer function and then massage it into a form suitable for plotting the frequency and phase response. Again, this illustrates the Laplace

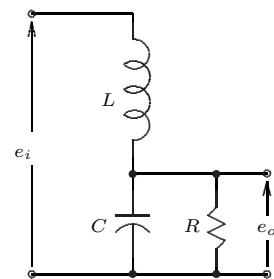


Figure 148: Second-Order Lowpass Filter

transform operator s as shorthand notation for $j\omega$.

Analysis

Treating the network of figure 148 as a voltage divider, the gain of the network is

$$\frac{e_o}{e_i} = \frac{Z_2}{Z_1 + Z_2}$$

where Z_1 and Z_2 are the upper and lower halves of the voltage divider.

$$\begin{aligned} Z_1 &= sL \\ Z_2 &= \frac{1}{sC} \parallel R \\ &= \frac{R}{1 + sRC} \end{aligned}$$

Then the transfer function is

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{\frac{R}{1 + sRC}}{sL + \frac{R}{1 + sRC}} \\ &= \frac{R}{sL + s^2RLC + R} \end{aligned}$$

Now we will manipulate this equation into a more useful form, in which the the s^2 term in the denominator has a coefficient of 1. Divide the denominator by the factor R in the numerator and then factor LC out of the denominator:

$$\frac{e_o}{e_i} = \frac{1}{LC \left(s^2 + \frac{1}{RC}s + \frac{1}{LC} \right)} \quad (285)$$

Now we can substitute for some of these quantities. The resonant frequency for an RLC circuit is

$$\omega_o = \frac{1}{\sqrt{LC}} \quad (286)$$

so we can substitute ω_o^2 for $1/LC$ in equation 285.

Recall the *Q factor*, which is the ratio of the resistance in the circuit to the reactance at resonance. At resonance, the inductive and capacitive reactance are equal, so we could choose either one. We'll choose the inductive reactance:

$$Q = \frac{R}{\omega_o L} \quad (287)$$

Since at resonance the inductive and capacitive reactances are equal:

$$\omega_o L = \frac{1}{\omega_o C} \quad (288)$$

Substituting for $\omega_o L$ in equation 287,

$$Q = \omega_o R C \quad (289)$$

and

$$\frac{1}{RC} = \frac{\omega_o}{Q} \quad (290)$$

so we can substitute ω_o/Q for $1/RC$ in equation 285. Then

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (291)$$

This is the *standard form* for the 2nd order lowpass filter.

5.2.7 Amplitude and Phase of the Lowpass RLC Filter

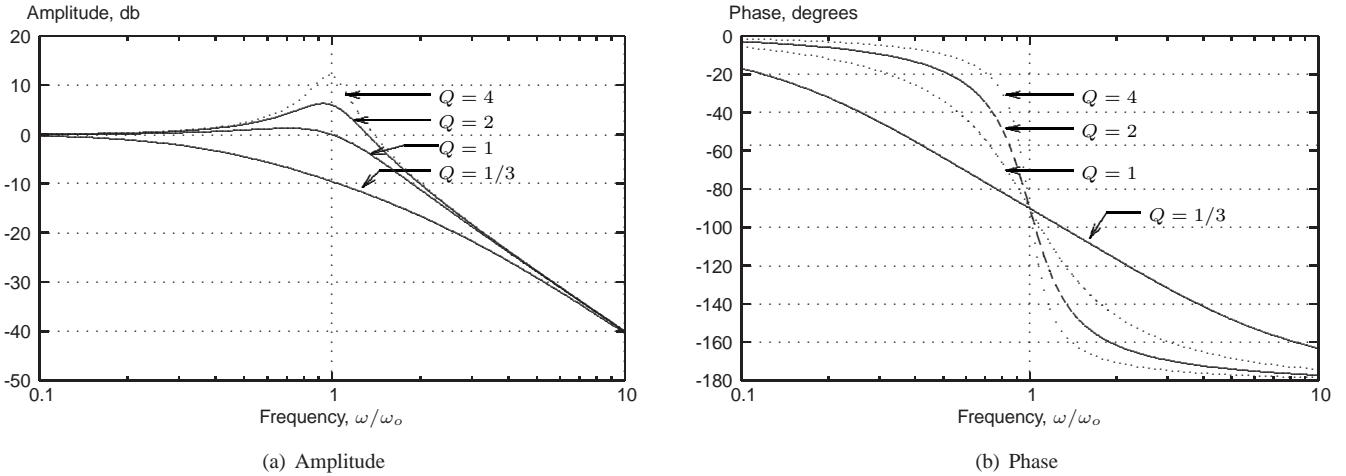


Figure 149: Amplitude and Phase of RLC Lowpass Network

In equation 291, move the ω_o^2 term in the numerator into the denominator.

$$\frac{e_o}{e_i} = \frac{1}{\left(\frac{s^2}{\omega_o^2} + \frac{s}{\omega_o Q} + 1\right)} \quad (292)$$

Now replace each occurrence of s by $j\omega$. Since $j = \sqrt{-1}$, $j^2 = -1$.

$$\frac{e_o}{e_i} = \frac{1}{\left(\frac{-\omega^2}{\omega_o^2} + \frac{j\omega}{\omega_o Q} + 1\right)} \quad (293)$$

For notational convenience replace ω/ω_o with the x axis variable, which I'll call x . Then x represents the ratio of frequency to cutoff frequency. We could plot the function over the range $x = 0.1$ to $x = 10$ to get an idea of the response in the two decades around the cutoff frequency.

Then we have

$$\frac{e_o}{e_i} = \frac{1}{-x^2 + \frac{j}{Q}x + 1}$$

Collecting real and imaginary terms, we have

$$\frac{e_o}{e_i} = \frac{1}{(1 - x^2) + j\frac{x}{Q}} \quad (294)$$

Magnitude

The magnitude is equal to the square root of the sum of the real and imaginary components each squared. As in the case of the RC filter, we'd like the result in decibels, so we take $20 \log_{10}$ of the result:

$$\left| \frac{e_o}{e_i} \right| = 20 \log_{10} \sqrt{\frac{1}{(1 - x^2)^2 + \left(\frac{x}{Q}\right)^2}} \quad (295)$$

Now, if you consider that

$$\begin{aligned} \log_{10} \sqrt{\frac{1}{A}} &= \log_{10} \left(A^{-\frac{1}{2}} \right) \\ &= -\frac{1}{2} \log_{10}(A) \end{aligned} \quad (296)$$

then we can write equation 295 as:

$$\left| \frac{e_o}{e_i} \right| = -10 \log_{10} \left[(1 - x^2)^2 + \left(\frac{x}{Q} \right)^2 \right] \quad (297)$$

Equation 297 is in a form that can be plotted, as shown in figure 149(a).

Phase

If a complex number a is written in rectangular format as for example $a = x + jy$, the phase angle is given by

$$\angle a = \tan^{-1} \left(\frac{y}{x} \right) \quad (298)$$

where x is the so-called *real* part and y is the *imaginary* part. A complex number may also be written in polar format as for example $a = R\angle\alpha$ where R is the *magnitude* and α is the phase angle.

Then it can be shown (see section 4.4 on page 158) that

$$\angle \frac{1}{x + jy} = -\tan^{-1} \left(\frac{y}{x} \right) \quad (299)$$

Applying this concept to equation 294, we have that

$$\angle \frac{e_o}{e_i} = -\tan^{-1} \left(\frac{x/Q}{1-x^2} \right) \quad (300)$$

Equation 300 is plotted in figure 149(b).

Some points of interest:

- The phase changes from 0° to -180° , passing through -90° at the resonant frequency ω_o
- The rate of change of phase in the vicinity of ω_o increases with increasing Q factor (which corresponds to decreasing values of damping δ).

5.2.8 Measurement

The experimental lowpass filter was constructed according to figure 148 on page 204 with $L = 30\text{mH}$, $C = 1\mu\text{F}$, $R = 620\Omega$ ⁹².

Then the cutoff frequency f_o for the filter (equation 286) is:

$$\omega_o = \frac{1}{\sqrt{LC}}$$

Since $\omega_o = 2\pi f_o$, then

$$f_o = \frac{1}{2\pi\sqrt{LC}} = 919 \text{ Hz}$$

The Q factor (equation 287) is:

$$Q = \frac{R}{\omega_o L} = \frac{620}{2\pi \times 919 \times (30 \times 10^{-3})} = 3.58$$

We expect to see a lowpass filter that rolls off above 919Hz with a significant peak in the region of the cutoff frequency.

The measurement results are shown in figure 150. Figure 150(a) shows the waveforms below cutoff frequency. As expected, they are in phase and nearly equal in magnitude.

Figure 150(b) shows the waveforms at f_o , the cutoff frequency. The waveforms are 90° out of phase and the output is larger than the input, due to peaking at this frequency.

Figure 150(c) shows the waveforms above cutoff. The output is now 180° out of phase with the input and much reduced in magnitude.

Figure 150(d) shows the network analyser measurement. As expected, there is peaking around f_o and the output drops at 40db/decade above the cutoff frequency. The phase display can only accommodate a range of $+90^\circ$ to -90° , so we see an abrupt change in phase at the resonant frequency. At high frequencies, the phase shift is 180° as expected.

⁹²The waveform generator has an internal resistance of 75Ω that should be taken into account. However, it turns out this resistance is small compared to the circuit impedances and does not have a major effect on the circuit operation so for our purposes can be ignored. If there is any doubt in the matter, its effect must be investigated.

As well, the inductor has some DC resistance and the capacitor some ESR (equivalent series resistance) that should be taken into account in a completely precise analysis. These also have been ignored in this measurement.

Including these components in the analysis complicates it to the point that a circuit simulation is probably the most efficient approach.

In a teaching environment such as an EE lab, it's up to the designer of the measurement exercise to ensure that these so-called parasitic components do not obscure the behaviour that is being illustrated.

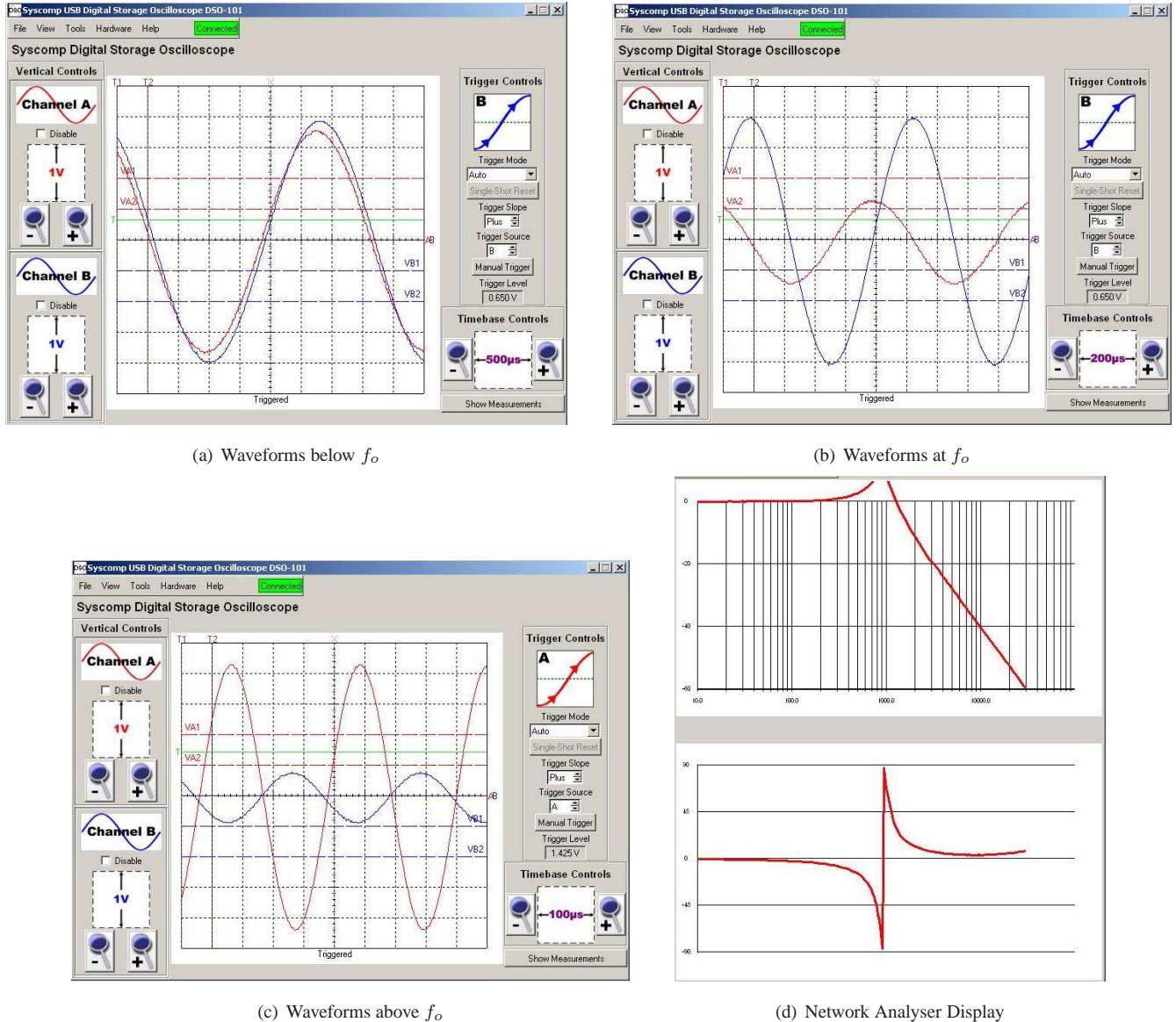


Figure 150: RLC Lowpass Measurement

5.2.9 Transfer Function of the Two Stage (Second Order) RC Lowpass Filter

The two-stage RC lowpass filter is shown in figure 151(a) and a rearrangement for analysis in figure 151(b).

We can treat this network as a two-stage voltage divider. Working backward from output to input:

- The output voltage e_o is divided down from the intermediate voltage e_a by resistor R_2 and capacitor C_2 .
- The intermediate voltage e_a is divided down from e_i by a voltage divider consisting of resistor R_1 and the network Z . The network Z is composed of capacitor C_1 in parallel with the series combination of R_2 and C_2 .

After much algebraic crank-turning, we obtain:

$$\frac{e_o}{e_i} = \left[\frac{1}{s^2 + s \frac{(R_1 C_2 + R_1 C_1 + R_2 C_2)}{R_1 R_2 C_1 C_2} + \frac{1}{R_1 R_2 C_1 C_2}} \right] \times \frac{1}{R_1 R_2 C_1 C_2} \quad (301)$$

On most occasions, the resistors and capacitors are equal, so let us make:

$$R_1 = R_2 = R, \quad C_1 = C_2 = C$$

Many things cancel at this point, and after some cosmetic surgery we have as a result:

$$\frac{e_o}{e_i} = \frac{\frac{1}{R^2 C^2}}{s^2 + \frac{3}{RC}s + \frac{1}{R^2 C^2}} \quad (302)$$

This is similar to something we've seen before, the *standard form* of the second-order lowpass transfer function, equation 816 on page 574. Here is the standard form again:

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \quad (303)$$

Equation 302 is equivalent to the standard form of equation 303 if we make:

$$\omega_o = \frac{1}{RC}, \quad Q = \frac{1}{3}$$

Then the magnitude and phase are as shown on the plots for the standard form, figure 149 on page 206. In the case of the two stage RC lowpass, the value of Q is $1/3$ which is one of the traces in figure 149.

Now let's think about what this means. Ideally, we'd like our filter to have no attenuation in the passband (below the cutoff frequency) and much attenuation in the stop band (above the cutoff frequency).

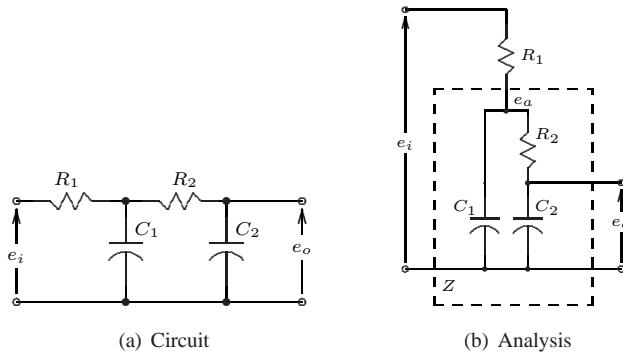


Figure 151: 2nd Order RC Lowpass Filter

According to figure 149 that ideal is achieved most closely for the second-order lowpass filter when the Q value is about 1. The response drops at 40db/decade above the corner frequency and the attenuation at the cutoff frequency is approximately zero.

In contrast, the dual-section RC filter makes the transition very gradually between passband and stop band. It attenuates the signal by 10 db at the cutoff frequency. It is only at ten times the cutoff frequency that the attenuation rate approaches 40db/decade. So the dual section RC filter is not a great performer in the frequency domain.

5.3 Time Domain Response

In section 5.2 we determined the frequency response behaviour of the four lowpass filters in figure 141 (page 196), in order to determine which would make the most effective filter for a noise signal. Conceptually, we applied to the input of the filter a sine wave of various arbitrary frequencies and then measured the relationship between the input and output sine waves at these frequencies.

The Laplace transform substitutions we used (such as $1/sC$ for a capacitance) are useful in determining the frequency response of networks. That work requires a fair amount of algebra, but nothing new relating to the Laplace transform.

Now we will consider what happens when we apply various signals that are arbitrary in the time domain: that is, they have arbitrary shapes. The Laplace transform is particularly useful for this type of analysis, when it is necessary to determine the time-response of a particular electrical circuit.

1. Choose an input signal: step, ramp, sine, whatever.
2. Determine the Laplace transform of the input signal. Some input signals are simple enough that the transform is known. For example, the Laplace transform of the *unit impulse* is simply 1. The Laplace transform of the *unit step* is $1/s$. In other cases, a more complex signal is specified in the problem and some work is required to determine the transform.
3. Determine the Laplace transform of the network transfer function as we showed in section 5.2.1 and section 5.2.7. This can usually be written out by inspection of the circuit.
4. Multiply the input signal of step 2 by the transfer function of step 3 to obtain the Laplace transform of the output signal.
5. Take the inverse Laplace transform of the output signal to obtain the output voltage as a function of time.

Next, we introduce two useful tools for this type of analysis: a *computer algebra* program (Maxima) (section 5.3.1) and a table of Laplace transforms (section 5.3.2).

In sections 5.3.4, 5.3.5 and 5.3.6 we use these tools to develop the Laplace transforms of the impulse, step and ramp input signals, or *forcing functions* as they are known.

Then we apply each of these signals to the RC lowpass network and determine the resultant output voltage-time signal.

5.3.1 Computer Algebra: Using Maxima

A *computer algebra system* (CAS) [65] manipulates the symbols of math equations according to the rules of algebra, calculus and other branches of mathematics. A computer algebra system can remove much of the labour from Laplace and Inverse Laplace transforms. In this section we use the open-source CAS *Maxima* program (figure 152).⁹³

A list of alternative CAS programs is in [66].

Apart from a few introductory examples given below, we show only commands that are relevant to the business at hand. A complete manual for Maxima (860 pages!) is available at the Maxima home page [67].

Here are some examples of Maxima at work:

Big Numbers

The (%i28) and (%o28) are Maxima input and output prompts.

```
(%i28) 12^26;
(%o28) 11447545997288281555215581184
```

Solving an Equation

First, we define the equation

```
(%i24) a*x^2+b*x+c;
(%o24) a x2 + b x + c
```

Then we ask Maxima to solve it. The % symbol means ‘the previous equation’.

```
(%i25) solve (% ,x);
(%o25) [x = -  $\frac{\sqrt{b^2-4 a c}+b}{2 a}$ , x =  $\frac{\sqrt{b^2-4 a c}-b}{2 a}]$ 
```

Differentiating

```
(%i30) diff (cos(x),x);
(%o30) - sin(x)
```

```
(%i32) diff ((sin(x))^3,x);
(%o32) 3 cos(x) sin(x)2
```

Integrating

```
(%i31) integrate(-sin(x),x);
(%o31) cos(x)
```

Complex Math

(%i is the complex operator, also known as j.)

```
(%i33) rectform(5.0 / (3.0+2.0*%i) + 4.0 / (8.0*%i+5.0));
(%o33) 1.378565254969749 - 1.128781331028522 i
```

⁹³Maxima is roughly comparable to the proprietary programs *Mathematica* and *Maple*. Notice that *Matlab* and open-source equivalent *Octave* (widely used in universities for simulation of electronic systems) is (to quote Wikipedia): *a numerical computing environment and programming language*. In other words, Maxima, Mathematica and Maple work with symbols. Matlab and Octave work with numbers. For a general-purpose solution, not tied to specific values in the circuit, you need a program that works with symbols.

The screenshot shows the wxMaxima 0.7.4 interface. The window title is "wxMaxima 0.7.4 [unsaved*]". The menu bar includes File, Edit, Maxima, Equations, Algebra, Calculus, Simplify, Plotting, Numeric, and Help. Below the menu is a toolbar with various icons. The main area displays a session history:

```
(%o15) t  
  
(%i16) ilt(1/s^3,s,t);  
(%o16)  $\frac{t^2}{2}$   
  
(%i17) a/(s^2 * (a+s));  
(%o17)  $\frac{a}{s^2(s+a)}$   
  
(%i18) ilt(%,s,t);  
(%o18)  $\frac{a e^{-at}}{a} + t - \frac{1}{a}$   
  
(%i19) partfrac(a/(s^2 * (a+s)),s);  
(%o19)  $\frac{1}{a(s+a)} - \frac{1}{as} + \frac{1}{s^2}$   
  
(%i20)
```

At the bottom, there is an "INPUT:" field and a toolbar with buttons for Simplify, Simplify (r), Factor, Expand, Solve..., Plot 2D..., Simplify (tr), Expand (tr), Reduce (tr), Rectform, Solve ODE..., and Plot 3D... A status bar at the bottom right says "Ready for user input".

Figure 152: Maxima Screen Shot

5.3.2 Table of Transforms

	Description	Time Domain Function	Laplace Transform
1	Impulse	$\delta(t)$	1
2	Unit Step	$u(t), f(t) = 1$	$\frac{1}{s}$
3	Unit Ramp	$r(t), f(t) = t$	$\frac{1}{s^2}$
4	Integration	$\int f(t)$	$\frac{1}{s} F(s)$
5	Differentiation	$\frac{d}{dt} f(t)$	$sF(s)$
6	Linearity	$kf(t)$	$kF(s)$
7	Exponential	$\exp^{\alpha t}$	$\frac{1}{s - \alpha}$
8	Superposition	$Af(t) + Bg(t)$	$AF(s) + BG(s)$
9	Sine Wave Oscillator	$\frac{1}{s^2 + a^2}$	$\frac{\sin at}{a}$
10	Sine Wave Oscillator	$\frac{s}{s^2 + a^2}$	$\cos at$
11	Decaying Sine Wave	$\frac{1}{(s - b)^2 + a^2}$	$\frac{e^{bt} \sin at}{a}$

5.3.3 Table of Transforms, Quadratic

	Description	Time Domain Function	Laplace Transform
1	Underdamped, $\delta < 1$	$\frac{1}{\omega_d} e^{-\delta \omega_n t} \sin \omega_d t$	$\frac{1}{s^2 + 2\delta\omega_n s + \omega_n^2}$
2	Critically damped, $\delta = 1$	$t e^{-\omega_n t}$	$\frac{1}{s^2 + 2\delta\omega_n s + \omega_n^2}$
3	Overdamped, $\delta > 1$	$\frac{1}{\omega_d} e^{-\delta \omega_n t} \sinh \omega_d t$	$\frac{1}{s^2 + 2\delta\omega_n s + \omega_n^2}$
4	Damped natural frequency	$\omega_d = \omega_n \sqrt{1 - \delta^2}$	
5	Damping Ratio	$\delta = \frac{1}{2Q}$	

The table of section 5.3.2 lists common transform pairs, which we will use in subsequent sections. The table of section 5.3.3 shows transforms for a quadratic⁹⁴. This is relevant to second-order systems such as a resonant RLC circuit, and active filters. The damping factor δ is equivalent to the Q factor (section 4.13) by $\text{delta} = 1/2Q$.

⁹⁴As adapted from Table 1 of [68].

Now we will discuss some of the transforms from the first table, section 5.3.2.

5.3.4 Forcing Function: Unit Impulse

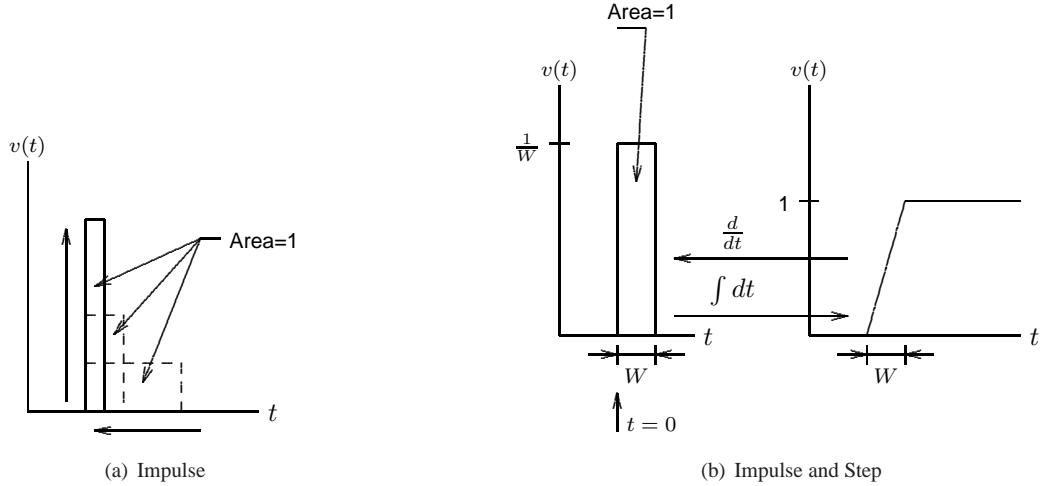


Figure 153: Impulse Signal

The simplest forcing function is the *unit impulse* $\delta(t)$. The theoretical abstraction of the unit impulse is a pulse with area unity, zero width and infinite amplitude.

One way to think of the unit impulse is shown in figure 153(a). A unit impulse is the limiting case of a pulse that is reduced in duration while keeping the area unchanged.

Another view is shown in figure 153(b). The unit impulse is the slope (differential d/dt) of the rising portion of a *unit step* waveform. As the slope increases, the width of the unit impulse decreases while its amplitude increases to maintain unity area. In the limit, as the rising portion of the step approaches vertical, the unit impulse approaches the theoretical abstraction of zero width, infinite height, unity area.

When is the unit impulse useful? The Laplace Transform of the unit impulse is simply 1. As a result, if you apply a unit impulse to a system, then the output is the system transfer function. This makes it very simple to determine the transfer function: apply an impulse signal to the input and record the output. That is the transfer function in the time domain.

For example, to measure the transfer function of a mechanical system apply an impulse by hitting it with a hammer. Record the output. Now, this impulse is probably not ideal: it has finite duration and finite amplitude. However, if the pulse is short compared to any of the time constants in the system, then it will appear as an impulse.

With the time-domain response to an impulse input you can have enough information to determine the response of a system to *any* input signal,

1. Apply an impulse to the system and record the corresponding output. This is the system impulse response in the time domain.
2. Use the Laplace transform to convert this time-domain system impulse response to the Laplace domain.
3. Choose an input signal and obtain its Laplace transform. Multiply the Laplace domain impulse response by the Laplace domain input signal. This yields the Laplace domain output signal.
4. Take the inverse Laplace transform of this output signal to determine the output signal in the time domain.

The impulse technique cannot be used in some practical applications. Perhaps the system cannot be disturbed with an impulse. Or an impulse drives the system into non-linear behaviour. Or it may be that there is a signal-noise problem: the output signal is too weak at the maximum allowable impulse signal. In those cases, *cross-correlation* may be more useful. Low-level random noise is fed into the system. The output signal is cross-correlated against the input. The resulting correlation function is the impulse response of the system.

The amplitude of an impulse may be scaled by some factor K . Then the Laplace transform of the scaled impulse is $K \times 1 = K$.

It is a bit tricky to prove that the Laplace transform of the unit impulse is 1. Maxima is not much help. A proof is shown in section 5.5.1 on page 240.

5.3.5 Forcing Function: Unit Step

The *unit step* is shown in figure 154. A step waveform of height 1 unit is the time integral of a unit impulse. In the Laplace domain, integration is accomplished by multiplying by $1/s$. Consequently, the Laplace transform of the unit step is $1/s$ times unity, or $1/s$. We can obtain the same result using the definition of the Laplace transform, as shown in section 5.5.2 on page 242.

Or we can use Maxima to do this. From the Maxima manual:

Figure 154: Unit Step

`laplace(expr,t,s)` attempts to compute the Laplace transform of *expr* with respect to the variable *t* and transform variable *s*.

In this case, a unit step, *expr* is simply 1.

```
(C22) laplace (1, t, s);
(D22)      1
           -
           s
```

This also works in reverse. Again, from the manual:

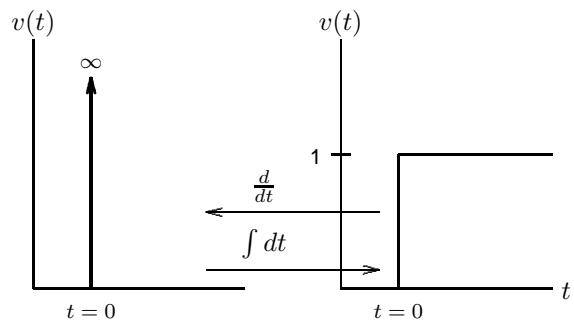
`ilt(expr,s,t)` computes the inverse Laplace transform of *expr* with respect to the *t* and variable *s*. *expr* must be a ratio of polynomials whose denominator has only linear and quadratic factors.

This time, *expr* is $1/s$ and the inverse transform yields a unit step waveform.

```
(C21) ilt (1/s, s, t);
(D21)      1
```

In practice, waveforms are rarely one unit in amplitude, in which case the amplitude of the step may be scaled by some factor K .

5.3.6 Forcing Function: Unit Ramp



5.3 Time Domain Response

The *unit ramp* is shown in figure 155, a ramp waveform of 1 unit increase per unit of time. As the step waveform is an integral of the impulse, so is the ramp an integral of the step. Again, in the Laplace domain integration is accomplished by multiplying by $1/s$. Consequently, the Laplace transform of the unit ramp is $1/s^2$.

This is confirmed by Maxima, for the transform and inverse transform:

```
(C15) laplace(t,t,s);
```

```
(D15)
```

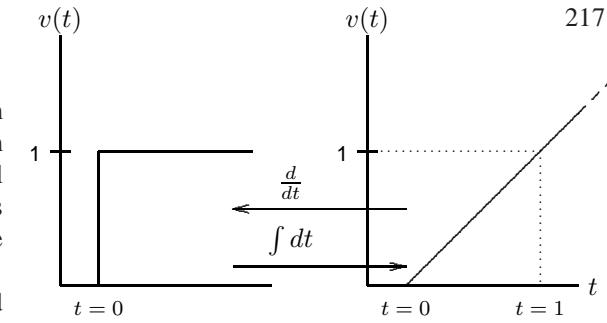


Figure 155: Unit Ramp

```
(C14) ilt (1/s^2, s, t);
```

```
(D14)
```

$$\begin{matrix} 1 \\ -- \\ 2 \\ s \end{matrix}$$
 t

As in the case of the impulse and step waveforms, the magnitude of the ramp may be scaled by some factor K .

Unit Impulse and RC Lowpass

Now we are in a position to analyse some simple circuits. Our first case is to determine the time domain output signal when the RC lowpass filter is driven by a unit impulse.

The Laplace transform of the transfer function of the RC lowpass filter was derived previously as equation 274 on page 200:

$$\frac{e_o}{e_i} = \frac{1}{1 + s\tau} \quad (304)$$

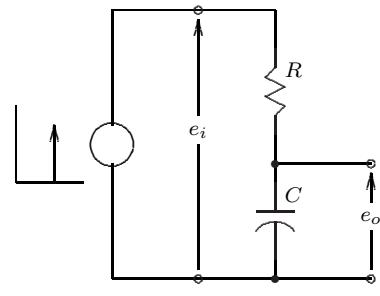
The Laplace transform of the output signal is the product of the input signal (unity) and the transfer function, equation 304. To find the output signal as a function of time, we need the inverse transform for equation 304. There are two ways to do this: with the computer algebra system (Maxima) and manually, with a table of transforms.

Inverse Transform, Using Maxima

Maxima can find the inverse transform:

```
(C18) ilt (a/(s+a), s, t);
```

```
(D18)
```



$$\tau = RC, \omega_c = 1/RC$$

Figure 156: Impulse with RC Lowpass

$$\begin{matrix} - a t \\ a \%E \end{matrix}$$

where $\%E$ is the constant e . Back substitute $a = 1/\tau$ and we obtain:

$$\frac{e_o(t)}{e_i(t)} = \frac{1}{\tau} e^{-\frac{t}{\tau}} \quad (305)$$

Inverse Transform, Using Tables

Consulting the table of Laplace transforms (section 5.3.2 on page 214), entry 7 looks as if it might be useful.

$$\frac{1}{s-a} \leftrightarrow e^{at} \quad (306)$$

We'll manipulate equation 304 into that form.

$$\begin{aligned} \frac{e_o(s)}{e_i(s)} &= \frac{1}{1+s\tau} \\ &= \frac{1}{\tau} \left(\frac{1}{s + \frac{1}{\tau}} \right) \end{aligned} \quad (307)$$

Then $-1/\tau$ in equation 307 is equivalent to a in the transform pair of equation 306. We also have a leading constant $1/\tau$ in equation 307 to take into account. According to entry 6 of the table, this remains unchanged from the Laplace to the time domain. Then the inverse transform of equation 307 is:

$$\frac{e_o(t)}{e_i(t)} = \frac{1}{\tau} e^{-\frac{t}{\tau}} \quad (308)$$

which matches equation 305.

Measurements

Figure 157(a) shows the theoretical result of an impulse measurement on an RC lowpass network. The initial magnitude of the output waveform is equal to K/τ , where K is the area of the impulse in volt-seconds and τ is RC , the time constant of the network. After the initial output, the output decays to $1/e$ of its original value in one time constant.

Figure 157(b) shows an impulse measurement on an RC lowpass network with $R = 9880\Omega$ ($10k\Omega$ nominal), $C=95.6nF$ ($100nF$ nominal). Then the time constant for this network is $\tau = RC = 9880 \times (95.6 \times 10^{-9}) = 0.944$ milliseconds.

We must check the source and load impedance of the measuring equipment to determine that it does not affect the measurement. The output impedance of the generator is 75Ω , much less than the resistance, so it may be neglected. The input impedance is the oscilloscope is $1M\Omega \parallel 20pF$, again large enough to be neglected.

In order that the input pulse be considered an impulse, the pulse width must be much less than the time constant of the network and there must be sufficient time for the output transient to die away. In this case, the generator is set to a duty cycle of 1% which yields a pulse of width $50\mu\text{Sec}$, one tenth of the time constant. The repetition rate is 200Hz so the time between pulses is 5mSec, 5 time constants. The pulse is unipolar so the generator output is offset by +4 volts with a pulse amplitude of 8 volts peak.

Figure 157(d) shows the control setup for the generator.

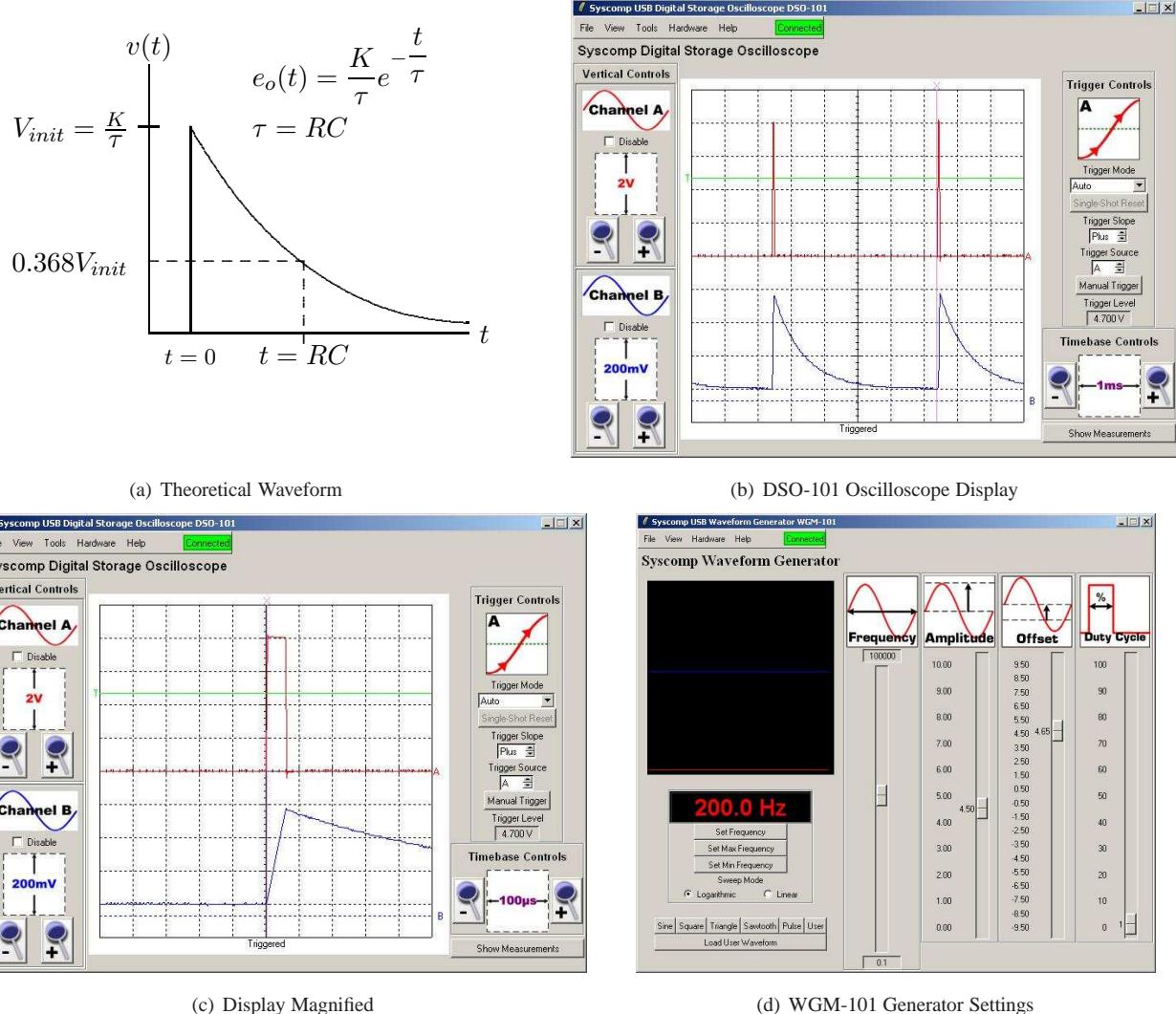


Figure 157: Impulse and RC Lowpass

Results

Figure 157(b) shows the measured input impulse (upper trace) and output waveform (lower trace).

The measurement cursors and adjustable trigger point of the DSO-101 oscilloscope make it very convenient to measure the voltage and time values of the output waveform.

The input area of the pulse is 8 volts times the pulse width of $50\mu\text{Sec}$ for a total of 400×10^{-6} volt-seconds. This is the value of K in the equation of figure 157(a). Then the peak value of the output pulse is theoretically $V_{init} = K/\tau = 400\text{mV}$. The measured value is 465mV.

The output waveform should decay to $1/e = 0.368$ of its original value in one time constant, 0.944mSec in this case. In fact, a decay from 465mV to 173mV occurs in $870\mu\text{Sec}$.

Figure 157(c) shows an expanded view of the pulse interval. Notice how the capacitor voltage ramps up

during the pulse. During the charging interval the waveform is exponential but because the capacitor voltage is always much less than the pulse voltage, the voltage across the resistor is approximately constant and the capacitor charging current is approximately constant. As a result, the capacitor charging waveform is approximately linear. This is a useful approximation to keep in mind for some circuit analysis problems.

5.3.7 Unit Step and RC Lowpass

Now we'll determine the *unit step response* of the RC lowpass filter, figure 158.

Again, we start with the Laplace transform of the transfer function of the RC lowpass filter, equation 274:

$$\frac{e_o}{e_i} = \frac{1}{1 + s\tau} \quad (309)$$

Again, it looks as if the transform

$$\frac{1}{s - a} \leftrightarrow e^{at} \quad (310)$$

might be useful. Rearrange equation 309 and put $1/\tau = a$ to simplify the notation:

$$\frac{e_o}{e_i} = \frac{a}{a + s} \quad (311)$$

The output signal as a function of time is the inverse Laplace transform of the product of the unit step input signal ($1/s$) and the transfer function, equation 311.

$$\frac{e_o(t)}{e_i(t)} = \mathcal{L}^{-1} \frac{1}{s} \left(\frac{a}{a + s} \right) \quad (312)$$

Now we must find this inverse transform. We'll show two methods, a computer algebra method using Maxima and a manual method.

Method 1: Using Maxima

Equation 312 is of the form

$$\frac{a}{s} \left(\frac{1}{s + a} \right)$$

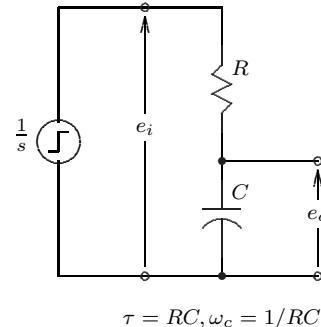
where $a = 1/\tau$. Here is Maxima finding the inverse transform:

```
(C25) ilt( (a/s)*(1/(s+a)),s,t);
(D25)                                - a t
                                         1 - %E
```

Back substitute $a = 1/\tau$ and the result is:

$$\frac{e_o(t)}{e_i(t)} = 1 - e^{-t/\tau} \quad (313)$$

This equation is the time response of an RC lowpass network to a unity-step input. If the step is K volts in magnitude, then the response is multiplied by K .



$$\tau = RC, \omega_c = 1/RC$$

Figure 158: Step with RC Lowpass

Method 2: Using Transform Table Entry

Here is the manual method. It's a lot more work than using Maxima, but nothing is hidden 'behind the curtain'.

We begin with equation 312. We need to apply *partial fraction expansion* to get equation 312 into a suitable form for the transforms in the table (section 5.3.2). Maxima can do partial fraction expansions, but we'll do it manually. In equation 315 we reorganize equation 312 into the sum of two terms. Each denominator is one that has a recognizable entry in the table of Laplace transforms. Now we need to determine the numerators x and y to satisfy this form.

It's also a good idea to move the numerator constant a out of the partial fraction expansion (as we'll see later), so we'll do that now.

$$\begin{aligned} \frac{1}{s} \left(\frac{a}{a+s} \right) &= a \left[\frac{1}{s} \left(\frac{1}{a+s} \right) \right] \\ &= a \left[\frac{x}{s} + \frac{y}{a+s} \right] \end{aligned} \tag{314}$$

$$\begin{aligned} &= a \left[\frac{xa + xs + ys}{s(a+s)} \right] \\ &= a \left[\frac{xa + s(x+y)}{s(a+s)} \right] \end{aligned} \tag{315}$$

Comparing the numerator of equation 315 with the numerator of equation 312, we can write the following:

$$\begin{aligned} xa &= 1 \\ x &= \frac{1}{a} \end{aligned} \tag{316}$$

Also, since the s term in the numerator in the numerator of equation 312 is non-existent:

$$\begin{aligned} x+y &= 0 \\ y &= -x \\ &= -\frac{1}{a} \end{aligned} \tag{317}$$

Check these values by back-substituting for x and y in equation 314:

$$\begin{aligned} \frac{1}{s} \frac{a}{a+s} &= a \left[\frac{x}{s} + \frac{y}{a+s} \right] \\ &= a \left[\frac{1}{as} + \frac{-1}{a} \frac{1}{a+s} \right] \\ &= \frac{a}{s(a+s)} \end{aligned} \tag{318}$$

This confirms that x and y are correct. Now we can proceed with the inverse transform.

$$\frac{e_o(t)}{e_i(t)} = \mathcal{L}^{-1} \left(\frac{a}{a+s} \right) \frac{1}{s}$$

$$\begin{aligned}
 &= \mathcal{L}^{-1} a \left(\frac{1}{as} - \frac{1}{a(a+s)} \right) \\
 &= \mathcal{L}^{-1} \left(\frac{1}{s} - \frac{1}{(a+s)} \right)
 \end{aligned} \tag{319}$$

Consulting the table of Laplace transforms (section 5.3.2 on page 214), entry 8 indicates that each of these terms can be treated separately. According to entry 2 of the table, the $1/s$ term transforms to 1. The -1 term in the numerator is a constant so it transforms unchanged. The $1/(a+s)$ term transforms to e^{-at} . Putting this together and back substituting $1/\tau = a$, we have:

$$\begin{aligned}
 \frac{e_o(t)}{e_i(t)} &= \mathcal{L}^{-1} \left(\frac{1}{s} - 1 \frac{1}{a+s} \right) \\
 &= +1 - 1e^{-at} \\
 &= 1 - e^{-t/\tau}
 \end{aligned} \tag{320}$$

which is the same result that Maxima found for us in equation 313.

Sanity Check

It's always a good idea to check the reasonableness of an equation against the physical behaviour of the circuit.

When the unity-value step is applied to the RC lowpass circuit, the capacitor will begin to charge and that process will continue until the capacitor voltage is equal to the input voltage. Let's examine equation 320 to see if that is true.

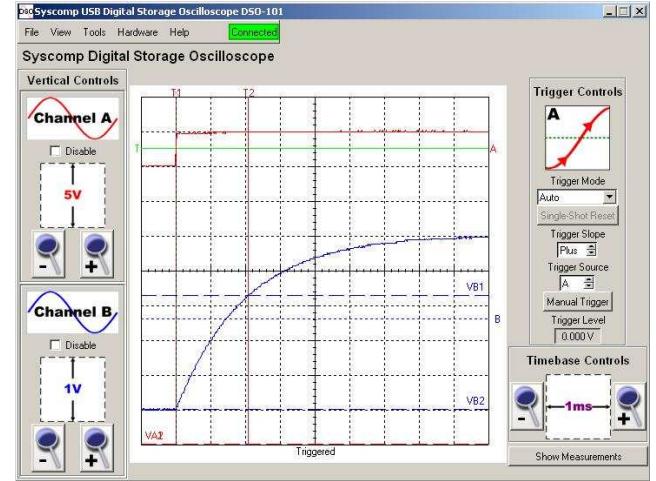
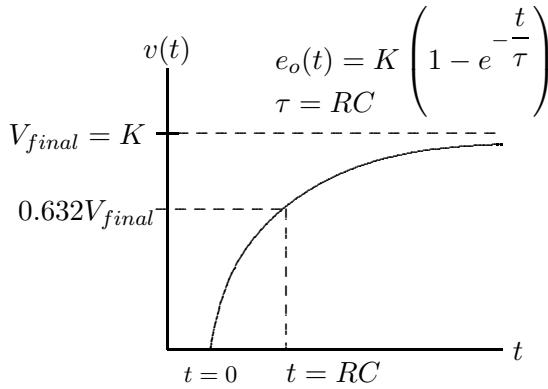
At time $t = 0$, the exponential term $e^{-t/\tau}$ is then $e^0 = 1$, so the output voltage is zero.

At time $t = [\text{Very Large Value Compared to } \tau]$, then $e^{-[\text{Large Value}]} \approx 0$ and the output voltage is equal to the input voltage.

Both of these agree with the physical situation.

We could also put $t = \tau$ in which case the output voltage is $1 - e^{-1} = 0.633$. That is, the charging is 63% complete after one time constant.

Measurements



(a) Theoretical Waveform

(b) DSO-101 Scope Display

Figure 159: Step and RC Lowpass

Figure 159(a) shows the theoretical step response of an RC lowpass network. The output voltage rises up to the input voltage at a rate defined by the time constant. Figure 159(b) shows the measured response of an RC lowpass network to a 5 volt input step, where $R = 2.075\text{k}\Omega$ and $C = 1\mu\text{F}$. The time constant is then $RC = \tau = 2.075\text{mSec}$. At 63% of the final voltage (3.15 volts) the time elapsed is about 2mSec, as predicted.

5.3.8 Unit Ramp and RC Lowpass

Now we'll do the *unit ramp response* of the RC lowpass filter, figure 160.

Again, we start with the Laplace transform of the transfer function of the RC lowpass filter, equation 274:

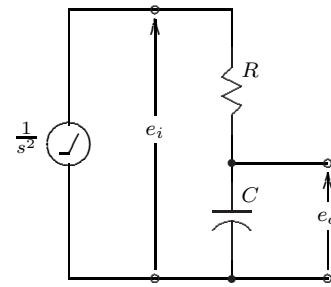
$$\frac{e_o}{e_i} = \frac{1}{1 + s\tau} \quad (321)$$

Again, it looks as if the transform

$$\frac{1}{s - a} \leftrightarrow e^{at} \quad (322)$$

might be useful. Rearrange equation 309 and put $1/\tau = a$ to simplify the notation:

$$\frac{e_o}{e_i} = \frac{a}{a + s} \quad (323)$$



$$\tau = RC, \omega_c = 1/RC$$

Figure 160: Ramp with RC Lowpass

The Laplace transform of the input ramp signal is $1/s^2$. The output signal as a function of time is the inverse Laplace transform of the product of the input signal and the transfer function, equation 324.

$$\frac{e_o(t)}{e_i(t)} = \mathcal{L}^{-1} \frac{1}{s^2} \left(\frac{a}{a+s} \right) \quad (324)$$

It simplifies things to move the a term in the numerator out in front of the expression as a constant.

$$\frac{e_o(t)}{e_i(t)} = \mathcal{L}^{-1} a \left[\frac{1}{s^2} \left(\frac{1}{a+s} \right) \right] \quad (325)$$

Inverse Transform, Using Maxima

Maxima can find the inverse transform:

$$(C26) \text{ ilt}((a/(s^2))*(1/(s+a)), s, t); \\ \frac{-a t}{a} + \frac{1}{a} \\ (D26) \quad \frac{1}{a}$$

Back substitute $a = 1/\tau$ and rearrange, then we have:

$$\frac{e_o(t)}{e_i(t)} = t - \tau \left(1 - e^{-t/\tau} \right) \quad (326)$$

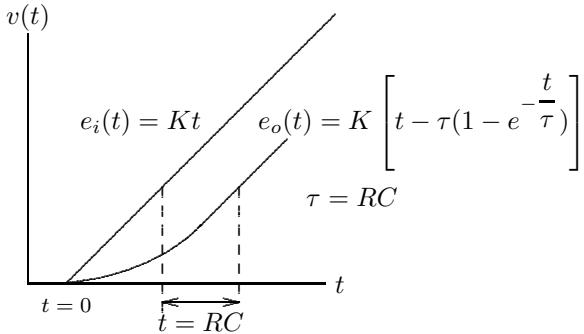


Figure 161: RC Lowpass Output vs Time with Ramp Kt Input

This equation is the time response of an RC lowpass network to a unity-ramp input. If the ramp is K volts/second in magnitude, then the response is multiplied by K (figure 161).

Let's think about this equation. It says that the ramp input generates a ramp output minus *something*. Looking at the *something* part, the exponential term disappears as time becomes large. In other words, everything after the ramp eventually becomes a constant equal to τ . So the output tracks the input with a constant difference (control systems people would call this an *error*) of $\tau = RC$.

Inverse Transform, Using Algebra and Transform Table

We can obtain the same result by hand. Start with equation 325 above.

If we extract a $1/s$ term from the right side, then what remains inside the square brackets is something we've already done:

$$\frac{e_o(t)}{e_i(t)} = \mathcal{L}^{-1} \frac{a}{s} \left[\frac{1}{s} \left(\frac{1}{a+s} \right) \right] \quad (327)$$

We previously showed (equation 319) that the expression inside the square brackets can be expanded in fractions as follows:

$$\frac{1}{s} \left(\frac{1}{a+s} \right) = \frac{1}{as} - \frac{1}{a(a+s)} \quad (328)$$

Back substitute from equation 328 into equation 327 and we have:

$$\begin{aligned}\frac{e_o(t)}{e_i(t)} &= \mathcal{L}^{-1} \frac{a}{s} \left[\frac{1}{as} - \frac{1}{a(a+s)} \right] \\ &= \mathcal{L}^{-1} \frac{1}{s} \left[\frac{1}{s} - \frac{1}{(a+s)} \right] \\ &= \mathcal{L}^{-1} \frac{1}{s^2} - \frac{1}{s} \frac{1}{(a+s)}\end{aligned}\quad (329)$$

We can do the same substitution on the second term in the equation, again from equation 319:

$$\begin{aligned}\frac{e_o(t)}{e_i(t)} &= \mathcal{L}^{-1} \frac{1}{s^2} - \left[\frac{1}{s} \frac{1}{(a+s)} \right] \\ &= \mathcal{L}^{-1} \frac{1}{s^2} - \left[\frac{1}{as} - \frac{1}{a(a+s)} \right] \\ &= \mathcal{L}^{-1} \frac{1}{s^2} - \frac{1}{as} + \frac{1}{a(a+s)}\end{aligned}\quad (330)$$

Now we can proceed with the inverse transform.

Consulting the table of Laplace transforms (section 5.3.2 on page 214), entry 8 indicates that each of these terms can be treated separately.

$$\mathcal{L}^{-1} \frac{1}{s^2} = t \quad (331)$$

This is the ramp signal.

$$\mathcal{L}^{-1} \frac{1}{as} = \frac{1}{a} \quad (332)$$

$$\mathcal{L}^{-1} \frac{1}{a} \left(\frac{1}{a+s} \right) = \frac{1}{a} e^{-at} \quad (333)$$

Putting this together and back substituting $1/\tau = a$, we have:

$$\begin{aligned}\frac{e_o(t)}{e_i(t)} &= \mathcal{L}^{-1} \frac{1}{s^2} - \frac{1}{as} + \frac{1}{a(a+s)} \\ &= t - \frac{1}{a} + \frac{1}{a} e^{-at} \\ &= t - \tau \left(1 - e^{-t/\tau} \right)\end{aligned}\quad (334)$$

This matches the result obtained using Maxima.

Measurement

Figure 162(a) shows the theoretical ramp response of an RC lowpass network. This measurement requires a waveform that increases during a *ramp* interval of time and then drops back to zero during a *reset* interval. Triangle and sawtooth waveforms commonly available from function generators provide a ramp but not the necessary reset interval.

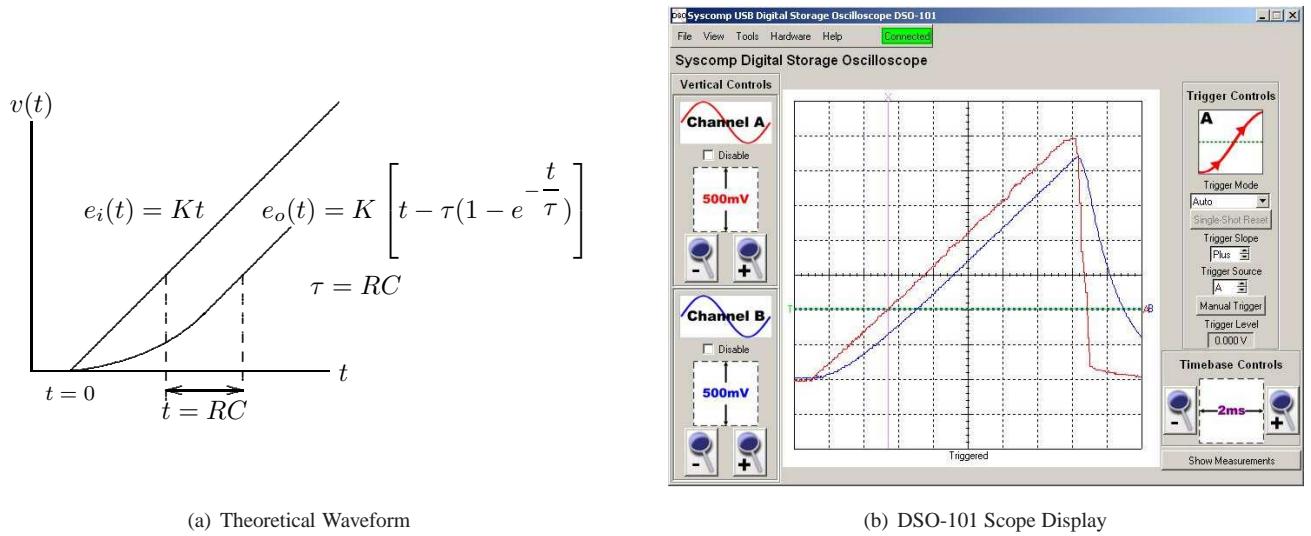


Figure 162: Ramp and RC Lowpass

In this case, the necessary ramp-reset waveform was created with equal time assigned for the ramp and the reset. The *Wavemaker* utility program was used to draw the waveform⁹⁵. Then this waveform was loaded into the WGM-101 waveform generator. The WGM-101 can then reproduce this arbitrary waveform at the desired frequency and amplitude.

The output voltage rises at the same rate as the input, but delayed by one time constant. Figure 162(b) shows the measured response when the time constant is $RC = \tau = 945\text{mSec}$. The measured delay is 937mSec.

5.3.9 Capacitor with Initial Charge

When a circuit is put into operation with an initial charge on a capacitor, the initial capacitor voltage may be modelled as a voltage source in series with the capacitor [69]. The voltage source produces a step-function of voltage at time zero. Figure 163 shows an example.

The original circuit is in figure 163(a). The switch starts in position (A) so that the capacitor is charged to supply voltage E . Then the switch moves to position (B) and the capacitor discharges through the resistor.

Figure 163(b) represents the transient condition, just after the switch moves to position (B). The charged capacitor is represented by the circuit inside the dashed box, a step voltage source E/s in series with an uncharged capacitor. The terminals X and Y are now at the edge of the dashed box.

Figure 163(c) represents the situation after the initial transient. The original capacitor and its charge are represented by voltage source E in series with the capacitor.

Behaviour

First, let's think about the physical behaviour of this circuit, based on the circuit of figure 163(a). When the switch is moved to position B, the voltage on the capacitor appears across the resistor. This drives current through the resistor. The current discharges the capacitor, and the capacitor voltage decreases. The current through the resistor decreases, so the rate of discharge decreases.

⁹⁵Wavemaker is available as a free download from the Syscomp web page, in the Downloads section.

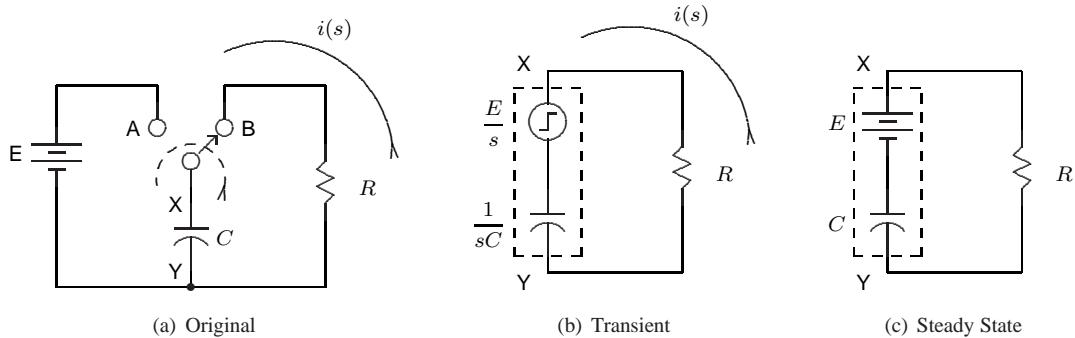


Figure 163: Capacitor with Initial Charge in RC Circuit

In summary, we'd expect to see the capacitor voltage and resistor voltage die away exponentially from an initial value of E to zero. The current would have a similar shape, since it's proportional to the voltage across the resistor. The Laplace analysis must correspond to this physical understanding of the circuit behaviour.

Laplace Analysis

In figure 163(b), the algebraic sum of voltages around the loop must equal zero.

$$+ \frac{E}{s} - V_r(s) - V_c(s) = 0 \quad (335)$$

where E is the initial voltage, $V_l(s)$ the voltage across the resistor and $V_c(s)$ is the voltage across the capacitor.

Use Ohm's law to relate the voltages to the circuit current:

$$V_r(s) = i(s)R \quad (336)$$

$$V_c(s) = i(s) \times \frac{1}{sC} \quad (337)$$

Substitute for $V_r(s)$ and $V_c(s)$ from equations 336 and 337 into equation 335:

$$+ \frac{E}{s} - i(s)R - \frac{i(s)}{sC} = 0 \quad (338)$$

Solve for the current:

$$i(s) = \frac{EC}{1+sRC} \quad (339)$$

Referring to the Laplace Transform table (section 5.3.2, page 214), entry #7 looks promising. We'll now manipulate the equation into that form. Based on that, we must get s by itself in the denominator. It also helps to recognize that $RC = \tau$, the time-constant in an RC network. After some algebra, we can obtain:

$$i(s) = \frac{E}{R} \left(\frac{1}{s + 1/\tau} \right) \quad (340)$$

Compare that with Laplace Transform pair #7:

$$\frac{1}{s-a} \iff e^{at}$$

By comparison of equation 340 with this transform pair, $a = -1/\tau$, so the time domain equation of current is:

$$i(t) = \frac{E}{R} e^{-t/\tau} \quad (341)$$

In words, the initial value of the current is E/R and it decays exponentially with a time constant $\tau = RC$. That's consistent with our understanding of the behaviour of the circuit.

Steady State

In the physical circuit of figure 163(a), the capacitor voltage V_{xy} decays to zero. How is that accomplished in the equivalent circuits of figures 163(b) and 163(c)?

In figures 163(b) and 163(c), the loop current flows into the bottom terminal of the capacitor. As a consequence, the lower terminal of the capacitor becomes more positive. This voltage is opposite in polarity to E , so the net voltage V_{xy} decreases.

After $t = \infty$ (if you're a mathematician) or $t \approx 5\tau$ (if you're an engineer), the voltage across the capacitor is exactly equal and opposite to E , so they cancel out and the output voltage V_{xy} is zero.

5.3.10 LC Resonant Circuit

Now will show how Laplace Transform applies to *resonance*, first introduced in section 4.10 on page 173. The circuit is repeated in figure 164 below. As in section 5.3.9 the capacitor initial voltage is represented by a voltage step, E/s . The capacitor and inductor are each indicated with their Laplace-transformed reactance.

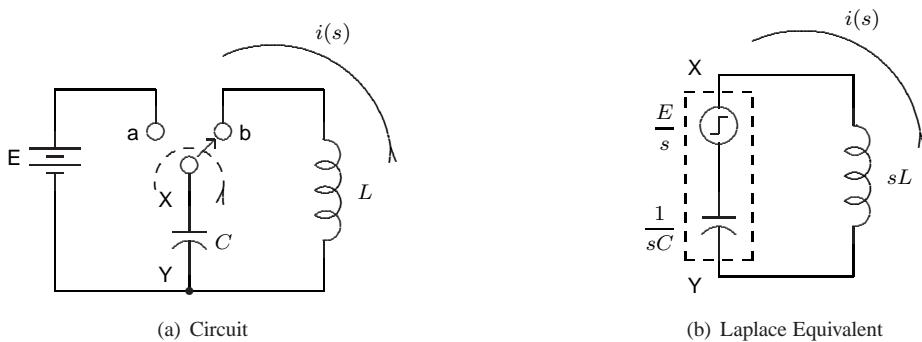


Figure 164: LC Resonant Circuit and Waveforms

We'll solve for the circuit current. We know in advance that it will have a sinusoidal shape. The equations should deliver the exact magnitude and frequency. We'll use the Laplace Transform Table, section 5.3.2 on page 214.

We begin by applying KVL to the circuit at the instant after the switch has been moved from (a) to (b).

$$+ \frac{E}{s} - V_c(s) - V_l(s) = 0 \quad (342)$$

where E is the initial voltage, $V_c(s)$ is the voltage across the capacitor and $V_l(s)$ the voltage across the inductor. That is, the initial voltage is balanced by the voltage drops across the capacitor and inductor. The voltages across the capacitor and the inductor are proportional to the loop current.

$$V_c(s) = i(s) \times \frac{1}{sC} \quad (343)$$

$$V_l(s) = i(s) \times sL \quad (344)$$

where $i(s)$ is the loop current. Using equations 343 and 344, substitute for $V_c(s)$ and $V_l(s)$ in equation 342 and solve for current:

$$i(s) = E \left[\frac{C}{1 + s^2 LC} \right] \quad (345)$$

Referring to the transform table, entry #9 looks promising. The trick now is to manipulate equation 345 into this form: that is, 1 in the numerator, and s^2 plus some constant in the denominator. The entire expression can be preceded by a constant.

After some algebra with this objective, we can obtain:

$$i(s) = \frac{E}{L} \left[\frac{1}{s^2 + \frac{1}{LC}} \right] \quad (346)$$

From other work (section 4.11, equation 203, page 203), we know that the natural resonant frequency ω_o is given by:

$$\omega_o = \frac{1}{\sqrt{LC}} \quad (347)$$

Substitute this in equation 347:

$$i(s) = \frac{E}{L} \left[\frac{1}{s^2 + \omega_o^2} \right] \quad (348)$$

Entry #9 of the Transform Table gives the transform pair:

$$\frac{1}{s^2 + a^2} \iff \frac{\sin(at)}{a} \quad (349)$$

Comparing this with equation 348, $a = \omega_o$ and we can write:

$$i(t) = \frac{E}{\omega_o L} \sin(\omega_o t) \quad (350)$$

Equation 350 says that the current waveform will have a sinusoidal shape, with frequency ω_o and maximum amplitude $E/\omega_o L$ amperes. Notice that there is no indication of decreasing amplitude. With ideal lossless components, the current will oscillate forever.

Voltage Waveform

In the Laplace domain, the voltage across the inductor is the product of the current $i(s)$ (equation 348 above) and the reactance of the inductor, sL :

$$\begin{aligned}
 v_l(s) &= i(s) \times sL \\
 &= \frac{E}{L} \left[\frac{1}{s^2 + \omega_o^2} \right] \times sL \\
 &= E \left[\frac{s}{s^2 + \omega_o^2} \right]
 \end{aligned} \tag{351}$$

Entry #10 of the Transform Table gives the transform pair:

$$\frac{s}{s^2 + a^2} \iff \cos(at) \tag{352}$$

By comparison with equation 351,

$$v_l(t) = E \cos(\omega_o t) \tag{353}$$

That is, the voltage acrosss the inductor has a maximum value of E volts and is phase shifted by 90° from the current waveform, which is a sine function of time (equation 350 above).

As a quick sanity check, substitute $t = 0$ in equation 353 and we obtain $v_l(t) = E$, which corresponds to the physical behaviour of the circuit: the initial voltage across the inductor is the precharge voltage E .

Figure 165 on page 230 shows the waveforms of current and voltage.

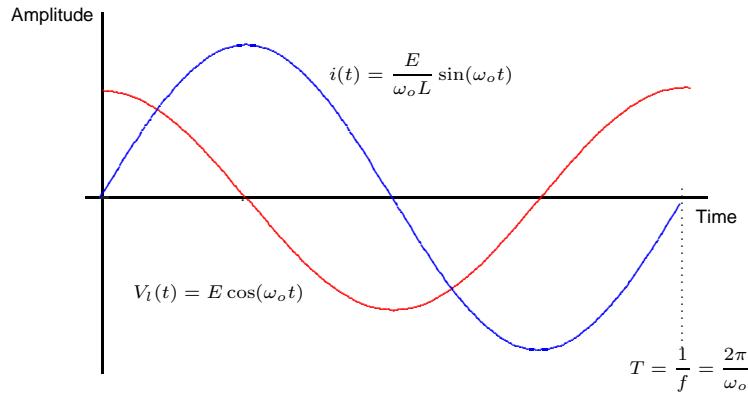


Figure 165: Current and Voltage in LC Resonant Circuit

Virtual Capacitor Voltage Waveform

Referring to figure 164(a) the voltage across the physical capacitor V_{xy} is equal to the voltage across the inductor. At this point, we have all the information we need about the circuit. We have determined the voltage across the inductor, the current through it: the amplitude, shape and frequency of the waveforms.

Consequently, what follows is academic curiosity. The curiosity arises from the equivalent circuit of figure 164(b). After it's initial step, the voltage source continues to produce a voltage E . Somehow, that has to be cancelled out in the circuit or the inductor would appear have a DC voltage across it. Does the math take care of this? As it turns out, it does. Here's the proof.

Just to be clear in the following, when we refer to a *capacitor*, we're referring to the virtual capacitor in figure 164(b), not the physical capacitor between terminals XY in figure 164(a).

We'll start in the Laplace domain. The voltage across the capacitor is the product of the current (equation 348) and the reactance of the capacitor, $1/sC$.

$$\begin{aligned} v_c(s) &= i(s) \times \frac{1}{sC} \\ &= \frac{E}{L} \left[\frac{1}{s^2 + \omega_o^2} \right] \times \frac{1}{sC} \\ &= \frac{E}{LC} \left[\frac{1}{s} \right] \left[\frac{1}{s^2 + \omega_o^2} \right] \end{aligned} \quad (354)$$

As usual, Maxima would make short work of this, but we'll do it by hand. A review of the Table of Transforms (page 214 and reference [4]) does not turn up anything that resembles the product of the two terms in square brackets. However, we do find individual transforms for each of these terms. Consequently, if we can manipulate the two squared-bracket terms into a sum rather than a product, we can find the overall transform. That requires a partial fraction expansion on the square bracket material as follows:

$$\left[\frac{1}{s} \right] \left[\frac{1}{s^2 + \omega_o^2} \right] = \left[\frac{A}{s} \right] + \left[\frac{B}{s^2 + \omega_o^2} \right] \quad (355)$$

To make this work, we need the values for A and B . Multiply the LHS and find a common denominator for the RHS of equation 355:

$$\begin{aligned} \frac{1}{s(s^2 + \omega_o^2)} &= \frac{A(s^2 + \omega_o^2) + B(s)}{s(s^2 + \omega_o^2)} \\ &= \frac{As^2 + A\omega_o^2 + Bs}{s(s^2 + \omega_o^2)} \\ &= \frac{(As^2 + Bs) + A\omega_o^2}{s(s^2 + \omega_o^2)} \end{aligned} \quad (356)$$

Now we can compare the numerators of the LHS and RHS to find the values of A and B . First, the constant term:

$$\begin{aligned} A\omega_o^2 &= 1 \\ A &= \frac{1}{\omega_o^2} \end{aligned} \quad (357)$$

Now the factors of s :

$$\begin{aligned} As^2 + Bs &= 0 \\ B &= -As \end{aligned} \quad (358)$$

Substitute the value for A that we found in equation 357:

$$B = -\frac{s}{\omega_o^2} \quad (359)$$

Now we can rewrite equation 354, imposing the expanded fraction:

$$\begin{aligned}
 v_c(s) &= \frac{E}{LC} \left\{ \left[\frac{A}{s} \right] + \left[\frac{B}{s^2 + \omega_o^2} \right] \right\} \\
 &= \frac{E}{LC} \left\{ \left[\frac{1}{\omega_o^2} \times \frac{1}{s} \right] + \left[-\frac{s}{\omega_o^2} \times \frac{1}{s^2 + \omega_o^2} \right] \right\}
 \end{aligned} \tag{360}$$

Recognize that $LC = 1/\omega_o^2$, substitute for LC and equation 360 simplifies to:

$$v_c(s) = E \left[\frac{1}{s} - \frac{s}{s^2 + \omega_o^2} \right] \tag{361}$$

Now we can make the Laplace Transforms:

$$\begin{aligned}
 \mathcal{L}^{-1} \left\{ \frac{1}{s} \right\} &= 1 \\
 \mathcal{L}^{-1} \left\{ \frac{1}{s^2 + a^2} \right\} &= \cos(at)
 \end{aligned}$$

Then by comparison, equation 361 transforms into the time-domain equation for capacitor voltage:

$$\begin{aligned}
 v_c(t) &= E [1 - \cos(\omega_o t)] \\
 &= E - E \cos(\omega_o t)
 \end{aligned} \tag{362}$$

Now we can test KVL in this circuit by summing voltages around the loop. Based on the direction of current flow, the KVL loop equation is:

$$+ E - v_l(t) - v_c(t) = 0 \tag{363}$$

Substitute for $v_l(t)$ from equation 353 and $v_c(t)$ from equation 362:

$$\begin{aligned}
 + E - v_l(t) - v_c(t) &= +E - (E \cos \omega_o t) - (E - E \cos(\omega_o t)) \\
 &= 0
 \end{aligned} \tag{364}$$

Evidently KVL is satisfied in the equivalent circuit.

5.3.11 RLC Resonant Circuit

Circuits inevitably have resistance. As a result, the oscillations in a resonant circuit die away unless something pumps more energy into the circuit. The circuit resistance also has a minor effect on the resonant frequency, causing it to shift slightly. We distinguish between the *undamped* resonant frequency ω_o (which occurred in the lossless LC circuit of section 5.3.10) and the *damped* resonant frequency ω_d which applies to this circuit. As the resistance decreases, ω_d approaches ω_o .

Figure 166 shows the RLC resonant circuit and its Laplace equivalent. As in section 5.3.10, the initial charge on the capacitor is modelled in the Laplace circuit by a step function generator, E/s .

We'll start by going after the circuit current in the Laplace domain, $i(s)$. Applying KVL to the circuit at the instant after the switch has been moved from (a) to (b), we have:

$$+ \frac{E}{s} - V_r(s) - V_c(s) - V_l(s) = 0 \tag{365}$$

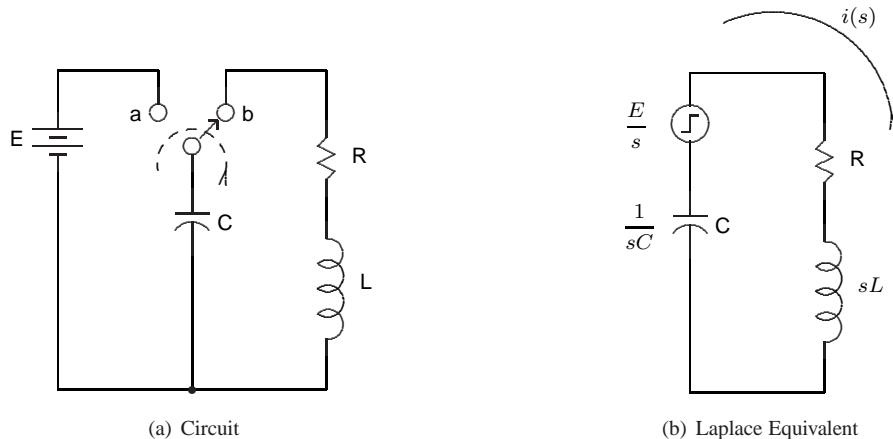


Figure 166: RLC Resonant Circuit

where the terms of the equation are:

- | | | |
|----------|------------------------------|------------|
| E | the initial voltage | |
| $V_r(s)$ | voltage across the resistor | $i(s)R$ |
| $V_l(s)$ | voltage across the inductor | $i(s)sL$ |
| $V_c(s)$ | voltage across the capacitor | $i(s)1/sC$ |

Substitute from the last column of the above list into equation 365 and after some manipulation we obtain:

$$i(s) = \frac{E}{L} \left[\frac{1}{s^2 + \frac{R}{L}s + \frac{1}{LC}} \right] \quad (366)$$

Now we will do some substitutions that move us away from specific component values to the general properties of resonant circuits. From previous work, we know that $\omega_0 L / R = Q$, where Q is the *quality factor* of the circuit, the ratio of inductive reactance to the resistance. A circuit with a larger Q factor has better selectivity in filter circuits, hence *higher quality*. Then we can substitute:

$$+ \frac{R}{L} = \frac{\omega_o}{Q} \quad (367)$$

Also, as usual by now, we know that $\omega_o = 1/\sqrt{LC}$, so we can substitute:

$$+ \frac{1}{LC} = \omega_o^2 \quad (368)$$

Substitute from equations 367 and 368 into equation 366:

$$i(s) = \frac{E}{L} \left[\frac{1}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \right] \quad (369)$$

Now we must obtain the inverse transform of this to determine $i(t)$, current in the time domain. The Table of Transforms (section 5.3.2) shows this for entry #11:

$$\frac{1}{(s - b)^2 + a^2} \iff \frac{e^{bt} \sin at}{a} \quad (370)$$

Now we'll try to get the denominator of equation 369 into the same form as the LHS of this transform pair. We will use the technique of *completing the square* [70]. This technique is illustrated in equation 371. We convert the form of the LHS to the form of the RHS of equation, while maintaining the equality.

$$s^2 + As + B = (s + X)^2 + Y \quad (371)$$

Where X and Y are unknown at this point. To find them, expand out the RHS of equation 371:

$$\begin{aligned} s^2 + As + B &= (s^2 + 2sX + X^2) + Y \\ As + B &= 2sX + (X^2 + Y) \end{aligned} \quad (372)$$

Equating like terms from equation 372 and solving for X and Y , we have $X = A/2$, $Y = B - A^2/4$. Back substitute these into equation 371:

$$s^2 + As + B = \left(s + \frac{A}{2}\right)^2 + \left(B - \frac{A^2}{4}\right) \quad (373)$$

This *completes the square*. Now use equation 373 as a guide to reformat equation 369, that is, $A = \omega_o/Q$, $B = \omega_o^2$.

$$\begin{aligned} i(s) &= \frac{E}{L} \left[\frac{1}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \right] \\ &= \frac{E}{L} \left[\frac{1}{\left(s + \frac{\omega_o}{2Q}\right)^2 + \left(\omega_o^2 - \frac{\omega_o^2}{4Q^2}\right)} \right] \end{aligned} \quad (374)$$

Now, by comparison of equation 374 with 370 we can find the inverse Laplace transform. From the transform pair,

$$b = -\frac{\omega_o}{2Q} \quad (375)$$

$$\begin{aligned} a &= \sqrt{\omega_o^2 - \frac{\omega_o^2}{4Q^2}} \\ &= \omega_o \sqrt{1 - \frac{1}{4Q^2}} \end{aligned} \quad (376)$$

The quantity a is the damped natural frequency, ω_d . Notice as Q becomes very large, ω_d approaches ω_o .

Substitute for a and b in the RHS of the transform pair:

$$\begin{aligned}
 i(t) &= \frac{E}{L} \left[e^{bt} \frac{\sin at}{a} \right] \\
 &= \frac{E}{\omega_d L} \left[e^{-\frac{\omega_o t}{2Q}} \sin \omega_d t \right] \\
 &= \left[\frac{E}{\omega_d L} e^{-\frac{t}{2Q/\omega_o}} \right] \sin \omega_d t
 \end{aligned} \tag{377}$$

where (to recap):

$$\omega_o = \frac{1}{\sqrt{LC}} \tag{378}$$

$$Q = \frac{\omega_o L}{R} \tag{379}$$

$$\omega_d = \omega_o \sqrt{1 - \frac{1}{4Q^2}} \tag{380}$$

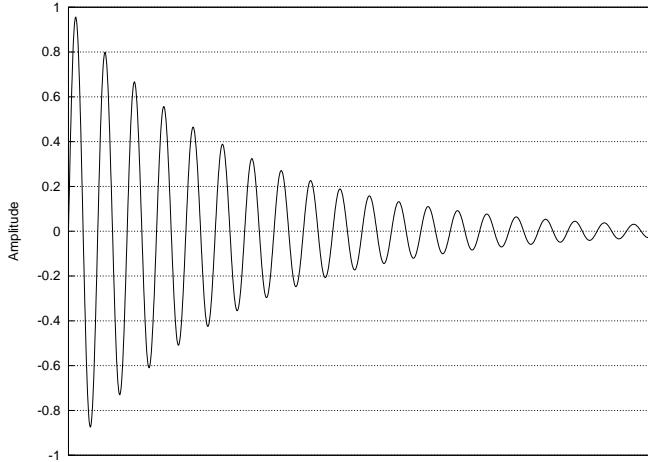


Figure 167: Damped Sine Wave

Equation 377 is known as a *decaying* or *damped* sine wave. Figure 167 shows an example. This particular waveform was obtained from a high Q circuit because it rings (oscillates) for many cycles before the amplitude decreases to zero.

The form of equation 377 suggests the following detailed interpretation: The waveform of current is a sine wave at a frequency ω_d , the damped natural frequency. The amplitude of this sine wave begins at a magnitude of $E/\omega_d L$ and then dies away exponentially. The time constant of the decay is given by $2Q/\omega_o$.

This waveform is common in many physical situations, such as the *boing* sound wave created by striking a gong. Increasing the damping (lowering the Q factor) turns the *boing* into more of a *thud* and the sine wave decays more quickly.

It's important to understand the operation of the RLC resonant circuit. However, the RLC circuit is usually not used in circuit designs to produce damped sine waves. Inductors are bulky and expensive, and it's difficult to control the resistance in the circuit to produce the desired Q factor and decay rate. If one needs an underdamped, decaying sine wave, a better approach is to synthesize equation 377 with an operational amplifier circuit (see section 13.13).

5.3.12 Two Stage (Second Order) Lowpass, Impulse Response

In section 5.2.9 we showed the frequency domain analysis of the two stage RC lowpass filter of figure 168.

We found that the transfer function is a second-order lowpass filter with a Q factor of $1/3$. Now, in accordance with our original scenario (section 5.1), let us investigate the effectiveness of this filter in attenuating a noise spike.

We can approximate a noise spike by an ideal impulse, so we need to know the impulse response of the filter. The impulse response is the inverse Laplace transform of the transfer function. Here is how we use Maxima to determine that:

1. We enter the transfer function of the second-order RC lowpass (equation 303). In this equation, w represents ω_o :

```
(%i48) w^2 / (s^2 + (w/q) * s + w^2);
```

2. Maxima formats and echoes this back:

$$(\%o48) \frac{w^2}{w^2 + \frac{s w}{q} + s^2}$$

3. We enter the command to find the inverse Laplace transform of the previous equation:

```
(%i49) ilt(% , s , t);
```

4. Maxima asks in effect whether the transfer function is overdamped, critically damped or underdamped. If the function is underdamped, the impulse response will be oscillatory, that is, there will be overshoot and ringing.

Is (2 q - 1) (2 q + 1) w² positive, negative, or zero?

In this case, the Q factor is $1/3$, so the correct answer is negative.

```
negative;
```

5. Maxima now determines the inverse Laplace transform.

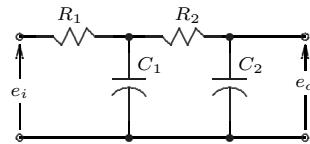


Figure 168: 2nd Order RC Lowpass Filter

$$(380) \quad \frac{2 q w e^{-\frac{t w}{2 q}} \sinh \left(\frac{\sqrt{1-4 q^2} t w}{2 q} \right)}{\sqrt{1-4 q^2}}$$

This is the time-domain response to an impulse.

6. We can simplify this equation somewhat before plotting:

- Replace $w (= \omega_o)$ with $1/\tau$
- Substitute $Q (= q) = 1/3$ in the expression

$$\frac{1}{2Q} \sqrt{1 - 4Q^2} = 1.13$$

Then we obtain the much more manageable:

$$f(t) = \frac{1}{\tau} \left[\frac{1}{1.13} e^{\frac{-1.52t}{\tau}} \sinh \left(\frac{1.13t}{\tau} \right) \right] \quad (381)$$

7. Replace t/τ by the variable x , so the horizontal axis is units of time relative to the time-constant.
8. Plot the amplitude vs time with Gnuplot or some equivalent. The result of plotting the quantity within the square brackets of equation 381 is shown as the solid-line trace in figure 169.

As in previous cases, this assumes an impulse signal that has an area of 1 volt-second. If that's not the case, then the output must be scaled by K , where K is the area of the impulse in volt-seconds.

5.3.13 Comparing the Filters

Now we are in a position to compare the single stage (first-order) RC lowpass with the two-stage (second-order) RC filter. We'll assume a unit impulse input signal.

The response of a single stage RC lowpass filter was shown in figure 157(a) on page 219. Assume a unit impulse ($K = 1$) and regroup the equation for comparison purposes:

$$f(t) = \frac{1}{\tau} \left[e^{-\frac{t}{\tau}} \right] \quad (382)$$

As we did with equation 381 we can put $t/\tau = x$ and plot everything inside the square braces. This is shown as the dashed trace in figure 169.

Two things are immediately evident:

- the second-order RC lowpass is much more effective (by a factor of four, approximately) than the first-order RC lowpass in reducing the amplitude of a noise pulse.
- the leading edge of the second-order RC lowpass pulse rises at a much lower rate. This could be significant if the filter is being used to reduce electrical noise, which is often related to the rate of change of signals⁹⁶.

⁹⁶Figure 169 shows an infinite rate-of-rise of the voltage. In practice, the rate of rise is determined by the amplitude of the pulse and the RC values, as shown in figure 157(c) on page 219.

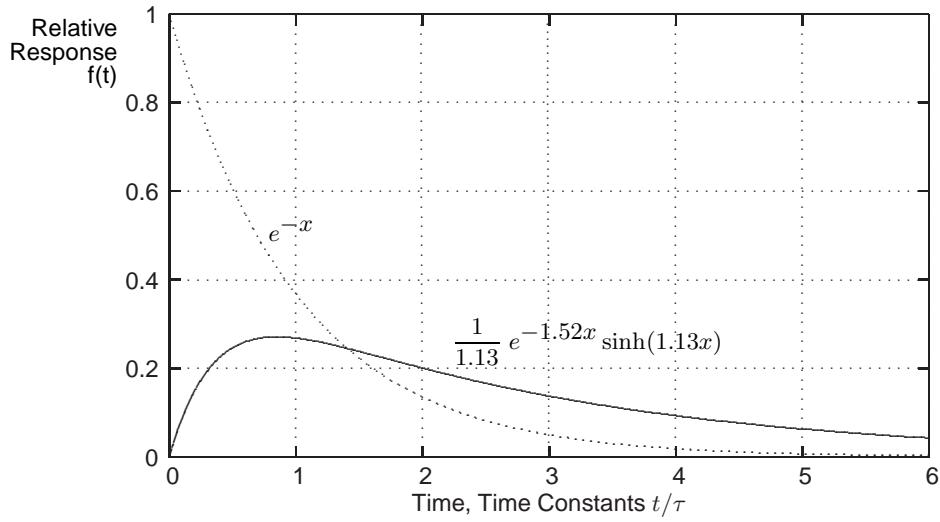


Figure 169: Impulse Response of 2nd Order Lowpass Filter

It should be emphasised that this is an analysis of ideal components and the non-ideal behaviour of capacitors in particular should be considered in the design of an effective noise filter. For example, the *equivalent series resistance* of a capacitor and its *lead inductance* may impact its effectiveness in this application.

As a practical consideration, the two series resistors must be proportioned so they do not cause excessive voltage drop for the DC current.

5.4 Arbitrary Waveforms

To this point, we have assumed that the input signal is an impulse, step or ramp function of time. Not infrequently one would like to determine the response of an electrical circuit to some arbitrary input waveform. Then it is necessary to obtain the Laplace transform for that input waveform. In this section we show various examples of such waveforms. We will need two tools: the *linearity* property and the *time-shifting* property.

5.4.1 Linearity Property

This is straightforward. The transform of the sum of two time functions is equal to the sum of the individual transforms.

$$\mathcal{L}\{f_1(t) + f_2(t)\} = f_1(s) + f_2(s)$$

Then to determine the Laplace transform of some time function:

- break it up into a sum of simpler components
- determine the transform of each of these components
- add the transformed components

More generally, these components can be scaled by constants, that is:

$$\mathcal{L}\{af_1(t) + bf_2(t)\} = af_1(s) + bf_2(s)$$

where a and b are constants.

5.4.2 Time Shift Property

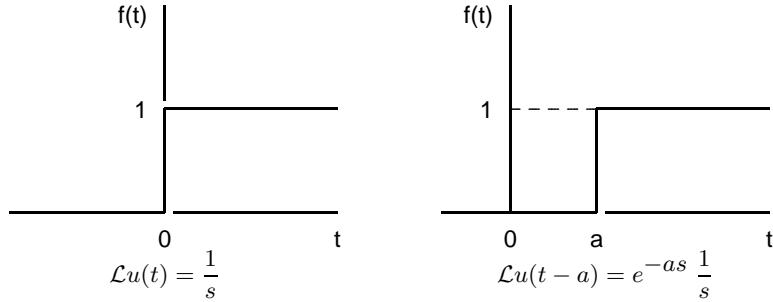


Figure 170: Shift Property

The time shift of a time function $f(t)$ by an amount a in the positive time direction is equivalent to multiplication of the Laplace transform $f(s)$ by e^{-as} .

The example of figure 170 shows the Laplace transform of a step function that has been time-shifted by an amount a .

5.4.3 Example: Square Pulse

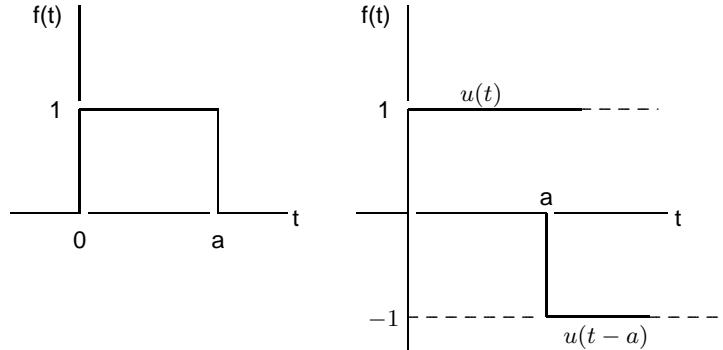


Figure 171: Pulse Waveform

Figure 171 shows how a pulse waveform may be constructed by adding together two step functions. A positive unity step initiates the waveform. After some time a an equal negative step brings the waveform back to zero. The time domain function is:

$$f(t) = u(t) - u(t-a) \quad (383)$$

Now we can find the Laplace transform of each of these step waveforms. From section 5.3.5 the transform of the first step function is:

$$\mathcal{L}\{u(t)\} = \frac{1}{s} \quad (384)$$

The second step waveform is multiplied by a constant -1 and shifted by an amount a , so the Laplace transform is:

$$\mathcal{L}\{-u(t-a)\} = -\left[e^{-as} \frac{1}{s}\right] \quad (385)$$

Putting these two together, we have

$$\mathcal{L}u(t) - u(t-a) = \frac{1 - e^{-as}}{s} \quad (386)$$

5.4.4 Example: Triangular Pulse

A triangular pulse is shown in the upper half of figure 172. Davis [71] suggests the following clever technique for determining the Laplace transform of this waveform.

Recall that a ramp waveform is the integral of a step waveform, which is in turn the integral of an impulse. Then a ramp is the double integral of an impulse. In the Laplace domain, integration is accomplished with a multiplication by $1/s$. So the upper waveform in figure 172 may be considered the double integral of the lower waveform, a series of impulses.

The first unit impulse starts a unit ramp which runs for 1 second. The second impulse, amplitude -2 units, cancels the first ramp and creates a negative-going unit ramp. That runs for a further second until it is cancelled by the third impulse, which creates a positive unit ramp.

Impulses are easy to deal with since the Laplace transform of a unit impulse is simply 1 . Then the first impulse is $+1$.

The second impulse is -2 shifted by one second: $-2e^{-s}$

The third impulse is $+1$ shifted by 2 seconds: $+1e^{-2s}$.

Putting this together, the Laplace transform of the impulses is:

$$f(s) = 1 - 2e^{-s} + e^{-2s}$$

The Laplace transform of the triangle pulse is the double integral of this:

$$f(s) = \frac{1 - 2e^{-s} + e^{-2s}}{s^2}$$

A similar technique, the single integration of a series of impulses, could be used to create a stepped waveform.

5.5 Appendices

5.5.1 Laplace Transform of Unit Impulse

The Laplace transform of some function $f(t)$ is given by⁹⁷:

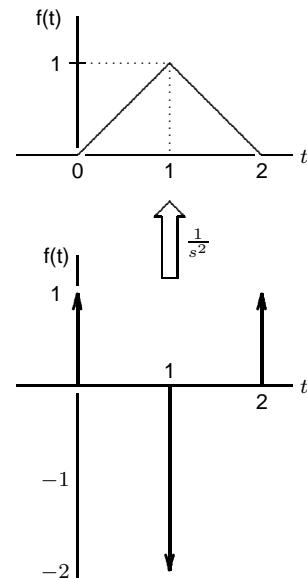


Figure 172: Triangular Pulse

⁹⁷The approach in this section is based on [72]

$$\mathcal{L}f(t) = \int_0^\infty e^{-st} f(t) dt \quad (387)$$

Referring to figure 153(b) on page 215, during the interval $t = 0$ to $t = W$ seconds the unit impulse time-domain function is:

$$f_\delta(t) = \frac{1}{W} \quad (388)$$

The impulse function is zero outside the region between $t = 0$ to $t = W$, so these can become the limits of the integration. Then:

$$\begin{aligned} \mathcal{L}f_\delta &= \int_0^W e^{-st} \left(\frac{1}{W} \right) dt \\ &= \frac{1}{W} \int_0^W e^{-st} dt \end{aligned} \quad (389)$$

Use the identity

$$\int e^{ax} = \frac{1}{a} e^{ax} \quad (390)$$

to expand equation 389:

$$\begin{aligned} \mathcal{L}f_\delta &= \frac{1}{W} \int_0^W e^{-st} dt \\ &= \frac{1}{W} \left(\frac{1}{-s} e^{-st} \Big|_0^W \right) \\ &= \frac{1}{W} \left(\frac{1}{-s} e^{-Ws} - \frac{1}{-s} e^{0s} \right) \\ &= \frac{1 - e^{-Ws}}{Ws} \end{aligned} \quad (391)$$

The final step is to make the width of the pulse W approach zero. Unfortunately, this puts zero in the numerator and denominator, which makes the expression indeterminate. However, because the numerator and denominator are *both* zero, L'Hôpital's Rule [73] allows us to replace both numerator and denominator by their derivatives⁹⁸.

Then

$$\begin{aligned} \lim_{W \rightarrow 0} \frac{1 - e^{-Ws}}{Ws} &= \lim_{W \rightarrow 0} \frac{0 - We^{-Ws}}{W} \\ &= \lim_{W \rightarrow 0} e^{-Ws} \\ &= 1 \end{aligned} \quad (392)$$

QED: the Laplace transform of the unit impulse is 1.

⁹⁸L'Hôpital's Rule also applies if the numerator and denominator approach infinity.

5.5.2 Laplace Transform of the Unit Step

The *unit step*, known as $u(t)$ has a value of unity for all times greater than zero. Plugging that into the definition of the Laplace transform (equation 387 above), we have:

$$\begin{aligned}
 \mathcal{L}f(t) &= \int_0^\infty e^{-st} u(t) dt \\
 &= \int_0^\infty e^{-st} 1 dt \\
 &= -\frac{1}{s} e^{-st} \Big|_0^\infty \\
 &= -\frac{1}{s} [e^{-\infty s} - e^{-0s}] \\
 &= -\frac{1}{s} [0 - 1] \\
 &= \frac{1}{s}
 \end{aligned} \tag{393}$$

That's it!

5.5.3 Gnuplot Magnitude and Phase Plotting Routines

The Linux version of the Gnuplot plotting program was used to create the magnitude and frequency response plots in this section. Gnuplot is also available to run under Windows operating systems.

The lines `set terminal fig` and `set output "mag-single-tc-lp.fig"` set up the program to dump its output into a data file suitable for the XFig drawing program⁹⁹.

On a Linux system, the line `set terminal x11` directs the output to the display. Under Windows, you may need to choose some other terminal. Check the Gnuplot manual under ‘terminal’.

For reasons known only to Gnuplot, directing the output to the display with the `set terminal x11` command has the effect of disabling the output to the Xfig formatted file, so you can't do both at once. Start with the display enabled so you can see the result. When you have a satisfactory display, you can create an XFig file (or some other format) by commenting out the lines `set terminal x11` and `replot` with a hash mark at the start, as shown in the phase plot listing below.

First Order Lowpass Amplitude Response

```
#List of commands to plot 1st order lowpass filter magnitude response
set terminal fig
set output "mag-single-tc-lp.fig"
set ylabel "Amplitude, db"
set xlabel "Frequency, $\omega/\omega_o$"
set grid
set nologscale y
set logscale x
plot [0.01:100] (10.0/2.302)*log(1/(1+(x)**2)) title '' with lines
# Factor 2.302 converts natural log to log-base-10
```

⁹⁹XFig is a drawing program available on Linux systems. The Windows equivalent is JFig.

```
set terminal x11
replot
pause -1
```

First Order Lowpass Phase Response

```
#List of commands to plot 1st order lowpass filter phase response
set terminal fig
set output "phase-single-tc-lp-a.fig"
set ylabel "Phase, degrees"
set xlabel "Frequency, $\omega/\omega_o$"
set grid
set nologscale y
set logscale x
plot [0.01:100] (180/3.14)*(-atan(x)) title '' with lines
# Factor 180/3.14 converts radians to degrees
# set terminal x11
# replot
pause -1
```

Second Order Lowpass Amplitude Response

```
#List of commands to plot 2nd order lowpass filter magnitude response
#There are three plots, each with different Q factor
set terminal fig
set output "mag-quadratic-lp.fig"
set ylabel "Amplitude, db"
set xlabel "Frequency, $\omega/\omega_o$"
set grid
set nologscale y
set logscale x
plot [0.05:50] \
    -(10.0/2.302)*log(((1-(x**2))**2)+((2*0.25*x)**2)) title '' with lines,\ 
    -(10.0/2.302)*log(((1-(x**2))**2)+((2*0.5*x)**2)) title '' with lines,\ 
    -(10.0/2.302)*log(((1-(x**2))**2)+((2*1*x)**2)) title '' with lines
set terminal x11
replot
pause -1
```

Second Order Lowpass Phase Response

```
# List of commands to plot 2nd order lowpass filter phase response
# Getting this to work was tricky. The actual function to plot
# vs angle is atan((x/1)/(1-x**2)). However, this takes the function
# into a quadrant with a negative x axis and positive y axis,
# and so the angle jumps as it passes through +90 degrees,
# to -90 degrees and then decreases back to zero. We want a function
# that smoothly decreases from 0 to -180 degrees. To get this
# function into the right quadrant, I inverted the expression
```

```

# and then added -90 degrees.
set terminal fig
set output "phase-quadratic-q-lp.fig"
set ylabel "Phase, degrees"
set xlabel "Frequency, $omega/omega_o$"
#set key 6.2, 250
set grid
set nologscale y
# set angles degrees
# set yrange [-180:0]
set logscale x
set angles degrees
plot [0.05:50] -90+atan((1-x**2)/(x/1)) title '' with lines,\ 
    -90+atan((1-x**2)/(x/2)) title '' with lines,\ 
    -90+atan((1-x**2)/(x/4)) title '' with lines
set terminal x11
replot
pause -1

```

Second-Order RC Lowpass ($Q = 1/3$) Impulse Response

```
#List of commands to plot impulse response of 2nd-order lowpass filter
#For explanation, see 'laplace-cookbook' section.
```

```

set terminal fig                      # direct graphical output to a file
set output "impulse-curve.fig"        # establish name of the file
set xlabel "Time Constants"          # label x axis
# set ylabel "Charge\\Fraction"       # label y axis
# set key 6.2, 250                  # position graph title
set grid                            # turn grid on
plot [0:10] (2.71818**(-1.52*x))*(sinh(1.13*x))      # show output on terminal
set term x11                         # redo the plot
replot
pause -1

```

5.5.4 Further Reading

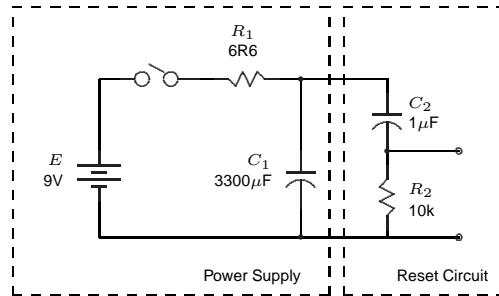
Books on electric circuit analysis and signal processing usually have a section on the Laplace transform. These tend to be math-based expositions and not particularly intuitive.

Fourier analysis can be understood as a correlation process that searches the original signal for its sine and cosine (or *in-phase and quadrature*) components. The Laplace transform is more difficult to put on that sort of physical intuitive basis. Two sources that try are Smith [74] and Lyons [75].

Smith's work is available on the internet, and is more detailed and extended. Lyon's text is available as a Low Price Edition through Abe Books [76]. Both books are excellent for studies in digital signal processing.

5.6 Exercises

1. The circuit shows a common technique for generating a reset pulse when the power supply is switched on.



As the power supply voltage rises, the voltage across capacitor C_2 stays constant for a long enough period that the voltage across resistor R_2 rises up to some maximum and then decays away to zero. The result is a pulse of limited duration. (You can also think of the $C_2 R_2$ network as an approximate differentiator for the rising portion of the supply voltage transient.) However, because of the power supply filter capacitor C_1 and the internal resistance R_1 of the power supply, the supply voltage does not rise instantly to +9V when the supply is turned on.

Use Laplace techniques to determine the shape and amplitude of the reset pulse.

6 The Semiconductor Diode

6.1 Semiconductor Diode

The basic operation of a semiconductor diode is very simple: it conducts current in the forward direction and blocks current in the reverse direction. It's the *one-way valve* of electronics.

It has other useful properties. When conducting¹⁰⁰, the forward voltage drop is approximately equal to 0.6 volts, independent of forward current. This forward voltage drop is a predictable function of temperature: it decreases at a rate of $2.2\text{mV}^{\circ}\text{C}$.

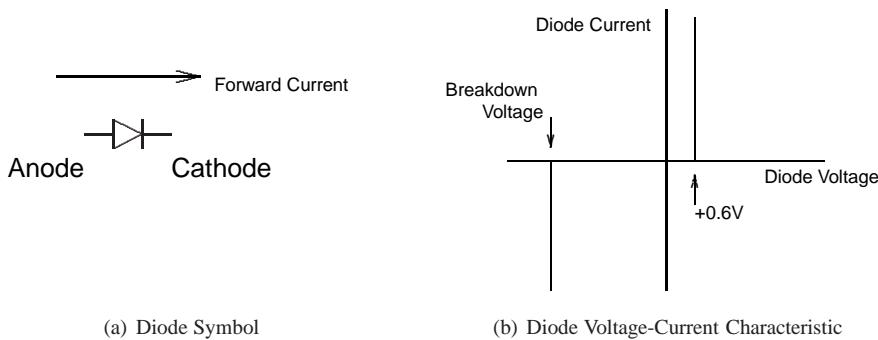


Figure 173: The Semiconductor Diode

The diode symbol is shown in figure 173(a). When the anode is made positive with respect to the cathode, the diode conducts in its forward direction.

The V-I characteristic of the diode is shown in idealized form in figure 173(b).

The reverse breakdown voltage ranges from 50 to thousands of volts. The forward current capability ranges from milliamperes to hundreds of amps. Both these properties depend on the construction of the diode.

Two very common diode types are the 1N4148 small-signal diode, and the 1N400x series of 1 amp power rectifier diodes.

Exact Diode Equations

For much electronics work, it is entirely satisfactory to characterize the silicon diode as an ideal diode with a fixed forward voltage drop of 0.6 volts. However, there are occasions when it is important to know the exact voltage-current function, which is

$$I_f = I_s(e^{V_f/NV_t} - 1) \quad (394)$$

where

¹⁰⁰These figures are for the silicon diode. Different values apply to other types of diodes. For example, the germanium diode forward drop is about 0.2V.

I_s is the *saturation current*, typically equal to 10^{-12} to 10^{-15} amps.

V_f is the forward voltage applied to the diode, volts.

V_t is the *thermal voltage* approximately equal to 0.026 volts at room temperature.

N is a constant between 1 and 2 that depends on the construction of the diode. For simplicity, we'll generally assume that N is equal to unity.

The thermal voltage V_t is given by

$$V_t = \frac{kT}{q_e} \quad (395)$$

where

k is Boltzmann's constant, 1.38×10^{-23} joules/kelvin

T is the absolute temperature in degrees Kelvin (ie, *kelvins*)

q_e is the electron charge, 1.6×10^{-19} coulomb

For most values of forward voltage the exponential term is much larger than 1. Making this approximation and assuming that N is unity, equation 394 can be written as:

$$I_f = I_s e^{V_f/V_t} \quad (396)$$

This is the form of the diode equation that we will assume throughout the remainder of the book.

The saturation current is far too small to measure directly, so it must be inferred using equation 396 and a set of values of forward-voltage and forward current at some known temperature. A typical diode function is shown in figure 174 for $I_s = 10^{-14}$ amps.

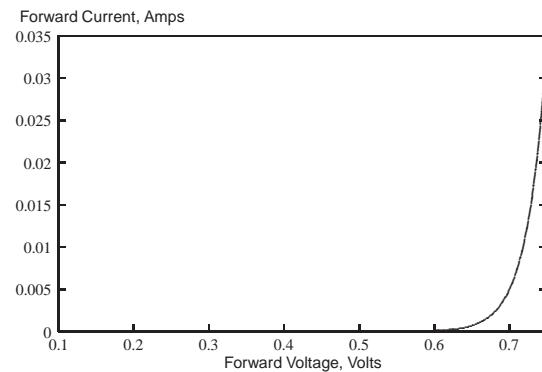


Figure 174: Diode Equation

6.2 Diode Logic OR Gate

A diode OR Gate, is shown in figure 175.

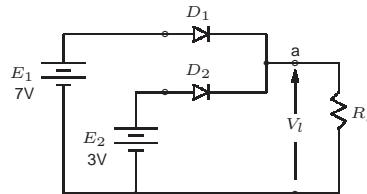


Figure 175: Diode OR Gate

This circuit *selects the larger of the two input voltages*, E_1 or E_2 , which then appears across the load resistor R_l . To analyse a circuit like this one, with multiple diodes, it is necessary to first decide which diodes are conducting. To start with, a conducting diode may be treated as a short circuit: a non-conducting diode as an open circuit.

In the absence of solid evidence which diodes are conducting, make a guess and test the assumptions. For example:

Assume D_2 is conducting and D_1 is not Then the E_2 source is connected to point a by a short circuit, forcing the output voltage V_l to be 3 volts. Then the voltage across D_1 must be $7 - 3 = 4$ volts in the forward direction, so D_1 must be conducting and the original assumption was incorrect.

Assume both diodes are conducting The point a is connected to both a 7V and 3V source. This would drive current from the 7V source into the 3V source, backwards through D_2 . So diode D_2 cannot be conducting and this assumption is also incorrect.

Assume D_1 is conducting and D_2 is not. Then point a is at 7 volts, reverse biasing D_2 . D_2 is an open circuit, and the output voltage is equal to 7 volts, the larger of the two input voltages.

This circuit is known as an OR gate because it can provide the Boolean OR operation when the input voltages represent logic signals. For example, suppose logic 0 is represented by 0 volts and logic 1 by +5 volts. The gate inputs are known as A and B, and the output F. Then (ignoring the 0.6 volt forward drop of the diode) we have:

A	B	F
0	0	0
0	1	1
1	0	1
1	1	1

which is the OR logic function¹⁰¹.

Large computers can and have been built using this kind of circuitry, but the 0.6 volt drop in the diodes requires that an amplifier follow the diode logic function.

The diode OR gate need not be limited to 2 inputs and can be extended to as many inputs as needed.

6.3 Diode Logic AND Gate

The AND Gate, which is the inverse of the OR Gate, is shown in figure 175.

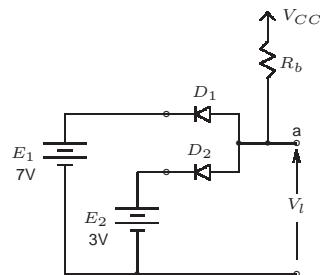


Figure 176: Diode AND Gate

The AND gate requires an additional source of bias current, which is provided by resistor R_b (for *bias*) and the positive supply voltage V_{CC} . The positive supply must be equal to or larger than the largest input voltage. The effect of a load resistor R_l complicates the analysis, so we'll leave it out for now¹⁰².

¹⁰¹If we define the logic variables differently (logic 0 is +5V and logic 1 is 0V) then the gate is a logic AND gate.

¹⁰²In a practical circuit, there would need to be a load resistance, but it would be made much larger than R_b .

The AND gate *selects the smaller of the two input voltages*.

Again, we treat a conducting diode as a short circuit and a non-conducting diode as an open circuit. Suppose neither diode is conducting and the power supply voltage V_{CC} is switched on. Both diodes are initially reverse biased and open circuits. No current flows through R_1 , and so there is no voltage drop across it. So the voltage at point **a** will rise toward V_{CC} . When point **a** rises above the smaller of the two input voltages E_2 , (3V in this case) diode D_2 conducts. At that point the voltage at point **a** cannot rise any higher, because the ideal voltage source E_2 holds it there.

The other diode D_1 , connected to a larger input voltage, remains reverse biased throughout this sequence of events.

This circuit is known as an AND gate because it can provide the Boolean AND operation when the input voltages represent logic signals. For example, suppose logic 0 is represented by 0 volts and a logic 1 by +5 volts. The gate inputs are known as A and B, and the output F. Then (ignoring the 0.6 volt forward drop of the diode) we have:

A	B	F
0	0	0
0	1	0
1	0	0
1	1	1

which is the AND logic function¹⁰³.

Like the OR gate, this construction requires that an amplifier follow the diode logic function to make up for the 0.6 volt diode forward voltage drop. The diode AND gate can also be extended to as many inputs as needed.

6.4 Half Wave Rectifier

The process of *rectification*¹⁰⁴ is the conversion of alternating current to direct current. The alternating current can be any waveform, but is often a 60Hz sine wave from the AC power line.

The circuit and waveform are shown in figure 177. Notice the effect of the 0.6V diode forward voltage drop.

6.5 Full Wave Rectifier

The half-wave rectifier (section 6.4) is useful for small load currents. However, for larger currents it is often desirable to generate a full-wave rectified waveform.

This can be accomplished with two half-wave rectifiers, each driven by a sine wave, where the two sine waves are inverted polarity with respect to each other. Then one rectifier conducts on the positive half-cycle of the first source and the other rectifier conducts on the positive half-cycle of the other source. The diodes function as an OR gate, selecting the larger of the two waveforms. Because the sources are opposite in polarity, the positive half-cycles fill in the dead spots. A filter capacitor is attached to the output to remove the AC content and save the DC content.

The circuit and waveforms are shown in figure 178.

Antiphase Waveforms

¹⁰³If we define the logic variables differently (logic 0 is +5V and logic 1 is 0V) then the gate is a logic OR gate.

¹⁰⁴The term *rectification* means *adjust or make right*. So why is it used here? Why not something more apropos like *one-way-ification*? In the early days of electricity, direct current was being promoted by Thomas Edison as the correct way of providing electrical power. Alternating current was being championed by Nicolai Tesla. Converting from the upstart AC to the correct DC was therefore referred to as rectification, and the name has stuck.

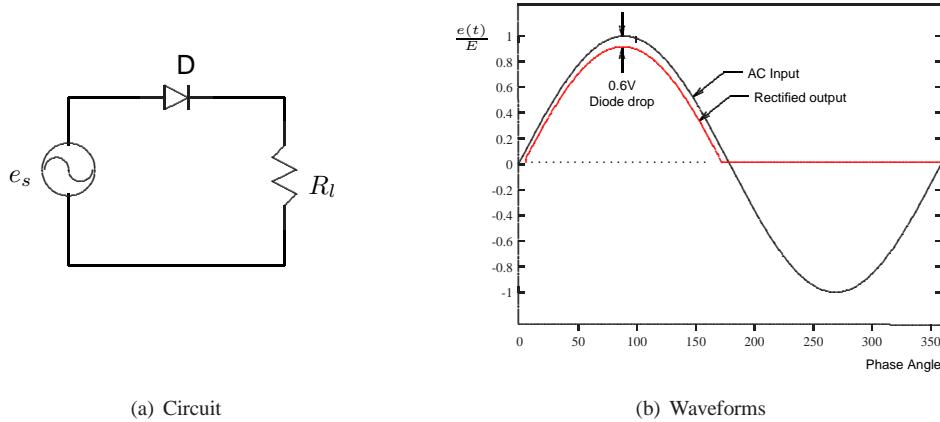


Figure 177: Half-Wave Rectifier

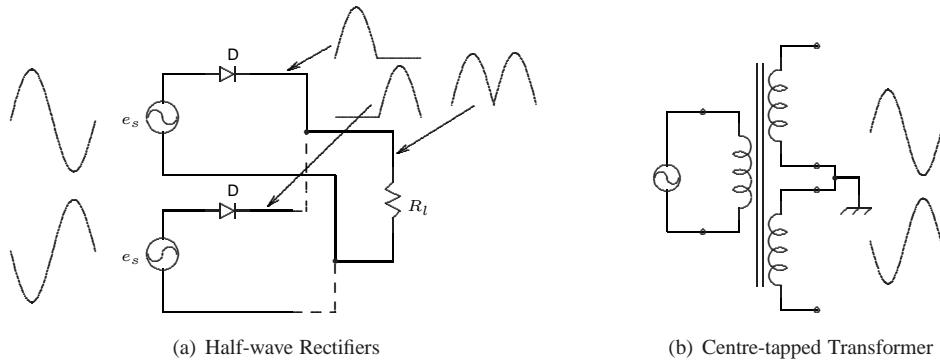


Figure 178: Full-Wave Centre-tapped Rectifier

The two input waveforms may be created by a center-tapped transformer arrangement, as shown in figure 178(b). There are two identical secondary windings on this transformer, tied together at the ground point.

Consider a snapshot in time such that the upper winding has E volts across it. Then its topmost point is at $+E$ volts with respect to ground.

The lower winding also has E volts across it. Because its uppermost terminal is connected to ground, its bottom output terminal must be at $-E$ with respect to ground.

When the upper winding voltage inverts, so does the lower winding, and so they produce antiphase sine waves as required.

A complete DC power supply, using a centre-tapped transformer, two halfwave rectifiers, and a filter capacitor, is shown in figure 179.

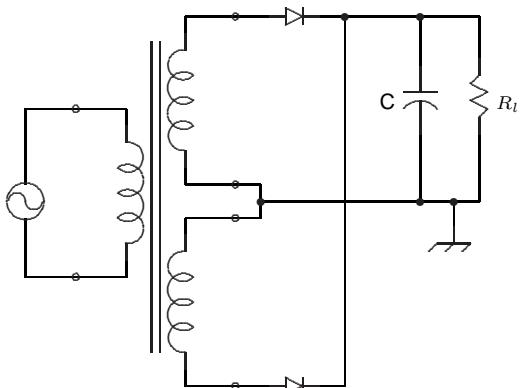


Figure 179: Full-Wave Centre-Tapped Supply

6.6 Diode Bridge Rectifier

The *bridge rectifier* circuit is shown in figure 180(a). It is essentially a steering circuit, in which current is forced to go the same direction through the load resistance regardless of the direction of the input current.

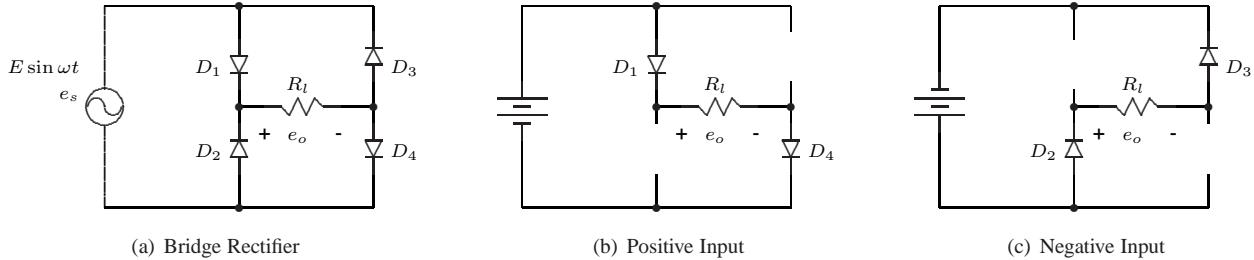


Figure 180: Bridge Rectifier

When the input voltage is positive, current flows through diodes D_1 and D_4 , figure 180(b). When the input voltage reverses, it flows through diodes D_2 and D_3 . In both cases, current flows from left to right through the load resistance.

- The current flows through two diodes as well as the load, so the load voltage is $2V_f$ volts less than the input voltage, where V_f is the diode voltage drop, 0.6 volts for a silicon diode.
- There can only be one ground point in this circuit. Typically either one side of the supply is grounded or one side of the load, but never both.

The bridge rectifier is very commonly used in a full-wave power supply, figure 181(a). The bridge generates pulsating direct current (figure 181(b)). The full wave rectifier is followed by a capacitor filter to reduce the ripple to some desired amount.

The input voltage to the power supply is generally produced by a transformer secondary winding, both ends of which are floating off ground. Then one side of the load resistance can be grounded. Notice the position of the common ground symbol in the schematic of figure 180(a).

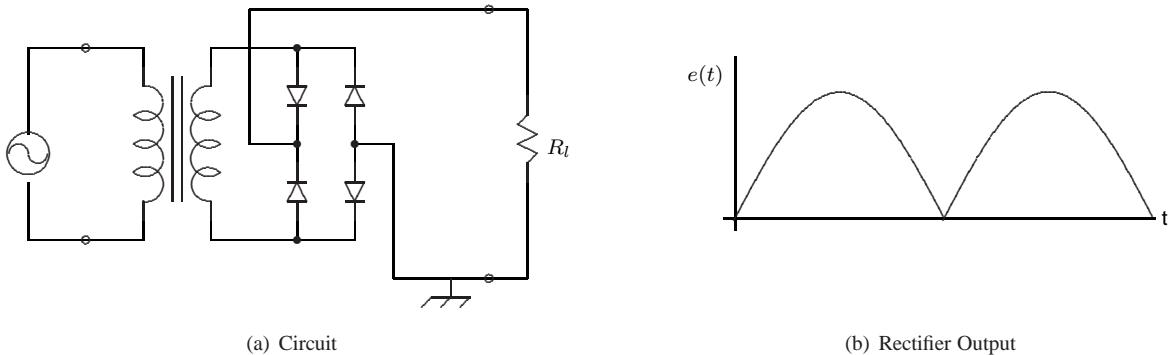


Figure 181: Bridge Rectifier Power Supply

6.7 Diode Clamp

A *clamp* circuit adjusts the offset of an AC signal so that its lowest or highest peak is set to zero volts (figure 182). The circuit is shown in figure 182(a) and the effect on a sine wave in figure 182(b).

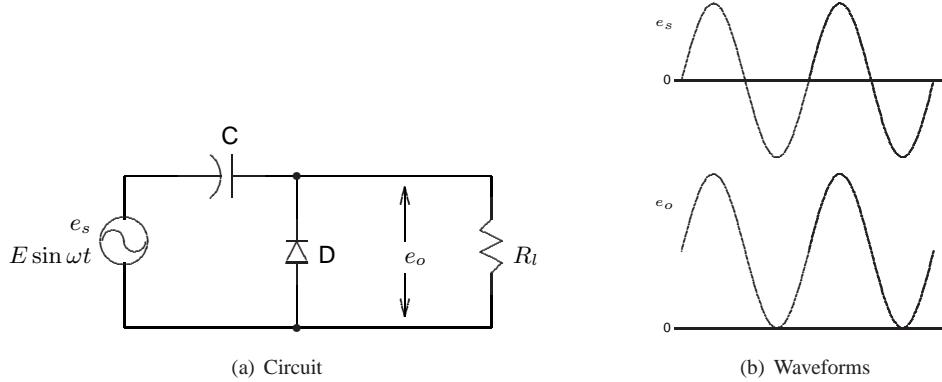


Figure 182: Clamp

The clamp circuit is also known as a *dc-restorer*, and is commonly found in video processing circuits, where it restores a waveform so that the most negative excursion is at the zero volts level. If the diode is reversed, then the output waveform is entirely negative and the most positive excursion is at the zero volts level.

How it works

The clamp circuit is a half-wave rectifier with the positions of the capacitor and diode interchanged. For simplicity, first assume R_l is large enough to be ignored and there is no voltage drop across the diode. Then the capacitor will charge up to the peak value E of the sine wave. The output waveform is equal to the sum of the two voltages E and e_s , that is, the sine wave is shifted upward by E volts. This puts its most negative excursion at zero volts.

Now consider the effect of the diode voltage drop. The forward voltage drop of the diode causes the capacitor to charge up to $E - 0.6$ volts rather than E . The vertical shift is then 0.6 volts less than the peak value of the waveform, and the waveform actually goes negative by an amount equal to the diode voltage drop, 0.6 volts.

With R_l in the circuit, the capacitor C and load resistance R_l form a highpass filter. The corner frequency of this filter must be well below the operating frequency or the sine wave will be phase-shifted and attenuated between input and output. That is

$$\begin{aligned} \omega &\gg \omega_o \\ &= \frac{1}{R_l C} \end{aligned} \quad (397)$$

Furthermore, the resistance R_l discharges between each AC cycle. For accurate restoration, the time-constant $R_l C$ must be much larger than the period of the waveform. This amounts to the same condition as expressed by equation 397.

The definitive discussion of clamping circuits is in [77], Chapter 8.

6.8 Diode Clipping Circuit

A *clip* circuit, also known as a *limiter*, restricts the maximum excursion of a signal.

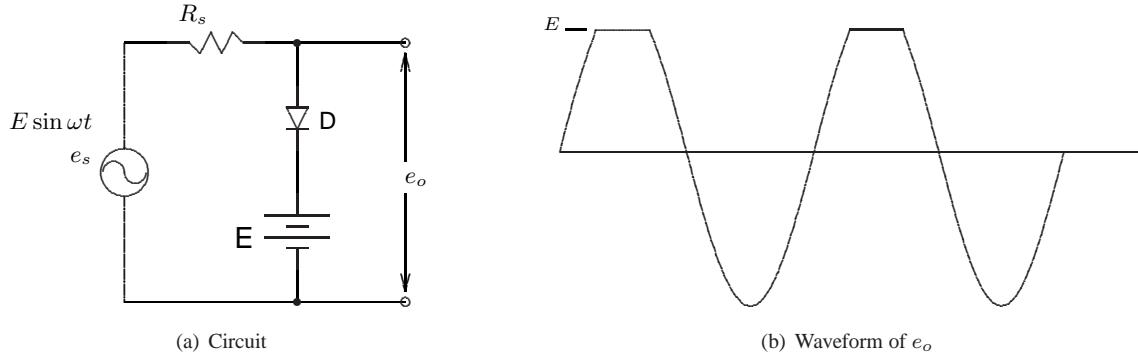


Figure 183: Clip Circuit

When the input voltage e_s is below the clip voltage E , the diode is reverse biased and may be considered an open circuit. Then the current through the resistor R_s is zero, the voltage drop across R_s is zero, and the output voltage e_o follows the input voltage e_s .

When the input voltage exceeds the clip voltage, the diode conducts and may be considered a short circuit. Then the clip voltage E is connected to the output, and the output voltage is held constant. The difference between the input voltage e_s and the output voltage E appears across the source resistance R_s . Current through R_s must flow through the clip voltage source.

The diode forward voltage V_f simply adds to the clip voltage E , so E should be reduced by that amount.

6.9 Bi-Directional Clipping

Both positive and negative peaks of the waveform may be clipped, as shown in figure 184. In this case, two diodes are connected in series to make a clip voltage E of about 1.2 volts.

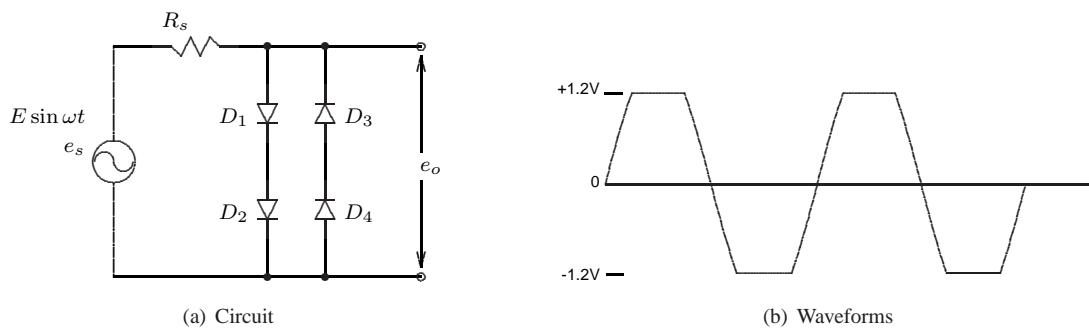


Figure 184: Bidirectional Clip Circuit

6.10 The Flyback Diode

In figure 53 (page 99), we showed how an inductor generates a large spike of voltage when the current is interrupted. That spike may cause component failure and electromagnetic noise, so there are occasions when it must be reduced or, in the jargon of the trade, *snubbed*. In the Kettering ignition system of figure 52 (page 52), this was accomplished by a capacitor across the terminals of the switch.

An alternative solution using a semiconductor diode is shown in figure 185. This scenario unfolds differently after t_2 in figure 53, when the switch is opened.

The voltage across the inductor reverses polarity and would like to become infinite, but is clamped at 0.6 volts by the flyback diode D_1 . The inductor current continues to flow through the inductor, but now it circulates around the diode-inductor loop. The 0.6 volts across the inductor cause the current to ramp downward at a rate of $\frac{0.6}{L}$ amps per second until the current reaches zero. This limits the magnitude of the spike. The circuit loop including the inductor and diode should be kept as compact as possible to minimize the generation of magnetic field noise by the transient.

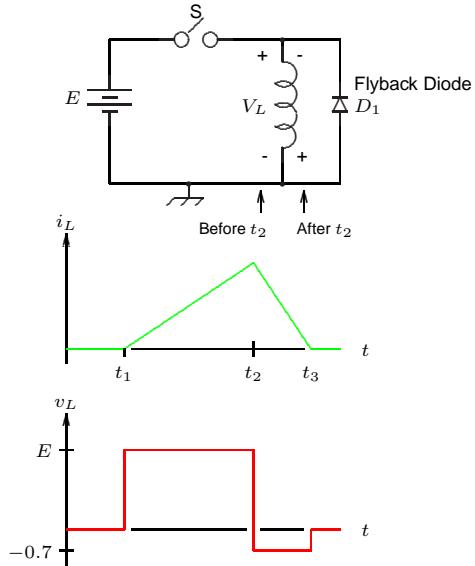


Figure 185: Inductor Flyback Diode

6.11 Diode Types

In Current Use

Diode Type	Example Part	Forward Voltage	Application
Germanium, Signal	1N34	0.2V@0.1mA	Small signal rectification with low forward voltage. Relatively high forward resistance. Not generally used unless low forward voltage is required.
Silicon, Signal	1N4148	0.6V@1mA	Small signal rectification, below 50mA. Most widely used for signal applications.
Silicon, Power	1N4001	0.8V @ 1A	Power rectifier, at and below 60Hz. General purpose, inexpensive, slow reverse recovery. Not useful in switching power supply applications.
Silicon, Power, Fast Recovery	MUR105	0.8V @ 1A	Power rectifier useful in high frequency switching power supply applications. Also available as <i>Ultra-Fast Recovery</i> .
Shottky, Power	1N5817	0.45V @ 1A	Power rectifier. No minority carriers allows rapid switching. Useful in switching power supply applications and where low forward voltage is required.

Figure 186: Diodes in Current Use

Obsolete

Diode Type	Forward Characteristic	Voltage Breakdown	Notes
Copper Oxide	0.4V @ 50mA/cm ²	6V	Essentially no forward threshold voltage made the copper oxide useful in rectifier applications in instrumentation. Poor reverse characteristic.
Selenium	1V @ 50mA/cm ²	25V	Stacked in series to increase reverse voltage breakdown. Widely used as power rectifier to replace vacuum tube rectifiers in equipment circa 1955.

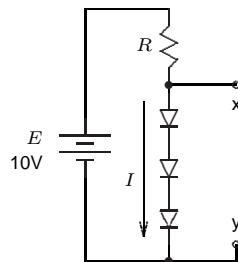
Figure 187: Obsolete Diodes

6.12 Exercises

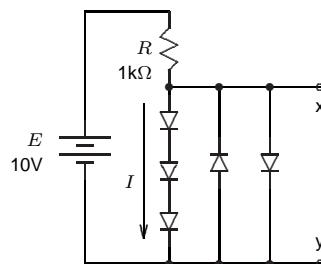
1. In all these diode problems, assume that the forward voltage drop of a conducting diode is 0.6 volts.

For the circuit shown,

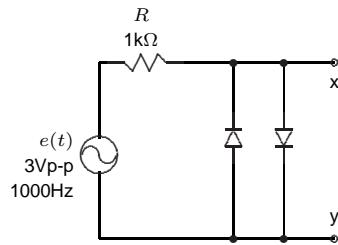
- (a) Calculate V_{xy} .
- (b) Calculate the resistance R so that the current I is 10mA.



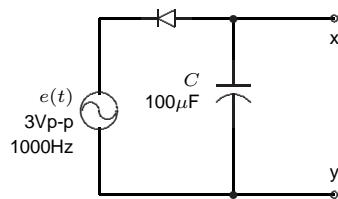
2. Calculate the current I and voltage V_{xy} for this circuit.



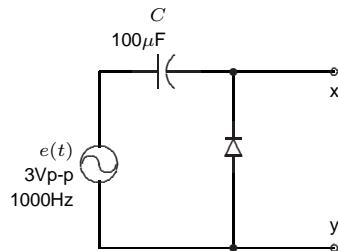
3. Sketch a scaled version of the input voltage $e(t)$ and the output voltage $V_{xy}(t)$.



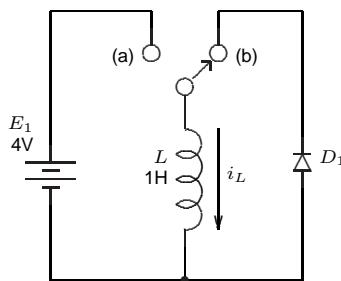
4. Sketch a scaled version of the input voltage $e(t)$ and the output voltage $V_{xy}(t)$.



5. Sketch a scaled version of the input voltage $e(t)$ and the output voltage $V_{xy}(t)$.



6. In the circuit shown below, the switch is initially in position (b). At time $t = 0$ the switch is moved to position (a), kept there for 2 seconds, and then returned to position (b).

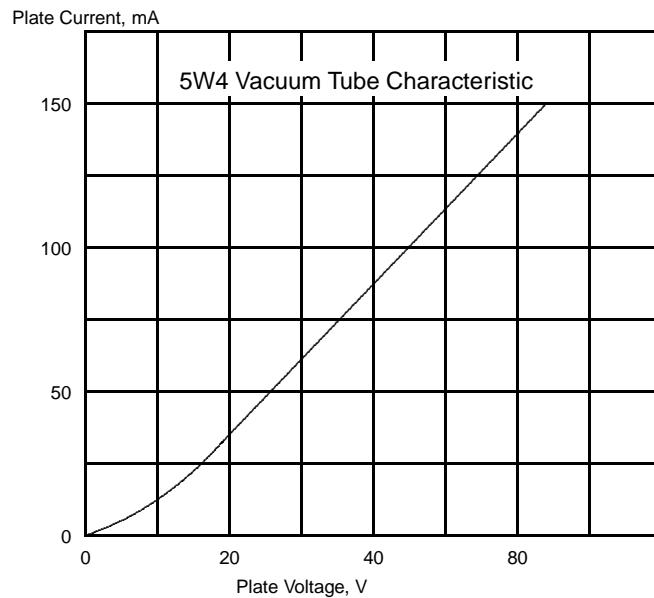


- Plot the inductor current i_L as a function of time if the diode is ideal with a zero forward voltage drop when conducting.
- Plot the inductor current i_L as a function of time if the diode has a fixed forward voltage drop of 0.6 volts when conducting.

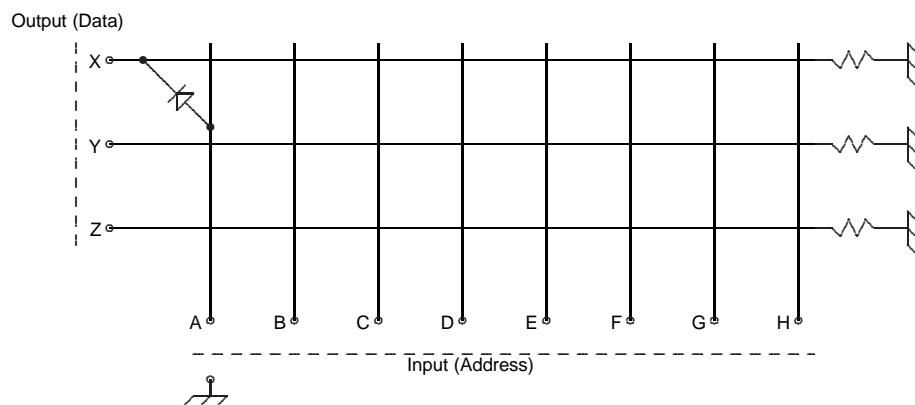
7. In the early days of radio, rectifier diodes were based on vacuum tubes. The voltage-current characteristic of a 5W4 vacuum tube rectifier diode is shown in the figure.

In what respects is a modern-day silicon rectifier diode (such as the 1N4004) an improvement over this device?

How does the threshold voltage of the 5W4 compare with a silicon diode?



8. The *read only memory* of digital technology is an interesting application of semiconductor diodes.



All the voltages are digital levels, 0V for logic zero and +5V for logic 1. A pattern of these signals on the address lines selects a particular pattern on the data lines. For this exercise, assume that the input is a logic 1 (ie, +5V) on any one of the lines A through H. The output is a 3 bit binary number on XYZ, according to the following logic table, where logic 1 is approximately +5V.

Z	Y	X	H	G	F	E	D	C	B	A
0	0	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0	1	0
0	1	0	0	0	0	0	0	1	0	0
0	1	1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0	0	0
1	0	1	0	0	1	0	0	0	0	0
1	1	0	0	1	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0

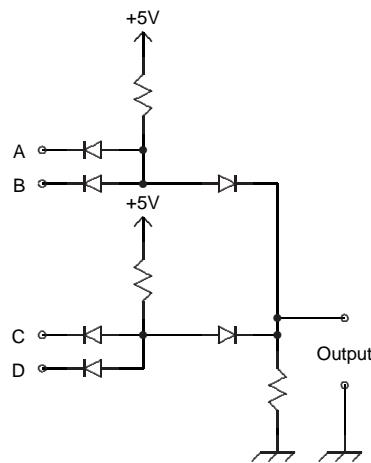
In essence, each of the three output line voltages is the logical OR of the voltages on certain input lines. By placing diodes at certain junctions in the matrix, we can create the appropriate OR gates. One diode (which may be in the wrong position) is shown as an example in the schematic diagram.

These diodes may also be regarded as preventing *sneak paths* that inadvertently connect one output line to another. The resistors ensure that the default voltage on an output line is zero volts (logic zero).

Correct the position of the diode shown on the schematic (if required) and fill in the remaining diodes.

In this particular example, the circuit is an *encoder*. A signal on one of 8 input lines creates the binary coded equivalent on the three output lines. Many other variations are possible.

9. For the logic gate shown in the schematic, you may assume that +5V corresponds to a logic 1 and zero volts to logic 0. What logic function of the input variables A,B,C and D appears at the output of this circuit?



7 The Zener Diode

7.1 The Zener Diode

A semiconductor diode has a large reverse breakdown voltage (at least 50 volts). If it is operated beyond its reverse breakdown voltage (figure 173 on page 246) it is usually destroyed. The *zener diode*, however, has been constructed so that

- it has a predictable reverse breakdown voltage V_z
- the diode can be safely operated in the reverse region at various currents as long as the total power dissipation is not exceeded.

These properties make the zener diode a very useful *constant voltage* device. For example, the 1N52xx series of zener diodes have a power dissipation of 500mW and various voltage ratings ranging from the 1N5221 at 2.4 volts through to the 1N5281 at 200 volts, in 10% increments of V_z .

In the forward direction, the zener diode operates like a conventional silicon diode, with a 0.6 volt forward voltage drop. The voltage-current characteristic is exactly the same as a silicon diode.

The symbol, equivalent circuit and VI characteristic are shown in figure 188.

The equivalent circuit indicates that the zener does not conduct in its 'forward' direction until the applied voltage exceeds 0.6 volts. It does not conduct in the reverse direction until the applied voltage exceeds V_z .

Since the region of main interest is the reverse breakdown region, the zener VI graph is flipped around so that region is in the first quadrant of the plot.

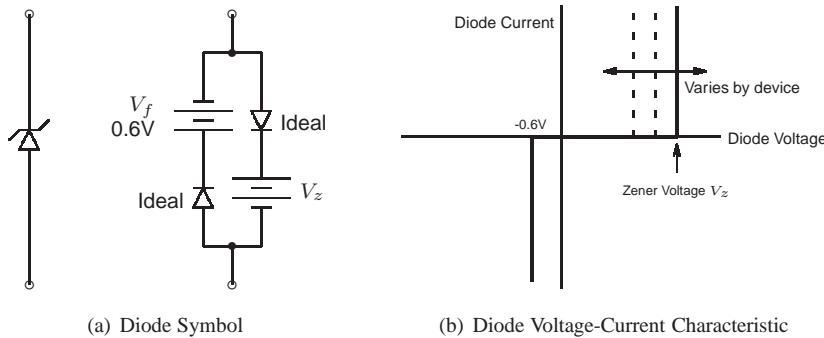


Figure 188: The Zener Diode

The zener cannot be powered by a voltage source, because it will conduct too heavily when the input voltage exceeds the device voltage. There must be a current-limiting resistance of some sort in the circuit.

7.2 Zener Regulator

The most common application of the zener diode is as a *voltage reference* device. A voltage reference is a circuit that receives an unstable voltage and produces a lower, more stable voltage. The voltage reference is used in power supply regulators and as the reference voltage for A/D or D/A conversions (see sections 26 and 25).

The zener reference circuit is shown in figure 189(a). The output voltage must supply some sort of load resistance, but we'll ignore that for the moment. The input is characterised as a fixed voltage supply E to which is added some time-varying supply ΔE .

The zener may be represented as a fixed voltage source V_z in series with an ideal diode¹⁰⁵, as shown in figure 189(b).

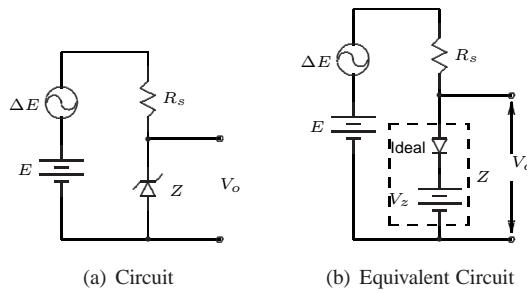


Figure 189: Zener Diode Reference Circuit

Consider the equivalent circuit of figure 189(b). If the input voltage $E + \Delta E$ is less than the zener voltage V_z , then the ideal diode will be reverse biased and can be treated as an open circuit. No current will flow through R_s and the voltage drop across it will be zero. Then the output voltage V_o will be equal to the input voltage.

If the input voltage is greater than the zener voltage V_z , then the ideal diode will be forward biased and can be treated as a short circuit. Then the output voltage will be V_z and the voltage drop between the input voltage $E + \Delta E$ and V_z will appear across the series resistor R_s .

The time-varying input voltage is generating a fixed output voltage equal to the zener voltage V_z .

The Effect of a Load Resistance

When a load resistance R_l is attached to this circuit, some current will flow through R_l . For a fixed input voltage, adding a load resistance has the effect of stealing current from the zener.

Within certain limitations, the value of load resistance (and its load current) can vary without affecting the output voltage. As more or less current flows through the load, less or more current flows through the zener diode. However, the zener voltage is constant irrespective of its current, so the output voltage stays constant.

However, if the load current is made too large, the zener current will drop to zero and the zener will *drop out of regulation*, ie, cease to conduct. The output voltage will then drop below V_z .

The zener regulator can be analysed by redrawing it as shown in figure 190 and then Thevenizing the voltage source E , series resistance R_s and load resistance R_l into an equivalent voltage source E_{oc} in series with an internal resistance R_{int} . If E_{oc} is greater than V_z , then the zener will be in conduction and the output voltage will be regulated.

Design Conditions

A zener diode reference circuit should be checked for the following conditions:

Dropout For minimum input voltage E_s and minimum load resistance R_l , the zener current is at minimum. Check that value of E_{oc} exceeds the zener voltage V_z .

Over Dissipation For the maximum value of E_{oc} and the largest value of R_l , the current in the zener is at maximum. Check that the power dissipated in the zener $P_z = V_z I_z$ does not exceed the power rating of the zener.

¹⁰⁵In this context, an *ideal* diode is one that conducts perfectly in its forward direction, with no forward voltage drop. The ideal diode is an open circuit in the reverse direction.

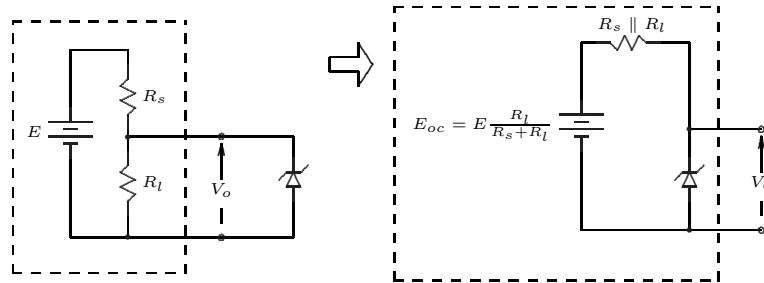


Figure 190: Zener Regulator with Load

7.3 Zener Impedance

Figure 188 on page 259 shows the zener regulator VI characteristic as a vertical straight line. In fact, this is not exactly correct. The VI characteristic has a slightly less than vertical slope, figure 191(a), representing a small *incremental resistance* in series with the zener voltage, figure 191(b).

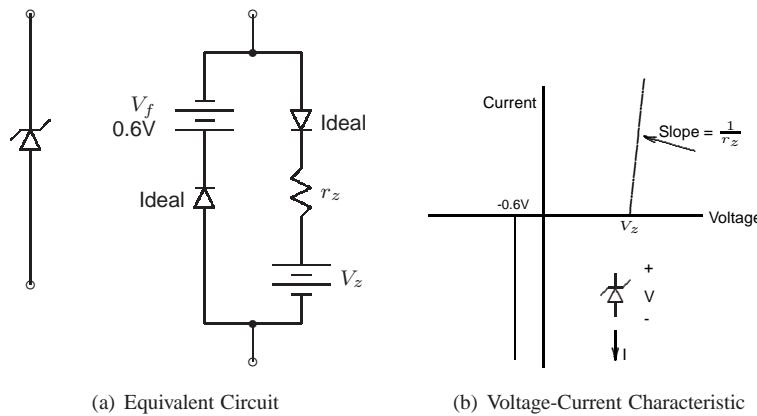


Figure 191: The Zener Incremental Resistance

The incremental resistance r_z varies from hundreds of ohms to a few ohms, depending on the zener voltage. Zeners in the 6 to 9 volt range have the lowest incremental resistance, under 10 ohms.

For most stable regulation, a zener diode should have low incremental resistance and the smallest possible temperature coefficient of zener voltage. This often turns out to be a 6.2 volt zener, so this is a common choice of device for a reference circuit.

For example, suppose the zener is a 1N5995 device, which is a 6.2 volt zener with 10Ω zener incremental resistance r_z . It is operated from a 12 volt input voltage E_s and a source resistance R_s of $1k\Omega$.

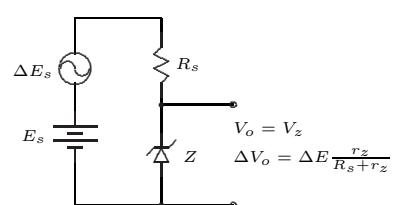


Figure 192: The Zener Incremental Resistance

For each volt that the supply voltage changes, the output voltage will change by

$$\Delta V_o = \Delta E_s \frac{r_z}{R_s + r_z} \quad (398)$$

$$= 1 \times \frac{10}{1000 + 10} \quad (399)$$

$$= 10 \text{ mV}$$

Similarly, if there is a load resistance and it changes, then this will change the zener current and as a result of the zener resistance the output voltage will change slightly.

Now 10 millivolts is No Big Deal in many situations, but if this circuit is being used as the reference for an A/D converter where one step is a millivolt, then a 10 millivolt drift might be unacceptable.

If this change in the output voltage is unacceptably large, then the circuit has to be modified to improve it. From equation 400, the output ΔV_o can be reduced by increasing the series resistance R_s . But this may be unacceptable because it has the effect of starving the zener of current.

The alternative is to operate the zener from a constant current source. This will supply the needed zener current but has a high incremental impedance, so the effective value of R_s is very large. Consequently, a precision voltage reference will consist of a zener diode driven by a constant current source.

7.4 Zener Regulator Amplifier

In a simple regulator circuit, the output of the zener regulator drives the load resistance directly. If instead the output of the zener regulator is used as the input to an amplifier, and the amplifier drives the load, there are two advantages to be obtained:

- The amplifier can be designed to supply much larger load currents than would be possible with a simple zener regulator circuit, and
- Variations in load current have less effect on the zener current, so the regulated output voltage is more constant.

The circuit of figure 193 uses an emitter follower (section 29.27, page 870) as a current amplifier for the basic zener reference circuit¹⁰⁶. The zener voltage must be chosen to take into account the 0.6 volt V_{be} drop of the emitter follower amplifier.

The circuit has an attractive simplicity and is often used. However, it does have some limitations:

- It's not a particularly stable circuit, because the base-emitter voltage of the transistor will change by $2.2\text{mV}/^\circ\text{C}$ change in temperature. However, for non-critical applications, this may be acceptable.
- The addition of the amplifier does not reduce any output variation due to changes in the input voltage. The voltage divider effect between R_s and r_z is unchanged. Any improvements in this regard must be obtained by some other circuitry.

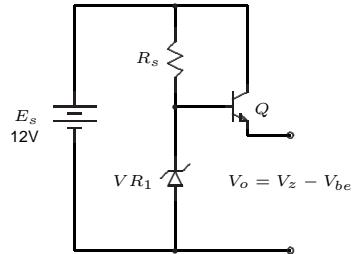


Figure 193: Zener with Current Amplifier

¹⁰⁶This section and following assume some familiarity with the BJT (junction transistor) so you may wish to visit that material first, starting with section 8.1, page 266.

- The circuit has no short-circuit protection. If the load resistance somehow becomes shorted, then the transistor will overheat and destroy itself trying to drive infinite current into the short, unless the transistor is massively oversized and mounted on a sufficiently large heat sink.

For all these reasons, new designs tend to use integrated circuit three-terminal regulators in place of the zener-emitter-follower circuit. The IC regulator overcomes all of these disadvantages, albeit at slightly higher cost.

7.5 Zener Current Drive

We pointed out in section 7.3 that the zener regulation is improved, with respect to variations in input voltage, if the source resistance is replaced by a constant-current source.

An example of a current-driven zener regulator circuit is shown in figure 194.

This is a two-stage regulator circuit. The first zener regulator is VR_1 , which establishes a fixed voltage of 3.3 volts across R_2 . Then R_2 and Q_2 function as a current source for the second zener diode, VR_2 , which produces the reference 6.2 volts.

The transistor Q_1 is diode connected to balance the base-emitter voltage of Q_2 . When the temperature changes, both base-emitter voltages will change in the same direction (by an amount equal to $-2.2\text{mV}/^\circ\text{C}$), and the effect is cancelled.

Because of the finite zener incremental resistance of VR_1 , the voltage across it will vary slightly as the input voltage ΔE_s changes. This will cause a slight change in the current out of the current source, and through the second zener diode. This change in zener current is further reflected in a change in the output voltage. Assuming that the transistors are high current gain, the output change in voltage is about 0.2mV per volt of input change, a substantial improvement over the simple zener regulator of figure 192.

Notice that constant current drive does not improve the performance of the regulator with respect to any variations in output current. To take best advantage of the constant current drive, the output load current should be as constant as possible (or very small, by driving an amplifier that drives the load.)

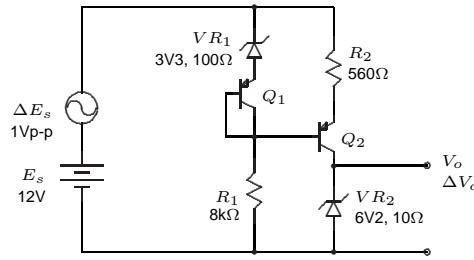


Figure 194: Current-Driven Zener Regulator

7.6 The Integrated Circuit Shunt Regulator (IC Zener)

It is very important to understand the operation of the zener diode, and there are analog circuits where it is a useful component. The *integrated circuit shunt regulator* is modern replacement for the zener diode, with some significant advantages. Figure 195(a) shows the symbol. The IC Zener has a third terminal, used for adjustment of the output voltage, as shown in figure 195(b).

The following table shows a comparison of a small signal zener diode and a typical integrated circuit shunt regulator.

	Zener	IC Shunt Regulator
Example	1N5221A Series	TLV431
Voltage range	2.4 to 200V	1.24 to 6.0V
Initial tolerance	$\pm 10\%$	$\pm 1.5\%$
Adjustment	Various fixed values	Adjustable
Current	5 mA typ	$80\mu A$
Incremental Resistance	7 to 2500Ω	0.25Ω
Tempco	-0.085 to +0.110 %/ $^{\circ}C$	0.006%/ $^{\circ}C$
Price	\$0.07 to \$3.00	\$0.44

Zener diodes are available in a wide variety of voltage ratings. The shunt regulator is usually regarded as a 1.25 volt device, although it can be trimmed in value by an external potentiometer. Notice the difference in initial tolerance of the reference voltage: $\pm 10\%$ for the zener vs $\pm 1.5\%$ for the IC zener¹⁰⁷.

The zener diode and shunt regulator are both *constant voltage* devices. Ideally the terminal voltage should not change with current, ie, the internal incremental resistance should be zero. The internal resistance of the shunt regulator is *much* lower than that of a zener diode. Consequently, a variation in device current creates a smaller change in voltage across the shunt regulator and it is a closer approximation of the ideal.

Both the zener and shunt regulator are suitable for use as a *voltage reference* for an analog-digital converter or power supply. Ideally, the reference should maintain a constant voltage regardless of changing temperature. Again, the shunt regulator is more stable with temperature by an order of magnitude.

The shunt regulator can operate from a much smaller current: 5mA vs 0.08mA. That would be critically important in battery-operated equipment.

A shunt regulator is more expensive than a zener diode, but where accuracy and stability are important it is a better choice.

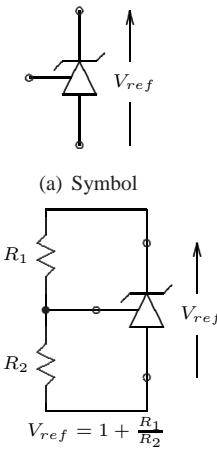
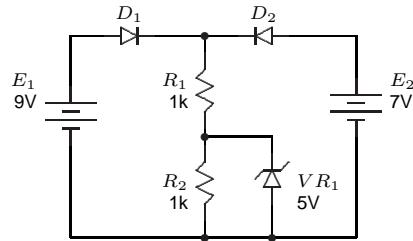


Figure 195: The IC Zener

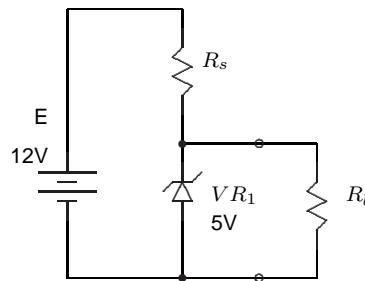
¹⁰⁷These figures are for the least expensive version: tighter tolerances are available for both devices – at a higher price, of course.

7.7 Exercises

1. For the circuit shown, the diodes have a forward voltage drop of 0.6 volts when conducting.



- (a) What is the current through resistor R_2 ?
 (b) What is the current through zener diode VR_1 ? Justify your answer.
 2. The zener regulator circuit shown below is to generate a 5 volt output from a 12 volt input. The load current through R_l varies from 0 to 100mA. The minimum zener current is 10mA.



- (a) Calculate a suitable resistance value and power rating for the series dropping resistor R_s .
 (b) Calculate a power rating for the zener diode.
 (c) If the incremental zener resistance is 8Ω , what is the change in output voltage when the load current changes from 0 to 100mA?
 (d) Assuming the same incremental zener resistance, what is the change in output voltage if the input voltage changes by 1 volt?
 3. (a) Determine the DC voltages and currents in the circuit of figure 194. Assume that the transistor current gain is very large (ie, ignore base currents).
 (b) Calculate the output ripple ΔV_o for the circuit.

8 Active Devices

Solid state electronics relies on various *active devices* to provide amplification and switching functions.

An active device is a semiconductor device which functions as a controlled source. The three controlled source devices we will study (in order of their invention) are tabulated in figure 196.

Name	Acronym	Function	Type
Bipolar Junction Transistor	BJT	Current-controlled current source	CCCS
Junction Field Effect Transistor	JFET	Voltage-controlled current source	VCCS
Metal Oxide Semiconductor Field Effect Transistor	MOSFET	Voltage-controlled current source	VCCS

Figure 196: Controlled Sources

CCCS means *Current controlled current source*, **VCCS** means *Voltage controlled current source*.

Subsequent sections examine each of these devices in detail. However, it's important first to understand the basics of the operation of each of these devices and how they compare with each other. This information is sufficient to design non-critical applications.

Power Gain

These active devices are useful because they provide *power gain* to an electronic signal. In effect, a small power input signal actuates a controlled source, which then modulates power from a supply. The net result is a signal that has the desired characteristics, but at a higher power level than the input.

This says that the product of voltage and current at the output will be larger than the product of voltage and current at the input. So for example the output voltage could be smaller than the input, and if the output current is much larger, the product of voltage and current at the output is larger than at the input, and the device has power gain.

8.1 The BJT: Bipolar Junction Transistor

The properties of the *bipolar junction transistor* (BJT) are summarized in figure 197.

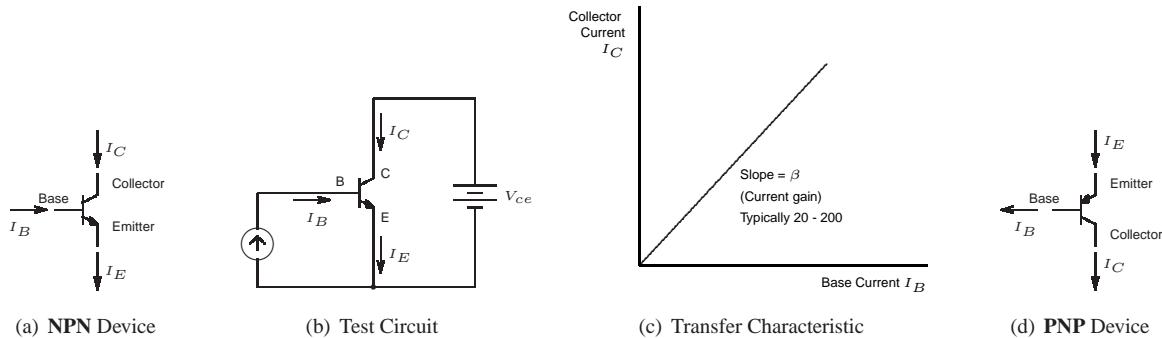


Figure 197: Bipolar Junction Transistor

Providing that the collector-emitter voltage is greater than the saturation voltage $V_{ce_{sat}}$, the collector current is equal to a constant, the current gain β , times the base current. So a small base current controls a much larger collector current, and the BJT can function as an amplifier.

There are two current loops: the base loop, I_B , and the collector loop I_C . The two currents join and flow out of the emitter and so the emitter current is

$$I_E = I_B + I_C \quad (400)$$

Base-Emitter Voltage

The base-emitter junction appears as a diode, where the diode current is I_E . The current-voltage relationship for the base-emitter junction is given by the diode equation 401.

$$I_f = I_s e^{V_f / N V_t} \quad (401)$$

where

I_s is the *saturation current*, typically equal to 10^{-12} to 10^{-15} amps.

V_t is the *thermal voltage* approximately equal to 0.026 volts at room temperature.

N is a constant between 1 and 2 that depends on the construction of the transistor. For simplicity, we'll generally assume than N is equal to unity and leave it out.

As shown in figure 197(c), the relationship between base current and collector current is linear, so in that respect it is a current-controlled-current-source.

It's also possible to regard the BJT as a voltage controlled-current device, because there is a small voltage across the base-emitter junction when there is current into the base. As is the case for any diode, the voltage is a very non-linear function of the current. Consequently, the BJT is best regarded as a current-controlled-current-source¹⁰⁸

Saturation in the BJT

In the test circuit of figure 197(b), consider that the collector-emitter voltage V_{ce} is gradually reduced. This is observed to have no effect on the collector current, which is solely¹⁰⁹ controlled by the base current and β .

However, as V_{ce} drops below about 0.1 volts (for a small signal transistor), the collector current I_C abruptly ramps downward until it reaches zero as V_{ce} reaches zero (figure 198). Below this transition value of V_{ce} , the transistor is said to be *saturated* and the transition voltage is known as the *saturation voltage* V_{cesat} . Notice that V_{cesat} is significantly less than V_{be} .

Putting it another way, the transfer characteristic of figure 197(c) requires that the transistor not be saturated and $V_{ce} > V_{cesat}$

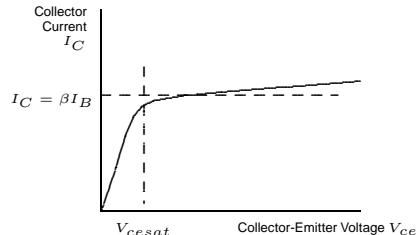


Figure 198: BJT Saturation

Complementary BJT

The NPN transistor has a polarity complement in the PNP transistor, figure 197(d). The behaviour is similar to the NPN device, but the polarity of the voltages and currents are reversed. (It's possible to buy *complementary*

¹⁰⁸This is an entry-level view of the BJT, and is suitable for getting started. In fact, the current gain varies to some extent with the operating current, and so the relationship between the input and output current is not perfectly linear. Furthermore, the relationship between input current and output current is not predictable a-priori. It depends on how the devices is constructed. On the other hand, the relationship between input voltage and output current is highly non-linear, but it is also much more predictable than current gain. Then, so the argument goes, the BJT is better regarded as a voltage-controlled-current device. For a first-time user of the BJT, it is best regarded as a current-controlled-current-source. The details can come later.

¹⁰⁹Well, not quite solely. There is a small change in current caused by the *Early Effect*, which is documented elsewhere (see page 29.11).

pairs of NPN and PNP transistors that have very similar characteristics.) It is often useful to use a combination of NPN and PNP devices in a circuit design.

The junction transistor is a very versatile component. It is available in a variety of designs, low and high power, low and high frequency operation.

8.2 The JFET: Junction Field Effect Transistor

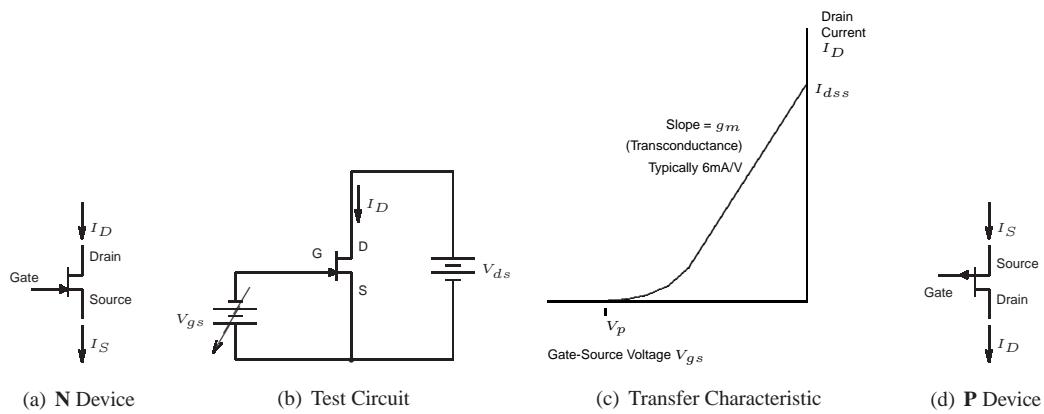


Figure 199: Junction Field Effect Transistor

The *JFET* transistor consists of a diode junction, the *gate*, embedded into a semiconductor channel that connects the *drain* and *source* terminals. The gate diode is normally reverse biased, so there is no gate current. Consequently, the drain current is always equal to the source current.

At zero gate-source voltage, the current in the channel is at its maximum value. This current is called I_{dss} , the drain saturation current, and is typically between 5 and 20mA (figure 199(c)).

In the test circuit of figure 199(b), as the gate is made more negative with respect to the source terminal the channel narrows down, reducing the drain current. The relationship is non-linear, but over a large part of the range the departure from non-linearity is minimal. The transfer relationship between drain current and gate-source voltage is characterised by the *transconductance* g_m , measured in μs , the reciprocal of the Ω .

Eventually, when the gate-source voltage is made sufficiently negative at V_p , the *pinch-off* voltage, the drain current is completely throttled off.

Because the JFET conducts with no gate-source voltage and is progressively turned off as the gate-source voltage is made more negative, it is described as a *depletion mode* device: increasing (negative) control voltage depletes the drain-source channel of carriers, reducing the drain current ¹¹⁰.

The equation relating drain current and gate-source voltage is

$$I_D = I_{dss} \left(1 - \frac{V_{gs}}{V_p} \right)^2 \quad (402)$$

where

- I_D is the drain current
- I_{dss} is the *drain saturation* current
- V_{gs} is the gate-source voltage
- V_p is the jfet *pinch-off* voltage

¹¹⁰The power MOSFET, described in the next section, is an *enhancement-mode* device, that is, increasing gate-source voltage causes the drain current to increase.

Like the BJT, the N-Channel JFET has a complementary P channel version, figure 199(d).

The JFET is a small signal device – there are no power JFET devices. The inter-terminal capacitances are small, so it is well suited to high frequency amplifier applications. Because the input circuit is a reverse-biased diode, the current required from the source is almost nil. As we will see, the output of the JFET can be a relatively low resistance source, and so the JFET makes a very effective high-to-low impedance converter.

The JFET also has some application as an analog switch (although CMOS switches are now preferred) and as a variable resistance device.

Saturation in the JFET

The JFET behaves in a similar fashion to the BJT. When the drain-source voltage is above the JFET saturation voltage, the drain current is independent of drain-source voltage V_{ds} . In the JFET however, the saturation voltage is given by

$$V_{ds_{sat}} = V_p + V_{gs} \quad (403)$$

where

$V_{ds_{sat}}$ is the saturation voltage

V_p is the *pinch-off* voltage, a property of the JFET

V_{gs} is the gate-source voltage (negative)

In effect, the saturation voltage increases as the gate-source voltage is reduced and the drain current increases, as shown in figure 200.

For example, suppose the pinch-off voltage V_p for a certain JFET is 6 volts. Then for a gate-source voltage V_{gs} of -2 volts the saturation voltage $V_{ds_{sat}}$ is 4 volts. That is, V_{ds} must be greater than 4 volts to ensure that the JFET operates as a controlled current source.

More details and applications for the JFET are in section 31 (page 976).

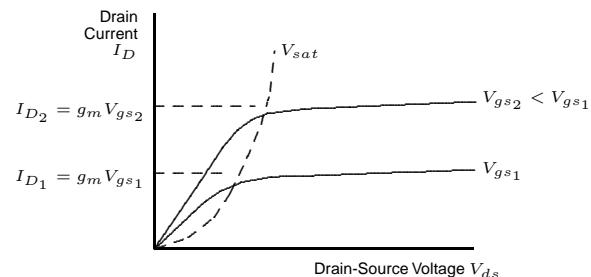


Figure 200: JFET Saturation

8.3 The MOSFET: Metal Oxide Semiconductor Field Effect Transistor

The MOSFET is pervasive in electronics. Millions are used in integrated circuit chips. The discrete power version has revolutionized the design of the switching power supply, a component of every portable computing device and desktop PC.

In this note, I will focus on the power MOSFET. A more general introduction to MOSFETs and their application in IC design is in [78] and [79].

First, the matter of the symbol. A commonly used symbol for the MOSFET is shown in figure 201. This is the one found in the commercial databooks for power MOSFETs. I've used the simpler version shown in figure 202(a). This is easier to draw and indicates the polarity of the device in a similar manner to the BJT.

The main characteristics of the MOSFET are shown in figure 202.

In both figure 202(a) and 201 the symbol indicates an important aspect of the MOSFET. The gate terminal is completely insulated from the conducting channel. As a result, the input current is effectively zero and a high-resistance source can turn a MOSFET on and off¹¹¹.

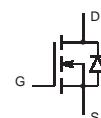


Figure 201: Alternative MOSFET Symbol

¹¹¹The gate-source capacitance of a power MOSFET is quite large – hundreds of picofarads – and the input signal must charge up this capacitance to change the gate voltage. Consequently, significant transient current drive is required to switch the device at high speed.

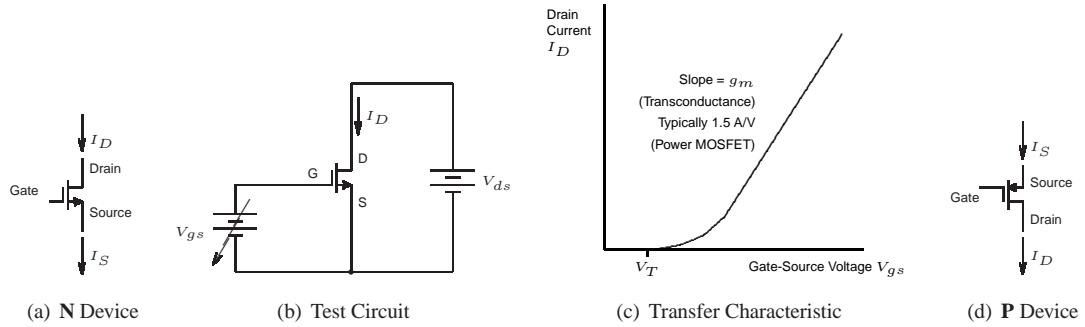


Figure 202: MOSFET

Notice also in figure 201 that there is an intrinsic diode built into the structure of the MOSFET. This can be useful as a free-wheeling (snubbing) diode when the MOSFET is driving an inductive load. The insulating layer between the gate and channel is extremely thin – in the order of 20 to 100 nanometers [78]. The maximum value of gate-source voltage is typically in the order of 20 volts. Careful measures may be required to prevent static electricity from destroying the device¹¹².

The power MOSFET is an *enhancement-mode* device. That is, without a gate-source voltage, it conducts no current. A positive gate-source voltage is required to initiate conduction. Notice the polarity of the gate-source voltage supply V_{gs} in figure 202(b).

This behaviour is also reflected in the transfer characteristic of figure 202(c). The gate-source voltage must reach a positive threshold V_T before any significant drain current flows. As the gate-source voltage increase beyond that, the drain current increases in a parabolic function:

$$I_D = k(V_{gs} - V_T)^2 \quad (404)$$

where

- I_D is the drain current
- k is a constant that depends on the construction of the device
- V_{gs} is the gate-source voltage (positive)
- V_T is the *threshold* voltage, a property of the device

The slope of the transfer characteristic is the transconductance g_m , same as the JFET. However, the transconductance is orders of magnitude larger in the power MOSFET.

Most applications of the power MOSFET are in switching mode, where the device is either completely off or conducting heavily. As a result, it's usually only necessary to know that the input gate voltage will cause sufficient drain current.

Saturation in the MOSFET

As in the case of the JFET, the saturation voltage of the MOSFET depends on the gate-source voltage:

$$V_{ds_{sat}} = V_{gs} - V_T \quad (405)$$

where $V_{ds_{sat}}$ is the drain-source saturation voltage. Thus, the saturation voltage increases with V_{gs} in much the same fashion as the JFET, figure 200.

As indicated in figure 202(c), there is a complementary version of the MOSFET, the p-channel device. In this case, the device is turned on by making the gate negative with respect to the source.

¹¹²Some devices have internal zener diodes that limit the gate voltage to protect the device.

8.4 Application Circuit

A simple motor drive circuit is shown in figure 203. The motor is assumed to be a small DC motor that operates from 12VDC maximum. It requires 100mA under no load, increasing to 300mA under full load.

The basic idea is to use potentiometer R_2 to control the speed of motor M . It would be nice to be able to use a small, low cost potentiometer for R_2 , but such a pot cannot carry the necessary current to operate the motor directly. However, a power MOSFET is inexpensive and can be used to provide the necessary control current for the motor.

The potentiometer supplies a control voltage to the gate terminal of the MOSFET, which is an open circuit. Consequently, we can expect the pot to behave properly as a voltage divider since it has effectively no load resistance. As the gate voltage increases, the current through the MOSFET increases, and the motor will run faster.

The motor windings are inductive and as the motor runs, the commutator¹¹³ briefly open-circuits these windings, creating pulses of noise voltage at the terminals of the motor. The diode D provides a path for any inductive pulses that appear as positive spikes at the lower terminal of the motor.

A typical power MOSFET conducts heavily when its gate voltage is about 6 volts, figure 202(c). R_1 drops the supply voltage so that 6 volts appears at the top end of the potentiometer R_2 . Then full rotation of the pot takes the motor from zero to full drive current.

This circuit does have some limitations:

- The MOSFET transistor operates in its active region, that is, there is voltage across it and current through it. Consequently, it will dissipate considerable power, and a substantial heatsink will be required to remove heat from the transistor (section 28).
- As figure 202(c) shows, the relationship between gate voltage and drain current is nonlinear, so the relationship between potentiometer shaft rotation and motor speed will also be non-linear. (Since the drain current is zero up to the gate threshold voltage, it would help to put a resistance in series with the lower end of R_2 so that the bottom end of the pot is at V_t volts.)
- The output of the MOSFET is a current drive, so the motor is *current-controlled*. The current through a DC motor controls its torque. For any given motor current the torque output will be constant and the motor speed will depend on the mechanical load. This may or may not be acceptable. If it's not acceptable, then one could add a negative feedback system of some sort that senses motor speed, compares it with a setpoint, and adjusts the motor current to keep the motor speed constant.

It is more common in motor speed control to use the MOSFET as a switch. The duty cycle (percentage ON time of the switch) determines the average current through the motor. This method dissipates far less power in the MOSFET¹¹⁴.

References

A useful overview of power MOSFET technology is in [80].

¹¹³The *commutator* is a rotary switch built into the motor. As the shaft of the motor rotates, the commutator switches the connections to the rotor winding, so that the motor continues to turn beyond one-half rotation.

¹¹⁴As a side effect, when the switch is off the voltage of the motor is proportional to its speed, which can be used in a negative feedback system for speed control.

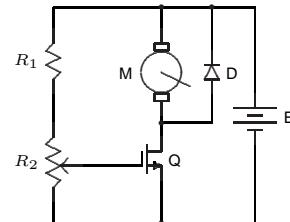


Figure 203: MOSFET Motor Drive Circuit

8.5 The MOSFET Switch

The BJT, JFET and MOSFET devices described in previous sections, can each function as a *switch* device.

When the device operates in switching mode, it is either fully on or off, so the design analysis is relatively simple. The MOSFET is a good example, because it is simple to use and is often operated in switching mode.

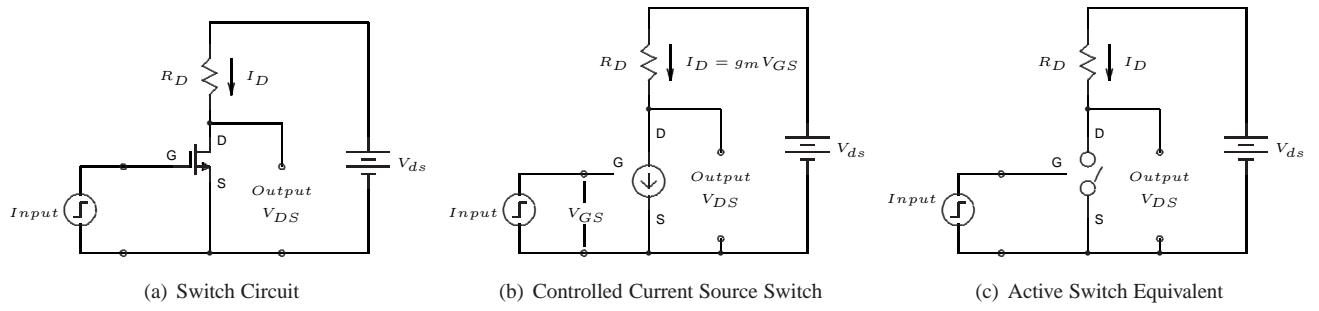


Figure 204: MOSFET Switch

The Controlled Current Source

The switching circuit, with an N MOSFET device, is shown in figure 204(a). The equivalent circuit for the MOSFET is a controlled current source, shown substituted into the circuit in figure 204(b).

The input voltage is V_{GS} , and since this is a switching circuit, the output has two possible states: V_{DD} and zero, or *high* and *low*.

- When the input voltage is low, the drain current I_D is zero, there is no voltage drop across the drain resistance R_D , and so by KVL the entire supply voltage V_{DD} appears across the MOSFET drain-source terminals. In switching parlance, we say that the output is high.
- When the input voltage is high, the MOSFET conducts heavily, there is a substantial voltage drop across R_D , and the output voltage is low.

Now we'll do this in a bit more detail. In a MOSFET, the drain current I_D is given by

$$I_D = g_m V_{GS} \quad (406)$$

where g_m is the transconductance.

The drain-source voltage V_{DS} is given by KVL:

$$V_{DS} = V_{DD} - V_{RD} \quad (407)$$

where V_{RD} is the voltage drop across the drain resistance R_D .

The voltage across the drain resistance is simply:

$$V_{RD} = I_D R_D \quad (408)$$

Combining these three equations, we can get the *transfer function* for the switch:

$$V_{DS} = V_{DD} - g_m R_D V_{GS} \quad (409)$$

The transfer characteristic is shown in figure 205, where the circuit input and output voltages are plotted against each other. There are several things to notice about the transfer characteristic:

- As the input voltage increases the output voltage decreases, so the device is an *inverter*.
- The two points of operation on the graph are at *cutoff* and *saturation* of the switch.
- At cutoff, the drain current is zero and the output voltage is equal to the drain supply voltage V_{DD}
- The *active region* occurs between saturation and cutoff, when the drain current and drain-source voltage are at some intermediate value between their maximum and minimum.
- At saturation, the drain current is limited by some external current component – in this case, the drain resistance and power supply voltage. The output voltage is zero.

In order to ensure that the device switches fully into the saturated state, referring to equation 409 the designer must choose R_D so that V_{DS} goes to zero for the minimum value of the MOSFET transconductance g_m and input voltage V_{GS} .

Switch Model

On the simplest level, the MOSFET may be regarded as a switch, figure 204(c).

Opening and closing of the switch is controlled by the input voltage, the gate-source voltage V_{GS} . When the input is high, the switch is closed and the output is low, and vice versa. So the circuit is an inverter.

It's also a power amplifier. There is no input current, because the gate circuit is effectively open, so there is no input power. There *is* output power, because a load can be attached to the output terminals and will receive both current and voltage. Consequently, the MOSFET switch has very large power gain.

However, because there is input capacitance between the gate and source terminals, the driving circuit must supply a pulse of current to raise the gate voltage. This represents a consumption of power, and so the power dissipation is dependent on how frequently the circuit switches from one state to another.

Switch Example

A simple example of a switching circuit using the N MOSFET is shown in figure 206. When the switch is in the upper position, Q_1 is ON, so its drain voltage is LOW. Then Q_2 is OFF and the LED is extinguished. When the switch is moved to the lower position, this turns Q_1 OFF, Q_2 ON, and the LED is illuminated.

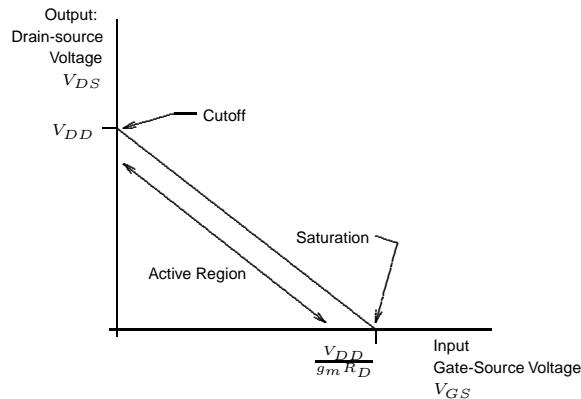


Figure 205: Transfer Characteristic

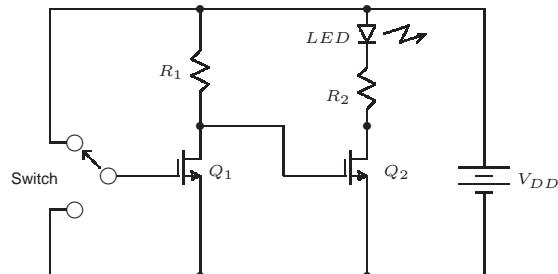


Figure 206: MOSFET Switch Example

8.6 The BJT Switch

The MOSFET switch described in section 8.5 is a voltage controlled current source. The BJT is a current-controlled current source, so the operation is slightly different.

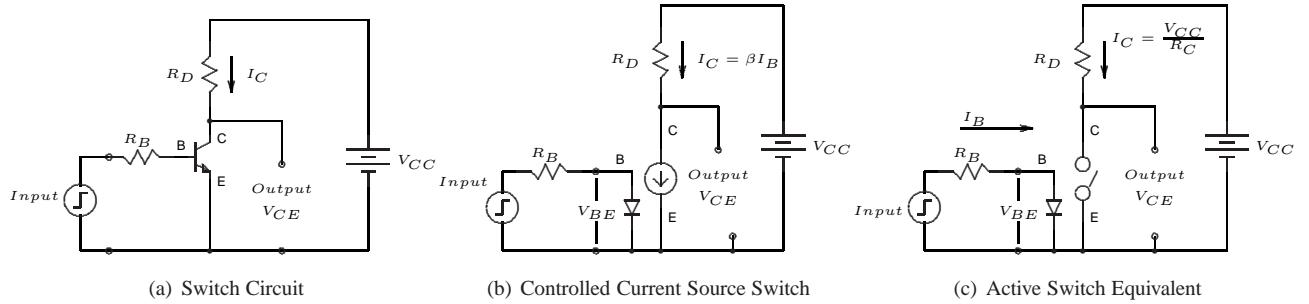


Figure 207: NPN BJT Switch

The BJT switching circuit is shown in figure 207(a).

The switch circuit is usually driven by a voltage source, in which case a resistor R_B is required in the base circuit. The base-emitter junction is a diode (figure 207(b)), and when it is conducting, the base-emitter voltage is limited to 0.6 volts. The input voltage must be greater than this to drive current around the input circuit loop, and there must be a resistor to absorb the difference between the input voltage and the base-emitter voltage.

The equivalent circuit for the BJT is a controlled current source, shown substituted into the circuit in figure 207(b).

Suppose that the input voltage is E_I , and since this is a switching circuit, it has two possible states: V_{CC} and zero, or *high* and *low*.

- When the input voltage is low (below 0.6 volts) the base current is zero, the collector current I_C is zero, there is no voltage drop across the collector resistance R_C , and so by KVL the entire supply voltage V_{CC} appears across the BJT collector-emitter terminals. In switching parlance, we say that the output is high.
- When the input voltage is high, a base current $I_B = (E_I - 0.6)/R_B$ flows, a substantial collector current flows, there is a substantial voltage drop across the collector resistance and the output voltage is low. The minimum voltage is known as the *saturation voltage* of the BJT, and is typically about 0.1 volts for small signal silicon transistors. (Notice that the saturation voltage is significantly less than the base-emitter voltage of a transistor when it is conducting. This turns out to be useful, as shown in the application below. We'll often assume that the saturation voltage is zero.)

Now we'll do this in a bit more detail.

In a BJT, the base current is given by

$$I_B = \frac{E_I - 0.6}{R_B} \quad (410)$$

providing that E_I is greater than 0.6 volts. (Otherwise the base current is zero.)

The collector current is β times this, or

$$I_C = \beta I_B \quad (411)$$

where β is the *current gain* of the transistor.

The collector-emitter voltage V_{CE} is given by KVL:

$$V_{CE} = V_{CC} - V_{RC} \quad (412)$$

where V_{RC} is the voltage drop across the collector resistance R_C .

The voltage across the collector resistance is simply:

$$V_{RC} = I_C R_C \quad (413)$$

Equations 410 through 413 can be used to design a transistor switch, as shown in the following example.

Example

A transistor switch is to be used in a 5 volt system, that is, the switching signal is 0 to 5 volts and the collector supply voltage V_{CC} is 5 volts. The collector resistor is $1k\Omega$ and the transistor current gain β is minimum 20.

Choose a suitable value for R_B .

Solution

When the input signal is +5 volts, the transistor must saturate (ie, be fully ON), that is, V_{CE} is approximately zero.

Then from equation 412, the voltage across R_C is V_{CC} : 5 volts.

The current through R_C is then (equation 413)

$$\begin{aligned} I_C &= \frac{V_{RC}}{R_C} \\ &= \frac{5V}{1k\Omega} \\ &= 5 \text{ mA} \end{aligned}$$

The current into the base terminal is β times smaller than this, according to equation 411.

$$\begin{aligned} I_B &= \frac{I_C}{\beta} \\ &= \frac{5 \text{ mA}}{20} \\ &= 0.25 \text{ mA} \end{aligned}$$

This is the current that must flow into the base terminal to drive the transistor into saturation. The value of R_B corresponding to this current is given by equation 410:

$$\begin{aligned} R_B &= \frac{E_I - 0.6}{I_B} \\ &= \frac{5 - 0.6}{0.25 \text{ mA}} \\ &= 17.2 \text{ k}\Omega \end{aligned}$$

In a practical design, you would choose R_B to be the next smaller standard value from this.

BJT Switch Transfer Characteristic

The transfer characteristic between E_I (input) and V_{CE} (output) is shown in figure 208. As in the case of the MOSFET switch transfer characteristic in section 8.6:

- As the input voltage increases the output voltage decreases, so the device is an *inverter*.
- The two points of operation on the graph are at *cutoff* and *saturation* of the switch.
- At cutoff, the collector current is zero and the output voltage is equal to the supply voltage V_{CC}
- At saturation, the collector current is limited by some external current component – in this case, the collector resistance and power supply voltage. The output voltage is zero.
- Using equations 410 through 413, (and making the assumption that E_I is much larger than 0.6 volts) we can determine that saturation occurs for an input voltage

$$E_I = \frac{V_{CC}}{R_C} \frac{R_B}{\beta} \quad (414)$$

This is not a particularly important equation in itself – it's better to be able to use equations 410 through 413 to derive it – but it shows how the various quantities in the BJT switch circuit interact at the saturation point. For example, if R_C is increased, the saturation point will move to the left, and a smaller value of input voltage E_I will saturate the transistor.

- The *active region* occurs between saturation and cutoff, when the drain current and drain-source voltage are at some intermediate value between their maximum and minimum.

Key Concepts

- When the BJT is used as a switch, it has two states: ON or OFF.
- In the OFF state, the collector current is zero.
- In the ON state, the collector current is at its maximum (saturation) value $I_{C_{sat}}$, which is limited by the external power supply and collector resistance.

$$I_{C_{sat}} = \frac{V_{CC}}{R_C} \quad (415)$$

- In switching applications, the base current is usually larger than necessary to ensure that the transistor saturates in the ON state. That is,

$$I_B > \frac{I_{C_{sat}}}{\beta} \quad (416)$$

- When the transistor is saturated, its base-emitter voltage is about 0.6 volts and its collector-emitter voltage is about 0.1 volts.

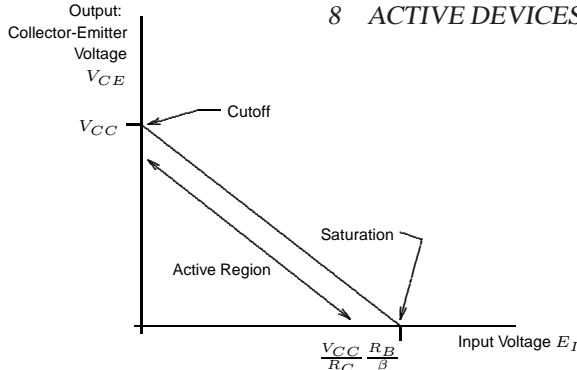


Figure 208: Transfer Characteristic

8.7 BJT Logic Devices

The BJT may be used to construct various logic circuits as shown in figure 209. These logic gates were first used in the era when computers were constructed with discrete component logic gates and the designs are now essentially obsolete. However, they continue to be sporadically useful in circuit design.

RTL Inverter

Figure 209(a) shows a BJT transistor inverter. This form of logic is known as *resistor-transistor logic*, or RTL. The circuit relies on the fact that the BJT saturation voltage (0.1 volts) is less than the threshold base-emitter voltage (0.6 volts). Consequently, it doesn't take much noise to trigger the inverter incorrectly, and the *noise immunity* of this circuit is poor. Nonetheless, it continues to be useful in simple applications.

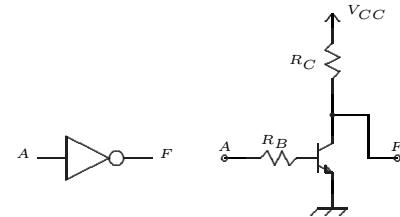
RTL NOR

Figure 209(b) shows an RTL circuit which is a NOR gate for positive logic. If input A OR input B is high, then the output will be low, which is the NOR gate function. Again, this circuit suffers from poor noise immunity. Since any logic circuit can be constructed entirely with NOR gates (or NAND gates), this logic circuit would in theory be sufficient to build an entire computer.

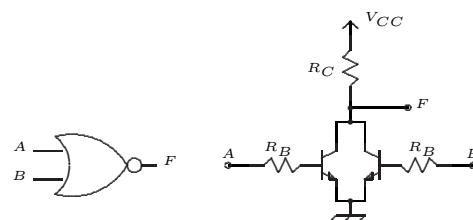
DTL NOR

Figure 209(c) shows a *diode-transistor* or DTL NOR logic circuit. Essentially, this is the diode OR gate of section 6.2 followed by an inverter. If either input A or B is high, then the transistor will be ON and the output will be low, the NOR gate function.

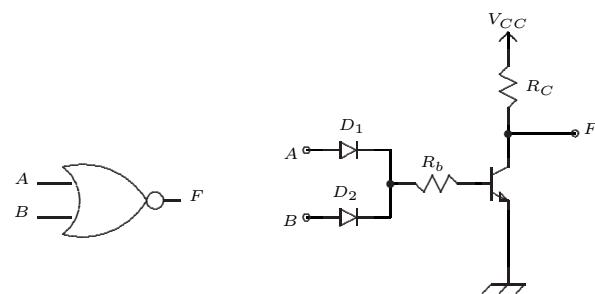
This circuit has better noise immunity than the RTL NOR gate by virtue of the voltage drop across the OR gate diodes. This requires that an input signal be at least 1.4 volts (the drop across the diode plus the V_{BE} of the transistor) to register as a logic high signal. Additional diodes are sometimes added in series with the transistor base to further increase the logic high threshold value¹¹⁵.



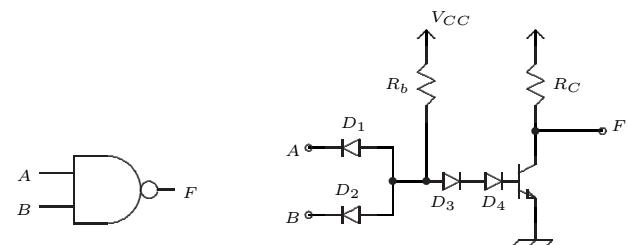
(a) RTL Inverter



(b) RTL NOR



(c) DTL NOR

(d) DTL NAND
Figure 209: BJT Logic Gates

DTL NAND

Figure 209(d) is a similar logic circuit but with the diodes reversed and R_B returned to the positive supply. Then both inputs must be high in order for the transistor to be turned ON. (Putting it another way, if either input is low,

¹¹⁵As a practical matter, this circuit would need a *pull-down* resistor between the diode cathodes and ground to ensure the transistor stays off when the diodes are reverse-biased. The DTL NAND is a much more reliable circuit.

this holds the transistor in the OFF state). This is the NAND function.

The diodes D_3 and D_4 are absolutely necessary to provide a 1.4 volt noise immunity for the gate. Without them, the noise immunity would be a disastrous 0 volts.

8.8 The BJT Flip Flop

An entertaining and useful BJT switch application is shown in figure 210. The *flip-flop* is a device that has two stable states, like a toggle light switch. In this case, either of the LEDs will illuminate, but not both. If the switch opposite the illuminated side is pushed, the flip-flop changes state and the other LED illuminates.

The collector resistors R_1 and R_2 limit the current through the LEDs and the transistors. They would be chosen for a suitable LED current, about 10mA. With an LED voltage of about 2 volts and a supply voltage of 5 volts, then R_1 and R_2 would be about 300Ω .

The base resistors R_3 and R_4 limit the current into the bases of their respective transistors, and must be much larger than R_1 and R_2 for the circuit to work. Assuming a minimum β value of 50, then these resistors could be about $15k\Omega$ each.

Now let's see how this works.

- Consider that side 1 is illuminated.
- Then Q_1 is ON (saturated) and its collector-emitter voltage is about 0.1 volt.
- This voltage appears at the base of Q_2 , and since it's less than 0.6 volts, Q_2 is OFF.
- Current flows from the power supply, down through LED_2 , through R_2 , through R_3 and into the base of Q_1 , keeping it in the ON state.
- There is current through the LED, but it's kept to a small value by the large value of R_3 , so that the LED_2 remains dark. (This is why the base resistors must be large.)

Now consider that pushbutton S_2 is pushed to change the state of the flip flop.

- This short-circuits the collector of Q_2 to ground, and makes the base-emitter voltage of Q_1 equal to zero.
- Q_1 now shuts off and its collector voltage rises.
- Current now flows down through LED_1 , through R_1 , through R_4 and into the base of Q_2 .
- When pushbutton S_2 is released, this base current holds it in the ON state, and LED_2 remains illuminated.

Switches S_1 and S_2 can be replaced with transistors. The inputs of these transistors can then be activated with other signals to cause the flip-flop to change state.

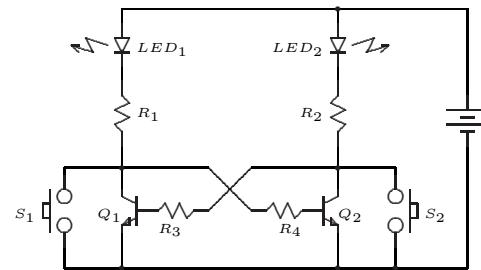


Figure 210: BJT Flip Flop

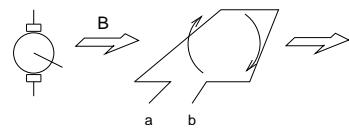
8.9 Exercises

1. With reference to figure 210 on page 278, determine the component values for the flip-flop as follows:
 - (a) LED_1 and LED_2 are each to illuminate with a current of 5mA. The power supply voltage V_{CC} is 5 volts and the voltage drop across a red conducting LED is 2.1 volts (section 27.2 on page 773).
 - (b) Assume that transistors Q_1 and Q_2 are small signal silicon NPN types with a conducting base-emitter voltage V_{BE} of 0.6V, a collector-emitter saturation voltage V_{cesat} of 0.1V and a minimum current gain β of 40. Calculate the value of the base drive resistors R_3, R_4 .
 - (c) A skeptic looks at the circuit of figure 210 and says *That won't work! When Q_1 and LED_1 are conducting, LED_2 is supposed to be OFF. But there is base current flowing through LED_2 into the base of Q_1 , which will cause LED_2 to illuminate.*
Is the skeptic correct? Explain.
 - (d) Examine the data sheet for the 2N4401 transistor. Does it meet the requirements for Q_1 and Q_2 of the flip-flop circuit?
2. The MOSFET version of the flip-flop is even simpler than its BJT counterpart. Draw it.

9 Motors and Generators

9.1 DC Motor and Generator

The symbol and basic construction of an electro-mechanical *DC generator* or *DC motor* is shown in figure 211. It turns out that these are the same device, which will function equally well as a generator or motor¹¹⁶



We can get at the behaviour of this device with two equations and a certain amount of inference.

Figure 211: Generator

1. First, let's consider the device as a generator. According to Lenz's Law, the voltage generated by a changing magnetic field is:

$$e_{ab} = N \frac{d\phi}{dt} \quad (417)$$

where N is the number of turns of wire immersed in the magnetic field and ϕ is the magnetic flux in webers. So there will be a voltage generated at terminals $a : b$, the magnitude proportional to the rate of change of the flux that the conductors are moving through.

The wire is moved through the magnetic field at a speed proportional to the rotational speed of the generator. Consequently, it's not too big a stretch to infer that *the voltage at the terminals of the generator is proportional to the rotational speed of the generator*.

2. Now let's consider it as a motor. The force on a conductor in a magnetic field is given by

$$F = BlI \quad (418)$$

where B is the magnetic flux density in webers/m², l is the length of the conductor in metres, and I is the current flowing through the conductor. This force is exerted at right angles to the conductor, so it tends to produce a *torque* in a rotational device. Consequently, we can imagine that *the torque is proportional to the current flowing through the motor*.

A motor is a device for transforming electrical energy into mechanical energy. A generator does the opposite: transforms mechanical energy into electrical energy. But in both cases, the torque is proportional to the device current.

It's possible to do the analysis in greater detail, but for our purposes this is sufficient: speed is proportional to terminal voltage and torque is proportional to current.

¹¹⁶The device shown here has a magnetic field that we assume is created by a permanent magnet. It is equally possible to use an electromagnet to create this field.

9.1.1 The Ideal DC Motor

For an ideal motor (one with no armature resistance or frictional losses), that's all there is to it: the speed of a motor may be controlled by varying its terminal voltage, the torque may be controlled by varying its current.

If an ideal motor is connected to a voltage source, it will run at constant speed. When there is no mechanical load, the motor current will be zero. As the mechanical load increases, so does the motor current.

The speed-torque characteristic for an such ideal motor is shown in figure 212.

The equations governing the operation of an ideal DC motor:

First,

$$E_m = K_m \omega \quad (419)$$

where

- E_m is the motor *back emf*, volts
- K_m is the motor *speed constant*, volts per rpm
- ω is the motor speed, rpm

Equation 419 says that the speed of the motor is directly proportional to the applied motor voltage.
Second,

$$I_m = K_t \tau \quad (420)$$

where

- I_m is the motor current, amperes
- K_t is the motor *torque constant*, amps per newton-metre
- τ is the motor output torque, newton-metres

Equation 420 says that if an ideal motor has no mechanical load, it requires no current to spin. In practice, of course, there *are* frictional losses, so the motor current is non-zero even when there is no mechanical load. As well, a motor inevitably has some armature resistance. The effect of armature resistance is to cause the motor armature voltage to drop when the armature current increases as the result of a mechanical load. When the armature voltage drops, the motor slows down.

An ideal motor will not slow down when a mechanical load is applied. Obviously, this is not the case, so the concept of an ideal motor has some limitations. However, the concept is still useful to provide some insight, especially if the frictional losses and armature resistance are small enough to be neglected.

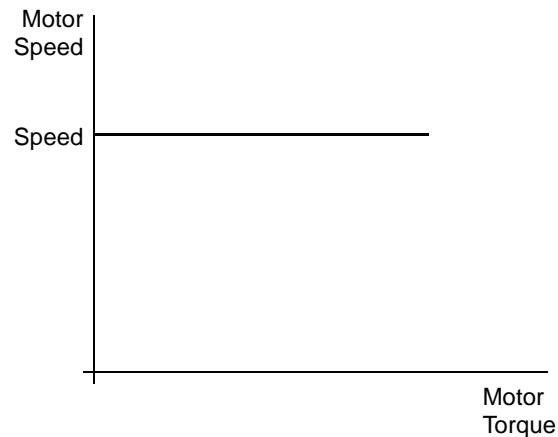


Figure 212: Ideal Motor, Speed-Torque Characteristic

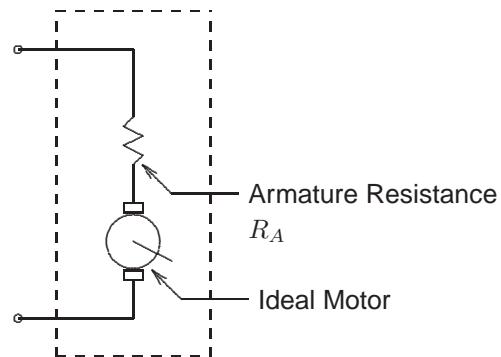


Figure 213: Real Motor, Equivalent Circuit

Real DC Motor Behaviour

There is always some finite resistance in the armature winding of a DC motor. A real DC motor consists of an ideal motor (in which the speed is exactly proportional to the voltage across it) in series with an armature resistance R_A .

As a result, when the motor is loaded and the motor current must increase to output mechanical torque, the voltage across the armature resistance increases, leaving less voltage for the armature itself, and the speed decreases. This corresponds to our experience of a DC motor: increasing the mechanical load causes the motor current to increase and the speed to decrease. The speed-torque characteristic is shown in figure 214.

Under no-load conditions, the motor runs at its *no-load speed*. The torque required to bring the motor to a complete stop is the *stall torque*. These two points define the end-points of the speed-torque characteristic.

An application of figure 214 is shown in figure 215. The speed-torque characteristic of a mechanical load – such as a fan in this case – is plotted on the motor characteristic. The point of intersection is the actual operating point of the motor-fan combination. This technique is particularly useful when either the motor or fan characteristic is non linear.

9.1.2 Real DC Generator Behaviour

An ideal DC generator behaves very simply as a pure voltage source where the output voltage is proportional to rotational speed. A real DC generator has some internal resistance, so the output voltage will decline as the output current increases.

9.1.3 Speed Control

In an ideal world, a motor may be made to run at a constant, predictable speed – regardless of load – if it is connected to a voltage source and the motor resistance is somehow removed or cancelled. As a mechanical load is applied, the motor current will increase. The ideal voltage source will supply whatever current is required to keep its terminal voltage constant – and to keep the motor running at the same speed.

It's not difficult to construct a voltage source that is very close to ideal - such as a regulated power supply. However a real motor has an internal resistance that causes it to slow down when producing a mechanical load torque (and therefore drawing current from the supply). Is there any way this can be overcome? It sounds glib to say, but the answer is to put a *negative resistance* in series with the positive resistance of the motor, so that they cancel out and the effective motor resistance is zero. Such a negative resistance is not readily available off the shelf, but it can be created with op-amp circuitry, as shown in section 37.2 on page 1107. In practice, it turns out to be difficult to make the negative resistance *exactly* equal to the motor resistance. The system runs away if

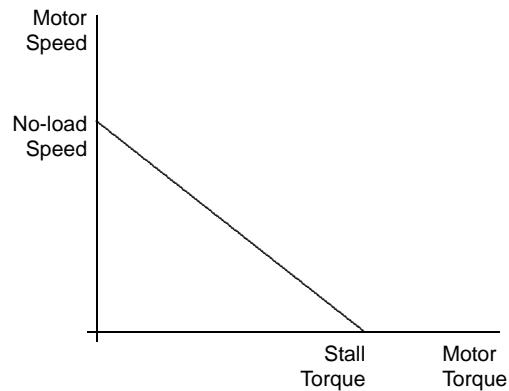


Figure 214: Real Motor, Speed-Torque Characteristic

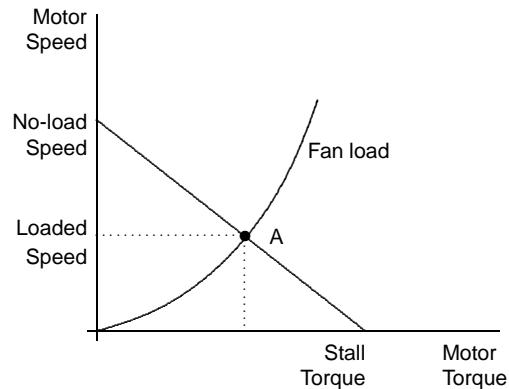


Figure 215: Real Motor with Fan Load

the negative resistance is larger than the motor resistance so, in practice, it's made slightly smaller. The voltage control of the speed is then very good but not perfect.

There is one possible downside to the *ideal voltage/ideal motor* speed control technique. If the motor runs up against a stop, it will try to produce infinite output torque, and that will cause mechanical damage.

9.1.4 Torque Control

DC motors are sometimes required to produce constant output torque, regardless of speed. Just as the speed of an ideal motor is adjusted by its terminal voltage, the torque of an ideal motor is adjusted by its current. Since the motor resistance and the ideal motor are in series, it is quite practical to control the motor torque by controlling the current. (The voltage produced across the motor must be within the *compliance limit* of the current supply).

The ideal combination is an *ideal voltage/ideal motor* to control speed, with a maximum output torque established by a current limit in the power supply. The current limit is set to produce a torque that is above any torque to be encountered in normal operation, so the motor will run at the desired speed regardless of small variations in load. However, if there is a mechanical jam, the torque is limited by the available current to some level below destruction of the machine.

9.1.5 Alternatives

These techniques are successful in practice, but there are some additional considerations.

- If the motor is changed (which is something that is always possible in a production situation, when there is a switch to a different vendor), there may be a change in winding resistance and that will affect the operation of the negative resistance compensation.
- Applying a variable DC voltage to control DC motor speed is inherently inefficient, so it may not be practical in a battery operated unit – such as the focus motor in a digital camera. It may be preferable to operate the motor from an electronic switch, such that the average voltage across the motor is determined by the relative ON time and OFF time.
- It is common to employ *limit switches* to disable a DC motor before the moving part reaches its mechanical limit. Mechanical switches can fail, so more robust alternatives such as Hall Effect switches are often used.

References

The textbooks by Davies (references [81] and [82]) contain much useful information on the selection and application of small electro-mechanical devices. Reference [82] shows several methods of measuring the torque and speed of small DC motors.

10 Negative Feedback

Block Diagrams

Many of the diagrams in this section will be *block diagrams*. Each individual block represents a subsystem, and often corresponds to some component in the original device. A complete block diagram describes the behaviour of a system.

The elements of a block diagram are shown in figure 216.

- A *signal* is indicated by an arrow. The arrow indicates the direction of flow on the diagram, that is, where the signal is applied.
- A *block* contains the value of the transfer function, that is, the ratio of output to input. Laplace transform notation is common, so an integrator would be indicated as a transfer function of $1/s$, for example.
- Signals may be combined in various ways. This is indicated by a circle containing the *operator* that combines the signals.

An example of a complete block diagram is shown in figure 217(b) below.

The Concept of Negative Feedback

The concept of *negative feedback* is illustrated in figure 217.

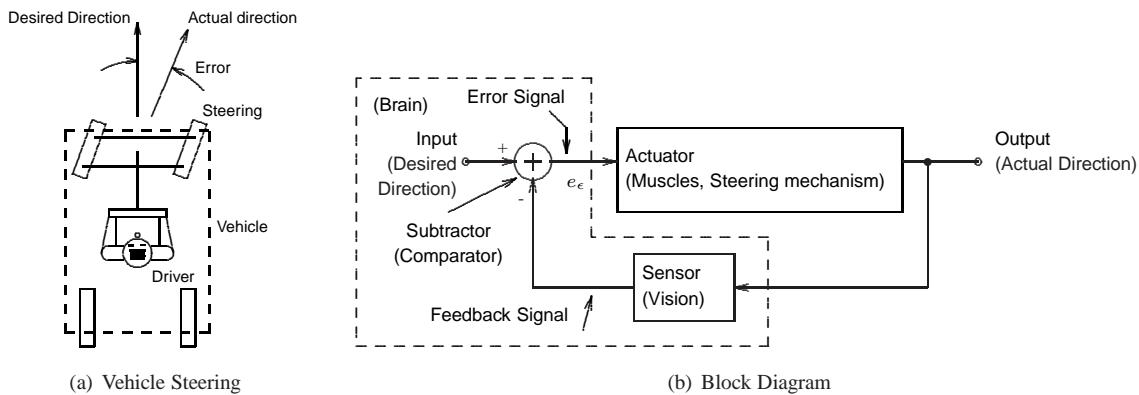


Figure 216: Block Diagram Elements

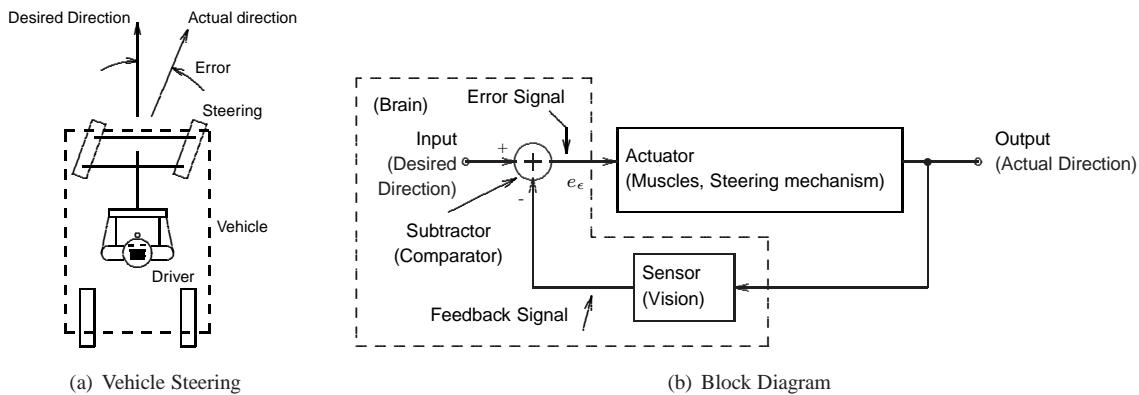


Figure 217: Driver and Vehicle

Figure 217(a) shows a schematic top view of a vehicle. The difference between the actual and desired vehicle direction is the steering *error*. The operator uses the error signal to actuate the arm muscles, which then (via the steering mechanism) aim the front wheels of the vehicle in the direction to reduce this error.

This system is drawn as a block diagram in figure 217(b). A negative feedback system can be implemented in a wide variety of ways, but it invariably includes the elements shown in this figure.

Input This signal establishes the desired output of the system. In process control applications, it's known as the *setpoint*.

Subtractor The subtractor finds the difference between the input and feedback signals, which then becomes the *error* signal for the system. The subtractor is often referred to as the system *comparator*. (This comparator is a completely different device than the comparator referred to in section 12.2.)

Actuator This block provides the muscle for the system. In a process control application, this block is referred to as the *plant*, since it may correspond to some large industrial process such as an oil refinery. In electronics systems, it's often an amplifier. The transfer function of this block is referred to as the *forward gain*.

Sensor This system detects the output and generates a suitable feedback signal for the system. Its transfer function (feedback signal divided by the output signal) is called the *sensor gain*.

Some observations on the steering system:

- Because there is a complete actuator-sensor loop this type of system is referred to as *closed-loop*. If a gust of wind blows the vehicle in the wrong direction, the driver will correct with a different error signal input to the actuator until the vehicle is back on track and the error is back to zero. A closed-loop system can react to changing circumstances, and it corrects for disturbances.
- If for some reason, the driver's vision becomes obscured (from mud on the windscreens, for example), then the vision system is no longer functional, the system has no feedback signal, and it becomes *open-loop*. As long as the road continues in the same direction and there are no disturbances, then the vehicle will continue in the correct direction. However, a disturbance will drive it off course. An open-loop system cannot correct for disturbances.
- Now consider that the vehicle has power steering and this malfunctions so that more effort is required to move the steering wheel. As long as the driver is capable of moving the steering, the system will continue to function and the vehicle will be steered correctly. Consequently, a change in the strength (gain) of the actuator (within certain limits) will not dramatically affect the operation of the system.
- On the other hand, if something affects the system sensor so that the vision of the driver is incorrect, then the direction of the vehicle will be affected. The sensor and subtractor (comparator) are critical in determining direction of the vehicle.
- If the driver is intoxicated, the effect is to introduce a *time delay* into this system. A time delay in the forward or feedback path makes a negative feedback system unstable, and the vehicle will oscillate (weave) around the desired path.

This example illustrates the strengths and shortcomings of a negative feedback system. On the one hand, it corrects for disturbances and is somewhat immune to the strength of the forward gain. On the other hand, if the sensor is compromised the system fails badly, and under certain conditions it can be made to oscillate. As Longfellow put it in his poem *There Was a Little Girl*:

*When she was good,
She was very good indeed,
But when she was bad she was horrid.*

This description applies to negative feedback.

10.1 Negative Feedback and Amplifiers

Negative feedback is easiest to understand in the context of an example, the requirement to construct an amplifier with predictable gain. It is not difficult to make a high gain amplifier. It is much more difficult to make a *stable* amplifier. The active components (BJTs, JFETs and MOSFETs) that provide the gain function tend to drift with changes in power supply voltage, temperature or the passage of time, changing the behaviour of the amplifier.

A negative feedback system takes a particular approach to making such a device. Supposed that the desired gain is G , where G is equal to the ratio of the output to input signal: $G = \frac{e_o}{e_i}$

- The amplifier is deliberately designed to have much more gain than is required by the application, by a factor of at least $10G$.
- A sample of the output signal, equal to $1/G$, is subtracted from the input signal.
- If the amplifier is functioning correctly, the feedback signal exactly matches the input signal, and the difference between them is zero. If they are not exactly equal, there is a difference signal, known as the *error*.
- The error signal is used to drive the input of the high-gain amplifier. The effect is to drive the output in the direction that reduces the error signal.

The block diagram of a negative feedback system is shown in figure 218.

The amplifier gain is A_{ol} , for the *open-loop* gain. We shall use A or A_{ol} for the amplifier gain, and G for the gain of the negative feedback system, also known as the *closed loop* gain. The adder device has one positive input and one negative input, so the effect is to generate an error signal equal to $e_\epsilon = e_i - e_o$.

If this is done properly, the net effect is that the closed-loop gain is equal to $1/B$, the reciprocal of the sensor gain B .

This is a Good Thing, because the sensor is a relatively simple device compared to the amplifier, and therefore is likely to be stable and predictable¹¹⁷.

Further, any signals introduced by the amplifier or by external disturbances will show up as error signals that drive the amplifier in such a way to correct them. Consequently, distortion is reduced and the system is less susceptible to external disturbances.

The basic idea can be encapsulated into a few lines of algebra:

Starting at the input, we have that

$$e_\epsilon = e_i - e_f \quad (421)$$

The amplifier produces an output that is

$$e_o = A_{ol}e_\epsilon \quad (422)$$

The feedback signal is

$$e_f = Be_o \quad (423)$$

¹¹⁷The symbol β is often used for sensor gain, but we'll reserve that for the current gain of the junction transistor.

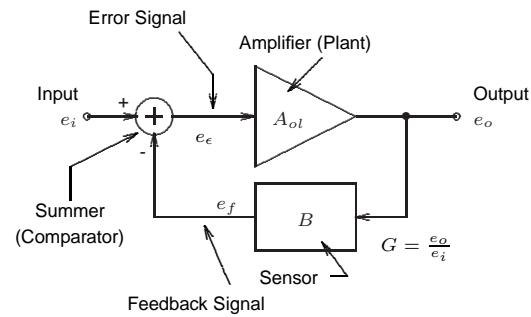


Figure 218: Negative Feedback System

Solving for the gain G , we have

$$\begin{aligned} G &= \frac{e_o}{e_i} \\ &= \frac{A_{ol}}{1 + A_{ol}B} \end{aligned} \quad (424)$$

Equation 424 may be rewritten as

$$G = \frac{1}{\frac{1}{A_{ol}} + B} \quad (425)$$

We can lose the A_{ol} term in the denominator if

$$\frac{1}{A_{ol}} \ll B \quad (426)$$

or what is the same thing, if

$$A_{ol} \gg \frac{1}{B} \quad (427)$$

Then we can say that

$$G \approx \frac{1}{B} \quad (428)$$

In words:

If the open-loop gain A_{ol} is much larger than the reciprocal of the sensor gain, then the gain of the negative feedback system is equal to the reciprocal of the sensor gain.

For example, if the value of B is 0.1, then the gain G of the negative feedback system will be 10 *providing* the open-loop gain A_{ol} is much larger than 10, eg, 100.

This is a very important concept and equation 424 should be committed to memory, along with the important approximation in equation 428.

Reference

A very interesting account of the invention of negative feedback in electronic amplifiers, by its inventor, is in [83].

10.2 Negative Feedback and the Operational Amplifier

The operational amplifier provides two of the main components for a negative feedback system: a subtractor and a high-gain amplifier.

A simple example that illustrates negative feedback using an operational amplifier is shown in figure 219. Compare this figure with figure 218 on page 286 to see the similarities to the general negative feedback system.

In this case, the sensor gain B is provided by a voltage divider, so that

$$\begin{aligned} B &= \frac{e_f}{e_o} \\ &= \frac{R_2}{R_1 + R_2} \end{aligned} \quad (429)$$

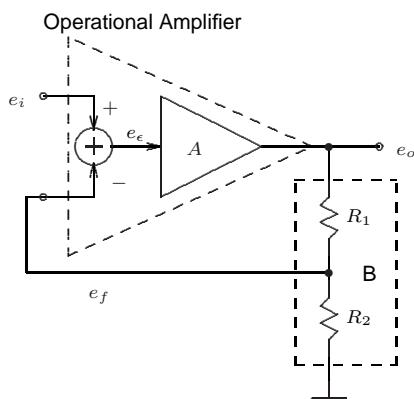


Figure 219: Negative Feedback using the Operational Amplifier

Providing that the open-loop gain A_{ol} is much larger than the closed-loop gain G the overall gain of the system is given by equation 428:

$$G = \frac{1}{B} \quad (430)$$

Substitute for B from equation 429 into equation 431, and

$$\begin{aligned} G &= \frac{R_1 + R_2}{R_2} \\ &= 1 + \frac{R_1}{R_2} \end{aligned} \quad (431)$$

This circuit is known as a *non-inverting amplifier* because the sign of the output is the same as the input. The minimum value of the gain is unity, when R_2 is made much larger than R_1 , or is open-circuited.

The analysis of a negative feedback system is often this simple. The key focus should be to identify the sensor and the sensor gain. Then the overall transfer function of the system is known.

10.3 Minimizing Disturbances

A negative feedback system minimizes disturbances. A *disturbance* is some signal that tends to cause the output of the negative feedback system to be incorrect. It could be caused by

- An imperfection in the amplifier, such as distortion or a dead spot in the transfer characteristic.
- Environmental influences, such as changing temperature or power supply voltage.
- Changes in the load on the system. As more current is required by the load, for example, the output voltage of the system tends to drop.

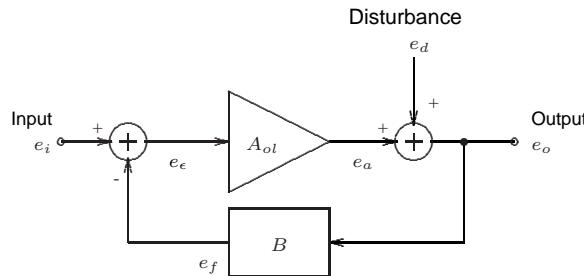


Figure 220: Modelling a Disturbance

Now let's do the block-diagram algebra on this system. The disturbance is represented by a signal e_d which adds to the output of the amplifier. The equations are very similar to those of section 10, with modifications for the effect of the disturbance:

$$\begin{aligned} e_\epsilon &= e_i - e_f \\ e_a &= A_{ol}e_\epsilon \\ e_o &= e_a + e_d \\ e_f &= Be_o \end{aligned}$$

After a certain amount of algebra, we can obtain:

$$e_o = \frac{A_{ol}e_i}{1 + A_{ol}B} + \frac{e_d}{1 + A_{ol}B} \quad (432)$$

For $A_{ol}B \gg 1$

$$\begin{aligned} e_o &\approx \frac{A_{ol}e_i}{A_{ol}B} + \frac{e_d}{A_{ol}B} \\ &= \frac{e_i}{B} + \frac{e_d}{A_{ol}B} \\ &= Ge_i + \frac{e_d}{A_{ol}B} \end{aligned} \quad (433)$$

In words:

The output signal e_o is composed of two parts: a desired signal equal to Ge_i due to the input signal and an undesired output $\frac{e_d}{A_{ol}B}$ due to the disturbance signal.

The effect of the disturbance signal is reduced by a factor equal to the loop gain, ie, by $\frac{1}{A_{ol}B}$.

A Numerical Example

Suppose a negative feedback system has the following parameters:

$$\begin{aligned} e_i &= 0.5 \text{ V} \\ B &= 0.1 \text{ so } G = 10 \text{ V/V} \\ A &= 10^5 \text{ V/V} \\ e_d &= 3 \text{ V} \end{aligned}$$

Calculate the effect of the 3 volt disturbance voltage on the output voltage.

Using equation 433, we have that:

$$\begin{aligned} e_o &= Ge_i + \frac{e_d}{A_{ol}B} \\ &= 10 \times 0.5 + \frac{3}{10^5 \times 0.1} \\ &= (5) + (3 \times 10^{-4}) \end{aligned}$$

The 5 volts is the desired output. The disturbance has been reduced to 3×10^{-4} volts, which is completely negligible.

How Does This Work?

In physical terms, where did the disturbance go? What happens is this:

1. A disturbance is added to the output of the op-amp.
2. The disturbance adds to the system output signal e_o .

3. This new output signal feeds back through the sensor device and to the adder.
4. The feedback signal is now no longer equal to the input signal.
5. The error signal e_ϵ changes.
6. This causes the output of the amplifier e_a to adjust itself in such a direction as to cancel the disturbance signal. In our numerical example, when the +3 volt signal adds to the output of the amplifier, the output of the amplifier moves down by -3 volts so that the system output e_o stays where it should be.

To summarize, the output of the amplifier adjusts in response to any disturbance to keep the output of the system where it should be, which is at Ge_i volts.

10.4 Negative Feedback Goes Bad

There are so many advantages to negative feedback that it is pervasive throughout electronic systems. However, when it goes wrong, it has the potential to go spectacularly wrong.

Saturation

If for some reason, the feedback signal is disabled (the sensor fails or a connection breaks), then the feedback signal goes to zero and the error signal becomes equal to the input signal. This finite signal is then amplified by the very large open-loop gain of the amplifier. The output would like to go to some very large value, but in practice bangs up against some limit of the system.

In the case of an operational amplifier, that limit is probably determined by the positive or negative power supply, and the amplifier is said to be *saturated*.

Now consider that the feedback system is used to steer an automobile, where the input is the position of the steering wheel operated by a driver, and the output is the pointing angle of the front wheels of the car. If the position sensor fails, the front wheels will go hard over to one side or another, clearly an unacceptably dangerous situation.

Consequently, it may be required to take special measures to ensure that the system fails in a *soft* manner, or undergoes *graceful degradation* rather than *catastrophic failure*. In the steering system example, the power assist might disappear making steering more difficult but still manageable.

A negative feedback *servo* system that controls the position of mechanical devices, often includes *limit switches* to disable the drive motor when it exceeds its range.

Oscillation

If the feedback signal is somehow modified so that it is *added* to the input signal rather than subtracted from it, the system will malfunction. This will occur if there is sufficient phase shift (180°) in the amplifier and sensor network. The basic equation governing the behaviour of the negative feedback system is:

$$G = \frac{A_{ol}}{1 + A_{ol}B} \quad (434)$$

If the term $1 + A_{ol}B$ approaches zero, then G blows up and becomes very large. If this happens at some particular frequency, then the system will oscillate at that particular frequency.

For this to happen, then

$$\begin{aligned} A_{ol}B &\rightarrow -1 \\ &\rightarrow 1\angle 180^\circ \end{aligned}$$

In practice, the magnitude of $A_{ol}B$ can be larger than unity when the phase angle of $A_{ol}B$ approaches 180° , and the system will burst into oscillation. The amplitude of the oscillations will grow until the amplifier or some other component goes into saturation, and that will effectively reduce the gain until the loop gain is unity. The oscillation waveform may appear as a clipped sine wave, but other shapes are possible.

Obviously, if this is a steering system, then an oscillation¹¹⁸ is highly undesirable.

When negative feedback was first invented by Harold Black, the patent office refused to issue a patent on the grounds that feeding the output back to the input would inevitably cause oscillation. Black showed that this was not the case by building amplifiers that used negative feedback and did not oscillate [83].

Where does this excess phase shift come from? There are inevitably RC lowpass filters in the $A_{ol}B$ path, each of which contributes some phase lag.

- In one common scenario, the output resistance of the amplifier drives a load capacitance. These two components form an RC lowpass filter which has a phase shift approaching 90° at high frequencies.
- Circuit impedances interact with stray capacitance to create lowpass filters that contribute phase shift to the signal [84]. This is a common problem in amplifiers that must operate at higher frequencies.
- A pure time delay creates a phase shift which increases with frequency, and this may be sufficient to cause oscillation¹¹⁹.

Preventing Oscillation

To prevent a negative feedback system from approaching an oscillatory condition, it is necessary to study the magnitude and phase responses of $A_{ol}B$. If these do not approach $1\angle 180^\circ$, then the system will be stable. Since it is usually impossible to prevent the phase angle from eventually reaching or exceeding 180° , the strategy is to ensure that the magnitude is rolled off to something less than unity at that frequency.

Depending on the characteristic of the amplifier and the frequency response of A_{ol} , it may be necessary to add one or more *phase compensation networks* that adjust the phase angle of A_{ol} . Section 24 covers this topic in detail.

Jiri Dostal is a useful reference on this subject, in [85].

10.5 Application: Temperature Controller

The circuit shown in figure 221 is a temperature controller for a small oven. We'll have a look at it here because it illustrates some important ideas about feedback.

Heat for the oven is produced by the resistor R_H , which is driven by the operational amplifier.

The basic sensing element is a *thermistor* R_T , which is a temperature sensitive resistor. The resistance-temperature characteristic of a typical thermistor is shown in figure 222. As the temperature increases, the thermistor resistance decreases in a non-linear fashion¹²⁰. The thermistor is attractive for temperature sensing applications because it is relatively sensitive to temperature, which simplifies the required electronic circuitry.

The heater and thermistor are both located in the oven. We assume that the amplifier can produce enough electrical power to heat the oven to the required maximum temperature. (If it did not, then one could increase the power output of the amplifier or insulate the oven more thoroughly.)¹²¹

¹¹⁸A mechanical engineer would call this *hunting*, since the system is hunting for its equilibrium point and never reaching it.

¹¹⁹Oscillation caused by a time delay is seen in many natural systems, such as the stock market and political policy-making.

¹²⁰Such a device is known as NTC: *negative temperature coefficient*. There are also PTC (positive temperature coefficient) devices.

¹²¹To keep things simple, the amplifier is shown as an operational amplifier in figure 221. However, a small-signal op-amp could not supply enough current to heat an oven. (Maximum output current of a typical small-signal op-amp is about 10mA.) However, a low-power op-amp can be made into a high-power op-amp with the addition of a buffer stage, as shown in section 12.3. In effect, the small-signal op-amp and the buffer become equivalent to an op-amp, so you can consider that is what is used in figure 221.

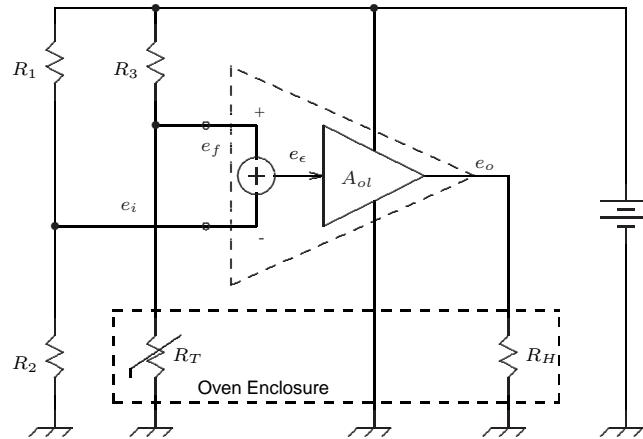


Figure 221: Temperature Controller

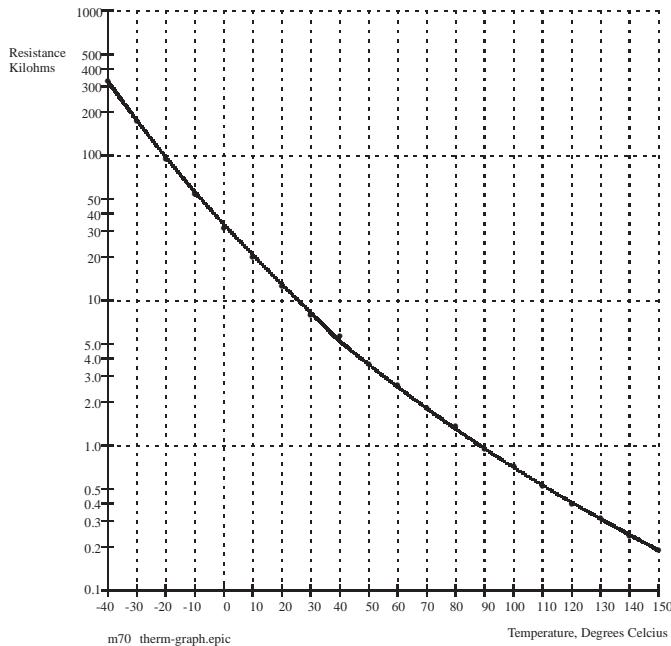


Figure 222: Thermistor Characteristic

Checking for Negative Feedback

The first task is to check that the system really is negative feedback, that is, that it operates such as to minimize the error signal. We do that by tracing the polarity of signals around the loop.

Suppose that the surroundings of the oven cool down. This constitutes a disturbance that causes the temperature of the oven to decrease. We know that the system should act in such a way as to counteract the effect of this

disturbance. Let's see if this is the case.

- The oven is caused to cool down. $T \downarrow$.
- This causes the thermistor resistance to increase: $R_H \uparrow$
- This raises the lower resistor in a voltage divider so that the feedback voltage e_f increases. $e_f \uparrow$
- The feedback signal is applied to the non-inverting input of the op-amp, so the output of the op-amp moves in the same direction, and so the output voltage increases: $e_o \uparrow$.
- If the output voltage increases, then the power in the heater resistance increases, increasing the temperature in the oven. $T \uparrow$.

Summarizing (the symbol \rightsquigarrow should be read as *leads to*):

$$T \downarrow \rightsquigarrow R_H \uparrow \rightsquigarrow e_f \uparrow \rightsquigarrow e_o \uparrow \rightsquigarrow T \uparrow. \quad (435)$$

Consequently, the action of the control system is to oppose (cancel) the effect of the disturbance, so it is indeed a negative feedback system.

Notice in this case that the feedback signal is connected to the non-inverting terminal of the op-amp, not the inverting signal. Consequently, *you cannot assume that the sensor signal goes to the inverting terminal* in a negative feedback system. It depends on the behaviour of the other components in the system.

Adjusting Output Temperature

How could you make the output temperature adjustable?

There are several possibilities, one of which is to make the setpoint voltage e_i adjustable. We could do that by replacing the R_1-R_2 voltage divider with a potentiometer.

Now consider this: what is the effect on the oven temperature T when the value of e_i is reduced?

We established that this is a negative feedback system, so the value of feedback voltage e_f will track the input voltage. If the input voltage decreases, so will the feedback voltage.

If the feedback voltage decreases, that implies that the thermistor resistance decreased, which further implies that the temperature increased.

So decreasing e_i will cause the oven temperature to increase.

Summarizing:

$$e_i \downarrow \rightsquigarrow e_f \downarrow \rightsquigarrow R_T \downarrow \rightsquigarrow T \uparrow. \quad (436)$$

This makes sense, and it accurately describes the steady state result of changing the input voltage. But it's a bit unsatisfying, because it does not entirely describe the chain of events. Let's do that:

- The setpoint voltage e_i is decreased: $e_i \downarrow$
- Assume that the system has not responded to this change, so at this particular time, the value of e_f is unchanged.
- Since e_i has decreased and e_f is unchanged, there is now a non-zero error signal equal to

$$\begin{aligned} e_\epsilon &= e^+ - e^- \\ &= e_f - e_i \end{aligned}$$

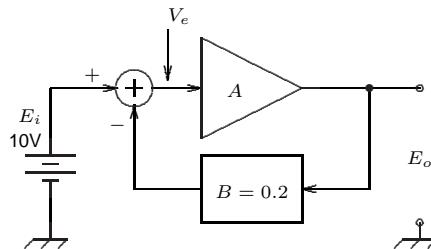
- This error signal is amplified by a factor of the open-loop gain (A_{ol} , typically 10^5V/V), which causes the output signal e_o to rise significantly. (In fact, the output of the op-amp will probably go up to its positive limit.)
- This increase in voltage across the heater resistor causes it to heat up and pump more energy into the oven.
- The oven temperature T increases.
- The increasing oven temperature causes the thermistor resistance R_T to fall, reducing the feedback voltage e_f .
- Eventually, the feedback voltage approaches the input voltage and as it does so, the error signal drops. The equilibrium position is reached when the difference between the input and feedback signals is equal to the output voltage divided by the open-loop gain. For all practical purposes, the error signal is very small and may be considered zero, so equilibrium occurs when the input and feedback voltages are again equal.

Summarizing:

$$e_i \downarrow \sim e_e \uparrow \sim e_o \uparrow \sim T \uparrow \sim R_T \downarrow \sim e_f \downarrow \quad (437)$$

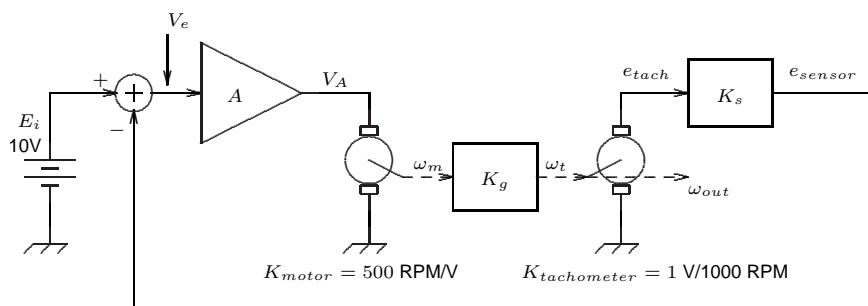
10.6 Exercises

1. The block diagram of a negative feedback system is shown in the figure. The amplifier may be considered to be ideal in all respects except for finite gain.



- (a) When the amplifier gain is 10^4 , determine the values of output voltage E_o and error voltage V_e .
 (b) When the amplifier gain is reduced to 10, determine the new values of output voltage E_o and error voltage V_e .

2. The block diagram of a motor speed control negative feedback system is shown in the figure. The transfer functions for the motor and tachometer are indicated on the diagram.



- For the purposes of this problem, the amplifier gain may be considered infinite and all components are operating in their linear regions – not saturated or cutoff.
- The output voltage of the amplifier drives a motor with a transfer function of 500 RPM/V:

$$K_m = \frac{\omega_m}{V_A} = 500 \text{ RPM/V}$$

- The output shaft of the motor drives a gearbox with a ratio of tachometer shaft speed to motor speed given by the transfer function:

$$K_g = \frac{\omega_t}{\omega_m}$$

- The speed of the output shaft of the gearbox is the output from the system. It is sensed by a tachometer with a transfer function of 1mV/RPM (1 volt/1000 RPM):

$$K_t = \frac{\omega_m}{V_A} = 0.001 \text{ V/RPM}$$

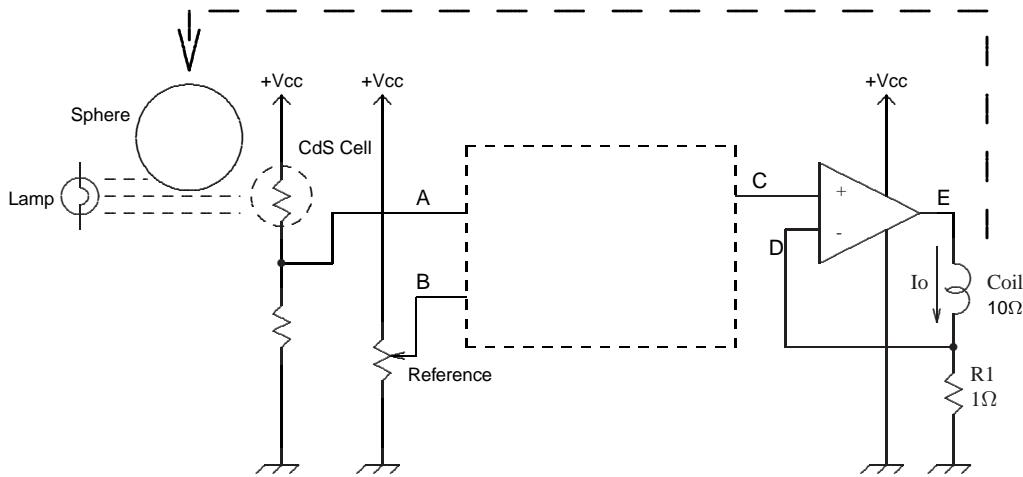
- The output voltage of the tachometer feeds a sensor scaling device with transfer function:

$$K_s = \frac{e_{sensor}}{e_{tach}}$$

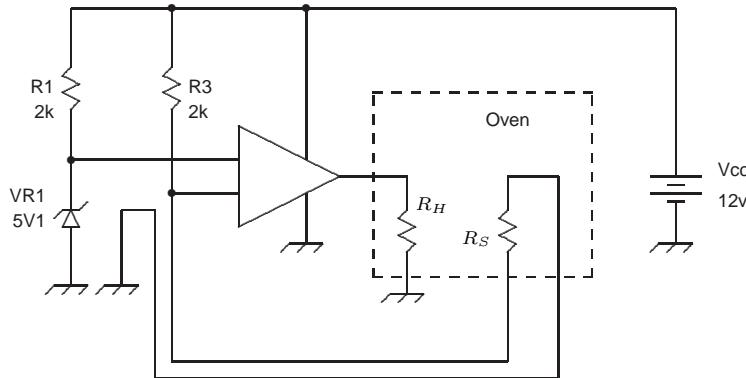
- (a) What is the value of the error voltage V_e ?
- (b) Assuming that the gearbox transfer function K_g and the sensor transfer function K_s are both unity, what is the rotational speed of the output shaft, ω_{out} ?
- (c) Now assume that the gearbox is changed to a stepdown ratio of 2:1 between the motor and tachometer (ie, $K_g = 0.5$). What is the effect on the tachometer shaft speed ω_t ? On the motor shaft speed ω_m ?
- (d) Assume that the gearbox ratio is restored to 1:1, and the sensor gain is reduced to $K_s = 0.5$. What is the effect on the tachometer shaft speed ω_t ? On the motor shaft speed ω_m ?
- (e) If the tachometer should fail in some way so that its output voltage drops to zero, without reflecting the value of the shaft speed, what effect does this have on the motor speed?
3. The circuit shown below is a device for suspending a magnetic sphere in space. The circuit pumps a current through the suspension coil, which generates a magnetic field just sufficient to balance the effect of gravity on the sphere. The negative feedback system ensures that the sphere stays in the same position.

The position of the sphere is sensed by a *cadmium sulphide photoresistor* (section 27.4). The resistance of the photoresistor decreases with increasing illumination. As the sphere falls, it reduces the light hitting the photoresistor, which should increase the current in the suspension coil, to increase the magnetic field, to pull the sphere back into its original position.

- (a) Assuming that a gain of 10V/V is required in this circuit between points A and C, design a suitable operational amplifier circuit to close the negative-feedback loop. Draw that circuit, with component values, inside the box on the diagram (or on the back of the previous page if it's too large). Your circuit should have provision for comparing the feedback signal with the reference voltage from the adjustable potentiometer.
- (b) As the coil conducts current, it will warm up, which will cause the resistance of the coil to increase. Assuming the negative feedback system is operating correctly, what voltages in the circuit (A through E) would change and in which direction (increase or decrease)? Justify your answer.



4. The circuit shown below is a negative-feedback system to regulate the temperature of a small, insulated enclosure for use in a very cold arctic environment. The heater resistor is R_H . The temperature sensor R_S is a *thermistor*. As the thermistor heats up, its resistance decreases. You may assume that the system is capable of heating the enclosure to whatever temperature is demanded by the electronics. You may also assume it is a correctly functioning negative feedback system (effectively no error).



The equation relating temperature to resistance of the sensor R_S is:

$$\frac{1}{T} = a + b(\ln R_S) + c(\ln R_S)^3$$

where

R_S is the sensor resistance, **in kilo-ohms**

a is a constant, 2.772×10^{-3}

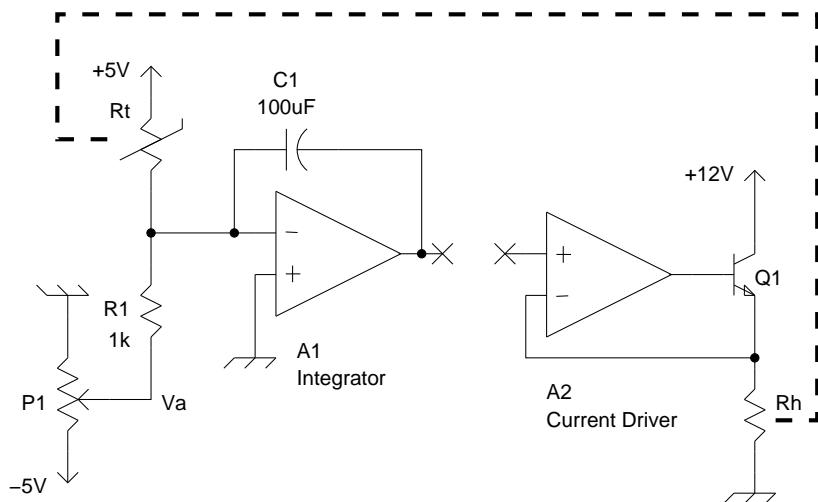
b is a constant, 2.524×10^{-4}

c is a constant, 3.042×10^{-7}

T is the temperature in degrees Kelvin (ie, degrees C plus 273)

- (a) Indicate the inverting and non-inverting terminals on the op-amp so that the circuit functions as a

- negative feedback system.
- Determine the current through the zener diode $VR1$.
 - Determine the resistance of the thermistor R_S .
 - Determine the temperature of the enclosure, degrees C.
 - Which of the resistors in the circuit would you adjust to vary the temperature in the enclosure? Explain.
5. A device for sensing wind speed is shown below. A thermistor R_t is heated by a resistor R_h and maintained at the same temperature by a negative feedback system. As air moves past the thermistor, it tends to cool it down and more heat is required from the resistor to keep it at the same temperature. Consequently, the amount of heat required is proportional in some fashion to wind speed.



- (a) Determine what should be installed in the interval XX to make the system a negative feedback system.
 (b) The parameters of the thermistor are as follows:

- Thermistor Equation: $R_t = R_{25}e^{\beta(\frac{1}{T} - \frac{1}{T_{25}})}$
- $\beta = 3977$
- $R_{25} = 10k\Omega$
- T is the operating temperature of the thermistor, K (kelvins)
- T_{25} is the temperature in kelvins corresponding to 25°C.

It is decided to operate the thermistor at 60°C. Determine its resistance at this temperature.

- Determine the voltage V_a when the thermistor temperature is 60°C.
- Calculate a suitable resistance value for potentiometer $P1$.
- On the diagram, indicate where you would attach a voltmeter to indicate wind speed.
- If you wanted to increase the operating temperature of the thermistor, in which direction would you move the wiper of P_1 ? Explain.

6. **Temperature Regulator using V_{be} Sensor.** The circuit diagram below (figure 223) shows a temperature regulator for a small enclosed, insulated box (the oven). Transistor Q_1 is used as a heater for the box interior and diode-connected transistor Q_2 senses the temperature inside the box.

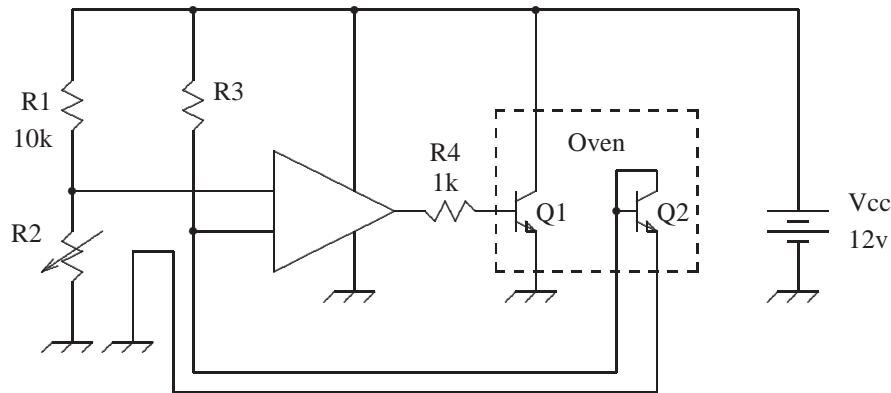


Figure 223: Oven with Diode Sensor

The temperature sensor Q_2 has a forward voltage V_{be} of 0.595 volts at 25°C and 1 mA of forward current. The value of V_{be} for Q_2 changes at a rate of $-2.2 \text{ mV}/^\circ\text{C}$. The interior of the heated box is to be maintained at 50°C and ambient temperature is 20°C .

- Determine the polarity of the op-amp inputs to make this a negative feedback system, and show this on the circuit diagram.
 - Determine a suitable value for R_3 if the current through the sensing transistor Q_2 is 1mA.
 - Calculate the value of R_2 to keep the interior of the box at 50°C .
 - (This problem requires some understanding of the BJT (section 29, page 798). How would you expect the collector current in Q_1 to change as the ambient temperature decreases? Explain.)
7. (This problem requires some understanding of the BJT (section 29, page 798)).

The circuit shown in figure 224 is an oven controller. Resistor R_2 is a thermistor of about $4\text{k}\Omega$ at 200°C , located inside the oven to sense its temperature. The thermistor changes resistance by approximately $10\Omega/\text{degree C}$ when at 200°C .

The LED is optically coupled to an optical SCR (switch) so that the oven heater turns on when the LED is ON.

- The transistors Q_1 and Q_2 are used as constant current sources. Calculate current I_{c2} .
- Should the thermistor be a positive or negative coefficient device?
- Explain how the circuit operates.
- How would you increase the set point temperature by 10°C ? (You must calculate a new component value.)

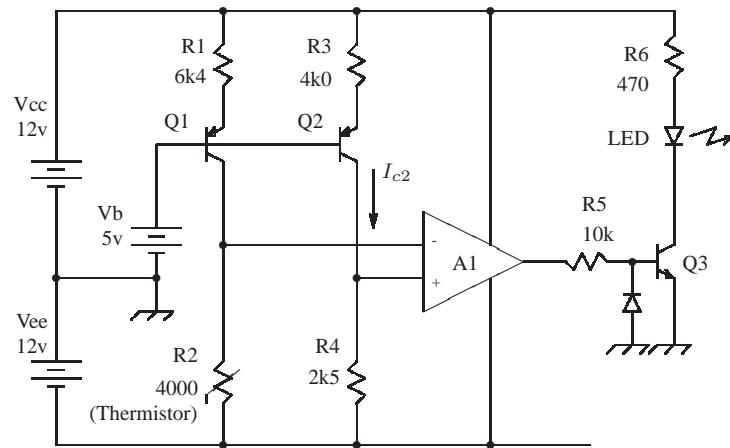


Figure 224: Thermistor Oven Controller

11 Reactance and Bode Plots

Introduction

The *Reactance Plot* and *Bode Plot* are graphical tools that are tremendously useful in analysing and designing circuits that have some frequency dependence. The term *frequency dependence* implies that some property, like the reactance of a network or the magnitude of its transfer function changes with frequency.

The reactance plot and the Bode plot have much in common: they are based on a log-log plot of some quantity versus frequency. The reactance plot shows the logarithm of reactance vs the logarithm of frequency. The Bode plot shows the gain in decibels (a logarithmic measurement) and phase (in linear form), both vs the logarithm of frequency.

An example of a reactance plot is shown in figure 112 on page 164, where the reactance of a capacitor is plotted vs frequency. An example of a Bode plot is shown in figure 120(a) and 120(b) on page 171, where the gain and phase of a lowpass RC network are plotted vs frequency.

Why are the Reactance and Bode Plots so Useful?

- The plots that result are curved lines but can be approximated by straight lines with sufficient accuracy that useful information can be derived from them. Consequently, reactance and Bode plots can be drawn by hand. Or, when they are drawn by a computer program, they are relatively simple to interpret. Bends in the curve (aka *breakpoints*) are directly associated with certain components in the circuit.
- When components are connected in series, the resultant frequency response plot is the product of the individual responses. Because the plot is logarithmic in the vertical axis, and multiplication is equivalent to adding logarithms, you can obtain the product by *adding* the individual responses. Since the individual responses are approximated by straight lines, this turns out to be very easy to do.
- When the frequency and/or phase must be modified in some way, it is possible to use the Bode plot or reactance plot to determine the type and value of the components that should be added, modified or removed.
- These impedance and frequency response plots may be generated with very simple lab equipment. It is tedious, but they can be hand plotted with the help of a signal generator and oscilloscope. Or there is automated equipment – the *network analyser* – that will generate these graphs directly¹²². In other words, there is a direct link between the behaviour of some component or network and its graphical representation.
- It turns out that the Bode plot represents quantities that have real meaning to human hearing. The logarithmic vertical and horizontal axes of a Bode plot closely represent the behaviour of human auditory equipment. A change of a certain given amount on the vertical axis, regardless of its position on the vertical axis, represents (approximately) the same apparent change in loudness to the human ear. Similarly, a constant amount of change on the horizontal axis is perceived (approximately) as the same change in pitch. For example, a doubling in frequency is always perceived as an octave, regardless of where it occurs on the frequency scale. Consequently, the Bode plot of an active filter used in audio equipment can be directly related to its acoustical effect.

What are the limitations?

The phenomenon of *resonance* must be identified independently of the Bode chart. For example, consider the case of a capacitor and inductor in series. If you plot the frequency characteristic of the capacitor and the inductor

¹²²For example, the Syscomp CircuitGear CGR-101 (www.syscompdesign.com) includes the necessary circuitry and software.

separately on a reactance chart, you would expect (as the frequency increases) that the series combination would decrease in reactance with the capacitor until the reactances were equal, and then increase with the reactance of the inductor, for a valley shaped curve. But in fact, at the point where the reactances are equal, the impedance of the series combination drops to zero! There is a clue in the plot of phase vs frequency: it changes abruptly through from $+90^\circ$ to -90° as the network changes from mainly capacitive to mainly inductive, and abrupt changes of phase are representative of resonance. However, there are better ways to understand resonance than a reactance plot.

Once it is understood that some network is a resonant system (also known as a *second-order* system), then its curves can be plotted onto a Bode plot. It is possible to do this by hand, but a computer plot is preferable. Once this is on the plot, then the amplitude and phase response may be added to other responses to get an overall response for the system.

Pattern Recognition

A complex reactance or Bode plot can be built up from simpler plots. All plots (we'll ignore the second-order resonance plot for the moment) are based on 3 patterns. However, two other patterns occur frequently enough that they are worth learning and recognizing, so there are 5 patterns in total. Each of these patterns includes a gain and a phase component. In the next section we'll introduce each of these components and then we will see how they fit together to make composite plots.

11.1 Reactance Plot

There are three basic types of plots that can occur as components on a reactance plot, representing resistance, inductance and capacitance.

- Resistance is constant with frequency.
- Inductive reactance is given by

$$\begin{aligned} X_l &= sL \\ &= j\omega L \\ &= |2\pi fL| \angle +90^\circ \end{aligned}$$

The magnitude of the inductive reactance is directly proportional to frequency: an increase in frequency by a factor of 10 increases inductive reactance by a factor of 10.

- Capacitive reactance is given by

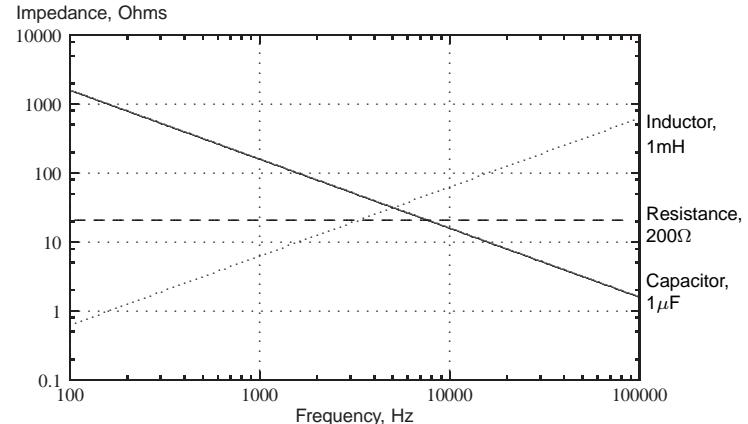


Figure 225: Reactance Types

$$\begin{aligned} X_c &= \frac{1}{sC} \\ &= \frac{1}{j\omega C} \\ &= \left| \frac{1}{2\pi fC} \right| \angle -90^\circ \end{aligned}$$

The magnitude of the capacitive reactance is inversely proportional to frequency: an increase in frequency by a factor of 10 reduces capacitive reactance by a factor of 10.

The magnitude plots of three reactances are shown in figure 225.

Reactances in Series and Parallel

With the help of some simple rules, we can represent the series and parallel combination of reactances on this same chart. Again, we'll ignore phase.

- When reactances are in series, the largest reactance dominates.
- When reactances are in parallel, the smallest reactance dominates.

Two examples are shown in figure 226.

When a resistor and capacitor are in series, then the capacitive reactance dominates at low frequencies. As the frequency increases, the reactance decreases until it is equal to the resistance. Beyond that frequency, the reactance is equal to the resistance. Because the slope of the reactance plot changes at this frequency, it is known as the *break frequency* and often given the name f_o or ω_o . A plot of reactance of the series pair follows the capacitance up to the break frequency and then follows the resistance.

When a resistor and capacitor are in parallel, the capacitive reactance is large at low frequencies so the resistance dominates. Consequently, a plot of the reactance of the parallel pair follows the resistance up to the break frequency and then follows the capacitance.

In this case, the break frequency occurs where the resistance and capacitive reactance are equal, so

$$\begin{aligned} R &= \frac{1}{\omega_o C} \\ \omega_o &= \frac{1}{RC} \\ f_o &= \frac{1}{2\pi RC} \end{aligned}$$

The Effect of Phase

The effect of phase is very simple: the phase of the impedance follows the phase of the dominant component. For example, if the dominant component is a capacitor, then the phase of the reactance is approximately -90° . If the dominant component is a resistance, then the phase of the reactance is approximately 0° . If there is a break frequency defining a transition from, say, -90° to 0° , then the phase angle of the impedance at the break frequency will be halfway between the two reactances, -45° .

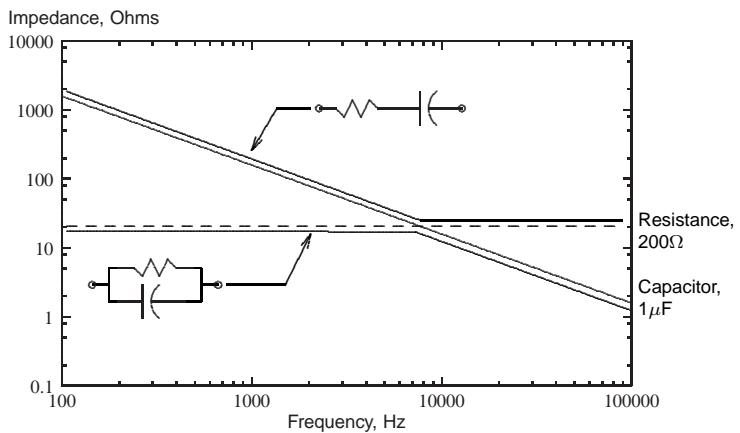


Figure 226: Reactance Plots, Series and Parallel

A plot of reactance of the series pair follows the capacitance up to the break frequency and then follows the resistance.

A plot of reactance of the parallel pair follows the resistance up to the break frequency and then follows the capacitance.

In this case, the break frequency occurs where the resistance and capacitive reactance are equal, so

$$\begin{aligned} R &= \frac{1}{\omega_o C} \\ \omega_o &= \frac{1}{RC} \\ f_o &= \frac{1}{2\pi RC} \end{aligned}$$

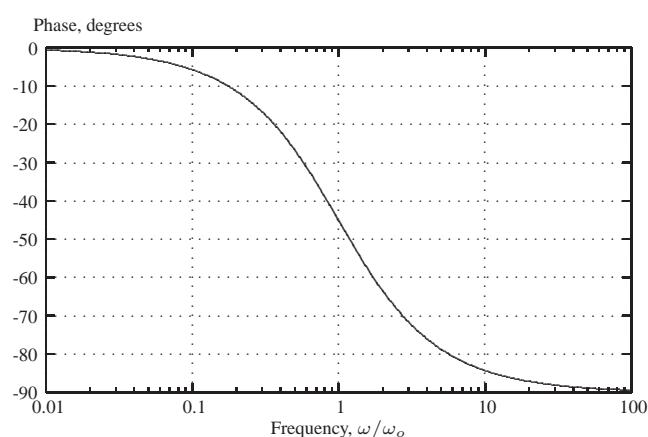


Figure 227: Reactance Phase Plot

The changeover occurs in the region $f_o/10$ to $10f_o$, as shown in figure 227.

This phase curve may be approximated by a straight line segments that are

- 0° below $\omega_o/10$
- between 0° and -90° over the frequency range $\omega_o/10$ to ω_o
- -90° above $10\omega_o$

An Example

In this example, we'll plot the magnitude and phase frequency response of the impedance Z_T for the network shown in figure 228.

1. The first step is to examine the network and get a feel for its behaviour at low and high frequencies.

- At low frequencies, the capacitor is an open circuit, so the total resistance is $R_1 + R_2$. But R_2 is ten times larger than R_1 , so $R_1 + R_2 \approx R_2$.
- At high frequencies, the capacitor is a short circuit and R_2 is bypassed. Then the total resistance is simply R_1 .
- At some intermediate frequencies, the total impedance has to step down from R_2 to R_1 . It's not immediately obvious where those break frequencies are. The reactance plot can help identify them.

2. Draw the reactance plots for each of the individual reactances, R_1 , R_2 and C , as shown in figure 229(a).

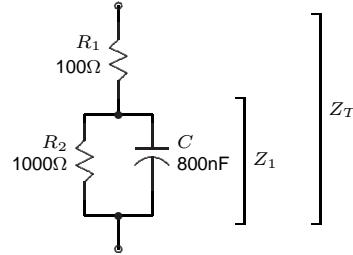
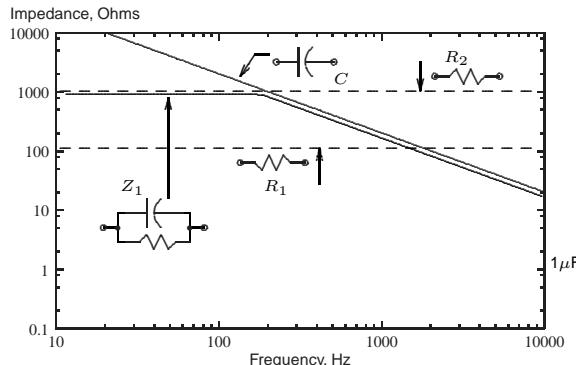
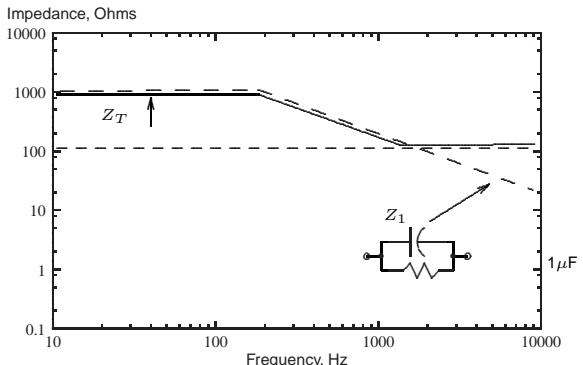


Figure 228: Example Network



(a) Step 1



(b) Step 2

Figure 229: Reactance Plots

The resistors are drawn as constant with frequency. The reactance of the capacitor passes through the point 1000Ω at 200Hz , and decreases by a factor of 10 with each factor 10 increase in frequency.

3. Now we can draw the impedance of the network Z_1 . The parallel combination of R_2 and C will follow the smaller of those two impedances, and the resultant curve is also shown in figure 229(a).

4. Next, we'll think about the total impedance Z_T of the network, which is the series combination of R_1 and Z_1 . The larger of these two predominates, so the trace for Z_T follows Z_1 at low frequencies, sloping down to follow Z_1 until it intersects with R_1 . At that point, R_1 is larger, so the trace for Z_T follows R_1 from that point onward. The result is shown in figure 229(b).

Plotting of the phase for this network follows a similar reasoning process.

The Important Thing

When dealing with a network that is made up of frequency-dependent reactances (inductors and capacitors), it is important to be able to picture the frequency response of the resultant impedance. It's often useful to draw a sketch on squared paper to get a general idea of the behaviour of the magnitude and phase against frequency.

Reference

The EDN article by Michael Street, [86], is the definitive paper on the reactance chart: plotting and applications.

11.2 The Bode Plot

Like the reactance plots of section 11.1, the Bode Plot is a graph of amplitude vs frequency. The frequency is always in logarithmic form. The amplitude may be either the magnitude of the transfer function of some network, or the phase of the transfer function of the network.

For an amplitude plot, the gain is a measure of the magnitude of the transfer function, expressed in decibels. For example, if the network is a voltage amplifier, the magnitude is

$$|G| = 20 \log_{10} \frac{e_o(s)}{e_i(s)} \quad (438)$$

where $|G|$ is the magnitude of the gain in *decibels* and $e_o(s)$ and $e_i(s)$ are the output and input voltages of the amplifier. Their ratio is the *transfer function* of the amplifier. Showing e_o and e_i as functions of s simply emphasizes that they are functions of frequency, that is, the magnitude of the transfer function changes with frequency¹²³.

One can apply equation 438 to determine the gain of a network. But certain values of gain in decibels are worth remembering (this comes with practice):

1. The magnitude of unity gain is 0 db (because the logarithm of 1 is zero).
2. A gain of $\sqrt{2}$ is +3db. This often occurs when the two voltages e_o and e_i are equal in magnitude but 45° out of phase.
3. Taking a reciprocal of some gain changes the sign of the gain value in decibels. For example, a gain of 100 volts/volt is +40db, a gain of 1/100 is -40db.
4. As a result of the previous rule, taking the reciprocal of some function $G(s)$ has the effect of reflecting its magnitude curve around the zero db axis of the plot. Where the function was formerly at $+x$ db, it will become $-x$ db, and vice versa. Applying this rule to all the points in the function causes it to be reflected.
5. A gain of $\times 2$ volts/volt is +6db. One-half is -6db¹²⁴.

¹²³In practice, $e_i(s)$ is likely to be a test signal that is constant in amplitude with frequency, and $e_o(s)$ is the one that is dependent on the network and changes with frequency.

¹²⁴A gain of 8db is perceived in an audio signal as approximately *twice as loud*.

6. A gain of $\times 10$ volts/volt is +20db.
7. Each additional multiplication of the gain by a factor of 10 increases the decibel value by 20db, so a gain of +80 db is 10^4 volts/volt.

For the phase plot, phase is plotted in linear form against frequency.

Poles and Zeros

I'll discuss these terms briefly because they are pervasive in the literature.

A complete transfer function can be written as the ratio of two polynomials in s .

$$G(s) = \frac{N(s)}{D(s)} \quad (439)$$

When the numerator $N(s)$ and denominator $D(s)$ are factored, they each become the product of one or more terms. When a term is in the numerator, it is referred to as a *zero*. When that term becomes zero, the overall function $G(s)$ becomes zero.

When the term is in the denominator, it is referred to as a *pole*. When that term becomes zero, the overall function $G(s)$ becomes infinite.

So the words *pole* and *zero* simply refer to terms in the numerator and denominator of the transfer function.

It is possible to plot poles and zeros on the *complex frequency plane*, and this can lead to some insight about the frequency response of the system. That will be explored more fully in section 17 on Active Filters.

References

Bode plots are a primary tool in the design and analysis of control systems, so any basic textbook on control system design will have a section on Bode plots. For example, see [87], [88] and [89].

11.3 Constant, Integrator and Differentiator Factors

In this and subsequent sections, we'll look at the component factors that may be used to build up a Bode plot.

Constant

A constant factor k shows on the Bode magnitude plot as a horizontal straight line at a value of

$$G(s) = 20 \log_{10} k \text{ db} \quad (440)$$

On the phase plot, the constant shows as 0° if positive or -180° if negative.

Taking the reciprocal of a constant factor simply reverses the sign of the decibel value:

$$\begin{aligned} G(s) &= 20 \log_{10} \frac{1}{k} \\ &= 20 \log_{10}(1) - 20 \log_{10}(k) \\ &= 0 - 20 \log_{10}(k) \\ &= -20 \log_{10}(k) \end{aligned}$$

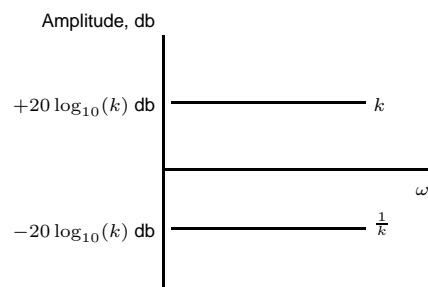


Figure 230: Constant Factor

The Bode amplitude plots for k and $1/k$ are shown in figure 230. The phase plots are not shown because they are trivial: zero degrees for a positive constant and -180 degrees for a negative constant.

Differentiator

The differentiator term has the form

$$G(s) = \frac{s}{1} \quad (441)$$

in Laplace notation form, or substituting $j\omega$ for s :

$$G(j\omega) = \frac{j\omega}{1} \quad (442)$$

The expression is shown as a fraction to emphasise that the $j\omega$ term is in the numerator.

This magnitude function slopes upward at the same rate as the frequency: an increase by a factor of 10 in frequency causes an increase by a factor of 10 in $G(j\omega)$. In decibels, an increase by a factor of 10 is equivalent to 20db, so the differentiator term slopes upward with frequency at a rate of 20db/decade.

At unity frequency, the function has a value of 1, or 0 db. So the function passes through the 0db axis at a frequency of 1 radian/second.

The phase function, due to the ' j ' term in the numerator, is a constant +90°.

Integrator

The integrator term has the form

$$G(s) = \frac{1}{s} \quad (443)$$

in Laplace notation form, or substituting $j\omega$ for s :

$$G(j\omega) = \frac{1}{j\omega} \quad (444)$$

As a consequence of rule 4 above, the integrator magnitude curve is obtained by reflecting the differentiator magnitude curve around the zero db axis of the plot.

The magnitude of the integrator slopes downward with frequency at a rate of 20db/decade, and again passes through the 0db axis at a frequency of 1 radian/second.

Since the j term is now in the basement of this expression, the phase of the integrator is a constant -90°.

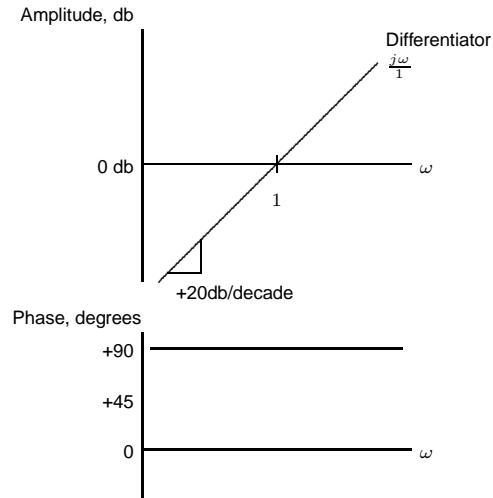


Figure 231: Differentiator

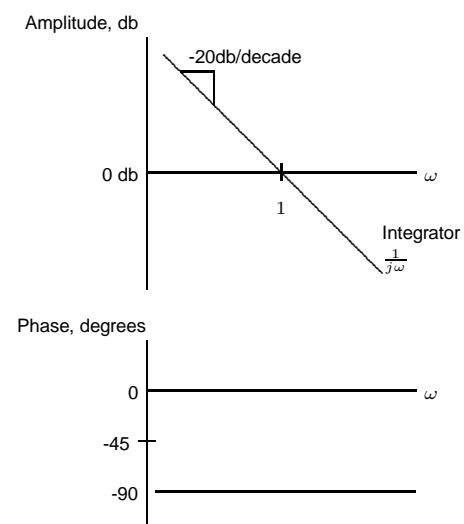


Figure 232: Integrator

11.4 First Order Lead and Lag Factors

11.5 First-order Lead

The *first-order lead* function in Laplace notation has the form

$$G(s) = 1 + \tau s \quad (445)$$

The constant τ is the *time constant* that we met while studying RC networks.

Its reciprocal is the *cutoff frequency* ω_o . Substituting $1/\omega_o = \tau$ and $j\omega = s$, then we can write:

$$G(j\omega) = 1 + j\frac{\omega}{\omega_o} \quad (446)$$

At low frequencies, where ω is small compared to ω_o , the magnitude is given by

$$G(j\omega) \approx 1 \quad (447)$$

so the function trails along the 0db axis and the phase is approximately 0° .

At high frequencies, where ω is large compared to ω_o , the magnitude is given by

$$G(j\omega) \approx j\frac{\omega}{\omega_o} \quad (448)$$

so the function ramps upward at a rate of 20db/decade. The phase, due to the j term in the numerator, is $+90^\circ$

The transition from one behaviour to the other takes place around the *corner frequency* ω_o . At the frequency $\omega = \omega_o$, the expression for $G(j\omega)$ is

$$\begin{aligned} G(j\omega) &= 1 + j\frac{\omega}{\omega_o} \\ &= 1 + j1 \\ |G(j\omega)| &= \sqrt{2} \\ &= +3 \text{ db} \\ \angle|G(j\omega)| &= +45^\circ \end{aligned}$$

So the gain of $G(j\omega)$ is +3db and the phase $+45^\circ$ at the corner frequency ω_o .

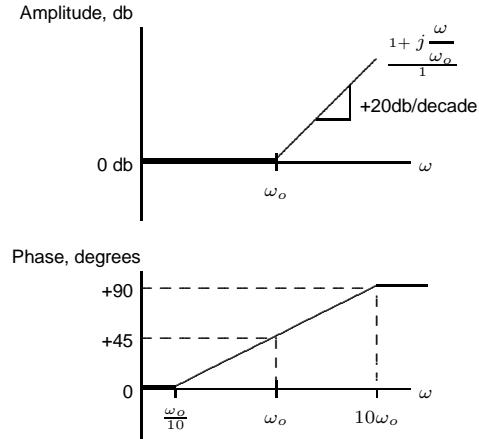


Figure 233: First Order Lead

Example

An example first-order lead¹²⁵ is shown in figure 234.

$$\begin{aligned} G(s) &= \frac{e_o}{e_i} \\ &= -\frac{Z_f}{Z_i} \\ Z_f &= R \\ Z_i &= R \parallel \frac{1}{sC} \end{aligned}$$

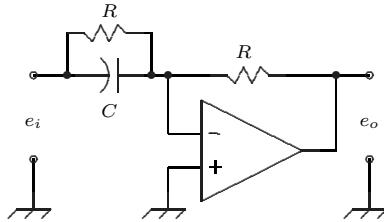


Figure 234: First-order Lead

Putting these together, we have

$$\begin{aligned} G(s) &= 1 + sRC \\ &= 1 + j\omega RC \\ &= 1 + j\frac{\omega}{\omega_o} \end{aligned}$$

where

$$\omega_o = \frac{1}{RC} \quad (449)$$

11.6 First-order Lag

The *first-order lag* function is the reciprocal of the first-order lead, and has the form

$$G(j\omega) = \frac{1}{1 + j\frac{\omega}{\omega_o}} \quad (450)$$

Consequently, according to rule 4 from section 11.2, we can obtain the characteristic of the first-order lag by reflecting the first-order lead about the x axis. Then first-order lag has the following properties:

- the magnitude trails along the 0db axis up to the region of ω_o
- the magnitude breaks downward at ω_o and thereafter decreases at a rate of 20db per decade
- the magnitude has a value of -3db at ω_o

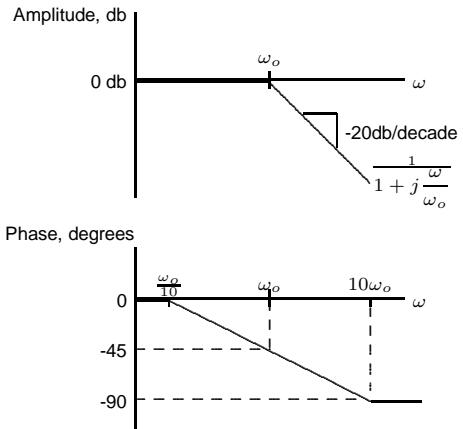


Figure 235: First Order Lag

¹²⁵This example requires some understanding of the operational amplifier, which is explained in section 12. At this point, it's sufficient to understand that an amplifier is necessary because the output is larger than the input signal.

Depending on the values chosen for R and C , this circuit is likely to be unstable and oscillate at some high frequency. The usual cure is to place a small resistor in series with the capacitor C , which changes the behaviour from a pure lead circuit. In section 24.4, we'll discuss why the circuit is prone to oscillation and why the cure works.

- the phase ranges from 0° at low frequencies (below $\omega_o/10$), transitions through -45° at ω_o and is equal to -90° at high frequencies (above $10\omega_o$).

Example

The RC lowpass network (section 4.8) shown in figure 236 behaves exactly as a first-order lag.

$$\begin{aligned}
 G(s) &= \frac{e_o}{e_i} \\
 &= \frac{1}{R + \frac{1}{sC}} \\
 &= \frac{1}{1 + sRC} \\
 &= \frac{1}{1 + j\omega RC} \\
 &= \frac{1}{1 + j\frac{\omega}{\omega_o}}
 \end{aligned}$$

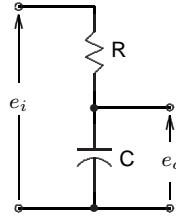


Figure 236: First-order Lag

where

$$\omega_o = \frac{1}{RC} \quad (451)$$

11.7 Example: PID (Proportional, Integral, Differential) Controller

In this section, we'll apply Bode plot analysis to the design of an important analog circuit, the PID Controller.

Many industrial systems are controlled with the help of negative feedback. The block diagram of such a system is shown in figure 237.

The *plant* is some device, like a chemical cooker or the motor in a paper-making machine, that must be controlled correctly. The actual behaviour of the device is sensed and compared with a *setpoint* signal. The difference between the feedback and setpoint signals generates an error ϵ . A *controller* acts on this error signal to produce any corrective signals for the plant.

The controller contains circuitry to ensure that the steady-state error is very small, and that the system responds properly to transient inputs. For example, the system should respond quickly to a step input, but not overshoot.

For critical applications, the transfer function of the plant must be known in detail to design the controller.

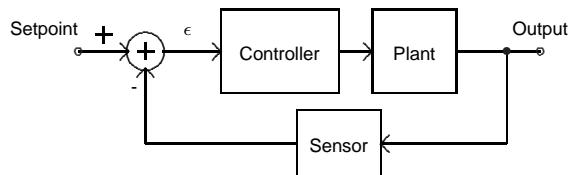


Figure 237: Industrial Control System

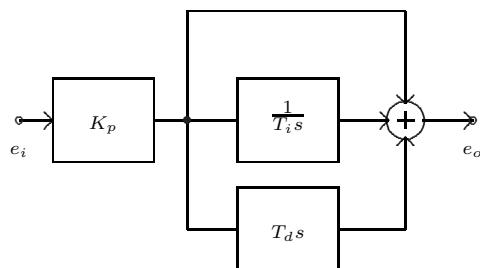


Figure 238: PID Controller

However, in many industrial applications, the behaviour of the plant is not known exactly and the controller can be tuned empirically, on the job, to make the system perform correctly.

A very common controller for this purpose is known as the PID controller - for *proportional, integral and differential*. The block diagram of a PID controller is shown in figure 238, and has a transfer function

$$\frac{e_o}{e_i} = K_p \left(1 + \frac{1}{T_i s} + T_d s \right) \quad (452)$$

The controller adjustments are:

- Proportional gain K_p .
- Integrator time-constant T_i
- Differentiator time-constant T_d

In a non-critical application, the operator increases the proportional gain until the system begins to evidence instability. Then she backs off on the proportional gain some amount and adjusts the integrator and differentiator gain for best transient response.

Example Controller Design

Our focus is the electronic design of the controller. In this section, we'll describe a plant, determine the controller settings, and design a suitable controller.

The Ziegler-Nichols rules are a systematic method of tuning a controller [90], [89]. For example, if the plant responds to a step input with a delay of D seconds, followed by a rise to the final value over R seconds, one possible group of settings is:

Proportional Gain	K_p	$1.2 \frac{R}{D}$
Integrator Time-constant	T_i	$2D$
Differentiator Time-constant	T_d	$0.5D$

Let's assume that a certain plant is measured to have a step response that includes a delay time $D = 0.1$ seconds and rise time $R = 0.5$ seconds.

Then according to the Ziegler-Nichols rules given above,

$$\begin{aligned} K_p &= 1.2 \frac{0.5}{0.1} \\ &= 6 \\ T_i &= 2(0.1) \\ &= 0.2 \text{ seconds} \\ T_d &= 0.5(0.1) \\ &= 0.05 \text{ seconds} \end{aligned}$$

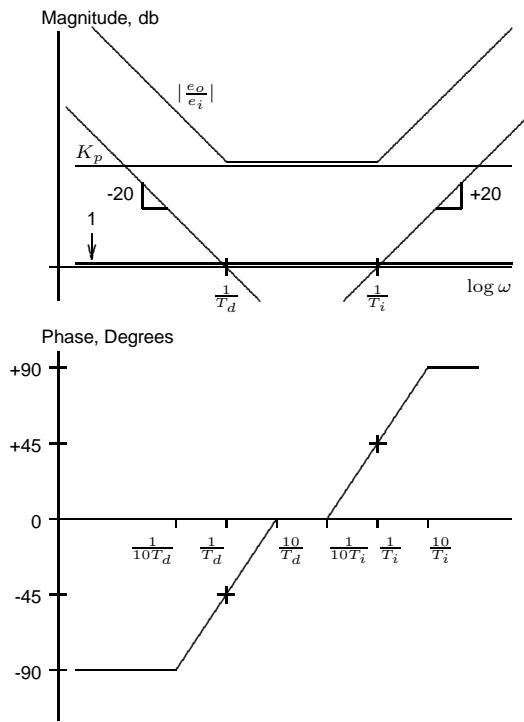


Figure 239: PID Controller Bode Plot

Controller Bode Plot

To draw the Bode plot, consider the various components in equation 452 and their effect on the magnitude of the transfer function.

- The constant term 1 has a value of 0db.
- The integrator term $1/T_i s$ slopes downward at 20db/decade, passing through the 0db axis at a frequency of $1/T_i$ radians/second.
- The differentiator term $T_d s$ slopes upward at 20db/decade, passing through the 0db axis at a frequency of $1/T_d$ radians/second.
- When these last three terms are combined by addition, the result is a Bode magnitude plot that follows the largest of the three of them. Let's call this result the bracketed factor.
- The proportional gain term K_p is a constant at $+K_p$ db. Its effect is to move the bracketed factor upward by $+K_p$ db.

The resultant magnitude plot is shown in the top half of figure 239. The phase may be determined according to the following rules:

- The integrator contributes a phase shift of -90° . This phase shift applies while the magnitude function is decreasing.
- When the magnitude function transitions from a negative slope to a horizontal slope, the phase angle transitions to 0° .
- The transition from -90° to 0° occurs over a frequency span of $1/10T_i$ to $10/T_i$. At the break frequency $1/T_i$, the phase angle is -45° .
- The differentiator contributes a phase shift of $+90^\circ$. This phase shift applies while the magnitude function is increasing.
- In an analogous manner to the integrator phase angle transition, the transition from 0° to $+90^\circ$ occurs over a frequency span of $1/10T_d$ to $10/T_d$. At the break frequency $1/T_d$, the phase angle is $+45^\circ$.

Putting all this together, we have the phase plot shown in the bottom half of figure 239.

11.8 Quadratic Factor

The quadratic term has the form

$$1 + 2\delta \left(j \frac{\omega}{\omega_o} \right) + \left(j \frac{\omega}{\omega_o} \right)^2$$

The constant δ is the *damping factor*¹²⁶. We'll show the effect of damping in a moment.

As a practical matter, the quadratic term usually appears in the denominator of the transfer function, eg, as

$$G(j\omega) = \frac{1}{1 + 2\delta \left(j \frac{\omega}{\omega_o} \right) + \left(j \frac{\omega}{\omega_o} \right)^2} \quad (453)$$

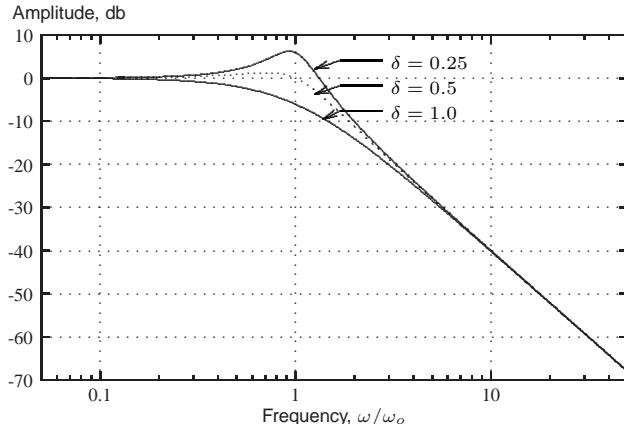


Figure 240: Quadratic Factor, Magnitude

To obtain an expression for the amplitude of this function, collect the real parts and imaginary parts. The amplitude is given by the formula for a hypotenuse, where the two sides of the triangle are the real and imaginary components. Then the amplitude of this term is given by

$$|G(j\omega)| = \frac{1}{\sqrt{\left(1 - \left(\frac{\omega}{\omega_o}\right)^2\right)^2 + \left(2\delta\left(\frac{\omega}{\omega_o}\right)\right)^2}} \quad (454)$$

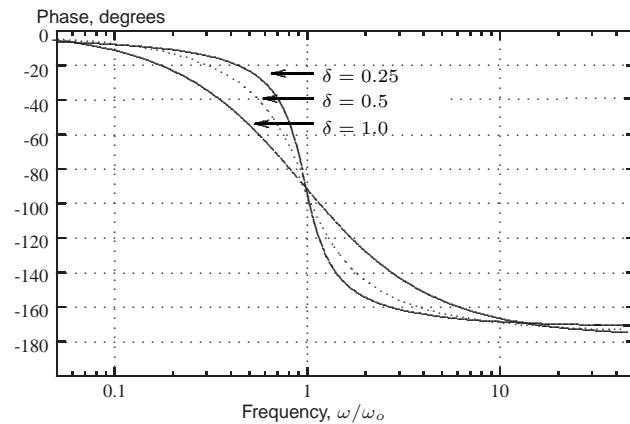
The tangent of the phase angle is equal to the imaginary component divided by the real component, so:

$$\angle G(j\omega) = -\tan^{-1} \left[\frac{2\delta\left(\frac{\omega}{\omega_o}\right)}{1 - \left(\frac{\omega}{\omega_o}\right)^2} \right] \quad (455)$$

The magnitude and phase are plotted in figures 240 and 241.

Applying the 'at-low-frequency', 'at-high-frequency' analysis to the formula second-order quadratic:

- At low frequencies, the amplitude trails along the 0db axis.
- At high frequencies, the amplitude rolls off at a rate of 40db per decade. (The term $(\omega/\omega_o)^2$ predominates.)
- In the region of ω_o , the shape of the curve depends on the damping factor, figure 240.



¹²⁶The damping factor is proportional to the inverse of the Q factor, section 4.1. $\delta = \frac{1}{Q}$

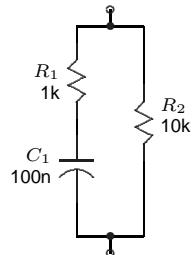
Figure 241: Quadratic Factor, Phase

- The lower the damping factor, the more pronounced the peaking in the transition region.
- At low frequencies, the phase is 0° .
- At high frequencies, the phase is -180° .
- The phase makes its transition between 0° and -180° in the region of the corner frequency ω_o .
- The lower the damping factor, the more abrupt the transition in phase (or, put another way, the smaller the range of frequencies over which the phase makes its transition.)

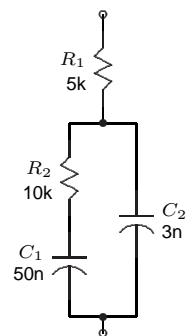
11.9 Exercises

1. With reference to the figure,

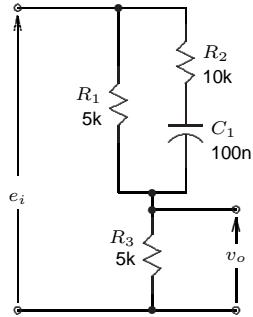
- What is the reactance of this network at low frequencies?
- What is the reactance at high frequencies?
- Sketch the reactance plot, showing break frequencies and impedance values



2. Repeat the previous question for the network shown below.



3. For the network shown below:



- (a) What is the gain v_o/e_i at low frequencies?
- (b) What is the gain at high frequencies?
- (c) Draw the Bode plot of gain v_o/e_i vs frequency.

12 The Operational Amplifier: Basics

12.1 Introduction

In section 10, we saw how negative feedback can benefit an electronic system.

The *operational amplifier* (op-amp) contains the core components of a complete negative feedback system:

- A *summer* that subtracts the feedback signal from the input signal to generate the error signal, and
- a high gain amplifier for the error signal.

Typically, the amplifier is a *voltage amplifier*. It presents a high impedance to the input and a voltage source (with low internal resistance) to the load. The voltage gain A is typically in the order of 10^5 volts/volt (100db).

A complete negative feedback system can be constructed with an operational amplifier plus a feedback network, and the behaviour of this system will be defined primarily by the feedback network. This was shown earlier in section 10.2, and an example from that section is shown in figure 242.

Everything inside the large dashed-outline triangle is contained in an integrated circuit operational amplifier. The two resistors define the **B** network and hence the closed-loop gain e_o/e_i of the system.

The op-amp is one of the great success stories of electronics. The $\mu\text{A}709$ from Fairchild, one of the first integrated circuit op-amps introduced in the 70's, cost over \$100. It was easily destroyed and more prone to oscillate than to amplify. Since then, op-amps have become more robust and easier to use. They have also dropped below 25 cents.

Bipolar Op-Amp Power

An operational amplifier is often used with a *bipolar* power supply, as shown in figure 243.

There are two power supplies: a positive supply V_{CC} and a negative supply V_{EE} . The point where these two supplies are joined is the ground reference point.

The output of the operational amplifier can swing as far positive as V_{CC} and as far negative as V_{EE} . In the days when operational amplifiers were used in analog computing, the signal ranged between plus and minus 10 volts, and the power supplies were typically plus and minus 15 volts. This is still a common choice of supply voltage for operational amplifiers¹²⁷.

It's often helpful to think of the output circuit of the operational amplifier as a potentiometer that is connected between V_{CC} and V_{EE} , with its wiper connected

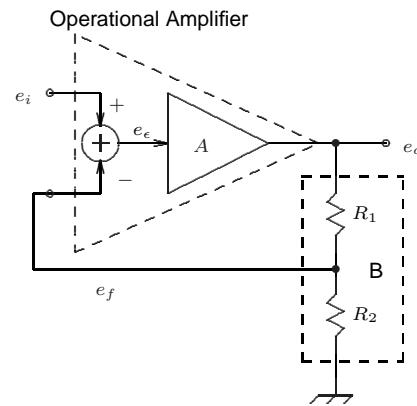


Figure 242: Negative Feedback using the Operational Amplifier

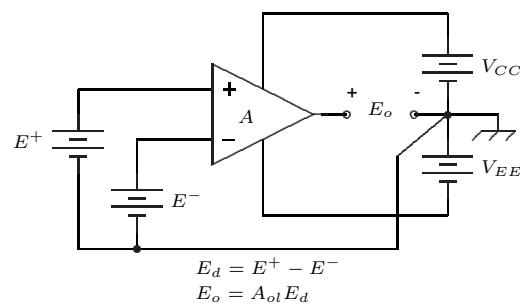


Figure 243: Op Amp Power and Input Voltages

¹²⁷This is a bit of a simplification. Many operational amplifiers can output a voltage only to within a volt or two of the power supply rails. However, the so-called *rail-rail* op-amp can drive its output to within a few millivolts of the supply rails.

to the output of the op-amp, as shown in figure 244. When the wiper is at the top of the pot, the output is $E_o = V_{CC}$. When it's at the bottom, the output is $E_o = V_{EE}$. (Notice that V_{EE} is negative with respect to the ground reference point.) When the wiper is in the center, $E_o = 0$

Alternatively, an op-amp may be powered from a single power supply rail and ground. This is explained in section 12.8.

In many op-amp circuits, we will omit drawing the power supply connections. However, they are always required¹²⁸.

Input Signals

The input signals to the op-amp are shown in figure 243, and their relationship to the output voltage is:

$$E_d = E^+ - E^- \quad (456)$$

$$E_o = A_{ol} E_d \quad (457)$$

The label on the + terminal should be read as *non-inverting*, not *plus*, and the label on the - terminal should be read as *inverting*, rather than *minus*.

Then *the output of the op-amp is proportional to the voltage at the non-inverting terminal minus the voltage at the inverting terminal*.

We can redraw the input circuitry of figure 243 in terms of the *differential* input voltage E_d and *common-mode* voltage E_{cm} as shown in figure 244, where the common-mode voltage E_{cm} is given by:

$$E_{cm} = \frac{E_i^+ - E_i^-}{2} \quad (458)$$

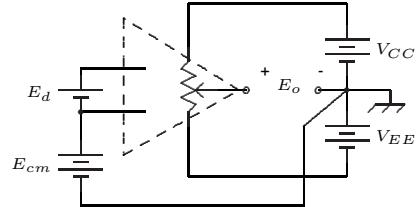
To a first approximation, the common-mode voltage E_{cm} has no effect on the output: it's only the differential voltage E_d that drives the output. However, there are limits to the common-mode voltage.

Typically, it cannot be above $V_{CC} - 2$ volts or below $V_{EE} + 2$ volts for the input circuitry to behave properly. In other words, the average value of the differential voltage must be centred somewhere between the supply rails.

Keep in mind that the output voltage can be positive or negative depending on the relative magnitudes of the two input voltages. Notice that no resistance is shown at the input, so that the current is zero into the op-amp input terminals.

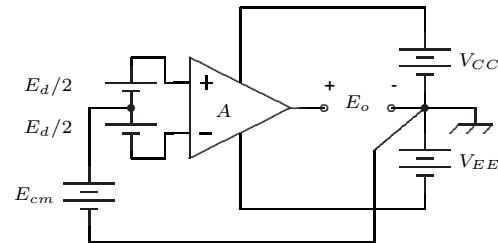
This is a very useful and important model for op-amp behaviour, but it has certain limitations that must be kept in mind:

- The output voltage cannot exceed specified values for the op-amp. These limits are within the boundaries established by the supply voltages, often by a volt or more.
- The common-mode input voltage cannot exceed specified voltages for the op-amp, again usually within the boundaries established by the supply voltages.



$$E_o = A_{ol} E_d$$

Figure 244: Op Amp Potentiometer Analogy



$$E_o = A_{ol} E_d = A_{ol}(E^+ - E^-)$$

Figure 245: Common Mode and Differential Inputs

¹²⁸This is something that escaped your author on the first circuit he built, a *starved current* vacuum-tube amplifier from a schematic that did not show a power supply. A helpful television repair technician pointed out the problem.

- The op-amp output is shown as a pure voltage source, which implies unlimited output current. In practice, the maximum output current is a few tens of milliamperes, and the source has some internal resistance.

There are other subtle imperfections to op-amp behaviour. These are discussed in section 21. However, the model of figure 244 is sufficient for a great many applications.

12.2 The Op-Amp Comparator

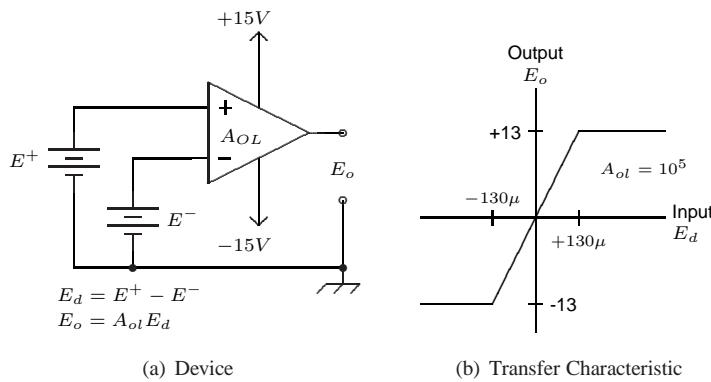


Figure 246: Op-Amp Comparator

The operational amplifier has a typical open-loop gain A_{ol} of 10^5 , so a few millivolts of difference voltage E_d at the input will drive the output to its maximum positive or negative value. The *transfer function* of the open-loop amplifier is shown in figure 246(b). Notice that the input is measured in microvolts and the output in volts.

In figure 246(a), the operational amplifier is operated *open-loop*, ie, without any feedback connection. Consequently, the state of the op-amp output is an indicator of the relative magnitudes of the two inputs. If the op-amp output is high, then the non-inverting terminal is at a greater potential than the inverting terminal. The circuit compares the two voltages and produces an output to indicate which is larger, so it is known as a *comparator*¹²⁹

Application: Zero Crossing Detector

There are circumstances where it is useful to *square up* a sine wave. The comparator circuit is shown in figure 247, along with the input and output waveforms. Anything more than a few millivolts of input voltage e_i will drive the op-amp to full output.

Without the network R_1 , D_1 and D_2 , the voltage at the non-inverting input follows the input voltage. With the network present, it limits the non-inverting input to ± 0.6 volts. This protects the op-amp from damage in the event of a very large input signal.

Although any operational amplifier can be used as a comparator, there is a specific *comparator* integrated circuit which is optimized for this type of operation. It is discussed in section 35.6.

¹²⁹The term *comparator* is also used in the context of a negative feedback system where it means something completely different. In that context, the comparator is a device that compares the desired and actual output signals from a system and produces an error signal that is proportional to the difference between the two. To avoid confusion, we will refer to this latter device as an *adder* or *subtractor*.

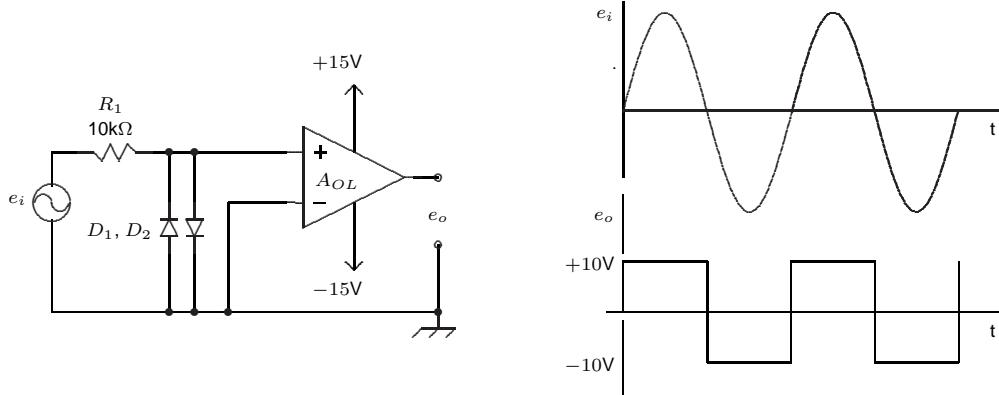


Figure 247: Zero Crossing Detector

12.3 The Buffer

If the feedback network is a wire connecting the output to the summer, then the transfer function \mathbf{B} is simply unity, and the closed-loop gain e_o/e_i is also unity.

$$G = \frac{e_o}{e_i} = \frac{A}{1 + AB} \approx \frac{1}{B} \text{ if } A \gg \frac{1}{B}$$

Also, for a connection between output and input, we have

$$B = \frac{e_f}{e_i} = \frac{e_o}{e_i} = 1, \text{ so } G = 1$$

This circuit is referred to as a *follower*, because the output follows the input. It's also known as a *buffer amplifier* because it *buffers* or *protects* the input from the effect of a load at the output.

The circuit is shown in figure 248. In this scenario, the source contains significant internal resistance R_{int} . If it is connected to the load, the source resistance R_{int} and the load resistance R_l form a voltage divider and the load voltage is some fraction of the input voltage. This is known as the *loading effect*: the voltage drops when a load is connected.

The input resistance of the operational amplifier is effectively infinite (open circuit) and its output resistance close to zero (voltage source). As a result, when the buffer is added between the source circuit and the load there is no loading effect at the input or output, and the source voltage e_i is transferred to the load e_o without any decrease in amplitude.

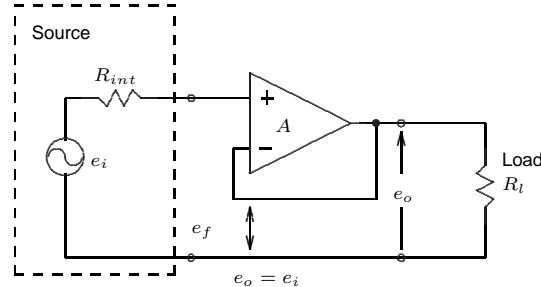


Figure 248: Operational Amplifier Buffer

Limitations

The buffer circuit should be designed with the following limitations in mind:

- Although the resistance looking into the input terminals of the op-amp is effectively infinite, a small amount of *bias current* must flow into both the inverting and non-inverting terminals (see section 21.2). As a result, there must be a DC path for this current into both inputs. If the source is AC coupled, then a bias resistor R_b must be provided as shown in figure 249.
- Although the output resistance of the op-amp is low (especially when enhanced by negative feedback) the output current capability is quite small – the output source of an op-amp typically limits at about $\pm 10\text{mA}$. If a larger current is required, as is often the case, then a special high-power amplifier must be used or a high-current output stage added to the op-amp.

12.4 The Non-Inverting Amplifier

The circuit for a non-inverting amplifier was first introduced in section 10.2 and is repeated here in figure 250. This is how the circuit is usually drawn.

The feedback gain is determined by the ratio of the voltage divider resistors, ie:

$$\begin{aligned} B &= \frac{e_f}{e_o} \\ &= \frac{R_2}{R_1 + R_2} \end{aligned}$$

According to feedback theory, the closed-loop gain G is the reciprocal of the sensor gain, that is:

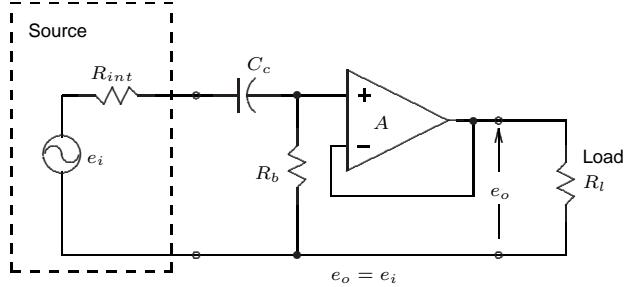


Figure 249: Capacitively Coupled Buffer

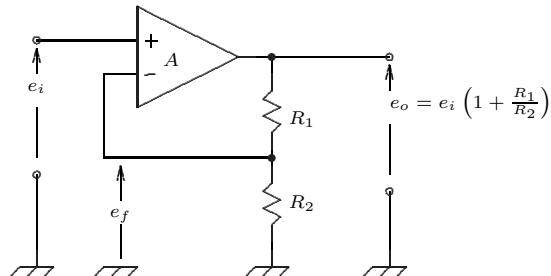


Figure 250: Non-Inverting Amplifier

$$\begin{aligned} G &= \frac{e_o}{e_i} \\ &= \frac{1}{B} \\ &= \frac{R_1 + R_2}{R_2} \\ &= 1 + \frac{R_1}{R_2} \end{aligned} \quad (459)$$

It's possible to use impedances instead of resistances, in which case the resistors are replaced with Z_1 and Z_2 and the amplitude response of the circuit may change with frequency.

The Lever Diagram

Consider the geometric construction of figure 251. Triangle OAB is similar to triangle OCD. Consequently, their sides are in proportion and we can write:

$$\frac{e_i}{R_2} = \frac{e_o}{R_1 + R_2} \quad (460)$$

Rearranging this equation, we can obtain equation 459, the gain equation for the non-inverting amplifier. Consequently, the triangle diagram of figure 251 represents the gain for the non-inverting amplifier.

The triangle diagram of figure 251 can be generalized as the *lever* diagram shown in figure 252, a useful visualization tool for the non-inverting amplifier. The hypotenuse of the triangles is a lever that pivots as the input and output voltage change. The ratio of input and output voltages is determined by the ratio of resistances.

In figure 252, R_1 and R_2 have been made equal, so the closed-loop gain is 2 volts/volt.

If the input voltage is +3 volts, then the output is +6 volts. As the input increases, the sloping line pivots counter-clockwise about the pivot point, increasing the output voltage.

Notice the effect as the ratio of the resistances is varied. If R_2 is made large compared to R_1 , the input scale moves to the right. The gain approaches unity, and it can never be less than unity.

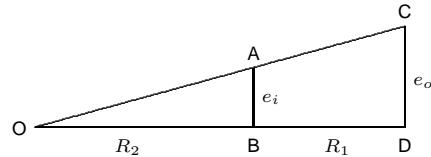


Figure 251: Geometric Construction

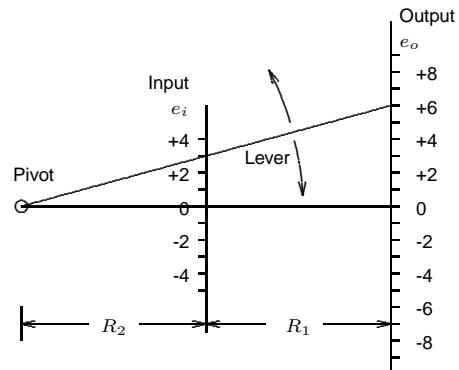


Figure 252: Non-Inverting Amplifier Lever Diagram

12.5 The Inverting Amplifier

Our consideration of the *inverting* amplifier begins with the circuit for a non-inverting amplifier that was first discussed in section 10.2. It is repeated here in figure 253(a).

An interesting thing happens if we move the input signal to another point in the circuit, as shown in figure 253(b).

To understand how this works, we'll describe the physical behaviour. A more rigorous explanation is given as *Alternative Analysis*, below.

- When the circuit is first turned on, the output voltage of the op-amp is zero.
- Consequently, the top end of the R_1 , R_2 voltage divider is at zero volts and the bottom is at e_i volts. This puts the junction of R_1 , R_2 (point B on the diagram) at some positive voltage. This is the feedback voltage.
- The input to the non-inverting input of the op-amp is zero, so the error voltage is equal to the negative of the feedback voltage, so e_e is some finite negative value.
- The op-amp multiplies this negative error signal by its large open-loop gain A and directs it to the output of the op-amp. Consequently, the output voltage e_o takes off in a negative direction.

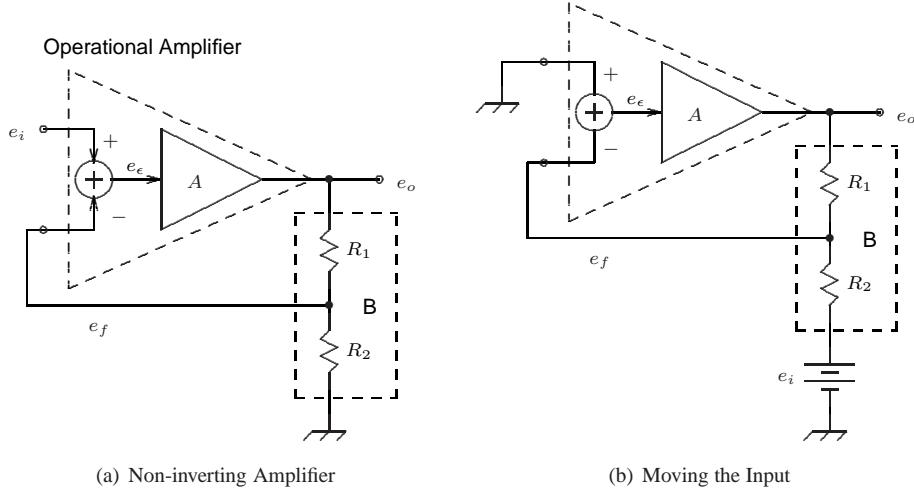


Figure 253: Amplifier Configurations

- This movement of e_o in a negative direction takes the R_1, R_2 voltage divider junction more negative as well.
- The output of the op-amp continues to go negative until the error signal becomes very small (ie, close to zero). At that point, the two signals into the R_1, R_2 voltage divider balance each other. If we assume that the centre point of the divider is at zero volts, then by superposition:

$$+ e_i \frac{R_1}{R_1 + R_2} + e_o \frac{R_2}{R_1 + R_2} = 0 \quad (461)$$

which can be simplified to

$$\frac{e_o}{e_i} = -\frac{R_2}{R_1} \quad (462)$$

To summarize:

- The output voltage has the opposite sign to the input voltage
- The magnitude of the output is determined by the ratio of the two resistors.

The standard method of drawing an inverting amplifier is shown in figure 254. Resistor R_2 is known as R_i , the *input* resistor, and resistor R_1 is known as R_f , the *feedback* resistor.

It's possible to use impedances instead of resistances, in which case these components are Z_i and Z_f and the amplitude response of the circuit will be change with frequency.

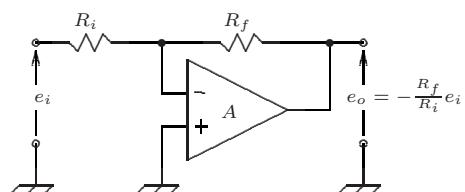


Figure 254: Inverting Amplifier

The Virtual Earth Point

The inverting input of the op-amp is held to within millivolts of zero potential by the negative feedback action of the operational amplifier. Consequently, for all intents and purposes, the inverting terminal may be regarded as being at zero volts. That is often a useful insight into the circuit operation. The inverting input terminal is then referred to as a *virtual earth* or *virtual ground* point¹³⁰.

The inverting input of the op-amp will not be at virtual earth if the feedback action fails for some reason or because the amplifier hasn't had time to respond. For example, if the output is driven against one of the supply rails then the inverting-input terminal will deviate from zero volts.

It's a useful test of the circuit to check the voltage at the virtual earth point. If it is non-zero, then negative feedback has failed.

If the output voltage of the op-amp is not zero, then the voltage across the input terminals, although very small, is not quite zero. Then a virtual earth point is actually at some positive or negative voltage with respect to the non-inverting terminal, that is, at a slightly different voltage than the ground. However, because the op-amp voltage gain is enormous (typically 10^6 volts/volt) the input voltage is small enough to be treated as zero. That said, there are a few situations where it is necessary to take it into account.

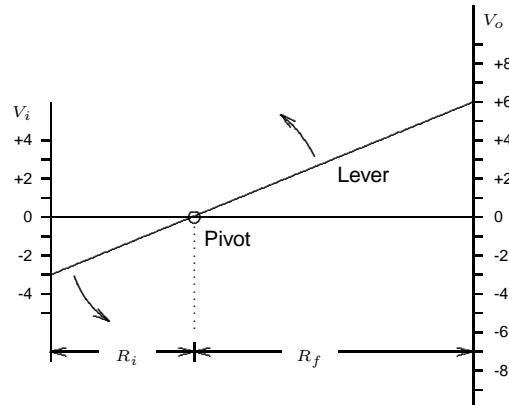


Figure 255: Inverting Amplifier Lever Diagram

The lever diagram for the inverting amplifier follows directly from equation 462 and is shown in figure 255.

In figure 255, the input voltage is -3 volts, the ratio of R_f/R_i is 2, and so the output is $+6$ volts. As the input increases, the sloping line pivots clockwise about the zero point, which is the virtual earth point.

Example Application: Pan-Pot

An interesting application of the current divider is shown in figure 256. The *pan-pot* is a control that routes an input signal in greater or lesser proportion into two output channels, and is commonly used in audio mixing consoles¹³¹.

The normal arrangement for a pan-pot consists of two potentiometer voltage dividers, mechanically linked. They are wired so that one pot increases its output voltage as the other one decreases its output. This works well, but requires two potentiometers. This circuit requires only one.

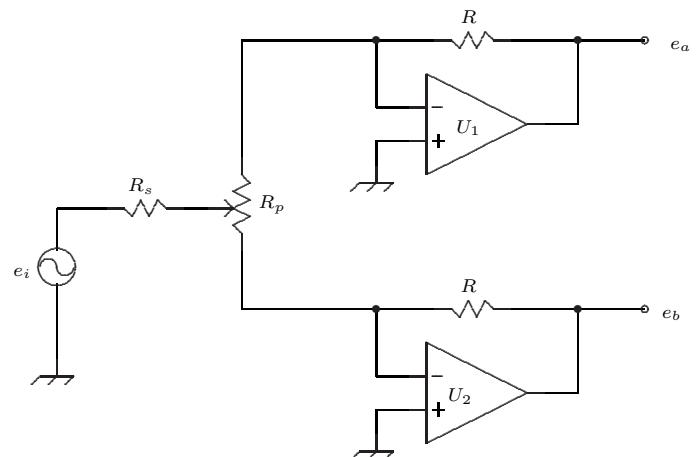


Figure 256: Pan Pot

¹³⁰According to the nomenclature for floating and grounded components in section 2.12, the term *virtual ground* is more correct. We use both terms interchangeably.

¹³¹The name *pan-pot* derives from the expression *panning*, which in this context means to move a source from one side to another. Hence *panning-potentiometer* has become *pan-pot*.

The design is proportioned so that the resistor R_s is at least 10 times the value of the potentiometer, R_p . Then the input circuit e_i and R_s effectively form a current generator that feeds current into the wiper of the pot. Both ends of the potentiometer are held at virtual earth (near zero volts) by the operational amplifiers. Consequently, the input current is very nearly e_i/R_s amps at all times. The input current is proportioned into the input of U_1 or U_2 depending on the position of the pot wiper. If the pot wiper is at its mid-point, then R_p behaves as a balanced current divider. Half the current goes into the virtual earth point of each amplifier, and the outputs are each $e_a = e_b = -e_i/2$. (The operational amplifier inverts the sign of the input.)

Now consider that the wiper is moved to the upper end of its travel. The voltage at both ends of the pot track is held by the op-amps at zero volts. Consequently, there is zero current through the pot track into U_2 , and its output is zero. All the input current flows into the virtual earth of U_1 and its output voltage would be $e_a = -e_i$.

Thus, as the wiper of the pot is moved from one extreme to the other, the input current is directed into one op-amp or the other in proportion to the pot position.

12.6 Formal Analysis of the Inverting Amplifier

To analyse an op-amp circuit, designers generally use the virtual earth technique described in section 12.7 on page 325. The virtual earth analysis is easy to understand and closely tied to the physical behaviour of the circuit.

The analysis shown in this section is a formal approach, a frontal attack on the circuit with basic circuit tools. This approach is used by computer-based circuit analysis tools such as Spice, and only rarely by humans. We include it here only to show that basic circuit tools can be used to analyse the circuit.

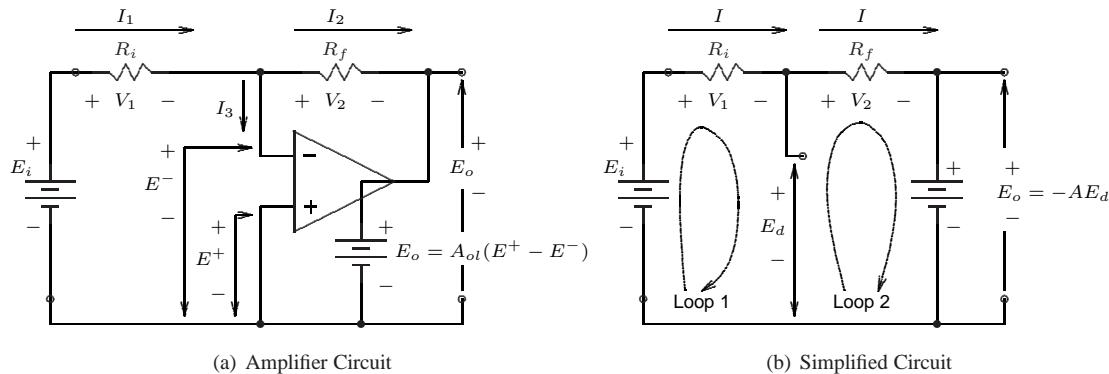


Figure 257: Inverting Amplifier Circuit Model

Starting with the inverting amplifier circuit of figure 254, substitute the op-amp circuit model from figure 244 on page 316. Then we obtain figure 257(a).

We can simplify and modify this circuit in several ways:

- Current does not flow into either input terminal of an op-amp. Consequently I_3 is zero, and by KCL that makes $I_1 = I_2$, which we'll simply call I .
- The voltage E^+ at the non-inverting terminal is evidently zero, so it may be eliminated.
- Consequently the expression $E_o = A(E^+ - E^-)$ can be simplified to $E_o = A(-E^-)$. For notational convenience we'll abbreviate this to $E_o = -AE_d$, where E^- has been shortened to E_d . The name E_d is appropriate because this is the difference voltage between the two input terminals of the op-amp.

- The circuit has the two loops shown in figure 257(b).

These simplifications and changes are incorporated in the circuit as figure 257(b).

Now we will set up the equations for this network and solve for the closed-loop gain e_o/e_i . For Loop 1, by KVL:

$$+ E_i - V_1 - E_d = 0 \quad (463)$$

For Loop 2, by KVL:

$$+ E_d - V_2 - E_o = 0 \quad (464)$$

Apply Ohm's law to the two resistors:

$$V_1 = IR_i \quad (465)$$

$$V_2 = IR_f \quad (466)$$

Finally, we have for the relationship between the input and output voltages of the amplifier:

$$E_o = -E_d A \quad (467)$$

These are the equations we have to work with. Now the algebra begins.

Substitute for E_d from equation 467 into equations 463 and 464.

Combine equations 465 and 466 to obtain:

$$V_1 = \frac{R_i}{R_f} V_2 \quad (468)$$

and use that to eliminate V_1 in equation 463. After some algebraic crank-turning to eliminate V_2 and obtain E_o/E_i , we have:

$$\frac{E_o}{E_i} = -\frac{R_f}{R_i} \left(\frac{A}{1 + \frac{R_f}{R_i} + A} \right) \quad (469)$$

We should be watchful for the expression:

$$\frac{R_i}{R_i + R_f} = B \quad (470)$$

where B is the sensor gain (figure 253 on page 321). Putting these ideas to work, we can manipulate equation 469 as follows:

$$\begin{aligned} \frac{E_o}{E_i} &= -\frac{R_f}{R_i} \left(\frac{A}{\frac{1}{B} + A} \right) \\ &= -\frac{R_f}{R_i} \left(\frac{AB}{1 + AB} \right) \end{aligned} \quad (471)$$

If the loop gain AB is much larger than 1, (or what is the same thing, $A \gg 1/B$) then the term in the brackets approximates to 1 and we obtain the previous result:

$$\frac{E_o}{E_i} = -\frac{R_f}{R_i} \quad (472)$$

In a later section (21.1 on page 640), we'll explore the implications of insufficient op-amp gain. For now, it's sufficient to understand that the loop gain AB must be large compared to 1 for the approximate relationship of equation 472 to be correct.

12.7 The Rules of Op-Amp Behaviour

It is entirely possible to use Kirchhoff's Voltage and Current Laws (Sections 2.9 and 2.10) together with a circuit model of the op-amp (figure 219 on page 287) to analyse any operational amplifier circuit. However, there are certain rules that can greatly simplify understanding and analysis.

If the circuit is known to use negative feedback, then

1. The inverting and non-inverting terminals must be at the same voltage. If the circuit is an inverting amplifier configuration and the non-inverting terminal is grounded, then the inverting terminal must be at zero volts, a so-called *virtual earth* or *virtual ground*.
2. No current flows into the input terminals of the op-amp.
3. The output of the op-amp is a perfect voltage supply.

Caveats and Elaborations

1. Actually, the input terminals are at very slightly different voltages – by an amount equal to the output voltage of the op-amp divided by its open loop gain. Usually this is small enough to be ignored.
2. To be precise, the amount of *signal current* flowing into the input terminals is small enough to be ignored. This is a consequence of the very large *differential input resistance* between the two input terminals. This is usually much larger than any practical external resistance, so it may be ignored.
However, keep in mind that there must be a path for the DC bias current into each terminal of the op-amp. This is modelled by a current generator between each input terminal and ground. The incremental resistance of a current source is infinite and it appears as an open circuit to the input and feedback signals, so for the purposes of the small signal analysis it may be ignored. (For more on bias current, see section 21.2.)
3. While the output resistance is usually low enough to be ignored, the available output current is limited to a few tens of milliamperes. Beyond that, the output voltage source goes into current limit.

Example: Inverting Op-Amp

The inverting amplifier circuit of figure 258 is very easily analysed with these rules.

- By Rule 1, the non-inverting terminal is grounded, so the inverting terminal must be at zero volts (*virtual earth* point).
- Consequently, the voltage across R_i is e_i and the current through it is

$$i_i = \frac{e_i}{R_i} \quad (473)$$

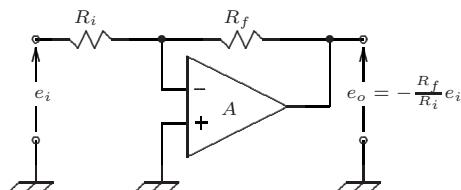


Figure 258: Inverting Amplifier

- By Rule 2, no current flows into the op-amp input terminals. Consequently, the current must flow through R_f . The current through R_f sets up a voltage across R_f of:

$$\begin{aligned} V_f &= i_i R_f \\ &= \frac{e_i}{R_i} R_f \end{aligned} \quad (474)$$

- By rule 1, the left end of R_f is at zero volts, so the right end, which is the output voltage e_o , is at $-V_f$, so

$$\begin{aligned} e_o &= -\frac{e_i}{R_i} R_f \\ \frac{e_o}{e_i} &= -\frac{R_f}{R_i} \end{aligned} \quad (475)$$

This is the same equation that was obtained by formal analysis in section 12.5 for the inverting amplifier. As you can see, this process is much simpler.

12.8 Single Supply Operation

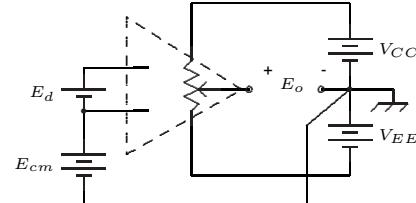
The output circuit of the operational amplifier can be represented as a potentiometer as shown in figure 259 (see section 12.1). In figure 259, the op-amp is powered from a split power supply and the ground reference point is taken as the tap between the two power supplies. The output of the op-amp is taken between the wiper of the potentiometer and the ground reference point. When the wiper is at its mid-point, the output voltage is equal to the ground reference voltage, or zero. When the wiper is at the top of its travel, the output voltage is at its maximum positive value. When the wiper is at the bottom, the output voltage is at its maximum negative value.

This arrangement provides for both positive and negative output voltages, which is useful in many systems. (Imagine an analog computing circuit where the output voltage represents some quantity. This arrangement allows positive and negative quantities and provides a symmetrical number line with zero at the centre.)

The split power supply arrangement shown in figure 259 requires a positive and a negative power supply. There are many systems where it is inconvenient to provide the negative power supply. For example, a predominantly digital system may be powered by a single 5 volt or 3.3 volt supply. It is possible to derive other supplies from these voltages, but it would be convenient not to have to do so. Single-supply operation of the op-amp is shown in figure 260.

Because there is only one supply and because it is much lower than 15 volts, the realities of the amplifier behaviour intrude more strongly into the design. One must be much more careful about how the problem maps onto the circuitry to ensure that the signal does not exceed the input or output limitations of the amplifier.

For example, ideally the common mode voltage E_{cm} has no effect on amplifier behaviour: the output voltage is solely dependent on the differential input voltage E_d and the amplifier voltage gain. In practice, E_{cm} must be somewhere between the positive supply rail V_{CC} and ground or the amplifier simply will not function. (In fact, in some cases the gain reverses sign when the common mode voltage exceeds the supply voltage! As you can imagine, this is problematic in a negative feedback system.) It is not unusual for the common mode voltage to be constrained even more: between 0V and $V_{CC} - 1.4V$, for example.



$$E_o = A_{ol} E_d$$

Figure 259: Split Supply Operation

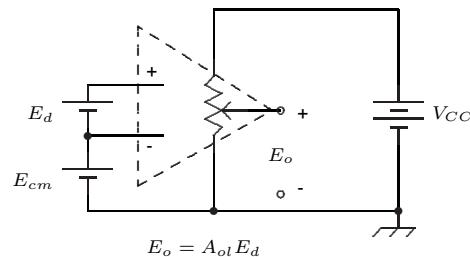
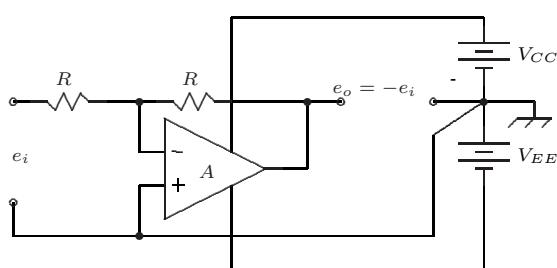


Figure 260: Single Supply Operation

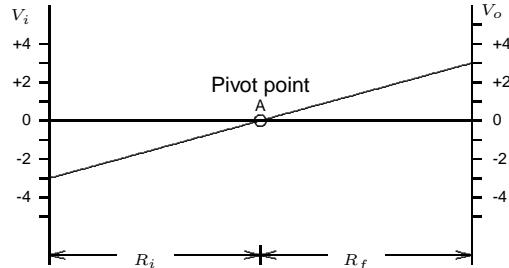
Furthermore, the output voltage should ideally be able to range between zero volts and the positive supply V_{CC} . This may be the case (in so called *rail-to-rail* op-amps) or the output voltage may be constrained to something like 0V to $V_{CC} - 1.4V$.

The moral is this: the dynamic range of the signal, between its most positive and negative excursions, must be constrained to be within the physical limitations of the chosen device.

The Pseudo-Ground



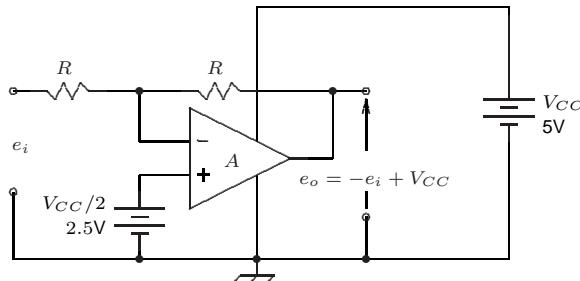
(a) Split Supply Inverting Amplifier



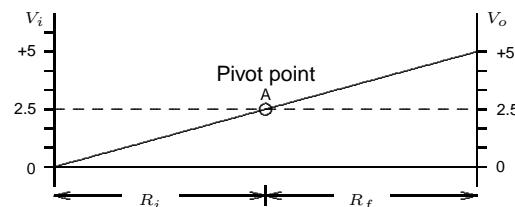
(b) Lever Diagram

Figure 261: Inverting Amplifier, Split Supply

A split supply inverting amplifier and its lever diagram are shown in figure 261. For simplicity, we assume that both resistors are equal to R and $G = -1$. As the input varies between $+/- 2.5V$, the output varies between $-/+ 2.5V$ and the diagonal trace rotates around the pivot point A. Both the input and output signals can be positive or negative.



(a) Single Supply Inverting Amplifier



(b) Lever Diagram

Figure 262: Inverting Amplifier, Single Supply

The single-supply version of this amplifier is shown in figure 262. We'll assume that the gain $G = -1V/V$ as before and the output voltage varies between $+5V$ and $0V$. Then it can be shown (see section 13.8) that the non-inverting terminal of the op-amp must be biased to a potential of $+2.5V$. This potential might be supplied by a voltage divider from the V_{CC} supply. As a result, an input that ranges between 0 to $+5$ volts generates an output between $+5$ and zero volts.

When there are a number of amplifiers in a single supply system, many points in the circuit may need to be offset by an amount equal to half the supply voltage. This potential serves the same purpose as the ground terminal in a split supply system, so it is known as the *pseudo-ground* point in the single supply system.

It's important that this point presents a low impedance to its loads, so it may be buffered by an operational amplifier. An example is shown in figure 263. The supply is split by the equal resistors R and then buffered by op-amp follower A_2 .

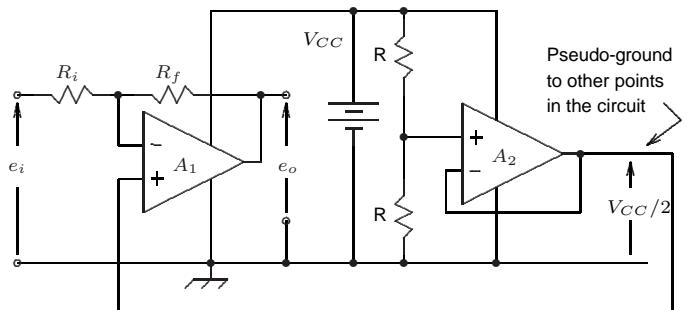


Figure 263: Pseudo Ground

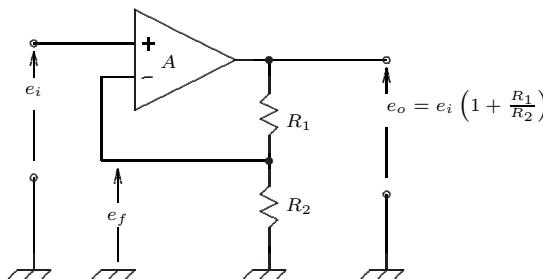
Reference

Single-supply op-amp circuits are described in Marsh [91], Mancini [92] and Carter [93]. Single-supply op-amp circuits specifically intended for audio applications are shown in section 13.18.

12.9 Exercises

Where an op-amp power supply is not shown, assume it is $\pm 15V$.

- For each of the following op-amp circuits, determine the output voltage E_o . The op-amp may be considered to be ideal. The output can swing between the op-amp power supply rails.
 - Using the op-amp rules of behaviour, analyse the non-inverting amplifier circuit shown in the figure. Show that the expression for voltage gain is the same as that obtained in section 12.4.



3. Draw each of the following circuits:
 - (a) Inverting Amp of gain -4V/V.
 - (b) Non-Inverting Amp of gain +3V/V
 4. Lever diagrams:
 - (a) Draw the lever diagram for an inverting op-amp circuit when $R_i = 10k\Omega$, $R_f = 15k\Omega$ and the input voltage is +2 volts.
 - (b) Draw the lever diagram for a non-inverting op-amp circuit when the gain is +3V/V and the input voltage is +2 volts.

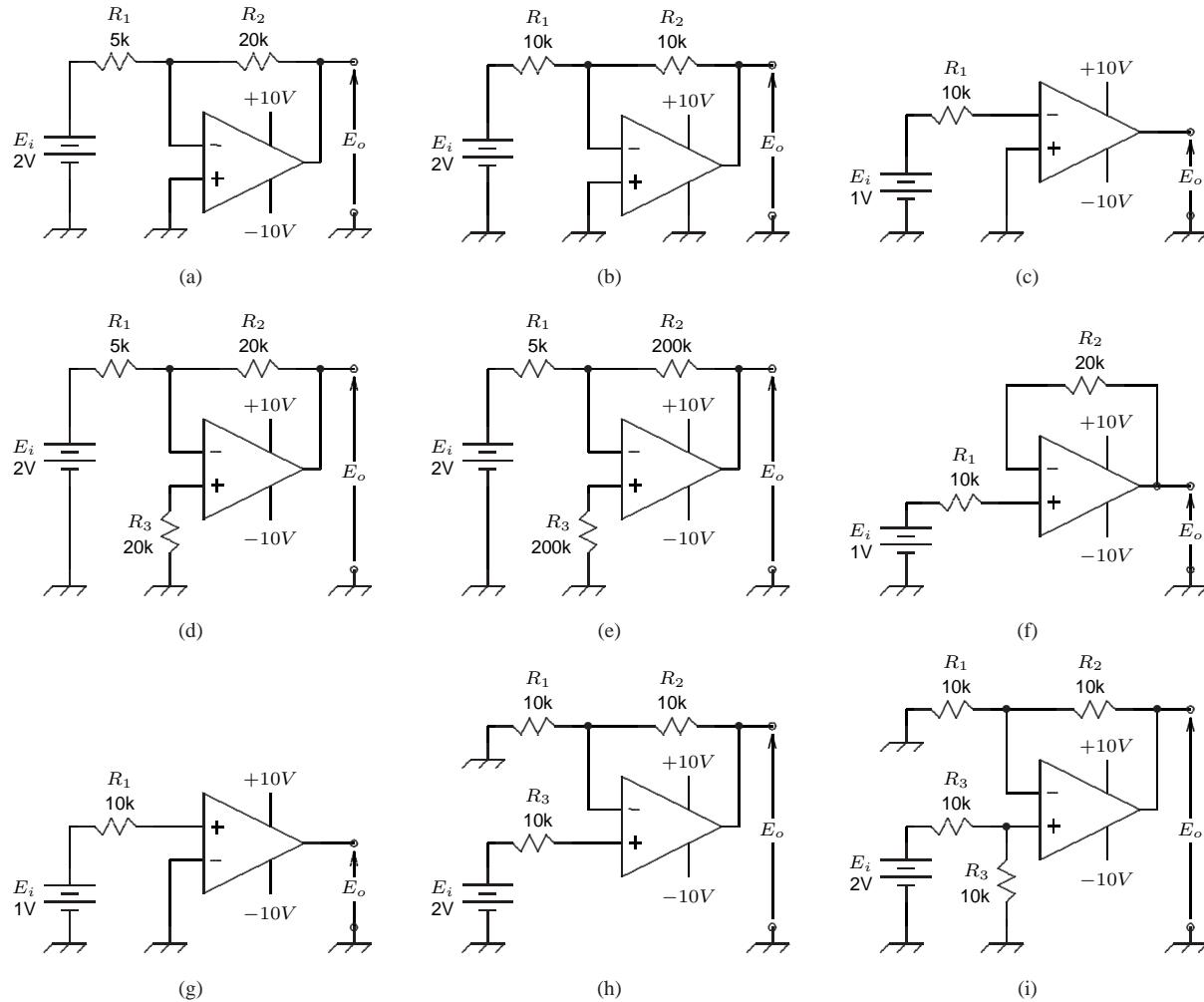
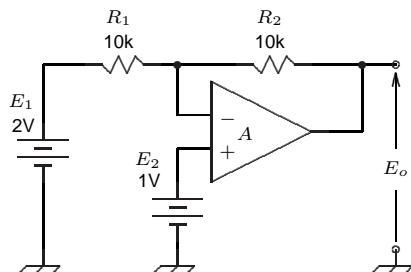


Figure 264: Operational Amplifier Circuits

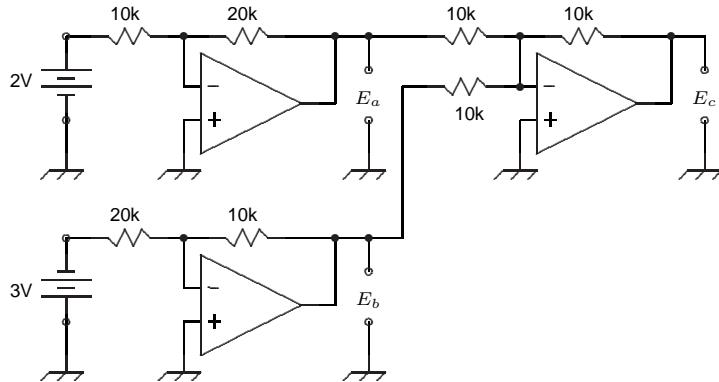
5. Use a lever diagram to show that the voltage gain for a non-inverting amplifier cannot be less than unity.
6. For the circuit shown:



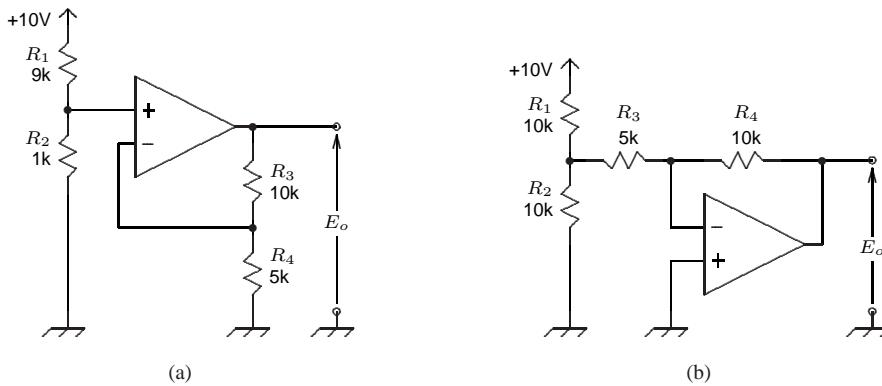
- (a) Draw the lever diagram and estimate the output voltage E_o .

(b) Confirm the value of E_o using the superposition theorem.

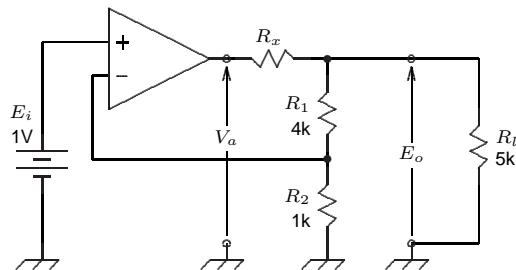
7. For the op-amp circuit shown, determine voltages E_a , E_b and E_c . The op-amps are ideal.



8. Find the output voltage E_o in each case.



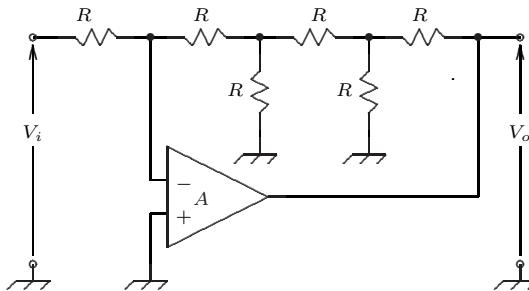
9. For the op-amp circuit shown,



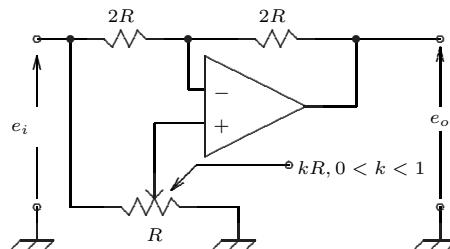
(a) For $R_x = 1\text{k}$, determine voltages V_a and E_o .

(b) Repeat for $R_x = 2\text{k}$.

- (c) If the op-amp is powered by $\pm 15V$ power supplies and the output of the op-amp can swing rail-rail, what is the maximum allowable value of R_x ?
10. For the op-amp circuit shown,
-
- (a) Sketch with scale factors the transfer characteristic v_o versus e_i of this circuit over the range of $-10V < e_i < +10V$.
- (b) Sketch to scale the waveforms of e_i and v_o if the input is a triangular wave with peak voltage of 10 volts and frequency 100 Hz.
11. For the op-amp circuit shown, use the rules of op-amp behaviour to determine V_o/V_i . (Hint: trace the currents, starting at the input. Simultaneous equations are not necessary.)

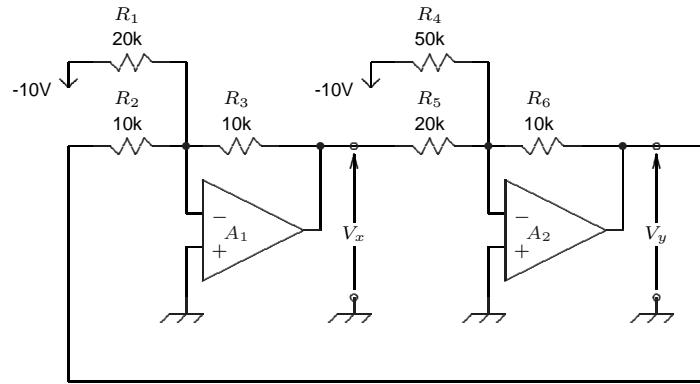


12. In the circuit shown below, the potentiometer has a resistance of R ohms. The wiper of the potentiometer is positioned so that the right half of the potentiometer is kR ohms, where $0 < k < 1$.



- (a) Using superposition, determine an expression for the output voltage e_o in terms of the input voltage e_i and the potentiometer position k .

- (b) Determine the output voltage e_o in terms of the input voltage e_i when:
- The potentiometer is at its left-most setting.
 - The potentiometer is in the middle.
 - The potentiometer is at its right-most setting.
13. For the circuit shown, determine the values V_x and V_y .



For this circuit not to lock up with the amplifiers at their bounds, the voltage gain around the entire loop (for example, from the output of the first op-amp back to that same point) must be less than unity. Verify that this is the case.

14. Design an op-amp circuit to implement the PID controller of section 11.7, page 310.

13 The Operational Amplifier: Applications

13.1 Miller Effect

The Miller Effect [94] provides us with a very useful tool for the simplification of many op-amp and discrete circuits. Consider the circuit shown in figure 265¹³².

A resistance R_f is connected between the inverting input and the output of a high-gain amplifier.

Miller's Theorem says that this combination of amplifier and resistance appears to the input source as a resistance r_i which is A times smaller than the feedback resistor R_f :

$$r_i \approx \frac{R_f}{A} \quad (476)$$

This is easy to show. Referring to the circuit, the voltage at the left end of the resistor is e_i . The voltage at the right end is $-e_i A$. So the total voltage across the feedback resistor is

$$\begin{aligned} v_f &= e_i - (-e_i A) \\ &= e_i(1 + A) \end{aligned}$$

Since (almost) no signal current flows into the input terminal of the op-amp, the current i_f through R_f is equal to the input current i_i , so

$$\begin{aligned} i_i &= i_f \\ &= \frac{e_i(1 + A)}{R_f} \end{aligned}$$

The input resistance seen from the source is then

$$r_i = \frac{e_i}{i_i} \quad (477)$$

$$= \frac{R_f}{1 + A} \quad (478)$$

which may be approximated as:

$$r_i \approx \frac{R_f}{A} \text{ for } A \gg 1 \quad (479)$$

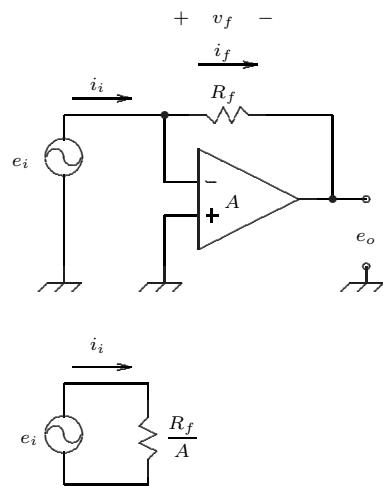


Figure 265: Miller Effect

¹³²This circuit isn't very practical because the value of e_i must be microscopic to not overdrive the amplifier into saturation. There are other possible circuits that can be used to illustrate the Miller Effect, but this one has the virtue of a pristine simplicity.

Example: The Inverting Amplifier Re-analysed

Now we'll apply this idea to the analysis of the inverting amplifier. The inverting op-amp configuration is shown in figure 266(a). From the input, the op-amp and its feedback resistance are equivalent to the circuit shown in figure 266(b).

Treating the input circuit as a voltage divider gives us a method of calculating the input error voltage e_ϵ :

$$e_\epsilon = e_i \cdot \frac{R_f/A}{R_f/A + R_i} \quad (480)$$

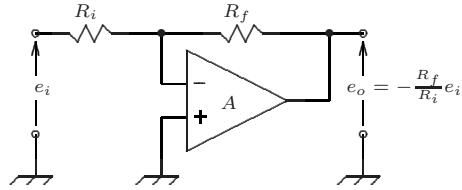
where A is the open-loop gain. Because A is large, then R_f/A is small and it is safe to assume that R_i is much larger than R_f/A . Then equation 480 can be written:

$$e_\epsilon = e_i \cdot \frac{R_f/A}{R_i} \quad (481)$$

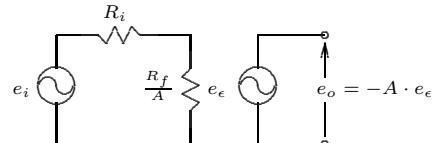
The output voltage is $(-A)$ times the error voltage:

$$\begin{aligned} e_o &= -A \left(e_i \cdot \frac{R_f/A}{R_i} \right) \\ &= -e_i \cdot \frac{R_f}{R_i} \\ G &= \frac{e_o}{e_i} \\ &= -\frac{R_f}{R_i} \end{aligned} \quad (482)$$

Equation 483 is the same result as we obtained previously in section 12.5 but it's a little more useful. Equation 481 shows that the error voltage is an inverse function of the open-loop voltage gain of the amplifier: as A increases, the error voltage e_ϵ decreases. Furthermore, if we wanted a more accurate equation for closed-loop voltage gain G that did not assume a large value of open-loop gain A , this method of analysis would provide it.



(a) Inverting Amplifier



(b) Equivalent Circuit

Figure 266: Applying the Miller Effect

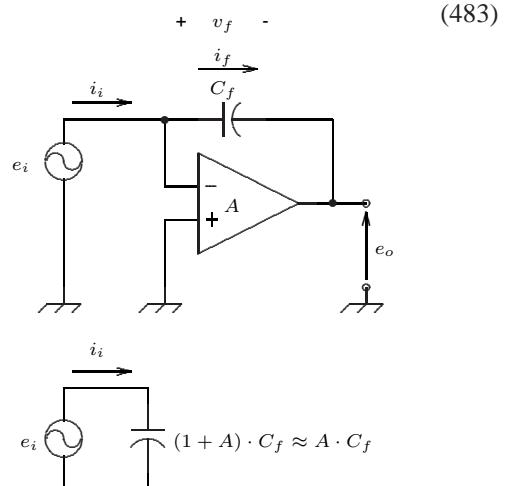


Figure 267: Miller Effect with Capacitance

Feedback Capacitance

If the feedback resistance is replaced with an impedance Z_f , then equation 478 becomes:

$$z_i = \frac{Z_f}{1 + A} \quad (484)$$

Now suppose that the feedback impedance Z_f is a capacitance C_f . The impedance of the capacitance is

$$\begin{aligned} Z_f &= X_c \\ &= \frac{1}{sC_f} \end{aligned}$$

Substitute for Z_f in equation 484 and we have

$$\begin{aligned} z_i &= \frac{1}{(1 + A)sC_f} \\ &= \frac{1}{s(1 + A)C_f} \end{aligned} \quad (485)$$

The consequences are shown in figure 267. The feedback capacitance C_f is equivalent to a capacitance $1 + A$ times as large across the input terminals. The effective value of the input capacitance is often referred to as the *Miller Capacitance*.

Miller Capacitance

Miller capacitance is an important factor limiting the high frequency performance of amplifiers. Figure 268(a) depicts a common scenario. An inverting amplifier of gain A has a feedback capacitance C_s . This capacitance is created by the unavoidable stray capacitance that is part of a device equivalent circuit, such as the collector-base capacitance C_{cb} of a BJT.

This amplifier circuit is driven by a source with an open-circuit voltage e_s and an internal source resistance R_s .

As shown in figure 268(b), the Miller effect creates an effective input capacitance to the amplifier that is A times the stray feedback capacitance. With the source resistance R_s this capacitance forms a lowpass RC filter (figure 120(a) on page 171) with a cutoff frequency given by:

$$\omega_o = \frac{1}{R_s A C_s} \text{ radians/sec} \quad (486)$$

The open-loop gain in the denominator of this equation has the effect of substantially reducing the cutoff frequency. The design of a high frequency amplifier must minimize this effect.

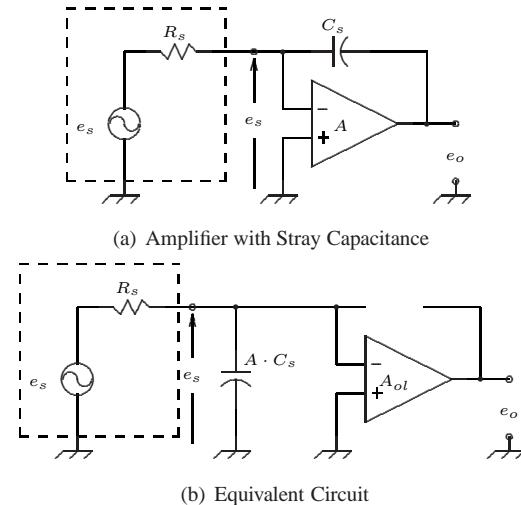


Figure 268: Miller Capacitance and High Frequency Response

13.2 The Op-Amp Adder

An operational amplifier *adder* is shown in figure 269. As indicated in the figure, when all the input resistors are equal, the result is the negative of the sum of the inputs.

This result may be determined as follows:

- Each of the input resistors is connected to a voltage e_i at the left end, and a virtual earth point (zero volts) at the right end.
- So the currents through the three resistors are each

$$i_i = e_i / R \quad (487)$$

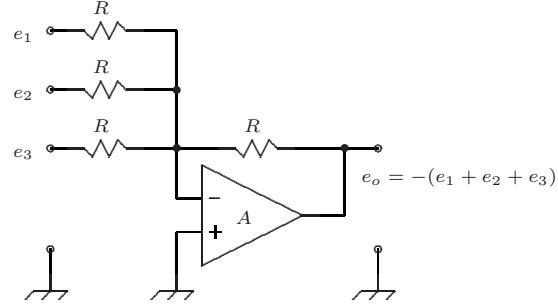


Figure 269: Op-Amp Adder

- These currents sum at the virtual earth point. They cannot flow into the inverting input terminal of the op-amp, so they must flow through the feedback resistor. Consequently, the current through the feedback resistor is

$$i_f = i_1 + i_2 + i_3 \quad (488)$$

- This current sets up a voltage across the feedback resistor of

$$v_f = R(i_1 + i_2 + i_3) \quad (489)$$

- Since the left end of the feedback resistor is held at zero volts by the op-amp (it's the virtual earth point), then the right end of the feedback resistor must be

$$e_o = -v_f \quad (490)$$

Combining equations 487 through 490, we obtain the result shown in equation 491.

$$e_o = -(e_1 + e_2 + e_3) \quad (491)$$

Offset Generator

It is often necessary to offset some signal by a fixed amount, either positive or negative. This is equivalent to adding a fixed voltage to the signal, and may be accomplished by connecting one adder input to a potentiometer, as shown in figure 270.

If V_{CC} is positive and V_{EE} negative, then the output may be shifted to have a DC offset that is either positive or negative.

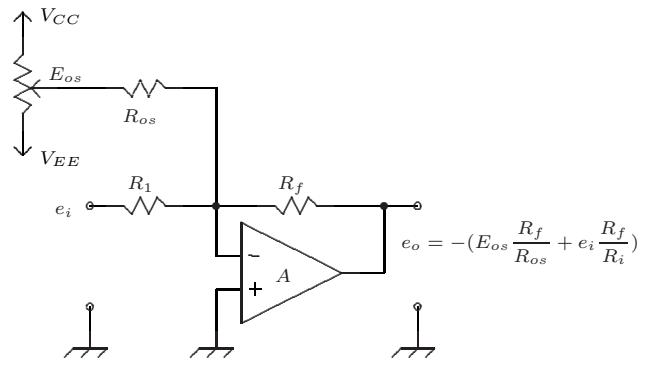


Figure 270: Offset-generator

13.3 Non-Inverting Adder

The traditional operational amplifier adder circuit is inverting: for a positive input voltage, the output is negative. This is a problem in systems where the op-amps are powered from a single supply voltage because the output of an amplifier cannot go negative.

A non-inverting amplifier circuit solves this problem. A circuit is shown in figure 271.

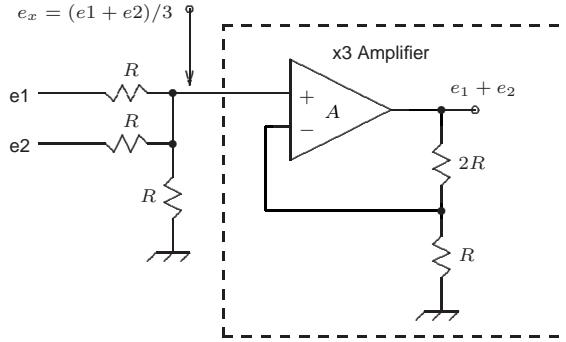


Figure 271: Non-Inverting Adder

By superposition, the output from the three-resistor network is

$$e_x = \frac{e_1 + e_2}{3} \quad (492)$$

Consequently, we could describe the three-resistor network as an adder with an overall gain constant of 1/3. In other words, there is a net loss of signal through the adder (see section 3.7).

This is followed by a non-inverting amplifier of gain 3 (see section 12.4). That is:

$$\frac{e_o}{e_x} = \frac{R + 2R}{R} = 3 \text{ V/V}$$

The passive resistor adder gain of 1/3 is cancelled by the non-inverting amplifier gain of 3, and the output of the amplifier is simply the sum of e_1 and e_2 .

This circuit can be extended beyond two inputs by adding resistors to the input network and increasing the gain of the non-inverting amplifier.

13.4 Sign Changer (Inverter)

If we have a *sign-changer* circuit, then we can fabricate a subtractor by changing the sign of the signal being subtracted, and adding. That is,

$$e_a - e_b = e_a + (-e_b) \quad (493)$$

A sign-changer is simply an inverting amplifier with a gain of unity, figure 272. When R_f and R_i are equal, the output voltage is the negative of the input voltage:

$$e_o = -e_i \quad (494)$$

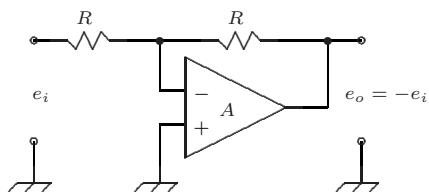


Figure 272: Op-Amp Sign Changer

This circuit is sometimes described as imparting a 180° phase shift to the signal, because an inverted sine wave is indistinguishable from one that is phase shifted 180° . However, this is not true of an asymmetrical waveform, and the circuit should be referred to as a sign changer or inverter¹³³. A better name would be *negator*, but we're stuck with inverter.

13.5 Switchable Sign Changer

Some electronics systems require a *switchable inverter*, that is, one that can be changed from a gain of -1 to $+1$ under control of an electronic signal. The function generator of figure 547 and the phase-sensitive-rectifier of section 16.5 both use this function.

The circuit of figure 273 accomplishes this very nicely. When switch S is closed, the non-inverting terminal of the op-amp is grounded and the circuit behaves as a sign-changer. The input voltage is inverted to produce the output and the gain is -1 .

When the switch is opened, there is no current through R_2 , so the non-inverting terminal is at the same potential as e_i . By negative feedback, the inverting terminal must also be at e_i . Then there is no voltage across the left resistor R_1 , no current through it, and therefore none through the right resistor R_1 . Consequently, the output voltage is the same as the input, and the gain is $+1$.

In practice, switch S is an electronic switch such as a BJT, JFET or analog gate, and the polarity can be switched at electronic rates.

The resistors marked R_1 must be equal. R_2 simply limits the current through the switch, so its value is not important. It's convenient but not necessary to make it equal to R_1 .

In some applications, it's convenient to regard this circuit as a multiplier, one that multiplies the input by $+1$ or -1 , depending on the switch control signal.

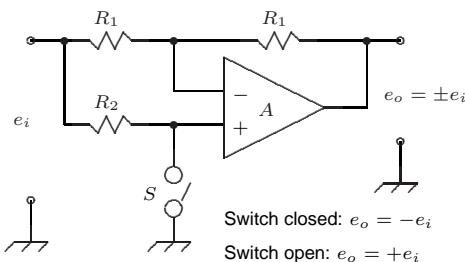


Figure 273: Op-Amp Switchable Sign Changer

13.6 The Subtractor

This section shows three different methods of subtracting electrical signals using op-amps. Each method has various advantages and disadvantages.

Negate and Add

A subtractor using a sign-changer is shown in figure 274. Op-amp A_1 changes the sign of input E_b and A_2 does the adding.

Then

$$\begin{aligned} E_o &= -(E_a + (-E_b)) \\ &= E_b - E_a \end{aligned}$$

If E_b is a positive voltage, then the output from A_1 is a negative voltage. This is not a

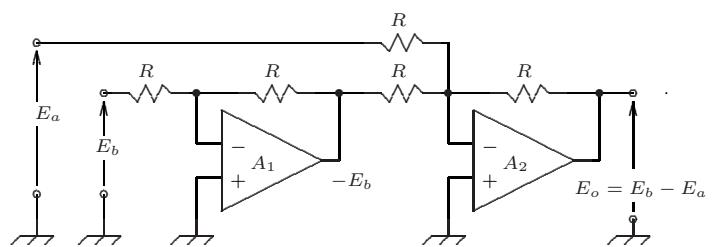


Figure 274: Op-Amp Subtractor #1

¹³³Even the term *inverse* is potentially misleading. In this context it means *flip the waveform upside down*. It doesn't mean *change the signal into its reciprocal*.

problem if the op-amps in this circuit are operated from a bipolar power supply. However, if the circuit is to be operated from a single positive power supply, then a negative output voltage is not possible.

The Difference Amplifier

A complete subtractor (or *difference amplifier* as it is more commonly known) can be created with one op-amp and some resistors, as shown in figure 275.

To understand the operation of this circuit, apply the superposition theorem:

- Assume that E_b is short-circuited. Then the output is

$$E'_o = -E_a \quad (495)$$

- Assume that E_a is short-circuited. Input voltage E_b drives an equal-valued voltage divider, so the voltage at the non-inverting input is $E_b/2$. The op-amp and two resistors constitute a non-inverting amplifier of gain $\times 2$. This gain cancels the gain of the voltage divider, so that

$$E''_o = E_b \quad (496)$$

- The total voltage is

$$\begin{aligned} E_o &= E'_o + E''_o \\ &= E_b - E_a \end{aligned}$$

If E_a , E_b and E_o are all positive, there are no negative voltages in this circuit and it could be operated from a single power supply.

This circuit must be driven from low impedance sources, ie, sources where the internal resistance is much smaller than R . Any source resistance affects the different inputs differently, effectively unbalancing the amplifier so that it does not subtract properly.

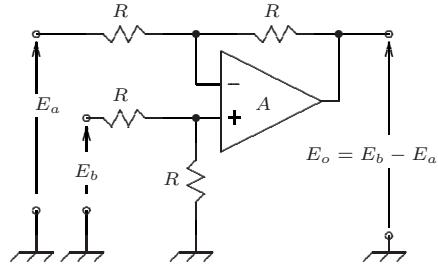


Figure 275: Difference Amplifier

High Impedance Inputs

The previous circuits have input impedance in the order of R ohms. An alternative circuit, which has very high input impedance, is shown in figure 276. Superposition can help us understand this circuit as well.

- First, consider that E_b is shorted and find the output voltage due to E_a . The first op-amp is a non-inverting amplifier of gain $\times 2$:

$$\begin{aligned} G &= 1 + \frac{R}{R} \\ &= 2 \text{ V/V} \end{aligned}$$

So the output of A_1 is

$$E_x = 2E_a \quad (497)$$

This signal is inverted by the following amplifier A_2 , so that

$$E'_o = -2E_a \quad (498)$$

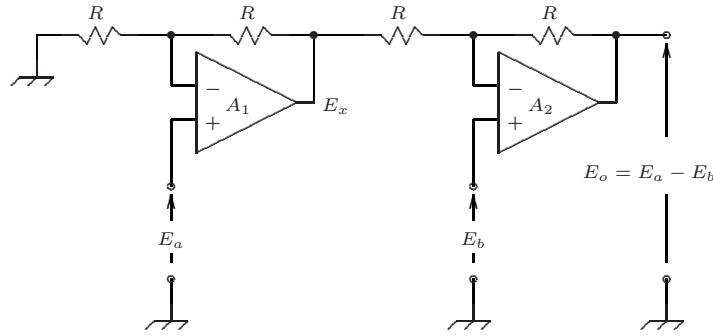


Figure 276: High Impedance Input Diff Amp

- Now consider that E_a is short-circuited and find the output voltage due to E_b . The value of E_x is now zero, and amplifier A_2 is a non-inverting amplifier of gain $\times 2$, so

$$E''_o = +2E_b \quad (499)$$

- The net result is

$$\begin{aligned} E_o &= E'_o + E''_o \\ &= 2(E_b - E_a) \end{aligned}$$

This circuit has very high input impedance for both inputs and can operate from a single supply as long as the result is a positive voltage (ie, $E_b \geq E_a$).

But it does have limitations. The maximum input voltage, into either op-amp, is limited by the operational amplifiers themselves. (This is not true for the circuit of figure 274, where the input terminals of the op-amp are held at zero volts and the input voltage can be very large.)

As well, the sources that produce E_a and E_b must be able to provide the bias currents for the op-amp inputs. The input circuits could not be capacitively coupled for example, unless a separate resistor was connected between each input and ground to provide a path for the bias current.

Because each input is multiplied by 2, one should be careful that the input signals do not overdrive the amplifiers into saturation.

Subtractor Accuracy

When the same signal is applied to both inputs of a subtractor, the output voltage should be zero. This is a good test of its accuracy, and requires good matching from the resistors. For example, in the case of the *negate-and-add* method of figure 274, this requires that the gain for the E_a input through to the output be precisely equal to the gain for the E_b input. In the case of the differential amplifier, it requires that the 4 resistors be matched.

While it is possible to build in gain for one or both of the inputs to these subtractor circuits, it's usually a better approach to build the subtractor with matched resistors. Their exact value is not important, but they should be matched to each other. This is relatively easy to achieve when all resistors have the same value. Gain should be provided in a separate op-amp stage.

13.7 The Instrumentation Amplifier

The difference amplifier of section 13.6 can be enhanced with the front-end circuitry shown in figure 277.

The input signal has been characterised as a common-mode signal E_{CM} , which is the average voltage into the two inputs, with a differential component, E_D . Ideally, the instrumentation amp should ignore the common-mode component and amplify the differential component.

The front end of the amplifier uses two operational amplifiers, A_1 and A_2 . Their responsibility is to present a high impedance to the input signal, some voltage gain, and a low output impedance to the subtractor circuitry around A_3 . The subtractor uses four equal-value resistors. This stage rejects the common-mode signal, so the resistors must be carefully matched.

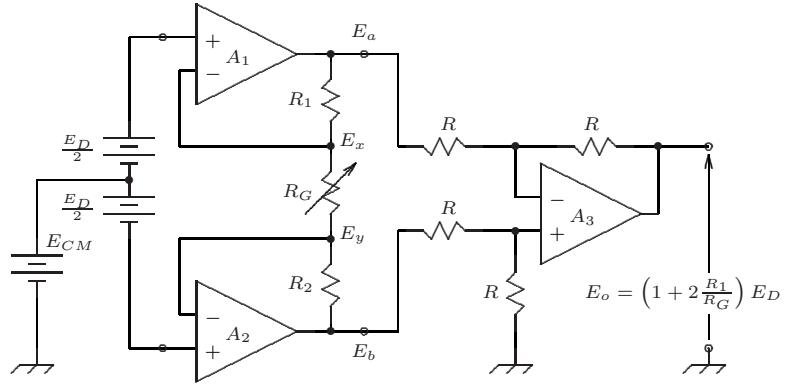


Figure 277: Instrumentation Amplifier

Analysis

The circuit is best understood using superposition. To start with, consider that the common-mode signal E_{CM} is zero.

The op-amps A_1 and A_2 are operating with negative feedback. Consequently the voltages at the input terminals of A_1 are equal, and the voltage at the top end of R_G may be determined to be

$$E_x = E_D/2 \quad (500)$$

Similarly, the voltage at the bottom end of R_G is

$$E_y = -E_D/2 \quad (501)$$

Then the voltage across R_G is simply E_D .

The current through R_G is

$$I_{R_D} = \frac{E_D}{R_G} \quad (502)$$

No current flows into the terminals of A_1 or A_2 , so that same current must flow through both R_1 and R_2 .

Then

$$\begin{aligned} E_a &= E_x + \left(R_1 \frac{E_D}{R_G} \right) \\ &= +\frac{E_D}{2} + \left(R_1 \frac{E_D}{R_G} \right) \end{aligned}$$

and

$$E_b = E_y - \left(R_2 \frac{E_D}{R_G} \right) \quad (503)$$

$$= -\frac{E_D}{2} - \left(R_2 \frac{E_D}{R_G} \right) \quad (504)$$

$$(505)$$

It's not necessary, but it simplifies the math to make R_2 equal to R_1 . Then the voltage at the output of the difference amplifier due to the difference signal E_D is then

$$E_{od} = E_A - E_B \quad (506)$$

$$= E_D \left(1 + 2 \frac{R_1}{R_G} \right) \quad (507)$$

That is, the gain for the differential signal E_D is

$$\frac{E_{od}}{E_D} = 1 + 2 \left(\frac{R_1}{R_G} \right) \quad (508)$$

Now consider that the differential input signal E_D is zero and determine the effect of the common-mode input signal E_{CM} :

The voltage at the top end of R_G will be E_{CM} and the voltage at the bottom end of R_G will be E_{CM} . Consequently, there will be no current through R_G . Then $E_A = E_{CM}$, $E_B = E_{CM}$ and both input terminals of the difference amplifier are at E_{CM} . The difference amplifier will then produce a zero result for this signal, rejecting the common-mode input.

$$\frac{E_{ocm}}{E_{CM}} = 0 \quad (509)$$

Summary

In its favour, this circuit

- Has high input impedance, so the internal resistance of the source is not important
- Can be adjusted for gain by one resistance, R_G
- Has common-mode rejection dependent on four equal-valued resistors R .

On the other hand,

- the amplifiers A_1 and A_2 limit the allowable range for the common-mode input voltage
- the maximum output voltages of amplifiers A_1 and A_2 limit the maximum values of the internal voltages E_A and E_B . If the gain is very large for differential signals ($\times 1000$, for example) then the output of A_1 or A_2 may saturate with the combination of the common-mode signal and the amplified differential signal.

In practice, where the instrumentation amplifier must reject a large common-mode input signal and provide high gain to the differential signal, there are special instrumentation amplifier designs that address these limitations. For a critical application, it would be best to use one of these specialized designs rather than construct an instrumentation amplifier from three ordinary op-amps.

13.8 Universal Amplifier

The inverting and non-inverting functions may be combined in the same circuit, as shown in figure 278. This gives us a *universal amplifier*, that can provide any linear transfer function.

This may be regarded as an inverting amplifier with input signal V_a and offset V_b , or as a non-inverting amplifier with offset V_a and input signal V_b .

When V_a is an input and V_b an offset voltage, we have a *generalized inverting amplifier*. When V_b is an input and V_a an offset voltage, we have a *generalized non-inverting amplifier*.

Design Example

Suppose that the design requirement is for an amplifier with an input voltage of ± 1 volt that corresponds to an output signal between 0 and +5 volts. The lever diagram for an inverting design is shown in figure 279(a), and for a non-inverting design in figure 279(b).

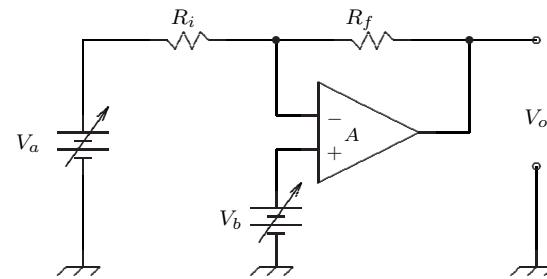


Figure 278: Generalized Inverting/Non-Inverting Amplifier

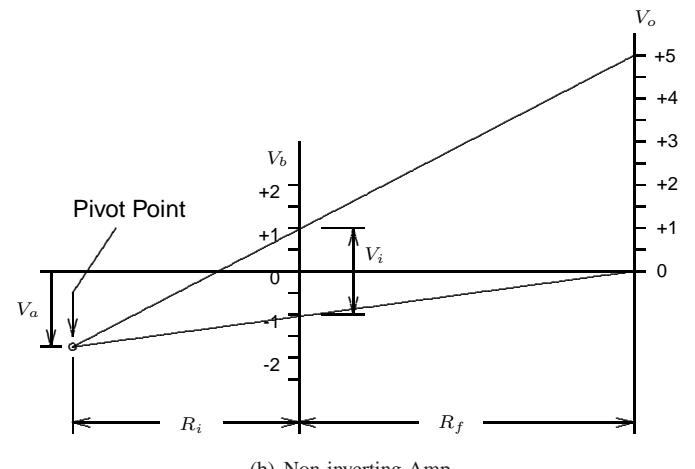
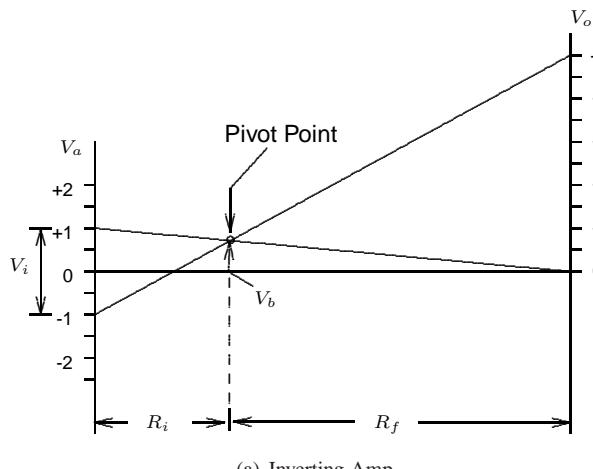


Figure 279: Lever Diagrams

Inverting Design

The inverting design lever diagram is constructed as follows:

1. Draw a trace between the maximum output voltage and the minimum input voltage.
2. Draw a second trace between the minimum output voltage and the maximum input voltage.
3. The point at which these two traces meet is the *pivot point*. As the input voltage varies between ± 1 volt, the two diagonal lines pivot around this point.

4. The vertical position of the pivot point defines the offset voltage V_b that must be applied to the non-inverting input terminal.
5. The horizontal position of the pivot point defines the ratio of $R_i : R_f$.

Reading values¹³⁴ off figure 279(a), $R_f/R_i = 2.55$ and $V_b = +0.75$ volts.

Non-Inverting Design

The lever diagram for a non-inverting amplifier (figure 279(b)) is constructed as follows:

1. Draw a trace between the maximum output voltage (+5) and the maximum input voltage (+1).
2. Draw a second trace between the minimum output voltage (0) and the minimum input voltage (-1).
3. Extend these two traces to the left until they meet. This is the pivot point for the non-inverting amplifier.
4. The vertical position of the pivot point defines the offset voltage V_a that must be applied to the left end of R_i .
5. As in the case of the inverting amplifier, the horizontal position of the pivot point defines the ratio of $R_i : R_f$.

Reading values off figure 279(b), $R_f/R_i = 1.40$ and $V_a = -1.7$ volts.

Comparing the Designs

Which design is better? Well, it depends. If it's important whether the signal is inverted or not, that may determine the circuit. In this numerical example, the inverting circuit has the advantage that the bias voltage V_b is positive compared to the non-inverting circuit which requires a negative value of V_b . In a single supply system, a positive bias voltage would be easier to provide.

Analysis: The Lever Diagram

Here's how we know that the lever diagram really does represent the behaviour of the generalized amplifier.

By superposition, the output voltage is equal to:

$$V_o = V_b \left(1 + \frac{R_f}{R_i}\right) - V_a \left(\frac{R_f}{R_i}\right)$$

The lever diagram is shown in figure 280.

The equation for a straight line is

$$y = mx + b$$

where m is the Slope and b is the Y-Intercept. In the case of figure 280,

$$m = \frac{V_b - V_a}{R_i}, \quad b = V_a$$

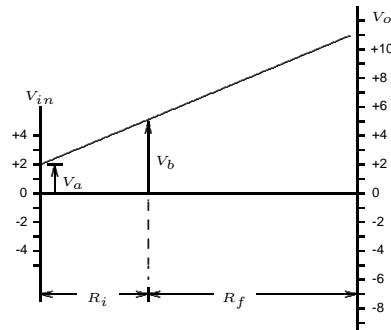


Figure 280: Voltage Distribution Diagram, Generalized Amplifier

¹³⁴We determined these values by scaling off the original xfig drawing.

Substituting these expressions for m and b in the general equation, we have

$$y = \left(\frac{V_b - V_a}{R_i} \right) x + V_a$$

One point on this line corresponds to $x = R_i + R_f$, $y = V_o$. Substitute these values for x and y and simplify:

$$\begin{aligned} V_o &= \left(\frac{V_b - V_a}{R_i} \right) (R_i + R_f) + V_a \\ &= V_b \left(1 + \frac{R_f}{R_i} \right) - V_a \left(\frac{R_f}{R_i} \right) \end{aligned}$$

This is the same as the previous equation obtained by superposition, so the lever diagram corresponds to the behaviour of the circuit.

13.9 Current to Voltage Converter

A current-to-voltage converter? Why not just use a resistor? Current flows through a resistor, sets up a voltage across it, done.

Yes, that's true. But there are two problems with a resistor as current-voltage converter:

- the voltage that is set up across the resistor when current flows through it may affect the source of the current, and
- the resistor probably represents a finite source impedance to drive whatever follows.

This device we're about to describe is a *precision* current-to-voltage converter. It has no effect on the source of the current because its input looks like a short circuit, and its output is a voltage source for output currents less than the current limit (see figure 282 below). So it's much closer to the ideal than a plain old resistor.

Consider the circuit shown in figure 281. This should look familiar, because it made a previous appearance in section 13.1.

The Miller Effect presents the source with a very low resistance,

$$r_i = \frac{R_f}{A_{ol}} \quad (510)$$

For example, if R_f is $10\text{k}\Omega$ and A_{ol} is 10^5 (both typical values) then the input resistance r_i is only 0.1Ω . Currents in the milliampere range will not create a significant voltage across this resistance, so it approximates the virtual earth point we promised in earlier sections on the operational amplifier. Effectively, then, the source sees a short circuit to ground, across which there is effectively no voltage to affect the source.

As for the conversion factor, we revert back to the original op-amp circuit. All the input current must flow through R_f , so for input current of the polarity shown in figure 281, the output voltage will be positive and given by

$$e_o = R_f i_i \quad (511)$$

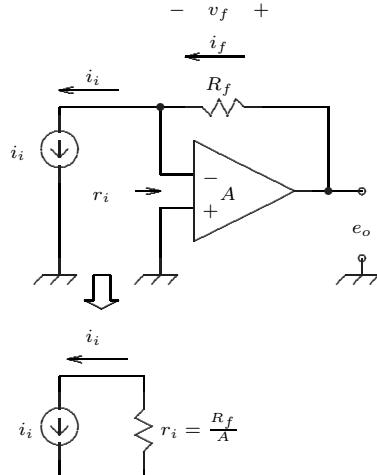


Figure 281: Current to Voltage Converter

So, if R_f is $10\text{k}\Omega$, then the output is 10 volts per milliampere of input current.

If it is important that the current-voltage converter be precisely accurate at low input currents, you should use a JFET or MOSFET input op-amp. The bias currents of a BJT op-amp may cause errors.

This amplifier accepts current and produces a proportional voltage, so in general its transfer function is an *impedance*, ie:

$$\frac{e_o}{i_i} = Z_t \quad (512)$$

For the circuit of figure 281 above, $Z_t = R_f$.

For that reason, the amplifier is known as a *transimpedance* amplifier, a contraction of the words *transfer* and *impedance*. The equivalent circuit for a current-voltage or transimpedance amplifier is shown in figure 282.

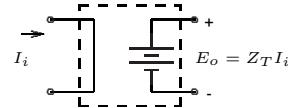


Figure 282: Equivalent Circuit

13.10 Integrator, Inverting

The *integrator* circuit, one of the mainstays of analog circuit design, is shown in figure 283.

This device integrates with respect to time. The integral of some function of time is the area under the voltage-time curve. Some typical electrical waveforms and their integrals are shown in figure 284 on page 347.

Analysis

Referring to figure 283: The inverting input terminal is at a virtual earth point, close to zero volts. So the input current is

$$i_i = \frac{e_i}{R_i} \quad (513)$$

This same current flows into the capacitor, charging it. The capacitor voltage is

$$V_c = \frac{1}{C_f} \int i_i dt \quad (514)$$

Substituting for i_i from equation 513 into equation 514, we have

$$\begin{aligned} V_c &= \frac{1}{C_f} \int \frac{e_i}{R_i} dt \\ &= \frac{1}{R_i C_f} \int e_i dt \end{aligned} \quad (515)$$

The left end of the capacitor is held at zero volts by the op-amp, so the output voltage is

$$\begin{aligned} e_o &= -V_c \\ &= -\frac{1}{R_i C_f} \int e_i dt \end{aligned} \quad (516)$$

The value

$$\frac{1}{R_i C_f}$$

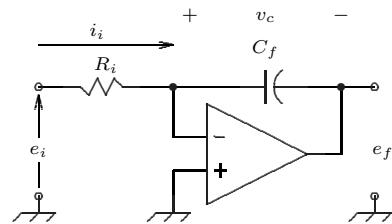


Figure 283: Integrator Circuit

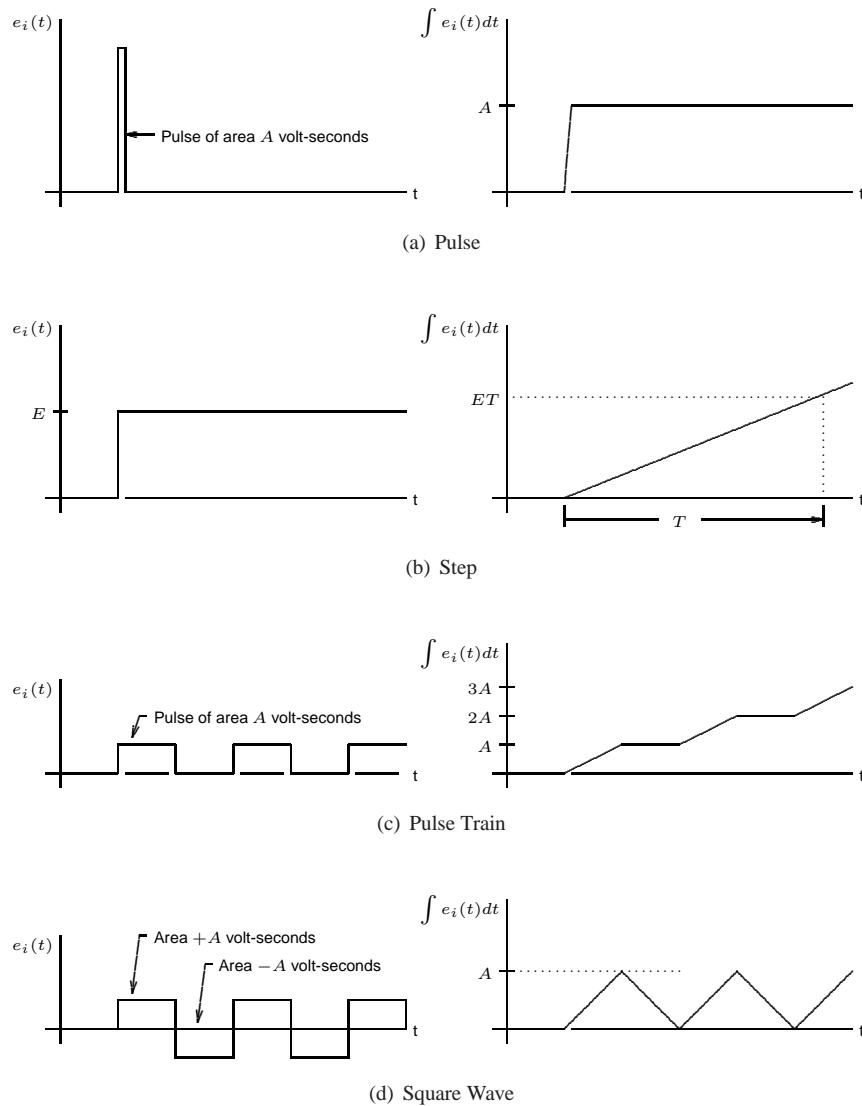


Figure 284: Waveforms and their Integrals

is the *integrator gain*. A common choice for these components is $1M\Omega$ for R_i and $1\mu F$ for C_f , in which case the gain is unity.

Integrator in the Frequency Domain

To determine the integrator amplitude and phase response versus frequency, the integrator circuit of figure 283 is redrawn as figure 285, where the components have been replaced by their impedances.

Then the transfer function is

$$\begin{aligned}\frac{e_o}{e_i} &= -\frac{Z_f}{Z_i} \\ &= -\frac{1}{sR_iC_f}\end{aligned}\quad (517)$$

The constant $1/R_iC_f$ can be treated as the crossover frequency ω_o . Then

$$\begin{aligned}\frac{e_o}{e_i} &= -\frac{1}{s/\omega_o} \\ &= -\frac{1}{j\frac{\omega}{\omega_o}}\end{aligned}\quad (518)$$

On a Bode plot, the amplitude-frequency graph of equation 518 is a straight line, decreasing at a rate of 20db/decade frequency, passing through the frequency ω_o , where

$$\omega_o = \frac{1}{R_iC_f} \text{ radians/sec.} \quad (519)$$

The phase-frequency graph of $1/sRC$ is a constant -90° . To see this, substitute $j\omega$ for s in equation 517 and convert from the rectangular to polar form:

$$\begin{aligned}\frac{e_o}{e_i} &= -\frac{1}{j\omega R_i C_f} \\ &= -\frac{1}{\omega R_i C_f \angle 90^\circ} \\ &= -\frac{1}{\omega R_i C_f} \angle -90^\circ\end{aligned}$$

The leading negative sign indicates that the op-amp integrator is inverting, which contributes an additional phase shift of 180° . Then:

$$\frac{e_o}{e_i} = \frac{1}{\omega R_i C_f} \angle +90^\circ \quad (520)$$

This has the Bode plot characteristic shown in figure 286. It is identical with figure 232 on page 306 except that the trace crosses the x axis at $\omega_o = 1/R_iC_f$.

The gain and phase plots for the op-amp inverting integrator are shown in figure 286.

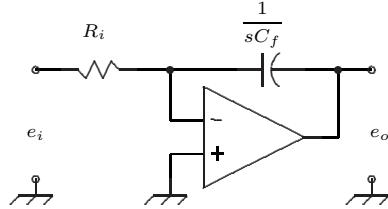


Figure 285: Integrator, Frequency Domain

Details and Enhancements

The integrator shown in figure 283 is not very useful as it is. If the input signal has a non-zero average value, the integrator output voltage will eventually run into the upper or lower limit of the operational amplifier. There are two ways of dealing with this:

- Provide some method, such as a shorting switch, to discharge the capacitor back to zero.
- Place the integrator inside a feedback loop that provides a corrective signal to ensure that the op-amp never reaches its limits.

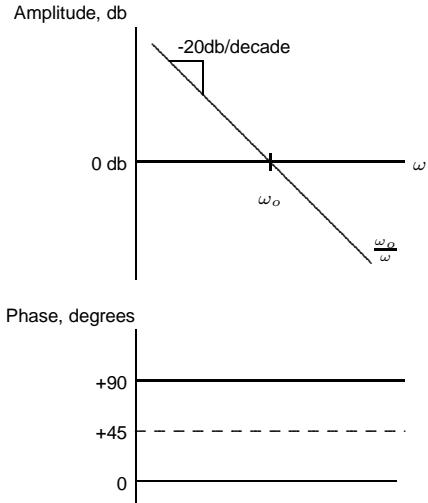


Figure 286: Op-Amp Integrator Frequency Response

13.11 Integrator, Non Inverting

The integrator that we studied in section 13.10 is an inverting device. In the frequency domain, the transfer function of the inverting integrator is:

$$\frac{e_o}{e_i} = -\frac{1}{sR_iC_f} \quad (521)$$

We'll now show that the integrator in figure 287 is a *non-inverting* integrator. Of course, we could create a non-inverting integrator with an inverting integrator followed by an inverter of gain -1 . But the approach shown in figure 287 saves us an op-amp at the expense of a second capacitor and resistor. As well, unlike the *inverting integrator followed by inverter* approach, the non-inverting integrator can operate from a single power supply. The transfer function of the non-inverting integrator is:

$$\frac{e_o}{e_i} = +\frac{1}{sRC} \quad (522)$$

To show this, we first determine the voltage e_a at the non-inverting terminal:

$$\begin{aligned} e_a &= \frac{1/sC_1}{R_1 + 1/sC_1} e_i \\ &= \frac{1}{1 + sR_1C_1} e_i \end{aligned} \quad (523)$$

Since this is a negative feedback system the non-inverting terminal is also at potential e_a . The current through R_2 is

$$\begin{aligned} i_2 &= \frac{e_a}{R_2} \\ &= \frac{1}{R_2} \frac{1}{(1 + sR_1C_1)} e_i \end{aligned} \quad (524)$$

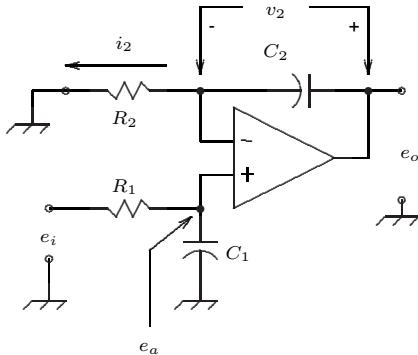


Figure 287: Non-Inverting Integrator

This current also flows through the feedback impedance $1/sC_2$. The voltage across the feedback impedance is

$$\begin{aligned} v_2 &= i_2 \frac{1}{sC_2} \\ &= \frac{1}{sR_2 C_2} \frac{1}{(1 + sR_1 C_1)} e_i \end{aligned} \quad (525)$$

Finally, by KVL the output voltage is the sum of e_a and v_2 : Substituting from equation 523 for e_a and equation 525 for v_2 , we have:

$$\begin{aligned} e_o &= e_a + v_2 \\ &= \frac{1}{1 + sR_1 C_1} e_i + \frac{1}{sR_2 C_2} \frac{1}{(1 + sR_1 C_1)} e_i \\ &= e_i \left(\frac{1 + sR_2 C_2}{1 + sR_1 C_1} \times \frac{1}{sR_2 C_2} \right) \end{aligned} \quad (526)$$

Put $R_1 C_1 = R_2 C_2 = RC$, and we have

$$\frac{e_o}{e_i} = \frac{1}{sRC} \quad (527)$$

Voila: non-inverting integrator.

The Howland Non-Inverting Integrator

The *Howland Circuit* can be configured as a non-inverting current generator circuit. (See figure 507 on page 598 for a more complete description of the Howland current generator.)

The output current is proportional to the input voltage,

$$i_o = e_i / R \quad (528)$$

where R is the Howland circuit resistance value.

The Howland current generator may be used as a non-inverting integrator, as shown in figure 288. Denoting the capacitor impedance as X_c , we have

$$\begin{aligned} e_o &= i_o X_c \\ &= i_o \frac{1}{sC} \end{aligned} \quad (529)$$

Substituting e_i/R for i_o from equation 528, we have

$$e_o = \frac{e_i}{R} \frac{1}{sC} \quad (530)$$

Rearranging to find the transfer function e_o/e_i ,

$$\frac{e_o}{e_i} = \frac{1}{sRC} \quad (531)$$

which is the equation for a non-inverting integrator.

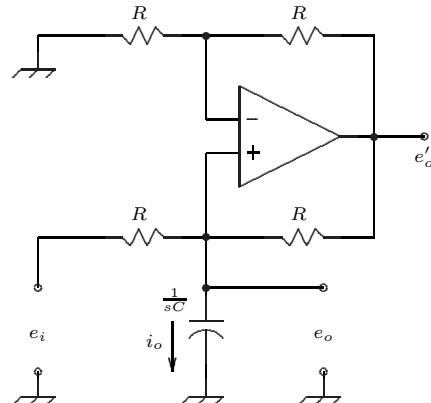


Figure 288: Howland Non-Inverting Integrator

Although the circuit of figure 288 will function in this fashion, the output voltage must not be loaded by any appreciable resistance or the integration current will be partially diverted into that resistance and integration action will be compromised. To connect the output to a following circuit, the output of figure 288 could be buffered with a unity-gain buffer, using a an op-amp with JFET or MOSFET devices at the input for small bias current.

Alternatively, the output can be taken from the output of the op-amp in figure 288. This point is a low source impedance and has no effect on the charging of the capacitor. It can be shown that the transfer function between input and output e'_o is given by

$$\frac{e'_o}{e_i} = \frac{2}{sRC}$$

This is the equation of an integrator with a gain of $2/RC$.

13.12 Differentiator

The basic *differentiator* circuit, figure 289, uses the same components as the integrator (section 13.10), with the capacitance and resistance interchanged. It implements the function of differentiation with respect to time, as illustrated in figure 290 for the Step and Ramp waveforms. (These waveforms assume a differentiator gain of unity).

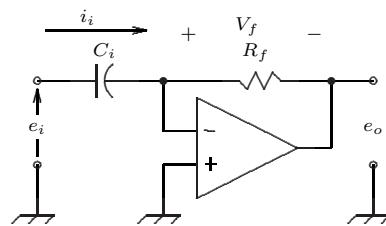
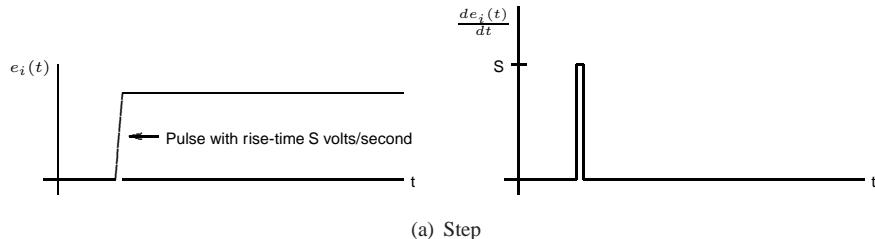
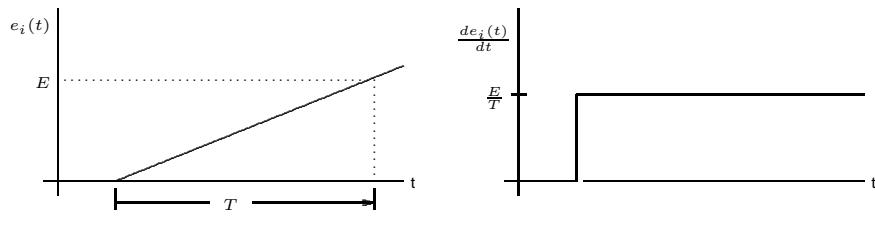


Figure 289: Differentiator



(a) Step



(b) Ramp

Figure 290: Waveforms and their Differentials

Analysis

The inverting input terminal is at a virtual earth point, close to zero volts. So the input current is

$$i_i = C_i \frac{de_i}{dt} \quad (532)$$

This same current flows through the feedback resistor R_f .

$$V_f = R_f C_i \frac{de_i}{dt} \quad (533)$$

The left end of the capacitor is held at zero volts by the op-amp, so the output voltage is

$$\begin{aligned} e_o &= -V_f \\ &= -R_f C_i \frac{de_i}{dt} \end{aligned} \quad (534)$$

The value $R_f C_i$ is the *differentiator gain*.

Differentiator in the Frequency Domain

In the frequency domain, referring to figure 291, the transfer function of the differentiator is:

$$\begin{aligned} \frac{e_o}{e_i} &= -\frac{Z_f}{Z_i} \\ &= -\frac{R}{1/sC} \\ &= -sRC \end{aligned}$$

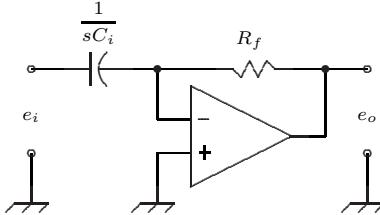


Figure 291: Differentiator, Frequency Domain

This which can be regarded as a differentiator term (s) multiplied by a constant RC . The minus sign causes a constant 180° contribution to the phase.

This combination of a constant gain RC and differentiator (the s term in the numerator) can be treated in various ways. For example, the value RC could be treated as a constant factor and shown as a separate trace on the Bode plot. In that case, the differentiator has the characteristic shown in figure 231 on page 306 and the constant factor RC decibels would be added to it.

It's also possible to view the constant in a different light. When $\omega = 1/RC$, the value of $G(s)$ is zero db, that is, the curve is crossing the zero db axis. We can call this crossover frequency ω_o , where

$$\omega_o = \frac{1}{RC} \quad (535)$$

Viewed in this way, the transfer function of the differentiator-with-gain is given by

$$G(s) = j \frac{\omega}{\omega_o} \quad (536)$$

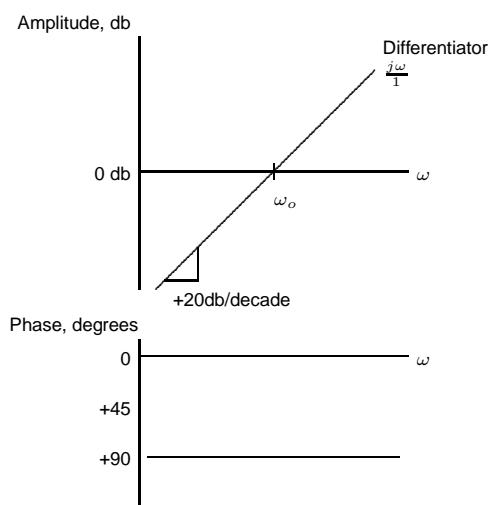


Figure 292: Op-Amp Differentiator with Gain

The Bode characteristic is identical to that of figure 231 on page 306, except that the magnitude trace crosses the zero-db axis at ω_o rather than unity. The result is shown in figure 292.

Notice the phase value in figure 292. A differentiator contributes $+90^\circ$, the constant RC contributes 0° , and the amplifier contributes -180° . The resultant phase angle is -90° .

Details and Enhancements

As Tom Waits says¹³⁵

You got it buddy: the large print giveth and the small print taketh away...

Here's the small print on the differentiator. There are two problems:

- First, the gain of a negative feedback amplifier, expressed in terms of input and feedback impedances, is

$$\frac{e_o}{e_i} = -\frac{Z_f}{Z_i} \quad (537)$$

In the case of the differentiator, Z_i is a capacitance which decreases in impedance at increasing frequencies. Consequently, the gain of the differentiator increases with frequency. Any high frequency noise in the input signal will be emphasized, and this can be a problem.

- The second problem is concerned with the issue of stability. As shown in section 24.4 on page 711, the differentiator circuit of figure 289 is unstable and will tend to oscillate.

The solution, it turns out, is to add small resistor R_i in series with the input capacitance, as shown in figure 293. The effect is to put a lower limit on the input impedance Z_i .

The frequency response of the improved differentiator is given by

$$\frac{e_o}{e_i} = -\frac{sCR_f}{1 + sCR_i} \quad (538)$$

This indicates that the circuit behaves as a differentiator for frequencies below

$$\omega_t = \frac{1}{CR_i} \quad (539)$$

At high frequencies, where the impedance of C_i is small compared to R_i , the gain flattens out to a constant value of

$$\frac{e_o}{e_i} \approx -\frac{R_f}{R_i} \quad (540)$$

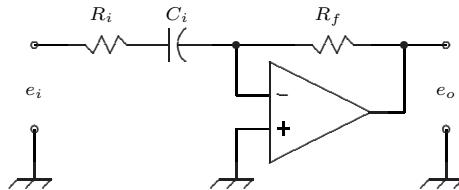


Figure 293: Improved Differentiator

¹³⁵Step Right Up: Asylum Records, 1975, Manifesto Records, 1995.

13.13 Application: Spring-Mass Simulation

In this section, we show an example application of op-amp circuits, simulating the motion of an oscillatory spring-mass system¹³⁶.

In the world of motion, displacement is the time integral of velocity and velocity is the time integral of acceleration. If some electrical signal is available that is proportional to acceleration then signals proportional to velocity and displacement may be obtained by electronic integration. This provides us with a very powerful method of simulating mechanical, thermal, hydraulic and electrical systems.

An example of a mass-spring system is shown in figure 294 (reference [95]).

The forces in this system are:

- Due to acceleration of a mass: $F_a = Ma$
- Due to velocity against a damping force: $F_v = Bv$
- Due to displacement of a spring: $F_s = Kx$
- Total input force F_i

These forces must sum to zero, that is:

$$+ Ma + Bv + Kx - F_i = 0 \quad (541)$$

or, dividing by M:

$$+ a + \frac{B}{M}v + \frac{K}{M}x - \frac{F_i}{M} = 0 \quad (542)$$

This may be rewritten as

$$\begin{aligned} a &= -\frac{B}{M}v - \frac{K}{M}x + \frac{F_i}{M} \\ &= -(\frac{B}{M}v + \frac{K}{M}x - \frac{F_i}{M}) \end{aligned} \quad (543)$$

That is, the negative of acceleration may be generated by adding together the velocity, displacement and force values, each of which is multiplied by some constant.

Step by step, here is how we create the circuit diagram shown in figure 295:

1. There is a trick in analog simulations: *assume some variable is known*. Usually the appropriate variable to choose is the highest differential of some quantity. In this case, the highest differential of displacement is acceleration.
2. Once the acceleration is known, it may be integrated to get a voltage proportional to velocity v . In the circuit, that's done by A_1 .
3. Velocity v can be integrated to generate a voltage proportional to displacement x . That's done by A_2 .
4. The output of integrator A_1 is negative velocity. The positive version of the velocity signal is obtained by processing this signal through sign-changer A_4 .

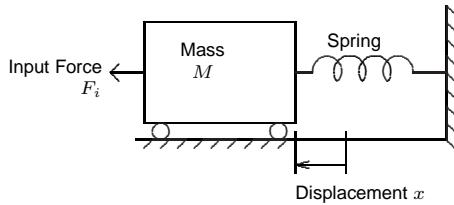


Figure 294: Mass-Spring System

¹³⁶Much more complex simulations are possible. The author recalls seeing a simulation of satellite motion around the planet. The author's brother (David Hiscocks) created a simulation of the population dynamics for a predator-prey system, rabbits and wolves.

5. As shown in equation 543, velocity must be multiplied by a constant value equal to $\frac{B}{M}$. In the circuit, that's done by potentiometer P_3 .
6. Similarly, displacement must be multiplied by a constant value equal to $\frac{k}{M}$. That's done by potentiometer P_2 .
7. We need a constant term equal to the force input. That's generated by a voltage source processed through potentiometer P_1 , which is set to a gain of $\frac{1}{M}$.
8. Now we have all the terms on the right-hand-side of equation 543. We sum them in adder A_1 to generate the acceleration a . This is the variable we assumed was available in the first place, so we have closed the loop on the circuit and made it a complete simulation of the mass-spring system.

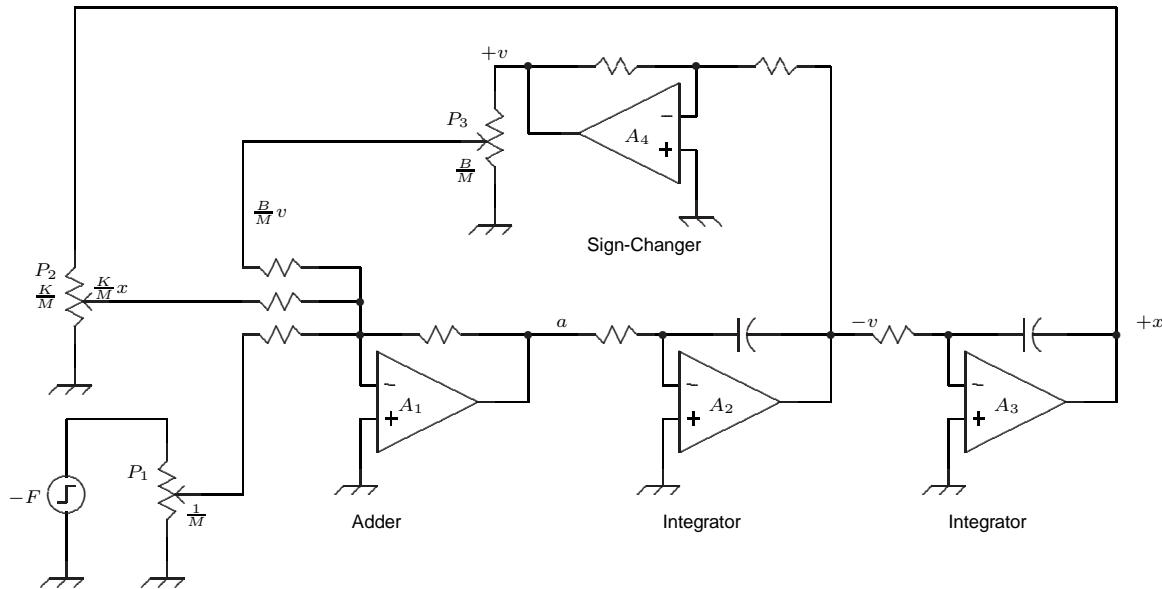


Figure 295: Mass-Spring Simulation

Further notes on the Simulation

- The physical constants are dialed into the various potentiometers. Obviously, the setting for the pot must be between zero and 1.0. If a quantity greater than unity is required, say 4.3 for example, one would insert an amplifier of gain 10 followed by a pot with setting 0.43.
- The problem has not been *scaled* in any way, so we assume that the physical constants are such that the voltages are directly equivalent. For example, if the displacement is in metres and the op-amps are capable of ± 10 volts, then the displacement must be limited to ± 10 metres. If this is not the case, the problem must be scaled. See reference [95] for the gory details.
- Furthermore, the simulation runs in *real-time*: the simulation time scale is the same as the physical device. It's also possible to scale the simulation so that it runs faster or slower than the physical system.

- The fixed resistors would be $1M\Omega$ and the capacitors $1\mu F$ to give integrator gains of 1 and an adder scale factor of 1.
- The potentiometers are shown as directly driving the input resistors of the following op-amps. The input resistors will load the pots and cause setting errors unless the pot resistance is much less than $1M\Omega$. (For example, $1k\Omega$ pots would work well.) Alternatively, the output of each pot could be buffered through a follower circuit (section 12.3).
- An oscilloscope may be attached to any of the amplifier outputs to read acceleration, velocity or displacement vs time.
- The forcing function F could be a step, impulse, continuous sine wave of some frequency, or some other waveform.
- The simulation should be checked for stability: every loop should have an odd number of inversions. Otherwise there is positive feedback and the loop will lock up. In this circuit, there are two loops, each with three op-amps, which is acceptable.

Once this simulation is set up and debugged, the various potentiometer settings can be varied to see the effect of changing the system constants.

Modern practice is to simulate this type of system on a digital computer, using numerical methods to perform the integrations. However, there are some arguments in favour of the analog approach described here. See for example [96], which shows an example of analog simulation using modern op-amps.

13.14 Booster Stage

The output current available from an integrated circuit amplifier is usually very limited, typically to $20mA$. Where more current is required, a *booster* or *buffer* stage can be added to the basic amplifier. If this is inside the feedback loop, then the booster stage becomes part of the amplifier and its imperfections are (somewhat) compensated by negative feedback.

A simple booster circuit is shown in figure 296. The booster is a complementary NPN-PNP emitter follower stage that will boost the output current by a factor of the current gain of the booster transistors. Resistors R_1 and R_2 help prevent the emitter follower circuits from bursting into operation with certain types of loads, and are typically set to a few hundred ohms. Resistors R_3 and R_4 limit the output current in the event of a short circuit, and they can be omitted if you like to live dangerously.

Resistors R_f and R_i show how you could provide negative feedback around the op-amp and buffer combination. Other negative feedback configurations, such as the non-inverting amplifier are also possible.

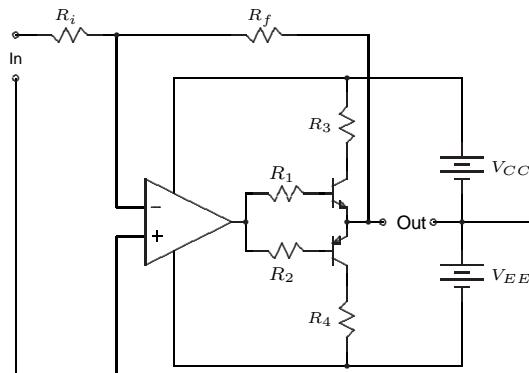


Figure 296: Basic Booster

Improved Booster Stage

The circuit configuration of figure 296 is subject to an effect known as *crossover distortion*. Suppose the input is being driven by a sine wave. As the output passes through zero in a positive direction, the output of the op-amp must switch from -0.6 volts while driving the lower PNP transistor, to $+0.6$ volts while driving the upper NPN

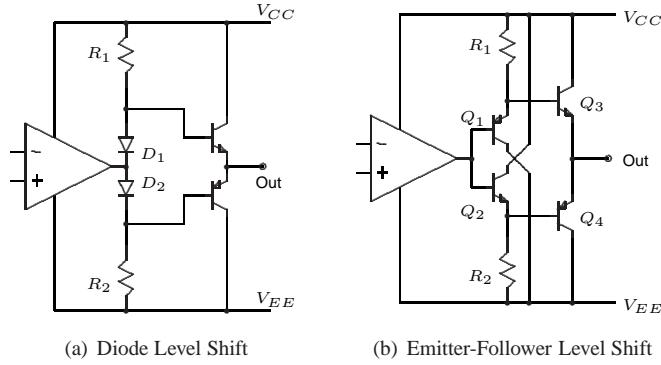


Figure 297: Improved Booster Stages

transistor. If the frequency is relatively low (a few hundred Hz, say) then the op-amp will rapidly switch through this region, and the output voltage will not be distorted. The negative feedback mechanism will automatically ensure that the output of the op-amp is sufficient to drive the buffer output to the correct value.

However, at high frequencies the op-amp may take a significant time to transition through this *dead zone* between -0.6 and +0.6 volts, distorting the output signal. So this circuit is suitable for non-critical low frequency applications but may not work well where low distortion is required (such as audio) or at high frequencies.

Two other possible buffer designs that reduce crossover distortion are shown in figure 297(a) and figure 297(b).

In figure 297(a), the output of the op-amp is shifted up by 0.6 volts and down by 0.6 volts for each of the driver transistors. The voltage across the diodes is unlikely to be exactly equal to the base-emitter drops of the transistors, but it's an improvement over the circuit of figure 296 and will reduce crossover distortion.

In figure 297(b), the diodes of figure 297(a) are replaced by emitter followers Q_1 and Q_2 that are complementary to the output buffer transistors Q_3 and Q_4 . The first emitter follower stage provides voltage shifting and possible current amplification to drive the output emitter follower stage. This circuit is capable of driving more output current at higher frequencies and less distortion than the preceding circuits.

In any of these booster circuit designs, the booster transistors may be carrying considerable current. The resultant power dissipation may require that they be attached to a heat sink. In any case, this should be checked as part of the design.

References

Buffer stages are often plagued by various oscillation problems. The underlying cause and cure are discussed in section 24.4.

The definitive sources of information on op-amp buffer stages are references [97] and [98], both by Jim Williams. Reference [98] has a valuable section on dealing with such problems, under the title *The Oscillation Problem: Frequency Compensation without Tears*.

13.15 Current Generator, Floating Load

When neither end of the load resistance R_l is connected to ground (ie, the load is *floating*), the circuit of figure 298 can be used to drive current through it. (A discussion of floating and grounded connections is in section 2.12.) The current through the load is given very simply by

$$I_l = \frac{E_i}{R_s} \quad (544)$$

The value of R_l does not appear in this equation so the current through the load is *constant*, ie, independent of load resistance. The output current and voltage must be within certain limits:

- the output current must not exceed the output current capability of the op-amp.
- the voltage developed across the load resistance must be within the output voltage capability of the op-amp.

Notice that the output current may be caused to reverse in direction if the input voltage E_i is reversed in polarity. This assumes that the op-amp is operated from bipolar power supplies and can produce the necessary positive or negative voltage to drive the correct current through the load.

Increased Output Current

We showed in section 12.3 how the output current could be increased with a buffer amplifier stage. If it is sufficient that the output current flow in one direction, a BJT or MOSFET transistor will provide the necessary muscle, as shown in figure 299.

In this case, the maximum output current is increased over that of the circuit in figure 298 by a factor equal to the current gain of the transistor.

Applications

There are several applications where the load can be floated off ground and these circuits used for constant current drive.

- In the vicinity of an electromagnet, the strength of the magnetic field is proportional to the current through the windings. In a magnetic levitation system, one would control the current through the electromagnet to position the levitated object [99].
- The torque produced by a DC motor is proportional to its drive current. So a constant-torque drive is created by driving constant current through the motor.
- In an electro-plating system, the rate of plating deposition depends on the current through the plating tank, so a constant current drive system is desirable.

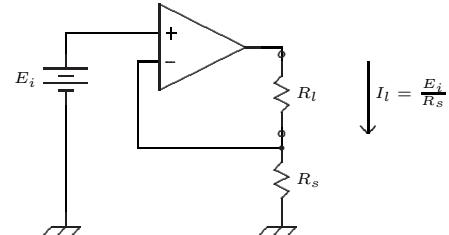


Figure 298: Current Generator, Floating Load

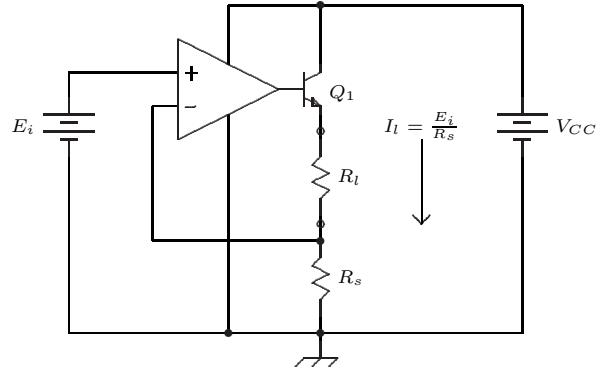


Figure 299: Current Generator, Increased Output

13.16 Current Generator, Grounded Load

In applications where one end of the load resistance must be connected to a power supply terminal, the circuit of figure 298 can't be used and we require a different configuration. One possibility is shown in figure 300. The operational amplifier will adjust its output so that the voltage V_s across R_s is equal to the input voltage E_i . Then the current into the source terminal S of the transistor is

$$I_s = \frac{E_i}{R_s} \quad (545)$$

For example, suppose that the voltage across V_s is less than E_i .

- An error voltage appears across the terminals of the op-amp.
- This error signal causes the output voltage of the op-amp to move towards its negative power supply.
- This increases the gate-source voltage of the MOSFET, causing it to conduct more heavily.
- This process continues until the error voltage approaches zero, that is, until E_i and V_s are equal.

Although other devices can be used for the transistor Q_1 , a P-Channel MOSFET is an attractive device because its drain current is exactly equal to its source current. The voltage at the drain terminal of the MOSFET has little effect on the drain current, so the load current is given by exactly E_i/R_s .

The maximum voltage that can appear across the load resistor (or capacitor) is limited by the maximum voltage that can appear at the drain of the transistor. This would be equal to the supply voltage V_{CC} minus the voltage across R_s (equal to E_i) minus the saturation voltage of the MOSFET.

Applications

This circuit has several common applications.

- Constant current into a capacitor creates a linear ramp voltage across the capacitor, which can be used for various comparison and timing functions.
- If a constant current is driven through a resistor, then the voltage across the resistor will be proportional to the resistance. This scheme is often used in a multimeter to measure resistance.

Moving the Ground Point

To emphasise that there is nothing magical about the *ground* point in a circuit, consider the floating-load current generator of figure 298 on page 358. Suppose that the lower end of the load resistance R_l is a ground point that cannot be changed. The circuit can accommodate that situation by floating the power supply, as shown in figure 301. If the circuit supply is produced via a transformer, the entire supply can be floated off ground.

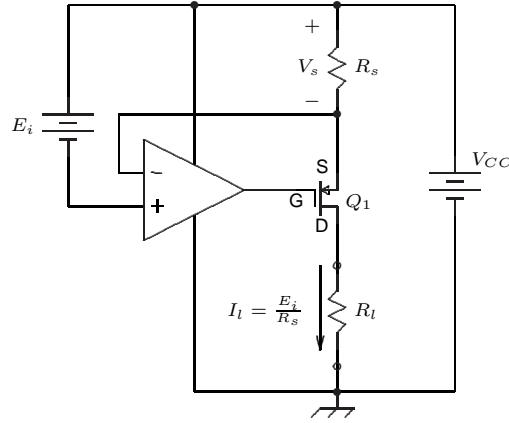


Figure 300: Current Generator, Grounded Load

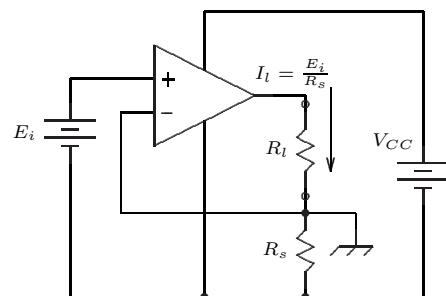


Figure 301: Current Generator, Grounded Load

13.17 Charge Amplifier

A *transducer* is a device which changes some physical quantity, such as *force*, into an electronic signal. A *charge transducer* is a device that converts some physical quantity into an electrical charge. To electronic circuits, a charge transducer appears as a charged capacitor. In figure 302, component X_1 is a charge transducer. As we shall see, the op-amp and capacitor function as a *charge amplifier* for the signal from the transducer.

Here are three examples of charge transducers:

Piezo-electric Crystal

A *piezo-electric crystal* is a material that produces a charge on its surface when it is mechanically compressed. (This effect is reciprocal: a charge between two surfaces of the crystal causes the crystal to compress.) If we place the crystal between a surface and a mass, we have an accelerometer – a device for measuring acceleration. If the surface is accelerated vertically, the force created by the inertial mass will be proportional to the product of the mass and acceleration. Consequently the charge on the surface of the crystal is proportional to acceleration.

The accumulation of a charge on the surface is connected to measurement electronics by metallic electrodes on the top and bottom surface of the crystal.

Electret Microphone

An *electret* is a material that retains an electrostatic charge in an analogous fashion to a permanent magnet, which permanently retains a magnetic field. The permanent electrostatic charge generates an electrostatic field, and may be used to create an electret microphone, as shown in figure 304.

In this case, the upper electrode is a flexible diaphragm such as a mylar film with a conductive metal coating. When sound waves strike the diaphragm, they cause it to flex. This changes the spacing between the two electrodes. Recall that capacitance is given by

$$C = \frac{\epsilon A}{s} \quad (546)$$

where C is the capacitance in farads, ϵ is the dielectric constant of the substance between the two plates of the capacitor, A is the area of the two plates, and s is their spacing. Consequently, sound waves change spacing s and hence the capacitance of the microphone.

Now, the voltage and charge on a capacitor are related by:

$$Q = CV \quad (547)$$

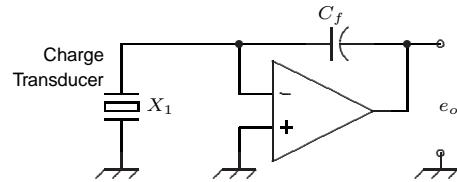


Figure 302: Charge-Amplifier

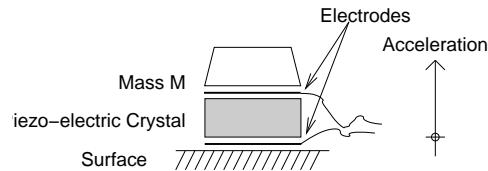


Figure 303: Piezo-Electric Accelerometer

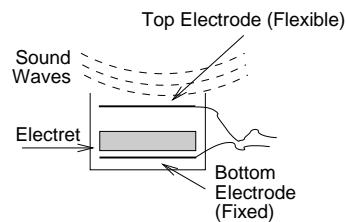


Figure 304: Electret Microphone

where Q is the charge and V the terminal voltage of the capacitor. When the capacitance C is modulated by a sound wave and the charge Q stays constant, the terminal voltage must vary.

Consequently, this device behaves as a microphone: a transducer that converts a sound wave signal into an electrical signal.

Solid-State X-Ray Detector

A silicon crystal doped with lithium atoms and kept at a low temperature becomes an efficient detector of X-Rays. Each X-Ray photon generates a pulse of charge that appears at the surface of the crystal. As in the case of the accelerometer and electret microphone, the device may be modelled as a capacitor in parallel with a charge-generator. (References [100] and [101] contain descriptions of this device.)

Charge Sensor Interface

Suppose that the charge sensor is to be connected to an amplifier to increase the output voltage signal. A first attempt at a circuit is shown in figure 305.

In response to some input (such as a force) the transducer crystal generates a charge. This charge then distributes over the total capacitance which is the sum of the sensor, cable and load capacitance.

This arrangement has two problems.

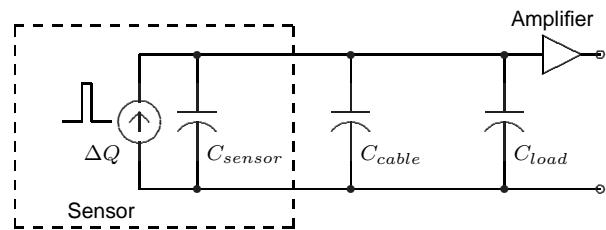


Figure 305: Charge Sensor Interface

- By the equation

$$Q = CV$$

where Q is the transducer charge, C the capacitance attached to the transducer and V the voltage into the amplifier, the magnitude of V is a function of the load capacitance. But the cable capacitance depends entirely on the wiring arrangement, such as the length of the cable. Consequently, the voltage into the amplifier depends on the cable and load capacitance, so it is not predictable and will change if the cable is changed¹³⁷.

- When producing an AC signal, the charge generator can be regarded as an AC current generator. This current generator can be thevenized into a voltage source in series with a very large resistor R_{int} . The internal resistance and the various capacitances will act as a lowpass filter for alternating current signals.

Consequently, the frequency response will roll off at 20db/decade above a very low corner frequency. When high frequency response is important (as in the case of an electret microphone) this is not acceptable.

Moving the buffer amplifier to the sensor (figure 306) improves matters. The buffer amplifier provides a high-impedance load to the sensor and a low source impedance to the cable. Furthermore, the direct connection of the transducer to the amplifier results in a low, fixed value of load capacitance for the transducer. This is the approach used in most electret microphones.

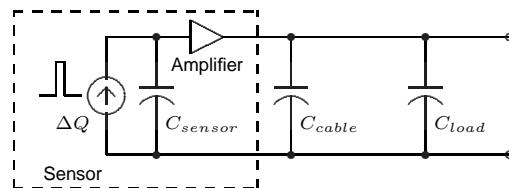


Figure 306: Charge Sensor Interface, Improved

¹³⁷This is particularly true because to prevent noise coupling, the connecting cable is likely to be coaxial with an inner conductor and an outer surrounding shield. Coaxial cable has a significant capacitance which is proportional to its length.

However, power must somehow be provided to the microphone amplifier. The microphone may have provision for an internal battery or a clever system known as *phantom power* [102] may be used to pipe power from an audio mixing console down to the microphone, using the audio signal and ground leads.

Charge Amplifier Design

A basic charge amplifier is shown in figure 307. The op-amp presents a virtual earth point to the sensor signal. Because the op-amp maintains this point at zero volts and current cannot flow into the op-amp input, any charge q_i from the sensor flows as current i_i into the feedback network.

For the sensor and cable capacitances, one terminal is at ground and the other at virtual earth (ground), so there is no voltage across them. Consequently, there is no current through those capacitances and they have no effect on the circuit behaviour.

Now let us consider what device or devices should be used for the feedback impedance Z_f .

$$v_o = -Z_f i_i$$

We require the output voltage v_o to be proportional to input charge q_i . A resistance would transform the input current i_i into an output voltage. However, current is the time differential of charge, so the output voltage would then be proportional to the *time differential* of charge, not what we require.

Now consider that Z_f is a capacitor C_f . The voltage across a capacitor is the time integral of the current into it. The time integral of current is charge. Consequently, the output voltage is then proportional to the input charge, as required. Then

$$q_i = -C_f v_o \quad (548)$$

In other words, the output change in voltage is a direct function of the generated charge and the feedback capacitance. The gain of the charge amplifier can be adjusted by changing the feedback capacitance C_f . Surprisingly, the sensor and cable capacitance now have no effect on the gain of the system.

The Function of R_f

As explained in section 21.2, all op-amp circuits must have some path for a small bias current into the input terminals. (This is a DC current that should not be confused with the signal current.) The circuit of figure 307 does not satisfy this requirement. Both the sensor and the feedback are open-circuits for DC currents. Consequently, a feedback resistance R_f is required.

Without R_f , the DC bias current of the op-amp will slowly charge up the input and feedback capacitances, causing the op-amp to drift to one of its saturation boundaries.

At high frequencies where the reactance of C_f is much smaller than R_f , the feedback resistance can be

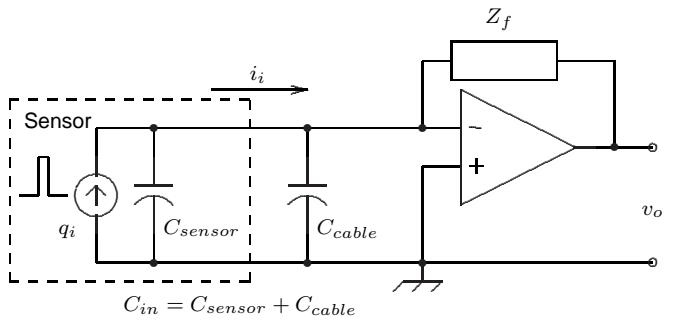


Figure 307: Basic Charge Amplifier

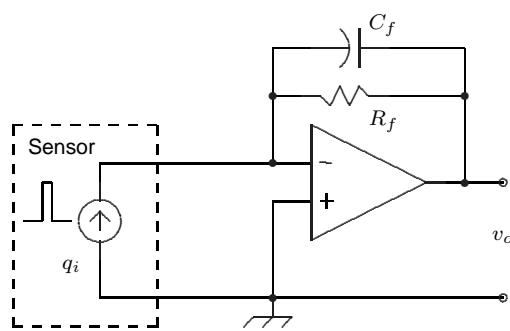


Figure 308: Complete Charge Amplifier

ignored and the circuit behaves as a charge amplifier. At low frequencies, where the reactance of C_f is much larger than R_f , the feedback capacitance can be ignored. Then the output is proportional to the differential of charge rather than the charge itself. The frequency at which the transition occurs between these two behaviours is the point at which the two reactances are equal, or

$$\omega_c = \frac{1}{R_f C_f} \quad (549)$$

where ω_c is the cutoff frequency.

If the capacitance C_f is chosen to suit the gain requirement, then R_f is chosen to provide the necessary low-frequency cutoff. In general, the value of R_f is quite large and consequently any significant bias current will create a sizeable offset voltage at the output of the amplifier. Fortunately, CMOS op-amps have a very small bias current, so that is the type of device to use in this circuit.

Design Example

A certain accelerometer¹³⁸ has the following specifications:

Sensitivity	40pC/g
Frequency Response	3 to 6000 Hz
Capacitance	350pF

Design a charge amplifier to generate an output of 100mV/g, working to a minimum frequency of 10Hz.

Solution

The circuit is shown in figure 309.

Applying equation 548 (and ignoring the minus sign, which only affects the phase), we can determine the value of the feedback capacitance C_f :

$$\begin{aligned} C_f &= \frac{q_i}{v_o} \\ &= \frac{40\text{pC/g}}{100\text{mV/g}} \\ &= \frac{40 \times 10^{-12}}{100 \times 10^{-3}} \\ &= 400 \text{ pF} \end{aligned}$$

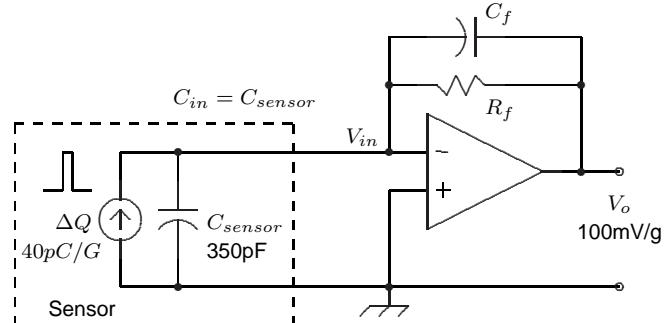


Figure 309: Charge Amplifier Example

The feedback resistance R_f is equal to the reactance of the feedback capacitance C_f at the cutoff frequency, 10 Hz. At 10Hz, the reactance of the feedback capacitance is given by:

$$\begin{aligned} X_{Cf} &= \frac{1}{2\pi f C_f} \\ &= \frac{1}{2\pi \times 10 \times (400 \times 10^{-12})} \\ &= 39\text{M}\Omega \end{aligned}$$

¹³⁸A acceleration of '1 g' is equal to 980 cm/sec². Do not confuse this use of the symbol 'g' with the symbol 'G', which is used in this text for closed-loop gain.

This is a very large value of resistance. Commonly available resistance values top out around $10M\Omega$. This resistance would be made up of a series of smaller values or obtained from a specialty supplier. Furthermore, the construction technique must ensure that there are no parallel resistance paths on the circuit board caused, for example, by contaminants on the board.

We should also check the effect of bias current. Assuming a CMOS op-amp such as the National LM660, which has a bias current of $4pA$ maximum, the DC output voltage is simply the product of the bias current and feedback resistance:

$$\begin{aligned} V_{oDC} &= I_b R_f \\ &= (4 \times 10^{-12}) \times (39 \times 10^6) \\ &= 156\mu V \end{aligned}$$

This is probably small enough to be ignored.

The frequency response of the charge amplifier is shown in figure 310. The output rises with frequency at a rate of 20db/decade to the cutoff frequency of 10Hz and thereafter is flat at $100mV/g$.

13.18 AC-Coupled Inverting Amplifier

To set the stage for this subject, consider that we wish to amplify a dynamic microphone signal to drive a permanent magnet loudspeaker. The microphone is a transducer which converts sound waves into an electrical signal. Sound waves strike a moveable diaphragm, which moves a coil in the presence of a magnetic field. Movement of the coil in the magnetic field generates a small voltage.

The permanent magnet loudspeaker is a transducer which converts a time-varying current into a sound wave. Its construction is similar to a dynamic microphone. A current flows through a coil which is in the presence of a magnetic field. This generates a force, which moves a diaphragm, which creates sound waves. Electrically, both the microphone and loudspeaker are a low resistance coil of wire.

The microphone signal is in the order of millivolts, which is not powerful enough to drive a loudspeaker and so an amplifier is required. In this example, we'll assume it's an inverting amplifier which is to be operated from a single supply.

A first attempt at a suitable design is based on the single supply circuit of figure 262 on page 327. The microphone is attached to the input, and the loudspeaker is attached to the output. The result is shown in figure 311. Alas, for a variety of reasons, this will not work.

The root of the problem is the bias supply $V_{CC}/2$, connected to the non-inverting terminal of the op-amp. This is an inverting amplifier, so point A in figure 311 must be sitting at the same potential. The microphone has

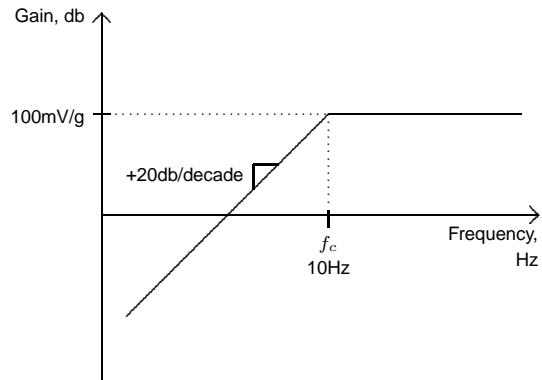


Figure 310: Charge Amplifier Frequency Response

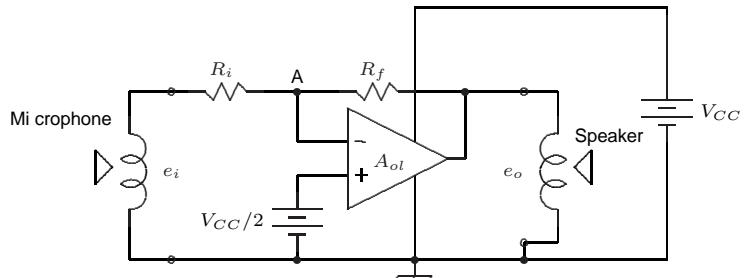


Figure 311: Defective Amplifier

a low internal resistance, which we'll treat as negligible compared to R_i . Then a DC current I_i flows through the microphone, where

$$I_i = \frac{V_{CC}/2}{R_i} \quad (550)$$

This is Not Good, because the direct current forces the diaphragm off centre which causes distortion.

This same current flows through R_f , which causes two more problems:

- The right end of R_f (the output of the op-amp) will now be at some positive potential. For example if $R_f = R_i$ the output is sitting at V_{CC} volts, that is, up against the positive power supply. This will clip a signal waveform.
- The positive DC potential across the loudspeaker will cause it to deflect inward or outward, and this will prevent it from operating properly.

As it is, this is not a usable design. One possible solution is to use a split supply for the op-amp, as shown in figure 329 on page 376. Then the non-inverting terminal of the op-amp can be grounded, point A is at zero volts, and there is no DC current through the microphone. The output voltage of the op-amp is near zero volts when there is no signal, so there is no DC current through the loudspeaker.

The other solution is to recognize that the signals of interest are alternating in nature. Audio signals span the frequency range from 20 to 20,000Hz. Capacitors can be used to block the troublesome DC currents in the circuit of figure 311 while allowing the audio AC currents to flow properly.

AC Amplifier, Improved

An improved circuit is shown in figure 312. A capacitor C_i is placed in series with the microphone. This capacitor appears as an open circuit for direct current. Consequently, there is no DC current through R_i or R_f . This eliminates the problem of DC current through the microphone. When the microphone generates an AC signal, (providing that the reactance of the capacitor is sufficiently small at that frequency), the AC current will flow through R_i and be amplified by the ratio $-R_f/R_i$ as in an inverting op-amp.

However, there is still a problem. There is no current through the feedback resistor R_f , so the output of the op-amp is sitting at a DC voltage equal to point A, that is, at $V_{CC}/2$ volts. This DC voltage will drive DC current through the loudspeaker. The solution is to put another capacitor C_o in series with the loudspeaker, as shown in figure 313.

Capacitor C_o prevents the DC output voltage of the op-amp from flowing through the loudspeaker. The output AC voltage sets up an AC current through the output capacitor and loudspeaker.

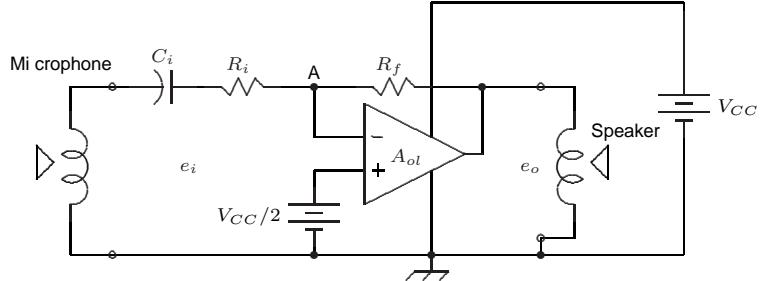


Figure 312: Better Amplifier

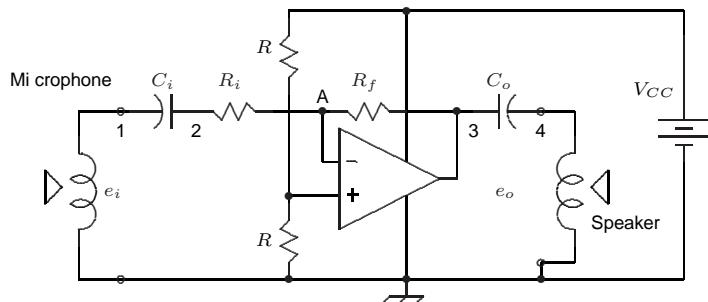


Figure 313: Working Amplifier

AC Amplifier, Working Version

In the final circuit of figure 313, the $V_{CC}/2$ voltage is created by a voltage divider. The amplifier accepts an input AC voltage from the microphone, amplifies it by the factor $-R_f/R_i$, and delivers that amplified AC signal to the loudspeaker. There are no DC currents through the microphone or loudspeaker. The output voltage of the amplifier sits halfway between the positive supply rail and ground, so the amplifier can deliver a peak AC signal of $V_{CC}/2$ volts into the loudspeaker before clipping. A representation of the signals in the amplifier is shown in figure 314. The numbered waveforms in figure 314 appear at the corresponding points in the schematic of figure 313.

DC Analysis

The circuit may be analysed in two stages, by considering the DC equivalent circuit (ie, the bias circuit) separately from the AC equivalent circuit. The equivalent circuit for the DC input is shown in figure 315(a).

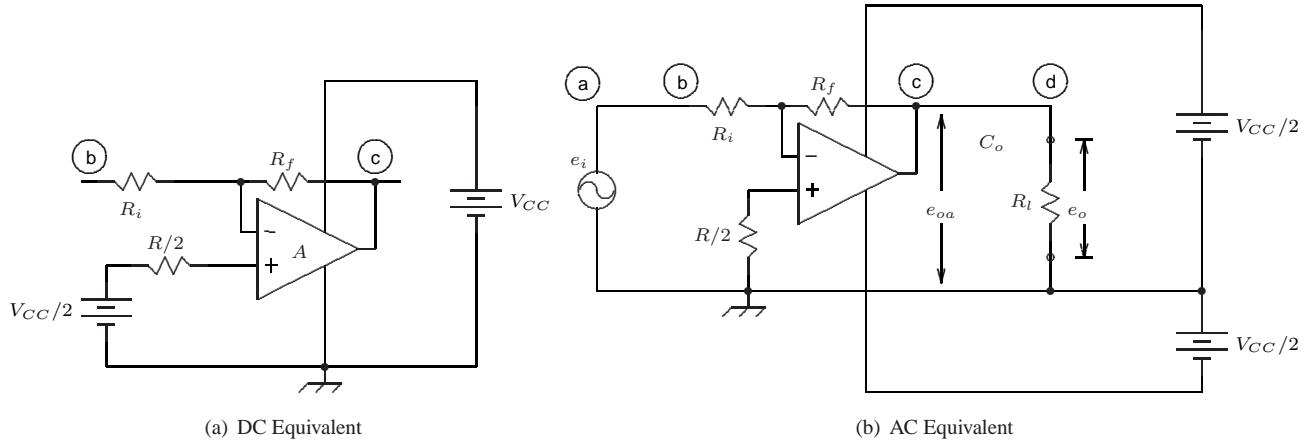


Figure 315: Inverting Amplifier, Single Supply, Equivalent Circuits

The capacitors have been replaced by open circuits, which how they appear to a direct current. The AC voltage source e_i could be replaced by a short circuit, but since it's in series with the input capacitance C_i , which is an open circuit, this is irrelevant. The bias supply voltage has been Thevenized into an open circuit voltage of $V_{CC}/2$ in series with a resistance $R/2$.

As required, the DC bias supply sets up a DC voltage of $V_{CC}/2$ at the non-inverting terminal, the inverting terminal, and points **b** and **c**.

AC Analysis

The equivalent circuit for the AC input is shown in figure 315(b). This time, the bias voltage source $V_{CC}/2$ has been replaced by a short circuit. The capacitors are assumed to be short circuits. (The values of capacitance

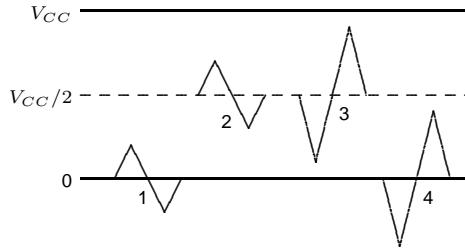


Figure 314: AC Coupled Inverting Amplifier Signals

required to achieve this are determined in section 13.19). Then the overall AC gain of the circuit is given by

$$\frac{e_o}{e_i} = -\frac{R_f}{R_i} \quad (551)$$

That is, providing that the capacitors are large enough, the AC gain is exactly the same as it would be for a split supply amplifier with resistors R_i and R_f (section 12.5).

13.19 AC-Coupled Inverting Amplifier, Frequency Response

In this section, we'll determine the frequency response of the amplifier that was described in section 13.18. For convenience, it's shown again in figure 316.

We'll focus on low frequencies which is the frequency region affected by the coupling capacitors. (The high frequency response is determined by completely different mechanisms.) We'll discuss the *amplitude* Bode plot, because that is most directly of interest in the design of an audio amplifier. Phase is a secondary concern (and it may not matter at all – this is an issue of ongoing controversy among audio professionals and enthusiasts). The expressions that we derive will provide the design equations for the coupling capacitors C_i and C_o .

The overall gain of the amplifier may be split into two parts, that is:

$$\begin{aligned} G &= \frac{e_o}{e_i} \\ &= G_{13} \cdot G_{34} \end{aligned} \quad (552)$$

where G_{13} is the gain between points **1** and **3** in figure 316 and G_{34} is the gain between points **3** and **4**. This is permissible because the output of the op-amp is a voltage source and the network attached to point **3** has no effect on the output voltage at **3**. This would not apply (to choose an example) for the cascade of two RC lowpass filters.

This is good news, because multiplication of gain factors implies that the Bode amplitude plot can be constructed by adding the individual gains of the two sections.

First Stage Gain

As we showed previously, when the reactance of the input capacitor C_i is small compared to the input resistance R_i , then it may be ignored. However at low frequencies the capacitive reactance increases until it becomes significant compared to R_i .

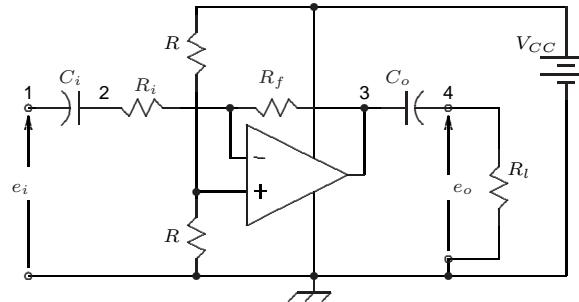


Figure 316: AC Coupled Inverting Amplifier

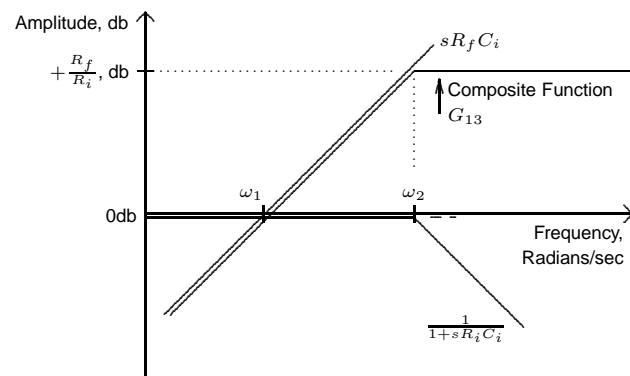


Figure 317: Bode Amplitude Plot for G_{13}

Then in figure 316 the overall gain of the amplifier from input (point 1) through to the output of the op-amp (point 3) is given by

$$\begin{aligned} G_{13} &= \frac{e_3}{e_1} \\ &= -\frac{Z_f}{Z_i} \end{aligned} \quad (553)$$

where

Z_i is the input impedance, equal to $X_{C_i} + R_i$
also equal to $\frac{1}{sC_i} + R_i$

Z_f is the feedback impedance, equal to R_f

Then, substituting for Z_i and Z_f in equation 553, we have:

$$\begin{aligned} \frac{e_3}{e_i} &= -\frac{R_f}{\frac{1}{sC_i} + R_i} \\ &= -\frac{sR_f C_i}{1 + sR_i C_i} \\ &= -\frac{s/\omega_1}{1 + s/\omega_2} \end{aligned}$$

where

$$\begin{aligned} \omega_1 &= 1/R_f C_i \\ \omega_2 &= 1/R_i C_i \end{aligned}$$

The numerator is a differentiator, rising at +20db/decade, passing through the zero db axis at $\omega_1 = 1/R_f C_i$. The denominator is a lowpass filter, which trails along the zero axis up to frequency $\omega_2 = 1/R_i C_i$ and thereafter breaks downward at a slope of -20db/decade. The composite function, (see section 11) forms a highpass filter, as shown on the Bode plot of figure 317.

For design purposes, we would put the lowest operating frequency (20Hz for audio) at ω_2 .

Second Stage Gain

The output capacitor C_o and load resistance R_l form an RC highpass filter (section 4.9).

$$\begin{aligned} G_{34} &= \frac{e_4}{e_3} \\ &= \frac{R_l}{X_{C_o} + R_l} \\ &= \frac{R_l}{\frac{1}{sC_o} + R_l} \\ &= \frac{sR_l C_o}{1 + sR_l C_o} \\ &= \frac{s/\omega_3}{1 + s/\omega_3} \end{aligned}$$

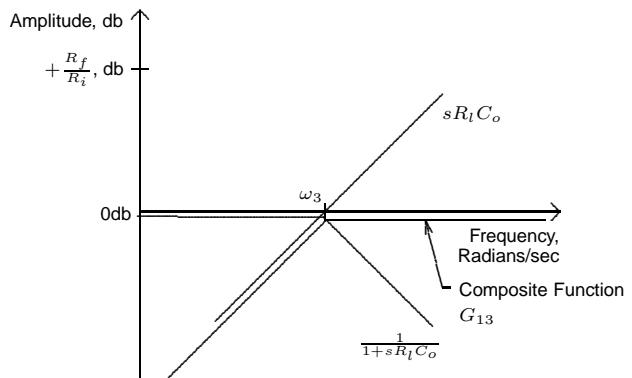


Figure 318: Bode Amplitude Plot for G_{34}

where

$$\omega_3 = 1/R_l C_o$$

The numerator is a differentiator, rising at +20db/decade, passing through the zero db axis at $\omega_3 = 1/R_l C_o$. The denominator is a lowpass filter, which trails along the zero axis up to the same frequency $\omega_3 = 1/R_l C_o$ and thereafter breaks downward at a slope of -20db/decade. The composite function G_{34} forms a highpass filter as shown on the Bode plot of figure 318.

Total Gain

The complete frequency response for the amplifier is obtained by adding the Bode plot G_{13} from figure 317 to the Bode plot G_{34} from figure 318. The two corner frequencies ω_2 and ω_3 would likely be placed to coincide with each other.

The final amplitude Bode plot is shown in figure 319 as the trace G_{14} . The rolloff of this curve is 40db/decade.

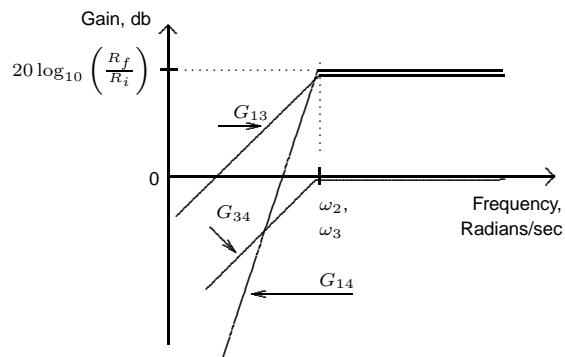


Figure 319: Completed Bode Amplitude Plot

13.20 AC-Coupled Non-Inverting Amplifier

Figure 320 shows the evolution of a non-inverting ac-coupled amplifier circuit from the inverting amplifier of section 13.18, figure 313.

- Figure 320(a) shows the original inverting amplifier from section 13.18.
- In figure 320(b), the input has been moved to the non-inverting side of the op-amp. The negative feedback network is grounded at point 1. The microphone is placed in series with the bias voltage supply $V_{CC}/2$, so that the non-inverting terminal of the op-amp receives the sum of the DC bias voltage and microphone signal AC voltage.

The amplifier non-inverting input terminal bias current must now flow through the microphone, but that is unlikely to be a problem because the bias current is very small. However, the microphone is floating at a DC voltage of $V_{CC}/2$, which may be undesirable. As well, the microphone AC current must flow through the bias supply.

- An alternative method of connecting the microphone is shown in part (c). Resistor R_B in series with the bias supply $V_{CC}/2$ prevents the AC current of the microphone from being short-circuited through its zero internal resistance. Notice that we have moved capacitor C_i to the new input position. This capacitor allows AC voltage from the microphone to reach the non-inverting terminal of the op-amp while blocking the bias supply current from flowing through the low-resistance microphone. The original capacitor C_i has been renamed to C_f since it's now part of the feedback network.
- Part (d) of the figure is a redrawn version of part (c) and shows how a voltage divider creates the bias supply (where $R = 2R_b$).

DC Equivalent Circuits

At low frequencies, capacitor C_f is an open circuit. The DC equivalent circuit is as shown in figure 321(a).

Then the feedback signal is equal to the output signal, the feedback gain B is unity, and the closed-loop gain, which is $1/B$, is also unity.

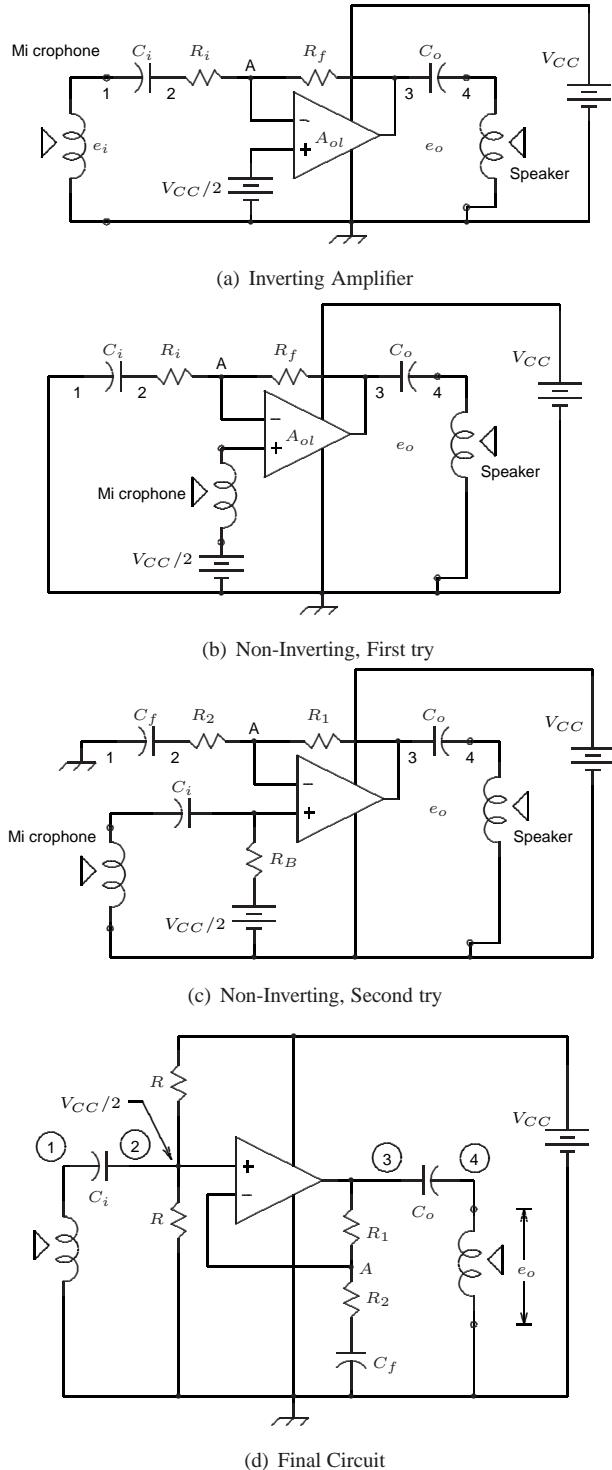


Figure 320: Evolution of the Non-Inverting AC Coupled Amplifier

Consequently the $V_{CC}/2$ DC voltage established at the non-inverting input will also appear at the output.

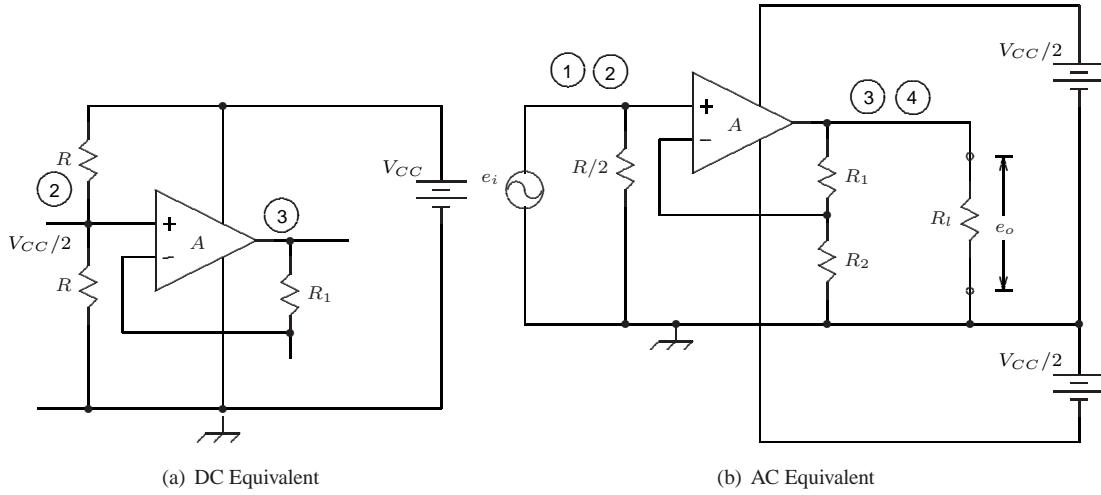


Figure 321: Non-Inverting AC Coupled Amplifier Equivalent Circuits

AC Equivalent Circuit

At high frequencies, capacitors C_i , C_f and C_o are all short circuits. Then the AC equivalent circuit is as shown in figure 321(b). The feedback factor is given by

$$B = \frac{R_2}{R_1 + R_2} \quad (554)$$

and the gain at that frequency is given by $1/B$, so

$$\begin{aligned} \frac{e_{o(ac)}}{e_{i(ac)}} &= \frac{R_1 + R_2}{R_2} \\ &= 1 + \frac{R_1}{R_2} \end{aligned}$$

which is the usual equation for voltage gain for the non-inverting amplifier.

The progression of signals through the amplifier is shown in figure 322, with the numbers keyed to the schematic of figure 320(d).

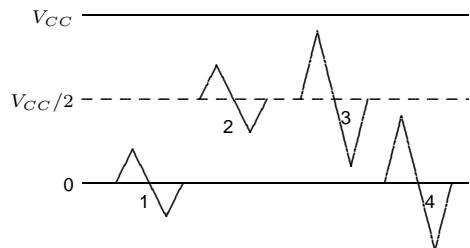


Figure 322: AC Coupled Non-Inverting Amplifier Signals

13.21 AC-Coupled Non-Inverting Amplifier, Frequency Response

Determining the frequency response of this amplifier follows the same general pattern as section 13.19.

The circuit is shown in figure 323. To determine the frequency response, we break it into three sub-circuits, determine the frequency response of each subcircuit, and then combine these responses to get the overall response.

$$\begin{aligned} G_{14} &= \frac{e_o}{e_i} \\ &= G_{12} \cdot G_{23} \cdot G_{34} \end{aligned}$$

where G_{14} is the overall frequency response, G_{12} , G_{23} and G_{34} are the responses of the individual sections in volts per volt as keyed to the numbers in figure 323.

If each of these gains is converted into decibel format, then

$$G_{14(db)} = G_{12(db)} + G_{23(db)} + G_{34(db)} \quad (555)$$

that is, we can find the individual responses on a Bode plot and then add them to get the total response.

Figure 323 shows the amplifier circuit and its three sub-sections. Above the circuit is a sketch of the Bode Amplitude Plot. The Bode Plot break frequencies (in radians/sec) are related to the circuit component values.

First Gain Section

The first gain section, G_{12} , is a highpass RC network. The resistor is equal to the equivalent input resistance of the bias divider, $R/2$. The capacitor is C_i . So the break frequency is equal to $1/\frac{R}{2}C_i$ radians per second.

Second Gain Section

The middle gain section, G_{23} , requires some analysis. The gain of this section is given by:

$$\frac{e_3}{e_2} = 1 + \frac{Z_1}{Z_2}$$

where

e_2 is the AC input voltage to the op-amp non-inverting terminal

e_3 is the AC output voltage of the op-amp

$$Z_1 = R_1$$

$$\begin{aligned} Z_2 &= R_2 + \frac{1}{sC_f} \\ &= \frac{1 + sR_2C_f}{sC_f} \end{aligned}$$

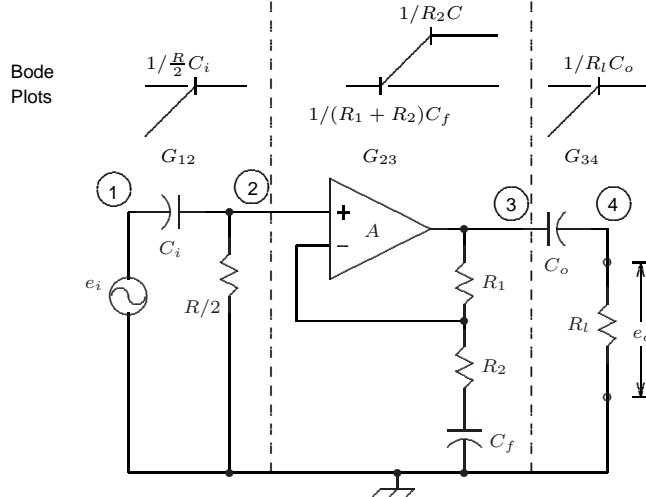


Figure 323: AC Coupled Amplifier Sub-Circuits

Then

$$\begin{aligned}\frac{e_3}{e_2} &= 1 + \frac{R_1}{\frac{1+sR_2C_f}{sC_f}} \\ &= \frac{1+s(R_1+R_2)C_f}{1+sR_2C_f}\end{aligned}\quad (556)$$

The Bode Plots for G_{23} are shown in figure 324. The Bode Plot for the numerator runs along the 0db axis up to frequency $1/(R_1 + R_2)C_f$. Then it breaks upward to a slope of +20db/decade. The Bode plot for the denominator runs along the 0 db axis up to a frequency $1/R_2C_f$. Then it breaks downward to a slope of -20db/decade. The net result is a *shelving characteristic*, as it is called, that has a gain of 0db at low frequencies and a gain of $(R_1 + R_2)/R_2$ db at high frequencies.

This can be confirmed by inspection of equation 556 above:

- Recall that $s = j\omega$. At frequencies where the s terms are much less than 1, equation 556 degenerates to 1/1, that is, 0db.
- At high frequencies, the s terms are much larger than 1. Then equation 556 becomes:

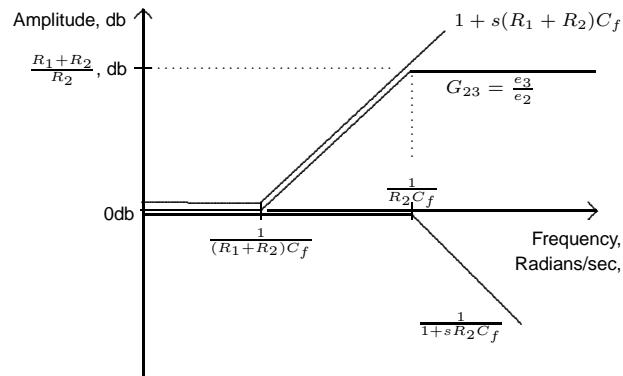


Figure 324: Bode Plot for G_{23}

$$\begin{aligned}\frac{e_3}{e_2} &= \frac{1+s(R_1+R_2)C_f}{1+sR_2C_f} \\ &\approx \frac{s(R_1+R_2)C_f}{sR_2C_f} \\ &= \frac{(R_1+R_2)}{R_2}\end{aligned}$$

Third Gain Section

The third gain section, G_{34} , is another highpass RC network. This time, the resistor is the load resistance R_l and the capacitance is C_o . So the break frequency is equal to $1/R_lC_o$ radians per second, as shown in figure 325.

Overall Amplitude Response

The completed amplitude response for an AC-coupled non-inverting amplifier, where the three cutoff frequencies have been made to coincide and the gain of the amplifier G_{23} is 20db, is shown in figure 325.

Design Example

Suppose we wish to construct an AC amplifier of the type shown in figure 323, to have a 3-db down frequency of 20Hz and a voltage gain of 20db. The load resistance is 1000Ω . The bias resistors R are each $1M\Omega$. The supply voltage is +12 volts.

- We need to give some thought to the placement of the cutoff frequencies of the three highpass filters. There's no reason to make the cutoff frequencies different, so they will occur at the same frequency.

If all three highpass filters are positioned with their cutoff frequencies at 20Hz, then each filter will have an attenuation of 3db at 20 Hz, with a cumulative attenuation of 9db. This doesn't meet spec, so we have to position the cutoff frequencies somewhat lower than 20Hz.

- A highpass filter is 3db down at the corner frequency, and progressively approaches the horizontal axis above the corner frequency. At twice the corner frequency, the curve is within 1db of the horizontal axis. So, if we establish the cutoff frequency at 10Hz, then each highpass filter will be 1db down at 20Hz, for a cumulative attenuation of 3db at 20Hz, which is what is required.

The Bode plot for the placement of the three gain curves and the total gain is shown in figure 325.

- The high frequency voltage gain is 20db, which is 10 volts/volt, so $R_1/R_2 = 9$.
- Now we can calculate the various capacitance values. The reactance of capacitor C_i is equal to $R/2$ at 10Hz, which puts its value at about $0.03\mu F$.

- The values of R_1 , R_2 and C_f are defined by two equations: one sets the gain, and the other sets the cutoff frequency. Consequently, we have two equations in three unknowns, and we must choose one of these values. Larger values of resistance result in smaller values of capacitance, which is always desirable. Since we're already using $1M\Omega$ for resistors R , let's choose that value for R_1 .

Since the voltage gain is $10v/v$, R_2 must be $R_1/9 = 111k\Omega$.

Referring to figure 325, the reactance of C_f is equal to R_2 at 10Hz, which makes it about $0.14\mu F$. (This could be written as $140n$, for 140 nanoFarads).

- The reactance of C_o is equal to R_l at 10Hz, which puts its value at $16\mu F$.

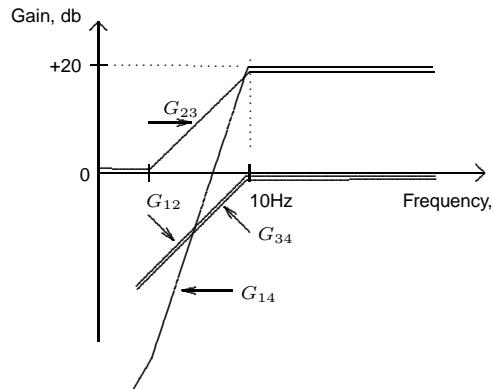


Figure 325: Design Example

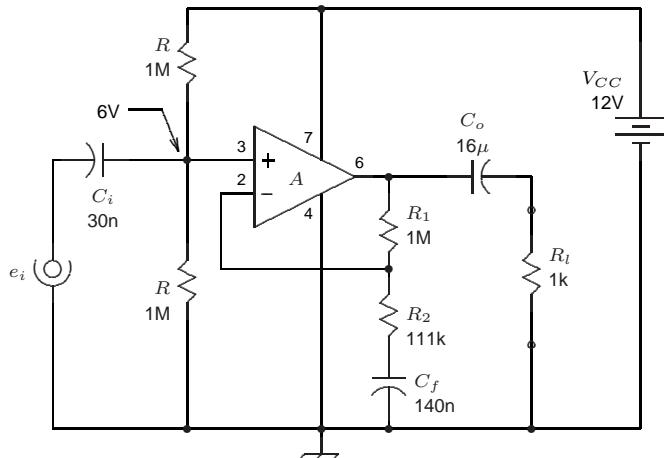


Figure 326: Example Design Completed

- As shown in figure 325, the gain falls off very rapidly below 10Hz. There are three highpass filters at work, so the gain will change at a rate of 60db/decade.

The complete circuit of the amplifier with component values is shown in figure 326.

The op-amp pin numbers shown in figure 326 are representative of many single op-amps in an 8-pin DIP package. The calculated values of the components are shown in figure 326. In practice, these would be changed to the nearest standard values.

13.22 AC-Coupled Amplifier, Miscellaneous Issues (Advanced)

Voltage Offset in the AC Coupled Amplifier

The *offset voltage* of an operational amplifier is a small voltage difference (typically a few millivolts) that appears at the input terminals when the output is zeroed (section 21.3). In a DC coupled amplifier, this offset voltage is indistinguishable from the signal so it creates an error term in the output voltage. As explained in section 21.3, offset voltage in a DC coupled amplifier is magnified by the noise gain G_N where:

$$G_N = 1 + \frac{R_f}{R_i} \quad (557)$$

This is problematic in a high gain DC coupled amplifier. Offset voltage changes with temperature, and this will cause the output of the amplifier to drift significantly.

In an AC coupled amplifier, offset voltage is not a problem. The equivalent circuit for the offset voltage is shown in figure 327. Offset voltage is modelled by a voltage source V_{os} which can be placed in series with either op-amp input. In this case, we've put it as in series with the non-inverting input. The offset voltage simply adds to the bias voltage. Since the offset voltage is much smaller than the bias voltage its effect is negligible.

Consequently a high gain AC coupled amplifier does not suffer from offset-induced drift problems and is therefore much easier to build.

Comparing the Inverting and Non-inverting AC Amplifiers

Both the inverting and non-inverting AC-coupled amplifier circuits are shown in figure 328. In many applications, whether the phase of the AC signal is inverted or not is unimportant, so either circuit could be chosen. Which is preferable?

Where fidelity is critical, the inverting circuit has an advantage. The input terminals of the op-amp are held at the DC bias voltage, $V_{CC}/2$. As a consequence, the op-amp is not required to deal with common-mode AC signals. As Williams points out in [103], the common-mode AC signal in the non-inverting circuit creates some distortion products that can be avoided in the inverting circuit.

When ultra-low distortion is not a requirement, either configuration could be used.

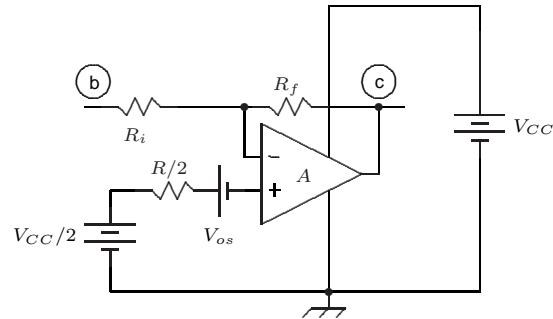


Figure 327: Offset Voltage in AC Coupled Amplifier

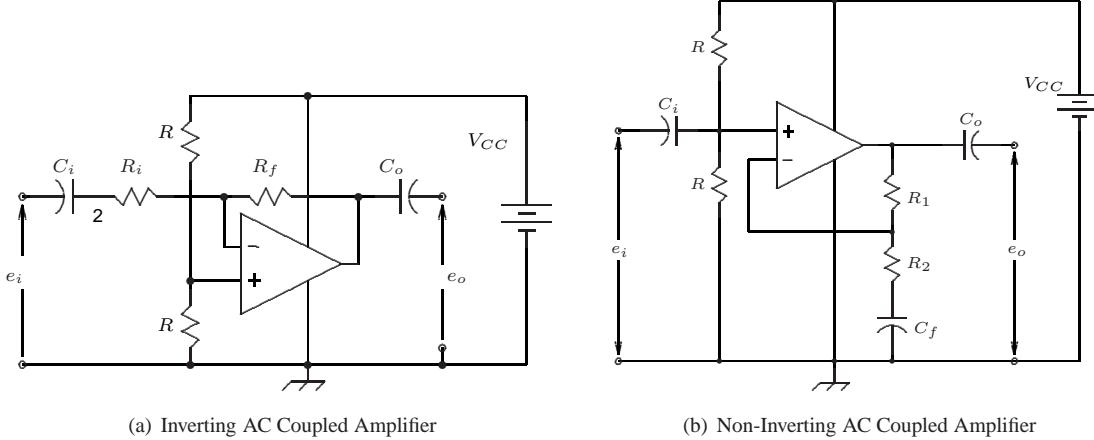


Figure 328: Inverting and Non-Inverting AC Coupled Amplifiers

Split Supply or Capacitor Coupling?

For the design of an audio amplifier, which of these two methods is better: split supply without coupling capacitors or single supply with coupling capacitors?

Historically, for reasons of cost audio amplifiers have used a single power supply with capacitor coupling. However, there has been a move toward split supplies because, in high fidelity applications, the coupling capacitors may introduce some distortion.

As well, a split supply system does not have a turn-on transient. To understand this last point, consider what happens when either of the amplifiers in figure 328 is first powered on. The input and output coupling capacitors must charge to a voltage of $V_{CC}/2$ volts and as they do so they conduct a transient current. In an audio system, this transient current appears as a 'pop' in the loudspeaker which is unattractive and potentially harmful to the loudspeaker.

Some audio systems use a split supply, which eliminates the requirement for coupling capacitors and helps prevent a power-supply transient in the output. The design deliberately rolls off the gain below 20Hz to avoid multiplication of the offset voltage by the amplifier noise gain. This reduces the DC component in the output voltage, which might cause problems in a loudspeaker load. A system of this sort, using the non-inverting amplifier configuration, is shown in figure 329. Capacitor C_f sets the low frequency cutoff.

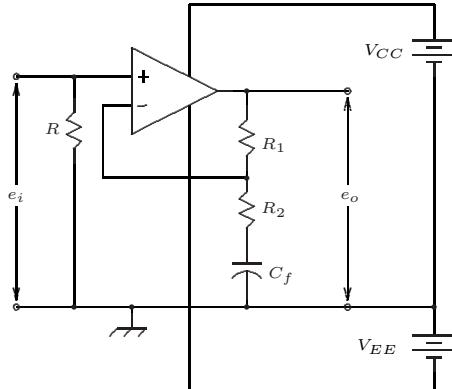


Figure 329: Split Supply with Limited Frequency Response

Reference

A number of single-supply audio circuits are shown in Carter [104], [105], [106].

13.23 Input and Output Resistance

In this section we see the effect of negative feedback on the input and output resistance¹³⁹ of different types of amplifiers. We see that negative feedback raises or lowers the input and/or output resistance by some factor related to A , the open-loop gain.

The op-amp open-loop gain A is extremely large at low frequencies. When some resistance is multiplied by A , the result is a value that is much larger than other resistances in the circuit – effectively, an open circuit. Conversely, when a resistance is divided by A , the result is effectively zero. In effect, the results are very close to the ideal and may usually be treated as zero or infinite.

However, the open-loop gain of the op-amp declines at high frequencies (section 23). If the input or output resistance must maintain their values at a high frequency, they should be calculated using the open-loop gain at that frequency.

In this section we examine:

- Input resistance of the non-inverting voltage amplifier
- Output resistance of the non-inverting voltage amplifier
- Input resistance of the transresistance amplifier (current to voltage generator)
- Output resistance of the transconductance amplifier (voltage to current generator)

These examples are not exhaustive but they are representative and show the technique.

Measurement Technique

The measurement of input resistance and output resistance is accomplished by applying a test voltage source, measuring the current, and calculating the ratio of the two. Alternatively, one can apply a current and measure the voltage.

Although it is possible in some circumstances to use a DC source for the measurement, there are two complications:

- The resistance of interest may be hidden behind a coupling capacitance, as shown in figure 330(a). In that case, a DC measurement will read open-circuit instead of the resistance.
- The resistance may be dependent on the operating conditions of the amplifier. Applying a DC source disturbs the operating condition and so the measurement may not be valid. Resistance r_x is written in lower case to indicate that it is an incremental resistance, that is, dependent on other currents and voltages in the circuit.

To avoid these problems, the test source should be an AC source as shown in figure 330(b). The frequency of the AC source is chosen to be high enough that the impedance of capacitor C is low compared to resistance r_x . The ratio of AC voltage to AC current will give the value of r_x . The test signals are kept small in order not to disturb the operating condition of the circuit.

Now, when doing an analysis of some circuit to determine input or output resistance, it is of critical importance to get the relative polarities of the various sources correct. If AC sources are used, then they must be labelled with a reference polarity.

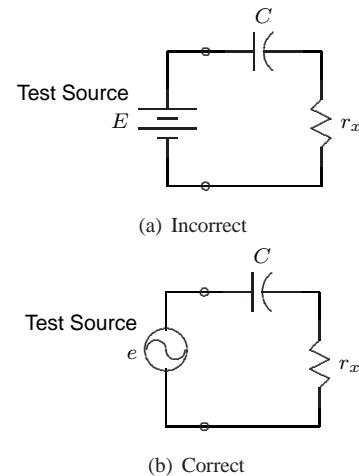


Figure 330: Measuring R_x

¹³⁹More generally, the input and output *impedance*.

The approach used here is to consider that the test source is a DC source that is changing in magnitude by some amount Δ . For example, a DC voltage source changes abruptly by an amount ΔE . There is a corresponding step in current ΔI , and the resistance of interest is

$$r_x = \frac{\Delta E}{\Delta I} \quad (558)$$

A step change in voltage is transparent to any coupling capacitors and the voltage step is chosen to be small enough not to materially affect the operating conditions of the amplifier.

This technique simplifies tracking the direction of the various currents and voltages that result from the test signal. The overall effect of feedback – to raise or lower some resistance – can be determined by tracing the signals on the schematic.

Non-inverting Voltage Amplifier: Input Resistance

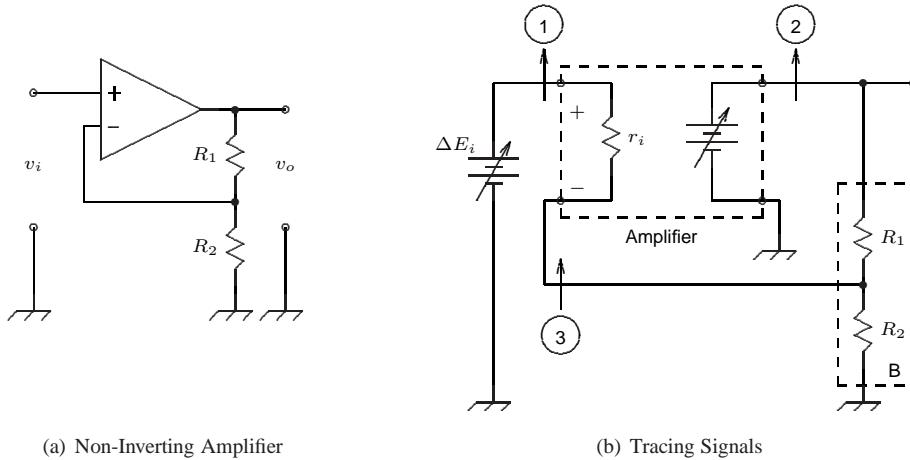


Figure 331: Non-Inverting Amplifier: Input Resistance

A non-inverting amplifier is shown in figure 331(a).

First, we determine the effect of the feedback. Does it raise or lower the parameter of interest? This can be determined by tracing signals around the feedback loop.

Figure 331(b) shows an equivalent circuit for the purpose of tracing the effect of an input signal. The input resistance r_i appears between the inverting and non-inverting terminals of the op-amp. It's usually large enough to be ignored, but in this example we'll take it into consideration.

Now consider that the input signal is changed by some amount $+\Delta E_i$. This raises the voltage at point 1. The op-amp senses this change in input voltage and raises the output at point 2. The feedback network samples the output and feeds some fraction of the output back to the input at point 3.

The signals at point 1 and point 3 move in the same direction. However, the change in the feedback signal at point 3 must be very slightly less than the change at point 1 in order to develop an error voltage that causes the output of the amplifier to increase. So there is a very small increase in the voltage across the input resistance r_i .

Now consider the effect of the feedback on the input current. When the input voltage increases, it tends to increase the current through resistance r_i . However, the feedback signal moves in the same direction as the input signal, so it tends to decrease the current in r_i . Consequently, feedback *reduces* the current in the input resistance, and this makes the effective input resistance r'_i larger than the actual physical resistance r_i .

Analysis

Figure 332 shows a circuit to determine the effective input resistance r'_i .

The equations are:

$$r'_i = \frac{\Delta E_i}{\Delta I_i} \quad (559)$$

$$\Delta I_i = \frac{\Delta E_i - \Delta E_f}{r_i} \quad (560)$$

Assuming that

- the input resistance r_i is much larger than the internal resistance of the R_1, R_2 voltage divider, and
- the output resistance r_o is much less than the resistance of the R_1, R_2 voltage divider,

then we can write:

$$\Delta E_f = B\Delta E_o \quad (561)$$

where the negative feedback sensor gain B is given by

$$B = \frac{\Delta E_f}{\Delta E_o} = \frac{R_2}{R_1 + R_2} \quad (562)$$

and the corresponding change in output voltage is given by

$$\Delta E_o = A(\Delta E_i - \Delta E_f) \quad (563)$$

Combining equations 559 through 563 to solve for r'_i , we obtain the final result:

$$\begin{aligned} r'_i &= (1 + AB)r_i \\ &\approx (AB)r_i \text{ providing } AB \gg 1 \end{aligned} \quad (564)$$

In words,

the effective input resistance r'_i is equal to the actual input resistance r_i multiplied by the loop gain AB .

Non-inverting Voltage Amplifier: Output Resistance

First, a general view: in section 10.3 on negative feedback, we saw that negative feedback reduces the effect of a disturbance by an amount equal to the loop gain AB . Consider that the amplifier has some output resistance r_o and that a load resistance is attached to the output of the amplifier in figure 331(a).

Load current will flow out of the amplifier into the load resistance, causing the output voltage to drop. This is seen by the negative feedback system as a disturbance, so the system will compensate in such a way that the effect is minimized. In this case, it means that the change in output signal is reduced by an amount equal to the loop gain AB or, what is the same thing, that output resistance is reduced by the loop gain AB .

Notice that the negative feedback need not *always* reduce the output resistance. On the contrary, if the system senses the output current, then the effect of negative feedback is to keep the output current as constant as possible and this *raises* the output resistance of the system.

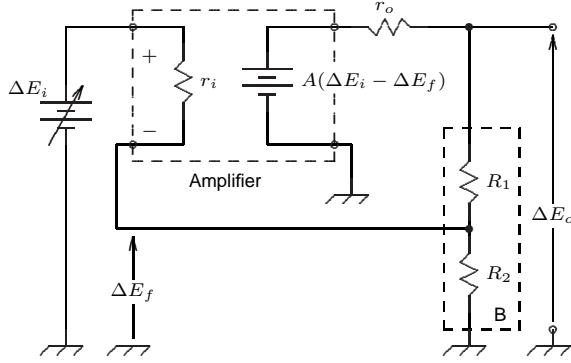


Figure 332: Input Resistance: Analysis

Signal Tracing

Now we will use figure 333(a) to trace the signals that determine the effective output resistance r'_o of the circuit.

Consider that the voltage source ΔE_o is applied to the output. This causes:

1. the output voltage to move in the positive direction
2. the feedback voltage ΔE_f to move in the positive direction
3. the amplifier output $A\Delta E_f$ to move in the *negative* direction.

The net effect of the feedback system is to *increase* the voltage across the output resistance r_o , and the current through it. This *reduces* the effective output resistance.

Analysis

Now a full-blown analysis, using figure 333(b). We'll assume that:

- the resistances R_1 and R_2 are much larger than r_o . Then most of the output current i_o flows into the op-amp and the current in the sensor network can be ignored. (Typically, r_o is in the order of a few tens of ohms and the divider resistances are in the order of thousands of ohms.)
- r_i is much larger than the internal resistance of the R_1, R_2 voltage divider.

Then the relevant equations are:

$$r'_o = \frac{\Delta E_o}{\Delta I_o} \quad (565)$$

$$+ \Delta E_o - \Delta I_o r_o + A\Delta E_f = 0 \quad (566)$$

$$+ \Delta E_f = B\Delta E_o \quad (567)$$

where (once again) the negative feedback sensor gain B is given by

$$B = \frac{R_2}{R_1 + R_2} \quad (568)$$

Combining these three equations and solving for the effective output resistance r'_o , we obtain:

$$\begin{aligned} r'_o &= \frac{r_o}{1 + AB} \\ &\approx \frac{r_o}{AB} \text{ providing } AB \gg 1 \end{aligned} \quad (569)$$

In words,

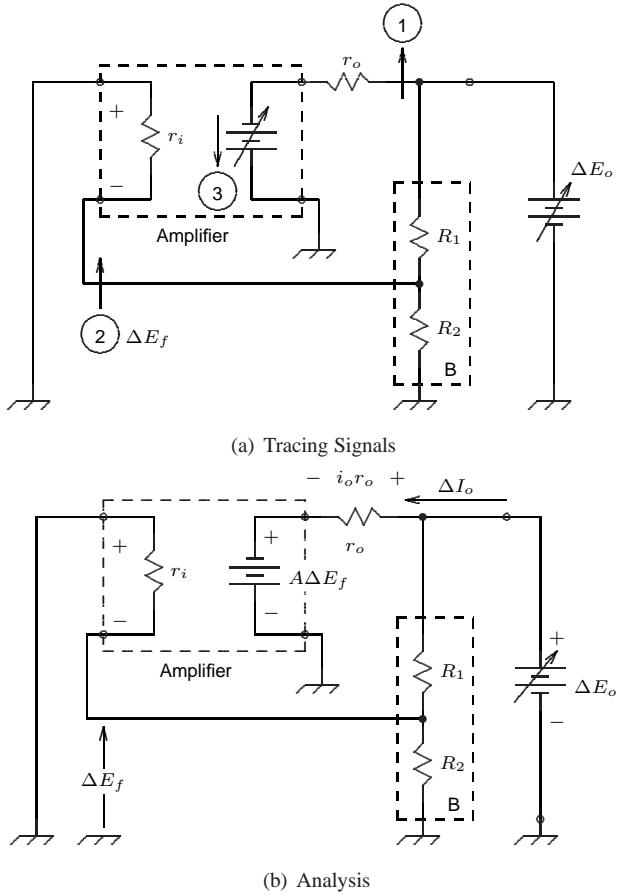


Figure 333: Output Resistance

the effective output resistance r'_o is equal to the actual output resistance r_o divided by the loop gain AB .

Current-Voltage Converter: Input Resistance

Section 13.9 introduced the current-voltage converter, which is shown here in figure 334. The output voltage E_o is given by:

$$E_o = R_f I_i \quad (570)$$

Ideally, in order to have no effect on the current source, the input resistance r'_i of the current-voltage converter should be much less than the internal resistance of the source.

As shown in section 13.1 on the Miller Effect, the effective input resistance r'_i is the feedback resistance R_f divided by $A + 1 \approx A$, the open-loop gain. (This resistance appears in parallel with the physical resistance r_i between the input terminals of the op-amp. Since the physical resistance is in the order of megohms, it can be ignored.)

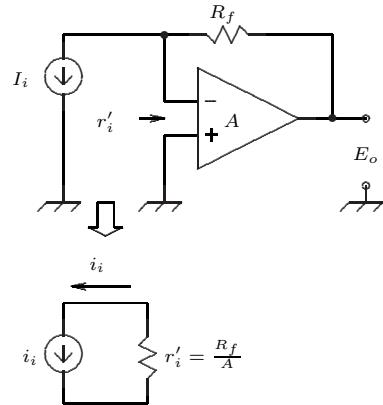
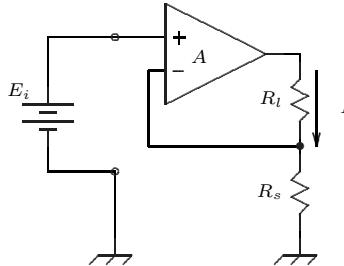


Figure 334: Current-Voltage Converter: Input Resistance

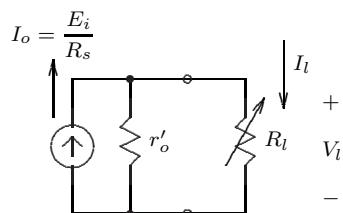
$$\begin{aligned} r'_i &= \frac{R_f}{1+A} \\ &\approx \frac{R_f}{A} \text{ providing } A \gg 1 \end{aligned} \quad (571)$$

For example, if the feedback resistance R_f is $10\text{k}\Omega$ and the open-loop gain A is 10^5 volts/volt, then the effective input resistance is 0.1Ω , very low indeed.

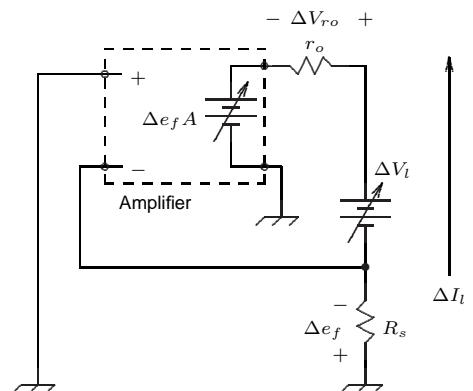
Voltage-Current Converter: Output Resistance



(a) Circuit



(b) Equivalent



(c) Analysis

Figure 335: Current Generator, Output Resistance

A constant current generator (section 13.15) is shown in figure 335(a). This generates a load current I_l which is given by

$$I_l = \frac{E_i}{R_s} \quad (572)$$

in the load resistance. Ideally, the load resistance R_l and the voltage V_l across the load have no effect on this current. However, as the load voltage V_l increases, some of the generator current will flow through the internal resistance r'_o that is in parallel with the load, and the load current will decrease. The larger the output resistance compared to the load resistance, the closer the source to an ideal current generator.

Analysis

The equivalent circuit for analysis is shown in figure 335(c). Let's check that this circuit makes sense.

- Suppose the load voltage increases in the direction shown.
- This drives the increased load current ΔI_l up the page.
- The load current creates a feedback voltage Δe_f which is negative with respect to ground.
- This negative feedback voltage is inverted and amplified in the op-amp. The amplified feedback is in such a direction as to *reduce* the original increase in load current. Consequently, the feedback action *raises* the output resistance of the circuit.

This seems reasonable, so now we can apply analysis to the circuit.

The definition of output resistance is

$$r'_o = \frac{\Delta V_l}{\Delta I_l} \quad (573)$$

By KVL,

$$-\Delta e_f + \Delta V_l - \Delta V_{ro} - \Delta e_f A = 0 \quad (574)$$

Also, by Ohm's law:

$$\Delta e_f = \Delta I_l R_s \quad (575)$$

$$\Delta V_{ro} = \Delta I_l r_o \quad (576)$$

Substitute for Δe_f and ΔV_{ro} from equations 575 and 576 into equation 574 and rearrange equation 574 into the form of equation 573:

$$\begin{aligned} r'_o &= \frac{\Delta V_l}{\Delta I_l} \\ &= r_o + R_s(1 + A) \\ &\approx AR_s, \text{ if } r_o \ll AR_s \text{ and } A \gg 1 \end{aligned} \quad (577)$$

That is, the output resistance is simply the voltage gain A times the sensing resistance R_s .

Example

The circuit shown in figure 336 is a 0.5mA current source. What is the variation in current as R_l slowly changes from zero to 16k Ω ?

Solution

For the LM324 op-amp, the voltage gain A is 10⁵ volts/volt. The output resistance r_o is not given but it's reasonable to assume zero.

Then by equation 577 the output resistance of this current generator is:

$$\begin{aligned} r'_o &\approx AR_s \\ &= 10^5 \times 2000 \\ &= 2 \times 10^8 \Omega \end{aligned}$$

With a load resistance of 16k Ω , the voltage across the load is 8 volts. As shown in figure 335(b), this voltage diverts current through the parallel resistance r'_o . The current through r'_o , (call it I_ϵ for error current) is then given by:

$$\begin{aligned} I_\epsilon &= \frac{V_l}{r'_o} \\ &= \frac{8}{2 \times 10^8} \\ &= 40\text{nA} \end{aligned}$$

This current subtracts from the load current. However, it is a negligible error – the error due to an offset voltage of 10mV (section 21.3) is orders of magnitude larger than this.

Generalizing

The examples of this section can be generalized, see for example [107] and [78]. The general methods rely on a precise identification of the amplifier, the feedback network, and their method of interconnection and this can be difficult for a specific circuit. On the other hand, injecting a test signal and measuring the ratio of test voltage to current is relatively straightforward and reliable, so that is the method presented here.

13.24 Summary of Amplifier Circuits

This section presents a generalized, black-box amplifier. This is relevant to the systematic consideration of amplifier designs.

Considering amplifiers in general and all the alternatives for input and output resistance in an amplifier, we have the amplifier taxonomy shown in the chart on page 387. There are four types of amplifiers, each of which has a different combination of voltage or current input and voltage or current output.

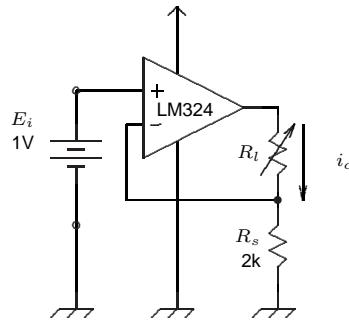


Figure 336: Output Resistance: Example

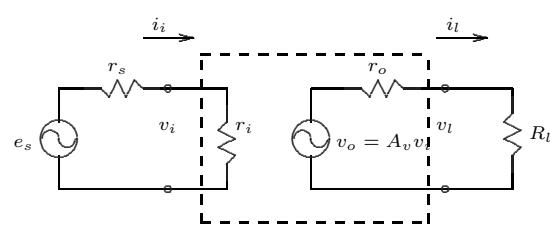


Figure 337: Generalized Amplifier

The *Example* column shows one possible implementation of the circuit using an operational amplifier. The op-amp is inherently a voltage amplifier but with the appropriate feedback network it can implement a high or low input resistance and a high or low output resistance.

In the interest of simplicity, the current generator output of the current amplifier and transconductance amplifier are shown with floating load resistances. It is possible with slightly more circuitry to drive current loads that have one terminal grounded. (Section 18.3 shows an example.)

Application: Transresistance Amplifier

The transresistance amplifier (more commonly and generally known as a *transimpedance* amplifier) presents a short circuit to the input signal and provides output from a low-resistance voltage source. Consequently, the transimpedance amplifier can be referred to as a *current to voltage converter* in which the constant of proportionality is the transimpedance gain R_m volts/amp. (More generally, when phase shifts are involved this would be the impedance Z_m). The gain may be expressed in volts/amp or more compactly in ohms.

An example application is shown in figure 338. (See also section 27.3 on page 774). The input device is a reverse-biased photodiode which detects pulsed illumination at the receiving end of a fiber-optic cable. The op-amp creates a virtual earth point at the inverting terminal, which appears as a short-circuit to ground to an external signal. Illumination of the diode causes current to flow into the virtual earth point, which then causes an output voltage. This arrangement speeds up operation, which allows a higher bandwidth for the system. Because there is no voltage swing across the diode, the junction capacitance of the diode is not charged and discharged and so the signal rise-time is faster than it would be with a resistive load. (The amplifier must have the requisite speed to take advantage of this effect.)

The transimpedance amplifier then outputs a voltage signal from a low-resistance source that can be applied to the input of an A/D converter or for some other purpose.

The Voltage Amplifier Approximation

Many signals are conveyed and processed as voltages. As a result, it is common to require an amplifier that will accept a signal voltage and produce a larger version of that voltage into a resistance.

A generalized amplifier is shown inside the dotted lines in figure 337. The input is driven by a source which has an open-circuit voltage of e_s and an internal resistance r_s . The output of the amplifier drives a load R_l .

The amplifier presents an input resistance r_i to the source. It amplifies the input voltage by a factor A_v and drives this voltage into the load via the amplifier output resistance¹⁴⁰ r_o .

If the amplifier of figure 337 has an input resistance r_i that is much larger than the internal resistance r_s of the source, then the signal voltage will be unaffected when the amplifier is attached to the source. That is, $v_i \approx e_s$. Similarly, if the output resistance r_o of the amplifier is

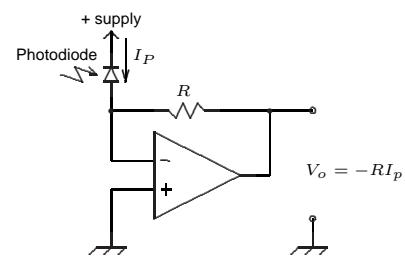


Figure 338: Transimpedance Amplifier Application

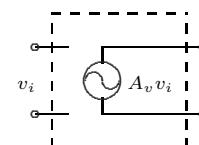


Figure 339: Ideal Voltage Amplifier

¹⁴⁰Some generalized amplifier models include a component that accounts for the effect of the output on the input, a *reverse transfer* parameter. This is an important effect at radio frequencies but generally negligible at low frequencies and DC.

much smaller than the load resistance R_L , then the output voltage of the amplifier will be unaffected when the amplifier is attached to the load resistance. That is, $v_l \approx v_o$.

This is the case for the operational amplifier: it has very large input resistance compared to most cases of source resistance, and it has a small output resistance compared to most loads. There are relatively few circuits where the input resistance and output resistance have an effect on the behaviour of the circuit, and so we often ignore them.

When these approximations hold, then $v_l/e_s \approx A_v$. That is, the overall voltage gain between the open-circuit voltage of the source and voltage across the load is approximately equal to the voltage gain A_v of the amplifier itself. Then the amplifier may be modelled as shown in figure 339. The input resistance is effectively infinite (an open circuit). The output resistance is zero, so the voltage source behaves as an ideal voltage source. The voltage gain is A_v volts/volt.

Application: Transconductance Amplifier

An application of the transconductance amplifier is shown in figure 340. The transconductance amplifier outputs a current that is proportional to the input voltage. The torque of a DC motor is proportional to its current. If the torque of a motor must be controlled then it should be driven from a current source, which could be a transconductance amplifier. The input voltage to the amplifier will then control the motor current and torque. This is particularly relevant for motors that are designed not to rotate but to apply a torque to some mechanical load.

As a further application, we will see that the BJT and JFET amplifier can be modelled as a transconductance amplifier.

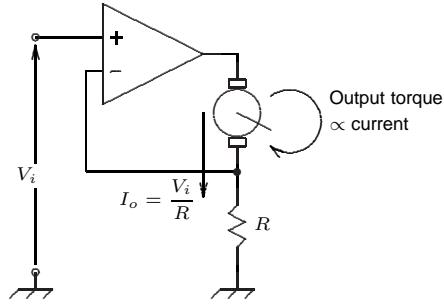


Figure 340: Transconductance Amplifier Application

Measurement Conditions

When we speak of the *voltage gain* of an amplifier, we are relating the *open-circuit* output voltage of the amplifier to the input voltage. Similarly, the *current gain* relates the *short circuit* output current of the amplifier to the input current. Both types of load cannot be applied at the same time so voltage gain and current gain are measured under different conditions.

Once a set of amplifier parameters has been determined the behaviour of a circuit with source resistance and load resistance can be determined.

Relating the Amplifier Parameters

A generalized amplifier is shown in figure 341(a). The amplifier is arranged for a measurement of voltage gain by connecting the output into an open circuit. By inspection, the amplifier parameters are:

Input resistance	r_i
Output resistance	r_o
Voltage gain	A_v

Now consider that we measure the current gain by connecting the same amplifier to a short-circuit load, figure 341(b). The input current i_i is:

$$i_i = \frac{v_i}{r_i} \quad (578)$$

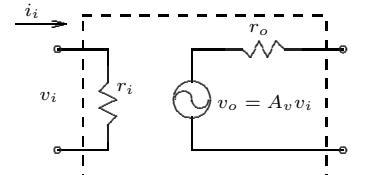
The output current i_o is:

$$i_o = \frac{A_v v_i}{r_o} \quad (579)$$

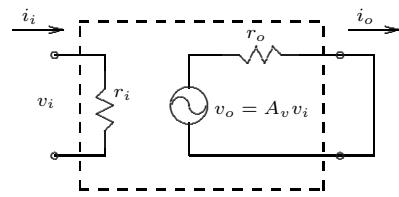
Substitute equations 578 and 579 in the definition for current gain A_i , and we have

$$\begin{aligned} A_i &= \frac{i_o}{i_i} \\ &= \left(\frac{A_v v_i}{r_o} \right) \div \left(\frac{v_i}{r_i} \right) \\ &= A_v \frac{r_i}{r_o} \end{aligned} \quad (580)$$

That is: **any three of the amplifier parameters determine the fourth.**



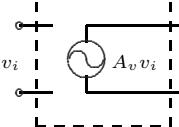
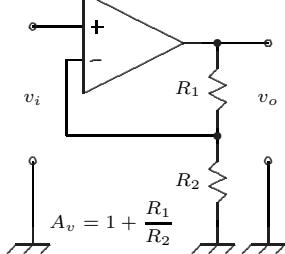
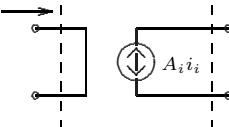
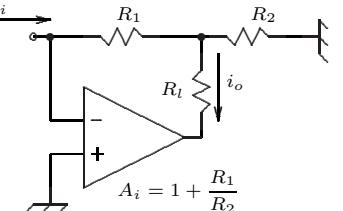
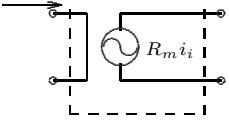
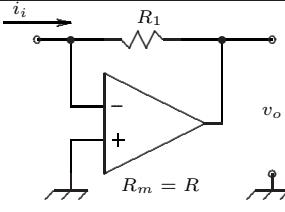
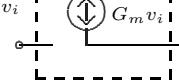
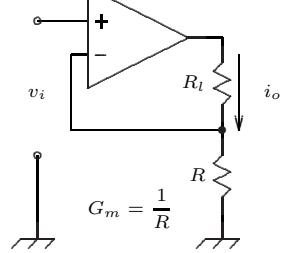
(a) Voltage Gain



(b) Current Gain

Figure 341: Relating Amplifier Parameters

Summary: Amplifier Classification

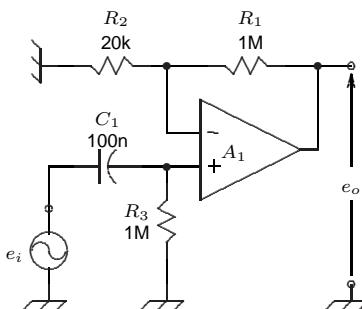
Type	Model	r_i	r_o	Gain	Example
Voltage		High	Low	$\frac{v_o}{v_i} = A_v \text{ Volts/Volt}$	
Current		Low	High	$\frac{i_o}{i_i} = A_i \text{ Amps/Amp}$	
Transresistance		Low	Low	$\frac{v_o}{i_i} = R_m \Omega$	
Transconductance		High	High	$\frac{i_o}{v_i} = G_m \text{ S}$	

Notes:

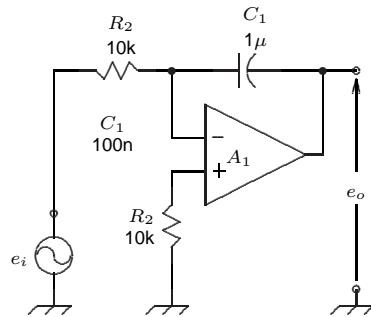
- 1 The current-amplifier example circuit is from reference [108].
- 2 Notice the symmetry between the voltage and current amplifiers, and between the transresistance and transconductance amplifiers.

13.25 Exercises

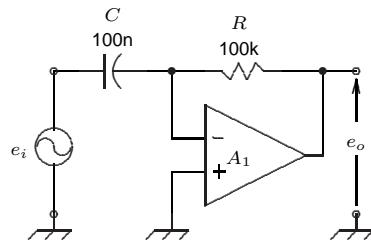
1. In the following parts to this question, you may use as many op-amps and other components as you require. However, economy improves the circuit.
- Draw an op-amp circuit that will generate the function $e_o = -0.5e_i$. Indicate the resistor ratios.
 - Draw an op-amp circuit that will generate the function $e_o = -3e_a - 2e_b$, where e_o is the output from the circuit and e_a and e_b are the inputs. Indicate the resistor ratios.
 - Draw an op-amp circuit that will generate the function $e_o = 5e_a - 5e_b$, where e_o is the output from the circuit and e_a , e_b are the inputs. Resistor values may be chosen arbitrarily.
 - Draw an op-amp circuit that will generate the function $e_o = -100 \int e_idt$, where e_o is the output from the circuit and e_i is the input. Resistor and capacitor values may be chosen arbitrarily.
 - Draw an op-amp circuit that will generate the function $e_o = e_a + e_b + e_c + e_d$, where e_o is the output from the circuit and e_a , e_b , e_c and e_d are the inputs. Resistor values may be chosen arbitrarily.
 - Draw an op-amp circuit that will generate the function $i_o = 1 \times 10^{-3}e_i$, where i_o is the output current from the circuit into some load resistance R_l and e_i is the circuit input voltage.
 - If the output voltage capability of the op-amp is ± 12 volts, what limits the maximum value of the load resistance R_l ? (Explain in words.)
 - What is the minimum allowable value of the load resistance R_l ?
 - Draw a circuit using one or more op-amps that will generate the signal $e_o = \int(e_a + e_b)dt$ where e_a and e_b are input voltages.
 - Draw an op-amp circuit that will operate from a single supply and generate the signal $e_o = 2(e_a - e_b)$. It is always true that $e_a \geq e_b$.
 - Draw an op-amp current generator that produces $i_o = e_i/100$ amps. You may assume that the input voltage e_i is always positive, the op-amp can produce unlimited current, and the load resistance R_l is floating. The op-amp is powered from ± 15 volts.
 - Draw the circuit.
 - For this same current generator, what is i_o when R_l is a short circuit?
 - For this same current generator, what determines the maximum allowable value of R_l ?
 - Draw an op-amp circuit that will generate the function $e_o = 0.25e_i$ and operate from a single positive supply voltage. Indicate the resistor ratios.
2. Assuming that the amplifier is ideal, sketch the frequency response of this circuit.



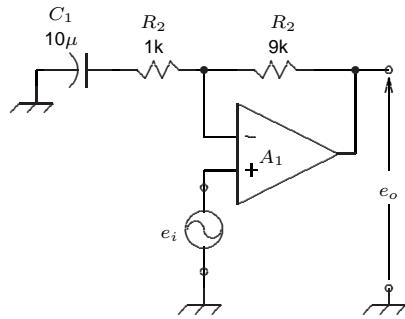
3. For the op-amp circuit shown below, sketch the amplitude and phase Bode plot. Mark the frequency of the gain-crossover (0db) point.



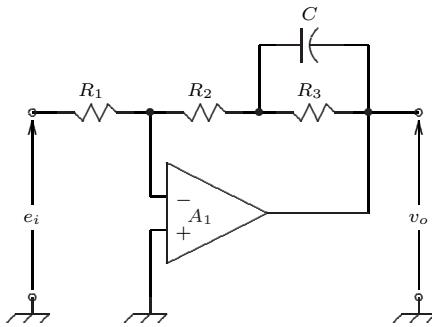
4. For the op-amp circuit shown below, sketch the amplitude and phase Bode plot. Mark the frequency of the gain-crossover (0db) point.



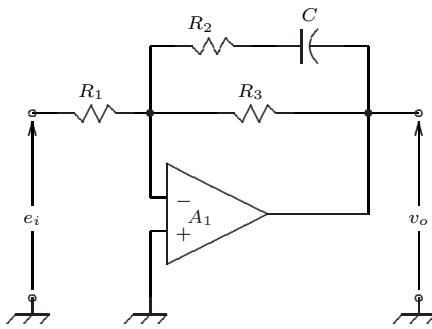
5. For the op-amp circuit shown below, sketch the amplitude and phase Bode plot. Mark the frequency of the gain-crossover (0db) point.



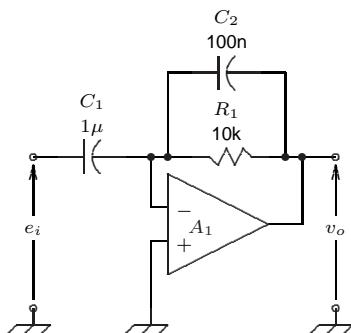
6. For the op-amp circuit shown below, all the resistors are equal. Sketch the amplitude and phase Bode plot.



7. For the op-amp circuit shown below, all the resistors are equal. Sketch the amplitude and phase Bode plot.

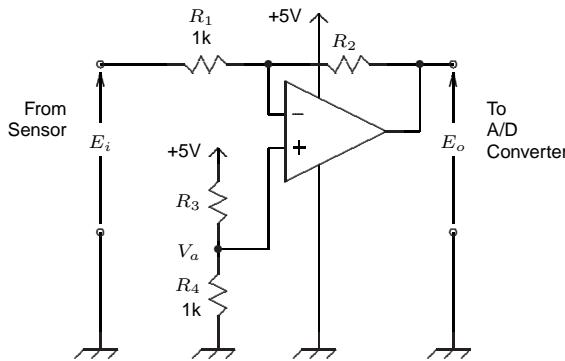


8. Sketch the amplitude and phase Bode plot for the op-amp circuit shown below.



9. A signal-conditioning amplifier is to be designed that will accept a sensor signal ranging between 100mV and 295mV. This signal is to be amplified and shifted to match the zero to 5 volt range of an 8 bit analog to digital converter. The schematic of the amplifier is shown in the figure.

- What is the voltage per step of the A/D converter?
- An input voltage of 100mV should correspond to an A/D reading N_{ad} of 230. An input voltage of 295mV should correspond to an A/D reading N_{ad} of 30. Calculate values for R_2 , V_a and R_3 .



- (c) Over what range of voltages must the output of the amplifier function properly?
- (d) What effect would a change in offset voltage of +2mV have on the A/D reading?
10. Draw each of the following op-amp circuits:
- Differential amplifier
 - Op-amp integrator
 - AC coupled, single supply, non-inverting amplifier
 - Non-inverting adder, 2 inputs
11. Given voltage sources E_1 , E_2 and E_3 , draw an op-amp circuit to generate a voltage equal to $E_1 + E_2 - E_3$. You may use as many op-amps as required.
12. Draw an op-amp circuit that accepts three voltages E_a , E_b and E_c and produces a voltage E_o that is the average of these three, ie:
- $$E_o = \frac{E_a + E_b + E_c}{3}$$
13. The circuit diagram below (figure 342) shows a temperature regulator for a small enclosed, insulated box (the oven). Transistor Q_1 is used as a heater for the box interior and diode-connected transistor Q_2 senses the temperature inside the box.
- The temperature sensor Q_2 has a forward voltage V_{be} of 0.595 volts at 25°C and 1 mA of forward current. The value of V_{be} for Q_2 changes at a rate of $-2.2 \text{ mV}/^\circ\text{C}$. The interior of the heated box is to be maintained at 50°C and ambient temperature is 20°C .
- Determine the polarity of the op-amp inputs to make this a negative feedback system, and show this on the circuit diagram.
 - Determine a suitable value for R_3 if the current through the sensing transistor Q_2 is 1mA.
 - Calculate the value of R_2 to keep the interior of the box at 50°C .
 - If the thermal resistance Θ from the enclosure interior to the ambient is $60^\circ\text{C}/\text{watt}$, what is the collector current of Q_1 ? (For thermal resistance, see section 28.1 on page 784).

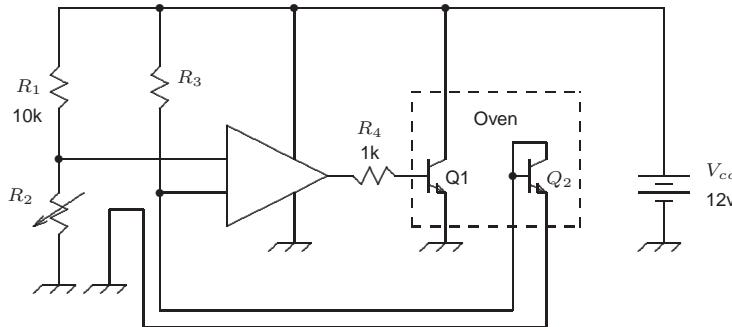


Figure 342: Temperature Regulator Circuit

- (e) How would you expect the collector current in Q_1 to change as the ambient temperature decreases? Explain.
- (f) If the offset voltage of the amplifier were to change by 4 millivolts, what effect would this have on the temperature of the oven?
14. An air pressure measuring circuit is shown in figure 343. The pressure sensor outputs a common mode voltage $V_x = V_y$ of half the supply voltage and a differential voltage V_{xy} which depends on the sensor air pressure at a rate of 0.6 millivolts per kilopascal of pressure.

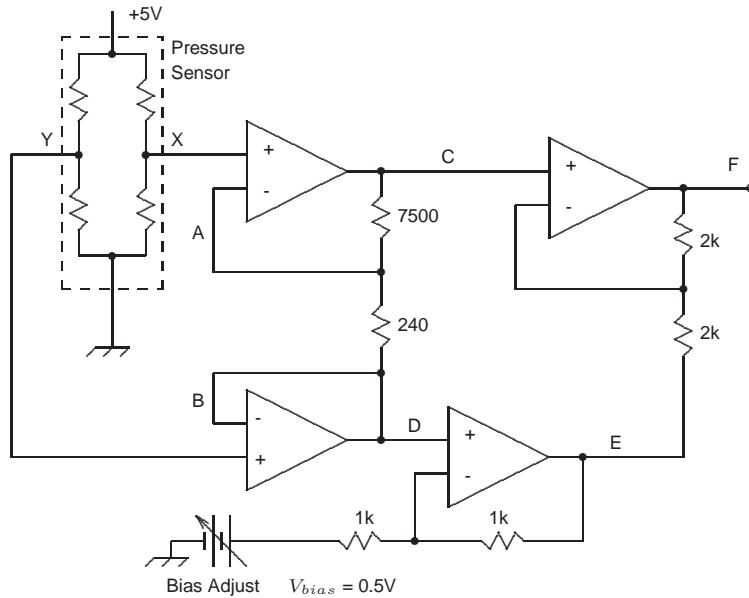


Figure 343: Pressure Measurement Circuit

- (a) Assuming initially that the pressure on the sensor is zero, determine the ground-referenced voltages at each of the points A through F, to the nearest millivolt.
- (b) Now assume that the air pressure is 100 kilopascals, and redetermine these voltages, again to the nearest millivolt.

- (c) Calculate the transfer function of the circuit $\Delta V_{out}/\Delta P_{in}$ where ΔV_{out} is the change in output voltage (volts) and ΔP_{in} is the change in input pressure (kilopascals).
- (d) Calculate the effective gain transfer function between the output voltage and the offset adjust voltage, $\Delta V_{out}/\Delta V_{bias}$.
15. The concept of the *4-20mA Current Loop* was introduced in question 19 on page 116. Design a circuit that will convert a sensor signal in the range 0 to +1VDC to a current between 4 and 20mA. The power supply voltage is derived from the current loop and varies from +8 to +24VDC.
- Hint: This circuit can be separated into two stages. The final stage is a current generator that produces one milliamp of current for (say) each volt of input. The next-to-final stage, driving the current generator, is a circuit that offsets and amplifies the 0 to +1VDC sensor signal to get it into the correct range to drive the current generator.
16. At the receiving end of a 4-20 mA current loop, the current is converted into a voltage across a floating 250Ω resistor. Design a circuit to accept this 4 to 20mA differential signal and convert it into a single-ended signal 0 to +1VDC. You may use as many op-amp stages as required.
- Based on the circuit shown in question 19 on page 116, what is the maximum common mode voltage into this circuit?

17. For the differential amplifier shown below:

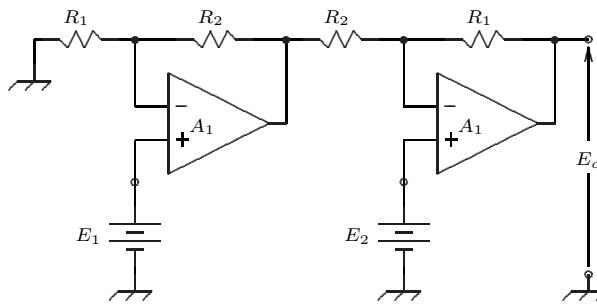
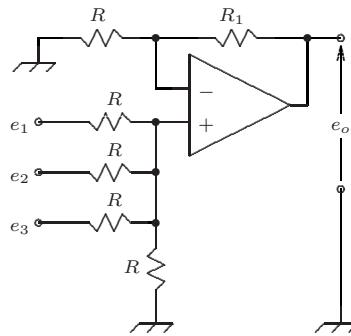
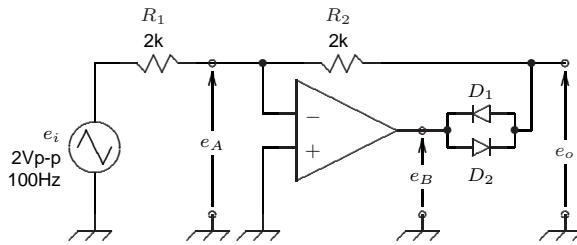


Figure 344: Differential Amplifier

- (a) Determine the output voltage E_o in terms of the input voltages E_1 and E_2 and the values of the resistors R_1 , R_2 .
- (b) How does the allowable maximum common mode input voltage compare in this circuit with the classical 4-resistor differential amplifier: is it greater or less? Draw the classical circuit and justify your answer.
18. For the circuit shown below calculate R_1 so that $e_o = e_1 + e_2 + e_3$.

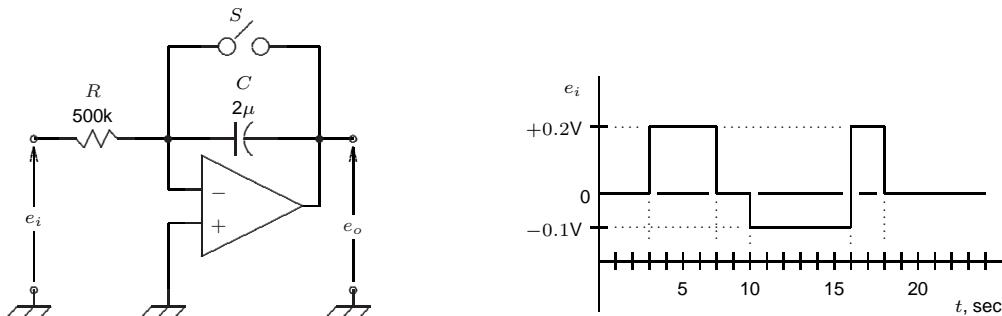


19. For the circuit shown below, the op-amp may be considered ideal. The diodes are silicon, with a forward voltage drop of 0.6 volts when conducting. The input waveform e_i is a 2 volt peak-peak triangle waveform at a frequency of 100Hz. Draw the corresponding waveforms for e_A , e_B and e_o .



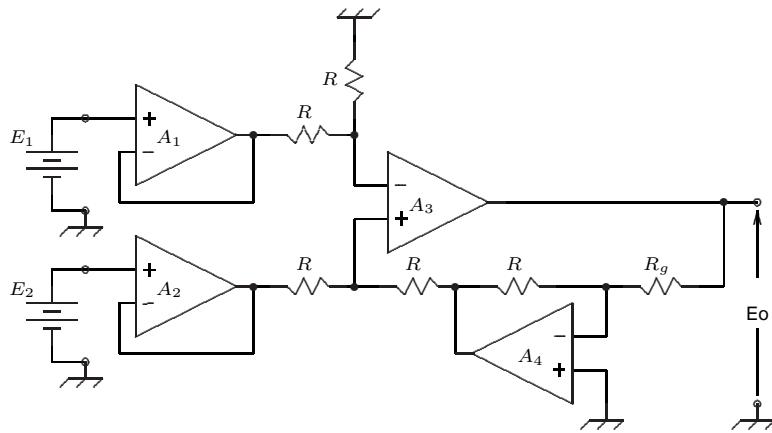
20. In the circuit shown below, the switch is opened at time $t = 0$. The input waveform is as shown.

Draw the plot of output voltage e_o as a function of time over the same interval as the input voltage.

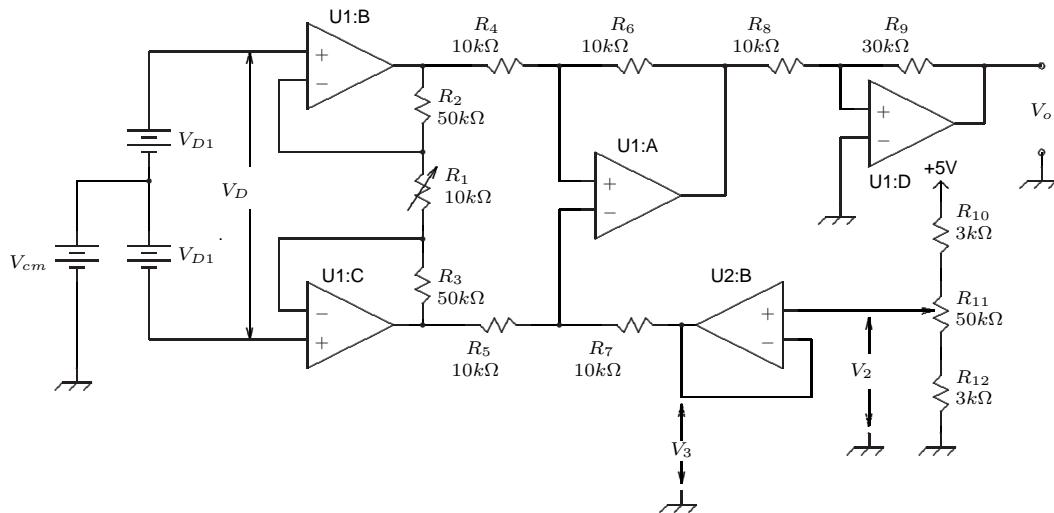


21. For this circuit,

- Determine the output voltage E_o in terms of E_1 , E_2 and the resistor values.
- Suggest an application for the circuit.

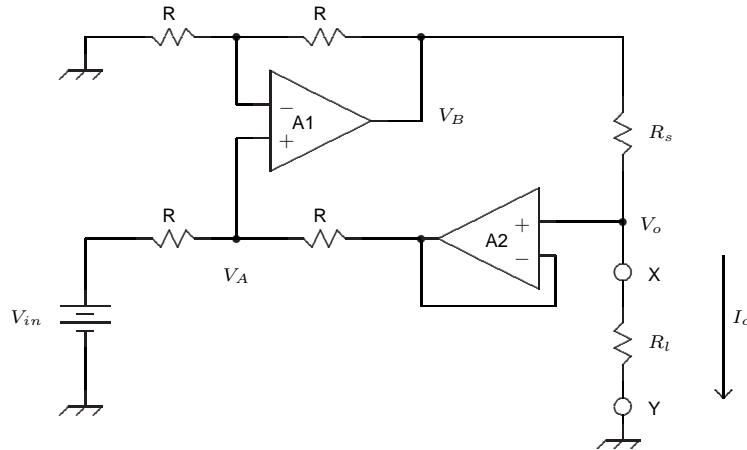


22. The circuit shown below is an amplifier for the signal from an air pressure sensor. The sensor outputs a common mode voltage and a differential voltage. The common-mode voltage is essentially constant. The differential voltage changes with air pressure. The output of this system drives a 5 volt span, 8 bit A/D converter. You may assume that all op-amps are ideal and clip at an output voltage of ± 12 volts.

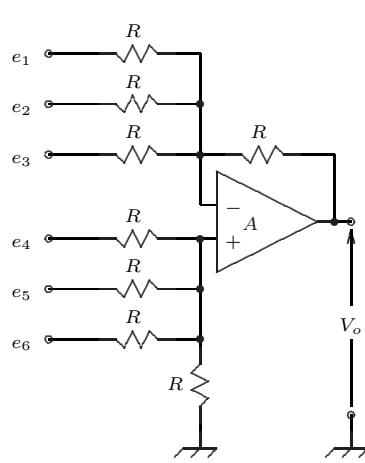


- List the resistors that should be very carefully matched to each other.
- Briefly explain why this is important.
- To start with, consider that the potentiometer R11 is at its midpoint. What is the magnitude of voltage V_2 (at the wiper of potentiometer R11)?
- What is the magnitude of voltage V_3 under the same condition?
- The differential input voltage V_D is equal to $V_{D1} + V_{D2}$. For each 1 millivolt change in V_D , what is the change in V_o ?
- For each 1 volt change in the common-mode voltage V_{cm} , what is the change in V_o ?
- Now consider the effect of R_{11} . This is used to adjust the output V_o of the system to exactly 2.5 volts when $V_D = V_{D1} + V_{D2}$ is zero. What should be the voltage V_2 when this is the case?

23. The circuit shown below is a *current source*, which outputs current I_o into the load resistance R_l . For this question, the operational amplifiers may be considered to be ideal.



- (a) What is V_A in terms of V_{in} and V_o ?
 - (b) What is V_B in terms of V_A ?
 - (c) What is I_o in terms of V_{in} and R_s ?
 - (d) Assuming the op-amps are not saturated or cut-off, what is the effect on I_o if R_l is doubled?
 - (e) What limits the maximum value of R_l ?
 - (f) What is the minimum value of R_l ?
24. For the circuit shown below:



- (a) Develop an expression for the output voltage in terms of the input voltages.
- (b) Generalize this result for n inputs into the upper half and n inputs into the lower half of the circuit.

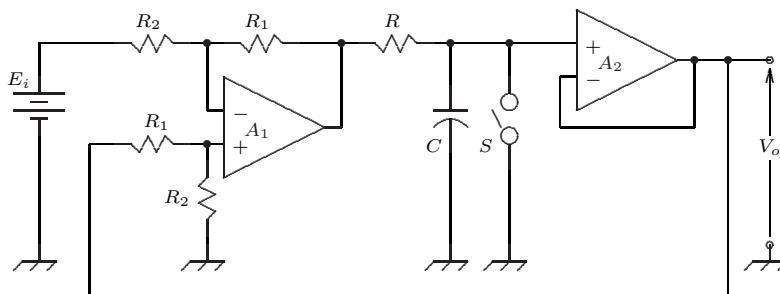
25. A lead acid battery has a terminal voltage of about 14.5 volts when fully charged. It is considered to be discharged when it reaches 11 volts.

Design a *battery condition meter* to indicate full scale when the input is 15 volts and zero scale when the input is 11 volts. This meter circuit must operate from the same battery supply, that is, the circuit must not require any negative voltages or voltages in excess of 11 volts. The meter movement that is used as an indicator is 1mA full scale with an internal resistance of 100Ω .

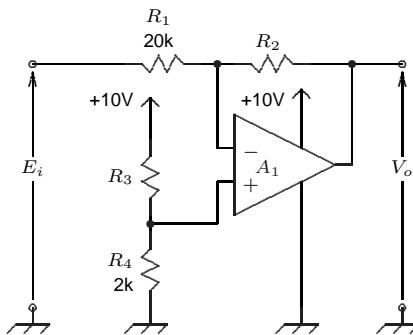
Design the circuit, using standard 1% or 5% resistors and a single-supply operational amplifier. (At this point, you can assume the op-amp is ideal).

Hint: One approach is to break the circuit into two stages. The first stage implements a transfer function such that input voltage range 11 to 15 volts generates a voltage between (say) 0 and 3 volts. A second stage is a voltage-current converter that accepts 0 to 3 volts and generates 0 to 1 mA. Any stable reference voltage should be generated with zener diode. The circuit should be designed so that it does not require any negative voltages or voltages in excess of 11 volts.

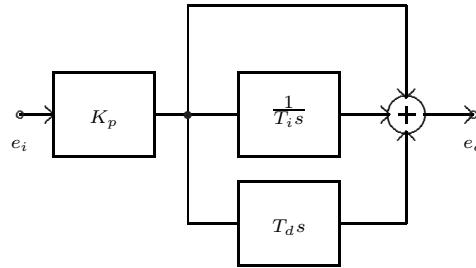
26. The circuit shown below functions as an integrator. Unlike the single op-amp integrator circuit shown in section 13.10, one terminal of the integrating capacitor is connected to ground. This simplifies the design of the reset switch S .



- (a) The switch S is originally closed. What is the behaviour of the output voltage V_o after the switch is opened?
- (b) Verify that the circuit generates an output voltage proportional to the integral of the input voltage.
- (c) Relate the gain of the integrator to the values of the circuit components shown in the diagram.
27. For the circuit shown, calculate R_2 and R_3 so that an input voltage E_i over the range $\pm 0.5V$ corresponds to a range of 0V to +5V in the output voltage V_o .



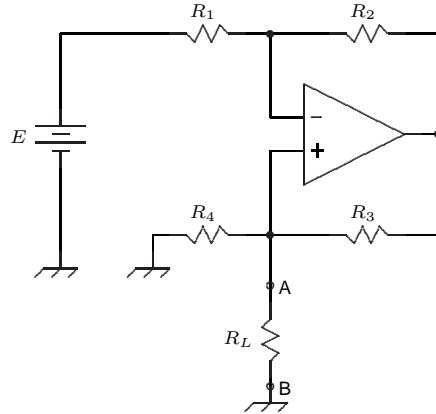
28. The block diagram for a PID controller is shown below.



Draw the schematic diagram of such a controller, using operational amplifier building blocks. The controller should

- have adjustable gain for the proportional, integral and differential factors.
- have a facility to individually disable the integral and differential factors so that the proportional gain can be adjusted in isolation.
- have additional inputs that provide for addition or subtraction of a feedback signal at the input of the controller.

29. For the Howland Current Source (section 18.3) shown below:



- (a) Prove that the output resistance is given by:

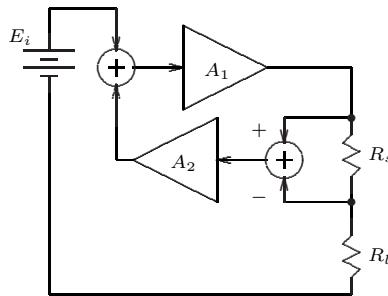
$$r_o = \frac{R_4}{\left(\frac{R_4}{R_1} \frac{R_2}{R_3} - 1\right)}$$

Hint: Replace the input voltage generator with a short circuit. Replace the load resistance R_L by a voltage generator e_o between terminals A and B. Calculate the current i_o into terminal A.

- (b) Determine the conditions required for the circuit to become a perfect current generator, ie, for the output resistance to become infinite.

30. For each of the amplifier circuits shown in section 13.24, determine the input and output resistance.

31. The figure shows the concept for a negative-feedback voltage-current generator, where one terminal of the load resistance R_l can be grounded. Resistance R_s is the current sensing resistance.



- (a) Label the inverting and non-inverting inputs of the adder associated with amplifier A_1 .
 - (b) Determine the output resistance r_o of the circuit.
 - (c) Using op-amp building blocks, redraw the circuit as a complete schematic diagram.
32. The adder circuit shown in section 13.2, figure 269 (page 336), has all resistors equal.
- (a) Determine an expression for the output voltage when all the input resistors have the same value, but the feedback resistance has a different value.
 - (b) Suggest a situation where this would be useful.
 - (c) Determine an expression for the output voltage when all the resistances are different.

33. Referring to the Howland Integrator circuit in figure 288 on page 350:

- (a) Draw the Bode plot for gain and phase for the circuit.
- (b) Show that the voltage at the output terminal of the op-amp is given by

$$\frac{e'_o}{e_i} = \frac{2}{sRC}$$

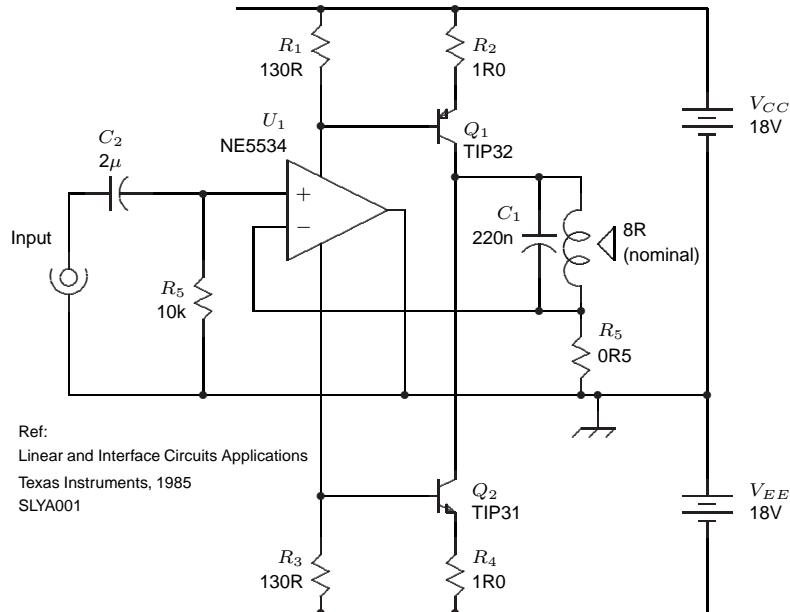
34. The schematic shows the circuit of an unusual audio power amplifier [109].

The output of the amplifier drives current directly into a ground connection. This current must be supplied via the positive or negative supply, so it appears in a supply line and creates a current in the corresponding power transistor. The quiescent power supply current of the op-amp establishes a non-zero current in these transistors to prevent crossover distortion. This system is claimed to have somewhat superior frequency response. In effect, the op-amp and power transistors behave as one operational amplifier¹⁴¹.

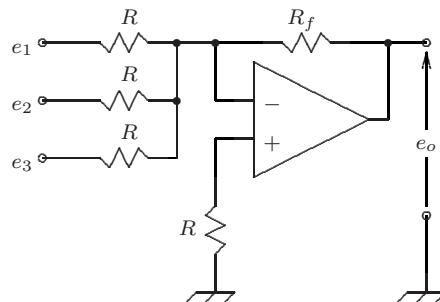
- (a) Confirm that the system uses negative feedback.
- (b) Does the amplifier and its feedback system create an output that is high impedance (voltage source) or low impedance (current source)? Explain.
- (c) The impedance of a moving coil loudspeaker can vary quite substantially over its frequency range. What effect does this have on the current through it? Explain.

¹⁴¹A tip of the analog design hat to Michel Gallant, who brought this circuit to my attention.

- (d) In the original note, the author claims that this circuit improves the damping factor of the amplifier. (See http://en.wikipedia.org/wiki/Damping_factor for more on the damping factor of audio amplifiers.) Do you agree? Explain.
- (e) Estimate the low-frequency cutoff for the amplifier, in Hz.



35. For the circuit shown below show that $e_o = \frac{R_f}{R_i} (e_1 + e_2 + e_3)$.



14 The Schmitt Trigger

14.1 Introduction

In section 12.2, we showed how an operational amplifier could be used as a *comparator* for two input voltages. In that case, there is no feedback around the op-amp – it operates open-loop. As a result, the output has two states: at the positive bound or at the negative bound, depending on the relative magnitudes of the two inputs. The comparator circuit is useful and used frequently as a circuit element.

However, when the input is noisy because it is contaminated with some undesired signal, the unadorned comparator will not work correctly.

Consider the situation shown in figure 345. In this case, the signals V_a and V_b are being compared. The output of the comparator should be high when $V_a > V_b$ and low when $V_a < V_b$.

However, in this case the non-inverting input signal is contaminated by noise (greatly exaggerated in the diagram). As the inverting input signal passes through the threshold region, the noise signal causes repeated switching of the output until the difference between the two inputs is larger than the amplitude of the noise. This series of transitions in the output is known as *chatter*¹⁴².

In some circuits, chatter is not a major problem. However, if the signal is a logic signal that drives a digital counter, the counter will quite happily count every transition in the chatter sequence instead of the one transition that is the desired result. Obviously, this has the potential of creating major errors¹⁴³.

Hysteresis

The Schmitt trigger solves this problem. It provides positive feedback so that once it begins to switch, the comparison level is driven in such a direction as to increase the input signal. This is a form of positive feedback that rapidly drives the output to one extreme or the other.

The degree to which the comparison level is adjusted in a switching transition is known as the *hysteresis*.

The toggle switch is a mechanical analogy of Schmitt trigger action. The handle of a toggle switch is moved to cause the switch to operate. Once the handle reaches

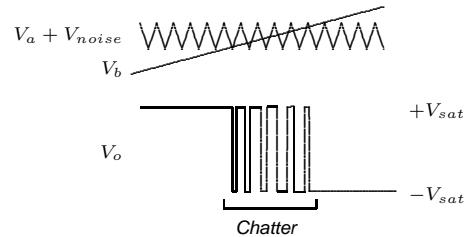
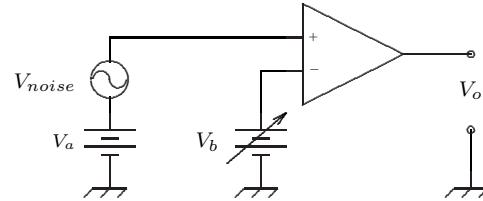


Figure 345: Noisy Input to a Comparator

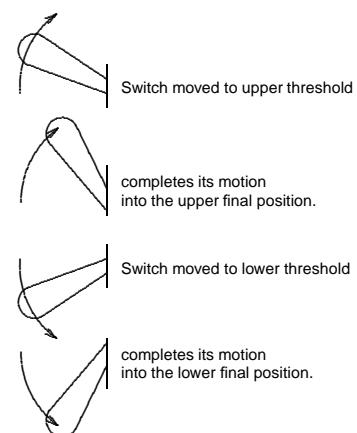


Figure 346: Mechanical Hysteresis

¹⁴²The name derives from similar behaviour in relay logic, where it can be heard as a buzzing noise.

¹⁴³For those familiar with microprocessor technology, picture that this signal is connected to an edge-triggered interrupt on a microprocessor. Unless the microprocessor is somehow programmed to ignore the chatter, it will cause errors.

its switching threshold, it rapidly carries itself through to the limit position. When the handle is moved in the opposite direction, it must be moved past the previous threshold to a more distant threshold before the switch transitions into its other state.

In this case, the magnitude of the hysteresis would be the distance between the two threshold positions.

Transfer Function

The effect of hysteresis may be visualized with a plot of the transfer function. The transfer function of a simple comparator is shown in figure 347(a).

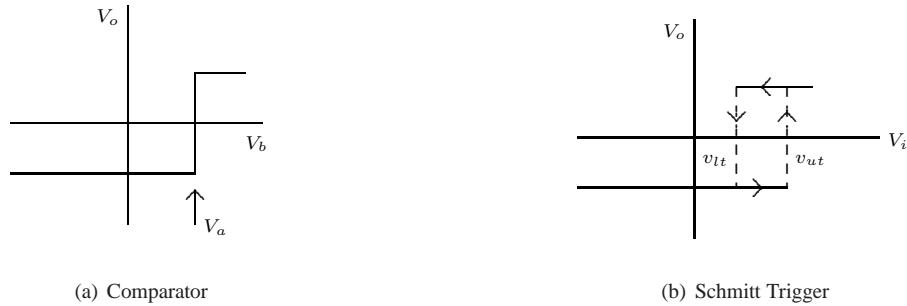


Figure 347: Comparator and Schmitt Transfer Function

The transfer function of a Schmitt trigger is shown in figure 347(b). The transfer function follows the arrows: as the input voltage increases, it must reach the upper threshold v_{ut} before the output switches from its low to its high state. Once the output has switched into the high state, the input voltage must move down to the lower threshold voltage v_{lt} before the output returns to its low state. The hysteresis voltage is the difference between these two thresholds:

$$V_h = v_{ut} - v_{lt} \quad (581)$$

Schmitt Trigger Circuit Design, Overview

There are two possible Schmitt configurations: inverting and non-inverting. The best circuit to use depends on the application and available supply voltages. It's possible to slog through the design of a Schmitt trigger using nothing more than an understanding of open-loop op-amp operation and Ohm's law. However, there is a graphical approach to the design that gives much more insight into these two circuits and allows one to make an educated choice between them. The graphical approach also gives a rough indication of the resistor and bias voltage values, so you can choose the circuit that best suits your application.

Following that, in section 14.4, we show the design equations, which allow you to calculate the resistor and bias voltages with greater precision.

14.2 Schmitt Trigger, Inverting

The lever diagram (section 12.5 on page 322) may be applied to positive feedback circuits. As in the case of linear amplifiers, there are two cases to consider: the inverting and non-inverting Schmitt Trigger.

The inverting Schmitt Trigger is shown in figure 348. It is identical to the non-inverting amplifier discussed previously except that the input terminals of the op-amp are exchanged. The corresponding lever diagram is shown in figure 349.

The amplifier receives positive feedback, so it will drive the output to one of its output limits. We refer to this configuration as the *inverting* Schmitt trigger because the output switches to its negative limit as the input voltage exceeds the positive input threshold, and switches to its positive limit as the input voltage exceeds (moving in a negative direction) the negative input threshold.

Let's assume that the output limits (bounds) are $V_{OU} = +3$ volts and $V_{OL} = -3$ volts. Assume that the input voltage is zero, and that the amplifier is currently at the upper bound. Resistors R_1 and R_2 are equal. Then the voltage at their junction, the non-inverting input terminal of the op-amp, will be at +1.5 volts. This is V_{IU} , the *upper threshold* of the Schmitt trigger, represented by the upper solid sloping line on figure 349.

The equation governing the generic behaviour of the op-amp is

$$V_o = A_{ol}(V^+ - V^-)$$

where A_{ol} is the open-loop gain of the op-amp, V^+ is the non-inverting input voltage (+1.5 volts), and V^- is the inverting input voltage (0 volts). In this circuit, with the output at its +3 volt positive limit, there is a net input voltage of 1.5 volts. This is multiplied by the large open-loop gain of the op-amp, driving the output up against its positive limit. This situation corresponds on the solid sloping line in the lever diagram of figure 349.

Now consider that the input voltage V_i at the inverting terminal increases positively to slightly more than the upper threshold V_{ut} . The net input to the op-amp will be negative ($V^+ - V^-$), and this will drive the output of the amplifier negative. As the output moves negative, it increases the effective input voltage between the op-amp input terminals, accelerating the movement of the output to its lower bound V_{lb} . On the lever diagram of figure 349, this new stable state corresponds to the dashed sloping line. In this state, the input must go beyond the lower threshold $V_{lt} = -1.5$ volts to cause the Schmitt Trigger to switch back to its positive state.

As in the case of the linear amplifiers, a bias voltage source V_b may be used to modify the behaviour of the inverting Schmitt trigger, as shown in the following design example.

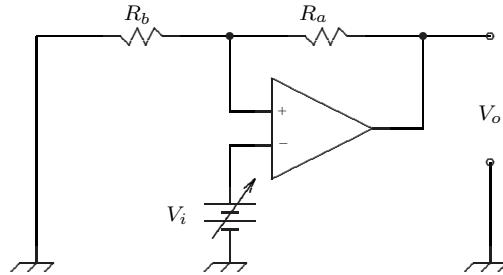


Figure 348: Inverting Schmitt Trigger

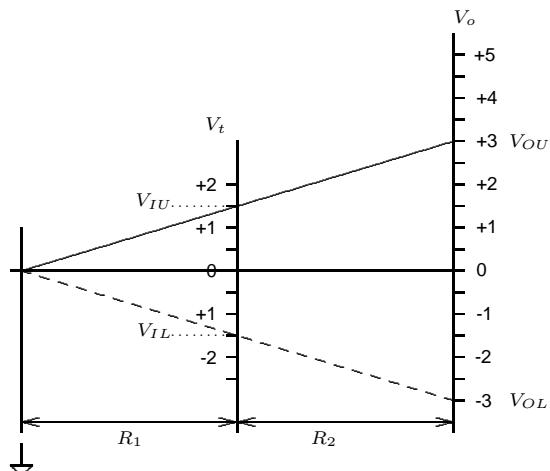


Figure 349: Lever Diagram, Inverting Schmitt Trigger

Design Example

Suppose we wish to design an inverting Schmitt Trigger circuit where the output limits at +5 volts and 0 volts. The input thresholds are ± 1 volt. This might correspond to an application where an AC coupled signal (one with zero DC component) is to be *squared up* for measurement by a microprocessor. The requirements are summarized in figure 350. Then the lever diagram must look like figure 351(a).

Parameter	Symbol	Value, volts
Input Upper Threshold	V_{IU}	+1
Input Lower Threshold	V_{IL}	-1
Output Upper Bound	V_{OU}	+5
Output Lower Bound	V_{OL}	0

Figure 350: Schmitt Trigger Design Requirements

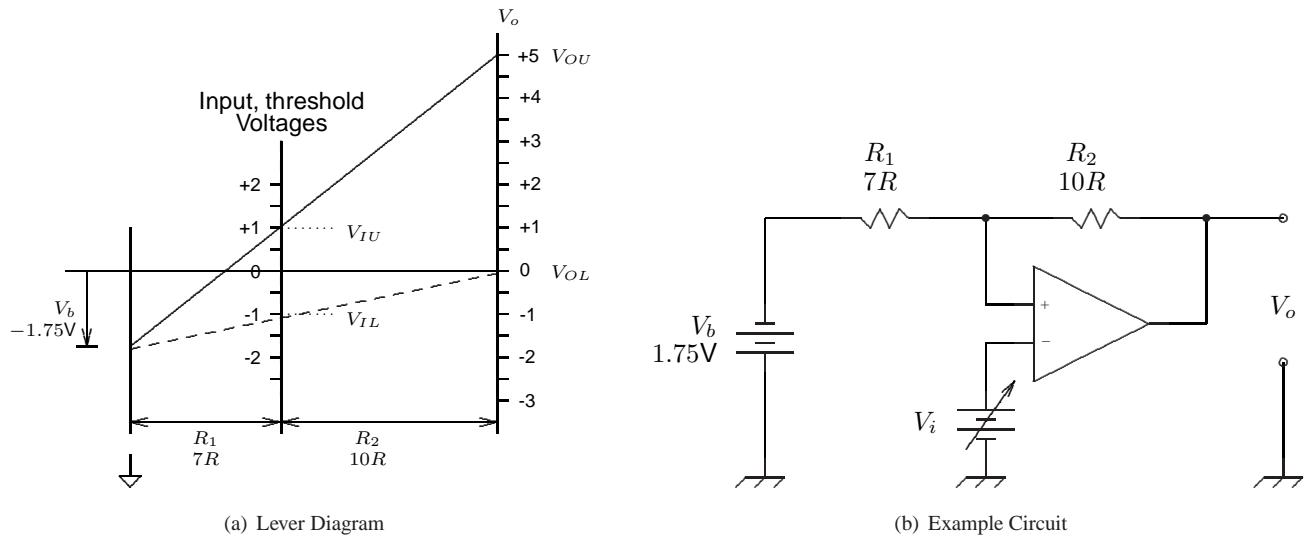


Figure 351: Inverting Schmitt Trigger

This example introduces an additional complication. In the lever diagram of figure 349, the input and output are symmetrical around zero volts. In this example, the inputs are not symmetrical and, as shown in figure 351(a), the two traces (levers) intersect off the horizontal axis. This represents a *bias voltage* V_b , as shown in the schematic of figure 351(B).

According to the lever diagram of figure 351(a), the design requires that the bottom end of the voltage divider be biased to $V_b = -1.75$ volts. The ratio of $R_2 : R_1$ is $10 : 7$. (This is a graphical solution, and so the values for V_b and the resistor ratio are approximate. We could find the exact values using the design equations for this circuit, shown later in section 14.7). The schematic is shown in figure 351(b). The exact values for R_1 and R_2 are not terribly important, as long as their ratio is correct. Values in the region of $10\text{K}\Omega$ to $1\text{M}\Omega$ are typical.

The lever diagram shows that a negative voltage is required for this circuit. If the circuit must operate from a single positive supply, the -1.75 volt bias signal may not be convenient to provide. Section 14.3 shows an alternative that uses a non-inverting Schmitt trigger circuit.

14.3 Schmitt Trigger, Non-Inverting

The non-inverting Schmitt Trigger is shown in figure 352(a). The input signal is moved to one end of the $R_1 : R_2$ voltage divider.

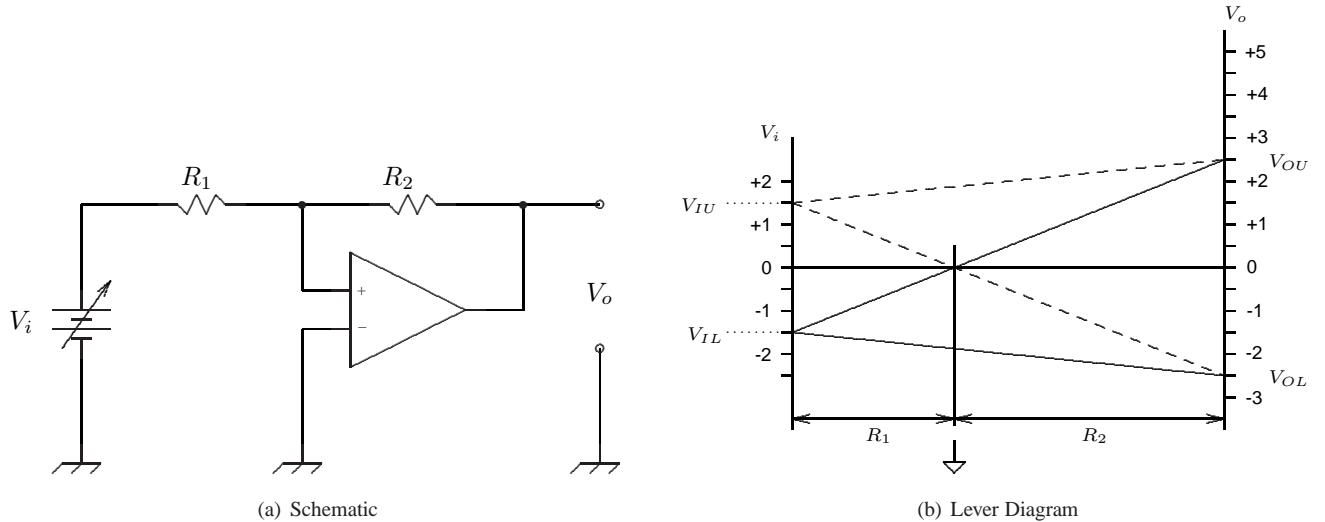


Figure 352: Non-Inverting Schmitt Trigger

The corresponding lever diagram is shown in figure 352(b). In this case, we have assumed that the amplifier output limits are the V_{ub} (upper bound) at +2.5 volts and V_{lb} at -2.5 volts.

In this circuit configuration, the amplifier changes state whenever the voltage at the centre tap of the $R_1 : R_2$ voltage divider moves through zero volts. From the straight line constructions on figure 352(b), we can determine that the upper threshold V_{TU} will be at +1.5 volts and the lower threshold V_{TL} at -1.5 volts.

Consider that the input voltage is positive and the Schmitt is in its *output high* state, with the output voltage V_o equal to the *upper bound* value V_{OU} , +2.5 volts. This corresponds to the upper dashed line in figure 352(b).

Now suppose that the input voltage moves negative through zero, ending up in the position indicated by the upper solid line in figure 352(b). When the input voltage (the left end of the upper solid line) passes through the lower threshold level V_{TL} of -1.5 volts, the voltage at the op-amp non-inverting terminal passes through zero. The op-amp immediately switches into the *output low* state, with V_{OL} equal to -2.5 volts. This new state corresponds to the lower solid line in figure 352(b), with the output voltage V_o equal to the *lower bound* value V_{OL} , -2.5 volts.

To revert back to the original state, output positive, the input voltage has to move positive to take the op-amp non-inverting terminal through zero volts, as indicated by the lower dashed line in figure 352(b). This requires that the input voltage move up through the upper threshold level V_{TU} of +1.5 volts.

The lever diagram is useful in visualizing the action of this circuit, but its main utility is in designing a circuit to some specific requirements, as described in the following design example..

Design Example

Now we will revisit the previous design, this time using the non-inverting Schmitt Trigger circuit. As previously summarized in figure 350, the output limits are at +5 volts and 0 volts, and the input thresholds are ± 1 volt. The lever diagram for a non-inverting Schmitt trigger with these thresholds and output limits is shown in figure 353.

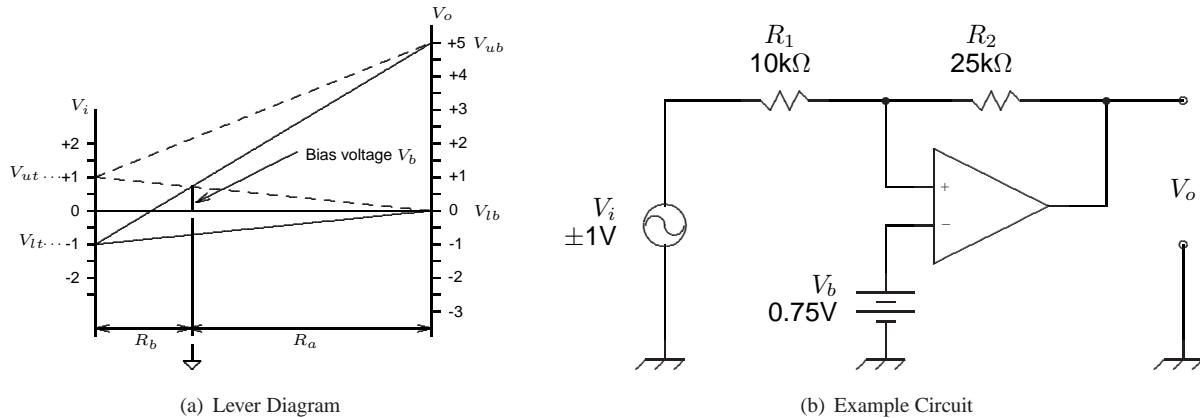


Figure 353: Non-Inverting Schmitt Trigger

The diagram indicates that a bias voltage of +0.75 volts is required at the inverting terminal of the op-amp. The ratio of $R_2:R_1$ is 2.53:1. The final circuit design is shown in figure 353(b).

Comparing the circuit designs in figures 351 and 353, the non-inverting version in figure 353 is probably preferable¹⁴⁴. In that design, the +0.75 volt bias voltage may be derived from the system positive supply voltage with another voltage divider.

14.4 Schmitt Trigger Design Equations

First, a preliminary step that will save us work. The circuit shown in figure 354 arises repeatedly in an analysis of either of the schmitt-trigger circuits, so it's useful to be able to analyse it easily. We can use the *superposition theorem* (section 3.6) along with the operation of a voltage divider (section 3.3). Our objective is to calculate the voltage V_A .

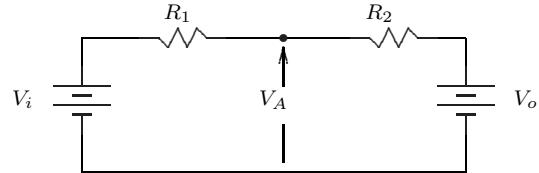


Figure 354: Schmitt Voltage Divider

- Using superposition, short-circuit voltage source E_o and determine the voltage V'_A due to voltage source E_i . Then resistors R_1 and R_2 form a voltage divider, so that:

$$V'_A = E_i \left(\frac{R_2}{R_1 + R_2} \right) = E_i \left(\frac{R_2}{R_T} \right) \quad (582)$$

where we have put the total resistance of the divider $R_1 + R_2$ equal to R_T for convenience.

2. Using superposition again, short-circuit voltage the other source E_i and determine the voltage V_A'' due to voltage source E_o . Again, resistors R_2 and R_1 form a voltage divider, so that:

$$V_A'' = E_o \left(\frac{R_1}{R_1 + R_2} \right) = E_o \left(\frac{R_1}{R_T} \right) \quad (583)$$

¹⁴⁴This assumes that the phase of the output with respect to the input is not a major issue and can be accounted for easily further in the signal chain. This is certainly the case if the signal is used by a microprocessor.

3. Determine the final voltage V_A by adding the effects of the independent sources.

$$V_A = V'_A + V''_A = E_i \left(\frac{R_2}{R_T} \right) + E_o \left(\frac{R_1}{R_T} \right) \quad (584)$$

14.5 Inverting Schmitt Trigger, Design Equations

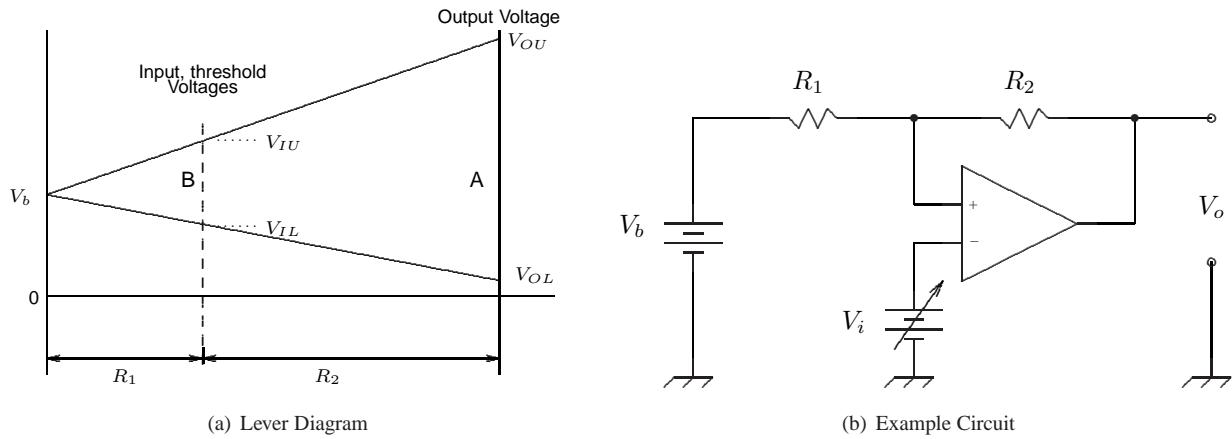


Figure 355: Inverting Schmitt Trigger

Here we develop design equations for the inverting Schmitt trigger. We assume that the upper and lower levels of the input and outputs are known, and we wish to solve for the ratio of $R_2 : R_1$ and the value of the bias voltage V_b ,

Referring to figure 355(a), by similar triangles the ratio of line segments A and B is given by:

$$\begin{aligned} \frac{A}{B} &= \frac{R_1 + R_2}{R_1} \\ &= 1 + \frac{R_2}{R_1} \end{aligned} \quad (585)$$

But:

$$A = V_{OU} - V_{OL} \quad (586)$$

$$B = V_{IU} - V_{IL} \quad (587)$$

Substitute from for A equation 586 and B from equation 587 into equation 585:

$$\frac{V_{OU} - V_{OL}}{V_{IU} - V_{IL}} = 1 + \frac{R_2}{R_1} \quad (588)$$

This yeilds the ratio of the two resistors. We would choose a convenient value for one and then calculate the other. Convenient implies values that are large, to reduce the amount of current used by the circuit, but not so large that the bias currents of the comparator or op-amp cause an significant voltage across them.

Now to calculate the bias voltage. We could choose either state of the Schmitt: here we'll choose the lower one, where the output voltage is at V_{OL} . Using the superposition theorem and voltage divider equation as shown in equation 584, we can write:

$$V_{IL} = V_b \frac{R_2}{R_1 + R_2} + V_{OL} \frac{R_1}{R_1 + R_2} \quad (589)$$

Solve for V_b :

$$V_b = V_{IL} \left(\frac{R_1}{R_2} + 1 \right) - V_{OL} \frac{R_1}{R_2} \quad (590)$$

This is the second of our two design equations.

Example

Design an inverting Schmitt with the following specifications:

Parameter	Symbol	Value, volts
Input Upper Threshold	V_{IU}	+1.45
Input Lower Threshold	V_{IL}	+1.05
Output Upper Bound	V_{OU}	+2.5
Output Lower Bound	V_{OL}	0

Figure 356: Schmitt Trigger Design Requirements

Solution

An input voltage greater than 1.45 volts causes the output to snap into its low state, 0 volts. An input voltage below 1.05 volts causes the output to snap into its high state, +2.5 volts. First we use equation 588 to calculate the ratio of resistors:

$$\begin{aligned} 1 + \frac{R_2}{R_1} &= \frac{V_{OU} - V_{OL}}{V_{IU} - V_{IL}} \\ &= \frac{2.5 - 0}{1.45 - 1.05} \\ &= 6.25 \end{aligned}$$

from which $R_2 = 5.25R_1$. We might choose $R_1 = 100k$, $R_2 = 525k$. Now to calculate the bias voltage V_b . From equation 590 and our previously calculated value for R_2/R_1 , we have:

$$\begin{aligned} V_b &= V_{IL} \left(\frac{R_1}{R_2} + 1 \right) - V_{OL} \frac{R_1}{R_2} \\ &= 1.05 (5.25 + 1) - 0 \times 5.25 \\ &= 1.25V \end{aligned}$$

This completes the basic design of the inverting Schmitt trigger.

14.6 Non-Inverting Schmitt Trigger, Design Equations

Here we follow a path similar to the previous section, to develop the design equations for the non-inverting Schmitt trigger. As before, we assume that the upper and lower levels of the input and outputs are known, and we wish to solve for the ratio of $R_2 : R_1$ and the value of the bias voltage V_b ,

Referring to figure 357(a), by similar triangles the ratio of line segments A and B is given by:

$$\frac{A}{B} = \frac{R_2}{R_1} \quad (591)$$

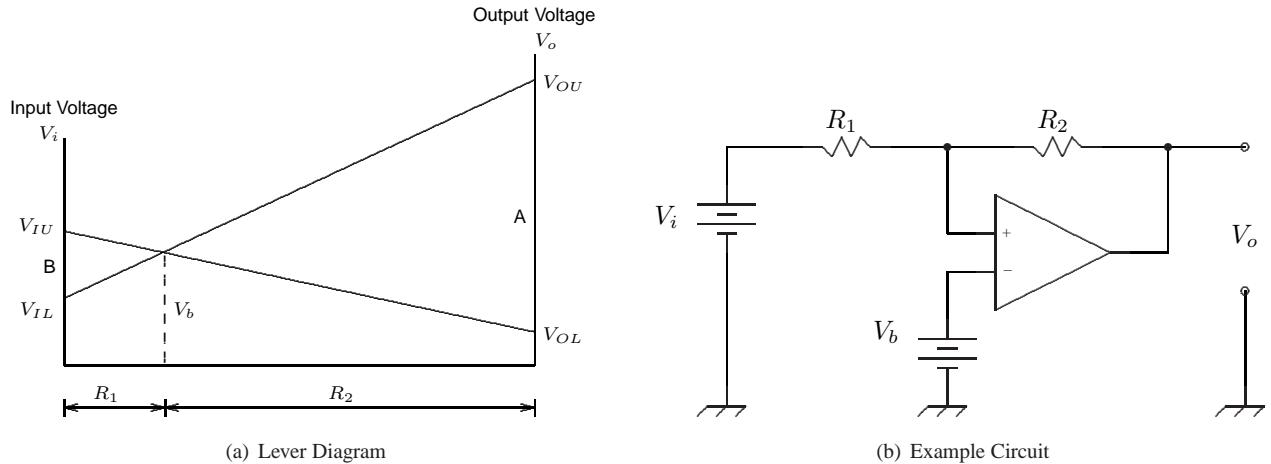


Figure 357: Non-Inverting Schmitt Trigger

But:

$$A = V_{OU} - V_{OL} \quad (592)$$

$$B = V_{IU} - V_{IL} \quad (593)$$

Substitute from for A equation 592 and B from equation 593 into equation 591:

$$\frac{V_{OU} - V_{OL}}{V_{IU} - V_{IL}} = \frac{R_2}{R_1} \quad (594)$$

This yeilds the ratio of the two resistors.

Now to calculate the bias voltage. We'll choose the lower state of the Schmitt, where the output voltage is at V_{OL} . We'll also use the Superposition-cum-Voltage-Divider approach of equation 584 previously.

$$\begin{aligned}
 V_b &= V_{IH} \frac{R_2}{R_1 + R_2} + V_{OL} \frac{R_1}{R_1 + R_2} \\
 &= \frac{V_{IH}}{\frac{R_1 + R_2}{R_2}} + \frac{V_{OL}}{\frac{R_1 + R_2}{R_1}} \\
 &= \frac{V_{IH}}{1 + \frac{R_1}{R_2}} + \frac{V_{OL}}{1 + \frac{R_2}{R_1}}
 \end{aligned} \tag{595}$$

The ratio R_2/R_1 is known from equation 594. Its inverse is R_1/R_2 . The values of V_{IH} and V_{OL} are known. Consequently, this equation can be used to calculate the bias voltage V_b . This is the second of our two design equations for the non-inverting Schmitt trigger.

Example

Design a non-inverting Schmitt with the specifications listed in the table of figure 356 (page 408).

Solution

Using equation 594, we calculate the ratio of $R_2 : R_1$.

$$\begin{aligned}\frac{R_2}{R_1} &= \frac{V_{OU} - V_{OL}}{V_{IU} - V_{IL}} \\ &= \frac{2.5 - 0}{1.45 - 1.05} \\ &= 6.25\end{aligned}$$

Using equation 595 and substituting the known values, we have:

$$\begin{aligned}V_b &= \frac{V_{IH}}{1 + \frac{R_1}{R_2}} + \frac{V_{OL}}{1 + \frac{R_2}{R_1}} \\ &= \frac{1.45}{1 + \frac{1}{6.25}} + \frac{0}{1 + 6.25} \\ &= 1.25V\end{aligned}$$

14.7 Summary of Design Equations

Inverting

Resistor Ratio $\frac{V_{OU} - V_{OL}}{V_{IU} - V_{IL}} = 1 + \frac{R_2}{R_1}$ Equation 588

Bias Voltage $V_b = V_{IL} \left(\frac{R_1}{R_2} + 1 \right) - V_{OL} \frac{R_1}{R_2}$ Equation 590

Non-Inverting

Resistor Ratio $\frac{V_{OU} - V_{OL}}{V_{IU} - V_{IL}} = \frac{R_2}{R_1}$ Equation 594

Bias Voltage $V_b = \frac{V_{IH}}{1 + \frac{R_1}{R_2}} + \frac{V_{OL}}{1 + \frac{R_2}{R_1}}$ Equation 595

14.8 Fan Thermostat, Design Example

Figure 358 shows an application of the Schmitt trigger, a thermostat for a small, 12VDC cooling fan. The fan should turn ON at a temperature of 45°C and turn off at 35°C, so there is 10°C thermal hysteresis. (If you set the hysteresis band too small, the fan will cycle ON and OFF rather frequently, and the noise of a cycling fan is quite distracting.) We assume that the fan can cool the temperature in this fashion when it's operating.

Here are the design steps.

1. We decide to use one section of an LM339 quad comparator, because it's inexpensive, popular (ie, readily available) and can operate off a single power supply. (See section 35.6 for more on the LM339).

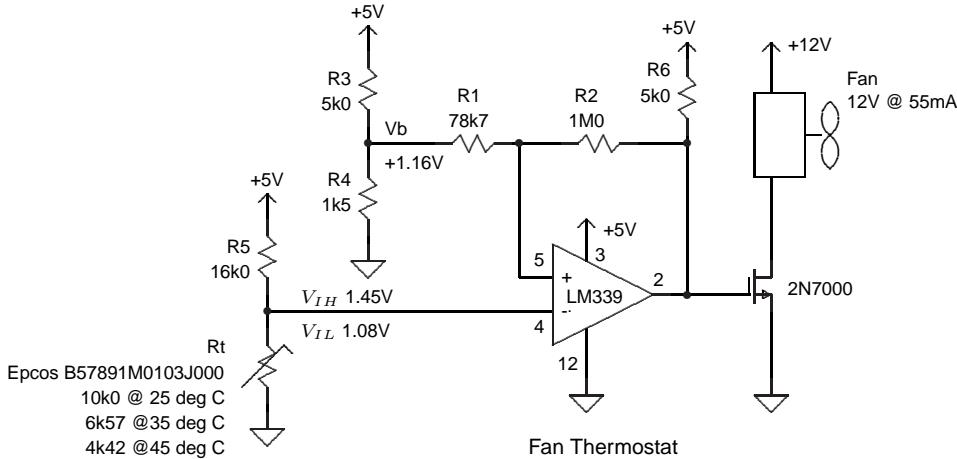


Figure 358: Fan Thermostat

2. As part of a larger system, we could operate the comparator itself from just about any supply voltage, but +5V is readily available so we'll use that.
3. The output of the comparator must be *pulled up* through a resistor to a supply voltage, to define the output voltage. This resistance should be low enough that it's much smaller than the combination of R_1 and R_2 , but not so small that the comparator has to sink a lot of current. In this case, we somewhat arbitrarily use a 5k0 pullup resistance for R_6 . We have now established the output voltages of the comparator as $V_{OU} = 5V$, $V_{OL} = 0V$.
4. The fan current is 55mA, which is beyond the capability of the LM339. We'll use a small, popular N-Channel MOSFET, the 2N7000, to drive the fan. As is typical for MOSFETs, the data sheet for the 2N7000 shows that a gate voltage of +5V from the comparator will turn the MOSFET completely ON. (We could equally well use a small NPN BJT, but that would require an additional resistor in the base of the BJT.) We assume a 12 volt supply is available in this system.
5. Now let's look at the temperature sensor. There are many devices that could be used, but the *thermistor* is inexpensive. It's resistance decreases rapidly with increasing temperature, simplifying the design of the sensing circuit. The disadvantage of the thermistor is its large tolerance, about $\pm 13\%$ overall. Perusing catalogues, we find the Epcos B57891M0103J000.

The Epcos thermistor data sheet shows the resistance is $10k\Omega$ at room temperature ($25^\circ C$), $6k57$ at $35^\circ C$, and $4k42$ $^\circ C$. We place the thermistor in the lower position of a voltage divider (R_t in the schematic) in order to create a voltage signal that changes with temperature. The upper resistance of the divider R_5 is somewhat arbitrary. A good starting point is to make this resistor equal to the thermistor resistance at room temperature. On the other hand, we should increase R_5 minimize the current through the thermistor to minimize self-heating. However, if we make R_5 too large, that reduces the output signal from the voltage divider. Somewhat arbitrarily, we've chosen the resistance to be $16k\Omega$.

The self heating of the thermistor could be an issue, so we should check that. The power dissipated in the thermistor is in the order of $370\mu W$. The thermal resistance is given on the data sheet as $1/(3.5 \times 10^{-3})$

degrees C/watt. Then the thermistor self-heating temperature rise is 0.1°C , which is acceptable.

6. Now we can calculate the output voltages from the $R_5 : R_t$ voltage divider, at the two threshold temperatures, using a supply voltage of +5V, $R_5 = 16k$, and R_t at the two different temperatures. These are the values of V_{IU} and V_{IL} for our design equations.

$$V_{IU} = 1.45V, V_{IL} = 1.08V$$

7. We have established the values of $V_{OU} = 5V$, $V_{OL} = 0V$, $V_{IU} = 1.45V$, $V_{IL} = 1.08V$, so we can calculate the ratio of the resistors R_1 and R_2 , using equation 588. We find $R_2 : R_1 = 12.6$. These resistors should be large compared to R_3 and R_6 , so we'll make R_2 as large as possible, but not so large as to cause problems with bias and other stray currents: say, $1M0\Omega$. Then resistor R_1 is 12.6 times smaller, $79k3\Omega$. The nearest standard 1% value is $78k7\Omega$.
8. Now we have sufficient information to calculate the bias voltage V_b . Using equation 590, we find $V_b = 1.16$ volts.
9. Finally, we can design the $R_3 : R_4$ voltage divider to produce 1.16 volts from the 5 volt supply. The resistances in this divider should be much less than the series combination of $R_1 : R_2$, in order that the divider will appear as a voltage source. If the divider resistances are too large, the value of bias voltage V_b will change when the comparator switches state.

We're already using a 5k resistor for R_6 , and for production economy it's a good idea to minimize the number of different component values. Choosing $5k\Omega$ for R_3 , we calculate $1k5\Omega$ for resistor R_4 .

This completes the basic design of the thermostat-controlled fan.

However, all of these components have tolerances, so the exact operation of the circuit (the switching temperature for example) will vary with the actual values of the components. It would be prudent, especially if accurate operation is a requirement, to create a spreadsheet analysis of the circuit, showing high and low values for the various components, and the effect on the circuit operation. This will ensure that the operation of hundreds or thousands of these circuits will work correctly regardless of component tolerance.

Testing and Debugging

It's common for a circuit not to work correctly on its first operation. The circuit can be subject to conceptual errors, miswirings, incorrect components, and calculation errors. In this circuit, you could check the following:

- Power supply voltages
- Bias voltage V_b .
- Input voltage, which should vary with temperature. If you have a temperature probe, check that the input voltage correctly corresponds to temperature.
- Voltage at the junction of R_1 and R_2 , which should be equal to V_{IU} or V_{IL} , depending on the state of the comparator output.
- Replace the thermistor by a 10k adjustable resistor. Vary the resistor and observe that the output of the comparator switches at the correct input voltages.

14.9 Appendix: Confirming the Schmitt Trigger Lever Diagram

In this section, we confirm the lever diagram for the inverting Schmitt trigger¹⁴⁵.

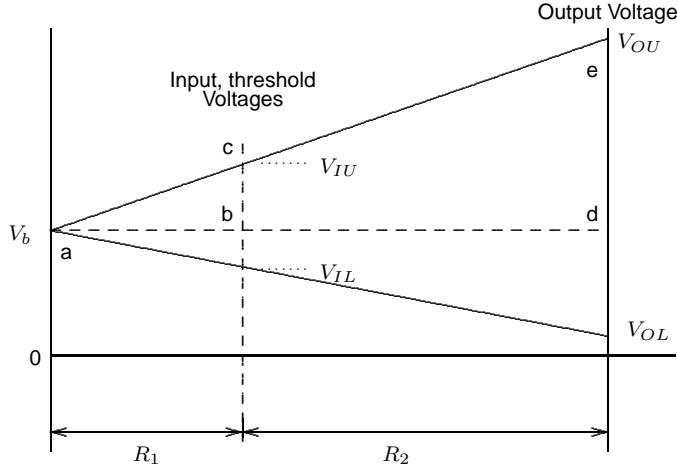


Figure 359: Inverting Schmitt Trigger Lever Diagram

Figure 359 shows a suitably labeled lever diagram for the inverting Schmitt trigger. Our objective is to use the circuit diagram for the inverting Schmitt, similar to figure 351, to establish that the ratio of R_1 to R_2 is as shown in the diagram of figure fig:confirming-lever-diagram, that is, the line ab is to R_1 as the line ad is to $R_1 + R_2$..

By geometry, triangles abc and ade are similar. Then their sides are in proportion, ie:

$$\frac{ab}{ad} = \frac{cb}{ed} \quad (596)$$

According to the diagram:

$$cb = V_{IU} - V_b \quad (597)$$

$$ed = V_{OU} - V_b \quad (598)$$

Using superposition, as shown in equation 584, we can write:

$$V_{IU} = \left[V_b \frac{R_2}{R_T} + V_{OU} \frac{R_1}{R_T} \right] \quad (599)$$

where $R_T = R_1 + R_2$. Substitute V_{IU} from equation 599 into equation 597. Then, using equation 598, we can write:

$$\frac{cb}{ed} = \frac{\frac{1}{R_T} (V_b R_2 + V_{OU} R_1 - V_b R_T)}{V_{OU} - V_b} \quad (600)$$

Replace R_T by $R_1 + R_2$ and after some algebra, we obtain:

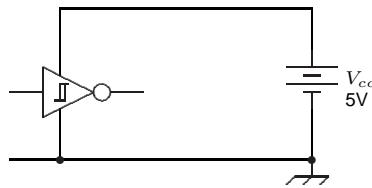
$$\frac{cb}{ed} = \frac{R_1}{R_1 + R_2} \quad (601)$$

Consequently, the diagram correctly represents the proportions of R_1 and R_2 , and the lever diagram is a correct representation of the circuit operation.

¹⁴⁵The non-inverting Schmitt circuit is left as an exercise.

14.10 Exercises

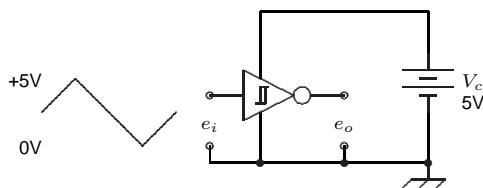
1. The Schmitt trigger is available as a digital integrated circuit inverter, type 74HC14. Its symbol and connection to the power supply are shown below. The output of this device switches to 0 volts when the input is above threshold and to V_{CC} (typically 5 volts) when the input is below threshold.



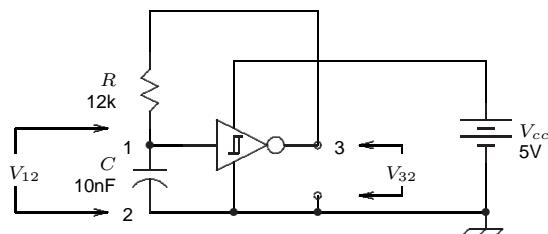
The threshold voltage depends on the current state of the output:

Output Voltage	Threshold Voltage
V_{CC}	$2/3V_{CC}$
Gnd	$1/3V_{CC}$

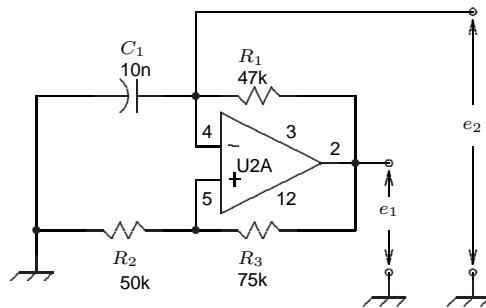
- (a) The 74HC14 schmitt trigger is supplied from a 5 volt supply, so that $V_{CC} = 5V$. A 5 volt peak-peak triangle waveform of frequency 1kHz is connected to the input of the schmitt trigger, as shown in the figure. Sketch the input and output voltage-time waveforms.



- (b) The same Schmitt trigger is set up in the oscillator circuit shown below. Sketch the amplitude-time waveforms V_{12} and V_{32} .



2. The circuit shows an op-amp Schmitt-trigger oscillator. The output of the op-amp can swing between the power supply rails, which are ± 10 volts.



Determine the amplitude-time waveforms e_1 and e_2 .

15 Power Supplies

Previous sections have described various power supply components. In this section, we'll look at complete power supply circuits and methods of their design.

The DC power supply, whether a battery or line-derived voltage, is a required part of any electronic circuit. Electronic systems typically require multiple voltages and currents, including both positive and negative polarities. Power supplies are an essential part of the electronics vocabulary. The cost and performance of the power supply can be critical to the success of the complete system.

In this section, we'll deal exclusively with line-derived power supplies, that is, power supplies for which the input is 117VAC RMS at 60Hz. Readers in other jurisdictions, which use other voltages and frequencies, can extrapolate to other standards.

When the requirements are not particularly critical, the power supply will consist of a transformer, rectifier (full or half wave) and a filter capacitor. This type of supply is commonly available as an *AC Adaptor* or the slang *Wall Wart*: a small epoxy brick which plugs into a wall outlet and provides a DC voltage. We'll have more to say about this device in section 15.7, page 435.

When the power supply voltages must be stable, it is common practice to use an integrated circuit regulator to stabilize the voltage and limit the output current. The regulator may be a linear device, which sustains a variable voltage drop between the raw, unregulated input voltage and the stable output voltage. Linear regulators are readily available at very low cost and trivial to apply. However, if they conduct appreciable current they require heat sinking. In some applications, the heat sink, generated heat, and inefficiency may be unacceptable.

In such cases, it may be more attractive to use a switching regulator which sustains the average voltage drop by varying the duty cycle of a switch and averaging the result. The switching regulator is much more efficient and requires a much smaller heat sink, but it also requires more circuitry. In cases where the circuitry deals with low-level signals, the electrical noise from the switching regulator may be problematic.

The description of the switching regulator in section 15.8 is quite lengthy and detailed. To illustrate that the design of a switching regulator is not excessively complicated, a worked design example is provided in section 15.9.

The switching regulator described in these sections is the *buck* or *step-down* regulator, the most common configuration. There are also switching regulator configurations that can step-up or invert the input voltage. Their circuits are provided in connection with problems 3 and 5 in the power supply exercises (page 488).

Finally in this section we look at voltage multipliers and voltage inverters that can be implemented without inductors. These circuits are useful where an inexpensive, low-current supply is required.

15.1 AC Power Distribution

The structure of a building's AC power distribution system affects the design of power supplies and has an impact on noise control in an electronic system, so it is important to understand. A representative schematic of a power distribution system (in North America) is shown in figure 360.

- The power distribution system begins at a hydro-company transformer, typically located on a pole on the street or in an underground street vault. The transformer steps down the local distribution voltage with a centre-tapped transformer that delivers a nominal 120VAC either side of the centre tap. Then a high power load (such as an oven or clothes-dryer in a house) can be connected between the two outer lines to receive a nominal 240 volts.
- The three leads from the transformer arrive at the power distribution box, typically found in the basement of a house. The two hot lines are first fed through a DPST circuit breaker of 200 Amp (or similar) capacity. (In older buildings this will consist of two cartridge fuses and a DPST disconnect switch.) This provides a mechanism for disconnecting power from the entire building or one area of a building.

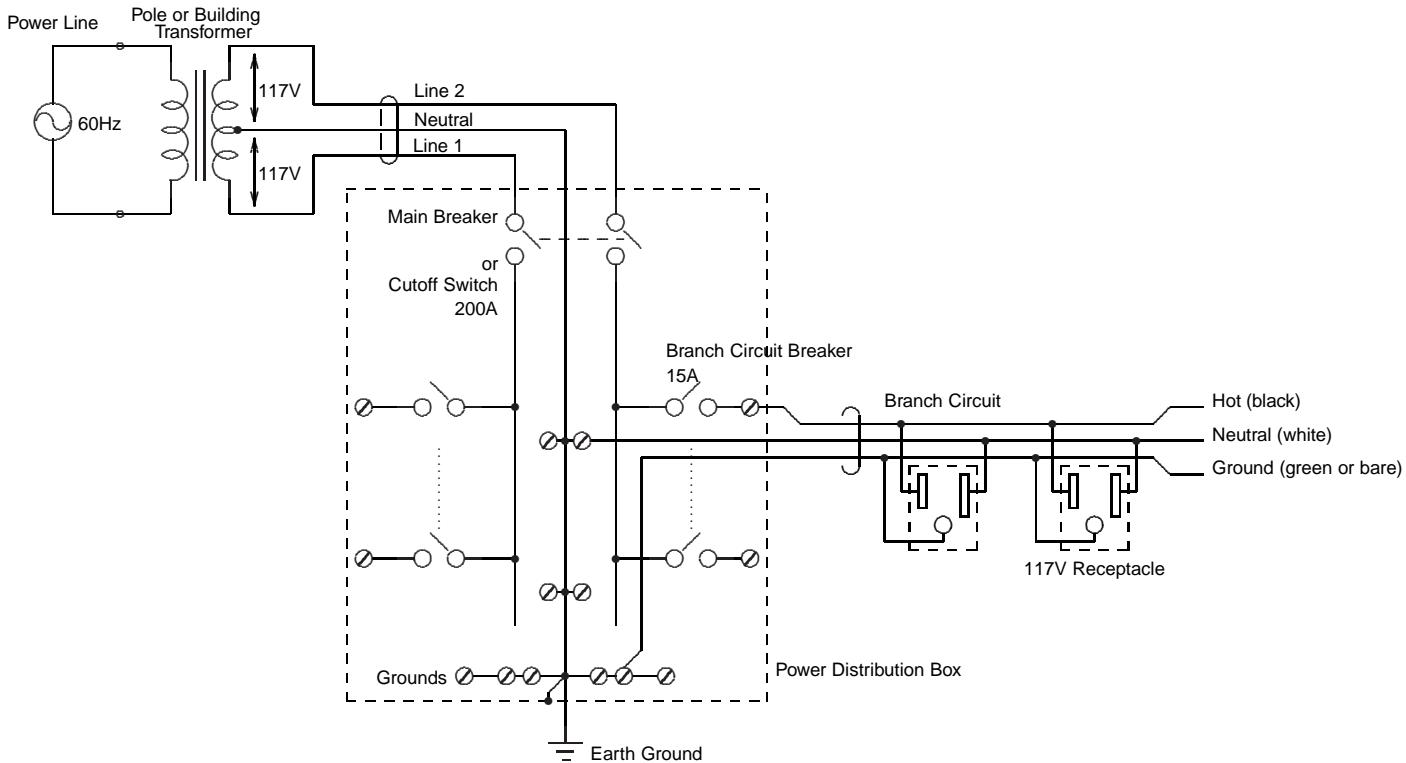


Figure 360: AC Power Distribution Wiring

- The hot lines and the neutral are then distributed to the various branch circuits, each of which has its own fuse or circuit breaker and provides a maximum current in the order of 15 amps or so.
- The branch circuit wiring has 3 conductors: hot, neutral and ground. The hot and neutral wires carry current. The ground connection is for safety and normally will not carry current.
- Back at the distribution box, the neutral and the ground wires are connected together. The neutral is also connected to an *earth ground*, which might be a connection to a cold water pipe.

This connection is an important safety feature. If nearby hydro lines are struck by lightning, then the ground connection will hopefully divert the current pulse to ground.

Safety and Grounding

The *grounding* safety feature is illustrated in figure 361. A three-wire appliance is powered from the AC line via a duplex receptacle¹⁴⁶. The metal enclosure of the appliance connects to the safety ground wire, which eventually leads back to the earth ground of the power distribution system.

Now consider that a fault occurs in the form of a connection between point A (the *hot* connection) and the enclosure. This will provide a short-circuit path for the 120V supply via the ground connection, and trip the circuit breaker.

¹⁴⁶Notice that the blade terminals are slightly different sized. The larger is the neutral connection [110].

Without the ground connection to the enclosure, a fault between the hot line and the enclosure would cause the enclosure to become live with 120V. If a human touches the enclosure, then the current finds a path to ground via the human, with possibly serious results.

The injury caused by the electric shock depends on the body current, which in turn depends on the resistance of the human connection to the enclosure and ground. For example, rubber-soled shoes act as a high resistance, reducing the current and consequent danger to the human participant.

Now consider that a fault occurs between the neutral connection at B and the enclosure. Return current from the device (and other devices on the same circuit) normally flows through the neutral connection back to the power distribution box. When there is a fault, the ground wire appears in parallel with the neutral and will share its current. Again, the ground connection to the enclosure ensures that it cannot rise to a dangerous potential.

Some outlets are equipped with a *ground-fault interrupter*, GFI. This device measures the currents on the hot and neutral lines. If there is no fault, then these currents will be exactly equal. If there is a fault, then some current will flow on the ground wire and the hot and neutral currents will be unequal. This causes the GFI to disconnect the circuit.

15.2 Power Supply Configuration

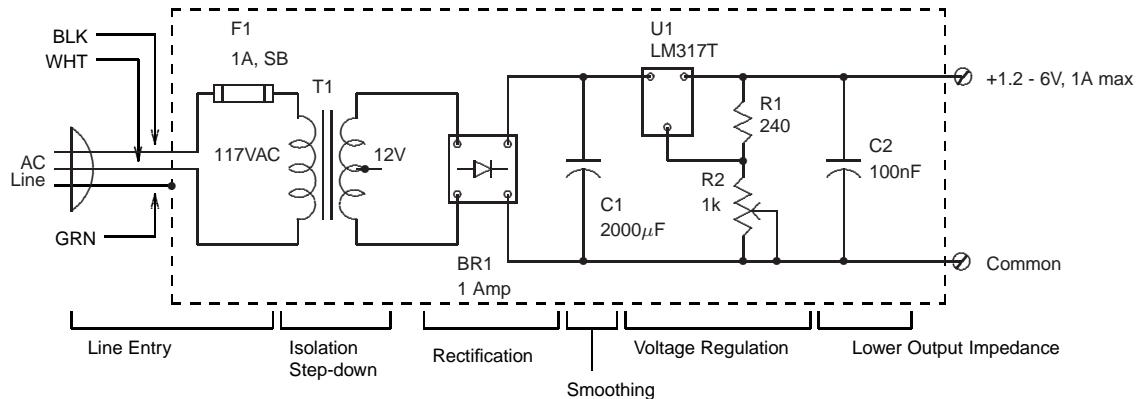


Figure 362: A DC Power Supply

In an electronics system, the power supply accepts the raw AC line voltage or battery voltage and generates the required DC supply voltages for the system electronics. An example of a simple, series-regulated linear power supply¹⁴⁷ is shown in figure 362. The function of each item in the power supply is as follows:

¹⁴⁷This supply is usable as shown as a variable voltage bench power supply. The regulator U_1 must be mounted on a substantial heatsink.

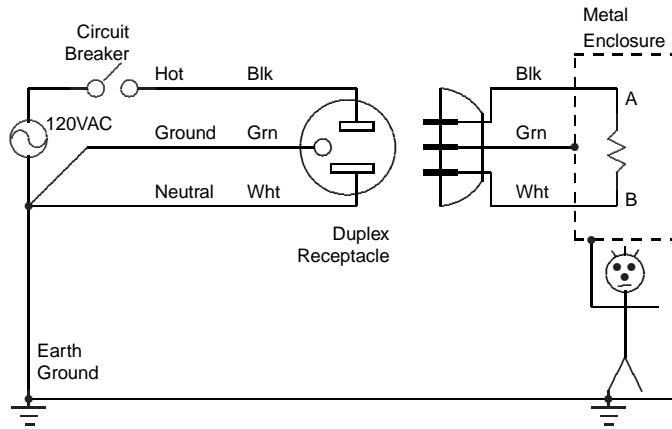


Figure 361: AC Power Wiring Safety

Line Entry This section of the power supply connects to the AC power outlet. At a minimum, it must include some sort of current-limiting device, typically a protective fuse or circuit breaker. It may also include filters (which prevent noise travelling from the AC line into the device or vice-versa), a power on-off switch, and an indicator lamp.

The diagram shows a 3-wire connection. Notice the colours of the line cord. The black (BLK) lead is HOT and goes to the fuse (and switch, if there is one). The white (WHT) lead is the neutral and carries the return current. The green (GRN) lead is connected to the metal chassis of the instrument. Via the AC line connection, it is eventually connected to an earth ground.

If the circuit is contained in a plastic box so that a human cannot touch the circuitry, then a 2-wire power connection is possible, and the third wire, the ground, is not required. However, even two-wire connections are polarized (one blade on the plug is wider than the other) so that the hot and neutral are defined correctly.

This section of the power supply is of interest to the regulatory agencies such as CSA (Canadian Standards Association), U/L (United Laboratories, USA), or IEC (Europe). They must be satisfied that the design does not constitute a shock or fire hazard. For example, *flaming debris* from the power supply must be safely contained in the enclosure.

Isolation Isolates the ground of the power supply system from the electronics ground, in order to reduce the shock hazard.

Step-up, Step-down In the days of electron tubes, this function would step up the line voltage, for example from 117VAC to 220VAC. These days of solid-state electronics, it's likely to be a step-down function, more like 117VAC to 8VAC. Isolation and step-up/down are often accomplished simultaneously in a transformer. When multiple voltages are required, they are may be created at this stage with multiple transformer secondary windings.

Rectification Changes the AC waveform into some form of DC, typically half or full-wave rectified, pulsating DC.

Filtering Provides a charge storage function which converts the pulsating DC waveform to a relatively constant DC voltage.

Regulation Regulates the output voltage against fluctuations due to variations in the input voltage (line regulation) and variations in the load current (load regulation). Voltage regulators may also serve the function of isolating certain noisy load currents from other sections of the circuit. For example, digital circuitry is electrically noisy and often must be isolated from analog circuitry.

Lower Output Impedance An ideal voltage source has zero internal impedance. This capacitor shorts together the output terminals of the supply for AC signals and provides a reservoir of charge for loads that make a sudden demand of current.

Distributed vs Local Regulation

Originally, a power supply regulator was constructed from discrete components. It was a centralized unit and the regulated output was distributed to the various circuits it serviced.

When three-terminal regulators became available in the 70's, another approach was possible: distributed regulation. A centralized power supply performed the step-down, rectification and filtering operations. The output, an unregulated DC voltage, was distributed to the various circuits, at which point a three-terminal linear regulator created the required regulated voltage. This approach has three advantages:

- Regulation is at the load point rather than some distance from the load. As a consequence, there is less wiring between the regulator and the load and the regulation is better.
- The various loads are electrically isolated by the various regulators from each other, reducing noise crosstalk problems.
- Instead of one large heatsink at a central regulator, the system uses a small heatsink at each of the three-terminal regulators. This may simplify the mechanical packaging of the system and improve access of the heatsink to the cooling air flow.

Multiple Voltage Supplies

It is usual for an electronic system to require a variety of different power supply voltages. In the recent past, for example, a mixed analog-digital system would typically require +5V for the logic and microprocessor, and $\pm 12V$ for the op-amps and linear circuitry. Multiple supply voltages add significant cost to a system so more recent system designs have been focussed on op-amps that can function with a single supply and rail-rail operation. A rail-rail op-amp uses the full range between the power-supply voltage and ground for signals, thereby improving the dynamic range.

However, digital logic supply voltage, traditionally at +5 volts, is migrating to lower values. As it does so, it is common for a system to require a mix of devices with different supply requirements. Once again, a system must accommodate a variety of supply voltages.

Traditionally, multiple supply voltages have been provided in a centralized transformer-linear regulator system similar to that of figure 362.

The current method favours one raw DC supply followed by many switching regulators, each with its own sensing and regulation system. Modern circuit integration can put several switching regulators and their ancillary circuitry on a single IC chip. Modern computers require the high efficiency of switching regulators in order to minimize cooling and heatsinking requirements.

15.3 Rectifier-Filter Configurations

Some useful rectifier configurations for line-operated power supplies of modest requirements are shown in figures 363 and 364. All these circuits are shown with a capacitor filter, which is the norm. We'll discuss the inductor filter in a moment.

Half-Wave

Figure 363(a) is a half-wave rectifier. Capacitor C charges through diode D on alternate positive half-cycles of AC voltage from the transformer. The capacitor charges to the peak value of the secondary voltage during the charging interval and then discharges between charging pulses, as shown in the upper half of figure 365, page 423.

The transformer secondary current is a brief positive pulse of charging current with each positive AC cycle, shown in the lower half of figure 365.

This circuit is usually only useful for small load currents. For a given ripple voltage and load current, the capacitor must be large – twice the size of the filter capacitor in the full wave rectifier, described below.

Dual Polarity Half-Wave

Figure 363(b) adds in a second rectifier diode and filter capacitor to generate a negative DC voltage. In this version, if the load currents are equal the current pulses alternate positive and negative polarity in the transformer secondary, so the average DC current in the secondary is zero and saturation is not a problem.

Full-Wave Centre Tapped

This configuration, figure 363(c), uses two equal secondary windings driving half-wave rectifiers. The voltages at the two diode inputs have opposite polarity, so the diodes conduct on alternate half cycles, generating a full-wave output.

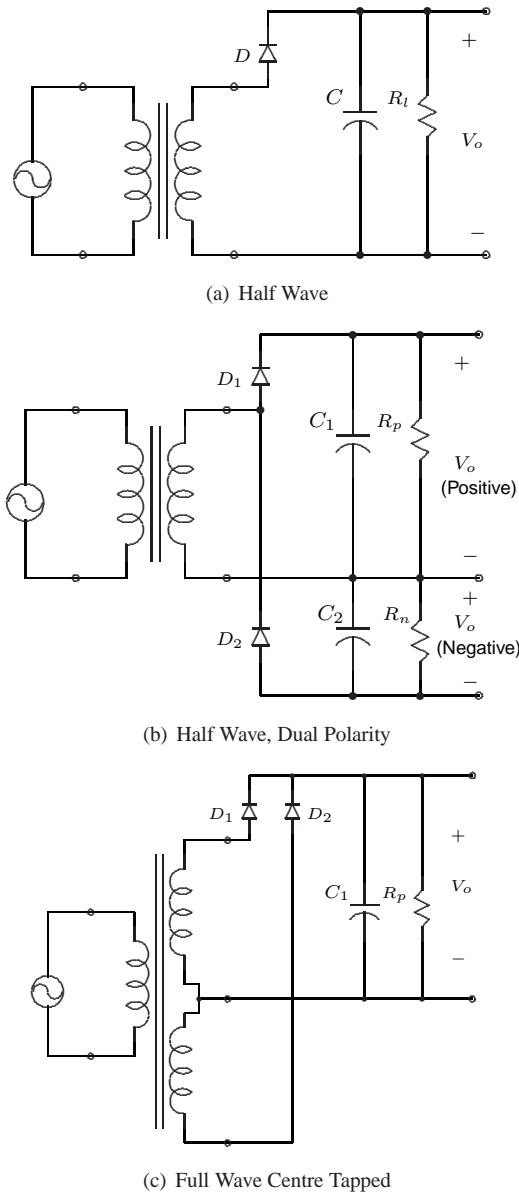


Figure 363: Configuration Set #1

Full Wave Bridge

In general, transformer windings are more expensive than diodes. The full wave bridge configuration in figure 364(a) eliminates one secondary winding and adds two diodes to have the same effect, and so it's the preferred configuration for a simple full-wave system.

The diode pair D_1, D_4 alternates conduction with D_2 and D_3 . The transformer secondary does not conduct DC, and the capacitor is recharged twice during each AC cycle. Notice that the peak value of the capacitor voltage is two diode voltage drops below the peak transformer secondary voltage.

Full Wave Bridge, Dual Polarity

The configuration of figure 364(b) generates full-wave voltage at both positive and negative polarities. It can be regarded as a centre-tapped bridge rectifier configuration or as two full-wave centre-tapped rectifiers.

Capacitor Filter: Idealized

Most electronic power supplies use a capacitor to filter the pulsating output from the rectifier system. The capacitor stores charge when the input voltage exceeds the output voltage, and then discharges when the input voltage drops below the output voltage.

First, we'll describe a simplified situation – where the resistance of the transformer windings is zero. Then we'll describe a more realistic situation for design purposes.

When the source resistance (the transformer winding resistances) is zero, the waveforms are as shown in figure 365 for a half-wave power supply.

Notice these points of interest:

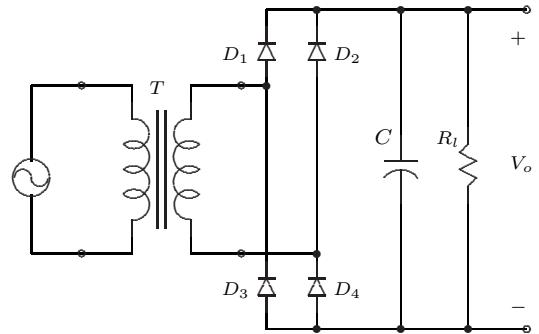
- Referring to the voltage waveforms in the upper half of the diagram:

The output voltage follows the input voltage, minus a small drop across the diode, when the input voltage exceeds the output.

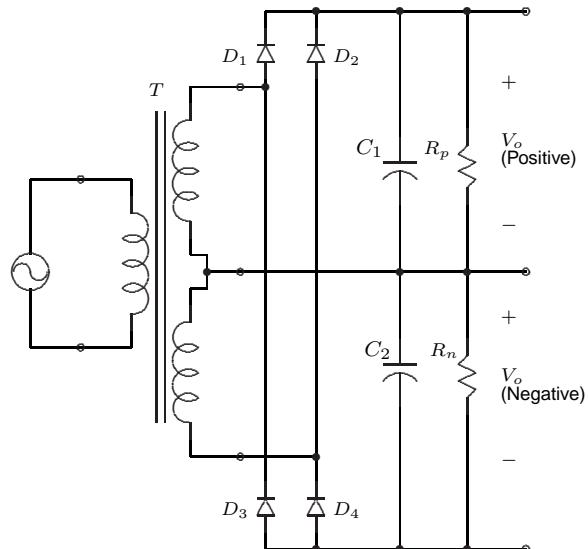
- The output voltage runs down between charging intervals. If the load is a voltage regulator that connects to a fixed load resistance, then the discharge current is constant and the discharge rate linear. If the load is a resistor, then strictly speaking the discharge rate is the usual exponential RC discharge curve.

However, because the time constant is large, the curve may be approximated as linear. Then the rate of discharge in volts/second is given by:

$$\frac{dV}{dt} = \frac{I}{C} \text{ volts/sec} \quad (602)$$



(a) Full Wave Bridge



(b) Full Wave CT, Dual Polarity

Figure 364: Rectifier Configurations #2

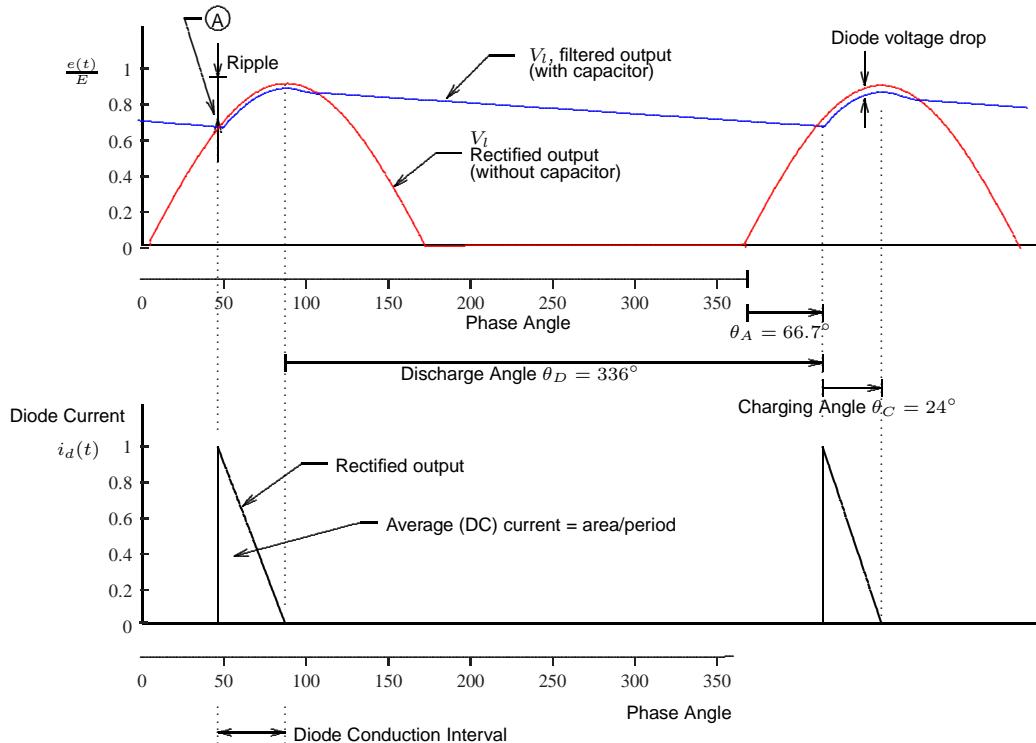


Figure 365: Half-Wave Rectifier, Waveforms

where I is the load current in amps, C is the filter capacitance in farads.

- The *peak-peak ripple voltage* is equal to the maximum output voltage minus the minimum output voltage. To reduce the ripple, according to equation 602 one would either reduce the load current (often not an option) or increase the filter capacitance.
- Referring to the diode current waveform in the lower half of the diagram:

This current occurs in triangular pulses [111]. The average value of these pulses must be equal to the DC output current. Since the pulses are of limited duration, their amplitude is *much* larger than the average output current: a ratio of 6:1 is not uncommon¹⁴⁸

Here's an important point: *if the filter capacitance is increased (to reduce the ripple voltage), then the diode conduction interval will decrease. That means the peak value of the diode current must increase.* This large pulse of current may cause voltage drops in the supply wiring that create noise in the circuit.

Capacitor Filter and Linear Regulator: Tradeoffs

Now let us consider a common scenario: the output of the capacitor filter driving a linear voltage regulator (figure 362 on page 418). The regulator compensates for a variety of shortcomings in the raw capacitor voltage. It

¹⁴⁸In one memorable incident, the author was present when a newly-designed DC power supply was being demonstrated to an important client. The power supply burst into flame because the peak diode current had not been calculated properly and the rectifier diodes were underrated. The client was not impressed.

removes the ripple voltage and produces a stable output voltage in the face of changes due to varying line voltage and varying load current. It also limits the output current if there is a catastrophic short circuit. It does all this for a few pennies, so it's a worthwhile investment in most electronic systems¹⁴⁹.

To work properly, a voltage regulator must have an input voltage that exceeds the output by an amount called the *dropout* voltage (section 15.6, page 433), which might be 2 volts. As a result, the capacitor filter must be designed to ensure that the capacitor voltage always exceeds the regulator output voltage plus the regulator dropout voltage. The worst-case minimum capacitor voltage occurs:

- At the end of the capacitor discharge interval, just before the diode starts conducting again
- For a *low line voltage condition*, when the line voltage may be low by 10%, and so the transformer secondary voltage is also low by 10%.
- At the maximum output current, which causes the most rapid discharge of the filter capacitor.

On the other hand, the power dissipation in a linear voltage regulator is equal to its input-output voltage differential times its output current. Larger power dissipation requires a larger heatsink – or in extreme cases, a fan. Heatsinks are big and expensive, fans are noisy and unreliable. So the dissipation in the regulator should be kept to the absolute minimum that ensures it will work correctly. Since there is usually no control over the output current the dissipation is determined by the regulator input voltage. As a consequence, the capacitor voltage should be large enough but no larger than necessary¹⁵⁰.

To summarize: in the classic paper by Schade [111], and in more recent papers such as [112], the emphasis is on the *average* voltage of the capacitor. When a linear voltage regulator is used, it's the minimum voltage that is critical.

Transformer Resistance

It turns out that the waveforms of figure 365 are unrealistic in a practical design. The culprit is source resistance: the winding resistances of the power transformer.

At first blush, it would appear that transformer winding resistance could be neglected since it only amounts to a few ohms. However, the current through the transformer winding is not the average load current, it is a short-duration high amplitude current pulse. For example, if the winding resistance is one ohm, the average load current one amp, and the peak current 6 amps, the maximum voltage drop across the winding resistance is a non-negligible six volts. This voltage drop significantly hinders the charging of the filter capacitor.

This effect is particularly evident for the half-wave rectifier, which must supply the average output current in pulses 1/60 seconds apart. As a result the rectifier system should be full wave whenever possible. Unfortunately in some cases, such as the single-transformer-secondary dual-output supply of figure 363(b), half-wave is the only option.

To provide some perspective on the winding resistance in small power transformers, figure 366 shows measurements on a group of similar Hammond power transformers that might be used in electronic power supplies. The table lists winding resistance as a function of secondary voltage and current.

The primary winding resistance is seen from the secondary as reduced by a factor N^2 . N is the transformer primary:secondary turns ratio, equal to the primary:secondary voltage ratio.

Then the effective secondary resistance is the sum of the reflected primary resistance plus the secondary resistance.

$$R_t = \frac{R_{pri}}{N^2} + R_{sec} \quad (603)$$

¹⁴⁹The stable output voltage of the common-garden-variety linear regulator is a great achievement in integrated device design, because the output voltage must be stable over the large range of temperatures that a linear regulator is regularly required to endure.

¹⁵⁰This is the case for a *linear* regulator, which controls the output by varying the resistance in the current path. A *switching regulator* (section 15.8 on page 437) works entirely differently and largely avoids the issue of power dissipation.

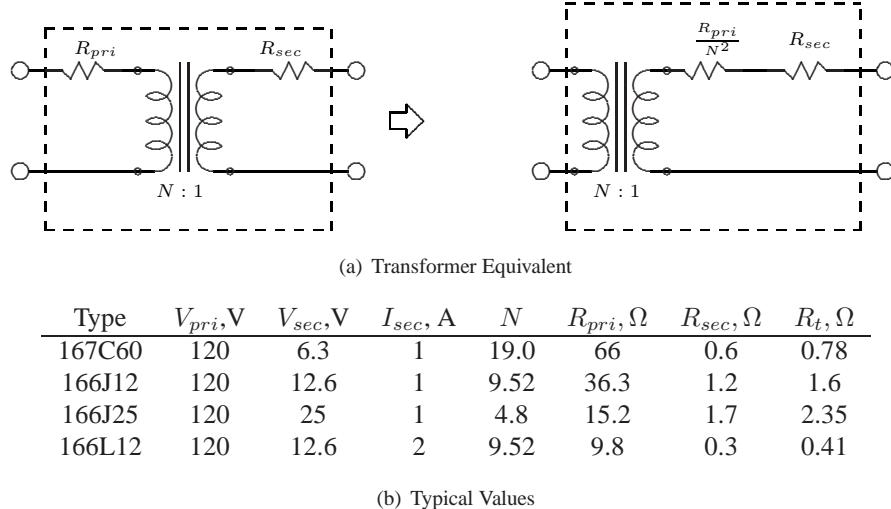


Figure 366: Transformer Resistance

where R_{pri} is the resistance of the primary winding, R_{sec} is the resistance of the secondary.

Notice:

- The first three entries of figure 366 compare transformers with the same current output and different secondary voltages. The total resistance increases with transformer secondary voltage. That makes sense: the wire size stays the approximately the same at the same current level, and the number of turns increases. As a result, it is somewhat counter-productive to compensate for transformer winding resistance by increasing the secondary voltage.
- The second and last entries compare two 12 volt transformers with 1 and 2 amp output. The ratio of resistance is inversely proportional to the *square* of the output current. Again, that makes sense. The power losses in the transformer are an I^2R effect. To double the output current and keep the I^2R losses equal, the resistance must decrease by a factor of 4.

The table gives some indication of the magnitude of the resistance and the effect of transformer secondary voltage and current. However, these figures should not be regarded as universal. The results could be quite different for another transformer manufacturer or for another type of transformer - such as an AC wall adaptor. Obtain specifications from the manufacturer or measure representative units.

Capacitor Filter and Transformer Resistance

Now that we have some idea of the magnitude of the source resistance, we can determine the behaviour of a given rectifier circuit. This is a non-linear circuit, so it must be solved by manual iteration or by simulation.

Circuit simulation must be used with caution. As someone said of computers, they are capable of generating incorrect answers really, really quickly. If the circuit component simulation models are incorrect, the simulation can be very misleading¹⁵¹.

In this case, the circuit models are relatively simple and manual or graphical calculation of the circuit operation is difficult and time consuming [111], so simulation is an attractive alternative.

¹⁵¹This suspicion of simulation by analog circuit designers is best exemplified by the picture of Bob Pease, a guru of the analog design world, hurling a computer monitor off the roof of a building [113], p145.

The simulation of a 12 volt, 1 amp half-wave rectifier circuit is shown in figure 367. The oscilloscope traces in figure 367(b) show the input and output voltages at startup of the circuit. The single trace in figure 367(c) is the voltage across the source resistance, which indicates the charging current.

Most obviously, the output waveform bears no resemblance to that shown in figure 365. With zero source resistance, we would expect the capacitor to charge to the peak input voltage, 17.8 volts. In this circuit, with a 1.6Ω source resistance, the steady-state maximum rectified voltage is 10.5 volts, much less. The simulation shows that the minimum rectified voltage, which is critical for a regulator input, is 7.0 volts.

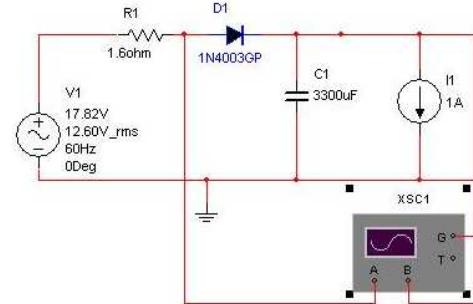
The voltage difference, between the peak input voltage of 17.8 volts and the output voltage, is caused by the pulse of charging current in the 1.6Ω source (transformer winding) resistance.

Notice the voltage drop across the rectifier diode: it's approximately one volt, significantly above the usually assumed 0.6 volts.

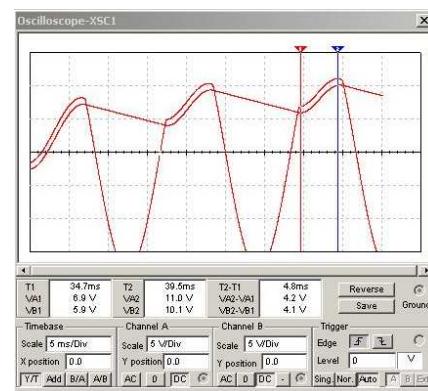
The charging current in figure 367(c) was measured with the simulated oscilloscope, by measuring the voltage across the source resistance. The current peaks at $10V/1.6\Omega = 6.25$ amps. The RMS value of the current waveform is 2.2 amps (section 15.20 on page 483). A conservative design would rate the transformer and diode somewhat larger than this value.

The simulation makes it possible to explore different measures to increase the output voltage. Here are the results for varying different circuit parameters:

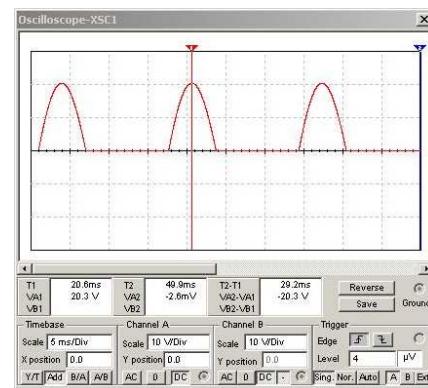
- Converting to a full-wave bridge rectifier increases both the minimum and maximum output voltage and (approximately) halves the peak-peak ripple voltage.
- Doubling the transformer current rating reduces the transformer resistance by a factor of 4, increasing the both the minimum and maximum output voltage. The peak-peak ripple voltage increases slightly.



(a) Circuit



(b) Voltage Waveforms



(c) Current Waveform

Figure 367: Rectifier Simulation

Circuit	Minimum V_o	Maximum V_o	Ripple, p-p V
Original circuit	7.0	10.5	3.5
Change rectifier to full wave	10.5	11.9	1.4
Double transformer current	11.5	15.5	4.0
Double transformer voltage	12.1	15.5	3.4
Double Capacitance	8.0	9.7	1.7

- Doubling the transformer voltage increases the minimum and maximum output voltage by somewhat less than a factor of 2 and has no effect on the peak-peak ripple voltage.
- Doubling the filter capacitance slightly increases the output voltages and halves the peak-peak ripple voltage.

In summary, converting to full-wave rectification is the most effective measure to increase output voltage and reduce ripple. Doubling the transformer current or voltage increase the output voltage by a similar amount and probably have similar implications in cost and size. Doubling the voltage may require increasing the voltage rating of the filter capacitor. Doubling the capacitance is effective in reducing ripple, but has little effect on the average output voltage¹⁵².

The Inductor Filter

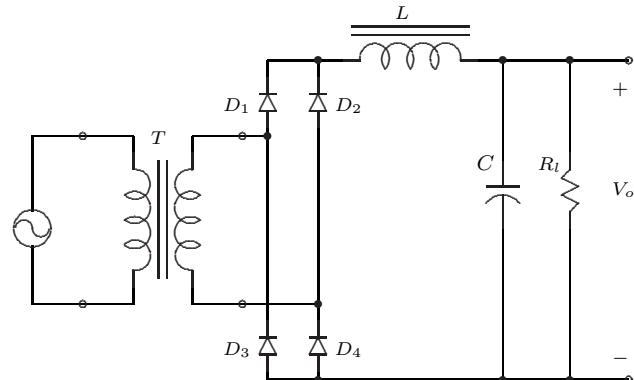
It is quite possible to use an inductive lowpass filter to extract the average from the rectified half or full wave voltage waveform.

At high frequencies this is feasible and common practice (section 15.8). However, at a 50 or 60Hz AC line frequency the inductor is large, heavy and expensive compared to a capacitor filter. Consequently, the capacitor filter is preferred.

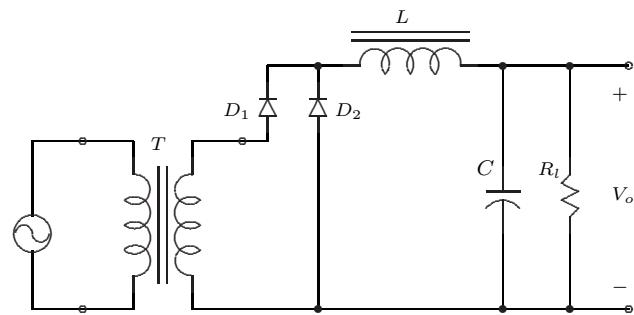
However, where very large load currents are required, the inductor filter comes into its own, even at line frequency. As we saw in the half-wave rectifier capacitor filter example of figure 365, the diode current consists of short pulses where the peak amplitude is many times larger than the average. For large average currents these peak currents are unacceptable. They cause large voltage drops in connecting wiring and make unacceptable demands of the rectifier devices. An inductive filter has the opposite effect – it tends to maintain a constant current through the diodes so that the peak diode current is similar to average current.

Salient points about the inductive power supply filter:

- The inductor filter may be regarded as an approximately constant current device. Consequently, the output DC current must flow continuously. (Open circuiting a constant current source creates a very large voltage.) The rectifier system must have provision for continuous DC current to flow. Full-wave bridge rectifiers with inductive filter as in figure 368(a) and full-wave centre-tapped rectifiers satisfy this requirement. A half-wave rectifier followed by an inductor filter requires a 'free-wheeling' (snubber) diode D_2 as shown in figure 368(b).



(a) Full Wave Bridge, Inductor Filter



(b) Half Wave Rectifier, Inductor Filter

Figure 368: Rectifier-Inductor Filters

¹⁵²Some further insight into this circuit can be obtained by interchanging the position of the diode and transformer resistance, and then thevenising the voltage divider formed by the transformer resistance and load resistance. The thevenized resistance and filter capacitance form an RC lowpass filter that extracts the average value of the half-wave rectified waveform.

- Because the inductor is a constant-current source, its internal impedance is large and this results in very poor load voltage regulation. A change in load resistance will cause a large change in load voltage. To lower the internal resistance of the supply, a filter capacitance is often connected in parallel with the load, as shown in both configurations of figure 368.
- As we saw earlier, the capacitor filter charges up to the peak value of the input voltage. However, the inductor filter extracts the average value of the pulsating input, so its DC output voltage is some fraction of the peak value. The Fourier series [114] for a full-wave rectified voltage consists of a DC term that is equal to $2V_p/\pi$, plus various AC terms that are removed (to a greater or lesser extent) by the inductor filter. For example, a 12 volt peak AC waveform produce an average voltage of:

$$\begin{aligned} V_{AV} &= \frac{2V_p}{\pi} \\ &= \frac{2 \times 12}{\pi} \\ &= 7.6 \text{ volts} \end{aligned}$$

For the modest supply currents required by electronic equipment, current practice for AC line operated transformer-rectifier power supplies is to use a capacitor filter which is allowed to produce significant ripple. This is followed by an electronic regulator that removes the ripple, provides a DC voltage that is immune to line voltage changes, and presents the load with a low source resistance (good load regulation.) We'll focus on that approach, without further investigation of line-frequency inductor filters. Even with the additional cost and complexity of electronic regulation, this is a far more compact and cost-effective approach than an inductor filter.

15.4 Series and Shunt Voltage Regulators

The *voltage regulator* is a circuit that provides a stable source of DC voltage to a load. It has two functions:

Line Regulation

In general, the voltage into the regulator will vary due to a variety of causes. The AC line voltage may vary by as much as $\pm 10\%$ from its nominal value, depending on the demands on the power grid. Furthermore, varying currents in the line cause a varying voltage drop across the line resistances. (This is why the lights in the kitchen may dim briefly when the refrigerator starts).

Load Regulation

A DC power supply has internal resistance. Consequently, any change in load current will tend to change the terminal voltage of the power supply. The voltage regulator senses this change and adjusts the output voltage to compensate, thereby effectively reducing the output resistance of the power supply.

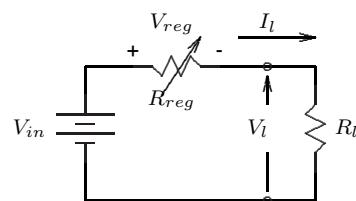


Figure 369: Series Regulator

Series Regulator

The *series regulator* circuit concept is shown in figure 369. The input voltage V_{in} must be greater than the load voltage. The series regulation element (usually a transistor, but shown here as a resistor) varies to adjust the voltage drop V_{reg} so that the output voltage is correct.

The regulation element is *dissipative*, that is, it turns electrical energy into heat to control the output voltage. This heat is waste energy that must be removed by a heatsink and cooling air, and this adds to the size of the power supply.

The series regulator may be open circuited, in which case the dissipation in the series element is at a minimum. Under short-circuit conditions, the output current must be limited or the regulation element is likely to overheat and be damaged. The heatsink must be designed for the worst case power dissipation, which occurs at maximum input voltage and maximum output current.

The efficiency of the series regulator is given by

$$\begin{aligned}
 \eta &= \frac{P_{out}}{P_{in}} \\
 &= \frac{V_l I_l}{V_{in} I_l} \\
 &= \frac{V_l}{V_{in}} \\
 &= \frac{V_l}{V_l + V_{reg}}
 \end{aligned} \tag{604}$$

Notice that the efficiency is independent of the load current and depends simply on the ratio of load to input voltage.

Shunt Regulator

The *shunt regulator* circuit concept is shown in figure 370.

The series resistance R_s , input voltage V_i and load voltage V_l are all fixed. As a result the current I_s is a constant value, which is chosen to be equal to or greater than the maximum load current. The regulation resistance R_{reg} adjusts to shunt more or less of this current into the load, thereby to adjust the load voltage.

If the load current decreases or the input voltage increases, the regulation resistance R_{reg} decreases so that the extra current bypasses the load.

Minimum dissipation occurs in the regulation element when the load current is at a maximum; maximum dissipation when the load current is zero. If there is a possibility that the load will be disconnected, then the regulation element must dissipate a power equal to the entire load current times the input voltage.

Current through the shunt regulation element R_{reg} generates heat and does not contribute to power in the load. Consequently, for the same input and output conditions the efficiency of the shunt regulator is always worse than a series regulator. When the load power is small efficiency is a minor consideration and a shunt regulator may be appropriate. However, once the load current is more than a hundred millamps or so, a series regulator is a better choice.

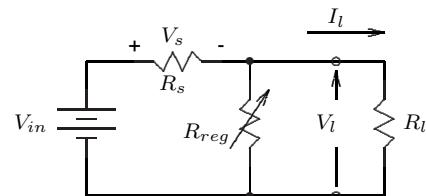


Figure 370: Shunt Regulator

Shunt Regulator Example

The zener regulator, discussed in section 7.2 and shown in figure 371, is a common example of a shunt voltage regulator. It's simple and is suitable for low-current loads where the line and load regulation are not too stringent.

The output voltage is defined by the zener voltage. The input current is chosen to be larger than the load current so that the excess flows through the zener diode, which establishes the load voltage.

Series Regulator Example

Figure 372, discussed originally in section 7.4, shows how the output of the zener regulator may be amplified by a transistor so that the regulator can supply more current. This circuit consists of a shunt regulator made up of R_s and the zener diode VR_1 . The regulated output from this circuit establishes a constant voltage at the base of the transistor Q .

The transistor is the series regulation element R_{reg} of figure 369. If the load voltage drops, then the base current of the transistor increases and this drives more current into the load.

This arrangement lowers the internal resistance of the shunt regulator circuit by an amount equal to the current gain of the transistor, and thereby improves the load regulation. However, the line regulation is unchanged from the basic zener regulator circuit and there is no current limiting. A short circuit load will probably destroy the transistor.

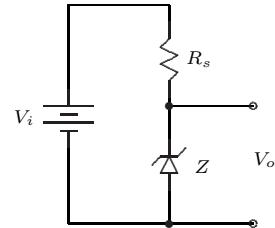


Figure 371: Zener Regulator

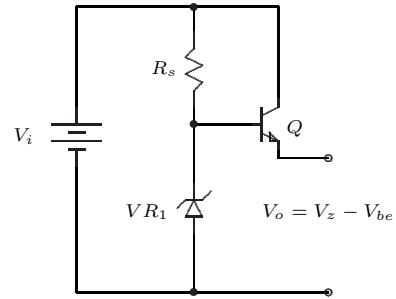


Figure 372: Amplified Zener Regulator

15.5 Feedback Voltage Regulator

At its most basic, a negative-feedback voltage regulator is simply an operational amplifier driven by a reference voltage, figure 373. In this arrangement, the regulator is a non-inverting amplifier driven by a reference voltage.

The reference voltage is shown as a zener diode driven by a constant-current source. The constant-current source insulates the zener current against changes in input voltage, thereby improving the line regulation. A typical value for the reference is 6.2 volts, since that voltage is most temperature stable. The output voltage is:

$$V_o = V_{ref} \left(1 + \frac{R_a}{R_b} \right) \quad (605)$$

The output voltage is determined by the ratio of R_a and R_b , but can never be less than the reference voltage V_{ref} .

For output voltages less than V_{ref} , the reference voltage must be reduced, possibly by deriving the input to the op-amp from some fraction of V_{ref} by means of a voltage divider.

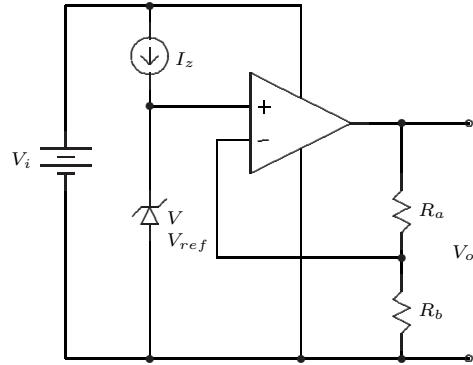


Figure 373: Op Amp Regulator

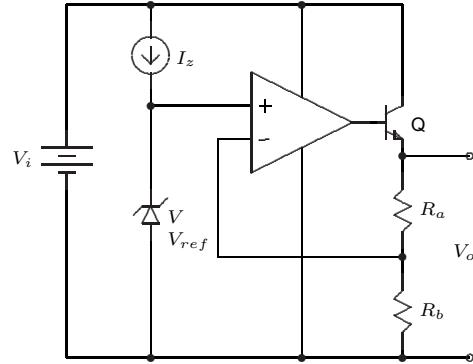


Figure 374: Regulator with Series Pass Transistor

Increasing the Output Current

An op-amp is capable of only a very small amount of current. As shown in figure 374, a *series-pass* transistor can provide current gain and larger output current. In effect, this transistor becomes part of the operational amplifier and increases the output power capability of the amplifier.

This power transistor may be required to dissipate considerable power. The power dissipation in Q is:

$$P_d = (V_{in} - V_o)I_o \quad (606)$$

The power dissipation is at a maximum for the specified output current into a short circuit ($V_o = 0$). Either the heat-sink must be designed to dissipate this power while keeping the junction of the transistor below its maximum, or some other measure must be taken to protect the transistor. One option is to sense the transistor temperature and shut off if it exceeds a limit.

The power transistor Q is frequently constructed as a Darlington Pair (section 30.1 on page 911) to increase the current gain so that a few millamps of op-amp output can control amperes of current.

The transistor may be regarded as part of the op-amp. Consequently, the base-emitter voltage drop will be compensated for by the feedback system and equation 605 for the output voltage continues to apply.

Short Circuit Current Limit

Because the internal resistance of a regulated power supply is extremely small, it is theoretically capable of a very large short-circuit current. In practice, a short circuit would cause a large output current to flow until some component overheats and is destroyed. A *current-limit* circuit sets the maximum output current and prevents damage to the supply.

One possible current-limit circuit is shown in figure 375. The circuit of figure 374 has been modified with the addition of a current sensing resistor R_{limit} and a second transistor Q_{sc} .

The current sensing resistor somewhat worsens the load regulation of the input voltage, but this will have no major effect on the regulation of the output because the negative feedback loop is unaffected. The negative feedback circuit continues to maintain the output voltage at some multiple of the reference voltage.

The load current must flow through the current-sensing resistor on its way back to the supply. The value of the current sensing resistor is chosen such that the maximum allowable output current sets up a voltage equal to the base-emitter threshold voltage of transistor Q_{sc} . When the voltage across R_{limit} reaches this value, it turns on transistor Q_{sc} , which then diverts the output

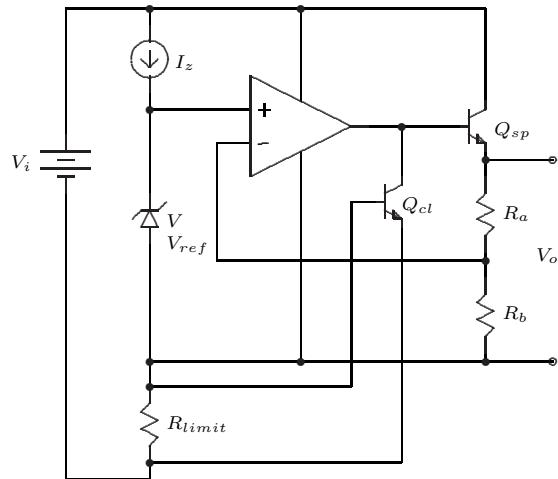


Figure 375: Regulator with Current Limit

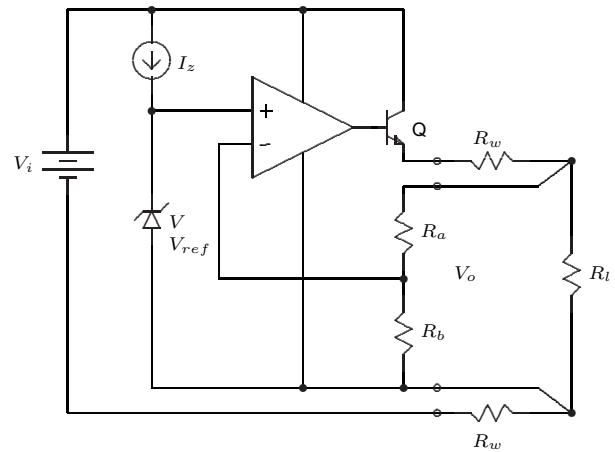


Figure 376: Regulator with Remote Sensing

current of the op-amp to ground. That disables the series pass transistor, and reduces the output voltage to zero, the desired effect. (We're counting on the fact that the output of the op-amp can be shorted to ground without ill effect, which is usually the case since it will have its own internal current limit.)

Remote Sensing

In some applications, there is a significant voltage drop in the leads between the output of the power supply and the load resistance. Then the output of the power supply may be perfectly regulated at its terminals, but the voltage at the load will decrease with increasing current.

This can be corrected by *remote sensing* the load voltage, as shown in figure 376.

The wire resistances are represented by the two resistors R_w . In this arrangement, the output voltage of the regulator will adjust itself until the voltage across the load is given exactly by equation 605. The voltage drops across the wire resistances may be regarded as a disturbance signal which is compensated for by the negative-feedback circuit.

However, if either one of the sensing wires breaks, the negative feedback system will sense that the output voltage has gone to zero and consequently set the output voltage to its maximum value. If this is potentially disastrous, some other safety measure such as an overvoltage protection circuit must be incorporated.

15.6 Three-terminal Linear Regulator

Section 15.5 introduced the negative feedback-based voltage regulator. In 1969, Robert Widlar of National Semiconductor designed the LM109, a high-power integrated circuit regulator that contains a complete voltage regulator – reference, error amplifier, series pass transistor and protection circuits. An integrated circuit regulator is a design challenge because it must be able to cope with a wide variety of loads without becoming unstable, and the voltage reference must not drift even with large changes in operating temperature.

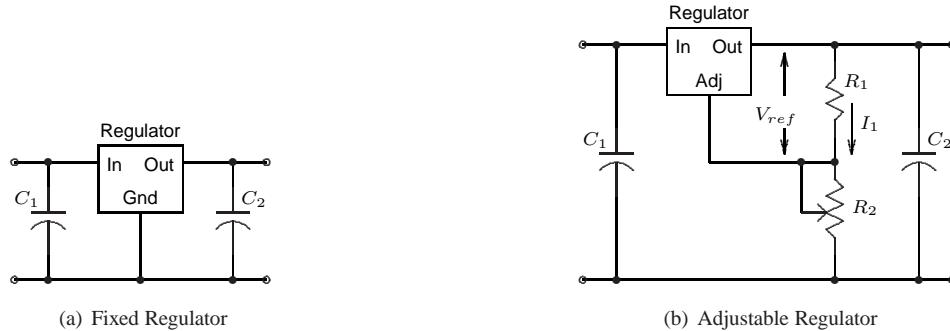


Figure 377: 3 Terminal Voltage Regulators

Voltage regulators come in several varieties:

- Positive or negative polarity
- Fixed or adjustable output voltage
- Various output voltages. In the fixed regulators, typically 5, 6, 9, 12 and 15 volts.
- Various output current capabilities, ranging from 100mA to 3 amps.
- Normal and *low dropout* versions.

Regulator, Fixed Voltage

The circuit for a fixed voltage regulator is shown in figure 377(a). A variable DC voltage supplies the input terminal, and regulated DC voltage appears at the output terminal.

For the regulator to function properly it must be provided with a sufficiently large input voltage. The minimum voltage across the regulator, known as the *dropout voltage*, is typically required to be at least 2 volts. For example, for a 5 volt regulator the input voltage must be at least 7 volts. The design of the power supply must be such that the input voltage to the regulator never drops below this minimum, even under worst case conditions: low line voltage, high output current, and high ripple across the filter capacitor.

The regulator must be on a heat sink that is large enough to keep the regulator from overheating. Typically, the junction temperature of the regulator must be kept below 100°C. The power dissipation is obtained from the application (maximum output current times maximum input-output voltage). Maximum allowable junction temperature and the thermal resistance are available from the device datasheet. Then the designer can calculate the size of the heatsink (see section 28, page 783).

Capacitor C_1 is often provided by the power supply filter capacitance. If the filter capacitance is some distance from the regulator, another capacitor is required near the regulator. Capacitor C_2 improves the output transient response but is not essential.

Adjustable Regulator

An adjustable voltage regulator is shown in figure 377(b). The regulator maintains a fixed voltage, typically 1.25 volts, between its OUT and ADJ terminals. This establishes a fixed current I_1 through resistor R_1 , typically 5mA.

The current out of the ADJ is small (typically 100 μ A) compared to the current I_1 , so the output voltage is established by the value of R_2 according to:

$$\begin{aligned} V_o &= V_{ref} + I_1 R_2 \\ &= V_{ref} \left(1 + \frac{R_2}{R_1} \right) \end{aligned} \quad (607)$$

If R_2 is adjusted to zero ohms, then the output voltage is at a minimum, which is V_{ref} .

Low Dropout Regulator: the LDO

The conventional 3-terminal regulator has an output stage similar to that shown in figure 378. Transistor Q_2 is a power transistor. Transistor Q_1 provides additional current gain, in the Darlington configuration.

In this arrangement, the base voltage of Q_2 is one V_{be} above the output voltage and the base voltage of Q_2 is another V_{be} above the output voltage. Typically, there must be about half a volt across the current source I_1 , so that the input voltage V_1 is a total of about 2 volts above the output voltage. Consequently, the input voltage must be at least 2 volts above the output voltage, and so the dropout voltage of this type of regulator is about 2 volts.

As we first saw in equation 604 on page 429, the efficiency of a series regulator can be expressed as:

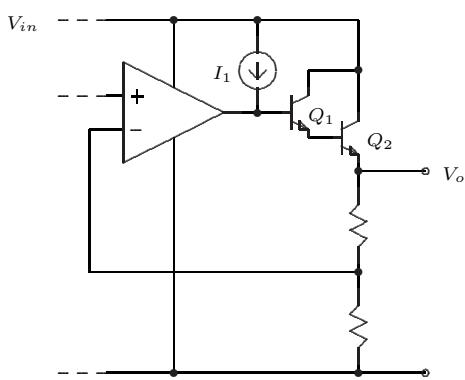


Figure 378: Regulator with NPN Output

$$\eta = \frac{V_o}{V_o + V_{reg}} \quad (608)$$

This indicates that a regulator with a low output voltage V_o and a large dropout voltage V_{reg} will have poor efficiency. In a device like a cellular phone, the supply voltage is low to minimise power dissipation, efficiency is an issue because it impacts battery life, and so the dropout voltage should be as small as possible.

An alternative configuration, the *PNP LDO Regulator* [115] [116] is shown in figure 379. The input voltage V_{in} must be larger than the output voltage by an amount equal to the saturation voltage V_{cesat} of the transistor. This is typically around 100mV at moderate currents, much less than the base-emitter voltage. Consequently, the minimum dropout voltage for this regulator is much less than the NPN output configuration.

There are two limitations to this configuration:

- The current gain of an integrated PNP transistor is much less than an NPN, and the PNP configuration is a single transistor. Consequently, the supply current to operate a PNP-LDO regulator is much larger than its NPN counterpart.
- In the NPN configuration, the load resistance is in the emitter of the series-pass transistor which inherently has a low output resistance. As a result, the nature of the load does not have much effect on the loop gain. In the PNP-LDO configuration, the load is in the collector circuit of the series-pass transistor, which is a high impedance point. A power supply load is largely capacitive which causes a phase lag in the loop phase response and can therefore destabilize the LDO. Consequently, the impedance of the load affects the regulator loop gain and stability and must be taken into account when designing the system.

There is a relatively simple cure to this problem. If the load capacitance has sufficient series resistance it will add a pole in the loop gain which stabilizes the loop. This resistance can be an inherent component in the capacitor, the *ESR, equivalent series resistance*. In some circuits, ESR is an undesirable property, but in this case it is essential. The manufacturers of the LDO integrated circuit specify the required ESR in the load capacitance¹⁵³.

15.7 The AC Adaptor

The *AC Adaptor* is a small, rectangular block that plugs into an AC wall outlet and delivers a DC voltage for powering an electronic circuit. Wall adaptors are not greatly beloved by the purchasing public, but they have some significant advantages from the standpoint of the equipment manufacturer.

- They eliminate the requirement for a line cord and a fuse.
- They are approved by regulatory agencies so no further certification of the power supply connection is required. This is a significant incentive because certification can require a significant investment of time and money, especially if the regulatory agency requests a design modification.

¹⁵³There is a potential problem in a production situation that someone will substitute a capacitor that has the correct capacitance and voltage ratings, but a different ESR. In the worst case, this would work correctly at room temperature but cause the LDO to oscillate at some extreme of temperature.

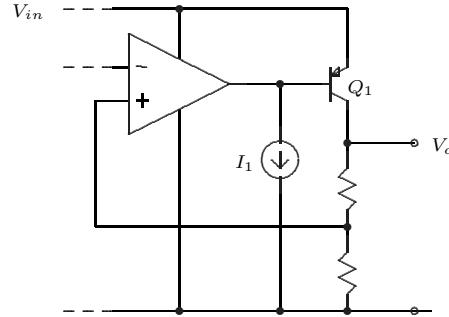


Figure 379: PNP LDO Regulator

- Because they are produced in bulk for a variety of customers, the cost is very low.

For a cost-sensitive consumer product, these are strong arguments in favour of using an AC adaptor.

There are two types of AC adaptors currently in use. Traditionally, an AC adaptor is a simple transformer-rectifier-capacitor filter, with unregulated output voltage. However, as the cost of electronics comes down, it is becoming more common for an adaptor to contain a switching power supply, complete with electronic regulation and current limit¹⁵⁴. These remarks refer to the unregulated version.

The output characteristic of a *12V, 800mA 16 Watt* adaptor is shown in figure 380. Notice that the open-circuit output voltage at 16V is considerably larger than the nameplate value of 12V. The output voltage drops significantly and is down to 10.5 volts at the nameplate current of 800mA.

This is deliberate. If the power supply is inadvertently short-circuited, then the short-circuit current is limited by the internal resistance of the supply. Presumably the approval agencies have ensured that it can survive this condition without bursting into flame. This behaviour eliminates the requirement for a fuse, but it does require consideration in the design.

At the maximum output current, the peak-peak ripple voltage is in the order of 3 volts, and the minimum output voltage, including the effect of the ripple, is about 7.5 volts. (If the value of ripple is excessive, it could be reduced by providing additional filter capacitance at the input to the electronic circuit).

It would be fair to say that this supply will indeed produce 12V and 800mA, but not both at the same time.

At the maximum output current, the output power is in the order of 8 Watts. Clearly the figure '16 Watt' does not describe the output capability of the adaptor.

For most electronic circuits, the load regulation of this power supply is too poor to use directly, and a regulator would be required. The voltage drop across a regulator is about 2 volts minimum. Consequently, this supply could be used to generate a regulated voltage of 5 volts, up to its rated current. On the other hand, a 12 volt regulated output would require that the output current be limited to less than about 130mA.

It is clear that the output characteristic of a particular AC adaptor should be measured carefully before it is designed into a product.

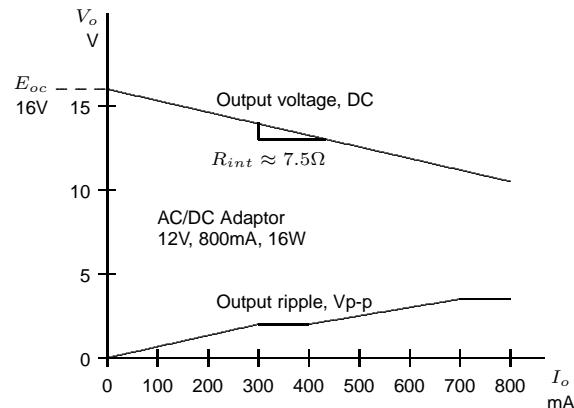


Figure 380: AC Adaptor Output

¹⁵⁴The cost of the regulator electronics is offset by material savings in the size of the transformer, which can be much smaller in the regulated unit (see figure 400). As well, a regulated unit weighs less than an unregulated unit, so the shipping cost is lower.

15.8 Switching Regulator

Background

In the 1970's, digital logic systems operated from 5 volt power and required considerable current. The usual supply configuration was a transformer, rectifier, capacitive filter and dissipative series regulator. The result was a massive power supply, typically 1.8 times the size of the electronics, with an efficiency [117] in the order of 35%. The Commodore PET computer, an early personal computer, used this type of power supply. Its competitor, the Apple II computer, used a new concept, the *switching power supply*. One had only to lift the two computers to realize the huge advantage of the switcher – the weight of the Apple II was a small fraction of the Commodore PET. The Apple II power supply was also smaller, since it was more efficient and therefore had smaller heatsinks. Since that time, switching power supplies have shrunk to about 20% of the size of the electronic systems they power, and increased in efficiency to about 90%.

Concept

Throughout the following discussion, the term *switching regulator* refers to a *step-down* or *buck* regulator. A step-down switching regulator, like the dissipative series regulator, accepts a voltage that is larger than the load voltage and adjusts it downward to a regulated value. There are other types of switching regulators, but this is the most common type and application.

The concept is shown in figure 381. Instead of using a dissipative control element to throttle the power into a load, the switching regulator uses a pair of switches. The switches operate out of phase – when S_1 is open, S_2 is closed and vice-versa. The duty cycle of the switch S_1 (ratio of on time to total period) determines the average voltage in the load. In formula form:

$$V_o = \frac{t_{on}}{T} V_i \quad (609)$$

where the variables are:

V_o	Output voltage, volts
V_i	Input voltage, volts
t_{on}	Duration of switch S_1 in the closed state, seconds
T	Period of the switching waveform, seconds

For notational compactness, we'll put

$$V_o = DV_i \quad (610)$$

where D is the *duty cycle*, equal to t_{on}/T .

Some sort of averaging device, either the characteristics of the load itself or some sort of lowpass filter, must extract the average value from the switched waveform.

This is the same technique used in the pulse width modulation technique of D/A conversion, section 25.7. However, the D/A conversion technique contends with relatively small voltages and currents. The currents and

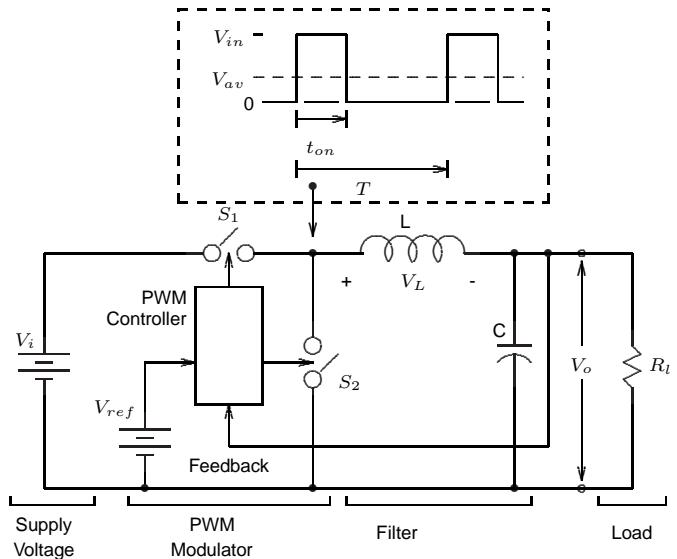


Figure 381: Switching Regulator Concept

voltages of a switching power supply are much larger, and so a different method is required to average the output signal. In figure 381 the averaging is provided by the inductor L and capacitor C .

The duty cycle is controlled by a negative feedback system that senses the output voltage, compares it to a reference value, and adjusts the duty cycle to make the reference and output voltages equal.

A Mechanical Analog

The primary energy storage device in this circuit is the inductor L_1 . Recall that an inductor is analogous to a mechanical flywheel, and the current through it the speed of the flywheel.

An early form of the internal combustion engine consisted of one cylinder connected via a crank mechanism to a flywheel. Every time the cylinder fired, it would put mechanical energy into the rotation of the flywheel. The flywheel coasted between these impulses, passing relatively constant rotational energy to some mechanical load. A mechanical negative feedback system sensed the rotational speed of the flywheel, causing the cylinder to fire whenever the speed dropped below the reference setting. Consequently, the cylinder would fire more frequently when the output was loaded. These engines are now known as 'hit and miss' engines and can be seen at some agricultural fairs¹⁵⁵.

An electrical engineer would refer to this system as a *pulse frequency* modulation scheme, in that the controller generates pulses of fixed width at a variable rate. Some switching power supplies do use this same scheme, but most switching power supplies modulate the duty cycle.

Operation

It greatly simplifies understanding an analysis of this circuit if we make the preliminary assumption that the load voltage is constant. This is approximately true if the filter capacitor is sized properly. Later, we'll verify that assumption. Now we'll develop equations for the general operation of the regulator.

The switching period T is a fixed interval chosen by the designer. For example, if the regulator switching frequency f_s is given as 20kHz, then the switching period is:

$$T = \frac{1}{f_s} \quad (611)$$

As the switching frequency increases, the size, weight and cost of the components decreases but the switching losses increase. Typical operating frequencies range between 20kHz and 1MHz.

As usual, the output power P_o is related to the input power P_i by the efficiency η :

$$P_o = \eta P_i \quad (612)$$

so

$$V_o I_o = \eta V_i I_i \quad (613)$$

For a step-down regulator of reasonable efficiency, the average input current will be less than the output current:

$$I_i = I_o \frac{V_o}{\eta V_i} \quad (614)$$

(The word *average* is important here because the input current has a pulse waveform, as we'll see in a moment.)

¹⁵⁵ http://en.wikipedia.org/wiki/Hit_and_miss_engine

Example

A switching regulator is to be designed to operate from a 28VDC aircraft supply bus and supply 2 amps at 5 volts into an electronic load. The operating frequency will be 20kHz, and the efficiency is estimated at 80%. Determine the input current and the *on* time of the switching waveform.

Solution

From equation 611, the switching waveform will have a period of

$$\begin{aligned} T &= \frac{1}{f_s} \\ &= \frac{1}{20 \times 10^3} \\ &= 50\mu\text{Sec} \end{aligned}$$

Rearranging equation 609 to solve for t_{on} , we have the *on* time:

$$\begin{aligned} t_{on} &= \frac{V_o}{V_i} T \\ &= \frac{5}{28} (50 \times 10^3) \\ &= 8.9\mu\text{Sec} \end{aligned}$$

From equation 614, the average input current will be:

$$\begin{aligned} I_i &= I_o \left(\frac{V_o}{\eta V_i} \right) \\ &= 2 \left(\frac{5}{28 \times 0.8} \right) \\ &= 0.446\text{A} \end{aligned}$$

The resultant regulator block diagram is shown in figure 382.

Switching Regulator Waveforms

Now we are in a position to examine the regulator waveforms in detail, which will lead us eventually to the component specifications.

A great many formulae can be derived for this circuit. However, rather than working with these formulae directly it is better to understand and remember the switching regulator waveforms. The relevant equations may be derived from the waveforms with KVL, KCL, and the differential equations for capacitance and inductance.

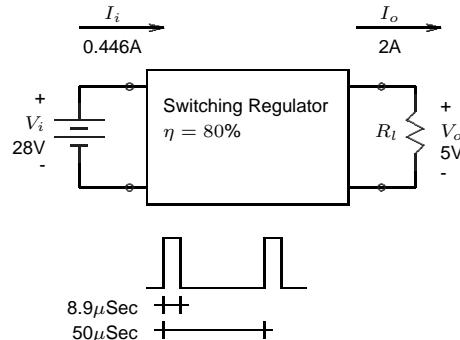


Figure 382: Switching Regulator Example

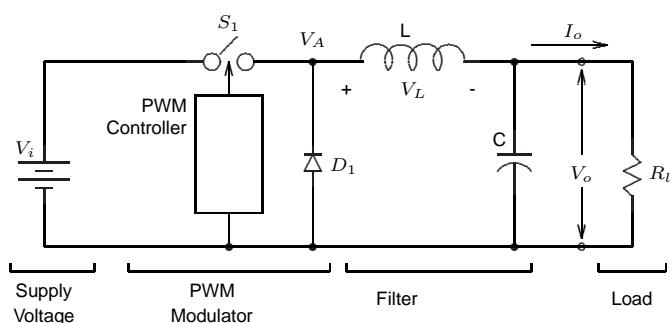


Figure 383: Switching Regulator Concept, With Diode

We will study the waveforms with reference to the schematic diagram of figure 383. This is figure 381 with switch S_2 implemented by a diode, D_1 . The waveforms are shown in figure 384.

During Phase 1, the switch S_1 is closed and this puts energy into the inductor.

- The voltage across the inductor is the difference between the input voltage and load voltage:

$$V_L = V_i - V_o \quad (615)$$

- The current in the inductor is given by

$$V_L = L \frac{di}{dt} \quad (616)$$

Consequently, during Phase 1 current ramps upward at a constant rate given by

$$\frac{\Delta I_L}{\Delta t} = \frac{V_i - V_o}{L} \quad (617)$$

During Phase 2 the switch is open and the inductor current freewheels through the inductor and diode D_1 .

- The voltage across the inductor reverses:

$$V_L = -(V_o - V_d) \quad (618)$$

where V_d is the forward voltage across the diode, about 0.6 volts for a silicon diode.

- This causes the inductor current to ramp downward at a constant rate given by:

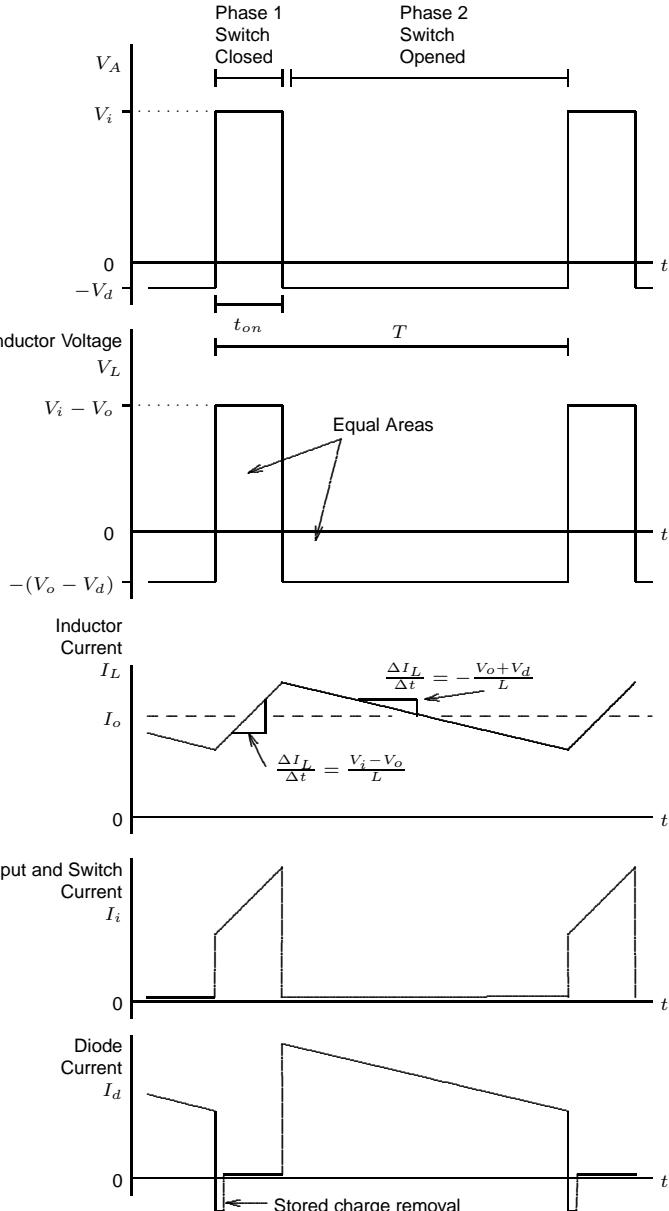


Figure 384: Switching Regulator Waveforms

$$\frac{\Delta I_L}{\Delta t} = -\frac{(V_o - V_d)}{L} \quad (619)$$

Touring the Waveforms

Now we will consider points of interest in these waveforms, and how they affect component selection.

- When the average current through an inductor is not changing, then the average voltage across the inductor must be zero. As a consequence, the inductor voltage waveforms during Phase 1 and Phase 2 must have equal areas. We can show that more thoroughly by the following derivation:

During Phase 1 the voltage across the inductor is constant. Starting with equation 616, we can write this expression for the change in inductor current:

$$\Delta I_L = \frac{1}{L}(V_i - V_o)t_{on} \quad (620)$$

Similarly, the change in current during Phase 2 is:

$$\Delta I_L = -\frac{1}{L}(V_o + V_d)(T - t_{on}) \quad (621)$$

In steady state, the upward change in inductor current during Phase 1 is exactly balanced by the downward change in Phase 2. Then the sum of the two changes in current is zero. Applying this to equations 620 and 621 and cancelling where possible, we have:

$$(V_i - V_o)t_{on} - (V_o + V_d)(T - t_{on}) = 0 \quad (622)$$

That is, the two areas under the voltage-time curve for the inductor, shown in figure 384, must be equal.

- The switch current and diode current (bottom two traces) are complementary. That is, they sum together (approximately) to make the inductor current (middle trace).
- At the instant when the switch closes, the diode has been conducting for an interval of time. When the switch closes, it reverse biases the diode and stops it from conducting. In practice, there is a certain amount of stored charge in the diode junction region, and this must be removed before the diode ceases conducting. The net effect is that the diode conducts in the reverse direction for a brief instant of time. This is indicated on the *Diode Current* trace of figure 384 as *stored charge removal*.

This stored charge removal interval creates heat in the diode. It is conducting substantial current (albeit for a brief instant) and the voltage across it is the full input voltage. Conventional power diodes, such as the ubiquitous 1N400x series, have too much stored charge to work properly in this circuit. Either a *fast recovery* or a *Schottky* diode (which has no stored charge at all) must be used.

- The input current is a pulse waveform. This can make the switching power supply into a rather inconsiderate neighbour for other equipment powered from the same source. The pulse current waveform may appear as line noise. At a minimum, this current must be supplied from a capacitor across the input terminals. It is likely that some sort of LC filter will be necessary to trap this current noise inside the switching power supply.
- Consider the inductor current waveform. The total excursion in current ΔI_L depends on the frequency of operation and the inductance, so it can be reduced by increasing the inductance or the frequency of operation.

If the average current drops below $\Delta I_L/2$, then the current waveform will run into the zero axis, that is, the inductor current drops to zero. This is referred to as *discontinuous operation*, which we'll discuss in a moment.

The Inductor

The inductor must meet the following criteria:

- The magnitude of the inductance determines the rate of change of current as energy is stored in the magnetic field of the inductor. Generally, we will choose an inductor which limits the peak swing in current to much less than the average inductor current. If this is the case, the inductor current flows continuously and the switcher is in continuous mode.
- For high efficiency, the resistance of the inductor winding must be small compared to the load resistance.
- The inductor must not reach magnetic saturation at the peak current (section 2.28).

Inductor core materials are available in two major categories: *ferrite* and *powdered iron*. Ferrite material has a high permeability and in a closed magnetic circuit (without any gaps in the path of the magnetic flux) is suitable for a transformer core. However, ferrite saturates easily. If it is to be used to construct an inductance for a switching power supply, then it must have a gap in the magnetic path to limit the magnitude of the flux in the core. The gap reduces the path reluctance and thereby reduces the flux density for a given MMF.

Powdered iron has a much lower magnetic permeability than ferrite, so it requires more winding turns for the same inductance. However, powdered iron can support a much higher flux density in an ungapped core. As a result, powdered iron is a common choice for a switching power supply inductor.

- The inductor may be wound on a variety of core shapes, ranging from a ferrite rod to a powdered iron toroid. However, a closed magnetic path (such as a toroid) is much to be preferred because it confines the magnetic field to the interior of the core. A rod inductor injects a large magnetic field into the surrounding space, and this may induce interfering currents in nearby circuitry.
- It is possible to purchase an off-the-shelf inductor, in which case the inductance, maximum current and series resistance will be specified.
- Otherwise, it is possible to use the equations of section 2.28 to design and construct a suitable inductor.

To do so:

- Peruse a catalog of core materials and guess at a suitable core size.
- Based on the maximum allowable flux density in the core (from the datasheet), the core area and the voltage across the inductor, determine the number of winding turns for the inductor. By operating the core at close to its maximum allowed flux density, it is being used as effectively as possible – ie, it is not under-utilized.
- Determine whether these windings will fit on the inductor.
- Determine the inductance and compare to the requirements of the switching regulator design.

It's usually necessary to iterate these equations several times in order to converge on a working design.

The Filter Capacitor

Now it is time to examine the capacitor more closely. First, we'll determine the capacitor current and then the capacitor voltage. These two waveforms will lead us to design parameters for the capacitor.

The capacitor current and voltage waveforms are shown in figure 385.

We can determine the current waveform by the following reasoning:

- The current through the inductor consists of two components: a DC component I_L and an AC component ΔI_L .
- The average voltage of the capacitor is not changing, so the DC current into the capacitor is zero. Consequently, the current waveform into the capacitor is symmetrical about the zero axis.
- If the capacitor is a low impedance compared to the load resistance, then most of the AC current will flow through the capacitor in preference to the resistance. Consequently, the capacitor AC current will be the same as the inductor AC current.
- By this reasoning, the capacitor current waveform shown in figure 385 is simply equal to the inductor current waveform of figure 384 with the DC component removed.

We don't yet know the magnitude of the capacitor voltage waveform, but it's useful to have some idea of the shape to guide the mathematics. We can sketch the shape of the waveform with the following reasoning:

- From the capacitor differential equation

$$i_c = C \frac{dV_c}{dt} \quad (623)$$

we have that the slope of the capacitor voltage-time curve is

$$\frac{dV_c}{dt} = \frac{1}{C} i_c \text{ volts/second} \quad (624)$$

That is, the slope of the capacitor voltage-time curve is proportional to the capacitor current.

- Referring to figure 385, we can determine that the slope of the capacitor voltage curve will be zero (horizontal) at points **b** and **d**.
- At point **a** the slope will have its maximum negative value and at **c** it will have its maximum positive value.
- The slopes that are defined at the previous points are joined by a continuous curve, since capacitor voltage cannot change instantaneously.

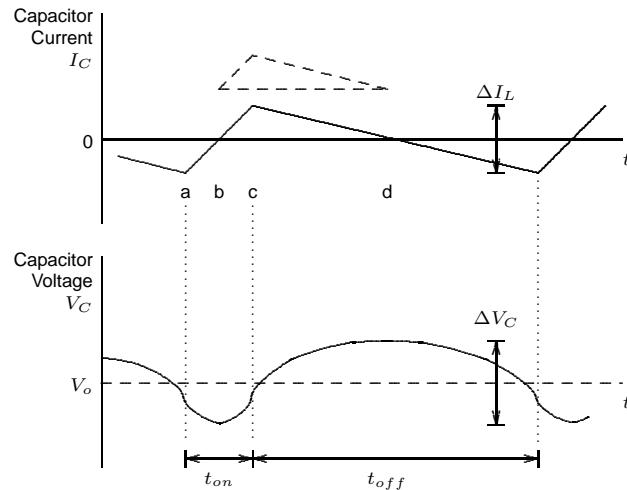


Figure 385: Capacitor Waveforms

Now we can determine the magnitude of the capacitor voltage waveform [118]. When current flows into a capacitor for an interval t_1 , the voltage on the capacitor at the end of that interval is determined by putting equation 623 into its integral form:

$$\Delta V_c = \frac{1}{C} \int_0^t i_c(t) dt \quad (625)$$

The quantity inside the integral is the area of the dashed triangle in figure 385. Recall that the area of a triangle is

$$A = \frac{1}{2}bh \quad (626)$$

where b is the base and h the height. Consequently, the peak to peak change in capacitor voltage (the peak-peak ripple) is equal to:

$$\begin{aligned} \Delta V_c &= \frac{1}{C} \times \frac{1}{2} \times \frac{\Delta I_L}{2} \times \left(\frac{t_{on}}{2} + \frac{t_{off}}{2} \right) \\ &= \frac{1}{C} \frac{\Delta I_L}{8} T \end{aligned} \quad (627)$$

The largest variation of current in the inductor occurs when the duty cycle is 0.5, so the largest possible output ripple occurs at a duty cycle of 0.5.

Reality Intrudes

It turns out in practice that the output filter capacitor must be *much* larger (by a factor of 1000 in some cases) than predicted by equation 627. To understand why, we need to look at the parasitic components of an electrolytic capacitor.

Capacitor Parasitic Components

A capacitor includes parasitic components of inductance and resistance as shown in figure 386. In many applications, these parasitic components can be neglected. However, there are circumstances where they must be taken into account.

The effect of the series inductance is discussed section 24.2. The equivalent series resistance R_{esr} is of interest in a switching power supply.

There are two common ways to specify the series resistance of a capacitor.

- Specify the tangent of the phase angle δ between the in-phase and quadrature components of capacitor impedance, as shown in figure 387.

Then

$$\tan \delta = \frac{R_{esr}}{X_c} \quad (628)$$

The value of $\tan \delta$ at some operating frequency may be specified in the capacitor datasheet, in which case the value of series resistance R_{esr} may be calculated.

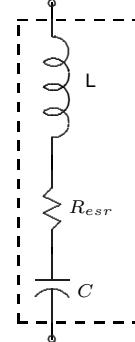


Figure 386: Capacitor Equivalent Circuit

- Alternatively, the *dissipation factor* DF may be specified on the capacitor datasheet, where

$$DF = \frac{R_{esr}}{X_c} \times 100\% \quad (629)$$

- The values of ESR and DF are often specified at 120Hz, which is the ripple frequency for a full-wave rectified power supply in North America. It's risky extrapolating these values to the operating frequency of a switching power supply, so some manufacturers give the impedance of the capacitor at a higher frequency, such as 100kHz. If the impedance of the capacitor at this frequency is known, then the ESR may be calculated from the relationship:

$$Z_c = \sqrt{R_{esr}^2 + X_c^2} \quad (630)$$

where Z_c is the total impedance of the capacitor in ohms, X_c is the impedance in ohms equal to $1/2\pi fC$, f is the test frequency (100kHz in this case) and C the value of the capacitance in farads. Then the value of R_{esr} may be calculated.

Now we are able to illustrate an with example that shows the effect of ESR on output ripple voltage.

Example

A 100kHz switching power supply that produces 5VDC output generates an inductor ripple current ΔI of 2 amperes peak-peak. Choose a suitable capacitor to reduce the output ripple to 50mV peak-peak.

Solution

The operating frequency is 100kHz, so the switching period is

$$\begin{aligned} T &= \frac{1}{f} \\ &= \frac{1}{100 \times 10^3} \\ &= 10 \times 10^{-6} \text{ sec} \end{aligned}$$

Rearranging equation 627 to solve for capacitance, we have:

$$\begin{aligned} C &= \frac{1}{\Delta V_c} \frac{\Delta I}{8} T \\ &= \frac{1}{50 \times 10^{-3}} \times \frac{2}{8} \times (10 \times 10^{-6}) \\ &= 50\mu\text{F} \end{aligned}$$

An examination of a capacitor manufacturer's catalog [119] for electrolytic capacitors turns up a $100\mu\text{F}$, 16V capacitor that is designed for operation in switching power supplies. Alas, the capacitor impedance Z_c at 100kHz is specified as 1.65Ω . At that rate, the ripple voltage is:

$$\begin{aligned} \Delta V_o &= \Delta I \times Z_c \\ &= 2 \times 1.65 \\ &= 3.3 \text{ volts peak-peak.} \end{aligned}$$

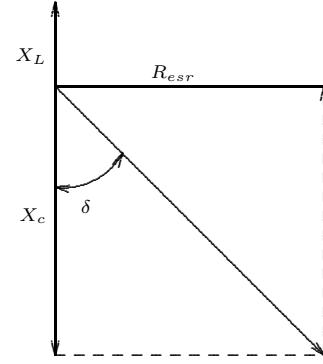


Figure 387: Vector Diagram

Of course, this is totally unsatisfactory. The ESR of the capacitor is dominating the capacitive reactance. To get the ripple voltage down to 50mV, we would require a capacitor impedance of:

$$\begin{aligned} Z_c &= \frac{\Delta V_o}{\Delta I} \\ &= \frac{0.05}{2} \\ &= 0.025\Omega \end{aligned}$$

Further study of the catalog reveals that the $4600\mu\text{F}$, 16V capacitor has an impedance of 0.055 ohms, so the required impedance could be obtained by paralleling two of these capacitors. Notice that we could reduce the size and/or number of capacitors by selecting a larger inductor, thereby reducing the peak-peak magnitude of the ripple current.

In summary

- the filtering of the output capacitor is totally dominated by the ESR of the capacitors
- a much larger filter capacitance is required than indicated by a simple calculation of capacitive reactance
- it may be necessary to reduce the peak-peak ripple current to achieve the desired output ripple voltage with a reasonable number and size of filter capacitors.

Discontinuous Operation

The waveforms for discontinuous operation are shown in figure 388. Starting with the top waveform:

- The average value of inductor current, \bar{I}_L , has decreased to a value where the inductor current decays to zero during the off interval. The effect is to create a discontinuous interval t_{disc} , when the inductor current is zero.

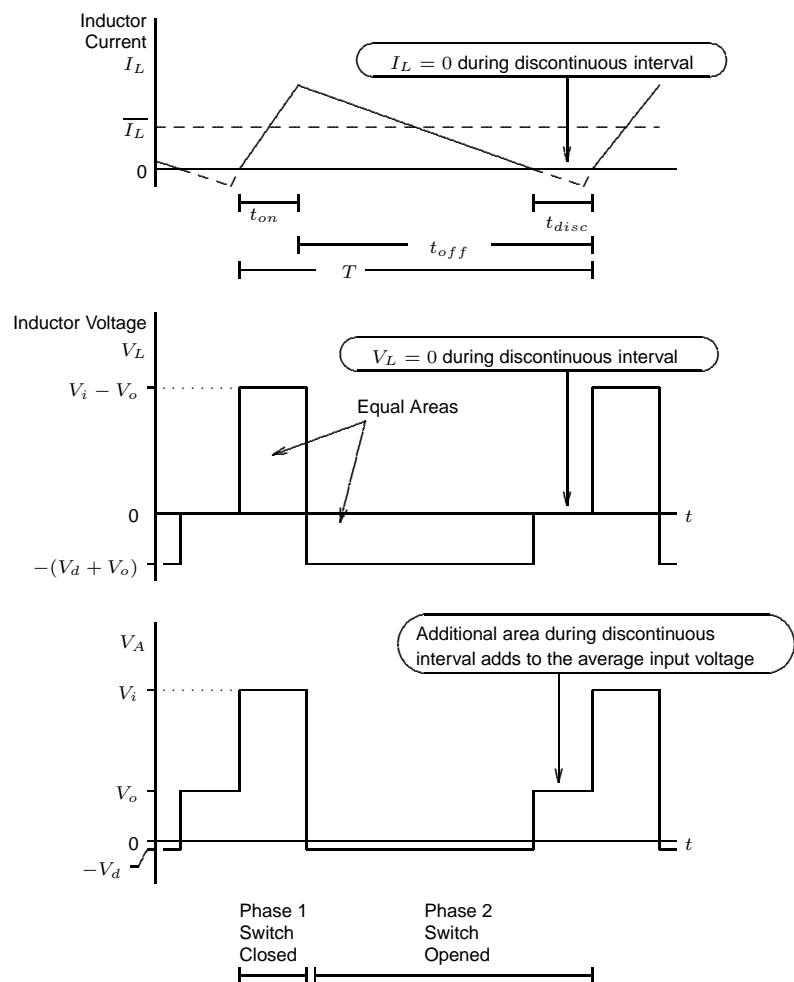


Figure 388: Switching Regulator, Discontinuous Mode Waveforms

- During the discontinuous interval there is no current through the inductor so there is no voltage across it. This is indicated in the second trace for inductor voltage.
- Since there is no voltage across the inductor during the discontinuous interval, during this interval the voltage V_A pops up to equal the output voltage as shown in the third trace.
- The average output voltage is determined by the area under the V_A waveform, divided by the period T . The added area during the discontinuous interval increases the average output voltage.

The main conclusion is this: for a constant switching duty cycle, the output voltage rises if discontinuous operation is allowed to occur. To put this in quantitative terms, consider the middle trace of inductor voltage. The average voltage across the inductor must be zero, so the two areas under the inductor voltage trace must be equal. That is:

$$(V_i - V_o)t_{on} = V_o(t_{off} - t_{disc}) \quad (631)$$

Solving for output voltage V_o , we have:

$$V_o = V_i \left(\frac{t_{on}}{t_{on} + t_{off} - t_{disc}} \right) \quad (632)$$

If the discontinuous interval is zero, then this expression becomes

$$V_o = V_i \left(\frac{t_{on}}{t_{on} + t_{off}} \right) \quad (633)$$

which we saw earlier applies to continuous mode operation.

If the discontinuous interval t_{disc} approaches the OFF interval t_{off} , and the switching duty cycle is held constant, equation 632 indicates that the output voltage approaches the input voltage [118]. This results in the behaviour shown in figure 389.

At large output current, the output voltage is simply equal to the duty cycle times the input voltage. At low output current, the output voltage rises until it is equal to the input voltage.

In practice, the switching duty cycle of a regulator of this type is not held constant but is determined by a feedback system that senses the output voltage and compares it with a reference. If the output current drops to the point where the regulator enters discontinuous mode, the negative feedback system compensates for the rise in output voltage by reducing the duty cycle. There may be a lower limit for the duty cycle, determined by the design of the controller. In that case, the regulator must be attached to a load resistance to ensure that the output current never drops below some lower limit.

Although it is important to understand what is happening, if the negative feedback system is working properly discontinuous operation does not have any effect on the output voltage. However, it is usual to design the regulator so that it operates in continuous mode under normal circumstances.

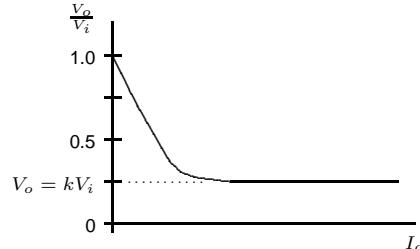


Figure 389: Output Voltage during Discontinuous Operation

Power Losses

For the following reasons, it is important to minimize power losses:

- High overall efficiency implies less heat dissipation which simplifies the cooling requirement for the supply. For example, the power losses might require a cooling fan. Fans are highly undesirable: they make noise and they are prone to failure, if for no other reason than the accumulation of dust¹⁵⁶.
- Lower power losses in the switching element result in a smaller heatsink or even, in the best of all possible worlds, no heatsink at all. This reduces size and cost.
- For battery-operated equipment, high efficiency is mandatory since it translates directly into battery life.
- Heating in the filter capacitor is a common source of failure in switching power supplies.

Switch Losses

Significant power losses occur in the switching device. For an ideal switch (infinite off resistance, zero on resistance, zero switching interval) the losses are zero.

Power loss in the switch is so-called *conduction loss*. For a BJT switch, the conduction power loss is equal to:

$$P_{switch} = V_{cesat} I_{switch} \quad (634)$$

where I_{switch} is the average current through the switch.

A MOSFET can switch at a higher frequency than a BJT, which allows smaller components in the regulator. For a MOSFET switch, the conduction power loss is in the drain resistance:

$$P_{switch} = I_{RMS}^2 R_{DS} \quad (635)$$

where I_{RMS} is the RMS current through the switch.

The operating frequency of a switching regulator is usually chosen so that the switching transition time, between ON and OFF states, is small compared to the total period and then the switching losses are small enough to be ignored. If that is not the case, then

- the switch waveforms of current and voltage are plotted
- the two waveforms are multiplied point by point to get a plot of the power dissipation vs time
- the waveform is averaged to get the steady-state power dissipation of the switch.

Capacitor Losses

The capacitor ripple current may be sufficient to create significant heating in the ESR of the filter capacitor.

For the output capacitor, the RMS current is developed in the section *Effective Value of Triangle Waveform*, page 485:

$$I_{RMS} = \frac{\Delta I}{\sqrt{12}}$$

where ΔI is the variation in current through the inductor.

¹⁵⁶ Apparently a large amount of dust enters the atmosphere from meteorites and micro-meteorites.

For the input capacitor, the RMS current is developed in the section *RMS Value of a Pulse Waveform* on page 486:

$$I_{RMS} = 0.5I_{ac}$$

where I_{ac} is the peak-peak value of the input current.

The RMS value of current creates heat in the capacitor when it flows through the equivalent series resistance R_{esr} of the capacitor.

$$P_{cap} = I_{RMS}^2 R_{esr} \quad (636)$$

where I_{RMS} is the RMS current through the capacitor.

The electrolytic capacitors that are commonly used as filter capacitors in switching power supplies. These capacitors are a common point-of-failure in switching power supplies, generally because they are overstressed by excessive current or excessive heating. For example, reference [120] indicates that the life expectancy of an electrolytic capacitor decreases from 20,000 hours at 40°C to 2000 hours at 80°C.

The calculation of power loss in a capacitor gives only a very approximate indication whether the capacitor will overheat. For example, if the power dissipation is 2 watts in a small package, it's likely to overheat. If the power dissipation is 10mW, then it's likely to run cold.

However, without definite figures for the thermal resistance from the interior of the capacitance to the outside environment it is impossible to more specific [121]. A prudent engineering approach would be to design the supply with a low-esr capacitor and then check the capacitor temperature in the prototype.

15.9 Switching Regulator, Design Example

For modest currents – up to 3 amperes or so – the design of a switching regulator has been greatly simplified with the availability of integrated circuits that contain the electronics in one package. The designer has only to choose a few additional discrete components to complete the design.

In this section, we'll go through the design of a switching regulator that uses the National Semiconductor LM2575-5 switching regulator. Similar devices are available from Linear Technology and Micrel.

We'll choose as an example design to meet the requirements for a 28 volt input, 5 volt, 1 amp output supply. The 28VDC supply is common in aircraft, and 5 volts is a common supply for digital equipment. The output ripple must be less than 50mV peak-peak.

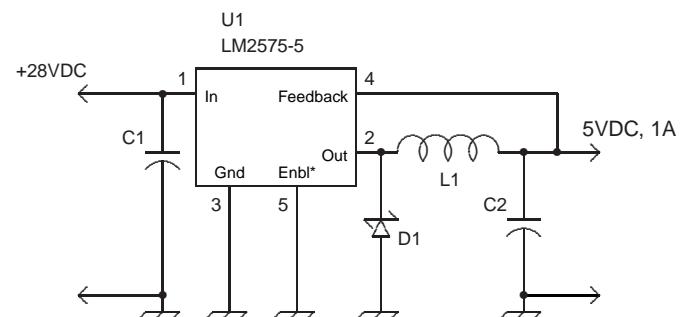


Figure 390: Switching Regulator, Basic Circuit

Rationale for a Switcher

It would be possible to construct a linear power supply to operate to these specifications but the efficiency would be terrible. To confirm this, we'll now calculate the power dissipation and efficiency of a series regulator and a switching regulator.

As we saw in section 15.4, the power dissipation in the series regulator transistor will be:

$$P_d = (V_{in} - V_l)I_L = (28 - 5)1 = 23 \text{ watts}$$

The efficiency of a series linear regulator is equal to the ratio of output power to input power. If the input and output currents are both equal to the load current, which is true if the regulator does not require significant operating current, (usually true) then the efficiency is equal to the ratio of output voltage to input voltage.

$$\eta = \frac{P_{out}}{P_{in}} = \frac{V_l I_l}{V_{in} I_l} = \frac{V_l}{V_{in}} = \frac{5}{28} = 0.17 = 17\%$$

This is very poor efficiency and the large heat dissipation in the series control transistor would require a physically large heatsink or fan cooling.

By comparison, the power dissipation in a switching regulator (assuming a BJT switch) will be approximately given by equation 634 on page 448. Assuming a saturation voltage for the switch of around 0.5 volts, then:

$$P_{losses} = V_{cesat} I_{switch} = 0.5 \times 1 = 0.5 \text{ watts}$$

Comparing the series linear regulator and the switcher power dissipations, the switcher will evidently require a much smaller heatsink. The efficiency for the switching regulator is:

$$\eta = \frac{P_{out}}{P_{in}} = \frac{P_{out}}{P_{out} + P_{losses}} = \frac{5 \times 1}{(5 \times 1) + 0.5} = 0.909 = 90.9\%$$

The efficiency of the switcher is obviously *much* higher than the series linear regulator.

The switching regulator has some compelling advantages in its favour: less heat dissipation, smaller size, higher efficiency. However, a switching regulator does generate electrical noise which can be conducted into the power supply bus and radiated in electromagnetic fields. If a switcher is used, it must be carefully shielded and filtered.

Circuit

The basic schematic diagram of a switching regulator circuit using the National Semiconductor LM2575-5 integrated circuit is shown in figure 390 on page 449. The internal block diagram is shown in figure 391.

There are versions of this device for fixed outputs of 3.3V, 5V, 12V, 15V, and there is an adjustable output version. The integrated circuit is in a package similar to the 3-lead TO-220 used for linear regulators, except that this one has 5 leads.

Referring to figure 391:

- The output voltage is detected at the Feedback terminal.
- The negative feedback system ensures that the feedback voltage V_f is equal to the 1.2 volt reference voltage V_{ref} .
- The voltage divider R_1-R_2 establishes the output voltage at:

$$V_o = V_{ref} \left(1 + \frac{R_1}{R_2} \right) \quad (637)$$

In the LM2575-5 these resistors are chosen so that the output voltage is 5 volts.

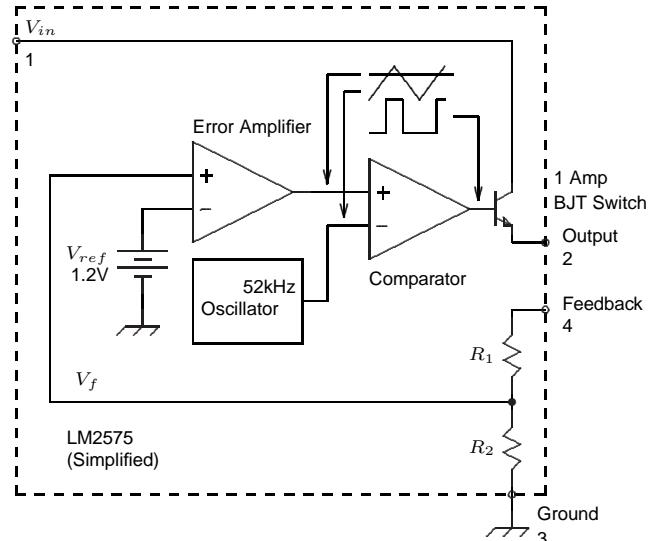


Figure 391: Switching Regulator, Internal Circuit

- The error amplifier compares the reference and feedback signals, and provides the open-loop gain for the negative feedback system.
- The 52kHz oscillator generates a triangle wave which is compared with the output of the error amplifier.
- The output of the comparator is a pulse waveform, the duty cycle of which establishes the output voltage.

Not shown in the block diagram are:

- A shutdown control signal at pin 5 that can be used to disable the output of the regulator
- Thermal shutdown circuitry that disables the output if the chip overheats
- Current limit circuitry that restricts the maximum output current, in order to protect the BJT switch.

In figure 390, the symbol for diode D_1 is for a Schottky diode, a type of rectifier diode that has no stored charge and therefore recovers quickly in applications like this one, where it must switch rapidly from the conducting to non-conducting states. To complete this design, we must choose values for the inductor and the two capacitors.

Inductor

As in most engineering designs, this one is an interaction between what is desirable and what is available. We'll proceed by getting a rough idea of what is required for the inductor and then find something suitable that is commercially available.

First, we'll get the basic switching parameters for this circuit. The datasheet for the LM2575 indicates that it switches at a fixed frequency of 52kHz. Then the switching period T is:

$$T = \frac{1}{f} = \frac{1}{52000} = 19.2\mu\text{sec}$$

The input and output voltages are respectively 28 and 5 volts. Then the duty cycle D is:

$$D = \frac{V_o}{V_i} = \frac{5}{28} = 0.178$$

Then the ON time t_{on} and OFF time t_{off} are:

$$t_{on} = DV_i = 0.178 \times 19.2\mu\text{sec} = 3.42\mu\text{sec}$$

$$t_{off} = (1 - D)V_i = (1 - 0.178) \times 19.2\mu\text{sec} = 15.8\mu\text{sec}$$

Now we can determine a value for the inductance. Ideally, we would like the peak-peak ripple current to be as small as possible because that will allow a smaller output filter capacitor for the same output ripple voltage. A smaller peak-peak ripple also makes it more likely that the regulator will be in continuous mode at light load currents.

The datasheet for the LM2575 suggests a value of $100\mu\text{H}$ for the inductor. This should be regarded as no more than a starting point for the design, and we do our own calculations to verify that this will work.

A catalog search turns up the Miller 2100 series of powdered iron toroids. The cost of the toroids is constant with increasing inductance up to a unit with $220\mu\text{H}$ inductance. Its specifications are as follows:

Part Number	Miller 2116
Inductance	$220\mu\text{H}$
Maximum DC Current	2.4 Amps
Inductance at rated current	$134\mu\text{H}$
DC resistance	0.12Ω
Outer Diameter	0.86 inches
Width	0.34 inches

Notice that two inductances are specified: the larger value, $220\mu\text{H}$, is measured at very low DC current in the core. The smaller value, $134\mu\text{H}$, is measured at rated current, 2.4 Amps. As the DC current in the core increases, the core approaches saturation and the inductance decreases. In our case, the rated current is 1 amp, so the minimum inductance is somewhere between these two values: we'll guess at $170\mu\text{H}$.

We can solve for the peak-peak variation in inductor current ΔI by using equation 617 on page 440, section 15.8 and substituting t_{on} for Δt :

$$\begin{aligned}\Delta I_L &= \left(\frac{V_i - V_o}{L} \right) \Delta t \\ &= \left(\frac{28 - 5}{170 \times 10^{-6}} \right) (3.42 \times 10^{-6}) \\ &= 0.462 \text{ amps, p-p}\end{aligned}$$

The waveforms of inductor voltage and current are shown in figure 392. The regulator will operate in continuous mode down to a minimum output current of $\Delta I/2 = 0.231$ amps.

Output Capacitor C_2

Based on the analysis in section 15.8, the main determinant of output voltage ripple is the inductor ripple current flowing through ESR of the output capacitor. Then

$$\begin{aligned}R_{esr} &= \frac{\Delta V_o}{\Delta I_L} \\ &= \frac{50 \times 10^{-3}}{0.462} \\ &= 0.108\Omega\end{aligned}$$

A search through reference [119] indicates the ESR is related more to the physical size of the capacitor than its capacitance or voltage rating. For an ESR of 0.108Ω , the smallest capacitor size is 10mm dia x 25mm length.

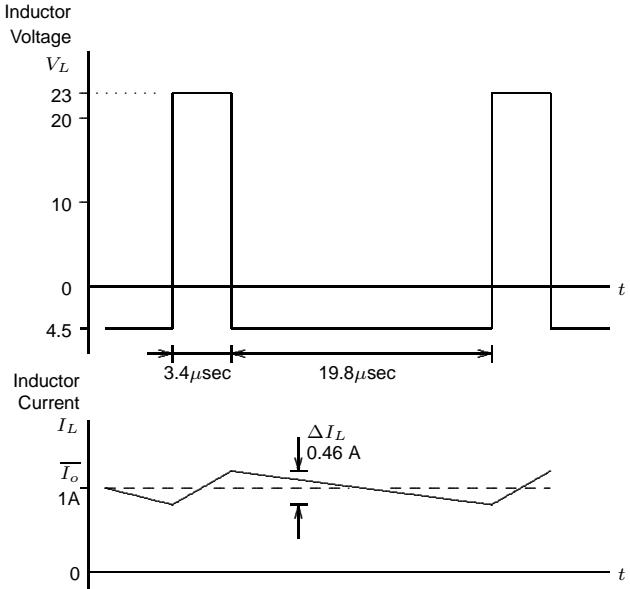


Figure 392: Inductor Voltage and Current

One possible combination of ratings for this size of capacitor is $820\mu\text{F}$, 16V. Then the specifications for the output capacitor C_2 are:

Part Number	UCC LXF16VB821M10X25LL
Capacitance	$820\mu\text{F}$
Working Voltage Max	16 V
Maximum impedance at $+20^\circ\text{C}$, 100kHz	0.052Ω
Maximum impedance at -10°C , 100kHz	0.11Ω
Outer Diameter	10mm
Width	25mm
Maximum ripple current	1.26Amps RMS at $+105^\circ\text{C}$, 100kHz

Input Capacitor C_1

Referring to the switching regulator waveforms of figure 384 on page 384 it can be seen that the input current is a pulse of duration t_{on} and amplitude approximately equal to the output current. Let's make the conservative assumption that all of this current is provided by the input capacitor during this interval.

The input capacitor will charge up to something in the order of 28VDC, with a drop in voltage during t_{on} equal to the capacitor impedance times the capacitor current. A check in the catalog shows that a capacitor of $120\mu\text{F}$ and 50VDC rating has an internal impedance of 0.28Ω . The capacitor current during t_{on} is one ampere, so the input voltage will drop from 28VDC to $28 - (1 \times 0.28) = 27.7\text{VDC}$ during this interval.

This voltage drop will communicate as noise to other equipment on the same power bus unless there is an inductive filter at the input to the switching supply. Otherwise, this is satisfactory and the specifications for the input capacitor C_1 are:

Part Number	UCC LXF50VB121M8X20LL
Capacitance	$120\mu\text{F}$
Working Voltage Max	50 V
Inductance at rated current	$134\mu\text{H}$
Maximum impedance at $+20^\circ\text{C}$, 100kHz	0.14Ω
Maximum impedance at -10°C , 100kHz	0.28Ω
Outer Diameter	8mm
Width	20mm
Maximum ripple current	610mA RMS at $+105^\circ\text{C}$, 100kHz

We should check that the RMS current rating of the capacitor (610mA) is not exceeded. From equation 684 on page 684, plugging in 1 ampere for I_{ac} and 0.178 for duty cycle D we have:

$$\begin{aligned} I_{RMS} &= I_{ac} \sqrt{D(1-D)} \\ &= 1 \sqrt{0.178(1-0.178)} \\ &= 0.382 \text{ amps RMS} \end{aligned}$$

This is well below the specified limit for this capacitor.

Freewheeling Diode D_1

The average diode current is somewhat less than the output current, 1 ampere, so a diode rated at 1 ampere continuous would be marginally acceptable in this circuit. When the BJT switch closes, the reverse voltage on the

diode is equal to the input voltage, so in this case the Peak Reverse Voltage must be greater than 28 volts. The 1N5822 is a 3 amp, 40PRV Shottky diode which is readily available at modest cost.

Part Number	1N5822
Average forward current	3 Amps
Working Voltage Max	40 V
Forward voltage drop	0.39V at 1 Amp
Package	Approx 5mm dia x 9.5 mm long

Notice that as a bonus the forward voltage drop is substantially less than a silicon diode. This is desirable since it reduces the losses in the diode.

Losses and Heatsink

Now we must determine whether the integrated circuit needs to be mounted on a heatsink. The datasheet for the LM2575 indicates that the saturation voltage of the BJT switch is about 1 volt at a current of 1 ampere. The power dissipation of the chip is given by:

$$P_d = V_{in}I_q + DI_oV_{cesat}$$

where

V_{in}	is the input voltage, 28V
I_q	is the quiescent current required by the integrated circuit, about 5mA according to the datasheet
D	is the duty cycle, 0.178 in this design
I_o	is the output current, 1 Amp
V_{cesat}	is the saturation voltage, 1 volt.

Then the power dissipation of the integrated circuit is:

$$\begin{aligned} P_d &= V_{in}I_q + DI_oV_{cesat} \\ &= 28 \times (5 \times 10^{-3}) + 0.178 \times 1 \times 1 \\ &= 0.318 \text{ watts} \end{aligned}$$

To determine the size of the heatsink, we need first to determine the necessary thermal resistance to keep the junction temperature below its maximum value. For the LM2575, the thermal parameters are:

T_{jmax}	maximum junction temperature, 125°C
θ_{jc}	thermal resistance junction to case, 2°C/W
θ_{ja}	thermal resistance junction to ambient with the package mounted horizontally, 65°C/W
T_A	maximum ambient temperature, assumed to be 40°C

Notice that the thermal resistance junction to case θ_{jc} is much lower than the thermal resistance junction to ambient θ_{ja} . For a device without a heatsink, we should use the value of θ_{ja} to determine the junction temperature. Since it is desirable not to require a heatsink, we will first calculate the junction temperature using θ_{ja} , and see if that is acceptable. Consulting section 28, we can write the heatsink equation:

$$T_j = T_A + P_d\theta_{ja}$$

$$\begin{aligned}
 &= 40 + (0.318 \times 65) \\
 &= 60.67^\circ\text{C}
 \end{aligned}$$

This is much lower than the maximum allowed junction temperature, so no heatsink is necessary.

To allow for a safety margin, the junction temperature should be kept below 100°C. If the value of junction temperature exceeded this in the previous calculation, we would have to use a heatsink. Then the thermal resistance θ_{ja} would be replaced by the sum of the thermal resistance junction to case θ_{jc} plus the thermal resistance of the heatsink θ_{sa} . Then for a maximum chosen junction temperature the thermal resistance of the heatsink can be calculated.

Once that is known, the required heatsink area can be calculated with equation 1121 on page 788.

Circuit Layout

A switching regulator circuit manages rapidly changing voltages and currents. Step changes in voltage may couple through stray capacitance to other circuits. Rapid changes in current may induce significant voltage spikes across wiring inductance. It is important to lay out the circuit diagram to minimize these effects. Figure 393 indicates some guidelines for the layout.

These are based on the following rules:

- Follow the path of high currents and provide separate traces for them.
- Minimize the area of a circuit node where there is a rapid change in voltage.
- Keep the layout as compact as possible.
- Make high-current traces heavier than signal traces.
- Separate signal leads away from leads that carry power level currents and voltages.

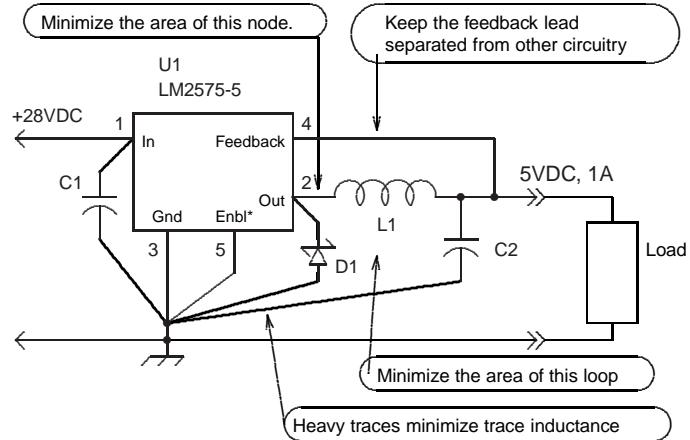


Figure 393: Switching Regulator Layout

15.10 Charge Pump

It is quite common in a circuit design to need a voltage that is somewhat larger than the supply voltage. For example:

- Most microcomputer systems work from a supply of +5 volts. The RS-232 signalling standard requires a swing of magnitude ± 5 volts minimum and ± 15 maximum. It's usual to design the driver circuit for something in the order of ± 10 volts. Consequently, the RS-232 driver circuit must have ± 10 volt power supplies.
- A device called the *varactor diode* changes capacitance based on its reverse voltage. It may be used in radios as a tuning element that can be varied electronically.

A varactor diode typically requires a maximum reverse voltage in the order of 20-30 volts, which is larger than the supply voltage for the rest of the circuit.

- A predominately digital circuit will be powered from a supply of about +5 volts. Many op-amp devices require a larger bipolar voltage, typically ± 15 volts.
- Certain EEPROM and EPROM devices are erased by an *overvoltage* which is typically +12 volts, compared to the normal operating voltage +5 volts.

All of these applications require an elevated and/or reversed supply, but at very small currents, in the order of a few tens of milliamperes maximum.

A *charge pump* is a circuit that provides an elevated or inverted voltage from an existing supply voltage [122], [123]. It does so with an arrangement of capacitors and switches, that is, without inductors.

Voltage Inverter

A charge pump voltage inverter circuit is shown in figure 394. The output voltage is equal to the negative of the input: +5 volts in produces -5 volts out. Capacitor C_1 is a so-called *flying capacitor*, that is, both ends are switched. The switches are SPDT units that are operated in unison: both are in position A and then moved to position B. They alternate in this fashion at some switching frequency.

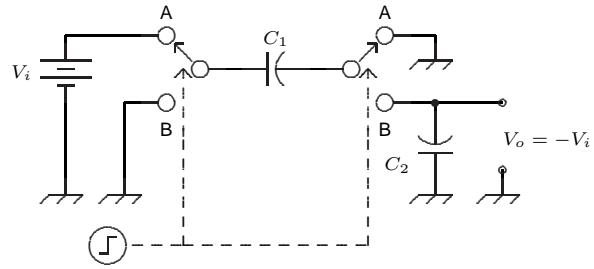


Figure 394: Charge Pump Inverter

- In position A, capacitor C_1 charges up to the supply voltage V_i with a charge $Q = C_1 V_i$ coulombs.
- Recall that capacitor voltage cannot change instantaneously. The switch moves from position A toward position B. The left plate of capacitor C_1 is now at zero volts, the voltage across C_1 is still V_i volts, so the right plate of C_1 must be at $-V_i$ volts. The capacitors are in parallel and so the charge on capacitor C_1 distributes itself across the two capacitors so that the charge on each capacitor is $Q/2$. Consequently, the voltage across each capacitor is $V_i/2$. This is left behind on capacitor C_2 when the switch goes back to the A position.
- The next time the switch moves from A to B, capacitor C_1 is again charged up to Q coulombs. When it joins C_2 , the total charge on the two capacitors is now 1.5 coulombs and the output voltage becomes $-0.75V$.
- Assuming that there is zero load current, each time the switch moves from A to B, the voltage on the output capacitor C_2 becomes more negative, asymptotically approaching $-V_i$ volts.

Charge Pump Doubler

In this circuit, the switches are caused to operate in unison between positions A and B at some repetition frequency. Then an input voltage V_i is doubled to create an output voltage of $V_o = 2V_i$. With the switches in position A, the two capacitors C_1 and C_2 are charged in parallel to V_i volts. Then, with the switches in position B, they are connected

- in series to create a total voltage of $2V_i$,
- to the output storage capacitor C_3 to store this voltage

Switch Configuration

The switches in figure 394 and 395 could be implemented by CMOS switches. For example, the 4066 device contains four SPST switches. The 4053 contains 3 SPDT switches. However, the switch resistance of these devices is significant and may require choosing a different unit.

Efficiency

The charge pump is very energy-inefficient, so it is not suitable for manipulating large amounts of power. To see why, recall that the energy stored in a capacitor is given by

$$W = \frac{1}{2}CV^2 \text{ joules} \quad (638)$$

where V is the capacitor voltage.

Now consider the scenario shown in figure 396. The capacitors are assumed equal in capacitance. In figure 396(a), capacitor C_1 is charged to voltage V_i , so its stored charge is

$$Q = CV_i \text{ coulombs} \quad (639)$$

and the total energy stored in the two capacitors is

$$\begin{aligned} W_t &= \frac{1}{2}C_1V_i^2 + \frac{1}{2}C_2V_i^2 \\ &= \frac{C_1V_i^2}{2} \text{ joules} \end{aligned} \quad (640)$$

Now consider that the switch is moved to position B. Capacitor C_1 will discharge into capacitor C_2 until they both have the same voltage. Since they are the same capacitance, the charge will distribute as $Q/2$ coulombs in each capacitor. Consequently, each capacitor voltage will be $V_i/2$ volts, and the new total energy W'_t will be:

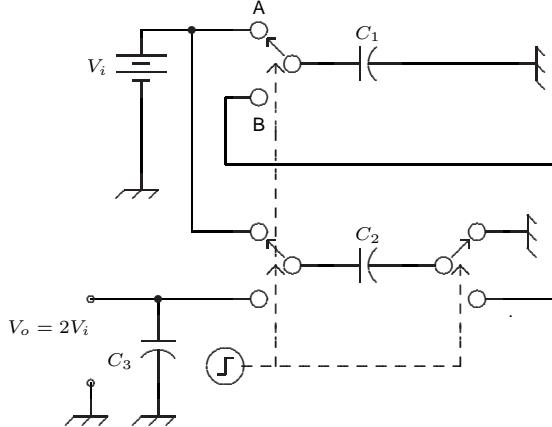


Figure 395: Charge Pump Doubler

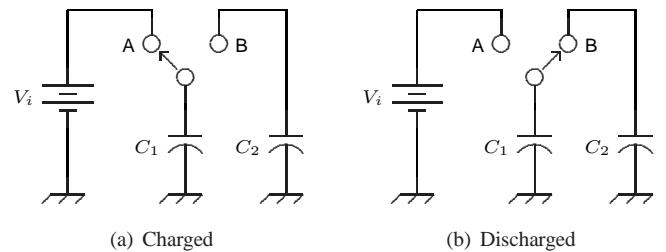


Figure 396: Charge Pump Efficiency

$$\begin{aligned} W'_t &= \frac{1}{2}C_1\left(\frac{V_i}{2}\right)^2 + \frac{1}{2}C_2\left(\frac{V_i}{2}\right)^2 \\ &= \frac{W_t}{2} \text{ joules} \end{aligned} \quad (641)$$

Evidently half the input energy has been lost, making the system 50% efficient. (In practice, the lost energy would be dissipated in the wiring resistance. It's interesting to speculate where the lost energy goes when the components are ideal. In a radiated electromagnetic pulse? If the components are ideal, then the charge is transferred as an infinite amplitude, zero duration impulse.).

In a real charge pump application, the difference in voltages between the flying capacitor and the output capacitor is much smaller, and so the losses are smaller. For example, reference [122] shows efficiencies in the range of 65 and 85% at an output current of 10mA.

Design Issues

The required capacitance of the output capacitor depends on the switching frequency, output current, and allowable ripple, according to

$$I \approx C \frac{\Delta V}{\Delta t} \quad (642)$$

where

- I is the load current in amps
- C is the capacitance
- ΔV is the peak-peak value of the output ripple
- Δt is the time of a half-period of the switching frequency

Consider for example the charge-pump inverter of figure 394. There are two intervals:

- when the capacitors are separated, C_1 is being charged by the power supply and C_2 is supplying the load on its own.
- when the capacitors are paralleled, and both capacitors are supplying the load.

The discharge curve will have two distinct slopes, the second one half the first, over the total time interval of one cycle of the switching frequency. Then, knowing the ripple voltage ΔV , switching time ΔT and load current I , equation 642 may be used to determine the size of the capacitance.

We've been assuming that the switches have zero resistance, which is definitely not the case for CMOS switches. Typical values can range into the hundreds of ohms. The time constant formed by a capacitance and the switch resistance must be much less than the charging interval.

15.11 Voltage Multiplier

The *voltage multiplier* circuit is a type of charge pump circuit that uses diodes rather than switches and is used to generate a voltage some multiple larger than the power supply voltage. Like the charge pump, the voltage multiplier is only capable of very small output currents because it has a high internal resistance.

To introduce the concept, consider the voltage doubler circuit in figure 397. Diode D_1 and capacitor C_1 form a diode clamp circuit as formerly seen in section 6.7. The clamp circuit shifts the input AC waveform upwards so that its maximum voltage is equal to twice the peak value of the original sine wave.

The second diode-capacitor pair, D_2 and C_2 , capture the peak value of this waveform in a storage capacitor so that the output is a DC value, with some ripple which depends on the load current.

This configuration can be extended over a number of stages. Figure 398 shows a 6 stage multiplier. (The component numbering system is consistent with reference [124]).

Components C_4 , C_5 , D_4 and D_5 have the same configuration as the voltage doubler in figure 397. Then capacitor C_3 and diode D_3 act as a clamp for the AC signal, so that the minimum value of the AC waveform is clamped to $2E$ volts. The maximum value of this waveform, $4E$, is captured by capacitor C_2 and diode D_2 . Similarly, capacitor C_2 and diode D_2 pump the voltage up to a maximum value of $6E$. Capacitor C_2 and diode D_2 capture the final output voltage.

Bugler [124] shows that there is little advantage in making the multiplying capacitors (all the capacitors except C_0) different sizes, so it is reasonable to assume that they are all the same value. Then Bugler shows that the output voltage is given by

$$V_o \approx NE - \frac{N^3}{12Cf} I_o \quad (643)$$

where

- V_o is the output voltage, volts
- N is the voltage multiplication factor
- E is the peak value of the AC input voltage, volts
- C is the multiplying capacitance, farads
- I_o is the load current, amps
- f is the frequency of the input AC signal, Hz

This equation indicates that the open circuit output voltage is simply NE volts. There is also a voltage drop

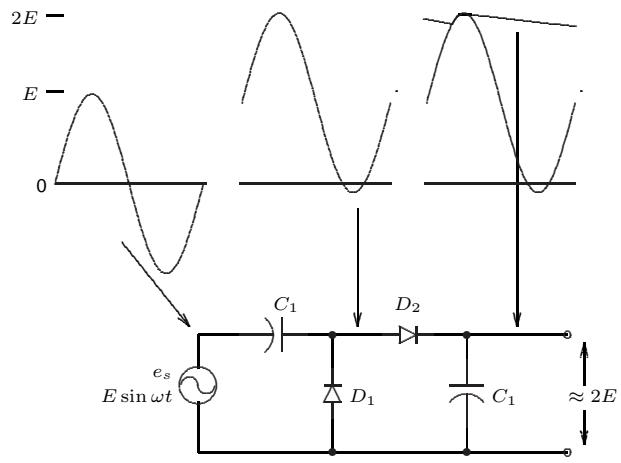


Figure 397: Voltage Doubler

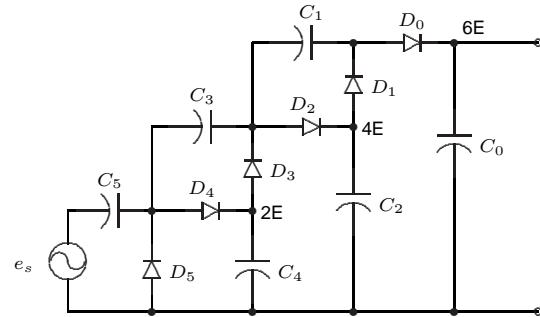


Figure 398: Voltage Multiplier ($N=6$)

across the internal impedance which is the product of the output current and a factor proportional to N cubed. As a consequence, the output impedance rises very rapidly with the number of multiplication stages, and this severely limits the available output current.

As well, since capacitance and frequency are located in the basement of the expression for internal impedance, increasing these quantities will reduce the internal impedance.

Bugler shows that for 10% ripple in the output waveform the value of the filter capacitance C_0 is:

$$C_o \geq C \frac{60}{N^4} \quad (644)$$

The Alternative Schematic

The voltage multiplier of figure 398 can be redrawn as shown in figure 399. This is the conventional way of drawing a voltage multiplier circuit.

The schematic has a pleasing symmetry, but it's more difficult to understand how it functions. Understanding is often significantly assisted, as it is in this case, by a redrawing a circuit schematic.

Applications

The voltage multiplier finds application where high voltages must be generated at very low currents, as in ionization generators and acceleration voltage for a cathode ray tube. Where extremely high voltages are to be generated, it is usual to step up the AC line voltage with a high-voltage transformer and then use the output of the transformer to drive a many-stage voltage multiplier. In this manner, it is possible to generate some tens of thousands of volts at a small current.

15.12 Transformer Design

Figure 362 on page 418 illustrates the application of a power transformer. In that circuit, the power transformer provides isolation from the 60Hz AC line and adjusts the voltage level into the power supply from 117VAC to 12VAC.

If the transformer operating frequency can be increased, it is possible to reduce the size, weight and cost of a transformer. Figure 400(a) shows a typical 60 Hz, 12 volt power transformer. Figure 400(b) shows a transformer which has similar specifications but operates at 100kHz.

The 60Hz transformer of figure 400(a) requires specialist materials and skills to construct so it is usually purchased off the shelf or obtained as a custom design from a transformer manufacturer. In contrast, the 100kHz transformer of figure 400(b) is easy to wind by hand to suit a particular application¹⁵⁷.

Figure 401 shows block diagrams for low and high frequency power supplies. The low-frequency supply in figure 401(a) accepts 60Hz power directly from the AC line. A 60Hz transformer isolates the output from the AC and steps the voltage up or down.

The high-frequency supply in figure 401(b) generates DC directly from the AC line or uses an existing supply of DC voltage. This DC supplies a high frequency oscillator. As in the low-frequency design – but at a much higher operating frequency – the transformer provides isolation and steps the voltage up or down as required. The output of the transformer is rectified by one of the diode configurations shown in section 15.3 on page 421. Garden-variety rectifier diodes (such as the 1N400x series) do not switch quickly enough to operate at frequencies

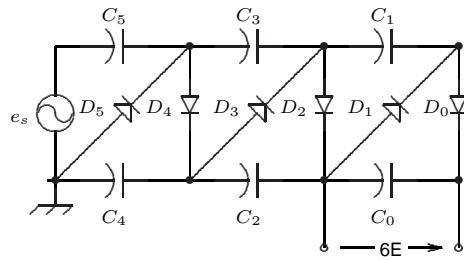


Figure 399: Voltage Multiplier Redrawn

¹⁵⁷The transformer shown in figure 400(b) was wound by a student for a lab project.



(a) 60Hz Transformer

(b) 100kHz Transformer

Figure 400: Power Transformers

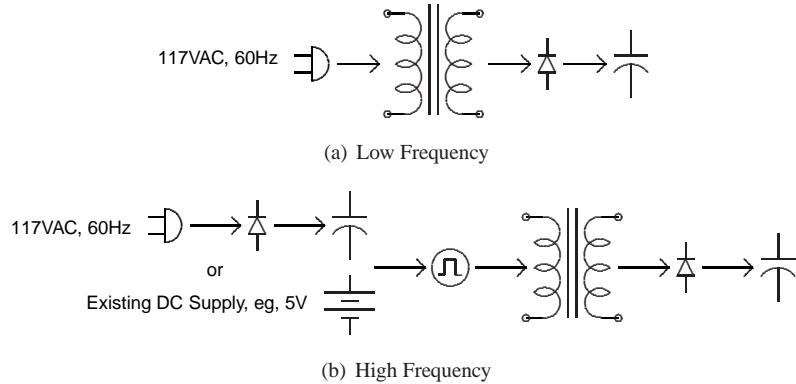


Figure 401: Power Supply Block Diagrams

higher than 60Hz. The rectifier diodes must be fast-recovery or Schottky types. The size of the transformer and output filter capacitor scale inversely with operating frequency, so the high-frequency units are much smaller than the 60Hz versions.

Operating Frequency

The operating frequency of the supply is a compromise. A higher operating frequency reduces the size of the transformer and filter capacitor. However, the switching losses of the oscillator driver transistors increase with frequency. As a result, the maximum operating frequency is limited by the switching speed of the driver transistors and rectifier diodes. Noise is also a consideration. Switching power supplies generally operate above 20kHz to avoid creating acoustical noise that will disturb humans. Fast switches allow operating at higher frequencies, but fast edges also generate high-frequency radio interference.

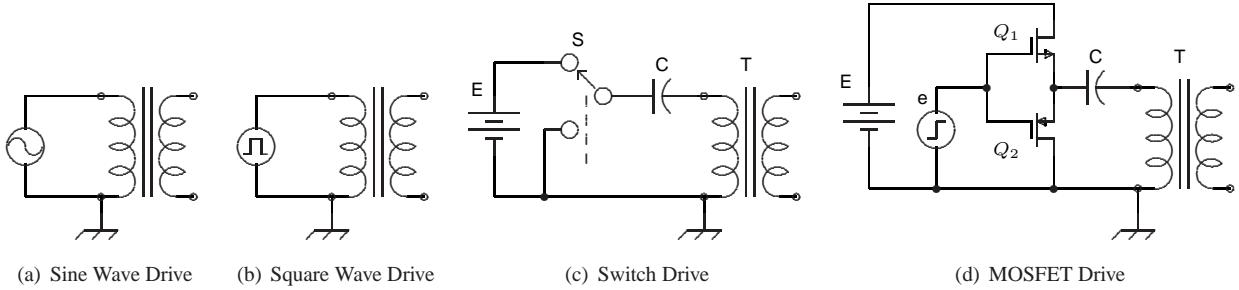


Figure 402: Simple Transformer Drive

Drive Methods

Single Ended Drive

The designer can choose from a variety of transformer drive circuits. The *single-ended drive* configuration is shown in figure 402(a), where the transformer is driven from a sine wave source. When implemented as a transistor circuit, sine wave drive is not efficient – typically some 70% of the input power appears in the transformer, the rest is dissipated as heat in the driver circuit.

It is more efficient to drive the transformer from a square wave source, as shown in figure 402(b). Then the driver transistors can operate as switches, either fully ON or OFF, with minimal dissipation¹⁵⁸.

When the square wave is produced from a single supply, it may be represented as shown in figure 402(c): voltage supply E and SPDT switch S . This square wave has an average value of half the supply voltage. Applying DC to a transformer winding causes a large DC current to flow, since the current is only limited by the winding resistance. This current may cause magnetic saturation in the transformer. To prevent this, figure 402(c) is shown with blocking capacitor C in series with the transformer primary. This capacitor removes the DC component in the primary drive voltage waveform. It must appear as a low impedance at the switching frequency.

Notice that the peak value of the drive waveform is $E/2$ volts.

In figure 402(d) a complementary pair of MOSFETs is configured as driver of the primary winding. The MOSFET drive waveform e must be sufficient to ensure that the MOSFETs operate in switching mode, ie, turn completely ON and OFF.

Bridged Drive

If the two ends of the transformer primary winding are driven in opposite directions by some voltage e , then the voltage across the winding is equal to $2e$. This is a useful technique for creating an AC voltage that exceeds the magnitude of the supply voltage. It is known as a *bridged* drive, and is a common technique in audio amplifiers to increase the power in a loudspeaker.

Figure 403(a) shows the basic circuit using two sine waves, one inverted with respect to the other¹⁵⁹.

Figure 403(b) shows the bridged drive with square waves, in a similar manner to the sine wave drive.

Figure 403(c) shows how a DPDT switch could be used to generate a reversing voltage across the transformer primary winding. Since the average value of the drive voltage is zero – if the waveforms have exactly 50% duty

¹⁵⁸Notwithstanding the lower efficiency, a sine wave drive is sometimes chosen because it generates less electromagnetic interference than the fast edges of a square wave. Alternatively, the rise and fall time of the square wave can be increased to reduce interference.

¹⁵⁹Colloquially, these sine waves are often referred to as '180 degrees out of phase'. It is only true for certain shapes of waveforms that the inverted version is the same shape as the phase-shifted version. A more general description is to describe the waveform as inverted.

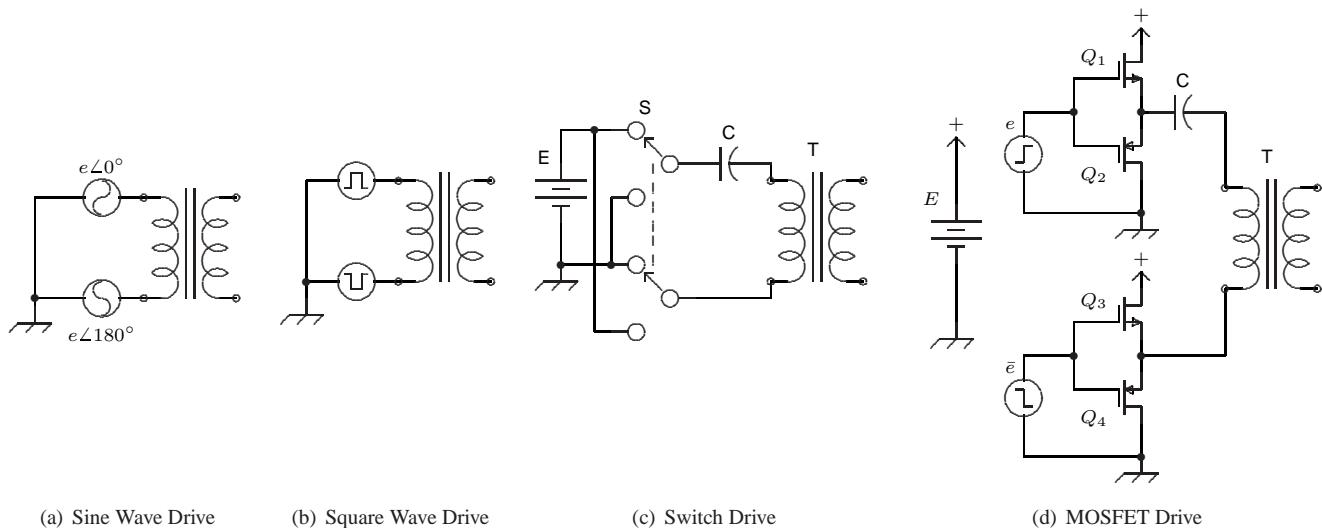


Figure 403: Bridged Transformer Drive

cycle – the blocking capacitor C is not strictly necessary. However, it's a useful precaution against a non-zero average input¹⁶⁰.

Figure 403(d) is an extension of Figure 403(b), showing how MOSFET transistors can be used to increase the drive level to the transformer. For the first half cycle, current flows from the power supply through Q_1 , down through the transformer primary, and then through Q_4 to ground. During the second half cycle, current flows from the power supply through Q_3 , upward through the transformer primary, and then through Q_2 to ground.

Again, capacitor C is not strictly necessary, but is a useful precaution against transformer saturation.

Push-Pull Drive

The *push-pull* drive system is shown in figure 404. Compared to bridged drive, push-pull uses simpler electronics and a more complex transformer. The primary winding requires a centre tap and the voltage developed across each half of the primary is E volts, half that of the bridged drive.

Figure 404(a) shows the basic circuit. The primary winding centre-tap is connected to the positive terminal of the power supply. Switches S1 and S2 alternate connecting the extreme ends of the primary to the negative of the power supply. Then NPN transistors or N-Channel FETs may be used as the switches, as shown in figure 404(c).

Alternatively, the primary centre-tap may be connected to the negative terminal of the power supply as shown in figure 404(b), and the switches implemented with PNP or P-Channel devices, figure 404(d).

In the bridged drive circuit of figure 403, the driving transistors must withstand the supply voltage E when they are not conducting. In the push-pull circuit of figure 404, each driving switch must be able to withstand *twice* the supply voltage E . For example, if the supply is 24 volts, then the drive transistors must have a breakdown voltage of somewhat more than 48 volts. To see why this is so, consider circuit 404(a). Consider that switch S1 is closed and switch S1 open. Then the upper half of the primary winding is connected to the power supply. As a consequence, in the transformer core magnetic flux is ramping upward at some rate $d\phi/dt$.

¹⁶⁰Historical footnote: In the days of tube circuits, car radios required a high voltage supply. This was provided by stepping up the 6 volt automobile supply in a circuit similar to figure 403(c). The switch S was driven by an electromechanical switch buzzer called a *vibrator*. Like tubes, vibrators failed periodically and had to be replaced, so they were a plug-in device. An additional set of contacts could function as an electromechanical rectifier in the secondary circuit.

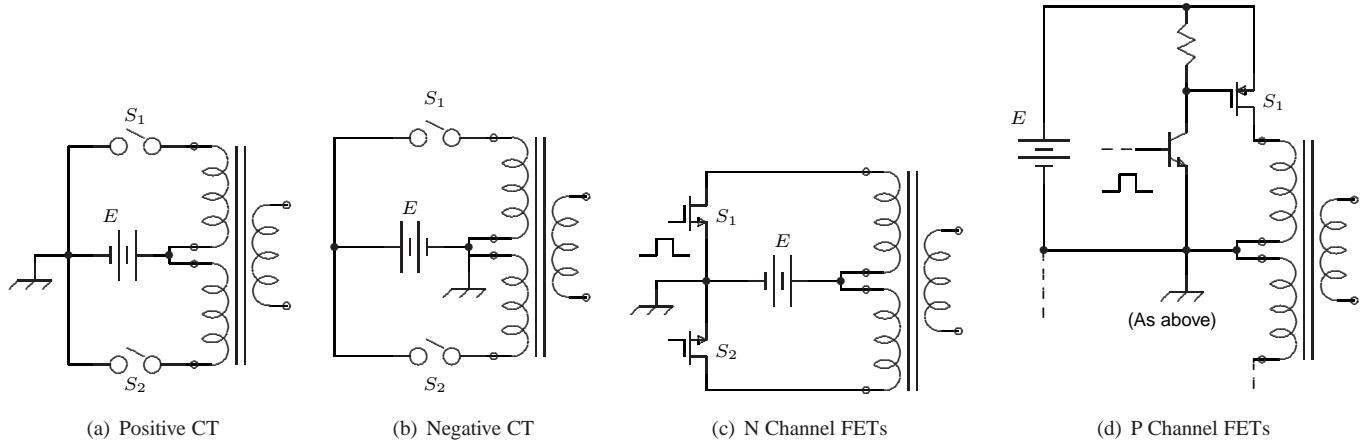


Figure 404: Push-Pull Transformer Drive

The lower half of the primary winding is wound on the same core, so the changing flux in the core induces a voltage E in that winding. The polarity of this voltage is such that it adds to the supply voltage at the centre tap. As a result, the voltage at the lower terminal of the primary winding is equal to $2E$ volts. This is not a particular consideration for small values of supply voltage. However, in a supply that operates directly from rectified line voltage the value of E is about 200 volts, and so the drive transistors must have a breakdown voltage of double that, 400 volts.

15.13 Transformer Design Parameters

Core Saturation

Consider that the primary winding of a power transformer is connected to a DC voltage, as shown in figure 405. When the switch is closed, according to Lenz's law the flux in the core of the transformer will ramp upwards. This changing flux induces a voltage in the core of the transformer and creates a *back emf* in the primary that balances the supply voltage. If the magnetizing inductance is large, which is the usual case, then the current in the primary winding is small enough to be ignored.

As shown in the graph of flux versus time, there is an upper limit to the flux density in the transformer core. At some point, all the magnetic domains are aligned and no further increase in flux can occur. The core is said to be *saturated*.

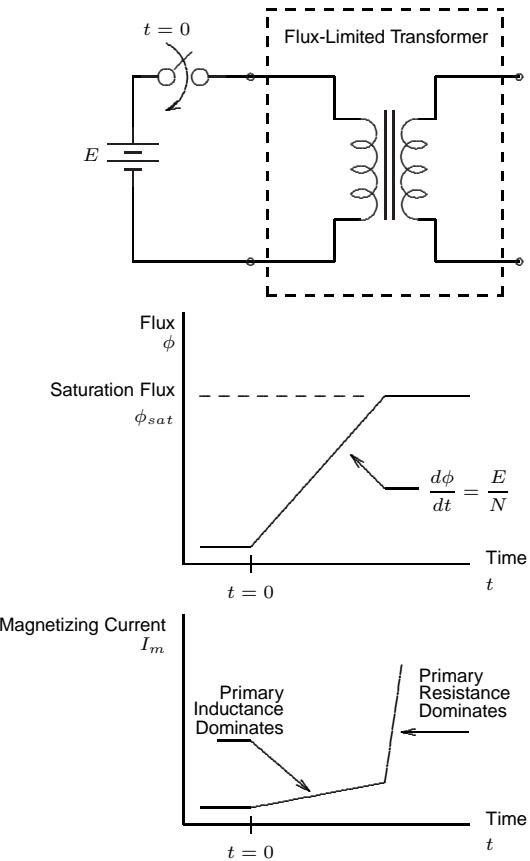


Figure 405: Transformer Saturation

At that point, the back emf disappears and the primary current is limited only by the primary winding resistance. As shown in the graph of magnetizing current versus time, the primary current rockets upward. In effect, at saturation the transformer ceases to function and so it must be configured so that the total flux never exceeds saturation.

Putting it another way, up to the point of saturation, the rate of rise of primary current is controlled mainly by the *magnetizing inductance* of the primary winding, which is much larger than the resistance of the primary winding. After saturation, the magnetizing inductance disappears and the current is limited only by the resistance of the primary winding.

The rate of increase of the flux in the core decreases with increasing primary turns, and the maximum excursion of the flux decreases if the time available decreases, which is the effect of increasing operating frequency. The allowable total flux increases with increasing core area. These factors must be juggled to ensure the flux density remains below saturation.

A good starting point is to assume a particular core size and then, on the basis of operating frequency and maximum allowed flux density, calculate the necessary number of turns.

Core Materials

Three types of core material are used in inductive devices:

Steel This is the traditional core material for low-frequency (60Hz) transformers. It's inexpensive, can be formed into a wide variety of shapes, and has the largest value of saturation flux density, in the range 8 to 24kGauss [125]. The maximum frequency of operation is limited to a few tens of kHz.

Powdered Iron Powdered iron and ferrite trade off permeability for saturation flux density.

Powdered iron has a relatively low magnetic permeability and a large saturation flux density, and is useful at frequencies into the 100's of kHz. Its main application is in switching regulators as an inductor that stores energy while conducting a DC current (section 15.8). As a result, saturation flux density is a critical parameter.

Ferrite A ferrite core has high permeability, a low saturation flux density and can be used at high frequencies. As a consequence of the low saturation flux density, it's not a useful material for energy storage as required for an inductor¹⁶¹.

A transformer system can be designed so that the windings conduct only AC current. Then the peak flux density can be controlled by the number of winding turns, and the relatively low value of the maximum saturation flux density is not a serious limitation.

Consequently, ferrite is the preferred material for high-frequency transformers.

Core Shape

The transformer core must create a closed magnetic path for the flux, and the transformer windings must encircle the magnetic flux. One of the simplest possible core shapes is the *toroid*¹⁶², shown in figure 406.

There is a major advantage to the toroid: theoretically and in practice there is little magnetic flux leakage from the core. As a result, the transformer is unlikely to create magnetic interference in nearby conductors. (The effect is reversible: the transformer is equally unlikely to respond to external magnetic fields.)

Some non-toroidal transformers are wound in two stages. First, the windings are spooled onto a bobbin. Then a two-part core is assembled through the bobbin. This two-step process facilitates automatic manufacturing.

¹⁶¹However, ferrite can be used as the basis for an energy storage device if the magnetic path includes an air gap. Then the ferrite guides the magnetic field into the air gap and the energy storage takes place in the gap. See for example [126].

¹⁶²Doughnut or donut shape, depending on your resident country.

A toroidal transformer core, however, cannot be disassembled, so the winding must be fed repeatedly through the opening in the toroid. This process can be automated, but it's more complex and expensive than a two-stage technique.

Alternatively, a high-frequency power transformer will often have very few winding turns and in that case it is quite feasible to wind a toroid by hand.

Operating Frequency

There is negligible dissipation in the drive transistors when they are fully ON or OFF. Most dissipation occurs when the drive circuitry is making the transition from one state to the other. Consequently, a lower switching frequency results in lower power dissipation.

On the other hand, a higher switching frequency reduces the amount of core material in the transformer and the number of winding turns, which determine the cost and size of the transformer.

Low-power switching circuits of this type generally operate at frequencies between 20kHz and 1MHz. Higher frequencies, above 1MHz, are limited by switching losses and the maximum frequency capability of magnetic material in the transformer core.

A useful approach is to operate the transformer at the highest frequency that will not cause excessive dissipation in the drive circuits or exceed the limitations of the core material.

For example, in the case study of section 15.14, it turns out that the core material losses become excessive at frequencies greater than 100kHz. The drivers transition in the order of tens of nanoseconds, a small fraction of the period of a 100kHz waveform, so driver dissipation should not be an issue at that frequency. Consequently, a drive frequency of 100kHz is a suitable choice.

The core losses – which increase with magnetic field and frequency – cause power dissipation in the core material. This core loss power and the $i^2 R$ losses in the windings will raise the internal temperature of the transformer. There are sophisticated simulation tools for predicting the internal temperature of a transformer. The usual approach is to ensure that the internal power dissipation is not excessive, and to verify this by temperature measurement of the transformer prototype.

Wiring Gauge

The currents in the windings and the allowable voltage drop determine the allowable winding resistance. Then the resistance and length of wire give the required wire size.

The designer then needs to verify that the required number of turns of that wire gauge will fit in the chosen core space. As well, the wire size must be sufficient for the current density.

If the windings do not fit, then the designer must choose another core size and iterate the design toward a solution.

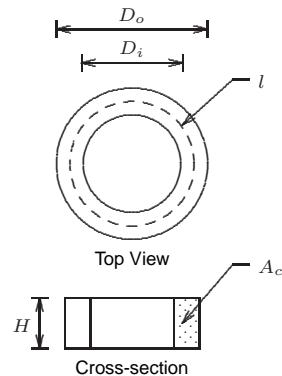


Figure 406: Toroid Core

15.14 Application: Transformer Isolated Interface

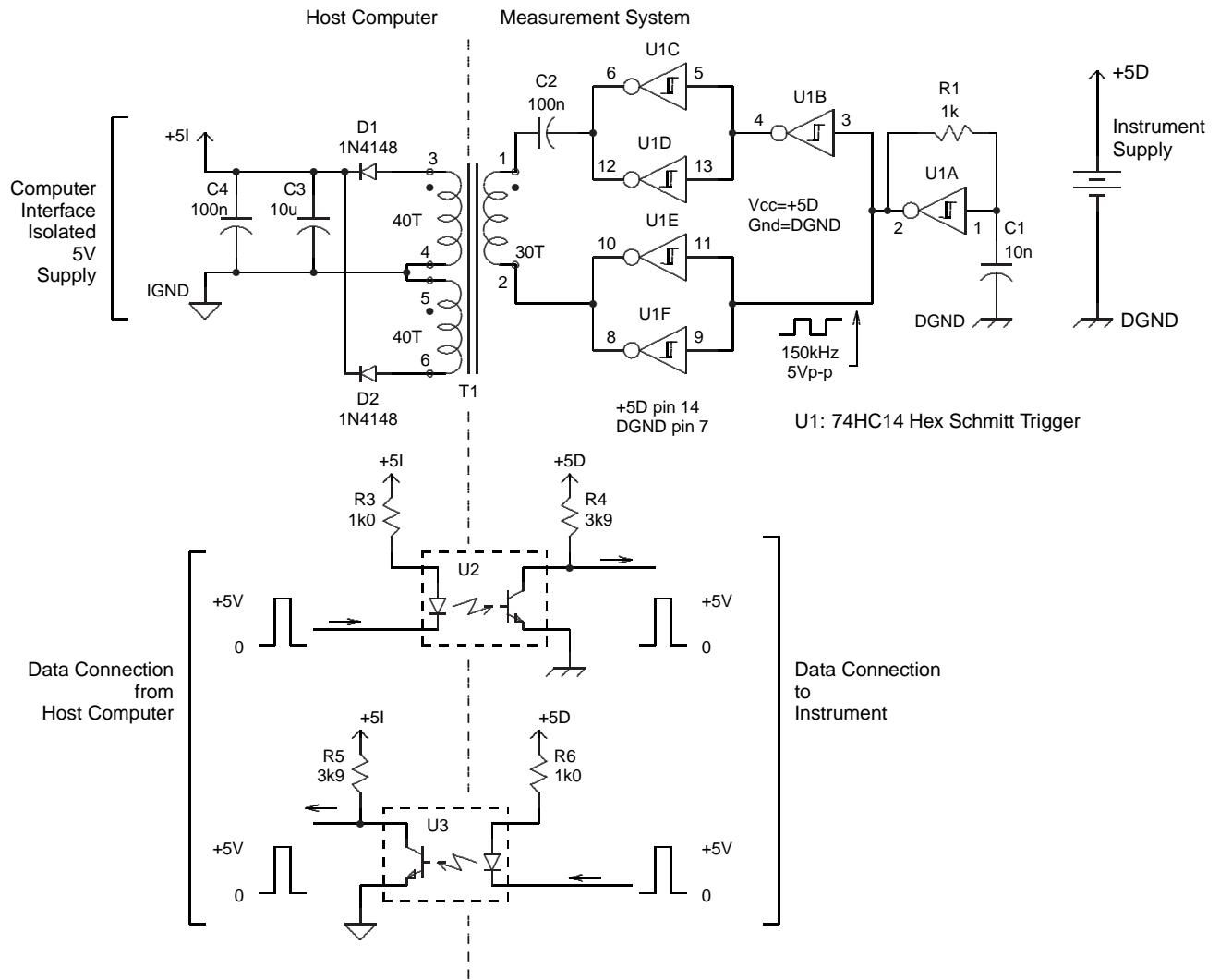


Figure 407: Transformer Isolated Interface

Now we will look at an example, the isolated power supply of figure 407. This application illustrates how a small transformer might be useful in an electronic circuit.

Rationale for the Design

At the left of the diagram is a computer, which exchanges data with some instrument, at the right of the diagram. If the instrument were to be connected directly to the computer, there is some possibility that a malfunction or electrical overload in the instrument could damage the computer. To prevent that, the instrument is electrically isolated from the computer – they do not share power supply or ground connections. Opto-isolators transfer the

digital signals from one side to the other.

Isolating the instrument from the computer has a further advantage – it helps isolate digital noise on the computer ground from analog measuring circuitry in the instrument.

This circuitry is referred to as the *interface* between the computer and instrument.

In this situation, as in many interfaces, the computer does not provide a +5 volt power supply connection to operate its side of the interface – the computer side of the two opto-isolators. The instrument has a source of power so the necessary power for the computer side must be supplied in an isolated fashion from the instrument side. Notice the two ground symbols in the diagram: DGND for the instrument and IGND for the isolated side of the interface – indicating that the grounds are isolated. Similarly, there are two isolated 5 volt supplies: +5D for the instrument and +5I for the isolated side of the interface.

Now we will examine the isolated power supply, and then the transformer itself.

Isolated Power Supply

The isolated power supply has three components: an oscillator, a transformer and a rectifier.

The oscillator is a Schmitt Trigger Astable, U1A (see section 20.6) which uses one of the 6 Schmitt Trigger Inverters in the 74HC14 package. Resistor R1 and capacitor C1 set the operating frequency. Section U1B inverts the oscillator waveform so that in-phase and inverted-phase waveforms are available to drive the transformer. The remaining sections of the IC are used as buffers to drive the transformer primary winding. Their outputs swing in anti-phase, each over a range of 0V to +5V. The net effect is to drive the transformer primary with a 10 volt p-p square wave. This arrangement – driving both ends of the winding in antiphase – is the *bridged* drive described in section 15.12 on page 462.

(A simpler method would be to ground one end of the primary winding and drive the other with a 5 volt square wave. But this would halve the primary drive voltage. A larger primary voltage reduces the required number of turns on the secondary.)

A review of the 74HC14 data sheet indicates that each inverter can supply in the order of 25mA with an internal voltage drop of 0.5 volts – assuming that is acceptable, then the pair can supply 50mA into the primary of the transformer.

Capacitor C2 is important – it prevents any DC voltage appearing across the primary winding. A DC voltage might occur as the result of a non 50% duty cycle in the drive waveform. A DC voltage will cause the core flux to increase and possibly saturate the core.

The secondary rectifier circuit is a full-wave centre-tapped arrangement. A bridge circuit, with a simple untapped secondary, could have been used instead. A full-wave rectified square wave theoretically needs no filtering at all. Capacitors C3 and C4 filter noise and transients that appear in the secondary supply.

15.15 Example: Transformer Design

Now we will go through the steps to design a small power transformer similar to T_1 in figure 407. This is a common application, where some input DC voltage is to be converted to another DC voltage with isolation between the two supplies. In this case, the input voltage is 12VDC and the output 18VDC.

It should be understood that the following sequence of steps illustrates a design. However, like many design illustrations, it implies that the design process is a linear sequence of steps. In fact, especially in transformer design, the steps must be iterated toward a successful design. As well, the design process can be greatly simplified by reference to previous designs. Then certain parameters can be chosen on the basis of experience. Engineers must rely extensively on experience in order to complete designs in a reasonable period of time. However, experience can be misleading and a competent engineer will periodically check her assumptions for validity.

Step 1: Design Requirements

Input Voltage	12VDC
Output Voltage	18VDC
Operating Frequency	100kHz Square Wave
Output Current	1 amp DC

Step 2: Core Specifications

As usual in electronic design, the choice of core is driven by what is available at reasonable price. Steel is out because it only works at low frequencies. Powdered iron is out because its low value of magnetic permeability would result in too low a value of magnetizing inductance. That leaves ferrite.

There are two parts to the core specification: the type of ferrite material, and the dimensions of the core.

The transformer is to be driven by a 100kHz square wave, so the chosen core material must have low losses at that frequency. A supplier of ferrite cores¹⁶³ suggests using the F9C ferrite core material, since it is readily available and has reasonably low losses at 100kHz.

Section 15.18 on page 478 develops a method of estimating core size based on the transformer power rating. From an analysis of this type (and other factors such as availability and cost), the core is chosen to be the MMG-Neosid type 28-519-C36, an epoxy-coated toroidal core using the F9C core material. The epoxy coating is important – it helps to prevent the partially conductive ferrite from causing inter-winding short circuits if the enamel coating on a winding is damaged. This choice of core may need to be modified based on the results of the design, but it's a starting point.

The specifications¹⁶⁴ for the F9C core material are as follows:

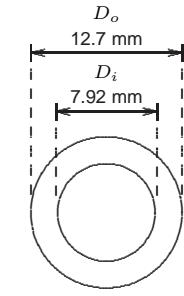
Parameter	Symbol	Value	Units
Permeability	μ_r	5000	
Saturation Flux	B_{sat}	350	mT (1 T = 1weber/m ²)
Power Loss	PLD	300	mW/Cubic Centimetre

The dimensions (figure 408) are:

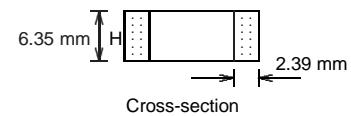
Outer Dia, mm	12.7
Inner Dia, mm	7.92
Height, mm	6.35
Path Length L_e , mm	31.22
Core Area, A_e , mm ²	14.9
Core Volume V_c , mm ³	465

These figures must be converted to the MKS system for use in the standard formulae.

With the requirements and a candidate core in hand, we can now design the transformer.



Top View



Cross-section

Step 3: Saturation

In this step, we determine the number of winding turns necessary to keep the core from saturating.

Figure 408: Toroid Core Dimensions

¹⁶³Neosid, Toronto, Canada

¹⁶⁴These are adapted from the datasheet for the F9C core material. The saturation flux density is given on the datasheet as 460mT but an examination of the hysteresis curve (figure 58 on page 104) shows that this is very optimistic and a safer value is much lower: hence 350mT.

From the design requirements, 12 volts is applied across the primary winding. The flux ramps upward at a constant rate. In the time interval equal to half the driving cycle, the core flux must be less than saturation.

The relationship of equation 645 links applied voltage E , flux ϕ , number of turns N and time interval:

$$\begin{aligned} E &= N \frac{d\Phi}{dt} \\ &= N \frac{\Delta\Phi}{\Delta T} \end{aligned} \quad (645)$$

where:

- E voltage across the primary winding, volts
- N number of turns on the primary winding
- Φ flux, webers
- t, T time, seconds.

Since the rate of rise of flux is a linear function of time, $d\Phi/dt$ can be replaced by $\Delta\Phi/\Delta T$. The core flux ramps by $\Delta\Phi$ from zero to its saturation value Φ_{sat} in Δt seconds. The operating frequency is 100kHz, so Δt is 5 μ Sec.

The core saturation flux Φ_{sat} webers is related to the saturation flux density B_{sat} by the core area A_e , which must be converted to square metres. ($1 \text{ mm}^2 = 10^{-6} \text{ metres}^2$)

$$\begin{aligned} \Phi_{sat} &= B_{sat} A_e \\ &= (350 \times 10^{-3}) \times (14.9 \times 10^{-6}) \\ &= 5.22 \times 10^{-6} \text{ webers} \end{aligned}$$

Rearrange equation 645 to solve for the number of turns and substitute numerical values:

$$\begin{aligned} N &= E \frac{\Delta T}{\Delta\Phi} \\ &= 12 \left[\frac{5 \times 10^{-6}}{5.22 \times 10^{-6}} \right] \\ &= 11.5 \text{ turns} \end{aligned}$$

This is the *minimum* number of turns on the primary winding to keep the core from saturating. The number of turns on the primary also determines the *magnetizing current*, which we'll investigate in the next step.

Step 4: Magnetizing Current

Consider that a transformer is operated from an AC source, figure 409. If the transformer is ideal and the secondary winding is open circuited, then its primary current will be zero. In practice, an AC current does flow, and this effect can be modelled by a magnetizing

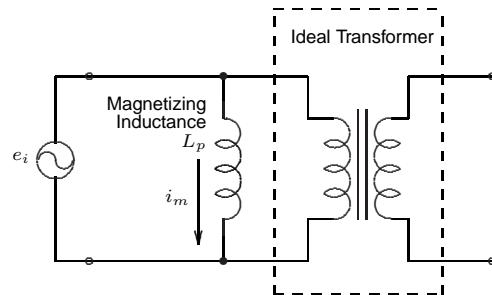


Figure 409: Magnetizing Inductance

inductance L_p in parallel with the primary winding of the transformer. The size of the magnetizing inductance determines the magnetizing current.

In this case, the primary voltage is a square wave. Then the magnetizing current is a triangle waveform, ramping upward when the primary voltage is positive and downward when the voltage is reversed.

For efficiency, the magnetizing current i_m should be much smaller than the load current. Somewhat arbitrarily, we'll choose the magnetizing current as 2% of the output current, or 20mA RMS. The RMS value of a triangle waveform is given by equation 676 in section 15.20:

$$I_{RMS} = \frac{\Delta I}{\sqrt{12}} \quad (646)$$

where ΔI is the peak-peak amplitude of the triangle waveform of magnetizing current. Substitute 20mA for I_{RMS} in equation 646 and solve for ΔI :

$$\begin{aligned} \Delta I &= I_{RMS} \times \sqrt{12} \\ &= 0.020 \times 3.46 \\ &= 70 \text{ mA, peak-peak} \end{aligned}$$

Now we must determine the number of turns on the primary to obtain this magnetizing current. Equation 647 leads to a value for the required magnetizing inductance, relating the applied voltage, inductance and rate of rise of current:

$$E = L \frac{di}{dt} \quad (647)$$

In this particular case, the relevant values are:

- E 12V voltage across the primary winding
- L L_m magnetizing inductance, henries
- di ΔI change in current, 70mA in one-half oscillation period
- dt ΔT time of one-half the oscillation period, 5μSec

Substitute these values in equation 647 and solve for the magnetizing inductance:

$$\begin{aligned} L_m &= E \frac{\Delta T}{\Delta I} \\ &= 12 \frac{5 \times 10^{-6}}{70 \times 10^{-3}} \\ &= 857 \mu\text{H} \end{aligned}$$

Now we can determine the number of turns required to create this inductance, using equation 98 from section 2.28:

$$L = \frac{N^2 \mu A}{l} \quad (648)$$

The permeability μ is the product of the core material *relative permeability* μ_r and the permeability of free space μ_o , which is $4\pi \times 10^{-7}$. Then equation 648 becomes:

$$L = \frac{N^2 \mu_r \mu_o A}{l} \quad (649)$$

The values for this calculation are:

L	L_m	magnetizing inductance, $857\mu\text{H}$
N	N_p	primary winding, turns
μ_r		core material relative permeability, 5000
μ_o		permeability of free space, $4\pi \times 10^{-7}$, Henries per metre
A	A_e	core cross sectional area in metres ²
l	L_e	core magnetic path length in metres, 31.22×10^{-3}

Rearrange equation 649 to solve for N and substitute values:

$$\begin{aligned} N_p &= \sqrt{\frac{L_m L_e}{\mu_r \mu_o A_e}} \\ &= \sqrt{\frac{(857 \times 10^{-6}) \times (31.22 \times 10^{-3})}{5000 \times (4\pi \times 10^{-7}) \times (14.9 \times 10^{-6})}} \\ &= 16.9 \text{ turns} \end{aligned}$$

To prevent the core from saturating requires a primary winding of 11.5 turns (page 470). To meet the magnetizing current requirement requires 16.9 turns. The appropriate strategy is to choose the largest of these two, in the order of 17 winding turns on the primary.

Step 5: Turns Ratio

The secondary voltage is to be 18 Volts. From the transformer equation 254 on page 186:

$$\frac{e_p}{e_s} = \frac{N_p}{N_s} \quad (650)$$

That is, the voltage is stepped up or down from primary to secondary in the same ratio as the number of turns. Consequently, the number of turns on the secondary will be:

$$\begin{aligned} N_s &= N_p \frac{e_s}{e_p} \\ &= 17 \left[\frac{18}{12} \right] \\ &= 25.5 \text{ turns} \end{aligned}$$

We'd round this up to 26 full turns.

Step 6: Winding Resistance

For the purposes of this design, we'll assume that #26AWG has been chosen as a starting point for the transformer windings. This allows us to calculate the winding resistances, which lead to an electrical model of the transformer.

The height of the core is 6.35mm, the thickness 2.39mm (figure 408 on page 469). Then each winding turn has a length of:

$$\begin{aligned} L_{turn} &= (2 \times 6.35) + (2 \times 2.39) \\ &= 17.48 \text{ mm} \end{aligned}$$

According to the wire table (section 15.19, page 482), the resistance of #26AWG is $40.82\Omega/1000\text{ft}$. ($1000\text{ft} = 1\text{kft}$). Converting¹⁶⁵ this to ohms per millimeter, we have:

$$\begin{aligned} r_w &= \left[\frac{40.82 \text{ Ohms}}{1 \text{ kft}} \right] \times \left[\frac{1}{1000} \frac{\text{kft}}{\text{ft}} \right] \times \left[\frac{1}{12} \frac{\text{ft}}{\text{inch}} \right] \times \left[\frac{1}{25.4} \frac{\text{inch}}{\text{mm}} \right] \\ &= 133 \times 10^{-6} \text{ Ohms/mm} \end{aligned}$$

Then the resistance of the primary winding is the product of the number of turns, length per turn, and resistance per length:

$$\begin{aligned} R_p &= 17 \text{ turns} \times \left[17.48 \frac{\text{mm}}{\text{turn}} \right] \times \left[133 \times 10^{-6} \frac{\text{Ohms}}{\text{mm}} \right] \\ &= 0.039 \Omega \end{aligned}$$

Similarly, the resistance of the secondary is $R_s = 0.059 \Omega$.

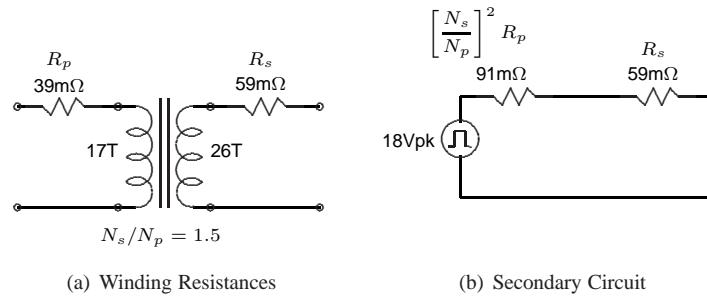


Figure 410: Transformer Equivalent Circuit

The equivalent circuit for the transformer, showing the winding resistances, is shown in figure 410(a).

From the viewpoint of the secondary winding, the primary resistance appears multiplied by the square of the turns ratio (page 487), as shown in figure 410(b). As a consequence of the winding resistance, each ampere current in the secondary winding will create a voltage drop of $114 + 75 = 189 \text{ mV}$. It may be necessary to increase the open-circuit voltage of the transformer secondary to allow for this voltage drop.

Core Winding Area

We must check that the windings will fit on the toroid core.

The available core window area (the hole in the toroid donut) is determined from the core opening diameter (figure 408):

$$\begin{aligned} A_o &= \pi R^2 \\ &= 3.14 \times \left(\frac{7.92}{2} \right)^2 \\ &= 49.2 \text{ mm}^2 \end{aligned}$$

¹⁶⁵Notice how including the units in this calculation ensures that the conversion factors are correct.

Some of this area is wasted. A typical value for *winding utilization factor* in toroids, the fraction of the total opening that is usable by windings, is 0.2 (page 481). Then the available area for windings A'_o is

$$\begin{aligned} A'_o &= 0.2 \times 49.2 \\ &= 9.84 \text{ mm}^2 \end{aligned}$$

The total area occupied by the windings is approximately equal to the total number of turns (primary and secondary) times the cross-sectional area of #26AWG wire¹⁶⁶. According to the wire table (section 15.19), the wire diameter is 0.4049mm. Then the area occupied by $17 + 26 = 43$ turns is:

$$\begin{aligned} A_w &= 43 \times (\pi R^2) \\ &= 43 \times 3.14 \times \left(\frac{0.4049}{2}\right)^2 \\ &= 5.53 \text{ mm}^2 \end{aligned}$$

Since the wire area is less than the available area, the windings should fit in the core opening¹⁶⁷.

15.16 Summary of Power Supply Design

The design of a power supply can be as much of a challenge as any other part of an electronic system. A designer must be aware of the various possible circuit elements (linear and switching regulator), and the cost and availability of those elements in the marketplace. The designer must then trade off issues such as cost, size and thermal management to arrive at a reliable and economic design.

Traditional Power Supply Design

For a traditional, line-operated, regulated power supply like that of figure 362, it's possible to map out a straightforward design process.

Identify the load voltage and current requirements. This will require estimating or measuring the current consumption of the *client* circuits.

Decide whether to make or buy the power supply. Regulated power supplies are available at very reasonable cost, for a wide variety of output voltages and currents. For a one-off or small production run which can be satisfied by a commercially available unit, it may make more sense to buy an off-the-shelf supply. This allows the engineers to concentrate on other aspects of the design.

If the cost is critical or there are special requirements – a particular shape that must fit around existing circuitry, or unusual voltages – then the supply must be designed from scratch.

Choose a regulator. Assuming that the supply voltage must be regulated, it is necessary to decide between a linear regulator and a switcher. Linears are less expensive and do not generate electrical noise. They may require a cooling system that adds to the cost and size. Switchers are a little more expensive but can fit into a smaller space. In general, a linear regulator is a better choice at low currents (say, less than 100mA) and switchers are preferable for larger output current.

¹⁶⁶The utilization factor quoted earlier accounts for the difference between round and square areas.

¹⁶⁷This result should be regarded as a guide and may need to be refined, based on the actual packing density of the wire.

Determine the minimum required input voltage. Use the regulator dropout specification to determine the minimum input voltage.

Design the raw DC supply. This could be a line-operated transformer-rectifier-capacitor system (section 15.3, page 421).

Or it could be a DC wall adaptor (section 15.7, page 435), which minimizes parts cost and eliminates the need for certification.

Generally speaking, a line-operated transformer-rectifier-capacitor system will be required for currents in excess of an amp or so.

Choose the input voltage. There is an interaction between the voltage supplied by the transformer or wall adaptor, and the size of the filter capacitor.

A larger capacitor results in lower ripple, which allows a lower supply voltage. It's necessary to iterate this calculation, trying various available transformers or adaptors, and various filter capacitors.

The key equation is 602 on page 422. The calculation must be done for worst-case input low voltage, that is, assuming that the voltage is 10% lower than its nominal value. Put another way, the nominal input voltage must be at least 10% higher than the minimum.

For a commercial design, it's quite possible to have a transformer custom designed to meet a particular requirement.

As shown in section 15.3 (page 424), the transformer resistance can have a dramatic effect on the operation of the rectifier-capacitor system. Unless the winding resistance is *much* less than the load resistance, it is advisable to take this into consideration using circuit simulation.

Check thermal dissipation. Now we have a rough design, we must determine the size of the required heatsink or fan. This time, assume the input voltage is worst-case high, that is, 10% above its nominal value. Calculate the dissipation in the regulator and design a heatsink (section 28, page 783). If the heatsink is too large, change the regulator to a switcher or, as a last resort, add a fan (section 28.6, page 791).

This process is necessarily iterative. A spreadsheet is useful to keep track of the variables and automate calculations. However, because a spreadsheet hides formulae, it is extremely important to do a *sanity check* on the results.

Measurement Checklist

There are many tests that one can apply to a new power supply. Assuming it doesn't catch fire when first powered on, these are the critical things to check:

1. Check the operating voltages in the circuit and ensure that the components are operating well within their limits.
2. Check the supply dissipation under *high line*, worst case output current. If this is marginal, repeat the test in an environmental chamber with the ambient temperature at its worst-case maximum.
3. Check for correct operation under *low line*.
4. Check the temperature of rectifier diodes.
5. View the output voltage using AC coupling and a sensitive oscilloscope setting. Characterise the output ripple and ensure it meets the requirements.
6. Check the RMS current in the filter capacitors and ensure that it does not exceed the capacitor rating.

15.17 Appendix: The Transformer Voltage Equation

In the design of power-line frequency transformers, a so-called *transformer voltage equation* (often referred to simply as the *transformer equation*) is used to estimate the number of turns required on each winding. It is useful to understand the development of an equation in order to understand its limitations.

In this section, we develop this equation and then show how the equation is applied.

Consider that a sinusoidal AC voltage $e(t)$ is applied to an N turn coil wound on a magnetic core. Then by Faraday's law, the flux in the core and the applied voltage are related as:

$$e(t) = N \frac{d\phi}{dt} \quad (651)$$

where ϕ is the magnetic flux in the core in webers.

We would like to relate the maximum value of flux in the core to the applied voltage. The maximum allowable flux is usually specified in terms of flux density, webers/metre², or *Teslas*. That is:

$$B = \frac{\phi}{A_c} \quad (652)$$

where B is the magnetic flux density in the core in Teslas. A_c is the cross-sectional area of the magnetic core in metre² (figure 406 on page 466).

Substitute for ϕ from equation 652 into equation 651 and we have:

Symbol	Unit	Description
ϕ	Webers	Magnetic flux
B	Teslas	Magnetic flux density, webers/m ²
B_{sat}	Teslas	Saturation flux density, webers/m ²
B_g	Gauss	Magnetic flux density, 10^{-4} Tesla
B_{gsat}	Gauss	Saturation flux density, gauss
μ		Magnetic Permeability
H	Amp-turns/metre	Magnetizing Force
L	Henries	Inductance
l_c	Metres	Core length
A_c	Metres ²	Core cross-sectional area, metres ²
A_o	Metres ²	Core window (open) cross-sectional area, metres ²
A_w	Metres ²	Wire cross-sectional area, metres ²
η	(none)	Transformer efficiency, P_o/P_i . A typical value is 0.9.
P_i	watts	Input power to the transformer
P_o	watts	Output power from the transformer
N_p	Turns	Number of winding turns on primary winding
N_s	Turns	Number of winding turns on secondary winding
N	Turns	Total number of windings, $N_p + N_s$
K	(none)	Winding utilization factor: ratio of area available for winding to the area of the core window, NA_w/A_o . A typical value for a toroidal core is 0.2 [127].
f	Hz	Frequency of the electrical waveform
ω	radians/sec	Frequency of the electrical waveform
E	volts	Peak value of a voltage, eg: $e_p = E \sin \omega t$
e_p	volts	Primary voltage
I	amps	Peak value of the current
C	metres ² /amp	Wire current carrying capability

Figure 411: Transformer Symbols

$$\begin{aligned} e(t) &= N \frac{dA_c B(t)}{dt} \\ &= NA_c \frac{dB(t)}{dt} \end{aligned} \quad (653)$$

If the applied voltage $e(t)$ is a sinusoidal wave shape, then the waveform of the flux density $B(t)$ must also be a sinusoidal wave shape, since the integration and differentiation of sine waves does not change their shape. Suppose that the flux density is a sine wave with the function

$$B(t) = B_{pk} \sin \omega t \quad (654)$$

where B_{pk} is the peak value of the flux density and ω the operating frequency in radians per second.

Substitute for $B(t)$ from equation 654 into equation 653:

$$\begin{aligned} e(t) &= NA_c \frac{d(B_{pk} \sin \omega t)}{dt} \\ &= NA_c B_{pk} \omega \cos \omega t \\ &= NA_c B_{pk} 2\pi f \cos 2\pi ft \end{aligned} \quad (655)$$

where f is the operating frequency in Hz.

Consequently, the voltage waveform is a sinusoidal waveform which can be written:

$$e(t) = E_{pk} \cos 2\pi ft \quad (656)$$

where E_{pk} is the peak value of the voltage waveform, given by:

$$E_{pk} = NA_c B_{pk} 2\pi f \quad (657)$$

This is the most basic form of the transformer design equation. It relates the winding voltage to the number of turns, core cross sectional area, peak allowable (saturation) flux density and operating frequency.

To make this formula more user friendly, the voltage is usually expressed in RMS volts, the area in square centimetres and the flux density in Gauss:

$$E_{RMS} = \frac{E_{pk}}{\sqrt{2}} \quad (658)$$

$$B \text{ Tesla} = 10^4 B \text{ Gauss} \quad (659)$$

$$A_c \text{ cm}^2 = 10^4 A_c \text{ metre}^2 \quad (660)$$

Substituting these values in equation 657, we have:

$$\begin{aligned} E_{RMS} &= \frac{1}{\sqrt{2}} N A_c B_{pk} 2\pi f \times 10^{-8} \\ &= 4.45 N A_c B_{pk} f \times 10^{-8} \end{aligned}$$

It's usual to rewrite the equation to determine the *volts per turn* of the transformer:

$$\frac{E_{RMS}}{N} = 4.45 A_c B_{pk} f \times 10^{-8} \text{ volts per turn} \quad (661)$$

Equation 661 applies to operation with sine waves. A similar analysis for *square waves* yields the result:

$$\frac{E_{RMS}}{N} = 4.0 A_c B_{pk} f \times 10^{-8} \text{ volts per turn} \quad (662)$$

Example

A 60 Hz power transformer similar to the unit shown in figure 400 is to be designed. Its primary voltage v_p is 117VAC. A single secondary winding has voltage v_s of 12VAC.

The core material is a silicon iron steel which has a saturation flux density of 12 kilogauss. The core is a square cross section, 2.5cm on each side.

Determine the volts/turn value for this transformer and the number of turns on each winding.

Solution

Applying equation 661, we have the following:

$$\begin{aligned}\frac{E_{RMS}}{N} &= 4.45 A_c B_{pk} f \times 10^{-8} \\ &= 4.45 \times (2.54)^2 \times (12000) \times 60 \times 10^{-8} \\ &= 0.206 \text{ volts per turn}\end{aligned}$$

The primary winding will therefore require:

$$\begin{aligned}N_p &= \frac{v_p}{E_{RMS}/N} \\ &= \frac{117}{0.206} \\ &= 566 \text{ turns}\end{aligned}$$

The secondary winding will require:

$$\begin{aligned}N_s &= \frac{v_s}{E_{RMS}/N} \\ &= \frac{12}{0.206} \\ &= 58 \text{ turns}\end{aligned}$$

Notice the large number of turns on the primary and secondary windings, compared to the high-frequency transformer designed in section 15.15 on page 468.

15.18 Appendix: Transformer Power Equation

In this section, we develop the *transformer power equation*. This equation relates the size of the transformer to its electrical specifications and provides an estimate of a suitable size of magnetic core¹⁶⁸.

We'll do this for a transformer driven by a square wave. Using the terms defined in figure 411 and repeating the development of the previous section, we can show that

$$E = 4.0 N A_c B_{sat} f \text{ volts per turn} \quad (663)$$

The *winding factor* K , the fraction of the core window area actually taken by the windings, is defined as:

$$K = N \frac{A_w}{A_o} \quad (664)$$

¹⁶⁸This equation development closely follows material in reference [127], with some elaboration and change of symbols.

Equation 664 assumes that both the primary and secondary windings have the same wire cross-sectional area A_w . That's not necessarily the case. However, where small currents are involved, it simplifies the construction.

Solve for N from equation 664 and substitute this in equation 663. Isolate the quantity $A_o A_c$:

$$A_o A_c = \frac{EA_w}{4KB_{sat}f} \text{ metres}^4 \quad (665)$$

The quantity $A_o A_c$ gives us some indication of the volume of the transformer. It would be helpful to relate the wire cross-sectional area A_w to the transformer current. The current-carrying capacity of the primary winding is defined as:

$$C = \frac{A_w}{I_p} \text{ metres}^2/\text{amp} \quad (666)$$

Solve for A_w and substitute in equation 665:

$$A_o A_c = \frac{EI_p C}{4KB_{sat}f} \text{ metres}^4 \quad (667)$$

Recognise that the quantity EI_p is the input power P_i to the transformer:

$$A_o A_c = \frac{P_i C}{4KB_{sat}f} \text{ metres}^4 \quad (668)$$

The input power is related to the output power by the efficiency η :

$$\eta = \frac{P_o}{P_i} \quad (669)$$

Substituting for input power from equation 669, we can rewrite equation 668 as:

$$A_o A_c = \frac{P_o C}{4K\eta B_{sat}f} \text{ metres}^4 \quad (670)$$

Example

A small power transformer is to be designed. It is to be operated at a frequency of 100kHz.

The core shown in figure 412 is suggested by a local supplier. It is a toroid with a maximum flux density of 2500 Gauss. Estimate the power-handling capability of this core.

Solution

The solution requires gathering together the various known quantities in equation 670 and then solving for the output power P_o . Equation 670 is in MKS units, so the information must be converted into that form.

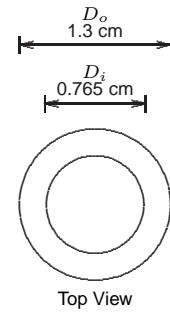


Figure 412: Candidate Toroid Core

Core Window Area A_o

From figure 412, the core window area is given by:

$$\begin{aligned} A_o &= \pi r^2 \\ &= 3.14 \times \left(\frac{0.765}{2}\right)^2 \\ &= 0.459 \text{ cm}^2 \end{aligned}$$

Using the units conversion technique described in section 1.2 (page 48), the conversion factor of cm^2 to metres^2 is given by:

$$\begin{aligned} 1 \text{ metre} &= 10^2 \text{ cm} \\ \text{Then } 1 \text{ metre}^2 &= 10^4 \text{ cm}^2 \end{aligned}$$

For the purposes of unit conversion, the last line may be rearranged so that cm^2 appears in the denominator:

$$1 = \frac{\text{metre}^2}{10^4 \text{ cm}^2}$$

Now we can use this expression to convert A_o to metres 2 :

$$\begin{aligned} A_o &= 0.459 \text{ cm}^2 \\ &= 0.459 \text{ cm}^2 \left[\frac{\text{metre}^2}{10^4 \text{ cm}^2} \right] \\ &= 0.459 \times 10^{-4} \text{ metre}^2 \end{aligned}$$

Core Cross-Sectional Area A_c

From figure 412, the core cross-sectional area is given by:

$$A_c = H \times (R_o - R_i) = H \times \left(\frac{D_o}{2} - \frac{D_i}{2} \right) = 0.65 \left(\frac{1.3}{2} - \frac{0.765}{2} \right) = 0.174 \text{ cm}^2$$

where H is the height of the core cross section, R_o the radius to the outer surface, R_i the radius to the inner surface.

Similarly to the conversion of A_o given above:

$$A_c = 0.174 \times 10^{-4} \text{ metres}^2$$

Current Carrying Capacity C

Reference [127] gives the maximum allowable current density in copper wire of $4.05 \times 10^{-3} \text{ cm}^2/\text{amp}$. Converting the area to square metres, similar to the conversion of A_o given above:

$$C = 4.05 \times 10^{-3} \left(\frac{\text{cm}^2}{\text{amp}} \right) \times 10^{-4} \left(\frac{\text{m}^2}{\text{cm}^2} \right) = 4.05 \times 10^{-7} \frac{\text{m}^2}{\text{amp}}$$

Winding Utilization K

Reference [127] suggests a utilization factor of $K = 0.2$ for a toroidal core.

Efficiency η

A typical value for transformer efficiency is 90%, so $\eta = 0.9$.

Maximum flux density B_{sat}

The maximum flux density is given as 2500 Gauss for this core. This must be converted to *Teslas*, the MKS unit for flux density:

$$1 \text{ Gauss} = 10^{-4} \text{ Tesla}$$

then

$$2500 \text{ Gauss} = 2500 \text{ Gauss} \times \left(\frac{10^{-4} \text{ Tesla}}{1 \text{ Gauss}} \right) = 0.25 \text{ Tesla}$$

Operating frequency f

The operating frequency has been chosen to be 100kHz, or 10^5 Hz.

Solving for power capability

Now we have all the necessary information, in MKS units, to plug into equation 670. Rearranging the equation to solve for output power P_o , we have:

$$\begin{aligned} P_o &= \frac{A_o A_c 4 K \eta B_{sat} f}{C} \\ &= \frac{(0.459 \times 10^{-4}) \times (0.174 \times 10^{-4}) \times 4 \times 0.2 \times 0.9 \times 0.25 \times 10^5}{4.05 \times 10^{-7}} \\ &= 28.5 \text{ watts} \end{aligned}$$

There are many approximations in this calculation, so the result should be regarded as a guide. A careful design should keep the actual transformer power output well below this value.

15.19 Wire Table

A.W.Gauge No.	Diameter, Mils	Cross-sectional Area, Circular Mils	Square Inches	Resistance Ohms per 1000 ft	Fusing Current, Amps
0000	460.0	211,600	0.1662	0.04901	
000	409.6	167,800	0.1318	0.06180	
00	364.8	133,100	0.1045	0.07793	
0	324.9	105,500	0.08289	0.09827	
1	289.3	83,690	0.06573	0.1239	
2	257.6	66,370	0.05213	0.1563	
3	229.4	52,640	0.04134	0.1970	
4	204.3	41,740	0.03278	0.2485	
5	181.9	33,100	0.02600	0.3133	
6	162.0	26,250	0.02062	0.3951	668
7	144.3	20,820	0.01635	0.4982	561
8	128.5	16,510	0.01297	0.6282	472
10	101.9	10,380	0.008155	0.9989	333
12	80.81	6,530	0.005129	1.588	235
14	64.08	4,107	0.003225	2.525	166
15	57.07	3,257	0.002558	3.184	140
16	50.82	2,583	0.002028	4.015	117
17	45.26	2,048	0.001609	5.064	98.4
18	40.30	1,624	0.001276	6.385	82.9
19	35.89	1,288	0.001012	8.051	69.7
20	31.96	1,022	0.0008023	10.15	58.4
21	28.46	810.1	0.0006363	12.80	
22	25.35	642.4	0.0005046	16.14	41.2
23	22.57	509.5	0.0004002	20.36	
24	20.10	404.0	0.0003173	25.67	29.2
25	17.90	320.4	0.0002517	32.37	
26	15.94	254.1	0.0001996	40.82	20.5
27	14.20	201.5	0.0001583	51.46	
28	12.64	159.8	0.0001255	64.90	14.4
29	11.26	126.7	0.00009953	81.84	
30	10.03	100.5	0.00007894	103.2	10.2
31	8.928	79.70	0.00006260	130.1	
32	7.950	63.21	0.00004964	164.1	7.19
33	7.080	50.13	0.00003937	206.9	
34	6.305	39.75	0.00003122	260.9	5.12
35	5.615	31.52	0.00002476	329.0	
36	5.000	25.00	0.00001964	414.8	3.62
37	4.453	19.83	0.00001557	523.1	
38	3.965	15.72	0.00001235	659.6	2.50
39	3.531	12.47	0.000009793	831.8	
40	3.145	9.888	0.000007766	1049	1.77
41	2.800	7.842	0.000006159	1323	
42	2.494	6.219	0.000004884	1668	
43	2.221	4.9321	0.000003873	2103	
44	1.978	3.911	0.000003072	2652	

A *mil* is 1/1000 inch.

A *circular mil* is a unit of area, equal to the area of a circle of diameter 1 mil [128]. In metric units, one circular mil is equal to 5.064×10^{-10} metres².

The *fusing current* is the current at which the conductor melts [129]. Current in a wire should be limited to well below the fusing current.

15.20 Appendix: Calculating the Effective (RMS) Value of a Waveform

In many applications, it is necessary to calculate the power generated by a voltage or current waveform in a resistance. In this section, we'll explain the concept of waveform *effective value* and show a general purpose method of calculating the effective value.

If a waveform is a constant DC value, then the power it creates in a resistor is given by:

$$\begin{aligned} P_R &= \frac{V^2}{R} \\ &= I^2 R \end{aligned} \quad (671)$$

where P_R is the power in watts generated in resistor R , V is the DC voltage, I is the DC current, R is the resistance in ohms.

When the voltage and current vary with time, we first determine the *effective* value of the waveform. **The effective value of a current, for example, is the value of the DC current that would cause the same power dissipation.** Then, if the effective value of a voltage or current is known, it's possible to use this in place of the DC value in equation 671:

$$\begin{aligned} P_R &= \frac{V_{eff}^2}{R} \\ &= I_{eff}^2 R \end{aligned}$$

This is very useful and convenient, and so sine waves – especially in power applications – are described by their effective value rather than peak or peak-peak value.

The effective value is also known as the *RMS* value. RMS stands for *Root-Mean-Square* which is handy, since it describes the mathematical method of calculating the effective value of a waveform¹⁶⁹.

For a repetitive waveform of period T , the effective or RMS value of a waveform $x(t)$ may be shown to be

$$x_{eff} = \sqrt{\frac{1}{T} \int_0^T x^2(t) dt} \quad (672)$$

Deconstructing the equation, starting at the inside:

- Square the function. The squaring operation removes the effect of negative voltage or current – power is generated regardless of the polarity of voltage or current.
- Find the average value of this squared function by integrating over one complete cycle and dividing by the period T of the cycle.
- Take the square root.

For a sine wave, the effective value is the familiar $1/\sqrt{2} = 0.707$ times the peak value of the waveform.

RMS Value of an Arbitrary Waveform, Graphical Method

If a waveform can be characterised as a mathematical function of time, then its effective value may be determined with equation 672. Substitute the formula for $x(t)$ and turn the mathematical crank (or reach for a mathematical handbook). This method is illustrated with the examples in section 15.20, page 485.

¹⁶⁹In the computer world, RMS is Richard M Stallman, a prominent figure in the world of free software.

In some cases, the waveform cannot be characterised as a mathematical function or the integral may be difficult to calculate. Then the integral may be determined graphically by finding the area under the curve. This uses the *graphical method of integration* and is now illustrated using a spreadsheet for calculations.

We'll use the current waveform shown in figure 413 to illustrate the graphical method.

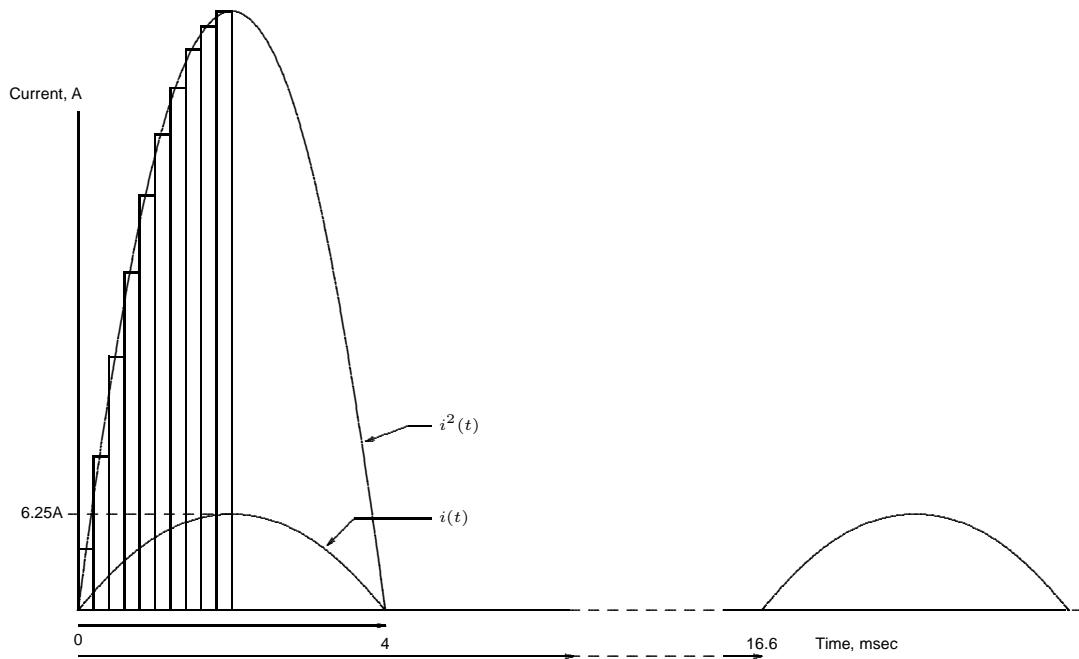


Figure 413: RMS Calculation, Graphical Method

The waveform, its square, and an approximation to the area under the squared waveform are shown in figure 413. The steps to calculate the RMS value of a current waveform are these:

1. From a simulation or measurement obtain the current waveform $i(t)$.
2. Blow up the current pulse waveform $i(t)$ to read off the current and time values.
3. Divide the waveform into a series of time segments. In this example, the duration of the pulse is 4mSec. It was arbitrarily divided into 10 segments, each 0.2mSec in duration.
Because the waveform is symmetrical, we can calculate the area under the first half of the waveform and then double the result.
4. At each time segment, copy down from the oscilloscope display or simulation the value of the time (column 1) and current value (column 2).
5. Square the current waveform for each time segment (column 3) and determine the area of that segment in amps²-mSec (column 4).
6. Add all the values in column 4 to determine the area in amps²-mSec under one-half the waveform. In effect, we are calculating the area of the rectangles of figure 413, which approximates the half the total area under the $i^2(t)$ waveform.

Double this value to get the total area under the waveform.

Time, mSec	Voltage, V	Current, A	Current squared	Current ² x time increment amps ² -mSec
0.00	0.77	0.48	0.23	0.05
0.20	2.50	1.56	2.44	0.49
0.40	4.20	2.63	6.89	1.38
0.60	5.60	3.50	12.25	2.45
0.80	6.80	4.25	18.06	3.61
1.00	7.80	4.88	23.77	4.75
1.20	8.60	5.38	28.89	5.78
1.40	9.20	5.75	33.06	6.61
1.60	9.60	6.00	36.00	7.20
1.80	9.70	6.06	36.75	7.35

- Divide the total area by the period T of the waveform. In this case, the pulse repeats every 60th of a second, so $T = 16.6\text{mSec}$. The result is the mean-squared current.
- Find the square root. This is the effective or RMS value of the waveform. The heating effect of this waveform in a resistor would be the same as a DC current of 2.38 amps.

Results

Total area under half the current-squared waveform:	39.67	amp ² -mSec
Total area under the entire current-squared waveform:	79.34	amp ² -mSec
Mean-squared current	4.78	amp ²
Root-mean-squared current	2.18	amps

Effective Value of Triangle Waveform

The inductor current in a switching regulator has a triangular shape. To determine the heating effect of the current waveform in the resistive components that carry that current we need to calculate the RMS value of the waveform. As it turns out, for a triangle waveform of constant height the duty cycle has no effect on the RMS value. Consequently, we can find the RMS value of the sawtooth waveform shown in figure 414, which is simpler to deal with.

The RMS value is given by

$$I_{RMS} = \sqrt{\frac{1}{T} \int_0^T i^2(t) dt} \quad (673)$$

To find the RMS current, we work backwards through this equation.

The equation of a straight line is $y = mx + b$, where m is the slope and b the y intercept. Substituting $i(t)$ for y , ΔI for m , t for x and $-\Delta I/2$ for b , the equation of current is:

$$\begin{aligned} i(t) &= \frac{\Delta I}{T}t - \frac{\Delta I}{2} \\ &= \Delta I \left(\frac{t}{T} - \frac{1}{2} \right) \end{aligned} \quad (674)$$

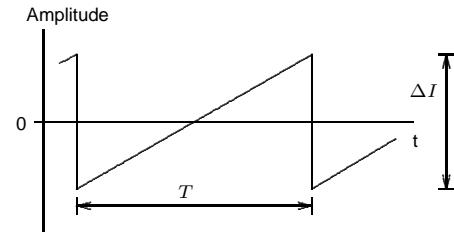


Figure 414: Sawtooth Waveform

Then the current squared is given by:

$$\begin{aligned} i^2(t) &= \Delta I^2 \left[\frac{t^2}{T^2} - 2 \left(\frac{t}{T} \times \frac{1}{2} \right) + \frac{1}{4} \right] \\ &= \Delta I^2 \left[\frac{t^2}{T^2} - \frac{t}{T} + \frac{1}{4} \right] \end{aligned} \quad (675)$$

Repeatedly using $\int x^n dx = \frac{x^{n+1}}{n+1}$, the integral over one cycle is given by:

$$\begin{aligned} \int_0^T i^2(t) dt &= \int_0^T \Delta I^2 \left[\frac{t^2}{T^2} - \frac{t}{T} + \frac{1}{4} \right] dt \\ &= \Delta I^2 \left(\frac{t^3}{3T^2} - \frac{t^2}{2T} + \frac{t}{4} \right)_0^T \\ &= \Delta I^2 \left(\frac{T}{3} - \frac{T}{2} + \frac{T}{4} \right) \\ &= \Delta I^2 \frac{T}{12} \end{aligned} \quad (676)$$

Substituting from equation 676 in equation 673, we obtain for the RMS current:

$$\begin{aligned} I_{RMS} &= \sqrt{\frac{1}{T} \int_0^{T/2} i^2(t) dt} \\ &= \sqrt{\frac{1}{T} \left(\Delta I^2 \frac{T}{12} \right)} \\ &= \frac{\Delta I}{\sqrt{12}} \end{aligned} \quad (677)$$

This is the RMS value of a sawtooth or triangle current waveform of peak-peak amplitude ΔI .

RMS Value of a Pulse Waveform

For a conservative design, it is desirable that the AC current into the switching regulator does not affect other devices on the same circuit. Ideally, then, all the AC input current flows through the capacitor C_{in} . The total input current is shown in figure 384 on page 440. The AC component of this current is the same waveform with the average value removed. In figure 415, to simplify the analysis this has been approximated by a rectangular waveform. (An exact expression is given in [121]).

First, we need to determine the positive and negative peak currents. The average

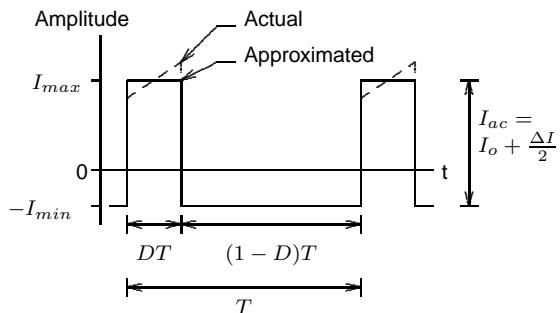


Figure 415: Pulse Waveform

current through the input capacitor is zero, so the positive and negative areas of figure 415 must sum to zero. That is:

$$\begin{aligned} I_{max}DT - I_{min}(1-D)T &= 0 \\ I_{max}DT &= I_{min}(1-D)T \end{aligned} \quad (678)$$

The sum of these two currents are equal to the peak-peak value I_{ac} :

$$I_{max} + I_{min} = I_{ac} \quad (679)$$

Putting equation 678 and 679 together we have:

$$I_{max} = (1-D)I_{ac} \quad (680)$$

$$I_{min} = -DI_{ac} \quad (681)$$

Now we can determine the RMS value of the current. Over the positive interval, the area value is

$$\int_0^{DT} i^2(t)dt = I_{max}^2 DT \quad (682)$$

During the negative interval, the area value is

$$\int_{DT}^T i^2(t)dt = I_{min}^2(1-D)T \quad (683)$$

Substituting these values into equation 673 for RMS current, along with the expressions of equation 680 for I_{max} and equation 681 for I_{min} , we have:

$$\begin{aligned} I_{RMS} &= \sqrt{\frac{1}{T} \int_0^T i^2(t)dt} \\ &= \sqrt{\frac{1}{T} I_{max}^2 DT + I_{min}^2(1-D)T} \\ &= \sqrt{\frac{1}{T} (1-D)^2 I_{ac}^2 DT + D^2 I_{ac}^2 (1-D)T} \\ &= I_{ac} \sqrt{D(1-D)} \end{aligned} \quad (684)$$

This value reaches maximum for a 50% duty cycle, in which case

$$\begin{aligned} I_{RMS} &= I_{ac} \sqrt{D(1-D)} \\ &= I_{ac} \sqrt{0.5(1-0.5)} \\ &= 0.5I_{ac} \end{aligned} \quad (685)$$

In words, the maximum RMS value of the current through the input capacitor is equal to half the peak-peak value of the input current.

References

A number of simple power supply designs are shown in every edition of the ARRL Handbook, [45].

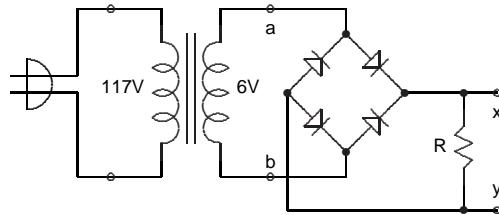
The *Linear & Switching Voltage Regulator Handbook* published by On Semiconductor [112] is a useful overview of power supply design. It is currently available on the Web.

Brown [126] and Pressmann [130] are both useful sources of information on the design of a wide variety of switching power supplies.

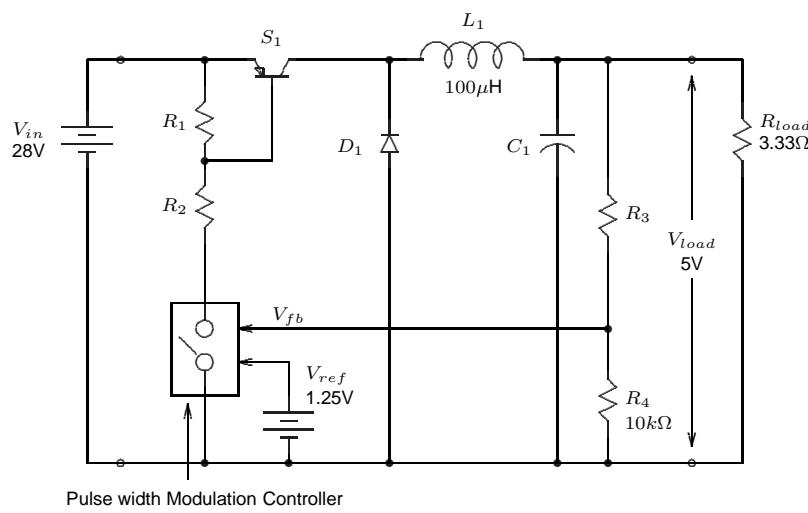
15.21 Exercises

Assume that the forward voltage drop of a conducting diode is 0.6 volts.

- For the circuit shown, sketch a scaled version of the input voltage $V_{ab}(t)$ and the output voltage $V_{xy}(t)$.



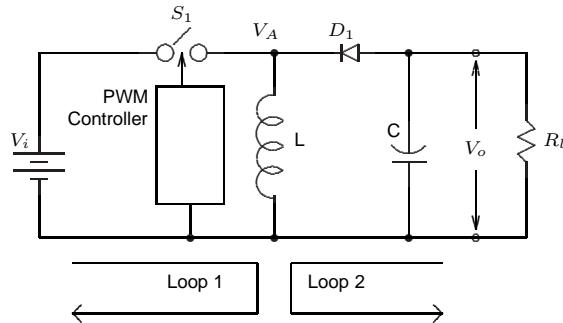
- Draw the circuit of a full-wave bridge rectifier, being driven from a 117:12VAC transformer, with a capacitor filter. What is the no-load voltage across the capacitor? What voltage rating would you recommend for the capacitor?
- Consider the switching regulator circuit shown in the figure.



- The frequency of oscillation is 50kHz.
 - The filter capacitor C is large enough that the ripple voltage across the load resistor is negligible.
 - The Pulse Width Modulation Controller compares the feedback voltage V_{fb} with the reference voltage and adjusts the duty cycle until they match.
- Calculate a suitable value for R_3 .
 - Calculate the duty cycle DC for the conditions shown on the diagram.
 - For what length of time during each cycle is the PWM controller switch closed, in microseconds?
 - Calculate the minimum value of the inductor current that occurs during a switching cycle, in amps.

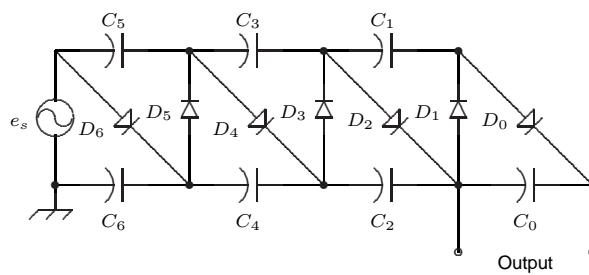
4. Section 15.8 described the *buck* or *step-down* switching regulator. The circuit shown in the figure given here can be used to generate an output voltage that is negative with respect to the input voltage.

When the switch is closed, current builds up in the inductor around Loop 1. When the switch is opened current continues around Loop 2, charging the output capacitor C .



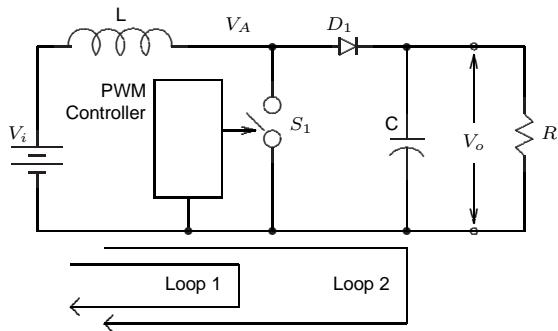
- (a) Assuming a 50% duty cycle, draw the waveforms for inductor voltage, inductor current, diode current and input current.
- (b) Develop an expression for the output voltage in terms of the input voltage and switching duty cycle.
5. The voltage multiplier configuration shown in section 15.11 was capable of even-numbered voltage multiplication.

For the voltage multiplier circuit shown below, determine the output voltage and show that it is an odd multiple of the input voltage.

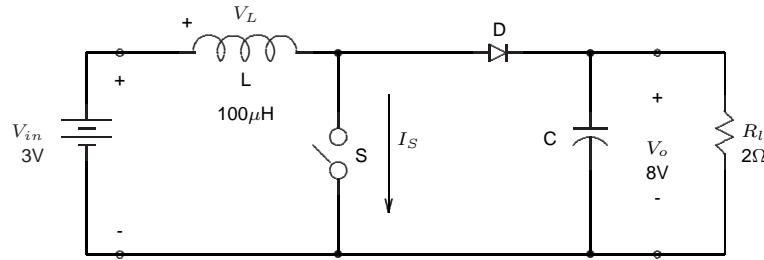


6. Section 15.8 described the *buck* or *step-down* switching regulator. The circuit shown in the figure given here can be used to generate an output voltage that is greater than the input voltage.

When the switch is closed, current builds up in the inductor around Loop 1. When the switch is opened current continues around Loop 2, charging the output capacitor C .

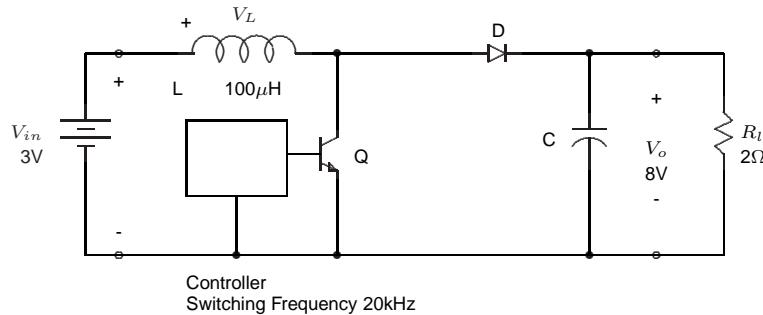


- (a) Assuming a 50% duty cycle, draw the waveforms for V_A , inductor voltage, inductor current and input current.
- (b) Develop an expression for the output voltage in terms of the input voltage and switching duty cycle.
7. The circuit of a low-current power supply is shown below. You may assume that the input is fixed at 117 volts RMS, 60 Hz. The transformer is wound so that the secondary voltage is 12.6 volts RMS. The internal resistance of the secondary is small enough that it can be ignored. The minimum value of V_B is 14 volts.
- The diagram shows a power supply circuit. It starts with a 117VAC 60Hz input connected to a center-tapped transformer. The primary winding is connected to ground. The secondary winding has two output lines. The top line is connected to a diode and then to the non-inverting input (+) of an operational amplifier (op-amp) labeled 'A'. The inverting input (-) of the op-amp is connected to ground through a zener diode labeled VR_1 . The output of the op-amp is connected to the base of a transistor Q_1 . The collector of Q_1 is connected to the bottom output line of the secondary winding. The top output line of the secondary winding is connected to a capacitor C_1 and then to the inverting input (-) of the op-amp. The output V_o is connected to a load resistor R_l of 240 ohms. The output V_o is also connected to a 12V 10k resistor R_2 and a 10k resistor R_3 in series. The junction of R_2 and R_3 is connected to ground.
- (a) Draw the waveforms V_A and V_B , showing the vertical and horizontal scale factors.
- (b) Calculate the value of capacitor C_1 .
- (c) Calculate the peak current through the rectifier diode. (You may assume that the diode current waveform is a rectangular pulse.)
- (d) Calculate the power dissipation in transistor Q_1 if the average value of V_B is 15.5 volts. The dissipation caused by base current can be ignored.
- (e) Choose a suitable voltage for the zener diode.
- (f) The suggestion is made that the half-wave rectifier could be replaced with a diode full-wave bridge, and that this would lead to some possible cost savings in other components. Is this true? Explain.
8. A Step Up switching regulator is shown in the figure. The switch S is driven by a controller with a duty cycle such that the output voltage V_o is 8 volts. The switching frequency is 20kHz. For this part of the question, the diode and switch may be considered ideal.

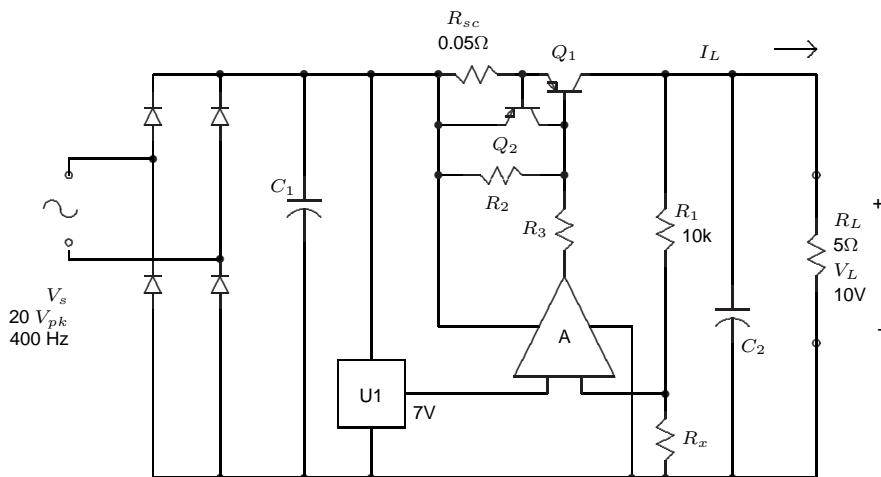


- Calculate the average load current I_{lav}
- Calculate the switch average current I_{Sav}
- Calculate the switch peak current I_{Spk} .

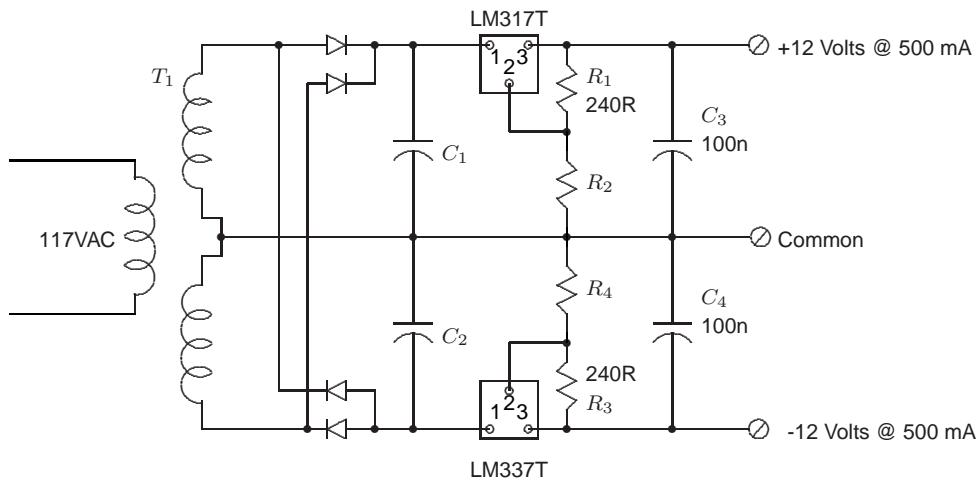
The same regulator circuit uses the non-ideal switching components Q and D as shown below. The average on-voltage of the switch Q is $V_{Qon}=0.5\text{V}$. The average on-voltage of the diode is V_{Don} is 0.7V .



- Calculate Duty Cycle D
 - Calculate the diode average current I_{Dav}
9. Consider the linear regulator shown below. Notice that the line frequency is 400Hz. V_{be} for both transistors is 0.6 volts. The integrated circuit U_1 generates a stable source of 7 volts.



- (a) Calculate the value of the resistor R_x .
- (b) Calculate the average output current I_L
- (c) Assuming a ripple voltage ΔV_C across the capacitor C_1 of 1 volt peak-peak, calculate a suitable value for the filter capacitor C_1 in microfarads.
- (d) Calculate the value of the short circuit current I_{sc} .
- (e) Label the polarity of the inputs on the operational amplifier A .
- (f) (Advanced) Explain the purpose of R_2 and R_3 .
10. The circuit diagram shows a power supply which supplies ± 12 volts at 500mA maximum output current.



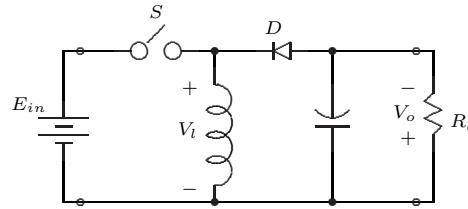
You may assume that

- the peak-peak ripple voltage across each filter capacitor (C_1, C_2) is 1.5 volts
- the dropout voltage of each regulator is 2 volts.
- the thermal resistance θ_{jc} of the regulator IC is 5°C per watt
- the ambient temperature is 40°C
- the maximum allowable junction temperature of the regulator IC is 100°C
- the regulators maintain 1.2 volts between terminals 2 and 3, with negligible current into or out of pin 2.

Determine

- (a) the required area for the heat sink on each regulator IC.
- (b) the capacitance in microfarads and voltage rating of the filter capacitors C_1 and C_2
- (c) the RMS value of the current in each bridge diode. (You may approximate the diode current waveform.)
- (d) the transformer secondary RMS voltage rating
- (e) the values of resistors R_2 and R_4 .

11. The circuit of an inverting switching regulator is shown below. The switch operates at a rate of at least some tens of kilohertz, with a duty cycle that controls the magnitude of the output voltage. The circuit produces an output voltage which is *negative* compared to the input voltage.



You may assume:

- All the components are ideal. For example, the diode has zero forward voltage when conducting and zero reverse current when shut off.
 - The output voltage is negative so that the diode D is open circuit when the switch is closed (during the interval T_{on}) and short circuited when the switch is opened (during the interval T_{off}).
 - The ripple voltage across the capacitor is small enough that the output voltage may be assumed to be constant.
 - When analysed, the system is in steady-state.
- (a) Determine the voltage across the inductor during T_{on} , in terms of E_{in} and V_o .
 - (b) Determine the voltage across the inductor during T_{off} , in terms of E_{in} and V_o .
 - (c) Using the fact that the steady-state average voltage across an inductor must be zero, find V_o in terms of E_{in} , T_{on} and T_{off} .
 - (d) Assume that:
 - $E_{in} = 5$ volts
 - $V_o = -12$ volts
 - $R_L = 250\Omega$
 - The switching frequency $f_{osc}=100\text{kHz}$
 - The peak current in the inductor is 1.5 times the average inductor current.

Determine the following:

- i. The duty cycle $T_{on}/(T_{on} + T_{off})$.
- ii. Draw the inductor current waveform.
- iii. Calculate a suitable value for the inductance L .

12. Determine the RMS value of a sine wave of 10 volts peak and period 2mSec, using a spreadsheet and the method of section 15.20 on page 483.

What is the exact value? Explain any discrepancy.

13. Using the approach shown in section 15.17, develop equation 662.
14. Reference [127] suggests a maximum current density in transformer windings of $4.04 \times 10^{-3} \text{ cm}^2/\text{ampere}$. Another reference suggests 420 amperes/cm². Brown [126] lists a rating of 1000 circular mils per ampere. (A *mil* is 1/1000 inch. A *circular mil* refers to the circumference of the wire. For example, a wire of 2 circular mils has a circumference of 2/1000 inch.)

- (a) Convert these ratings to $\text{cm}^2/\text{ampere}$ and compare them.
- (b) What would be the effect on transformer power rating if the 2nd and 3rd current capacities were used? (See for example the calculation on page 479.)
15. A power supply is being designed to use a linear regulator integrated circuit. The requirements are:
- Input voltage to the regulator is 18VDC.
 - Under short circuit conditions (the worst case for dissipation in the regulator) output voltage from the regulator is 0VDC and the output current is 1 ampere.
- The relevant specifications for the regulator IC are:
- Maximum Junction Temperature $T_{j_{max}} = 100^\circ\text{C}$
 - Junction to case thermal resistance $4^\circ\text{C}/\text{watt}$
 - Case to ambient thermal resistance $35^\circ\text{C}/\text{watt}$
 - Maximum ambient temperature $T_{a_{max}} = 40^\circ\text{C}$
- (a) Show that it is not possible to meet these specifications with this IC.
- (b) Assuming that the performance specifications cannot be adjusted, suggest an alternative design.
16. Design a charge pump inverter to generate -5 volts from a supply of +5 volts. The switching frequency is 50kHz, the output current 2mA, and the allowable ripple in the output voltage 100mV.

16 Precision Rectifiers

In measurement and control applications, there are many situations where we are interested in the properties of a signal. Signals have number of properties – statistical, spectral and so on. Here we will focus on the *amplitude*.

When we measure the amplitude of an AC signal, we are measuring a particular amplitude property of that waveform. For example, an AC voltmeter typically displays the RMS (effective) value of the waveform.

In this section, we'll look at circuits capable of generating those measurements, but first a brief exposition to explain their significance.

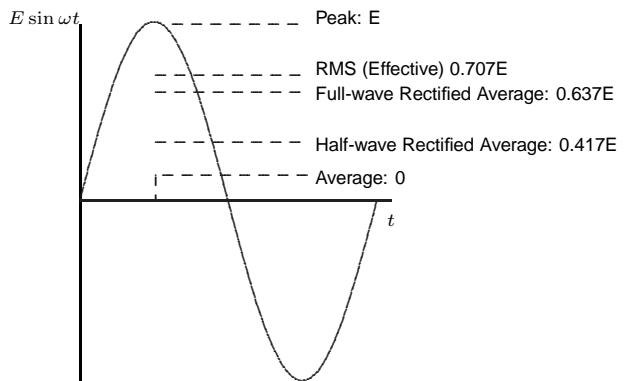


Figure 416: Sine Wave Amplitude Measurements

Average

For a time-varying waveform, the *average* value is found by integrating the signal over a period of time and dividing by the measurement interval. If the waveform is periodic, one cycle of the waveform is sufficient. If the waveform is random, then there is a tradeoff between the accuracy of the measurement and the time required to take the measurement. Longer measurements yield more accurate results.

Mathematically, the average value of a waveform is given by

$$v_{av} = \frac{1}{T} \int_0^T v(t) dt \quad (686)$$

where T is the averaging interval and $v(t)$ is the waveform function of time.

The average value of a waveform is often characterised as *the DC component* and is measured with a DC instrument, such as a moving-coil ammeter. This device has an inherent lowpass filter built into its mechanism inasmuch as the movement has inertia and behaves as a mechanical lowpass filter. When the AC component is at a high frequency compared to the mechanical response of the meter, the meter eliminates the AC component. In electronics, the average component of a waveform may be generated by lowpass filtering the waveform.

Notice that a pure AC waveform (one without a DC component) has an average value of zero.

Effective

The *effective* value of a waveform is related to its ability to produce power. Since alternating current generates power in a resistor regardless of the direction of current flow through the resistor, the effective value of an AC waveform is non-zero.

The effective value of a waveform is the value of the DC voltage (or current) that would generate the same average power in a resistor. For example, effective voltage is given by:

$$P_{av} = \frac{v_{eff}^2}{R} \quad (687)$$

where P is the power in watts generated in resistor R by waveform v_{eff} .

The instantaneous power in a resistor is

$$P(t) = \frac{v^2(t)}{R} \quad (688)$$

The average power is obtained by averaging the instantaneous power over one complete cycle of the waveform:

$$P_{av} = \frac{1}{T} \int_0^T \frac{v^2(t)dt}{R} \quad (689)$$

Equating equations 687 and 689 and solving for v_{eff} , we have:

$$v_{eff} = \sqrt{\frac{1}{T} \int_0^T v^2(t)dt} \quad (690)$$

That is, the effective value of a voltage is obtained by squaring, averaging and then square rooting the signal. This is complicated process but the only way to measure effective voltage for any shape waveform. A device which generates a signal proportional to the effective value of any waveform is known as a *true rms* measuring device. The effective value of a waveform is usually known as the *RMS value*.

Rectified Average

The *rectified average* value of a waveform is obtained, as the name suggests, by rectifying the waveform and then averaging the result. For a waveform of a particular shape, there is a fixed scale factor between the rectified average value and the RMS value. Consequently, if the shape of the waveform is known, then the RMS value may be determined by measuring the rectified average value and multiplying it by a conversion factor. In the case of a sine wave, the average rectified value is multiplied by 1.11 to obtain the RMS value.

Rectification and averaging are much simpler than computing RMS, so this greatly simplifies the design of a measuring instrument. Since many measurement waveforms are sinusoidal, most voltmeters make this assumption and automatically apply the appropriate conversion factor. However, the readings will be completely incorrect if the waveform is not sinusoidal in shape.

Notice that we could also measure the *peak* value of the waveform, assume that it is a sine wave, and compute the effective value by multiplying by a constant. Again, this would be totally incorrect for waveforms of other shapes.

Peak

It is possible to extract a voltage proportional to the peak value of the waveform. This can be stored in an analog memory device (a capacitor) until manually reset, or it can gradually discharge away so that the circuit is self-resetting. This might be useful for monitoring a power supply line for noise spikes. In audio recording, certain waveforms have a very high peak-average ratio (they are very *spiky*), and it is useful to know the peak value to avoid overloading the recording equipment.

Peak to Peak

Picture two peak detectors, one for positive peaks, the other for negative, and a differential amplifier that measures the distance between them. Then the output of the differential amplifier is proportional to the peak-peak value of the waveform. Of course, the waveform need not be symmetrical.

Many of the rectifier and peak-detector circuits in this section rely on the unidirectional conduction properties of the semiconductor diode. Section 6 described a number of simple circuits using the semiconductor diode. In all these circuits, the forward drop of the diode was relevant and had to be taken into consideration.

This forward voltage drop is inconvenient in the signal processing applications listed above. Fortunately, the currents in these circuits are modest, and so an operational amplifier may be used to cancel out the voltage drop of the diode. The result is a circuit which closely approximates the ideal rectifier. Since circuits using this technique generate more accurate results than an unadorned diode, the op-amp/diode combination is known as a *Precision Diode* and rectifiers using this technique are known as *Precision Rectifiers*.

Microprocessor Computation

In considering the various circuits shown in this section (especially the circuit for RMS conversion, which is relatively complex) it is tempting to consider using a microprocessor to do these computations. A microprocessor with an A/D converter can read the waveform into the microprocessor and then a software algorithm can compute average, peak and effective values, along with other signal statistics that might be useful.

This will work for clean signals that are a good match for the dynamic range of the A/D converter, but it won't work well for signals that are small or contaminated by AC noise. A typical microprocessor A/D converter is an 8 bit device. Powered by 5 volts and operating in ratiometric fashion, that makes each of the 256 steps equivalent to 19.5 millivolts. For any sort of resolution, the signal must be much larger than the step size. Consequently, the microprocessor A/D converter must be preceded by a gain-switched amplifier in order to read small signals.

In the design of a voltmeter test instrument, it is a viable approach to use a different type of A/D. The dual-slope A/D converter is accurate for small signals and can be designed to inherently reject AC line noise. The dual-slope converter is strictly a DC device, however, and so this A/D must be preceded by a precision rectifier (and possibly an RMS conversion circuit) to prepare AC signals for measurement.

That said, there may come a time when a high-resolution A/D converter plus microprocessor is the more economical approach. At the time this is being written (mid 2010) the price of a simple microprocessor is approaching that of an op-amp. An analog rectifier circuit requires ancillary components: resistors, capacitors, diodes and circuit-board area, each of which has an associated purchase and assembly cost. A microprocessor-based system requires software, which has essentially zero purchase and assembly cost.

16.1 Precision Diode

In section 6 we saw how a semiconductor diode could be used to generate half-wave rectified waveforms and other functions. A half-wave rectifier circuit, with its transfer function and representative waveforms, is shown in figure 417.

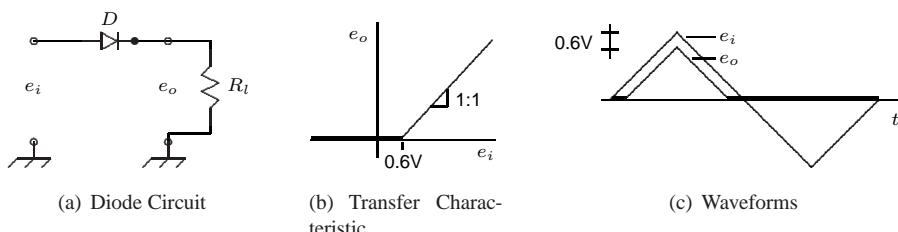


Figure 417: Half-Wave Rectifier

Notice how the fixed 0.6 volt forward drop of the diode shows on the transfer characteristic of figure 417(b) and on the output waveform of figure 417(c). In addition to the 0.6 volt offset, the transfer characteristic would be somewhat nonlinear, the extent depending on the current through the diode. (A smaller load resistance creates a large variation in diode current, causing a variation in diode forward voltage, distorting the load voltage.)

The equivalent *precision diode* circuit is shown in figure 418.

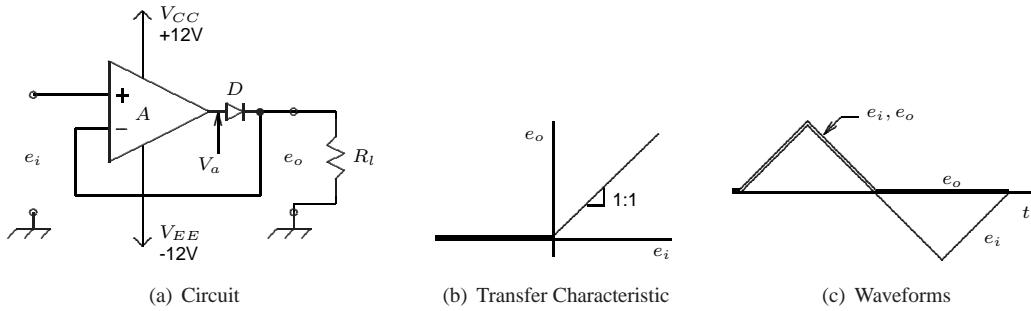


Figure 418: Precision Diode

The op-amp removes the 0.6 volt offset and (though it's not evident from the diagram) linearizes the output so that it follows the input exactly. The output is an ideal half-wave rectified waveform.

Now lets examine how this circuit works.

As is the case for many of these circuits, it's useful to consider the operation in two stages. Consider first that the input voltage is slightly larger than the output voltage. Then the error voltage at the input to the op-amp will be slightly positive, and the output of the op-amp will move in a positive direction. This process continues until the diode is conducting and the error voltage is reduced to zero. At that point, the output voltage of the op-amp must be one diode-voltage-drop above the load voltage.

In effect, the negative feedback action of the circuit causes the output of the op-amp to move to a voltage that drives the error to zero, and to do this, it must compensate for the voltage drop across the diode.

Now consider operation when the input voltage is slightly below the output voltage. The error voltage is now negative, so the output of the op-amp begins to move in a negative direction. The diode becomes reverse biased, which makes it an open circuit. Now the op-amp is operating open-loop, the error voltage at the input is not corrected to zero, and the output of the op-amp stays at negative saturation throughout the remainder of the cycle.

The complete waveform of op-amp output is shown in figure 419, where the negative saturation voltage of the op-amp is indicated as $V_{a_{sat}}$.

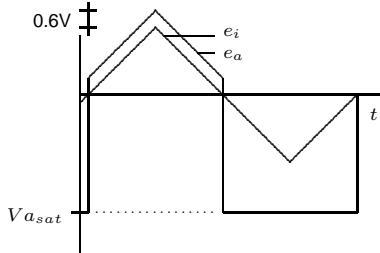


Figure 419: Op Amp Output Waveform

Slew Rate Limiting

As figure 419 indicates, the op-amp must slew to its negative saturation voltage and back, once each cycle of the input waveform. As we'll see in a subsequent section, real op-amps (as opposed to ideal ones) have a finite slew rate which limits how rapidly this can occur. For example, the 741 has a slew rate of 0.5 volts per microsecond, so a trip to the negative limit of -10 volts will consume 5 microseconds on the negative transition and 5 on the return trip. This 10 microsecond interval must be a small fraction of the total cycle time, or the waveform will be distorted. So the slew rate limits the upper frequency at which this circuit is fully functional.

16.2 Precision Half-Wave Rectifier

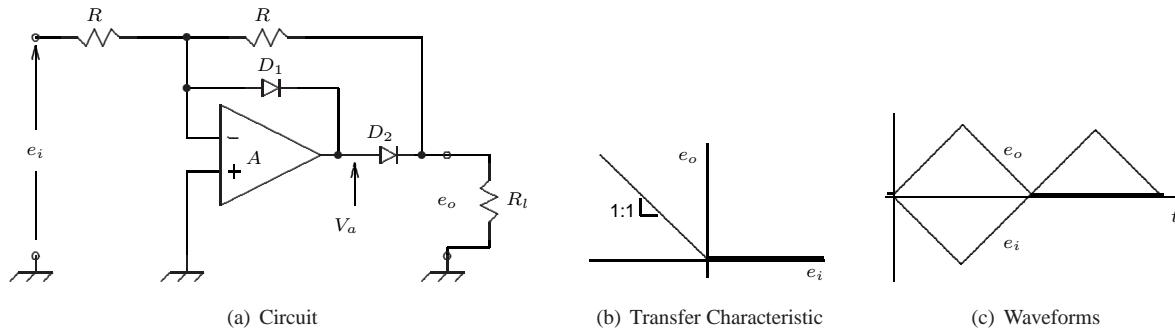


Figure 420: Precision Half Wave Rectifier

The precision diode of section 16.1 can be used in many applications that require an ideal diode. However, because of slew rate limitations of the op-amp, it does not work well at high frequencies.

Where the object is to produce a half-wave rectified waveform, the circuit of figure 420 is an improvement over the ideal diode.

- This circuit may be regarded as an inverting amplifier with the addition of diodes D_1 and D_2 .
- Diode D_2 serves the same purpose as the diode in the precision diode circuit of figure 418 on page 498. It ensures that the output follows the input while the output is positive.
- Diode D_1 prevents the op-amp from slewing to its negative limit when the output tries to go negative. While the input signal is positive, the output of the op-amp is driven negative. Diode D_2 becomes reverse biased, so the feedback resistor is disconnected from the summing junction, and there is no feedback via that path. So the output slews rapidly negative.

When the output of the op-amp reaches -0.6V, diode D_2 conducts, diverting current away from the inverting input terminal. Consequently, the output of the op-amp is limited at -0.6V during that part of the cycle. We say that diode D_2 clamps the output of the op-amp to -0.6 volts during that part of the cycle.

- Since the op-amp does not need to slew great distances during one-half of the waveform, this circuit will work properly to a much higher frequency than one using the ideal diode circuit of figure 418.
- The polarity of the rectification can be reversed by reversing both diodes.
- As shown in figure 420, the gain of the circuit is unity. As usual for the inverting amplifier circuit, the gain may be adjusted by changing the ratio of the two resistances.

The input waveform and the corresponding waveform at the output of the op-amp are shown in figure 421.

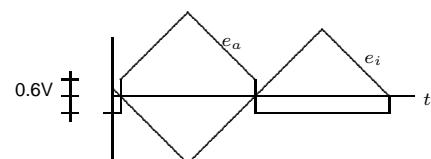


Figure 421: Op Amp Output Waveform

16.3 Precision Full-Wave Rectifier

There are many ingenious precision full-wave rectifier circuits. The block diagram for one of the most common is shown in figure 422.

The circuit half-wave rectifies the waveform, doubles the magnitude, and adds that to the original. The effect is to overpower the negative halves of the input signal, thereby effectively flipping their polarity. The positive halves of the input signal are unaffected.

A complete circuit of this type is shown in figure 423.

Amplifier A_1 and its associated circuitry are the precision half-wave rectifier. Amplifier A_2 functions as the adder. Notice that the gain is doubled for the rectified input to the adder by scaling that adder input resistor by one-half.

The adder inverts, so the actual output of the circuit of figure 423 will be an inverted version of waveform D in figure 422 above.

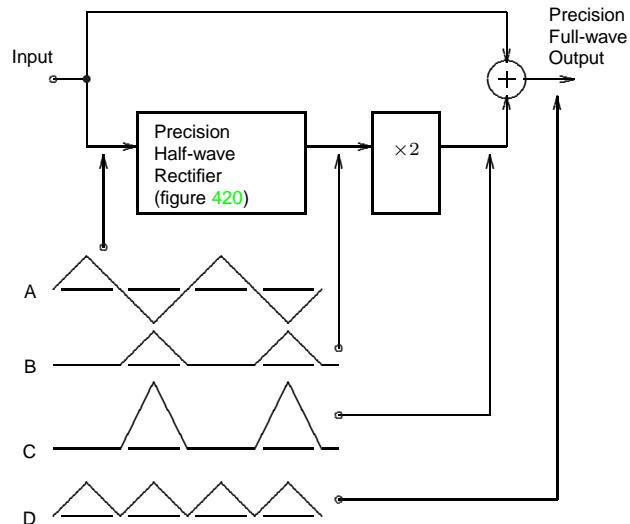


Figure 422: Precision Full Wave Rectifier, Block Diagram

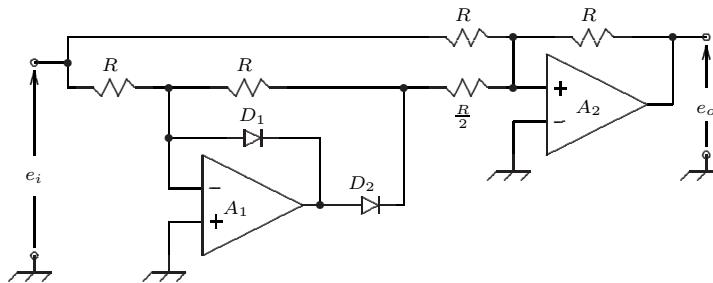


Figure 423: Precision Full Wave Rectifier, Circuit

Switchable-Amplifier Based Full Wave Rectifier

A third absolute-value amplifier circuit (aka full-wave rectifier), taken from [131], is shown in figure 424.

Amplifier A_1 and diode D form an ideal diode (section 16.1). By grounding the non-inverting terminal of A_1 , the cathode of the ideal diode is held at zero volts.

Amplifier A_2 and its three resistors form a switchable inverting-amplifier (section 16.1). In this case, the

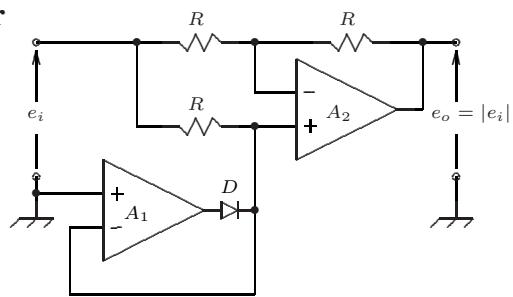


Figure 424: Full-Wave Rectifier #3

ideal diode becomes a type of switch: open circuit when the input voltage is positive, and short circuit when the input voltage is negative. This causes A_2 to be an inverting amplifier when the input voltage is negative, and a non-inverting amplifier when the input voltage is positive. Consequently, the output voltage is the absolute value of the input.

Amplifier A_1 must switch in synchronism with the input waveform. When the input voltage is negative, the output voltage of A_1 will be at +0.6 volt. When the input voltage is positive, the output voltage of A_1 will be at negative saturation. In the reference, amplifier A_1 is shown as powered from a positive power supply and ground (rather than a negative power supply) to limit the magnitude of the negative saturation voltage. Otherwise, the frequency response would be very limited. (The reference claims a bandwidth exceeding 1MHz when using the CA3140 op-amp.)

Alternative Full Wave Rectifier

Another precision full wave rectifier circuit, described in [132], is shown in figure 425. The first op-amp is basically an inverting amplifier. There are two feedback resistors, selected by diodes D_1 and D_2 .

When the input voltage is negative, D_2 conducts, generating a positive half-cycle of the waveform. When the input voltage is positive, D_1 conducts, generating a negative half-cycle of the waveform. The second op-amp amplifier inverts the negative half-cycle and adds it to the positive half cycle to generate a full-wave rectified waveform.

The output of A_1 is required to slew through two diode voltage drops at the beginning and end of each half-cycle. This is a relatively modest requirement, so the circuit will perform correctly over a wide frequency range.

An analysis of this circuit takes a bit of care but is illustrative. The general approach is to treat positive input voltages and negative input voltages as two separate cases. In both cases, we can assume that the inverting input to A_1 remains at virtual earth point, 0V.

Figure 426(a) shows the case when e_i is positive. Diode D_2 is an open circuit.

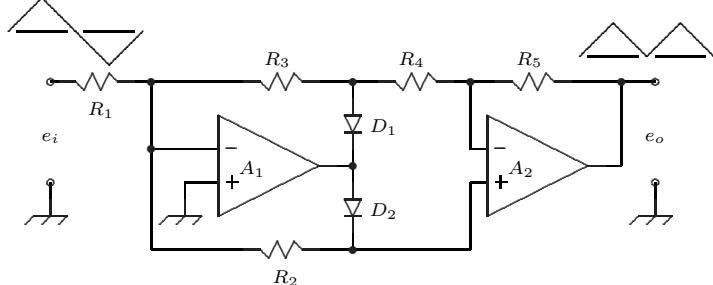


Figure 425: Alternative Full Wave Rectifier

$$V_a = -\frac{R_3}{R_1}e_i \quad (691)$$

There is no current through R_2 , so $V_b = 0$ volts and

$$\begin{aligned} e_o &= -\frac{R_5}{R_4}V_a \\ &= +\frac{R_3}{R_1}\frac{R_5}{R_4}e_i \\ &= e_i \end{aligned}$$

if all the resistors are equal.

Figure 426(b) shows the case when e_i is negative. Diode D_1 is an open circuit. This case is more complicated because a negative input voltage causes current through R_1 which then splits and flows through both the R_3, R_4 path and the R_2 path. In that case, we can use KVL to write:

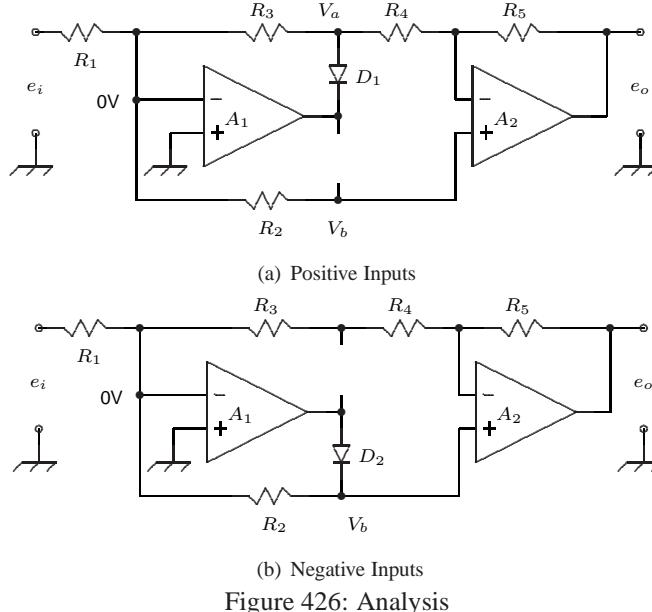


Figure 426: Analysis

$$\frac{e_i}{R_1} = \frac{V_b}{R_3 + R_4} + \frac{V_b}{R_2} \quad (692)$$

As well, we can relate V_b to the output voltage by considering the voltage at the inverting terminal of A_2 :

$$e_o = V_b + \frac{V_b}{R_3 + R_4} R_5 \quad (693)$$

Combining equations 692 and 693 to eliminate V_b , we obtain the charming

$$e_o = \frac{R_2(R_3 + R_4)}{R_2 + R_3 + R_4} \left(\frac{R_3 + R_4 + R_5}{R_3 + R_4} \right) \frac{e_i}{R_1} \quad (694)$$

For equal resistors, this collapses to $e_o = e_i$.

Since both polarities of e_i cause a positive value of e_o , then the circuit is a full-wave rectifier, or absolute value circuit.

Like the Howland Circuit of section 18.3, it is not too difficult to prove that the circuit works. It is much more difficult to imagine how it was conceived in the first place. Circuits like this one are far from obvious.

References

There are many other precision full-wave rectifier circuits. Chapter 5 of [133] is entirely devoted to these circuits.

16.4 Full-Wave Meter Driver

The circuits of figure 427 are useful when a moving-coil meter must be driven by an AC signal source. This is commonly required where an audio signal must drive a moving-coil VU meter. The VU meter is typically a 1mA meter movement, which may or may not have a bridge rectifier included.

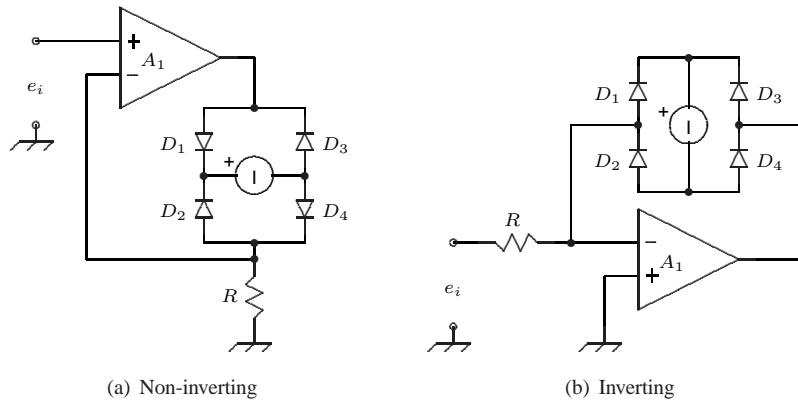


Figure 427: Full Wave Meter Driver

The first circuit of figure 427 is the non-inverting version. This may be viewed as a current-source circuit similar to figure 298 on page 358, where the load resistance is the meter movement and bridge rectifier. An input voltage of e_i volts will drive a current of e_i/R amps through the bridge diodes and meter movement. When e_i is positive, the op-amp drives current through D_1 and D_4 . When e_i is negative, the op-amp drives current through D_2 and D_3 . In both cases, current flows into the positive terminal of the meter movement. Consequently, the waveform of current in the meter movement is full-wave rectified.

The diode voltage drops have no effect on the meter current because they and the meter movement are driven by a current source. The output of the op-amp adjusts itself to produce the correct current.

The inverting version of this circuit is essentially the same as the non-inverting version. The meter and rectifier diodes are in the feedback path, so any input current must flow through them, and the output of the op-amp will adjust to make that occur. Again, the meter current is equal to e_i/R amps.

Filtering

The current through the meter movement is a full-wave rectified version of the input current, equal to e_i/R amps. This has a large ripple component unless there is some filtering. In practice, there is no need for a filter capacitor in this circuit. The mechanical inertia of the meter movement will average out the reading, as long as the input signal is higher in frequency than the mechanical resonant frequency of the meter movement. This is normally the case with a moving meter movement and audio signals.

Consequently, the meter movement will respond to the *rectified average* value of the input waveform.

16.5 Phase-Sensitive Rectifier

A *phase-sensitive rectifier* produces an output voltage that is proportional to the magnitude of the input AC signal (like most rectifiers), but also proportional to the *phase angle* of the AC signal.

Phase angle has to be defined relative to some reference, so some other AC signal of the same frequency must be available to act as the reference.

Various devices can serve as the heart of the PSR, but we'll focus on the switchable sign-changer of section

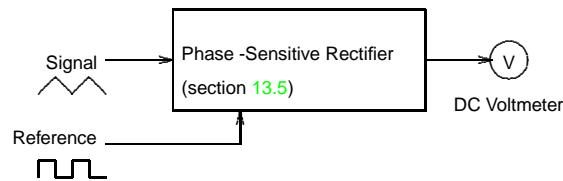


Figure 428: Phase-Sensitive Rectifier

13.5.

Consider that the input to the switchable sign-changer is a triangle wave. The polarity of the sign changer is controlled by a square wave, which serves as the reference signal. When the reference signal is positive, the circuit is non-inverting with a gain of +1. When the reference signal is negative, the circuit is inverting with a gain of -1.

The output of the sign-changer is monitored by a DC voltmeter, which displays the average voltage of the output waveform.

The waveforms for various phase shifts between the signal and reference are shown in figure 429.

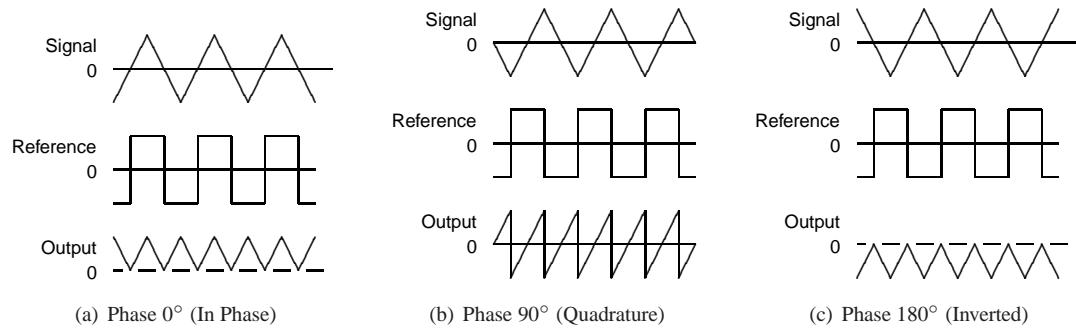


Figure 429: Phase Sensitive Rectifier, Waveforms

In Phase When the signal and reference are in phase, the output is a full-wave rectified waveform. This has some positive average voltage, so it would read as a positive voltage on the voltmeter.

Quadrature When the signal and reference are in quadrature, the output is an AC waveform which is symmetrical about the zero axis, so it has an average voltage of zero.

Inverted When the signal and reference are inverted, the output is a full-wave rectified waveform of negative polarity. This has some negative average voltage and would read as a negative voltage on the voltmeter.

Consequently, the average output voltage is some function of the phase angle between the reference and signal waveforms. This circuit could be used as a measurement system for phase angle. The precise relationship between output voltage and relative phase angle depends on the shape of the signal waveform.

The phase-sensitive rectifier is really a special case of a multiplier, where the signal is being multiplied by a square wave of amplitude ± 1 . We'll discuss this in greater detail in section 36.1.

16.6 RMS Conversion

An *rms converter* is a circuit that produces an output voltage proportional to the effective or root-mean-square value of the input waveform. In figure 430, some arbitrary waveform of voltage is generating power in load resistor R . The output of the RMS converter is proportional to the magnitude of the power. If the power in the load must be known, it is simply

$$P = \frac{V_{RMS}^2}{R} \quad (695)$$

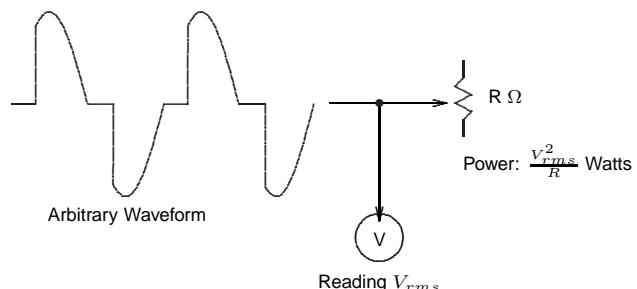


Figure 430: RMS Conversion

16.6.1 Obvious Implementation

The obvious implementation is shown in figure 431.

The squarer can be implemented using a multiplier with both inputs tied together, or with a special purpose squaring device such as a JFET [134]. Similarly, the square root circuit can be implemented with a squaring device in the feedback circuit of an op-amp. An integrator would be an ideal method of calculating the average, but is difficult to implement as an open-loop system. So integration is usually approximated by a single-pole lowpass filter.

However, as Sheingold points out in [132] and Gilbert in [135], this circuit is limited to a dynamic range of about 10:1. That is, for a maximum input voltage of 1 volt the minimum allowable voltage for satisfactory accuracy would be about 100mV. As the input signal decreases in amplitude by some factor Y , the output of the squarer decreases by Y^2 . For small input signals the output signal of the squarer becomes comparable to the squarer output offset voltage, and the accuracy becomes unacceptable.

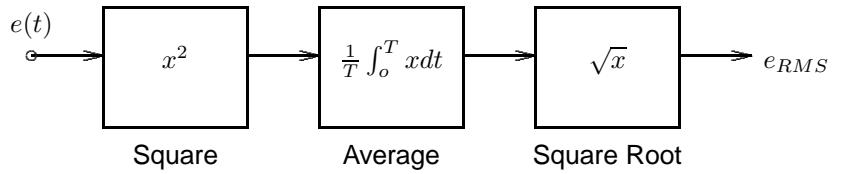


Figure 431: RMS Conversion Circuit

16.6.2 The Implicit Method

It turns out that it is relatively straightforward using so-called *translinear techniques* to generate the function

$$Z = \frac{X^2}{Y}$$

where X and Y are inputs to the module and Z is the output. Such devices are available as an integrated circuit *multiproportional module* or may be constructed from a quad of matched transistors [136]. This gives us another way to implement the RMS converter: by *implicit* computation. As in the analog computing methods described in sections 17.12 and 37.1, we assume that the output signal is available, and then use it to compute the output signal. In the case of RMS conversion,

$$V_{RMS} = \sqrt{\overline{e_i^2}} \quad (696)$$

where the overbar indicates *average*. Then:

$$V_{RMS}^2 = \overline{e_i^2}, \text{ and} \quad (697)$$

$$V_{RMS} = \frac{\overline{e_i^2}}{V_{RMS}} \quad (698)$$

Equation 698 implies the block diagram shown in figure 432.

This is an interesting concept, because it reduces the dynamic range at the output of the multifunctional module. As the input signal reduces in amplitude, the computed value of V_{RMS} also decreases. As it does so, the Y value into the multifunctional module decreases and, because Y appears in the denominator, this effectively increases the gain of the multifunctional module.

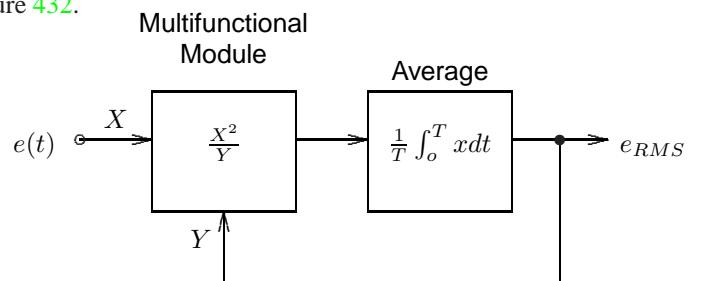


Figure 432: RMS Conversion, Implicit Method

Consequently, the output of the multifunctional module must vary over approximately the same range as the input voltage, rather than its square.

Using this technique, Gilbert [135] reports an error of only 0.04% over a dynamic range of 40:1. Cate and Handler [137] show a complete circuit of an implicit RMS conversion circuit.

16.7 Thermal Method

The thermal method of RMS measurement is based on the definition: the RMS value of a waveform is the value of the DC voltage (or current) that would cause the same heating effect in a resistor. In this method, the unknown waveform is caused to heat a resistor R_1 . A feedback system applies DC voltage to an equal resistance R_2 until both resistors are at the same temperature. Then the DC voltage is equal to the RMS value of the unknown.

In figure 433, adapted from [138], the temperature sensing function is carried out by two semiconductor diodes D_1 and D_2 , by virtue of the silicon diode temperature coefficient of $-2.2\text{mV}^{\circ}\text{C}$. Resistor R_1 converts the incoming waveform into heat, which is sensed by D_1 . Amplifier A drives resistor R_2 until diode D_2 is at the same temperature as D_1 .

Resistor R_1 and diode D_1 are thermally coupled, as are resistor R_2 and diode D_2 . These assemblies are then thermally isolated from each other. As a result of this arrangement, the ambient temperature affects each assembly equally, and so has no effect on the measurement.

Unlike the direct and implicit methods described above which are limited by the frequency response of their op-amps, this method can be applied to waveforms well into the RF region, some 10's of MHz. The dynamic range for a commercially available device, the LT1088 available from Linear Technology, is given as 10:1. Consequently, any system that must measure a wider range of amplitudes will require a variable gain preamplifier and/or attenuator to condition the signal into this device.

16.8 Peak Detector

A *peak-detector* circuit stores an output voltage equal to the peak value of the input waveform. A simple peak detector circuit is shown in figure 434.

The op-amp circuit is a precision rectifier. The capacitor C_m behaves as an analog memory device, storing the largest positive voltage appearing at the input.

The circuit of figure 434 implies that the capacitor is connected to an open circuit. Consequently, the capacitor will never discharge, which is not a practical arrangement. In practice, the capacitor might be connected to a load resistance, which gradually discharges the capacitor after a peak voltage has been stored. If the capacitor voltage is read (by a microprocessor A/D converter, for example) in a time interval that is short compared to the discharging time constant, then the error is minimal and the resistor provides a simple method of resetting the capacitor voltage.

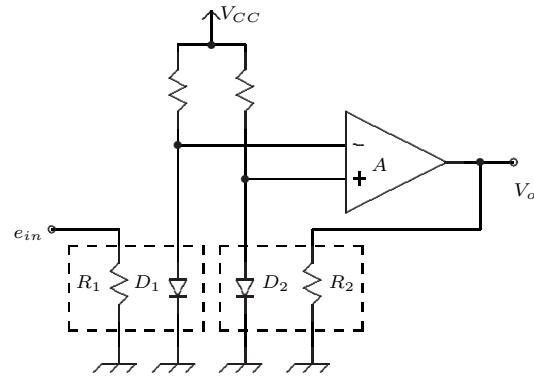


Figure 433: RMS Conversion, Thermal Method

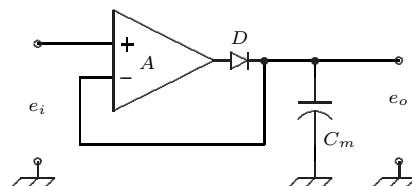


Figure 434: Simple Peak Detector

Alternatively, there may be an electronic *reset* switch across the capacitor. The microprocessor would read the capacitor voltage and then short-circuit the capacitor for a brief interval to discharge it back to zero.

The Capacitor

The magnitude of the capacitance is a compromise. On the one hand, it should be small enough that the op-amp can charge it quickly. To work correctly, the op-amp output current must be able to charge the capacitor up to the peak input voltage in a time equal to the duration of this voltage. A smaller capacitance will charge more quickly for a given charging current.

On the other hand, depending on the load resistance, the capacitor should be large so that the discharge time constant is lengthened. However, if the load resistance is buffered by an op-amp with a high input impedance and low bias currents (JFET or MOSFET inputs), then the capacitor can be smaller in size for the same discharge time-constant.

The capacitor *type* is also important. Certain capacitors exhibit the phenomenon of *soakage* or *dielectric absorption* [139]. Once the capacitor is charged, the capacitor dielectric absorbs some of the charge on the capacitor, and this then reduces the magnitude of the stored voltage. This effect can also be demonstrated in reverse: a charged capacitor is shorted, to discharge it, and then open circuited and connected to a voltmeter. The capacitor then appears to slowly recharge itself back up to a small voltage, as the charge is released by the dielectric.

These effects are small, but if high accuracy is required, then they must be taken into consideration. Reference [140] indicates that a ceramic capacitor may exhibit as much as a 1% hysteresis effect due to soakage. Polypropylene and polystyrene capacitors exhibit much lower hysteresis and are preferred in this application.

16.8.1 Improved Peak Detector

The circuit of figure 434 suffers from a major disadvantage. Whenever the input voltage drops below the capacitor voltage, the op-amp slews to its negative saturation voltage. To record a positive peak that exceeds the existing capacitor voltage, the output of the op-amp must slew from its negative saturation voltage to the new voltage. This limits the ability of the op-amp to capture short transient pulses.

The circuit of figure 435 deals with that problem by limiting the output of the op-amp to -0.6 volts. First, an overview:

- Whenever the input voltage exceeds the voltage on memory capacitor C_m , amplifier A_1 drives the voltage upward until the two are equal.
- Amplifier A_2 is simply a unity-gain buffer. Its purpose in life is to provide the capacitor with a load that is close to an open-circuit, and the feedback signal (via R) with a low impedance drive point. Consequently, the feedback voltage e_f is equal to the voltage e_m stored in the memory capacitor.
- Amplifier A_1 does the comparison between e_f (ie, the memory voltage e_m) and the input voltage and drives the capacitor voltage appropriately.
- Diode D_1 restricts the excursion of A_1 when the circuit is in *hold* mode. Diode D_2 provides the switching function between *hold* and *track* mode.

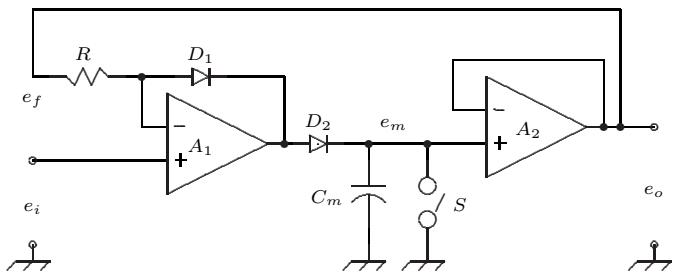


Figure 435: Improved Peak Detector

- Switch S is operated periodically to reset the capacitor voltage back to zero (by a nearby computer, for example). Or the switch could be replaced by a large-value resistor so that the capacitor slowly discharges after each peak.

Now let's look a little more closely at the operation, first when the input voltage is less than the voltage on the memory capacitor.

Input voltage less than memory voltage

For this situation, the feedback voltage e_f is *greater* than the input voltage e_i . This creates a net negative error voltage above the input terminals of the op-amp, so its output takes off in a negative direction. Because the non-inverting terminal of A_1 is held at e_i volts, the inverting terminal will also ride there. This process will terminate when the output voltage of A_1 is one diode drop below e_i . Then diode D_2 is reverse biased (ie, an open circuit) and the capacitor voltage does not change.

Input voltage greater than memory voltage

In this case, the feedback voltage e_f is *less* than the input voltage. This creates a net positive error voltage across the input terminals of the op-amp, so its output takes off in a positive direction. As the output moves positive, it first reverse biases D_1 , which becomes an open circuit. At some point, the output of the op-amp reaches a voltage equal to one diode drop across the current memory voltage e_m , and diode D_2 starts to conduct. The capacitor charges, raising the memory voltage e_m and the feedback voltage e_f until the net input voltage into A_1 becomes very slightly negative.

How far negative? Enough to ensure that diode D_2 is no longer conducting. But the voltage across D_2 need only be slightly reversed to stop it conducting, and because amplifier A_1 has very large voltage gain, the memory voltage on the capacitor is only very slightly larger than the input voltage.

References

Franco [139] contains a very clear description of the operation of the improved peak detector circuit. There is also useful information in [141]. Guinta [142] has an excellent description of the different types of capacitors, their advantages and disadvantages.

16.9 Sample-Hold Circuit

The sample-hold circuit is a memory device for an analog voltage. The simplest possible sample-hold circuit is shown in figure 436(a). During the *sample* interval, the switch is closed and the voltage on the memory capacitor C_m tracks the input voltage. During the *hold* interval, the switch is opened. The capacitor voltage remains at the value it had when the switch opened. An example of sample-hold waveforms is shown in figure 436(b).

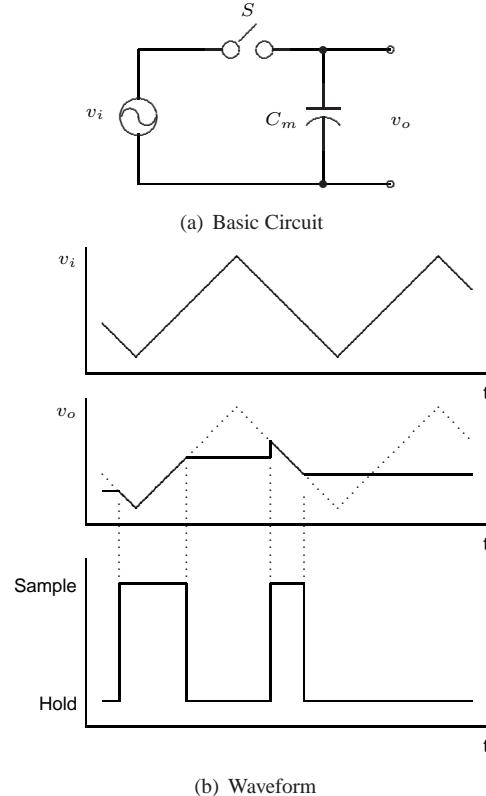


Figure 436: Basic Sample Hold

There are two limitations of the basic circuit shown in figure 436(a).

- The input voltage will generally have some internal resistance. This resistance limits the charging rate of the capacitor and consequently limits the speed with which the capacitor voltage can move to a new value during *sample* mode.
- Any load resistance across the capacitor will discharge it during the *hold* interval.

Alternative sample-hold circuits that address these issues are shown in figure 437.

In the open-loop configuration of figure 437(a), amplifier A_1 provides a low-impedance drive source for rapidly charging the capacitor during *sample* mode, without loading the input source v_i . Amplifier A_2 presents a high-impedance load to the capacitor so that it is not discharged by a load resistance during *hold* mode.

Amplifier A_2 would be a JFET or MOSFET input device for minimum bias current. Because both amplifiers operate with local feedback loops, they tend to have relatively predictable dynamics under transient conditions. However, the offset voltages of A_1 and A_2 are additive: they both contribute to the accuracy of the circuit.

In the closed-loop configuration of figure 437(b), the feedback encloses both amplifiers. During *sample* mode, amplifier A_1 will drive the capacitor voltage to whatever voltage is necessary to equalise its two input voltages. The offset voltage of amplifier A_2 is irrelevant. As a result, it is only the offset voltage amplifier of amplifier A_1 that affects the accuracy of the circuit. However, because the feedback loop encloses two amplifiers, the circuit is more likely to experience overshoot and ringing during *sample* mode.

Furthermore, when the switch is open in *hold* mode, there is no feedback around amplifier A_1 and so it will saturate at one of its power-supply bounds. This can be prevented by adding additional switches that cause the configuration to revert to figure 437(a) when in *hold* mode.

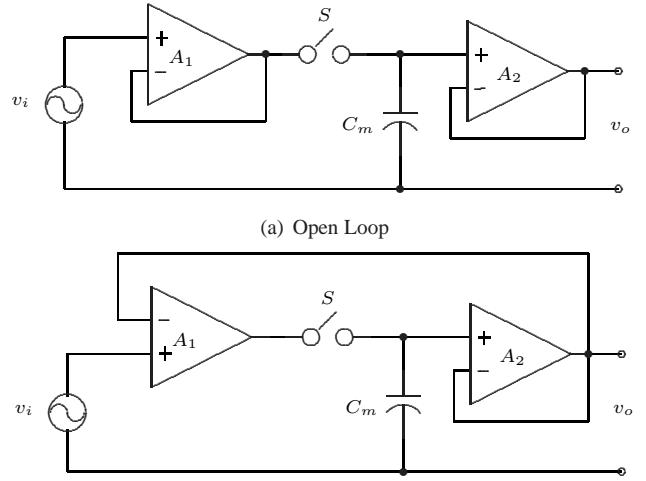


Figure 437: Sample Hold Circuits

Application: Analog-Digital Converter

Some A/D converters such as the dual-slope device (section 26.6) average an input voltage over a period of time. If the input voltage changes, then the reading will be the time-average of the input over the measurement interval.

Other A/D converters such as the successive-approximation converter (section 26.4) require that the input voltage be constant during the measurement interval. If the input voltage may change during the measurement, a *sample-hold* circuit must precede the A/D converter to hold the A/D input voltage constant.

Many A/D converters have a sample-hold circuit built in or the architecture of the converter may be such that a sample-hold is inherent in the device. For example, the *charge-redistribution* technique of successive approximation conversion uses storage capacitors, rather than resistors, and so inherently contains an analog memory that makes a sample-hold unnecessary. The A/D converter in the 68HC11 microprocessor uses this technique, described in [143].

Sample-Hold and Track-Hold

These terms tend to be used interchangeably. However, strictly speaking, a sample-hold device takes a brief sample of an input signal and spends most of its time in the *hold* mode. A *track-hold* can spend considerable

lengths of time tracking the input signal, or can spend significant time in the hold mode. So they are similar devices, but the sample-hold circuit probably uses a brief sampling pulse.

Dielectric Absorption

As mentioned in section 16.8, for accuracy it is important that the phenomenon of *dielectric absorption* not occur in the memory capacitor. For that reason, capacitors using a polypropylene or polystyrene film dielectric are preferable.

Sample-Hold Specifications

Where high accuracy and/or high speed are required, there are a number of aspects of sample-hold behaviour that should be scrutinized [144]. These are in addition to the usual performance metrics of an analog circuit, such as small-signal bandwidth and offset voltage.

Droop Rate The rate at which the output voltage drifts from its original value during *hold* mode. Assuming that this drift is caused by leakage of current, droop rate can be decreased by increasing the size of the memory capacitance. However, this may impinge on the rate at which the capacitor can be charged during *sample* mode. Consequently, the size of the memory capacitor is a compromise.

Aperture Delay The delay time between the time the *hold* command is given and the sample-hold switch is completely open.

Aperture Jitter is the uncertainty of the aperture delay. The jitter sets an upper limit on the frequency that can be sampled. If the following circuit is an analog-digital converter, and the input waveform is changing during the sample time, then the aperture jitter will cause a noise-like variation of voltage into the A/D converter. Consequently, the interaction of the aperture jitter and the rate of change of the input signal should be less than one LSB of the A/D converter.

For a sinusoidal waveform input, the aperture jitter sets an upper limit on the frequency that can be digitized.

Hold Step When the sample-hold makes a transition from *sample* to *hold* mode, the switching signal may cause a voltage step to occur in the output. This represents an error term in the output voltage. It is particularly problematic when the hold step is proportional to the input voltage, because the error is then a distortion of the signal.

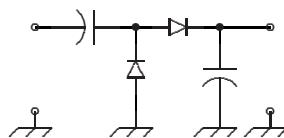
Hold Mode Settling Time The time required for the output voltage to settle to within a specified error band, from the time of the hold command.

Reference

A variety of sample-hold circuits are described in Chapter 8 of Johns and Martin [79].

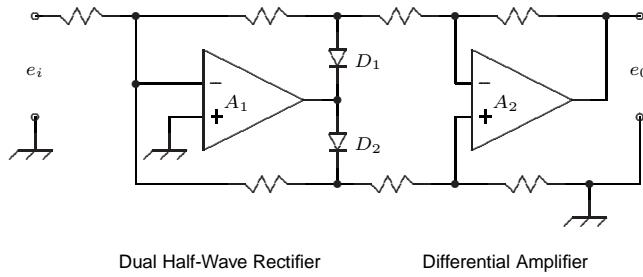
16.10 Exercises

1. The circuit shown below is a *voltage doubler* for AC signals. For example, if the AC input signal is E volts peak, the output is approximately $2E$ volts DC.



There are some losses in this circuit due to the fixed voltage drop across the diodes. As a consequence, it is not very accurate for small input voltages.

- (a) Show how the circuit could be modified to use *precision diodes* in place of the ordinary diodes [145].
 - (b) Could single supply op-amps be used for the precision diodes in this circuit?
 - (c) Why or why not?
2. The circuit shown below has been proposed as a full-wave precision rectifier. The concept is this: the first op-amp generates two half-wave rectified waveforms. One is the positive half of the waveform, the other is the negative half. These are then combined by the second op-amp, which acts as a differential amplifier. All the resistors may be assumed to be equal.



- (a) Show that this reasoning is fallacious, and the circuit will not work as planned.
- (b) The design can be fixed by putting two unity-gain buffers into the circuit. Show the modified circuit and verify that it works correctly.

17 Active Filters

17.1 Introduction

A *filter* (also known as a *wave filter*) processes a signal with a frequency dependent amplitude and/or phase response. Certain frequencies are transformed in some manner. Some frequencies may be emphasised, while others are diminished. Or certain frequencies may obtain a particular phase shift, while other frequencies are unaffected.

Figure 438 shows the effect of a *lowpass filter* on two sine waves of different frequencies. The lower frequency waveform is largely unaffected. The higher frequency waveform is reduced in amplitude. (In practice, it would also be shifted in phase compared to the original.)

In general, this filtering action may be accomplished by analog or digital means. In an analog circuit, some combination of circuit elements processes the signal. Frequency selectivity may be obtained by a combination of passive components – the inductor and capacitor – or with the help of an amplifier of some sort. In any case, there must be some sort of energy storage element that can delay a signal, such as a resistor/capacitor combination that can effect a phase shift on an AC signal.

When filtering is to be accomplished by digital means, the signal must first be converted from analog to digital form by an A/D converter, then the resulting stream of numbers filtered in some way, and then possibly converted back to analog form by a D/A converter. The filtering action is accomplished by various numerical memory units that can delay additions, subtractions and multiplications to obtain a frequency sensitive characteristic.

While it is possible to construct very elaborate filters using analog means, this is a specialty skill and beyond the scope of this section. Furthermore, a very complex filter may be better implemented using digital signal processing, rather than an analog hardware circuit. As a consequence, most analog active filters will be relatively simple, and that will be our focus.

17.1.1 Applications

Active filters are widely used in audio applications for equalization, tone control, active crossover networks, and noise removal from old recordings [146].

They are used in communication receivers to control the width of the audio spectrum, thereby improving the signal-noise ratio. (This is particularly useful in the reception of morse code, which is a pulse-modulated audio tone that can be processed through a very narrow bandwidth.)

In more specialized applications, active filters are used to select a particular segment of the spectrum after a modulation or demodulation process. For example, [147] uses this technique to effectively synthesize a narrow bandwidth filter that can be swept over the audio band.

The literature on active filters is vast and often shrouded in mathematical complexities. Entire textbooks [148], [149] are devoted to the subject, and there are thousands of literature references to different filter designs. To deal with this complexity and diversity, we will limit ourselves to a few active filter types and designs. The choice of these designs will be determined entirely by their usefulness in practical circuit designs.

There are a number of dimensions to the universe of filter designs.

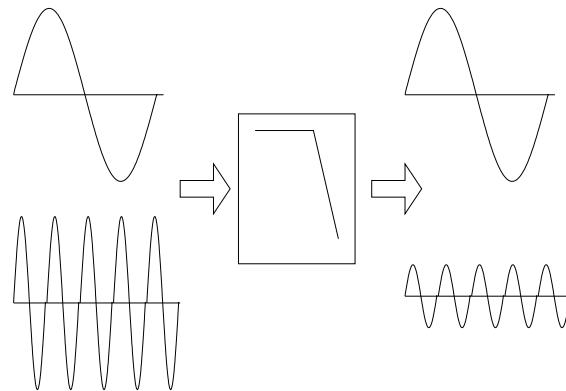


Figure 438: A Lowpass Filter

17.1.2 Function

There are four major functional types: lowpass, highpass, bandpass or band reject¹⁷⁰. Figure 439 shows the amplitude response $|e_o/e_i|$ of the various filter types.

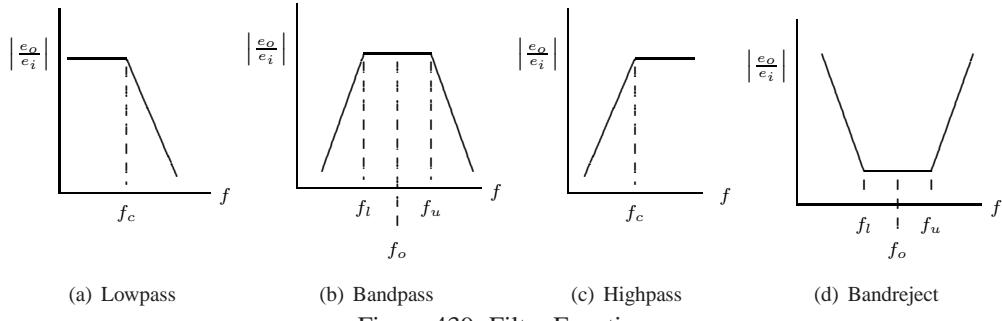


Figure 439: Filter Functions

The region where the filter passes a signal is the *pass band*. The region where the signal is rejected is the *stop band*. The region in between is the *transition band*. For the lowpass and highpass filters, the band edge is defined by the *cutoff frequency* f_c . For the bandpass and bandreject filters, the band edges are defined by the *lower cutoff frequency* f_l and the *upper cutoff frequency* f_u . The *centre frequency* f_o is the geometric mean of the upper and lower cutoff frequencies:

$$f_o = \sqrt{f_l \cdot f_u} \quad (699)$$

The choice of filter function is driven by a knowledge of the spectrum of the signal, what frequencies in that spectrum are to be passed, and which are removed. Even within a particular function there are choices. For example, a bandpass filter can be created by a resonator circuit (the passive circuit shown in figure 452 for example, or its active equivalent) or by means of a cascade of highpass and lowpass filters, with their cutoff frequencies staggered to create a pass band.

17.1.3 Symbols

	Lowpass	Bandpass	Highpass	Bandreject
American				
European				

The IEEE standard for electronic circuit symbols [150] includes symbols for a wide variety of devices – ranging from *Telegraph Key* to *Hydroelectric Generating Station* – but not for filters. The filter symbols in common usage in North America and Europe are tabulated above.

¹⁷⁰A narrow band reject filter is often known as a *notch* filter. Some authors include the all-pass filter in this list. I've treated it separately in its own section on phase shifters, section 19.

The North American versions contain a stylized representation of the amplitude response and are the ones that will be used in this text.

17.1.4 Order

The filter *order* is a measure of the complexity of the filter. Increasing the order of a filter results in a steeper transition between pass and stop band. As a consequence, a higher order filter is more *selective*, that is, it is more effective at separating closely spaced frequencies.

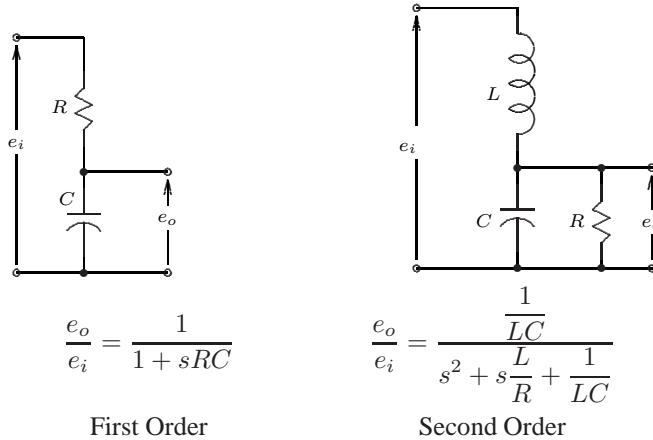


Figure 440: First and Second Order Lowpass Filters

Put another way, if we want to pass 100 Hz and block 10kHz, a relatively simple, low-order filter will suffice. If we want to pass 100Hz and block 200Hz, then the filter must be more complex and of a higher order.

Two lowpass filters are shown in figure 440. We'll have more to say about these filters later. For the moment, accept and consider the transfer functions shown next to each filter.

For the first-order filter, the denominator is a first-order expression in s . For the second-order, the denominator is a second-order expression in s . *The order of the filter is thus a function of the denominator polynomial in s .* It is also equal to the number of energy storage elements in the filter. In the first-order filter, the one energy storage element is the capacitor. In the second-order filter, both the inductor and capacitor store energy.

A high-order filter can be constructed with a series of first and/or second-order sections. For example, a cascade of one first-order section followed by two second-order sections would be a 5th order filter.

The order of the filter determines the ultimate rapidity with which the filter makes its transition from passband to stopband. For example, a lowpass filter rolls off at $n \times 20$ decibels/decade, where n is the order of the filter. A fifth order filter would roll off at 100 db per decade. (The word 'ultimate' is important, because a variety of characteristics can be chosen for the transition region between passband and stopband.)

17.1.5 Shape

When 2nd-order sections are cascaded to make up a higher order filter, the Q factors and the cutoff frequencies of these sections may be chosen to achieve a specific result. An example is shown in figure 441.

Both these characteristics are for 4th-order lowpass filters. Both ultimately roll off at 80db/decade. However, the Chebyschev response (section 17.19 on page 557) accepts a certain amount of ripple in the passband in return for a steeper transition at cutoff.

The filter shape is chosen on the basis of factors such as

- the desired rate of change in the transition region
- allowable ripple in the filter passband and stopband
- phase-frequency response
- transient response

For example, some applications require a filter transient response that has very little overshoot and ringing, in which case the Bessel response (section 17.20 on page 564) may be a good choice.

17.1.6 Sensitivity

A filter design must meet its requirements in spite of component tolerances and drift. When a component in a filter changes value, this change has some effect on the performance of the filter. It may alter the Q factor, the DC gain, the cutoff frequency, or all of these and other properties as well. Obviously, it is desirable that this change in value have a minimal effect on the filter properties. Furthermore, if the filter is to be manufactured in production quantities, it must meet its performance specification in the face of component tolerances.

In both cases, we should have some idea of the effect of component variations on the behaviour of the filter. This concept is expressed as the *sensitivity* of the filter to some component variation.

From [78], the sensitivity of some property of the filter y to some component value x is written as:

$$S_x^y \approx \frac{\Delta y/y}{\Delta x/x} \quad (700)$$

For example, if

$$S_{R1}^{\omega_o} = 1/2$$

then a 5% tolerance in the value of R_1 could be expected to change the cutoff frequency ω_o by 2.5%.

In the past, certain filter configurations have been proposed that have unacceptable sensitivities and as a result were not usable in practice.

For example, if

$$S_{R1}^Q = Q$$

then a filter with a Q factor of 10 and tolerance in R_1 of 5% could be expected to have a variation in Q of 50%! Obviously, this is not a usable circuit.

The approach here will be to suggest filter configurations that are likely to have satisfactory sensitivities. However, for any critical design or for production, the sensitivity of the filter should be checked more carefully.

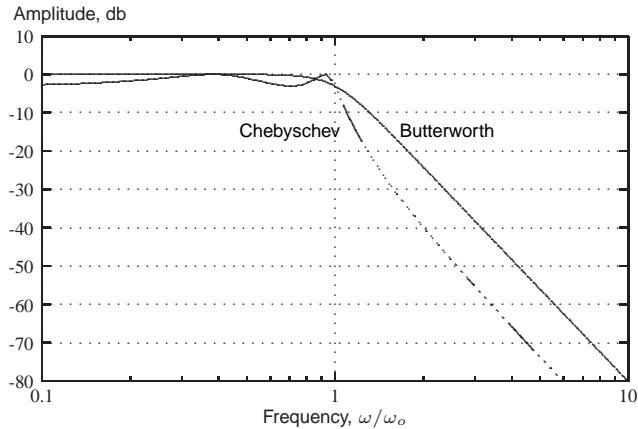


Figure 441: Chebyschev and Butterworth Characteristics

17.1.7 Implementation

The filter *implementation* is the specific circuit that creates the desired response. For example, a second-order lowpass filter can be implemented with a circuit using *passive* components such as the LRC circuit shown in figure 442(a). The same filter can be implemented with the operational amplifier *active* filter circuit shown in figure 442(b).

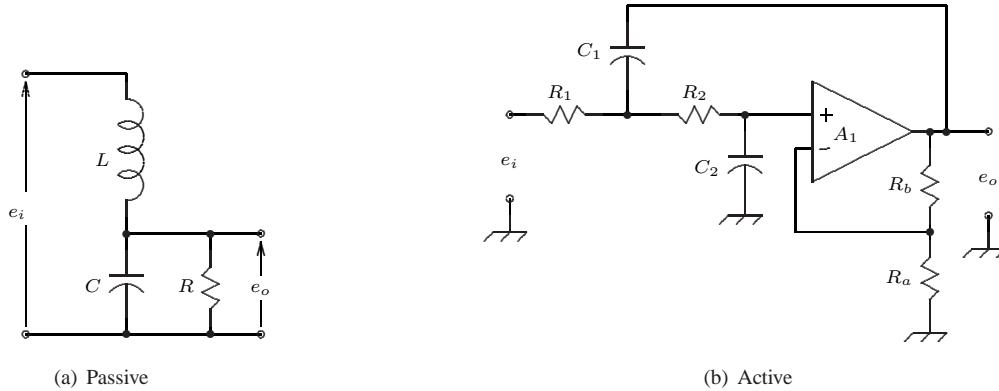


Figure 442: Filter Implementations

A filter composed of passive devices is appropriate at radio frequencies because these frequencies are beyond the capability of op-amps. At audio frequencies, inductors generally become too large and expensive to be practical and an op-amp active filter is a better choice.

The Sallen-Key op-amp filter makes an excellent lowpass or highpass filter because it is easy to design and relatively stable in the face of component variations. (It has *low sensitivity* to component tolerances and drift.) On the other hand, though it is possible to build a Sallen-Key bandpass filter, the *state-variable* filter is a better choice because it is easier to design and adjust and also has low component sensitivities.

These types of observations drove the selection of circuits presented in this section.

17.1.8 Roadmap to the Section

A brief roadmap to the section on active filters may help illuminate the relationship of the various sections.

1. The first sections, 17.4 through 17.7 develop the equations for the lowpass, bandpass, highpass and notch second-order filters, in general terms and without reference to any specific circuit. These equations can be implemented by a passive RLC circuit or by some operational amplifier circuit using only resistors and capacitors. These functions are the building blocks for filter designs and may be used in various combinations in filters higher than second order. Strictly speaking, the allpass filter belongs with these filters. However, its function is to change phase without affecting amplitude, so its application domain is different and it is treated separately in section 19.
2. Section 17.7 shows passive and active circuits that can implement the notch function.
3. Section 17.9 shows the *Bronzite* lowpass and highpass filter circuit. It's the first circuit studied in this section because it is relatively simple to analyse.
4. Sections 17.10 and 17.11 show circuits for the Sallen-Key lowpass and highpass filters, which can be used to implement the second-order lowpass and highpass sections. These are perhaps the most commonly used lowpass and highpass circuit designs.

5. Section 17.12 shows how the second-order bandpass filter may be implemented with a *state-variable* filter. There are other possible circuits, but the state-variable has properties that make it a preferred choice in many designs.
6. In sections 17.16 through 17.24, we introduce higher-order filters, show possible characteristics, and provide tables of coefficients for the designs, and give a design example. These higher-order filters are usually lowpass or highpass filters, so the discussion is restricted to those configurations.

17.2 First-Order Filters

A first-order filter may be implemented using *passive* components like the inductor, capacitor and resistor that have no energy sources. Or the filter may use *active* components such as transistors or operational-amplifiers, that have energy sources – in combination with passive components.

Active filters have the advantage that they can compensate for losses in any passive components. They can also pump energy into the filter, which allows for the elimination of inductors. On the other hand, active filters need a power supply.

First, we discuss passive filters.

17.2.1 First-Order Passive Filters

First-order passive RC and RL filters are summarized in figure 443 (section 4.8, page 169).

Example

A certain audio system is amplifying signals in the range 300Hz to 3000Hz (voice quality). A highpass RC filter is to be placed at the input of this amplifier to reduce an interfering signal of 120Hz.

1. If a suitable value of resistance is 1000Ω , calculate the value of capacitance so that the corner frequency f_c is at 300Hz.
2. Determine the attenuation of the interfering signal with respect to the audio signal.

Solution

1. The corner frequency is given by equation 189 on page 171.

$$\omega_c = \frac{1}{RC}$$

Recall that $\omega = 2\pi f$, where f is the frequency in Hz. Then

$$C = \frac{1}{\omega_c R} = \frac{1}{2\pi f_c R} = \frac{1}{2 \times \pi \times 300 \times 1000} = 0.53 \times 10^{-6} F = 0.53 \mu F$$

2. According to the table on page 519, the transfer characteristic of the RC highpass filter is given by

$$\frac{e_o}{e_i} = \frac{\frac{s}{\omega_c}}{1 + \frac{s}{\omega_c}}$$

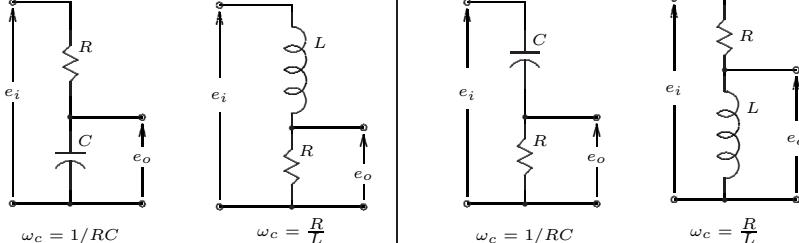
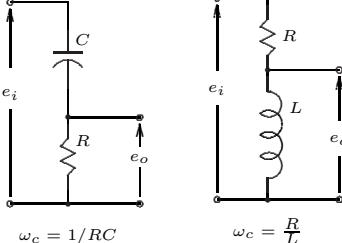
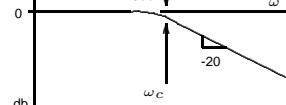
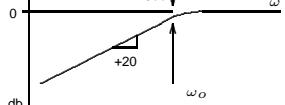
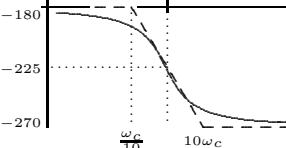
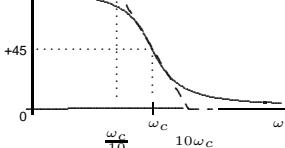
	Lowpass	Highpass
Circuit	 <p style="text-align: center;">$\omega_c = 1/RC$ $\omega_c = \frac{R}{L}$</p>	 <p style="text-align: center;">$\omega_c = 1/RC$ $\omega_c = \frac{R}{L}$</p>
Equation	$\frac{e_o}{e_i} = \frac{1}{1 + \frac{s}{\omega_c}}$	$\frac{e_o}{e_i} = \frac{\frac{s}{\omega_c}}{1 + \frac{s}{\omega_c}}$
Magnitude		
Phase		

Figure 443: First Order Passive Filter Characteristics

Substitute $j\omega$ for s and $2\pi f$ for ω , and we have

$$\frac{e_o}{e_i} = \frac{\frac{j2\pi f}{2\pi f_c}}{1 + \frac{j2\pi f}{2\pi f_c}} = \frac{\frac{jf}{f_c}}{1 + \frac{jf}{f_c}}$$

Plugging $f = 120$ and $f_c = 300$ into this equation, we have:

$$\frac{e_o}{e_i} = \frac{\frac{j120}{300}}{1 + \frac{j120}{300}} = \frac{j0.4}{1 + j0.4} = \frac{0.4\angle 90^\circ}{1.07\angle 21^\circ} = 0.37\angle 69^\circ$$

In words, at 120Hz the RC highpass filter produces a signal which is 0.37 times the magnitude of the input

signal. The output is also phase-shifted by +69 degrees with respect to the input signal.

In decibels, the gain of the filter is:

$$\begin{aligned}\frac{e_o}{e_i} \text{ db} &= 20 \log_{10} \left| \frac{e_o}{e_i} \right| \\ &= 20 \log_{10}(0.37) \\ &= -8.6 \text{ db}\end{aligned}$$

A reduction of 8 db in psychoacoustic terms is roughly equivalent to 'half as loud', so this filter halves the apparent loudness of the hum signal.

17.2.2 First Order Active Filters

	Lowpass	Highpass
Circuit		
Equation	$\frac{e_o}{e_i} = k \left(\frac{1}{1 + \frac{s}{\omega_c}} \right)$ $k = -R_f/R_i$ $\omega_c = 1/(R_f C)$	$\frac{e_o}{e_i} = k \left(\frac{\frac{s}{\omega_c}}{1 + \frac{s}{\omega_c}} \right)$ $k = -R_f/R_i$ $\omega_c = 1/(R_i C)$
Magnitude		
Phase		

Figure 444: First Order Active Filter Characteristics

Operational amplifiers may be used to construct first order lowpass and highpass filters, as shown in figure 444. These filters have the same basic shape as the filters shown in figure 443: notice the similarity between the equations. However, there are the following differences in detail:

- The maximum possible gain of the passive filters of figure 443 is 0db (1 volt/volt). The active filters can have gain – that is, the output voltage can be larger than the input.
- The op-amp introduces an inversion into the signal, and this must be taken into account when calculating the phase relationship between input and output: a phase of -180° must be added to the phases in figure 443.
- The output of the op-amp active filters is a voltage source, so the internal resistance of the filter is low compared to the passive versions.

The op-amp circuits of figure 444 are based on the inverting amplifier configuration. Alternatively, the non-inverting op-amp configuration can be used. This is explored in problem 17 on page 592.

17.3 Example: A Simple Bandpass Filter

It is possible to construct simple active filters using the Reactance Plot and Bode Plot techniques discussed in section 11. Figure 445 shows a bandpass filter based on these ideas.

The concept of the circuit is as follows:

- The overall gain of the circuit is determined by the ratio of the feedback and input impedances. By selecting the appropriate impedances, the overall transfer function can be made to assume a bandpass shape.
- At low frequencies, both capacitors are open circuits and the gain is very small (negative infinite decibels).
- As the frequency increases, the reactance of capacitor C_1 decreases below the resistance of R_1 . Capacitor C_2 is still an open circuit compared to R_2 . Then the overall gain is $-R_f/R_i$, which is -10 volts per volt. (In log format, this is $+20\text{db}$, with a 180° phase shift due to the minus sign.)
- At higher frequencies, the reactance of capacitor C_2 decreases below the resistance of R_2 , shorting it out, and the gain then decreases back into the decibel basement.

It's not obvious where the break frequencies will fall until one does the calculations, so this intuitive concept must be checked with an analysis and Bode plot.

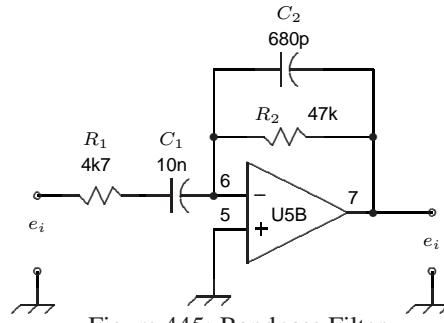


Figure 445: Bandpass Filter

The Transfer Function

Overall, the transfer function is

$$G(s) = -\frac{Z_f(s)}{Z_i(s)} \quad (701)$$

where $Z_i(s)$ is the input impedance and $Z_f(s)$ the feedback impedance.

$$\begin{aligned} Z_f(s) &= R_2 \parallel \frac{1}{sC_2} \\ &= \frac{R_2}{1 + sR_2C_2} \end{aligned} \quad (702)$$

$$\begin{aligned} Z_i(s) &= R_1 + \frac{1}{sC_1} \\ &= \frac{1 + sR_1C_1}{sC_1} \end{aligned} \quad (703)$$

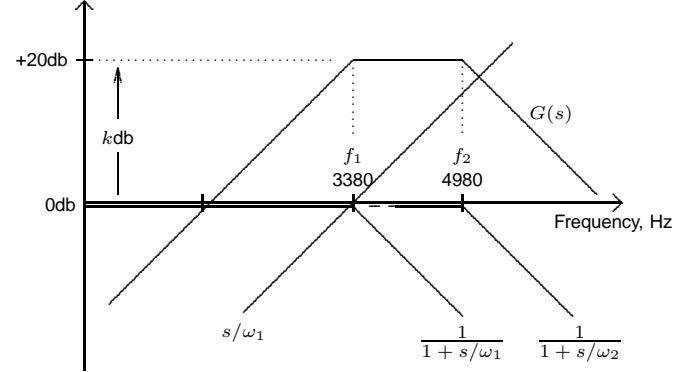


Figure 446: Filter Bode Plot

Substitute for $Z_i(s)$ and $Z_f(s)$ in equation 701, and

$$\begin{aligned} G(s) &= -\frac{\frac{R_2}{1 + sR_2C_2}}{\frac{1 + sR_1C_1}{sC_1}} \\ &= -\frac{sR_2C_1}{(1 + sR_1C_1)(1 + sR_2C_2)} \end{aligned} \quad (704)$$

The passband of this amplifier occurs between f_1 and f_2 in figure 446, and in the bandpass region C_1 is a short-circuit compared to R_1 and C_2 is an open circuit compared to R_2 . By examination of the circuit diagram we would expect that the mid-band gain would be:

$$G(s) = -\frac{R_2}{R_1} \quad (705)$$

We can extract the factor R_2/R_1 out of equation 704 if we multiply it by R_1/R_1 .

$$\begin{aligned} G(s) &= -\frac{R_1}{R_1} \frac{sR_2C_1}{(1 + sR_1C_1)(1 + sR_2C_2)} \\ &= -\frac{R_2}{R_1} \frac{sR_1C_1}{(1 + sR_1C_1)(1 + sR_2C_2)} \end{aligned} \quad (706)$$

We can put

$$\begin{aligned} k &= R_2/R_1 \\ \omega_1 &= 1/R_1C_1 \\ \omega_2 &= 1/R_2C_2 \end{aligned}$$

Then equation 706 can be written as:

$$G(s) = -k \frac{s/\omega_1}{(1 + s/\omega_1)(1 + s/\omega_2)} \quad (707)$$

There are four terms of interest here:

- The frequency independent gain $-k$, where $k = R_2/R_1$
- The factor s/ω_1 in the numerator which crosses the frequency axis at

$$\omega_1 = \frac{1}{R_1 C_1} \text{ radians/sec}$$

- The factor $(1 + s/\omega_1)$ in the denominator, which is a first-order lag (lowpass) that breaks downward at ω_1 radians/sec.
- The factor $(1 + s/\omega_2)$ in the denominator, which is another first-order lag, this time breaking downward at ω_2 radians/sec.

Next, attach numbers to the constant and the frequencies. Working in Hz for the frequency, we have:

$$\begin{aligned} -k &= -R_2/R_1 \\ &= -47k/4k7 \\ &= -10 \text{ volts/volt} \\ &= +20\text{db}, \angle 180^\circ \\ f_1 &= \frac{1}{2\pi R_1 C_1} \\ &= \frac{1}{2\pi(4.7 \times 10^3)(10 \times 10^{-9})} \\ &= 3.38 \text{ kHz} \\ f_2 &= \frac{1}{2\pi R_2 C_2} \\ &= \frac{1}{2\pi(47 \times 10^3)(680 \times 10^{-12})} \\ &= 4.98 \text{ kHz} \end{aligned}$$

Now we can construct the individual components of the Bode plot.

- The term $-k$ is a +20db constant, with a 180° phase shift.
- The term sR_2C_1 is a differentiator which passes through the 0db axis at f_1 , 3.38kHz.
- The term $1 + sR_1C_1$ is a first-order lag which breaks downward at 3.38kHz.
- The term $1 + sR_2C_2$ is another first-order lag which breaks downward, this time at 4.98kHz.

These individual terms are shown plotted in figure 446. When these curves are added together, we obtain the resultant frequency response, the magnitude of the transfer function $G(s)$.

For example, the result follows the differentiator up to frequency f_1 . At that point, the first-order lag causes a break downward by 20db/decade, cancelling the rising characteristic of the integrator and creating a characteristic with zero slope. At frequency f_2 , the second lag causes another break downward by 20db/decade, causing $G(s)$ to decrease at this rate.

Now it should be evident how the various parameters affect the design of the filter. The passband gain is determined by the ratio of resistors R_1, R_2 . The frequency of the lower limit of the passband is determined by resistor and capacitor R_1, C_1 . The upper limit is determined by resistor and capacitor R_2, C_2 .

17.4 Second-Order Low-Pass Filter

Now let us consider how we would design a lowpass filter where the rate of transition from passband to stopband is greater than the 20db/decade provided by a first-order lowpass filter.

We might consider cascading several RC filters, as shown in figure 447. Each filter contributes a rolloff of -20db/decade, so the entire filter will contribute a rolloff of -40db/decade. The filters are separated by buffer amplifiers of unity gain that prevent each filter being loaded by its successor.

A Bode plot, using the straight-line approximations, suggests that this circuit will create a filter that rolls off at a rate of -40 db/decade above the corner frequency, $\omega_o = 1/RC$. Unfortunately, there is a problem. Each lowpass section is down by 3db at the cutoff frequency, so the cascade of two lowpass sections is down by 6db. For many applications, we would prefer that the gain remain constant up to the cutoff frequency and then transition abruptly to the -40 db/decade slope above the corner frequency.

Each further addition of a first-order lowpass filter worsens the depression in the region of the cutoff frequency by 3db.

An examination of the transfer function of the circuit in figure 447 gives us an important clue how to improve the frequency response curve.

17.4.1 Transfer Function of the First Order Cascade

The transfer function of each RC lowpass filter section is given by

$$\begin{aligned} \frac{e_o}{e_1} &= \frac{1/sC}{R + 1/sC} \\ &= \frac{1}{1 + sRC} \end{aligned} \quad (708)$$

The cascade of two of these sections is then given by:

$$\begin{aligned} \frac{e_o}{e_i} &= \left(\frac{1}{1 + sRC} \right)^2 \\ &= \frac{1}{1 + 2sRC + s^2(RC)^2} \end{aligned} \quad (709)$$

The *standard form* for a second-order lowpass filter, which is also the quadratic factor of section 11.8, is:

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \quad (710)$$

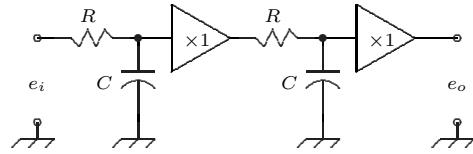


Figure 447: Cascade of First-Order Lowpass Sections
The figure shows a circuit diagram of a second-order low-pass filter. It consists of two cascaded first-order sections. Each section is a series R-C combination followed by a buffer amplifier with a gain of 1. The overall transfer function is the square of the product of the two first-order terms.

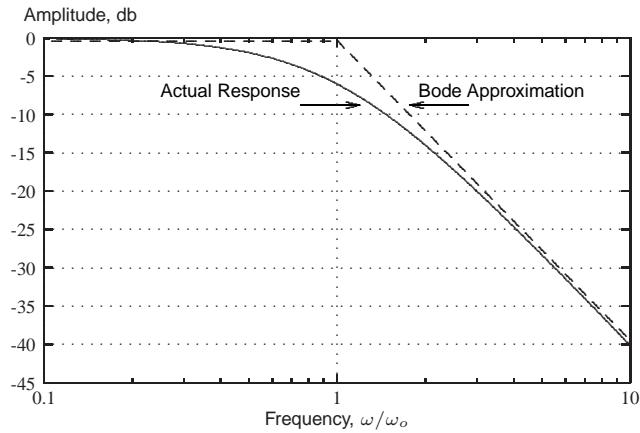


Figure 448: Response of Cascaded RC Sections

Massaging equation 709 into this form, we have:

$$\begin{aligned}\frac{e_o}{e_i} &= \frac{1}{(RC)^2 \left(s^2 + \frac{2}{RC}s + \frac{1}{(RC)^2} \right)} \\ &= \frac{\left(\frac{1}{RC}\right)^2}{s^2 + \frac{2}{RC}s + \left(\frac{1}{RC}\right)^2}\end{aligned}\quad (711)$$

Comparing equation 711 with the standard form of equation 710, we have for the cutoff frequency and Q factor:

$$\omega_o = 1/RC \quad (712)$$

and

$$Q = 1/2 \quad (713)$$

When the amplitude of the standard low-pass equation 710 is plotted for different values of Q , we obtain the family of curves shown in figure 449. Notice the curve for $Q = 0.5$: it corresponds to the amplitude curve shown in figure 448. Furthermore, figure 449 shows us that we would obtain a much more ideal characteristic with a value of $Q = 1$. A cascade of first-order sections has a poor response in the region of the cutoff frequency because its Q value is too low.

Furthermore, the Q factor is not dependent on R or C so it cannot be increased by tinkering with the values of resistance and capacitance.

To increase the value of Q , we must modify the circuit configuration of figure 447. But before we do that, we will look at the passive RLC lowpass filter, in which the value of Q can be adjusted to suit different requirements.

Second-Order RLC Filter

The circuit shown in figure 450 is a second-order lowpass filter: it passes low frequencies without attenuation and it attenuates high frequencies at a rate of 40db/decade.

As we did in the previous section, we'll develop the transfer function and relate it to the Bode plot for the quadratic factor of section 11.8.

Analysis

Treating the network as a voltage divider, the gain of the network is

$$G(s) = \frac{e_o}{e_i}$$

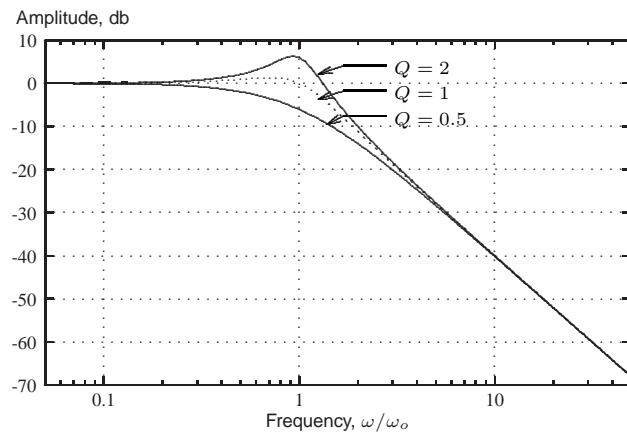


Figure 449: Lowpass Filter, Magnitude Plot

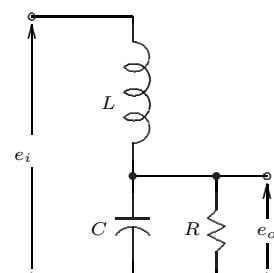


Figure 450: Second-Order Lowpass Filter

$$= \frac{Z_2}{Z_1 + Z_2}$$

where Z_1 and Z_2 are the upper and lower halves of the voltage divider.

$$\begin{aligned} Z_1 &= sL \\ Z_2 &= \frac{1}{sC} \parallel R \\ &= \frac{R}{1 + sRC} \end{aligned}$$

Then the transfer function is

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{\frac{R}{1 + sRC}}{sL + \frac{R}{1 + sRC}} \\ &= \frac{R}{sL + s^2RLC + R} \end{aligned}$$

To get this into the standard form, we need a fraction with 1 in the numerator and a 2nd order polynomial in s in the denominator. Divide the denominator by the factor R in the numerator and then factor LC out of the denominator:

$$\frac{e_o}{e_i} = \frac{1}{LC \left(s^2 + \frac{1}{RC}s + \frac{1}{LC} \right)} \quad (714)$$

It helps at this point to recognize that the inductor, resistor and capacitor form a parallel resonant circuit. To see that, consider that the voltage source produces a short impulse and then goes quiet. The internal impedance of a voltage source is zero, so the voltage source is now a short circuit, and the inductor is in parallel with the capacitor and resistor.

Now we can substitute for some of these quantities. From section 4.12 on page 176, we know that the resonant frequency for an RLC circuit is

$$\omega_o = \frac{1}{\sqrt{LC}} \quad (715)$$

so we can substitute ω_o^2 for $1/LC$ in equation 714.

Furthermore, for a parallel resonant circuit, the Q factor is given by

$$Q = \frac{R}{\omega_o L} \quad (716)$$

(see equation 208 on page 177).

We also know that at resonance the inductive and capacitive reactances are equal:

$$\omega_o L = \frac{1}{\omega_o C} \quad (717)$$

Substituting for $\omega_o L$ in equation 716,

$$Q = \omega_o RC \quad (718)$$

and

$$\frac{1}{RC} = \frac{\omega_o}{Q} \quad (719)$$

So we can substitute ω_o/Q for $1/RC$ in equation 714. Then

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (720)$$

This is the *standard form* for the 2nd order lowpass filter and should be placed away in the visual memory for electronics. To see that it is exactly equivalent to the quadratic factor of equation 453 in section 11.8, move the ω_o^2 term from the numerator, divide it into the denominator and replace s by $j\omega$. Then one obtains

$$\frac{e_o}{e_i} = \frac{1}{\left(j\frac{\omega}{\omega_o}\right)^2 + \frac{1}{Q}j\frac{\omega}{\omega_o} + 1} \quad (721)$$

Equation 453 in section 11.8 (page 11.8) is equivalent to equation 721 above when

$$Q = \frac{1}{2\delta} \quad (722)$$

Consequently, the response curve of figure 240 in section 11.8 describes the behaviour of this second-order lowpass filter network.

Equation 722 illustrates an equivalence between two terms used in slightly different disciplines. Control systems engineers use the term *damping factor* δ to describe the peaking effect in a second-order system. Electronic filter designers and radio engineers use the term *Q factor* to describe the same phenomenon. A competent engineer can go back and forth between these two representations.

Adjusting the Q Factor

According to equation 716 above, the Q factor is a function of the resonant frequency ω_o , the inductance and the resistance. To adjust the Q factor without affecting the resonant frequency, one would adjust the circuit resistance R .

Resonant Frequency ω_o and Cutoff Frequency ω_c

Refer again to the magnitude response of figure 449. As the value of the Q factor increases, the peak of the amplitude curve moves toward $\omega/\omega_o = 1$, that is, toward ω_o . If the cutoff frequency ω_c is defined as the point at which the response is down by some amount (3db, say), then the cutoff frequency varies substantially with different values of Q factor. Consequently, it is usually more convenient when characterising a second-order lowpass filter to refer to the *undamped resonant frequency* ω_o .

Phase Response

The phase response of the second-order lowpass filter is shown in figure 241 on page 312. The essential points are:

- The phase changes from 0° to -180° , passing through -90° at the resonant frequency ω_o
- The rate of change of phase in the vicinity of ω_o increases with increasing Q factor (which corresponds to decreasing values of damping δ).

Plotting the Amplitude and Phase Curves

In the event that you need to know the response curve of a lowpass filter for some value of Q other than the values shown in figure 449, it's useful to have the amplitude and phase curves in a form that can be plotted.

Magnitude

To obtain the magnitude plot, start with equation 721 and replace ω/ω_o with the x axis variable, which I'll call x . Then x represents the ratio of frequency to cutoff frequency.

We could plot the function over the range $x = 0.1\omega_o$ to $x = 10\omega_o$ to get an idea of the response in the two decades around the cutoff frequency.

Then we have

$$\begin{aligned}\frac{e_o}{e_i} &= \frac{1}{(jx)^2 + \frac{1}{Q}jx + 1} \\ &= \frac{1}{-x^2 + \frac{1}{Q}jx + 1}\end{aligned}$$

Collecting real and imaginary terms, we have

$$\frac{e_o}{e_i} = \frac{1}{(1-x^2) + j\frac{x}{Q}} \quad (723)$$

The magnitude is equal to the square root of the sum of the real and imaginary components each squared. As well, we'd like the result in decibels, so we take $20 \log_{10}$ of the result:

$$\begin{aligned}|G(w)| &= 20 \log_{10} \sqrt{\frac{1}{(1-x^2)^2 + \left(\frac{x}{Q}\right)^2}} \\ &= 20 \log_{10} \left[(1-x^2)^2 + \left(\frac{x}{Q}\right)^2 \right]^{1/2} \\ &= -10 \log_{10} \left[(1-x^2)^2 + \left(\frac{x}{Q}\right)^2 \right]\end{aligned} \quad (724)$$

Equation 724 is in a form that can be plotted.

Phase

The phase angle of a complex number is given by

$$\angle G = \tan^{-1} \left(\frac{Im}{Re} \right) \quad (725)$$

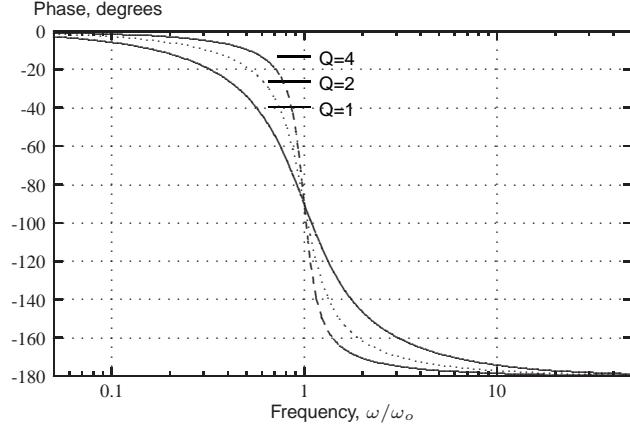


Figure 451: Lowpass Filter, Phase Plot

where R_e is the real part and I_m is the imaginary part.

We can also use the fact that

$$\frac{1}{\theta} = -\theta \quad (726)$$

Applying these concepts to equation 723, we have that

$$\angle G(\omega) = \tan^{-1} \left(\frac{x/Q}{1-x^2} \right) \quad (727)$$

which is the equation to plot for the phase of the second-order lowpass filter.

17.5 Second-Order Band-Pass Filter

Now that we have the second-order lowpass filter under control, we'll look at variants of that circuit. The bandpass version is shown in figure 452. At very low or very high frequencies, the reactances have low impedance, so the output voltage is reduced. Around the resonant frequency ω_o , the parallel combination of L and C resonate and become a high impedance, raising the output voltage e_o .

First, we'll work through the development of the transfer function. Then we'll show how this function relates to the lowpass function developed in section 17.4.

Analysis

Treating the network as a voltage divider, the gain of the network is

$$\begin{aligned} G(s) &= \frac{e_o}{e_i} \\ &= \frac{Z_2}{Z_1 + Z_2} \end{aligned} \quad (728)$$

where Z_1 and Z_2 are the upper and lower halves of the voltage divider.

$$\begin{aligned} Z_1 &= R \\ Z_2 &= \frac{1}{sC} \parallel sL \\ &= \frac{sL}{1+s^2LC} \end{aligned}$$

Substituting for Z_1 and Z_2 in 728, and cranking the algebra machine¹⁷¹, the transfer function is

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{sL}{R + s^2RLC + sL} \\ &= \frac{1}{RC} \left(\frac{s}{s^2 + \frac{s}{RC} + \frac{1}{LC}} \right) \end{aligned}$$

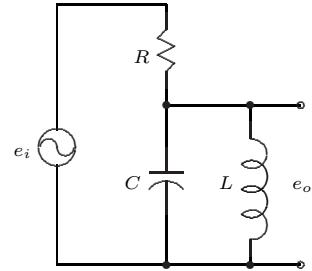


Figure 452: Bandpass Filter

¹⁷¹The challenge, as always in algebraic manipulations, is knowing where one wants to go with the expression. It helps to know (a) the form of the desired result and (b) useful substitutions that will get us there.

As shown in section 17.4, we can substitute

$$\begin{aligned}\frac{1}{RC} &= \frac{\omega_o}{Q} \\ \frac{1}{LC} &= \omega_o^2\end{aligned}$$

Then

$$\frac{e_o}{e_i} = \frac{\omega_o}{Q} \left(\frac{s}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \right) \quad (729)$$

This is a common formulation for the 2nd order bandpass filter. The frequency response form is obtained by dividing the numerator and denominator by ω_o^2 and replacing s by $j\omega$

$$\frac{e_o}{e_i} = \frac{1}{Q} \left(\frac{j\frac{\omega}{\omega_o}}{\left(j\frac{\omega}{\omega_o}\right)^2 + \frac{1}{Q}j\frac{\omega}{\omega_o} + 1} \right) \quad (730)$$

Plotting the Amplitude and Phase Curves

Magnitude

We could generate the magnitude and phase plots from first principles, but there's a shortcut. Starting with equation 730, notice that much of the denominator is the same as the lowpass case, equation 721. Comparing the lowpass and bandpass equations, the bandpass can be seen to be equal to a lowpass filter, multiplied by the term

$$\frac{j\frac{\omega}{\omega_o}}{Q}$$

Consequently, we can plot the magnitude function by adding to it a quantity equal to this term. Substituting x for ω/ω_o as we did previously for the lowpass case, we have that

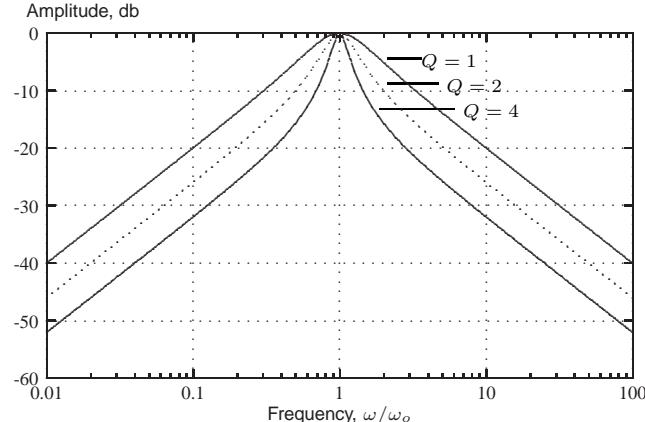


Figure 453: Bandpass Filter Amplitude

$$|G(w)| = 20 \log_{10} \left(\frac{x}{Q} \right) - 10 \log_{10} \left(1 - x^2 \right)^2 + \left(\frac{x}{Q} \right)^2 \quad (731)$$

A plot of this amplitude response for various Q factors is shown in figure 453.

In terms of the Bode plot, to convert a lowpass characteristic to a bandpass characteristic, we multiplied the lowpass magnitude transfer function by s , that is, we added a differentiator to the system. The effect is to rotate the lowpass characteristic counter-clockwise. The region below the resonant frequency goes from horizontal to

a slope of +20 db/decade. The region above the resonant frequency goes from a slope of -40 db/decade to -20 db/decade.

Moving from a lowpass to a bandpass also multiplies the lowpass function by $1/Q$. The effect is to shift the curve downward by Q db so that it always passes through 0db at ω_o .

Phase

The phase curve may be determined very easily. It is exactly the same shape as the lowpass case, but shifted up by 90° because of the added j term in the numerator of the bandpass characteristic. Consequently, we can generate the phase characteristic by adding 90° to equation 727, the phase characteristic for the lowpass filter:

$$\angle G(\omega) = 90 + \tan^{-1} \left(\frac{x/Q}{1-x^2} \right) \text{ degrees} \quad (732)$$

Notice that the phase characteristics for the bandpass filter have the same shape as the lowpass filter. For the lowpass filter, the phase curves run between 0° and -180° . For the bandpass filter, they run between $+90^\circ$ and -90° .

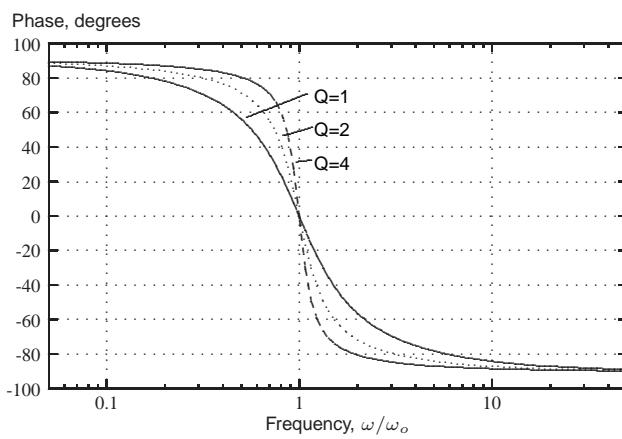


Figure 454: Bandpass Filter Phase

17.6 Second-Order High-Pass Filter

The circuit shown in figure 455 is a highpass filter: it passes high frequencies without attenuation, and attenuates low frequencies.

In this example, we'll develop the transfer function of the network and relate it to the Bode plot for the quadratic factor of section 11.8.

Analysis

Treating the network as a voltage divider, the gain of the network is

$$\begin{aligned} G(s) &= \frac{e_o}{e_i} \\ &= \frac{Z_2}{Z_1 + Z_2} \end{aligned} \quad (733)$$

where Z_1 and Z_2 are the upper and lower halves of the voltage divider.

$$\begin{aligned} Z_2 &= R \parallel sL \\ &= \frac{sRL}{R + sL} \end{aligned}$$

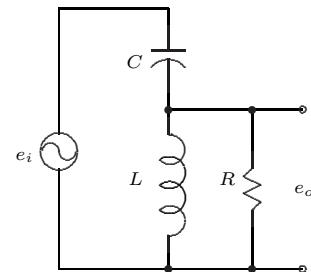


Figure 455: Highpass Filter

$$\begin{aligned} Z_1 + Z_2 &= \frac{1}{sC} + R \parallel sL \\ &= \frac{R + sL + s^2 RLC}{sC(R + sL)} \end{aligned}$$

Substituting for Z_1 and Z_2 in equation 733 the transfer function is

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{\frac{R + sL + s^2 RLC}{sC(R + sL)}}{\frac{sRL}{R + sL}} \\ &= \frac{s^2}{s^2 + \frac{s}{RC} + \frac{1}{LC}} \end{aligned}$$

As usual by now (see sections 17.4 and 17.4), we'll substitute

$$\omega_o = \frac{1}{\sqrt{LC}} \quad (734)$$

and

$$\frac{1}{RC} = \frac{\omega_o}{Q} \quad (735)$$

Then

$$\frac{e_o}{e_i} = \frac{s^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (736)$$

Plotting the Amplitude and Phase

The highpass characteristic of equation 720 is a lowpass filter characteristic multiplied by two differentiator terms (the s^2 term in the numerator). The effect of this multiplication is to rotate the amplitude function of the lowpass filter counter-clockwise by 40db per decade. The horizontal section of the lowpass filter becomes the rising section (+40db/decade) of the highpass filter. The decreasing section of the lowpass filter becomes the horizontal section of the highpass filter.

The phase response is simply equal to the phase response of the lowpass filter with $2 \times 90^\circ$ added to it.

Magnitude

Replace ω/ω_o in equation 736 with the x axis variable x .

Then we have

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{(jx)^2}{(jx)^2 + \frac{1}{Q}jx + 1} \\ &= \frac{-x^2}{-x^2 + \frac{1}{Q}jx + 1} \\ &= \frac{1}{1 - \frac{j}{Qx} - \frac{1}{x^2}} \end{aligned}$$

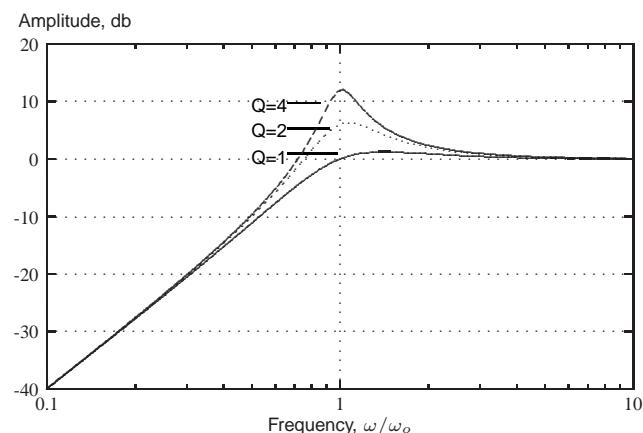


Figure 456: Highpass Filter, Magnitude Plot

Collecting real and imaginary terms, we have

$$\frac{e_o}{e_i} = \frac{1}{\left(1 - \frac{1}{x^2}\right) - \left(j\frac{1}{Qx}\right)} \quad (737)$$

Take the magnitude by squaring the real and imaginary parts and then taking the square root. Then convert to decibels:

$$|G(x)| = -10 \log_{10} \left[\left(1 - \frac{1}{x^2}\right)^2 + \left(\frac{1}{Qx}\right)^2 \right] \text{ db}$$

The magnitude plot is shown in figure 456.

Phase

The equation of phase is obtained by adding 180° to the phase curve for the lowpass filter (equation 727):

$$\angle G(\omega) = 180 + \tan^{-1} \left(\frac{x/Q}{1-x^2} \right) \text{ degrees} \quad (738)$$

The phase plot is shown in figure 457.

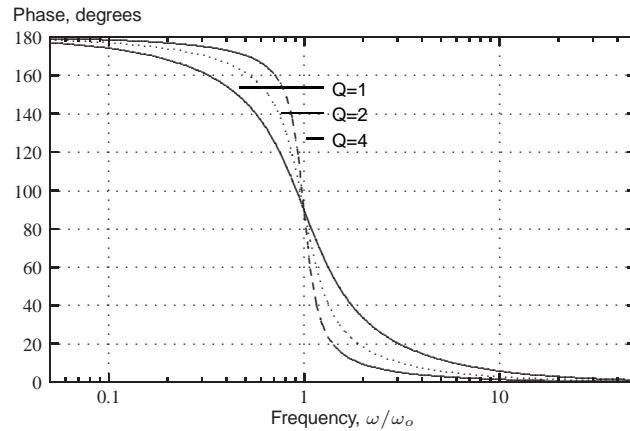


Figure 457: Highpass Filter, Phase Plot

17.7 Second-Order Band-Reject (Notch) Filter

The circuit shown in figure 458 is a *band-reject* or *notch* filter. At low frequency, the capacitor looks like an open circuit, and $e_o \approx e_i$. At high frequencies, the inductor looks like an open circuit, and again, $e_o \approx e_i$. At the resonant frequency, the series combination of the inductor and capacitor becomes low impedance, effectively shorting out the output signal. The output is substantially attenuated from its low and high frequency values.

As in the bandpass filter case, the Q factor of the circuit determines the width of the notch: higher Q results in a narrower notch.

The notch filter is useful when a single frequency must be removed from a wide-spectrum signal, with minimum impairment to the original spectrum. This is a requirement when a 60Hz or 120Hz interfering signal (due to electromagnetic coupling from the power line, or from a power supply) has contaminated an audio signal and must be removed. A highpass filter will work, but a notch is preferable if frequencies below the interfering signal are to be preserved.

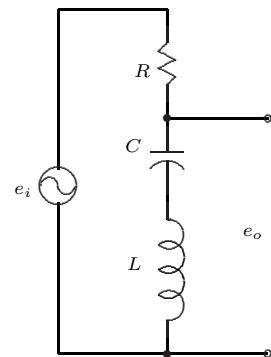


Figure 458: Band-Reject Filter

Ideally, in order to remove a single frequency, the notch filter should be infinitely narrow and consequently have the highest possible Q factor. In practice, this presents difficulties in centering the notch on the interference, and some width to the notch is required.

Analysis

As usual by now, we treat the network as a voltage divider. The gain of the network is

$$\begin{aligned} G(s) &= \frac{e_o}{e_i} \\ &= \frac{Z_2}{Z_1 + Z_2} \end{aligned} \quad (739)$$

where Z_1 and Z_2 are the upper and lower halves of the voltage divider.

$$\begin{aligned} Z_1 &= R \\ Z_2 &= \frac{1}{sC} + sL \\ &= \frac{s^2LC + 1}{sC} \end{aligned}$$

Substituting for Z_1 and Z_2 in equation 739 the transfer function is

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{\frac{s^2LC + 1}{sC}}{R + \frac{s^2LC + 1}{sC}} \\ &= \frac{s^2 + \frac{1}{LC}}{s^2 + \frac{R}{L}s + \frac{1}{LC}} \end{aligned}$$

As usual by now, we'll substitute

$$\omega_o = \frac{1}{LC} \quad (740)$$

and

$$\frac{R}{L} = \frac{\omega_o}{Q} \quad (741)$$

in equation 740. Then

$$\frac{e_o}{e_i} = \frac{s^2 + \omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2 \right)} \quad (742)$$

Equation 742 is the standard form for the transfer function of the notch filter.

As a quick check, consider the magnitude of this function at low frequencies, the resonant frequency, and high frequencies.

- At low frequencies, the s terms (equal to $j\omega$) will be small enough to be ignored. Then equation 742 reduces to ω_o^2/ω_o^2 , which is unity, or 0 db.

- At high frequencies, the s terms are dominant compared to the ω_o terms, and equation 742 reduces to s^2/s^2 , again unity volts/volt or 0 db.
- At the resonant frequency, $\omega = \omega_o$, so the numerator of equation 742 becomes $-\omega_o + \omega_o = 0$ volts/volt, or $-\infty$ db.

Plotting the Amplitude and Phase Curves

To derive the plot functions for the notch, begin with equation 742 above, and replace s with $j\omega$. Then

$$\begin{aligned} G(w) &= \frac{-\omega^2 + \omega_o^2}{-\omega^2 + \frac{\omega_o}{Q}j\omega + \omega_o^2} \\ &= \frac{1 - \left(\frac{\omega}{\omega_o}\right)^2}{1 - \left(\frac{\omega}{\omega_o}\right)^2 + \frac{1}{Q}\frac{j\omega}{\omega_o}} \quad (743) \end{aligned}$$

To simplify the notation, substitute x for ω/ω_o :

$$\begin{aligned} G(x) &= \frac{1 - x^2}{1 - x^2 + \frac{1}{Q}jx} \\ &= \frac{1}{1 + j\frac{x}{Q(1 - x^2)}} \quad (744) \end{aligned}$$

The magnitude of the function is equal to the square root of the real part squared plus the imaginary part squared. Then

$$|G(x)| = \sqrt{\left(1 + \left(\frac{x}{Q(1 - x^2)}\right)^2\right)^{-1}} \quad (745)$$

In decibels, this is

$$\begin{aligned} |G(x)| \text{ dB} &= 20 \log_{10} \sqrt{\left(1 + \left(\frac{x}{Q(1 - x^2)}\right)^2\right)^{-1}} \\ &= -10 \log_{10} \left(1 + \left(\frac{x}{Q(1 - x^2)}\right)^2\right) \quad (746) \end{aligned}$$

This is the function plotted in figure 459.

Notice:

- The Q factor determines the width of the notch.
- With ideal components, the notch null extends to $-\infty$ regardless of the Q factor. However, with real components the notch depth is finite.

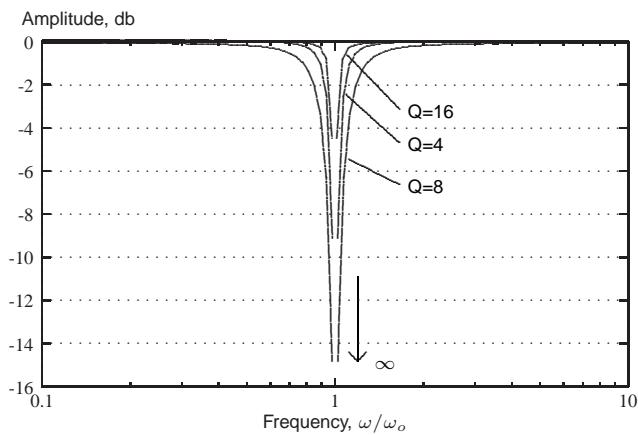
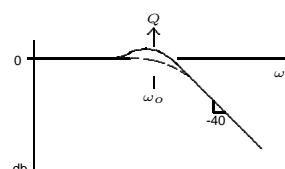
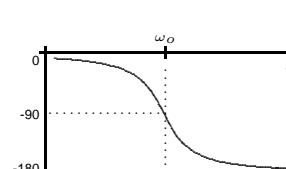
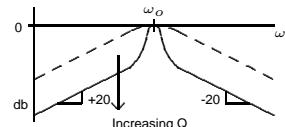
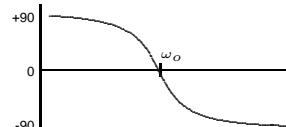
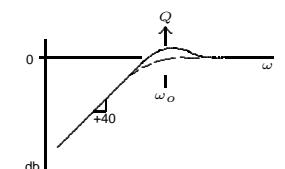
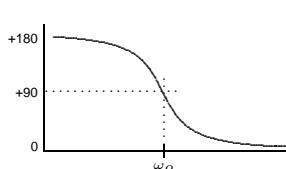
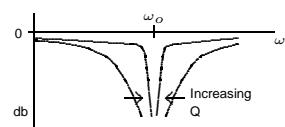
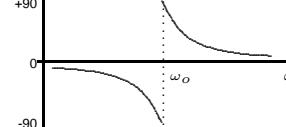


Figure 459: Notch Filter, Magnitude Plot

17.8 Summary, Second-Order Filters

Basic Type	Transfer Function	Amplitude	Phase
Lowpass	$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)}$		
Bandpass	$\frac{e_o}{e_i} = \frac{\omega_o}{Q} \frac{s}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)}$		
Highpass	$\frac{e_o}{e_i} = \frac{s^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)}$		
Bandreject	$\frac{e_o}{e_i} = \frac{s^2 + \omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)}$		

ω_o Undamped natural resonant frequency

Q Quality Factor

The preceding sections have shown the general equations and response curves for active filters. Now we will look at *implementation*: how these filters may be created using op-amps, resistors and capacitors.

17.9 The Bronzite Circuit

Lowpass Circuit

The circuit shown in figure 460 is one of several possible circuits to implement a second-order lowpass filter. Since it apparently has no other official name attached to it, I'll refer to it as the Bronzite filter after the reference [151] that first describes it. It's also briefly described in [152].

It's straightforward to calculate the component values for this circuit, as we shall see.

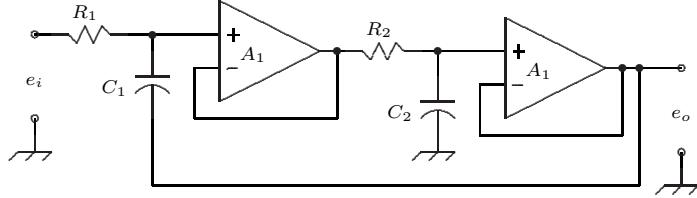


Figure 460: The *Bronzite* Lowpass Filter

Lowpass Analysis

We'd like to relate the circuit components to the parameters Q and ω_o of the second-order lowpass circuit. Then we can run those equations backwards and find circuit values for whatever Q and ω_o we require. We'll do that in two steps:

1. Determine the transfer function $G(s) = e_o/e_i$. This will contain terms in s^2 , s with the various resistances and capacitances.
2. Compare this transfer function to the standard lowpass transfer function in Q and ω_o . Then the values of the resistors and capacitors can be related to Q and ω_o .

Unfortunately, there are no magic shortcuts to analysing this circuit. We can't apply superposition to the outputs of the amplifiers, because the amplifiers are dependent sources – their output depends on their input.

The circuit of figure 460 is redrawn in figure 461 for the analysis. Now, a brief adventure in KVL, KCL and algebra:

$$e_i - i_1 R_1 - \frac{i_1}{sC_1} - e_o = 0 \quad (747)$$

$$e_a = e_i - i_1 R_1 \quad (748)$$

$$e_a = i_2 R_2 + e_o \quad (749)$$

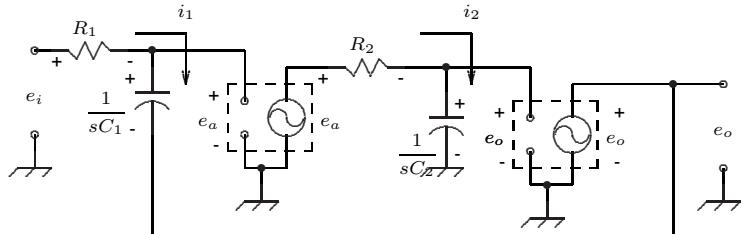


Figure 461: Bronzite Filter Analysis

$$i_2 = e_o s C_2 \quad (750)$$

Collapsing these equations and simplifying, we eventually obtain:

$$\begin{aligned} G(s) &= \frac{1}{s^2 R_1 R_2 C_1 C_2 + s R_2 C_2 + 1} \\ &= \frac{1}{R_1 R_2 C_1 C_2} \left(\frac{1}{s^2 + \frac{1}{R_1 C_1} s + \frac{1}{R_1 R_2 C_1 C_2}} \right) \end{aligned} \quad (751)$$

The standard second-order lowpass form (section 17.4 on page 524 is:

$$G(s) = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (752)$$

Comparing equations 751 and 752,

$$\omega_o = \sqrt{\frac{1}{R_1 R_2 C_1 C_2}} \quad (753)$$

and

$$\frac{\omega_o}{Q} = \frac{1}{R_1 C_1} \quad (754)$$

Equations 753 and 754 are the design equations for this circuit.

Lowpass Design Procedure

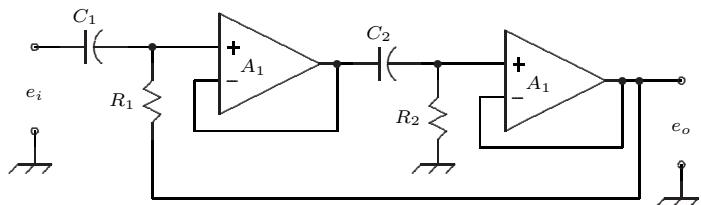
Once we have selected values for Q and ω_o , we can calculate the resistor and capacitor values.

1. Since there is no reason to do otherwise, choose the capacitors to be equal at some arbitrary value: $C_1 = C_2 = C$.
If the resistances turn out to be too small or too large (see below), change the capacitor value by some amount and redo the calculations. Of course, a spreadsheet is a useful tool for this type of design tweaking.
2. Substitute the values of ω_o , Q and C into equation 754 to determine R_1 .
3. Substitute the values of ω_o , C , and R_1 into equation 753 to determine R_2 .

Highpass Circuit

The highpass version of the Bronzite filter is shown in figure 462. It is obtained by interchanging the resistors and capacitors in the low-pass circuit.

After a similar analysis to the above, the transfer function of the circuit is



$$G(s) = \frac{s^2}{s^2 + \frac{1}{R_2 C_2} s + \frac{1}{R_1 R_2 C_1 C_2}} \quad (755)$$

Figure 462: The *Bronzite* Highpass Filter

The standard second-order highpass form (section 17.6 on page 531 is:

$$G(s) = \frac{s^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (756)$$

Comparing equations 755 and 756,

$$\omega_o = \sqrt{\frac{1}{R_1 R_2 C_1 C_2}} \quad (757)$$

and

$$\frac{\omega_o}{Q} = \frac{1}{R_2 C_2} \quad (758)$$

Highpass Design Procedure

The design procedure is similar to the lowpass case:

1. Choose the capacitors to be equal to C and choose a value for C .
2. Substitute the values of ω_o , Q and C into equation 758 to determine R_2 .
3. Substitute the values of ω_o , C , and R_1 into equation 757 to determine R_1 .

Notes

- Every input to an operational amplifier must have a DC path for bias current. Even when the bias currents are very small (as in the case of a MOSFET input op-amp) this path for current must exist. For example, in figure 460: a path for DC current must be available into the non-inverting terminal of the op-amp. The source e_i may conduct DC current. If not, then a large-value resistor across the input terminals or between the op-amp non-inverting terminal and ground, as shown in figure 463, should be added to provide this DC current path.

- The amplifiers shown in this circuit are connected as unity gain buffers. In the highpass circuit, the output is capacitively coupled back into the input. Consequently the amplifiers can be allowed to have a DC level shift between input and output (which simplifies their design), and that will have no effect on the operation as a highpass filter. References [151] and [153] show low-gain discrete component amplifiers being used in this type of circuit.

17.10 The Sallen-Key Lowpass Circuit

The *Sallen-Key* circuits [154] are the classic method of implementing second-order lowpass or highpass sections. They have reasonable component count and low sensitivity to variations in component tolerances. One version of the Sallen-Key lowpass filter is shown in figure 464. It's very similar to the Bronzite configuration, which it predates. However, it uses one op-amp instead of two. This complicates the design procedure, but reduces the parts count in cost-critical applications.

To understand how the filter works, notice that it consists of two RC lowpass sections in series with some positive feedback from the output of the amplifier back in to the filter section. Without the positive feedback, the two RC filters would

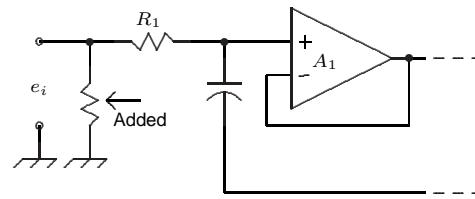


Figure 463: The *Bronzite* Lowpass Filter, DC Current Path

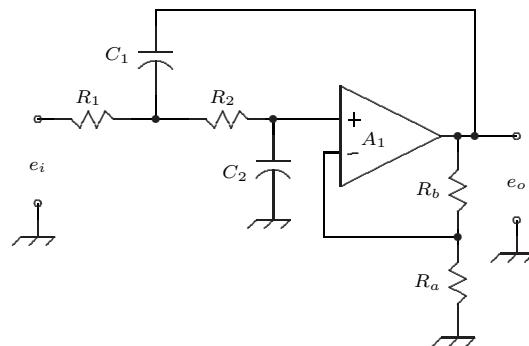


Figure 464: Sallen-Key Lowpass Filter

indeed function as a lowpass filter, but the response would be down significantly at the cutoff frequency – see figure 447 on page 524. This positive feedback is only effective around the cutoff frequency:

- At low frequencies, C_1 is an open circuit, blocking the positive feedback path
- At high frequencies, C_2 is a short circuit, shunting the positive feedback signal to ground

Consequently, the positive feedback increases the gain of the filter in the region of the cutoff frequency. The amount of positive feedback controls the degree of peaking at cutoff, that is, the Q factor of the filter – see figure 449 on page 525.

Analysis

The overall transfer function of the lowpass circuit in figure 464 is given by

$$G(s) = \frac{k \left(\frac{1}{R_1 R_2 C_1 C_2} \right)}{s^2 + \left[\frac{1}{R_1 C_1} + \frac{1}{R_2 C_1} + \frac{1}{R_2 C_2} (1 - k) \right] s + \frac{1}{R_1 R_2 C_1 C_2}} \quad (759)$$

where k is the gain of the amplifier, a non-inverting configuration:

$$k = 1 + \frac{R_b}{R_a} \quad (760)$$

The standard form of the lowpass (equation 720 in section 17.4) is:

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q} s + \omega_o^2 \right)} \quad (761)$$

By comparison of equation 761 with equation 759, we can see that equation 759 is a second order lowpass with a frequency-constant gain factor of k . (This factor k is sometimes called the *DC gain*, but it applies to all frequencies.) By comparison we can determine the relationships for ω_o and Q :

$$\omega_o = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}} \quad (762)$$

and

$$\frac{1}{Q} = \sqrt{\frac{R_2 C_2}{R_1 C_1}} + \sqrt{\frac{R_1 C_2}{R_2 C_1}} + (1 - k) \sqrt{\frac{R_1 C_1}{R_2 C_2}} \quad (763)$$

Design Options

Within the Sallen-Key family of circuits, there are several design options: the *unity gain* configuration, the *equal component* configuration and the *gain of two* configuration.

Unity Gain Configuration

For a unity gain configuration ($k = 1$), the value of R_A becomes infinite (open circuit) and R_B becomes a short circuit. The lowpass filter simplifies to the circuit shown in figure 465.

The resistors R_1 and R_2 can be chosen to be equal, in which case wonderful simplifications happen to equation 763.

$$\frac{1}{Q} = 2\sqrt{\frac{C_2}{C_1}} \quad (764)$$

In words, the Q factor is determined by the ratio of the capacitor values.

The design then proceeds as follows. As usual, we assume Q and ω_o are known.

1. Rewrite equation 764 into the form:

$$C_1 = 4Q^2 C_2 \quad (765)$$

Choose a value for C_2 and use the specified Q value in equation 765 to calculate C_1 .

2. Substitute the values for ω_o , C_1 and C_2 in equation 762 and solve for $R = R_1 = R_2$

Equal Component Configuration

The resistors are made equal to R and the capacitors equal to C . This is attractive for production, since it reduces the required inventory. The filter gain k depends on the value of Q and cannot be selected independently in this configuration.

Substitute R for the resistors and C for the capacitors in equations 762 and 763, and our design equations become:

$$\omega_o = \frac{1}{RC} \quad (766)$$

$$Q = \frac{1}{3-k} \quad (767)$$

In general, the values of ω_o and Q will be specified in the design requirements. One possible design procedure is:

1. Substitute the value of Q into equation 767 and determine the value of gain k . Typical values of Q are in the range 1.7 to 0.9. The gains corresponding to these Q values are 1.3 to 2.1.
2. Using equation 760, choose R_A (for example) and calculate R_B .
3. Choose a value of capacitance C . Substitute this value and the cutoff frequency ω_o into equation 766 to determine R .

Gain of 2 Configuration

Current feedback amplifiers are useful in high-frequency applications such as video amplifiers, where they must work at some frequency in the tens of MHz. When the amplifier drives a load resistance via some length of coaxial cable the output resistance of the amplifier must be matched to the cable impedance with a *buildout* resistor R_{buildout} in series with the output (figure 466).

The output then drives a coaxial cable of the same impedance into a load resistor which also matches the cable impedance. The buildout resistor and load resistor form a voltage divider with a gain of 0.5. If the amplifier has a gain of 2, then the signal at the input to the amplifier will arrive with the same amplitude at the load, and it is

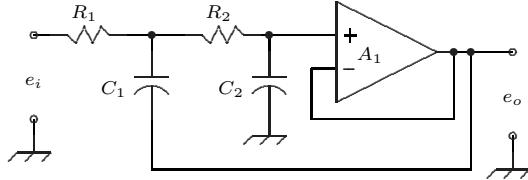


Figure 465: Sallen-Key Lowpass Filter, Unity Gain

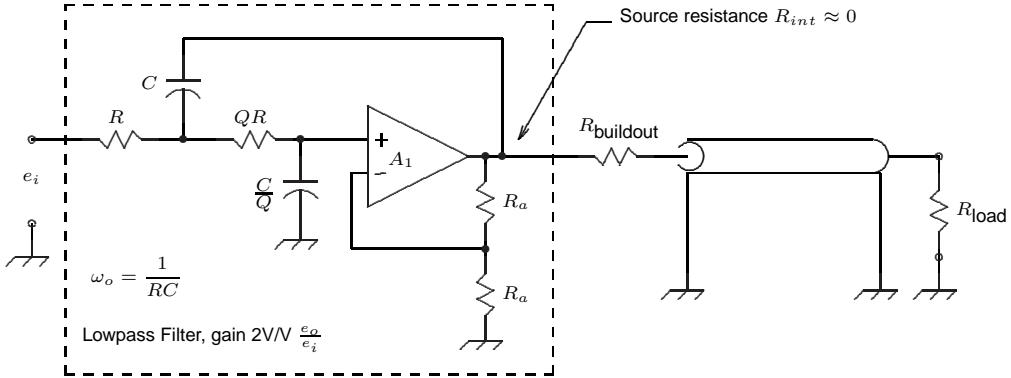


Figure 466: Sallen-Key Filter Drives Coaxial Cable

matched to the cable impedance at both ends of the cable. When such an amplifier is also to provide lowpass or highpass filtering, it's useful to be able to design the filter for a gain of 2, shown next.

First, regarding the filter gain: to obtain the design equations for this filter configuration, first rearrange equation 760, relating the value of gain k to resistors R_A and R_B . Since $k = 2$ in this case,

$$\begin{aligned} \frac{R_b}{R_a} &= k - 1 \\ &= 1 \end{aligned} \quad (768)$$

that is, $R_a = R_b$.

Now, regarding the Q factor, defined in equation 763: since so many of the terms in this equation are in the form of ratios, it turns out to be useful to choose the resistor and capacitor values in the following manner¹⁷²:

$$\begin{aligned} R_1 &= R \\ R_2 &= \alpha R \\ C_1 &= C \\ C_2 &= C/\alpha \end{aligned}$$

That is, the resistance of R_2 is larger than R_1 by a factor α . The impedance of C_2 is larger than C_1 by the same factor. Substituting these values and $k = 2$ in the Q equation 763, we obtain:

$$\begin{aligned} \frac{1}{Q} &= \sqrt{\frac{\alpha RC/\alpha}{RC}} + \sqrt{\frac{RC/\alpha}{\alpha RC}} + (-1)\sqrt{\frac{RC}{\alpha RC/\alpha}} \\ &= \frac{1}{\alpha} \end{aligned} \quad (769)$$

Our scale factor α is simply the value of Q , and so

$$\begin{aligned} R_2 &= QR_1 \\ C_2 &= C_1/Q \end{aligned}$$

¹⁷²As I've said before, if this looks to you like magic, you're not alone. It helps to know where you're going to know what substitutions are appropriate.

Now for the cutoff frequency ω_o : Substituting QR_1 for R_2 and C_1/Q for C_2 in equation 762, we obtain

$$\begin{aligned}\omega_o &= \frac{1}{\sqrt{R_1\alpha R_1 C_1 C_1/\alpha}} \\ &= \frac{1}{R_1 C_1}\end{aligned}\quad (770)$$

The design procedure:

1. The gain k is 2, and both the cutoff frequency ω_o and Q factor are known.
2. Choose the gain-setting resistors R_a and R_b . Since the gain is 2 they must be equal, but the absolute value is arbitrary. We could minimize inventory by choosing them to be equal to one of the filter resistances such as R_1 .
3. Arbitrarily choose R_1 . Use equation 770 for ω_o to calculate C_1 .
4. Using the Q factor, calculate $R_2 = QR_1$, $C_2 = C_1/Q$.

Summary

A summary of the three Sallen-Key lowpass filter designs is shown in figure 467 (adapted from [141]).

References

The design equations for the Sallen-Key lowpass are based on material in [141]. Similar material is in [139].

The unity-gain configuration can be designed very quickly and easily using the tabular method given in reference [153].

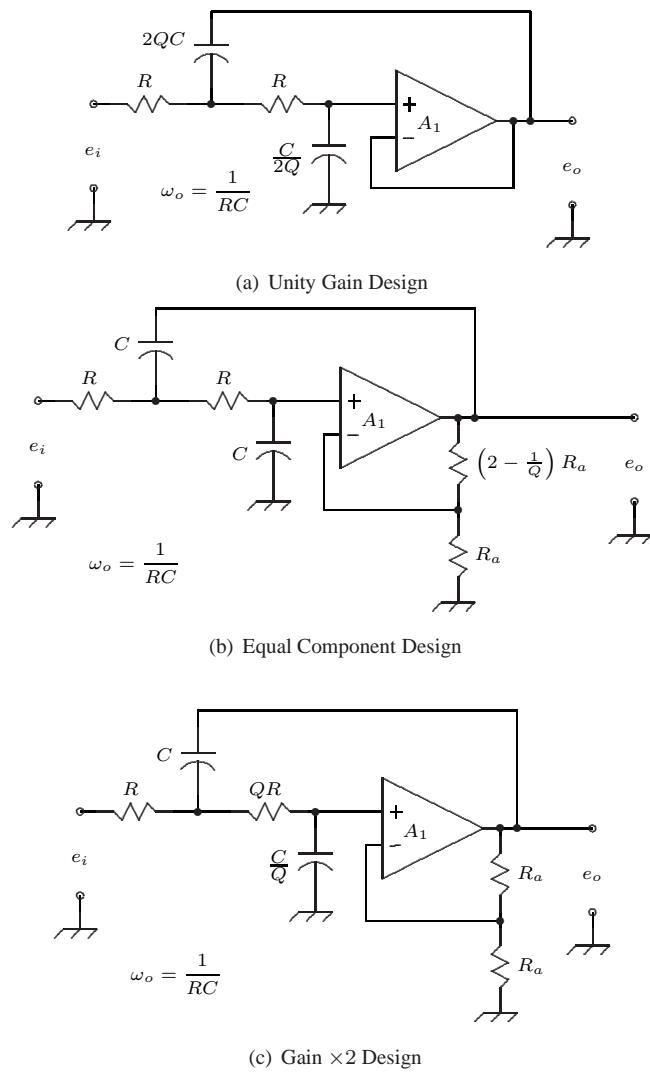
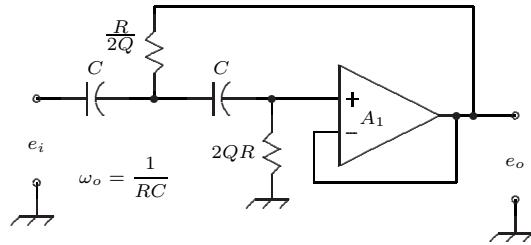
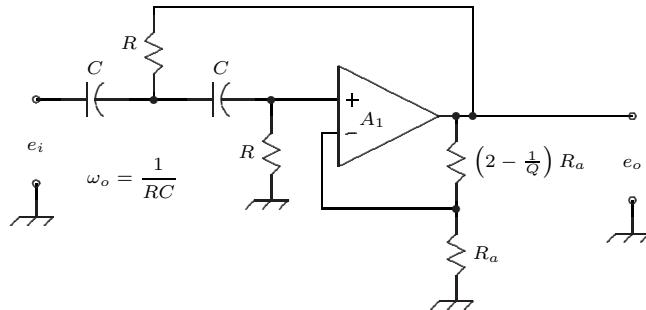


Figure 467: Sallen-Key Lowpass Filter Designs

17.11 The *Sallen-Key* Highpass Circuit



(a) Unity Gain Design



(b) Equal Component Design

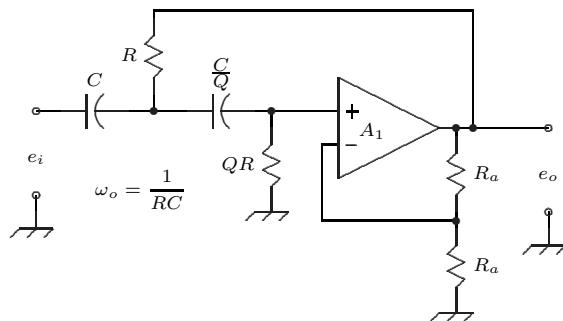
(c) Gain $\times 2$ Design

Figure 468: Sallen-Key Highpass Filter Designs

It would be entirely possible to work through a development for the Sallen-Key highpass filter as we did for the lowpass filter in section 17.10. However, there is an easier way.

With the application of two rules, we can transform the lowpass filters shown in figure 467 into highpass filters with the same cutoff frequency and Q factor:

- Resistors become capacitors and vice-versa.
- The impedance ratios are maintained. That is, a capacitor of value C/Q would become QR . In both cases, the impedance is increased by a factor of Q . Similarly, QC would become R/Q , because both impedances are lowered by a factor Q .

Applying those rules to the lowpass filters of figure 467, we obtain the designs shown in figure 468.

17.12 The State Variable Circuit

The state variable filter (figure 471 on page 548) is a combination lowpass, bandpass and highpass filter. Its primary application is as a bandpass filter since it can support large values of Q factor, that is, very narrow bandwidths. It's also relatively simple to adjust the centre (resonant) frequency ω_o , to make a tracking filter or spectrum analyser. The state variable filter circuit uses four op-amps, which may seem extravagant. However, four op-amps are readily available in one package at low cost and the advantages of this filter circuit often outweigh the cost.

To understand the state variable design, consider the series resonant circuit shown in figure 469 (see also section 4.11). Suppose that the frequency of the source is swept over a range that includes the resonant frequency of the network.

- At very low frequencies, the capacitive reactance will dominate the impedance and so most of the supply voltage will appear across the capacitor.
- As the frequency increases, the voltage across the capacitor will decrease and the voltage across the inductor will increase. At very high frequencies, the inductive reactance will dominate and most of the supply voltage will appear across the inductor.
- At resonance, the inductive and capacitive reactance cancel and the entire supply voltage appears across the resistance.

Based on this understanding of the behaviour, it would appear that the a plot of voltage across the capacitor, vs frequency, will have a lowpass filter characteristic. The voltage across the inductor will have a highpass characteristic and the voltage across the resistor will have a bandpass characteristic. We can confirm that the voltage across the resistor has a bandpass characteristic as follows:

Treating the network as a voltage divider where Z_T is the total impedance, the frequency response is given by

$$\begin{aligned}
 G(s) &= \frac{R}{Z_T} \\
 &= \frac{R}{R + \frac{1}{sC} + sL} \\
 &= \frac{R}{L \left(s^2 + \frac{R}{L}s + \frac{1}{LC} \right)}
 \end{aligned} \tag{771}$$

We can put this into the standard form by recognizing that:

$$\omega_o = \frac{1}{\sqrt{LC}} \tag{772}$$

and

$$Q = \frac{\omega_o L}{R} \tag{773}$$

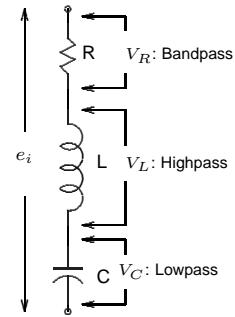


Figure 469: The Series Resonant Circuit

Consequently, we can substitute ω_o^2 for $1/LC$ and ω_o/Q for R/L in equation 771. Then

$$G(s) = \frac{\omega_o}{Q} \frac{s}{(s^2 + \frac{\omega_o}{Q}s + \omega_o^2)} \quad (774)$$

This is exactly the form shown in section 17.5, equation 729 for the bandpass filter.

Similarly, one can show that the output voltage across the inductor is a 2nd order highpass filter and the output voltage across the capacitor is a 2nd order lowpass filter.

Analog Computer Implementation

The circuit of figure 469 is used as a bandpass filter at radio frequencies, where the inductance and capacitance are both small components. However, at low frequencies such as the audio band, the inductance is inconveniently large and it is better to simulate the circuit with operational amplifiers.

We begin with KVL for the network in figure 469:

$$e_i = v_c + v_l + v_r \quad (775)$$

where v_c , v_l and v_r are the voltages across the individual components.

Next, we write the integral and differential equations for these voltages:

$$v_c = \frac{1}{C} \int idt \quad (776)$$

$$v_l = L \frac{di}{dt} \quad (777)$$

$$v_r = iR \quad (778)$$

Here's the trick: we assume that di/dt is available, then generate i and $\int idt$ by successive integrations. We then subtract these results from the input voltage, thereby generating the di/dt we assumed was available. In effect, we are forcing the circuit to converge on a state whereby the output voltage satisfies the input voltage. (A similar technique is shown in section 37.1.)

Substituting into equation 775 and re-arranging, we have

$$e_i = L \frac{di}{dt} + iR + \frac{1}{C} \int idt \quad (779)$$

$$\frac{di}{dt} = \frac{1}{L} \left(e_i - \frac{1}{C} \int idt - iR \right) \quad (780)$$

This form of the equation suggests that it could be implemented by the analog computing circuit [155], [96] shown in figure 470. (The circles contain multipliers, the boxes contain integrators and the triangle is an adder.)

This block diagram suggests that the series resonant circuit could be simulated with 2 opamp integrators and an opamp adder, and that is indeed the case. The *state variable* filter circuit based on this concept is shown in figure 471.

The name *state variable* is a reference to the similarity of this circuit to those used in analog computing and analog control systems, where the state of the system can be characterised by a set of variables (circuit voltages or currents) [156]. (For a similar example, see section 11.7 on the PID Controller.)

The design shown in figure 471 uses four op-amps, in contrast to some designs which use 3. However, this design has the advantage of independently adjustable gain (k) and Q factor. As indicated on the diagram, the highpass, bandpass and lowpass functions are available simultaneously, so this filter design is sometimes referred to as the *universal filter*. We will see in a moment that the filter can be tuned over a wide range by varying the R and C values. Finally, the state variable filter has low sensitivity figures, so its performance is not sensitive to component drift or tolerance.

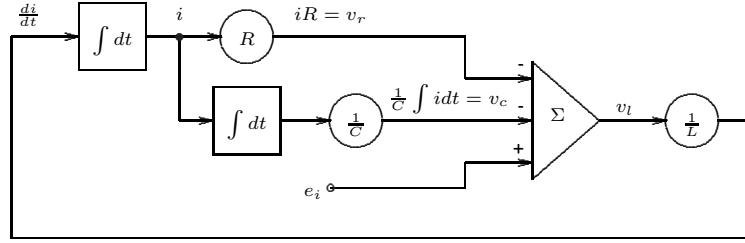


Figure 470: Series Resonant Circuit Simulation

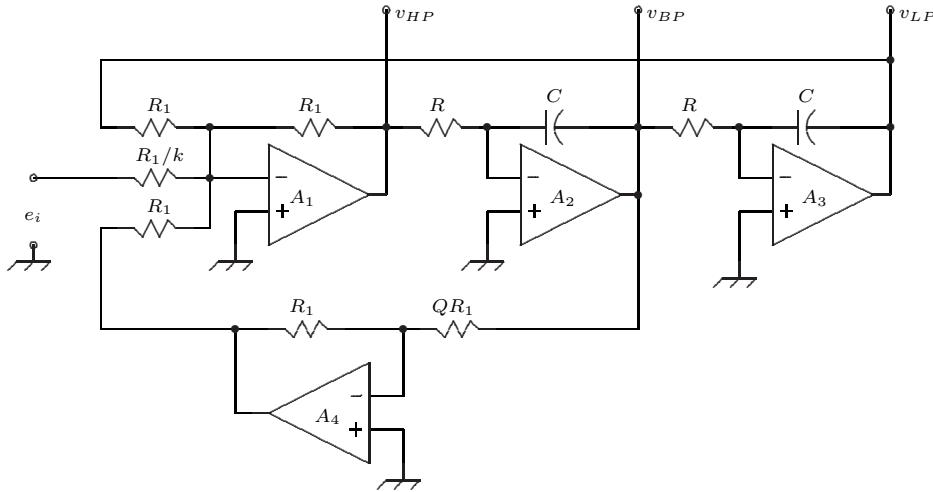


Figure 471: State Variable Filter

Analysis

As mentioned at the outset of this section, the state variable can provide lowpass, bandpass and highpass filtering functions. However, the primary use is as a bandpass filter, so we will focus on the output of op-amp A_2 in figure 471.

This op-amp circuit is an inverting summer (section 13.2), so we can write:

$$v_{HP} = - \left(v_{LP} + k e_i - \frac{v_{BP}}{Q} \right) \quad (781)$$

Opamp A_2 is configured as an integrator, so we can write

$$v_{BP} = - \frac{1}{sRC} v_{HP} \quad (782)$$

or, solving for v_{HP} , we have

$$v_{HP} = -sRC v_{BP} \quad (783)$$

Opamp A_3 is another integrator, so

$$v_{LP} = - \frac{1}{sRC} v_{BP} \quad (784)$$

Now substitute for v_{HP} from equation 783 and v_{LP} from equation 784 into equation 781:

$$-sRCv_{BP} = -\left(-\frac{1}{sRC}v_{BP}\right) - ke_i + \frac{v_{BP}}{Q} \quad (785)$$

Manipulating this to solve for the transfer function, we have

$$\begin{aligned} G_{BP}(s) &= \frac{v_{BP}}{e_i} \\ &= k \frac{s}{RC \left(s^2 + \frac{s}{RCQ} + \frac{1}{R^2C^2} \right)} \end{aligned} \quad (786)$$

To get the denominator into the standard form, put

$$\omega_o = \frac{1}{RC} \quad (787)$$

Then the final equation for the state variable bandpass output is:

$$\frac{v_{BP}}{e_i} = k\omega_o \left(\frac{s}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \right) \quad (788)$$

Compare this with the standard form for the second order bandpass filter (equation 729 from section 17.5):

$$\frac{e_o}{e_i} = \frac{\omega_o}{Q} \left(\frac{s}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \right) \quad (789)$$

The state variable bandpass is the same as the standard form, except that the magnitude is increased by a factor kQ . The factor k is the 'gain' of the filter. The Q factor can be quite large for this type of filter, and the voltage at the bandpass output will be larger than the input by a factor kQ , so care must be taken that clipping of the output does not occur for signals in the passband of the filter.

Electronic Tuning

If in figure 471 the two frequency-setting resistors R are replaced with some sort of electronic gain control element, then the frequency of the filter may be tuned by an electronic signal. Some of the possible devices that may be used as gain control elements are the multiplier (section 36.1), the *operational transconductance amplifier* (section 35.5), or a dual *multiplying D/A converter* (section 25).

References

Kennedy [141], Franco [139] and Lancaster [148] all contain readable analysis of the state variable filter and show useful variations on the circuit design. Sparkes [157] describes multiplier-controlled state-variable filter designs.

17.13 Notch Filter Circuits

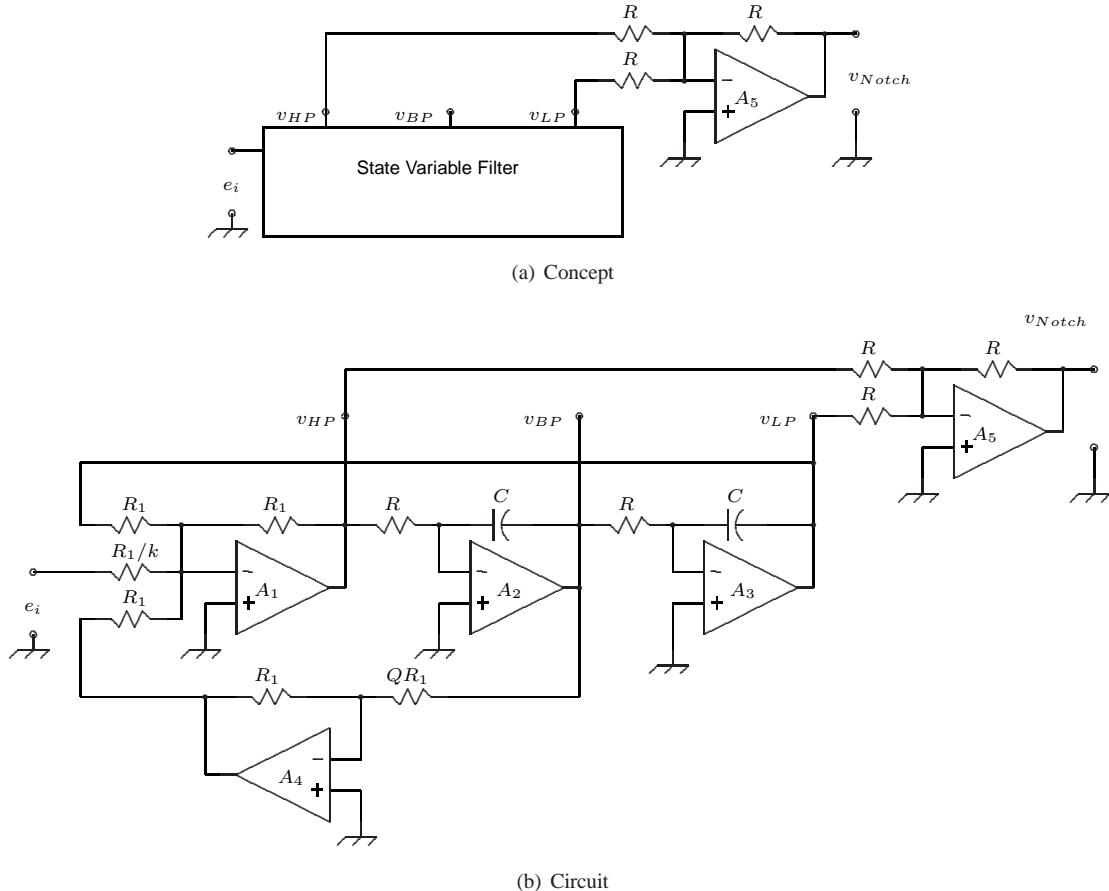


Figure 472: State Variable Notch Filter

There are occasions when a signal, comprising a range of frequencies, is contaminated by one localized frequency. This single frequency can be removed, without much effect on other frequencies in the signal, by a notch filter. An example of notch filter frequency response is shown in figure 459 on page 535.

An AC signal is cancelled (reduced to zero) when it is added to its inverted version – that is, subtracted from itself. If this can be made to occur in the region of one particular frequency, then the spectrum will have a narrow notch or band-reject region.

17.14 State Variable Notch

Consider what happens when we add the lowpass and highpass outputs of the state variable filter, as shown in figure 472(a). First consider phase:

- At low frequencies, the lowpass phase is 0° , the highpass is 180°
- At resonance (frequency ω_o), the lowpass phase is -90° , the highpass is $+90^\circ$.

- At high frequencies, the lowpass phase is -180° , the highpass is 0° .

So it would seem reasonable to expect that these signals differ by a constant 180° , that is, are out of phase.

Now consider the amplitudes. In the region of the resonant frequency ω_o , they have equal amplitudes. Consequently, they will cancel in the region of ω_o .

This is easy to show mathematically: from section 17.4, equation 720, the transfer function of the lowpass filter is

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (790)$$

From section 17.6, equation 736, the transfer function of the highpass filter is

$$\frac{e_o}{e_i} = \frac{s^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (791)$$

When these are added together, the result is

$$\frac{e_o}{e_i} = \frac{s^2 + \omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (792)$$

which is equation 742 of section 17.7, the standard form for the notch filter.

Consequently, the addition of the lowpass and highpass filter results in the notch characteristic. In circuitry, this could be accomplished by starting with the state-variable filter, section 17.12 figure 471. The outputs v_{LP} and v_{HP} are added together in an op-amp adder circuit (section 13.2) to produce the notch signal. The Q factor of the notch is given by the Q factor of the lowpass or highpass filter. The complete circuit is shown in figure 472 above.

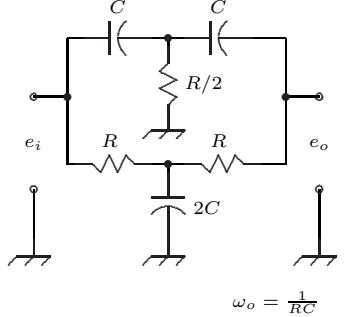
This seems like a lot of circuitry to produce a notch filter, but the state variable is a stable and predictable circuit design and it will reliably generate high-Q filter characteristics. Furthermore, it is easy to adjust for centre-frequency and Q factor.

17.15 The Twin-Tee Network

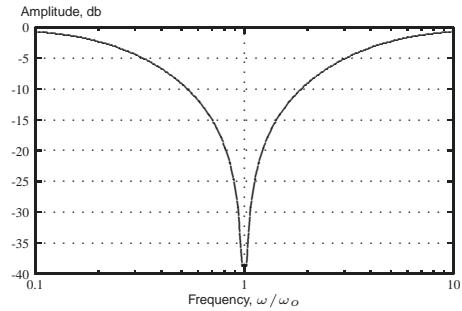
The Twin-Tee network shown in figure 473 is completely passive – no op-amps or power supply required [158], [159].

It can be shown that the transfer function e_o/e_i for the twin-tee is the standard form of equation 792 where Q is equal to $1/4$ and ω_o is equal to $1/RC$.

The upper and lower T sections each function as a phase-shifting network. The upper network causes the output current to lead the input voltage; the lower network causes the output current to lag the input voltage. These two output currents vary in amplitude and are a constant 180° out of phase with each other. At the notch frequency, the two currents are equal and so they cancel.



(a) Twin-Tee Circuit

(b) Twin-Tee Response
Figure 473: Twin-Tee Notch Filter

Analysis

To determine the notch frequency, we use a trick [160]. Since the output voltage is zero at the notch frequency, we can place any impedance across the output terminals without affecting the output current. As shown in figure 474, it's convenient to choose a short circuit. A short circuit across the output terminals ensures that the output current from the upper tee section is independent of the lower tee section and vice-versa, so we can analyse the sections independently.

Treating the upper tee section as a voltage divider, we have that

$$\begin{aligned} e_a &= e_i \left(\frac{R/2 \parallel 1/sC}{R/2 \parallel 1/sC + 1/sC} \right) \\ &= e_i \left(\frac{sRC}{2 + 2sRC} \right) \end{aligned} \quad (793)$$

Then the output current from the upper tee section is:

$$\begin{aligned} i_a &= \frac{e_a}{1/sC} \\ &= e_i \left(\frac{sRC}{2 + 2sRC} \right) sC \end{aligned} \quad (794)$$

Similarly, treating the lower tee section as a voltage divider:

$$\begin{aligned} e_b &= e_i \left(\frac{R \parallel 1/s2C}{R \parallel 1/s2C + R} \right) \\ &= e_i \left(\frac{1}{2 + 2sRC} \right) \end{aligned} \quad (795)$$

Then the output current from the lower tee section is:

$$i_b = \frac{e_b}{R}$$

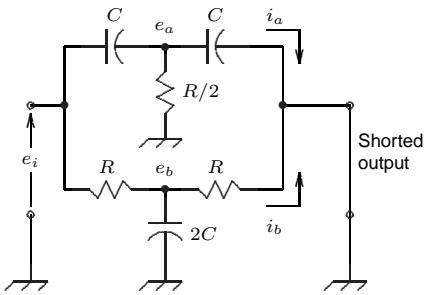


Figure 474: Twin-Tee Notch Filter, Analysis

$$= e_i \left(\frac{1}{2 + 2sRC} \right) \frac{1}{R} \quad (796)$$

At the notch frequency, the output currents i_a and i_b are equal, so we can use equations 794 and 796 to write:

$$e_i \left(\frac{sRC}{2 + 2sRC} \right) sC = e_i \left(\frac{1}{2 + 2sRC} \right) \frac{1}{R} \quad (797)$$

Simplifying,

$$s = \frac{1}{RC} \quad (798)$$

Substitute $j\omega$ for s, and then the null frequency ω_o is equal to $1/RC$.

17.15.1 Tuneable Passive Notch Filter

A tuneable version of this network [159] is shown in figure 475. The potentiometer adjusts the amplitude of the signal in the lower tee section, and this changes the frequency at which cancellation occurs.

The notch frequency is given by:

$$\omega_o = \frac{1}{\sqrt{kRC}} \quad (799)$$

where k is the pot setting: 1 at the top, 0.5 at the mid-point.

17.16 Higher Order Filters

In many filter applications, it is desirable to have an abrupt demarcation between the filter passband and the filter stopband. For example, consider the case of a *graphic equalizer*, used in audio applications. This device consists of a number of bandpass filters, each one with a different centre frequency in the audio spectrum. The input signal is fed to all these filters, passed through the filters and then combined with a volume control for each filter output (figure 476). Consequently, the frequencies in each third-octave segment of the spectrum can be adjusted up or down in order to modify the audio spectrum.

Ideally, each gain control will affect a certain range of frequencies and no others. This would imply band-pass filters with vertical cutoff slopes ($-\infty$ db/decade). It turns out that this *brick wall* characteristic, as it is called, is impossible to achieve exactly in practice. However, the ideal can be approached by *cascading* a number of filter sections, that is, processing the signal through a series of filters.

Higher order filters may be constructed with a cascade of first and second-order filter sections¹⁷³. The composite response is then the product of the individual responses.

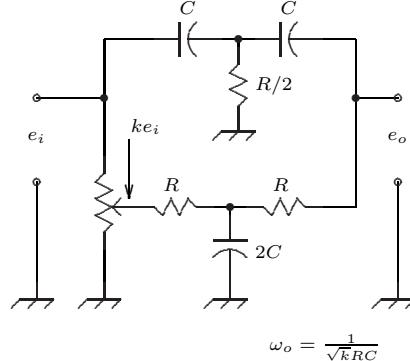


Figure 475: Tuneable Twin-Tee Notch Filter

Figure 475: Tuneable Twin-Tee Notch Filter

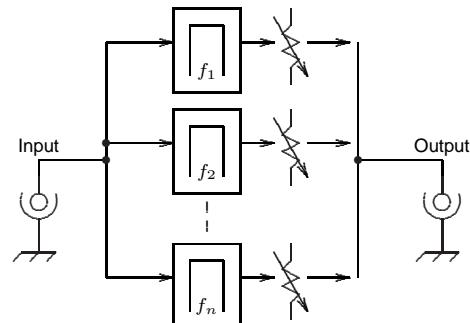


Figure 476: Graphic Equalizer Block Diagram

¹⁷³There are other methods of constructing high-order filters, but this method is simple to understand and simple to apply.

The precise shape of the higher-order filter (the transfer characteristic) can be selected to meet certain requirements: most rapid transition into the stop band, flattest response in the passband, or best transient response.

The overall shape is then controlled by the Q factors and the resonant frequencies of the individual sections (section 17.21).

17.17 The Butterworth Characteristic

One useful characteristic for high-order filters is the *Butterworth Approximation* [161]. According to [139], [78], the magnitude curve for this function is:

$$|G(w)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}} \quad (800)$$

where N is the order of the filter, and ω_c is the cutoff frequency of the filter. Figure 477(a) shows the Butterworth magnitude function for filters order 2 through 6.

Significant points of interest:

- The passband is flat and without ripples, decreasing smoothly toward the cutoff frequency.
- At the cutoff frequency, all curves pass through the -3db point (a numerical gain of $1/\sqrt{2}$ volt/volt.)
- The ultimate slope is $20N$ db/decade where N is the order of the filter. For example, the 5th order filter has an attenuation of 100db at $10\omega_c$.

Above the cutoff frequency, in the region where ω/ω_c is greater than 1, equation 800 can be approximated by:

$$|G(w)| \approx \left(\frac{\omega_c}{\omega}\right)^N \quad (801)$$

The phase response for the Butterworth filter is shown in figure 477(b). It was obtained with the coefficients of ω and Q for the Butterworth filter in section 17.24 and the method of section 17.16 for plotting the phase of high-order filters.

17.18 Designing A Butterworth Filter

1. Choose the required order of the filter based on the required attenuation in the stopband. For example, suppose that we require a filter that will process a 1kHz signal and a 3kHz signal, which are equal in magnitude. At the output of the filter, the 3kHz signal should be reduced to 1/100th of the 1kHz signal.

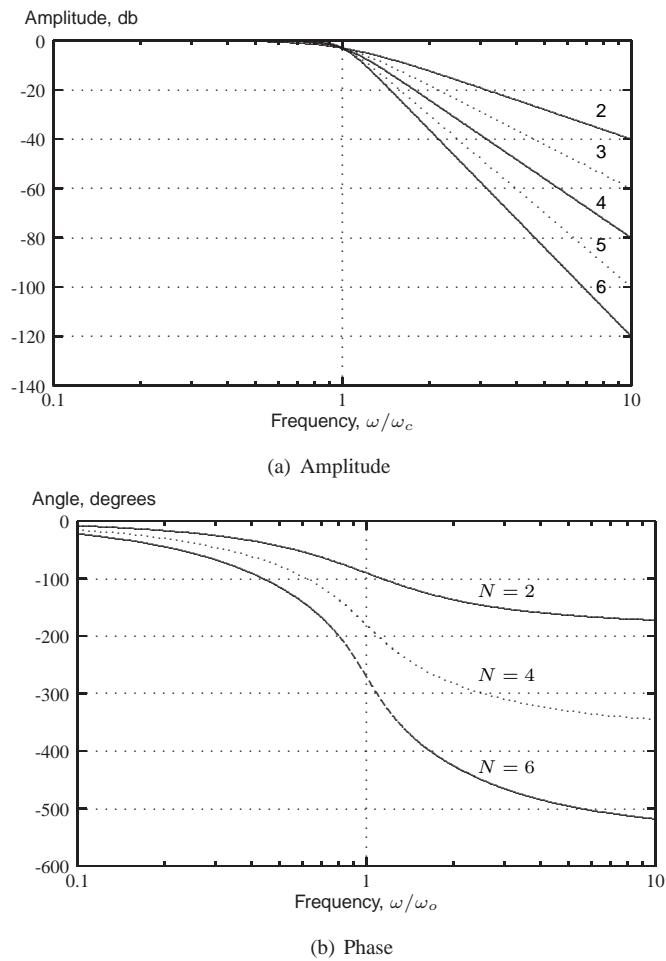


Figure 477: Butterworth Characteristic

Then the required relative attenuation is

$$\begin{aligned} G_{db} &= 20 \log_{10} G_{v/v} \\ &= 40 \text{ db} \end{aligned}$$

Now refer to figure 477(a). The filter passes the 1kHz signal and attenuates the 3kHz signal by 40db. We'll put the cutoff frequency f_c at 1kHz. To obtain an attenuation of 40db at 3kHz, $\omega/\omega_c = 3$, and a 5th order filter is about right. If the 1kHz signal is at ω_c , then it will be attenuated by 3db.

As a 5th order filter, it will consist of one first-order section and two second-order sections in cascade (series).

2. Consult the normalized Butterworth lowpass design table (section 17.24 on page 571) to identify the normalized values for ω_o and Q . From the table, the normalized fifth order filter coefficients are given as:

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
5	1	0.618	1	1.620	1	

These are the Q factors and cutoff frequencies for the filter functional blocks. The frequency factors ω_1 , ω_2 and so on are the resonant frequencies ω_o of the various second order sections as a ratio of the cutoff frequency ω_c of the filter. Then the first second-order filter has a cutoff frequency $\omega_o/\omega_c = 1$, so ω_1 for this filter would be at 1kHz. The Q factor of the first section is 0.618. The second, second-order filter also has a cutoff frequency $\omega_o/\omega_c = 1$ and a Q factor of 1.620. The third section, which is first order, has a cutoff frequency $\omega_o/\omega_c = 1$.

In fact for the Butterworth filter for all the second-order sections $\omega_o/\omega_c = 1$, ie, $\omega_o = \omega_c$. That is,

$$\begin{aligned} \omega_c &= 2\pi f_c \\ &= 2\pi 1000 \\ &= 6280 \text{ radians/sec} \\ &= \omega_o \end{aligned}$$

3. The block diagram for the fifth-order Butterworth filter is shown in figure 478(a). There are two, second-order lowpass sections in series, followed by a first-order section. The sequencing of these sections is unimportant¹⁷⁴.
4. Any first order or second order filter circuit could be used to implement this block diagram. However, for an active lowpass filter, the Sallen-Key lowpass filter (section 17.10) is a suitable choice. Then the complete filter schematic consists of two Sallen-Key lowpass filters, followed by a simple RC lowpass filter. The final op-amp is a voltage follower to provide a low impedance source for whatever load is connected to the filter. The complete schematic diagram is shown in figure 478(b).

Notice that all the resistors are the same value. This occurs because the resistors of a unity-gain Sallen-Key lowpass filter are equal. In general, the resistors of a Sallen-Key filter will have different values.

5. Using the design equations for the Sallen-Key lowpass filter from section 17.10, we could calculate the component values as follows:

¹⁷⁴In general, the sequencing of filter sections can affect the overload behaviour and noise floor of the filter. For example, if the first section magnifies the amplitude of a range of frequencies and the second section does the reverse, the output of the first section may overload. If they are reversed in sequence, then that is less likely to happen, but may compromise the signal-noise performance of the filter.

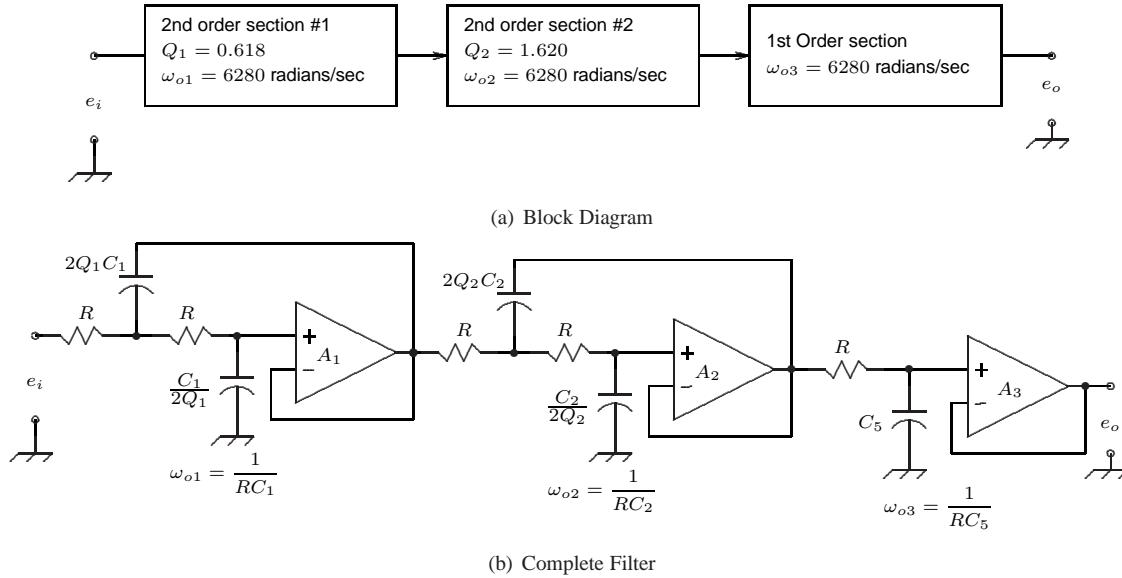


Figure 478: Fifth Order Lowpass Filter

- (a) Choose a suitable value for resistor R . In the absence of other constraints, you could choose some standard value like $10\text{k}\Omega$. A larger value of R results in smaller capacitors. However, very large values of R may cause excessive bias-current offset (section 21.2) or excessive noise (section 22), so this is another engineering tradeoff.
 - (b) For the final first-order section, use
- $$\omega_o = \frac{1}{RC}$$
- to calculate the value of the capacitor for the values of R and ω_{o3} .
- (c) Similarly, for the first of the second-order sections, calculate the value of C_1 from $R \omega_{o1}$. Now use the value of Q_1 to calculate the actual capacitor values for that section.
 - (d) Repeat this step for the second of the second-order sections.
 - (e) Now review the calculated component values. The capacitors will be oddball values that must be adjusted to their nearest standard value. If the capacitances are too large (resulting in capacitors that are physically too large), increasing resistor R by some factor will reduce the capacitance values by the same factor.
 - (f) Once the design is completed, use a circuit simulation program to determine the overall frequency response with the chosen resistor and capacitor values. This is a check against errors and shows the effect of the final capacitor values.

Butterworth Highpass Characteristic

The highpass transfer function of any filter may be obtained from the lowpass transfer function by substituting $1/f$ for f . So equation 800 for the lowpass transfer function is modified into the highpass transfer function by inverting ω/ω_c . Then we obtain for the Butterworth Highpass filter:

$$|G(w)| = \frac{1}{\sqrt{1 + \left(\frac{\omega_c}{\omega}\right)^{2N}}} \quad (802)$$

This is plotted in figure 479 for various values of filter order N .

The figures for Q and ω_o that correspond to some filter order are obtained from the same table as the lowpass filter. The filter is then implemented with first and second order highpass sections such as the Sallen-Key highpass circuit shown in section 17.11.

17.19 The Chebyshev Characteristic

The Chebyshev¹⁷⁵ characteristic trades passband ripple for steepness of descent around the cutoff frequency. A comparison of the Butterworth 4th order and Chebyshev 4th order lowpass filters is shown in figure 480.

The characteristics of the Chebyshev filter are:

- The response in the passband is not flat, but composed of ripples.
- The height of these ripples is controlled by a parameter ϵ in the design. In figure 480, these ripples have a magnitude of 3db peak-peak.
- Larger magnitude ripples occur with a steeper rolloff above the cutoff frequency.
- The number of inflection points (flat spots, points with zero slope) in the passband is equal to the order of the filter. In figure 480, the filter is fourth order and there are two valleys and two peaks, each of which is an inflection point.

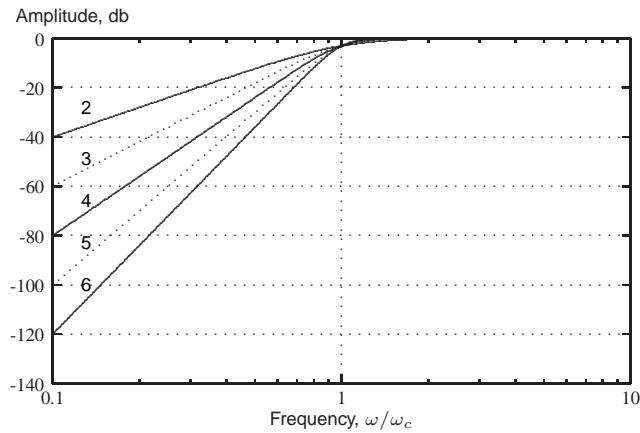
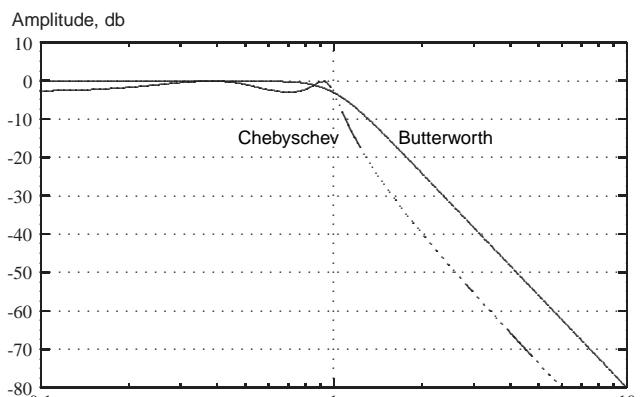


Figure 479: Butterworth Highpass Characteristic



(a) Magnitude

Figure 480: Chebyshev and Butterworth Characteristics

¹⁷⁵After Pafnuty Lvovich Chebyshev, Russian mathematician of the 19th century. There is a certain latitude in spelling his surname. Derbyshire [162] found a total of 32 different versions. Chebyshev's work on active filters actually derives from his study of the motion of linkages in steam locomotives.

- The amplitude response exits the ripple band at frequency ω_c . In the case shown, the ripple band is 3db, so the filter is 3db down at ω_c , the same as the Butterworth case. That's a coincidence in this case – they don't necessarily coincide.
- Ultimately, the lowpass filter rolls off at the same rate as any other lowpass filter of the same order ($20n$ db/decade, where n is the order of the filter). But because the Chebyschev filter rolls off more quickly in the region just above the cutoff frequency, the design requirement may be satisfied with fewer filter sections (ie, a lower order filter).
- The cutoff frequency ω_c occurs at the frequency where the amplitude response exceeds the passband variation established by ϵ .

Chebyschev Design Formulae

The magnitude of the passband ripple is related to ϵ according to:

$$A_{max}, \text{db} = -20 \log_{10} \frac{1}{\sqrt{1 + \epsilon^2}} \quad (803)$$

For example, for a Chebyschev filter with 3db passband ripple, $\epsilon = 0.997$.

Plotting of the magnitude function of the Chebyschev filter must be done over two regions: below and above the cutoff frequency ω_c . According to [139], [78] for a filter of order N and $\omega \leq \omega_c$ the magnitude function is given by:

$$|G(jw)| = \frac{1}{\sqrt{1 + \epsilon^2 \cos^2 \left[N \cos^{-1} \left(\frac{\omega}{\omega_c} \right) \right]}} \quad (804)$$

For $\omega \geq \omega_c$:

$$|G(jw)| = \frac{1}{\sqrt{1 + \epsilon^2 \cosh^2 \left[N \cosh^{-1} \left(\frac{\omega}{\omega_c} \right) \right]}} \quad (805)$$

When the hyperbolic cosine function \cosh and its inverse \cosh^{-1} are not available, use the following identities [4]:

$$\cosh x = \frac{e^x + e^{-x}}{2} \quad (806)$$

$$\cosh^{-1} x = \ln(x + \sqrt{x^2 - 1}) \quad (807)$$

Equations 804 and 805 were used to plot the Chebyschev magnitude characteristic in figure 480.

The phase characteristic is shown in figure 480(b). It was obtained using the tabulated coefficients of ω and Q for the Chebyschev filter, 3db passband ripple in section 17.24 and the method of section 17.16 for plotting the phase of high-order filters.

Designing A Chebyschev Filter

1. Decide the allowable ripple A_{max} db in the passband, bearing in mind that a larger allowable ripple will reduce the required order of the filter to achieve a given attenuation in the stopband. Use equation 803 to calculate the corresponding value of ϵ .

2. Determine the required order of the filter. There are two possible ways to do this:
 - Use equations 804 and 805 to plot the Chebyschev characteristic for your chosen value of ϵ and various values of order N . Examine these plots and choose the order that meets the requirement. Or,
 - Plug various values of N into equation 805 to determine the attenuation $|G|$ at some frequency ω/ω_c .
3. Select the Chebyschev table of coefficients according to the ripple factor (0.1db, 3db, whatever). Within that table, look up the coefficients for Q factor and ω_o/ω_c for each second-order section. (If the filter order is odd, there will also be a value of ω_o/ω_c for the first-order section.)
4. Use the coefficients for Q factor and ω_o/ω_c in the design of the filter circuits.

Example

According to [163]: *Electroencephalography is a medical imaging technique that reads scalp electrical activity generated by brain structures.* It is used by the medical profession to identify abnormalities in the physiology of the brain. The electrical signals detected at the scalp surface are very small – in the order of microvolts. The first stage of signal processing is a highpass filter removing frequencies below 0.5 Hz.

On the basis of other considerations, we have decided that this filter should have an attenuation of at least 80 db between 0.1 and 0.5Hz. Extreme flatness in the passband is not generally required, and so the allowable ripple in the passband can be as much as 1db. This permits the use of a Chebyschev filter for maximum rate of attenuation.

Design the filter.

Solution

It is common practice to design the lowpass version of a particular filter and then transform it into the highpass version at implementation. In this case, we will first design a lowpass filter with a cutoff frequency of 0.5Hz and 80 db attenuation at 2.5Hz. This provides us with a lowpass design where the attenuation is 80db at frequency 5 times the cutoff frequency. Then we will transform this filter into a highpass filter where the attenuation is 80db at 1/5 the cutoff frequency.

1. Determine the ripple factor ϵ .

Using equation 803 we have:

$$1_{db} = -20 \log_{10} \frac{1}{\sqrt{1 + \epsilon^2}}$$

Solving for ϵ , we obtain:

$$\epsilon = 0.5088$$

2. (Optional) Estimate the order N of the filter

The value N must be found by trial and error using the Chebyschev equations 805 through 807. To close in on likely values of N , it helps to have some approximate idea of the filter order.

The Chebyschev filter rolls off at a faster rate than the Butterworth (figure 480). Consequently, we could use the Butterworth magnitude equation 801 (page 554) as a rough guide to the requirements for the Chebyschev.

In this case, we have $f_c/f = 1/5$ for an attenuation of 80db. Then

$$\begin{aligned} 20 \log_{10} |G(w)| &= -80 \text{ db} \\ |G(w)| &= 10^{-4} \end{aligned}$$

Now we can get an indication of N , using equation 801:

$$\begin{aligned} |G(w)| &\approx \left(\frac{\omega_c}{\omega}\right)^N \\ 10^{-4} &= \left(\frac{2\pi f_c}{2\pi f}\right)^N \\ 10^{-4} &= \left(\frac{f_c}{f}\right)^N \end{aligned}$$

Take the \log_{10} of both sides and substitute $f_c/f = 1/5$:

$$-4 = N \log_{10} 1/5$$

and

$$N = 5.7$$

Consequently, we'd need a 6th order Butterworth filter to meet this requirement. That suggests that a 4th or 5th order Chebyschev filter will do the trick.

3. Determine the order N of the Chebyschev filter

We will plug the frequency ratio ω_c/ω and a guess at order N into the highpass version of the Chebyschev amplitude function, and see what attenuation results. If that is insufficient, we'll up the order number and try again.

We previously determined that $\epsilon = 0.5088$

The frequency ratio is $\frac{\omega_c}{\omega} = 1/5$

To start with, we'll try $N = 4$.

Using equation 805,

$$\begin{aligned} |G(jw)| &= \frac{1}{\sqrt{1 + \epsilon^2 \cosh^2 \left[N \cosh^{-1} \left(\frac{\omega}{\omega_c} \right) \right]}} \\ &= \frac{1}{\sqrt{1 + 0.5088^2 \cosh^2 [4 \cosh^{-1} (5)]}} \end{aligned}$$

Using equation 807 for \cosh^{-1} , we have

$$\begin{aligned} \cosh^{-1} 5 &= \ln(5 + \sqrt{5^2 - 1}) \\ &= 2.292 \end{aligned}$$

Substituting that in the calculation of $|G(jw)|$, we have:

$$\begin{aligned}|G(jw)| &= \frac{1}{\sqrt{1 + 0.5088^2 \cosh^2(4 \times 2.292)}} \\ &= \frac{1}{\sqrt{1 + 0.5088^2 \cosh^2(9.169)}}\end{aligned}$$

Using equation 806 for $\cosh x$, we have

$$\begin{aligned}\cosh^2(9.169) &= \left(\frac{e^{9.169} + e^{-9.169}}{2}\right)^2 \\ &= 4797.4^2 \\ &= 23.02 \times 10^6\end{aligned}$$

Back to $|G(jw)|$ again, we have

$$\begin{aligned}|G(jw)| &= \frac{1}{\sqrt{1 + 0.5088^2 \cosh^2(9.169)}} \\ &= \frac{1}{\sqrt{1 + 0.5088^2(23.02 \times 10^6)}} \\ &= 4.096 \times 10^{-4}\end{aligned}$$

In db, this is

$$\begin{aligned}|G(jw)| &= 20 \log_{10} 4.096 \times 10^{-4} \\ &= -67.75 \text{ db}\end{aligned}$$

Evidently, 4 sections will give us an attenuation of about 67db at 5 times the cutoff frequency, which is not sufficient. Repeating this calculation for $N = 5$ yields

$$|G(jw)| = -87.6 \text{ db}$$

which does meet the requirements. Consequently, a 5 section filter is required.

4. Determine the Filter Coefficients

From section 17.24 on page 570, the coefficients for a lowpass, 5th order, 1db Chebyschev filter are:

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
5	0.994	5.556	0.655	1.399	0.289	n/a

The Q factor is not indicated for the third section, so it is the first-order section.

Each of the frequencies ω_1 , ω_2 and ω_3 is given as a ratio of the cutoff frequency for a lowpass filter. For a highpass filter, each value must be inverted, so that $\omega_{hp} = 1/\omega_{lp}$:

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
5	1.006	5.556	1.527	1.399	3.460	n/a

The actual cutoff frequency of the filter is 0.5Hz, so we must multiply each of the frequency ratios in this table by $0.5 \text{ Hz} = 3.14 \text{ radians/sec}$ to get the resonant frequency of each filter section.

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
5	3.158	5.556	4.79	1.399	10.86	n/a

Now we are ready to design hardware.

5. Choose a Filter Configuration

A suitable highpass filter second-order building block is the unity-gain Sallen Key filter, section 17.11, figure 468(a). This filter design is known to have good sensitivity figures. That is, it is stable and not excessively sensitive to component tolerances.

As well, the fact that it is unity-gain (in the passband) between input and output simplifies concerns about signal overload.

This second-order highpass schematic is reproduced in figure 481(a).

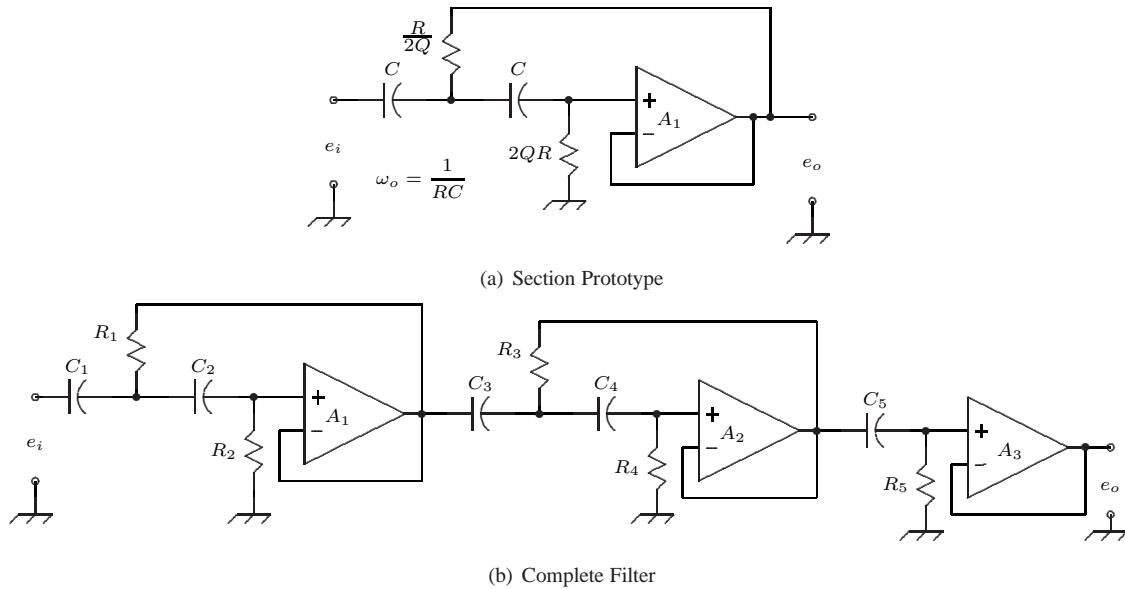


Figure 481: Fifth Order Highpass Filter

The complete fifth-order highpass filter is shown in figure 481(b). There are two second-order sections, followed by a first-order section, followed by a buffer amplifier A_3 . The buffer amplifier is not essential but it prevents the output from being affected by a load resistance.

6. Calculate the Component Values

(a) Calculate the Capacitance

In this particular configuration, the capacitors may be made equal, which is convenient. The frequency domain of this filter is quite low, so the RC product will be large. To avoid excessively large resistances, the capacitance should be chosen to be as large as possible. They are also required to be non-polarized, so a suitable capacitance value is $1\mu\text{F}$.

(b) Calculate the Resistances

From the table, the resonant frequency ω_1 of the first section is

$$\omega_1 = 3.158 \text{ radians/sec}$$

Then, for this section

$$\frac{1}{RC} = 3.158 \text{ radians/sec}$$

The capacitance is $1\mu\text{F}$. Substitute and solve for the resistance:

$$\begin{aligned} R &= \frac{1}{\omega_1 C} \\ &= \frac{1}{3.158 \times (1 \times 10^{-6})} \\ &= 316 \times 10^3 \Omega \end{aligned}$$

Referring to figure 481(a), we have:

$$\begin{aligned} R_1 &= \frac{R}{2Q_1} \\ &= \frac{316 \times 10^3}{2 \times 5.556} \\ &= 28 \times 10^3 \Omega \end{aligned}$$

$$\begin{aligned} R_2 &= 2Q_1 R \\ &= 2 \times 5.556 \times (316 \times 10^3) \\ &= 3.51 \times 10^6 \Omega \end{aligned}$$

Similarly, we use ω_2 and Q_2 to find that $R_3 = 74.6 \times 10^3 \Omega$, $R_4 = 581 \times 10^3 \Omega$.

Finally, we use ω_3 , obtained from the table as 10.86, to calculate R_5 :

$$\omega_3 = \frac{1}{R_5 C}$$

Then

$$\begin{aligned} R_5 &= \frac{1}{\omega_3 C} \\ &= \frac{1}{10.86 \times (1 \times 10^{-6})} \\ &= 92.08 \times 10^3 \Omega \end{aligned}$$

In summary, the resistor and capacitor values are:

R_1	R_2	R_3	R_4	R_5
28k Ω	3.51M Ω	74.6k Ω	581k Ω	92.08k Ω
C_1	C_2	C_3	C_4	C_5
1 μ F	1 μ F	1 μ F	1 μ F	1 μ F

7. Verify the Design with Circuit Simulation

A circuit simulation with ideal op-amps and the exact resistor and capacitor values should yield an amplitude response that is exactly as predicted. In this case, the simulation indicated that:

- The filter gain is -87.8db at 0.1Hz
- The gain is -1db at 0.501Hz
- The ripples in the passband are between 0 and +1db

These results correspond to the predicted values.

8. Do the Final Design Checks

- As shown in figure 17.21 (page 566) for a multi-section Chebyschev filter, the output from the high-Q section (amplifier A_1 in this case) is much larger than the overall output voltage. Assume the maximum input signal and check that the amplifier does not clip at the peak frequency for that section.
- Rerun the simulation with the resistors and capacitors at their nearest standard values and the op-amps replaced by real op-amp models.
- Vary the each component value in a random fashion within its tolerance band and determine the effect on the frequency response.

17.20 The Bessel Characteristic

To this point, we have been concerned primarily with the amplitude response of each higher-order filter. The Bessel filter [164] addresses the issue of phase response.

Consider an AC signal that is passed through a device which delays it by a certain amount irrespective of frequency. (You could picture a loudspeaker separated by some distance from a microphone. Compared to the speaker signal, the microphone signal will be delayed by a constant amount.) Now, as frequency increases, there are more cycles in a given time interval. So a constant time interval delay appears on a phase-frequency plot as a linear characteristic. Both axes must be linear for this to be true, so frequency must be plotted on a linear scale, not logarithmic.

The Butterworth and Chebyschev filters have non-linear phase responses. As a result, different frequencies are delayed by different amounts as they pass through the filter, and the overall shape of a time-domain waveform is changed substantially.

The Bessel filter sacrifices some slope in the transition band in return for a linear phase response in the passband. The result is an approximation to a linear phase-frequency characteristic.

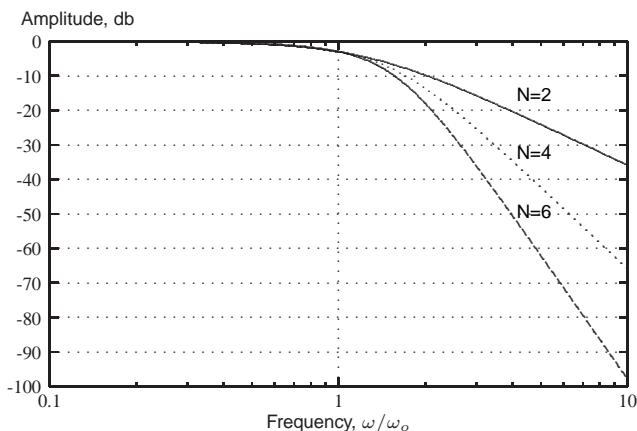


Figure 482: Bessel Magnitude Characteristic

Properties

- The phase response is approximately linear in the passband of the filter, resulting in a constant time delay.
- The time delay increases with the order of the filter.
- The rate of rolloff is less steep than the Butterworth or Chebyschev filters, but better than cascaded first order sections.

Plotting the Bessel Characteristic

It is possible to develop an expression for the Bessel Filter transfer function (see [165], for example). However, it is simpler to use the coefficients tabulated in section 17.24 for the Bessel filter, with the procedure outlined in section 17.16 for plotting high-order filters.

Using that approach, we obtain the amplitude response curves shown in figure 482 above.

The phase curves are shown in figure 483. Notice that the phase-frequency characteristic is much closer to a straight line than the Butterworth or Chebyschev filters.

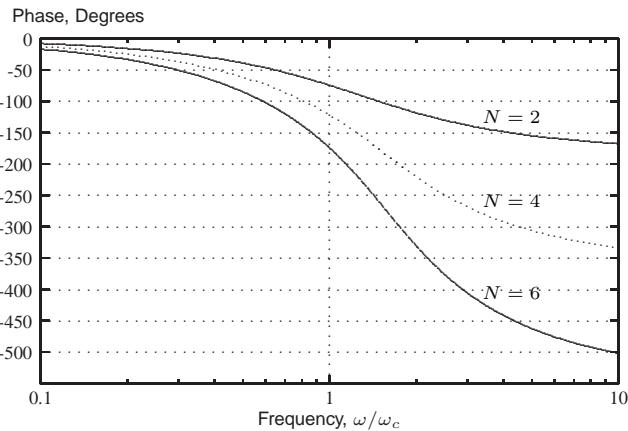


Figure 483: Bessel Phase Characteristic

17.21 Plotting High-Order Filter Characteristics

In the land of the Bode plot, multiplication becomes addition. Consequently, a Bode plot for the magnitude and phase response of a high-order filter may be obtained by plotting the responses of the individual first and second order sections, and adding the result. This is tedious to do manually but is straightforward with the help of a computer plotting program.

Example: Fourth Order, 3db Chebyschev Section

As an example, we'll plot the characteristic of a fourth order 3db ripple Chebyschev lowpass filter. The Chebyschev filter characteristic is discussed in detail in section 17.19. For this discussion you need only know that the frequency response is as shown in figure 480 on page 557.

The fourth order Chebyschev filter will be composed of two second-order sections in series. The values for Q and ω_o for the two sections are obtained from the table in section 17.24 for the Chebyschev filter, 3db ripple, $N=4$.

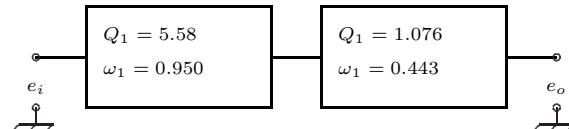


Figure 484: Fourth Order Chebyschev Filter

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
4	0.950	5.58	0.443	1.076		

Figure 484 shows a block diagram of the filter.

Magnitude

The magnitude function for a second order lowpass filter, in a form suitable for a Bode plot, is given by equation 724 in section 17.4:

$$|G(w)| = -10 \log_{10} \left[(1 - x^2)^2 + \left(\frac{x}{Q} \right)^2 \right] \quad (808)$$

where $x = \omega/\omega_o$.

We have one more detail to take care of. The values ω_1 and ω_2 are equal to ω_o/ω_c , where ω_o is the resonant frequency of the second-order section and ω_c is the cutoff frequency of the composite filter. We'd like the responses to be plotted in terms of ω/ω_c , so we have to modify equation 808. For example, in the first of the sections, wherever x occurs, we replace it by x/ω_1 . Then x represents ω/ω_c .

Substitute the values for Q_n and ω_n in equation 808, once for each second-order section, and add the results to get the composite filter magnitude response. Then the magnitude response for the 4th order, 3db ripple Chebyschev lowpass filter is given by:

$$|G(w)| = -10 \log_{10} \left[\left(1 - \left(\frac{x}{\omega_1} \right)^2 \right)^2 + \left(\frac{x}{\omega_1 Q_1} \right)^2 \right] - 10 \log_{10} \left[\left(1 - \left(\frac{x}{\omega_2} \right)^2 \right)^2 + \left(\frac{x}{\omega_2 Q_2} \right)^2 \right] \quad (809)$$

Providing the coefficients are available, this formula can be extended to any order as required.

The individual second-order responses and their sum, the fourth-order Chebyschev filter, are shown in figure 485.

Notice that the individual sections have quite different cutoff frequencies and Q factors. The resultant sum is the same as the magnitude plot shown in figure 480, section 17.19, which was derived directly from the transfer function for the Chebyschev filter.

These plots indicate one potential problem with a composite filter. Section 1 of the filter has a large Q factor of 5.58. This will tend to magnify any frequency components in the region of ω_o for that section. This section would overload and clip the signal before Section 2. Consequently, even though the composite filter characteristic is not high Q, individual sections may be, and they should be checked for overload under large signal conditions.

Phase

The phase of a second-order lowpass section is given as equation 727 in section 17.4:

$$\angle G(\omega) = \tan^{-1} \left(\frac{x/Q}{1 - x^2} \right) \quad (810)$$

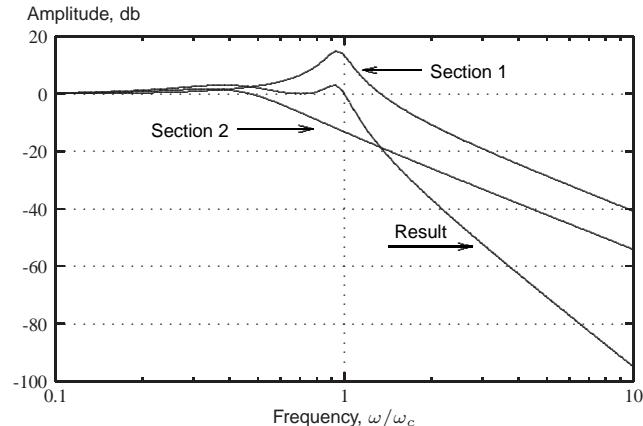


Figure 485: Chebyschev Components and Result

As we did previously, to generate a plot centred on ω_c rather than ω_o , we must replace x by x/ω_1 and so on. The composite phase curve is simply equal to the sum of the two individual phase curves, so that we have for the Chebyschev fourth order 3db ripple lowpass filter the following equation for phase vs frequency:

$$\angle G(\omega) = \tan^{-1} \left(\frac{\frac{x}{\omega_1 Q_1}}{1 - \left(\frac{x}{\omega_1} \right)^2} \right) + \tan^{-1} \left(\frac{\frac{x}{\omega_2 Q_2}}{1 - \left(\frac{x}{\omega_2} \right)^2} \right) \quad (811)$$

17.22 First-Order Transient Response

To this point, we have been focusing on the frequency response of various filters. However, there are occasions when the time domain response is of interest. Once a filter is designed, it can be simulated using some variant of the Spice program (Multisim, for example) to determine the step response. However, it's useful to have some ideas in hand to be able to modify the transient response.

A collection of common first and second-order *low Q* filters is shown in the table on page 568. Each entry shows a representative circuit, frequency-domain amplitude response, and time-domain step response.

The buffer amplifiers are assumed to have unity gain. They provide a high impedance load to the preceding stage and a low impedance source to the following stage.

The equations are shown in terms of the transition frequencies. For those who prefer to work with time constants, $\tau = 1/\omega_o$.

The bandpass response is particularly interesting. It would apply to circuits such as those shown in section 17.3. The transient response starts with an exponential rise that is defined by the lowpass time constant $\tau_2 = 1/\omega_2$, and then transitions to a much longer exponential decay defined by the highpass time constant $\tau_1 = 1/\omega_1$.

This same response would apply to an AC coupled amplifier (such as an audio amplifier), where the low-frequency cutoff corresponds to ω_1 and the high-frequency cutoff to ω_2 .

Reference

The entries in the table of page 568 are based on work in [166], which shows 18 frequency-domain amplitude responses with their corresponding time-domain step response.

17.23 Second-Order Transient Response

Unlike a first-order filter, the step response for a second-order filter can include overshoot and oscillation (aka *ringing*). Representative step responses for the second-order lowpass filter are shown in figure 486.

Ideally, the response would exactly follow the input step waveform. That is, the output should transition from 0 to 1 immediately and stay at 1 thereafter. In practice, the output shows overshoot and ringing that depends on the Q factor.

The chart on page

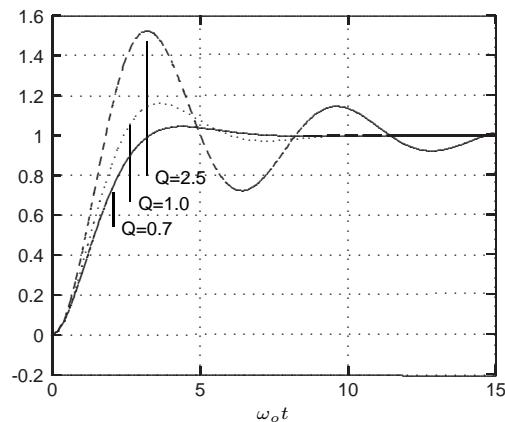
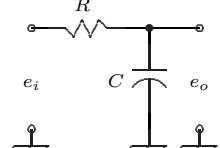
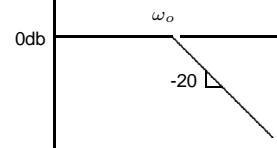
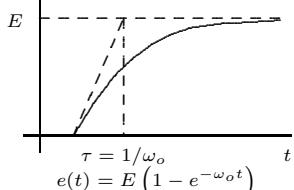
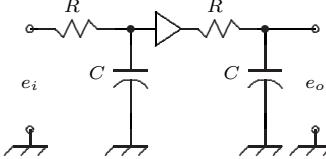
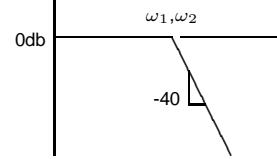
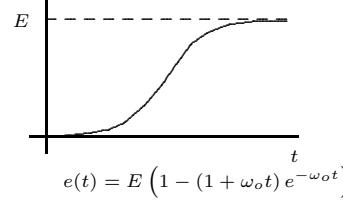
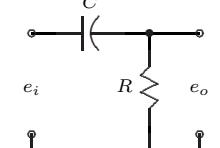
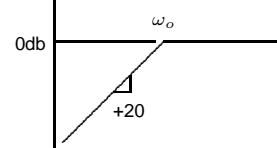
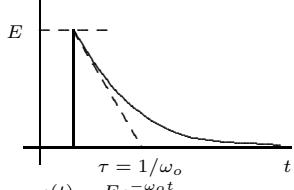
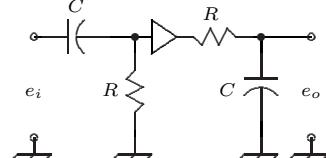
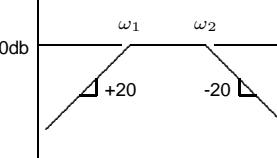
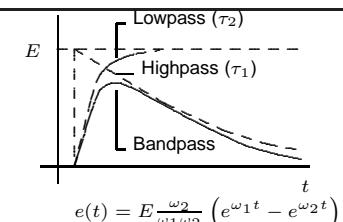


Figure 486: Second-Order Lowpass, Transient Response

Basic Type	Circuit	Amplitude	Step Response
Lowpass			
LP Cascade			
Highpass			
Bandpass			 <p style="text-align: center;"> Lowpass (τ_2) Highpass (τ_1) Bandpass </p> <p style="text-align: center;"> $e(t) = E \frac{\omega_2}{\omega_1 \omega_2} (e^{\omega_1 t} - e^{\omega_2 t})$ </p>

Other features:

- For a Q factor of $\sqrt{2}$, the overshoot is about 5%, and this is often taken as an ideal.
- Increasing Q factor (or reducing damping, the controls-system people would say) increases the overshoot.
- The oscillations that occur at high Q factor are approximately equal to ω_o , the resonant frequency in a second-order system. (To be precise, the oscillation frequency is called the *damped resonant frequency*, and it is somewhat less than ω_o , which is the *undamped resonant frequency*, ie, the frequency that occurs when the damping is zero, and the Q factor infinite. This is usually a minor detail.)
- The horizontal axis is normalized by making it equal to $\omega_o t$, which has the units *radians*. One complete cycle corresponds to $\omega_o t = 2\pi$. This graph then applies to a second order system with any resonant frequency.

These curves were generated by the equation

$$c(t) = 1 - \frac{\exp^{-\delta\omega_o t}}{\sqrt{1-\delta^2}} \sin\left(\sqrt{1-\delta^2}\omega_o t + \cos^{-1} \delta\right) \quad (812)$$

where

- $c(t)$ is the response to a step function
- δ is the damping factor, equal to $1/2Q$
- ω_o is the undamped natural frequency of the system

The independent variable (x axis) is equated to $\omega_o t$ and then $c(t)$ plotted vs x.

(Most textbooks on control systems have similar information. This is taken from Savant [87].)

Higher Order Responses

The transient response of a higher-order filter is complex and best dealt with by simulation, either of the transfer function using a mathematical toolbox (such as Matlab or Maple), or by simulation of the circuit itself. However, by looking at the step response of a second order section, we can generate some rules of thumb that will guide us.

A comparison of step responses for second-order filters, for the Chebyschev, Butterworth and Bessel characteristics is shown in figure 487.

Harkening back to the frequency responses for these filters, we can see that the transient response is related to the steepness of the filter in the transition band. *Steeper cutoff in the frequency response is associated with greater overshoot and ringing in the transient response.*

Figure 487 shows this comparison for second-order filters: it's clear that the Chebyschev response has significant

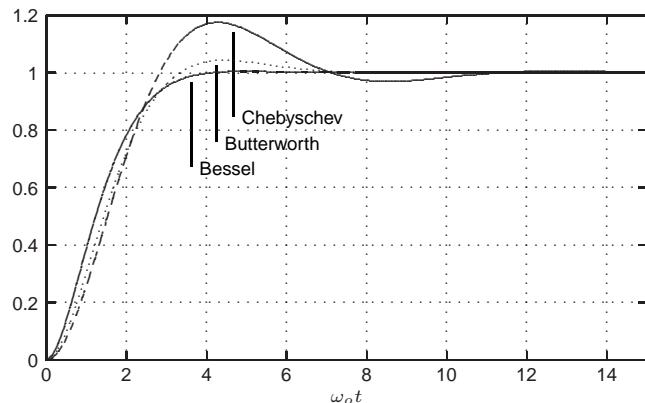


Figure 487: Chebyschev, Butterworth and Bessel Step Response

overshoot and ringing, and the Bessel response has none. Consequently, if transient response is important, the Bessel filter characteristic is a suitable choice.

Both the Butterworth and Chebyschev filters exhibit increasing delay as the order of the filter increases. However, the Bessel filter exhibits the same delay regardless of the order of the filter, with slightly decreasing rise time as the order increases. Figure 488 conveys the concept. For an exact rendition, see [165] or [149].

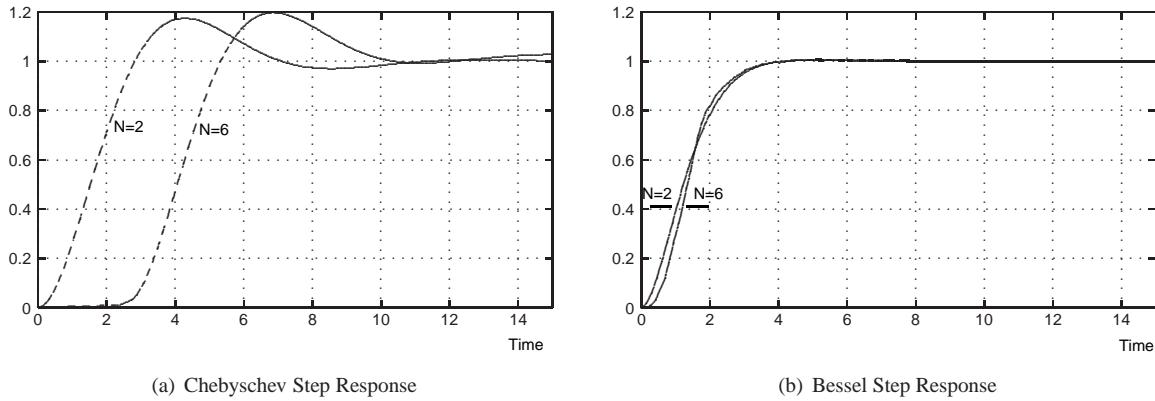


Figure 488: High-Order Step Responses

17.24 Filter Tables

Reading the Tables

In this section, I have tabulated the values of Q and ω_o for the second order filters used in high-order lowpass filters. Franco [139] shows the data in this form.

To obtain the highpass data, simply invert the values shown for ω_n for that entry. (An example is shown in section 17.19, page 559.)

It's useful to be able to derive these from the forms of the tables that are found in other references. In several cases such as [167] and [148], the values tabulated are in the form of *damping coefficient d*, where

$$Q = \frac{1}{d} \quad (813)$$

I limited the table entries to sixth order maximum, which corresponds to a cascade of 3 second-order sections. It's unusual in the practice of designing active filter circuits to require a more elaborate filter than sixth order. Furthermore, the component tolerances become progressively more stringent (and impractical) as the order of the filter increases¹⁷⁶.

If a higher order filter is required, then it may be worthwhile to revisit the system design and see if the approach can be changed to eliminate the need for such a complex filter.

For example, the original audio A/D converter systems sampled at a frequency that was very close to the $2f_{max}$ Nyquist aliasing frequency and as a consequence required very steep and elaborate anti-aliasing lowpass filters. Subsequently, digital technology evolved to the point where a much higher sampling rate could be used, and over-sampling A/D converters were created. These require only a very simple anti-aliasing filter.

¹⁷⁶It is much more common to require high-order filters in radio frequency design. However, these are implemented with inductors and capacitors, so the design methods are somewhat different.

Section frequency and Cutoff Frequency

We have been using the symbol ω_o to represent the cutoff frequency (3db point) for a first order section and for the resonant frequency of a second-order section. A high order filter has an aggregate cutoff frequency which I will refer to as ω_c . The resonant frequencies of the second-order sections that make up this filter may or may not be equal to ω_c .

For the Butterworth filter, $\omega_o = \omega_c$. For the Chebyschev filter, the filters of the individual second-order sections are staggered around ω_c , so in general ω_o is not equal to ω_c .

Butterworth Coefficients

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	1	0.707				
3	1	1.0	1			
4	1	0.541	1	1.306		
5	1	0.618	1	1.620	1	
6	1	0.518	1	0.707	1	1.932

Chebyschev Coefficients

In the case of the Chebyschev coefficients, the values given are the resonant frequency ω_o of the second order section, as a ratio of the aggregate cutoff frequency ω_c of the completed higher-order filter. In the tables, ω_c is assumed to be unity.

0.1db Ripple

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	1.820	0.767				
3	1.300	1.341	0.969			
4	1.153	2.183	0.789	0.619		
5	1.093	3.282	0.797	0.915	0.539	
6	1.063	4.633	0.834	1.332	0.513	0.599

From [139]

0.5db Ripple

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	1.2313	0.863				
3	1.0689	1.706	0.6265			
4	1.0313	2.9403	0.5970	0.705		
5	1.0177	4.545	0.6905	1.1778	0.3623	
6	1.0114	6.515	0.7681	1.810	0.3962	0.6836

Adapted from [167]

1.0 db Ripple

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	1.050	0.957				
3	0.997	2.018	0.494			
4	0.993	3.559	0.529	0.785		
5	0.994	5.556	0.655	1.399	0.289	
6	0.995	8.004	0.747	2.198	0.353	0.761

Adapted from [167]

2.0 db Ripple

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	0.907	1.129				
3	0.941	2.552	0.369			
4	0.964	4.594	0.471	0.929		
5	0.976	7.232	0.627	1.775	0.218	
6	0.983	10.462	0.730	2.844	0.316	0.902

From [168]

3.0 db Ripple

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	0.841	1.03				
3	0.916	3.06	0.299			
4	0.950	5.58	0.443	1.076		
5	0.967	8.849	0.614	2.136	0.178	
6	0.975	12.78	0.722	3.460	0.298	1.044

Adapted from [148]

Bessel Coefficients

n	ω_1	Q_1	ω_2	Q_2	ω_3	Q_3
2	1.2736	0.577				
3	1.4524	0.691	1.327			
4	1.5912	0.805	1.419	0.522		
5	1.7607	0.916	1.561	0.564	1.507	
6	1.9071	1.023	1.691	0.611	1.606	0.510

17.25 Design Method Overview

Now that the building blocks of an active filter are in hand, here is an overview of the design of an active filter.

Usually, the filter designer has in mind a particular filter function - lowpass, highpass, bandpass or band reject. The designer must then choose a general approach to the filter configuration. Some common and practical approaches to this decision are as follows:

Lowpass Filter The lowpass filter can consist of a cascade of one or more second-order lowpass sections. The Sallen-Key second-order lowpass circuit is suitable for implementing the filter sections.

Highpass Filter The highpass filter is closely related to the lowpass configuration and, like the lowpass, can be implemented as a cascade of second-order highpass sections using the Sallen-Key second-order highpass circuit.

Bandpass Filter The designer has several choices in the configuration for a bandpass filter.

- If the shape of the filter must be carefully controlled or it requires steep transitions between the pass and stop bands, it can be implemented with a highpass filter with cutoff frequency f_l followed by a lowpass filter with cutoff frequency f_u , where $f_u > f_l$. The filters would be implemented as a cascade of Sallen-Key lowpass and highpass sections.

For example, this is a popular approach for an audio octave or third-octave frequency analyser. The analyser consists of a bank of narrow bandpass filters with steep skirts and very closely specified cutoff frequencies.

- If a narrow bandwidth is required, then a high-Q second-order bandpass filter may be appropriate. This can be implemented with a state-variable or *biquad* circuit, a near-relative of the state variable filter. A high-Q bandpass filter is particularly suitable if the filter has a constant Q factor and must be tuned over a range of frequencies. Although each filter section requires 3 or 4 op-amps, the performance and stability are excellent. Op-amps are very inexpensive components these days, so it's a reasonable choice to use several of them if it minimizes the cost of other components.

While it's technically possible to cascade two or more high-Q second-order bandpass filters, it would require high-accuracy components to align the centre-frequencies of the two sections. For the same reason, it would be difficult to create a design that would successfully track over a range of frequencies. As a consequence, most filters of this design are single second-order sections.

- If there is a need to economize on parts count and the shape of the filter is not particularly critical (a 20db/decade rolloff is acceptable), then the single-amplifier, multiple feedback circuit shown in figure 497 of the exercises may be suitable. Since the Q of this bandpass filter is very sensitive to amplifier gain [78], it should not be used with Q factors greater than 10.

Band Reject Filter As in the case of the bandpass filter, it is possible to use a cascade of lowpass and highpass filters to pass low frequencies, block a range of mid-frequencies, and pass high frequencies. Alternatively, when a single frequency must be removed from a signal with minimum impact on the rest of the spectrum (removing a 120 Hz interfering signal is a common application) a high-q notch filter can be tuned to remove a narrow band of frequencies. There are both passive and active versions of the notch filter.

These are approaches that will work in most circumstances and are commonly used in real designs. However, this does not preclude other approaches. For example, Sedra & Smith [78] show how the Antoniou Inductance Simulation circuit (section 18.5) can be used to fabricate any of the second-order filter sections.

17.26 The Complex Frequency Plane

The material that has been presented to this point is sufficient for the design of many useful active filters. This section fills in certain additional details on filter design and provides some vocabulary for those who wish to pursue the topic.

17.26.1 Poles and Zeros

For a network composed of components that are describable by ordinary differential equations (eg, the inductor, capacitor and resistor), the transfer function $G = v_o/v_i$ can be written as the ratio of two nth order polynomials in s :

$$\frac{v_o}{v_i} = \frac{a_m s^m + a_{m-1} s^{m-1} + \dots + a_1 s + a_o}{b_n s^n + b_{n-1} s^{n-1} + \dots + b_1 s + b_o} \quad (814)$$

If the numerator and denominator are factored, then we can obtain a function of the form:

$$\frac{v_o}{v_i} = \frac{a_m}{b_n} \frac{(s - z_1)(s - z_2)\dots(s - z_m)}{(s - p_1)(s - p_2)\dots(s - p_n)} \quad (815)$$

where $H_o = a_m/b_n$ is the *scaling factor*.

When $s = z_1$ or z_2 (etc), then the function goes to zero. The z_i values are consequently called the *zeros* of the function. Similarly, when $s = p_1$ or p_2 (etc), then the function becomes infinite. The p_i values are called the *poles* of the function.

In Bode plot terms, a zero results in an additional upward bend of the amplitude function and an additional phase lead. A pole results in an additional downward bend and additional phase lag. The terms *zero* and *pole* have become synonymous with *break frequency upward on the Bode plot* and *break frequency downward on the Bode plot*.

17.26.2 The Second-Order Lowpass Filter, Normalized

In section 17.4 we developed equation 720 for the second-order lowpass filter:

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (816)$$

It is common practice to *normalize* this equation by specifying a natural-resonant frequency ω_o of 1 radian per second. This translates the filter to the frequency region around 1 radian per second without changing the shape, and leads to certain mathematical conveniences. Then equation 816 becomes:

$$\frac{e_o}{e_i} = \frac{1}{\left(s^2 + \frac{1}{Q}s + 1\right)} \quad (817)$$

In order to determine the poles and zeros of this equation, we need to factor the numerator and denominator. In this case, the numerator has no zeros, so we can focus on the denominator.

Applying the quadratic formula to the denominator of equation 817 to find the roots (which are also the poles), we have:

$$\begin{aligned} p_1, p_2 &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-\left(\frac{1}{Q}\right) \pm \sqrt{\left(\frac{1}{Q}\right)^2 - 4(1)(1)}}{2(1)} \\ &= -\frac{1}{2Q} \pm \sqrt{\frac{1}{4Q^2} - 1} \end{aligned} \quad (818)$$

Consequently, the poles p_1 and p_2 of the lowpass filter $G(s)$ are given by equation 818. Equation 817 could be written as:

$$\begin{aligned}
\frac{e_o}{e_i} &= \frac{1}{\left(s^2 + \frac{1}{Q}s + 1\right)} \\
&= \frac{1}{(s - p_1)(s - p_2)} \\
&= \frac{1}{\left(s - \frac{1}{2Q} + \sqrt{\frac{1}{4Q^2} - 1}\right) \left(s - \frac{1}{2Q} - \sqrt{\frac{1}{4Q^2} - 1}\right)}
\end{aligned} \tag{819}$$

We could play with this, but instead let's plug some numbers into the denominator function of equation 817 and see what results. The factors of the denominator depend on the value of the Q factor. To get a feel for the function, we will vary the value of Q , starting with $Q = 0.25$ and working upwards.

$$\begin{aligned}
s &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
&= \frac{-1/0.25 \pm \sqrt{(1/0.25)^2 - 4}}{2} \\
&= -3.73, -0.27
\end{aligned} \tag{820}$$

Similarly, for $Q = 0.4$ we have $s = -2, 0.5$.

Notice the pattern: we have two real roots which are moving towards each other. At $Q = 0.5$, an interesting thing happens: they coincide at $s = -1$.

Now the situation changes. For $Q = 1$ we have the following:

$$\begin{aligned}
s &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
&= \frac{-1/1 \pm \sqrt{(1/1)^2 - 4}}{2} \\
&= \frac{-1 \pm \sqrt{-3}}{2} \\
&= -0.5 \pm j0.86
\end{aligned} \tag{821}$$

In other words, there are two complex roots, conjugates of each other. Similarly, for $Q = 2$ we have $s = -0.25 \pm j0.968$.

In summary, the roots we have found are:

Pole Number	Q Factor	Part A	Part B
1	0.25	-3.37	-0.27
2	0.4	-2	-0.5
3	0.5	-1	n/a
4	1	-0.5+j0.86	-0.5-j0.86
5	2	-0.25+j0.968	-0.25-j0.968

Next, we'll plot these on the *Complex Frequency Plane*.

17.26.3 Diagram of the Complex Frequency Plane

Now it's time for a diagram. An AC voltage or current can be treated as a rotating vector, which in turn can be represented by a complex number. Then we can plot the vector on a cartesian graph in which the horizontal axis is the real part and the vertical axis the imaginary part¹⁷⁷. A similar approach will work here.

The plotting area is the *complex frequency plane*. The horizontal axis represents the magnitude of the real part, the vertical axis the magnitude of the imaginary part. Plotting the real poles from our table as crosses, we have figure 489(a). As Q increases, these poles move together until they coincide at $Q = 0.5$. As Q increases beyond 0.5, the poles become complex and move vertically on the plane.

Similarly, the zeros of the function can be plotted on this same plane.

17.26.4 Zeros at Infinity

Certain transfer functions become zero as the frequency becomes very large. Then the function is said to have *zeros at infinity*. For example, consider the normalized bandpass filter, which is given by:

$$\frac{e_o}{e_i} = \frac{s}{\left(s^2 + \frac{1}{Q}s + 1\right)} \quad (822)$$

Substitute $j\omega = s$ and let ω become very large. Then:

$$\begin{aligned} \frac{e_o}{e_i} &\rightarrow \frac{s}{s^2} \\ &= \frac{1}{s} \\ &\rightarrow 0 \end{aligned} \quad (823)$$

We say that this function *has a zero at infinity*.

Similarly, consider the normalized lowpass filter of equation 817:

$$\frac{e_o}{e_i} = \frac{1}{\left(s^2 + \frac{1}{Q}s + 1\right)} \quad (824)$$

Substitute $j\omega = s$ and let ω become very large. Then:

$$\begin{aligned} \frac{e_o}{e_i} &\rightarrow \frac{1}{s^2} \\ &\rightarrow 0 \end{aligned} \quad (825)$$

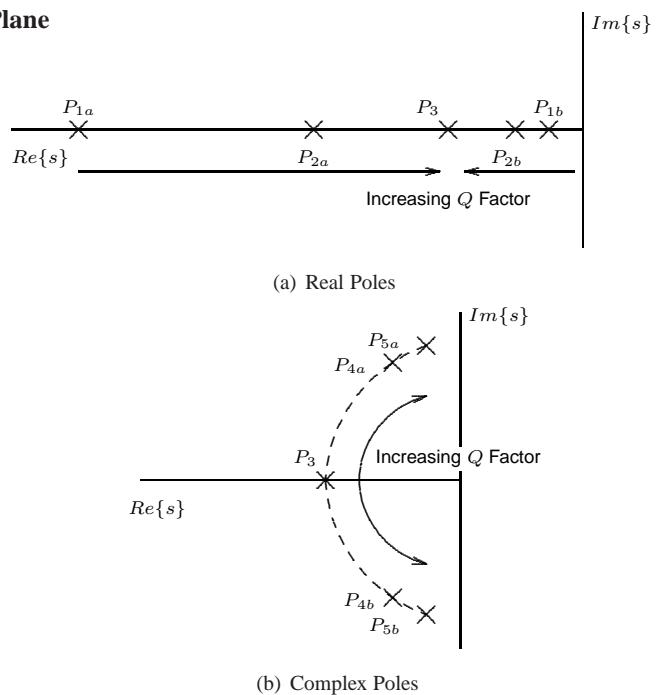


Figure 489: Pole Migration

¹⁷⁷The term *imaginary* is unfortunate, since it implies that this component is somehow less real than the real component. A better term for the vertical component would be the *quadrature* component. Quadrature implies *at right angles*, which is exactly correct.

This function is said to have *has two zeros at infinity*.

17.26.5 The Membrane Model

The number and position of the poles and zeros defines the shape of the amplitude and phase responses of the system.

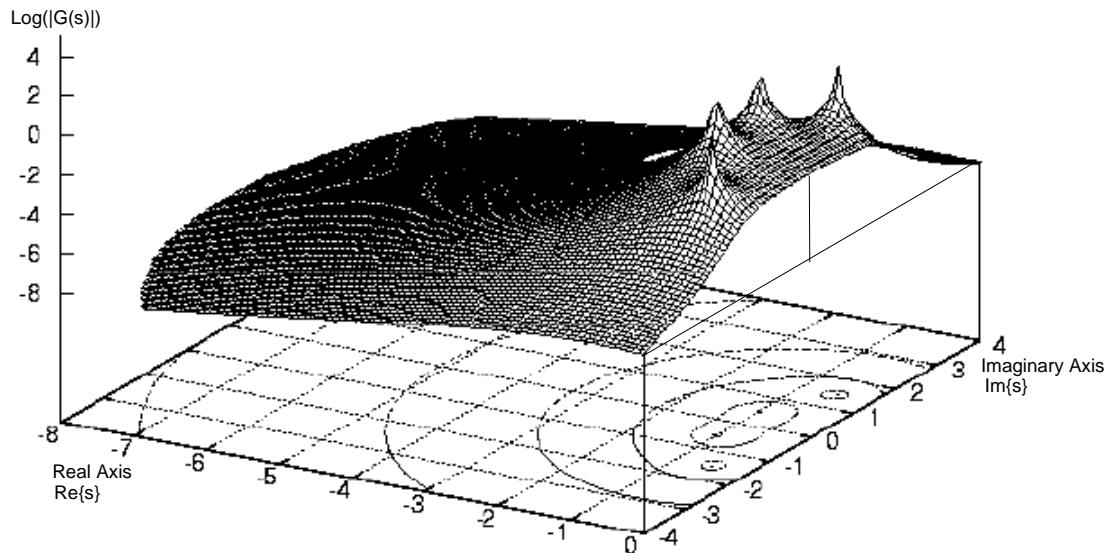


Figure 490: Membrane Model, Lowpass

The effect of the poles and zeros may be visualized using the *membrane model*, as shown in figure 490 for a 4-pole lowpass filter. Imagine that the s -plane is covered by a flexible membrane, the height of which is proportional to the magnitude of $G(s)$.

In figure 490, there are two complex-pole pairs: one pair at -0.5 ± 1.5 , the second pair at -1 ± 0.5 . (Their locations are also visible on the contour map at the base.)

At the pole locations, the height goes to infinity which provides some height to the membrane. At a zero location the height goes to zero¹⁷⁸. As indicated above in equation 825, the second-order lowpass function has two zeros at infinity, so a fourth order lowpass function would have four of them. These *zeros at infinity* nail the membrane to the zero plane, which forces the membrane to assume a tent shape. (The membrane diagram for a highpass filter is nailed to the zero plane at the origin, is raised by poles outside that area, and then floats above the zero plane thereafter.)

The cross section through the membrane, along the imaginary ($j\omega$) axis, is the amplitude response of the filter. It's symmetric around the mid point of the axis, so either half will do.

This model helps to visualize the effect of poles and zeros. For example, as a pole-pair moves closer to the imaginary axis, it causes a peak to appear in the frequency response. Depending on their position, more poles will fill in the gaps between these peaks or put more peaks in the frequency response. A zero close to the axis puts a dip in the frequency response because the membrane is nailed down at that point. For example, a zero at the mid-point of the imaginary axis creates a highpass filter characteristic, as shown in figure 491.

¹⁷⁸This is true for the usual case, in which the height is a linear function of $G(s)$. If the vertical scale is logarithmic (ie, db), then a zero goes to minus infinity.

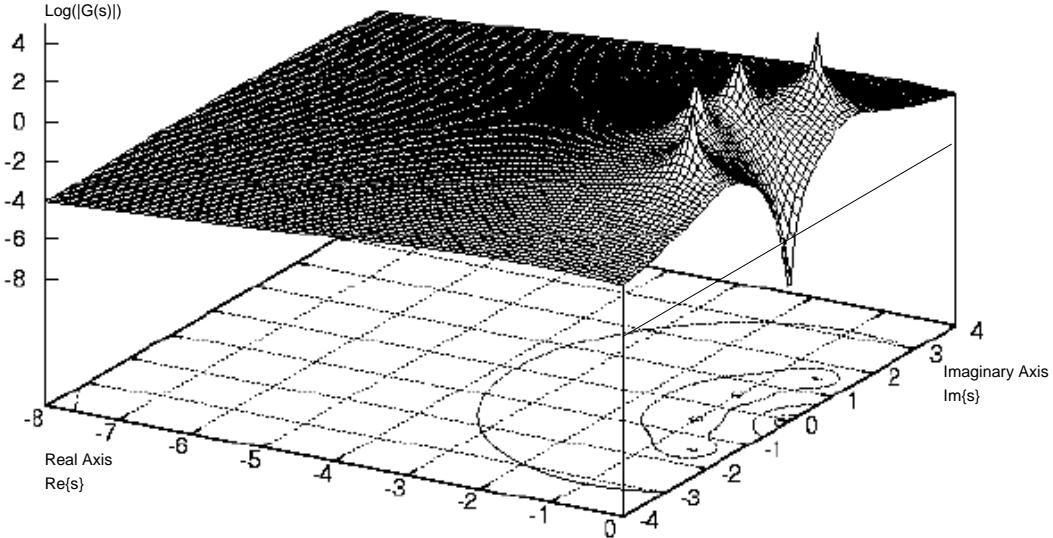


Figure 491: Membrane Model, Highpass

17.26.6 Amplitude and Phase Response from the Complex Frequency Plane

The membrane model is a useful tool for visualization of the amplitude response, but it does not generate quantitative results. Now we will look at a technique using pole and zero position to generate values for amplitude and phase. As usual, we'll start with a specific example and generalize from that.

Consider that we have a second-order normalized lowpass function that can be written as

$$G(s) = \frac{1}{(s - P_1)(s - P_2)} \quad (826)$$

This function has no zeros in the numerator and 2 poles in the denominator. Suppose the two poles are complex (and therefore conjugates), and that they are given by:

$$\begin{aligned} P_1 &= -\alpha + j\omega_1 \\ P_2 &= -\alpha - j\omega_1 \end{aligned} \quad (827)$$

Then they are located on the complex frequency plane at $P_1(-\alpha, +\omega_1)$, $P_2(-\alpha, -\omega_1)$, as shown in figure 492 on page 579.

A particular test frequency is shown as $j\omega$ on the imaginary axis. We will show that the product of the vectors L_1, L_2 is equal to the denominator of equation 826. That is, *the amplitude and phase of the function $G(j\omega)$ is the inverse of the product $L_1 \cdot L_2$* .

That being the case, we can slide the frequency $j\omega$, starting at the X-axis and moving vertically along the Y-axis. As we do so, the inverse of the product of the two vectors L_1, L_2 gives us the amplitude and phase of the filter transfer function.

Here's the proof: Start with the denominator of equation 826 and substitute the values of P_1 , P_2 from equations 827:

$$\begin{aligned}(s - P_1) \cdot (s - P_2) &= \{s - (-\alpha + j\omega_1)\} \cdot \{s - (-\alpha - j\omega_1)\} \\ &= \{s + \alpha - j\omega_1\} \cdot \{s + \alpha + j\omega_1\}\end{aligned}\quad (828)$$

Substitute $j\omega$ for s :

$$\begin{aligned}(s - P_1) \cdot (s - P_2) &= \{j\omega + \alpha - j\omega_1\} \cdot \{j\omega + \alpha + j\omega_1\} \\ &= \{\alpha + j(\omega - \omega_1)\} \cdot \{\alpha + j(\omega + \omega_1)\}\end{aligned}\quad (829)$$

Now, if you examine figure 492, you can see that the vectors L_1 , L_2 are defined on the complex plane as:

$$\begin{aligned}L_1 &= \alpha + j(\omega - \omega_1) \\ L_2 &= \alpha + j(\omega + \omega_1)\end{aligned}\quad (830)$$

For example, the real part (horizontal component) of L_1 is α . The imaginary part (vertical component) of L_1 is $\omega - \omega_1$. Consequently, we can rewrite equation 826 as:

$$\begin{aligned}Gs &= \frac{1}{(s - P_1)(s - P_2)} \\ &= \frac{1}{L_1 \cdot L_2}\end{aligned}\quad (831)$$

where L_1 , L_2 are as defined in equation 830, that is, as shown in figure 492. We can also write L_1 , L_2 in polar notation form, where:

$$\begin{aligned}L_1 &= |L_1| \angle \theta_1 \\ L_2 &= |L_2| \angle \theta_2\end{aligned}\quad (832)$$

Then the function can be written as:

$$\begin{aligned}G(j\omega) &= \frac{1}{(|L_1| \angle \theta_1)(|L_2| \angle \theta_2)} \\ &= \frac{1}{|L_1| \cdot |L_2| \angle (\theta_1 + \theta_2)}\end{aligned}\quad (833)$$

In words, the denominator is equal to the product of the lengths of vectors and the sum of their angles.

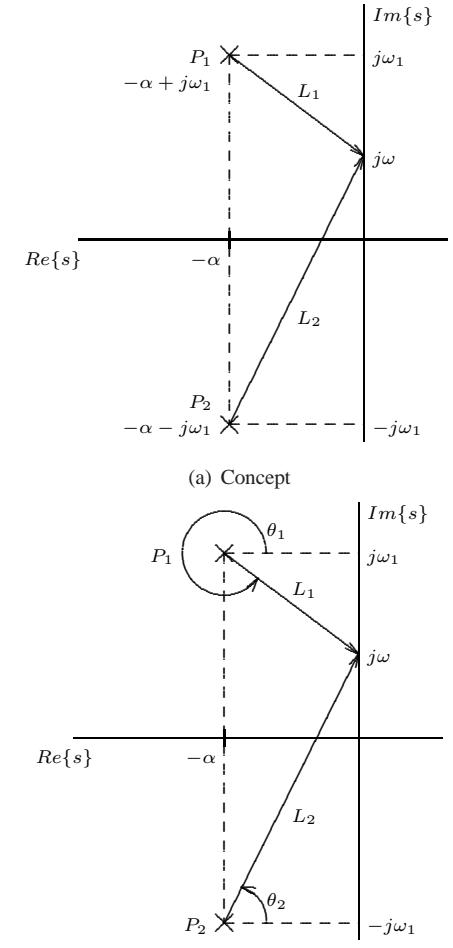


Figure 492: Response Calculation

17.26.7 Generalizing the Amplitude and Phase Response

The zeros of the function apply to the numerator. Apart from that, they are treated in the same fashion as the poles. With both zeros and poles in the function, the function is then given by:

$$G(j\omega) = \frac{(|M_1|\angle\gamma_1)(|M_2|\angle\gamma_2)\dots}{(|L_1|\angle\theta_1)(|L_2|\angle\theta_2)\dots} \quad (834)$$

where $|M|$ is the length of a zero vector and γ is the angle of that vector.

Figure 493 shows an example, where the zeros Z_1, Z_2 are on the right-half plane of the diagram. (They can be on the left-half plane, the right-half plane, or on the $j\omega$ axis. For a stable system, the poles must be on the left-half plane.)

To determine the complete amplitude and phase response of a system, equation 834 must be evaluated at different frequencies, that is, as the $j\omega$ point slides along the vertical axis¹⁷⁹.

This method of calculating frequency and phase response is the basis of a computer tool that allows one to place poles and zeros on the complex frequency plane. The program then automatically calculates and displays the corresponding amplitude and phase response. For example, see [170].

17.26.8 Example: The Highpass Filter

With the preceding calculation method in mind, it is possible to imagine an amplitude and phase plot by inspection of the corresponding pole-zero pattern.

The pole-zero table on page 582 of section 17.27 contains information about the single-pole highpass filter. For this filter there is a single real pole and a zero at the origin, as shown in figure 494. Let's see how the amplitude and phase response are derived from the pole-zero configuration.

- The frequency sweeps from a low value near the origin vertically along the $j\omega$ axis. There are two vectors in play: the pole vector L and the zero vector M . The magnitude of the frequency response is simply the ratio of these two, M/L . The phase is the difference between their angles, $\angle M - \angle L$.
- **Low Frequency** When $j\omega$ is at the origin, the value of M is zero, the value of L is 1, so the function value M/L is zero. At low frequency M is still small so the function magnitude will be small. As the frequency increases and the $j\omega$ point moves up along the axis, M is increasing while L continues at approximately equal to 1, so the function value increases.

At the origin, $\angle M = 90^\circ$ and $\angle L = 0^\circ$ so the function phase $\angle M - \angle L$ is $+90^\circ$ degrees. As the frequency increases and the $j\omega$ point moves up along the axis the value of $\angle M$ stays constant at $+90^\circ$ while the value of $\angle L$ is increasing. Consequently, the angle of the function decreases.

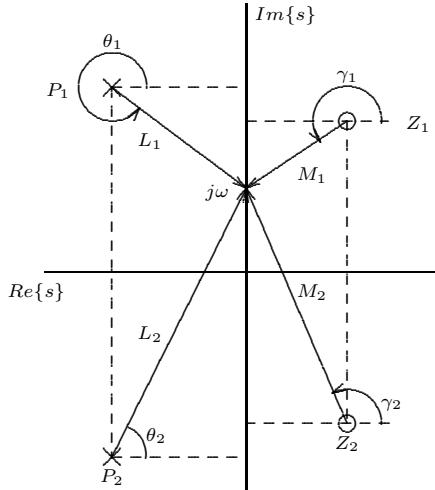


Figure 493: Pole-Zero Response

¹⁷⁹In pre-computer days, a device called the *spirule* – a mixture of protractor and ruler – helped with these calculations [169].

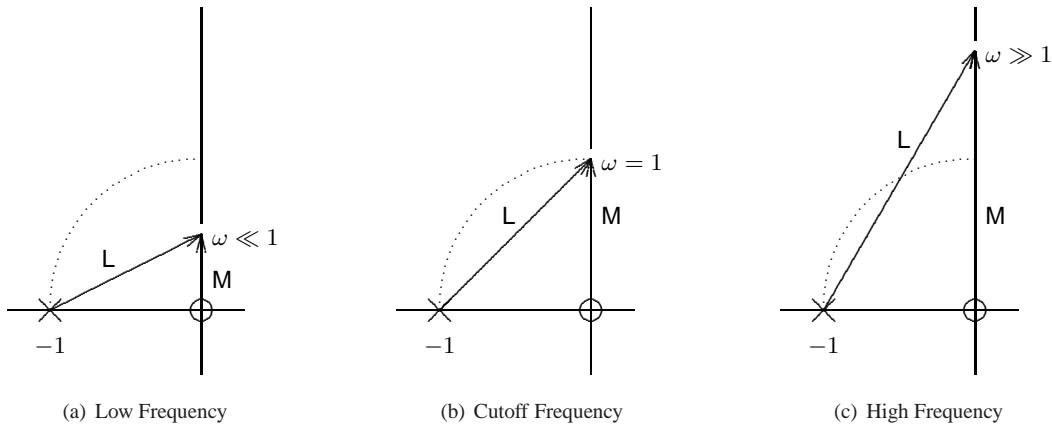


Figure 494: Highpass Filter Response from Pole-Zero Pattern

- **Mid Frequency, $\omega = 1$** At a frequency of 1 radian per second, which is the cutoff frequency of the frequency normalized highpass filter, the $j\omega$ point is at $+1$ on the vertical axis. The two vectors form a right-angle triangle, so $L = 1.41$ units, $M = 1$ unit. Then the amplitude is $1/1.41 = 0.707$, the so-called -3db point on the magnitude curve.

The angle of M is still $+90^\circ$ and $\angle L$ is $+45^\circ$, so the angle of the function is $+45^\circ$.

- **High Frequency** At high frequencies, the lengths of the L and M vectors approach each other in length, so the magnitude function approaches unity (that is, 0db).

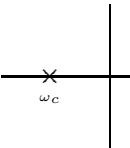
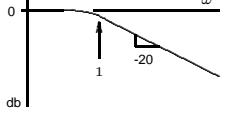
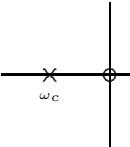
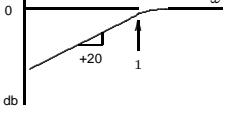
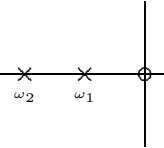
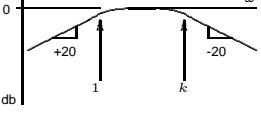
The angle of the L approaches $+90^\circ$ and as it does so, the angle of the function approaches $+0^\circ$.

These values correspond to the amplitude and phase of the first-order highpass filter. See for example figure 443 on page 519.

17.27 Summary of Common Pole-Zero Configurations

This section is a catalogue of some common pole-zero configurations. The most important consequence of the placement of poles and zeros is the shape of the response curve. Therefore the transfer functions are normalized to a frequency of 1 radian per second and the scale factor H_o is assumed to be 1.

Poles and Zeros on the Real Axis

Type	Function	Pole-Zero Plot	Characteristic	Section
Lowpass	$\frac{e_o}{e_i} = \frac{1}{(s + 1)}$			4.8, 17.2
Highpass	$\frac{e_o}{e_i} = \frac{s}{(s + 1)}$			4.9, 17.2
Bandpass	$\frac{e_o}{e_i} = \frac{s}{(s + 1)(s + k)}$			17.3

The bandpass filter is equivalent to the series connection of a highpass filter followed by a lowpass filter. The corner frequency of the highpass filter is at 1 radian/second and the lowpass filter is at k radians/second.

Second Order, Complex Poles and Zeros

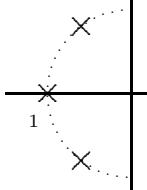
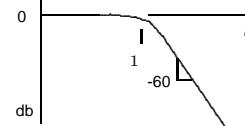
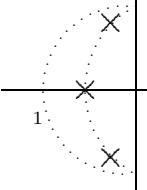
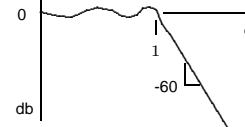
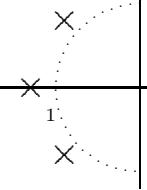
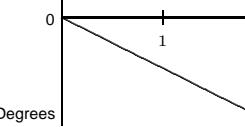
Type	Function	Pole-Zero Plot	Characteristic	Section
Lowpass	$\frac{e_o}{e_i} = \frac{1}{(s + \frac{1}{Q}s + 1)}$			4.8, 17.2
Bandpass	$\frac{e_o}{e_i} = \frac{s}{(s + \frac{1}{Q}s + 1)}$			17.5, 17.8
Highpass	$\frac{e_o}{e_i} = \frac{s^2}{(s + \frac{1}{Q}s + 1)}$			17.6, 17.8
Bandreject (Notch)	$\frac{e_o}{e_i} = \frac{s^2 + 1}{(s + \frac{1}{Q}s + 1)}$			17.7, 17.8
Allpass	$\frac{e_o}{e_i} = \frac{(s - \frac{1}{Q}s + 1)}{(s + \frac{1}{Q}s + 1)}$			19

Notes:

- **Highpass** There are two zeros at the origin.
- **Bandreject** The poles and zeros lie on a circle of radius 1 unit.
- **Allpass** Notice the symmetry of the pole-zero placements: the zeros mirror the pole positions. The filter has no effect on the amplitude of the output: the amplitude function is a flat, horizontal line. The phase does change with frequency, as shown in the diagram.

Higher Order Filters

In this chart, we compare the third order Butterworth, Chebyschev (2db Ripple) and Bessel pole-zero diagrams.

Type	Function	Pole-Zero Plot	Characteristic	Section
Butterworth	$\frac{e_o}{e_i} = \frac{1}{(s+1)(s^2 + s + 1)}$			17.17
Chebyschev	$\frac{e_o}{e_i} = \frac{s}{(s+0.4)(s^2 + 0.4s + 8)}$			17.19
Bessel	$\frac{e_o}{e_i} = \frac{s^2}{(s+1.3)(s^2 + 2s + 2)}$			17.20

Notes:

- **Equations** The coefficients of these equations are also available in the Filter Tables, section 17.24 on page 570.
- **Butterworth Poles** The poles lie on a circle of radius 1 unit.
- **Chebyschev Poles** The poles lie on an ellipse *inside* a circle of radius 1 unit.
- **Bessel Poles** The poles lie *outside* a circle of radius 1 unit.
- **Bessel Response** The *Characteristic* of the Bessel filter is *linear phase shift with frequency*. For the Butterworth and Chebyschev filters, the amplitude plots are log-log: decibels gain vs frequency on a logarithmic scale. For the Bessel filter, the amplitude characteristic is similar to the Butterworth. The phase characteristic is linear when plotted as linear degrees vs linear frequency.

17.28 Exercises

1. For the filter circuit shown in figure 495:

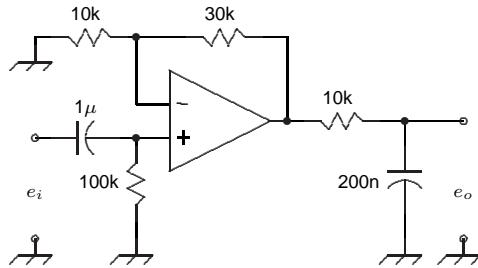
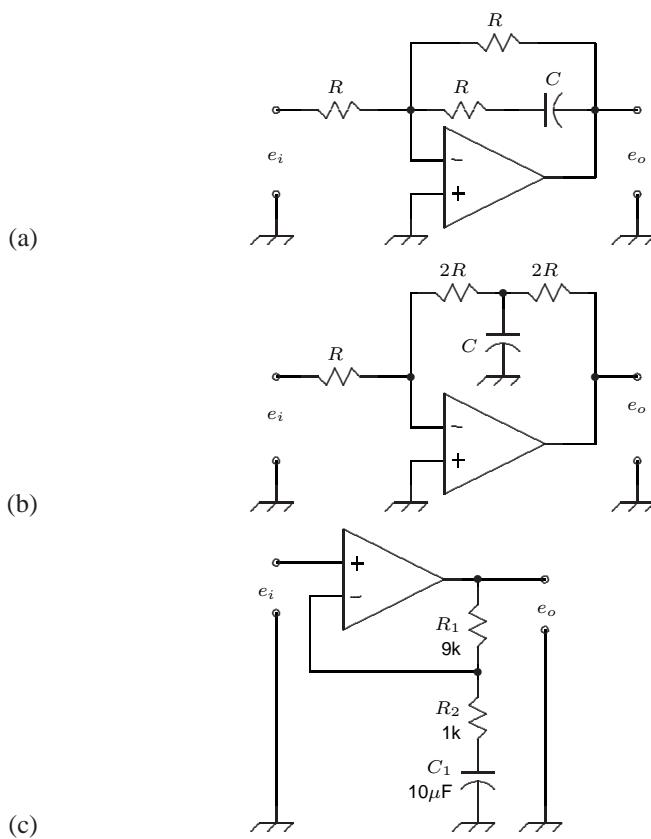


Figure 495: Filter Circuit

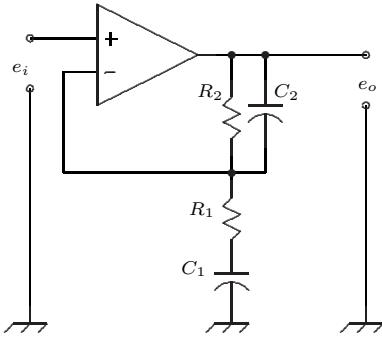
- (a) Sketch the amplitude response in decibels gain vs frequency.
 - (b) Sketch the phase response in degrees vs frequency.
 - (c) Estimate the passband gain.
2. For each of the op-amp circuits shown below, determine the transfer function, Bode amplitude plot and Bode phase plot.



3. In section 17.3, we showed that the inverting-amplifier filter of figure 445 has the transfer function:

$$\begin{aligned} G(s) &= -\frac{Z_f}{Z_i} \\ &= -\frac{sR_2C_1}{(1+sR_1C_1)(1+sR_2C_2)}. \end{aligned}$$

The non-inverting form of that filter is shown below:



- (a) Show that the transfer function e_o/e_i is given by

$$\frac{e_o}{e_i} = 1 + \frac{Z_2}{Z_1}$$

where

$$\begin{aligned} Z_1 &= R_1 + 1/sC_1 \\ Z_2 &= R_2 \parallel 1/sC_2 \end{aligned}$$

- (b) Assuming that $R_2C_2 \gg R_1C_1$, plot the Bode magnitude plot for this filter.

4. Design a Sallen-Key lowpass filter circuit with the Butterworth characteristic that has a cutoff frequency f_c of 1kHz and an attenuation of 60db at 8kHz. Verify your design with a circuit analysis program, showing the amplitude-frequency response over the range 800Hz to 80kHz.
 5. Acoustical noise is often processed through an *octave-band filter set* to determine the spectral content of the noise.

The filter set consists of a set of bandpass filters, each separated by one octave from its neighbours. The centre frequencies are at 63, 125, 250, 500, 1k, 2k and 8k Hz. (There is very little acoustical noise energy at 31 Hz or 16 kHz so these filters are omitted.) The filters are operated in parallel, an arrangement that is known as a *filter bank*. The input signal is fed into the input of all the filters. The output from each filter is displayed on an AC voltmeter which is calibrated in decibels.

To meet certain noise requirements, all of the filter outputs must be below the corresponding *noise contour* [171].

The frequency response of an octave filter¹⁸⁰ is shown in figure 496. The filter response must lie between the two sets of lines. You may assume that the upper cutoff frequency f_u of one filter corresponds to the lower cutoff frequency f_l of the next higher filter.

¹⁸⁰The octave filter response is adapted from [172]. For the official version, consult IEC standard 225.

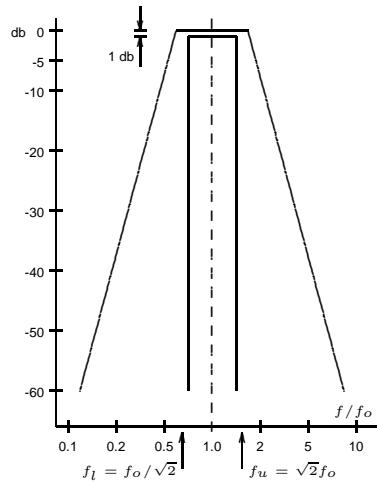


Figure 496: Octave Filter Response

- (a) The bandpass filter could possibly be implemented with an arrangement of Sallen-Key highpass and lowpass filters in series. Alternatively, a state-variable high-Q bandpass filter may be suitable. Determine which of these arrangements will work for this filter type. Take into account the fact that 7 of these octave filters are required for one filter bank, and that cost is an issue.
- (b) Design the 1kHz octave filter, showing component values for the filter. Plot the frequency response and compare it to the requirements.
- (c) Write a computer program to generate the component values for an octave filter, given the centre frequency value.
6. The circuit diagram of the *multiple-feedback* bandpass circuit is shown in figure 497.

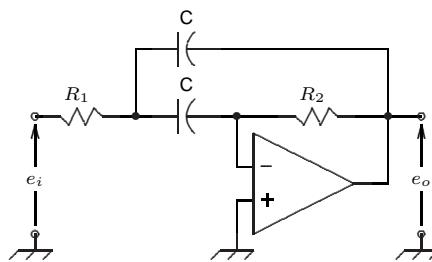


Figure 497: Multiple Feedback Bandpass Filter

The multiple feedback circuit is suitable for low-Q bandpass applications. Because it uses only one op-amp, it is particularly attractive when a bank of filters is required. Lancaster [148] recommends it be considered where the filter Q is less than 5. For higher Q applications, a multiple op-amp circuit such as the state-variable (see Section 17.12) or the biquadratic (Problem 10) is more suitable.

The two capacitors are equal, so they are each shown simply as C in the schematic.

(a) Show that the transfer function may be written as:

$$\frac{e_o}{e_i} = -\frac{1}{R_1 C} \frac{s}{\left(s^2 + \frac{2}{R_2 C} s + \frac{1}{R_1 R_2 C} \right)} \quad (835)$$

The *standard form* for the bandpass filter may be written as:

$$\frac{e_o}{e_i} = H_o \frac{\omega_o}{Q} \left(\frac{s}{s^2 + \frac{\omega_o}{Q} s + \omega_o^2} \right) \quad (836)$$

This equation is similar to equation 729 previously given on page 530, with the addition of the factor H_o which represents the gain of the filter at the resonant frequency ω_o .

(b) By comparison of equations 835 and 836, show that the undamped natural resonant frequency ω_o is given by:

$$\omega_o = \frac{1}{\sqrt{R_1 R_2 C^2}} \quad (837)$$

(c) Show that the Q factor is given by

$$Q = \frac{1}{2} \sqrt{\frac{R_2}{R_1}} \quad (838)$$

(d) Show that the gain H_o is

$$H_o = -2Q^2 \quad (839)$$

(e) Determine the resistor values for a bandpass filter of this type when the centre frequency is 1000 Hz and the Q factor is 5. You may assume a capacitor value of 100nF.

- i. What is the gain of this filter at 1000 Hz?
- ii. Assuming that the circuit is powered from ± 5 volt supplies, what is the maximum allowable value of the input voltage e_i , in peak-peak volts?
- iii. Sketch a diagram of the amplitude frequency response over the range $\omega_o/10$ to $10\omega_o$.

7. A common variant of the multiple feedback circuit is shown in figure 498.

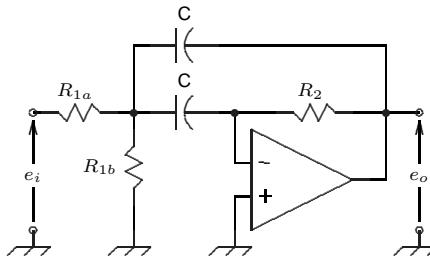


Figure 498: Multiple Feedback Bandpass Filter, Modified

Resistor R_1 in the multiple feedback filter of figure 497 is converted into a voltage divider R_{1a}, R_{1b} such that

- the internal resistance of the divider is equal to the original value of R_1

- the input signal is attenuated by some factor, call it k .

The frequency response, centre frequency and Q factor, are not affected but the circuit can cope with a larger input signal. For example, if $1/k = H_o$, then the gain of the circuit at the resonant frequency is equal to unity rather than H_o and clipping is less likely to occur. This modification also has the effect of increasing the resistance looking into the input, since R_{1a} must necessarily be larger than R_1 .

Using the value of R_1 that you calculated for the circuit in figure 498, determine values for R_{1a} and R_{1b} to make the overall gain at the resonant frequency equal to unity.

8. A tuneable bandpass filter is to be designed to make a simple swept-frequency spectrum analyser for audio applications. It is proposed to use the state variable filter (section 17.12) as the starting point.

The variable tuning may be accomplished with a dual-section $10\text{k}\Omega$ potentiometer that varies the tuning resistances. A DP3T rotary switch selects the tuning capacitances. The tuning should be in three bands: 20-200Hz, 200-2000Hz, 2000-20,000 Hz.

The Q factor for the design is to be 10. The maximum overall gain between input and output should be unity.

Where possible, use 10k fixed resistors.

- Draw the revised circuit with its component values.
- Using a circuit analysis program, plot the amplitude response when the filter is tuned to 1kHz.

9. A variant of the state variable filter is shown in figure 499.

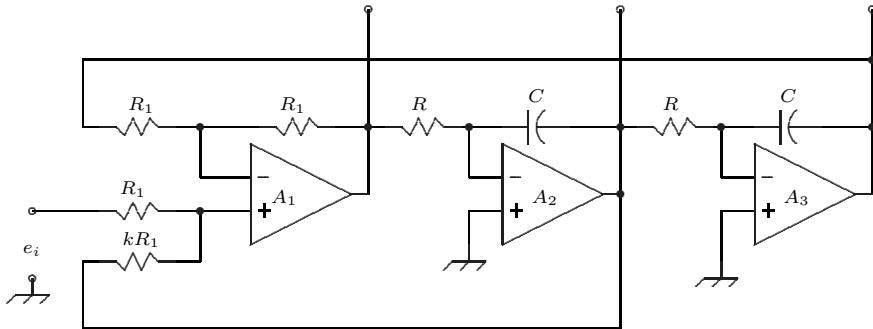


Figure 499: State Variable Filter Variant

- Determine the transfer function for each of the three outputs.
- Indicate which output is bandpass, lowpass and highpass.
- Determine an expression for the centre frequency ω_o .
- Determine an expression for the Q factor.
- Determine an expression for the gain H_o .
- Calculate component values if the circuit centre frequency is 5000Hz and the Q is 3. The capacitors should be 100nF .
- Determine the passband gain H_o of the completed filter.

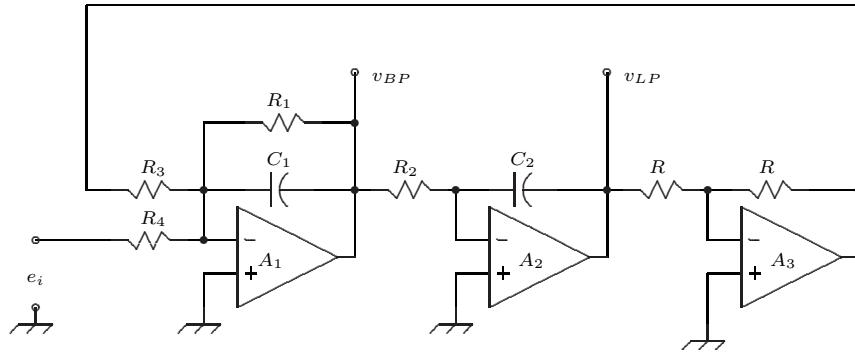


Figure 500: Biquad Filter

10. The *biquad* or *Tow-Thomas* filter circuit shown in figure 500 (adapted from [149]) is a near relative of the state-variable, which was discussed in section 17.12.

According to Huelsman [149], the parameters of the bandpass output v_{BP} are:

$$\omega_o = \frac{1}{\sqrt{R_2 R_3 C_1 C_2}} \quad (840)$$

$$\frac{1}{Q} = \frac{1}{R_1} \sqrt{\frac{R_2 R_3 C_2}{C_1}} \quad (841)$$

$$|H_o| = \frac{R_1}{R_2} \quad (842)$$

(H_o is the gain at the centre of the bandpass.)

It is usual to make $C_1 = C_2 = C$ and $R_2 = R_3 = R$. As in the case of the state-variable filter, the biquad centre frequency can be tuned by varying resistors R_2, R_3 with a dual section variable resistor.

- (a) What is the effect on the filter Q when the biquad is tuned in this manner?
- (b) How would this affect the bandpass output? Under what circumstances would this be acceptable?
Hint: $Q = \omega_o / \Delta\omega$
- (c) How would this affect the lowpass output? Would this be acceptable?
- (d) If your company was to develop a 'universal filter' for general commercial use in electronics labs, would it be better to base it on the state-variable filter or the biquad? Explain your answer.

11. An audio signal is contaminated with interfering signals at 60Hz and 120Hz.

- (a) Using the *twin-tee* circuit, design a circuit to eliminate both those frequencies.
- (b) Assuming ideal components, sketch the frequency response of the resultant filter, over the range 6Hz to 1200Hz.
- (c) Using a circuit simulation program, determine the effect of component tolerances on the notch depth of the filter. The capacitors may be assumed to have a 5% tolerance, the resistors 1% tolerance.

12. The *lag-lead* network is shown in section 24.8. The transfer function is given by:

$$G(s) = \frac{1}{k} \left(\frac{s + k\omega_1}{s + \omega_1} \right) \quad (843)$$

where $k > 1$.

This can be written in frequency normalized form as:

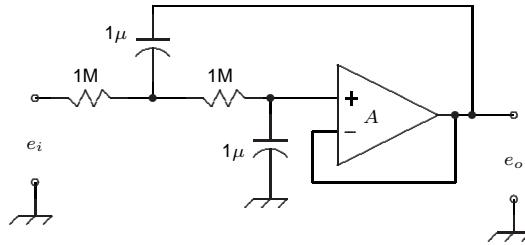
$$G(s) = \frac{1}{H_o} \left(\frac{s + k}{s + 1} \right) \quad (844)$$

where $H_o = k$.

Using the approach shown in section 17.26 on page 580:

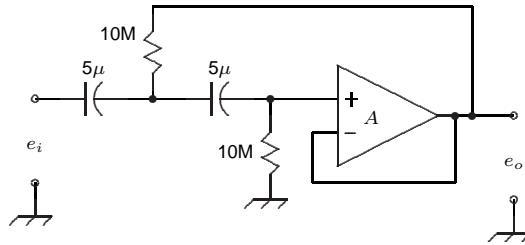
- (a) Draw the pole-zero plot for the lag-lead network.
- (b) Sketch the frequency and phase responses according to the pole-zero positions. (Indicate the amplitude and phase response at the frequency extremes and 2 other frequencies.)
- (c) Show how this frequency response corresponds to the frequency response obtained in section 24.8.

13. The lowpass filter shown below is used in a seismograph [173].



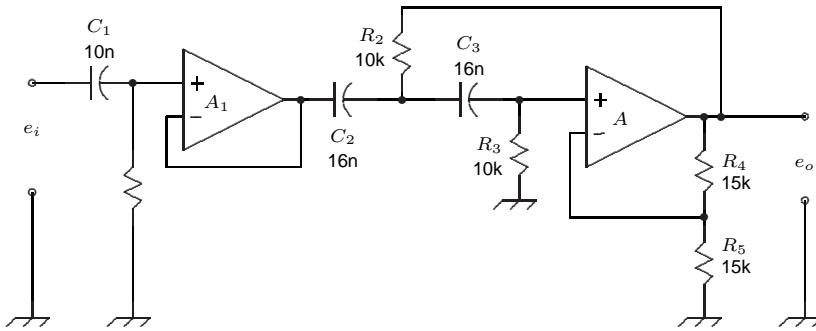
- (a) Determine the cutoff frequency ω_o and the Q factor for this circuit.
- (b) Determine new values for the capacitors such that the Q factor is critically damped ($Q = 0.707$).

14. The highpass filter shown below is used in the same seismograph as mentioned in question 13.

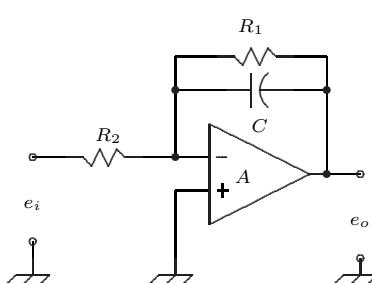


- (a) Determine the cutoff frequency ω_o and the Q factor for this circuit.
- (b) Determine new values for the capacitors such that the Q factor is critically damped ($Q = 0.707$).

15. Regarding the circuit shown below:



- (a) What type of filter is this: lowpass, bandpass, highpass or bandreject?
 (b) What is the *order* of the filter?
 (c) What is the slope of the transition region in db/decade?
 (d) What is the passband gain of the filter?
 (e) Calculate the 3db cutoff frequency f_c of the filter.
16. Design a 5th-order Butterworth lowpass filter using the Sallen-Key circuit configuration for the second-order building blocks, and using the compensated integrator (shown below) for the first-order building block. The voltage gain of the filter in dc steady state is unity and its cutoff frequency is 1000 radians/second.



- (a) Determine the transfer function of the compensated integrator.
 (b) Using the filter tables, determine the normalized values for the components used in the compensated integrator circuit.
 (c) Complete the design of the 5th order Butterworth filter and show the complete schematic of the designed filter with values of the components.
 (d) If the passband gain of the filter needed to be adjusted, which of these resistors should be chosen to make the adjustment? Explain.
17. Figure 444 on page 520 shows first-order lowpass and highpass filters based on the *inverting* op-amp configuration.
- (a) Draw the corresponding first-order lowpass and highpass filters based on the *non-inverting* op-amp configuration.
 (b) Determine the transfer function for each filter.
 (c) Draw the Bode Amplitude and Phase plots for each filter.

18 The Zoo: Unusual Circuits

If the circuits of the previous section are the domestic animals of the circuit menagerie, then this group of circuits is the *zoo*: a group that has unusual and somewhat counterintuitive properties. However, in the same way that the animals of a zoo tell us something about evolution, these circuits give us hints on evolving other useful circuits. And, like zoo animals, in their own environment they have useful and relevant properties.

Many of the circuits in other sections are so useful that they should be committed to memory and be recognisable as subsystems in larger schematics. However, the circuits in the zoo are curiosities, and unlikely to appear elsewhere with any frequency. However, it is useful to know (for example) that it is possible to transform a resistor into its negative, a capacitor into a much larger effective capacitor, or a capacitor into an inductor, and how to analyse a circuit that purports to accomplish this particular feat.

18.1 NIC: Negative Impedance Converter

First, let us consider the properties of a *negative resistance*, $-R$. When a voltage E is impressed across its terminals, a current I flows, given by

$$\begin{aligned} I &= \frac{E}{-R} \\ &= -\frac{E}{R} \end{aligned}$$

In other words, the magnitude of the current is unchanged, but the direction is reversed. The current flows out of positive terminal of the resistor. This behaviour is shown in figure 501(a).

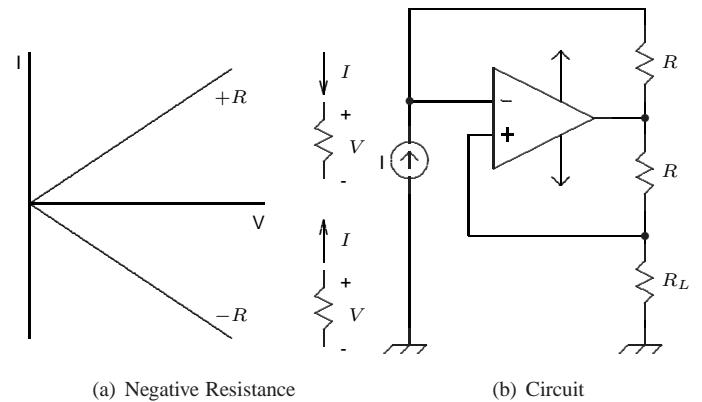


Figure 501: Negative Impedance Converter

Circuit

One cannot purchase a negative resistance at the parts depot¹⁸¹, but it can be synthesized with the op-amp circuit shown in figure 501(b).

The effect of the op-amp and resistors R is to transform resistor R_L so that it appears as $-R_L$ to the current generator. That is, the current out of I will be found to create a negative voltage at the op-amp inverting input terminal.

Before we launch into the analysis, notice that this circuit has a combination of negative feedback (between the output and inverting input) and positive feedback (between the output and non-inverting input). It's reasonable to ask whether the circuit behaves as a negative or positive feedback system.

If positive feedback predominates, then the circuit must be analysed as a kind of Schmitt trigger, and the input terminals of the op-amp are definitely not at the same potential. If negative feedback predominates, then we can assume that the input terminals are at the same potential.

The negative feedback factor (the gain between the output of the op-amp and the inverting input of the op-amp) is determined by the voltage divider comprised of the upper resistor R and the internal resistance of the current generator, which is infinite. Consequently the negative feedback factor is unity.

The positive feedback factor (the gain between the output of the op-amp and the non-inverting input of the op-amp) is determined by the voltage divider comprised of the lower resistor R and the load resistance R_L . As a result the positive feedback factor is somewhat less than unity.

¹⁸¹And if you could, it would be rather bulky, since it would have to contain batteries.

Overall then, for the configuration shown, the negative feedback factor predominates and we can take the two inverting terminals to be at the same potential¹⁸².

Analysis

The steps in the NIC analysis are numbered in figure 502. Notice that the op-amp must be powered by a bipolar power supply: as we'll see shortly, it must produce a negative output voltage. The usual rules of NFB op-amp analysis apply:

- No current flows into the op-amp input terminals
- The two input terminals are at the same voltage.

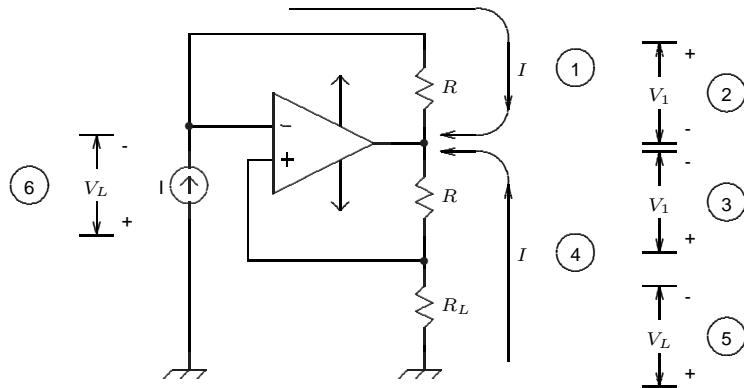


Figure 502: NIC Analysis

Now the analysis:

1. The supply current I flows through the upper resistance R
2. setting up a voltage V_1 across that resistor, of the polarity shown.
3. This is a negative feedback system, so the inverting and non-inverting terminals are at the same voltage, so the total voltage across the two resistors R is zero. Then the voltage across the lower resistor R is also equal to V_1 , polarity as shown on the diagram (opposite to the upper resistor voltage).
4. Consequently, a current I flows through the lower resistance R in the direction shown. Since the upper resistor conducts I down the page and the lower resistor conducts I up the page, the output terminal of the op-amp must be sinking a current of $2I$.
5. The lower current I sets up a voltage V_L across the load resistor, of the polarity shown on the diagram.
6. The input terminals of the op-amp are at the same voltage, so this voltage V_L must also appear at the non-inverting op-amp input terminal, and across the current generator. This voltage is negative with respect to ground, so current into the NIC produces a negative voltage, that is, the NIC input appears as a resistance $-R_L$. *QED.*

Application: Negative Resistance Oscillator

Consider a tank resonant circuit, consisting of an inductor and capacitor in parallel (section 4.12). If some energy is injected into this system (by starting with a charged capacitor, for example), the energy will shuttle back and forth between the electric field of the capacitor and the magnetic field of the inductor. In the process, a sine wave of voltage will appear across the tank. In any real circuit, both components (especially the inductor) will have some lossy behaviour, and the sine wave will eventually die away to zero.

The dissipative behaviour of the tank circuit can be characterized as a real positive resistance in parallel with the inductor and capacitor. Now, if the tank circuit is paralleled with an equal negative resistance, the total

¹⁸²Notice that the NIC would be a positive feedback system if we attempted to drive it by a voltage source. We could then call it an NNIC (Not-NIC).

equivalent resistance becomes zero. Consequently, any oscillation in the tank circuit will persist indefinitely, and the arrangement has become an oscillator.

In practice, the high frequency performance of op-amps is quite limited. This limits the maximum frequency of an oscillator implemented with a tank circuit attached to a NIC, but it works satisfactorily at low frequencies.

Negative Incremental Resistance

The negative resistance behaviour shown in figure 501 does not occur naturally but can be synthesised with an operational amplifier. For completeness and to avoid confusion, we mention that another type of negative resistance does occur intrinsically in certain devices, and that is *negative incremental resistance*, as shown in figure 503.

Over a small region, the device current decreases as its voltage increases (or what is the same thing, its current increases as its terminal voltage decreases). If the device is biased into the negative resistance region, it may be used to cancel a positive resistance in the circuit. Some devices which exhibit negative resistance:

- certain gas-filled tubes
- tunnel diode
- unijunction transistor

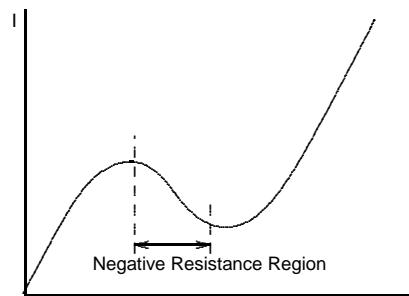


Figure 503: Negative Incremental Resistance

The NIC in Active Filters

It is possible to use the NIC to construct active filters. However, it turns out that these filters are very sensitive to component tolerances, and so the NIC is not used as a component in active filter designs. From [174], on the history of active filters:

The concept of using an NIC was interesting and simple. It therefore attracted researchers' attention in the development of active filters using NIC's in the 1950's. ... Nevertheless, the problems were still so severe as to make the practical use of these filters unattractive. Thus this research area was later abandoned and current technologies do not use filters with NIC's.

Application

Reference [175] describes how a negative electrical impedance can be used to cancel out certain properties of a transducer so that a more accurate measurement of the remaining properties can be made.

18.2 Capacitance Multiplier

A *capacitance multiplier* is an electronic circuit that makes a relatively small capacitance appear much larger. We have to be painfully honest: this is not a hugely useful accomplishment. It's usually more useful to be able to make a small capacitance appear even smaller, because the capacitance limits the high-frequency performance of some circuit.

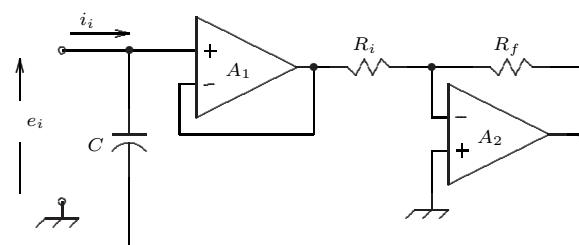


Figure 504: Capacitance Multiplier

The energy required for this large pseudo-capacitance must be supplied by an operational amplifier circuit of some sort, so this *capacitance on steroids* cannot be used in a power supply filter, where such an accomplishment would be truly useful.

Nonetheless, it is interesting just to see how it's done.

Recalling the Miller Effect (section 13.1), you will remember that a feedback capacitance C_f around an amplifier of gain $-A$ appears at the input of the amplifier as an effective input capacitance of

$$C_i = (A + 1)C_f \quad (845)$$

(See figure 267 on page 334).

In most cases, the voltage gain A is not that well defined, being the open-loop gain of an op-amp or the voltage gain of a transistor stage. But if we construct an amplifier with a well-defined gain, then we can construct a predictable capacitance multiplier. A suitable circuit is shown in figure 504 above.

Suppose that some voltage e_i is applied to the input of the circuit. Then the top end of the capacitor is at

$$v_c^+ = e_i \text{ volts} \quad (846)$$

The bottom end of the capacitor is at

$$v_c^- = -\frac{R_F}{R_I}e_i \text{ volts} \quad (847)$$

Consequently, the total voltage across the capacitor is

$$\begin{aligned} v_c &= e_i - \left(-\frac{R_F}{R_I}e_i \right) \\ &= e_i \left(1 + \frac{R_F}{R_I} \right) \text{ volts} \end{aligned} \quad (848)$$

The current into the capacitor is equal to this voltage divided by the impedance of the capacitor:

$$\begin{aligned} i_c &= \frac{e_i \left(1 + \frac{R_F}{R_I} \right)}{1/sC} \\ &= e_i \left(1 + \frac{R_F}{R_I} \right) sC \end{aligned} \quad (849)$$

The impedance seen at the input terminal is

$$\begin{aligned} X_c &= \frac{e_i}{i_c} \\ &= \frac{1}{\left(1 + \frac{R_F}{R_I} \right) sC} \\ &= \frac{1}{s \left(1 + \frac{R_F}{R_I} \right) C} \end{aligned} \quad (850)$$

This is equivalent to a capacitor of magnitude

$$C' = \left(1 + \frac{R_F}{R_I} \right) C \quad (851)$$

In words, the size of the capacitor is effectively multiplied by the gain of the inverting amplifier stage.

A Limitation

Not only is this circuit not that useful, but it has a serious limitation. The circuit requires that the input voltage be multiplied by some factor and then applied to the far end of the capacitor C. However, for any significant multiplication factor (1000, say), the output of amplifier A_2 is liable to saturate unless the input voltage is kept very small. So the signal handling capability of this circuit is very limited.

18.3 Howland Current Generator

The *Howland* circuit¹⁸³, like the circuit of section 13.16, create a constant current in a grounded load. It's more useful than its inclusion in the *Zoo* section might imply: it's here because it's a near relative of the negative impedance converter (section 18.1).

The circuit is shown (as it is usually drawn, in its simplest form) in figure 505. Operation is quite simple: the output current is

$$I_L = -\frac{E}{R} \quad (852)$$

Like the grounded-load current generator of section 13.16, the Howland circuit can create a linear ramp in a capacitor. It could also be used as the basis for an ohmmeter, by forcing constant current through an unknown resistor, and measuring the voltage across the resistor.

Like the NIC, the Howland circuit is a combination of negative and positive feedback, and we need to determine which predominates. As long as some load is connected to the Howland generator, the negative feedback will predominate and we can analyse it as a negative feedback circuit.

Analysis

A brute-force circuit analysis, similar to the reasoning used in section 18.1 to analyse the NIC, can be used to verify the operation of the Howland circuit¹⁸⁴. But more insight can be obtained by treating the circuit as an enhanced NIC.

The op-amp and resistors R_2 , R_3 serve to transform R_4 and R_L into a negative resistance. Applying that approach, the circuit of figure 505 is equivalent to that of figure 506.

We make $R_1 = R_4$. Referring to circuit 506, we can write:

$$I_4 = \frac{V_L}{R_4} \quad (853)$$

$$I_L = \frac{V_L}{R_L} \quad (854)$$

$$I_1 = \frac{V_L - E}{R_1} \quad (855)$$

$$I_1 = I_4 + I_L \quad (856)$$

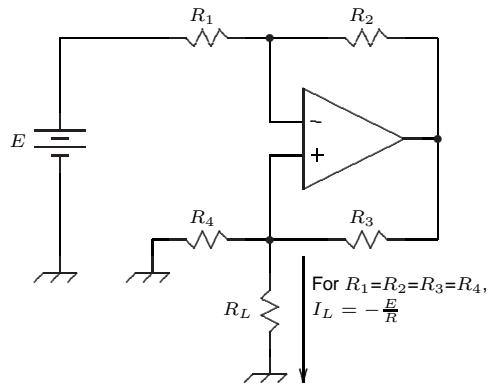


Figure 505: Howland Current Generator

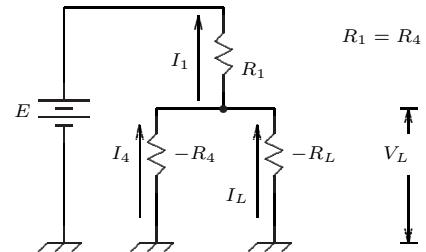


Figure 506: Howland Circuit Equivalent

¹⁸³Invented in the late 50's by Brad Howland at MIT, according to Robert Pease in [176].

¹⁸⁴For a truly general analysis of the Howland circuit that generates frightening equations on the way to a simple conclusion, see reference [177].

Substitute from equations 853, 854 and 855 into equation 856:

$$\frac{V_L}{R_1} - \frac{E}{R_1} = \frac{V_L}{R_4} + \frac{V_L}{R_L} \quad (857)$$

So

$$-\frac{E}{R_1} = \frac{V_L}{R_L} \quad (858)$$

But

$$\frac{V_L}{R_L} = I_L \quad (859)$$

Substituting from 859 into 858, we have

$$I_L = -\frac{E}{R_1} \quad (860)$$

QED

Notes on the Design

- The output of the op-amp may make substantial positive or negative excursions in driving current through the load resistance. Check the design to ensure that the op-amp output does not saturate.
- Output current must flow through the NIC resistors (R_2 and R_3 in figure 505), and this can be very inefficient if significant amounts of current are required. The circuit of section 13.16 might be a better choice in such a case.
- Careful matching of the resistors is required for obtaining high output impedance. The degree to which these resistors must be matched will depend on the requirements of the application.

Differential Input

It is also possible to apply the input voltage E to a different point in the circuit, as shown in figure 507. In that case the output current is the same magnitude but opposite sign:

$$I_L = +\frac{E}{R_1} \quad (861)$$

Both inputs may be used simultaneously, in which case the output current is proportional to the difference between the two input voltages.

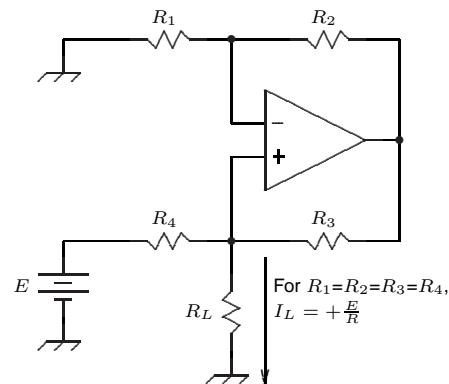


Figure 507: Alternate Input

References and Opinions

In the course of researching the Howland current source, I came across the following sources and their quotations:

... it is apparent that the (Howland) operational amplifier current source is a high quality current source in all respects. Not only can the current be accurately controlled but the high output impedance is easily achieved. [177]

This sounds great, but there's a hitch: the resistors must be matched exactly ... performance is limited by the CMRR of the op-amp ... at high frequencies, the output impedance can drop ... to as little as a few hundred ohms As clever as it looks, the Howland current source is not widely used. [100]

This (Howland current source) circuit suits many applications ... Circuits such as this (application) find wide use in automatic test equipment [176]

And it's not that any of these authors are incorrect, they are looking at different aspects and applications of the circuit. The moral? Understand how it works, understand the strengths and weaknesses, and make up your own mind.

All of these references have useful information on the circuit.

18.4 Simple Gyrator

A *gyrator* is a circuit that transforms an impedance into its inverse. This turns out to be useful in the design of electronic filters. There is a vast literature on the design of such filters using capacitors and inductors. Unfortunately, inductors have a number of limitations - especially at audio frequencies, where they are large, heavy and expensive.

A gyrator can transform a capacitance into an inductance, so an inductance can be simulated in the circuit. The circuit has the design and performance advantages of a filter that includes inductances, but has the size, weight and cost advantages of using capacitors.

For example, the impedance of a capacitor is given by

$$X_C = \frac{1}{j\omega C} \quad (862)$$

The gyrator transforms this into its reciprocal:

$$X_{in} = K j\omega C \quad (863)$$

where K is some constant that depends on the circuit. This is the behaviour of an inductor with equivalent inductance L_{eq} equal to

$$L_{eq} = KC \text{ Henries} \quad (864)$$

A simple example of a gyrator [178] is given in figure 508.

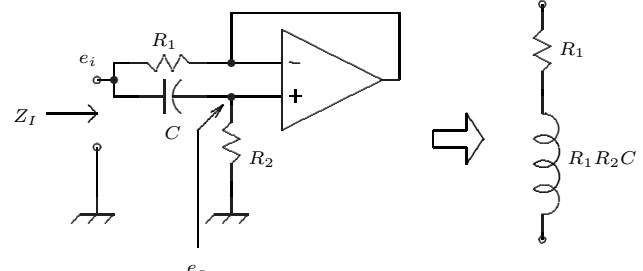


Figure 508: Gyrator

18.4.1 Transient Analysis: A Reality Check

The behaviour of this circuit can be understood intuitively with the analysis shown in figure 509. Consider that a step of voltage is applied to the input of the gyrator. It should mimic the behaviour of a resistance in series with an inductance. To see that it does do this, work through the following steps:

1. Apply a voltage step to the input of the circuit.
2. The voltage waveform at the non-inverting terminal is a sharp transient that dies away to zero.
3. This is a negative feedback circuit, so the op-amp causes the inverting terminal to follow the same waveform: a sharp transient that dies away to zero.
4. The voltage across resistor R_1 (and hence the current through it) is the difference between waveforms 1 and 3: an exponential rise toward some final value. This is the expected result: a step of voltage across the equivalent resistor-in-series-with-an-inductor should cause a current that starts at zero and rises toward a final value equal to the input voltage divided by the series resistance R_1 .

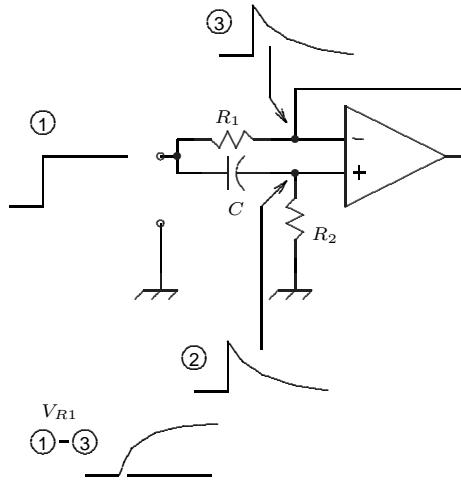


Figure 509: Gyrator Transient Response

AC Analysis

Now let us determine the behaviour of the gyrator in the frequency domain. This will give us some insight into the design of the circuit.

Refer to figure 508. To determine the effective impedance at the input terminals of this circuit, we'll apply an input voltage e_i to the circuit of figure 508 and calculate the resultant input current i_i . Then

$$Z_i = \frac{e_i}{i_i} \quad (865)$$

With a voltage e_i at the input terminals of the circuit, the voltage e_a appearing at the non-inverting terminal of the op-amp is

$$\begin{aligned} e_a &= e_i \left(\frac{R_2}{R_2 + 1/sC} \right) \\ &= e_i \left(\frac{sR_2C}{1 + sR_2C} \right) \end{aligned} \quad (866)$$

The current into the input has two components: through the resistor R_1 , and through the capacitor C .

$$\begin{aligned} i_i &= \frac{e_i - e_a}{R_1} + \frac{e_i - e_a}{1/sC} \\ &= e_i \left(\frac{1 + sR_1C}{R_1} \right) - e_a \left(\frac{1 + sR_1C}{R_1} \right) \end{aligned} \quad (867)$$

Substitute for e_a from equation 866 into equation 867:

$$\begin{aligned} i_i &= e_i \left(\frac{1 + sR_1C}{R_1} \right) - e_i \left(\frac{sR_2C}{1 + sR_2C} \right) \left(\frac{1 + sR_1C}{R_1} \right) \\ &= e_i \frac{1}{R_1} \left(\frac{1 + sR_1C}{1 + sR_2C} \right) \end{aligned} \quad (868)$$

So that

$$\begin{aligned} Z_i &= \frac{e_i}{i_i} \\ &= R_1 \left(\frac{1 + sR_2C}{1 + sR_1C} \right) \end{aligned} \quad (869)$$

Let's do a sanity check on this equation:

- According to equation 869, at very low frequencies where $s \approx 0$, the input impedance $Z_i \approx R_1$. Does that make sense from an examination of the circuit? The capacitor appears as an open circuit and the non-inverting terminal is at zero volts. Then the inverting terminal is also at zero, and the input impedance is R_i .
- According to equation 869 at very high frequencies, $sR_2C \gg 1$ and $sR_1C \gg 1$. Then the input impedance $Z_i \approx R_2$.

Does that make sense according to the circuit? At high frequencies, the capacitor appears as a short circuit, so there is zero voltage across R_1 , and no current through it. The input circuit sees R_2 connected to ground, so that forms the input impedance Z_i .

In both cases, at low and high frequencies, equation 869 for Z_i gives results that correspond to the behaviour of the circuit. That gives us some confidence in the equation.

Now, how does equation 869 indicate that the input impedance behaves as an inductor, as we promised at the beginning?

We can get some insight into this equation if we sketch the frequency response of the impedance, using the techniques of section 11 (page 300). As illustrated in figure 510 on page 602, the frequency response of input impedance (equation 869) can have one of two possible shapes, depending on the relative magnitude of the two resistors. In Case 1, the response breaks downward at ω_1 and then upward at ω_2 . In Case 2, the breaks are reversed.

For the input impedance to behave as an inductance, the impedance should *increase* with frequency. Then figure 510(b) is the correct one and we consequently require that $R_2 > R_1$. When that is true, figure 510(b) indicates that the circuit will appear as an inductance over the frequency range $\omega_1 = 1/R_2C$ to $\omega_2 = 1/R_1C$.

What is the magnitude of this inductance? We can get at that if we substitute $s = j\omega$, $1/R_2C = \omega_1$, $1/R_1C = \omega_2$ in equation 869 for input impedance:

$$Z_i = R_1 \left(\frac{1 + j\omega/\omega_1}{1 + j\omega/\omega_2} \right) \quad (870)$$

If we assume that the frequency of operation ω is much less than the second break frequency ω_2 , then equation 870 simplifies nicely:

$$Z_i = R_1 + j\omega R_1 R_2 C \quad (871)$$

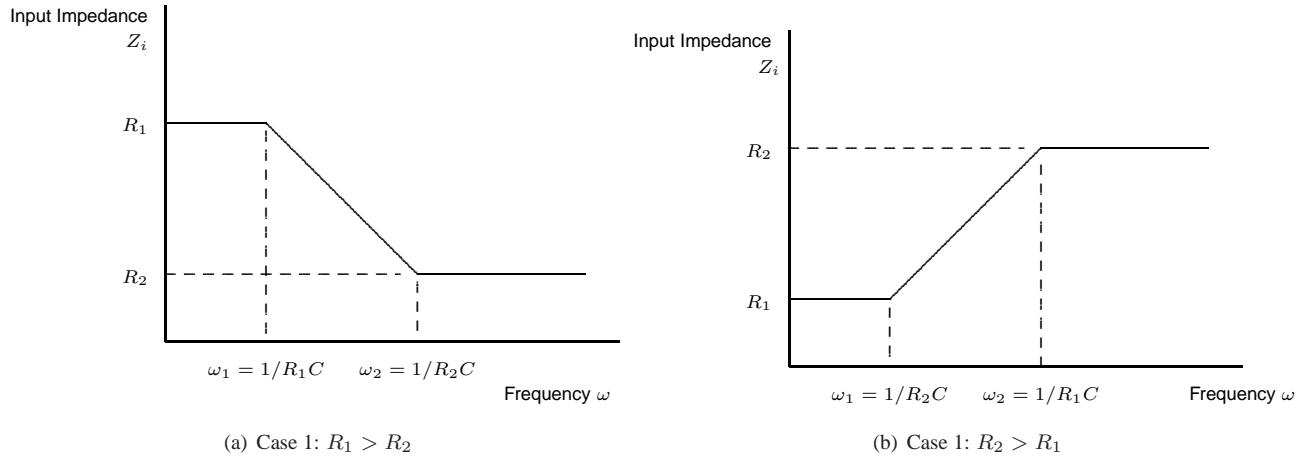


Figure 510: Gyrator Responses

This is the equation of an impedance comprised of a resistance R_1 ohms in series with an inductor of $R_1 R_2 C$ henries¹⁸⁵.

Example

An example gyrator circuit is shown in figure 511. (See reference [178] where it is used in an audio circuit).

The gyrator synthesizes a 6.8 Henry inductor with a series resistance of 680 ohms¹⁸⁶. As a winding on a magnetic core, this inductor would be huge, heavy, expensive and essentially impractical to build. The synthesized inductance is in parallel with C_O , a $1\mu\text{F}$ capacitor, to form a tank circuit that is resonant at 61Hz with a Q factor of 3.8.

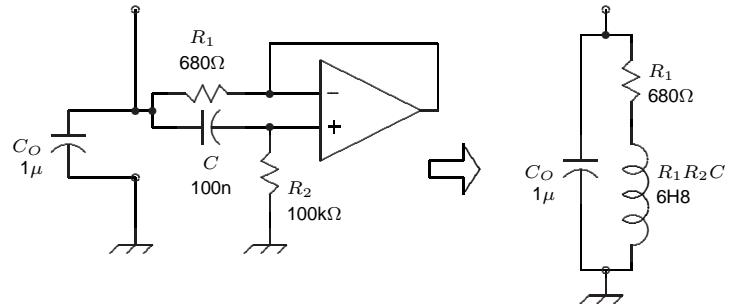


Figure 511: Gyrator Example

18.4.2 Assessment

This simple gyrator is a useful circuit and straightforward to design and understand. It does require that there be a resistance in series with the inductor, so the synthesized inductor cannot be ideal (zero resistance). It also has the limitations of all circuits synthesized with op-amps – the frequency response is limited to the region where the op-amp is operating properly. For example, the circuit could be limited by the slew rate of the op-amp for large signals at high audio frequencies.

¹⁸⁵This rather long-winded analysis can be greatly simplified with the assumption that R_2 is much greater than R_1 . Unfortunately, there is no easy way to determine that is a requirement, apart from doing the full analysis.

¹⁸⁶You should be able to verify these figures using the equations given above.

18.5 Improved Gyrator

An improved gyrator¹⁸⁷ circuit is shown in figure 512.

The effective inductance is given by

$$L_{eq} = \frac{R_1 R_3 R_5 C}{R_2} \quad (872)$$

Notice that there is no resistance in series with the simulated inductance, so this device can be much closer to the ideal than the circuit shown in section 18.4.

Applications

This gyrator circuit, along with the FDNR (section 18.6) is widely used in the design of complex active filters. The filter is first designed with inductors and capacitors, and then the inductors are replaced with this gyrator circuit.

Analysis

Providing that the op-amps may be regarded as ideal (zero voltage between the input terminals, no current into the input terminals) the analysis of this circuit is surprisingly straightforward.

Consider the diagram of figure 513.

As usual in determining the effective impedance across the input terminals, we apply a voltage e_i to the input and calculate the resultant current i_i . Then

$$Z_i = \frac{e_i}{i_i} \quad (873)$$

We will start at the top of the diagram and work our way down to R_5 . Because both the op-amps are negative feedback systems,

$$v_{13} = 0 \quad (874)$$

$$v_{35} = 0 \quad (875)$$

Consequently by KVL

$$v_{56} = e_i \quad (876)$$

and the current through R_5 is given by

$$i_5 = \frac{e_i}{R_5} \quad (877)$$

Now we can work our way back up the diagram. No current flows into the op-amp inputs, so this same current i_5 flows through the capacitor C , setting up a voltage across it of

$$v_{45} = \frac{e_i}{R_5} \frac{1}{sC} \quad (878)$$

¹⁸⁷Sedra and Smith [78] refer to this as an Antoniou Inductance Simulation circuit after its inventor. Ghausi [174] refers to it as a GIC: Generalized Impedance Converter. Lynch [179], Dent [180] and Troughton [181] refer to it as a Gyrator.

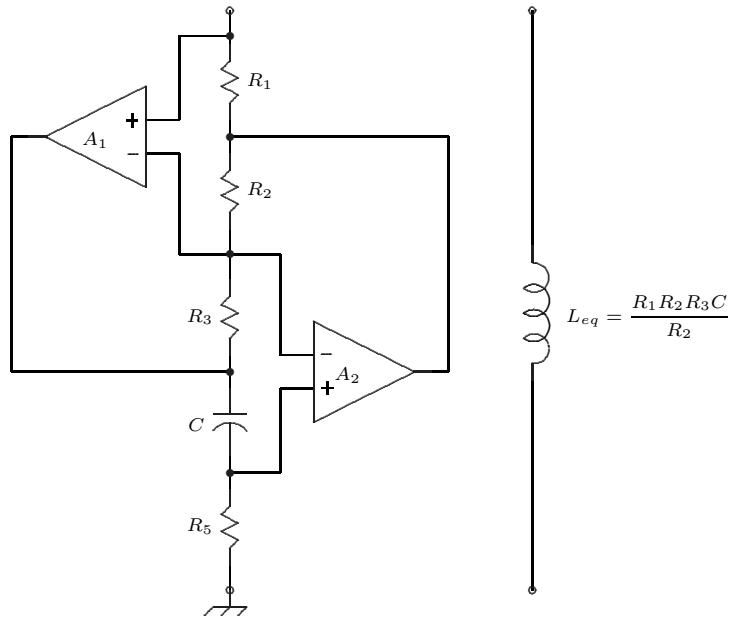


Figure 512: Gyrator

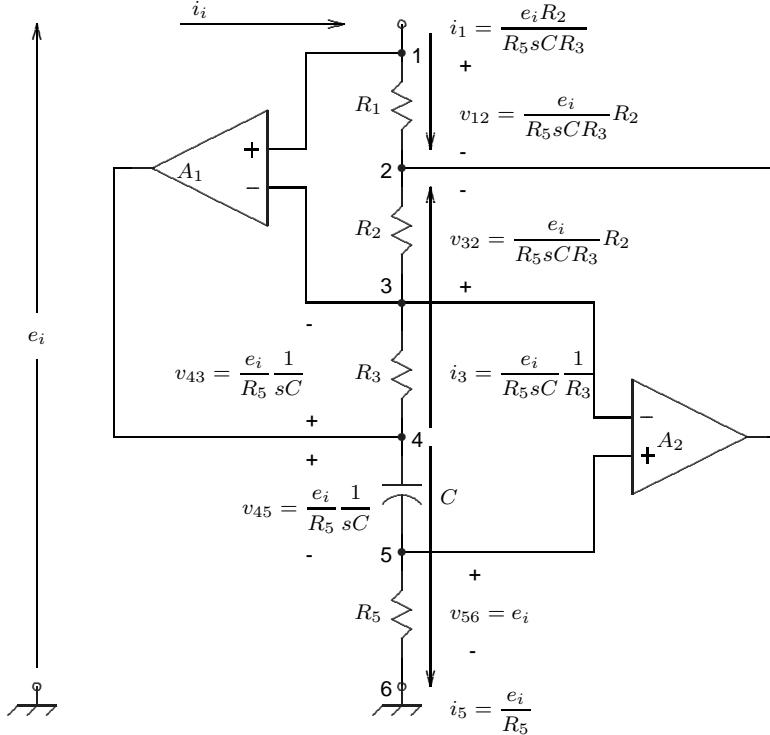


Figure 513: Gyrator Analysis

Now, the voltage v_{35} is zero because it's between the inputs of op-amp A_2 . Consequently, it must be true that

$$\begin{aligned} v_{45} &= v_{43} \\ &= \frac{e_i}{R_5} \frac{1}{sC} \end{aligned} \quad (879)$$

(Notice the polarities of these voltages, as indicated on the diagram.) So the current i_3 is given by

$$\begin{aligned} i_3 &= \frac{v_{43}}{R_3} \\ &= \frac{e_i}{R_5} \frac{1}{sC} \frac{1}{R_3} \end{aligned} \quad (880)$$

Now we do similar steps in the upper half of the diagram (Second verse, same as the first . . . :)
The current i_3 must also flow through R_2 , so the voltage v_{32} is given by

$$v_{32} = \frac{e_i}{R_5} \frac{1}{sC} \frac{1}{R_3} R_2 \quad (881)$$

This voltage must be equal and opposite to the voltage v_{12} , so

$$v_{12} = \frac{e_i}{R_5} \frac{1}{sC} \frac{1}{R_3} R_2 \quad (882)$$

with the polarity shown on the diagram.

Finally, the current into the input terminal is equal to this voltage divided by R_1 :

$$i_i = \frac{e_i}{R_5} \frac{1}{sC} \frac{1}{R_3} \frac{R_2}{R_1} \quad (883)$$

Substituting for i_i in equation 873 (and rearranging slightly), we have:

$$Z_i = s \left(\frac{R_1 R_3 R_5}{R_2} \right) C \quad (884)$$

As a result, the input impedance appears as an inductance of value

$$L_{eq} = \frac{R_1 R_3 R_5 C}{R_2} \quad (885)$$

If all the resistors are equal to R , then this simplifies to

$$L_{eq} = R^2 C \quad (886)$$

Not that difficult, once you get the hang of it, eh?

18.6 FDNR: Frequency Dependent Negative Resistance

This circuit creates a D element: a device that is a negative resistance that varies with frequency. The impedance-vs-frequency characteristic (adapted from Hickman [182]) is shown in figure 514.

The magnitude of a capacitive impedance decreases in proportion to frequency, and the phase angle of voltage lags behind the current by 90° . The magnitude of the D element impedance decreases as the *frequency squared* and the current is 180° out of phase with the applied voltage. A positive voltage applied to this device causes current to flow out of it back into the source, the same as the negative resistance described in section 18.1.

The impedance of the D element is given by

$$Z_i = \frac{1}{s^2 D} \quad (887)$$

That is, the magnitude is

$$|X_i| = \frac{1}{\omega^2 D} \quad (888)$$

and the phase angle 180° .

18.6.1 Application

It turns out that the D element is useful in the implementation of certain types of active filters. Here is a simple example that conveys the basic idea.

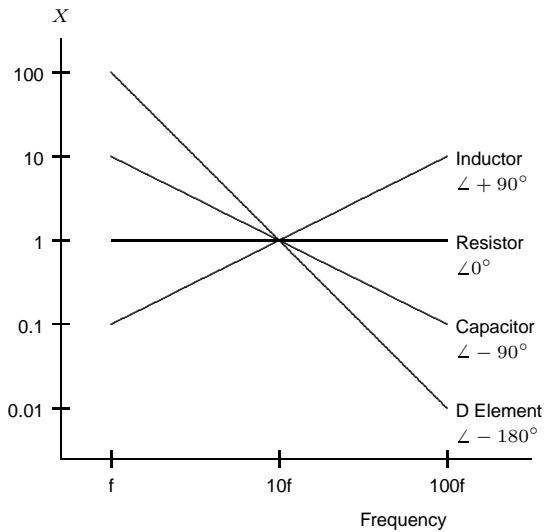


Figure 514: Response of the D Element

Consider the lowpass LC filter shown in figure 515. We'll show that it is equivalent to the RD filter shown in the same figure.

Treating it as a voltage divider, the transfer function of the LC filter is given by

$$\begin{aligned}\frac{e_o}{e_i} &= \frac{1/sC}{sL + 1/sC} \\ &= \frac{1}{1 + s^2LC}\end{aligned}\quad (889)$$

The transfer function of the RD network is

$$\begin{aligned}\frac{e_o}{e_i} &= \frac{1/s^2D}{R + 1/s^2D} \\ &= \frac{1}{1 + s^2RD}\end{aligned}\quad (890)$$

Consequently, these two lowpass filters are equivalent if we make

$$LC = RD \quad (891)$$

The RD filter has no inductors and so it has many practical advantages.

18.6.2 Circuit

A circuit that synthesizes the D element is shown in figure 516. Notice the symbol for the D element. The circuit is obviously a near relative of the gyrator (section 18.5), and can be analysed using much the same technique.

Using that technique, the value of the FDNR can be shown to be

$$D_{eq} = \frac{R_2 R_3 C_1 C_2}{R_4} \quad (892)$$

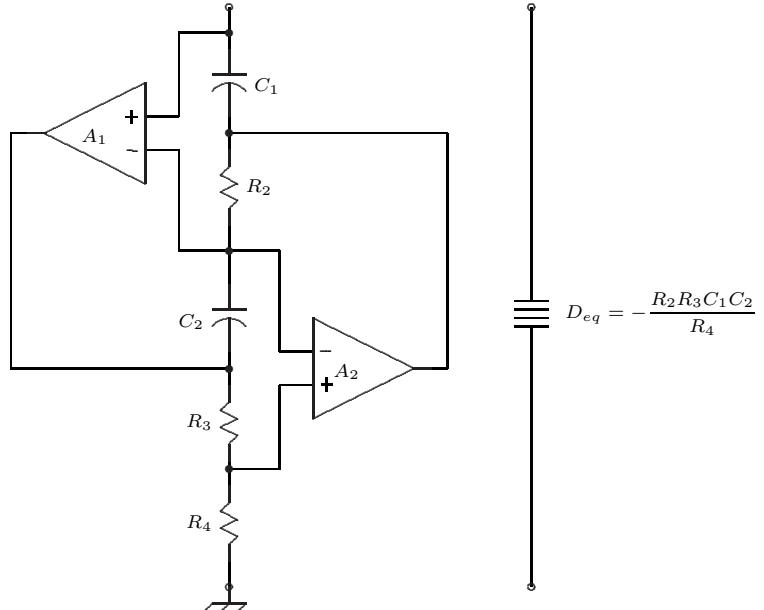


Figure 516: Frequency Dependent Negative Resistance

18.7 Circulator

The *circulator* is a device containing three or more ports¹⁸⁸. A signal fed into a port N appears as an output at port $N + 1$ and only that port. The signal from the last port transfers to the first.

Figure 517 shows the symbol for a three-port circulator. A signal input at Port 1 transfers to Port 2, and only port 2. An input at 2 transfers to 3, and so on. The original circulators used ferrite materials and were operational at microwave radio frequencies [183]. It proved useful and was subsequently extended to operate at lower radio frequencies. The version described here uses op-amps and can operate down to DC (zero frequency) [184]. With the appropriate op-amps, this version of the circulator can operate at frequencies from DC to 500MHz [185], [186].

Figure 518 shows applications of the circulator.

In figure 518(a), a transmitter and receiver share the same antenna. The input of the receiver deals with very small signals and must be a sensitive, low-noise circuit. Consequently it's important to block the large transmitter signal from reaching the input of the receiver. The circulator is an elegant solution that does not require a transmit-receive switch.

Figure 518(b) shows an intercom system. The circulators route the signal from microphone #1 to earphone #2, and from microphone #2 to earphone #1. Notice that there need only be a single connection between the two locations.

Figure 518(c) shows a system for measuring the impedance of a two-terminal component over a range of frequencies. If port #2 is terminated by its *characteristic impedance*, then the output signal will be totally absorbed at port #2. However, if this impedance differs from the port characteristic impedance, then some of the energy output at #2 will be reflected and appear at port #3. This might be useful to determine whether an antenna presents a matched impedance, over a range of frequencies [186].

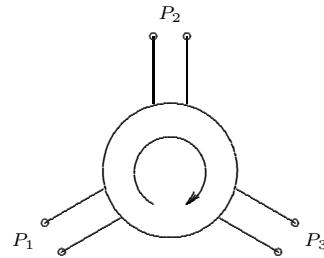
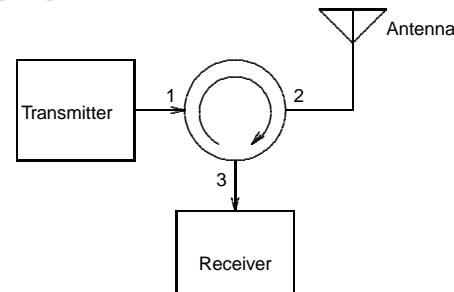
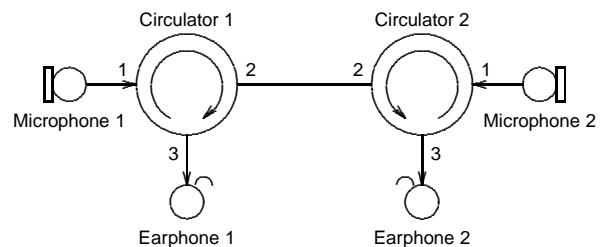


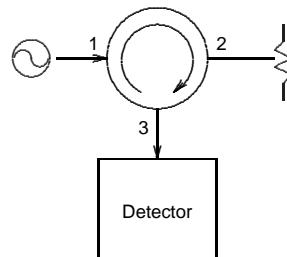
Figure 517: Three Port Circulator



(a) Antenna Switch



(b) Intercom



(c) Impedance Measurement

Figure 518: Circulator Applications

¹⁸⁸A port is a connection point that can be an input or an output.

Implementation

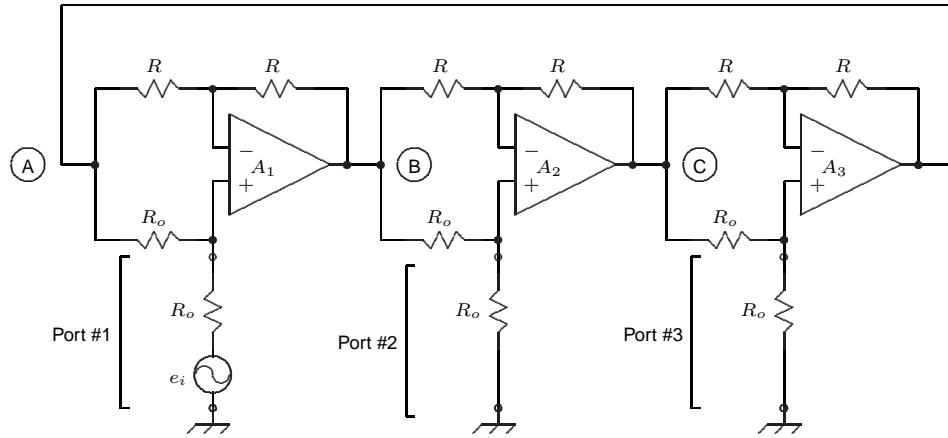


Figure 519: Op-Amp Circulator

Figure 519 shows one possible op-amp implementation of the three-port circulator [184]. As is common in radio-frequency circuitry, each signal source and load must have an internal resistance of some characteristic impedance, designated as R_o on figure 519. In radio practice, this is typically 50Ω . In the case of this circuit, it can be any convenient value.

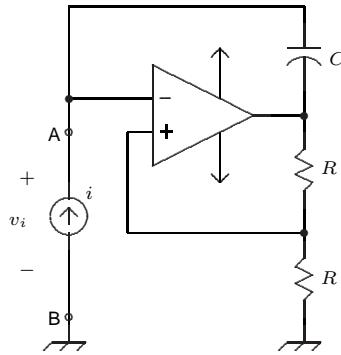
Consider first the A_1 stage: the output of A_3 is a voltage source so point A is at an AC ground. The voltage at the non-inverting input is therefore $e_i/2$. Amplifier A_1 is configured as a non-inverting gain of 2 stage for this signal, so the output of amplifier A_1 at B is simply e_i volts.

The e_i volts at point B appears at port #2 as $e_i/2$ volts, as expected. (A terminated signal is always half the open-circuit voltage, so the open circuit voltage at port #2 is e_i volts.) Consequently the circulator passes on the input signal at port #1 to port #2. However, the amplifier A_2 is configured as a differential amplifier for the signal at B , so its output voltage (at C) is zero. Consequently the signal e_i is *not* passed on to the third port and its output is zero.

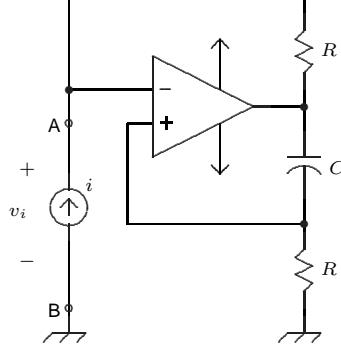
Notice that port #2 *must* be terminated by R_o for the circulator to work correctly. If the termination is different from R_o , then some of the signal will appear at port #3, as in the application of figure 518(c) above.

18.8 Exercises

1. The circuit shown below is a *negative capacitance* circuit [187].



- (a) With the current source removed, what is the impedance that appears between terminals A and B?
 (b) When driven by a current source, describe the relationship between voltage and current in the current generator.
2. The circuit shown below is a *negative inductance* circuit [187].

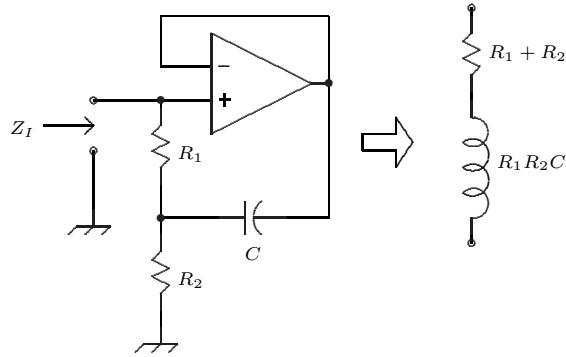


- (a) With the current source removed, what is the impedance that appears between terminals A and B?
 (b) When driven by a current source, describe the relationship between voltage and current in the current generator.
3. Show that the circuit shown in figure 516 on page 606 does synthesise a D element of value

$$D_{eq} = \frac{R_2 R_3 C_1 C_3}{R_4} \quad (893)$$

4. Show that the circuit shown below creates an impedance at its input terminals equal to:

$$Z_i = R_1 + R_2 + sR_1 R_2 C \quad (894)$$



5. The circuit given below uses two *operational transconductance amplifiers* (OTA). (see section 35.5). Each OTA has a normal op-amp input. The output is a current generator where

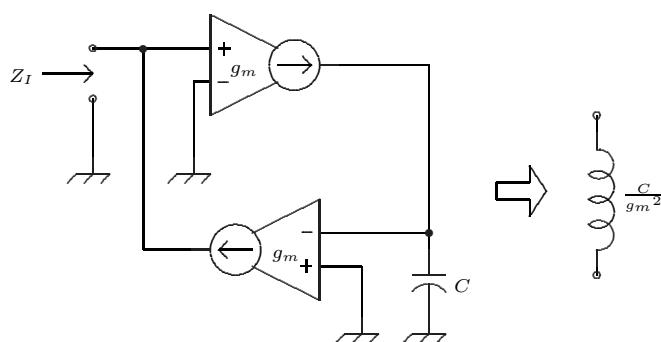
$$i_o = g_m v_d \quad (895)$$

where

i_o Output current, amps

g_m Transconductance, amps per volt (mhos, Ω) Positive current is in the direction of the arrow.

v_d Voltage between the inverting and non-inverting input terminals.



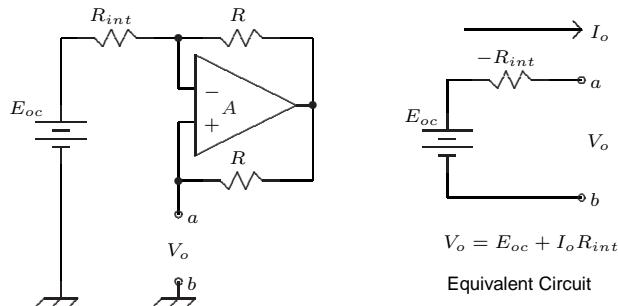
- (a) Show that the impedance seen at the input terminals of the circuit is an inductance of value

$$Z_i = \frac{C}{g_m^2} \text{ Henries} \quad (896)$$

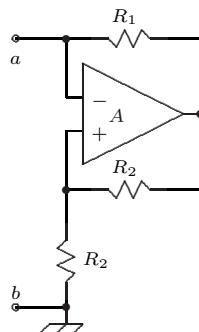
- (b) Determine the value of the inductance seen at the input terminals if $g_m = 1m\Omega$ and $C = 1\mu F$.

6. For the op-amp circuit shown below, the voltage source at the terminals **a** and **b** has a *negative* internal resistance, that is:

$$V_o = E_{oc} + I_o R_{int}$$

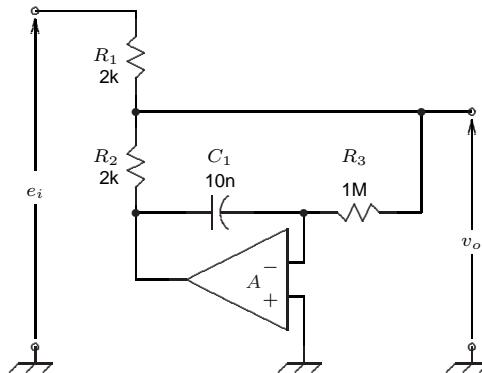


- (a) Consider that a load resistance R_l is attached to the circuit output and then use circuit analysis to verify that the equivalent circuit is correct.
- (b) Draw the output voltage-current characteristic V_o vs I_o for the circuit.
7. The circuit shown below synthesizes a negative resistance between terminals a and b . Analyse the circuit to verify that this is the case and to determine the magnitude of the negative resistance.



Suggest an application for this circuit.

8. The circuit shown below functions as a highpass filter. The effect of the op-amp feedback network is to synthesize a large virtual capacitor from a small physical unit [188].



- (a) Determine the overall transfer function v_o/e_i and verify that it is a single-pole highpass filter.
- (b) What is the cutoff frequency of the filter?

- (c) By what factor does this circuit effectively magnify the value of capacitor C_1 ?
- (d) What is the gain in the passband of this filter?
- (e) Resistors R_1 and R_2 function as a passive adder to combine the input and feedback signals. Suggest another configuration that does not cause any passband loss between input and output.

19 Phase Shift Networks

It is common in signal processing applications to need both the sine version of a signal and the cosine version. These signals are referred to as the *in phase* and *quadrature* signals. The term quadrature indicates a right-angle or 90 degree relationship between the two signal vectors.

In other cases, the phase of a signal is used to carry information, and so the creation of a variable phase waveform and the measurement of its phase become important. In this section we show how to generate a controlled change in the phase of a signal.

19.1 Passive Network Phase Shifter

An RC lowpass or highpass voltage divider may be used as a simple phase-shifting network (see sections 4.8 and 4.9). The output of the divider is a phase-shifted and smaller magnitude version of the input. The exact values of phase shift and magnitude scaling depend on the input frequency and the cutoff frequency ω_o of the network. If the input frequency changes, both the output amplitude and phase shift will change. If the frequency is fixed, then a fixed-gain amplifier may be used to compensate for the attenuation of amplitude in the network. These networks are limited to phase shifts in the range of $+90^\circ$ to 0° for the RC highpass network, or 0° to -90° for the RC lowpass network.

19.1.1 Improved Phase Shifter

Figure 520 shows a somewhat improved circuit. The two ends of the RC network are driven by anti-phase voltages e_i and $-e_i$. As the frequency changes, the phase of the output voltage changes but the amplitude stays constant at e_i . For this reason, the network is called a *phase shifter* or *all pass* network.

The output phase angle is given by

$$\angle \frac{e_o}{e_i} = -2 \tan^{-1} \frac{\omega}{\omega_o} \quad (897)$$

where ω is the frequency of the input waveform and

$$\omega_o = \frac{1}{RC} \quad (898)$$

Horowitz and Hill [100] show a clever phasor analysis of this circuit. We'll take a different tack and analyse it with the help of some math and Bode plots.

19.1.2 Analysis of the All-Pass Network

We'll use the superposition theorem to determine the output due to each of the sources.

The output due to the left source e_i will be called e_o' . Then applying the voltage divider equation:

$$\begin{aligned} e_o' &= e_i \left(\frac{1/sC}{R + 1/sC} \right) \\ &= e_i \left(\frac{1}{1 + sRC} \right) \end{aligned} \quad (899)$$

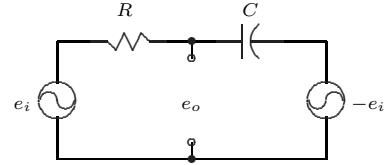


Figure 520: Phase Shifter, Passive

The output due to the right source $-e_i$ will be called e_o'' . Then

$$\begin{aligned} e_o'' &= -e_i \left(\frac{R}{R + 1/sC} \right) \\ &= -e_i \left(\frac{sRC}{1 + sRC} \right) \end{aligned} \quad (900)$$

Combining,

$$\begin{aligned} e_o &= e_o' + e_o'' \\ &= e_i \left(\frac{1}{1 + sRC} \right) - e_i \left(\frac{sRC}{1 + sRC} \right) \\ &= e_i \left(\frac{1 - sRC}{1 + sRC} \right) \end{aligned}$$

So the gain of the network is

$$\frac{e_o}{e_i} = \frac{1 - sRC}{1 + sRC} \quad (901)$$

19.1.3 Bode Plots

First, let us examine the magnitude functions for the all-pass network, shown on the Bode plot of figure 521.

The magnitude of the denominator term in equation 901,

$$1 + sRC$$

is a lowpass network which has a flat response up to the corner frequency $1/RC$ and then a response that decreases at a rate of 20db/decade thereafter.

The magnitude of the numerator term

$$1 - sRC$$

has a flat response up to the corner frequency $1/RC$ and then a response that increases at a rate of 20db/decade thereafter. Evidently the two magnitude terms cancel each other.

Now consider the phase responses, shown in figure 522.

Both the numerator and denominator have the same phase response: 0° at low frequencies (below $\omega_o/10$), -45° at ω_o , and -90° at high frequencies (above $10\omega_o$). The net effect is a phase that depends on frequency. The phase shift is 0° at low frequencies, -90° at ω_o , and -180° at high frequencies.

For some frequency inside the range $\omega_o/10$ to $10\omega_o$, the phase shift of the network can be adjusted by changing R or C in the network.

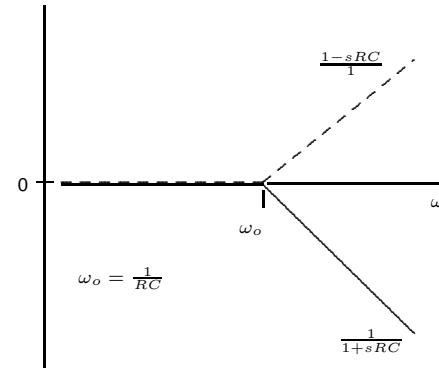


Figure 521: All-Pass Network, Magnitude Response

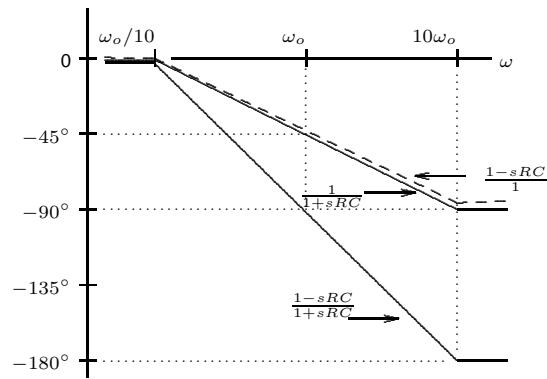


Figure 522: All-Pass Network, Phase Response

19.2 Op-Amp Phase Shifter

Consider the circuit of figure 523(a). For analysis, the input circuit has been redrawn as two sources in figure 523(b). Using the superposition theorem, we can say that the voltage at the output is the sum of the effects of e'_i and e''_i .

That is,

$$e_o = e'_o + e''_o \quad (902)$$

where e'_o is the output due to e'_i and e''_o is the output due to e''_i .

The output e'_o is given by

$$e'_o = -e'_i \quad (903)$$

The output e''_o is given by

$$e''_o = +2e_a \quad (904)$$

But

$$e_a = \frac{1}{1 + sRC} e''_i \quad (905)$$

Then

$$\begin{aligned} e_o &= e'_o + e''_o \\ &= -e_i + 2 \left(\frac{1}{1 + sRC} \right) e_i \\ &= \left(\frac{1 - sRC}{1 + sRC} \right) e_i \end{aligned}$$

so the transfer function is

$$\frac{e_o}{e_i} = \frac{1 - sRC}{1 + sRC} \quad (906)$$

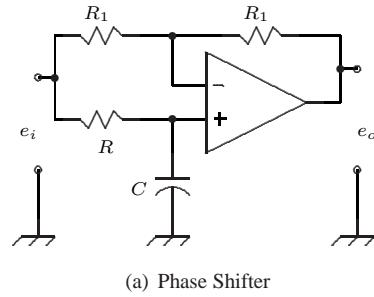
This is the equation that was developed for the passive all-pass network in section 19.1, so the same behaviour applies to this network. The output amplitude is constant at e_i . The output phase angle is given by

$$\angle \frac{e_o}{e_i} = -2 \tan^{-1} \frac{\omega}{\omega_o} \quad (907)$$

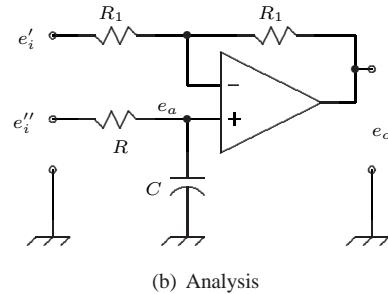
where ω is the frequency of the input waveform and

$$\omega_o = \frac{1}{RC} \quad (908)$$

Resistor R is often made adjustable to change the output phase angle at some given frequency.



(a) Phase Shifter



(b) Analysis

Figure 523: Op-Amp Phase Shifter

19.3 Wideband Phase Shifter

The circuits in section 19.1 and 19.2 provide the desired phase shift only over a limited range of frequencies. They have the convenience (compared to a simple lowpass or highpass RC network) that the output amplitude does not depend on frequency¹⁸⁹. But the phase shift does depend on frequency.

The *phasing method* of single-sideband generation in radio transmitters requires a phase-shifter that will provide a constant phase shift to audio signals *regardless of frequency* [45]. A block diagram of this system is shown in figure 524.

Audio signals are processed through the wideband phase-shifting network to generate two audio signals that are 90° out of phase. These are multiplied with RF signals that are also 90° out of phase. The outputs of the multipliers are two DSB (double sideband) signals with a cancelled carrier. When the DSB signals are added together, one of the sidebands is cancelled and the final output is a single-sideband modulated RF signal. The 0° signal is known as the *in-phase* component; the 90° phase-shifted signal is known as the *quadrature* component.

It's relatively simple to generate the in-phase and quadrature components of the RF signal, because it is at a fixed frequency. (At low frequencies, a *quadrature oscillator* can be used for this.) But the generation of the in-phase and quadrature components of the audio signal is quite a different matter. For reasonable speech intelligibility, the phase shifting network should operate over a 300 to 3kHz frequency range.

The circuit of figure 525 (adapted from [45]) gives an idea of the required complexity.

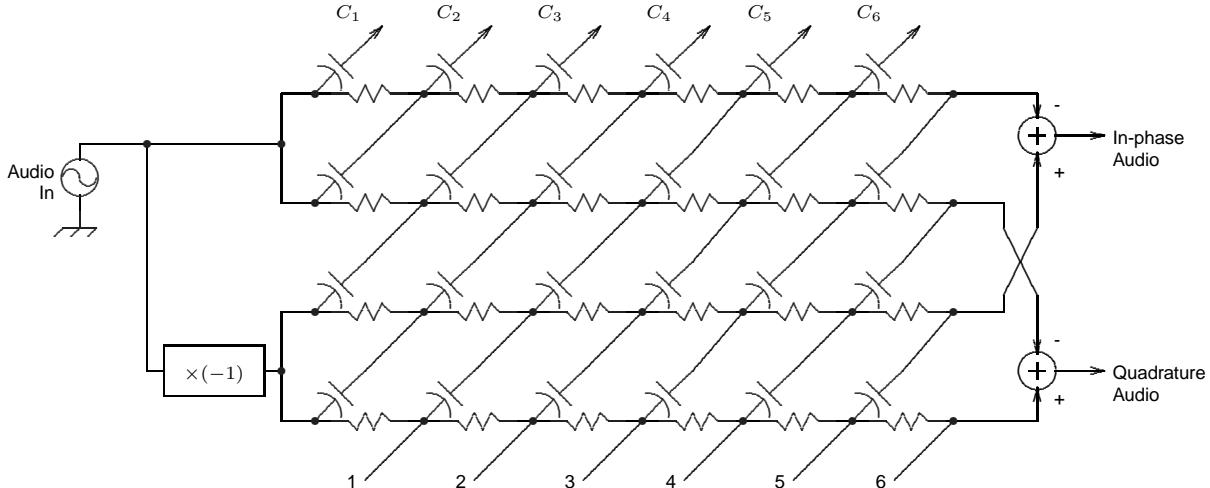


Figure 524: Single Sideband Generation, Phasing Method

The network is driven from the audio signal and its inverted version. The output is then extracted by a pair of differential amplifiers.

In the version shown in [45], the resistors are equal valued at $12\text{k}\Omega$. The capacitor values are given as:

¹⁸⁹In practice, it's more useful that the output amplitude does not change as the phase-shift components are adjusted.

$$C_1 = 44n \quad C_2 = 33n \quad C_3 = 20n \quad C_4 = 10n \quad C_5 = 5n6 \quad C_6 = 4n7$$

A very sophisticated analysis of this network, not for the faint-hearted, is given in reference [189].

Audio Applications

Another application for this network is in *audio frequency shifting*. For example, if the input audio signal is modulated with a 3 Hz sub-audio frequency in place of the RF signal in figure 524, the output is the same audio signal but frequency shifted by a constant 3 Hz. Selection of the sideband (by interchanging one of the in-phase and quadrature inputs to the multipliers) allows the frequency shift to be an increase or decrease. The effect of this unusual device is generally inaudible but has a dramatic effect in reducing acoustical feedback in a sound reinforcement system¹⁹⁰.

The definitive article on this device is in [190], with similar descriptions in [191] and [192].

The wideband phase-shifting network from [190] is shown in figure 526. According to this paper, a phase difference between the in-phase and quadrature outputs of $90^\circ \pm 7^\circ$ is obtained from 200Hz to 15kHz.

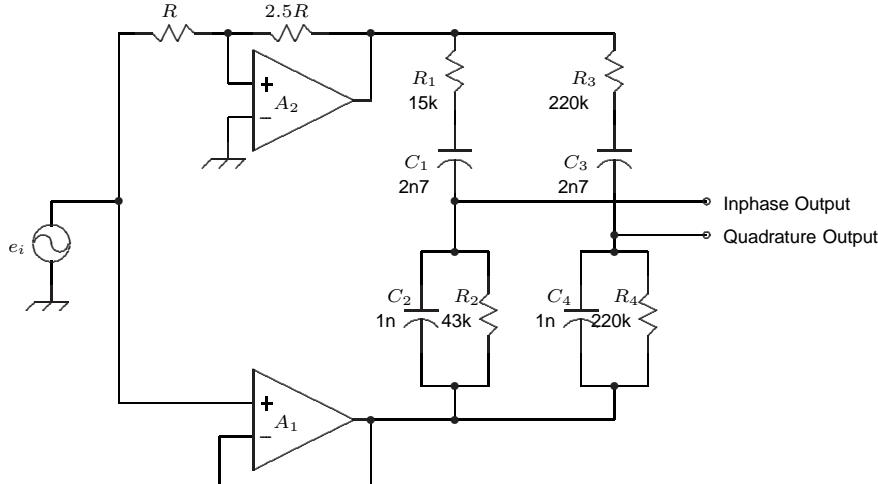


Figure 526: Wideband Phase Shifter 2

For this network, we have two Wien phase shifting networks (see section 20.2), with centre frequencies of 720Hz and 4000Hz corresponding to ω_1 and ω_2 .

$$\begin{aligned} \omega_1 &= 1/R_1 C_1 \\ &= 1/R_2 C_2 \\ \omega_2 &= 1/R_3 C_3 \\ &= 1/R_4 C_4 \\ \omega_2 &= 5.55\omega_1 \\ Q &= \frac{1}{\frac{R_2}{R_1} + 1} \\ &= 0.21 \end{aligned}$$

¹⁹⁰I used it to good effect for the sound design of a production of the Rocky Horror Picture Show, before radio microphones became readily available.

Then the phase response of each Wien network is given by

$$\phi = \tan^{-1} \left[2Q \left(\frac{\omega_o}{\omega} - \frac{\omega}{\omega_o} \right) \right] \quad (909)$$

The phase responses¹⁹¹ of the two Wien networks (the In-Phase and Quadrature outputs) are shown in figure 527.

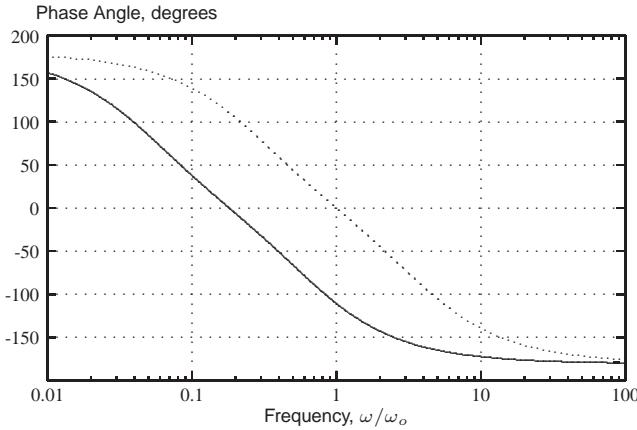


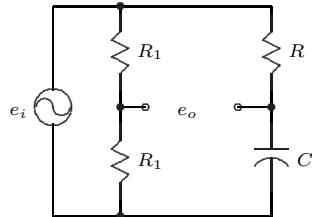
Figure 527: Wideband Phase Shifter 2, Phase Response

A similar frequency shifter, with deliberate selection of higher frequencies so that the effects are audible, has been used on occasion in electronic music equipment¹⁹².

19.4 Exercises

1. The network shown below is an alternative form (from reference [78]) of the allpass network discussed in section 19.1.

Develop the transfer function and confirm this.



¹⁹¹Sharp-eyed readers will notice that the plotted responses differ by somewhat more than 90° . These are the curves obtained with the standard component values shown in reference [190]. Because the component values are standard values, the curves are an approximation of the ideal.

¹⁹²Google search term: *electronic music frequency shifter*. See for example [193].

20 Oscillators

An *oscillator* circuit is one that produces a continuous alternating signal. In most cases, this signal is of constant peak-peak amplitude. In some cases, the signal is a sinusoid. In others, it may be a square or triangle wave.

There are several different ways to view an oscillator circuit.

Oscillator as Negative Feedback System

Section 10.4 showed how a negative feedback system could become an oscillator. If the magnitude of the loop gain AB is greater than unity and the loop phase shift (including the amplifier and sensor network) is some multiple of 360° , then the signal will grow in magnitude with each trip around the loop.

In any real system this can't go on forever, and at some point a *limiting mechanism* applies. (This could be the maximum output from an op-amp or some explicit clipping circuitry.) What happens next depends on the system. If the positive feedback is DC coupled, then the system will stay stuck up against that limit, and it is said to be saturated or cutoff.

If the feedback signal is AC coupled, then hitting the limit mechanism will cause the error signal to reverse polarity, and the output voltage will then reverse and head off to the opposite limit. In effect, since the signal can no longer grow in amplitude, the gain around the loop is forced to unity.

The amplitude and phase conditions are usually stated as the Barkhausen Criterion¹⁹³, that the phase of the loop gain should be zero and the magnitude should be unity [78].

Oscillation and Negative Resistance

Consider an LC tank circuit in which an inductor, capacitor and resistor are all connected in parallel. The resistance represents the losses in the inductor and capacitor. If the circuit is pulsed with a brief impulse function, it will oscillate (ring) with a decaying waveform. The inductance and capacitance are lossless, but the energy originally entered into the tank circuit will eventually be dissipated in the resistor.

If the tank circuit is now paralleled by a negative resistance equal to the positive resistance, the circuit becomes lossless and will oscillate forever. Consequently, it is often useful to regard an oscillator as a tuned circuit in parallel with a negative resistance [175], [196], [197].

Comparator Driven Oscillator

If a sine wave is fed into an operational amplifier comparator (section 12.2), then the output of the comparator will be a square wave at the same frequency of the sine wave. The square wave is then fed into a bandpass or lowpass filter, which extracts the fundamental frequency sine-wave component of the square wave. This sine wave is used to trigger the comparator, closing the loop.

This oscillator has a mechanical analog: the pendulum driven clock. The pendulum is a resonant system, which is driven from side to side by a force equal to the clock weight. The escapement mechanism directs the force to the correct side of the pendulum to produce sustained oscillations.

Simulation Note

Should you decide to simulate any of the oscillator circuits in this section, you may find you have difficulty getting them to start. The reason: ideal operational amplifiers (which are only found in electronic simulation) are noise free. In practice, it is the small amount of electronic noise in the amplifiers and resistors that provides the seed of a signal that can be amplified as it travels around the loop, and grow into the oscillation signal.

¹⁹³As stated in [194], *Barkhausen's criterion is a necessary condition for oscillation, not sufficient. This means there are some circuits which satisfy the criterion but do not oscillate.. A more reliable indicator of stability is the Routh Stability Criterion [195].*

In place of that, the simulation needs something to start it going. The oscillator may be shocked into oscillation by applying a step signal to some point in the loop. For example, using a circuit simulation program such as Multisim or LTSpice, one can start the oscillator of figure 539 by including a battery and pushbutton switch in the simulation. When the pushbutton is actuated, it applies a 10 volt step into a resistor connected to one of the op-amp summing junctions. This is enough to get the oscillator started.

20.1 Phase Shift Oscillator

An oscillator is a feedback system that at some oscillation frequency f_o has a loop gain greater than or equal to unity and a cumulative loop phase shift of 0 degrees (or some multiple of 360 degrees.) If the loop gain is greater than unity, the signal at the oscillation frequency will grow until some limiting mechanism establishes a maximum for the signal.

If the amplifier is non-inverting, we require a feedback network that has a phase shift of zero degrees at the oscillation frequency. Alternatively, we can use an inverting amplifier that effectively contributes 180 degrees of phase shift. Then the feedback network must contribute another 180 degrees of phase shift at the oscillation frequency.

An RC lowpass network provides a very simple building block for the phase shifting network. The circuit in figure 528 provides somewhere between 0° and 90° of phase lag, depending on frequency (section 4.8 on page 169). The unity gain amplifier acts as a buffer so that these networks can be cascaded, as shown in figure 529. Each RC network also attenuates the signal, so the amplifier A must provide sufficient gain to make the loop gain unity.

This is easiest to illustrate with an example.

20.1.1 A Design Example

We wish to design a 1KHz oscillator with an inverting gain stage and three of the RC lowpass networks shown in figure 528.

If f_{osc} is the oscillation frequency in Hz:

- The oscillation frequency ω is

$$\begin{aligned}\omega &= 2\pi f_{osc} \\ &= 6.28 \times 10^3 \text{ radians/second}\end{aligned}$$

- The amplifier will provide 180° of phase shift. Consequently, each of the 3 RC networks will provide 180°/3 = 60° of phase shift.
- Now we need to determine the corner frequency ω_o of the RC network. Phase shift in a lowpass RC network is given by

$$\theta = -\tan^{-1} \frac{\omega}{\omega_o} \quad (910)$$

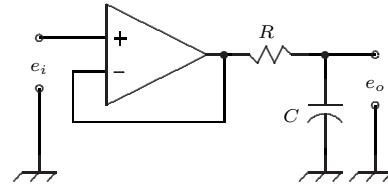


Figure 528: Phase Shift Network

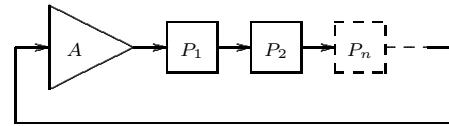


Figure 529: Phase Shift Oscillator

where $\omega_o = RC$. Substituting 60° for θ in equation 910, we have that

$$\begin{aligned}\frac{\omega}{\omega_o} &= \tan \theta \\ &= 1.73\end{aligned}$$

Substitute the value for oscillation frequency ω and we have that

$$\begin{aligned}\omega_o &= \frac{\omega}{1.73} \\ &= 3.63 \times 10^3 \text{ radians/second} \\ &= 578 \text{ Hz}\end{aligned}$$

This makes sense if you look at figure 120(b), which shows the phase vs frequency curve for an RC lowpass network. The corner frequency ω_o must be well below the operating frequency to generate a 60° phase shift. Now we need to determine the required gain of the amplifier.

- The gain of an individual RC lowpass network is given by

$$\left| \frac{e_o}{e_i} \right| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_o} \right)^2}} \quad (911)$$

In this case, we have $\frac{\omega}{\omega_o} = 1.73$, so

$$\begin{aligned}\left| \frac{e_o}{e_i} \right| &= \frac{1}{\sqrt{1 + 1.73^2}} \\ &= 0.5 \text{ v/v} \\ &= -6 \text{ db}\end{aligned}$$

A quick check of figure 120(a) shows that a gain of -6db is approximately right for $\frac{\omega}{\omega_o} = 1.73$.

- With three lowpass stages, the total gain G_s through the phase-shifters is

$$\begin{aligned}|G_s| &= 0.5^3 \\ &= 0.125 \text{ v/v}\end{aligned}$$

The product of the amplifier gain G_A and the phase shift network gain G_s must be greater than or equal to unity. Consequently,

$$\begin{aligned}G_A &\geq \frac{1}{G_s} \\ &\geq 8 \text{ v/v}\end{aligned}$$

so our amplifier must have a minimum inverting gain of 8 volts/volt.

The completed phase shift oscillator is shown in figure 530.

The design equations are shown in figure 530. Aside from meeting these requirements, the choice of the resistors and capacitors is arbitrary.

This circuit requires 4 op-amps, which would have made it hopelessly uneconomical a few years ago. But quad op-amps are now available at very low cost, so it's not a silly design.

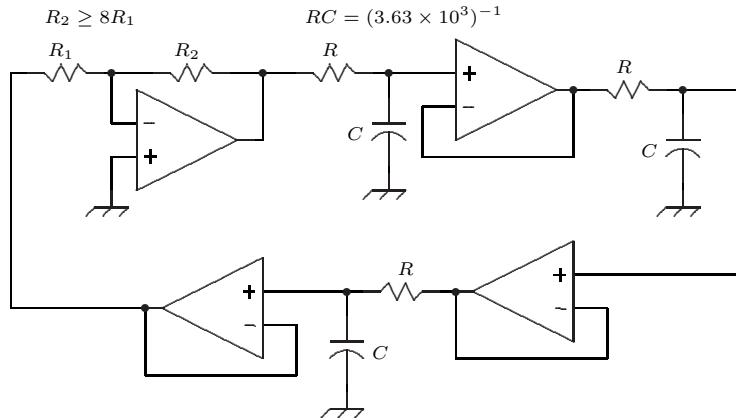


Figure 530: 1 KHz Phase Shift Oscillator

Reference

Ron Mancini of Texas Instruments has put together a very useful summary of sine wave oscillator circuits in [198].

20.2 Wien Bridge Oscillator

The Wien Bridge oscillator is the basis for many audio oscillator circuits. Referring to figure 531, the frequency of oscillation is given by

$$\omega = \frac{1}{RC} \quad (912)$$

For coverage of the audio band (20Hz to 20kHz), it is common to use a dual potentiometer to vary frequency continuously over a 10:1 range. The capacitors are switch selected to select one of 3 ranges¹⁹⁴.

The bridge network has a gain of 1/3 at the oscillation frequency, so the amplifier must provide a gain of slightly more than 3 to satisfy the Barkhausen criterion.

In figure 531, there is no explicit mechanism to adjust the loop gain to unity. The output signal will increase until the op-amp clips the tops of the sine-wave. For lower distortion, there must be a more sophisticated gain mechanism, and we'll look at that in a moment.

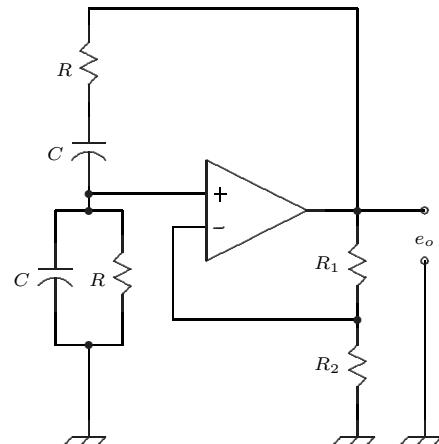


Figure 531: Wien Bridge Oscillator

¹⁹⁴But not always. The original HP200A oscillator used three sets of switched resistors and a massive variable dual-section variable capacitor [103].

Analysis

We need to set up the system so that the loop gain is equal to unity. The amplifier is a non-inverting unit, with a gain

$$\frac{e_o}{e_a} = \left(1 + \frac{R_1}{R_2}\right) \quad (913)$$

The gain of the bridge section is

$$\frac{e_a}{e_o} = \left(\frac{Z_p}{Z_p + Z_s}\right) \quad (914)$$

where

$$\begin{aligned} Z_p &= \frac{1}{sC} \| R \\ &= \frac{R}{1 + sRC} \end{aligned} \quad (915)$$

and

$$\begin{aligned} Z_s &= R + \frac{1}{sC} \\ &= \frac{1 + sRC}{sC} \end{aligned} \quad (916)$$

Substitute for Z_p and Z_s in 914, and after a certain amount of algebraic manipulation,

$$\begin{aligned} \frac{e_a}{e_o} &= \left(\frac{sRC}{1 + 3sRC + s^2R^2C^2}\right) \\ &= \left(\frac{1}{\frac{1}{sRC} + 3 + sRC}\right) \end{aligned} \quad (917)$$

Put

$$\omega_o = \frac{1}{RC} \quad (918)$$

and

$$j\omega = s \quad (919)$$

Substitute these in equation 917 and recall that

$$\frac{1}{j} = -j \quad (920)$$

then

$$\frac{e_a}{e_o} = \frac{1}{3 + j\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)} \quad (921)$$

At the frequency $\omega = \omega_o$, the gain of the bridge network is

$$\frac{e_a}{e_o} = \frac{1}{3} \quad (922)$$

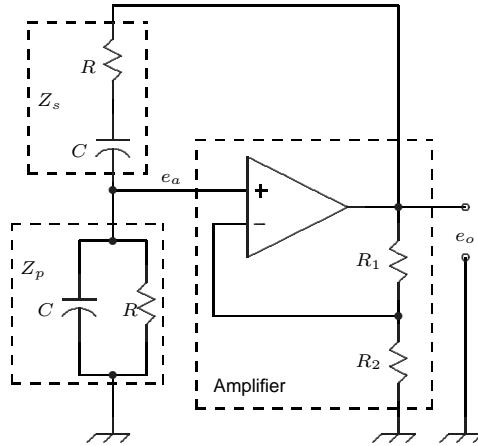


Figure 532: Wien Bridge Analysis

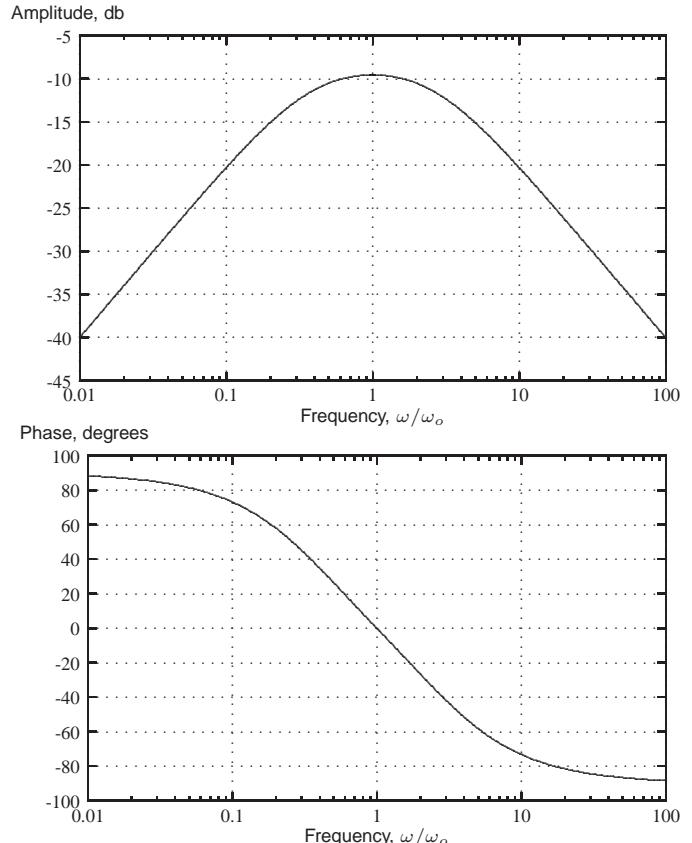


Figure 533: Wien Bridge Magnitude and Phase

Consequently, the gain of the amplifier must be at least 3 to cause the circuit to oscillate. That is

$$R_1 \geq 2R_2 \quad (923)$$

The amplitude and phase responses of the Wien network are shown in figure 533. The amplitude peaks at a value of

$$20 \log_{10}(1/3) = -9.5 \text{ db} \quad (924)$$

at ω_o . The phase transitions from -90° to $+90^\circ$, passing through 0° at ω_o .

20.2.1 References

The Wien Bridge oscillator is often used as a signal source to measure the distortion of audio components. In that type of application, the distortion of the output sine wave is critical, and significant care must be taken to reduce it to the absolute minimum. The references given below show how this is done.

- Jim Williams brings his inimitable style to an analysis of the Wien Bridge oscillator. One of the definitive works on the circuit. [103].
- Sedra and Smith [78] analyse the circuit and show two methods of amplitude stabilization.
- Application Note AN31 in [199] shows a simple light-bulb stabilization circuit.
- Vanderkooy and Koch [200] show a Wien bridge oscillator with high-power output, using a discrete component op-amp. Includes an analysis of the distortion contributed by incandescent lamp stabilization.
- Oliver and Cage [166] contains an alternative analysis of the oscillator circuit with some implementation details.

20.3 Wien Bridge Oscillator, Amplitude Stabilization

The wien-bridge amplifier must have a gain greater than 3 for oscillations to build up. However, once those oscillations have reached the desired amplitude, the gain must be reduced to *exactly* 3. Slightly more than 3, and the signal will grow until it is limited by clipping. Slightly less than 3, and the signal will die away. Consequently, there must be some sort of *amplitude stabilization* system that ensures an exact gain of 3.

The classic solution, invented by William Hewlett, is shown in figure 534. An incandescent lamp replaces one of the resistors in the sensor network of the negative feedback loop.

The incandescent lamp is a non-linear resistance. As the lamp current increases, the filament temperature increases, along with the resistance.

When the circuit is first switched on, the lamp resistance is low, and so the negative feedback signal is reduced. Consequently, the gain is larger than 3 and the output signal increases.

As the lamp heats up, it increases in resistance. This reduces the loop gain until the lamp resistance is exactly equal to $R_1/2$. If the output signal decreases, the lamp resistance again increases to increase the loop gain.

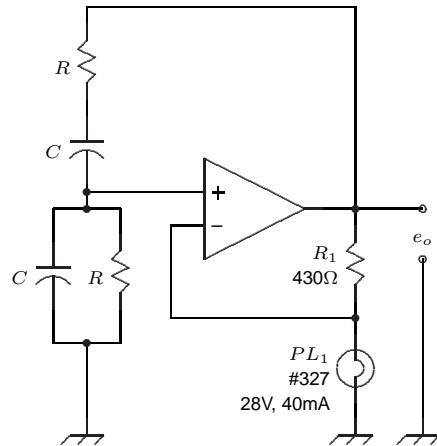


Figure 534: Lamp Stabilization

Over most of the audio frequency range, the thermal inertia of the lamp filament acts to average the magnitude of the oscillator AC output voltage. However, at very low frequencies, the lamp begins to heat up and cool down with the output waveform, distorting the output waveform. As a result, there is a conflict between the settling time of the amplitude stabilization network and the lowest frequency of the oscillator. If the thermal inertia of the lamp is increased, it will work without significant distortion at a lower frequency, but the amplitude stabilization takes longer to settle. A long settling time shows up as a bouncing amplitude as the frequency is adjusted. Some oscillators switch select a different lamp or multiple lamps for operation at low frequency¹⁹⁵.

The values for R_1 and the lamp PL_1 are from [103].

Diode Limiter

Another amplitude stabilization technique (see [78] for example) is shown in figure 535. The sensor resistors R_1 and R_2 are chosen for a gain somewhat greater than 3. When the oscillator is switched on, the output sine wave builds up. When the amplitude of the voltage across R_1 exceeds the threshold of the diodes, they appear as a parallel resistance to R_1 . The effect is to increase the sensor gain, ie, reduce the loop gain. The loop amplitude stabilizes with the peak-peak voltage across R_1 approximately equal to 0.6 volts.

This is simple and reliable, but the nonlinear conduction of the diodes substantially increases the distortion of the output waveform.

Integrator-Comparator

This amplitude stabilization technique uses a full-fledged negative feedback system with no shortcuts or compromises. The output amplitude is compared against a reference and the result used to drive a variable-resistance device which adjusts the forward gain of the oscillator amplifier. The basic idea, based on one of the circuits in [103], is shown in figure 536.

Amplifier A_2 is the amplifier in the gain control negative feedback system. Diode D_1 half-wave rectifies the AC output signal, causing pulses of current through resistor R_2 to be integrated in the capacitor C_f . This cur-

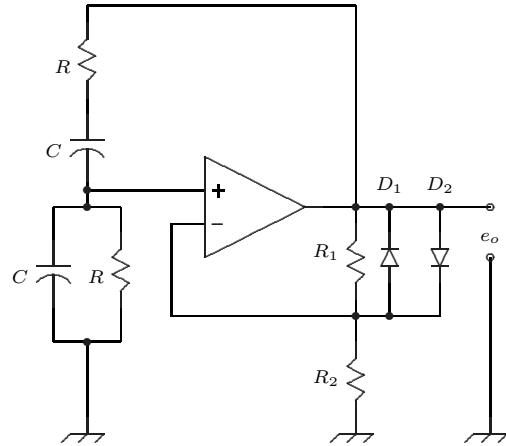


Figure 535: Diode Limiter

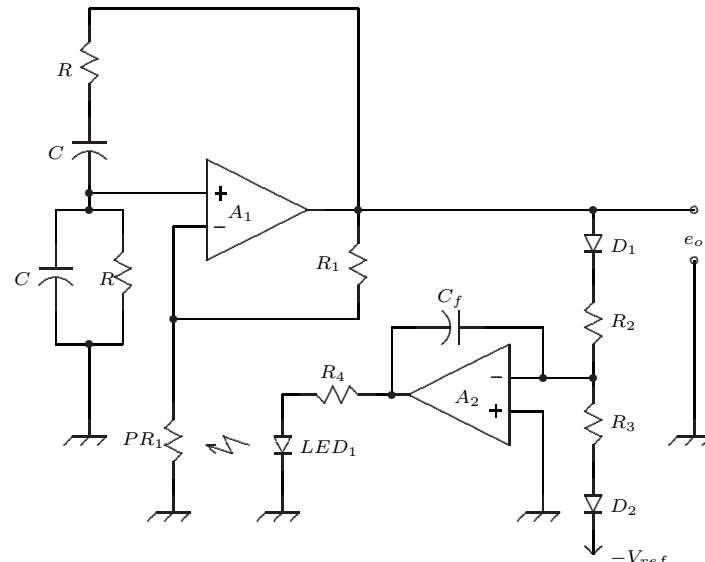


Figure 536: Integrator-Comparator Limiter

¹⁹⁵ Reference [201] shows the lamp in the R_1 position. One wonders how this circuit worked: heating of the lamp would *increase* the loop gain.

rent must on average be balanced by the reference current flowing out of the integrator input and down through R_3 . Diode D_2 provides temperature compensation for D_1 .

The output of the integrator drives light emitting diode LED_1 , which shines on photoresistor PR_1 , which adjusts the gain of the Wien-Bridge oscillator around A_1 .

If the output sine wave is too small to balance the reference current to the integrator, the output of the integrator will ramp upwards, reducing PR_1 , increasing the gain of the Wien-Bridge oscillator. The speed of response of the gain-control mechanism is determined by the R_2C_f time constant.

It is also possible to replace the LED-Photoresistor mechanism with a JFET. The gate voltage is controlled by the output of the integrator, and this controls the effective drain-source resistance, which then adjusts the oscillator gain.

A Comparison

A low-distortion sine wave is required for distortion measurements on audio equipment. It's not absolutely essential because there are work-arounds, but it is convenient if the sine-wave generator distortion is well below that of the audio equipment. Then the distortion contributed by the device under test is relatively easy to measure.

Consequently, in the competition to produce a low-distortion sine wave oscillator, it's interesting to compare the distortion of the sine wave when the amplifier is stabilized by these different methods.

Source	Method	Result
Original HP Circuit [103]	Tube amplifier with lamp	0.5%
Williams [103]	IC op-amp with lamp	0.0025%
Williams [103]	Op-amp with JFET	0.0018%
Vanderkooy-Koch [200]	Discrete op-amp with lamp	0.001%
Williams [103]	Integrator-photoresistor	0.0008%

These figures are interesting in view of the current revival of interest in electron-tube based audio amplifiers, which are claimed to have better listening characteristics than solid-state based designs. The original tube-based HP oscillator produced quite a high distortion level compared to modern solid-state oscillators.

20.4 Quadrature Oscillator

A *quadrature oscillator* produces two sine waves, differing in phase by 90° . The outputs are often referred to as the sine or in-phase component, and the cosine or quadrature component.

An integrator is the obvious building block for such an oscillator: depending on the design of the integrator, it has a constant phase shift between input and output of $+90^\circ$ or -90° . The output amplitude depends on frequency, so the required gain establishes the operating frequency.

In order to satisfy the Barkhausen criterion, there must be $n \times 360^\circ$ of phase shift around the oscillator loop, where n is an integer $0, 1, \dots$. We must have at least one integrator in the loop. Let's say that contributes a phase shift of -90° . What else has to be in the loop for oscillation to occur?

- One possibility is to complete the loop with a circuit element that has a phase shift of $+90^\circ$. Then the net phase shift around the loop is zero. If the loop gain is greater than unity, then the circuit will oscillate.
- Another possibility is to include an identical circuit element that generates an additional phase shift of -90° . Then the two circuit elements generate 180° of phase shift. An inverter, which contributes 180° of phase shift, completes the loop. Then the net phase shift around the loop is again zero and, if the loop gain is greater than unity, the circuit will oscillate.

Basic Quadrature Oscillator

Figure 537 shows the most widely used quadrature oscillator circuit [202], [108], [198]. Amplifier A_1 and its associated components are a non-inverting integrator (see section 13.11), so it contributes -90° of phase shift.

Amplifier A_2 is an inverting integrator (see section 13.10), which contributes $+90^\circ$ of phase shift. So the net phase shift around the loop is zero, satisfying one part of the Barkhausen Criterion.

In theory, all the resistors are equal to R , and the capacitors are equal to C . The gain of each integrator is given by

$$\frac{e_o}{e_i} = \frac{1}{sRC} \quad (925)$$

so the loop gain AB through two of these circuits is

$$AB = \frac{1}{s^2 R^2 C^2} \quad (926)$$

The oscillation frequency occurs where $AB = 1$, in which case

$$\omega = \omega_o = \frac{1}{RC} \text{ radians/sec} \quad (927)$$

or

$$f_o = \frac{1}{2\pi RC} \text{ Hz} \quad (928)$$

In practice, to guarantee that the oscillator will start, the loop gain must be made slightly larger than unity. This is usually accomplished by making the resistor marked * slightly smaller than the calculated value [202], [203].

The zener diodes limit the peak value of the *sine* output to $V_z + 0.6$ volts. This will introduce some distortion into the sine output. If the lowest possible distortion is required, the *cosine* output is preferable.

This circuit is simple and appropriate for a fixed oscillator. But it would be complicated to tune, since three resistors and/or three capacitors need simultaneous adjustment.

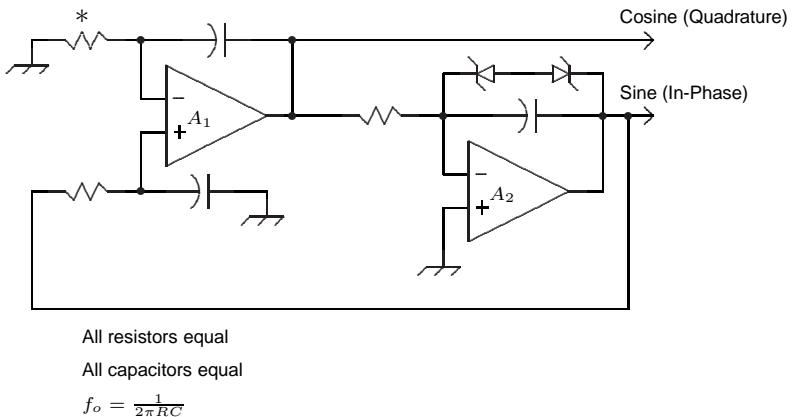


Figure 537: Quadrature Oscillator

Quadrature Oscillator using Howland Integrator

An alternative design for a quadrature oscillator [139], [78] is shown in figure 538.

This is very similar to the circuit shown in figure 537. The non-inverting integrator in figure 537 has been replaced with a non-inverting integrator based on the Howland current source (section 13.11). As in the previous circuit, one resistor (indicated again with an *) is made slightly smaller than calculated in order to create a loop gain slightly greater than unity, thereby guaranteeing that the oscillator will start.

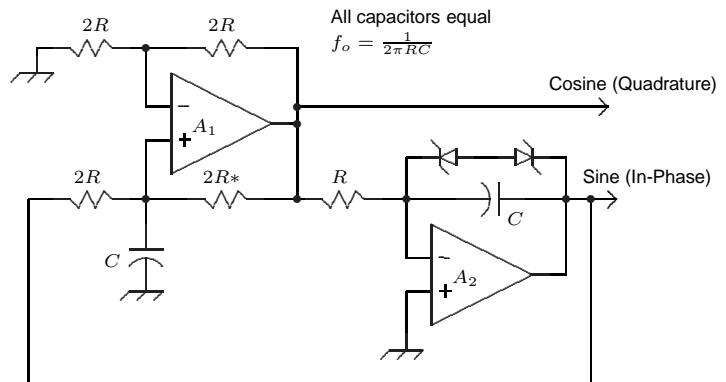


Figure 538: Quadrature Oscillator, Howland Type

Quadrature Oscillator based on State Variable Filter

The second of the two possible approaches to building a quadrature oscillator is shown in figure 539.

The inverting integrators A_2 and A_3 each contribute 90° of phase shift. The remaining 180° of phase shift is contributed by the inverter A_1 . To ensure that the loop gain is slightly larger than unity, R_{1*} should be made slightly smaller than R_2 . As in the previous case, the frequency of oscillation is given by

$$f_o = \frac{1}{2\pi RC} \text{ Hz} \quad (929)$$

As usual by now, the zener diodes limit the amplitude of oscillation.

This oscillator is relatively easy to tune. Resistors R are replaced by a dual-section variable resistance and Capacitance C is replaced by a dual-section switch to select capacitors¹⁹⁶.

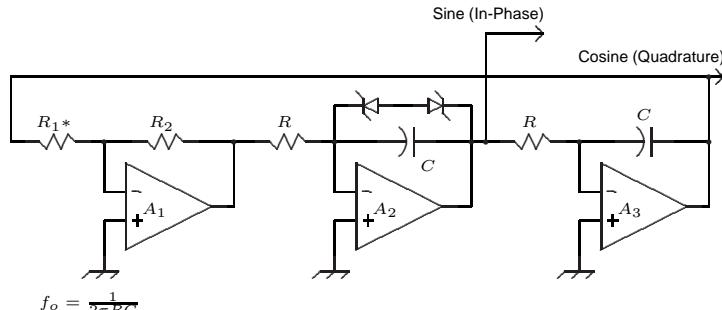


Figure 539: Quadrature Oscillator, State Variable Type

Applications

Two quadrature sine waves are required in a single-sideband modulator using the phasing method, which is the basis for the audio frequency shifter shown in [190]. Certain types of AC servomotors require quadrature sine waves in order to create a rotating magnetic field that carries the rotor with it. In a synchronous demodulator, the in-phase and quadrature components of a signal may be detected by multiplying the signal separately with two quadrature sine waves. Combining the in-phase and quadrature components in different proportions and polarities can be used to generate a sine wave of some arbitrary phase angle.

¹⁹⁶This type of quadrature oscillator is often based on the *state variable* active filter [108], [141]. In that type of implementation, another resistor provides a path for positive feedback in the circuit. However, this complicates the analysis and appears to be unnecessary.

20.5 Three-Phase Oscillator

The primary use of a three-phase oscillator is to synthesize waveforms for three-phase power systems. It produces three sine waves of equal amplitude but equidistant in phase angle, 120° from each other.

One's first impulse might be to build a phase-shift oscillator, as shown in section 20.1, with three sections. Alas, this won't work because the maximum phase shift of an RC section is 90°, and we need 120°. Then suppose we used 6 RC sections, each with 60° phase shift? That satisfies the phase problem, but won't work because the amplitude of the signal decreases as it passes through each RC phase-shifting network. The loop gain is made unity in one amplifier stage. So the RC phase-shift oscillator cannot easily produce the required waveforms.

However, the circuit shown in figure 540, based on a circuit shown in [108], will work.

20.5.1 Analysis

The analysis is useful because it shows how a Bode plot is generated from the transfer function. The Bode plot is most useful because it enables you to visualize how the amplitudes and phases of the waveforms are generated. And the analysis gives us an exact value for the oscillation frequency.

Our method of attack is to

1. Determine the transfer function of one of the three stages
2. Draw the Bode magnitude and phase plots for the stage, in order to get an idea of the function of each stage.
3. Use the fact that the magnitude of each stage transfer function must be unity (0db) at the oscillation frequency.
4. Verify that this frequency corresponds to a phase shift in each stage of 120°.
5. Determine an expression for oscillation frequency.

Ignoring the amplitude stabilization zeners around A_2 , the oscillator is made up of three identical stages. The transfer function of each stage is

$$\frac{e_o}{e_i} = \frac{Z_f}{Z_i} \quad (930)$$

$$\begin{aligned} Z_f &= R \parallel 1/s2C \\ &= \frac{R}{1 + sR2C} \end{aligned} \quad (931)$$

$$Z_i = \frac{R}{2} \quad (932)$$

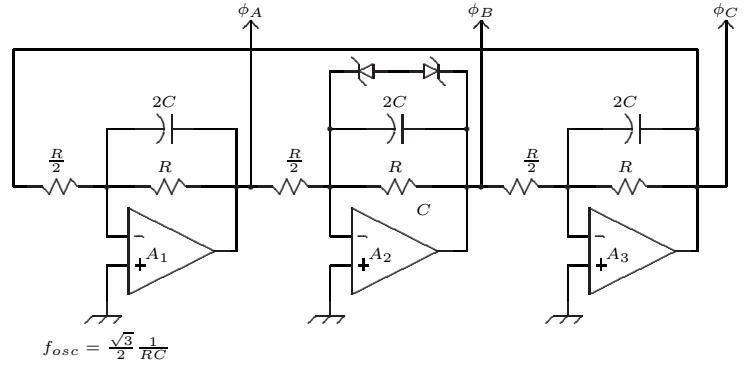


Figure 540: Three Phase Oscillator

Substitute for Z_f and Z_i in equation 930 and after a wee bit of algebra:

$$\frac{e_o}{e_i} = \frac{2}{1 + s2RC} \quad (933)$$

Since we're interested in frequency and phase response, substitute $j\omega$ for s , and

$$\frac{e_o}{e_i} = -\frac{2}{1 + j\omega 2RC} \quad (934)$$

As usual, we'll put $\omega_o = 1/RC$. For many of the oscillator circuits we've been discussing, $\omega_o = 1/RC$ is the oscillation frequency, but not in this case. To distinguish, we'll call the oscillation frequency ω_{osc} . Then

$$\frac{e_o}{e_i} = -\frac{2}{1 + j2\frac{\omega}{\omega_o}} \quad (935)$$

A quick check on our math: at low frequencies, where the capacitor may be considered an open circuit and the j term in equation 935 is small, this reduces to

$$\frac{e_o}{e_i} \approx 2 \quad (936)$$

which does correspond to the behaviour of the circuit.

The amplitude and phase responses of equation 935 are shown in figure 541. The magnitude plot is

$$\begin{aligned} \left| \frac{e_o}{e_i} \right| &= \frac{2}{\sqrt{1^2 + \left(j2\frac{\omega}{\omega_o} \right)^2}} \\ &= \frac{2}{\sqrt{1 - 4\left(\frac{\omega}{\omega_o}\right)^2}} \end{aligned} \quad (937)$$

The phase plot (we'll omit the leading minus sign for the moment) is given by

$$\angle \frac{e_o}{e_i} = \tan^{-1} \left(2\frac{\omega}{\omega_o} \right) \quad (938)$$

As a quick check for errors, notice that the magnitude plot is +6db at low frequencies, which correctly corresponds to a gain of 2v/v. Also, as we'd expect, the phase swings from 0° at low frequencies to 90° at high frequencies.

This is essentially the same as a single time-constant lowpass filter, somewhat shifted horizontally and vertically. From these plots, we can make some useful predictions about the oscillation frequency.

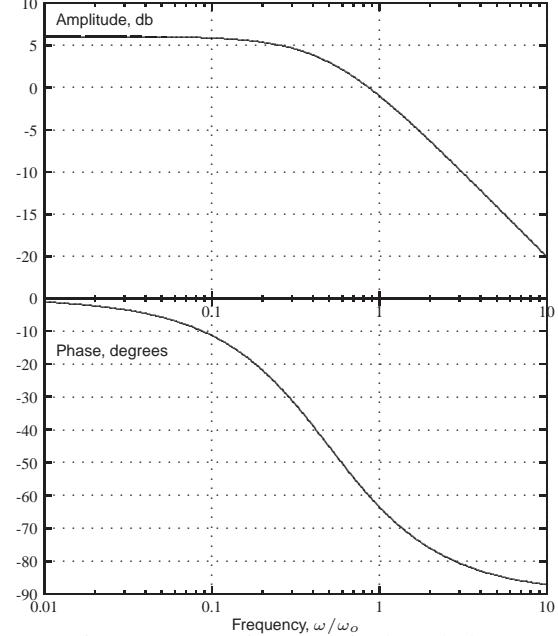


Figure 541: Section Magnitude and Phase

The gain of each of the three stages must be the same and their product must be unity, so each stage must have unity gain, or 0db on the magnitude plot. The amplitude plot passes through 0db at around $\omega/\omega_o = 0.8$. On the phase plot, this frequency corresponds to -60° . In addition to the phase shown on the plot, there is also a phase inversion, which contributes 180° . So the net phase shift through each stage is 120° , as required by the Barkhausen Criterion.

To find the oscillation frequency, we'll use the fact that the magnitude function is unity at that frequency, as required by the Barkhausen Criterion. Then, using equation 937 again:

$$\begin{aligned} \left| \frac{e_o}{e_i} \right| &= \frac{2}{\sqrt{1 - 4 \left(\frac{\omega_{osc}}{\omega_o} \right)^2}} \\ &= 1 \end{aligned} \quad (939)$$

Rearranging and squaring both sides,

$$\begin{aligned} 1 + 4 \left(\frac{\omega_{osc}}{\omega_o} \right)^2 &= 4 \\ \omega_{osc} &= \frac{\sqrt{3}}{2} \omega_o \\ &= 0.866 \omega_o \end{aligned} \quad (940)$$

This correctly corresponds to the zero-crossing of the magnitude curve.

20.6 Schmitt Trigger Astable

A very useful oscillator configuration is shown in figure 542.

The waveforms for the Schmitt Trigger Astable are shown in figure 543.

The operational amplifier uses positive feedback with R_1 and R_2 to make an inverting Schmitt trigger (see section 14.2).

To make the explanation more tractable, we'll assume certain things about the circuit. The op-amp output can swing between 0V and +5V, and the resistors R_1 and R_2 are equal.

Then, applying the superposition theorem to determine the effect of the output voltage and V_b , we can determine that the trip levels for the Schmitt Trigger are

$$v_{ut} = 3.75 \text{ V} \quad (941)$$

for the output voltage $V_o = +5V$, and

$$v_{lt} = 1.25 \text{ V} \quad (942)$$

for the output voltage $V_o = 0V$.

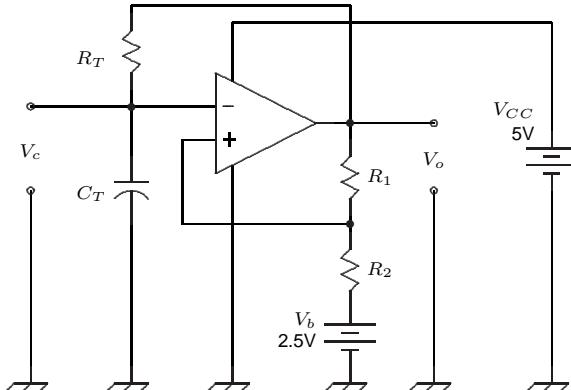


Figure 542: Schmitt Trigger Astable

Let's assume that the output voltage is at +5V. Then the voltage at the non-inverting terminal V_+ of the comparator will be $v_{ut} = +3.5$ volts.

When the voltage at the inverting terminal exceeds the upper trip level, the comparator will rapidly switch into its other state, with V_A at 0V. This sets the voltage at the non-inverting terminal V_+ of the comparator to $v_{lt} = 1.25$ volts, the lower trip level of the Schmitt trigger.

Components R_T and C_T are the *timing* components that make the Schmitt trigger into an oscillator. When the output is HIGH (+5 volts), capacitor C_T charges through R_T until the voltage at the inverting terminal V_- of the comparator reaches the upper trip level of the Schmitt trigger. Then the output switches to zero volts.

When the output is low, the capacitor C_T discharges down to the lower trip level, and the Schmitt trigger changes state to make the output high again. The resultant waveform is shown in figure 543.

Oscillation Frequency

In order to estimate the frequency of this waveform, we need to calculate the time intervals T_1 and T_2 . During T_1 , the voltage across the capacitor follows the exponential charging equation, where V_{CC} is 5 volts.

$$V_c = V_{CC}(1 - e^{-t/R_T C_T}) \quad (943)$$

1. Determine t_1 by substituting 1.25 volts for V_c in equation 943. We find that $t_1 = 0.287R_T C_T$.
2. Determine t_2 by substituting 3.75 volts for V_c in equation 943. We find that $t_2 = 1.38R_T C_T$.
3. Determine T_1 by $T_1 = t_2 - t_1$, and we find that $T_1 = 1.09R_T C_T$.
4. By similar reasoning, $T_2 = 1.09R_T C_T$, so the total period is $T = T_1 + T_2 = 2.18R_T C_T$.
5. The radian frequency $\omega = 1/T$, and $\omega = 2\pi f$ where f is the oscillation frequency. Then

$$f = \frac{1}{2\pi 2.18R_T C_T} \text{ Hz} \quad (944)$$

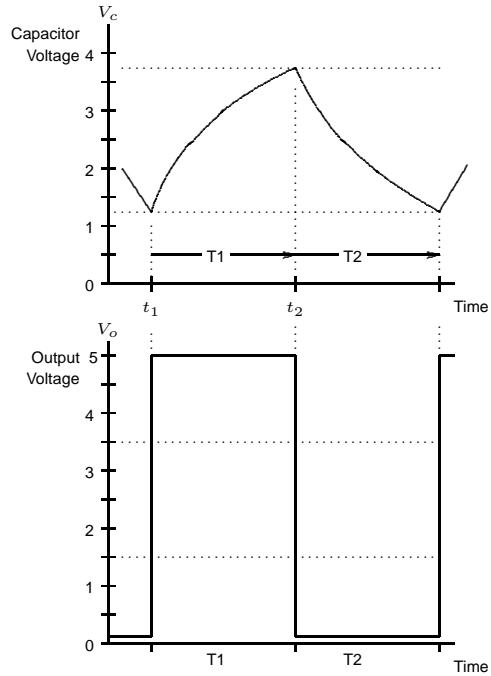


Figure 543: Astable Waveforms

20.7 Function Generator

The *function generator* is a useful oscillator circuit that inherently produces triangle and square waves. The triangle wave can be shaped into a sine wave, and the function generator can be tuned over a wide range of frequencies. The function generator is often the basis of the test signal generator used in an electronics lab.

The function generator does not use positive feedback, but an integrator-schmitt trigger combination that forms a type of astable multivibrator. The basic circuit is shown in figure 544.

Amplifier A_1 is configured as an inverting integrator (section 13.10). Amplifier A_2 is an inverting Schmitt trigger (section 14.2). (Notice that the feedback is to the non-inverting input on A_2).

Circuit Operation

The Schmitt trigger has two stable states. In figure 544 there is no explicit limiting mechanism shown for the op-amps, so we'll assume that the output limit for the Schmitt trigger is established at $\pm V_{sat}$ by the op-amp itself. To simplify things, we'll also assume that R_1 and R_2 are equal.

- Consider that the Schmitt trigger is in its upper state, producing $+V_{sat}$.
- If the input and output of the integrator are e_i and e_o , then

$$e_o = -\frac{1}{RC} \int e_i dt \quad (945)$$

In this case,

$$e_i = V_{sat} \quad (946)$$

so

$$e_o = -\frac{V_{sat}}{RC} t \quad (947)$$

Then the output of the integrator (the triangle wave signal) is ramping in a negative direction with a slope equal to:

$$\frac{e_o}{t} = -\frac{V_{sat}}{RC} \text{ volts/sec} \quad (948)$$

- As the output of the integrator ramps in a negative direction, it carries the bottom end of the R_1, R_2 voltage divider with it. Consequently, the voltage at the non-inverting terminal of the Schmitt trigger is moving negatively as well.

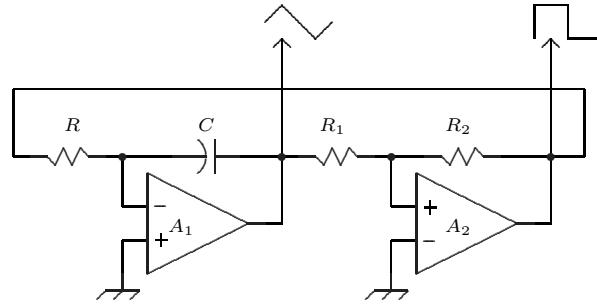


Figure 544: Function Generator

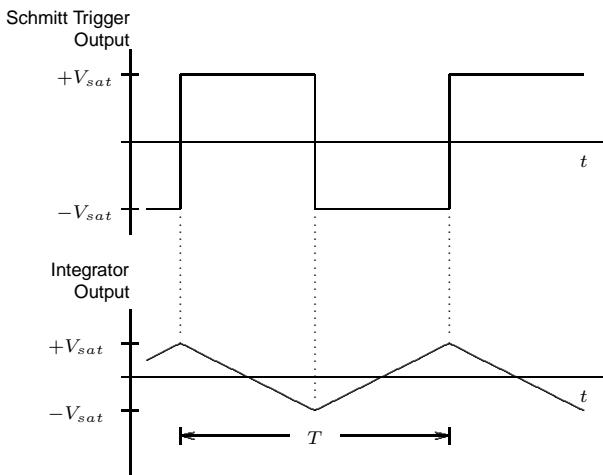


Figure 545: Function Generator Waveforms

- Eventually, when the amplitude of the triangle wave reaches just beyond $-V_{sat}$, the voltage at the non-inverting terminal of the Schmitt trigger passes through zero. This creates a negative error voltage at the input to the Schmitt, so it triggers rapidly into its other state, where the output is $-V_{sat}$.
- The integrator now sees a negative input voltage so its output signal reverses and ramps in a positive direction. This continues until the Schmitt trigger reaches its positive threshold. Then it reverses again and the cycle repeats.

The circuit waveforms are shown in figure 545.

If resistor R or capacitor C change, the slope of the triangle changes, changing the period and the frequency. It is usual to use an adjustable resistor for R to adjust the frequency continuously by a factor of 10:1. The capacitor C is switch selected in decade steps to select the frequency range.

Defining the Square Wave Amplitude

It is important that the amplitude of the square wave be stable, predictable and symmetrical. Any change in the square wave will be reflected in the amplitude and frequency of the triangle wave. The nonlinear circuit used to shape the triangle wave into a sine wave requires that the triangle wave not change amplitude, or significant distortion will ensue.

Consequently, it is usual to have a specific level clamping circuit to define the levels of the square wave. An example is shown in 546.

The voltage into the integrator is now fixed at $\pm V_Z + 0.6$ volts, where V_z is the voltage of the zener diodes.

Alternatively, the output of the Schmitt trigger may be used to switch a reference voltage into the input of the integrator, figure 547. Operational amplifier A_3 is a switched inverter (section 13.5), which produces at point A either $-V_{ref}$ or $+V_{ref}$, depending on the state of the analog switch S^{197} . The job of the Schmitt trigger is then reduced to driving the JFET switch. A bonus of this particular configuration: V_{ref} may be controlled by some other circuitry, which will then have the effect of sweeping the frequency of the generator.

Shaping the Triangle into a Sine Wave

A variety of circuits can be used to shape the triangle wave into a sine wave:

- multi-diode network relying on voltage-divider of reference voltages [204]
- diode network relying on the curved voltage-current characteristic of a silicon diode in the threshold region [205]
- the variable transconductance of a differential amplifier [202].
- square-law characteristic of a JFET [206], [207].

A variety of techniques of sine wave generation and shaping are shown in [208].

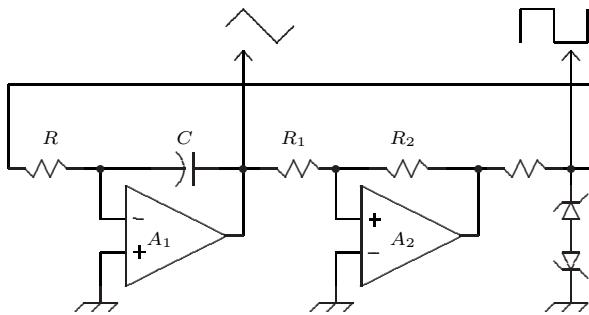


Figure 546: Level Clamping Circuit

¹⁹⁷The 4066 quad analog switch is a suitable device for this.

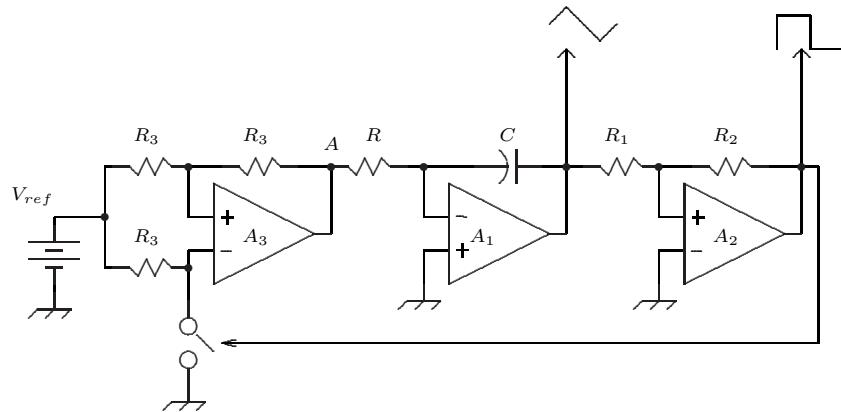


Figure 547: Function Generator with Controlled Inverter

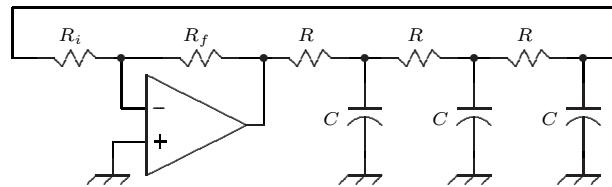
20.8 Exercises

1. The circuit shown below is a *quadrature oscillator*. You may assume that
 - All op-amps are ideal and clip at an output voltage of ± 12 volts.
 - The integrators operate at a frequency where the magnitude of their gain G is unity.

(a) Draw a suitable circuit in the *amplifier* box, showing the ratios of any resistors.
 (b) What will be the frequency of oscillation, in Hz.?
 (c) What is the phase of the signal e_2 as compared to e_1 ?
 (d) If this circuit is to be used to produce a low-distortion sine wave, how should it be modified? Explain your answer briefly.
2. For the phase shift oscillator shown in figure 530:
 - (a) The typical variation of resistors is 5%. If the gain of the amplifier section changes by 10% (from slightly above 8.0 to 8.8, say) how much does the frequency of the oscillator change?
 - (b) What would be required to tune this oscillator over the audio range of 20Hz to 20KHz? Compare the tuning of this oscillator to that of the Wien bridge oscillator, section 20.2.

- (c) Redesign the phase-shift oscillator to use four phase-shift sections and one inverting amplifier section. The output of the last phase-shift section can drive the inverting amplifier directly if R_1 is made much larger than the resistance of the preceding RC phase shift section, in which case a quad op-amp could be used [198].

3. For the phase-shift oscillator shown below:



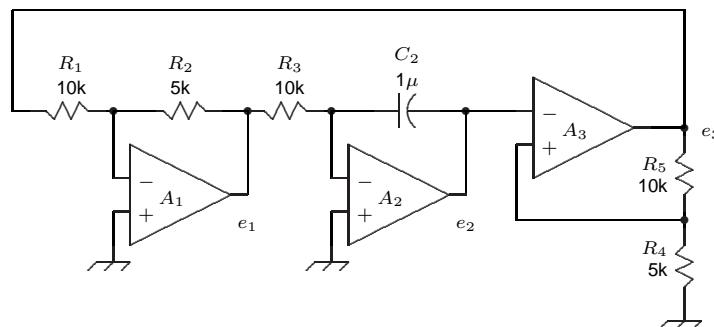
- (a) Will this oscillate at the same frequency as the 3-RC section phase shift oscillator shown in figure 530?
 (b) Estimate the closed-loop amplifier gain $-R_f/R_i$ required of the circuit to make it oscillate.

4. For the amplitude stabilization network shown in figure 536

- (a) Determine the peak-peak value of the oscillator output voltage. You may assume the diodes are ideal (zero volts forward voltage drop.)

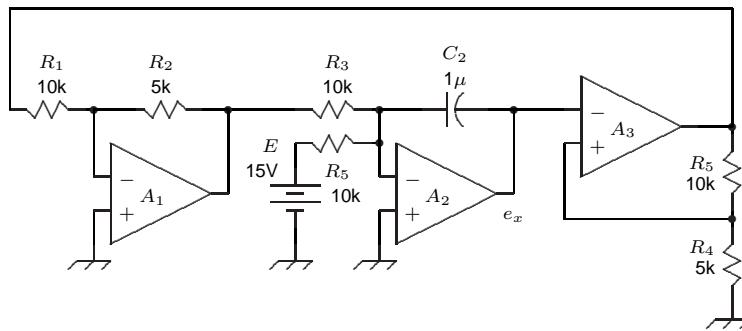
Hints: The average current into the integrator must be zero. You will need to find the average value of a half-wave rectified sine wave.

5. For the function generator shown in figure 544 on page 633, draw the waveform at the non-inverting input terminal of the Schmitt trigger.
 6. The basic function generator of figure 544 is made up of an inverting integrator and an inverting Schmitt trigger. Draw the circuit diagram of a function generator using a non-inverting integrator and a non-inverting Schmitt trigger. Draw the expected waveforms.
 7. Consider the circuit shown below. Assume that all op-amps are ideal and the maximum (saturation) voltages of the op-amps are $\pm 10V$.

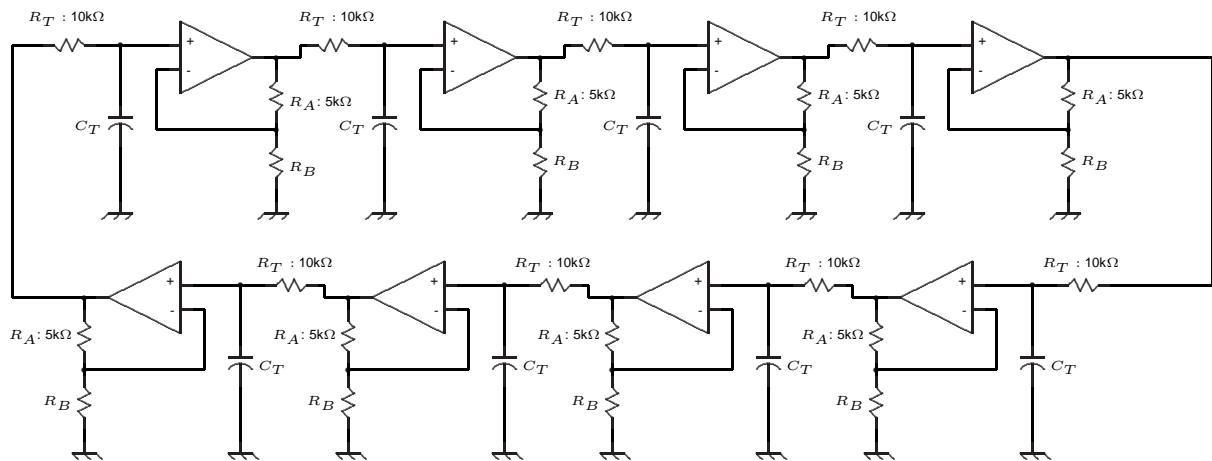


- (a) Draw the waveforms e_1 , e_2 and e_3 .

- (b) Determine the oscillation period.
 (c) If the circuit is modified as shown below, sketch the waveform e_x .

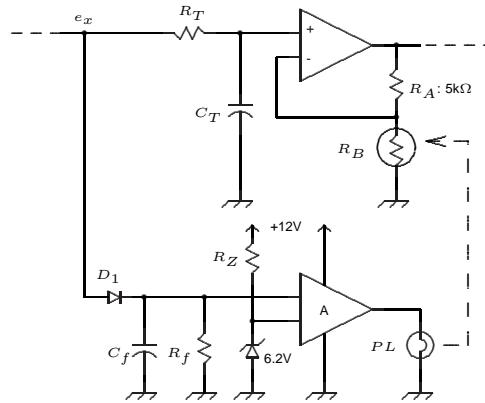


8. The oscillator circuit shown below has 8 identical stages (it's an *octo-oscillator*.) The power supply for each of the 8 op-amps (not shown on the diagram) is ± 12 volts.

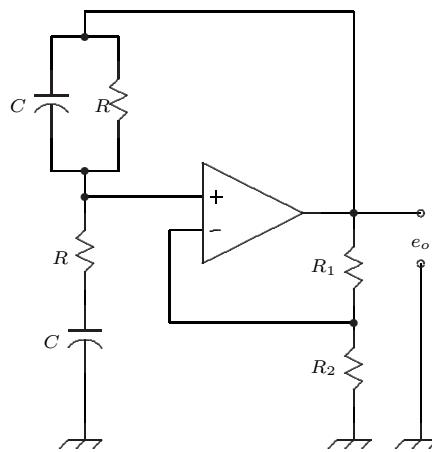


- (a) What phase shift ϕ , in degrees, would occur in each stage?
 (b) Calculate the value of capacitance C_T for an oscillation frequency of 60Hz.
 (c) Calculate the maximum value of R_B for which the circuit oscillates.

The loop gain is stabilized with the negative feedback circuit shown below. The photocell R_B decreases in resistance as the lamp PL increases in brightness. Notice that the op-amp A power supplies are $+12V$ and ground. You may assume that the components in the circuit are such that the circuit operates correctly. Diode D_1 has a forward drop of 0.6 volts when conducting and C_f may be considered a short circuit for AC.



- (a) What is the RMS value of the sine wave at point e_x ?
 (b) Label the polarity of the op-amp inputs so that the amplitude stabilization network is a negative feedback system. Justify this choice.
 9. One version of the Wien bridge oscillator was given in figure 531, page 622. The circuit shown below is an alternative.



Analyse this circuit and determine

- (a) The frequency of oscillation.
 (b) The required closed-loop gain of the amplifier for oscillation to occur.

21 The Non-Ideal Operational Amplifier

In the previous sections of this book, we have generally treated the operational amplifier as an ideal device: input an open circuit, output an ideal voltage source, infinite gain, unlimited frequency response.

In fact, the op-amp is limited in a variety of ways. Fortunately, in many applications these imperfections do not intrude significantly and the ideal approximation is sufficient. This domain of ideal behaviour is constantly expanding as designers improve on the devices in a variety of ways.

It is nearly always possible to find an operational amplifier or combination of op-amps that can fulfill the requirements of an application. However, op-amps are still optimized for certain applications. For example, a device that consumes little power is likely to be constrained to low frequencies. A high-frequency op-amp will probably have relatively relaxed specifications in offset voltage and bias current.

Consequently, it is important to be able to recognize where a particular op-amp parameter should be emphasised at the expense of others, and to be able to choose a suitable device to fulfill that requirement.

In this introduction, we list the parameters that affect the non-ideal behaviour of the op-amp. In following sub-sections, we explain each of these effects in detail and give examples.

Bias Current In the ideal model of an op-amp, we assume that no DC or AC current flows into the input of the op-amp. In fact, a small DC current does flow into (or out of) the input terminals of the op-amp, and this is known as *bias current*. There must be a DC path for this current into the inputs or the amplifier will not work correctly.

Input Resistance We have been assuming that no AC current flows between the input terminals of the op-amp.

In practice, a small amount of current does flow between the terminals, and this can be modelled by the *input resistance* of the amplifier. However, this is large enough that its resistance can be neglected in all but the most extreme amplifier designs.

Offsets There are small current and voltage mismatches in the input stage of an operational amplifier, which are modelled by voltage and current sources. In the early days of op-amps, one would add circuits to compensate for bias current and the offset voltages and currents. That's generally not done any more: one simply chooses an op-amp with sufficiently small values of bias and offsets.

Drifts Not only are there bias currents, offset currents and offset voltages to contend with, but these change with temperature. In critical applications, one must choose an amplifier that has satisfactory drift specifications.

Power Supply Rejection Ratio Ideally, a change in the power supply voltage should have no effect on the output signal of an op-amp. In practice, there is a small effect. It can be modelled in the same way as offset voltage: by a small voltage source in series with one of the inputs.

Common Mode Rejection Ratio Ideally, when both inputs of the op-amp are tied together and driven by an AC source, there should be no effect on the output signal. The output should be proportional only to the difference voltage difference between the two input terminals, not the common mode voltage. In practice, there is an effect, and it can be modelled in the same way as power supply rejection ratio: by a small voltage source in series with one of the inputs.

Finite Open Loop Gain It is a requirement for a negative feedback system to work correctly that the open-loop gain is much larger than the closed-loop gain. At low frequencies, this is probably true, because the open-loop gain of a modern op-amp is typically in excess of 10^4 v/v (80db). However, the gain decreases at higher frequencies, and this 20db/decade rolloff might start at 10Hz. Consequently, the gain at 10kHz will have fallen by 60db (three decades times 20db/decade), leaving an open loop gain of 20db (10v/v). This may very well be insufficiently large compared to the closed-loop gain.

Gain-Bandwidth Product The open-loop gain decreases with increasing frequency. At some point, this gain-frequency characteristic crosses the 0db axis of the Bode plot. That frequency is known as the *gain-bandwidth product*. If the slope of the gain-frequency characteristic is -20db/decade, then one can show that the product of gain and frequency is the same at any point on that characteristic. Consequently, the GBP is a useful indication of the ability of the amplifier to function in a negative feedback system at high frequencies.

Finite Slew Rate An operational amplifier has internal capacitances attached to the signal path in such a way that they must be charged and discharged by the signal currents. If the amplifier is commanded with a large step input, it should ideally respond with a perfect step at the output. In practice, the capacitors charge at a rate limited by the available current at that point in the circuit, and the output responds with a linear ramp at the so-called *slew rate* of the amplifier.

Stability There are certain op-amp behavioural characteristics related to the gain-frequency response and the output impedance that may cause an op-amp to oscillate. Most modern operational amplifiers have characteristics that make them relatively easy to use without the danger of oscillations. Nonetheless, it is a useful skill for the analog circuit designer to understand how oscillations occur and how to cure them.

Noise An operational amplifier is not entirely noiseless. The internal components generate some amount of current and voltage noise, and may be modelled as equivalent noise sources at the input of the amplifier. When the amplifier is configured for high gain, these noise sources may be significant. Or, put another way, if the input signals are comparable to the magnitude of these noise sources, the signal-noise ratio will be compromised. Noise is treated in a separate section, [22](#).

21.1 Finite Gain

Here we illustrate the effect of finite open-loop gain on the closed-loop gain.

Example

What is the exact value of the voltage gain of an inverting amplifier if the ratio R_2/R_1 is 100 and the open-loop gain A is 80db?

Solution

- Convert the open-loop gain to a voltage ratio:

$$20 \log_{10} A = 80 \text{ db} \quad (949)$$

so $A = 10^4$ volts/volt.

- Determine the sensor gain B :

$$\begin{aligned} B &= \frac{R_1}{R_1 + R_2} \\ &= \frac{1}{\frac{R_1 + R_2}{R_1}} \\ &= \frac{1}{1 + \frac{R_2}{R_1}} \end{aligned}$$

$$= \frac{1}{101} \quad (950)$$

3. Apply equation 471 to determine the exact value of the gain:

$$\begin{aligned} \frac{E_o}{E_i} &= -\frac{R_2}{R_1} \left(\frac{AB}{1+AB} \right) \\ &= -100 \left(\frac{10000/101}{1+10000/101} \right) \\ &= 99.990 \text{ volts/volt} \end{aligned} \quad (951)$$

21.2 Bias Current

Definition

The *bias current* of an op-amp is the average input current.

Model

If you picture an op-amp with a plain BJT differential pair in the first stage, the bias current is the base current of these transistors. If the transistors are NPN, the bias current flows into the op-amp. If they are PNP, it flows out.

In either case, bias current is approximately independent of the input voltage, so it may be modelled by constant current sources or sinks, as shown in figure 548.

In this figure, the bias currents are shown as flowing out of the op-amp input terminals.

Typical Values

Type	Part Number	Value
General Purpose BJT	741	$500 \times 10^{-9} A$
General Purpose JFET	TL074	$50 \times 10^{-12} A$
General Purpose MOSFET	LMC660	$2 \times 10^{-12} A$

These are worst case values. Typical values are about one order of magnitude less. Clearly, the JFET and MOSFET devices are superior to the BJT op-amp. The bias current of a MOSFET or BJT device is roughly constant with temperature. The bias current of a JFET device increases exponentially with temperature. At elevated temperatures (100°C, for example) the BJT may actually have a lower bias current than a JFET device. This might be important in an instrumentation system near a high-temperature machine such as an automobile engine.

Critical Applications

Bias current is critical in any circuit

- supplying the bias current through large-value resistors,
- circuits with capacitive feedback (such as the integrator or charge amplifier), and

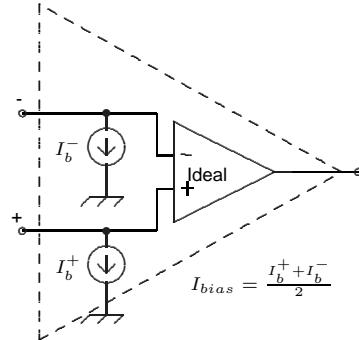


Figure 548: Bias Current Model

- any circuit with small signal currents, such as the photodiode amplifier.

Example

To see the effect of bias current, we'll consider the inverting-amplifier circuit of figure 549(a). The input voltage is reduced to zero to determine the effect of bias current. To do the analysis, the op-amp in figure 549(a) is replaced with the bias-current model of figure 549(b).

The non-inverting terminal is grounded, so the bias current i_b^+ is short circuited and the voltage at the non-inverting terminal is zero volts. The ideal op-amp holds the inverting terminal at the same potential, zero volts. Both ends of R_i are at zero volts, and there is no current through R_i . Then the entire bias current i_b^- must flow through the feedback resistor R_f . The output voltage must be at

$$\begin{aligned} e_o &= i_b^- R_f \\ &= (500 \times 10^{-9}) \times (1 \times 10^6) \\ &= 0.5 \text{ volts} \end{aligned}$$

This is an error voltage, and it is a long way from the desired output, which is zero.

The simplest way to reduce this error voltage is to use a different op-amp, either a JFET or MOSFET device.

Bias Current Correction

When a low-bias current op-amp cannot be used in the circuit, then the bias current may be corrected with the addition of resistor R_c as shown in figure 550. If R_c is chosen correctly and the bias currents are equal, then the output voltage e_o will be zero for zero input voltage. The bias-created output error voltage is eliminated. Now we'll determine the value for R_c :

If the output voltage is zero and the op-amp maintains the input terminals of the ideal op-amp at the same voltage, by KVL we can write

$$\begin{aligned} V_i &= V_f \\ &= V_c. \end{aligned}$$

Applying Ohm's law to these voltages:

$$\begin{aligned} I_i R_i &= I_f R_f \\ &= I_b R_c \end{aligned}$$

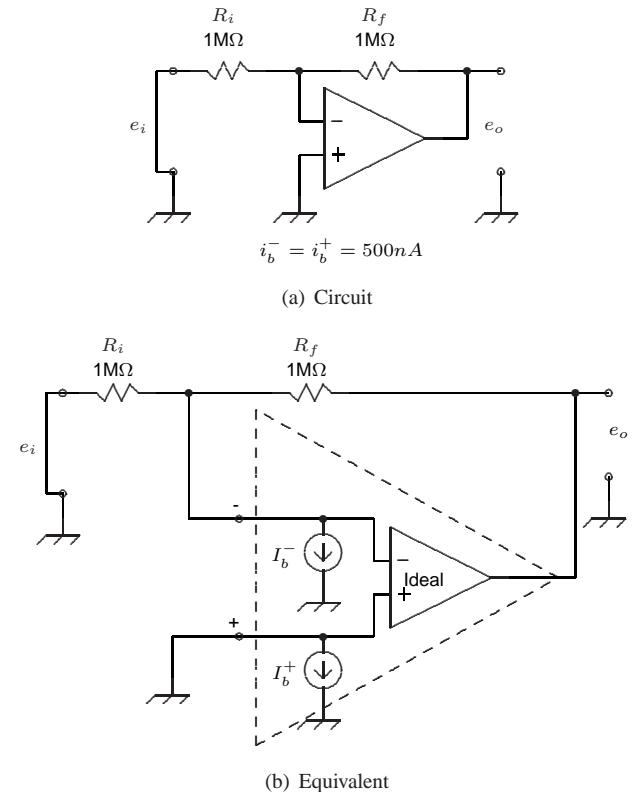


Figure 549: Bias Current Example

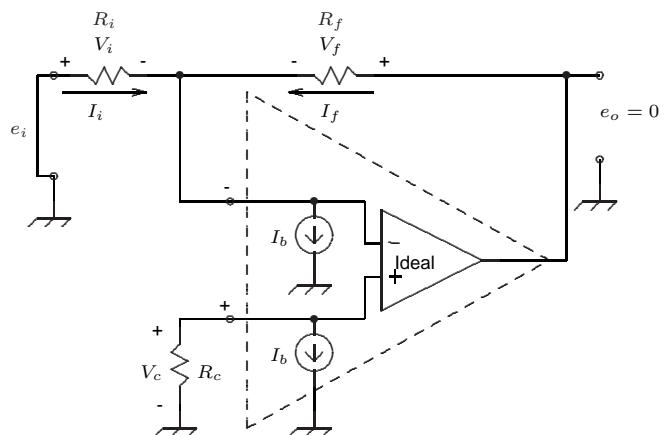


Figure 550: Bias Current Correction

We can rewrite these to put I_i and I_f in terms of I_b :

$$I_i = \frac{R_c}{R_i} I_b \quad (952)$$

$$I_f = \frac{R_c}{R_f} I_b \quad (953)$$

We can also apply KCL to the inverting input terminal:

$$I_b = I_i + I_f \quad (954)$$

Substituting for I_i and I_f from 952 and 953 into 954:

$$I_b = \frac{R_c}{R_i} I_b + \frac{R_c}{R_f} I_b \quad (955)$$

Cancel I_b and solve for $1/R_c$, and we obtain

$$\frac{1}{R_c} = \frac{1}{R_i} + \frac{1}{R_f} \quad (956)$$

That is

$$R_c = R_i \parallel R_f \quad (957)$$

Consequently, R_c should be made equal to the parallel combination of R_i and R_f to cancel the effect of bias current. This is a specific solution for a particular circuit, and other circuits will require other methods of bias compensation.

21.3 Offset Voltage

Definition

The *offset voltage* of an op-amp is a voltage that must be applied to the input terminals, through two equal resistors, to obtain zero output voltage [209].

Model

The output voltage of an ideal op-amp is given by $e_o = A \times e_d$ where A is the open-loop voltage gain of the amplifier and e_d the voltage between the two input terminals. Consequently, if the input terminals are shorted, reducing e_d to zero, the output should become zero as well.

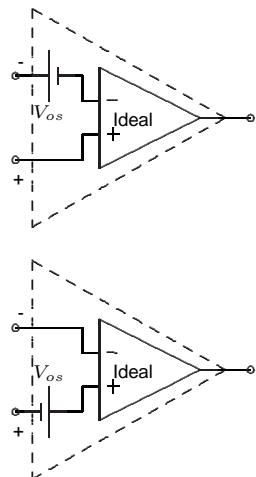


Figure 551: Offset Voltage Model

In practice, this is not quite the case. Because of mismatches in the input transistors of the op-amp, it is necessary to put a small voltage between the two terminals, opposite in polarity to the input offset voltage, to set the output to zero volts.

As shown in figure 551, the offset voltage may be modelled as a small voltage in series with either of the inputs. The two model circuits are equivalent, and either one can be used to represent a given offset voltage.

The magnitude of the offset voltage (worst case) is specified by the manufacturer, but the polarity is not predictable.

Typical Values

Type	Part Number	Offset Voltage	Offset Voltage Drift
		V_{os} μVolts	$\Delta V_{os}/\Delta T$ μVolts/°C
General Purpose BJT	741	±15000	15
General Purpose JFET	TL074	±5000	10
General Purpose MOSFET	LMC660	±1000	1.3
Precision BJT	LM637	±25	0.3
Chopper Stabilized	LTC1052	±5	0.05

The so-called *precision* operational amplifiers have been optimized for low noise and voltage offset, and are usually significantly more expensive than general purpose devices. The *chopper stabilized* amplifier uses a special technique (section 35.10) to obtain very small offset voltages.

Offset Voltage Drift

It is not difficult to compensate for the offset voltage of an op-amp. However, the offset voltage changes with temperature, and this is much more difficult to deal with. As shown in the last column of the table, the drift declines with the offset voltage itself. So a low-offset voltage op-amp probably has a low drift figure as well.

Critical Applications

In a DC amplifier, offset voltage and offset voltage drift are indistinguishable from the signal. Consequently, offset voltage is important where the input signal is comparable to the magnitude of the offset. Put another way, to obtain a satisfactory signal to noise ratio, the offset voltage must be small compared to the signal.

Offset voltage is generally not important in AC coupled amplifiers. In such amplifiers, it is possible to have a low DC gain accompany a large AC gain. The offset voltage is multiplied by the DC gain, but because the DC gain is small offset is generally not a problem.

Example

To see the effect of offset voltage, we'll consider the inverting-amplifier circuit of figure 552.

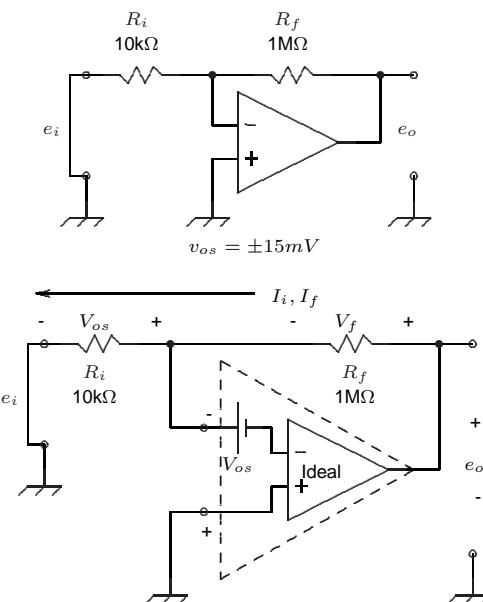


Figure 552: Offset Voltage Example

To do the analysis, the op-amp in figure 552 is replaced with the one of the offset voltage models of figure 551. The choice of model from figure 551 is arbitrary: we can choose whichever model yields the simplest analysis.

The offset voltage source V_{os} creates a current I_i through input resistor R_i that is equal to V_{os}/R_i . Since no current can flow into the input terminals of the op-amp, this same current must flow through the feedback resistor, setting up a voltage V_f across it, where V_f is given by

$$\begin{aligned} V_f &= I_i R_f \\ &= \frac{V_{os}}{R_i} R_f \end{aligned}$$

By KVL around the outer loop of the circuit

$$\begin{aligned} e_o &= V_{os} + V_f \\ &= V_{os} + \frac{V_{os}}{R_i} R_f \\ &= V_{os} \left(1 + \frac{R_f}{R_i} \right) \end{aligned}$$

Put another way, the gain for the offset voltage is:

$$\frac{e_o}{V_{os}} = 1 + \frac{R_f}{R_i} \quad (958)$$

Noise Gain

The quantity on the right side of equation 958 is known as the *noise gain* G_n , since it is the amount by which an input noise voltage is multiplied to appear at the output (section 22.4).

In words, *the offset voltage appears at the output multiplied by the noise gain of the amplifier*. Consequently, the offset voltage becomes more critical as the noise gain (and the inverting gain R_f/R_i) increase.

Offset Voltage Correction

In general, offset voltage is a problem in DC coupled amplifiers with a large voltage gain. Assuming that the voltage gain cannot be reduced the options are:

- Choose an op-amp with a smaller offset voltage. This is the preferred solution in an production situation, because it does not require any human intervention.
- Use AC coupling in the signal path (see sections 13.18 to 13.22). This is only possible if the signal spectrum does not extend down to zero frequency. For example, this is true for audio signals, where the minimum signal frequency is usually taken as 20Hz. However, for very slowly changing signals, such a temperature measurement system, DC coupling must be used.

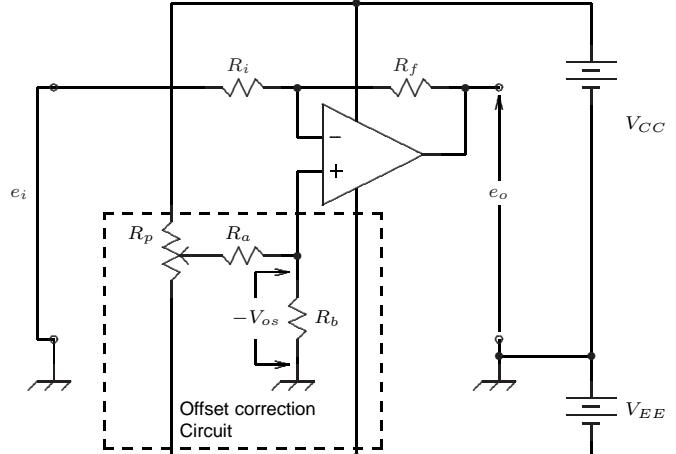


Figure 553: Offset Voltage Correction

When AC coupling is used, the AC gain can be different from the DC gain. Specifically, the AC gain can be large while the DC gain is small. The offset is a DC signal, so it is a function of the small DC gain, and therefore not a problem.

- Use an *offset-null* pot. Some amplifiers (such as the venerable 741) have terminals to which an offset-null potentiometer can be connected. This pot is then adjusted to reduce the offset voltage to zero. The input source is replaced by its internal resistance and the offset-null pot adjusted for zero output voltage. Nulling the offset generally requires the services of a human being so it tends to be expensive. As well, in the hands of someone who doesn't understand the circuit, a potentiometer allows for the possibility of improper adjustment.
- Add an offset-null circuit, as shown in figure 553. This circuit adds an expensive pot, the requirement for human intervention, and two resistors, so it is usually not the most economical solution. However, when the op-amp has other desirable characteristics (such as wide bandwidth) in the face of excessive offset voltage and no terminals for an offset pot, this circuit may be useful.

The potentiometer R_p creates a variable voltage, which is then divided down by the R_a, R_b voltage divider to create an adjustable voltage equal to and opposite polarity to the offset voltage. R_a is much larger than R_b in order to obtain the correct divider ratio. For example, with ± 15 volt power supplies and a $\pm 15\text{mV}$ value of offset voltage, the ratio of R_a/R_b would be set to 1000 : 1. R_p would be chosen to be approximately equal to $R_a/10$.

21.4 Power Supply Rejection Ratio

Design Context

Ideally, a change in the power supply voltage of an op-amp should have no effect on the output voltage. In practice, a change of x volts in the power supply voltage causes a change in the input voltage of x/PSRR , where PSRR is the *power supply rejection ratio*. Consequently, there are situations where drift, AC ripple, or other power-supply line noise will appear at the output of an amplifier.

Here are some examples where this might be important:

- An audio amplifier is intended for the consumer market, for which manufacturing cost is an important consideration. It is expensive to regulate the power supply, so the designer would like to provide an unregulated power supply. Along with the desired DC component, an unregulated supply produces some level of AC ripple riding on the DC component. The designer would like to rely on the PSRR of the amplifier to ignore the ripple signal. Since the amplifier is probably AC coupled, DC changes on the power supply line (caused by the $\pm 10\%$ variation of line voltage) will not affect the output signal directly. (They might affect the amplifier bias conditions, however, so this should be checked.) However, the AC ripple at 120Hz must be kept to a very small level in the output signal or it will be audible, and this will require excellent PSRR performance in the amplifier.
- A computer data logging system is being designed in which an op-amp preamplifier must co-exist with digital logic. For reasons of efficiency, the power supply is a switching type. This mixed signal device will undoubtedly have some level of digital switching noise on the power supply lines, from the logic or switching power supply, or both.

Ideally, the PSRR of the op-amp will reject noise on the power supply line. However, many specifications of an op-amp degrade at high frequencies and since switching noise contains high frequencies, this should be carefully checked. (There are several strategies which may be used to minimize the amount of digital noise on the amplifier power supply lines, and these should be the first line of defence, rather than relying on the amplifier PSRR. See section 34).

- An instrumentation system contains a DC coupled amplifier of gain 1000V/V. Any change in offset voltage (due to finite PSRR, in this case) will cause the output voltage to change. Since this is a high gain amplifier, the input offset voltage cannot be allowed to change substantially. The designer should use the regulation figure for the power supply and the PSRR figure for the op-amp to ensure that the output drift is within the required specification.

Definition

The power supply rejection ratio (PSRR) is the ratio of the change in power supply voltage to the equivalent change in input offset voltage [209].

That is,

$$\text{PSRR(db)} = \frac{\Delta V_{ps}}{\Delta V_{os}} \quad (959)$$

where

ΔV_{ps} is the variation in power supply voltage, volts

ΔV_{os} is the equivalent variation in input offset voltage due to variations in supply voltage, volts

The PSRR is often expressed in decibels, in which case the definition is:

$$\text{PSRR(db)} = 20 \log \frac{\Delta V_{ps}}{\Delta V_{os}} \quad (960)$$

Model

The power supply rejection ratio can be modelled as shown in figure 554. The power supply has been modelled as a DC voltage source V_{ps} in series with an AC voltage source e_{ps} . The AC source represents both the slow variations in the power supply voltage and any ripple or noise component.

The AC voltage source generates an equivalent input offset voltage v_{os} . From there, the effect of v_{os} on the output voltage depends on the feedback circuitry around the op-amp.

Using our understanding of the effect of offset voltage from section 21.3, we can predict that the effect on the output voltage will be proportional to the voltage gain of the circuit, and this effect will be problematic when the voltage gain is large.

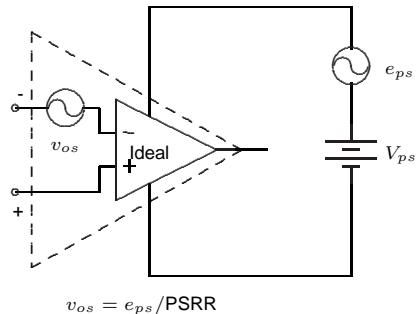


Figure 554: Power Supply Rejection Ratio, Model

Typical Values

Type	Part Number	PSRR db	PSRR V/V	Supply Type
General Purpose BJT	741	80	10000	Split
Single Supply BJT	LM324	65	1778	Single
Low noise (audio) BJT	SE5532A	86	100000	Dual
General Purpose JFET	TL074	70	3162	Split
General Purpose MOSFET	LMC660	68	2511	Single
General Purpose BIFET	TLC074	130	3162277	Single
Precision BJT	LM637	117	707000	Split
Chopper Stabilized	LTC1052	120	1000000	Split

Pitfalls of the PSRR

Split and Single Supply An operational amplifier may be operated from a split power supply (positive and negative with respect to ground) or from a single power supply.

When the op-amp is intended to be operated from a split supply, the PSRR is measured by varying *both* power supplies by the same amount, in opposite directions. For a single power supply, the single voltage is varied in one direction. The split supply case results in better PSRR figures, because the average voltage across the op-amp circuitry is not changing, and this is easier for the circuitry to ignore. A single supply amplifier often has a poorer PSRR figure because the average supply voltage is changing. This shows up in the Typical Values of the table shown above.

However, notice the spec for the TLC074, a recent design (2000). The PSRR is much larger than the other single or split supply designs.

Were one to measure the PSRR of a split supply amplifier by changing one of the supply magnitudes and not the other, the PSRR figure is certain to be significantly worse.

Frequency Dependence of PSRR The power supply rejection ratio is invariably specified at low frequencies, which is fine for slow variations and AC ripple in the power supply voltage. However, the PSRR decreases at high frequencies, and the amplifier is less able to ignore noise on the power supply lines at high frequencies.

A typical graph of PSRR vs frequency shows that PSRR decreases with frequency in much the same fashion as the open-loop gain: 20db/decade. For example, the PSRR for the TLC074 amplifier is 130db up to 10Hz, and then decreases to a value of unity at 10MHz, as shown in figure 555, adapted from [210].

Controlling PSRR Problems

As in the case of offset voltage, the finite value of PSRR is most evident when the closed-loop gain is large and there is significant noise on the power supply lines. The options are:

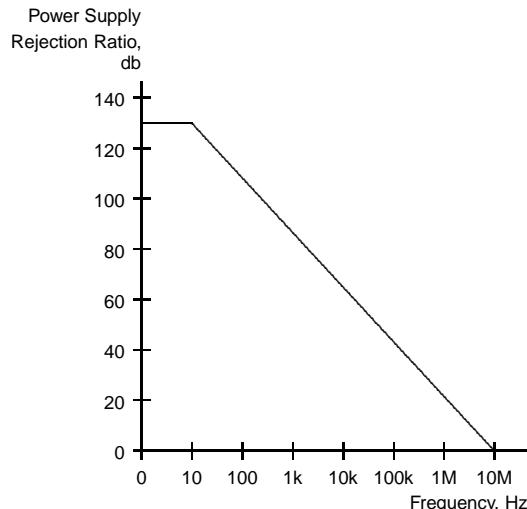


Figure 555: PSRR vs Frequency

- Reduce the noise on the power supply lines through regulation, filtering or rearrangement of the power supply and ground wiring.
- Choose an amplifier with a higher value of PSRR. A split supply amplifier will have a better PSRR than a single supply amplifier, but only if the noise on both power supply lines is identical and opposite polarity. This might be the case for AC hum, but would not be the case for noise induced from nearby digital circuitry.

21.5 Common Mode Rejection Ratio

First we'll look at the definition and description of Common Mode Rejection Ratio and then we'll see how this affects a circuit design.

It is fundamental to the operation of the op-amp that the output is proportional to the *difference* between the voltages at the two inputs.

The equation governing the output of an ideal op-amp illustrates this:

$$E_o = A_{ol}(E^+ - E^-) \quad (961)$$

where E^+ and E^- are the voltages at the two input terminals. This model is represented in figure 556. (The power supplies V_{CC} and V_{EE} , which are often left out of an op-amp diagram, are shown in this case in order to clarify the location of the ground reference point.)

The same model may be redrawn in terms of a *difference voltage* e_D and *common-mode voltage* e_{CM} as shown in figure, as shown in figure 557, where:

$$\begin{aligned} E_D &= E^+ - E^- \\ E_{CM} &= \frac{E^+ + E^-}{2} \end{aligned}$$

In words:

- The difference voltage is the voltage between the input terminals of the op-amp
- The common-mode voltage is the average input voltage at the input terminals of the op-amp.

Then the equation governing the output of an ideal op-amp becomes

$$e_o = A_{ol}E_D \quad (962)$$

Ideally the common-mode voltage is totally ignored, but in practice the common mode voltage effectively creates an offset voltage, which then affects the output, figure 558.

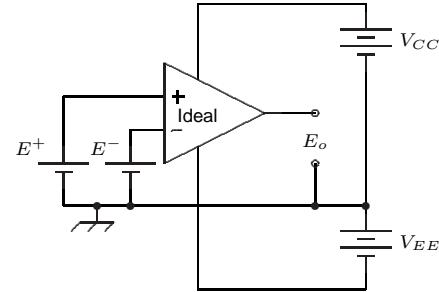


Figure 556: Common-Mode Rejection Ratio, Model 1

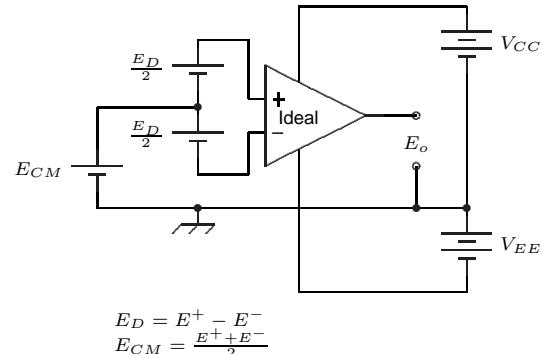


Figure 557: Common-Mode Rejection Ratio, Model 2

Definition

The common-mode rejection ratio (CMRR) is the ratio of the change of input common-mode voltage to the change input offset voltage [209].

$$CMRR = \frac{E_{CM}}{E_{os}} \quad (963)$$

This is often expressed in decibels.

$$CMRR_{db} = 20 \log_{10} \left(\frac{E_{CM}}{E_{os}} \right) \quad (964)$$

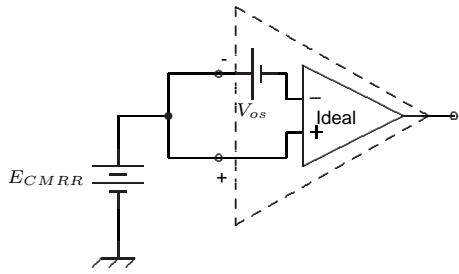


Figure 558: Common-Mode Rejection Ratio, Model 3

Typical Values

Type	Part Number	CMRR db
General Purpose BJT	741	70
Single Supply BJT	LM324	65
Low noise (audio) BJT	SE5532A	80
General Purpose JFET	TL074	70
General Purpose MOSFET	LMC660	68
General Purpose BIFET	TLC074	100
Precision BJT	LM637	126
Chopper Stabilized	LTC1052	120

Design Context

There are at least three situations where the CMRR of an op-amp may be important: difference amplifier, low-distortion amplifier, and high-frequency amplifier.

The Difference Amplifier

The *subtractor* or *difference amplifier* was discussed in section 13.6. One possible difference amplifier is shown in figure 559. The output of this circuit should be proportional to the difference between the two input signals. However, in addition to the desired difference signal, there will also be a small error term in the output proportional to the common-mode input signal.

In the circuit of figure 559, any mismatch between the resistors will add to this error term. However, even if the resistors are identical there will be an error term due to the finite common-mode rejection ratio of the op-amp. For example, the CMRR of the LM324 op-amp is given as 65db worst case, or a voltage ratio of 1778V/V. Consequently, in the circuit of figure 559 a common-mode input signal of 10 volts at the input to the circuit will create a common-mode voltage of 5 volts at the input terminals to the op-amp.

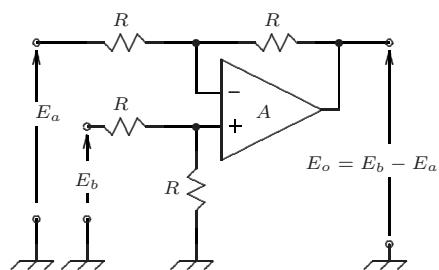


Figure 559: Difference Amplifier

This creates an offset voltage that is $5/1778 = 2.81\text{mV}$ due to the CMRR. This offset voltage is then amplified by a factor of 2 and appears at the output as an error voltage of 5.6mV .

This may appear to be a negligible error. However, in many practical cases the output of the difference amplifier is multiplied by a large gain factor, such as 1000V/V . In that case, a 5.6mV error becomes a 5.6 volt error, intolerably large.

Low Distortion Amplifier

In [103], Jim Williams describes the development of a low-distortion Wien Bridge oscillator. In his quest for the ultimate in signal purity, he replaced the usual non-inverting amplifier configuration with an inverting amplifier. The distortion went from 0.0015% to better than 0.0003%, the limit of his measuring equipment.

Williams characterises this effect as Williams Rule: *always invert*. From [103]:

This rule, promulgated after countless wars with bizarre, mysterious and stubborn effects in a variety of circuits, is designed to avoid the mercurial results of imperfect common mode rejection. Common mode-induced effects are often difficult to predict and diagnose, let alone cure. A zero volt summing point is a very friendly, very reassuring place. It is (nominally) predictable, mathematically docile, and immune from the sneaky common mode dragons.

Consider the non-inverting and inverting unity-gain amplifiers shown in figure 560.

For the non-inverting amplifier, both input terminals of the op-amp must track the input signal. Consequently, the common-mode signal is equal to the input signal. If there are any defects in the common mode rejection ratio, this will distort the output signal in some manner.

For the inverting amplifier, the common-mode input signal is zero. Both input terminals are held at or very close to ground potential. Consequently, any limitations in common-mode rejection ratio will have no effect on the signal.

For example, if one were designing an audio signal processing circuit where low distortion is a major requirement, the inverting configuration should be the first choice.

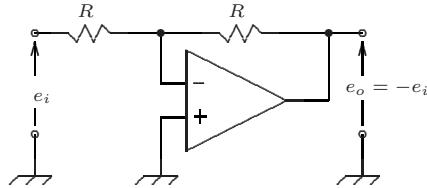
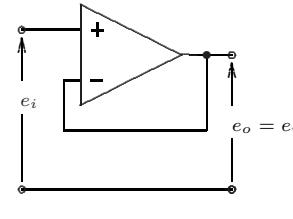


Figure 560: Inverting and Non-Inverting Amplifiers

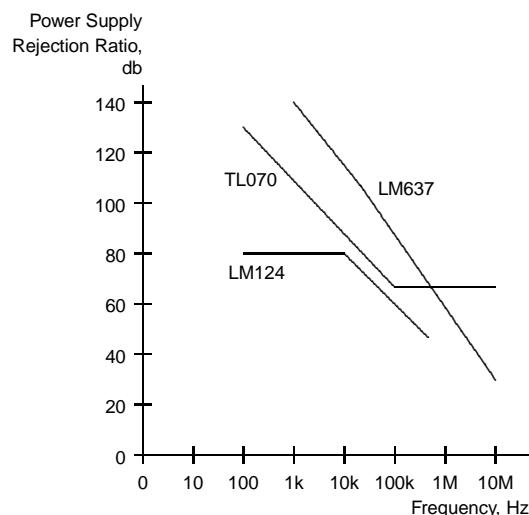


Figure 561: CMRR vs Frequency

High Frequency Amplifiers

Like the PSRR, the CMRR degrades at high frequencies. Figure 561 shows the approximated frequency dependence for three different operational amplifiers (adapted from their datasheets). If common-mode rejection ratio of high frequencies is important, then the cmrr-vs-frequency specification should be used, rather than a single figure which will apply to the low-frequency behaviour.

21.6 Output Voltage and Current Limit

In the ideal model of the op-amp, the output signal is generated by a voltage source. Based on the physical size of most op-amps, it's intuitively obvious that this voltage source cannot produce infinite current or voltage.

The reality is shown in figure 562, the positive output characteristic for a typical op-amp. This information is often provided on the datasheet for a particular op-amp, and would apply for a power supply voltage ± 15 volts. The positive output characteristic defines the region of positive output voltages and positive currents (ie, currents sourced by the op-amp) that can be produced by the op-amp. It includes the entire area to the left of the characteristic: anything to the right of the characteristic is impossible.

It is important to recognize that this graph shows the *maximum* capabilities of the amplifier: for example, under no circumstances, even a short circuit, will the amplifier produce more than 20mA. To determine the effect of a particular load, it can be plotted on the same graph. A 500Ω load is shown intersecting the output characteristic at point A on the graph. Consequently, the maximum output of this op-amp into this load is approximately 9 volts at 18mA.

For a split-supply op-amp, a similar characteristic would define the negative output capabilities of the op-amp, that is, when the op-amp is producing a negative voltage and sinking current into its output.

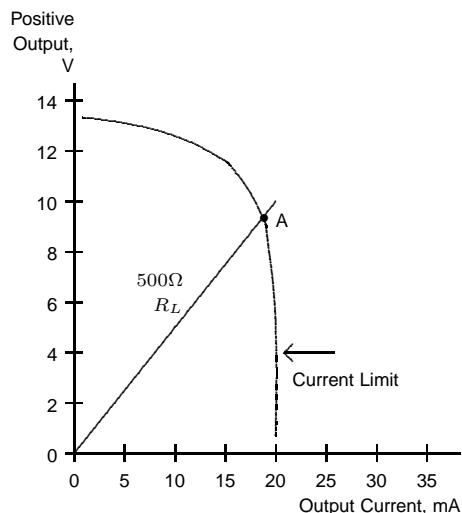


Figure 562: Positive Output Characteristic

Typical Values

As indicated in the table below, general purpose split supply op-amps are designed to operate with supplies of maximum ± 18 volts. The output swing is limited to something less than this, as listed in column 4 of the table.

The trend in op-amp applications has been to single supply operation at lower voltages. To be fully useful at a lower supply voltage, the op-amp output voltage should be able to swing as close as possible to the supply rail. The LMC660 is representative of this type of device, able to swing within 300mV of the supply rail.

In some cases, the maximum output current is limited by the power dissipation of the package and the effectiveness of the heatsink.

Type	Part Number	Supply Voltage Max	Peak Output Voltage	Output Current mA
General Purpose BJT	741	± 18	$V_{CC}-2$	10
Single Supply BJT	LM324	32	$V_{CC}-1.5$	10
General Purpose JFET	TL074	± 18	$V_{CC}-3$	10
General Purpose MOSFET	LMC660	16	$V_{CC}-0.3$	16
General Purpose BIFET	TLC074	± 18	$V_{CC}-1.5$	35
High speed current feedback	LM6181	± 18	$V_{CC}-3$	100
High power	LM675	± 30	$V_{CC}-5$	4000

When large current must be supplied from an op-amp, the usual approach is to split the system into a low-power op-amp driver (the brains) followed by a simple high-power driver stage (the muscle). Williams [98] shows some examples of this approach. It is possible to integrate an op-amp on the same chip with the high-power drivers (see the LM675 listed in the table), but the resultant circuit is likely to be more expensive than splitting the circuit into two parts. On the other hand, an integrated circuit containing both signal and driver circuitry is less likely to present stability and oscillation problems and requires less assembly labour.

21.7 Output Resistance

The output characteristic of figure 562 in section 21.6 defines the op-amp limits of the output voltage and current. However, even inside this region, the output voltage decreases as the output current increases. This may be characterised as the internal resistance of the output voltage source for the op-amp, and is known as r_o , the *output resistance*. To incorporate this effect, the op-amp model is modified with a resistance r_o in series with the output, as shown in figure 563.

This resistance is exactly equivalent to the internal resistance R_{int} that we discussed in section 2.20.

Typical open-loop values of output resistance range from 0.6Ω for the LM675 power op-amp, to 100Ω or so for small-signal op-amps. The exact value is not particularly relevant, since it is usually decreased substantially by the application of negative feedback.

A negative feedback system that is designed to control output voltage (as opposed to output current) will resist changes in output voltage that are caused by changes in output current. The system sees any change in output voltage, due to the voltage drop across the output resistance, as a disturbance that must be minimized. Providing that the output voltage and current fall within the limits of the device, the effect of the negative feedback system is to reduce the output resistance from r_{int} to r_o where:

$$r_o = \frac{r_{int}}{AB} \quad (965)$$

where

r_{int} is the open-loop value of the internal resistance

r_o is the effective output resistance with negative feedback

A is the open-loop voltage gain of the amplifier

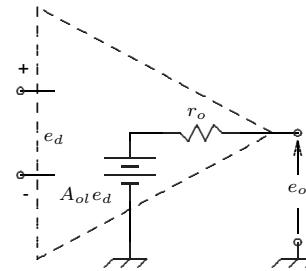


Figure 563: Output Resistance Model

B is the feedback factor of the system.

AB is the *loop gain* of the negative feedback system.

(When it is required to drive a constant current into a load, it is possible to *raise* the output impedance of the op-amp. However, in that case the output impedance is caused to be a function of some external resistance times the loop gain.)

There is one situation where the output resistance may be critically important. When the load is a capacitance C_L , the output resistance r_o and load capacitance C_L form an RC lowpass filter which introduces phase shift to the output signal. This phase shift may be sufficient to affect the stability of the amplifier. This situation and its cure are discussed more extensively in section 24.9.

Derivation: Negative Feedback and Output Resistance

To determine the effect of negative feedback on the output resistance, we'll use figure 564 and the *test voltage* method described in section 2.20. To simplify things, the input voltage e_i is reduced to zero.

The effective output resistance, with the feedback network connected, is the ratio of the test voltage e_T to the resultant current i_T .

$$r_o = \frac{e_T}{i_T} \quad (966)$$

Now we'll trace the effect of the test voltage.

If the output voltage is forced to be e_T , then the feedback voltage e_f is

$$e_f = Be_T \quad (967)$$

where B is the feedback ratio (see section 12.4). This feedback voltage becomes the error voltage into the op-amp. Since

$$v_o = A(e_i - e_f) \quad (968)$$

and the input voltage e_i is zero, then the internal supply of the op-amp goes to

$$\begin{aligned} v_o &= A(0 - e_f) \\ &= A(0 - (Be_T)) \\ &= -ABe_T \end{aligned} \quad (969)$$

In words, a voltage drive of e_T volts at the output causes the internal voltage supply of the op-amp to move in the opposite direction by an amount ABe_T .

Consequently, the total change in voltage across r_{int} is:

$$\begin{aligned} v_{int} &= e_T + AB e_T \\ &= (1 + AB)e_T \end{aligned} \quad (970)$$

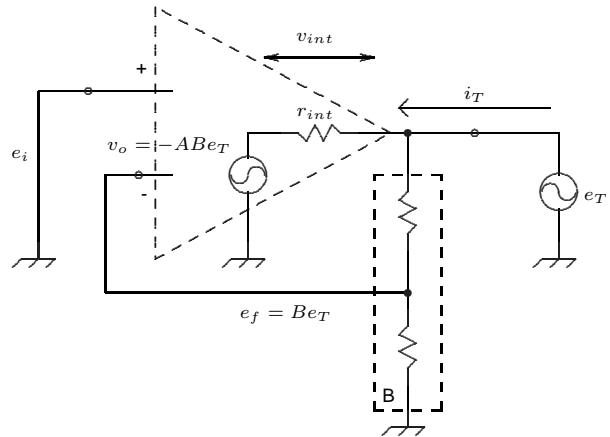


Figure 564: Output Resistance Analysis

The current i_T into the output terminal is equal to this voltage divided by the internal resistance:

$$\begin{aligned} i_T &= \frac{v_{int}}{r_{int}} \\ &= \frac{(1+AB)e_T}{r_{int}} \end{aligned} \quad (971)$$

Substituting for i_T from equation 971 into equation 966, we have

$$\begin{aligned} r_o &= \frac{e_T}{i_T} \\ &= \frac{e_T}{(1+AB)e_T} \\ &= \frac{r_{int}}{1+AB} \\ &\approx \frac{r_{int}}{AB} \end{aligned}$$

In words, the internal resistance r_{int} is reduced by an amount approximately equal to the loop gain AB .

At low frequencies where the open-loop gain A is large, the loop gain is also large, and the output resistance will be very small. As the frequency increases, the open-loop gain decreases, and the output resistance will rise, eventually approaching r_{int} .

21.8 Slew Rate

Definition

The *slew rate* of an op-amp is the maximum rate of change of the output when a large step voltage is applied to the input.

As illustrated in figure 565, there is an upper limit to the rate at which the output voltage can change. When requested to make an abrupt transition, the output *slews* to the new commanded voltage at the *slew rate* of the device.

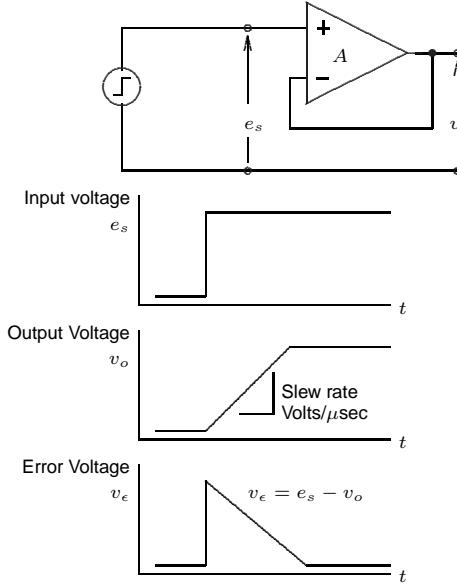


Figure 565: Slew Rate Limiting

Typical Values

Type	Part Number	Slew Rate Volts/ μ Sec
Low-cost BJT	LM324	0.25
General Purpose BJT	741	0.5
General Purpose JFET	TL074	13
General Purpose MOSFET	LMC660	1.1
High Speed BJT	LM6161	300
Current Feedback	LM6181	1000

Mechanism

The signal path of an op-amp usually includes a *frequency compensation capacitor* C_c which must be charged and discharged at the signal frequency. A current source drives this capacitor and must supply a peak current that is proportional to the rate of change of the signal:

$$i = C_c \frac{dV}{dt} \quad (972)$$

At low slew rates, the current requirement is modest and the current source charges the capacitor at the correct rate. However, at some maximum rate of signal change, the current source limits and the capacitor is charged in a linear fashion at the slew rate:

$$\frac{dV}{dt} = \frac{i_{limit}}{C_c} \quad (973)$$

(Notice that the i_{limit} current is an internal current, not the output current.) This is a non-linear phenomenon. During the slew period, the amplifier cannot respond to any other signals and the output signal is massively distorted compared to the input.

Increasing Slew Rate

Until very recently, the slew rate of most op-amps was quite limited. A fast op-amp required a custom design or an expensive device from a specialty manufacturer. In the last part of the previous century an extensive market developed for high-speed analog cable drivers and video amplifiers. The bandwidth of the standard NTSC video signal is in the order of 5MHz. An analog signal driver for a computer Cathode Ray Tube display must deal with signal bandwidth in the order of 100MHz. Fast op-amps arrived to meet this need.

As shown in the table of slew rates on page 656, amplifier slew rates range from 0.25 to 1000 volts/ μ Sec. As a result, it's possible to find an op-amp with a suitable slew rate for most applications. To guide your search, here are some simple rules of thumb:

- The simplest solution is to buy a faster op-amp. However, other parameters may suffer in the pursuit of speed: the noise, offset and power-consumption may be sacrificed in such an op-amp. Check the specifications carefully. If necessary, pair the high-speed device with a second device that has the necessary DC specifications, in a composite amplifier configuration.

Where possible, consider the *current feedback* amplifier configuration (section 35.3) rather than the classical voltage-driven op-amp. The current feedback amplifier is inherently much faster than an op-amp. However, in general, it has poorer DC characteristics such as offset voltage and bias current, so it may not be entirely suitable.

- Be aware that the minimum closed-loop gain establishes the severity of the frequency compensation. A $\times 1$ closed-loop-gain amplifier is the worst case, requires the most severe compensation, and will have the poorest slew rate.

If the minimum amplifier gain is somewhat larger, the size of the compensation capacitor can be reduced and the slew rate improved. For example, the LF155 and LF157 from National Semiconductor are two such variants, as shown in the following table.

Part Number	Minimum Closed Loop Gain	Slew Rate Volts/ μ Sec
LF155	$\times 1$	3
LF157	$\times 5$	40

The LF157 is a decompensated version of the LF155, so it cannot be used in applications where the closed-loop gain is less than 5. However, the slew rate is substantially increased.

- In theory (see [139], [100], [84]), a utility-grade JFET or MOSFET op-amp is likely to have a higher slew rate than a utility-grade BJT op-amp. However, slew rate depends on the overall design topology and many other factors, so this is not an entirely reliable guide.
- To remove the need for engineering the op-amp frequency response, most op-amps are equipped with *dominant pole compensation*. The open-loop frequency response is a single time-constant lowpass filter which rolls off at 20db/decade until the gain is below 0 db. With this frequency response, the op-amp is unlikely to oscillate but the slew rate is severely restricted. The op-amp is said to be *internally compensated*.

An op-amp which has access to the internal circuitry for custom compensation can be tailored to a particular requirement, thereby increasing the slew rate. However, this substantially increases the engineering effort. The National LM118 device is one such amplifier that allows such tinkering, where *feedforward compensation* can approximately double the slew rate [211].

Slew Rate and Sine Wave Signals

When the amplifier is processing a sine wave signal, the maximum rate of change occurs during the zero-crossing of the signal.

Suppose the signal is given by

$$e = E \sin 2\pi f t \quad (974)$$

where E is the peak value of the signal in volts and f the frequency in Hz.

Then the slew rate of this signal is

$$\begin{aligned} SR &= \frac{de}{dt} \\ &= \frac{d}{dt}(E \sin 2\pi f t) \\ &= 2\pi f E \cos 2\pi f t \end{aligned} \quad (975)$$

where SR is the slew rate in volts/second.

The maximum value of the slew rate occurs when the cos term is unity, in which case

$$SR_{max} = 2\pi f E \quad (976)$$

The product fE must not exceed the slew rate or the sine wave signal will be distorted. By rearranging equation 976, we see that the maximum undistorted output voltage is limited by the slew rate:

$$E = \frac{SR_{max}}{2\pi f} \quad (977)$$

Example

The slew rate of the 741 op-amp is 0.5 volts/ μ Sec. It is operated from a power supply of ± 15 volts. What is the maximum frequency for which it can produce its full output signal?

Solution

With ± 15 volt supplies, the 741 can produce a maximum output signal of around ± 13 volts, or 13 volts peak. Rearranging equation 977 to solve for f , we have

$$\begin{aligned} f &= \frac{SR_{max}}{2\pi E} \\ &= \frac{0.5 \times 10^6}{2\pi 13} \\ &= 6.12 \times 10^3 \text{ Hz} \end{aligned}$$

Notice that this is *well* below the maximum audio frequency of 20kHz, and so a 741 type audio amplifier cannot be expected to produce a full undistorted output voltage over the entire audio spectrum.

Large Signal Frequency Response

The maximum amplitude of a sine wave is slew-rate limited at high frequencies, and one way of presenting that information is in the form of a plot of large signal frequency response. An example is shown in figure 566 (adapted from the datasheet for the LM324 op-amp).

From this graph, slew rate limiting first appears at about 5KHz where the peak voltage is about $13.5/2 = 6.75$ volts peak. Above that frequency, the maximum possible amplitude decreases. Plugging these values into equation 977 and solving for the slew rate, we obtain

$$\begin{aligned} SR_{max} &= 2\pi fE \\ &= 2\pi \times 5000 \times 6.75 \\ &= 0.25 \text{ volts}/\mu\text{sec} \end{aligned}$$

References

The definitions for non-ideal op-amp parameters are taken for the most part from [209].

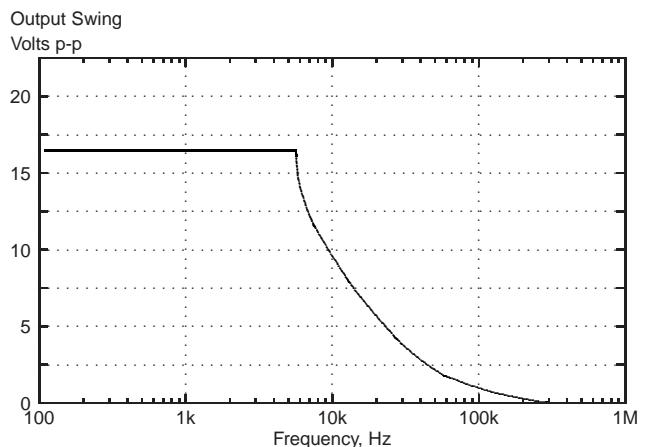


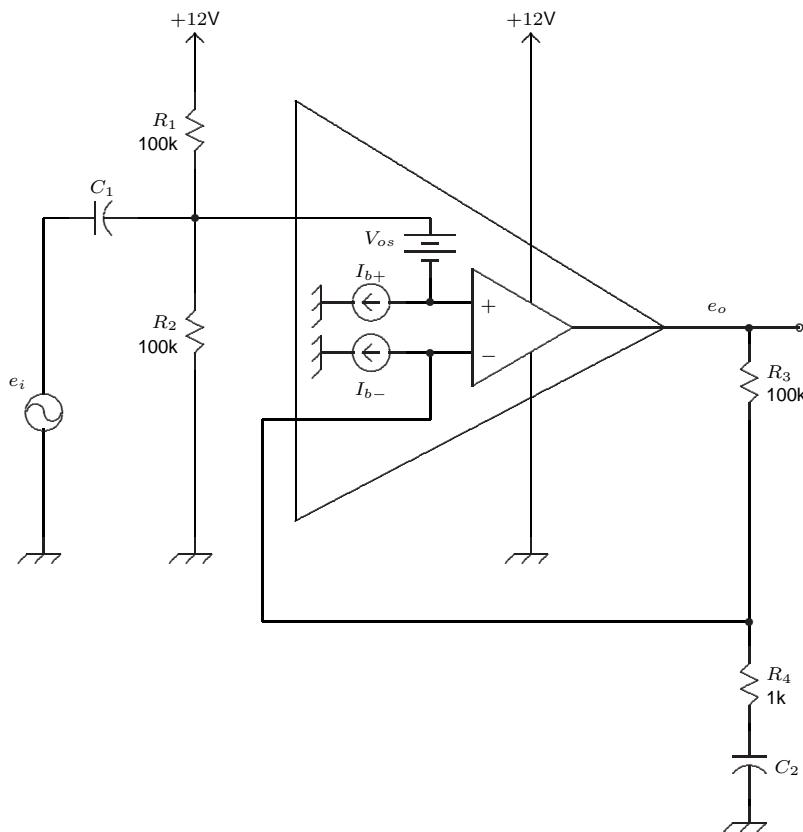
Figure 566: Large Signal Frequency Response

21.9 Noise and Bandwidth

There are two additional real-world imperfections of operational amplifiers: *noise* and *bandwidth*. Each of these subjects is shown as a separate section: noise in section 22, beginning on page 667, and bandwidth in section 23 beginning on page 691.

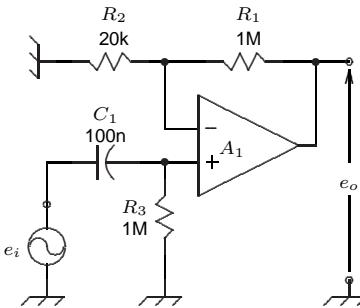
21.10 Exercises

- For the negative feedback operational amplifier circuit shown in the figure, $I_{b+} = I_{b-} = 100\text{nA}$ and $V_{os} = 3\text{mV}$.

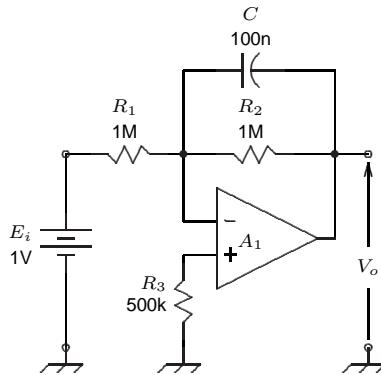


- What is the DC output voltage E_o , ignoring e_i , to the nearest millivolt?
 - What is the AC voltage gain e_o/e_i of this circuit, assuming the capacitors can be treated as AC short circuits?
 - Calculate values for C_1 and C_2 such that the low frequency response extends to 100Hz.
2. Referring to exercise 13 on page 391, answer the following:
If the offset voltage of the amplifier were to change by 4 millivolts, what effect would this have on the temperature of the oven?
- A certain op-amp has a slew rate of 0.5 volts/ μSec . At what frequency does slew-rate-limiting begin to affect a sine wave of 10 volts peak-peak?

4. For the amplifier shown below, the bias currents are 100nA into the op-amp terminals. The offset voltage is 3mV.

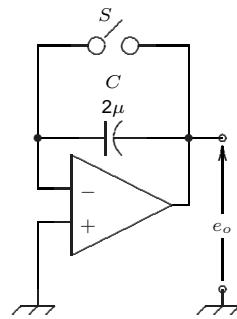


- (a) Calculate the average output voltage due to the bias currents.
 - (b) Calculate the average output voltage due to the offset voltage.
 - (c) How would you modify this circuit to reduce the output voltage created by the bias currents?
5. Consider the compensated integrator circuit shown below.

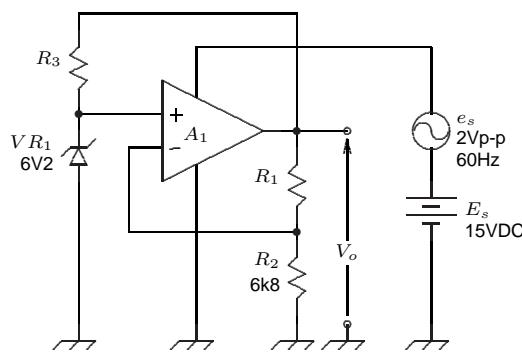


Compute the steady-state output voltage V_o

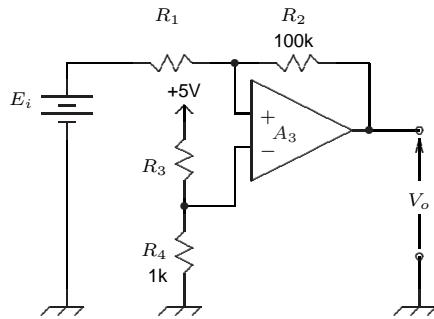
- (a) if the op-amp is ideal (no offset voltages, bias current or offset voltage).
 - (b) if the offset voltage of the op-amp is 100mV
 - (c) if the input bias currents of the op-amp are $I_b^+ = 0.9\mu\text{A}$ and $I_b^- = 1.1\mu\text{A}$. Bias currents flow into the op-amp terminals.
 - (d) if the previous two conditions apply simultaneously.
6. The operational amplifier in the circuit shown below has a bias current of $100\mu\text{A}$ and an offset voltage of 3mV. The input stage of the op-amp uses PNP transistors.



- (a) Calculate the output voltage when the switch is closed.
- (b) Calculate and plot the output voltage in the first 10 seconds after the switch is opened.
7. The circuit shown below is to supply a stable voltage of 10 volts with a low internal resistance.



- (a) Calculate a suitable value for R_1 .
- (b) If the zener current is 7.5mA, calculate the value of R_3 .
- (c) The op-amp specifications are:
- $V_{os}=3\text{mV}$
 - $I_{bias} = 1\mu\text{A}$
 - $\text{CMRR} = 120\text{db}$
 - $\text{PSRR} = 60\text{db}$
 - Slew Rate = $0.5\text{V}/\mu\text{Sec}$
 - Gain Bandwidth Product = 10^6Hz
- (d) The zener internal impedance is negligible. Estimate the DC and AC errors in the output.
8. In the amplifier shown below:
- For $E_i = 0.1\text{V}$, $E_o = 4.5\text{V}$
 - For $E_i = 0.2\text{V}$, $E_o = 0.5\text{V}$



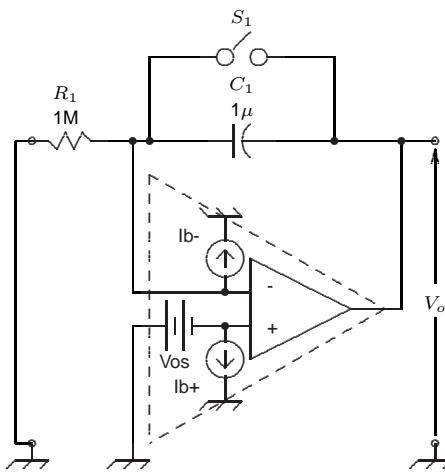
- (a) Calculate R_1 and R_3 .
 (b) If the temperature coefficient of offset voltage is

$$\frac{\Delta V_{os}}{\Delta T} = \frac{10\mu\text{V}}{\text{°C}}$$

what will be the effect on V_o of a temperature change ΔT of 50°C ?

9. In the amplifier shown below (with its offset voltage and bias current generators):

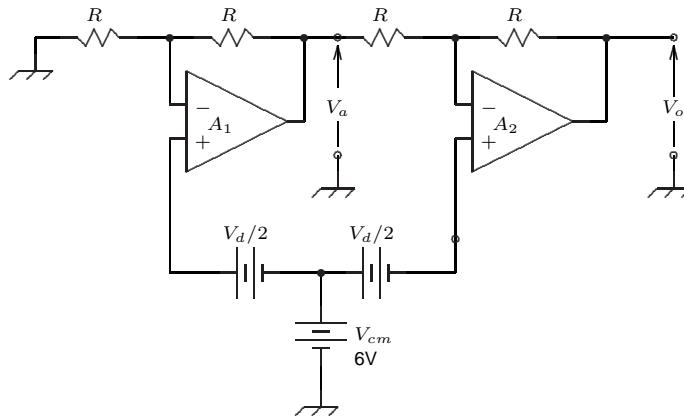
- $V_{os} = 2\text{mV}$
- $I_B^+ = I_B^- = 17\text{nA}$



(a) What is the value of V_o when the switch S_1 is closed?

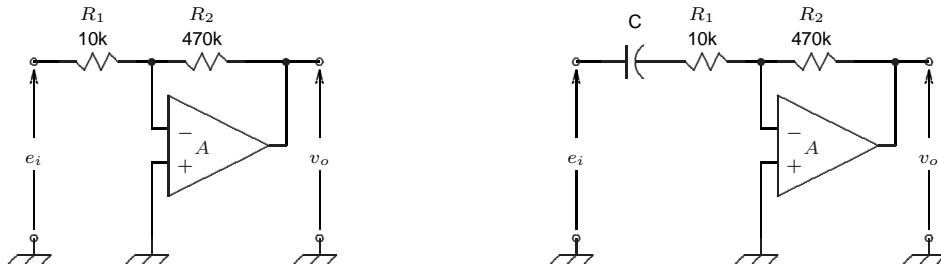
(b) What is the value of V_o when the switch S_1 is open?

10. A differential amplifier circuit is shown below.

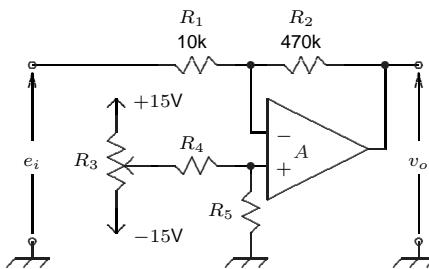


- (a) Assume that the op-amps are ideal except that they have a CMRR of 60db. What is the output V_a of A_1 , to the nearest millivolt, due to this common-mode voltage?
- (b) To the nearest millivolt, what is the output signal V_o due to the common-mode voltage?
- (c) Now determine the effect of V_d : If V_d is 3 volts, what is the output voltage V_o ?

11. The operational amplifier shown in these figures may be considered to be ideal except for an offset voltage of 5mV.



- (a) For the circuit on the left, what is the DC output voltage due to the offset voltage?
 (b) For the circuit on the right, choose a suitable value for capacitor C if the amplifier is to operate down to a low frequency of 100Hz.
 (c) What is the DC output voltage of this circuit due to the offset voltage?
 (d) What conclusion can you draw from a comparison of these two circuits?
12. The operational amplifier shown in this figures may be considered to be ideal except for an offset voltage of 10mV.

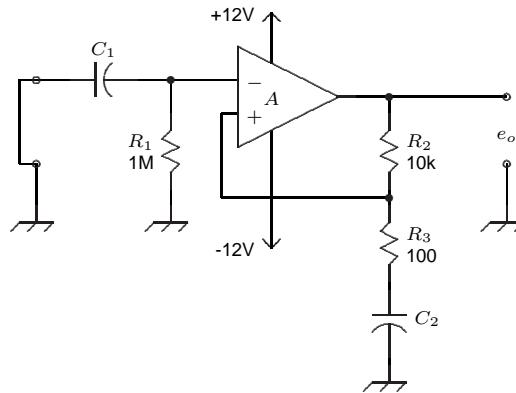


The network R_3 , R_4 and R_5 must be designed to cancel the maximum possible offset voltage, which can be of either polarity. If R_3 is chosen to be $10k\Omega$, calculate R_4 and R_5 .

13. For the circuit shown below, the following applies for the operational amplifier:

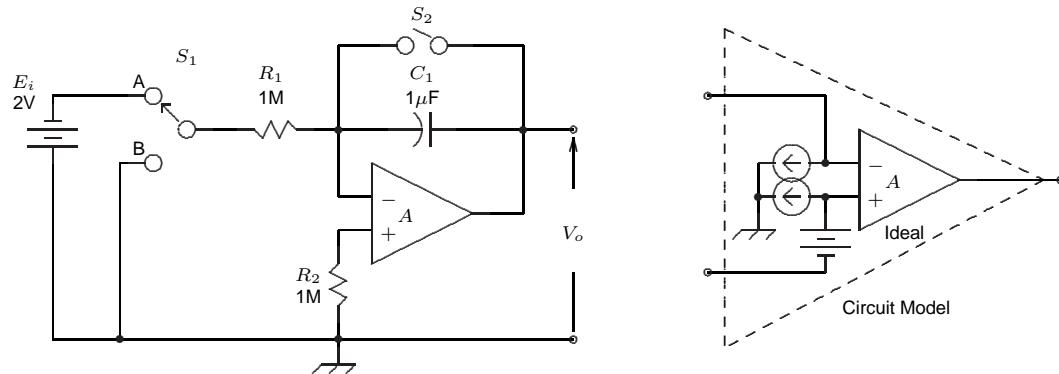
Offset Voltage	V_{os}	1mV
Bias Current	I_b	100nA

The amplifier has NPN BJT transistors at its input. Other than the specifications given above, the op-amp may be considered to be ideal. The capacitors may be considered open short circuits at zero frequency (DC).



What is the DC value of the output voltage?

14. The circuit diagram of an integrator, along with the equivalent circuit of the op-amp, is shown below.



The operational amplifier has a bias current I_b of 200nA and offset voltage V_{os} of 3mV. In other respects it may be considered to be ideal.

The integrator is cycled through the following series of events:

- At time T0, the integrator output is reset to an initial condition by closing switch S2. (The position of switch S1 is unimportant.)
 - At time T1, the integrator is allowed to integrate the input voltage by putting switch S1 into the input voltage position A and opening switch S2. This state continues for 4 seconds.
 - At time T2, the integrator is put into the 'hold' state by moving switch S1 to the B position. Switch S2 remains open. It remains in this state for 10 minutes, until time T3.
- To begin with, assume that the op-amp is ideal and determine the output voltage at the end of each state.
 - Now, as listed below, determine the overall operation of the integrator including the effects of bias current and offset voltage. In each state, the starting value of the output voltage is the ending value of the previous state.

- i. *Reset* state, T0 to T1
 - A. Draw an equivalent circuit for the integrator, including the bias current and offset voltage.
 - B. Calculate the output voltage at the end of this state.
 - ii. *Integrate* state, T1 to T2
 - A. Draw an equivalent circuit for the integrator, including the bias current and offset voltage.
 - B. Calculate the output voltage at the end of this state.
 - iii. *Hold* state, T2 to T3
 - A. Draw an equivalent circuit for the integrator, including the bias current and offset voltage.
 - B. Calculate the output voltage at the end of this state.
- (c) What are the errors, compared to the operation of an ideal op-amp, that are caused by I_b and V_{os} in the operation of this circuit?
- (d) Can the effect of either I_b or V_{os} be ignored in this circuit? Explain in detail.

22 Operational Amplifier Noise

Introduction

Noise is the unwanted guest of electronic circuitry. It is an undesired signal that adds itself to the desired signal. At a minimum, it worsens the fidelity of the signal. In severe cases, it may make the signal unreadable.

Two types of noise will concern us: *interference* and *device* noise. Interference results when an unwanted signal from some other source (electronic equipment, the sun, radio stations) hitch-hikes on the desired signal. Interference signals may be random in nature, but are more likely to be periodic: the clock noise of a digital circuit is one common example. We will deal with that problem in section 34.

Device noise results from the intrinsic properties of the electronic device and the physics of electron flow, and it is random in nature rather than periodic. This type of noise concerns us here.

Amplitude of a Noise Signal

The waveform of a random noise signal, as it might be observed on an oscilloscope, is shown in figure 567. This waveform was generated by creating a random number between ± 0.5 , and linking the values. It could equally well be the result of sampling a noise waveform 200 times and plotting the result.

Unlike a sine wave signal for example, a noise signal cannot be characterised by a simple predictive equation. The amplitude can and does assume any value, but certain amplitudes are more likely than others. In fact, the frequency of occurrence of the amplitudes is a gaussian distribution, as shown in figure 568, which is the *probability density function* (PDF). This PDF indicates that the probability of an amplitude being exceeded decreases with amplitude (large amplitudes are less likely).

To clarify the physical meaning of this function, consider that we have a string of random numbers, each of which is an integer between the values 0 and 9. After a period of time, when we have accumulated a significant sample, we plot the number of occurrences of each integer against the value of the integer. The X axis of such a plot consists of the possible values of the integers, 0 through 9. The Y axis shows the number of times each integer occurred in our sample. This plot is the discrete probability distribution function, because it deals with discrete values.

For a continuous signal, the amplitude is specified as between two magnitudes. For example, we might divide up the amplitude range into 10 equally spaced intervals. The waveform is sampled at regular intervals (under control of a clock, for example). Then each sample is categorized by its amplitude interval at the time of

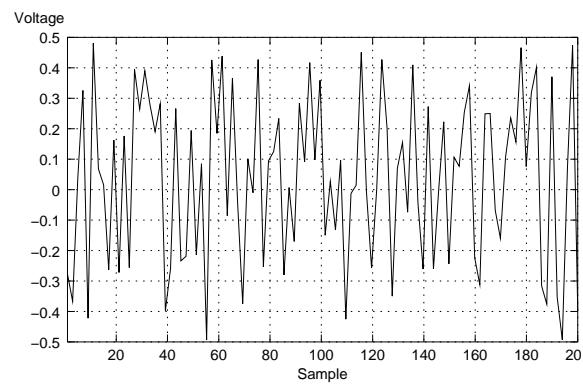


Figure 567: Noise Waveform

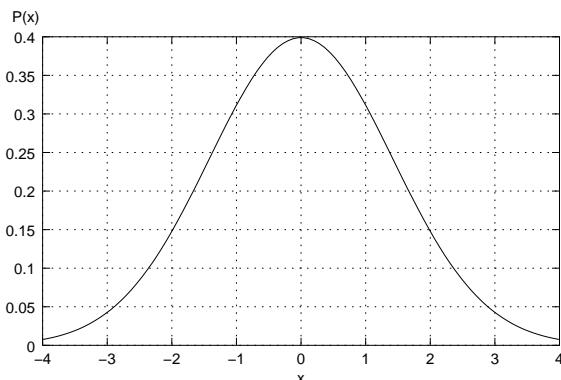


Figure 568: Probability Density Function

sampling, and then the samples are plotted as for the discrete PDF.

As a design criterion in dealing with a random noise signal, we may wish to know the probable maximum level of the signal in order to avoid clipping in the signal processing circuitry. It turns out [212] that the signal will exceed 4 times the rms value less than 0.01 percent of the time. For example, certain acoustic signals such as the snare drum resemble random noise. If an amplifier for this signal were designed to clip at 4 times the rms level of the signal, then it could be expected to clip less than 0.01 percent of the time.

Measuring noise amplitude

Since the peak amplitude of a noise waveform is not predictable, the general approach is to measure its effective or heating value in a resistor, that is, the RMS value of the waveform. For a given noise source and measurement bandwidth, this value is constant. If you are fortunate enough to have access to a 'true rms' voltmeter, then you can measure the RMS value of the noise waveform directly.

Otherwise, using a conventional AC voltmeter, you must apply a correction factor. On the 'AC' setting most voltmeters respond to the rectified average of an AC waveform. The meter assumes that the waveform is a sine wave and multiplies the actual rectified average value by 1.11 so that the reading is the RMS value for a sine wave (section 16 on page 496). This correction factor is different for different waveforms, and so a rectified average responding meter will not read the correct RMS value of a noise waveform. As a consequence, you must multiply the reading of a rectified-average meter by a factor of 1.13 to get the RMS value.

Alternatively, you can use an oscilloscope and the *tangential* method [213]. Connect the noise waveform to both the A and B channels of the oscilloscope. Set both channels to the same vertical gain and the timebase so that the noise blurs each trace into a vertical range of values.

Now, adjust these blurred traces vertically so that there is a clear gap between them, that is, you can still see the two individual traces. Then move the vertical position of the two traces toward each other until the gap between them disappears and they form one uniform band. Remove the noise signals. The separation in volts between the two traces is twice the RMS value. According to the reference, this method is capable of about 10% accuracy.

Signal to Noise Ratio

Noise is always present to some degree in an electronic signal. The key issue, which is application dependent, is the allowable noise with respect to the signal, the so-called *signal-noise ratio* (SNR). The SNR is the ratio of the signal power to the noise power, expressed in decibels:

$$SNR = 10 \log_{10} \frac{P_s}{P_n} \text{ decibels} \quad (978)$$

where P_s and P_n are the signal and noise powers in watts, respectively.

Now, the power in a resistor R is equal to

$$P_r = \frac{e^2}{R} \text{ watts} \quad (979)$$

where e is the RMS value of the voltage.

Consequently, if we measure the RMS value of the signal voltage e_s and the RMS value of the noise voltage e_n across the same value of resistor R , we can rewrite equation 978 as:

$$\begin{aligned} SNR &= 10 \log_{10} \frac{e_s^2/R}{e_n^2/R} \\ &= 10 \log_{10} \frac{e_s^2}{e_n^2} \end{aligned}$$

$$= 20 \log_{10} \frac{e_s}{e_n} \quad (980)$$

In a high-quality audio application, we might require the signal-noise ratio to be 80db. In a photodiode amplifier that must recover a digital signal in the presence of noise, a signal-noise ratio of 6db might be satisfactory. Communication links to space probes routinely recover a signal that is *below* the noise level, that is, the SNR is negative.

Adding Noise Signals

The RMS or *effective* value of a voltage is given by the expression:

$$e_{RMS} = \sqrt{\frac{1}{T} \int_0^T e^2(t) dt} \quad (981)$$

Suppose two voltage sources $e_1(t)$ and $e_2(t)$ (which we will nickname e_1 and e_2) are added together. Then

$$\begin{aligned} e_{RMS} &= \sqrt{\frac{1}{T} \int_0^T [e_1 + e_2]^2 dt} \\ &= \sqrt{\frac{1}{T} \int_0^T [e_1^2 + 2e_1e_2 + e_2^2] dt} \end{aligned} \quad (982)$$

Now, for many sources all three of these terms are significant. For example, if the sources e_1 and e_2 are a DC source of value E , then:

$$\begin{aligned} e_{RMS} &= \sqrt{\frac{1}{T} \int_0^T [E^2 + 2E^2 + E^2] dt} \\ &= \sqrt{\frac{1}{T} \int_0^T (2E)^2 dt} \end{aligned} \quad (983)$$

In other words, the RMS value can be obtained simply by adding the two DC voltages.

For noise signals, the situation is different. For reasons explained in a moment, the crossproduct term in equation 982 is zero. Then equation 982 may be rewritten as:

$$\begin{aligned} e_{RMS} &= \sqrt{\frac{1}{T} \int_0^T [e_1^2 + e_2^2] dt} \\ &= \sqrt{\frac{1}{T} \int_0^T e_1^2 dt + \frac{1}{T} \int_0^T e_2^2 dt} \\ &= \sqrt{e_{rms1}^2 + e_{rms2}^2} \end{aligned} \quad (984)$$

where e_{rms1} and e_{rms2} are the RMS values of the two noise signals.

This is an important result: *Noise signals add as a pythagorean sum, that is, the square root of the sum of the squares of the individual signals*, where all the amplitudes are in RMS volts. It's not proven here, but this result extends to multiple sources beyond 2.

Small Sources Don't Count

As a consequence of this *Pythagorean addition* process, we may be able to eliminate the smaller of two sources in the addition. For example, if the smaller source is 10% of the larger, then it contributes only 1% to the total. This approximation is very useful in noise calculations.

Why the Crossproduct Term is Zero

Equation 982 may be rewritten as:

$$e_{RMS} = \sqrt{\frac{1}{T} \int_0^T e_1^2 dt + \frac{1}{T} \int_0^T 2e_1 e_2 dt + \frac{1}{T} \int_0^T e_2^2 dt} \quad (985)$$

The middle term, an integral containing the product of the two noise voltages, is what is called the *correlation* function of e_1 and e_2 .

$$K_{correlation} = \frac{1}{T} \int_0^T 2e_1 e_2 dt \quad (986)$$

The expression generates a number which is related to the similarity of the two waveforms. For example, if e_1 and e_2 are identical waveforms, the result is a large number. In this particular case, the noise waveforms are totally uncorrelated, that is, they have no similarity at all. Consequently, over a long averaging period, the correlation will tend to zero.

References

The primary references for this section were Motchenbacher & Connelley [214] and Ott [215]. Franco [139] has useful applications information, including a section on noise in photodiode amplifiers. Dostál [85], Clayton [216] and Horowitz-Hill [100] filled in some of the details. Kennedy [141] and Texas Instruments Application Note SLVA043A [217] show example op-amp noise calculations.

Other references are mentioned in the text.

22.1 Resistor Thermal Noise

As a consequence of the thermal motion in a resistor¹⁹⁸, a small noise voltage appears at the terminals of the resistor. When the noise is measured by a true-rms voltmeter of bandwidth B , the noise voltage is found to be [218]:

$$e_n = \sqrt{4KTBR} \text{ volts} \quad (987)$$

where

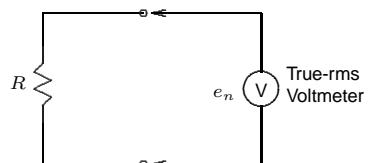
K is Boltzmann's Constant, 1.38×10^{-23} joules/kelvin

T is the absolute temperature, kelvins (${}^\circ\text{C} + 273$)

B is the measurement bandwidth, Hz.

R is the resistance, ohms

The amount e_n is the value of voltage that would appear on the true-RMS voltmeter of figure 569.



¹⁹⁸Depending on the construction of a resistor, there may be other sources of noise in addition to thermal agitation. However, in modern devices at room temperature, the thermal noise predominates.

It follows that the noise current into a short circuit is

$$i_n = \sqrt{\frac{4KTB}{R}} \quad (988)$$

Equation 987 implies that an infinite-bandwidth system will have an infinite noise level [218]. In practice, systems are always band-limited, by the stray inductance and capacitance if by no other means. An ideal noise signal, with infinite bandwidth, is a useful theoretical concept but unrealistic in practice.

A noise-generating resistor may be represented by the Thevenin or Norton model, whichever is more convenient, as shown in figure 570.

Example

What is the noise voltage generated by a $10\text{k}\Omega$ resistor at room temperature, over the audio spectrum (20Hz to 20kHz)?

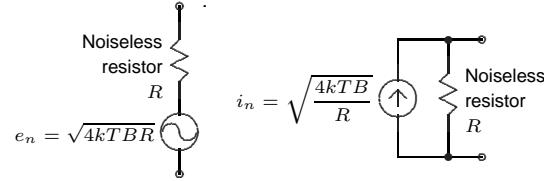


Figure 570: Resistor Noise Equivalents

Solution

We'll take room temperature as 23°C . Then $T = 296\text{k}$.

$$\begin{aligned} e_n &= \sqrt{4KTB R} \\ &= \sqrt{4 \times (1.38 \times 10^{-23}) \times 296 \times (20000 - 20) \times (10 \times 10^3)} \\ &= 1.80 \times 10^{-6} \text{ V} \end{aligned}$$

This solution assumes a *brick wall* spectrum characteristic between 20 and 20kHz, that is, the skirts have infinite slope at these two frequencies. In practice, this will not be the case, and the spectrum will roll off at some progressive rate like 20 or 40db per decade.

22.2 Noise Spectrum

A sine wave is a single frequency spectrum. Other periodic waveforms, such as a square wave, generate a harmonic spectrum, with frequencies at various multiples of the fundamental. Noise generates a *continuous* spectrum, that is, all frequencies are present.

Although a noise waveform generates all frequencies, it takes some time to do so. The noise power is spread over a range of frequencies. If the measurement is restricted to a small range of frequencies, the measured power will be less than if the entire range of frequencies is included. That is, the measured noise power is proportional to the measurement bandwidth.

This provides us with a powerful tool in improving the signal-noise ratio. If the information in the signal is restricted to a small range of frequencies, it can be bandpass filtered to this frequency range to restrict the noise density.

We can rewrite equation 987 as follows:

$$\frac{e_n}{\sqrt{B}} = \sqrt{4KTR} \text{ volts}/\sqrt{\text{Hz}} \quad (989)$$

This characterises the *voltage noise density*, which we will refer to as e_{nd} . In this case, it is a constant, $\sqrt{4KTR}$. In other cases, such as the noise generated in an operational amplifier, it may vary with frequency.

There is a similar measurement of *current noise density* i_{nd} which is measured in amps/ \sqrt{Hz} . (Be aware that data sheets often use the symbols e_n and i_n for noise density. However, the units of measurement make it clear that these are noise density values.)

If the voltage noise density is constant over the measurement bandwidth B , the value of the voltage noise over that bandwidth may be obtained by multiplication with \sqrt{B} . In formula form:

$$e_n = e_{nd}\sqrt{B} \text{ volts (rms)} \quad (990)$$

Similarly,

$$i_n = i_{nd}\sqrt{B} \text{ amps (rms)} \quad (991)$$

White Noise

When the noise spectrum value e_{nd} or i_{nd} is constant with frequency (described as *flat*), it is called *white* noise by analogy with the visible spectrum.. Measured with a constant-bandwidth filter, the noise power is everywhere the same. That is, the noise power is the same in the 10-20Hz band as it is in the 19980-19990 Hz band. To a human ear, white noise sounds like a hiss with a high-frequency edge.

When the noise density is not constant over the bandwidth, then a form of integration must be done over each separate region.

Pink Noise, 1/f Noise

An optical spectrum that has a greater component at low frequencies will appear red. A pink noise spectrum (also known as a 1/f spectrum) similarly emphasises lower frequencies. When plotted on a log-log scale, the voltage spectrum declines at a rate of 3db/octave, causing the power spectrum to decrease at 6db/octave. As a result, the noise power is constant when measured with a *constant percentage bandwidth* filter. Constant percentage bandwidth filters span a range of f to $n.f$. When n is 2, the filters are known as *octave* filters. A set of octave bandpass filters to cover the audio band would have the cutoff frequencies as shown in figure 571.

Pink noise sounds familiar to the human ear, because our biological spectrum analyser is a constant percentage device. Pink noise resembles the sound of rushing water from a tap or the inter-station hiss on the AM radio band. Pink noise is widely used as a test signal in testing loudspeakers and audio sound reinforcement systems.

In solid state devices, pink noise is referred to as 1/f noise or *flicker* noise. It tends to predominate at low frequencies.

f_{min}	f_{max}
31.25	62.5
62.5	125
125	250
250	500
500	1k
1k	2k
2k	4k
4k	8k
8k	16k

Figure 571: Octave Filter Band Edges

Popcorn Noise

Popcorn noise is a type of electrical noise in which bursts of square steps are added to the normal thermal noise at random times. Popcorn noise occurs rarely these days, but unfortunately it's not at 0%, not even with the cleanest processing and the best manufacturers. [113].

The width of the noise pulses varies from microseconds to seconds. The repetition rate varies from several hundred pulses per second to less than one pulse per minute. Typically, the amplitude (is fixed and) from 2 to 100 times the thermal noise. [215].

This type of noise was a serious problem in the first operational amplifiers. With improvements in integrated circuit technology, it has since become relatively insignificant. An op-amp with significant levels of popcorn noise may be regarded as defective.

Excess Noise

The theoretical minimum noise voltage in a resistor is given by equation 989. In practice, so-called *metal film* resistors (a modern resistor technology) exhibit a noise output equal to the theoretical minimum. However, other types of resistors such as the older technology *carbon composition* resistor, exhibit a larger noise than predicted by equation 989. This noise is known as the *excess noise*. It is proportional to the DC current through the resistor and, like pink noise, exhibits a 1/f spectrum (ie, is more prominent at low frequencies).

22.3 Effective Noise Bandwidth

The concept of *effective noise bandwidth* is illustrated in figure 572.

Consider that a noise source of infinite bandwidth is simultaneously applied to a brick wall filter of bandwidth f_o and a first-order lowpass filter of cutoff frequency f_o . The first order lowpass will pass more energy because the brick wall filter has no output above the cutoff frequency. The first order lowpass has a decreasing but non-zero output above the cutoff frequency.

To equalize the two outputs, we would have to make the bandwidth of the brick wall filter some amount greater than the cutoff frequency of the lowpass. It turns out as shown in [219], [214], [215] that this amount is a factor of $\pi/2$, or 1.57. As illustrated in figure 572, the two filters pass the same noise energy when

$$B = \frac{\pi}{2} f_o \quad (992)$$

$$= 1.57 f_o \quad (993)$$

where f_o is the 3db cutoff frequency of the lowpass filter and B is the bandwidth of the brick wall filter, both in Hz.

(In figure 572 these two filters are plotted on a log-log characteristic, decibels vs log frequency. If the two filters are plotted with linear scales, then the areas under the two filters are equal when $B = \frac{\pi}{2} f_o$.)

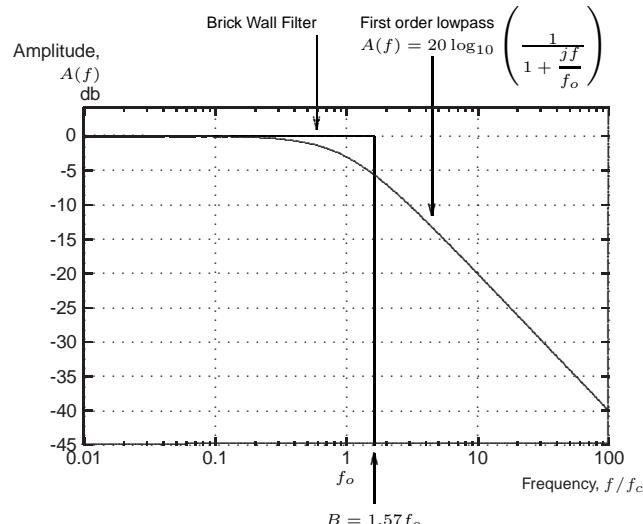


Figure 572: Noise Bandwidth

Similarly, as illustrated in figure 573, when a brick wall filter must pass the same energy as a bandpass filter, the brick wall bandwidth B is equal to 1.57 times the bandwidth $\Delta f = f_u - f_l$, where f_u and f_l are the upper and lower 3db frequencies of the bandpass filter. (Caveats: the filter cutoff frequencies must be widely separated, and the skirts must roll off at 20db/decade.) For example, this case would apply to the noise bandwidth of an amplifier with a bandwidth Δf .

As the order of a filter increases from first to second and so on, it approaches the shape of the brick wall filter. As a result, the difference factor between them becomes smaller. As tabulated by Ott [215]:

Filter Order	Rate of Rolloff db/decade	B/f_o
1	20	1.57
2	40	1.22
3	60	1.15
4	80	1.13

Example

A certain audio amplifier has a frequency response characteristic that ranges at the 3db points between 20Hz and 20,000Hz. The skirts slope at -20db/decade. What is the noise bandwidth?

Solution

$$\begin{aligned}\Delta f &= f_u - f_l \\ &= 20000 - 20 \\ &= 19980 \\ B &= 1.57 \cdot \Delta f \\ &= 1.57 \times 19980 \\ &= 31368 \text{ Hz}\end{aligned}$$

This is the value of noise bandwidth B that would be used in noise calculations for the amplifier.

22.4 Op-amp Noise

The sources of noise in an operational amplifier may be modelled as two noise current sources and a noise voltage source, as shown in the circuit model of figure 574. All of these sources are assumed to be uncorrelated, so voltages and currents add as an RMS sum, (equation 985 on page 670).

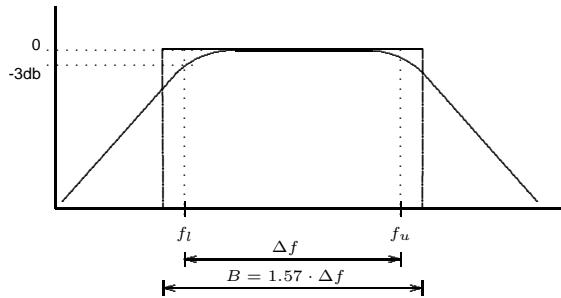


Figure 573: Noise Bandwidth of a Bandpass Filter

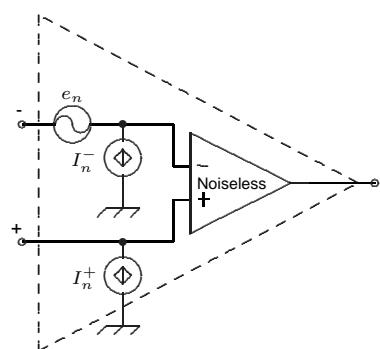


Figure 574: Op-Amp Noise Model

When the noise density is constant over the frequency band of interest, the effective or RMS value of an op-amp noise voltage or current is obtained from the product of the noise density and the square root of the effective bandwidth.

$$e_n = e_{nd} \sqrt{B} \quad (994)$$

$$i_n = i_{nd} \sqrt{B} \quad (995)$$

where the variables are

- e_n noise voltage, RMS
- e_{nd} noise voltage density, volts/ $\sqrt{\text{Hz}}$
- i_n noise current, RMS
- i_{nd} noise current density, amps/ $\sqrt{\text{Hz}}$
- B system bandwidth, Hz

Example

A certain op-amp has a spectral noise density of $5\text{nV}/\sqrt{\text{Hz}}$. What is the effective noise voltage e_n when this amplifier is used in an audio application?

Solution

Audio application implies a 3db bandwidth of 20 to 20,000 Hz. Then from equation 993 the noise bandwidth is:

$$\begin{aligned} B &= \frac{\pi}{2}(f_2 - f_1) \\ &= \frac{3.14}{2}(20000 - 20) \\ &= 31368 \text{ Hz} \end{aligned}$$

The noise voltage is given by equation 995:

$$\begin{aligned} e_n &= e_{nd} \sqrt{B} \\ &= (5 \times 10^{-9}) \sqrt{31368} \\ &= 885 \times 10^{-9} \text{ V} \end{aligned}$$

Typical Noise Specifications

Typical values for noise voltage and current are shown in the following table:

Type	Part Number	Noise Voltage $\text{nV}/\sqrt{\text{Hz}}$	Noise Current $\text{pA}/\sqrt{\text{Hz}}$
General Purpose BJT	741	22	1
General Purpose JFET	TL074	25	0.01
General Purpose MOSFET	LMC660	22	0.0002
Precision BJT	LM627	3.1	1
Low Noise Audio	NE5532	5	0.7

In general, op-amps based on junction-transistor (BJT) design have lower voltage noise and higher current

noise. JFET and MOSFET amplifiers tend to have larger voltage noise and smaller current noise. Consequently, if the source is high impedance or if the feedback resistors need to be large values, then a JFET and MOSFET amplifier is the preferred type. If the resistances are small, then a BJT device is probably superior.

In the case of certain devices, the noise specifications are not given. This is a warning that the manufacturer does not regard this as a suitable device for a low-noise application. The noise specification is relatively expensive to perform because it must take place long enough to get a reliable average value. Consequently, it's not performed or specified for low-cost devices. Low noise costs money. (For example, the noise voltage and current are not specified on the 741 datasheet. The noise voltage and current values are from [216]) This doesn't mean that the manufacturer is trying to hide the noise values – they may be quite satisfactory. But they don't measure or guarantee them.

For example, it turns out that the transistors 2N4401 and 2N4403 have quite excellent noise characteristics, even though they are identified as general purpose switching devices. As a result, some audio designers have incorporated them into their preamp designs – with the numbers erased to prevent the designs from being reverse engineered.

A Typical Noise Characteristic

The NE5534 is often used as a low-noise operational amplifier in audio systems. The voltage noise spectrum (re-drawn from the data sheet and approximated by straight line segments) is shown in figure 575.

It would be most convenient if the IC manufacturer would provide a single number figure for the noise voltage. However, to do that they would need to know the bandwidth of the application. This particular op-amp could be used in audio applications with a bandwidth of 20-20kHz, an instrumentation application with a bandwidth of 0.1-100Hz, or some other application. As a result, they usually supply a curve. It's up to the application engineer to specify the bandwidth and thereby generate the noise voltage figure.

The noise density is not constant over the entire spectrum. Above 200Hz, it has a *white* characteristic (constant noise density). Below 200Hz, it has a *pink* or $1/f$ characteristic. Consequently, generating an effective noise voltage for this is a bit more complicated.

Over the white noise portion, the relationship of equation 994 applies. As shown in section 22.8, equation 1008, noise in the $1/f$ section is given by:

$$e_n = e_{ndw} \sqrt{f_{cw} \ln \frac{f_2}{f_1}} \quad (996)$$

where f_1 and f_2 are the frequency limits of the measurement, f_{cw} is the corner frequency and e_{ndw} is the noise density at the corner frequency.

Example: Effective Noise Voltage of the 5532

Determine the effective noise voltage over the range 10Hz to 10kHz for the 5532 op-amp (see table, page 675).

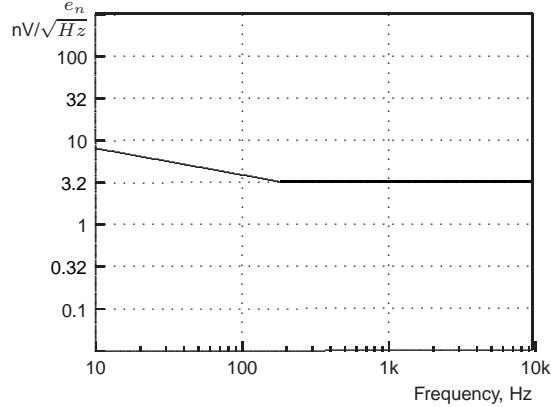


Figure 575: NE5534 Voltage Noise Spectrum

Solution

The noise corner frequency f_{cw} is at 200Hz. Consequently, the frequency spectrum divides into two regions: the 1/f region below 200Hz and the white noise region above 200Hz.

For the region below 200Hz, the effective noise voltage is given by equation 996:

$$\begin{aligned} e_{nf} &= e_{ndw} \sqrt{f_{cw} \ln \frac{f_2}{f_1}} \\ &= (3.2 \times 10^{-9}) \sqrt{200 \ln \frac{200}{10}} \\ &= 78.3 \times 10^{-9} \text{ V} \end{aligned}$$

For the region 200 to 10kHz, the effective noise voltage is given by equation 994:

$$\begin{aligned} e_{nw} &= e_{nd} \sqrt{B} \\ &= e_{nd} \sqrt{f_2 - f_1} \\ &= (3.2 \times 10^{-9}) \sqrt{10000 - 200} \\ &= 316 \times 10^{-9} \text{ V} \end{aligned}$$

The total noise is the Pythagorean sum of the two noise voltages:

$$\begin{aligned} e_n &= \sqrt{e_{nf}^2 + e_{nw}^2} \\ &= \sqrt{(78.3 \times 10^{-9})^2 + (316 \times 10^{-9})^2} \\ &= 325 \times 10^{-9} \text{ V} \end{aligned}$$

This is the value of e_n that would be used in noise calculations for this op-amp.

A Coda

Just out of curiosity, what result would we have obtained if we had ignored the rise of noise in the flicker region and assumed that the white noise region extended over the entire bandwidth?

$$\begin{aligned} e_{nw} &= e_{nd} \sqrt{B} \\ &= (3.2 \times 10^{-9}) \sqrt{10000 - 10} \\ &= 320 \times 10^{-9} \text{ V} \end{aligned}$$

In other words, ignoring the flicker noise changes the answer by 5 parts in 300. This approximation would have got us to within 1% of the correct answer. Consequently, it is probably reasonable to make this assumption in most noise calculations.

This is particularly true in audio applications. A 1db change in sound level is just noticeable by the human ear and corresponds to a factor of 1.12, ie, 12%. So anything less than a 12% difference in level will be unnoticeable.

22.5 Noise Gain

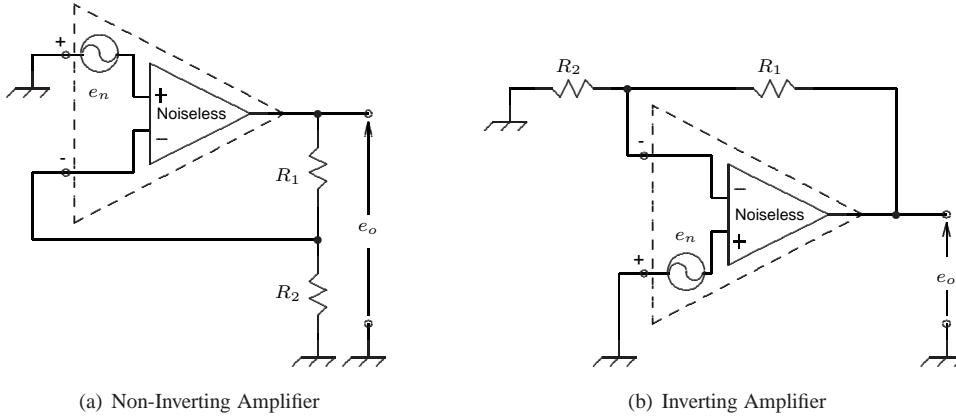


Figure 576: Noise Gain

Non-Inverting Amplifier

Consider the non-inverting amplifier with op-amp noise source e_n , shown in figure 576(a). (For simplicity, we'll ignore the contributions of the op-amp noise currents and resistor noise.) The *noise gain* G_n for this circuit (section 10.2) is:

$$\begin{aligned} G_n &= \frac{e_o}{e_n} \\ &= \frac{R_1 + R_2}{R_2} \text{ volts/volt} \end{aligned} \quad (997)$$

Now, treating this circuit as a negative feedback system, the sensor gain B is given by:

$$\begin{aligned} B &= \frac{e_f}{e_o} \\ &= \frac{R_2}{R_1 + R_2} \text{ volts/volt} \end{aligned} \quad (998)$$

Then the noise gain is related to the sensor gain as:

$$G_n = \frac{1}{B} \quad (999)$$

When a signal e_i is applied to the input of this circuit, it is attached to the non-inverting terminal. The signal gain G_s is given by:

$$\begin{aligned} G_s &= \frac{e_o}{e_i} \\ &= \frac{R_1 + R_2}{R_2} \text{ volts/volt} \end{aligned} \quad (1000)$$

That is, the noise gain and the closed-loop gain are equal.

Inverting Amplifier

Now consider the non-inverting amplifier of figure 576(b). In the absence of an external input signal, the circuit of figure 576(b) is a redrawn version of figure 576(a), so the noise gain of the inverting amplifier is the same as the non-inverting case. That is:

$$\begin{aligned} G_n &= \frac{e_o}{e_n} \\ &= \frac{R_1 + R_2}{R_2} \text{ volts/volt} \\ &= \frac{1}{B} \end{aligned} \quad (1001)$$

An input signal would be connected to the left end of R_2 . From section 12.5, the signal gain G_s is:

$$\begin{aligned} G_s &= \frac{e_o}{e_i} \\ &= -\frac{R_2}{R_1} \text{ volts/volt} \end{aligned} \quad (1002)$$

In this case, the noise gain and the closed-loop gain are *not* equal.

Summary

- In general, the term *noise gain* is a synonym for the quantity $1/B$ in a negative feedback system.
- As indicated in the example of the non-inverting and inverting amplifiers shown above, the noise gain must be distinguished from the gain (between input and output). They are different quantities.
- If one or both of the resistors in figure 576 is replaced by a reactance, the noise gain will become a function of the frequency.

The noise gain will turn up again in connection with stability analysis of op-amp circuits, section 24.

22.6 Noise Example: Voltage Follower

In circuit 577, a voltage buffer (follower) (see section 12.3) is driven from a source with internal resistance R_s .

For this example, we'll use the noise specifications of an amplifier like the TL081 JFET input op-amp. The noise voltage density is $e_{nd2} = 25nV/\sqrt{Hz}$ and the noise current density is $i_{nd1,2} = 0.01pA/\sqrt{Hz}$. We'll assume the temperature is 23°C, and the effective bandwidth is 20,000Hz.

Determine the output noise voltage.

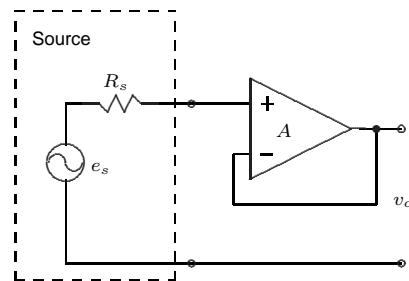


Figure 577: Operational Amplifier Voltage Follower

Solution

The equivalent circuit for noise analysis is shown in figure 578. Since we are interested in the output noise, the signal source e_s has been removed. The resistor is replaced with its noise thevenin equivalent and the op-amp with its noise model.

We evaluate each of the four noise sources separately (using the superposition principle) to determine its effective noise voltage, and the contribution of that noise voltage to the output noise. Then we generate a total noise voltage figure by doing an RMS sum of these outputs.

Step 1, Resistor Noise e_{n1}

The first step is to determine the effective noise voltage e_{n1} for the resistor. The temperature is 23°C, so $T = 293\text{K}$. Using equation 987, we have:

$$\begin{aligned} e_{n1} &= \sqrt{4KTBR} \\ &= \sqrt{4 \times (1.38 \times 10^{-23}) \times 293 \times 20000 \times (1 \times 10^6)} \\ &= 17.98 \times 10^{-6} \text{ V} \end{aligned}$$

The other noise voltage source (e_{n2}) is shorted and the noise current sources are open circuited. Then it is clear that this source contributes an output voltage:

$$e_{n1o} = 17.98 \times 10^{-6} \text{ V}$$

(Regarding notation: if the noise source is e_{n1} , the effect on the output has an **o** appended to make e_{n1o} .)

Step 2, Amplifier Voltage Noise e_{n2}

Next, we can determine the effect of the amplifier noise voltage e_{n2} . The noise voltage density is $e_{nd2} = 25\text{nV}\sqrt{\text{Hz}}$ over a bandwidth of 20,000Hz, so

$$\begin{aligned} e_{n2} &= e_{nd2}\sqrt{B} \\ &= (25 \times 10^{-9}) \times \sqrt{20000} \\ &= 3.53 \times 10^{-6} \text{ V} \end{aligned}$$

Like the resistor noise of Step 1, this will appear with the same magnitude at the output.

$$e_{n2o} = 3.53 \times 10^{-6} \text{ V}$$

Step 3, Amplifier Current Noise i_{n3}

This noise current source drives current through the output voltage. There is no resistance in this path, so the developed voltage is zero and this noise current source has no effect on the output.

(It is common to put a resistor R_f equal to the value of R_s in the feedback path. This has the effect of cancelling a voltage drop created by the input bias currents. If present, that resistor would convert the i_{n1} noise current into a noise voltage.)

Step 4, Amplifier Current Noise i_{n4}

This noise current source drives current through the source resistance R_s , so it creates a noise voltage. The effective value of the noise current is determined from the noise density and bandwidth:

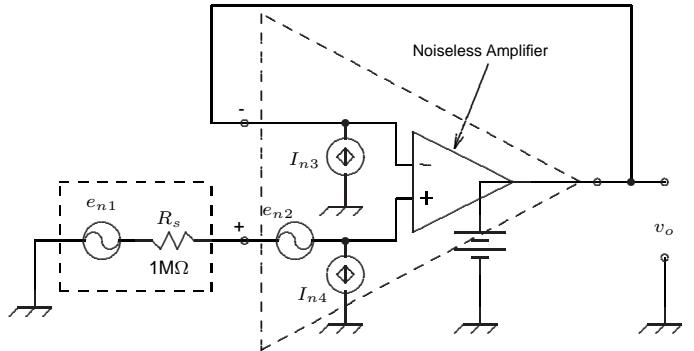


Figure 578: Voltage Follower Noise Model

$$\begin{aligned}
 i_{n4} &= i_{nd4}\sqrt{B} \\
 &= (0.01 \times 10^{-12}) \times \sqrt{20000} \\
 &= 1.41 \times 10^{-12} \text{ A}
 \end{aligned}$$

This noise current generates a noise voltage in R_s of:

$$\begin{aligned}
 e_{n4} &= i_{n4}R_s \\
 &= (1.41 \times 10^{-12}) \times (1 \times 10^6) \\
 &= 1.41 \times 10^{-6} \text{ V}
 \end{aligned}$$

As in the previous cases, this voltage appears at the output.

$$e_{n4o} = 1.41 \times 10^{-6} \text{ V}$$

Step 5, Total Noise Voltage e_{nT}

Finally, we calculate the RMS sum of the output voltages:

$$\begin{aligned}
 e_{nT} &= \sqrt{e_{n1o} + e_{n2o} + e_{n4o}} \\
 &= \sqrt{(17.98 \times 10^{-6})^2 + (3.53 \times 10^{-6})^2 + (1.41 \times 10^{-6})^2} \\
 &= 18.38 \times 10^{-6} \text{ volts}
 \end{aligned}$$

Notice that the result, $18.38\mu\text{V}$, is not that much larger than the largest noise source, $17.98\mu\text{V}$ because the other sources are relatively modest in comparison to it.

This example illustrates that large resistance values tend to contribute noise (a) because their inherent voltage noise is proportional to resistance, and (b) noise currents generate larger noise voltages when they flow through large resistances.

22.7 Noise Example: Inverter

Circuit 579 shows the standard inverting amplifier configuration. We'll assume the source resistance is zero in this case.

The noise density specifications for the NE5532 op-amp are:

$$\begin{aligned}
 e_{nd} & 5nV\sqrt{\text{Hz}} \\
 i_{nd} & 0.7pA\sqrt{\text{Hz}}
 \end{aligned}$$

As in the previous example, we'll assume the temperature is 23°C , and the effective bandwidth is 20,000Hz.

Determine the output noise voltage.

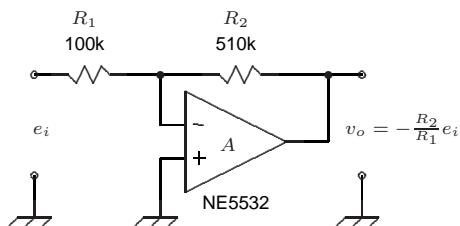


Figure 579: Operational Amplifier Inverter

Solution

The noise equivalent circuit for the inverter circuit is shown in figure 580.

As in the previous example, we'll evaluate each of the five noise sources separately and then generate a total noise voltage figure by doing an RMS sum.

Step 1, Noise Sources

The first step is to determine the effective noise voltages and currents. The table shows the effective noise voltages and currents, the formulae used to calculate them, and the page on which this formula is presented.

Label	Value	Equation	Page
e_{n1}	$5.68 \times 10^{-6} \text{V}$	$e_n = \sqrt{4KTBR}$	670
e_{n2}	$12.84 \times 10^{-6} \text{V}$	$e_n = \sqrt{4KTBR}$	670
e_{n3}	$0.70 \times 10^{-6} \text{V}$	$e_n = e_{nd}\sqrt{B}$	675
i_{n4}	$98.7 \times 10^{-12} \text{A}$	$i_n = i_{nd}\sqrt{B}$	675
i_{n5}	$98.7 \times 10^{-12} \text{A}$	$i_n = i_{nd}\sqrt{B}$	675

Sources i_{n4} and i_{n5} are equal in magnitude but uncorrelated.

Step 2, Output Noise due to e_{n1}

The noise voltage source e_{n1} appears at the input of an inverting amplifier of gain $-R_2/R_1$. Since we're only concerned with amplitude and phase is irrelevant in this situation, we'll drop the minus sign. Consequently, the output contribution of e_{n1} is

$$\begin{aligned} e_{n1o} &= e_{n1} \left(\frac{R_2}{R_1} \right) \\ &= (5.68 \times 10^{-6}) \left(\frac{510k}{100k} \right) \\ &= 28.9 \times 10^{-6} \text{ V} \end{aligned}$$

Step 3, Output Noise due to e_{n2}

The noise voltage source e_{n2} is located in the feedback path of the amplifier. The amplifier will maintain the voltage between the two inputs at zero. Consequently, whenever e_{n2} is at $+x$ volts, the output must be at $-x$ volts. As a result, this source contributes to the output voltage with unity gain.

$$\begin{aligned} e_{n2o} &= e_{n2} \\ &= 12.84 \times 10^{-6} \text{ V} \end{aligned}$$

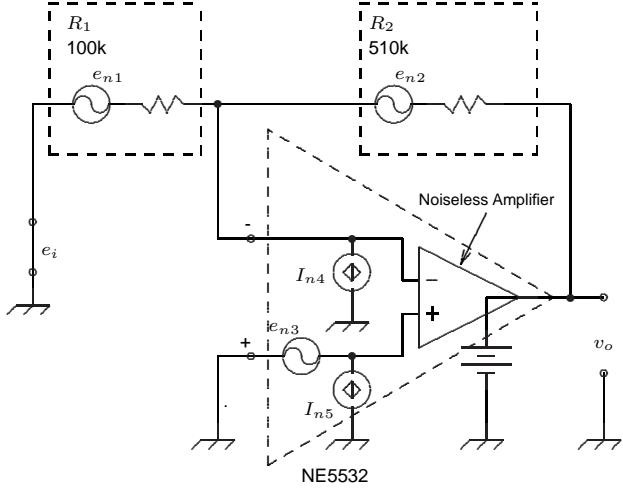


Figure 580: Inverter Noise Model

Step 4, Output Noise due to e_{n3}

The noise voltage source e_{n1} appears at the non-inverting input of what is effectively a non-inverting amplifier of gain

$$G = 1 + \left(\frac{R_2}{R_1} \right)$$

Since this is the amount by which the noise is amplified in both the inverting and non-inverting amplifier configurations, this gain is referred to as the *Noise Gain* (section 22.5) of the amplifier. (It could equally well be referred to as the Voltage Offset Gain or the Non-Inverting gain, but the name Noise Gain has stuck.)

$$\begin{aligned} e_{n3o} &= e_{n3} \left(1 + \frac{R_2}{R_1} \right) \\ &= (0.70 \times 10^{-6}) \left(1 + \frac{510k}{100k} \right) \\ &= 4.27 \times 10^{-6} \text{ V} \end{aligned}$$

Notice that this noise voltage contribution by the op-amp is much smaller than the noise contributed by resistors R_1 and R_2 . If we wanted to reduce the overall noise, a good place to start would be by reducing R_1 and R_2 .

Step 5, Amplifier Current Noise i_{n4}

With all other sources removed, the non-inverting terminal is grounded and the voltage at the non-inverting terminal is zero volts. The ideal op-amp holds the inverting terminal at the same potential, zero volts. Both ends of R_1 are at zero volts, and so there is no current through R_1 . Then the entire noise current i_{n3} must flow through the feedback resistor R_2 . The output voltage must be at

$$\begin{aligned} e_{n4o} &= i_{n3} R_2 \\ &= (98.7 \times 10^{-12}) \times (510 \times 10^3) \\ &= 50.3 \times 10^{-6} \text{ volts} \end{aligned}$$

(If this argument seems familiar, it's because it is similar to the argument advanced for the effect of bias current in section 21.2. We give you a deal: two circuit understandings for the price of one analysis.)

Step 6, Amplifier Current Noise i_{n5}

This noise current source drives current through the noise voltage source e_{n3} . Noise source e_{n3} has no internal resistance and is replaced in the superposition process by a short circuit. Consequently, the noise current source i_{n5} develops no noise voltage of its own and has no effect on the output.

$$e_{n5o} = 0$$

Step 7, Total Noise Voltage e_{nT}

Finally, we calculate the RMS sum of the output voltages:

Noise Voltage	Magnitude	Magnitude ²
e_{n1o}	28.9×10^{-6}	835×10^{-12}
e_{n2o}	12.84×10^{-6}	165×10^{-12}
e_{n3o}	4.27×10^{-6}	18×10^{-12}
e_{n4o}	50.3×10^{-6}	2530×10^{-12}
e_{n5o}	0×10^{-6}	0×10^{-12}
Total		3548×10^{-12}

Then the resultant output voltage is the square root of this total:

$$\begin{aligned} e_{nT} &= \sqrt{3548 \times 10^{-12}} \\ &= 59 \times 10^{-6} \text{ V} \end{aligned}$$

This result could be substantially reduced by reducing the values of the input and feedback resistors.

22.8 Noise Formulae

In this section, we develop the formulae to obtain noise voltage from noise voltage density, first for white noise and then for pink noise.

White Noise

The nature of an op-amp noise spectrum is shown in figure 581.

When the noise density is constant with frequency, it's a simple matter to determine the effective noise voltage: multiply the noise density by the square root of the equivalent bandwidth (equations 994 and 995). However, when the noise spectrum increases at low frequencies as shown in figure 581, (and the noise bandwidth includes those frequencies), a different technique must be used.

To do this more accurately, we must break up the noise spectrum into two segments: the $1/f$ noise segment below f_{cw} , and the white noise segment above f_{cw} .

The total power over some frequency range f_1 to f_2 is found by integrating the power noise density function over that range.

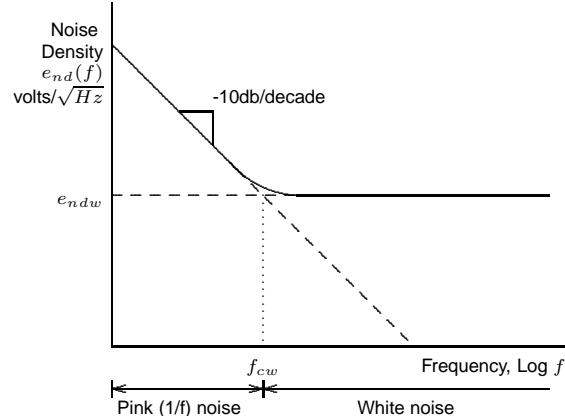


Figure 581: Op-Amp Noise Spectrum

$$P_{n12} = \int_{f_1}^{f_2} P_{nd} df \quad (1003)$$

Substituting e^2/R for P in equation 1003, we have

$$\begin{aligned}\frac{e_{n12}^2}{R} &= \int_{f1}^{f2} \frac{e_{nd}^2}{R} df \\ \text{and} \\ e_{n12}^2 &= \int_{f1}^{f2} e_{nd}^2 df \\ \text{or} \\ e_{n12} &= \sqrt{\int_{f1}^{f2} e_{nd}^2 df}\end{aligned}\tag{1004}$$

When the noise density is a constant value, as it is (e_{ndw}) during the white noise section above f_{cw} , then the integration is very simple:

$$\begin{aligned}e_{n12} &= e_{ndw} \sqrt{f2 - f1} \\ &= e_{ndw} \sqrt{B}\end{aligned}\tag{1005}$$

Pink Noise

Below the transition frequency f_{cw} , the noise power is inversely proportional to the frequency. Referring to figure 582, the noise power density P_{nd} at some frequency f is given by the relationship

$$\frac{P_{nd}}{P_{ndw}} = \frac{f_{cw}}{f}\tag{1006}$$

Substituting e^2/R for P in equation 1006, we can find that

$$e_{nd}^2 = e_{ndw}^2 \frac{f_{cw}}{f}\tag{1007}$$

Now, suppose that we wish to find the effective noise between frequencies f_1 to f_2 . We substitute for e_{nd}^2 from equation 1007 into equation 1004:

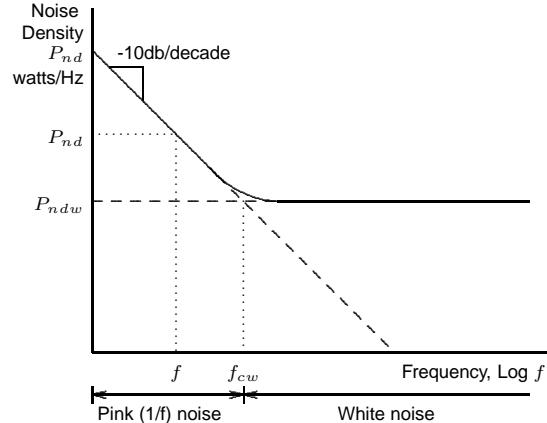


Figure 582: Flicker Noise Power

$$\begin{aligned}e_{n12} &= \sqrt{\int_{f1}^{f2} e_{ndw}^2 \frac{f_{cw}}{f} df} \\ &= e_{ndw} \sqrt{f_{cw} \ln \frac{f_2}{f_1}}\end{aligned}\tag{1008}$$

Symbols

There are many symbols to keep track of in this section. There is little consistency between the various references, so I have used symbols that help understanding and do not conflict with others in this text.

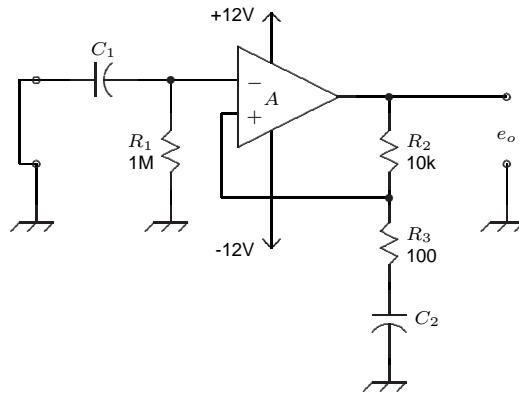
Symbol	Units	Notes
P_s	watts	Power in the signal waveform
P_n	watts	Total power in the noise waveform, over some bandwidth
P_{nd}	watts/Hz	Noise density, over a small range of the bandwidth.
e_s	volts RMS	Effective (RMS) voltage in the signal waveform
e_n	volts RMS	Total effective (RMS) voltage in the noise waveform over some bandwidth B
e_{nf}	volts RMS	Effective (RMS) voltage over the flicker noise section of the bandwidth.
e_{nw}	volts RMS	Effective (RMS) voltage over the white noise section of the bandwidth.
e_{n12}	volts RMS	Total effective (RMS) voltage in the noise waveform over some frequency range f_1 to f_2
i_n	amps RMS	Total effective (RMS) current in the noise waveform over some bandwidth B
e_{nd}	volts RMS/ $\sqrt{\text{Hz}}$	Noise voltage density, over a small range in the bandwidth
i_{nd}	amps RMS/ $\sqrt{\text{Hz}}$	Noise current density, over a small range in the bandwidth
e_{ndw}	volts RMS/ $\sqrt{\text{Hz}}$	Noise voltage density of the white noise segment (see figure 581)
f_c	Hz	Cutoff frequency of a highpass or lowpass filter
f_{cw}	Hz	Transition frequency from $1/f$ to white noise (see figure 581)
B	Hz	Bandwidth (brick wall characteristic)
B_e	Hz	The equivalent brick wall bandwidth, for a non-brick-wall filter characteristic
f_l	Hz	Lower cutoff frequency of a bandpass filter
f_u	Hz	Upper cutoff frequency of a bandpass filter

22.9 Exercises

- For the circuit shown below, the following applies for the operational amplifier:

$$\begin{array}{ll} \text{Noise Voltage} & e_n \quad 10\text{nV}/\sqrt{\text{Hz}} \\ \text{Noise Current} & i_n \quad 1\text{pA}/\sqrt{\text{Hz}} \end{array}$$

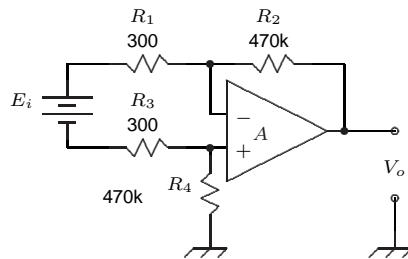
Other than the specifications given above, the op-amp may be considered to be ideal. The resistors may be considered to be noiseless, and the measurement bandwidth is 100kHz. The capacitors may be considered AC short circuits at the frequencies of interest. For measurement purposes, the input is shorted to ground as shown.



What is the RMS value of the AC component of the op-amp output voltage?

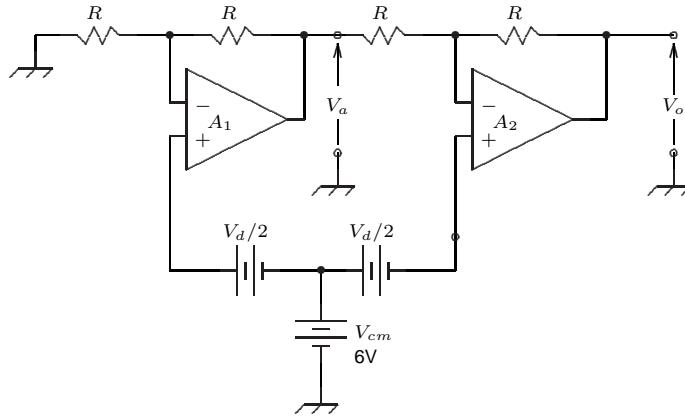
2. For the amplifier shown below:

- $e_n = 18\text{nV}/\sqrt{\text{Hz}}$
- $i_n = 1.2\text{nA}/\sqrt{\text{Hz}}$
- Bandwidth=20kHz
- Resistor noise= $\sqrt{4KTBR}$ volts
- $K = 1.38 \times 10^{-28}$



- What is the voltage gain of this amplifier V_o/E_i ?
- Draw the equivalent noise circuit with the op-amp and the resistors.
- Calculate the total noise output voltage.

3. A differential amplifier circuit is shown below.



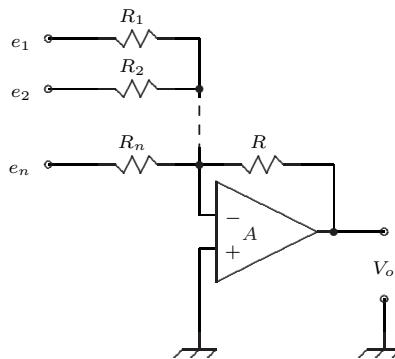
- (a) Each operational amplifier has an equivalent input noise voltage of $25\text{nV}/\sqrt{\text{Hz}}$. The resistors are $1\text{M}\Omega$. The op-amp equivalent input noise current may be ignored. We wish to determine the output noise voltage v_o .

Start by assuming U1 and its resistors are noiseless and determine the output noise voltage due to U2 and its two resistors.

- (b) Now assume U2 and its resistors are noiseless and determine the output noise voltage due to U1 and its two resistors.

- (c) Combine these two effects to get the total v_o .

4. For the adder circuit shown below:



- (a) What is the effect on the noise gain as the number of inputs is increased?

- (b) Assuming that all the resistances are equal to R , develop an expression for the noise gain in terms of the number of inputs n .

5. The circuit of a *magnetometer amplifier* is shown in figure 583. The coil is used to detect weak alternating (AC) magnetic fields. The coupling capacitor C_1 removes DC from the output signal.

The circuit should amplify the open-circuit output of the detector coil by a factor of 10^4 volts/volt. The amplifier to be used in this circuit has an offset voltage V_{os} of 2 mV and a bias current I_b of 400 nA . The change of offset voltage with temperature is $0\text{ }\mu\text{volts/}^\circ\text{C}$ and the offset current I_{os} is zero.

- (a) Calculate the voltage at the wiper of the pot for the worst case value of bias current and offset voltage.

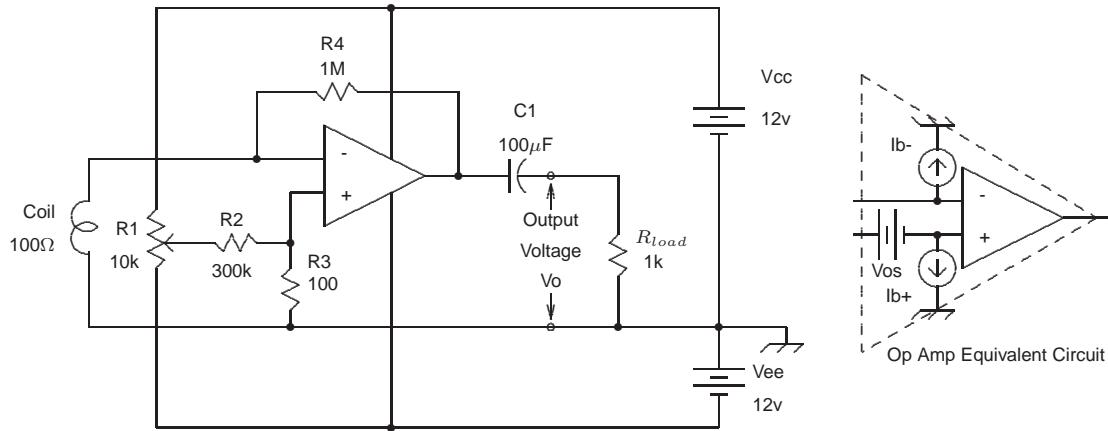


Figure 583: Magnetometer Circuit

- (b) What is the *primary* purpose of the potentiometer R_1 ?
- (c) Are components R_1 and R_2 really necessary? (Your answer must be justified with a calculation).
- (d) Two operational amplifiers are being considered for the circuit:
- The NE5534 is a BJT input operational amplifier with $e_n = 4 \text{ nv}/\sqrt{\text{Hz}}$ and $i_n = 1 \text{ pA}/\sqrt{\text{Hz}}$.
 - The TL081 is a FET input operational amplifier with $e_n = 25 \text{ nv}/\sqrt{\text{Hz}}$ and $i_n = 0.01 \text{ pA}/\sqrt{\text{Hz}}$.
- Calculate the total noise output contributions of each amplifier when it is used in this circuit, assuming a measurement bandwidth of 100 Hz.
- (e) An EE graduate joins the company manufacturing this circuit and proposes an alternative circuit:

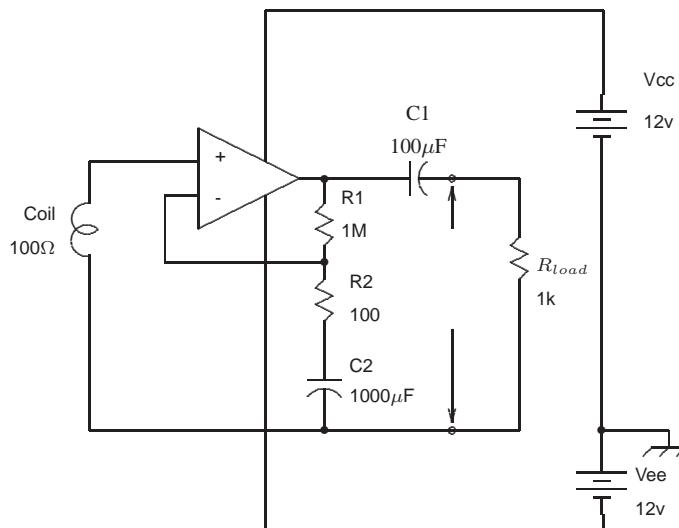
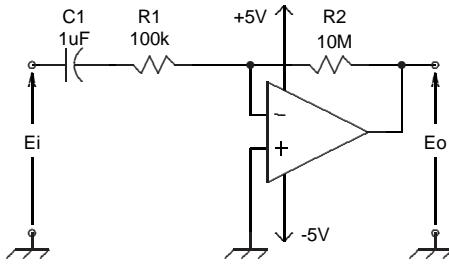


Figure 584: Alternative Magnetometer Circuit

Assuming that the offset voltage and bias currents are at their maximum values, calculate the DC output voltage due to both at the output of the op-amp in this new circuit.

- (f) Which of these two circuits, figure 583 or figure 584, is superior? Explain.

6. For the amplifier circuit shown below:



- (a) If the offset voltage V_{os} of the amplifier is 2mV, what is the magnitude of the output voltage caused by this offset?
 (b) If the bias current of the amplifier is 200pA, what is the output voltage caused by this current?
 (c) The noise specifications for the amplifier are:

- $e_n = 25\text{nV}/\sqrt{\text{Hz}}$
- $i_n = 1\text{pA}/\sqrt{\text{Hz}}$
- Bandwidth = 10^4 Hz
- Ambient Temperature = 25°C

Calculate the output noise due to the noise sources in the amplifier and resistors.

7. Regarding the noise model of figure 578 on page 680: The noise source e_{n2} can be located in series with either input: either position should give the same result. In figure 578 the noise source is placed in series with the non-inverting input. Suppose we had drawn the op-amp noise model so that the noise source e_{n2} was located in series with the op-amp inverting input. This should produce the same output voltage as it does in the other position. Why does that happen?)

23 Operational Amplifier Frequency Response

23.1 Introduction

For a negative feedback system to work properly, the open-loop gain A must be much larger than the closed-loop gain $1/B$ (section 10). The negative feedback op-amp circuits shown in earlier sections all made that assumption. For the op-amp this is a valid assumption at very low frequencies (so-called *DC* signals), where the open-loop voltage gain is at its maximum.

However, the voltage gain of an op-amp is constant at its open-loop value up to a few tens of hertz, and then decreases (*rolls off* in the parlance), typically at a rate of 20db per decade. That is, the voltage gain decreases by a factor of 10 as the frequency increases by a factor of 10. A typical example of this behaviour is shown in figure 585.

Now consider the case of an amplifier in a circuit where the feedback factor B is constant with frequency. (This would be the case for an amplifier that is to have a constant closed-loop gain of $1/B$ at all frequencies.) As the open-loop gain A falls, so does the loop gain AB . Negative feedback reduces errors (such as disturbances) by an amount equal to the loop gain. As a result, at higher frequencies a smaller loop gain results in a less effective reduction of errors and there is a very definite upper limit to the maximum frequency of operation of an op-amp based negative feedback circuit.

For the utility op-amp at moderate gains, this limit can be as low as some tens of kilohertz. When a circuit must operate correctly (that is, the negative feedback must be effective) then the open-loop gain and loop gain must be checked to ensure that they are sufficient at the highest operating frequency.

- Section 23.2 explains the origin and modelling of op-amp open-loop gain-frequency response, and the *gain-bandwidth* specification.
- Section 23.4 shows how the gain-bandwidth and closed-loop gain determine the loop gain and maximum usable frequency of the circuit.
- For reasons that will become clearer when we discuss stability in section 24, most op-amps are designed to have an amplitude response that rolls off at 20db/decade. Section 23.5 shows that the combination of a 20db/decade rolloff in gain with resistive feedback results in a step response that resembles the step response of an RC lowpass filter.
- Some op-amps have a frequency response that rolls off at 20db/decade at low frequencies and then a steeper 40db/decade at high frequencies. Section 23.6 shows that resistive feedback with this type of op-amp results in a 2nd order step response which may or may not be acceptable.

In Section 24 we will see that the frequency response not only determines the open-loop and loop gains, but may also affect the stability of a negative feedback circuit. Certain combinations of feedback and open-loop frequency response may cause the circuit to oscillate, a very ugly problem.

In this section we discuss the frequency response behaviour of the standard *voltage feedback* op-amp. In section 35.3 we will discuss the *current-feedback* amplifier, which has very different behaviour.

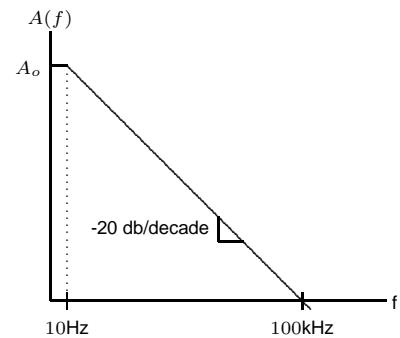
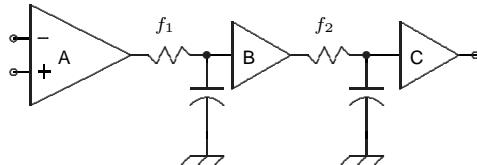


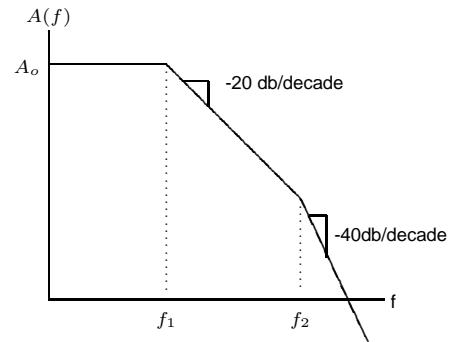
Figure 585: Open-Loop Gain vs Frequency

23.2 Gain-Bandwidth

In our ideal concept of an operational amplifier, the open-loop gain is constant with frequency. In the real world, the amplifier may be viewed as a series of stages A, B and C as shown in figure 586(a). Amplifier A is an ideal operational amplifier of low-frequency gain A_o volts/volt. Amplifiers B and C are *unity-gain buffers*. They present a high impedance at their input and a low impedance at their output¹⁹⁹, and have a voltage gain of 1 volt/volt. The buffers eliminate the effect of inter-stage loading, so that the frequency response of each stage may be added to obtain the overall frequency response. The device capacitances and resistances form lowpass filters in the signal path between each stage, here represented by discrete component lowpass RC filters.



(a) High Frequency Model



(b) Multiple Pole Response

Figure 586: Op Amp Open Loop Response

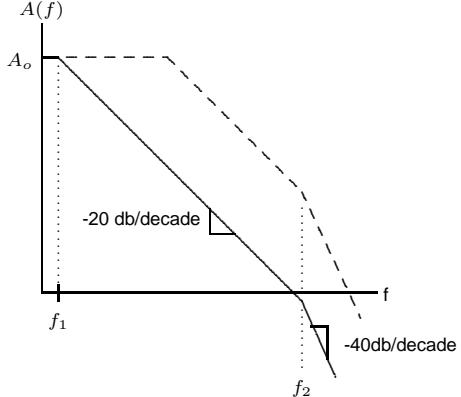
The effect of these lowpass filters is to break the open-loop gain downward, first at f_1 to -20db/decade and then at f_2 to -40db/decade, as shown in figure 586(b).

The first break frequency f_1 is often made much lower than any other break frequencies. This shapes the frequency response curve so that the break frequency f_2 (and higher frequencies) occur after the open-loop gain has passed through the 0db axis, that is, the open-loop gain is less than 1. This technique is referred to as *dominant pole compensation* and makes the op-amp stable in many common negative feedback circuits. An example of the open-loop frequency response is shown in figure 587. As a consequence of dominant pole compensation, the designer need not be concerned with stability issues in the design of many op-amp circuits.

Gain Bandwidth Product

In figure 587, the open-loop gain characteristic $A(f)$ is composed of a constant factor A_o equal to the low-frequency gain, and a single time-constant lag factor with a corner frequency equal to f_1 . The second break frequency f_2 has been moved to a point where it has no effect on the operation of the circuit, so it may be ignored. Then the open-loop gain is the product of the DC open-loop gain A_o and a first-order lowpass (lag) function with corner frequency at f_1 :

Figure 587: Dominant Pole Response



¹⁹⁹Notice that amplifier B is independent of the feedback factor B .

$$A(f) = A_o \left[\frac{1}{1 + \frac{jf}{f_1}} \right] \quad (1009)$$

On the sloping portion of the curve, where $f \gg f_1$, then

$$\begin{aligned} A(f) &\approx \frac{A_o}{\frac{jf}{f_1}} \\ &= \frac{A_o f_1}{jf} \end{aligned} \quad (1010)$$

The quantity $A_o f_1$ is known as *the gain-bandwidth product* f_{GBP} for the op-amp.

At the point where the open-loop gain crosses the zero axis, the value of $A(f)$ is zero db (unity volts/volt). Substituting 1 for $A(f)$ in equation 1010 and ignoring phase for the moment, we have:

$$\begin{aligned} 1 &= \frac{A_o f_1}{f} \\ f &= A_o f_1 \\ &= f_{GBP} \end{aligned} \quad (1011)$$

In words,

the frequency at the axis crossing is equal to the gain-bandwidth product.

The gain-bandwidth-product frequency f_{GBP} is marked on figure 588.

Example

The 741 type op-amp uses dominant pole compensation. The National Semiconductor data sheet for the 741 specifies an open-loop gain of 50V/mV. With no other indication of the frequency at which this gain was measured, we'll assume that it's a maximum value for the op-amp and corresponds to A_o in figure 587. The gain-bandwidth product (the National Semiconductor data sheet calls it the *bandwidth*) is 437kHz.

Draw the open-loop frequency response curve, showing the low-frequency gain A_o , the open-loop break frequency f_1 and the gain-bandwidth-product frequency f_{GBP} .

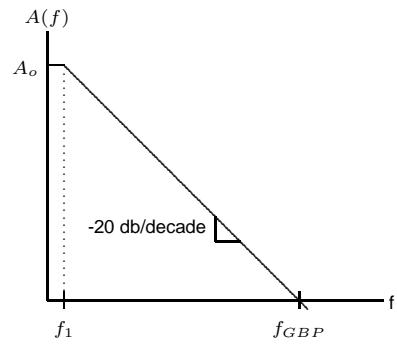


Figure 588: Gain-Bandwidth

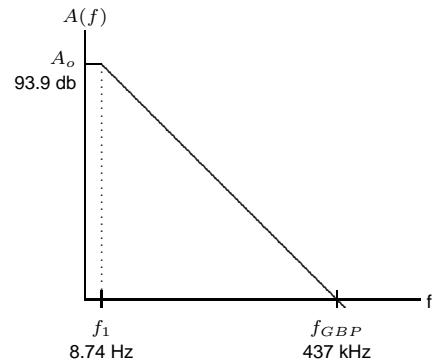


Figure 589: 741 Gain-Bandwidth

Solution

The open-loop gain A_o is given as 50V/mV, that is, 50×10^3 V/V. In decibels, this is:

$$\begin{aligned} A_{odb} &= 20 \log_{10}(50 \times 10^3) \\ &= 93.9 \text{ db} \end{aligned}$$

Over the sloping portion of the open-loop curve, the gain and frequency are constant and equal to the gain-bandwidth product. Then, from equation 1011, we have

$$\begin{aligned} f_1 &= \frac{f_{GBP}}{A_o} \\ &= \frac{437 \times 10^3}{50 \times 10^3} \\ &= 8.74 \text{ Hz} \end{aligned}$$

Typical Values of Gain-Bandwidth

Type	Part Number	Open Loop Gain Volts/Volt	Gain-Bandwidth Product MHz
Low-cost BJT	LM324	10^5	1
General Purpose BJT	741	50×10^3	0.437
General Purpose JFET	TL074	2×10^5	3
General Purpose MOSFET	LMC660	10^6	1.4
High Speed BJT	LM6161	750	50
Current Feedback	LM6181	n/a	100
High Speed MOSFET	OPA4354	3×10^5	100

23.3 Clarifying Gain

This section and section 24 on Stability rely heavily on a clear understanding of open and closed-loop gain factors, so before we proceed, a summary and review.

Open-Loop Gain A

The open-loop gain of an amplifier, A_{ol} or A is simply the ratio of the output signal to the input signal of an amplifier that is operated without feedback, figure 590(a).

Closed-Loop Gain G

Placing negative feedback around an amplifier is referred to as *closing the (feedback) loop*. The non-inverting amplifier configuration is shown as an example in figure 590(b). A sensor network B samples the output v_o . It then feeds back a sensor signal v_f to the input comparator. Negative feedback then acts to drive the feedback signal v_f always matches the input signal e_i .

For the closed-loop gain v_o/e_i we use the symbol G .

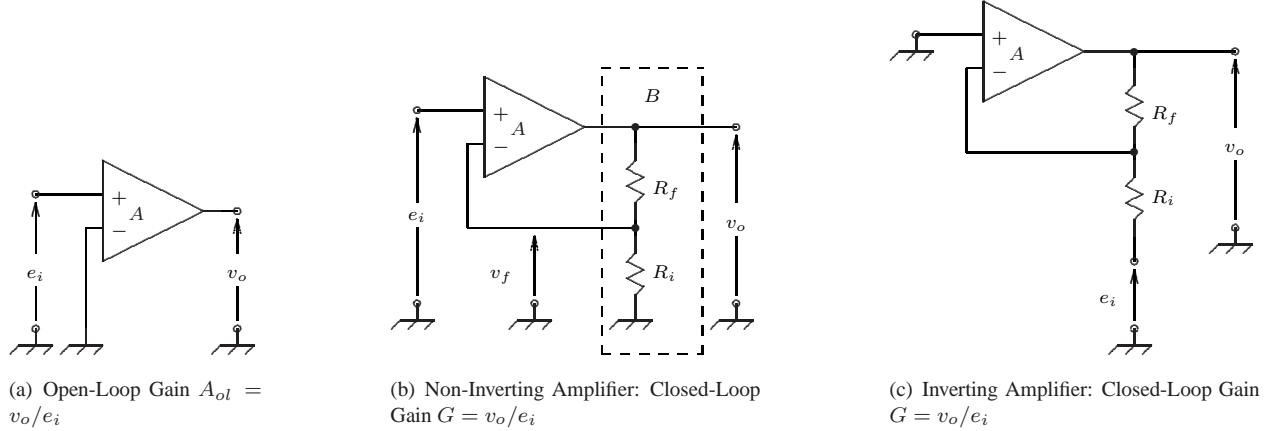


Figure 590: Open and Closed Loop Gain

Sensor Gain B

The gain B of the sensor network is the ratio of the feedback signal to the output signal, v_f/e_o . For the resistive divider that makes up the sensor network in figure 590(b), the sensor gain is:

$$B = \frac{v_f}{e_o} = \frac{R_i}{R_f + R_i} \quad (1012)$$

In the circuit of figure 590(b), it turns out (section 10.2 on page 287) that the closed-loop gain is equal to the reciprocal of the sensor gain²⁰⁰:

$$G = \frac{v_o}{e_i} = \frac{1}{B} = \frac{R_f + R_i}{R_i} = 1 + \frac{R_f}{R_i} \quad (1013)$$

Loop Gain AB

The *loop gain* AB is the product of the gains around the complete feedback loop. As we will see, the frequency characteristic of the loop gain is important in issues of frequency response and stability. The relationship between open-loop gain, sensor-gain and closed-loop gain is indicated in section 23.4 and figure 591 on page 696.

Noise Gain $1/B$

The *noise gain*, which is $1/B$ for a negative feedback system, is also the amount by which a noise signal is amplified before it appears at the output, section 22.5. For the circuit of figure 590(b), the noise gain and the closed-loop gain are equal, but that is not true in all cases.

Inverting Amplifier: Noise Gain not equal to Closed Loop Gain

For the inverting amplifier circuit of figure 590(c), the closed-loop gain G is given in section 12.5 and figure 253(b) by:

$$G = \frac{v_o}{e_i} = -\frac{R_f}{R_i} \quad (1014)$$

²⁰⁰As usual in a negative feedback system, the open-loop gain must be much larger than the closed-loop gain for this to be true.

However, the sensor gain B and noise gain $1/B$ are unchanged from the non-inverting case of figure 590(b).

Consequently, whenever in this text that the closed-loop gain G is taken as equal to the noise gain $1/B$, then the circuit must be the non-inverting type shown in 590(b).

These ideas are summarized in the following table.

Amplifier type	Non-Inverting	Inverting
Sensor Gain B	$\frac{R_i}{R_i + R_f}$	$\frac{R_i}{R_i + R_f}$
Noise Gain $1/B$	$\frac{R_i + R_f}{R_i}$	$\frac{R_i + R_f}{R_i}$
Closed-Loop Gain G	$\frac{1}{B} = \frac{R_i + R_f}{R_i}$	$-\frac{R_f}{R_i}$

23.4 Maximum Useful Frequency

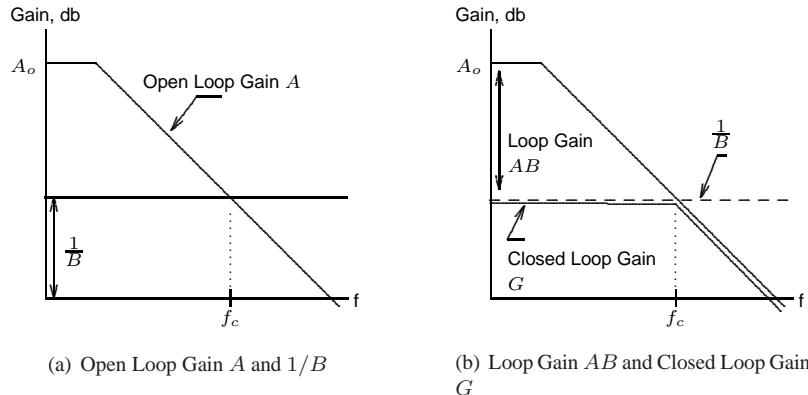


Figure 591: Open and Closed Loop Response

Consider the case of the non-inverting amplifier. When the open-loop gain A of a negative feedback system is much larger than the closed-loop gain G , the closed-loop gain is approximately equal to $1/B$, the reciprocal of the sensor gain, also known as the *noise gain*.²⁰¹

Then for low frequencies we could state:

$$G \approx 1/B \text{ when } A \gg 1/B \quad (1015)$$

In figure 591(a), the magnitude-frequency function of $1/B$ is superimposed on the function of the amplifier open-loop gain A . At the intersection frequency f_c , the open loop gain A is equal to $1/B$, and so the approximation $A \gg 1/B$ no longer holds. Above the frequency f_c , the closed-loop gain is defined by the open-loop gain, so that

$$G \approx A \quad (1016)$$

²⁰¹The open-loop gain, closed-loop gain and loop gain are all functions of frequency. For example, the open-loop gain should be written as $A(f)$. For notational compactness we will leave off the (f) and assume that it is understood. Then the open-loop gain is A , the sensor gain B and the closed-loop gain G .

This is indicated in figure 591(b), where the closed-loop gain G is shown as following $1/B$ at low frequencies and then A at high frequencies. That is, the system acts properly as a negative feedback system over the frequency range DC to somewhat below f_c , where A is still sufficiently large (whatever that may be) with respect to $1/B$.

Also indicated on figure 591(b) is the *loop gain* AB , which is equal to the difference between the open-loop gain A and the closed-loop gain $1/B$.

$$\begin{aligned}\text{Loop Gain} &= AB_{db} \\ &= A_{db} - 1/B_{db}\end{aligned}$$

This gives us another criterion for the maximum frequency of operation for this op-amp and closed-loop gain: the loop gain AB must be much larger than 1.

At frequency f_c , the closed-loop gain intersects with the open-loop gain, so they are equal. The loop gain is then 0db, or a factor of 1. Suppose that we define *much larger* as a factor of 10. The gain A rolls off at a rate of 20db/decade, or by a factor of 10 with each decade in frequency. Then a loop gain of 10 is obtained at one decade below f_c , at $f_c/10$, and the maximum useful frequency for this system is $f_c/10$.

Example

Suppose that the 741 op-amp of the example on page 23.2 is being considered for a high-fidelity application. Determine the open-loop gain at a frequency of 20kHz.

Solution

Using equation 1010, (and ignoring the phase, again) we can solve for the gain A at 20kHz:

$$\begin{aligned}A &= \frac{A_o f_1}{j f} \\ &= \frac{437 \times 10^3}{20 \times 10^3} \\ &= 20.85 \text{ volts/volt}\end{aligned}$$

In decibels,

$$\begin{aligned}A_{db} &= 20 \log_{10}(20.85) \\ &= 26.7 \text{ db}\end{aligned}$$

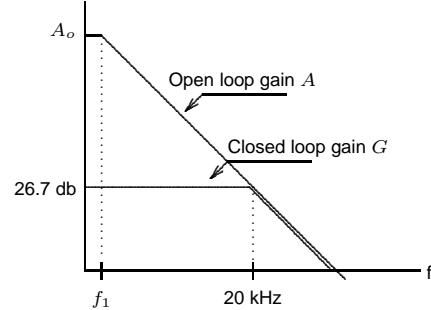


Figure 592: 741 Gain-Bandwidth

The resultant open-loop and closed-loop gain functions are shown in figure 592.

Notice the very limited voltage gain available from the 741 op-amp at the top of the audio band (20kHz).

Example

A 741 op-amp is to be used in an audio application over the frequency range 20Hz to 20kHz. Assuming that the open-loop gain A should be at least ten times the closed-loop gain G , what is the maximum allowable closed-loop gain? The amplifier is used in the non-inverting configuration.

Solution

The open-loop gain A decreases with frequency and, over the frequency range 20Hz to 20kHz, is at a minimum at 20kHz. In the previous example we determined that the open-loop gain is 20.85 volts/volt at that frequency. Consequently, the closed-loop gain G should be no larger than a tenth of this value, or 2.085 volts/volt.

Consequently, for any audio application requiring a closed-loop gain greater than 2 volts/volt, the 741 op-amp is not suitable.

23.5 Single Pole Response

We can substitute radian frequency ω radians/second for frequency f in Hz, using the identity $\omega = 2\pi f$. Then we can draw the open and closed-loop frequency responses with ω on the frequency axis as shown in figure 593.

We can see that

- the closed-loop gain extends out to frequency ω_c , and that ω_c will increase as the closed-loop gain decreases.
- beyond ω_c the closed-loop frequency response rolls off with the open-loop gain, at -20db/decade.

This system is equivalent to a gain block A in series with a first-order lowpass filter (like an RC lowpass filter) with a cutoff frequency of ω_c radians/second. Then it's reasonable to expect that the transient response will resemble the transient response of a lowpass filter.

Here is the algebraic analysis that confirms this:

The overall open-loop gain of the op-amp can be modelled as a constant term A_o in series with a lowpass filter of cutoff frequency ω_1 . Then

In equation 1009 on page 693 we had the frequency response for this system:

$$A(f) = \frac{A_o}{1 + \frac{jf}{f_1}}$$

Substitute ω for $f/2\pi$, ω_1 for $f_1/2\pi$ and s for $j\omega$, and we can write:

$$A(s) = \frac{A_o}{1 + \frac{s}{\omega_1}} \quad (1017)$$

The closed-loop gain is given by the negative feedback equation 425 on page 287. In this case, the open-loop gain A and closed-loop gain G are functions of frequency, so we'll write them as $A(s)$ and $G(s)$:

$$G(s) = \frac{A(s)}{1 + A(s)B} \quad (1018)$$

Substituting for $A(s)$ from equation 1017 into equation 1018,

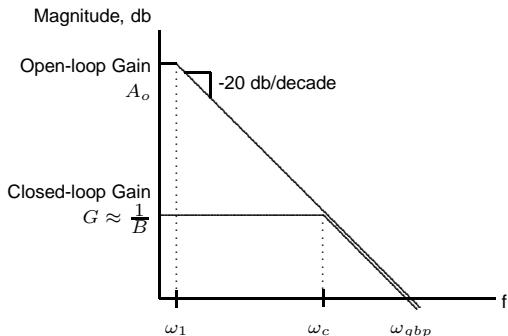


Figure 593: Single Pole Response

$$\begin{aligned}
G(s) &= \frac{\frac{A_o}{1+s/\omega_1}}{1 + \left(\frac{A_o}{1+s/\omega_1}\right)B} \\
&= \frac{A_o}{1 + A_o B + \frac{s}{\omega_1}} \\
&= \frac{1}{\frac{1}{A_o} + B + \frac{s}{A_o \omega_1}}
\end{aligned} \tag{1019}$$

If $A_o \gg 1/B$, the usual case, then $1/A_o \ll B$ and the term $1/A_o$ in the denominator can be neglected. Then:

$$\begin{aligned}
G(s) &\approx \frac{1}{B + \frac{s}{A_o \omega_1}} \\
&= \frac{1}{B} \left(\frac{1}{1 + \frac{s}{A_o B \omega_1}} \right)
\end{aligned} \tag{1020}$$

This can be written as

$$G(s) = \frac{1}{B} \left(\frac{1}{1 + \frac{s}{\omega_c}} \right) \tag{1021}$$

where

$$\omega_c = A_o B \omega_1 \tag{1022}$$

Consequently, the closed-loop gain is a single-pole lowpass function with low frequency gain equal to $1/B$ and cutoff frequency equal to ω_c , as shown in figure 593.

The original cutoff frequency ω_1 of the open-loop gain is extended out to the cutoff frequency ω_c by a factor equal to the loop gain, $A_o B$.

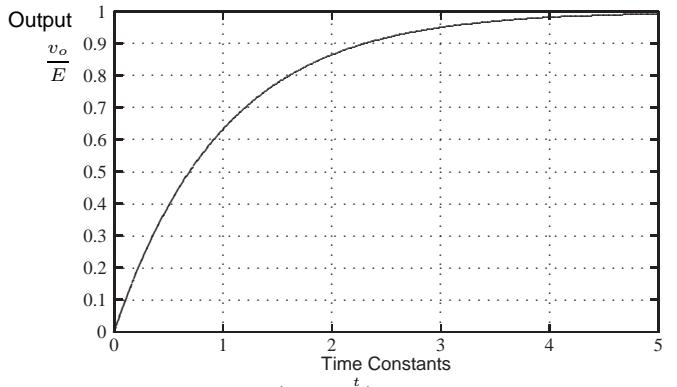


Figure 594: Single Pole Transient Response

The transient response of the closed-loop system is that of a single-pole lowpass filter, the single-time-constant exponential function. The response to a step input is shown in figure 594. The output voltage v_o approaches the final output E in an exponential fashion with a time constant τ where $\tau = 1/\omega_c$.

This is the *small signal* response to a step input and assumes that the output signal is not slew rate limited (section 21.8).

Example

A 741 operational amplifier, with open-loop gain 50×10^3 volts/volt and first break frequency f_1 at 8.74 Hz, is operated with a closed-loop gain of 10. Describe the step response.

Solution

In radian measure, the first break frequency ω_1 is given by:

$$\begin{aligned}\omega_1 &= 2\pi f_1 \\ &= 2 \times \pi \times 8.74 \\ &= 54.8 \text{ radians/sec}\end{aligned}$$

The closed-loop gain is 10 so $B = 1/10$.

Now we can apply equation 1022:

$$\begin{aligned}\omega_c &= A_o B \omega_1 \\ &= (50 \times 10^3) \times (1/10) \times 54.8 \\ &= 274 \times 10^3 \text{ radians/second}\end{aligned}$$

This is the cutoff frequency of the amplifier-lowpass filter system. The time constant τ of this filter is simply the reciprocal of ω_c :

$$\begin{aligned}\tau &= \frac{1}{\omega_c} \\ &= \frac{1}{274 \times 10^3} \\ &= 3.6 \times 10^{-6} \text{ seconds}\end{aligned}$$

That is, the step response will resemble figure 594 with a time-constant of $3.6\mu\text{Seconds}$.

(Caveat: for this to be true, the initial slope of the voltage-time curve in figure 594 must be less than the slew rate of the 741 amplifier, which is 0.5 volts/microsecond.)

23.6 Quadratic Response

In this section, we'll develop an expression for the closed-loop gain of a system of the type shown in figure 595.

First, a simplification: the frequency response of the op-amp will resemble the open-loop gain curve shown as the upper solid line in figure 595. However, it simplifies the algebra if we assume that the frequency response extends upward, in the fashion of an integrator, at frequencies below ω_1 , the dashed line in figure 595.. This is a legitimate approximation because the region of our interest is always well above this frequency.

Then the open-loop gain can be represented by the combination of a constant term A_o with an integrator at crossover ω_1 and a single-pole lowpass filter at ω_2 . (We will use A and G for open-loop gain and closed-loop gain respectively. It should be understood that these variables are functions of frequency $A(s)$ and $G(s)$):

$$\begin{aligned} A &= A_o \times \frac{1}{s/\omega_1} \times \frac{1}{1+s/\omega_2} \\ &= \frac{A_o\omega_1}{s(1+s/\omega_2)} \end{aligned} \quad (1023)$$

To determine the closed-loop frequency response G , we substitute for A in the standard equation:

$$\begin{aligned} G &= \frac{A}{1+AB} \\ &= \frac{\frac{A_o\omega_1}{s(1+s/\omega_2)}}{1+\frac{A_o\omega_1}{s(1+s/\omega_2)}B} \\ &= \frac{A_o\omega_1}{s^2 + \frac{s^2}{\omega_2} + A_oB\omega_1} \end{aligned} \quad (1024)$$

Now what? The final form of G should be of the form [closed-loop-gain] \times [second-order-lowpass-response]. This suggests that G will be equal to $1/B$ at low frequencies, will peak in the region of the resonant (natural) frequency ω_o , and will roll off thereafter at 40db/decade.

The closed-loop gain is $1/B$. The second-order-lowpass function was developed in section 17.4. From equation 720:

$$\frac{e_o}{e_i} = \frac{\omega_o^2}{\left(s^2 + \frac{\omega_o}{Q}s + \omega_o^2\right)} \quad (1025)$$

which can be put in the alternate form:

$$\frac{e_o}{e_i} = \frac{1}{\frac{s^2}{\omega_o^2} + \frac{s}{\omega_o Q} + 1} \quad (1026)$$

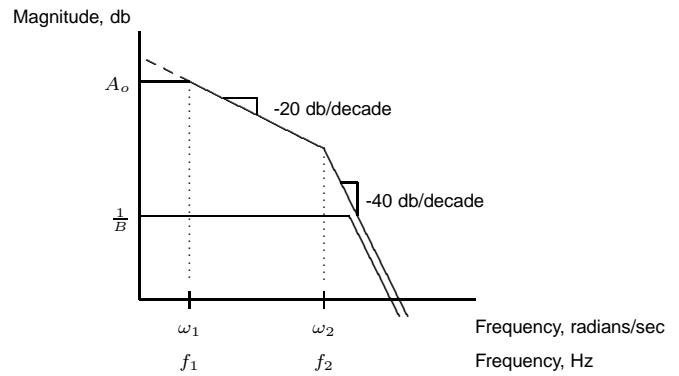


Figure 595: Two Pole Op-Amp Frequency Response

Frequency, radians/sec
Frequency, Hz

With those clues, we can manipulate the transfer-function as follows:

$$\begin{aligned} G &= \frac{A_o \omega_1}{s + \frac{s^2}{\omega_2} + A_o B \omega_1} \\ &= \frac{1}{B} \left(\frac{1}{\frac{s^2}{A_o B \omega_1 \omega_2} + \frac{s}{A_o B \omega_1} + 1} \right) \end{aligned} \quad (1027)$$

To develop expressions for ω_o and Q , we compare equations 1026 and 1027. Then

$$\begin{aligned} \omega_o^2 &= A_o B \omega_1 \omega_2 \\ \omega_o &= \sqrt{A_o B \omega_1 \omega_2} \end{aligned} \quad (1028)$$

We can also express this in terms of the various frequencies in hertz:

$$f_o = \sqrt{A_o B f_1 f_2} \quad (1029)$$

Again, by comparison of equations 1026 and 1027, we have that

$$\omega_o Q = A_o B \omega_1 \quad (1030)$$

so

$$Q^2 = \frac{A_o^2 B^2 \omega_1^2}{\omega_o^2} \quad (1031)$$

Substitute for ω_o^2 from equation 1028:

$$\begin{aligned} Q^2 &= A_o B \frac{\omega_1}{\omega_2} \\ Q &= \sqrt{A_o B \frac{\omega_1}{\omega_2}} \\ &= \sqrt{A_o B \frac{f_1}{f_2}} \end{aligned} \quad (1032)$$

Summarizing

Let's summarize what we've learned to this point.

- Equation 1027 indicates that the closed-loop gain will approximate $1/B$ at low frequencies, peak in the region of the resonant frequency ω_o by an amount proportional to the Q factor, and then roll off at a rate of 40db/decade above ω_o , in a manner similar to the response shown in figure 596 for a second-order lowpass filter.

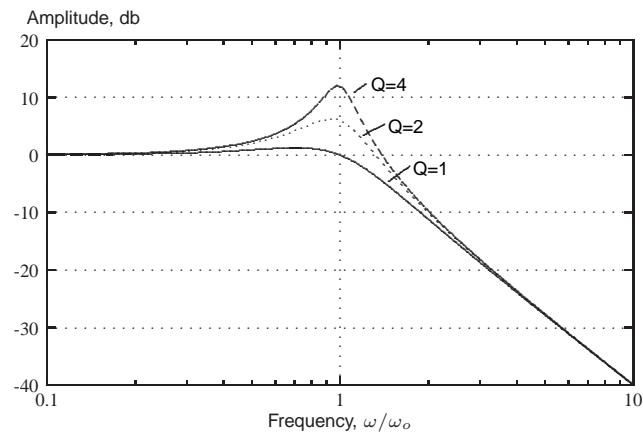


Figure 596: Lowpass Filter, Magnitude Plot

- The resonant frequency is given by equation 1029.
- The Q factor is given by equation 1032. Probably what we're most interested in is the relationship between the low-frequency closed-loop gain (also known as the noise gain) $1/B$ and the Q factor. A Q factor greater than 1 implies peaking in the frequency domain and overshoot and ringing in the time domain, neither of which are desirable.

Notice that a smaller closed-loop gain (and in the worst case, the follower configuration with $1/B = 1$) will cause the largest values of Q factor. This should be remembered as a rule that *the worst case for stability is unity gain feedback*.

Example

Suppose that the frequency characteristic of an op-amp is as shown in figure 597. Determine the values of Q factor and resonant frequency f_o at closed-loop gains of 40db ($\times 100$), 20db ($\times 10$) and 0db ($\times 1$).

Solution

We can use equations 1029 and 1032 to determine the resonant frequency f_o and Q factor at various values of closed-loop gain. For the 40db closed-loop gain case, we have that

$$\begin{aligned} 1/B &= 40\text{db} \\ &= 100 \text{ volts/volt} \end{aligned}$$

so

$$\begin{aligned} B &= 1/100 \\ &= 0.01 \text{ volts/volt} \end{aligned}$$

Calculating the resonant frequency f_o , we have from equation 1029:

$$\begin{aligned} f_o &= \sqrt{A_o B f_1 f_2} \\ &= \sqrt{10^5 \times 0.01 \times 100 \times (100 \times 10^3)} \\ &= 100 \text{ kHz} \end{aligned}$$

The Q factor is

$$\begin{aligned} Q &= \sqrt{\frac{A_o B f_1}{f_2}} \\ &= \sqrt{10^5 \times 0.01 \left(\frac{100}{100 \times 10^3} \right)} \\ &= 1 \end{aligned}$$

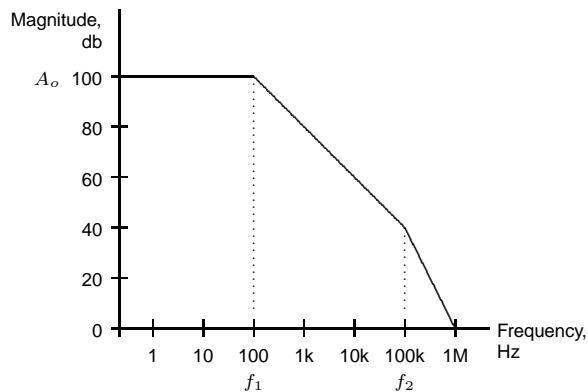


Figure 597: Two Pole System Example

Similar calculations for the 20db and 0db case yield the table of results shown below:

	Closed-loop Gain, G db	Sensor Gain B	Resonant Frequency Hz	Q factor
G_1	40	0.01	100 kHz	1.0
G_2	20	0.1	316 kHz	3.16
G_3	0	1.0	1.0 MHz	10

The resultant closed-loop responses are sketched in figure 598. Notice that

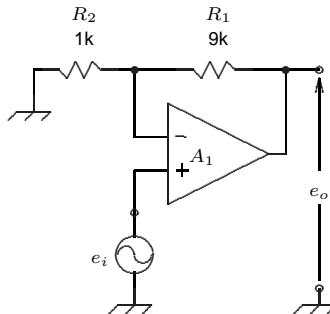
- the resonant frequency corresponds (approximately) to the frequency where the closed-loop gain $1/B$ intersects with the open-loop gain A .
- the Q factor increases as the closed-loop gain decreases. For unity gain (voltage follower), the Q factor is 10, which would lead to intolerable overshoot and ringing to a step input.
- as it currently stands, this op-amp is not suitable for use at closed-loop gains less than 40 db (100 volts/volt) because of the peaking in the region of the resonant frequency at lower values of closed-loop gain.
- the magnitude of the peak is given by

$$20 \log_{10} Q$$

For example, a Q value of 4 results in a 12db peak in the frequency response.

23.7 Exercises

1. The amplifier shown below has a gain-bandwidth product of 1MHz and a slew rate of 0.5 volts/ μ Sec.



- (a) Assuming that the open-loop gain must be at least 10 times the closed-loop gain, what is the maximum operating frequency of this circuit?
- (b) What is the largest sine wave in volts peak-peak that this circuit can amplify at a frequency of 20kHz without inducing slew-rate distortion into the waveform?

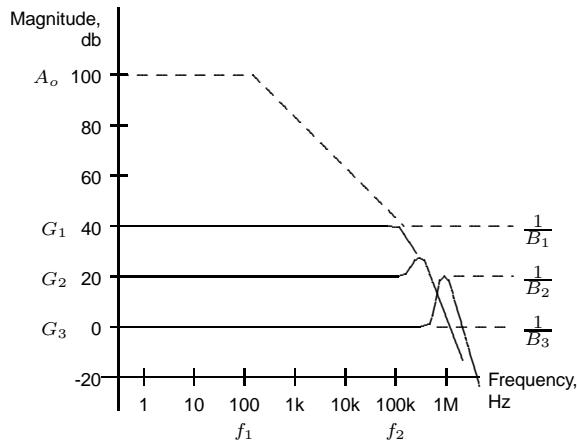
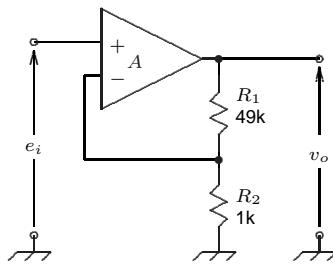


Figure 598: Closed Loop Responses

2. An operational amplifier has an open-loop gain of 80db and a gain bandwidth product of 1MHz.
- Draw the amplitude and phase response of the open-loop gain A_{ol} , showing the 3db cutoff frequency and the crossover frequency.
 - The circuit shown below uses this op-amp:



- What is the closed-loop gain $G = v_o/e_i$ for a sine wave of frequency 10Hz?
- What is the closed-loop gain $G = v_o/e_i$ for a sine wave of frequency 10kHz?
 - Draw the amplitude and phase response of the closed-loop gain G , showing the 3db cutoff frequency and the crossover frequency.

24 Stability

24.1 Introduction

In its natural state, the op-amp open-loop frequency response contains several poles (break frequencies) that correspond to phase shifts (figure 586 on page 692). In a negative feedback system, these phase shifts add to the total phase shift around the negative feedback loop. If the total phase shift around the loop reaches 360° and the total loop gain is greater than one, the system will oscillate.

As a consequence, such an op-amp is highly likely to oscillate in even the simplest of circuits. The Fairchild 709, introduced in 1965, had this characteristic and required two capacitors and a resistor external to the device to compensate for these phase shifts and prevent oscillation.

With some sacrifice in high-frequency performance it is possible to compensate the op-amp so that it is much simpler to use. In the 741 op-amp, introduced in 1968, an on-chip capacitor causes the frequency response to roll off at a constant rate until the gain is unity, figure 587. This *dominant pole compensation* technique makes the op-amp stable in most applications and as a result greatly simplifies the engineering design process. In most applications external components are not required for compensation.

Most modern op-amps use dominant-pole compensation. When that is the case, it is straightforward to predict both the closed-loop frequency response and the small-signal transient response of a particular op-amp circuit. This was discussed in section 23.5.

Even though dominant-pole compensation reduces the tendency for an op-amp to oscillate, it is still important to understand the mechanism of oscillation.

With the preceding ideas in hand, this section shows examples of oscillation-related problems, and their solutions.

- Section 24.2 emphasises the importance of *decoupling* in stabilizing an op-amp circuit. Op-amps should always be decoupled with a capacitor at each power supply terminal.
- The *phase margin* of a negative feedback system, discussed in section 24.3, is an important measure of its stability.
- A quick assessment of stability can be made on the basis of the *intercept rule*, in section 24.4.
- Section 24.5 shows how the noise gain can be adjusted to improve stability.
- A small-signal op-amp is frequently used in conjunction with a buffer stage to provide additional power output into the load. Depending on the frequency characteristics of the op-amp and buffer, the combination may be unstable. The cause and cure are described in section 24.6.
- Input capacitance can also have a destabilizing effect. It is an important effect to understand, especially in high-frequency op-amp circuits. Section 24.7.
- The *lag-lead* network of section 24.8 is a useful tool in compensating a negative feedback system.
- A capacitive load can destabilize an operational-amplifier circuit. The cause and cures are described in section 24.9.

24.2 Decoupling

A signal can move from the output of a circuit back to the input by means of the stray capacitances, inductances and resistances of the *hidden schematic* [220]. Under the right conditions, this causes an undesired oscillation. The process of blocking these unwanted paths is called *decoupling*, that is, removing the coupling between one part of the circuit and another.

Decoupling is particularly critical at high frequencies and at high gain. High frequencies couple well through stray capacitances. Wires and capacitors that should be good conductors for AC become inductive and can approach an open circuit. With a high forward gain, the path from output back to the input must have high attenuation to keep the loop gain less than unity, and oscillation from occurring.

We discuss the subjects of grounding, shielding, decoupling and circuit layout at greater length in section 34. However, there is one critical rule that should always be followed with respect to an operational amplifier: *decouple the power supply leads*²⁰².

An illustrative diagram is shown in figure 599.

Assume first of all that the decoupling capacitor C_D is not present. If the amplifier A drives some signal into the output transistor Q , then it will drive current into the load resistance R_L . That will be reflected in a transient drop of voltage across the wiring inductance L_W .

This transient then propagates through the coupling network to the input of the amplifier. If there is a condition where the loop gain via the coupling network and the amplifier is larger than one and the phase shift approaches 360° around the loop sustained oscillations will occur.

Now assume that the decoupling capacitor C_D is added to the circuit. An ideal capacitor will act as a constant-voltage device at the power supply terminal. It reduces the level of the AC signal at the power supply terminal of the device and unintentional feedback via the coupling network. This reduces the probability of oscillation occurring.

In general, the value of capacitance is not particularly critical, and a value in the region 10nF to 100nF is effective for op-amps with gain bandwidths in the 1MHz region. However, there are some caveats:

- A capacitor inevitably includes a certain amount of series inductance and resistance. The inductance and resistance are not shown (they are elements in the hidden schematic) but they do affect the operation of the

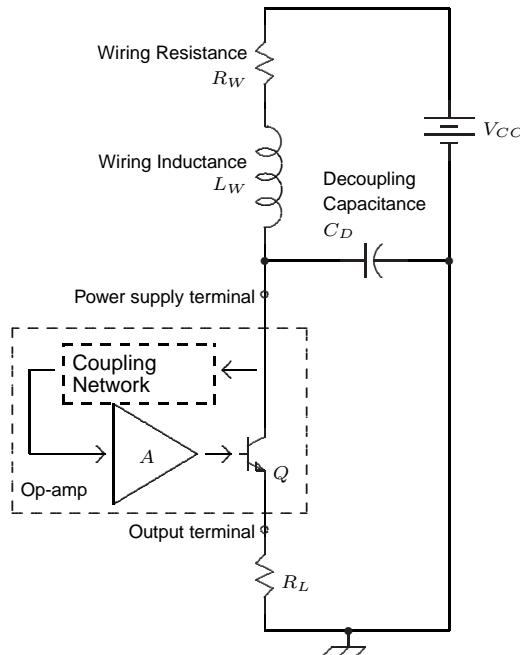


Figure 599: Single Supply Decoupling

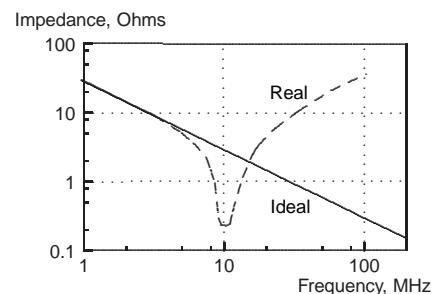


Figure 600: Capacitive Reactance vs Frequency

²⁰²This is equally important for most integrated circuits, regardless of type, and their discrete component equivalents as well.

capacitor. An ideal capacitor decreases in reactance with increasing frequency. A real capacitor resonates at some frequency with its series inductance. Above that resonant frequency, the so-called capacitor behaves as an inductor and its reactance increases with frequency [221].

The effect of this is shown in figure 600 (redrawn from [221]), which shows the reactance vs frequency of an ideal and real 5nF capacitor. The resonant frequency for this capacitor is about 10MHz. Above that frequency, the reactance increases due to the parasitic series inductance of the capacitor.

- A larger capacitance than 100nF is not an acceptable substitute because the resonant frequency of a capacitor decreases as capacitance increases. So, for example, a $10\mu\text{F}$ capacitor is not likely to work as well as 100nF.
- A larger capacitance in parallel with a smaller capacitance will improve the effectiveness of the decoupling. For example, a 100nF ceramic capacitor is often parallelled with a $10\mu\text{F}$ electrolytic. The larger capacitor is effective at lower frequencies, so there is a broadbanding effect. As well, the greater series resistance of the larger capacitor will tend to provide a damping effect on any resonances that occur.
- Lead lengths are critical, and should be kept to an absolute minimum. A bypass²⁰³ capacitor 6 centimetres from the IC package will have no effect at all, because of the intervening wiring inductance. The bypass capacitor must be as close as possible to the IC package.
- For very high-frequency op-amps (those with a GBP in the order of 100MHz, for example), bypassing becomes critical. Stray capacitance and inductance have a much more dramatic effect at these high frequencies. The manufacturer's data sheets will advise on specific bypassing requirements, and will indicate a suitable printed circuit board layout, which is also important. In general, the bypass capacitances should be surface-mount devices to keep lead lengths to an absolute minimum. When prototyping, the standard protoboard device will have excessive stray capacitance, and some other method of prototype construction will be required, as described in [222], [223].

Reference

Reference [224] is a useful guide for the application of decoupling capacitors.

24.3 Phase Margin and Stability

In section 23.2, we showed how the op-amp can be regarded as a series of stages A, B and C that are coupled by lowpass networks, as shown in figure 601(a). The effect of these lowpass filters is to break the open-loop gain downward, first at f_1 to -20db/decade and then at f_2 to -40db/decade, as shown in figure 601(b).

We also showed that the closed-loop gain $1/B$ may be plotted on this same graph, as shown in figure 601(b). The difference between the open-loop gain A and the closed-loop gain $1/B$ is equal to the loop gain AB . It is the loop gain, we will show, that determines the stability of the op-amp circuit.

Since the open-loop gain A is a function of frequency, we could write it as $A(\omega)$ or $A(s)$. For notational convenience, we'll stick to A .

²⁰³The terms *bypass* capacitor and *decoupling* capacitor are treated in this section as being synonymous. Bypass implies that a problematic signal (which might cause oscillation) is routed to ground, bypassing the circuit. Decoupling implies that the effect of the capacitor is to separate the operation of different sections of the circuit, thereby preventing oscillation.

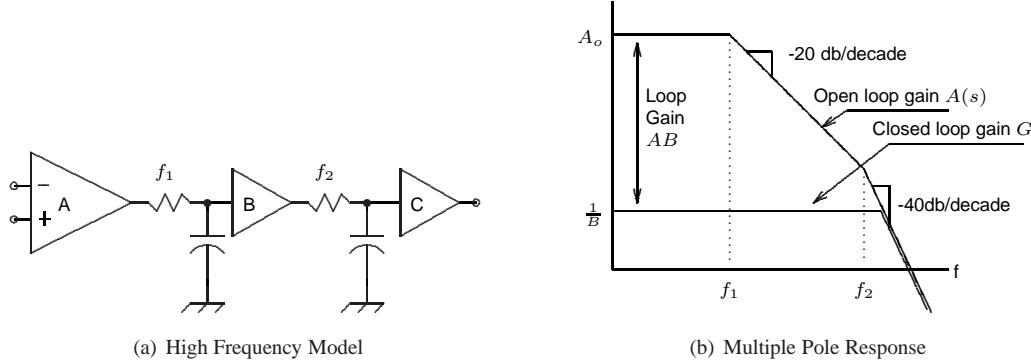


Figure 601: Op Amp Open Loop Response

Loop Gain and Stability

We can derive a Bode plot of the magnitude of loop gain AB from figure 601(b). It's shown in the upper half of figure 602.

To determine the phase of the loop gain, recall from section 11.4 on Bode plots that each break frequency downward adds another -45° of phase shift at that frequency, and ultimately an additional -90° . Putting all this together, we have the phase plot for the loop gain shown in the lower half of figure 602.

The circuit will oscillate if the loop phase becomes 180° while the gain is still greater than unity. At high frequencies the total phase shift of this circuit loop gain approaches -180° . In figure 602, the phase is within about 22° of the dreaded 180° when the gain passes through 0db. This phase difference is known as the *phase margin* of the system and is given the symbol ϕ_m .

The phase margin is a useful measure of the stability of the system. As the phase margin approaches zero, the system becomes progressively less stable.

Figure 603(a) shows what happens to the closed-loop frequency response as the phase margin decreases. The closed-loop gain $1/B$ shows peaking in the region where it intersects with the open-loop gain A . Figure 598 on page 704 shows this effect.

The optimum Q factor is often considered to be $\sqrt{1/2} = 0.707$. This corresponds to a phase margin of 65° . In figure 602, the phase margin is about 22° , so the performance is not optimum and the closed-loop frequency response would show significant peaking.

The relationship between phase margin and Q factor is shown in figure 603(b) (adapted from [79] and [87]).

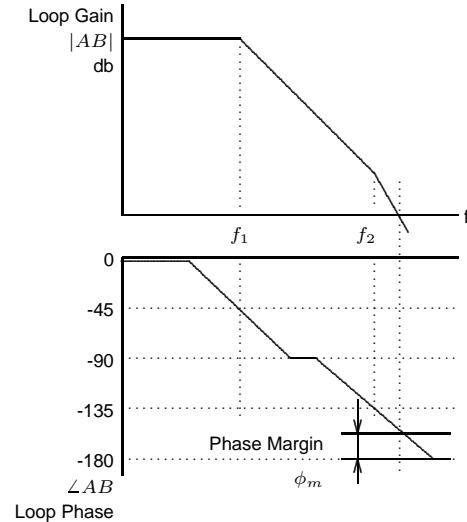


Figure 602: Loop Gain and Phase

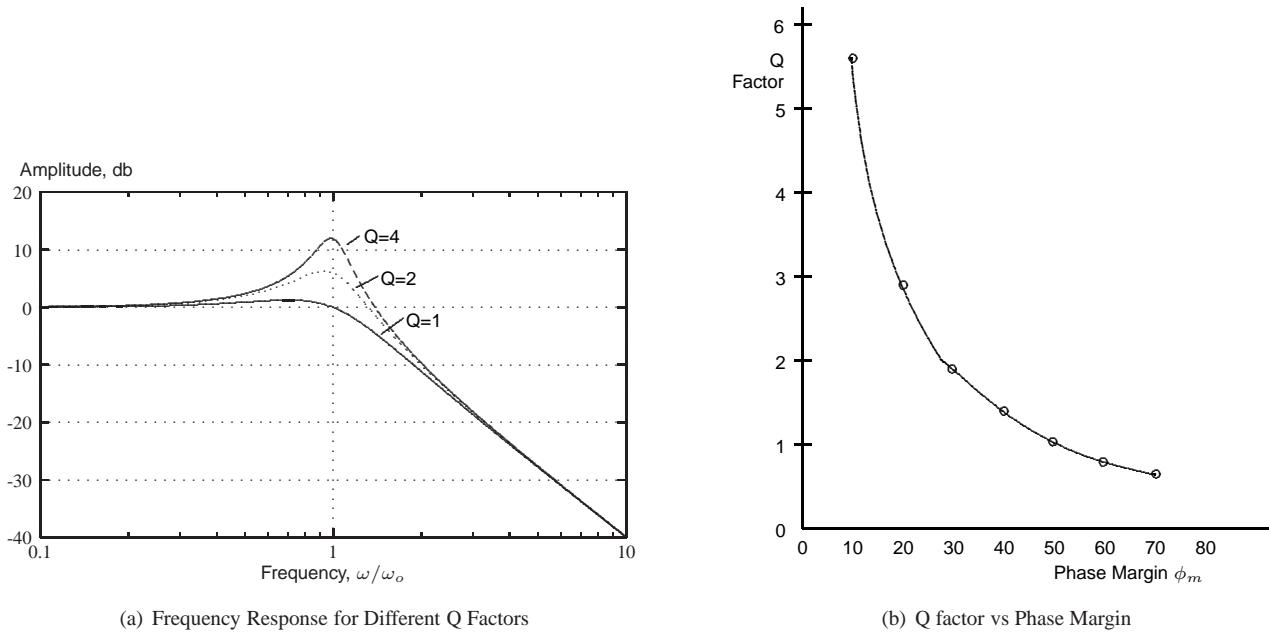


Figure 603: Q Factor, Phase Margin and Frequency Response

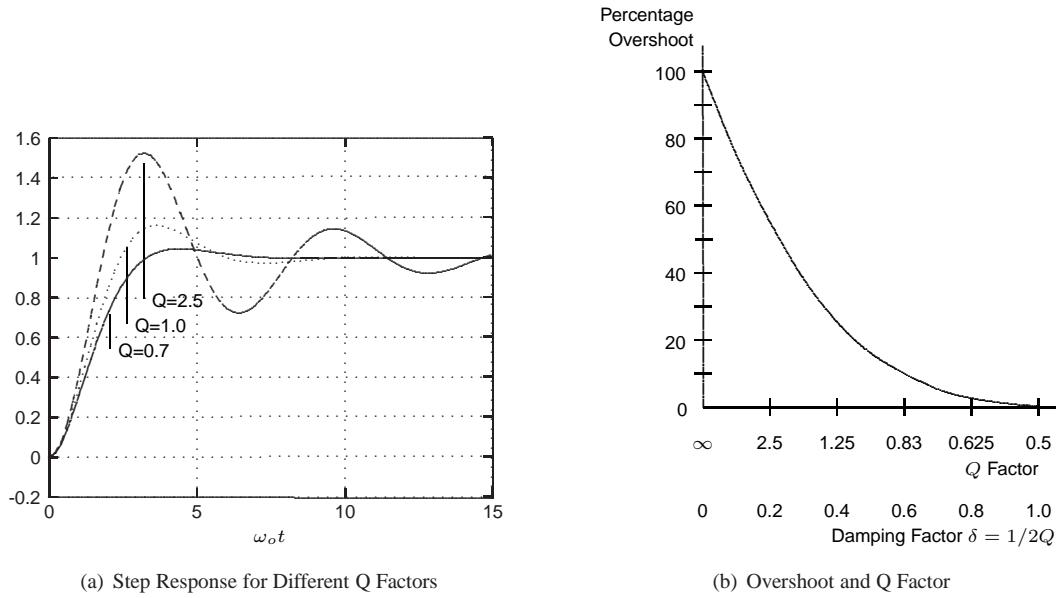


Figure 604: Q Factor and Step Response

Phase Margin and Transient Response

It is possible to measure peaking in the closed-loop frequency response and thereby relate that to the phase margin and the stability of the system. However, it's usually simpler to apply a step input to the system and extract this

same information from the transient response at the output. As the closed-loop frequency response shows greater peaking around the cutoff frequency with smaller phase margin and higher Q, the step response shows greater overshoot and ringing.

The effect of Q factor on transient response is shown in figure 604(a). The relationship between transient overshoot (in percent) and Q factor is shown in figure 604(b).

Phase Margin and Closed Loop Gain

Now let us consider the effect of closed-loop gain on phase margin.

Increasing the closed-loop gain $1/B$ reduces the loop gain AB . Consequently, referring to figure 602, the loop gain will cross through the zero db axis at a lower frequency. The loop phase is unchanged. As a result the phase margin is increased. This leads us to two important concepts:

- For a given operational amplifier, the phase margin is smallest at the smallest value of closed-loop gain, which is unity (the voltage follower circuit).
- If the op-amp is stable at unity gain, it will be stable at all other closed-loop gains.

24.4 The Intercept Rule

Now consider what happens if the closed-loop gain is increased so that it intersects the open-loop curve between f_1 and f_2 rather than after f_2 (figure 605). In that case, the loop gain will include one break frequency, f_1 , before it reaches zero db. Then, if the intercept point is before the second break frequency, the maximum phase of the loop gain will be -135° , corresponding to a phase margin of 45° . The system will be stable. Whether the resultant phase margin, gain peaking, overshoot and ringing are acceptable will depend on the application.

Consequently, we can formulate the general rule:

A negative feedback system will be stable if the noise gain $1/B$ intersects the open-loop gain A in such a way that the slope changes by 20db/decade. It will be unstable if the slope changes by 40db/decade or more.

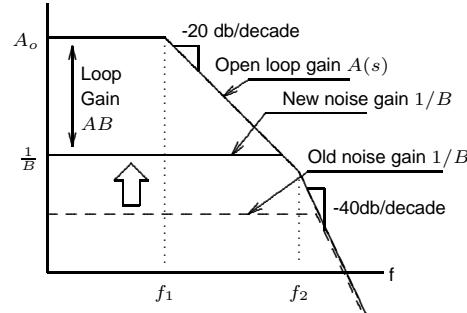


Figure 605: Increased Closed Loop Gain

Example: Differentiator Stability

The op-amp differentiator circuit shown in figure 606 illustrates the intercept rule.

As we indicated in section 13.12, the op-amp differentiator is potentially unstable.

The noise gain rises at a rate of +20db/decade and intersects the op-amp open-loop gain, which has a slope of -20db/decade. Consequently, based on the stability rule given above the change in slope is -40db/decade and the circuit is liable to be unstable.

The circuit can be stabilized by modifying it as shown in figure 607(a), with the addition of resistor R_i . The effect of R_i on the Bode plot is shown in figure 607(b).

The effect of R_i is to flatten out the noise gain beyond break frequency f_3 . At the intersection point, the change in slope is now only from 0db/decade to -20db/decade, a change of -20db/decade, so the circuit should be stable.

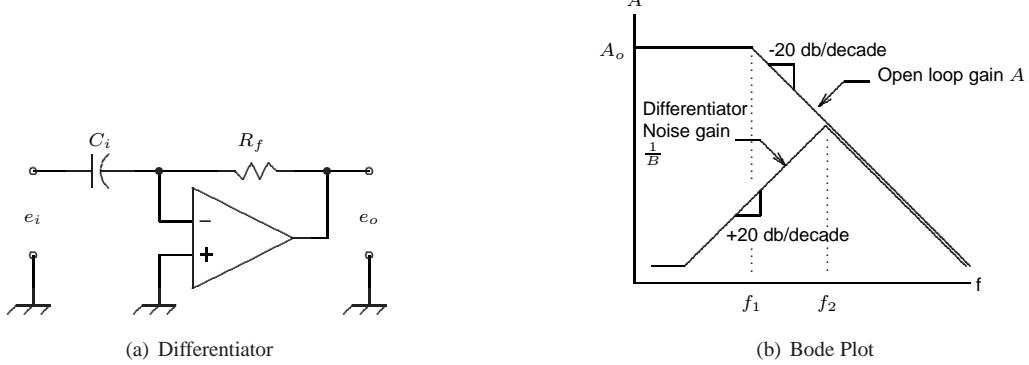


Figure 606: Differentiator

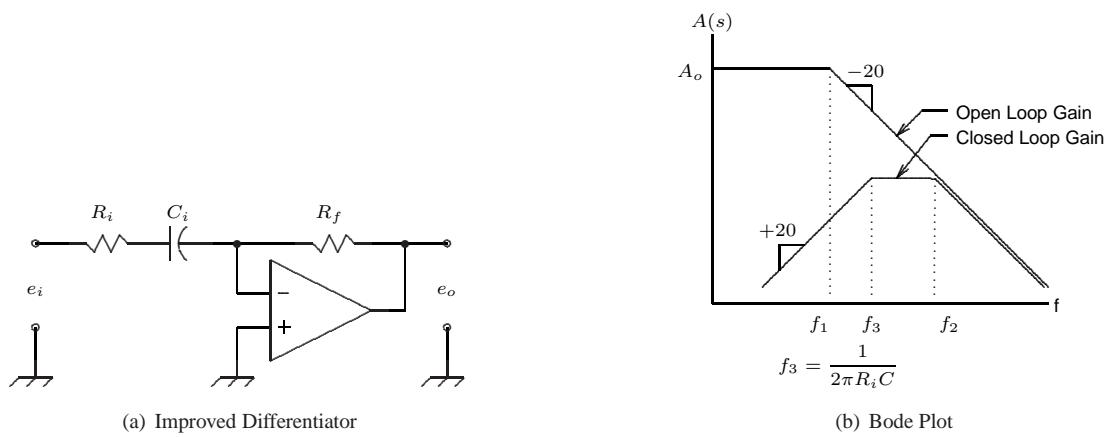
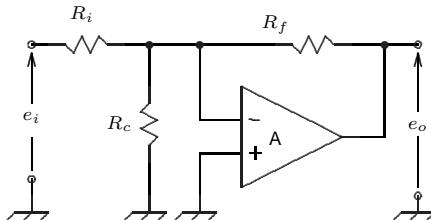
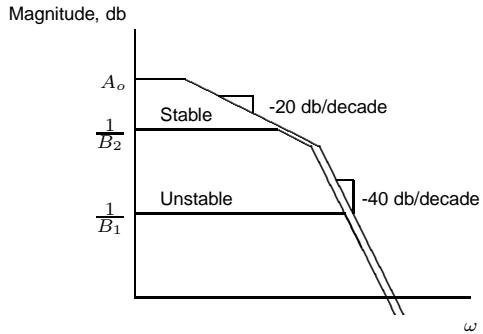


Figure 607: Improved Differentiator

24.5 Enhanced Noise Gain



(a) Circuit



(b) Open and Closed Loop Gain

Figure 608: Reduced Noise Gain

This method of compensation is based on the idea illustrated in figure 608(b).

Suppose that the desired gain of the system puts the noise gain at the lower of the two levels, $1/B_1$, shown in the figure. According to the intercept rule, the noise gain closes with the open-loop gain with a slope change of 40db/decade, and the system will be unstable. However, if we could move the noise gain to the upper level $1/B_2$, then the interception slope change is only 20db/decade and the system would be stable.

An inverting amplifier with a suitable circuit modification, the addition of the compensation resistor R_c , is shown in figure 608(a), redrawn in figure 609.

This resistor appears across the inputs of the op-amp, so the voltage across it is very nearly zero. Consequently, it conducts very little current and has no significant effect on the closed-loop gain of the circuit, which is the usual:

$$\frac{e_o}{e_i} = -\frac{R_f}{R_i} \quad (1033)$$

However, insofar as the noise gain and stability are concerned, the situation is quite different. In the original circuit the noise gain (call it B_1) is

$$\begin{aligned} B_1 &= \frac{e_f}{e_o} \\ &= \frac{R_i}{R_i + R_f} \end{aligned} \quad (1034)$$

In the compensated circuit, resistor R_c appears in parallel with R_i , and so the noise gain (call it B_2) is

$$\begin{aligned} B_2 &= \frac{e_f}{e_o} \\ &= \frac{R_i \parallel R_c}{(R_i \parallel R_c) + R_f} \end{aligned} \quad (1035)$$

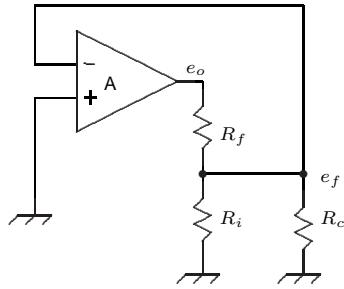
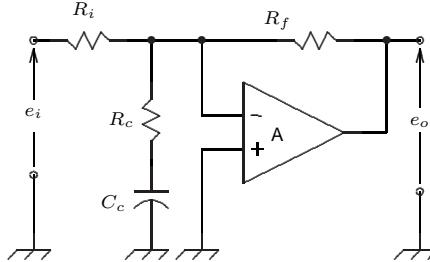


Figure 609: Stability Analysis Circuit

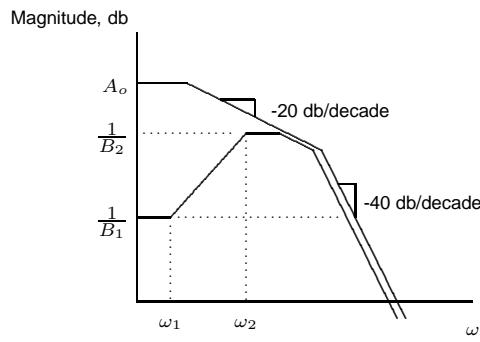
Consequently, the new noise gain is less than the original, and the trace moves upward on the plot of gain vs frequency.

There has to be a catch to this, and there is: reducing the noise gain reduces the loop gain AB . As a consequence, the op-amp is less effective in reducing output resistance, disturbances and other properties that depend on the loop gain. However, like the nesting Russian dolls, there is a fix for this shortcoming in the fix. The compensating resistor R_c is put in series with a capacitor C_c , as shown in figure 610(a). The effect on the closed-loop gain is shown in figure 610(b).

Then:



(a) Circuit



(b) Open and Closed Loop Gain

Figure 610: Improved Circuit

- At low frequencies (where phase margin and stability are not an issue), the capacitor C_c is an open circuit and the compensating network has no effect. As a result the loop gain is restored to its original value and low frequency values of output resistance and disturbance reduction are unaffected from the original circuit.
- At high frequencies, where it is needed, the capacitor appears as a short circuit and the compensation resistor reduces the noise gain, as required.

The Break Frequencies

Now let us determine where the break frequencies, ω_1 and ω_2 in figure 610(b), are located on the noise-gain curve.

A frontal attack on this network gets ugly very quickly. However, if the network is Thevenized, as shown in figure 611, it's relatively straightforward. The Thevenin parameters are:

$$\begin{aligned} e_{oc} &= K e_o \\ &= \frac{R_i}{R_i + R_f} e_o \end{aligned} \quad (1036)$$

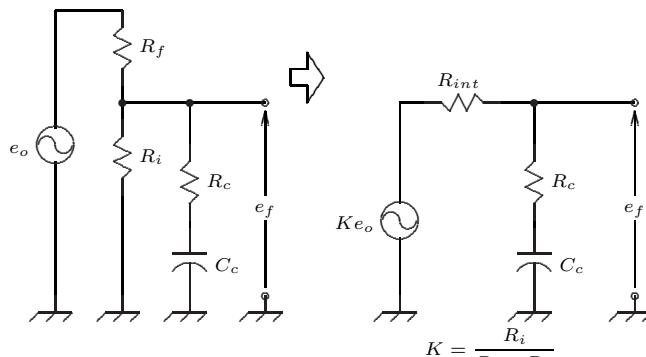


Figure 611: Noise Gain Network

$$R_{int} = R_i \parallel R_f \quad (1037)$$

Consequently, using equation 1036, we can write

$$\begin{aligned} B &= \frac{e_f}{e_o} \\ &= K \frac{e_f}{e_{oc}} \\ &= K \frac{Z_2}{Z_1 + Z_2} \end{aligned} \quad (1038)$$

where

$$Z_1 = R_{int} \quad (1039)$$

$$\begin{aligned} Z_2 &= R_c + 1/sC_c \\ &= \frac{1 + sR_c C_c}{sC_c} \end{aligned} \quad (1040)$$

Substituting for Z_1 and Z_2 from equations 1039 and 1040 into equation 1038, we have:

$$\begin{aligned} B &= K \frac{\frac{1 + sR_c C_c}{sC_c}}{R_{int} + \frac{1 + sR_c C_c}{sC_c}} \\ &= K \frac{1 + sR_c C_c}{1 + s(R_{int} + R_c)C_c} \end{aligned} \quad (1041)$$

At low frequencies (using the terms in figure 608) where the s terms are much less than 1, this reduces to

$$B_1 \approx K \quad (1042)$$

At high frequencies, where the s terms become large compared to 1, we have

$$\begin{aligned} B_2 &\approx K \frac{sR_c C_c}{s(R_{int} + R_c)C_c} \\ &= K \frac{R_c}{(R_{int} + R_c)} \end{aligned} \quad (1043)$$

The Bode plot for this network first breaks at low frequencies at

$$\omega_1 = \frac{1}{(R_{int} + R_c)C_c} \quad (1044)$$

and flattens out again at high frequencies at

$$\omega_2 = \frac{1}{R_c C_c} \quad (1045)$$

Equations 1042 through 1045 are the design equations for the feedback network.

24.6 Output Buffer Stage

The integrated circuit operational amplifier has a very limited output voltage and current capability. As a result, it is common to add a *power gain stage* or *buffer* to the output, to boost the output power capability (section 13.14).

To take advantage of the beneficial effects of negative feedback, the buffer stage is usually included in the overall feedback loop, so that the buffer effectively becomes part of the op-amp forward gain.

An example system is shown in figure 612. The op-amp drives a unity-gain non-inverting buffer. Overall feedback of using 2 equal resistors establishes a non-inverting gain of 2 volts/volt between input and output.

The purpose in life of the output buffer is to supply power gain at significant output currents. The amplifying devices used in such a circuit tend to be physically large and have significant inter-terminal capacitances. Consequently, buffer circuits are inherently relatively slow, that is, have a lower maximum frequency of operation than the driver op-amp. As Williams points out in [225], this leads to instability and oscillations.

The physical explanation is straightforward: the op-amp driver stage can respond more quickly to an error signal than can the buffer stage. Consequently, the op-amp output overshoots its correct value before the buffer stage output is correct. This excessive output from the driver stage then causes the buffer stage to overshoot, which causes the driver to overshoot in the other direction.

There are two possible solutions: speed up the buffer stage or slow down the driver. Usually, the second option is a more practical choice. Slowing down the driver is accomplished by lowering the dominant pole.

Stabilizing a Buffer

Figure 613 shows the system Bode plots for the circuit of figure 612.

First, we need to determine the overall amplitude-frequency response of the amplifier. The op-amp and the buffer are in cascade, so their transfer functions multiply, and their Bode plots add. Figure 613(a) shows the individual Bode plots for the op-amp and the buffer, and the result of adding their amplitude responses.

Notice that the combined result for the op-amp+buffer has two segments at high frequencies: a section that rolls off at 20db/decade and then beyond f_1 a section that rolls off at 40db/decade. That's a warning that the phase curve will ultimately approach -180° , so instability is a possibility.

Figure 613(b) shows the overall transfer function for the op-amp+buffer and the noise gain $1/B$, which is 2 volts/volt (+6db).

The noise gain intersects the op-amp+buffer open-loop gain well beyond the break frequency f_1 , at a point where the open-loop gain is decreasing at 40db/decade. Consequently, according to the intercept rule the system will have an insufficient phase margin and be only marginally stable.

An improved circuit is shown in figure 614(a). (The circuit has been simplified by representing the buffer as

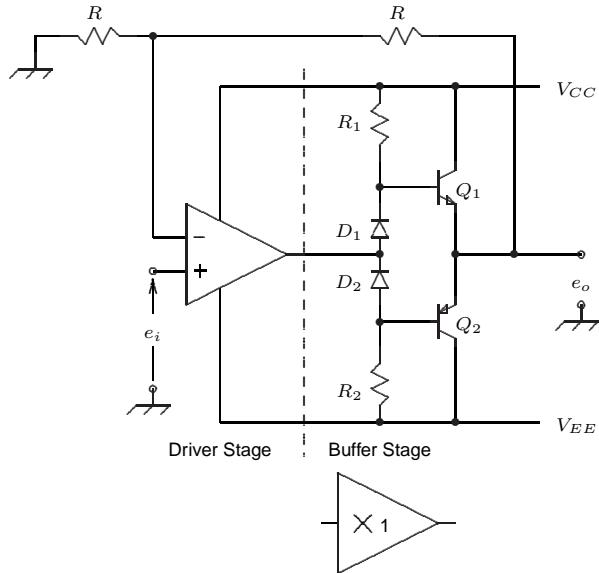


Figure 612: Op-Amp with Buffer Stage

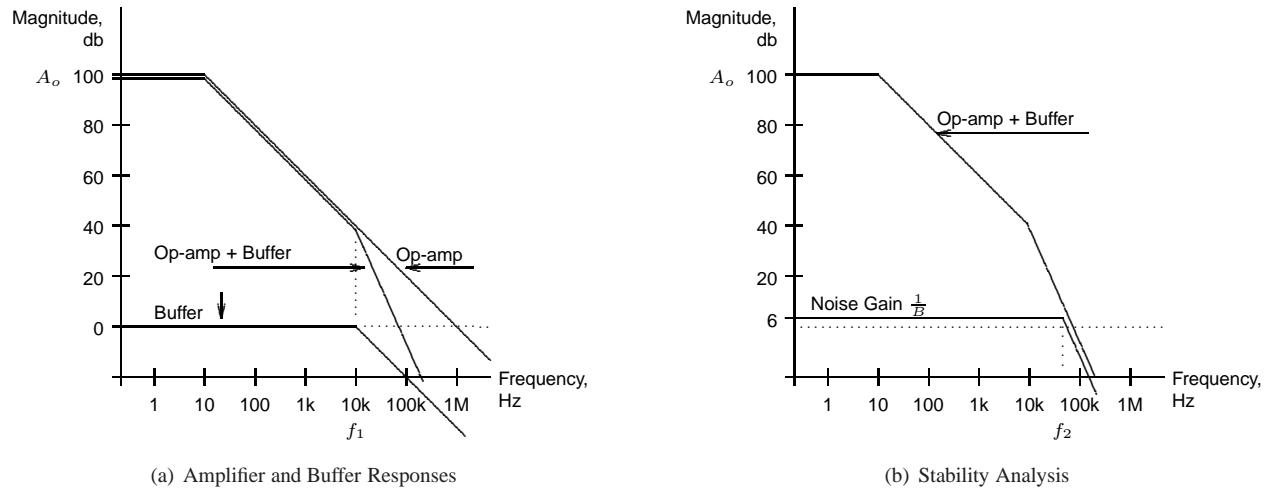


Figure 613: Amplifier and Buffer Bode Plots

an amplifier of unity gain and hiding the power supply connections.) In effect, the op-amp is changed into an integrator, with the addition of feedback (compensation) capacitance C_c . This rolls off the gain of the amplifier at a much lower frequency than previously. This puts the f_2 break frequency below the zero db axis, and the noise gain intersects with the composite plot where it has a slope of -20db/decade. Consequently, this system will be stable.

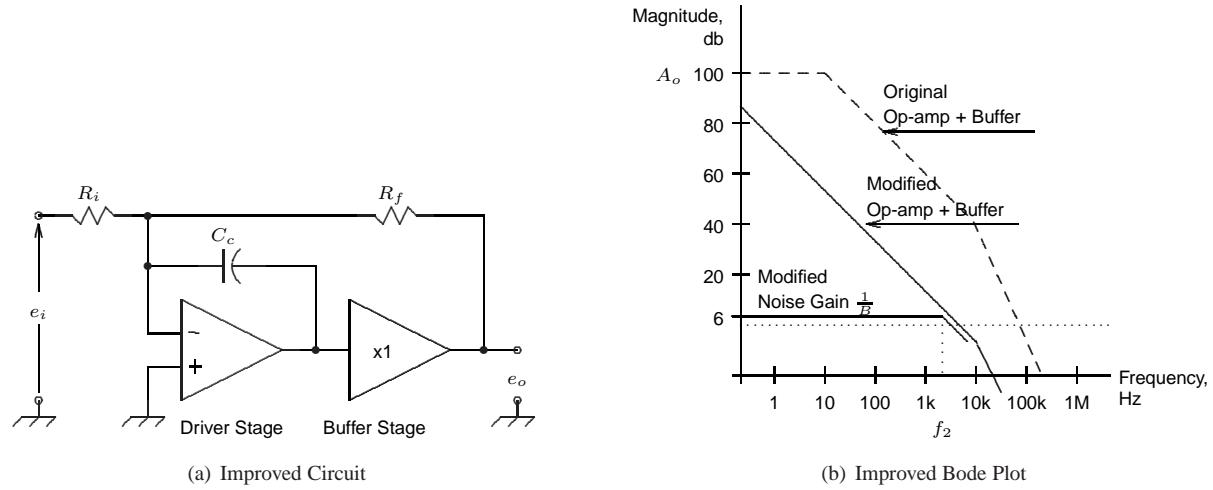


Figure 614: Stabilized Amplifier+Buffer

This system will not be as fast as it could be. The integrator naturally rolls off the gain at high frequencies, so it's a heavy-handed method of compensation. A more sophisticated approach would be to create a lead function (a transfer function zero) at frequency f_1 in figure 613(a). This lead would cancel the lag due to the buffer response, and straighten out the composite response so that it remains at the same slope down to the zero-db axis. Then the intersection point f_2 , where the noise gain and open-loop gain meet, would be at a much higher frequency, and the overall system would have faster response.

24.7 Input Capacitance

A small input capacitance appears between the inverting and non-inverting terminals of an operational amplifier. This is a so-called *stray* capacitance which is a by-product of the construction of the integrated circuit and the layout of the circuit board. In figure 615(a), this is represented by the capacitance C_i . Capacitance C_i is one of the components of the *invisible schematic*, so it does not appear on drawings, but it is there nonetheless. The precise value is dependent on the layout of the circuit, but will typically be less than 10pF.

Figure 615(b) shows the circuit redrawn for analysis of the stability. The input capacitance interacts with the Thevenin resistance of R_i and R_f to create an RC lowpass network that adds a phase lag in the feedback network path. This contributes to the phase shift around the loop and reduces the stability of the circuit.

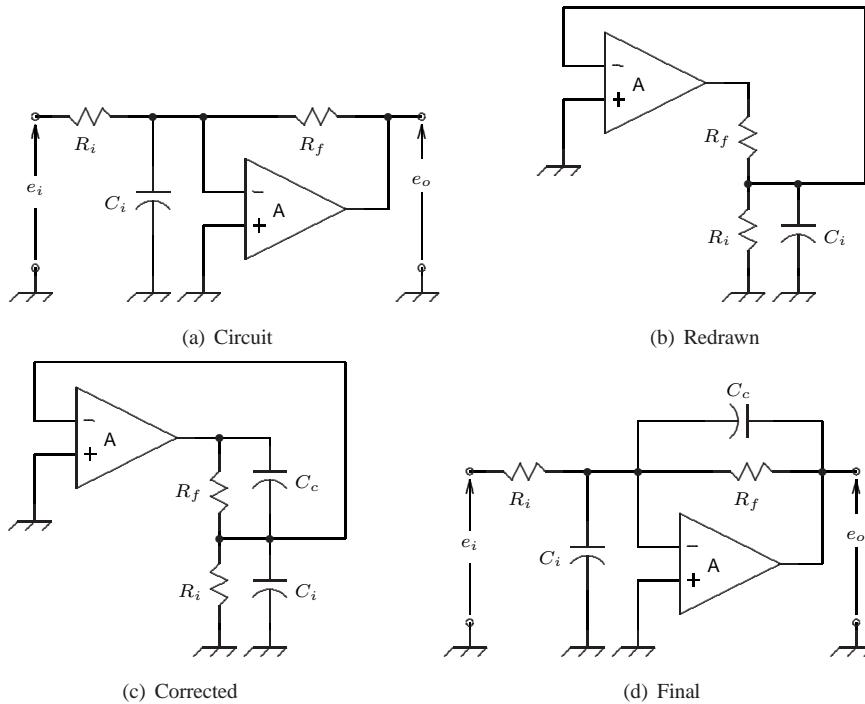


Figure 615: Input Capacitance

The cure is very simple: add a small capacitance C_c in parallel with the feedback resistance. With the correct value of feedback capacitance, the resistors and capacitors form a compensated attenuator, for which the response is flat with frequency (section 4.15).

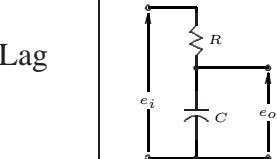
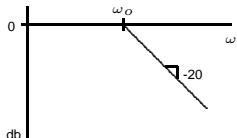
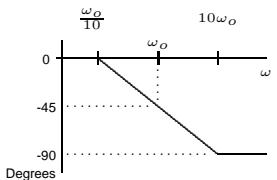
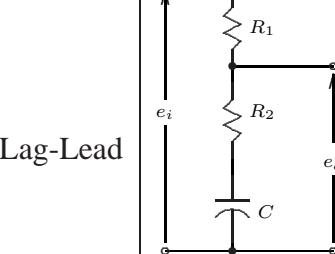
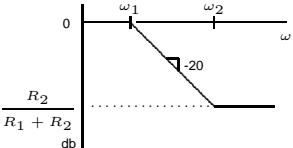
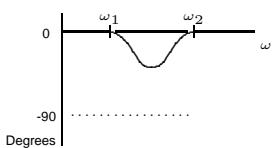
The gain of the compensated attenuator can be made to be independent of frequency when (using the circuit names from figure 615(b)):

$$R_f C_c = R_i C_i \quad (1046)$$

This provides us with a design equation for feedback capacitance when the resistances and input capacitance are known. However, because the input capacitance is usually not known exactly it is difficult to predict the exact value for C_c . It's much simpler to drive the op-amp with a square wave and adjust C_c for the optimum square wave response. If the value is too small the output square wave will exhibit overshoot and ringing. If the value

is too large, the output will be overdamped. The value of C_c is adjusted until the rise time is as small as possible without objectionable overshoot.

24.8 Frequency Compensation: The Lag-Lead Network

Name	Circuit	Transfer Function	Amplitude	Phase
Lag		$\frac{e_o}{e_i} = \frac{1}{1 + s/\omega_o}$		
Lag-Lead		$\frac{e_o}{e_i} = \frac{1 + s/\omega_2}{1 + s/\omega_1}$		

Break Frequencies:

$$\omega_o = \frac{1}{RC}$$

$$\omega_1 = \frac{1}{(R_1 + R_2)C}$$

$$\omega_2 = \frac{1}{R_2C}$$

In this section, we'll look at a technique for stabilizing a negative feedback system, by modifying the shape of the loop gain magnitude and phase. There are several possible ways to accomplish this. We will focus on the lag-lead network.

The circuit for a Lag-Lead Network is shown in the *Circuit* column of the second row of the table above. Its behaviour in magnitude and phase is compared in the table with the Lag Network.

- The magnitude response of the lag network rolls off forever at 20db/decade. The lead-lag network response rolls off for an interval of frequencies but then flattens out again above an upper frequency.
- A lag network has a phase shift of 0° at low frequencies and 90° at high frequencies. The lead-lag network generates a lagging phase shift over an interval of frequencies, but then returns back to 0° phase shift at high frequencies.

The maximum extent of the phase lag, between the two break frequencies, depends on their separation. A wider separation of break frequencies results in a larger excursion in phase.

We'll look at the development of the transfer function and then show why this is useful.

Transfer Function of the Lag-Lead Network

The transfer function of the network begins with the voltage divider:

$$\frac{e_o}{e_i} = \frac{Z_2}{Z_1 + Z_2} \quad (1047)$$

where Z_1 and Z_2 are the impedances in the upper and lower halves of the divider. In this case,

$$\begin{aligned} Z_1 &= R_1 \\ Z_2 &= R_2 + 1/sC \end{aligned}$$

Substituting in equation 1047 for Z_1 and Z_2 , we have

$$\frac{e_o}{e_i} = \frac{R_2 + 1/sC}{R_1 + R_2 + 1/sC} \quad (1048)$$

$$= \frac{1 + sR_2C}{1 + s(R_1 + R_2)C} \quad (1049)$$

The Bode plot of magnitude and phase for the network is shown in figure 616.

The numerator of equation 1049 is a first-order lead that breaks upward at ω_2 , where

$$\omega_2 = \frac{1}{R_2C} \quad (1050)$$

The denominator is a first-order lag that breaks downward at ω_1 , where

$$\omega_1 = \frac{1}{(R_1 + R_2)C} \quad (1051)$$

The gain of the transfer function steps down by an amount equal to the resistive attenuation of the network. The phase goes from zero to some negative value and back to zero again. This effect is often useful when an adjustment is required to a phase curve, and the maximum value of the phase bump is placed where the adjustment is required.

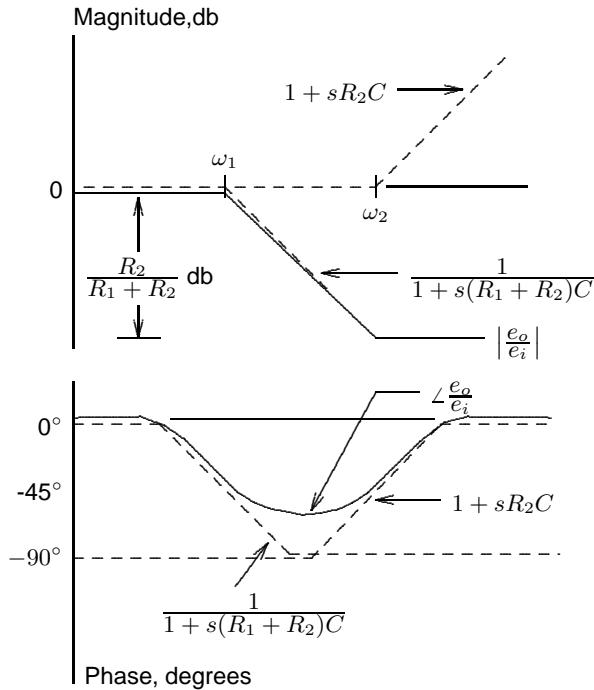


Figure 616: Lag-Lead Network Bode Plot

The transfer function can be further massaged as follows:

$$\begin{aligned}\frac{e_o}{e_i} &= G(s) \\ &= \frac{1 + s/\omega_2}{1 + s/\omega_1} \\ &= \frac{\omega_1}{\omega_2} \left(\frac{s + \omega_2}{s + \omega_1} \right)\end{aligned}\tag{1052}$$

Since ω_2 is larger than ω_1 , we can put

$$\omega_2 = k\omega_1\tag{1053}$$

Then equation 1052 becomes:

$$G(s) = \frac{1}{k} \left(\frac{s + k\omega_1}{s + \omega_1} \right)\tag{1054}$$

Example: Lag-Lead Network

Consider that the loop gain of a system resembles figure 617(a) on page 722. At the point where the magnitude crosses through the zero axis (unity loop gain), the phase angle is about 15° , much less than the 65° required for stability. Now we will consider various methods of compensating the system so that it has a larger phase margin.

- **Attenuator Compensation** If the loop gain is reduced by some factor (by adding a voltage divider to the loop, for example) figure 617(b) shows the result. The magnitude curve is moved downward. The phase curve is unaffected. This improves the phase margin, but it also reduces the loop gain. This is undesirable: larger loop gain brings a number of benefits to the performance of the system.
- **Lag Compensation** The effect of a simple lag network is shown in figure 617(c). This has no effect on the low-frequency open-loop gain, which is desirable. However, the lag network adds phase lag to the loop phase curve, 45° at the corner frequency of the lag, increasing to 90° at high frequencies. The effect is a disaster – the phase margin totally disappears and the system is guaranteed to oscillate.
- **Lag-Lead Compensation** If a lag-lead network is added to the loop, then the loop gain changes as shown in figure 617(d). The break frequencies of the lag-lead compensator are f_3 and f_4 . The magnitude takes a step downward at frequency f_3 and then resumes its normal rate of descent at frequency f_4 , moving the gain crossover frequency to the left.

Now, this does not help the situation if the downward step in gain is accompanied by an additional phase shift that persists up to the crossover frequency. The lag-lead network does put an additional negative frequency shift bump in the phase response. However, if the lag-lead frequencies are positioned correctly, *the effect of that additional lag has worn off by the gain crossover frequency*. Consequently, the lag-lead network introduces attenuation without additional phase shift.

The phase margin in figure 617(d) is only about 45° , still well below the required value, but it's a movement in the right direction.

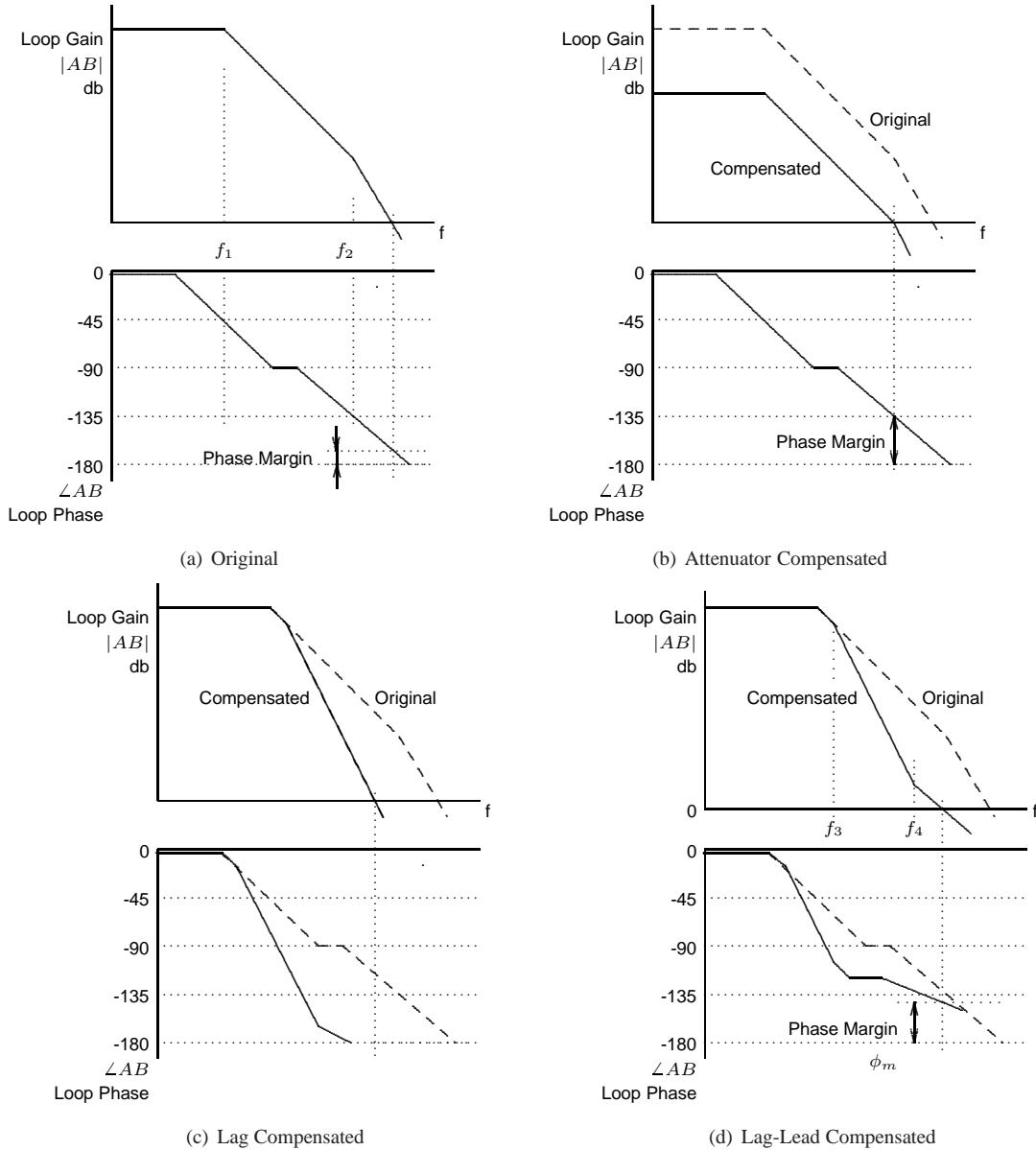


Figure 617: Lag-Lead Compensation

24.9 Capacitive Loading

It is common for an op-amp to be required to drive a capacitive load. For example, the capacitance of RG-174 coaxial cable is 101 pF/metre. A 10 metre length of the cable would present the driving amplifier with a capacitance of 1 nF.

As shown in the model of figure 618(a), this capacitance C_L and the output resistance of the op-amp r_o form a lowpass filter which contributes between 0 and 90° of phase lag, 45° at the corner frequency. This phase lag can

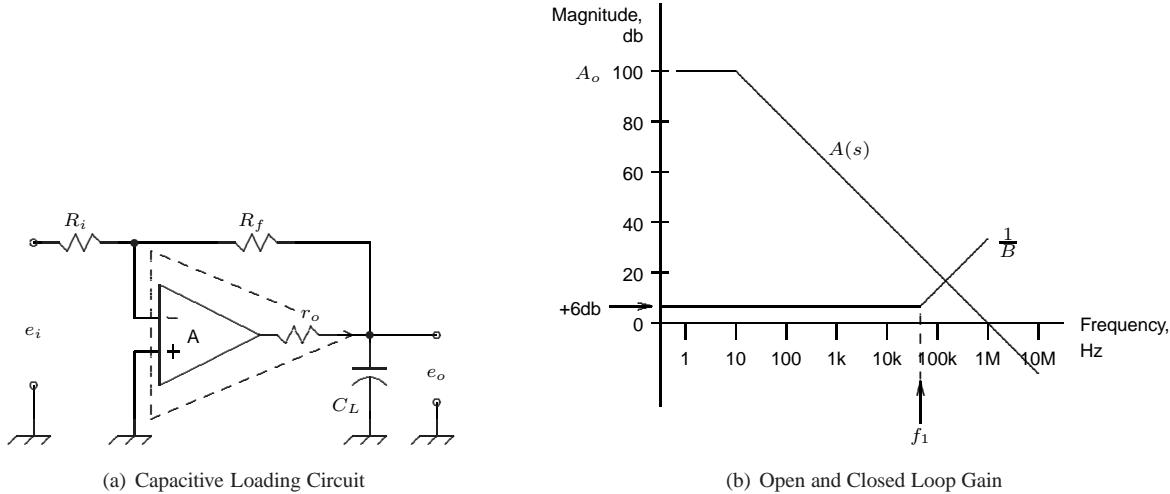


Figure 618: Capacitive Loading

seriously destabililize the op-amp, even if it uses dominant-pole compensation.

This is a relatively simple problem to diagnose. If the oscillations occur or worsen when the load is connected, then load capacitance is the likely culprit.

Components r_o and C_L are part of the *hidden schematic*. They don't appear explicitly in the schematic diagram, but they have an effect on the behaviour of the circuit and must be modelled when analysing the behaviour of the circuit.

The Bode plots for the system are shown in figure 618(b). The op-amp uses dominant-pole compensation, so its open-loop gain A rolls off at 20db/decade down to the gain-bandwidth frequency at 1MHz.

To keep things simple, we'll assume that the gain setting resistors R_i and R_f are equal (although they could have any value), in which case the sensor gain B is 0.5 at low frequencies. Above the corner frequency f_1 created by r_o and C_L , the feedback gain decreases at 20db/decade. Consequently, the closed-loop gain $1/B$ is +6db at low frequencies and then ramps upward at 20db/decade beyond f_1 , as shown in figure 618(b).

This system violates the intercept rule, section 24.4. The change in slope is from +20db/decade to -20db/decade so we can expect a stability problem.

We can confirm this by plotting the loop gain AB , which is the difference between the upper and lower traces in figure 618(b). The magnitude plot of loop gain is shown in the upper half of figure 619. Knowing that each 20db downward slope in the magnitude curve creates a phase lag in the phase curve, we can derive the

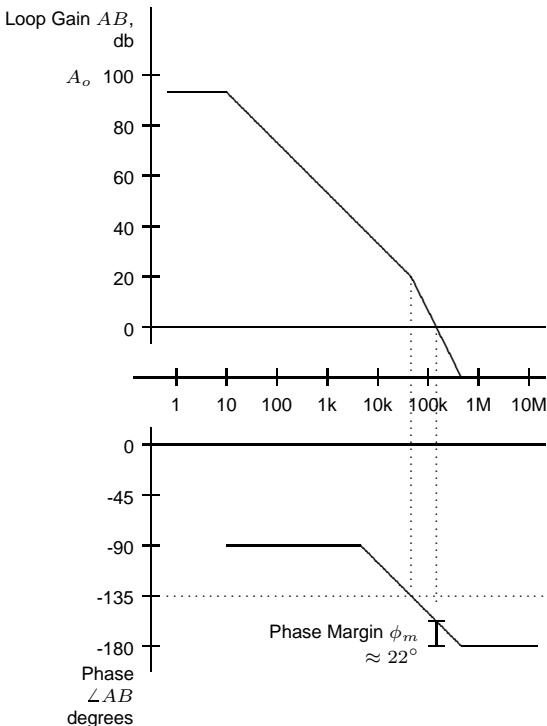


Figure 619: Loop Gain

phase plot shown in the lower half of figure 619. The phase margin ϕ_m is about 22° . This is well below the recommended minimum of 65° . Referring to figure 601 on page 709, the Q factor would be approximately 3, indicating significant peaking in the frequency domain and overshoot in the time domain.

A Caveat

The output resistance of an op-amp is often undefined: it doesn't appear on the op-amp data sheet. It may be highly non-linear, changing values when the op-amp is sourcing and sinking current. And it may rise or fall at high frequencies, becoming an output *impedance* rather than resistance.

Some op-amp datasheets deal directly with the issue of destabilizing load capacitance and suggest specific circuitry to compensate.

For all these reasons, precise calculations of a compensation network are not likely to be possible or yield the desired predictive result. However, it is still useful to review compensating circuits and understand how they function. With that understanding and a starting point for the design, the values can be tweaked in the lab to obtain the desired result.

In the next subsections, we show three different methods of compensation for capacitive loading: *buildout resistance*, *enclosed buildout resistance* and *compensated attenuator*.

Capacitive Loading Solutions: Buildout Resistance

If the requirements are not too demanding, it may suffice to put a small resistor R_s (the *build-out* resistance) in series with the output, as shown in figure 620. The effect of this is to turn the r_o , R_s and C_L network from a simple RC lag to a lag-lead network (see section 24.8).

Substituting the component names from the circuit of figure 621 into equation 1049 on page 720, the transfer function of this lag-lead network will be:

$$\begin{aligned} \frac{e_o}{e_i} &= \frac{1 + sR_2C}{1 + s(R_1 + R_2)C} \\ &= \frac{1 + sR_s C_L}{1 + s(r_o + R_s)C_L} \end{aligned} \quad (1055)$$

A simple choice of values is to make R_s much larger than r_o . (In the absence of other information, a value of 100Ω is often a good starting point for R_s .) Then the numerator and denominator of equation 1055 become approximately equal. The transfer function becomes unity magnitude and zero phase shift. In figure 618(b), the trace for the noise-gain $1/B$ flattens out, and the stability problem is solved.

This is a quick fix to an oscillation problem caused by capacitive loading. However, there are two possible disadvantages:

- The series resistance R_s worsens the output load regulation of the circuit. If the load resistance changes, the load voltage will change.

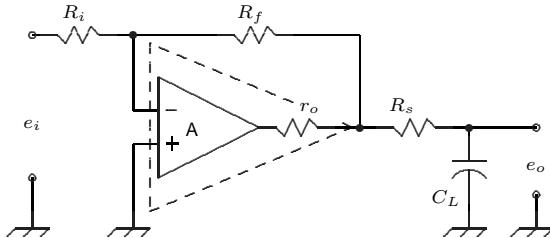


Figure 620: Compensation for Load Capacitance, 1

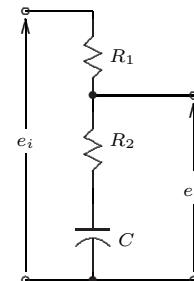


Figure 621: Lag-Lead Network

- The series resistance $r_o + R_s$ and load capacitance C_L form a lowpass circuit. The resultant rise-time of this circuit may be too slow to meet the circuit requirements.

In many practical situations, these are not significant concerns and the buildout resistance is acceptable.

Capacitive Loading Solutions: Enclosed Buildout Resistance

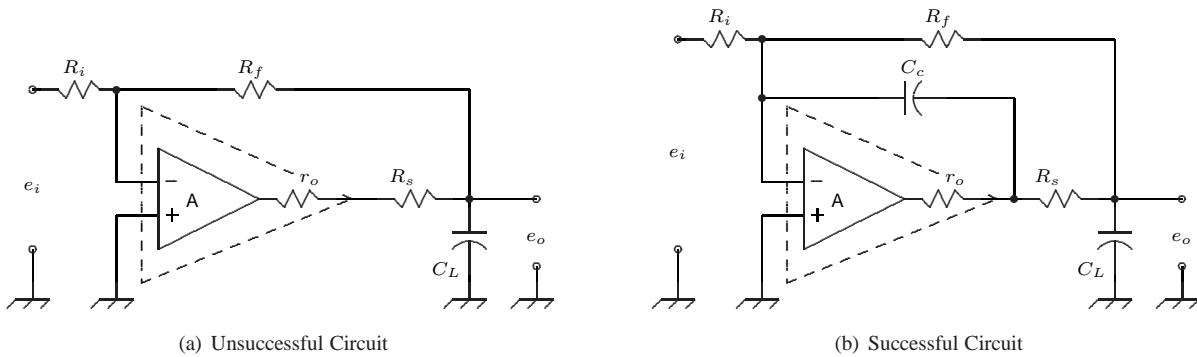


Figure 622: Enclosed Buildout Resistance

In section 21.7, we saw how the output resistance could be reduced by placing it inside the feedback loop.

$$r_o \approx \frac{r_{int}}{AB} \quad (1056)$$

It's natural to ask whether the beneficial effect of the buildout resistance can be retained and the output resistance lowered by placing the buildout resistance *inside* the negative feedback loop, as shown in figure 622(a). The answer, alas, is *no*. We're back to the original situation of figure 618, with r_o replaced by $r_o + R_s$. However, it turns out that the situation can be saved with the addition of a feedback capacitance C_c , as shown in figure 622(b).

At low frequencies, where stability is not an issue, capacitance C_c (which is a relatively small value, like 10 or 100pF) looks like an open circuit when compared to R_f . Consequently, the negative feedback action reduces the effect of the buildout resistance according to equation 1056.

At high frequencies, where stability is a problem, capacitor C_c takes effect. In general terms, it provides a phase lead that compensates for the phase lag created by the load capacitance.

This is an opportunity to illustrate a useful trick in analysing negative feedback systems: redraw the schematic so that the amplifier network $A(s)$ and feedback network $B(s)$ are separate and explicit. The input signal e_i is irrelevant in a stability analysis, so it may be considered as a ground point. Then figure 622(b) can be redrawn as shown in figure 623.

Now, by any measure this feedback network is complicated. However, some patterns can be detected.

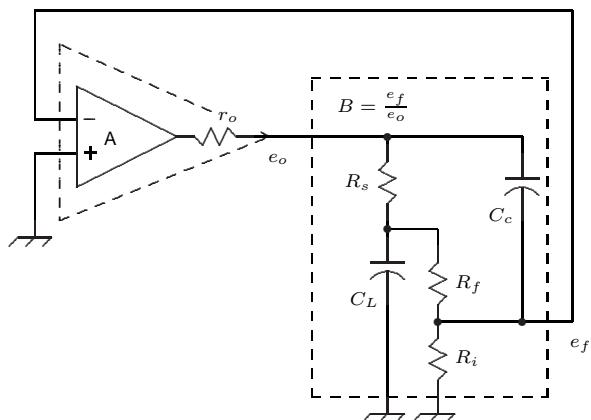


Figure 623: Figure 622(b) Redrawn

- Resistor R_s and capacitor C_L form a lag network, as we've pointed out previously.
- Providing that R_s is much smaller than R_f , then resistors R_i and R_f determine the low-frequency gain of the network.
- Finally, capacitor C_c and resistor R_f form a phase lead network that cancels the effect of the phase lag.

Graeme [226] discusses this network and its design in detail. Franco [139] provides the following design equations:

Configuration	Design Equations	3db Bandwidth
Inverting Amplifier	$R_s = (R_i \parallel R_f)r_o$ $C_c = (1 + R_i/R_f)^2(R_o/R_f)C_L$	$f_o \approx 1/2\pi R_f C_c$
Follower	$R_s = 30r_o$ $C_c = \sqrt{C_L/18\pi r_o B f_{GBP}}$	$f_o \approx \sqrt{B f_{GBP}/18\pi r_o C_L}$

Since there are many unknowns in this design exercise, such as the exact values of output resistance and load capacitance, a simulation of the response of the circuit would confirm that the design approach is correct. Then the design should be verified by experiment in the lab.

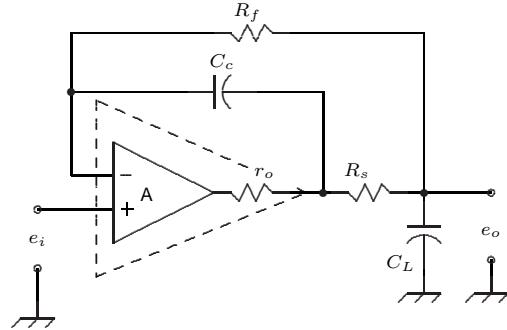


Figure 624: Compensated Follower

Capacitive Loading Solutions: Compensated Attenuator

Another method of load-capacitance compensation, suggested by Graeme in [226], is shown in figure 625(a). This is an interesting circuit, because it is relatively easy to understand – not always the case in feedback stabilization circuits. Figure 625 is drawn with the load resistance R_L included because it directly affects the circuit operation.

Redrawing the circuit for a stability analysis as shown in figure 625(b), we can see that the components inside the dashed box form a *compensated attenuator*. As we showed in section 4.15, the gain of the compensated attenuator can be made to be independent of frequency when (using the circuit names from figure 625(b))

$$R_s C_c = R_L C_l \quad (1057)$$

Then the trace for $B(s)$ in figure 618(b) becomes a horizontal straight line – constant with frequency, and the stability problem is solved.

There are some additional requirements that must be met for this to work.

- The R_f , R_i divider resistor values must be much larger than R_L (which is the usual case) so that

$$R_L \parallel (R_f + R_i) \approx R_L \quad (1058)$$

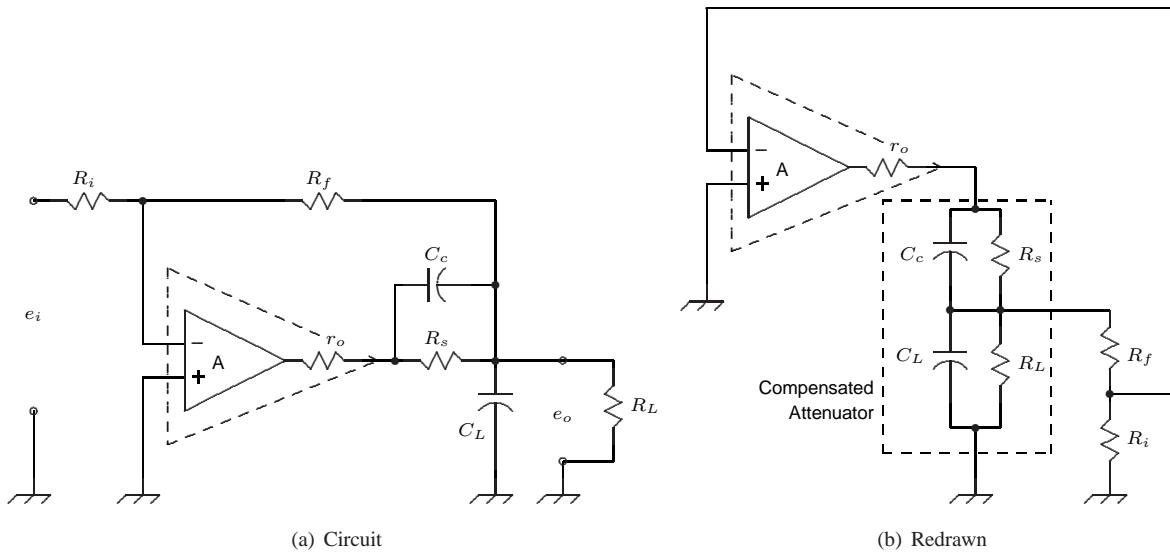


Figure 625: Compensation using Compensated Attenuator

- The load resistance R_L and capacitance C_L must be predictable. If the load capacitance were to change (because, for example, a coaxial cable length is changed) then the compensation is no longer exact.
 - As usual when a resistance is in series with the output as is the case for R_s in this circuit, the negative feedback will compensate for the drop across the resistor. However, in doing so, the op-amp must raise its output voltage and one must ensure that this does not lead to clipping of the output signal.

Reference

An excellent discussion of differentiator circuit stability (and stability of op-amp circuits in general) is in Jiří Dostál, Operational Amplifiers, Second Edition [219].

24.10 Case History: Oscillating Follower

Here is a case history that illustrates the debugging of an oscillation problem.

The case revolves around a high-speed CMOS op-amp, the Texas Instruments OPA356, used as an op-amp follower as shown in figure 626(a). This particular op-amp has a bandwidth of 200MHz, so it is an ideal choice to act as an *impedance converter* between a high impedance source and a lower impedance load. The bandwidth requirement in the particular application was 20MHz, so we would appear to have lots of gain in hand.

The schematic is disarmingly simple: the signal drives the non-inverting input pin. The output is directly connected to the inverting input pin. At these frequencies, it is important to use effective bypass capacitors on the power supply leads: the data sheet recommends 100nF in parallel with $10\mu\text{F}$.

There are some common-sense precautions. A socket for the op-amp would introduce too much parasitic capacitance and inductance so the IC must be soldered directly to the circuit board. The power supply bypass capacitors should have the shortest possible leads, so surface mount packages are preferable. Likewise, the leads that carry the input and output signals should be as short and direct as possible. A ground plane (cut away under the high impedance lead to avoid stray input capacitance) is advisable.

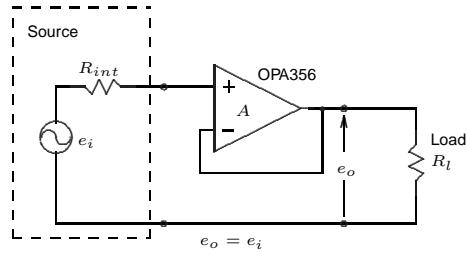
We constructed the circuit following all these guidelines. On testing the circuit, we found that the op-amp output had a half-volt DC offset under certain conditions of input. Our first suspect was a bias or leakage current generating voltage across the input impedance. However, touching the op-amp leads with a voltmeter probe caused the offset to disappear – a sure sign that an oscillation is occurring. The probe introduces enough capacitance into the circuit to stabilize it. Unfortunately, one can't leave it there to solve the problem. The oscillation, as is often the case, was at too high a frequency to see on our 150MHz bandwidth oscilloscope²⁰⁴.

In such a case, the first stop is the manufacturer's data sheet to see if there are any useful clues. A careful reading shows three relevant points:

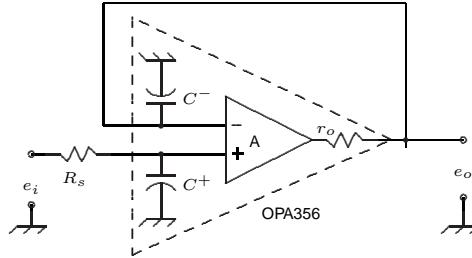
- The input of the op-amp appears as a large resistance in parallel with a 1.5pF capacitance to ground. The equivalent circuit is shown in figure 626(b).
- The frequency response for the non-inverting unity-gain case shows a 3db peak at around 300MHz, followed by an abrupt decrease in gain at higher frequencies. In this case, as in our circuit, the feedback resistance was zero ohms. The peak is a warning that the stability is decreasing in that frequency region.
- The frequency response for larger non-inverting gains, with a 604 ohm feedback resistor, shows a flat response (no peaking) up to the rolloff point.

These clues suggest a working hypothesis: the output impedance of the amplifier is interacting with the input capacitance C^- to cause sufficient phase shift to put the amplifier into oscillation. The data sheet implies that a zero ohms feedback resistor will work, but it's less stable than the case for a 604 ohms feedback resistor. Furthermore, the input capacitance is probably increased by the various strays from the printed circuit board wiring, which would make the problem worse.

But, if the output resistance interacts with the input capacitance to destabilize the follower, why does *increasing* the feedback resistor from 0 to 604Ω make it more stable? Because the feedback resistor has a small amount of parasitic capacitance C_f (about 1.5pF for an 0603 surface mount resistor), as shown in figure 627. This capacitance interacts with the 604 ohm feedback resistance to provide a lead term that partially compensates for the lag previously formed by the output resistance and input capacitance.



(a) Follower



(b) Equivalent

Figure 626: Oscillating Follower

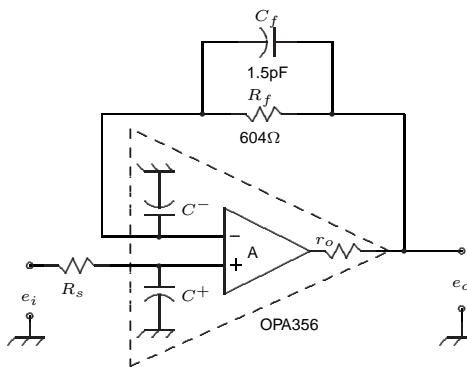


Figure 627: Oscillating Follower, Fixed

²⁰⁴We do have a Tektronix 7104 oscilloscope with 7A29 1GHz plug-ins, but the input impedance of those plug-ins is 50 ohms. Even using a $\times 10$ probe, this low an impedance would upset the high-impedance input circuit. If we were desperate to see this signal, we could have constructed a JFET follower buffer [113] for the 7A29 plug-in, but that would have been a rather time-consuming digression. And given Murphy's law (and the propensity for JFET followers to oscillate), that one might have been a problem as well.

As it turned out, the 604 ohm resistor in the feedback path – between the output and inverting input – killed the oscillation, and the mysterious offset disappeared. That's not the end of the engineering, however. Knowing that a follower has a propensity for oscillation, one would be advised to watch for overshoot and ringing on a pulse test, or a peaked response in a frequency sweep. Either of those symptoms indicates borderline stability and insufficient phase margin. In such a case, adding a few picofarads of capacitance across the feedback resistor might help stabilize the circuit.

24.11 Exercises

1. Draw the op-amp circuit for an inverting amplifier with a signal gain e_o/e_i of -1 and a noise gain of +10. Hint: Use the *Enhanced Noise Gain* technique of section 24.5.
2. Figure 628 (based on Pease [113]) shows a voltage follower and three different schemes for stabilizing the follower.

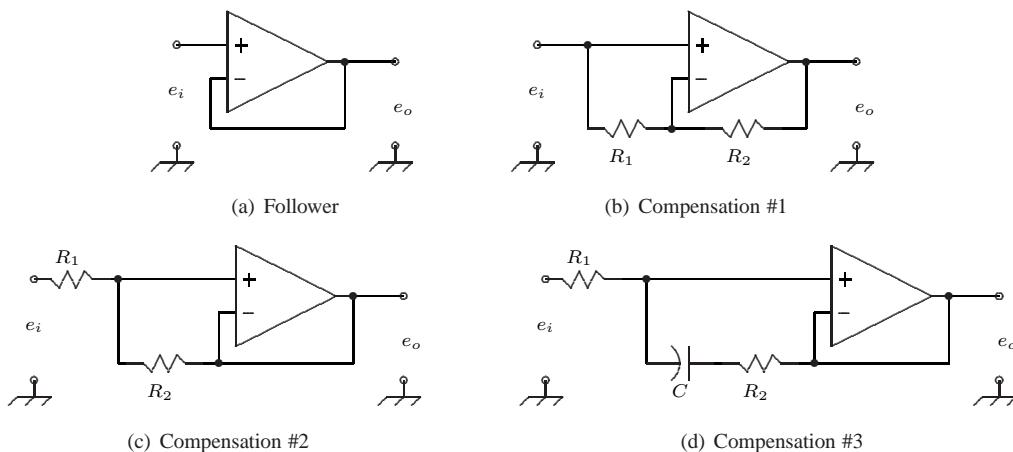


Figure 628: Stabilizing the Voltage Follower

- (a) Describe the open-loop frequency-gain characteristic of an op-amp which is likely to be unstable at unity gain.
- (b) For the circuits of figure 628(b) and 628(c), calculate the gain for a signal (the *signal gain* e_o/e_i) and the *noise gain*.
- (c) In figure 628(d), determine the signal gain and the noise gain as a function of frequency.
3. The National Semiconductor LF157 operational amplifier has the following open-loop gain characteristics:

Low-frequency Open Loop Gain	A_{ol}	105db
First break frequency	f_1	50 kHz
Second break frequency	f_2	30 MHz

It is recommended that this amplifier not be used below a closed-loop gain of 5 volts/volt. Determine the phase margin for that closed-loop gain.

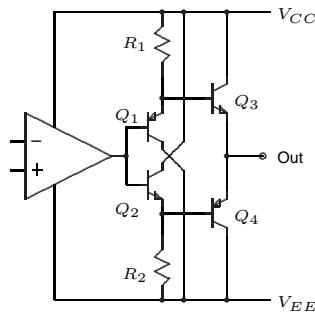


Figure 629: Amplifier with Booster

4. The circuit of figure 629 shows an op-amp with a *booster* output stage, comprised of transistors Q_1 through Q_4 . The booster stage increases the output current capability of the circuit (section 13.14 on page 356). The op-amp uses dominant-pole compensation. The open-loop gain has a first break frequency at 5Hz. The open-loop gain decreases to unity at 500kHz. The buffer has unity voltage gain and its first break frequency at 5kHz. (These are *very slow* transistors.)
- Draw the gain and phase response over the frequency range 1Hz to 10^6 Hz, for the amplifier, the buffer, and the combination.
 - Show how resistors and ground point would attach to this circuit to create a unity-gain inverting amplifier.
 - Determine the phase margin of this unity-gain amplifier and comment on the stability of the circuit.
 - If the circuit is determined to be unstable, show two methods of improving the stability.
5. An inverting amplifier of gain -2 volts/volt has the step response shown in figure 630.

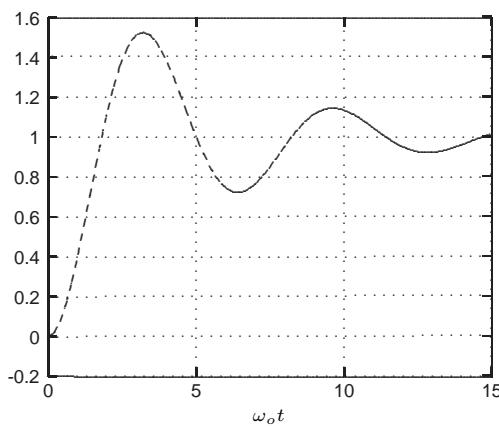


Figure 630: Step Response

- Based on the percentage overshoot, determine
 - the Q factor of this system
 - the phase margin in degrees.

- (b) One complete cycle of the oscillation corresponds to $33\mu\text{Sec}$. What is the approximate value of the frequency where the open and closed-loop gain intersect?

25 Digital to Analog Conversion

Accuracy and Precision

To begin with, we clarify three important concepts: *accuracy*, *resolution* and *precision*. These terms are often used interchangeably but in a strict sense they are completely different quantities [227].

Consider that we take a set of measurements of some length. The degree to which the mean value of the measurements conforms to the truth (the standard measurement) is the *accuracy*. The number of significant digits in the measurement (think of a dressmaker's measuring tape versus a micrometer) is the *resolution*. The spread of the measurements (their variability) is the *precision* of the measurements.

Consequently, it is possible to take a group of measurements that have excellent precision but poor accuracy. The measurements have very little spread to them, but their mean value is not close to the truth. Conversely, it is possible to take a group of measurements that have poor precision (vary widely) but high accuracy, in that their mean value is close to the truth²⁰⁵.

With that in hand, we are prepared for a consideration of analog and digital signals.

Analog Variable

An *analog* signal is one in which a continuous variable – voltage or current, for example – represents some physical quantity such as temperature or speed. If the signal is noiseless, it can be measured to any degree of precision desired. However, in any practical situation, the system noise sets a lower limit to the precision with which the signal can be measured.

Consider the case of a thermometer where the temperature is indicated by the expansion of mercury in a tube. The top of the column of liquid moves continuously (ie, not in steps) to represent temperature.

We could magnify the scale and thereby get a more precise measurement of the temperature, whether or not the reading is accurate.

Digital Variable

There are many reasons to involve a microprocessor in a signal processing or generating application. The microprocessor can perform complex calculations, store data, or display it in an informative fashion. Consequently, the ultimate destination for many analog signals is a microprocessor. In the converse, a microprocessor may be the source of an analog signal.

Put another way, most signals in nature are continuous, and so there must be a conversion process between the continuous analog signal and its discrete numerical representation in a microprocessor. This is *analog-digital conversion*, and *digital-analog conversion*.

In the process of conversion from the analog to digital domain, the analog signal is represented by a sequence of numbers, each one representing the amplitude of the signal at some instant of time – each number a so-called *sample*. If the samples are taken rapidly enough, the stream of numbers is an accurate representation of the original analog signal.

Each sample is a number which represents the amplitude of the analog signal at some instant. The sample has a finite number of digits and as a consequence the digital representation adds so-called *quantization noise*. As the number of digits increases, the precision increases and the quantization noise decreases. The precision of each sample must be sufficient that the noise is insignificant.

The emphasis in this material is a functional description of the various D/A and A/D converters. Johns and Martin [79] contains information for those who are interested in the internal circuit design of these devices.

²⁰⁵At a local hardware store, I came across a shelf displaying barometers for sale. They were all reading substantially different air pressures. Clearly, most if not all, were inaccurate, even though one could read the dial to considerable precision. It is interesting to speculate whether one could buy a number of such barometers, average the result, and thereby obtain improved accuracy.

25.1 D/A Specifications

The circuit and ideal transfer function for a digital-analog converter are shown in figure 631. This is a 3 bit²⁰⁶ converter, so it accepts 3 binary digits as an input and has $2^3 = 8$ possible outputs. In practice, A/D converters range from 6 to 16 bit devices, with some specialized devices extending to as many as 24 bits.

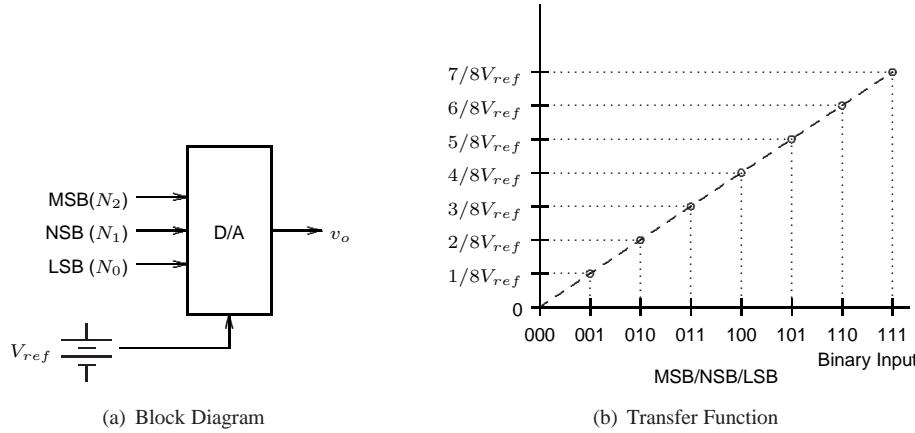


Figure 631: 3 Bit D/A Converter

Generalizing the output of the D/A converter, it is given by:

$$V_o = V_{ref} \left(\frac{N}{2^M} \right) \quad (1059)$$

where

- V_o Output Voltage
- V_{ref} Reference Voltage
- N Input binary number
- M Number of binary digits (in this example, 3)

This example is a *voltage* D/A converter, because it accepts a voltage reference and outputs a voltage. There are also *current* D/A's that accept a current reference and output a current.

Resolution

In this context *resolution* of the D/A refers to the change in output with a change of the least-significant bit. For example, a 12 bit converter has a resolution of $1/2^{12}$, that is, approximately 1 : 4000, or 0.025%.

$$\text{Resolution} = \frac{1}{2^M} \quad (1060)$$

Reference Stability

In applications where the magnitude of the output must be stable and predictable, the voltage reference is key. The stability of the output is only as good as the stability of the reference. If it drifts, the output will drift proportionally.

²⁰⁶We assume the reader is familiar with basic digital terminology. Any introduction to microprocessors or digital electronics defines these terms.

For example, we might require that the reference be as stable as the voltage represented by the least significant bit.

In audio applications, the emphasis is on linearity and absence of distortion. Stability of the reference is less critical and the magnitude can vary over a range without much effect.

The step size of the DAC output voltage is given by:

$$V_o(\text{minimum step}) = V_{ref} \frac{1}{2^M} \quad (1061)$$

For example, a 12 bit D/A converter with a 10 volt reference would generate a minimum step size of:

$$\begin{aligned} V_o(\text{minimum step}) &= V_{ref} \frac{1}{2^M} \\ &= 10 \frac{1}{2^{12}} \\ &= 2.44 \text{ mV} \end{aligned}$$

For many applications, the reference must be much more stable than this over the operating temperature of the circuit. This is not difficult to achieve with an 8 bit converter, but becomes progressively more difficult as the number of bits increases. If the D/A is in the same area with digital circuitry, then it may be difficult to keep the circuit noise below this minimum level, as well.

Non-Linearity

The points in the D/A transfer characteristic of figure 631 lie on a straight line, so the D/A output is linear. If the points lie on a curve of some sort, then the output is non-linear. This is an important issue when the D/A processes audio signals because non-linearity generates distortion products in the waveform.

Monotonicity

It is very important that every increase in the D/A input be matched by an increase in the output. If the output actually decreases at some point, as shown in figure 632, then the D/A is said to be *non-monotonic*. This is a serious error. In a control system, it reverses the sense of the feedback and will lead to instability.

Settling Time

When there is a step change in input code, the D/A will slew to a new output voltage. The time required for the new output voltage to be accurate within some tolerance (typically equivalent to 1/2 LSB), is known as the *settling time* T_s . An example is shown in figure 633.

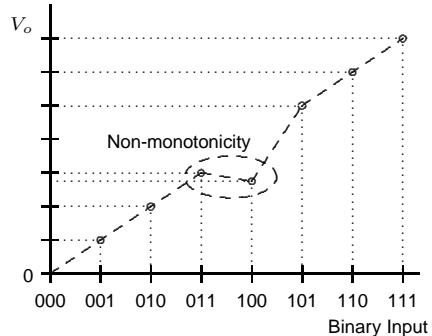


Figure 632: Monotonicity

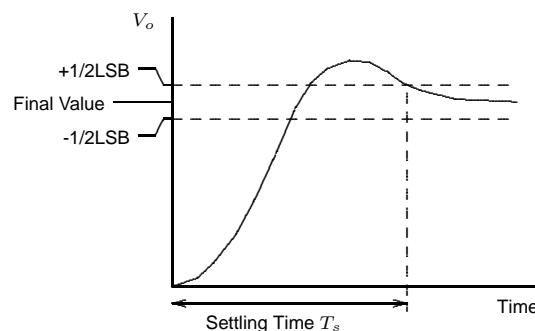


Figure 633: Settling Time

Glitches

Consider that an 8 bit D/A converter is being driven by a binary up-counter. At the midpoint of the count, the digital word will go from 01111111 to 10000000. Seven voltage or current sources switch off and one switches on. If the ON and OFF switching times are not exactly equal, then there will be a brief instant where the input number is different from either of these numbers. The result is a brief transient at the output, which is known as a *glitch*.

Glitches are an important consideration when the D/A is used to generate signal waveforms. They may be minimized by lowpass filtering the output (a severe compromise between waveform purity and speed), or by resampling the output with a switchable analog memory at the output of the D/A. The memory unit retains the last output voltage on a capacitor while the D/A is moving to a new output voltage. Then, when the output of the D/A has settled to its new value, the output from the D/A is connected to the memory capacitor to update its voltage. This adds a significant complication to the D/A circuitry.

25.2 Weighted Resistor D/A

A simple D/A converter is shown in figure 635. The circuit is basically an op-amp adder, with the input resistors weighted in a binary sequence. The switches are operated by the binary number: *open* for binary 0 and *closed* for binary 1. (As a practical matter, an electronic analog switch like the CD4066 device could be used for the switches.) Then the current flowing into the summing junction is proportional to the binary number, and the output voltage is a scaled version of that current.

More formally:

$$V_o = -R_f I_f \quad (1062)$$

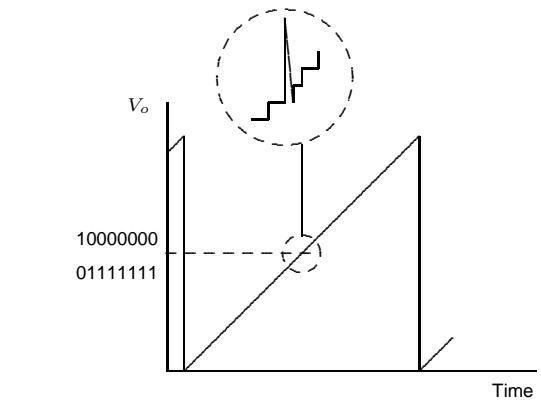


Figure 634: Output Glitch

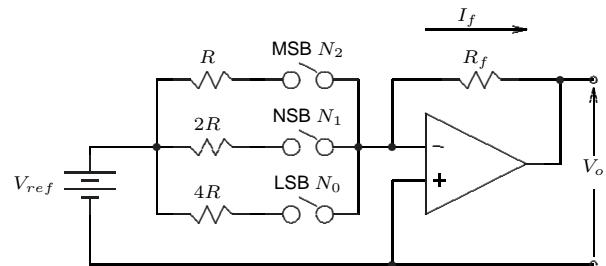


Figure 635: Weighted Resistor D/A

$$\begin{aligned} I_f &= \frac{V_{ref}}{R} N_2 + \frac{V_{ref}}{2R} N_1 + \frac{V_{ref}}{4R} N_0 \\ &= V_{ref} \left(\frac{N_2}{R} + \frac{N_1}{2R} + \frac{N_0}{4R} \right) \end{aligned} \quad (1063)$$

where the binary number is $N_2 N_1 N_0$ from MSB (most significant bit) to LSB (least significant bit) and N_x (a binary digit) is 0 or 1.

Substitute from equation 1063 into 1062:

$$\begin{aligned}
 V_o &= -RV_{ref} \left(\frac{N_2}{R} + \frac{N_1}{2R} + \frac{N_0}{4R} \right) \\
 &= -\frac{V_{ref}}{8} (4N_2 + 2N_1 + 1N_0) \\
 &= -V_{ref} \left(\frac{N}{2^M} \right)
 \end{aligned} \tag{1064}$$

where N is the binary number $N_2N_1N_0$ and M is the number of bits in the binary number which in this case is 3. This equation describes a digital-analog converter.

This will work quite nicely if a negative output is acceptable (which requires that a negative supply be present for the op-amp) and the number of bits M is not too large. However, as the number of bits increases, the range increases for the values of the input resistors and it becomes more difficult to ensure that the resistors are in a precise binary sequence.

25.3 R-2R D/A: Current Steering Mode

The weighted resistor D/A converter of section 25.2 is limited to a small number of bits. A better technique is the so-called *R-2R network*, based on the resistor network shown in figure 636.

Step 1

To analyse this network, start at the far right end. Collapse the $2R$ resistors in parallel into a resistance R . This appears in series with another resistance R , so they collapse into resistance $2R$. Continue with that process until the network is compressed into Step 1. Note that the voltage at A , V_{AG} , continues to be equal to the reference voltage V_{ref} .

Step 2

Now re-expand the network as shown in Step 2. The two equal resistors form a voltage divider which produces half the voltage at node A , so that the voltage at B , V_{BG} , is equal to $V_{ref}/2$.

Step 3

Re-expanding again to the network shown in Step 3, another voltage divider produces half the voltage at node B , so that the voltage at C , V_{CG} , is equal to $V_{ref}/4$.

This process can be continued to generate successive stages. Consequently, an R-2R network generates a binary series of voltage along its length, and it does so with two resistance values. Now we'll look at turning this into a digital-analog converter.

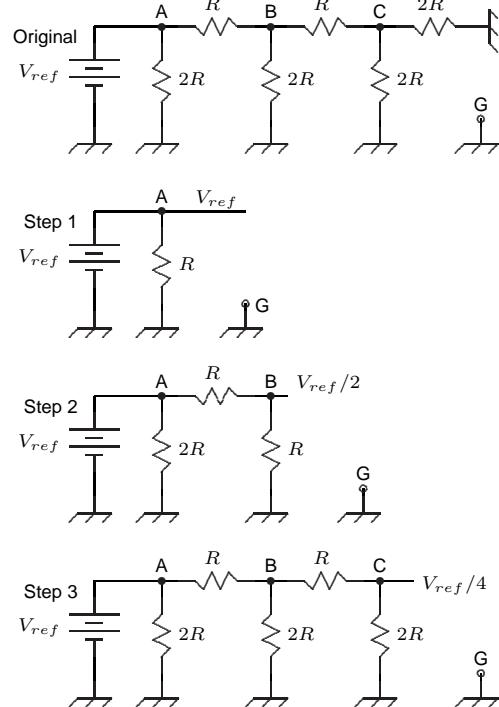


Figure 636: R-2R Network

Current Steering D/A

A complete digital-analog converter is shown in figure 637. The SPDT switches represent a 3 bit binary number with the MSB (most significant bit) at the left (S_A) and the LSB (least significant bit) at the right (S_C). If a binary digit is 0, the switch is in the left position. If it's a one, the switch is in the right position. In figure 637, the binary number is 011.

Notice that the switch position has no effect on the R-2R ladder. The pole of each switch is either connected to ground or the virtual earth of the op-amp, which is at ground potential. Consequently, voltages distribute along the R-2R network as they did in figure 636.

The currents through the switches are in a binary sequence: the current through S_B is half that through S_A , and so on. The op-amp sums up the currents that are enabled by 1's in the binary number, and thereby produces an output which is proportional to the binary number.

The circuit inside the dotted line (extended to a more useful number of bits) is available as a CMOS integrated circuit. The switches are actuated by control lines from the circuitry that generates the binary number.

As in the case of the binary-weighted-resistor D/A of section 25.2, the output voltage V_o is negative and the op-amp must be supplied with both positive and negative power.

25.4 R-2R D/A: Voltage Switching Mode

It's possible to reverse the connections to the current-steering D/A of figure 637, to make the arrangement shown in figure 638. This configuration is called the *voltage switching mode*, and has the advantage that the output voltage and power supply voltage are positive. A negative supply is not required.

Analysis

The general approach is to treat each switch as a separate source. Then we can determine the effect of each source on the output and combine the result according to the superposition principle.

Step 1

The circuit relevant to the MSB source is shown in figure 639.

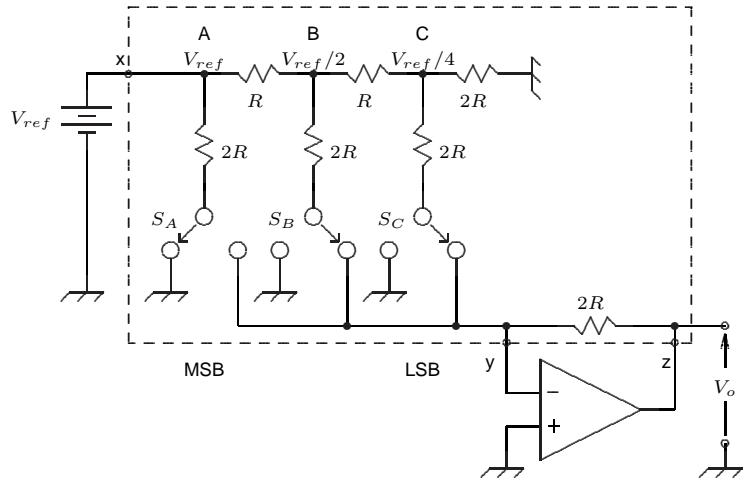


Figure 637: R-2R Current Steering D/A

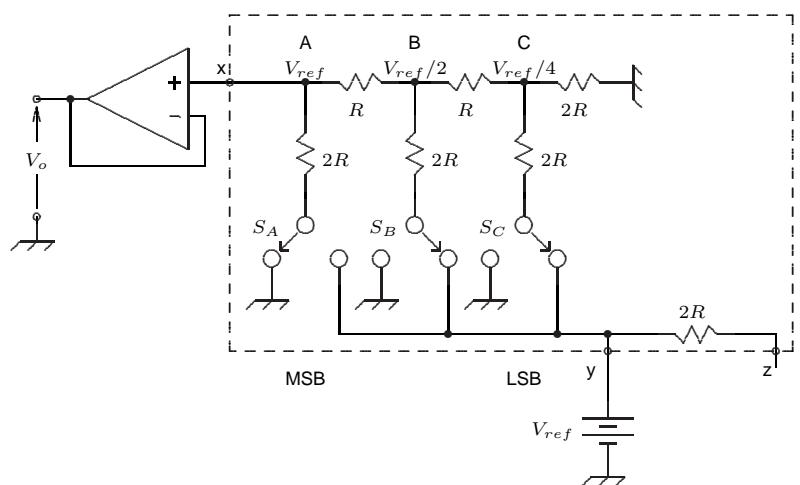


Figure 638: R-2R Voltage Switching D/A

When the resistors are collapsed, the circuit becomes an equal-resistance voltage divider. Consequently, the output voltage due to this source is $V_{ref}/2$

Step 2

The circuit for the NSB is shown in figure 640.

Applying our collapsing resistor trick to everything within the dashed-line box, it can be thevenized into a single resistor R and voltage source $V_{ref}/2$. This interacts with the remaining resistors outside the box to further divide the voltage by a factor of two, and the output voltage is $V_{ref}/4$.

Step 3

The LSB stage can be treated in much the same manner, and it outputs a voltage equal to $V_{ref}/8$.

Putting these steps together, the voltage sources contribute to the output in a binary sequence, that is, the outputs are proportional to their position in the binary number. Consequently, the circuit functions as a D/A converter, where

$$\begin{aligned} V_o &= N_2 \frac{V_{ref}}{2} + N_1 \frac{V_{ref}}{4} + N_0 \frac{V_{ref}}{8} \\ &= \frac{V_{ref}}{8} (4N_2 + 2N_1 + 1N_0) \\ &= V_{ref} \left(\frac{N}{2^M} \right) \end{aligned} \quad (1065)$$

where, as in previous 3-bit D/A converters, N is the binary number $N_2N_1N_0$ and M is the number of bits in the binary number.

Internal Resistance

In figure 638, the output of the voltage switching D/A is provided by a unity-gain follower circuit. This presents the R-2R network with a high impedance load (almost infinite) and provides a low impedance source for whatever output current is required into the load resistance.

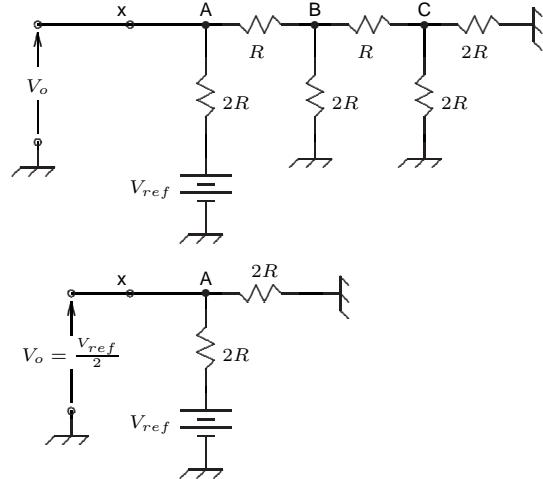


Figure 639: MSB Analysis

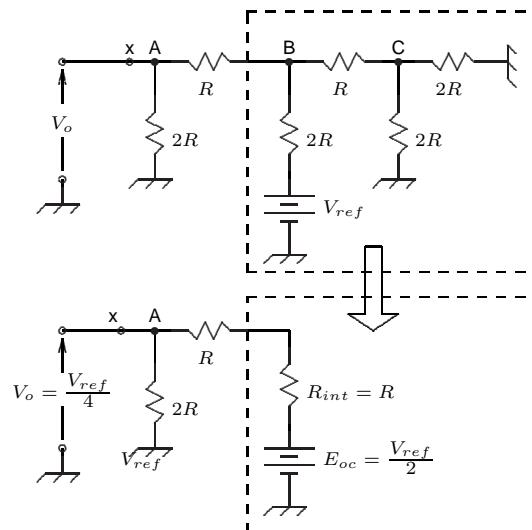


Figure 640: NSB Analysis

If we look back into the terminal x of the D/A, we see a *constant* internal resistance, regardless of the binary number setting, of R ohms. This is fairly evident from the equivalent circuit of figures 639 and 640, and it applies in general. Consequently, the voltage switching D/A can be used as the input resistance R_i in an inverting amplifier, as shown in figure 641. (This figure is drawn with the output on the left rather than the right to emphasize its relationship to figure 638).

Inexpensive D/A for CMOS Logic

CMOS digital logic, such as the 74HC series of logic devices, switches between levels that are very nearly equal to ground level and the supply level for the logic. For a low-resolution D/A it is possible to use these levels to drive an R-2R ladder network directly. For example, if a 5 bit D/A is required for some application, the circuit shown in figure 642 will work. Depending on the precision required, the resistor network can be constructed with (say) $10\text{k}\Omega$ and $20\text{k}\Omega$ discrete resistors. Alternatively, resistor networks, which typically match resistors to better than 0.5%, can be used.

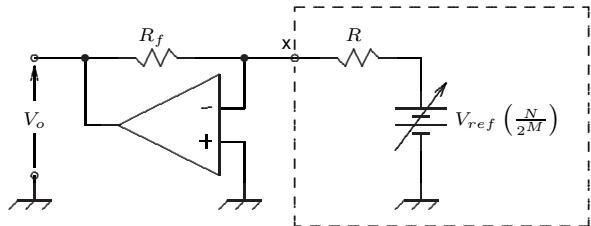


Figure 641: R-2R Voltage Switching D/A, Inverting Buffer

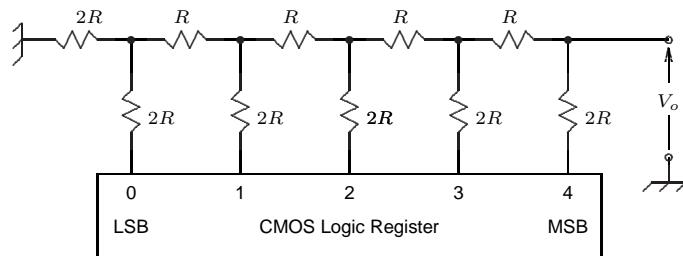


Figure 642: CMOS Driven Voltage Switching D/A

25.5 The MDAC: Multiplying Digital to Analog Converter

As we have seen in previous section, the defining equation for a digital-analog converter is

$$V_o = V_{ref} \left(\frac{N}{2^M} \right) \quad (1066)$$

where N is the digital value controlling the D/A, and M is the number of bits in the digital control word. For example, the equation for an 8 bit D/A is

$$V_o = V_{ref} \left(\frac{N}{256} \right) \quad (1067)$$

If the reference voltage V_{ref} and the digital number N are regarded as the inputs to the D/A, then the output voltage is proportional to the product of these two variables.

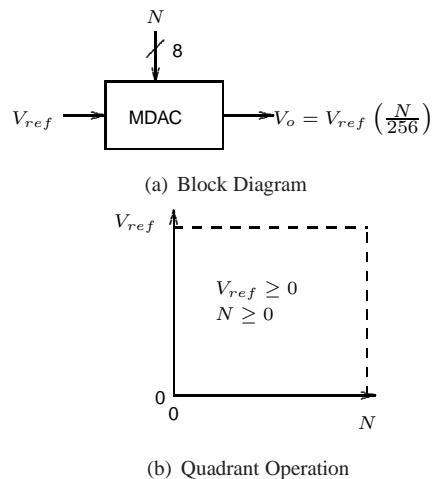


Figure 643: Basic MDAC, 8 bits

Then the D/A performs a multiplying operation, and is known as an *MDAC*, for *Multiplying Digital to Analog Converter*, with the block diagram shown in figure 643(a).

In some sense, all D/A's are MDAC's, since they all follow this equation. However, some D/A's are more suited than others for multiplication. This is dependent on the quadrants of operation available for the MDAC, as we'll see next.

Quadrant of Operation

Single Quadrant

The region of operation of a *one-quadrant* MDAC is shown in figure 643(b). V_{ref} must be positive. The binary number N is interpreted as an integer value greater than zero, ie, from 0 to 255 for an 8 bit converter. The output from the MDAC could be positive or negative, but not both.

Gain Control with a Single Quadrant MDAC

Suppose we wanted to use this type of MDAC to control the magnitude of an AC signal. As shown in figure 644, could an alternating voltage be used as the V_{ref} input, if we were to offset the AC by an amount $V_{ref}/2$?

Only with difficulty. The problem is this: as the value of the digital input N changes, it changes the magnitude of the AC input (which is desired), but it also changes the magnitude offset voltage in the output (which is undesired). Notice the output waveforms in figure 644. As the signal amplitude changes, so does the offset.

Now, the coupling capacitor removes the DC component in the output of the MDAC, and so only the AC signal appears after the capacitor. However, whenever the offset changes, it couples a transient through the capacitor, and this would be heard as a *thump* in the output audio.

It's possible to fix this design by subtracting from the output a voltage proportional to the DC component of the input (possibly using a second MDAC), but it's easier to use a Two Quadrant MDAC.

Two Quadrant

The region of operation of a *two-quadrant* MDAC is shown in figure 645. The reference voltage can be positive or negative and the binary number ranges from 0 to some positive value. This is a very useful configuration. The reference voltage input may be some AC source, in which case the MDAC is an electronic gain control for the AC signal, where the binary number sets the gain factor. For a standard MDAC, the output is directly proportional

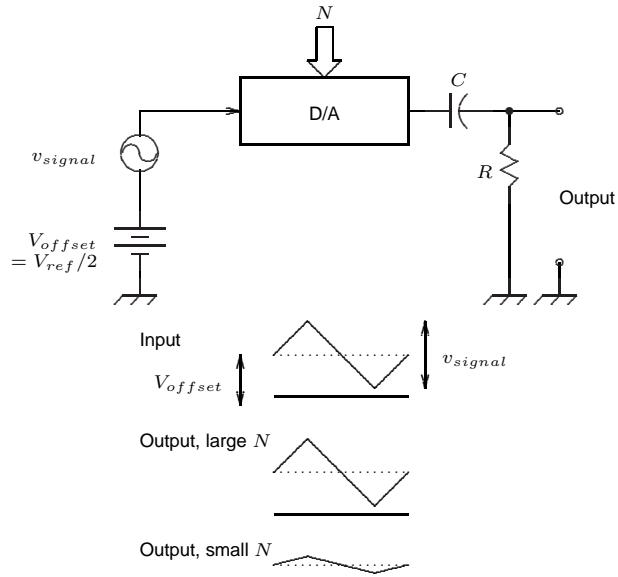


Figure 644: Gain Control with Single Quadrant MDAC

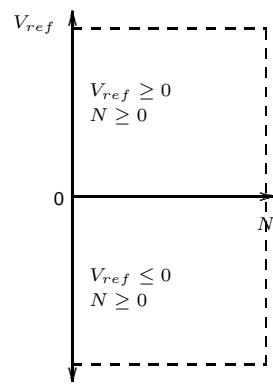


Figure 645: Two-Quadrant MDAC

to the value of the binary number and the control function is therefore linear. (For many audio applications a logarithmic control function is preferable, and there are specialized MDACS to provide that function.)

Four Quadrant

The region of operation of a *four-quadrant* MDAC is unrestricted in sign: both V_{ref} and N may be positive or negative. Most MDAC's do not have this capability built in. However, an external switchable sign-changer can be used to control the polarity of the output voltage. An example of this approach is shown in figure 646. The binary control value is in sign-magnitude format, and the sign-changer is driven from the sign bit of the binary control value.

25.6 MDAC Applications

In this section, we'll look at two representative applications for the MDAC: digitally controlled gain in an amplifier, and digitally controlled frequency in a function generator.

Gain Control

There are many applications where it is useful to be able to control the gain of an amplifier from a remote location or via another electronic signal. The MDAC provides one method of digital gain control. Figure 647 shows a possible circuit. The input and feedback resistors of an inverting amplifier are replaced by MDAC's.

The MDAC's are configured so that they generate an output *current* that is proportional to the input voltage and digital number. That is, using the MDAC equation:

$$i_i = \frac{v_i}{R} \left(\frac{N_i}{2^M} \right) \quad (1068)$$

and

$$i_f = \frac{v_o}{R} \left(\frac{N_f}{2^M} \right) \quad (1069)$$

where R is a property of the construction of the MDAC.

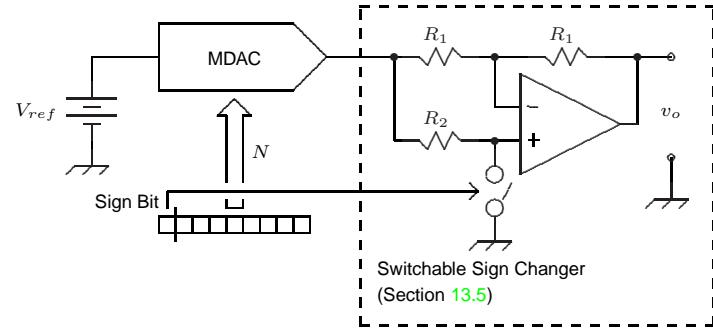


Figure 646: Four-Quadrant MDAC Example

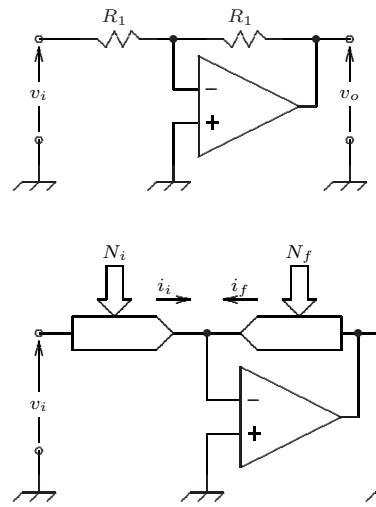


Figure 647: MDAC Gain Control

The op-amp will adjust its output voltage v_o so that all the input current flows through the feedback path, that is

$$i_f = -i_i \quad (1070)$$

Substituting equations 1068 and 1069 in equation 1070, we have

$$\begin{aligned} v_i \left(\frac{N_i}{2^M} \right) &= -v_o \left(\frac{N_f}{2^M} \right) \\ \frac{v_o}{v_i} &= -\frac{N_i}{N_f} \end{aligned} \quad (1071)$$

That is, the closed-loop gain is defined by the ratio of the two digital numbers. The gain is proportional to N_i so the relationship between gain and N_i is linear. The gain is inversely proportional to N_f so the relationship between gain and N_f is non-linear: highly so, since as N_f approaches zero the gain approaches infinity.

Of course, it is possible to replace either of the MDACS by a fixed resistor, depending on what is to be accomplished with the circuit.

Frequency Control

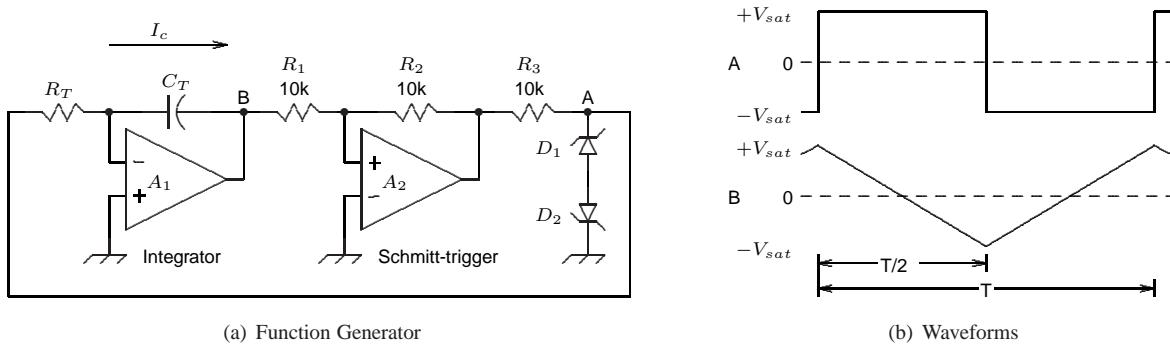


Figure 648: MDAC Controlled Function Generator

Now we'll look at a method of controlling frequency by means of an MDAC. The circuit of a simple function generator (section 20.7) is given in figure 648(a). The waveforms at points **A** and **B** are shown in figure 648(b). The output of the Schmitt trigger is limited by zener diodes at an amplitude of $\pm V_{sat}$.

First, we'll develop an equation for the frequency of the output waveform of the circuit of figure 648(a). The capacitor current is given by

$$I_c = \frac{V_{sat}}{R_T} \quad (1072)$$

We can also write that

$$\begin{aligned} I_c &= C \frac{dV_c}{dt} \\ &= C_T \frac{\Delta V_c}{\Delta T} \end{aligned} \quad (1073)$$

where $\Delta V_c / \Delta T$ is the slope of the capacitor-voltage waveform in volts per second, waveform **B** in figure 648(b).

Substitute $T/2$ for ΔT and $2V_{sat}$ for ΔV_c in equation 1073. Then combine equations 1072 and 1073, and we have:

$$\frac{V_{sat}}{R_T} = C_T \frac{2V_{sat}}{T/2} \quad (1074)$$

From which

$$T = 4R_t C_t \quad (1075)$$

The oscillation frequency is the inverse of this,

$$\begin{aligned} f_{osc} &= 1/T \\ &= \frac{1}{4R_t C_t} \text{ Hz} \end{aligned} \quad (1076)$$

It's always a happy moment in engineering when some variable cancels out of an equation, because that means that variable has no effect on the final result and does not need to be controlled. In this case, the saturation voltage cancels, and consequently has no effect on the output frequency. (This is another simplification: the value of V_{sat} has to be within limits that do not require the integrator to produce an output voltage beyond its power supply limits.)

The result in equation 1076 assumes (for simplicity) that the Schmitt trigger resistors R_1 and R_2 are equal. A more general result would not make that assumption.

Substituting an MDAC

In this case, the output current from the MDAC (which is the capacitor current I_c) is given by

$$I_c = \frac{V_{ref}}{R} \left(\frac{N}{N_{max}} \right) \quad (1077)$$

where, again, R is a property of the MDAC.

For this circuit, $V_{ref} = V_{sat}$. Making that substitution in equation 1077 and then simplifying as we did previously in equations 1073 through 1076, we obtain

$$f = \frac{N}{N_{max}} \frac{1}{4RC_T} \text{ Hz} \quad (1078)$$

The output frequency is proportional to the digital number input to the MDAC.

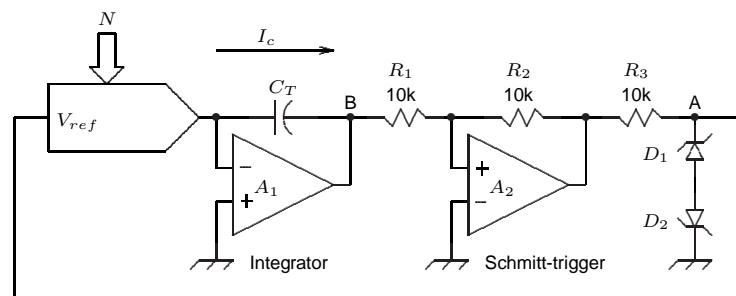


Figure 649: MDAC Frequency Control

25.7 Pulse Width Modulation DAC

It is a relatively simple matter to program a microprocessor to generate a pulse-width modulated waveform. Most microprocessors have hardware support for this task. The hardware is set up to produce a particular duty cycle and thereafter requires only a minimum of intervention from the software. This allows the microprocessor to focus on other functions that cannot be so easily mechanized.

The average value of the waveform depends on the duty cycle, and this may be extracted by a lowpass filter. This is particularly attractive when the device being controlled inherently provides the lowpass filtering action. For example, it is common practice to control the speed of a small DC motor by a duty-cycle modulated pulse waveform. The inductance of the motor windings and the mechanical inertia of the rotor average the driving waveform.

The basic concept is shown in figure 650. The digital circuit produces a pulse waveform with amplitude A . The duration of the pulse is αT , where T is the period of the waveform and α varies over some range between 0 and 1. The duty cycle is the ratio of time in the high state to the total time for one period of the waveform. That is:

$$\begin{aligned} \text{Duty Cycle} &= \frac{\alpha T}{T} \\ &= \alpha \end{aligned} \quad (1079)$$

Then the average (DC) value of the output voltage is given by the area under the pulse, divided by the time for one cycle:

$$\begin{aligned} V_{average} &= A \frac{\alpha T}{T} \\ &= \alpha A \end{aligned} \quad (1080)$$

Suppose that a microprocessor can control an output pulse with 8-bit resolution, that is, with a resolution of 1 part in 256. Then the average output voltage can be varied from the fractions 0/256 to 255/256 of the peak value, in steps of one unit at a time. Then, assuming that the average value can be extracted properly, this is effectively an 8 bit D/A converter.

As shown in figure 650, the averaged output value has a steady component and some degree of variation, caused by imperfect lowpass filtering, called the *ripple* component.

Time-Domain Ripple Analysis: First Order Filter

The lowpass filter input and output waveforms are shown in figure 651. Assume that the filter is a first-order RC lowpass filter with resistance R and capacitance C and consider the charging interval.

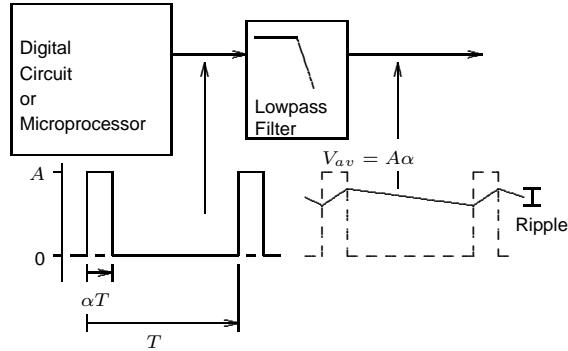


Figure 650: D/A Conversion using Pulse Width Modulation

Assuming the ripple voltage is small, the current charging the capacitor is given approximately by

$$I_c = \frac{V_m - V_{av}}{R} \quad (1081)$$

From the capacitor charging equation, we also have that

$$I_c = C \frac{\Delta V_o}{\Delta t} \quad (1082)$$

Substituting αT for Δt in equation 1082, and combining equations 1081 and 1082,

$$C \frac{\Delta V_o}{\alpha T} = \frac{V_m - V_{av}}{R} \quad (1083)$$

from which

$$\Delta V_o = \frac{(V_m - V_{av})\alpha T}{RC} \quad (1084)$$

But $V_{av} = \alpha V_m$ and $RC = \tau$, the time-constant of an RC lowpass filter. Substitute those expressions in equation 1084, and a little more algebra yields:

$$\frac{\Delta V_o}{V_m} = \frac{\alpha(1 - \alpha)T}{\tau} \quad (1085)$$

We need to design the lowpass filter to meet the specification under the worst case situation. The expression $\alpha(1 - \alpha)$ reaches a maximum at $\alpha = 0.5$. That is, the ripple will be a maximum when the duty cycle is 50%.

Example

Design a PWM D/A filter if the output voltage switches between 0 and +5V. The duty cycle varies between 0/256 to 255/256. Ripple must be less than 1/256 of the output voltage. Based on an analysis of the software loop that generates the PWM waveform, the period of the PWM waveform is expected to be around 1 millisecond.

Solution

The unknown quantity is τ , the time constant of the filter. Rearrange equation 1085 to solve for τ , set α equal to 0.5 and substitute the other known quantities:

$$\begin{aligned} \tau &= \alpha(1 - \alpha)T \frac{V_m}{\Delta V_o} \\ &= 0.5(1 - 0.5)1 \times 10^{-3} \frac{5}{5/256} \\ &= 64 \times 10^{-3} \text{ seconds} \end{aligned}$$

That is, the RC time constant of the filter must be 64 milliseconds (or greater) to meet the design requirements. For example, the filter component values could be $10k\Omega$ and $6.4\mu F$.

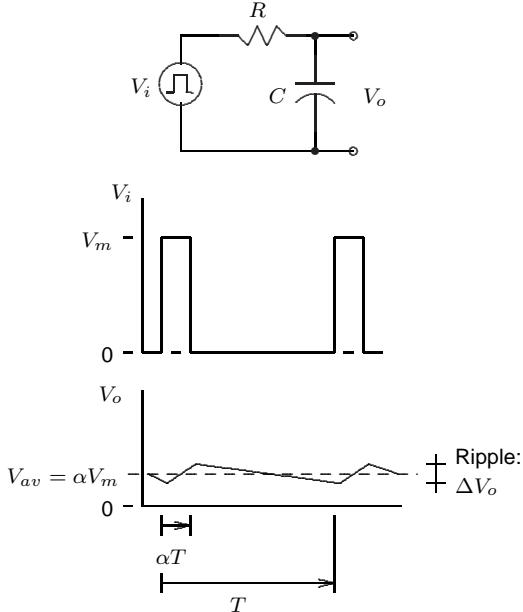


Figure 651: The Ripple Waveform

PWM Spectrum: The one that got away

It seems reasonable enough that we could predict the output ripple with an analysis in the frequency domain.

- Start with a knowledge of the PWM spectrum.
- Determine how that spectrum is modified by a lowpass filter.
- Reconstruct the output waveform by adding together the output harmonics of the lowpass filter.

The spectrum of a PWM waveform [160] is shown in figure 652. The signal consists of a series of harmonics occurring at multiples of the base frequency $1/T$. The envelope defining the magnitude of these harmonics is a $\sin(x)/x$ function with a null at frequency $1/\alpha T$.

The f_o component is the average or DC value. The other components at f_1 , f_2 and so on cumulatively form the AC ripple on the output waveform. The lowpass filter must pass the f_o component and sufficiently attenuate the AC harmonics from f_1 upwards.

In general, the duty cycle will change, changing the shape of the spectrum. Consequently, we need to determine the worst-case spectrum (the one with the largest magnitude of ripple frequency) and then design the filter to attenuate those signals sufficiently.

It would be nice if we could focus on the first harmonic, f_1 , and ignore the rest of the spectrum. Unfortunately, the harmonics are too close together for that to work. According to a simulation of the circuit, ignoring the higher order harmonics leads to an estimate of the peak-peak ripple that is a factor of two or so too low.

Considering the higher order harmonics is made more difficult by the phase shift of the lowpass filter. The phase shift of the filter depends on frequency, so the harmonics are phase shifted by different amounts. Adding them together would require taking these phase angles into consideration.

No doubt this can be done, but it's simpler to do the analysis in the time domain. However, considering the frequency-domain spectrum does give us an alternative view into the function of the lowpass filter.

Higher Order Filters

As described in the preceding paragraphs, the function of the lowpass filter is to remove the higher-order harmonics from the PWM spectrum.

If a first-order lowpass filter circuit is used to remove these harmonics the ripple may be reduced by reducing the cutoff frequency of the filter. The effect of this is to increase the filter time constant. As a consequence, if the duty cycle changes it will take longer for this change to appear in the average output voltage.

Alternatively, a higher-order lowpass filter may be used to filter the PWM waveform. A higher-order filter has a steeper descent into the stop band and consequently, for the same cutoff frequency, will have better ripple reduction than a first-order filter. Alternatively, for the same ripple reduction, a higher order filter will respond more quickly to changes in duty cycle.

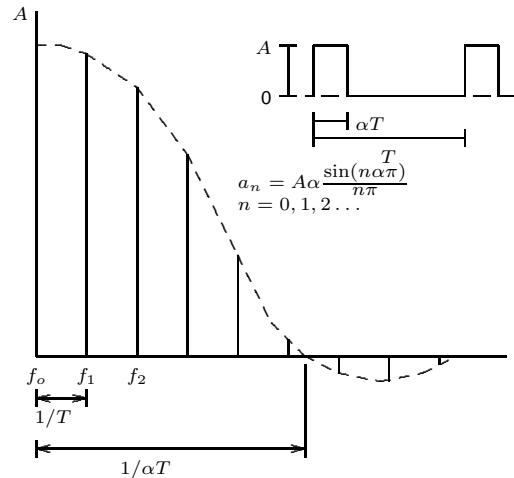
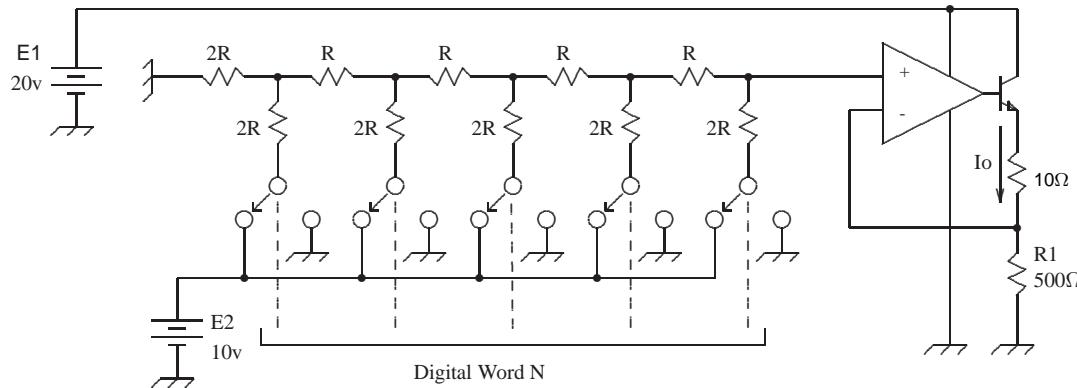


Figure 652: PWM Spectrum

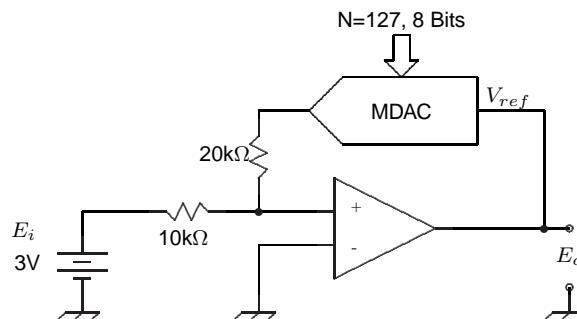
The exact effect of a higher-order lowpass filter depends on the order of the filter and the filter shape: Bessel, Butterworth or Chebyschev. Design the filter for some arbitrary cutoff frequency and then, using a circuit simulator, vary the input pulse frequency and duty cycle. Observe the output ripple. When the required ripple is achieved, note the relationship between filter cutoff frequency and pulse frequency. Then design the final filter circuit at the correct cutoff frequency with respect to the required pulse frequency.

25.8 Exercises

1. Consider the D-A control circuit shown in the figure.

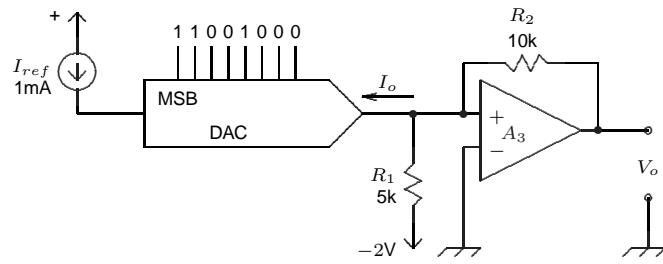


- (a) What is the *resolution* of the D-A converter?
 (b) Calculate the current in the 10Ω resistor when the digital word is set to 10101 (MSB to LSB).
2. For the circuit shown below, the 8-bit MDAC is driven by a digital signal N of 127. The reference voltage V_{ref} is derived from the output of an op-amp as shown. The op-amp is ideal.



Calculate E_o .

3. For the circuit shown below:



- (a) What is the value of the DAC output current I_o ?
- (b) What is the value of the op-amp output voltage V_o ?

26 Analog to Digital Conversion

26.1 Quantization

The process of converting a signal from analog to digital form introduces *quantization noise*, as illustrated in figure 653. Assuming a triangular input signal, the output waveform will have the step-like characteristic shown in figure 653(b). The time between each step is defined by some sort of clock timing signal that determines when each A/D and D/A conversion takes place.

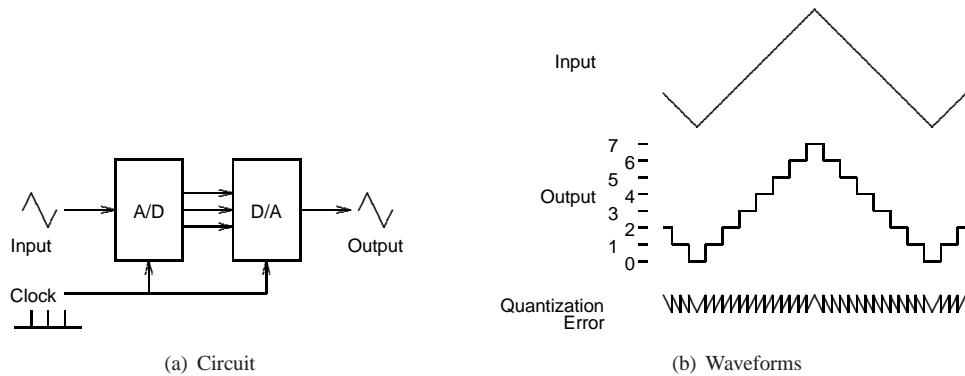


Figure 653: Signal-Noise Ratio

The example of figure 653 is a 3-bit A/D, D/A system, so the steps are very coarse. For example, an 8 bit system would contain $2^8 = 256$ steps and the steps would be much smaller in size. The error between the actual analog system and the digital approximation is known as the *quantization error* of the system. This is represented by the ugly waveform at the bottom of figure 653(b).

The RMS Value of Quantization Noise

The quantization noise shown in figure 653(b) is expanded and labelled in figure 654. Using the method outlined in [228] we can determine the RMS value of the noise by running the RMS definition backwards over one complete cycle of the waveform: *square, mean (average), square root*.

The equation of the waveform over one cycle (ie, time $t = -q/2s$ to $t = +q/2s$) is

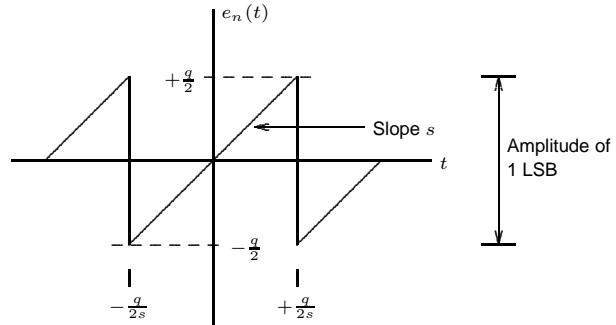


Figure 654: Quantization Noise

$$e(t) = st \quad (1086)$$

where s is the slope of the function in volts per second. Then the squared voltage is:

$$e^2(t) = (st)^2 \quad (1087)$$

We find the mean value by integrating over one cycle and dividing by the time for one cycle.

$$\begin{aligned} e^2(t)_{\text{average}} &= \frac{1}{T} \int_0^T (st)^2 dt = \frac{s}{q} \int_{-q/2s}^{+q/2s} (st)^2 dt = \frac{s^3}{q} \int_{-q/2s}^{+q/2s} t^2 dt \\ &= \frac{s^3}{q} \left[\frac{t^3}{3} \right]_{-q/2s}^{+q/2s} = \frac{s^3}{3q} \left[\frac{q^3}{8s^3} + \frac{q^3}{8s^3} \right] = \frac{q^2}{12} \end{aligned} \quad (1088)$$

The RMS value is the square root of the result:

$$e_{\text{rms}} = \sqrt{\frac{q^2}{12}} \quad (1089)$$

Example

What is the quantization noise expected from an 8 bit ($M=8$), 5 volt-full-scale A/D converter?

The LSB step voltage of a 5 volt 8 bit converter is given by:

$$\begin{aligned} V_{\text{lsb}} &= \frac{V_{\text{max}}}{2^M} \\ &= \frac{5}{2^8} \\ &= 0.0195 \text{ volts} \end{aligned}$$

Then the RMS noise voltage is:

$$\begin{aligned} V_{\text{noise}} &= \frac{V_{\text{lsb}}}{\sqrt{12}} \\ &= \frac{0.0195}{\sqrt{12}} \\ &= 5.6 \times 10^{-3} \text{ volts} \end{aligned}$$

26.2 Signal to Noise Ratio

We can gain some insight by considering the signal-noise ratio in a system like that of figure 653, using M bits to quantize a sine wave.

Suppose the sine wave is just large enough to exercise the conversion through all the possible levels of the converter, that is, its peak-peak value is V_m . Then its peak value is $V_m/2$, and its RMS value is:

$$v_{\text{signal}} = \frac{V_m}{2\sqrt{2}} \quad (1090)$$

According to equation 1089, the corresponding noise level is

$$v_{\text{noise}} = \frac{q}{\sqrt{12}} \quad (1091)$$

where q is the magnitude of the input voltage corresponding to a change of one significant bit in the digital code (the *step size*). It is also true that the maximum possible input voltage is 2^M times the step size:

$$V_m = 2^M q \quad (1092)$$

Substitute for V_m from 1092 into 1090 and find the signal to noise ratio SNR:

$$SNR = \frac{v_{signal}}{v_{noise}} = \frac{\frac{V_m}{2\sqrt{2}}}{\frac{q}{\sqrt{12}}} = \frac{\frac{2^M q}{2\sqrt{2}}}{\frac{q}{\sqrt{12}}} = 2^M \sqrt{1.5} \quad (1093)$$

Now express this in decibels:

$$\begin{aligned} SNR_{db} &= 20 \log_{10}(SNR) = 20 \log_{10}(2^M \sqrt{1.5}) = M \times 20 \log_{10}(2) + 20 \log_{10}(\sqrt{1.5}) \\ &= M \times 6.02 + 1.76 \text{ db} \end{aligned} \quad (1094)$$

That is, the signal-noise ratio is approximately equal to 6db per bit of resolution in the converter. An 8 bit converter would therefore have a signal-noise ratio of about 48 db.

This makes sense. If the number of bits is increased by one and the signal continues to fill the full scale of the system, then the number of steps will double and the step size will halve. The quantization noise is consequently reduced by a factor of 2 and the signal-to-quantization-noise ratio improves by a factor of 2. In decibels, this improvement is:

$$20 \log_{10} 2 = 6.02 \text{ db}$$

As a result of this, it is a useful rule of thumb that *the signal-noise ratio of an A/D or D/A system improves by 6db with each additional bit in the digital word*.

Notice that this signal-noise ratio is only realized when the input signal fills the entire input range of the system. In practice, the input signal is always less than the maximum allowable in order to avoid overloading and clipping in the A/D converter. Then the signal-noise ratio is worse than the best possible. Strictly speaking, our result is only true for the SNR of a sine wave, but it turns out to be approximately true for other waveform shapes as well.

In the worst case, consider a signal that is just smaller than one step on the A/D. Then the A/D outputs a constant digital value, this signal is totally ignored, and the signal-noise ratio is negative infinity: the signal is buried in noise.

Consequently, the system designer must ensure that

- there are enough bits available in the A/D and D/A conversion process to provide the desired signal-noise ratio²⁰⁷.
- the input signal is large enough to take advantage of a sufficient number of steps. This may require some sort of preamplification at the front end of the system.

Example: Oscilloscope Preamplifier

There is a technological tradeoff in the number of bits and speed of A/D conversion. For example, an 8 bit converter (at the same parts cost) is likely to be significantly faster than a 12 bit converter. Consequently, the first choice for a fast A/D converter would appear to be an 8 bit device. However, suppose we specify (somewhat arbitrarily) that we would like always to have at least 100 vertical points in our waveform display.

²⁰⁷This assumes that the path to improving signal-noise ratio is to increase the number of bits in the conversion. It's also possible in some applications to *oversample* the signal (sample the signal at a very high rate) and then filter the result through a lowpass filter. The oversampling operation spreads the quantization over a large spectrum. Lowpass filtering then restricts the spectrum and consequently reduces the quantization noise. As a result a very low resolution A/D converter can have very high effective resolution (see section 26.7). This technique is useful for high resolution converters for low frequency signals, and for audio applications.

The maximum input will be $2^8 = 256$ points. The minimum input is 100 points. Then the *allowable dynamic range* of the A/D converter is approximately 2.5:1. For example, if the maximum input signal is 10 volts peak-peak, then the minimum input signal is $10/2.56 = 3.90$ volts, or about 4 volts.

If we would like to observe signals below 4 volts, then they have to be amplified up to the maximum input voltage of the A/D, which requires an analog amplifier of gain factor 2.5 before the A/D converter.

Notice that amplifying the signal *after* digitization does not compensate for the lost information, and is no substitute for front-end amplification.

With a 12 bit A/D converter and a minimum of 100 vertical points, the dynamic range is

$$2^{12}/100 = 40.9 : 1$$

Consequently, with a 10 volt maximum input, the smallest displayable input would be

$$10/40.9 = 0.240 \text{ volts}$$

If this is acceptable as a minimum displayable signal, an adjustable gain preamplifier might not be required at all.

Now consider the effect of step size on noise susceptibility. A 10V, 8 bit A/D converter has a step size of

$$10/2^8 = 39 \text{ millivolts}$$

Compare this to a 10V, 12 bit converter, which has a step size of 2.4 millivolts. Consequently, the 12 bit converter will be much more susceptible to noise from the digital section of the circuit. For the same speed rating as an 8 bit device, it will also be more expensive.

It's a classic engineering tradeoff: adjustable gain preamp plus 8 bit A/D or simplified front end plus 12 bit A/D.

26.3 Aliasing

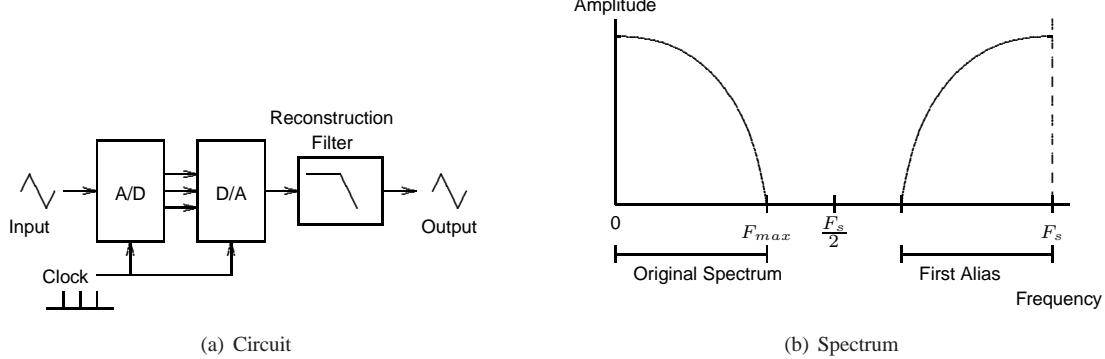


Figure 655: Normal Sampling

To illustrate the concept of *alias*, consider the hardware arrangement shown in figure 655(a). Input voltages are converted, one with each clock pulse, by an A/D converter. The next clock pulse causes the numerical output of the A/D to be converted back to an analog voltage in a D/A converter.

The effect of sampling the original spectrum at a frequency of F_s is to generate a *first alias* spectrum as shown in figure 655 [75]. The original waveform is reconstructed without error by passing the output of the D/A

converter through a lowpass filter, which is known as a *reconstruction filter*. The cutoff frequency of this filter is located somewhere between F_{max} and $F_s/2$. The reconstruction filter removes the first alias spectrum, leaving the original spectrum.

Notice how the original spectrum is folded about the frequency $F_s/2$ to create the first alias. We'll refer to $F_s/2$ as the *folding frequency*.

Now consider the effect when the maximum frequency of the input signal approaches the sampling frequency. If the two spectra overlap as shown in figure 656(a), then the waveform is said to be *undersampled* and a phenomenon known as *aliasing* occurs.

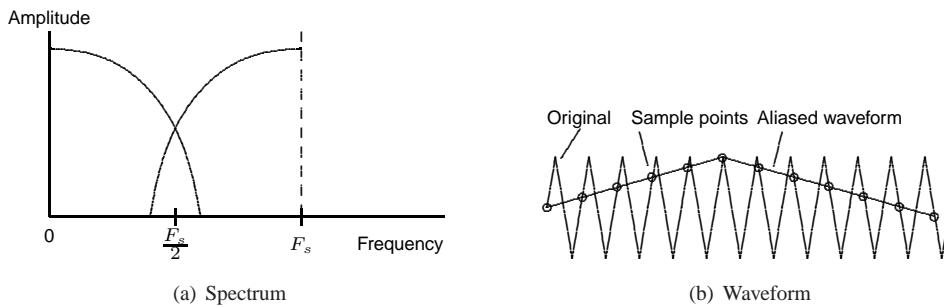


Figure 656: Undersampling

Figure 656(b) shows a time-domain example of aliasing. In this case, the sampling of a high-frequency triangle waveform generates a much lower frequency triangle waveform. The low-frequency waveform is an incorrect artifact of the sampling process.

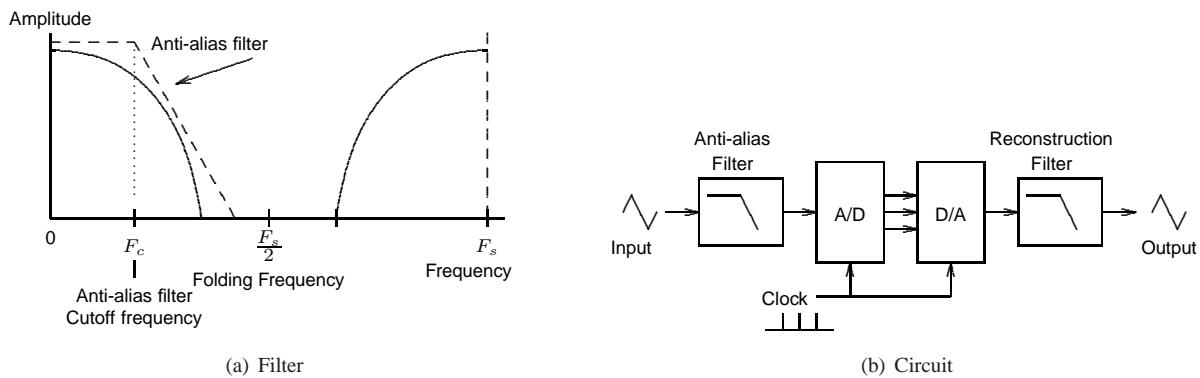


Figure 657: Anti-alias Filter

If there are frequency components in the input spectrum that exceed the folding frequency, then one option is to provide an *anti-aliasing filter* at the front end of the system. The effect on the spectrum is shown in figure 657(a) and the circuit in figure 657(b). This lowpass filter ensures that input signals are reduced to an inconsequential level at and above the folding frequency. In some circumstances (digital audio, for example), the cutoff frequency of the anti-alias filter is only slightly below the folding frequency. In such a case, the anti-alias filter must have a very steep descent from the passband into the stop band.

26.4 Feedback Converters

A great many circuits have been proposed for analog-digital conversion, of which a small subset have survived into practical use. Some of these are based on the block diagram shown in figure 658, an 8 bit device.

The input voltage V_x is applied to one input of a comparator. A feedback voltage V_f is generated by a digital-analog converter and applied to the other input of the comparator. The digital logic senses the state of the comparator via its C_{in} terminal and adjusts its digital output until the V_f is as close as possible to V_x .

A transition at the digital *start conversion* input causes the logic to begin the comparison process. When the conversion has completed, the logic asserts the *end of conversion* signal and the digital value is available at the terminals D0..D7.

There are three different algorithms that may be used by the digital logic in the process of matching V_f to V_x .

Up-counting Converter

The digital logic is based on a binary up-counter. At the start of a conversion, the counter is reset into its zero state. It then counts upward until the state of the comparator goes high, and stops. The waveforms are shown in figure 659.

In the worst case the counter has to reach its maximum count and the conversion time is given by

$$T_{conversion} = T_{clock}(2^M - 1) \quad (1095)$$

where T_{clock} is the period of the conversion clock and M is the number of bits in the conversion.

For example, an 8 bit converter could require 255 clock pulses to complete a conversion.

This A/D converter has the virtue of simplicity, but suffers from a long conversion time.

Tracking (Up-Down) Converter

The digital logic is based on an up-down counter. The counter runs continuously and the output of the comparator changes the direction of the count. The count continuously tracks the input voltage as shown in figure 660. Providing the counter can up and down count faster than the rate of change of the input voltage, the digital value is always within one count of the correct value. This converter has no *start* or *end-of-conversion* signals, since it runs continuously.

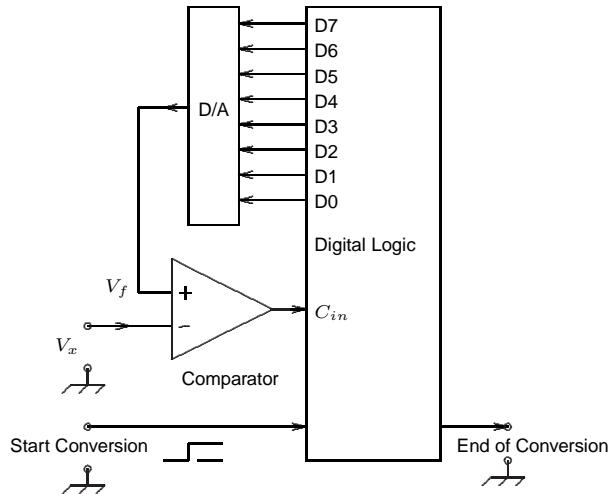


Figure 658: A/D Feedback Converter Concept

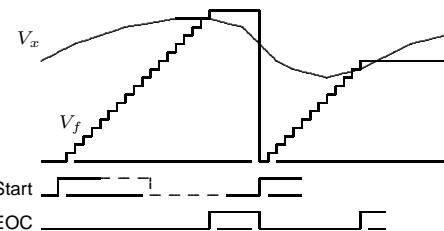


Figure 659: Up-Counter Converter Waveforms

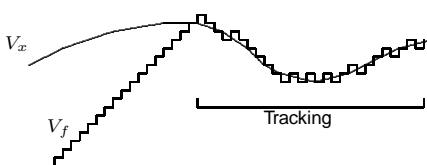


Figure 660: Up-Down Counter Converter Waveforms

Successive Approximation Converter

In this case, the logic implements a binary search algorithm. The search proceeds in M steps, where M is the number of bits in the binary word. With each step, it divides the search space in half and generates the corresponding feedback voltage. The comparator advises the logic whether this test voltage is larger or smaller than the input.

Consider the case of a balance-beam weight scale with a maximum weight value of 15 units.

An unknown weight is on the right side, and it's our job to put weights on the left side until the scale is balanced (or close to it). The balance weights are in a binary sequence: 8,4,2 and 1. How would we determine the correct combination of balance weights?

In general terms, we *approach from below*.

We start with the largest balance weight. If the weigh scale indicates that this is too much, we remove it, mark down a 0 for the most significant binary digit and proceed to the next smaller balance weight. On the other hand, if the largest weight is too little, we leave it in place, mark down a 1 for the most significant digit and then try the next smaller balance weight. We continue this process until we have tested every balance-weight. Then the final binary number is the value of the weight²⁰⁸.

Figure 661 shows this process as a waveform of the comparison values where the input voltage V_x has a value of 11.5.

- The first test level is half the maximum, a value of 8. The comparator indicates that the test level is less than the input, so the most significant digit of the conversion is a 1 and the test voltage includes a value of 8 units.
- The second level adds 4, for a test level of 12. The comparator indicates that the test level is greater than the input, so the next digit of the conversion is a 0. We discard the 4 units, so the test value remains at 8 units.
- The third level adds 2, for a test level of 10. The comparator indicates that the test level is less than the input, so the next digit of the conversion is a 1 and the test level remains at 10.
- The fourth level adds 1, for a test level of 11. The comparator indicates that the test level is less than the input, so the last (least significant) digit of the conversion is a 1.

The final conversion value is 1011 in binary, which is 11 in decimal.

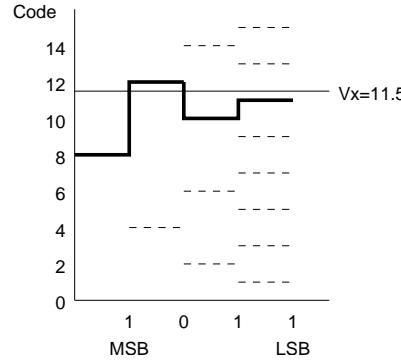


Figure 661: Successive Approximation Converter Waveforms

²⁰⁸This binary search procedure is useful in other situations as well. Suppose you have a computer program with a bug at an unknown location. Divide the source code in half and rerun. If the error is in this test half, then further subdivide that code in half and rerun. If it's not in the test half, it must be in the other half. Divide that sample of the code in half and rerun. And so on. The search process for N lines of code will lead you to the exact line of code in $\log_2 N$ tries. For example, if the source code is 4096 lines of text, then you will find the exact line in 12 tries. That technique was used in the debugging of the L^AT_EXcode for this book.

John Foster points out that in general this will not work with compiled programs, a good reason to stick with interpreted languages.

26.5 Single Slope Technique

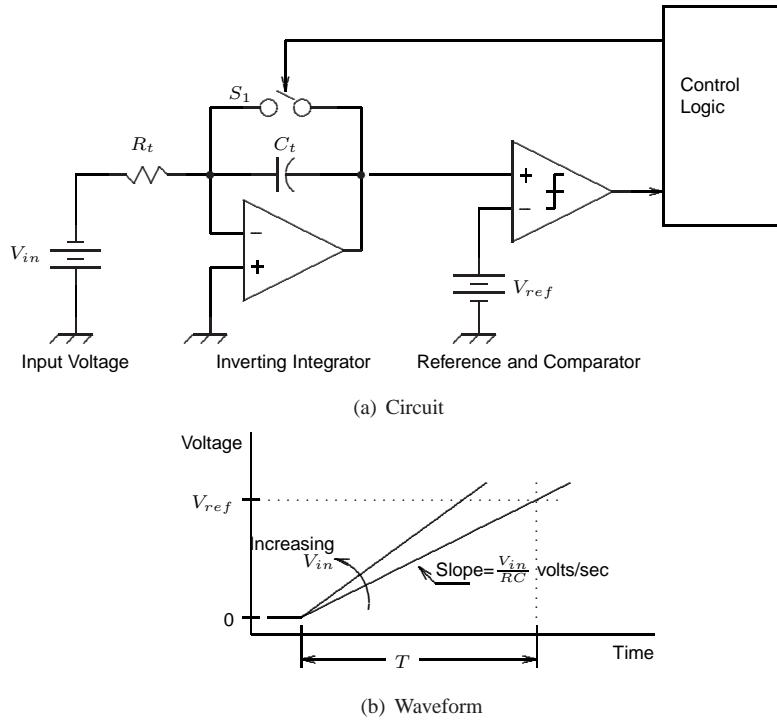


Figure 662: Single Slope Technique

In sections 13.10 and 13.11, we saw that a constant voltage input to an integrator causes the output of the integrator to ramp at a constant rate. This may be used as the basis for a *single-slope* analog-digital conversion technique.

As shown in figure 662(a), the output of the integrator is compared with a reference voltage. At the beginning of a measurement cycle, switch S_1 is opened and, with a negative input voltage V_{in} , the integrator output ramps upward. The measurement cycle concludes when the ramp equals the reference voltage. Then the rate of change of the integrator output over the measurement cycle is given by

$$\frac{\Delta v_o}{\Delta t} = \frac{v_{in}}{R_t C_t} \text{ volts/second} \quad (1096)$$

The control logic closes switch S_1 to set the output of the integrator to zero volts. Then it opens switch S_1 and measures the elapsed time T until the comparator changes state. In equation 1096 we can substitute V_{ref} for Δv_o and T for Δt . Then, solving for the input voltage v_{in} , we have

$$\frac{V_{in}}{V_{ref}} = \frac{R_t C_t}{T} \text{ volts} \quad (1097)$$

This *voltage-to-time* conversion technique is very useful in conjunction with a microprocessor as the control logic. It is straightforward for a microprocessor to measure time. If it is not interrupted, a simple counting loop will do. The elapsed time is then equal to the count times the number of microseconds for each iteration of the

count loop. However, calculation of the input voltage is complicated by the fact that it is proportional to the reciprocal of the time measurement.

This technique is widely used. In fact, it is not even essential that the ramp voltage be linear. Providing the shape is known, anything is possible. An exponential charging curve is convenient, because the ramp circuit then degenerates to a simple RC circuit. The microprocessor controller is then required to do a more complex computation, but that may be finessed by using a lookup table. A circuit is shown in figure 663. Some microprocessors include a comparator, which eliminates the requirement for an external device.

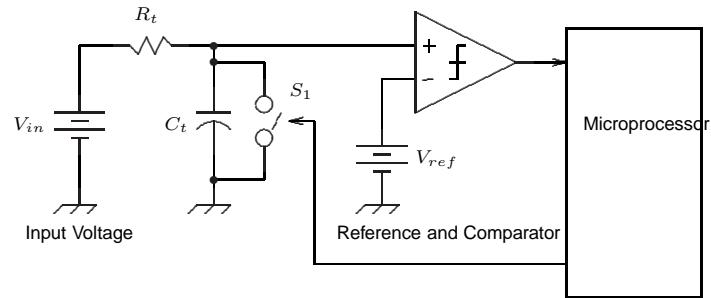


Figure 663: Single Slope A/D, Simplified

Single Slope Accuracy

As equation 1097 indicates, the measurement of input voltage depends on the timing resistance, timing capacitance, reference voltage, offset voltage of the comparator and time interval measurement (ie, stability of the time measurement clock). Even if any initial error can be *calibrated out* with microprocessor software (one of the strengths of using a microprocessor in this type of system) any changes in the component values will directly affect the accuracy of the conversion.

Noise

Voltage measurements are often contaminated with some type of noise. To illustrate the behaviour of the single-slope converter with a noisy input signal, consider an alternative single-slope configuration, shown in figure 664.

- The unknown input voltage and the reference voltage are interchanged, so we can interchange V_{in} and V_{ref} in equation 1097, whereby we obtain:

$$\frac{V_{in}}{V_{ref}} = \frac{T}{R_t C_t} \text{ volts} \quad (1098)$$

Consequently, as the input voltage V_{in} increases, the time measurement T increases. That is, the input voltage is directly proportional to the measured time interval, which may simplify the calculation of input voltage.

- The previous circuit has some ability to ignore input noise. This circuit has none. To see why, consider the circuit of figure 664(a). The precise instant of comparison depends on the amplitude of the input voltage V_{in} . If that varies by x volts of noise, then the A/D reading will vary by a proportional amount. Put another way, the variance in the A/D reading will be the same as the variance of the noise voltage.

Now consider the previous circuit of figure 662. Let us also consider the noise voltage as an AC component riding on the DC component we wish to measure. The time integral of a period waveform is zero. (Think of a sine wave integrated over one complete cycle). Consequently, a periodic waveform that has an integral number of cycles over the measurement period T will be averaged to zero. While this is unlikely to happen exactly, there is still some noise reduction effect due to the action of the integrator. As well, the frequency response of an integrator falls off at high frequencies – it is a lowpass filter, in effect. Consequently, high-frequency noise tends to be ignored.

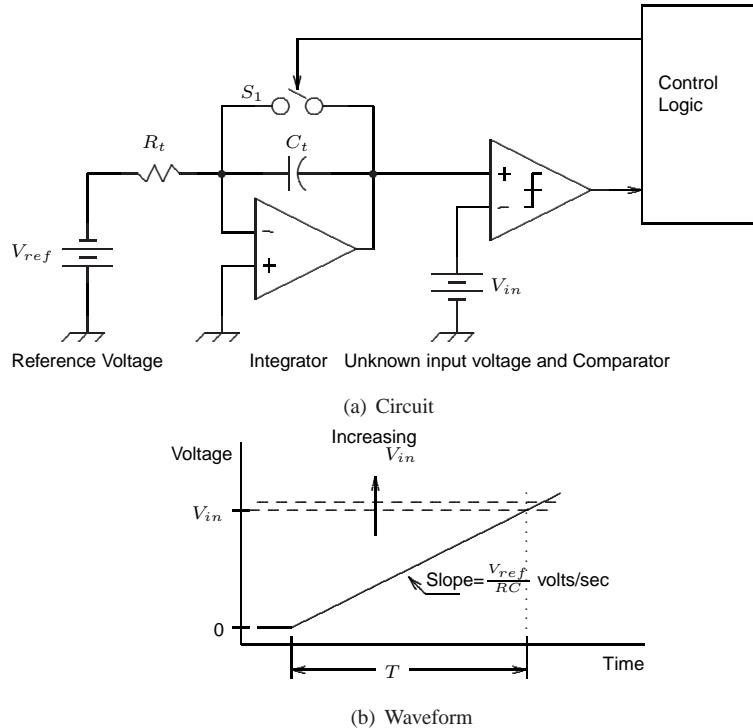


Figure 664: Single Slope A/D, Alternate Configuration

26.6 Dual Slope Technique

The *dual-slope* analog-digital conversion technique [229] eliminates some of the error sources in the single-slope technique of section 26.5.

The circuit is shown in figure 665. As in the single-slope technique, we assume that the input voltage is negative. The input and reference voltages are now selected by switch S_1 and both are processed through the integrator. The comparator detects when the output of the integrator crosses through zero volts.

A measurement cycle proceeds as follows:

1. The control logic closes switch S_2 to set the output of the integrator to zero volts.
2. The control logic moves switch S_1 to the V_{in} position. It then opens switch S_2 , and the integrator output begins ramping upward at a rate which depends on the input voltage:

$$\frac{\Delta v_o}{\Delta t} = \frac{v_{in}}{R_t C_t} \text{ volts/second} \quad (1099)$$

3. This phase is allowed to continue for a fixed interval of time T_1 .
4. At the end of time T_1 , The control logic moves switch S_1 to the V_{ref} position. The output of the integrator reverses direction and ramps downward at a fixed rate

$$\frac{\Delta v_o}{\Delta t} = -\frac{v_{ref}}{R_t C_t} \text{ volts/second} \quad (1100)$$

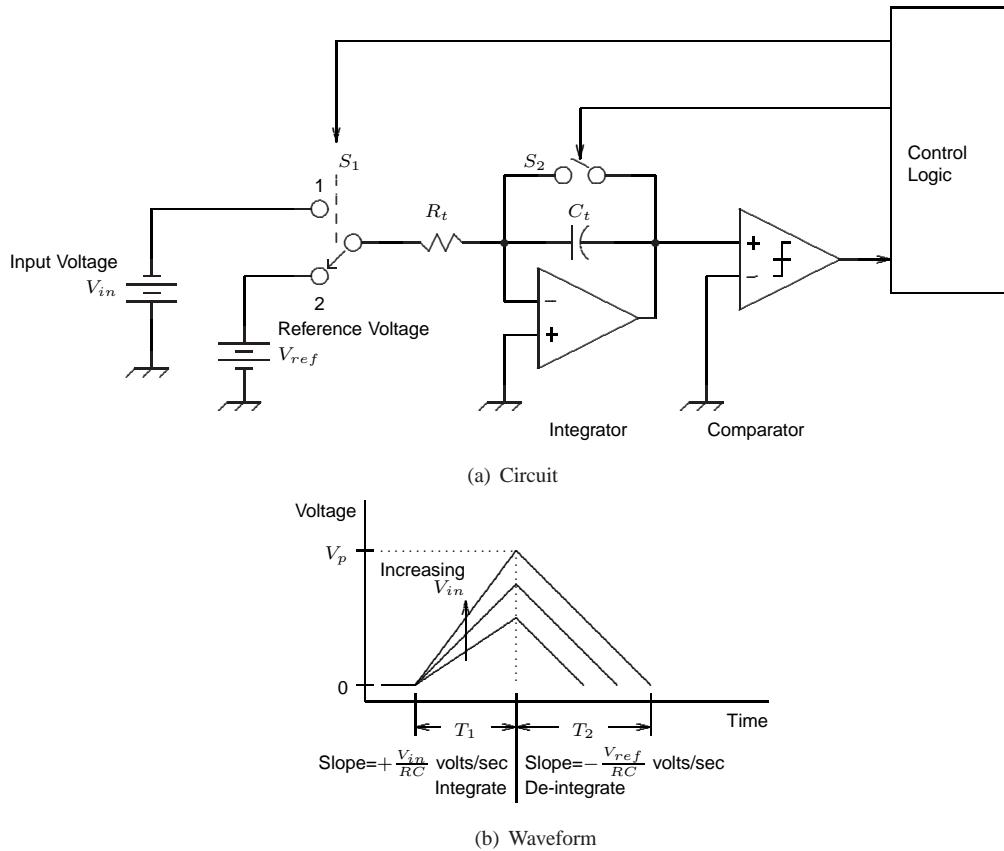


Figure 665: Dual Slope Technique

5. When the output of the integrator crosses through zero, the comparator changes state and the digital logic records the duration of the second time interval T_2

To summarize,

- the first part of the measurement results in a variable slope for a fixed interval
- the second part of the measurement results in a fixed slope for a variable interval

The net effect is to cancel the effect of component parameters, as we'll now see. The peak value V_p of the integrator output is given by

$$\begin{aligned} V_p &= \frac{V_{in}T_1}{R_tC_t} \\ &= \frac{V_{ref}T_2}{R_tC_t} \end{aligned}$$

Then

$$\frac{V_{in}}{V_{ref}} = \frac{T_2}{T_1} \quad (1101)$$

This is a joyful result, because it indicates that the timing resistor and capacitor have no effect on the result (providing that they are stable throughout the measurement interval).

The timing intervals are usually digital counts based on a clock of period T_c . If we put

$$T_1 = MT_c \quad (1102)$$

and

$$T_1 = NT_c \quad (1103)$$

then equation 1100 can be rewritten as

$$\frac{V_{in}}{V_{ref}} = \frac{N}{M} \quad (1104)$$

In other words, the exact value of the clock period T_c is not important (again, as long as it does not vary during a measurement) and the result is the ratio of the two integers N and M .

Now suppose that we set the integer M to 1000, and integrate during the first phase for 1000 clock cycles. Then the value of N may be interpreted as the value of the input voltage in *millivolts*.

Accuracy

The accuracy of the conversion is primarily dependent on the accuracy of the reference voltage source V_{ref} . It is also required that the offset voltage of the comparator be much less than the resolution of the measurement. This may be accomplished by so-called auto-zero circuits which store the offset voltage on a capacitor and subtract its effect during the measurement cycle [139]. With that modification, a dual-slope A/D converter is capable of 20 bit accuracy, that is, approximately 1 part in 10^6 .

Speed

The dual slope technique requires thousands of clock cycles to complete, so the conversion rate is typically in the region of 10's of conversions per second. That makes it suitable for the hand-held voltmeter DVM, and for precise measurements of variables, such as temperature, that change relatively slowly.

Noise

The mains power distribution system is often a source of interfering noise. However, the dual slope voltmeter can have considerable inherent rejection of such signals. If the first measurement period T_1 is made to be some multiple of the power line waveform period, the power frequency waveform will average to zero during the integration. The AC input will have no effect on the peak value V_p . As a result, the A/D ignores this noise source and accurately measures the DC component of the input.

26.7 Sigma-Delta Oversampling Converter

The sigma-delta oversampling converter uses a totally different approach to analog-digital conversion. The circuitry is a combination of a high-speed one-bit A/D converter and a digital filter. The combination produces a converter that

- simplifies the design of the anti-aliasing filter and
- converts at the necessary speed and precision for digital audio applications

The original impetus for development of the sigma-delta converter came from the need for a precise low-cost method of A/D conversion in digital audio applications. Subsequently sigma-delta converters have proven to be useful in instrumentation applications (such as digital weighsscales) where high resolution is required at modest conversion rates.

The Aliasing Problem

Consider the case of consumer digital audio. For high fidelity, the audio band should extend from 20Hz to 20kHz. The standard digital audio sampling rate f_s is 44.1kHz. Consequently, any frequency above $f_{max} = f_s/2 = 22.05\text{kHz}$ will be aliased back into the passband of the audio. The anti-aliasing filter must have zero attenuation at $f_c = 20\text{kHz}$ and lots of attenuation at $f_{max} = 22.05\text{kHz}$.

To quantify *lots*, consider that consumer digital audio is at least 16 bits for a dynamic range of approximately $6 \times 16 = 96\text{db}$. Ideally, the anti-aliasing filter should drop by 96 db in 1/24 of a decade. In practice, there is very little signal energy at 20kHz and a 9 to 13 pole lowpass filter is used in this type of system [230]. This is a very complex lowpass filter and would require great precision among the resistors and capacitors.

The Linearity Problem

Again considering the case of digital audio, it is of great importance that the A/D converter be completely linear. Any non-linearity in the A/D converter will appear as harmonic and intermodulation distortion in the output digital audio stream. This places great demands on the matching of the internal resistance network.

The Sigma-Delta Method

The sigma-delta converter uses a completely different method of A/D conversion, thereby avoiding some of these difficulties. A block diagram of a typical sigma-delta converter (based on reference [231]) is shown in figure 666.

There are two major sections to the converter: the modulator and the decimator. First, we'll explain their overall role in the circuit, and then we'll examine the circuitry in more detail.

The modulator converts the incoming analog voltage v_i into a stream of binary digits, logical 1's and 0's. The rate of this stream is determined by the clock rate, Kf_s , where K is the *oversampling ratio*. The average value of this digital stream represents the input voltage.

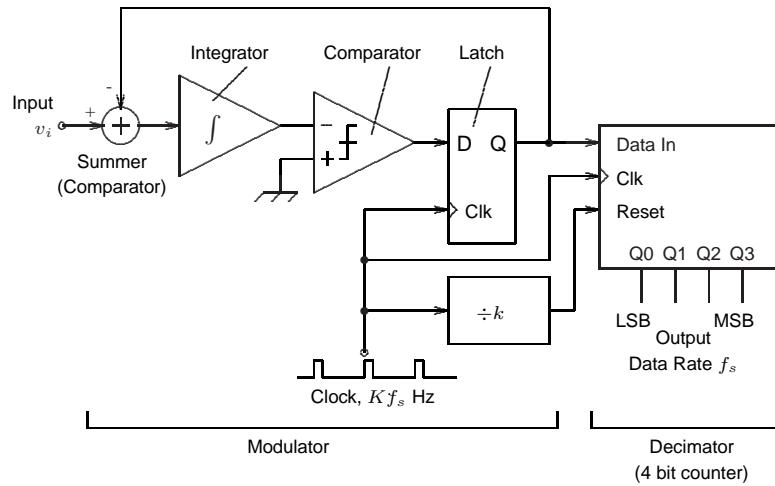


Figure 666: Sigma-Delta Converter, Block Diagram

Let's assume that the span of the A/D converter is 5 volts, that is, the input can vary between 0V and +5V. For example, if the input voltage is 2.5V, there will be an equal number of 1's and 0's.

If the input value is at its positive maximum, then the output of the modulator will have many more 1's than zeros, representing a positive voltage.

The decimator acts as an averaging device and reduces the output rate back to f_s , the original sampling frequency. In figure 666, the decimator is shown as a 4 bit up-counter. The counter functions as a digital integrator. Over a fixed interval (16 clock pulses in this case), it counts the number of 1's from the modulator. This value becomes an output digital word. Then the counter resets to zero for the next counting interval.

For example, if the input voltage is at its mid-point value, then the modulator will produce an equal number of 1's and 0's over the counting interval. A 4 bit counter will then show a count of 8 at the end of the counting interval. Since an input of 2.5 volts causes an output of 8 units, the A/D converter has a scale-factor gain of 8/2.5.

In practice, the decimator is often a digital FIR filter, which is a long shift register with multiple taps that determine the final digital output. In either case, counter or FIR filter, there is a delay between the time that the input voltage changes and when that change is reflected in the output digital word. The time delay is referred to as the *latency* of the converter.

Resolution Enhancement

The modulator of the converter can be regarded as a feedback type based on a one-bit D/A converter, the comparator. At the output of the decimator²⁰⁹, the resolution has become 4 bits, so the overall structure of the comparator enhances the resolution of the front end. Now we'll examine in more detail how this magic is accomplished.

The resolution of an analog-digital converter is related to its quantization noise. In section 26.2, we established that the quantization noise is equal to

$$V_{noise} = \frac{V_{lsb}}{\sqrt{12}} \text{ volts}$$

As shown in figure 667, the spectrum of this noise extends from zero up to the folding frequency. For the moment, we'll assume that the spectrum is flat (white) and the noise is evenly distributed.

Now consider the case shown in figure 668. In this situation, the signal is oversampled at a rate somewhat greater than necessary to prevent aliasing. (In figure 668 the factor is something in the order of 2 because of the difficulties of drawing a much larger and more realistic factor of 512.) Oversampling at a high rate would present difficulties with a conventional A/D converter, but the sigma-delta converter is relatively simple and so can be clocked at a high rate.

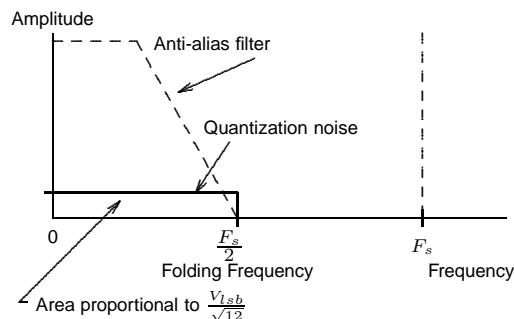


Figure 667: Quantization Noise, Normal Sampling

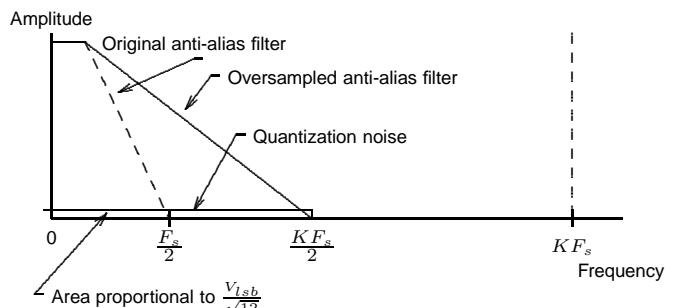


Figure 668: Quantization Noise, Oversampling

²⁰⁹My next book is an action story based around the use of sigma-delta converters in counter-terrorism, to be called *The Decimator*. If you are interested, call my agent.

The total quantization noise is the same magnitude in figures 667 and 668, but it is spread over a wider bandwidth in figure 668, so the noise density (section 22.2) is much lower in the second case.

The spectrum of figure 668 is generated by the modulator of the sigma-delta converter. The noise density is lowered by oversampling, and the sample rate is K times larger than required to prevent aliasing. This spectrum then passes through the decimator, which reduces the sample rate by a factor of K . In doing so, the decimator functions as a lowpass filter.

Without venturing into the realm of digital signal processing, this last statement requires a leap of faith. However, consider the behaviour of the counter-decimator in our original block diagram of figure 666:

- if the input to the counter is zero, it does not increment and the output count is constant
- if the input is a binary 1, the output count increments upward

This behavior is similar to that of a non-inverting analog integrator: if the input is zero, its output stays constant. If the input is some positive value, the output ramps upwards. In the frequency domain, an analog integrator has a lowpass gain characteristic (section 11.3), so it's not too much of a stretch to expect a counter to exhibit the lowpass frequency response of an integrator. Further supporting evidence for this behaviour may be found with *rate-multiplier* circuits [50], [232] in which counters are used as integrators.

The decimated spectrum is shown in figure 669. The amplitude of the quantization noise remains lowered by oversampling, and the extent of the quantization noise is limited by the decimator. Consequently, the total quantization noise is reduced.

Noise Shaping

We previously assumed that the noise spectrum was flat with frequency. In fact, the noise spectrum at the output of the sigma-delta converter rises with frequency. To see why, consider figure 670.

The modulator may be modelled as a negative feedback system with a disturbance, where the disturbance is the quantization noise. The system operates to minimize the amount of the disturbance that appears in the output (section 10.3). The degree to which the disturbance is reduced in the output signal is proportional to the loop gain, that is:

$$e_{on} = \frac{e_q}{AB} \quad (1105)$$

where e_{on} is the quantization noise in the output signal, e_q is the quantization noise, A is the forward gain, and B is the sensor gain. In this case, the forward gain A is the gain of an integrator, and the sensor gain B is unity.

The gain of an integrator is extremely large at low frequencies and decreases (at 20db/decade) as the frequency increases. Consequently, the output quantization noise is the opposite of this: low at low frequencies and increasing with frequency. The noise is said to be *shaped* with most of the energy at high frequencies. As a result, the action of the decimator is much more effective than if the noise spectrum were flat.

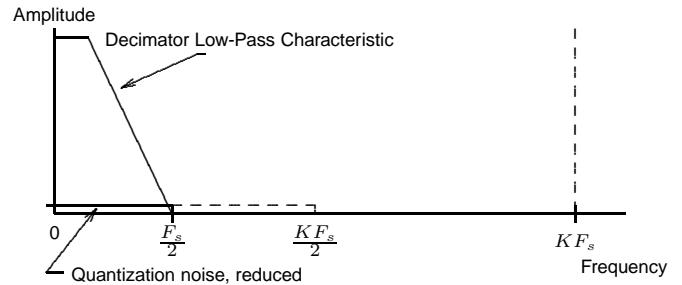


Figure 669: Quantization Noise, Oversampled and Decimated

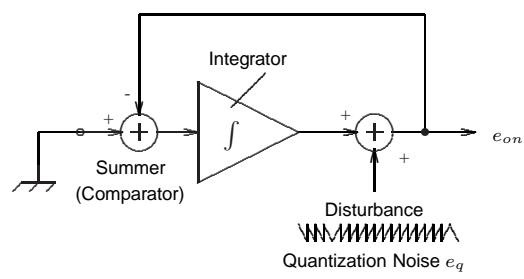


Figure 670: Delta-Sigma Circuit, Noise Analog

Other more elaborate sigma-delta converters are possible, in which even more energy is moved to high frequencies (and therefore removed). The effectiveness of this approach can be gleaned from the fact that some delta-sigma A/D converters have resolution of 24 bits. That's equivalent to a step resolution of 60 *nanovolts*.

Applications

Many sensors (strain gauges and thermocouples, for example) output a small signal. When such a signal must be digitized it has been traditional to amplify the signal in a low-drift operational amplifier and then convert it to digital form in an A/D converter with a span in the order of volts. The required gains can be very large, in the order of 1000V/V. Consequently, the amplifier must exhibit extremely low offset-voltage drift.

With the advent of a reasonably-priced 24-bit sigma-delta A/D converter, it becomes feasible to connect the signal directly to the input of the A/D converter and set the total span to a few millivolts. This entirely eliminates the requirement for an amplifier. On the other hand, many instrumentation applications have severe noise-pickup problems and one would have to ensure that the A/D converter could adequately reject radio interference and 60Hz signals that are coupled into the signal leads of the sensor.

References

The paper and application notes in references [233], [234], [231] and [235] are comprehensible accounts of the operation of the sigma-delta converter.

26.8 Voltage to Frequency Converter

The *voltage to frequency* converter generates a train of pulses, the frequency of which is directly proportional to an input voltage. It is useful as a relatively low-precision, low-cost A/D converter. It's also useful, as we will see, in certain industrial applications.

A block diagram is shown in figure 671(a). In essence, this is a negative feedback system in which an input charge Q_{in} is balanced by a feedback charge Q_f . For this reason, the circuit is often known as a *charge-balancing converter*.

In overview, the continuous outflow of charge via the timing resistor R_t and input voltage is balanced by pulses of charge from the reference current I_{ref} . The integrator ensures that, on average, the two charges are balanced. When this is the case, the rate of the pulses of current is proportional to the input voltage.

Detailed operation of the V to F converter is as follows:

- Assume for the moment that switch S_1 is in the **B** position, so that the reference current is being steered to ground.
- The input voltage V_{in} causes a current I_{in} through resistor R_t .
- This causes the output of the integrator to ramp in a positive direction, figure 671(b) (lower trace).
- When the output of the integrator passes through the threshold (trip) voltage, the output of the comparator switches abruptly from its low state to its high state.
- This positive-going transition triggers the monostable into its ON state. The ON state lasts for a fixed period of t_1 seconds.
- During the ON state of the monostable, the switch is driven into position **A**, which causes the reference current I_{ref} to be dumped into the virtual earth point of the integrator.

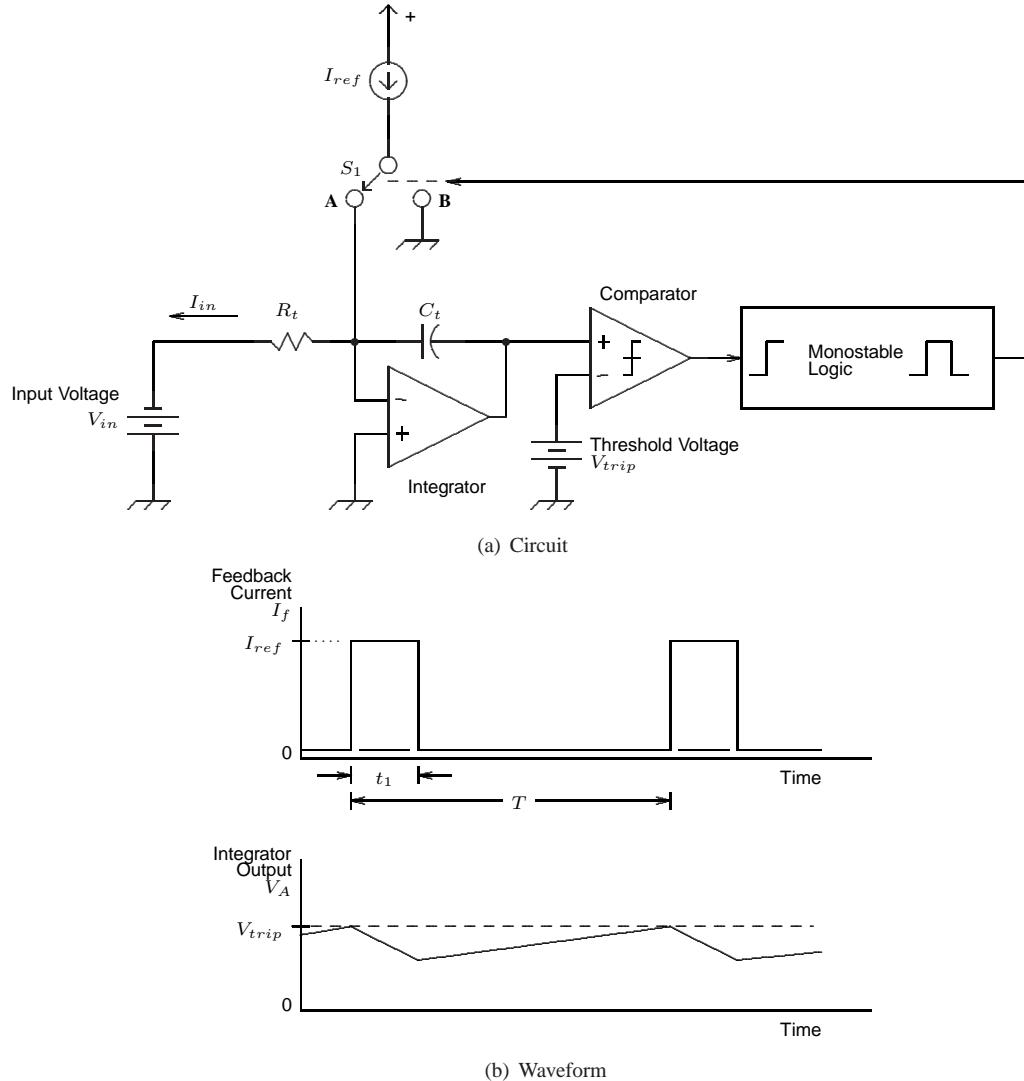


Figure 671: Voltage to Frequency Conversion

- The reference current is chosen to be larger than the maximum value of input current. The difference between these two currents is integrated and now drives the output of the integrator in a negative direction for the duration t_1 of the monostable pulse.

Now we'll derive the transfer function of the converter.

Here is a key idea: the output of the integrator varies over the short term by ramping upwards and downwards, but its average value stays constant. That being the case, at the virtual earth terminal of the integrator, the outflow of charge must equal the inflow of charge. Now, charge is the time-integral of current. In this case, the currents are constant over two different time intervals, so over those time intervals, the charge is simply the product of the current times the time interval.

Referring to figure 671(b) (upper trace), the charge outflow occurs over the interval between the end of the

pulse interval t_1 and the end of the period T :

$$Q_{out} = I_{in} \times (T - t_1) \quad (1106)$$

The charge inflow occurs over the interval t_1 and is equal to the difference between I_{ref} and I_{in} times that time interval:

$$Q_{in} = (I_{ref} - I_{in}) \times t_1 \quad (1107)$$

Equating 1106 and 1107, simplifying and solving for the period T , we have:

$$T = \frac{I_{ref}}{I_{in}} t_1 \quad (1108)$$

We can also write that

$$I_{in} = \frac{V_{in}}{R_t} \quad (1109)$$

and the frequency of the pulse train

$$f = \frac{1}{T} \quad (1110)$$

Substituting from equations 1109 and 1110 into equation 1108 and solving for frequency f , we have:

$$f = \frac{V_{in}}{I_{ref} R_t t_1} \quad (1111)$$

That is, the output frequency f is directly proportional to the input voltage V_{in} . Notice that the magnitudes of the integration capacitance C_t and the threshold voltage V_{trip} have no direct impact on the scale factor of equation 1111.

For example, if we were to choose $I_{ref} = 1\text{mA}$, $R_t = 1\text{M}\Omega$ and $t_1 = 1\mu\text{sec}$, then

$$\begin{aligned} f &= \frac{V_{in}}{I_{ref} R_t t_1} \\ &= \frac{V_{in}}{(1 \times 10^{-3})(1 \times 10^6)(1 \times 10^{-6})} \\ &= 1000 V_{in} \end{aligned}$$

That is, the voltage to frequency converter has a scale factor of 1kHz per volt. For an input voltage of 3 volts, the output frequency would be 3kHz.

Analog to Digital Conversion by Voltage to Frequency Conversion

The voltage-frequency converter can be used as an analog-digital converter by measuring the output frequency. For example, a microprocessor could count the number of pulses in a measurement interval (the measurement *timebase*) of one second. Then the count is equal to the frequency in Hz. This technique has some immunity to jitter in the frequency because the count is accumulated over some period of time and jitter will be reduced by the averaging effect of this accumulation.

Alternatively, the microprocessor could measure the time interval between pulses and then calculate the reciprocal to determine the frequency. This method has no immunity to jitter noise unless a series of period measurements is collected and averaged.

Direct measurement of frequency is most attractive at higher frequencies. Measurement of period (or *reciprocal frequency*, as it is sometimes called) is attractive at low frequencies, where the interval between pulses can be measured easily.

Accuracy and Dynamic Range

The *dynamic range* of the V-F converter is the range of input voltages that the device can accept and still achieve its specified accuracy.

For an accuracy of 0.01% the Analog Devices AD537 claims 80db while the National Semiconductor LM331 claims 100db. For example, if the maximum operating frequency is 10KHz and the dynamic range 100db:

$$20 \log_{10} \left(\frac{f_{max}}{f_{min}} \right) = 100 \quad (1112)$$

Then

$$\frac{f_{max}}{f_{min}} = 10^5 \quad (1113)$$

For example, if the maximum operating frequency is 10KHz, then the minimum frequency is

$$\begin{aligned} f_{min} &= \frac{10 \times 10^3}{10^5} \\ &= 0.1 \text{ Hz} \end{aligned}$$

Since the input voltage scales with the frequency, we can also write

$$\frac{V_{max}}{V_{min}} = 10^5 \quad (1114)$$

For a maximum input voltage of 10 volts, the minimum input voltage would be:

$$\begin{aligned} V_{min} &= \frac{10}{10^5} \\ &= 100 \mu\text{V} \end{aligned}$$

Application

For many industrial measurement situations, the diagram shown in figure 672(a) applies. The sensor, which generates a DC voltage proportional to some quantity such as temperature, is located some distance from the measurement recording device. Both the sensor and the recording device are connected to the so-called *ground* terminal. However, because of currents through the grounding system, the two ground terminals A and B are not at the same potential. Their difference in potential is represented by the noise generator e_{noise} .

The noise generator appears in series with the sensor voltage E_{sensor} and so the noise and sensor voltage add together at the input to the recording system A/D converter. The noise signal is said to corrupt the sensor signal.

In figure 672(b), the system has been modified to use a V to F converter. The signal is converted from analog form to a variable frequency pulse train by the V to F converter. The pulse output from the V to F converter drives the LED side of an opto-isolator. The transistor side of the opto-isolator is connected to the system ground and regenerates the pulse signal which is ultimately sent to the recording system.

In the second arrangement, the sensor ground and V to F ground are completely isolated from the system ground. (They would require an isolated power source, provided by a transformer-rectifier-filter-capacitor circuit,

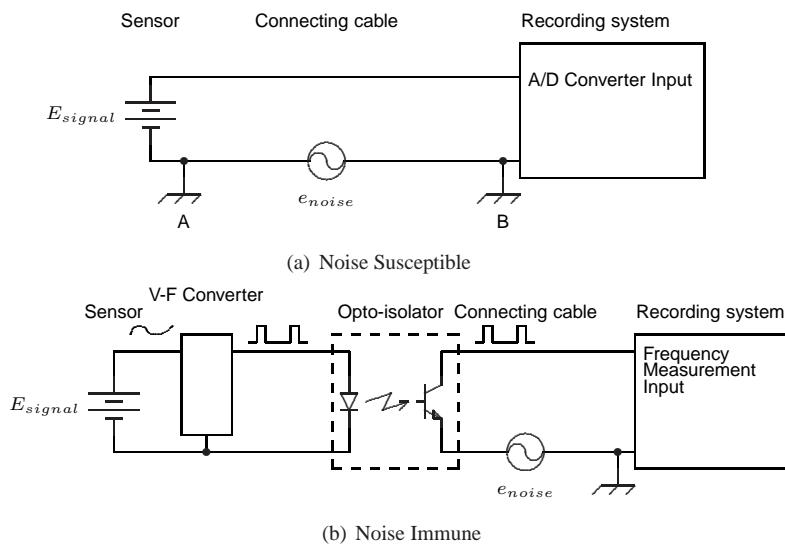


Figure 672: V to F Application

for example.) The noise source now appears in series with the pulse-generating transistor of the opto-isolator. Providing that the pulse amplitude is large compared to the noise voltage, the pulse counting circuit in the recording device will ignore the noise voltage.

26.9 Exercises

1. Name 3 types of A/D converters.
2. What type of A/D converter is most suitable for use in a hand-held digital voltmeter? Why?
3. What type of A/D converter is most suitable when the data must be transmitted over a fiber-optic cable? Why?
4. The circuit of figure 673 implements a microprocessor controlled dual slope A/D converter.

During the first phase of a measurement cycle (phase 1), the switch is in position **a** for 12 milliseconds. During the second phase of the measurement cycle (phase 2), the switch is in position **b**.

- (a) Describe the function of each operational amplifier in this circuit.
- (b) The operational amplifiers are operated from ± 15 volt supplies, and the output of each op-amp can swing to within 2 volts of a power supply rail. If the input voltage E_{in} has a maximum value of 5 volts, what is the longest time allowed for phase 1 of the measurement cycle?
- (c) If the input voltage is 3 volts, and phase 1 of the measurement is 12 milliseconds in duration, what is the total time for a complete measurement cycle, phases 1 and 2?
- (d) For the circuit shown in figure 673:
 - phase 1 is 12 milliseconds in duration.
 - maximum (full scale) input voltage is 5 volts.
 - the clock period is $0.5 \mu\text{Seconds}$

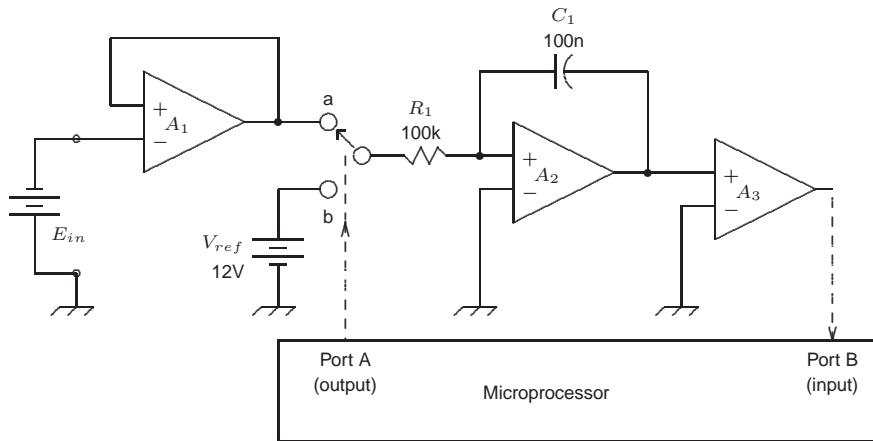


Figure 673: Dual Slope A/D Converter

What is the resolution of the A/D converter, in volts per timer count?

- (e) How would you modify this circuit and/or the microprocessor program to reject 60 Hz noise in the input signal? Assume that you can change any circuit component or any aspect of the computer program.
- (f) List and explain which offset voltages and bias currents will affect the accuracy of the conversion, and which can be ignored.

Amplifier:	A1	A2	A3
Offset Voltage			
Bias Current			

5. A voltage-frequency converter is shown in figure 674. The monostable U_2 produces a pulse of width 2 milliseconds every time it is triggered with the positive edge of a pulse from the comparator A_2 . The switch moves from position **a** to position **b** while the output of the monostable is high. The circuit may be assumed to have been initialized by friendly aliens who have now left the planet. The circuit is operating continuously when observed here.

- (a) Explain the term *charge balance* with reference to figure 674.
- (b) Explain briefly how the circuit converts E_{in} to a frequency.
- (c) What are the maximum and minimum values of E_{in} ?
- (d) If E_{in} is 1 volt, what is the output frequency F_{out} ?
- (e) Recalculate the output frequency if E_{in} is 1 volt and the offset voltage of the op-amp A_1 is 4 millivolts.
- (f) Recalculate the output frequency if E_{in} is 1 volt, the offset voltage of the op-amp A_1 is zero, and the bias current of the op-amp A_1 is 100 nA. (Bias current flows **out** of the input terminals of the op-amp.)

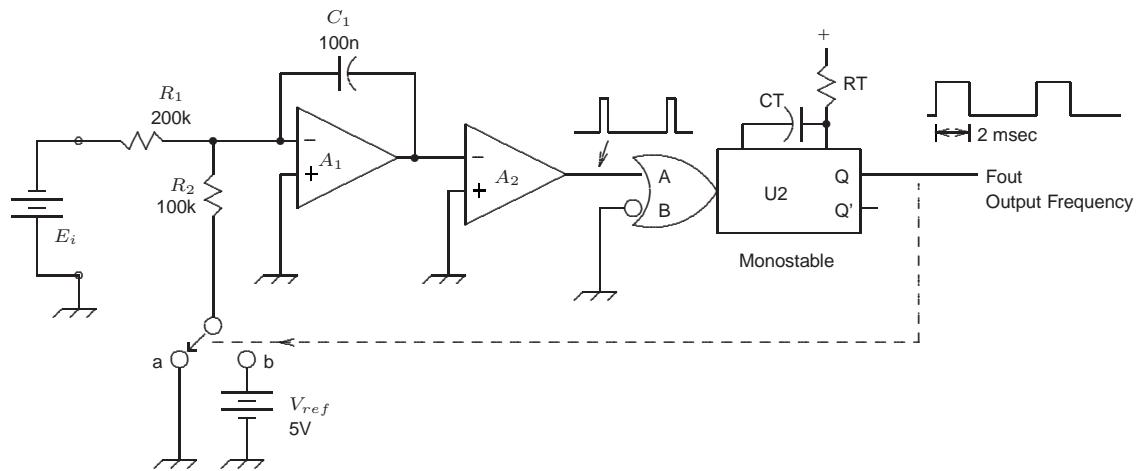
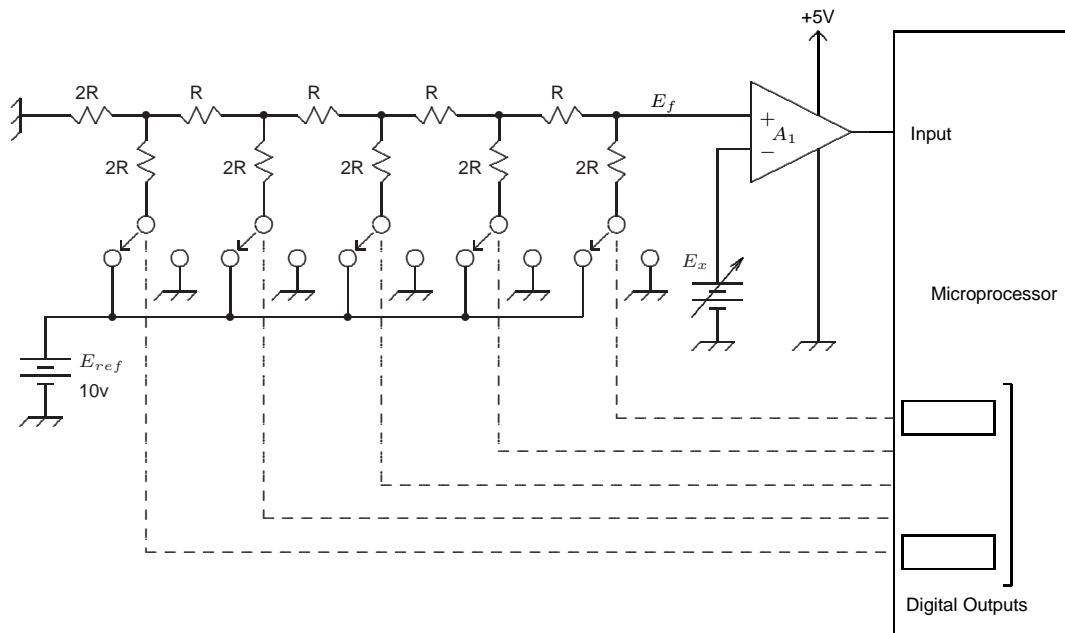


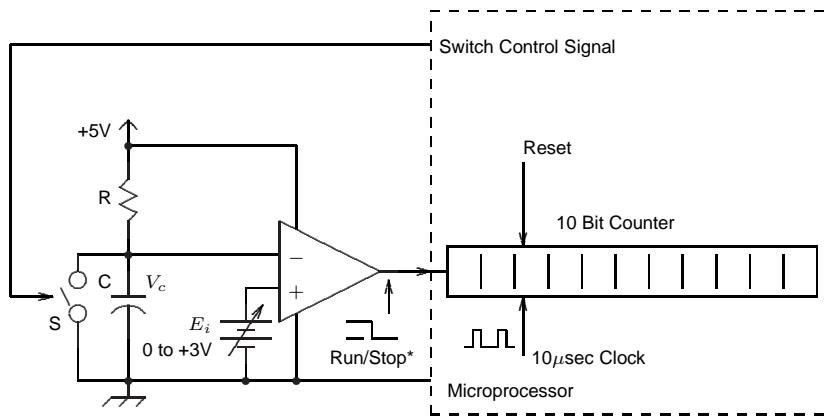
Figure 674: Voltage-Frequency Converter

6. The circuit shown below is a microprocessor-controlled A/D converter. Amplifier A_1 is a comparator that accepts input from an unknown voltage E_x and compares that with a feedback voltage E_f generated by the D/A converter.



- (a) What is the maximum value of E_f ?
 (b) On the diagram in the boxes, indicate the MSB and LSB of the digital word.

- (c) What is the resolution of the system, in volts?
- (d) If the unknown input voltage E_x is 2.6 volts and the microprocessor uses the up-counting algorithm, what will be the binary number at the output of the microprocessor when the conversion is complete?
7. The circuit diagram of an A/D converter is shown below.



Circuitry inside the dotted line is inside a microprocessor. The op-amp is used as a comparator. To make an A/D conversion,

- The microprocessor opens switch S.
- The microprocessor 10-bit counter begins counting 10 μ Sec clock pulses.
- When the capacitor voltage V_c reaches equality with the input voltage E_i , the comparator switches state.
- A flag bit is connected to the output of the comparator. (For simplicity, the flag bit is not shown on the diagram). When the comparator changes state this stops the counter and indicates to the microprocessor that the conversion is complete.
- The microprocessor reads the counter value.

The range of the signal E_i into the A/D converter is 0 to +3 volts.

- Calculate the time constant RC so that the maximum input signal +3V corresponds to the maximum A/D output reading.
- Treating this A/D as a linear device, what is the approximate resolution in volts per step?
- Explain why the relationship between input voltage and output count is not completely linear.
- Very briefly describe how would you linearize this system. (Any feasible method is allowable.)
- A particular sensor signal assumes values between +3 volts and 3 volts. Design an op-amp circuit that can convert the sensor signal into the input voltage range of this A/D converter. The circuit should use the +5 volt power supply. It is sufficient to show resistor ratios. The op-amp may be considered to be ideal.

27 Opto-Electronic Devices

27.1 Introduction

In this section, we'll look at some useful opto-electronic devices. These are devices that either emit light or detect it. The term *light* is used loosely here, and refers also to electromagnetic radiation at wavelengths that are not visible, such as ultraviolet and infrared, but are close to visible light.

For reference purposes, the visible spectrum and its neighbours are shown in figure 675. The *standard observer* curve shows the normal sensitivity of the human vision system [236].

The response of a phototransistor is shown on the same graph. (The sensitivities have been normalized to 100%). It's clear that the phototransistor sees a different spectrum than the human eye, and is much more sensitive to infrared radiation.

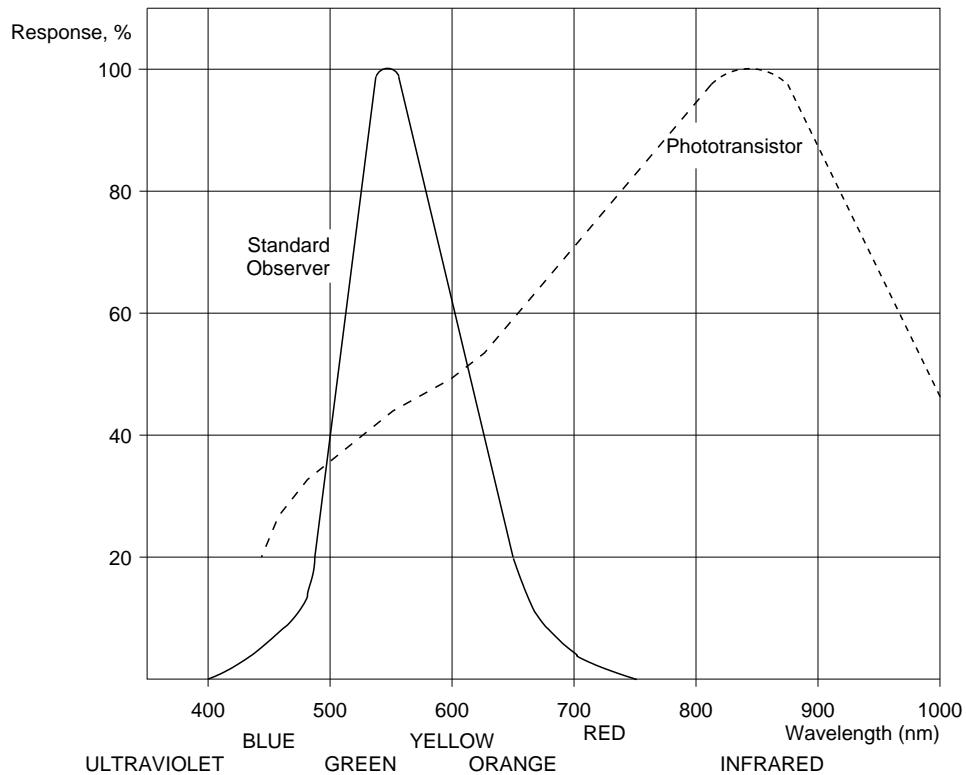


Figure 675: The Visible Spectrum and Neighbourhood

27.2 Light Emitting Diode

The *light-emitting diode* is one of the great contributions of semiconductor physics. Current flows through a semiconductor and light is emitted.

Although the LED is a diode, it differs from the small-signal silicon diode in two respects:

- The forward voltage drop of an LED is significantly larger than the small-signal diode, typically 1.6 volts to 4 volts compared to 0.6 volts for a small signal diode.
- The maximum allowable reverse voltage is much smaller for an LED, typically 5 volts compared to at least 50 volts for a small signal diode.

In many applications, we simply want an LED indicator. An appropriate current for the LED is between 5 and 20mA, and the forward voltage drop at those currents is around 2 volts.

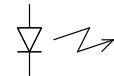


Figure 676: Symbol, Light Emitting Diode

Colour	Wavelength, nM	V_f at I_f
Infrared	940	1.25V at 50mA
Red	700	2.1V at 20mA
Orange	635	2.0V at 20mA
Green	565	2.1V at 20mA
Yellow	585	2.1V at 20mA
Blue	475	3.0V at 20mA
Ultraviolet	390	4.2V at 30mA
White	400-700	3.6V at 20mA

Infrared LEDs are commonly used in television remote controls, to transmit a coded light beam that is detected at the television set.

Ultra-violet LEDs are used in detectors for counterfeit paper money and for medical applications.

Because they are much more efficient and reliable than incandescent sources of light, white LEDs were very quickly adopted into flashlights and other lighting applications²¹⁰.

LEDs are available in a wide variety of package styles and with various lenses to disperse or concentrate the light output. For example, so-called *high intensity LEDs* include a lens that focuses the light energy into a narrow beam, and this restricts the viewing angle.

There are also *flasher LEDs*, which include a small electronic circuit to interrupt the current at a regular rate. These typically require at least 3 volts to function and the flash rate varies with supply voltage.

The maximum allowable average current is typically 50 to 100mA, with peak currents to as much as 2 amps if the duty cycle is restricted so that the average current is not exceeded.

In the large outdoor displays that show moving images in colour²¹¹, it is common to use a triad of red, green and blue LEDs at each pixel²¹² location to create the various colours.

Driving Multiple LEDs

There are many applications where multiple LEDs must be operated from a single power supply. Should they be connected in parallel or series?

²¹⁰As this is being written, November 2009, street lighting using white LEDs is being tested in Toronto.

²¹¹To my knowledge, the dystopian film *Blade Runner* previewed these displays, which are now common in cities and by the highway.

²¹²A *pixel* is a *picture element*, effectively one dot of resolution on the display.

The light output of an LED is proportional to its current, not its terminal voltage. Furthermore, the LED is (approximately) a constant voltage device. If LEDs are connected in parallel the inevitable small unit-unit variation in forward voltage will result in a large difference between the individual LED currents and their light output.

Connecting LEDs in series guarantees that they all conduct the same current and therefore have similar luminescence. The available supply voltage then determines the number of LEDs that can be connected in series. Integrated circuit power supplies are available specifically for this application. They incorporate a switching power supply and constant current driver optimized for driving a string of LEDs.

27.3 Photodiode

When light falls on the junction of a diode, a current is generated and the diode becomes forward-biased. Many small-signal diodes are encapsulated in transparent glass, and external illumination generates a small voltage across the junction. Light emitting diodes generate a significant voltage when exposed to light.

A *photodiode* is a diode that has been optimized to detect illumination. As a photodetector, it has the following properties:

- predictable behaviour over a wide range of illumination levels
- fast response (in the appropriate circuit)
- broad response: available for wavelengths from x-rays through to infrared

However, unlike the CdS photoresistor and phototransistor, the output current or voltage signal must be amplified to be useful.

The VI characteristics of a photodiode are shown in figure 678. First, notice the reference polarities for diode voltage and current, shown next to the symbol.

Now we can interpret this voltage-current characteristic.

- In region A, which occurs at about 0.6 volts, the diode is connected to a voltmeter (an open circuit) so its output current is zero. The voltage across the diode increases with increasing light level and the response is logarithmic. In terms of the model, the internal diode begins to conduct, limiting the maximum voltage across the device. This is known as *photovoltaic* operation, since the diode voltage increases (in a non-linear fashion) with light level.
- In region B, the diode is connected to an ammeter (a short circuit) so the output voltage is zero. The current through the short circuit is proportional to light level. This is known as *photoconductive* operation, since the diode current is proportional to light level.



Figure 677: Photodiode

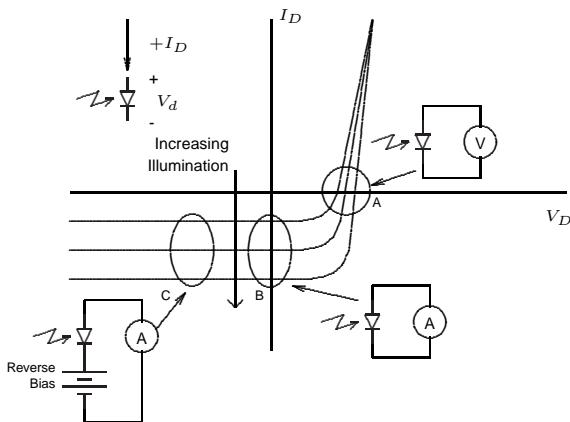


Figure 678: Voltage-Current Characteristic

- In region C, the diode is reverse biased (V_D is negative) and this is the photoconductive mode of operation. The reverse current is proportional to light level. Reverse biasing is used in high-speed applications because it reduces the internal capacitance, allowing the photodiode to switch at higher speeds. However, this reverse bias also increases the noise current of the diode, effectively making it less sensitive.

The simple equivalent circuit for a photodiode is shown in figure 679.

The output of the current generator I_P is proportional to light level. If the load is an open circuit (a voltmeter), then the output voltage will be developed across the diode and constrained to be about 0.6 volts, and a logarithmic function of the current. If on the other hand the load is constrained to be a short circuit, then the diode will never conduct because the voltage across it is fixed at zero. Then the current into the load will be proportional to light level.

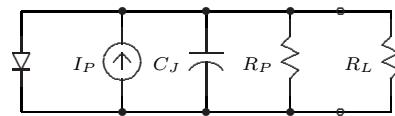


Figure 679: Equivalent Circuit

Amplifier Circuits

Various photodiode amplifier circuits are shown in figure 680.

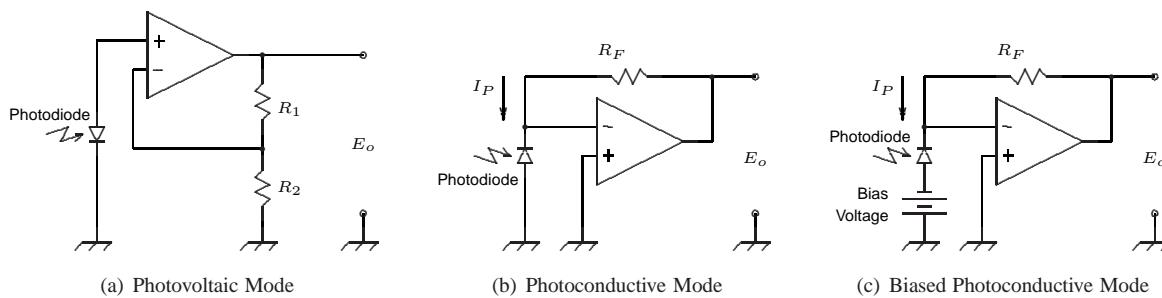


Figure 680: Photodiode Amplifiers

A photovoltaic mode amplifier is simply a non-inverting amplifier with the photodiode voltage as input, figure 680(a). Since the diode is essentially open-circuited, its response will be logarithmic and the voltage across the diode about 0.6 volts.

The photoconductive amplifier, figure 680(b), is a current-voltage converter circuit that maintains the voltage across the photodiode at zero volts and translates the diode current into an output voltage according to

$$E_o = R_f I_P \quad (1115)$$

For the diode polarity as shown in figure 680(b) the output voltage is positive.

In applications where fast switching is required, the photodiode may be reverse biased by connecting the anode to a negative voltage as shown in figure 680(c).

27.4 Photoresistive Cell

The *photoresistive* cell, aka *cadmium-sulphide*, *CdS* cell, aka *photoresistor* is a resistor that decreases in resistance in the presence of light.

- It is very sensitive to light and therefore often requires little in the way of external circuitry.
- The spectral response is over a wide range of visible light, so it's attractive for devices that must respond like the human eye.
- The speed of response is very slow, so it is essentially a *steady state* light detector.

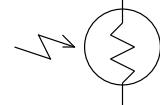


Figure 681: Photoresistor

Spectral Response

The spectral responses of a Cadmium-Sulphide (CdS) cell are shown in figure 682 along with the spectrum of a high-intensity red (gallium-arsenide) LED for comparison purposes (from references [237], [238], and [239]).

Transient Response

The CdS cell responds relatively slowly to changes in light level: a few milliseconds to respond to light, about 50 milliseconds to return to the dark state.

Application: Photorelay

The circuit of figure 683 shows a simple photosensitive relay circuit.

- When the photoresistor R_P is in darkness, its resistance is high. R_P forms a voltage divider with R_1 . We choose R_1 so that under these conditions the voltage at the gate of the MOSFET will be small, below the threshold voltage V_T of the MOSFET.
- When light falls upon the photoresistor, its resistance becomes very low, and the voltage at the gate of the MOSFET rises above V_T .
- The MOSFET conducts, turning on the relay RY_1 , and this activates some external circuit via the relay switch S_1 .
- The diode D_1 provides a path for inductive current to follow when the MOSFET turns off.

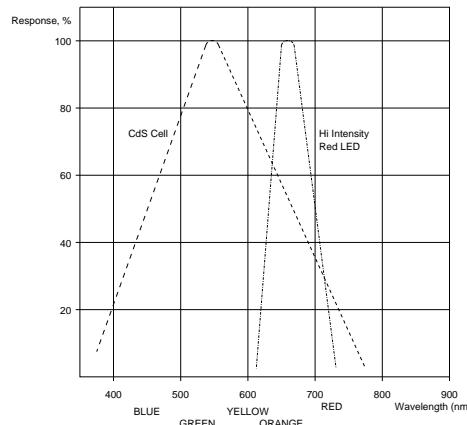


Figure 682: CdS Cell Spectral Response

This circuit will work well if there are large differences between *light on* and *light off*, and the light levels switch quickly from one to the other so that the MOSFET is either cutoff or saturated. However, if the light level is marginal, the MOSFET may be somewhere in the active region between cutoff and saturation, and the relay may *chatter*.

If this is a problem, you would add a Schmitt trigger between the voltage divider and the MOSFET gate. This will ensure that the MOSFET is fully on or fully off, and provide some hysteresis between the trigger levels.

27.5 Phototransistor

A *phototransistor* may be regarded as an NPN BJT transistor where the base current is replaced by illumination. Illuminating the transistor causes collector current to flow.

Phototransistors are much more sensitive than photodiodes (section 27.3) but are slower in response time and less predictable in their transfer characteristic. A related device, the photodarlington transistor (two cascaded transistors, figure 686, described in section 30.1), is more sensitive and slower still.

The sensitivity of a phototransistor relates the optical power input to the collector current output. Typical sensitivities and switch speeds are tabulated in figure 685 below. (Values are taken from representative devices in reference [240]).

Device	Sensitivity, $\mu\text{A}/\text{mW}/\text{cm}^2$	Sensitivity Ratio	Switching Speed T_R/T_F	Speed Ratio
Photodiode	0.6	1	50/50nSec	5000
Phototransistor	30	50	8/50 μSec	5
Photodarlington	16000	26667	45/250 μSec	1

Figure 685: Device Comparison

The tradeoff is very clear: speed must be traded for sensitivity among these photodetectors.

The base lead of the transistor may or may not be available, depending on the device. If speed is not an issue, then it's not necessary. However, the speed of a phototransistor can be improved by connecting a resistor between base and emitter, so the base lead may be useful [113].

Spectral Response

The spectral response of a phototransistor peaks in the infrared region, but there is some overlap with the visible spectrum, figure 687 (adapted from [240]). The spectrum is a good match for an infrared LED.

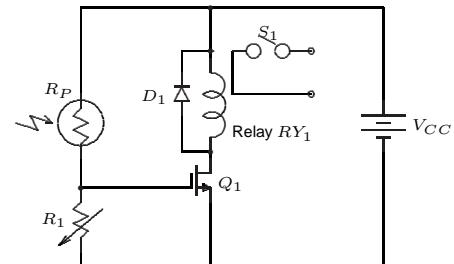


Figure 683: Photosensitive Relay

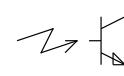


Figure 684: Phototransistor

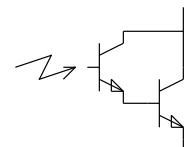


Figure 686: Photodarlington

When the phototransistor is intended to detect the light from an infrared LED, the lens of the phototransistor is coated with a dark film that attenuates visible light but allows infrared light to reach the transistor.

Reference

An excellent description of phototransistor characteristics with applications is in reference [236].

27.6 Optocoupler

An *optocoupler* (also known as an *opto-isolator*) is an LED-phototransistor pair arranged in a single package so that light from the LED activates the phototransistor. Two examples are shown in figure 688(a): an LED-BJT pair, and an LED-JFET pair.

The optocoupler is useful where a signal (usually a switching signal, but also possibly an analog signal) must be transferred between two circuits but the input circuit and output circuit must be isolated from each other. A simple example is shown in figure 688(b).

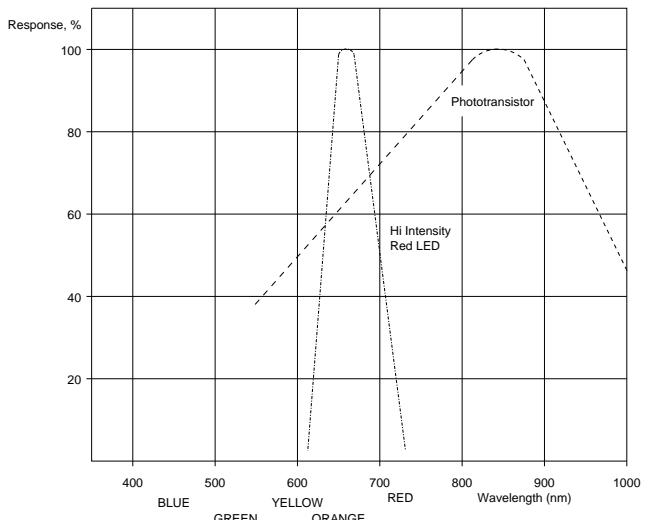
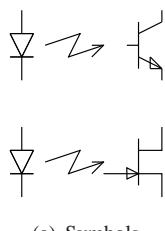
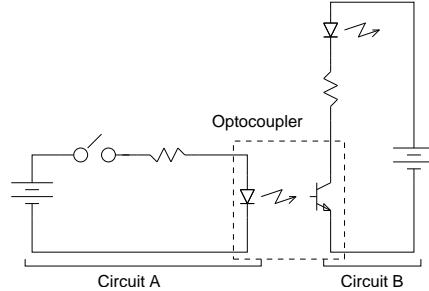


Figure 687: Phototransistor Spectral Response



(a) Symbols



(b) Example Circuit

Figure 688: Optocoupler

In this circuit, closing the switch in Circuit A will cause the LED to illuminate in Circuit B. However, there is no common ground connection between them. The signal is transferred by photons between the LED and phototransistor. In fact, (depending on the isolation specification of the opto-isolator), there could be a voltage difference between the two circuits, that is, the negative lead of battery A could be raised hundreds of volts above the negative lead of battery B, without affecting the circuit operation.

The optocoupler is also used where high currents or voltages (such as those that drive a large motor) are being switched by some logic circuit. If there is a fault in the motor power circuitry, it has the potential of doing serious damage to the drivers unless the drivers are opto-isolated.

Current Transfer Ratio

The key specification for an optocoupler is the *current transfer ratio* or *CTR*. This is the ratio of collector current in the phototransistor to diode current in the LED.

$$CTR = \frac{I_{cpt}}{I_{led}} \quad (1116)$$

where I_{cpt} is the collector current in the phototransistor and I_{led} is the current in the LED.

A typical value of CTR is 10%, that is, 5mA in the LED will create 1mA collector current in the phototransistor. The CTR differs for different optocouplers, so it should always be taken from the spec sheet for the device.

Opto-isolators are not particularly speedy devices, and this is a serious limitation when they must transfer digital data at high speed. However, there are various tricks that can be used to make them switch more quickly. In particular, if the circuit configuration holds the voltage across the phototransistor at a constant value, then the Miller capacitance of the transistor (the junction capacitance C_{CB} between collector and base terminals) does not need to charge and discharge and the device will switch more quickly.

As mentioned in connection with the phototransistor, the switching speed is reduced by connecting the base terminal of the phototransistor to its emitter, although this will also reduce the CTR. And there are special optocouplers that are optimized for high speed operation – they use a photodiode detector with an amplifier, rather than a phototransistor.

Opto-interrupter

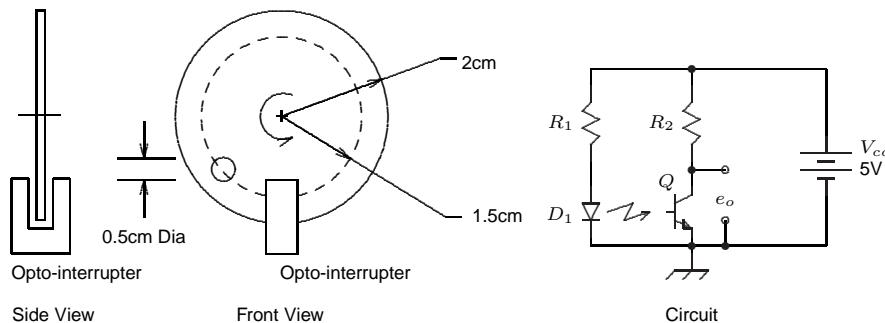
An opto-interrupter is an optocoupler which is constructed so there is a slot in the package between the LED and phototransistor. The light beam can be interrupted by a shutter of some kind. For example, if the shutter is a wheel with holes in it, then the optical beam will be pulsed as the wheel rotates and the output signal will be a series of pulses for which the frequency is proportional to the speed of the wheel, or a microprocessor can count the number of pulses and thereby determine the amount the wheel has rotated.

Opto-interrupters tend to be more reliable than mechanical switches, so they are often used to detect mechanical position in printers, copiers and other devices.

As in the case of the optocoupler, the specification of the CTR of an opto-interrupter should be consulted to ensure that the illuminating LED can saturate the detector phototransistor when the light path is uninterrupted.

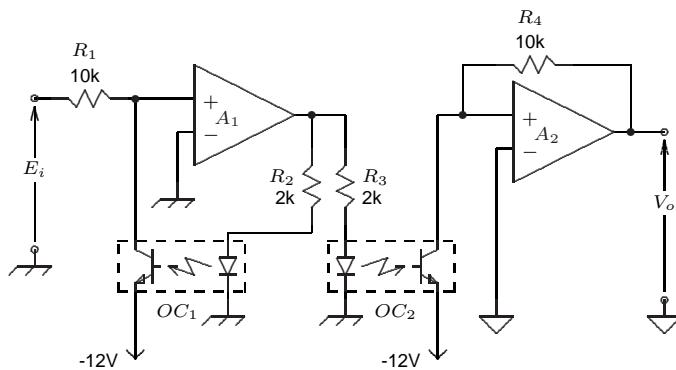
27.7 Exercises

1. The system shown below is to generate a pulse with each revolution of the disk. The small hole on the disk is aligned so that it passes through the optical axis of the photo-interrupter.

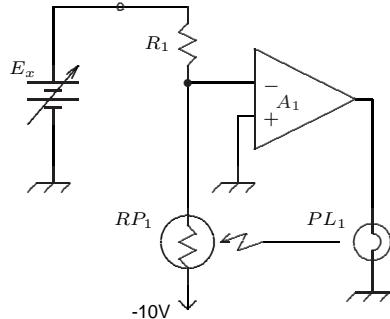


- (a) Assuming diode D_1 has a forward voltage drop of 1.7 volts when it is conducting, calculate R_1 so that the diode current is 20mA.

- (b) Assuming that the CTR (current transfer ratio) for the photo-interrupter is 0.2, calculate R_2 so that the transistor conduction is maximum (saturated) when the opto-interrupter gap is not blocked.
- (c) If the wheel rotates at 100RPM, draw the voltage-time waveform for e_o .
2. The circuit of a *linear opto-isolator* is shown below. The voltage E_i at the input is replicated as V_o at the output. For this to work,
- The input voltage must be positive.
 - The opto-isolators must be matched, with equal CTR (current transfer ratio) figures.
 - The op-amp power supplies must be isolated from each other.

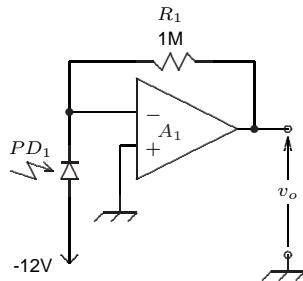


- (a) Amplifier A_1 is in a negative-feedback loop. Show why that is the case.
- (b) Verify that the circuit works as described above.
- (c) The opto-isolator relationship between input (led) current and output (phototransistor) current is somewhat non-linear. What effect will this have on the relationship between input voltage and output voltage?
- (d) What sets an upper limit on the input voltage E_i ?
- (e) Suggest an application for this circuit.
3. Consider the circuit shown below.

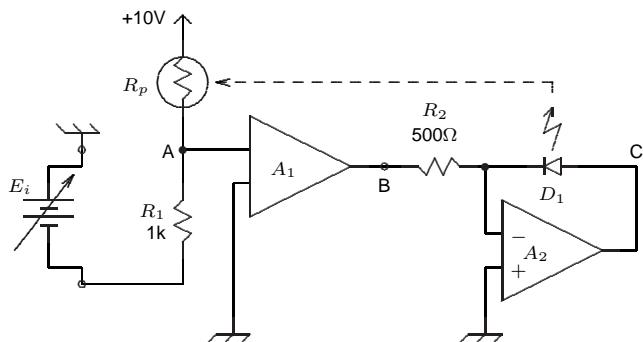


Resistor RP_1 is a photoresistive cell (section 27.4) that is illuminated by the incandescent lamp PL_1 . The resistance of the cell is a non-linear function of its illumination. The op-amp may be considered to be ideal. Hint: This is a negative feedback system.

- (a) Derive an expression for R_P showing its relationship to the input voltage.
 (b) What would be the effect of swapping the inputs to amplifier A_1 ?
 (c) Why could this circuit be described as a *linearization* circuit?
4. The circuit of a photodiode amplifier is given below. The diode is reverse biased and its reverse current is a function of the incident light.



- (a) According to the *National Semiconductor Linear Applications Handbook* [199],
the photodiode should be biased with a fixed voltage and work into a low-impedance load.
 Explain how this circuit satisfies both those requirements.
- (b) Which of the following amplifiers would result in less output noise:
 - Amplifier A, JFET inputs: $e_n = 20\text{nV}/\sqrt{\text{Hz}}$, $i_n = 0.01\text{pV}/\sqrt{\text{Hz}}$
 - Amplifier B, BJT inputs: $e_n = 4.5\text{nV}/\sqrt{\text{Hz}}$, $i_n = 0.7\text{pV}/\sqrt{\text{Hz}}$
 Justify your answer.
- (c) The noise specifications for the amplifier used in the circuit are $e_n = 12\text{nV}/\sqrt{\text{Hz}}$, $i_n = 3\text{nA}/\sqrt{\text{Hz}}$.
 The photodiode PD_1 may be treated as a noiseless constant current source. If the ambient temperature is 20° and the measurement bandwidth is 10kHz , determine the RMS value of the output noise voltage v_o .
5. The circuit shown below is a negative feedback system. The resistance of CdS photoresistor R_p is kept constant by feedback from a light emitting diode D_1 . You may assume that the system is in steady state and operating correctly – not saturated or cutoff. The voltage source E_1 is the setpoint or reference for the system. The CdS photoresistor decreases in resistance as the intensity of the light falling on it increases.



- (a) On the schematic, mark the inverting and non-inverting terminals of A_1 so that the system is negative feedback. Hint: it's not obvious.

- (b) What is the function of the amplifier A_2 and its associated components? (Be precise in your description.)
- (c) The LED is expected to operate at a current around 10mA. What voltage readings would you expect at A, B and C? Explain.
- (d) Derive an expression for the resistance R_p of the photoresistor in terms of E_i and the other circuit components.

28 Cooling

It is important to keep the temperature of electronic circuits within reasonable limits. Components deteriorate more quickly at elevated temperatures, and so the cooling of electronic equipment (*thermal management* is the technical term) is an important aspect of the design.

There are both *passive* and *active* cooling techniques. Passive cooling is preferred since it is relatively simple, inexpensive and reliable. The device to be cooled is attached to a conducting material that has a large surface area, known as a *heatsink*. The heatsink may be an integral part of the electronic enclosure, or it may be a separate unit as shown in figure 689.

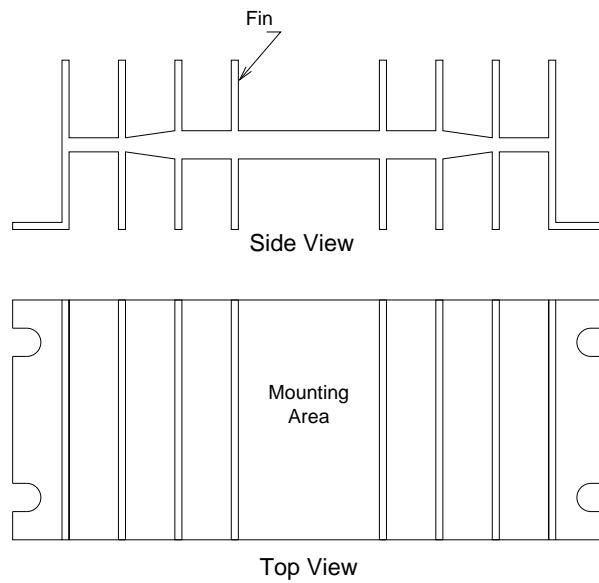


Figure 689: Heat Sink

Heat from the device transfers to the heatsink by conduction and then to the surrounding environment, primarily by convection. When the heatsink is surrounded by air, most heat transfer occurs by convection so it is important that there be a clear path for air to flow over the heatsink. When the heatsink is in a vacuum – in outer space, for example – the only heat-transfer mechanism is radiation.

When passive cooling is not sufficient, the heatsink air flow can be increased by a fan, which substantially increases the cooling effect (section 28.6, page 791).

Other, more exotic cooling techniques are possible. For example, certain models of the Cray computer circulated freon through the heatsinks and a refrigerator [241].

Thermal management often turns out to be a *killer problem* that is discovered in the final stages of a project when the system is being debugged and commissioned. The overall size and design of the case may have been decided at this point and so adding some sort of cooling mechanism can be very difficult. It is essential that cooling be considered early in the design.

For example, let us assume that marketing has chosen a particular attractive plastic case for a microprocessor system. An engineering calculation establishes that the power given off by the circuitry will raise the internal temperature well above 100°, which is unacceptable. Then the plastic case must be designed to have an external heatsink or to contain a fan. Alternatively, the case can be constructed of aluminum, which can act as a heatsink

without fans or additional cost. Whatever the decision, it must be made early in the design.

28.1 Thermal Circuits

Before launching into the details of the heatsink design, we will review *thermal equivalent circuit* concepts. Using this approach, we can model thermal behaviour with electric circuits familiar to electrical engineers.

Fourier's Law of heat transfer states that the flow of heat through a heat conducting substance, due to thermal conduction, is proportional to the temperature gradient across the conductor.

This may be stated as

$$P = k\Delta T$$

where

P is the power (rate of flow of energy) in watts

ΔT is the temperature gradient (temperature difference) between the ends of the conductor, degrees C

k is the constant of proportionality, known as the *thermal conductivity*, watts per degree C

For our purposes, it is convenient to rewrite this equation as

$$P = \frac{1}{\theta} \Delta T \quad (1117)$$

where θ is the inverse of k , and is called the *thermal resistance*, degrees C per watt.

This form of the equation resembles Ohm's law, where voltage is similar to temperature gradient, power is similar to electrical current, and thermal resistance is similar to electrical resistance. For an electrical engineer, this helps to provide an intuitive understanding of thermal behaviour. For example, if the thermal resistance is increased for a given power dissipation, the temperature gradient will increase.

Like voltage, the temperature gradient is measured between two points. In many cases, the reference temperature (similar to voltage *ground*) is the ambient temperature, T_a .

This equivalent circuit assumes that the system is in thermal steady state, that is, the devices have reached their final operating temperature.

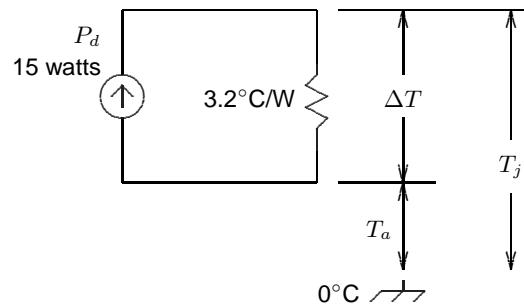


Figure 690: Thermal Circuit Example

Example

A certain power transistor²¹³ is dissipating 15 watts and has a thermal resistance to ambient of 3.2°C/watt. Calculate the temperature rise above ambient of the junction.

Solution

The equivalent circuit is shown in figure 690.

²¹³Throughout this section, we assume that the device to be cooled is a transistor. However, exactly the same principles apply to any device – resistor, zener diode, integrated circuit, whatever – that must be cooled.

Rewriting equation 1117 to solve for the temperature gradient, we have

$$\begin{aligned}\Delta T &= P \times \theta \\ &= 15 \times 3.2 \\ &= 48^\circ\text{C above ambient}\end{aligned}$$

For example, if the ambient temperature is 40°C , then the transistor junction is at $40 + 48 = 88^\circ\text{C}$.

Example

What is the effect on the thermal equivalent circuit of adding an 'insulating washer' between the transistor and the heatsink? The thermal resistance of the washer is 0.5°C/W .

Solution

The thermal washer appears in series with the thermal resistance of the device. By analogy with resistors in series, the total thermal resistance is the sum of the individual thermal resistances, or $3.2 + 0.5 = 3.7^\circ\text{C/W}$. Then the calculations proceed in a similar manner to the previous example.

28.2 The Derating Curve

The thermal resistance of a device is vital information in designing a suitable heatsink. In practice, the thermal resistance is often given in the form of a *power derating curve* for the device. For example, the table shown below is derived from the datasheet for the 2N4401 transistor (section 29.41 on page 907).

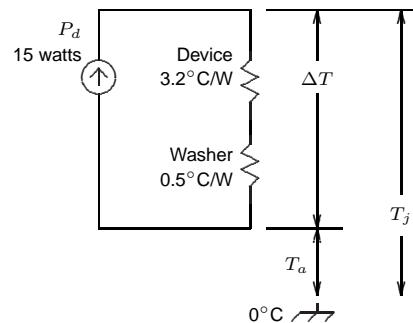


Figure 691: Effect of Insulating Washer

Rating	Symbol	Value	Unit
Operating and storage junction temperature	T_j	+150	$^\circ\text{C}$
Total device dissipation at $T_c = 25^\circ\text{C}$	P_d	1500	mW
Derate above 25°C		12	$\text{mW}/^\circ\text{C}$
Thermal resistance, junction to case	$R_{\theta jc}$	83.3	$^\circ\text{C/W}$
Total device dissipation at $T_a = 25^\circ\text{C}$	P_d	625	mW
Derate above 25°C		5	$\text{mW}/^\circ\text{C}$
Thermal resistance, junction to ambient	$R_{\theta ja}$	200	$^\circ\text{C/W}$

The power derating curve for this device is shown in figure 692. This derating curve shows two curves. The upper trace is the maximum allowable power dissipation versus *case* temperature. The lower trace is the maximum allowable power dissipation versus *ambient* (surrounding air) temperature.

- Both curves show that the allowable device dissipation decreases with temperature above 25°C.
- The maximum allowable power is 1500mW if the case temperature is maintained at 25°C. It is 625mW if the ambient temperature never exceeds 25°C.
- Both curves pass through the maximum allowable junction temperature, 150°C.
- The slope of the derating curve is the inverse of the thermal resistance. So a slope of -12mW/°C for the relationship between power and case temperature is equivalent to a thermal resistance θ_{jc} of 85°C/Watt junction to case.
- The slope of -5mW/°C for the relationship between power and ambient temperature is equivalent to a thermal resistance of 200°C/Watt junction to ambient θ_{ja} . As expected, the thermal resistance junction to ambient is much larger than junction to case.
- Should it be required, the thermal resistance θ_{ca} case to ambient may be determined from the foregoing by:

$$\theta_{ca} = \theta_{ja} - \theta_{jc} \quad (1118)$$

Applications

The derating curve is especially useful when it can be used directly. For example, suppose we need to know the maximum power dissipation allowable in a free-standing 2N4401 if the worst case air temperature is 40°C. From the derating curve, the maximum power is given by

$$\begin{aligned} P_d &= 625\text{mW} - 5\text{mW} \times (40 - 25)\text{°C} \\ &= 550\text{mW} \end{aligned}$$

However, when a heatsink is required, then the thermal resistance of the device and the maximum allowable junction temperature must be used to determine the size of the required heatsink.

Notice that it is highly unlikely that a 2N4401 could actually dissipate 1500mW. That would require that the case temperature is maintained at or below 25°C. To conduct the heat away from the case, the cold end of the heat conducting path would be at a temperature below 25°C. Ambient air is usually at or above 25°C, so this arrangement would require a refrigeration unit.

28.3 Designing a Heatsink: Calculating Thermal Resistance

A heatsink and power transistor assembly is shown in figure 693 on page 787. We now show how to calculate the necessary thermal resistance and thereby design a suitable heatsink.

Based on the previous example, we can write the following equation for junction temperature:

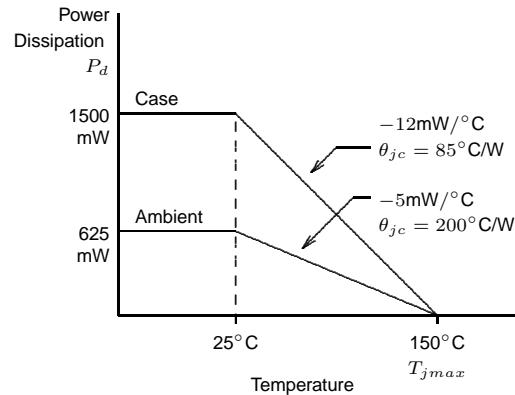


Figure 692: Derating Curve

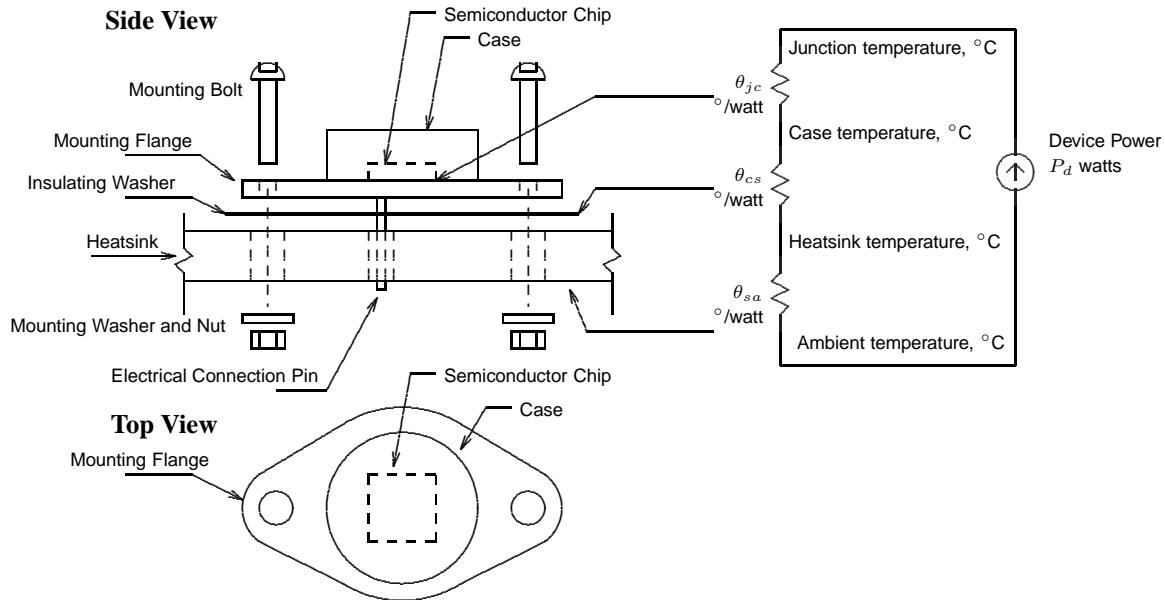


Figure 693: Device to Heatsink Assembly

$$T_j = T_a + \theta_t P_d \quad (1119)$$

where

- T_j is the junction temperature, $^{\circ}\text{C}$
- T_a is the ambient temperature, $^{\circ}\text{C}$
- θ_t is the total thermal resistance between the junction and ambient, $^{\circ}\text{C}/\text{watt}$
- P_d is the power dissipation in the transistor, watts

In figure 693, the total thermal resistance θ_t is made up of several thermal resistances in series: θ_{jc} between the junction and case, θ_{cs} between the case and heatsink, and θ_{sa} between the heatsink and ambient air. Like electrical resistors in series, these resistances add to form the total resistance:

$$\theta_t = \theta_{jc} + \theta_{cs} + \theta_{sa} \quad (1120)$$

Equations 1119 and 1120 are sufficient for the design of a heatsink. The design proceeds as follows:

1. Determine the maximum allowable junction temperature T_j . This is available from the transistor data sheet, and might be something like 150°C . To allow for some safety margin and prevent long-term deterioration of the device, derate this by at least 10%.
2. Choose a worst case (maximum) ambient temperature. In open air in a temperate climate, that might be something like 40°C . In a closed space (a locked car, for example) this will be higher.
3. Determine the worst-case maximum power dissipation P_d in the device.
4. Use equation 1119 to solve for the total thermal resistance θ_t . This is the maximum thermal resistance that will meet the constraints of junction temperature, power dissipation and ambient temperature.

5. From the transistor data sheet, determine the junction-case thermal resistance θ_{jc} .
6. The case to heatsink thermal resistance θ_{cs} depends on the method of mounting the transistor to the heatsink. For example, if the equipment chassis is used as a heatsink, then it will be at ground potential. If the case of the transistor is at some other potential (which is often required) then the case of the transistor must be insulated from the heatsink. An insulating washer increases the thermal resistance between case and sink. On the other hand, a *thermal compound* can be used to fill in the air voids between the transistor case and the heatsink, reducing this thermal resistance. In many cases, the case-sink thermal resistance is much less than the other thermal resistances, so its exact value is not critical. A typical value is about 0.5°C/W .
7. Plug the values for θ_t , θ_{jc} and θ_{cs} into equation 1120, and calculate the remaining thermal resistance θ_{sa} , the thermal resistance of the heatsink. Then use this value of heatsink resistance to identify a suitable commercial product from heatsink catalogues, or, as shown below, calculate the required area of the heatsink.

28.4 Heatsink Area

One of the most useful references on heatsinks for the cooling of semiconductors is reference [242]. The authors measured the thermal resistance of a large number of commercial heatsinks and then fitted equations to the resultant graphs of thermal resistance vs heatsink area. The results are shown in figure 694.

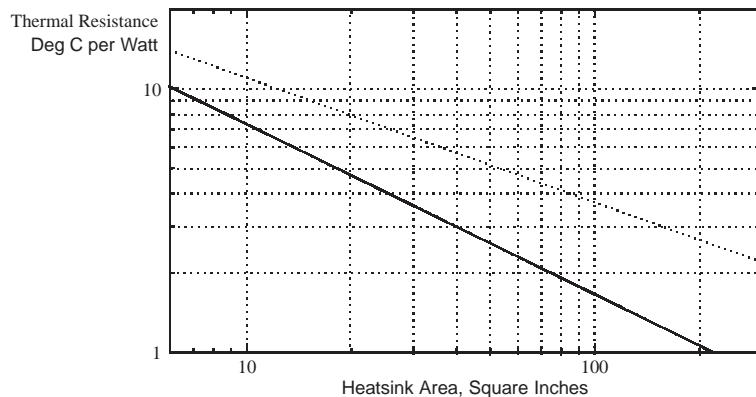


Figure 694: Heatsink Equation

The upper trace represents *Bright, 1/8 Inch Aluminum*, for which the thermal resistance is

$$\theta_{sa} = 32.6A^{-0.472} \quad (1121)$$

where A is the area of the heatsink in square inches. The lower trace is for *Black Aluminum, Vertical Fins*,

$$\theta_{sa} = 34.2A^{-0.645} \quad (1122)$$

A black surface radiates heat more effectively than a shiny silver one, so the thermal resistance is significantly less for a given surface area. The thermal resistances apply to heat sinks with the fins mounted vertically in free air and wouldn't apply to a heat sink inside a sealed enclosure, for example.

Given a value for the required thermal resistance, equation 1121 or 1122 may be used to calculate the required surface area. (Notice that the results are in square inches.) Each heatsink fin has two surfaces, so the total surface area of a 1" by 3" fin is 6 square inches.

28.5 Thermal Capacitance

Injecting energy into a material causes its temperature to rise at a rate that depends on the specific heat of the material and its mass. This effect can be modelled as a capacitance in the thermal equivalent circuit. For a given input power to the material its rate of temperature change is inversely proportional to the thermal capacitance, that is, a large thermal capacitance changes temperature more slowly than a small thermal capacitance²¹⁴.

To determine the units of thermal capacitance, we start with the electrical equation for capacitance:

$$I_c = C \frac{dV_c}{dt} \quad (1123)$$

where

I_c is the current into the capacitor, amps

C is the capacitance in farads

dV_c/dt is the rate of change of capacitor voltage with time, volts/sec

Using the thermal analogy, power substitutes for current and temperature for voltage. Then thermal capacitance would be defined by the equation:

$$P_c = C_t \frac{dT_c}{dt} \quad (1124)$$

where

P_c is the power into the material, watts

C_t is the thermal capacitance

dT/dt is the rate of change of material temperature with time

The units of thermal capacitance are then *watts per °C/sec.*

Example

A design engineer is considering a small plastic enclosure for a circuit. The circuit dissipates significant power and the enclosure is sealed so it is possible that the interior of the enclosure will overheat. The design team decides to run a heating test on the enclosure to determine the thermal resistance between the inside of the enclosure and ambient. Then the internal temperature may be predicted for any given power dissipation in the circuit.

In steady state, the thermal equivalent circuit is as shown in figure 690 on page 784. The experiment is set up to dissipate a known amount of power in the enclosure (in a power resistor, for example) and measure the interior and ambient temperatures²¹⁵. Then the thermal resistance may be calculated.

This works, but it turns out that the mass of the enclosure and the internal air volume are such that the interior temperature takes a long time to reach steady state. Is there any method by which this measurement process can be speeded up?

Solution

The first thing to observe is that the internal temperature-time curve has the same shape as the voltage-time curve of voltage in a capacitor in an RC charging circuit, figure 695(a). The temperature differential ΔT (between ambient T_a and internal T_{int}) rises in an exponential curve towards a final value T_{max} . This suggests that the thermal circuit can be modelled by some combination of electrical source, resistor and capacitor.

This is the case, and a suitable thermal equivalent circuit, with thermal capacitance, is shown in figure 695(b).

²¹⁴An item with a large thermal capacitance is often referred to as having a high *thermal inertia* or *thermal index*.

²¹⁵The temperature measurement may be done using thermistors (section 10.5)

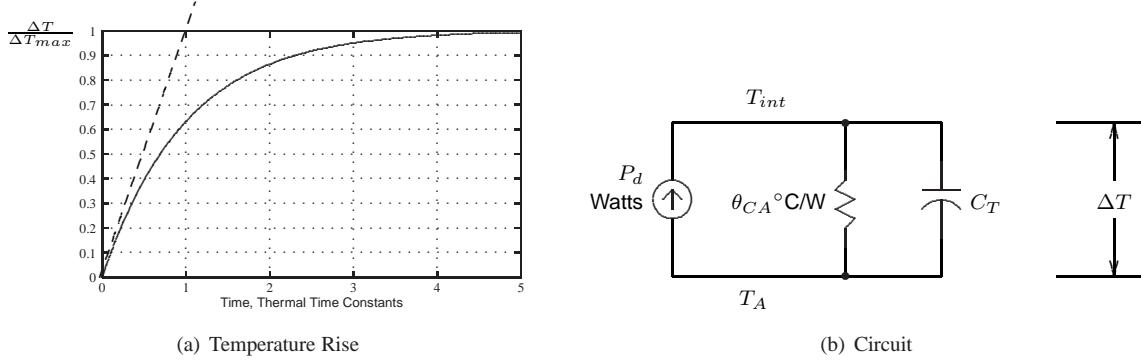


Figure 695: Thermal Circuit with Capacitance

We can analyse the electrical circuit and then convert the results to thermal quantities per the following table:

Quantity	Electrical Unit	Thermal Unit
I	Source current, amps	Power dissipation, watts
R	Resistance, ohms	Thermal resistance, $^{\circ}\text{C}/\text{watt}$
C	Capacitance, farads	Thermal capacitance, watts/ $(^{\circ}\text{C}/\text{sec})$
V_c	Voltage (across capacitor), volts	Temperature differential between interior and ambient (ΔT)

We're after the voltage across the capacitor, which is given by the usual exponential charging equation²¹⁶:

$$V_c = V_p \left(1 - e^{-t/RC} \right) \quad (1125)$$

where V_p is the peak voltage, the voltage across the capacitor once it is charged. The peak voltage is simply equal to the current times the thermal resistance:

$$V_p = IR \quad (1126)$$

The rate of change of V_c provides a simple way to calculate the thermal resistance and capacitance. Substitute equation 1126 into equation 1125 and take the differential with respect to time:

$$\frac{dV_c}{dt} = \frac{d}{dt} \left[IR \left(1 - e^{-t/RC} \right) \right] \quad (1127)$$

$$= \frac{I}{C} e^{-t/RC} \quad (1128)$$

At time $t = 0$, the exponential becomes 1, so that

$$\frac{dV_c}{dt}_{t=0} = \frac{I}{C} \quad (1129)$$

In words, the slope of the temperature-time curve, when the power is first applied, is equal to I/C . This makes sense: at time zero, the voltage across the resistor is zero, so the current through it is zero. Then all the current

²¹⁶It may help to convert current source I and resistor R to their Thevenin equivalent.

flows into the capacitor and the rate of change of voltage is simply I/C volts/second. In thermal terms this is the input power divided by the thermal capacitance. The input power is known, so the thermal capacitance can be calculated.

This gives us a simple method of calculating thermal capacitance and resistance:

- Apply power to the device and plot the internal temperature against time²¹⁷.
- Measure the initial slope of the temperature-time plot and plug this and the value of input power into equation 1129. Calculate the thermal capacitance C .
- Now choose some time after $t = 0$. Substitute values for the time, temperature and thermal capacitance into equation 1128. Solve for the value of thermal resistance R .

28.6 Fan Cooling

A cooling fan will dramatically reduce the temperature of electronic equipment for the same power input or allow much greater power dissipation for the same temperature rise in the enclosure. However, fan cooling is strictly a last resort as a cooling measure. Although their reliability has improved tremendously, fans are still prone to failure, and the results of such a failure can be dramatic. Fans ingest dust into the equipment unless the air is filtered, and if filters are used they must be cleaned regularly or the airflow is compromised²¹⁸.

Moving cooling air through an equipment enclosure is a problem in mechanical engineering, so the details are not given here. However, one point should be noted. Axial fans (ducted propellers) are suitable where there is little resistance to the flow of air, but cannot develop much static pressure. Where the air will encounter significant flow resistance, you should use a centrifugal fan.

A simple analysis, based on Pressman [130], yields a useful insight and formula for the effect of fan cooling. The main challenge is getting it dimensionally consistent and the conversion factors correct. We'll use the CGS system here because specific heat is given in calories per gram.

The relevant constants are

$$\begin{aligned} \text{Specific Heat of Air} &= 0.238 \text{ Calories per gm}^{-\circ}\text{C} \\ \text{Density of Air} &= 1.12 \times 10^{-3} \text{ Grams per cm}^3 \end{aligned}$$

If we have an enclosure containing a source of heat of P_{in} watts, and a fan is caused to move air through the enclosure, then the input power is equal to the increase in the air temperature ΔT times the specific heat, times the mass of air. The mass of air is equal to the density of air times the volume of air. Putting this together in formula form, we have

$$P_{in} = \text{Temperature Rise}^{\circ}\text{C} \times \text{Specific Heat} \frac{\text{calories}}{\text{gm}^{-\circ}\text{C}} \times \text{Density} \frac{\text{gm}}{\text{cm}^3} \times \text{Volume} \frac{\text{cm}^3}{\text{sec}} \quad (1130)$$

$$= \Delta T \times 0.238 \times (1.12 \times 10^{-3}) \times V_{fan} \quad (1131)$$

$$= 266 \times 10^{-6} V_{fan} \Delta T \text{ calories/second} \quad (1132)$$

Now, one watt-second of energy is equivalent to 4.186 calories, or, put another way, one watt is equivalent to 4.186 calories per second. Converting equation 1132 to watts, we have:

²¹⁷The Mastech 8226 available from [5] is a suitable instrument for this type of measurement. The instrument includes a temperature measurement probe and an opto-isolated connection to a host PC.

²¹⁸This is a very common problem with computer power supplies. A colleague reports the story of a machine that was running slowly and prone to rebooting. He was about to retire it, when the light dawned that it might be a cooling problem. Removing the covers and vacuuming out the dust restored the machine to its former sprightly self.

$$\begin{aligned}
 P_{in} \text{ watts} &= 4.186 P_{in} \text{ calories/second} \\
 &= 4.186 \times (266 \times 10^{-6}) V_{fan} \Delta T \\
 &= 1.113 \times 10^{-3} V_{fan} \Delta T
 \end{aligned} \tag{1133}$$

where V_{fan} is the fan air volume in cm^3/sec and ΔT is the temperature rise in $^\circ\text{C}$ of the air as it passes through the enclosure, that is, the temperature rise above ambient inside the enclosure.

Usually, we're interested in the temperature rise in the enclosure for a given power input and air flow. Rearranging equation 1133, we have

$$\Delta T = 898 \frac{P_{in}}{V_{fan}} \tag{1134}$$

Example: Cooling Fan Calculation

Here's a real life example.

A small enclosure contains electronics that consume 9 watts. Before a fan is added, the interior of the enclosure climbs to 42°C above the ambient, which is considered too warm to be reliable. A small processor cooler fan, 1.25 inches on a side, is being considered for cooling the enclosure. The output of the fan, estimated by comparison with known fans, is thought to be about 1.7 cubic feet per minute (CFM, a common specification unit for fans). Estimate the enclosure temperature rise with this fan.

Solution

First, we need to convert the output of the fan from CFM (cubic feet per minute) to cubic centimetres/sec.

$$\begin{aligned}
 V_{fan} &= 1.7 \frac{\text{ft}^3}{\text{minute}} \times \frac{\left(12 \frac{\text{inches}}{\text{foot}} \times 2.54 \frac{\text{cm}}{\text{inch}}\right)^3}{60 \frac{\text{seconds}}{\text{minute}}} \\
 &= 802 \text{ cubic centimeters/sec}
 \end{aligned}$$

Next, plugging this value and the input power, 9 watts, into equation 1134, we have

$$\Delta T = 898 \frac{P_{in}}{V_{fan}} \tag{1135}$$

$$\begin{aligned}
 &= 898 \times \frac{9}{802} \\
 &= 10.07^\circ\text{C}
 \end{aligned} \tag{1136} \tag{1137}$$

As a matter of interest, when the fan cooling effect was tested in the real-life example, the interior of the box measured 10.6°C above the ambient temperature, so the estimate is very close.

28.7 Power Dissipation in an Operational Amplifier

Op-amp circuitry is evolving to meet the needs of the market and its output current capability is increasing. Original op-amps were capable of 10mA or so. Contemporary op-amps (circa 2009) routinely exceed 100mA. As well, package size is decreasing, reducing the surface area. Running the chip at elevated temperature may reduce the reliability of the device, and it causes bias currents and offset voltages to change.

There are two sources of power dissipation in the amplifier: the *quiescent dissipation* due to the op-amp no-load operating current and the *load dissipation* due to the current through the op-amp into the load.

The quiescent dissipation is caused by the op-amp operating current, with no output current, times the supply voltage.

The load dissipation circuit is represented in figure 696. The op-amp is supplying DC current I_o to a load resistor R_l at some output voltage V_o , from a supply of V_{CC} volts. This power dissipation in the chip is caused by the voltage drop between supply and output, times the output current. The op-amp circuit is in figure 696(a) and equivalent circuit in figure 696(b).

Assume that the load resistance is fixed at some value, and the output voltage adjusted. When V_o is at its maximum, there is no internal voltage drop in the op amp, and so the internal dissipation is small. When V_o is at its minimum, zero volts, there is no current in the load or op-amp, and again, the internal dissipation is small. As the output voltage is varied between these two extremes, we see that the power dissipation increases to a maximum and then decreases again. The maximum power is the worst case, so that must be our design condition.

To determine how that situation occurs, a little mathematics.

The power dissipation in the op-amp is given by:

$$P_{int} = V_{int} \times I_o \quad (1138)$$

Let's get everything in terms of the output voltage V_o :

$$V_{int} = V_{CC} - V_o \quad (1139)$$

Also

$$I_o = \frac{V_o}{R_l} \quad (1140)$$

Back substitute for I_o and V_{int} in equation 1138, and we obtain:

$$P_{int} = \frac{V_{CC}V_o}{R_l} - \frac{V_o^2}{R_l} \quad (1141)$$

Notice that P_{int} drops to zero for $V_o = V_{CC}$ and $V_o = 0$, which confirms our intuitive analysis earlier. If we use equation 1141 to plot the power P_{int} against output voltage V_o , we obtain a function that increases from zero,

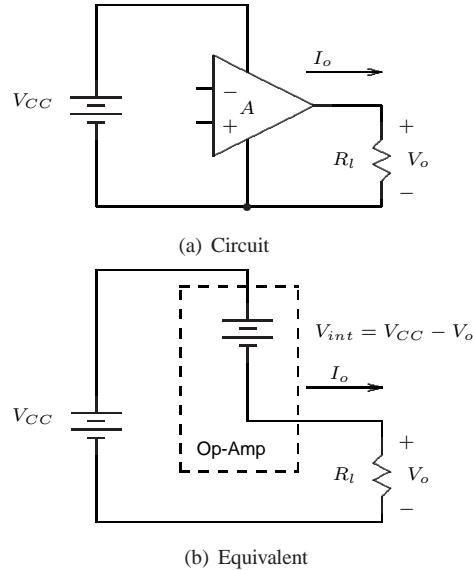


Figure 696: Power Dissipation in Op Amp

goes through a maximum, and then decreases back to zero. We'd like to find the maximum value. That's where the function flattens out, ie, the slope is zero and the derivative of power against output voltage becomes zero. Then:

$$\begin{aligned}\frac{dP_{int}}{dV_o} &= \frac{V_{CC}}{R_l} - \frac{2V_o}{R_l} \\ &= 0\end{aligned}$$

Solve for V_o and we obtain:

$$V_o = \frac{V_{CC}}{2} \quad (1142)$$

In words: maximum power in the op-amp occurs when the load voltage is half the supply voltage.

28.7.1 Example

An LM7171 operational amplifier is to be operated from $\pm 12V$ supplies, driving a worst-case minimum resistance of 150Ω . Determine whether the operating temperature will be exceeded.

28.7.2 Solution

From the datasheet for the LM7171, the relevant parameters are these:

I_s	Supply Current	9.5mA (max)
θ_{ja}	Thermal Resistance, junction to ambient, SOIC package	$172^\circ/\text{W}$
θ_{ja}	Thermal Resistance, junction to ambient, MDIP package	$108^\circ/\text{W}$
T_{jmax}	Operating Temperature Maximum	$+85^\circ$
	Absolute Maximum	$+150^\circ$

There are two available packages for this device: 8 pin DIP (*dual inline package*)²¹⁹ and 8 pin SOIC (*small outline integrated circuit*)²²⁰. The SOIC package is preferable since it takes up less printed circuit board space and can easily be placed by a robot, but it has a larger thermal resistance. The *Operating Temperature Maximum* we interpret as *desirable not to exceed during normal operation*. The *Absolute Maximum* we interpret as *this is where the chip dies*.

The quiescent power dissipation is the product of the total supply voltage (24 volts) and the supply current I_s :

$$\begin{aligned}P_q &= [V_{CC} - V_{EE}] \times I_s \\ &= [12 - (-12)] \times (9.5 \times 10^{-3}) \\ &= 0.228 \text{ watts}\end{aligned} \quad (1143)$$

The power dissipation due to the load occurs when the output voltage and internal voltage are equal, at half the supply voltage, either $+6V$ or $-6V$.

$$\begin{aligned}P_{int} &= V_{int} \times I_o \\ &= V_{int} \times \frac{V_o}{R_l}\end{aligned}$$

²¹⁹http://en.wikipedia.org/wiki/Dual_in-line_package

²²⁰http://en.wikipedia.org/wiki/Small-outline_integrated_circuit

$$\begin{aligned}
 &= 6 \times \frac{6}{150} \\
 &= 0.240 \text{ watts}
 \end{aligned} \tag{1144}$$

Total power dissipation is the sum of these two:

$$\begin{aligned}
 P_d &= P_q + P_{int} \\
 &= 0.228 + 0.240 \\
 &= 0.468 \text{ watts}
 \end{aligned} \tag{1145}$$

Assuming the SOIC package, the temperature rise across the junction to ambient thermal resistance is:

$$\begin{aligned}
 \Delta T_{ja} &= P_d \times \theta_{ja} \\
 &= 0.468 \times 172 \\
 &= 80.5^\circ \text{ C}
 \end{aligned} \tag{1146}$$

This is bad news, because the junction temperature is this value plus the ambient temperature. Any ambient temperature over 4.5°C will cause the junction to exceed its maximum operating temperature of 85°C . (Room temperature is about 20°C , so the circuit would only be safe inside a refrigerator.)

Redoing this calculation with the MDIP package:

$$\begin{aligned}
 \Delta T_{ja} &= P_d \times \theta_{ja} \\
 &= 0.468 \times 108 \\
 &= 50.5^\circ \text{ C}
 \end{aligned} \tag{1147}$$

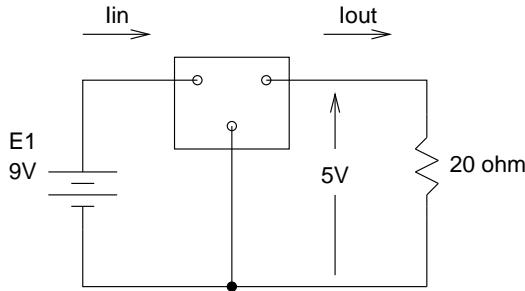
Now the maximum ambient temperature T_{Amax} is:

$$\begin{aligned}
 T_{Amax} &= T_{jmax} - \Delta T_{ja} \\
 &= 85 - 50.5 \\
 &= 34^\circ \text{ C}
 \end{aligned} \tag{1148}$$

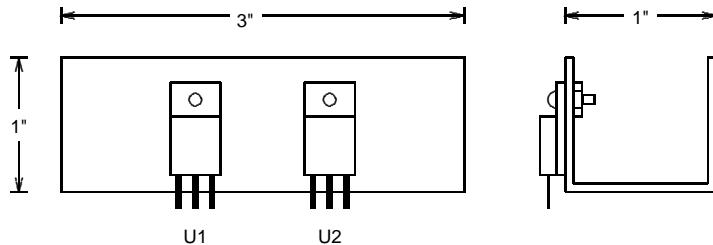
This is more reasonable. Whether this is completely acceptable is an engineering judgement based on the application and the likelihood that these factors – high ambient temperature, prolonged operation with this load current – are likely to occur at the same time.

28.8 Exercises

1. The circuit shows a power supply regulator with a 9 volt input, and a 5 volt output. The regulator input and output current are equal.



- (a) What is the magnitude of the current through the regulator?
- (b) What is the power dissipation in the regulator?
- (c) For the voltage regulator, the thermal resistance junction to ambient θ_{ja} is $30^\circ\text{C}/\text{Watt}$ and the ambient temperature T_a is 40°C . What is the junction temperature inside the regulator?
2. A certain power supply regulator device is in a TO-220 type package. The tab (the case) appears to be very hot when the circuit is in operation²²¹ and when measured is found to be at 95°C . The ambient temperature T_a during this measurement is 25°C .
- For this particular device, the thermal resistance junction-case θ_{jc} is $5^\circ\text{C}/\text{W}$. The thermal resistance case to ambient θ_{ca} is $65^\circ\text{C}/\text{W}$.
- (a) What is the junction temperature T_j of the regulator in $^\circ\text{C}$ for the situation described above?
- (b) Calculate the power dissipation in the regulator, in watts.
- (c) Calculate the area of a heatsink that will reduce the junction temperature to 80°C when the maximum (ie, worst case) ambient temperature is 50°C . You may assume that the heatsink is black aluminum. A typical value for case-heatsink thermal resistance θ_{cs} is $0.5^\circ\text{C}/\text{W}$.
3. Two TO-220 voltage regulators U1 and U2 are mounted on a shiny aluminum channel heatsink as shown in the figure front view and side view. This heatsink is mounted on a printed circuit board so the bottom surface is not in contact with air.



²²¹That is, the design engineer burned her finger on it.

The following specifications apply to the regulators:

Power Dissipation in U1	2.2 watts
Power Dissipation in U2	1.1 watts
Thermal Resistance, Junction to Case, each regulator	5°C/W
Thermal resistance, Case to Heatsink, each regulator	0.5°C/W
Ambient temperature	45°C

You may assume that the heatsink is at the same temperature throughout, that is, the heatsink body has negligible thermal resistance.

- (a) Draw the thermal equivalent circuit for the regulators and heatsink.
 - (b) Calculate the junction temperatures T_{j1} and T_{j2} of the two regulators.
4. Two transistors Q_1 and Q_2 , each in a TO-92 type case, are used in a discrete-component current mirror. It is desirable that the two transistor junction temperatures be as equal as possible. The relevant specifications are as follows:
- | | | |
|--|---------------|---------|
| Ambient Temperature | T_a | 25°C |
| Power Dissipation in Q_1 | P_{D1} | 24mW |
| Power Dissipation in Q_2 | P_{D2} | 1.3mW |
| Thermal resistance of TO-92 case,
junction to case | θ_{jc} | 83°C/W |
| Thermal resistance of TO-92 case,
junction to ambient | θ_{ja} | 200°C/W |
- (a) Determine the junction temperatures T_{j1} , T_{j2} of the two transistors if they are both mounted separately in free air.
 - (b) Draw the thermal equivalent circuit if the two transistors are forced into contact with each other. Hint: $\theta_{ja} = \theta_{jc} + \theta_{ca}$.
 - (c) Calculate junction temperatures T_{j1} , T_{j2} when the two transistors are forced into contact with each other.
 - (d) Which arrangement is preferable in making the junction temperatures as equal as possible: keeping the transistors separated, or placing them in contact with each other?

29 The Single Stage BJT Amplifier

29.1 Introduction

The operational amplifier is a key functional block in circuit design. The bipolar junction transistor is one of the atoms of that functional block, as well as being a very useful device on its own. In this section, we'll study the BJT as an amplifying device for alternating voltages and currents.

The BJT can be used in one of three possible configurations as an amplifier, and we'll study the properties of those different configurations. Because there is an unavoidable DC level shift between the input and output signals, the single BJT amplifier is difficult to use as a DC coupled amplifier. (It can be done, but it's not what one would characterise as an elegant design.)

A DC coupled amplifier generally uses two or more BJTs, and we will look at those circuits in section 30. For a single stage BJT amplifier AC coupling is necessary to input and extract the signals without disturbing the DC conditions.

A single stage AC coupled amplifier is not a particularly useful function in and of itself. It's usually simpler and less expensive to use an operational amplifier. However, it is necessary to understand the single stage circuits before grappling with more complicated circuits. The circuits in this section are primarily useful as a vehicle for teaching various concepts.

Furthermore, these concepts – abstracting the device into a model, biasing the device, determining the characteristics of a single stage amplifier – turn out to be applicable to other devices. For example, the characteristics of the JFET amplifier may be determined quite simply by analogy with the BJT circuits.

Ultimately, a knowledge of BJT operation is useful in the following areas:

- There are some circuits, notably audio power amplifiers, that are implemented with discrete BJT's. An audio power amplifier is essentially a high-power low-distortion op-amp. See for example [243], [244], [245].
- Many analog circuits require one or more discrete transistors to act as the equivalent of *glue*, to provide a function that is not available as an integrated circuit.
- A basic understanding of BJT operation supports further study on the design of integrated circuits.

Conceptual Design and The Role of Device Models

Developing a new circuit design is an exercise in putting together BJT's and other components to provide the desired function. At the outset it's a creative activity that requires choosing devices (BJT, JFET, MOSFET) and their configuration (for the BJT: common-emitter, emitter follower or common-base) in traditional and new ways. This creative activity is best done with a very simple model of the device.

The *device model* for a BJT (or other semiconductor device) is a simplified abstraction of reality. Depending on the degree to which the device model must match the behaviour of the real thing, the model can be more or less complex, but it is composed of simple circuit elements such as voltage and current generators, resistors and capacitors.

Once a circuit concept has been roughed out, an analysis using this very simple model can determine whether the circuit has any hope of meeting its requirements. For example, an inspection of a circuit may determine that the input resistance is too low to be useful and a different configuration is required.

For both the conceptual design and preliminary analysis stages of the design a complicated model is inappropriate because it is difficult to manipulate and it burdens the designer's mind with unnecessary details. Furthermore, some BJT specifications have huge unit-unit variations, so a precise model is pointless. Current gain, for example, varies over a 3:1 range. To manage this, concept design is done with the worst-case value of the transistor parameters (minimum current gain, for example).

If the preliminary analysis looks hopeful, then the circuit may be subjected to simulation. The accuracy of the simulation is dependent on an accurate model of the device, but a complex more device model can be used because it does not burden the designer.

Roadmap to the Section

Based on the preceding,

- it is reasonable to spend some time learning to analyse single stage BJT amplifier circuits, and
- the model of the BJT should be as simple as can support a reasonable facsimile of reality.

The section begins with an overview of BJT operation, the diode model of the transistor, and some simple circuits that illustrate this operation. A ramp-generator circuit is used to illustrate several features of the BJT.

The concept of *transconductance* g_m is fundamental to the BJT and other amplifying devices. In the BJT, it is useful also to deal with the emitter incremental resistance r_e , the reciprocal of transconductance.

This then leads to the T model of the transistor. Some circuit models require different parameter values for each circuit configuration and the component values in the model depend on the circuit. In the T model the component values are largely independent of the circuit configuration, and this makes it a convenient design tool.

The three different amplifier configurations are then studied for their properties as amplifiers: input and output resistance, current and voltage gain.

It would be convenient if transistor datasheets provided the T model parameters directly. Unfortunately these parameters must be determined from other information in the datasheet. A final section shows how the T parameters are obtained.

29.2 The Diode Model

A BJT may be visualized as a sandwich of semiconductor layers, as in figure 697. There are in fact several physical configurations for the transistor which do not resemble this model, but the semiconductor sandwich is a convenient fiction.

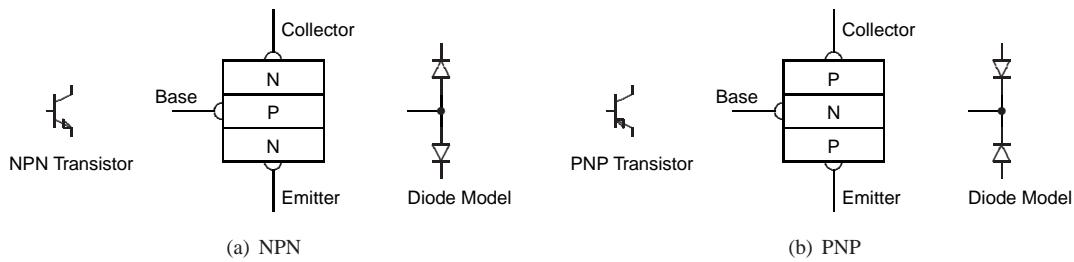


Figure 697: Diode Model of the BJT

For the purpose of applying the transistor at this stage, we need only know that an 'N-type' semiconductor is a crystal with an excess of negatively charged electrons. A 'P-type' semiconductor has a deficiency of electrons. When an N-type and P-type semiconductor are fused together, a diode junction occurs at the boundary between the two materials. Current flows easily in the *forward* direction, once it overcomes a small forward voltage, and the junction is said to be *forward biased*. If the junction is reversed biased, the reverse current is essentially zero.

Now to the transistor. The BJT consists of a sandwich of semiconductor materials, NPN or PNP, as shown in figure 697. (In practice, the base region is much thinner than indicated in the figure. This construction suggests that the BJT can be treated as two back-to-back diodes, also as shown in the figure.)

As shown in figure 697, the three terminals of the transistor are known as the *collector*, the *base* and the *emitter* (see page 800).

The emitter terminal is indicated by the arrow in the symbol. The direction of the arrow indicates the polarity of the BJT: out of the package for NPN, into the package for PNP²²².

The diode model shown in figure 697 accounts for the following behaviour:

- If a diode is forward-biased, it will conduct with a small voltage drop, about 0.6 volts for a silicon transistor.
- If a diode is reverse biased, its current is very small, negligible in most situations²²³.
- A diode can be destroyed by a sufficiently large reverse voltage. The base-collector and base-emitter diodes are constructed somewhat differently. The reverse breakdown voltage of the base-collector diode is much larger (typically 30 volts) than the base-emitter diode (typically 6 volts).
- The transistor can operate in different modes, and these are associated with the biasing of the diodes. For example, when operating as an amplifier, the base-emitter diode is normally forward biased and the base-collector diode reverse biased.

Application: A Simple Test for the BJT

The diode model of the transistor provides us with a simple method of checking a transistor²²⁴. Using the *ohms* setting on a multimeter, measure between the base-collector terminal in both directions. One direction should show some conduction, the other direction should show an open circuit. Similarly, testing the base-emitter diode should show some conduction in one direction and none in the other direction. Measurement between the collector and emitter terminals should show open circuit in both directions.

- If the transistor is destroyed, one or both of the diodes will be an open circuit or short circuit in both directions.
- If the polarity of the multimeter during resistance measurement is known, then this test can identify the polarity of the transistor: NPN or PNP.
- By a process of elimination, it is possible to identify the base terminal. However, the test is limited inasmuch as it cannot readily distinguish the emitter terminal from the collector. That requires an additional measurement.

29.3 BJT Nomenclature

BJT Terminals

The names of the transistor terminals, *base*, *collector* and *emitter* derive from the first transistor, constructed at Bell Laboratories in 1947 [246].

The test circuit of the inventors is shown in figure 698. The BJT is a PNP device in the common base configuration, that is, with the base terminal connected to the circuit ground.

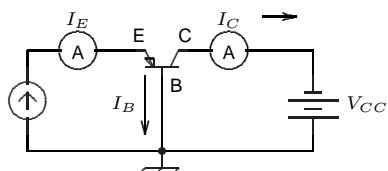


Figure 698: BJT Invention Test Circuit

²²²One large computer manufacturing company used the sandwich diagram as its symbol for the transistor long after everyone else was using the arrowed form.

²²³Not always, though. We recently had a situation where diode leakage current was causing a significant voltage across a $1M\Omega$ resistor. The solution was to choose a lower-leakage diode.

²²⁴This method has been known to destroy the base-emitter junction of certain RF transistors, ensuring that all the transistors that are tested are found to be defective. However, it is completely safe to use on most small-signal and power transistors.

In this arrangement, the names make some sense. The base is in fact the physical base that the device was built upon. The emitter is the terminal that is sending electrons into the base region. Some of the injected emitter current flows out the base as base current. The collector absorbs the remaining current.

Referring to figure 697 as an example, the convention is to double the subscript for a supply that operates a particular terminal. So the collector supply is V_{CC} , the emitter supply would be V_{EE} and a base supply V_{BB} .

The voltage at a particular terminal, with reference to ground, is given a single subscript. So the voltage between the collector and ground would be V_c .

When a voltage is measured between two terminals of the transistor, both subscripts are specified, the first one assumed to be more positive than the second. So the collector-emitter voltage would be V_{ce} , which has a positive value when the collector is at a more positive voltage than the emitter.

Figure 699 shows an example circuit with various voltages labelled.

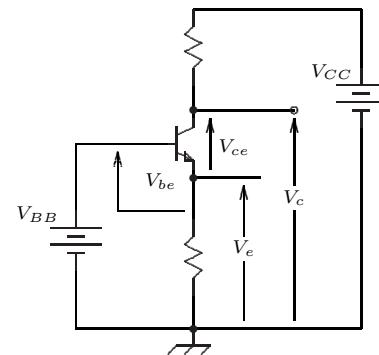


Figure 699: Example Circuit

29.4 The Current Generator Model

First, let us look at the Big Picture concerning the BJT. The Bilateral Junction Transistor is a device that behaves as a *current-controlled current-generator*. A certain input current causes a predictable output current, as shown in the block diagram of figure 700.

There are several points to notice about this arrangement:

- The input current can be created as shown in figure 700 by a voltage source in series with a resistance. Varying the voltage E_i or resistance R_i controls the input current I_i . The input circuit appears as a short circuit.
- The output current is some constant value times the input current. As it turns out, there are two possible input terminals to the BJT, the base and the emitter. When the input terminal is the base, the constant K is a large value, such as 200. When the input terminal is the emitter, the constant K is a number slightly less than unity. (Why this is useful will be explained later.)
- There must be a voltage source (shown as E_s in figure 700) to provide a source of current for the BJT to control. However, within wide limits **varying the supply voltage has no effect on the output current**. This property is extremely important.

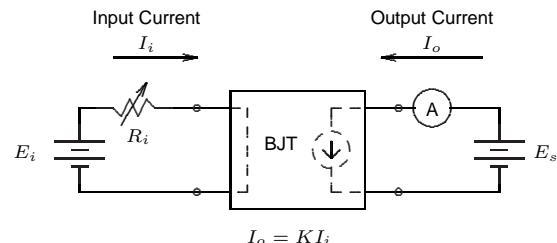


Figure 700: Current Gain vs Collector Current

This model is only an approximation of the behaviour of the BJT. In fact, the input circuit is not quite a short circuit – it's a forward-biased diode. The output circuit is not a pure current generator because the supply voltage has a slight effect on the output current. However, it is a useful and important model for the operation of the BJT, and sufficient for much circuit design.

29.5 Base Current Control

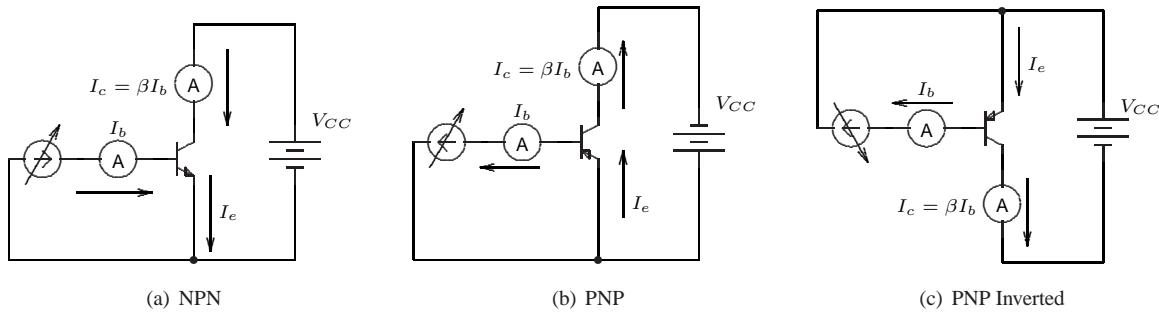


Figure 701: Base Current Control of the BJT

When the base current is used to control the output (collector) current, the BJT is configured in a circuit so that the base-emitter diode is forward biased and the base-collector diode is reverse biased. According to the diode model, the collector current I_c would be zero. However, there is an interaction of the two diode junctions that causes something more interesting and useful to occur.

When a small amount of current is injected into the base region, this triggers a much larger current in the collector. In this respect, the BJT is the current controlled current source, where the output current is much larger than the input current. The factor of *current amplification* is

$$\frac{I_c}{I_b} = \beta \quad (1149)$$

For a small-signal BJT, a typical value for β is 100.

The test setup for the NPN transistor is shown in figure 701(a). The behaviour of the collector current in this circuit is shown on the *collector characteristic* for an NPN BJT in figure 702. As the graph shows, once the collector-emitter voltage is large enough, the collector current is independent of the collector-emitter voltage: it's a function of the base current only. In this case, beta is equal to 100.

Any circuit containing NPN transistors may be converted into a PNP version by reversing all the voltage supplies and current generators, as shown in figure 701(b). An alternative representation of figure 701(b) is shown in figure 701(c).

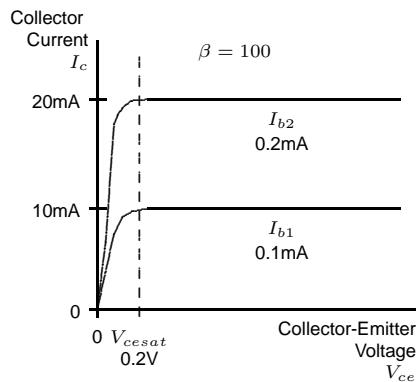


Figure 702: Collector Characteristic

This is a preferable way to draw the schematic of a PNP transistor amplifier, because it maintains consistency with the polarity of the power supply and direction of current flow with figure 701(a). Signals flow from left to right and power-supply currents flow down the page. This arrangement is especially preferable when there is a mixture of NPN transistors and PNP transistors in the same circuit.

Notice that the ‘ground’ (the reference terminal) in these circuits can be placed anywhere in the circuit without affecting operation of the circuit. A frequent location for ground is at the junction of input and output circuits, that is, at the emitter of the transistors in figure 701.

Practical Circuits

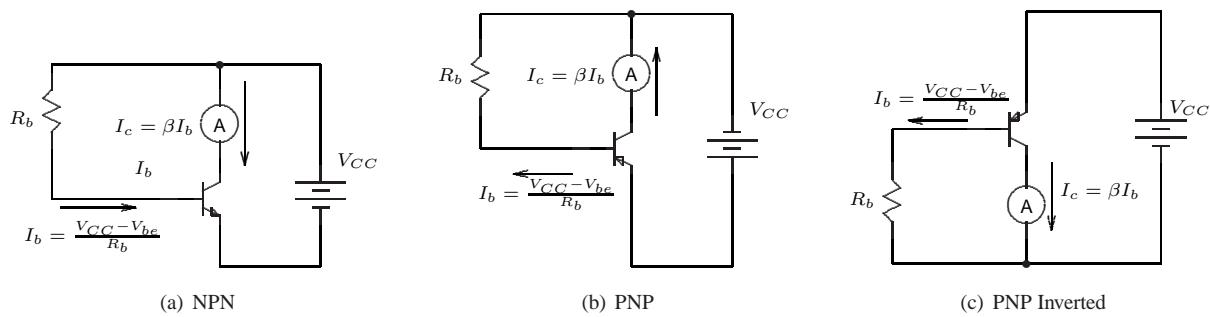


Figure 703: Practical Circuits

Practical versions of the circuits of figure 701 are shown in figure 703. The base current source is replaced by a resistor R_b driven from the same power supply as the collector. The base-emitter diode is forward biased by the base current, so the value of V_{be} is equal to the forward voltage drop of a silicon diode, about 0.6 volts. The base current is then known to be

$$I_b = \frac{V_{CC} - V_{be}}{R_b} \quad (1150)$$

If β is known, then we can predict the collector current. Or, if the collector current is measured, we can calculate the transistor β .

The Current Gain Model

The diode model of the transistor does not predict this phenomenon of current gain, and so we need to modify our model. We begin with an important concept: **The collector current is (largely²²⁵) independent of the collector-emitter voltage. It depends only on the base current and the current gain beta.** There are some reasonable conditions for this to be true:

- The collector-emitter voltage must be of the correct polarity and slightly greater than zero. In fact, it must be greater than the *saturation voltage* of the transistor (typically about 0.2 volts, much less than the forward biased base-emitter voltage).
- It must not exceed the collector-emitter breakdown voltage of the transistor or the current will rise to very large values.

²²⁵There is actually a small increase in current with collector-emitter voltage, but we will reserve that refinement for later.

- At an extremely elevated temperature, the reverse-biased collector-emitter diode begins to conduct a significant *leakage current*, masking the normal current generation process. At normal temperatures, this is not an issue²²⁶.

This is indicated in figure 703 where the collector current is indicated to be a function **only** of the base current and the beta of the transistor. Put another way, the collector supply V_{CC} can be changed over a wide range without having any significant effect on the collector current.

This behaviour implies that the collector circuit of the transistor contains a current generator. The new model circuit is shown in figure 704.

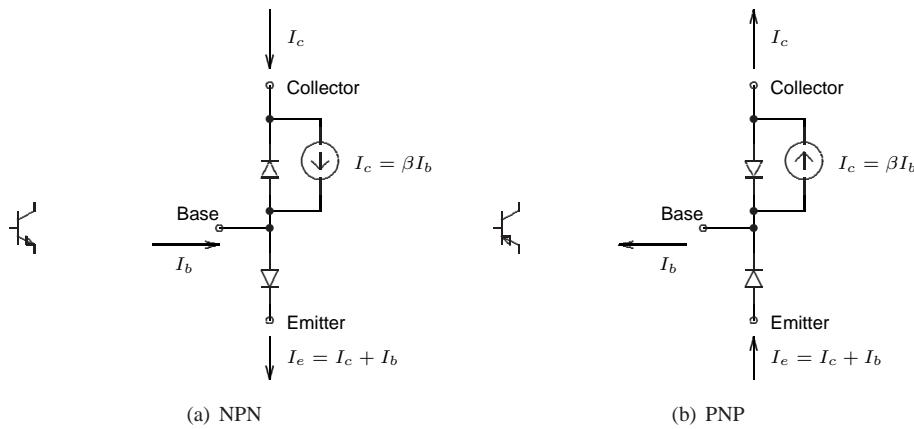


Figure 704: Current Generator Model

In normal operation, the collector-base diode is reverse-biased and so it may be treated as an open circuit and eliminated from the model.

The base-emitter diode on the other hand is forward biased so it may be treated as an (approximately) constant voltage source V_{be} . Incorporating these ideas, we have the circuit models²²⁷ for the BJT shown in figure 705.

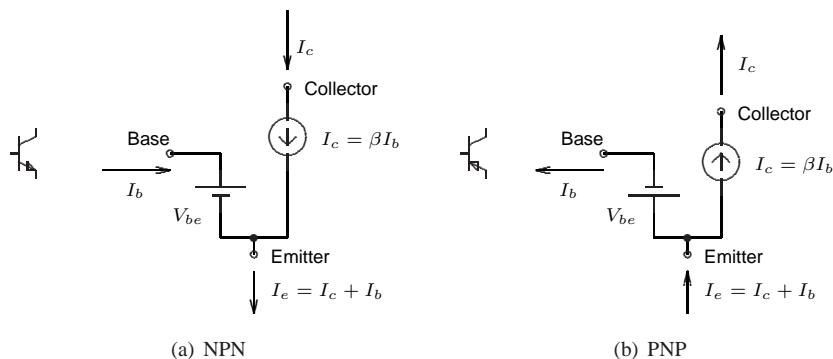


Figure 705: Current Generator Model, Simplified

²²⁶The earliest transistors were constructed with germanium and the collector-base diode had significant leakage current. Books on circuit design from this period spend considerable time showing how to minimize this effect. Modern transistors use silicon and collector-base leakage current is insignificant except at very high temperatures.

²²⁷For an additional discussion of models, see the Appendix on page 900.

The models shown in figure 705 will serve in many circuit design applications.

Variation in Current Gain, Unit to Unit

The unit-unit variation of current gain for a given transistor can vary over a 3:1 range. For example, for the 2N4401 NPN general-purpose small-signal transistor, the spec sheet shows a possible variation between 100 minimum and 300 maximum. As a result, a reliable design must function correctly at the extremes of these values of current gain. The usual design approach is to

- design for the minimum value of current gain
- configure the circuit so that larger values of current gain do not upset the operation of the circuit.

This usually involves some form of negative feedback to establish the bias point and overall gain of the circuit.

Variation of Current Gain with Operating Current

A BJT is optimized to operate at a particular collector current. This value is indicated on the data sheet in a somewhat oblique fashion: it's shown as the operating current where the parameter tests were made.

For example, if the datasheet shows that many measurements were made at around 10mA collector current, the device is appropriate to use at that collector current.

This is not a critical value, and the transistor will be usable over a range of operating currents 30 times less and more than this value. For example, the 2N4401 can be used at collector currents between 0.3 mA and 300mA, a total range of 1000:1, as shown in figure 706.

The transistor is still usable outside this range of collector currents, but with progressively degraded current gain.

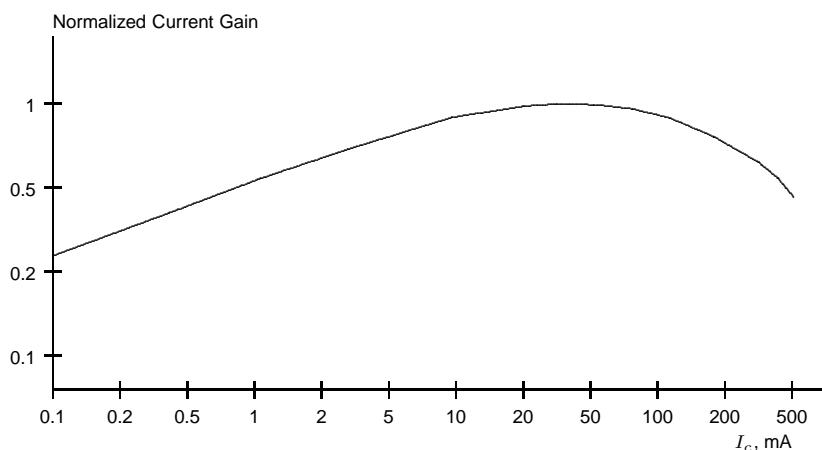


Figure 706: Current Gain vs Collector Current

29.6 Base Voltage Control

In section 29.6, we saw how the BJT collector current can be controlled by a much smaller base current. The constant relating these two is beta, the ratio of collector to base current, which has a typical value of 100 or so.

On the one hand, it is reasonable to view the BJT as a current-controlled-current-generator because the output (collector) current is indeed directly proportional to the input (base) current by an amount equal to beta. However, beta is a very unpredictable constant. It varies by a factor of 3:1 between units and it increases with temperature. Beta is optimum for some particular value of operating current, and it falls off at currents that are very large or very small compared to this optimum value²²⁸. It is bad practice to design a circuit so that it requires specific values of beta.

²²⁸It is a variable constant, in other words. It's a constant only under very restricted conditions.

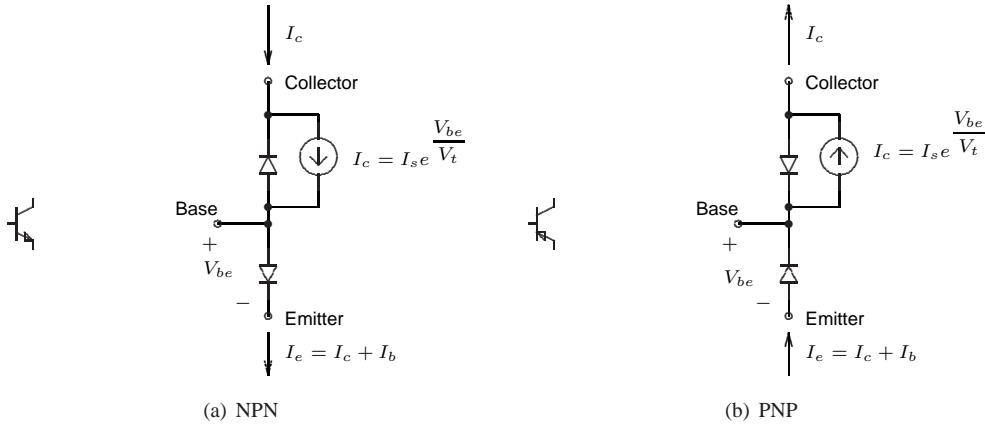


Figure 707: Voltage-Controlled Current Generator Model

There is another possible view of the BJT – as a voltage controlled current generator, in which the collector current is controlled by the base-emitter voltage. When the base-emitter diode is forward biased and conducting, its current is an exponential function of its forward voltage, as shown in equation 394 on page 246.

Applying this equation to the base-emitter diode of the BJT, we have:

$$I_e = I_{es} e^{V_{be}/V_t} \quad (1151)$$

where I_{es} is the *reverse saturation current* of the base-emitter diode. It arises from the diffusion processes in the semiconductor and has a value in the range 10^{-15} to 10^{-12} amperes²²⁹.

The term V_t is approximately 26 millivolts, given by the relationship:

$$V_t = \frac{kT}{q_e} \quad (1152)$$

where k is Boltzmann's constant, T is the temperature in Kelvins and q_e is the charge on an electron, in coulombs.

The Constant Alpha

Now we'll relate the emitter current to the collector current. It is useful to define a new current gain parameter *alpha*, where the alpha of a transistor is the ratio of collector (output) current to emitter (input) current.

$$\alpha = \frac{I_c}{I_e} \quad (1153)$$

KVL indicates that the emitter current is the sum of the base and collector currents.

$$I_e = I_b + I_c \quad (1154)$$

Substitute for I_e from equation 1154 into equation 1153, and then substitute $I_c = \beta I_b$:

$$\alpha = \frac{I_c}{I_e} = \frac{\beta I_b}{I_b + I_b} = \frac{\beta}{\beta + 1}$$

²²⁹In an ideal world, the reverse-biased diode current (leakage current) would be equal to the reverse saturation current. However, there are other effects (some difficult to predict) that determine the actual leakage current.

$$\begin{aligned}
 &= \frac{I_c}{I_c + I_b} \\
 &= \frac{\beta I_b}{\beta I_b + I_b} \\
 &= \frac{\beta}{\beta + 1}
 \end{aligned} \tag{1155}$$

The value of alpha is a number that is slightly less than unity and approaches unity as beta increases. For example:

	β	α
20	0.952	
100	0.990	
300	0.997	

29.6.1 The Ebers-Moll Equation

The collector current of a transistor is alpha times the emitter current. Substituting for emitter current from equation 1151, we have:

$$\begin{aligned}
 I_c &= \alpha I_e \\
 &= \alpha(I_{es} e^{V_{be}/V_t})
 \end{aligned} \tag{1156}$$

The quantity αI_{es} is defined as the *saturation current* of the transistor, I_s ²³⁰.

Substitute this in equation 1151 and we have

$$I_c = I_s e^{V_{be}/V_t} \tag{1157}$$

That is, the BJT collector current is an exponential function of the base-emitter voltage. This is known as the *Ebers-Moll* equation for the transistor.

It is useful to rearrange equation 1157 to solve for V_{be} :

$$V_{be} = V_t \log_e \left(\frac{I_c}{I_s} \right) \tag{1158}$$

This equation is accurate over a very large range of currents. As shown in figure 708 for the 2N4401 transistor, the base-emitter voltage is a logarithmic function of collector current as predicted by equation 1158, over a range of collector current from 0.1 to 500mA.

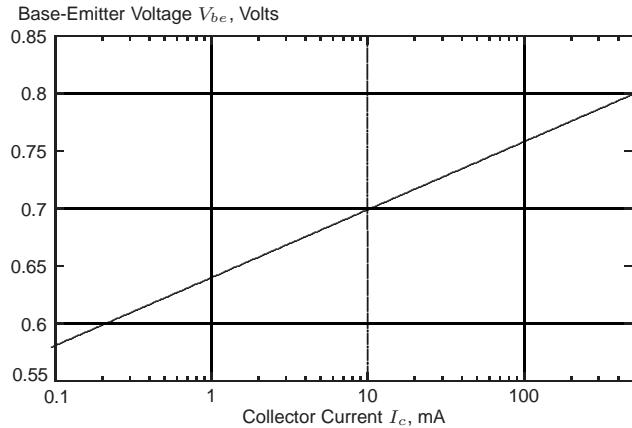


Figure 708: Collector Current vs Base-Emitter Voltage

Example

The datasheet for the 2N4401 NPN transistor shows that V_{be} is 0.8 volts at a collector current I_c of 500mA. What is the reverse saturation current I_s for this transistor?

²³⁰Unfortunately, *saturation* is an overloaded term in semiconductor physics, and you have to determine from the context exactly what it means. Section 8.6 shows another (more common) use of the term.

Solution

Plug $V_{be} = 0.8V$ and $I_c = 500mA$ into equation 1157 with $V_t = 26mV$ and solve for I_s to obtain $I_s = 2.16 \times 10^{-14}A$.

Is the BJT Current or Voltage-Controlled?

There has been some debate in the on-line electronics community on this issue. Here are the points of view:

Current Controlled	Voltage Controlled
As a current-controlled device, the output is (approximately) proportional to the input. That is, the transfer function is linear: $I_c = \beta I_b$. However, the value of beta varies from unit to unit, varies with collector current (figure 706), and changes with temperature.	As a voltage-controlled device, the collector current is an exponential function of the base-emitter voltage over a wide range of collector currents (figure 708). The relationship is ruler-straight (on a log scale), so for a given transistor it is a much more predictable function than the relation between collector and base current. However, this is a non-linear function and the value of saturation current I_s (which establishes the vertical position of this trace) is a function of temperature.

So, is the BJT current or voltage controlled? It's both.

29.7 The Ramp Generator

In this section, we introduce the *linear ramp-generator*. The ramp generator is a useful circuit in its own right, but it makes a particularly suitable vehicle for the exploration and demonstration of various properties of the BJT.

The circuit is shown in figure 709. The transistor generates a constant collector current. When the switch is closed, this current flows to ground and the capacitor is short-circuited, forcing the output voltage to zero. When the switch is opened, the capacitor charges up at a constant rate until the transistor is saturated.

Exercise

If the beta of the transistor is 50 and the saturation voltage of the transistor is 0.2V, draw the output voltage-time function after the switch is opened.

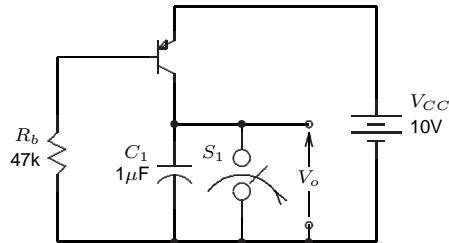


Figure 709: Ramp Generator

Solution

First, replace the PNP transistor in figure 709 with its equivalent from figure 705(b), obtaining the circuit of figure 710(a).

The collector current is determined by the base current, which is determined by the voltage across the base resistor R_b :

- The voltage across R_b is

$$\begin{aligned} V_{Rb} &= V_{CC} - V_{be} \\ &= 10 - 0.6 \\ &= 9.4 \text{ volts} \end{aligned} \quad (1159)$$

- Then the current through R_b is:

$$\begin{aligned} I_b &= \frac{V_{Rb}}{R_b} \\ &= \frac{9.4}{47 \times 10^3} \\ &= 0.2 \text{ mA} \end{aligned} \quad (1160)$$

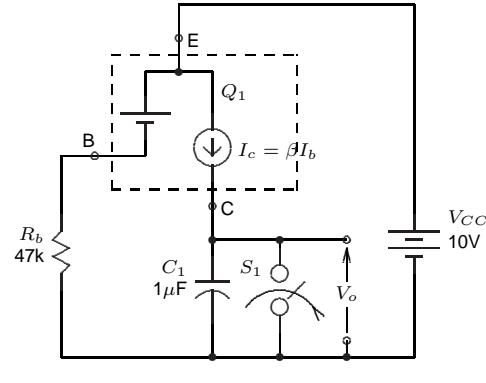
- Providing the transistor is not saturated, the collector current is simply β times the base current. (The transistor is not saturated at this point because the switch is closed and so the collector-emitter voltage is 10 volts, far more than the saturation voltage.)

$$\begin{aligned} I_c &= \beta I_b \\ &= 50 \times 0.2 \\ &= 10 \text{ mA} \end{aligned} \quad (1161)$$

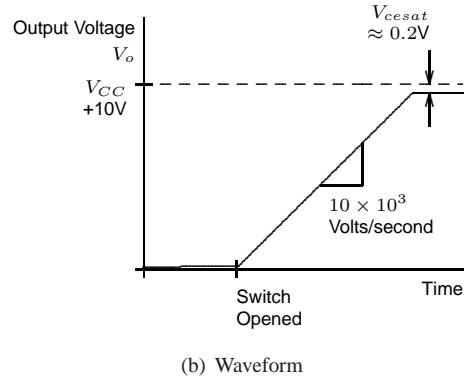
- This is the current flowing through the switch while it is closed. When the switch is opened, the capacitor begins to charge at a rate given by:

$$\begin{aligned} \frac{dV_c}{dt} &= \frac{I_c}{C} \\ &= \frac{10 \times 10^{-3}}{1 \times 10^{-6}} \\ &= 10 \times 10^3 \text{ volts/second} \end{aligned} \quad (1162)$$

The output waveform is shown in figure 710(b). When the switch opens, the capacitor ramps upward at 10×10^3 volts/second. When the collector-emitter voltage drops below the saturation voltage of the transistor (typically 0.2 volts), the charging process stops.



(a) Circuit



(b) Waveform

Figure 710: Ramp Generator

Improving the Ramp Generator

The ramp generator of figure 709 is not a suitable circuit for production. The base current of the transistor is essentially constant, and the collector current is beta times as large. The value of beta can vary over a unit-unit range of 3:1. In a production run, the ramp current and as a result the rate of charging of the capacitor would show this same range of variation. This is unlikely to be acceptable and a production design must somehow ensure a more predictable ramp current.

In the next sections we'll explore how that can be accomplished.

29.8 Emitter Current Control

When the BJT is controlled by current into the emitter, it is useful to use the current gain parameter *alpha* (section 29.6, page 806), where the alpha of a transistor is the ratio of collector (output) current to emitter (input) current.

$$\alpha = \frac{I_c}{I_e} \quad (1163)$$

The NPN and PNP circuits for emitter current control are shown in basic form in figure 711 and in more practical form in figure 712. For example, in figure 712(b), the emitter current is given by

$$I_e = \frac{V_{EE} - V_{be}}{R_e} \quad (1164)$$

The value of emitter current could be controlled by the voltage source V_{EE} or the emitter resistance R_e .

The collector current is simply

$$I_c = \alpha I_e \quad (1165)$$

Notice once again that the collector current is independent of the collector supply voltage V_{CC} . It is a function of alpha and the emitter current.

This is useful when a circuit requires a predictable current source. Notice in the table of alpha values given above that as beta varies dramatically (as it will from BJT unit to unit) the value of alpha is always close to unity. In the next section we will use this to design an improved ramp generator.

Base Current under Emitter Control

When the emitter current is controlled, it's useful to regard the base current as dependent on the emitter current.

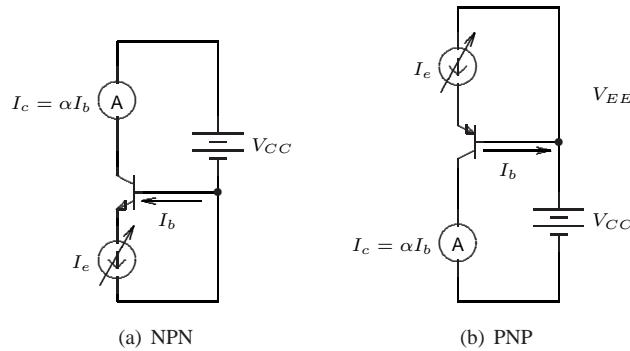


Figure 711: Emitter Current Control of Collector Current

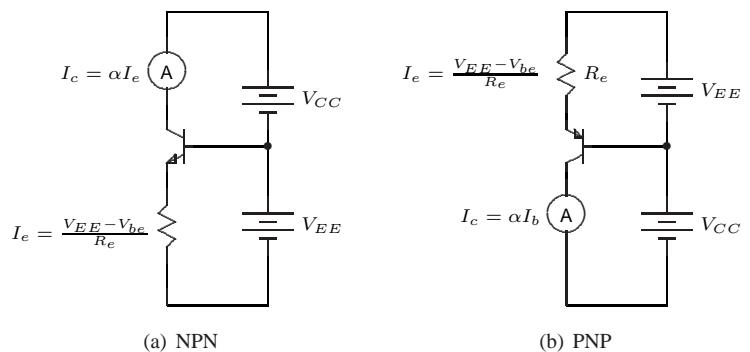


Figure 712: Emitter Current Control, Practical Circuits

Start with

$$I_c = \beta I_b \quad (1166)$$

By KVL,

$$I_c = I_e - I_b \quad (1167)$$

Substituting for I_c from equation 1167 into equation 1166 and simplifying, we have

$$I_b = \frac{I_e}{\beta + 1} \quad (1168)$$

Since beta is usually much larger than 1, it's useful to think of the base current as simply beta times smaller than the emitter current.

29.9 The Ramp Generator: Improved

In section 29.7, we looked at the design of a linear ramp generator. This circuit had a serious flaw: the current generator in the circuit depends on the beta of the transistor, and beta is subject to substantial unit-unit variations. As a result, the circuit shown in figure 709 is not usable.

In this section, we'll evolve a new current generator which solves that problem. The steps in the evolution are shown in figures 713 through 718.

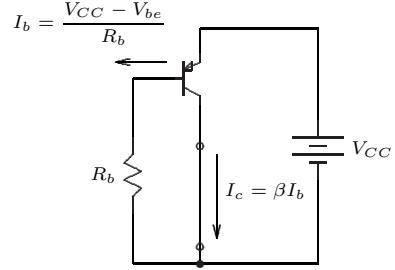


Figure 713: Evolution of the Current Generator

Original Current Generator

Figure 713 shows the original beta-dependent current generator, reduced to its essentials. To recap, the base current is a fixed value determined primarily by the supply voltage V_{CC} and R_b . The collector current is beta times the base current.

The output circuit is indicated as a short circuit, since (within wide limits) the output current is independent of what is in series with the collector. A short circuit is one simple possibility.

Emitter Current Control

Figure 714 shows a current generator controlled by its emitter current, which will be the starting point for a new design. The collector current I_c is then alpha times the emitter current. Since alpha is approximately unity over wide values of beta, the output collector current is predictable even in the face of wide variations in the transistor beta.

Voltage supply E_1 ensures that the collector-base diode of the transistor is reverse biased, as required to make the transistor function correctly.

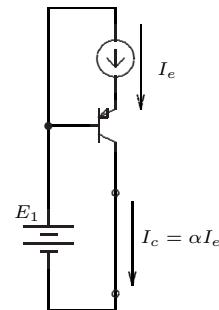


Figure 714: Evolution of the Current Generator

Emitter Resistor Control

Figure 715 replaces the emitter current generator of figure 714 with an emitter resistor R_e and voltage supply E_2 , moving us closer to a practical circuit.

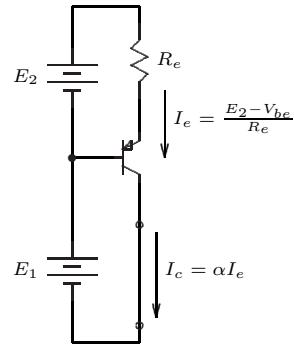


Figure 715: Evolution of the Current Generator

Rearrange Voltage Sources

In figure 716, the voltage sources have been rearranged. There is one supply voltage V_{CC} which is equal to the sum of the E_1 and E_2 voltage sources of figure 715. The voltage across R_e is still fixed, so its current is predictable and constant.

In addition, a resistor R_b is shown in series with the voltage source E_1 . This modification will allow us to morph the voltage E_1 into a voltage divider with a non-zero internal resistance. The current into E_1 is the base current of the transistor, which is smaller by a factor of $\beta + 1$ than the emitter current. Consequently the resistance has a negligible voltage drop across it, and the voltage at the base is essentially as it was in figure 715.

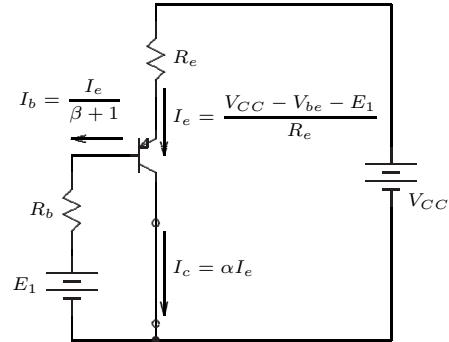


Figure 716: Evolution of the Current Generator

Voltage Divider Source

In figure 717, voltage source E_1 is replaced by a voltage divider. The internal resistance of this divider is equivalent to resistance R_b in figure 716. The ratio of R_1 and R_2 establish the voltage E_1 in figure 716. As usual for a voltage divider, the internal resistance R_b is equal to $R_1 \parallel R_2$.

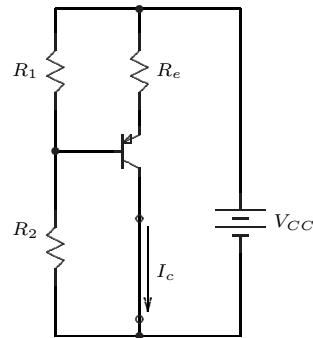


Figure 717: Evolution of the Current Generator

Completed Circuit

Now we can complete the linear ramp generator circuit, which is shown in figure 718.

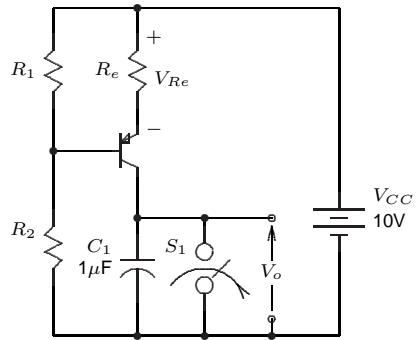


Figure 718: Completed Current Generator

Numerical Design

Now that we have evolved a suitable circuit, let's calculate the component values. As in the circuit of section 29.7, we will assume a supply voltage V_{CC} of 10 volts and an output current I_c of 10mA. Assume beta varies between 100 and 300.

Step 1: Choose the maximum value of the ramp voltage The R_1, R_2 voltage divider establishes a constant voltage across the emitter resistance R_e . When the switch is opened, the capacitor charges upward until the collector voltage of the transistor reaches the emitter voltage, the transistor becomes saturated, and the process stops. In the circuit of section 29.7, the emitter was at V_{CC} volts, so the ramp ran upward until it hit the positive power supply voltage, 10 volts.

In this case, there is a voltage across the emitter resistance so the emitter is at a somewhat lower voltage than the supply. Consequently, the emitter voltage sets the maximum value of the linear ramp voltage.

Arbitrarily, we'll assume that the ramp voltage must go up to 8 volts.

That leaves 2 volts to appear across the emitter resistance R_e . We'll call this voltage V_{re}

Step 2: Calculate the emitter current. We previously showed that

$$\begin{aligned} I_c &= \alpha I_e \\ &= \frac{\beta}{\beta + 1} I_e \end{aligned}$$

Then we can find the emitter current I_e :

$$\begin{aligned} I_e &= \frac{\beta + 1}{\beta} I_c \\ &= \frac{101}{100} 10 \text{ mA} \\ &= 10.1 \text{ mA} \end{aligned}$$

(Notice that assuming alpha equal to unity yields an emitter current equal to the collector current, 10mA, which is reasonably close to the correct answer.)

Step 3: Calculate the emitter resistance R_e .

$$R_e = \frac{V_{re}}{I_e}$$

$$\begin{aligned}
 &= \frac{2}{10.1 \times 10^{-3}} \\
 &= 198\Omega
 \end{aligned}$$

Step 4: Calculate the Thevenin Equivalent of the Voltage Divider The relevant circuit diagram for this calculation is figure 716. The voltage E_1 is the Thevenin open-circuit voltage of the divider, and R_b is the Thevenin internal resistance of the divider. We'll start by assuming that the voltage across R_b is negligible, and then we'll make that so. Then by KVL we can calculate the voltage E_1 :

$$\begin{aligned}
 E_1 + V_{be} + V_{re} - V_{CC} &= 0 \\
 E_1 &= V_{CC} - V_{re} - V_{be} \\
 &= 10 - 2 - 0.6 \\
 &= 7.4 \text{ volts}
 \end{aligned}$$

Now we need to calculate a suitable value for R_b . The voltage drop across R_b should be much less than the open circuit voltage of the divider. Let us somewhat arbitrarily put the voltage drop V_{rb} across at R_b at 5% of E_1 , or about 0.37 volts. Then the voltage at the base of the transistor will be $7.4 + 0.37 = 7.77$ volts.

The base current I_b is given by:

$$\begin{aligned}
 I_b &= \frac{I_e}{\beta + 1} \\
 &= \frac{10.1}{101} \\
 &= 0.1 \text{ mA}
 \end{aligned}$$

Now we have the voltage across R_b and the current I_b through it, so

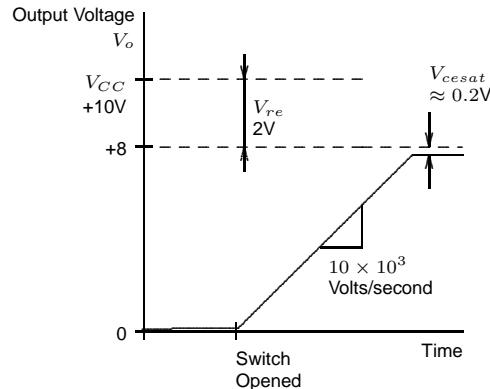


Figure 719: Improved Generator Waveform

$$\begin{aligned}
 R_b &= \frac{V_{rb}}{I_b} \\
 &= \frac{0.37}{0.1 \times 10^{-3}} \\
 &= 3700\Omega
 \end{aligned}$$

Step 5: Calculate the Voltage Divider Resistances.

Treating the voltage divider as a source with an open circuit voltage $E_{oc} = E_1$ and internal resistance $R_{int} = R_b$, we can write two equations:

$$\begin{aligned}
 E_{oc} &= E_1 \\
 &= V_{CC} \left(\frac{R_2}{R_1 + R_2} \right) \\
 &= 7.4 \text{ volts}
 \end{aligned} \tag{1169}$$

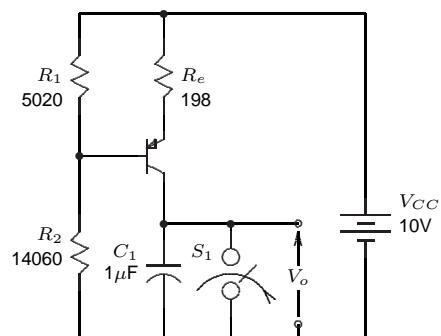


Figure 720: Improved Generator Circuit

$$\begin{aligned}
 R_{int} &= R_b \\
 &= R_1 \parallel R_2 \\
 &= 3700\Omega
 \end{aligned} \tag{1170}$$

Substitute 10 for V_{CC} in equation 1169 and then solve equations 1169 and 1170 simultaneously to determine R_1 and R_2 . We draw the curtain in front of this crank-turning exercise for a few moments and then open it to find:

$$\begin{aligned}
 R_1 &= 5020\Omega \\
 R_2 &= 14060\Omega
 \end{aligned}$$

The final waveform is shown in figure 719 and the circuit in figure 720.

The Effect of a Change in Beta

Changes in beta have very little effect on this circuit. Suppose that the beta goes from our design value, 100, to its maximum value, 300. The base current I_b will decrease by a factor of approximately 3, but this only affects the voltage across R_b , which we made a small fraction of the voltage E_1 . So the emitter voltage of the transistor is still established primarily by V_{CC} and E_1 . Consequently, the emitter current is largely unchanged.

The new value of alpha will determine the new collector current. However, when beta changes from 100 to 300, alpha changes from 0.990 to 0.997, not much of a change at all. Consequently, the collector current is not substantially changed, even with a dramatic change in beta.

If this hand-waving analysis is not entirely satisfactory and we'd like to have better numbers, there is a straightforward way to calculate them, and that is shown in the next section.

29.10 Reflecting Resistance

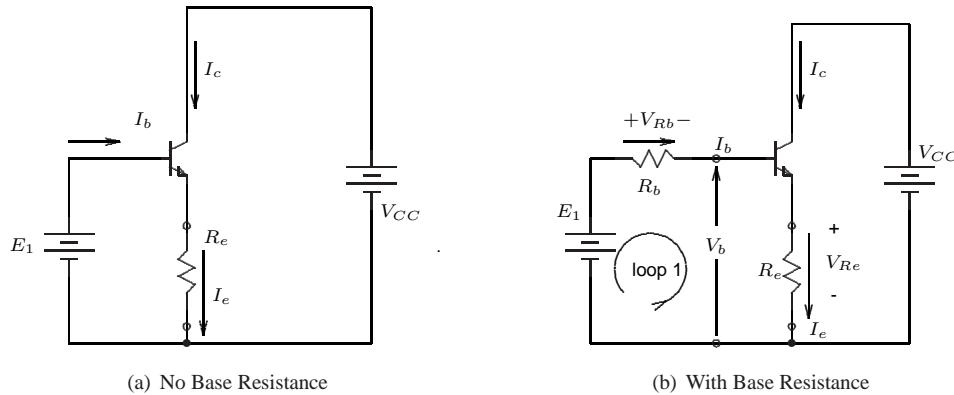


Figure 721: Reflecting Resistance

In section 29.9, we needed to be able to calculate the operating currents and voltages in a transistor circuit for various values of the current gain. This is a situation that arises frequently in the design of a BJT circuit and there is a useful shortcut in the calculations: the so-called *reflecting resistance* trick.

To illustrate this, we'll consider the circuit in figure 721. This is similar to the ramp circuit dissected in section 29.9, but with an NPN transistor.

No Base Resistance

In figure 721(a), the base resistance is zero. This makes the calculation of the circuit currents and voltages relatively simple.

- By KVL, the voltage V_{re} across the emitter resistor R_e is:

$$V_{re} = E_1 - V_{be} \quad (1171)$$

- This voltage defines the emitter current:

$$I_e = \frac{V_{re}}{R_e} \quad (1172)$$

- The collector current is alpha times the emitter current:

$$I_c = \alpha I_e \quad (1173)$$

- The base current is ($\beta+1$) times smaller than the emitter current:

$$I_b = \frac{I_e}{\beta + 1} \quad (1174)$$

That's it, we're done. This is so straightforward that we often wish away any base resistance that might be in the circuit.

With Base Resistance

When the base resistance cannot be ignored, as in figure 721(b), the situation is a little more complicated. Let's assume we want to calculate the base current. Then we have to take into account the voltage drop across the base resistor R_b as well as the voltage across the emitter resistor R_e .

Applying KVL to loop #1, we have:

$$+ E_1 - V_{Rb} - V_{be} - V_{Re} = 0 \quad (1175)$$

The voltages across R_b and R_e are:

$$V_{Rb} = I_b R_b \quad (1176)$$

and

$$V_{Re} = I_e R_e \quad (1177)$$

Finally, we have that the base and emitter current are related by:

$$I_e = (\beta + 1)I_b \quad (1178)$$

Collapsing equations 1176, 1177 and 1178 into equation 1175 and solving for the base current I_b , we have:

$$I_b = \frac{E_1 - V_{be}}{R_b + (\beta + 1)R_e} \quad (1179)$$

This is interesting and useful, because equation 1179 implies that **the effect of the emitter resistor on the base current is multiplied by ($\beta+1$)**. (You can remember this as effectively multiplying the emitter resistance by beta since beta is much larger than 1.)

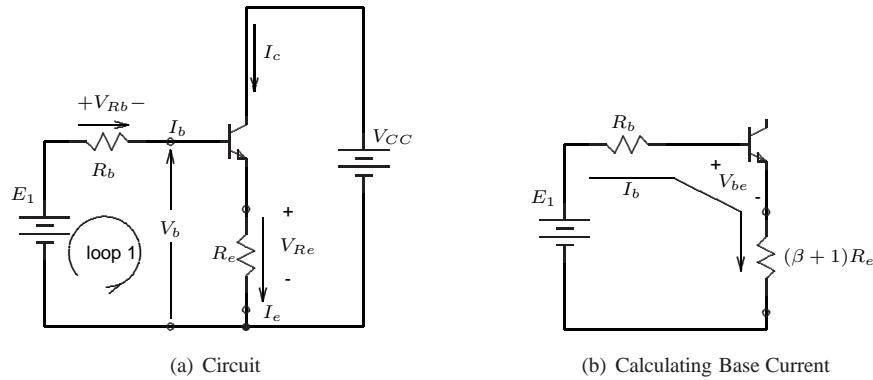


Figure 722: Reflecting Resistance

This is illustrated in figure 722. The original circuit is shown in figure 722(a), and the equivalent circuit for the base current in figure 722(b). The base circuit *sees* an effective value of emitter resistance that is $(\beta + 1)$ times the actual resistance. Knowing that is the case, the value of base current in figure is given by equation 1179, which may now be written down by inspection.

Knowing the base current then quickly leads to the collector current, the emitter current, and the voltages at the terminals of the BJT in the circuit.

Furthermore, this leads us to a simple method of determining whether the base resistance can be ignored. The base resistance R_b and the effective emitter resistance $(\beta + 1)R_e$ are in series. If R_b is much less than $(\beta + 1)R_e$, then it may be ignored. (You can decide what constitutes *much less*, based on the required accuracy of the calculation.)

29.11 Collector Resistance

In many circumstances, the current generator model of figure 704 on page 804 is sufficiently accurate. However, this model implies that the collector current is entirely independent of collector-emitter voltage, and that is not completely accurate. In fact, there is a slight increase in collector current with collector-emitter voltage, as indicated in figure 723.

Figure 723(a) shows the measuring circuit. The transistor is operated by a constant base current. The collector-emitter voltage is adjusted and the collector current measured.

The behaviour of a pure current source is shown in figure 723(b), the so-called *collector characteristic* of a BJT. As the collector-emitter voltage is increased from zero, there are two regions of operation:

- In the *saturation* region, as the collector-emitter voltage increases from zero up to the saturation voltage of the transistor (typically 0.2 volts for a small signal transistor) the current increases very rapidly up to beta times the base current.
 - In the *active* region, the collector current flattens out and becomes constant at beta times the base current, irrespective of the collector-emitter voltage. Changing the base current moves the collector current trace up or down.

More realistic behaviour is shown in figure 723(c). As the collector-emitter voltage increases, the collector current increases slightly. The inverse of the slope²³¹ of this characteristic is the *incremental collector resistance* of

²³¹The slope itself is known as the *incremental collector conductance* g_c and is measured in mA/V or $m\mathcal{U}$.

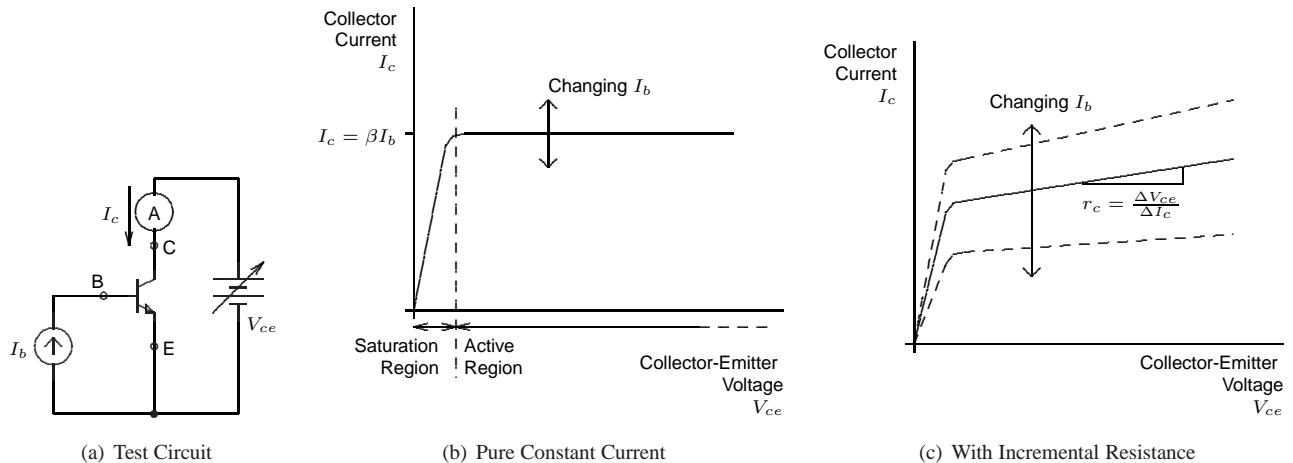


Figure 723: Collector Incremental Resistance

the transistor, known as r_c . For a small-signal transistor, the value of r_c can range from $1M\Omega$ to $10k\Omega$ depending on the operating current and the type of transistor. In general, larger values are preferred. For example, in amplifiers, larger values of collector resistance result in larger voltage gain.

BJT Circuit Models which Include Collector Resistance

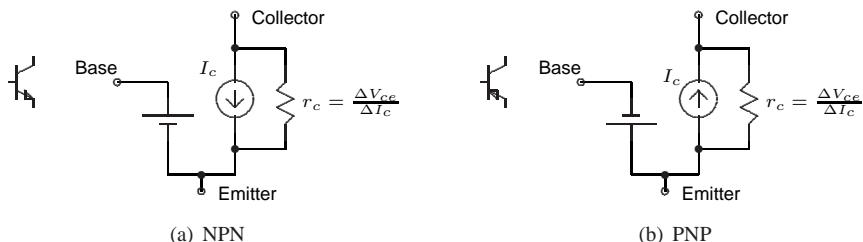


Figure 724: BJT models, including collector incremental resistance

Collector incremental resistance can be modelled in the NPN and PNP Transistor as shown in figure 724. The collector incremental resistance r_c appears in parallel with the collector current generator. When should collector incremental resistance be included in the model?

- For certain BJT amplifier configurations, collector incremental resistance sets the maximum voltage gain of the stage, and so its effect must be included.
- Again depending on the circuit configuration, it may affect the output resistance of a BJT amplifier stage.
- When the BJT is being used as a constant-current generator, r_c affects the output resistance of the generator. This, in turn, establishes how close the current generator is to the ideal, which could be important in a circuit such as a ramp generator. And with that in mind, we'll now see the effect of collector resistance in our previous example of the constant-current linear ramp circuit.

Ramp Generator with Collector Incremental Resistance

Another version of the linear ramp circuit is shown in figure 725. This version is based on the circuit of figure 731 on page 824. In this case, the transistor emitter resistors have been eliminated so that transistors Q_1 and Q_2 form a current mirror (section 30.5). The current set up in the diode-connected transistor Q_2 (10mA in this case) will be conducted by the constant-current source Q_1 .

Assuming that the collector incremental resistance of the transistor is $10k\Omega$, what is the effect on the shape of the ramp circuit?

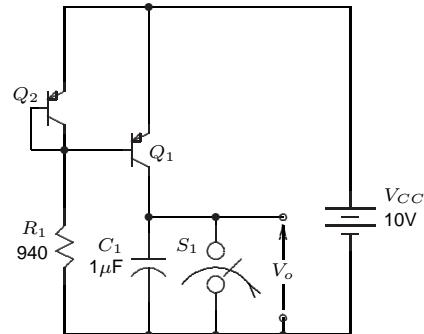


Figure 725: Ramp Generator Circuit

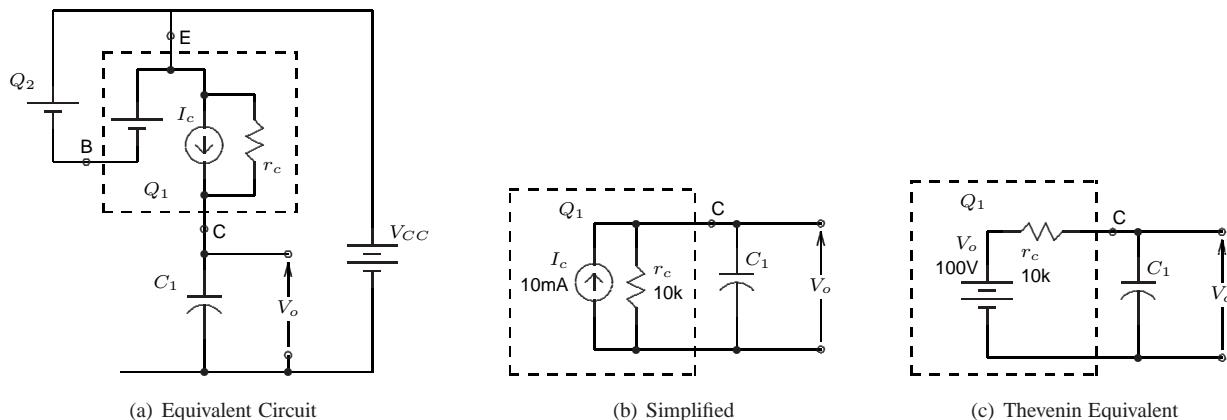


Figure 726: Ramp Generator Equivalent Circuits

A progression of equivalent circuits is shown in figure 726.

- In figure 726(a), the current source transistor Q_1 has been replaced with its equivalent from figure 724(b). Transistor Q_2 acts as a constant-voltage source of about 0.6 volts at the base of Q_1 , so it's simply shown as a voltage source.
- In figure 726(b), the circuit has been further simplified to focus on the current generator transistor output and the capacitor. The power supply voltage V_{CC} is not required in this circuit if we keep in mind that the Q_1 current source is equal to 10mA. (V_{CC} has no direct effect on the voltage of the capacitor).
- In figure 726(c), the current generator and its internal resistance r_c have been transformed into their Thevenin equivalent: an open circuit voltage source of $10\text{mA} \times 10k\Omega = 100$ volts in series with r_c . Of course, the capacitor cannot charge to more than 10 volts – at that point the current source transistor saturates. However, as it charges between 0 and 10 volts, the capacitor charges as if it were heading for 100V.

The charging curve is shown in figure 727. Figure 727(a) shows the full charging curve. If it were allowed to go from zero to 100 volts, it would exhibit the usual RC charging characteristic with a time-constant of 10mSec. In fact, the charging curve is limited to 10 volts. This region, around the origin, is shown expanded in figure 727(b).

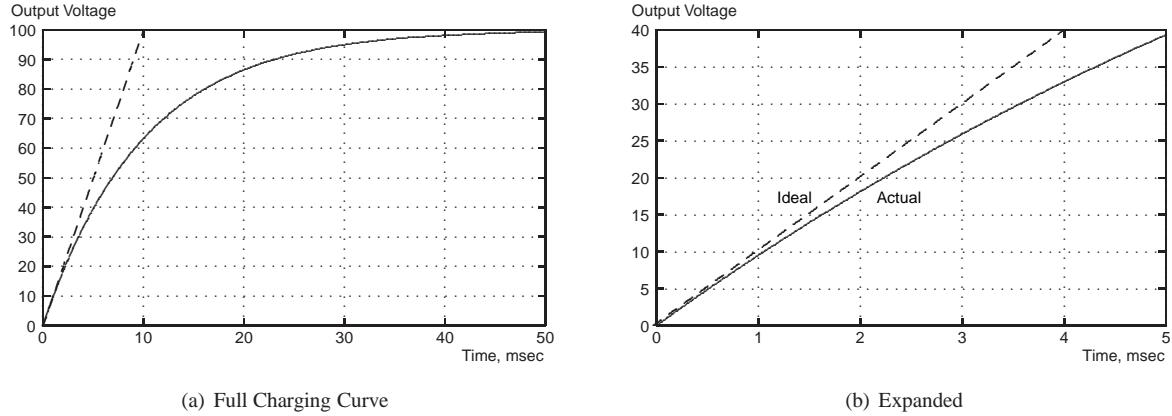
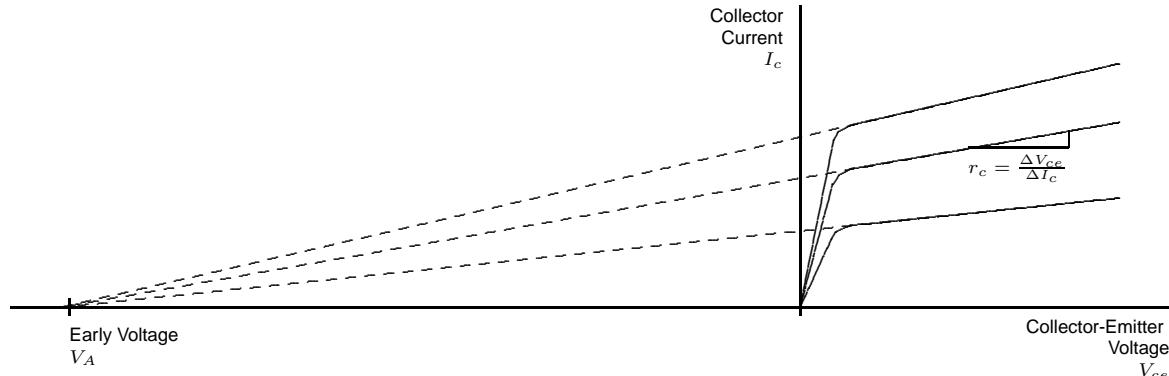


Figure 727: Ramp Generator Charging Curve

For comparison, a completely linear charging rate would appear as the dashed line in figure 727(b). Over the range of 0 to 10 volts, the deviation from the linear characteristic is slight but visible. Whether this deviation is acceptable or not would depend on the requirements of the application. To further linearize the ramp charging characteristic, it would be necessary to increase the output resistance r_o of the current generator.

Early Effect, Early Voltage and Collector Resistance

Figure 728: *Early Voltage* V_A

When the collector characteristics of figure 723(c) are projected to the left, they intersect at a point on the V_{ce} axis known as the *Early voltage*, named after Jim Early, who first identified this effect at Bell Laboratories [247], [248].

When the transistor is operated at some collector current I_c , then:

$$r_o = \frac{1}{\text{slope}} \quad (1180)$$

$$\approx \frac{V_A}{I_c} \quad (1181)$$

where r_o is measured at some collector current I_c .

In other words, V_A is identically equal to the Thevenin voltage we identified in the equivalent circuit for the transistor of figure 726(c). Notice that the incremental collector resistance r_o decreases as collector current increases.

Output Resistance and Emitter Current Control

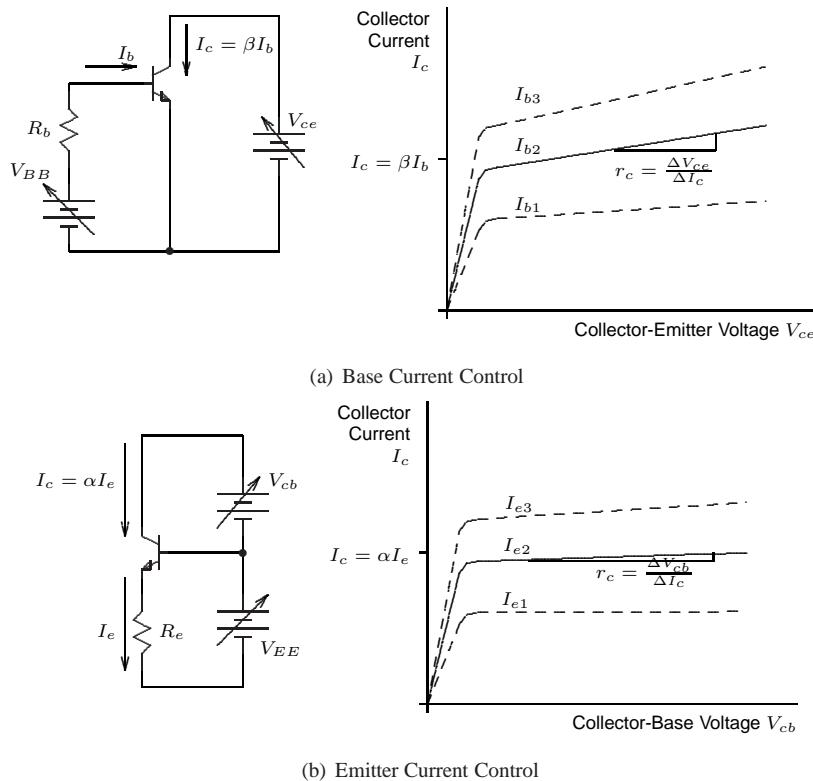


Figure 729: Base vs Emitter Current Control

In the previous section we examined the output resistance of the transistor when it is controlled by its base voltage. The output resistance was $r_o = r_c$, where r_c is typically in the 10's of kilohms. In this type of circuit, when the collector-emitter voltage increases it effectively narrows the base region which increases beta. This, in turn, increases the collector current. The result is a slope to the collector characteristic that represents a relatively modest value of output resistance.

Alternatively, when the transistor is controlled via its emitter current, (section 29.8) it turns out that the output resistance of the transistor is larger by a factor of beta, that is, $r_o = \beta r_c$ (see references [249], [250]). This puts the output resistance into the 100's of kilohms or the megohms.

This is summarized in figure 729 which compares the collector current characteristics of the two transistor configurations: base current control and emitter current control.

- Notice that the slope of the collector characteristic is much larger in figure 729(a), base current control, than in figure 729(b), emitter current control.

- In figure 729(a), the base current is:

$$I_b = \frac{V_{BB} - V_{be}}{R_b} \quad (1182)$$

where V_{be} is approximately 0.6 volts since the base emitter junction is a forward biased diode.

- In figure 729(b), the emitter current is:

$$I_e = \frac{V_{EE} - V_{be}}{R_e} \quad (1183)$$

- In figure 729(b), the collector-base voltage is controlled by voltage supply V_{cb} . The collector-emitter voltage, which is plotted in the diagram, is larger than V_{cb} by an amount equal to V_{be} , which is approximately 0.6 volts since the base emitter junction is a forward biased diode.
- Keep in mind that beta is much larger than alpha, so it takes a much larger emitter current than base current to cause the same collector current.
- We previously presented the Ebers-Moll equation (equation 1157) for the transistor, which relates base-emitter voltage to collector current:

$$I_c = I_s e^{V_{be}/V_t} \quad (1184)$$

This equation does not take into account the effect of output resistance. To incorporate that effect [251], the equation becomes:

$$I_c = \left(1 + \frac{V_{cb}}{V_A}\right) I_s e^{V_{be}/V_t} \quad (1185)$$

where V_A is the Early voltage identified in figure 728.

To see why the collector resistance is larger when emitter-current control is used, recall that the collector current in this configuration is alpha times the emitter current where alpha is slightly less than unity.

$$I_c = \alpha I_e \quad (1186)$$

Also recall from section 29.8 that a large change in beta results in only a very slight change in alpha.

$$\alpha = \frac{\beta}{\beta + 1} \quad (1187)$$

As a result, although an increase in collector-emitter voltage increases beta it does not much effect alpha or the collector current. This results in a large output resistance.

Comparing Ramp Generators

The circuit of figure 725 on page 819 uses base voltage control. The circuit of figure 718 on page 813 uses emitter current control. We'd expect the current generator transistor in this latter circuit to have a much larger output resistance and hence the circuit will generate a ramp with better linearity.

29.12 Temperature and Base-Emitter Voltage

Most BJT's spend their life working at reasonable temperatures, and temperature effects are not important. But there are some situations where temperature must be taken into account. In this and the next section, we'll look at these effects.

Base-Emitter Voltage

The base-emitter junction is a forward-biased diode. For silicon, its forward voltage drop is approximately 0.6 volts. As shown in Appendix B (section 29.40) the base-emitter forward-voltage drop changes by about $-2\text{mV}/^\circ\text{C}$. The minus sign indicates that the forward voltage drop *decreases* with temperature.

Now consider the circuit in figure 730. For a constant value of input voltage E_1 , we would like the collector current to have a predictable value. If the base-emitter voltage is constant, as we assumed in earlier sections, then the voltage across the emitter resistor is fixed by the input voltage and this in turn fixes the emitter current and the collector current.

However, if the base-emitter voltage changes with temperature, the voltage across the emitter resistor changes and this affects the collector current. But at $-2\text{mV}/^\circ\text{C}$, how bad can it be? Under the right circumstances, *really bad*, as it turns out.

An Unstable Circuit

To choose an extreme example, suppose E_1 is set to 1V. Suppose furthermore that the circuit must operate over the extremes of the climate in Canada, which might be -10°C to $+50^\circ\text{C}$. The base-emitter voltage at -10°C is 0.6 volts. The value of R_e is chosen so that it generates an emitter current of 1mA under these conditions.

What is the effect when the transistor changes temperature to $+50^\circ\text{C}$?

- The base-emitter voltage changes by

$$\begin{aligned}\Delta V_{be} &= 60^\circ\text{C} \times 2\text{mV}/^\circ\text{C} \\ &= 120\text{mV}\end{aligned}$$

- This change in voltage appears at the top end of the emitter resistor, so the emitter current changes by

$$\begin{aligned}\Delta I_e &= \frac{\Delta V_{be}}{R_e} \\ &= \frac{0.120}{400} \\ &= 0.3\text{mA}\end{aligned}$$

This is a change in the original emitter (and collector) current of some 30%, which is unlikely to be acceptable.

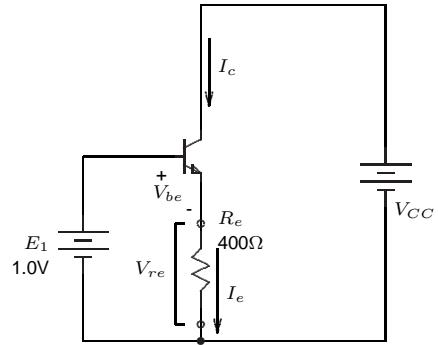


Figure 730: The Effect of V_{be}

Stabilizing for ΔV_{be}

The drift problem of the previous example arises because the voltage across the emitter resistor is comparable to the base-emitter voltage. A simple fix is to increase the voltage across R_e by increasing the base supply voltage E_1 . (Of course, the emitter resistance R_e must increase to maintain the emitter current at the same value, 1mA.). The larger this voltage, the less the effect of a change in base-emitter voltage.

Increasing the value of the base supply voltage reduces the collector-emitter voltage across the transistor. This is unacceptable in some applications. An alternative method of stabilization that keeps the emitter voltage unchanged is shown in figure 731(a).

In this circuit, a second transistor Q_2 compensates for changes in the base-emitter voltage of Q_1 .

Resistor R_1 drives current through Q_2 that is equal to the emitter current of Q_1 . The base of transistor Q_2 is connected to its collector. As a result, its collector-emitter voltage is held at its base-emitter voltage, about 0.6 volts.

If the two transistors are identical, operated at the same emitter current, and held at the same temperature then their base-emitter voltages will be equal. The exact value of the base-emitter voltage is no longer important because they cancel.

By KVL around the base loop:

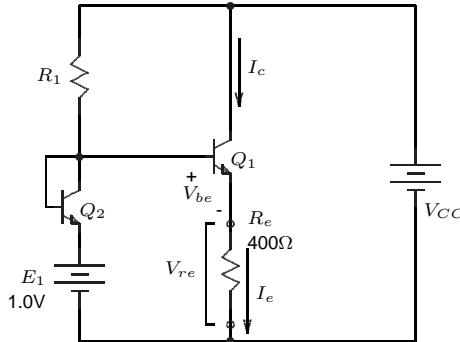
$$+E_1 + V_{be2} - V_{be1} - V_{re} = 0$$

and

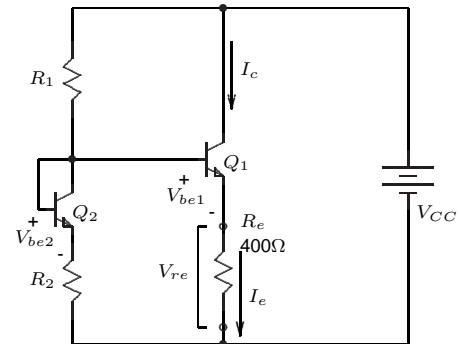
$$V_{re} = E_1$$

A practical circuit is shown in figure 731(b). The value of R_2 is made equal to R_e , and R_1 is proportioned to dump a current equal to I_e through Q_2 and R_2 .

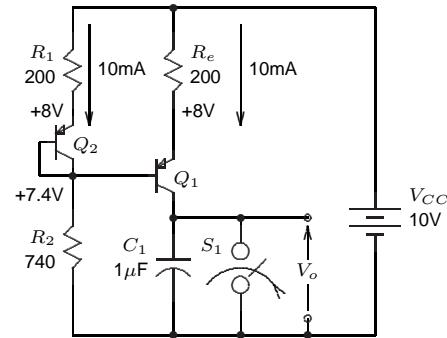
This technique of V_{be} cancellation can be applied to the ramp generator circuit of section 29.9. The revised circuit is shown in figure 731(c). Resistors R_1 and R_2 are proportioned to drive 10mA through the compensating transistor Q_2 . Then (assuming that the base current is negligible) the current through the current-generator transistor Q_1 will also be 10mA, independent of temperature.



(a) Compensating ΔV_{be}



(b) Practical Circuit



(c) Final Circuit

Figure 731: Ramp Generator with V_{be} Compensation

29.13 Temperature and Leakage Current

To a first approximation, a reverse-biased diode behaves as an open-circuit. However, there are actually two components of reverse current.

- One of these components is called *saturation current*²³² and given by the symbol I_s . It is called saturation current because it is constant and essentially independent of the reverse voltage up to the breakdown voltage of the device. A typical value for saturation current in a silicon diode is in the order of 10^{-14} amps. Saturation current figures prominently in the function relating diode forward voltage to current, as we will see later.
- The other component of reverse current is *leakage current*, which is typically much larger than the saturation current. Typical values are in the order of 10^{-10} amps.

Both these current values are quoted for room temperature and increase exponentially with temperature. Leakage current dominates. It is important to note that **leakage current doubles in magnitude at some rate between 8°C increase in temperature and 10°C increase in temperature** [252], [77].

Example

A certain diode has a reverse leakage current of 0.1nA when the junction is at 20°C. If leakage current doubles with each 8°C increase in temperature, what is the leakage current at 100°C?

Solution

In formula form, the relationship of leakage current and temperature is:

$$I_T = I_o 2^{\left(\frac{T - T_o}{8}\right)} \quad (1188)$$

where

I_o is the leakage current at temperature T_o

I_T is the leakage current at temperature T

Then

$$\begin{aligned} I_T &= I_o 2^{\left(\frac{T - T_o}{8}\right)} \\ &= (0.1 \times 10^{-9}) 2^{\left(\frac{100 - 20}{8}\right)} \\ &= 102.4 \text{nA} \end{aligned}$$

Leakage Current in the BJT

To see why this leakage current might be important in the BJT, consider figure 732(a) and its equivalent circuit 732(b). In this circuit, the collector-base diode is reverse biased and conducts leakage current, which is known as I_{cbo} (current, collector-base, with the emitter terminal open circuit or not conducting, as we shall see.) The leakage current flows through any resistance R_b in the base, creating a base-emitter voltage V_{be} .

²³²This saturation current is unrelated to the *saturation voltage* of the BJT.

For the transistor to conduct, the base-emitter diode must be forward biased, that is, the base-emitter voltage must be at or over 0.6 volts, the threshold voltage for a silicon diode. For small values of leakage current, the base-emitter voltage is negligible, the base-emitter diode is not conducting, and there is no collector or emitter current.

At room temperature, this is the situation. Now consider that the junction temperature is raised.

- The leakage current doubles with each 8°C increase in temperature.
- If the leakage current becomes sufficient to create 0.6 volts or more across the base-emitter diode, the transistor will begin to conduct emitter and collector current.
- This collector current increases the power dissipation in the transistor, which causes it to heat more. The result is a vicious circle in which the transistor is conducting when it should be shut off, and in extreme cases could result in the destruction of the device.

In small signal transistors, where the devices operate near room temperature and the leakage currents are small to start with, this is unlikely to be a problem. However, power BJTs have a large junction area and therefore large leakage currents. As well, they are commonly operated at high temperatures. The combination of these two factors makes leakage current a potential problem in a power transistor.

Example

A certain NPN power transistor has a leakage current I_{cbo} of $1.6\mu\text{A}$ when the junction is at 25°C . If the maximum allowable junction temperature is 125°C and we wish to ensure that the leakage current never causes the device to conduct, what is the maximum value of R_b ?

Solution

First, we will determine the worst case leakage current, which occurs at a temperature of 125°C . Invoking equation 1188, we have:

$$\begin{aligned} I_T &= I_o 2 \left(\frac{T - T_o}{8} \right) \\ &= (1.6 \times 10^{-6}) 2 \left(\frac{125 - 25}{8} \right) \\ &= 9.26\text{mA} \end{aligned}$$

Let us assume that the maximum allowable voltage across R_b will be 0.4 volts, well below the threshold of 0.6 volts. Then the maximum allowable value of the base resistance is:

$$R_b = \frac{V_{be}}{I_T}$$

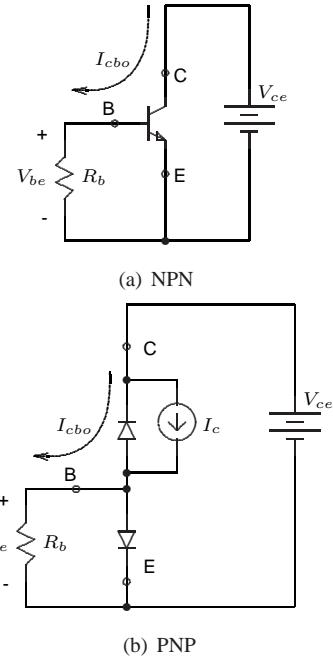


Figure 732: Leakage Current I_{cbo} in the BJT

$$\begin{aligned}
 &= \frac{0.4}{9.26 \times 10^{-3}} \\
 &= 43\Omega
 \end{aligned}$$

If this resistance is inconveniently small, there are a couple of options. One is to return the base to a negative voltage when the transistor is OFF. Then the value of R_b can be increased and there is a larger safety margin before the transistor conducts due to leakage current.

Other options are to lower the maximum junction temperature by using a larger heatsink, or to use a different device entirely. Power MOSFETs do not suffer from this particular problem, one reason of several why they are used in preference to the power BJT in modern designs.

Leakage Current in Small Signal Transistors

Leakage current is a negligible problem in small signal silicon transistors²³³. Nonetheless, when a transistor must be firmly switched OFF, it is advisable to ensure that there is some resistance between base and emitter.

²³³Older textbooks on transistor circuit design devote significant material to the issue of leakage current. The transistors of that day were germanium, in which leakage current is an order of magnitude larger than silicon, and therefore a real problem even at slightly elevated temperature.

29.14 DC BJT Circuits and Equations

It is often necessary to establish certain currents and voltages in a circuit that contains a BJT. This section shows a collection of single-BJT circuits and the corresponding circuit equations. We'll start with a circuit that contains one resistor and then move in steps to more complicated circuits.

	CIRCUIT	EQUIVALENT	EQUATIONS
1			$+V_{BB} - V_{Rb} - V_{be} = 0 \quad (1189)$ $V_{Rb} = I_b R_b \quad (1190)$ $I_c = \beta I_b \quad (1191)$ $V_{ce} = V_{CC} \quad (1192)$
2			$+V_{BB} - V_{be} - V_{Re} = 0 \quad (1193)$ $V_{Re} = I_e R_e \quad (1194)$ $I_c = \alpha I_e \quad (1195)$ $+V_{CC} - V_{ce} - V_{Re} = 0 \quad (1196)$
3			$+V_{BB} - V_{Rb} - V_{be} - V_{Re} = 0 \quad (1197)$ $V_{Rb} = I_b R_b \quad (1198)$ $V_{Re} = I_e R_e \quad (1199)$ $I_e = (\beta + 1) I_b \quad (1200)$ $+V_{CC} - V_{ce} - V_{Re} = 0 \quad (1201)$
4			<p>Same as Case 1 above except equation 1192:</p> $+V_{CC} - V_{Rc} - V_{ce} = 0 \quad (1202)$ $V_{Rc} = I_c R_c \quad (1203)$

DC BJT Circuits (continued)

5			<p>The emitter and collector current are the same as Case 2.</p> $+ V_{CC} - V_{Rc} - V_{ce} - V_{Re} = 0 \quad (1204)$ $V_{Rc} = I_c R_c \quad (1205)$ $V_{Re} = I_e R_e \quad (1206)$
6			<p>The emitter and collector current are the same as Case 3. The collector current is the same as Case 5.</p>
7			<p>The emitter source V_{EE} assumes the function of V_{BB} in Case 2 so the emitter current I_e is calculated in a similar manner. Then:</p> $V_{Re} = I_e R_e \quad (1207)$ $+ V_{EE} + V_{CC} - V_{ce} - V_{Re} = 0 \quad (1208)$
8			<p>The calculation of base and emitter current is similar to Case 3, with V_{BB} replaced by V_{EE}. The collector-emitter voltage is similar to Case 7.</p>

DC BJT Circuits (continued)

9			<p>The emitter current is determined as in Case 7. Then:</p> $I_c = \alpha I_e \quad (1209)$ $V_{Rc} = I_c R_c \quad (1210)$ $+V_{EE} + V_{CC} - V_{Rc} - V_{ce} - V_{Re} = 0 \quad (1211)$
10			<p>The emitter current is determined as in Case 8. The collector-emitter voltage is determined as in Case 9.</p> <p>For the base-emitter loop:</p> $+V_{EE} - V_{Rb} - V_{be} - V_{Re} = 0 \quad (1212)$ $V_{Rb} = I_b R_b \quad (1213)$ $+V_{Re} = I_e R_e \quad (1214)$ $I_e = I_c / \alpha \quad (1215)$ <p>For the Collector-emitter loop:</p> $+V_{CC} - V_{Rc} - V_{ce} - V_{Re} + V_{EE} = 0 \quad (1216)$ $V_{Rc} = I_c R_c \quad (1217)$ $I_b = I_c / \beta \quad (1218)$

Note:

In all cases, we assume that the transistor is in its active region, that is, neither cutoff nor saturated. For this to be true, the base-emitter junction must be forward biased (there must be current into the base) and the collector-emitter voltage V_{ce} must be greater than the saturation voltage of the transistor.

Bjt Bias Example: Case 10

While the bias equations are a major component of the design, there is also a certain element of engineering judgement involved – especially if the design is to be done quickly. Certain quantities are usually known at the start of the design:

Quantity	Function	Example value
V_{CC}, V_{EE}	Power supply voltages, usually specified by other requirements of the design.	12 volts
V_{ce}	Collector-emitter voltage. Depends on the application, but a typical choice is at approximately half the supply voltage.	6 volts
β, α	Beta is determined by the choice of transistor. Alpha then follows.	$\beta = 100$, so $\alpha = 0.990$
V_{be}	Determined by the type of transistor, usually 0.6V for silicon	0.6 volts
I_C	Collector current, determined by optimum current gain, current drive to following stage, or required transconductance	1mA

The objective of this exercise is to put the collector-emitter voltage at some value, and ensure that it stays there in spite of tolerances in beta and V_{be} . The collector-emitter voltage is determined by the supply voltage, collector resistance and collector current. Consequently, the objective is to put the collector current at its bias value and keep it there.

Emitter Loop Equation

We are almost ready to begin calculating, but there is an additional detail missing. Equation 1212 is underdetermined: we need to know the relative proportions of base resistance voltage drop V_{Rb} and emitter resistance voltage drop V_{Re} . The transistor parameter beta has a large tolerance. The voltage across resistor R_b is directly affected by the beta of the transistor, (equation 1213) so we should expect a substantial variation. Consequently *to keep the transistor bias unaffected by variations in beta, we should make the voltage across the base resistance as small as possible*. That way, its substantial variation is still only a small fraction of the other voltages in that loop.

It's possible to relate the tolerance in beta to the allowed tolerance in collector current, but for now we'll just make the voltage drop across the base resistance a small fraction of the other voltages, say $V_{Rb} = 0.1$ volts.

Now equation 1212 is determined and we can solve it for the unknown quantity, which is V_{Re} :

$$\begin{aligned} V_{Re} &= +V_{EE} - V_{Rb} - V_{be} \\ &= 12 - 0.1 - 0.6 \\ &= 11.2 \text{ volts} \end{aligned}$$

Notice that there is little error in ignoring V_{Rb} entirely – at least to get a first cut at the design. It's a short step to get the emitter resistance, using equation 1215.

$$\begin{aligned} I_e &= I_c/\alpha \\ &= 1 \times 10^{-3}/0.990 \\ &= 1.01 \text{ mA} \end{aligned}$$

We could have made the usual approximation that $\alpha \approx 1$ without much error: the collector and emitter currents are very nearly equal.

Now we can use equation 1219 to calculate the emitter resistance R_e .

$$\begin{aligned} R_e &= V_{Re}/I_e \\ &= 11.2/(1.01 \times 10^{-3}) \\ &= 11\text{k}\Omega \end{aligned}$$

Collector Resistor

Using equation 1216 we can determine the voltage across the collector resistance R_c and then its resistance:

$$\begin{aligned} V_{Rc} &= +V_{CC} - V_{ce} - V_{Re} + V_{EE} \\ &= 12 - 6 - 11.2 + 12 \\ &= 6.8 \text{ volts} \end{aligned}$$

Equation 1217 gives us the collector resistance R_c :

$$\begin{aligned} R_c &= V_{Rc}/I_c \\ &= 6.8/(1 \times 10^{-3}) \\ &= 6\text{k}8\Omega \end{aligned}$$

Base Resistance

The base current is related to the collector current by beta, equation 1218:

$$\begin{aligned} I_b &= I_c/\beta \\ &= 1 \times 10^{-3}/100 \\ &= 10\mu A \end{aligned}$$

The base resistance is given by equation 1213:

$$\begin{aligned} R_b &= V_{Rb}/I_b \\ &= 0.1/(10 \times 10^{-6}) \\ &= 10\text{k}\Omega \end{aligned}$$

The final circuit is shown in figure 733. Notice that KVL is satisfied around the outer loop and base-emitter loop. KCL is satisfied at the transistor.

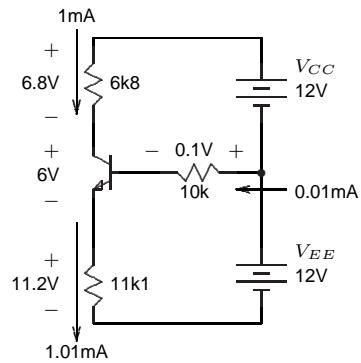


Figure 733: Final Circuit

29.15 Bias Stability and Beta

In this section, we'll show for the previous circuit that the dependence of collector current on beta is minimized for $R_b/\beta \ll R_e$.

The collector current is controlled with the base and emitter currents, so we begin with equation 1212, the KVL equation for the emitter loop:

$$+ V_{EE} - V_{Rb} - V_{be} - V_{Re} = 0 \quad (1219)$$

We'd like to get this into a form that includes collector current and beta. Then we can find the dependence of collector current on beta, $dI_c/d\beta$, and see how best to minimize that. The following equations move us in that direction:

$$V_{Rb} = I_b R_b \quad (1220)$$

$$I_b = I_c/\beta \quad (1221)$$

$$V_{Re} = I_e R_e \quad (1222)$$

$$I_e = I_c/\alpha \quad (1223)$$

$$\begin{aligned} \alpha &= \beta/(\beta + 1) \\ &\approx 1 \end{aligned} \quad (1224)$$

Collapse equations 1224 through 1220 back into equation 1219:

$$+ V_{EE} - V_{be} - \frac{I_c}{\beta} R_b - I_c R_e = 0 \quad (1225)$$

Solve for I_c :

$$I_c = (V_{EE} - V_{be}) \left(\frac{\beta}{R_b + \beta R_e} \right) \quad (1226)$$

Now determine the dependence of collector current on beta:

$$\frac{dI_c}{d\beta} = \frac{d}{d\beta} (V_{EE} - V_{be}) \left(\frac{\beta}{R_b + \beta R_e} \right) \quad (1227)$$

Applying the identity $d(u \cdot v) = u \cdot dv + v \cdot du$, we have:

$$\frac{dI_c}{d\beta} = (V_{EE} - V_{be}) \frac{d}{d\beta} \left(\frac{\beta}{R_b + \beta R_e} \right) + \left(\frac{\beta}{R_b + \beta R_e} \right) \frac{d}{d\beta} (V_{EE} - V_{be}) \quad (1228)$$

Since the term $(V_{EE} - V_{be})$ is a constant its differential is zero. Then:

$$\begin{aligned} \frac{dI_c}{d\beta} &= (V_{EE} - V_{be}) \frac{d}{d\beta} \left(\frac{\beta}{R_b + \beta R_e} \right) \\ &= (V_{EE} - V_{be}) \frac{d}{d\beta} \left(\frac{1}{\frac{R_b}{\beta} + R_e} \right) \end{aligned} \quad (1229)$$

If the value of R_b/β is made small compared to R_e , then the differential of the last term becomes the differential of $1/R_e$, which is zero. That is, the collector current is not dependent on the value of beta. Consequently, if R_b/β is made small compared to R_e then the value of beta does not change the bias point.

Pursuing this a bit further, we can show that this is equivalent to making the voltage across R_b small compared to the voltage across R_e . Starting with:

$$\frac{R_b}{\beta} \ll R_e \quad (1230)$$

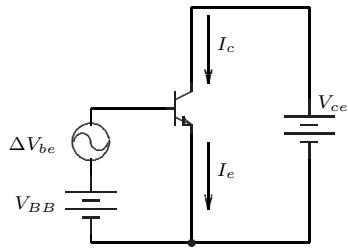
Multiply both sides by I_e .

$$I_e \frac{R_b}{\beta} \ll I_e R_e \quad (1231)$$

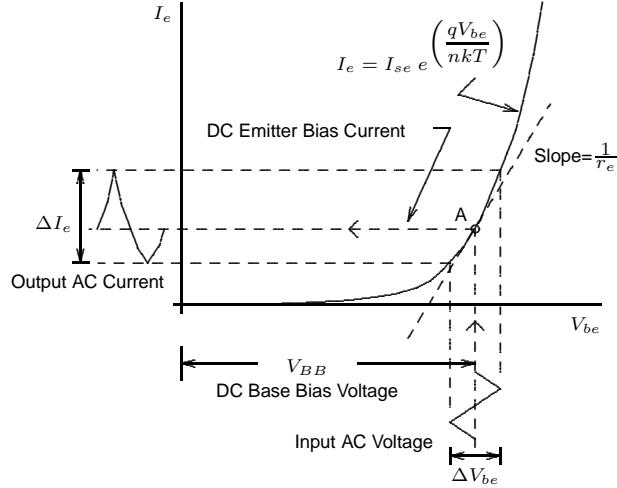
But $I_e/\beta \approx I_b$. Also, $V_{Rb} = I_b R_b$ and $I_e R_e = V_{Re}$ Substitute these in equation 1231, and we have our result:

$$V_{Rb} \ll V_{Re} \quad (1232)$$

29.16 Emitter Incremental Resistance and BJT Transconductance



(a) Circuit



(b) Collector Current vs Base Voltage

Figure 734: Base Voltage and Collector Current

We saw in section 29.5 that the input base current controls a much larger output collector current, where the constant of proportionality is simply β . In this section, we'll look at the relationship between input base *voltage* and emitter current. The collector current is alpha times the emitter current, where $\alpha \approx 1$, so the collector current has a similar relationship.

A test circuit is shown in figure 734(a). A fixed voltage V_{BB} approximately equal to 600mV forward-biases the base-emitter diode of the transistor to point A on the base-voltage emitter-current characteristic of figure 734(b).

The $I_e : V_{be}$ curve of figure 734(b) has an exponential slope that reflects the behaviour of the base-emitter diode of the transistor.

A small AC voltage ΔV_{be} is added to the fixed bias voltage V_{BB} . The resultant AC emitter current ΔI_e is shown to the left of the graph. A number of important points can be derived from this graph:

- As the peak-peak value of the input AC voltage ΔV_{be} is increased, the waveform of emitter current becomes progressively more distorted.

- As the DC bias voltage V_{BB} increases, this moves the bias point **A** further to the right and up the characteristic. Since the characteristic becomes steeper at larger values of V_{BB} , the same value of AC input voltage ΔV_{be} corresponds to a larger value of ΔI_e .
- The slope of the tangent at the operating point has the units of amps/volt, or $1/\Omega$ and is equal to the reciprocal of the incremental resistance of the base-emitter diode.

Now we'll determine a relationship for the slope of this tangent: The equation relating the base-emitter voltage to emitter current in a BJT transistor is:

$$I_e = I_{es} e^{\left(\frac{qV_{be}}{nkT}\right)} \quad (1233)$$

Assuming that the temperature is 25°C and $n \approx 1$, we can put

$$\begin{aligned} V_t &= \frac{nkT}{q} \\ &= 0.026V \end{aligned} \quad (1234)$$

This simplifies the notation. Then

$$I_e = I_{es} e^{\left(\frac{V_{be}}{V_t}\right)} \quad (1235)$$

The slope of this characteristic is

$$\frac{dI_e}{dV_{be}} = \frac{d}{dV_{be}} \left[I_{es} e^{\left(\frac{V_{be}}{V_t}\right)} \right] \quad (1236)$$

We can apply the identity

$$\frac{d}{dx} e^u = \frac{du}{dx} e^u \quad (1237)$$

in which case equation 1236 becomes

$$\begin{aligned} \frac{dI_e}{dV_{be}} &= \frac{d}{dV_{be}} \left(\frac{V_{be}}{V_t} \right) \left[I_{es} e^{\left(\frac{V_{be}}{V_t}\right)} \right] \\ &= \frac{1}{V_t} \left[I_{es} e^{\left(\frac{V_{be}}{V_t}\right)} \right] \end{aligned} \quad (1238)$$

The term in the square brackets is simply the DC value of the emitter current, I_e , so

$$\frac{dI_e}{dV_{be}} = \frac{I_e}{V_t} \text{ mhos} \quad (1239)$$

Expressing this as a resistance, we invert both sides:

$$\frac{dV_{be}}{dI_e} = \frac{V_t}{I_e} \Omega \quad (1240)$$

Now let us invest this with a physical interpretation. Equation 1240 says that the inverse of the slope can be interpreted as a resistance, the *incremental emitter resistance*, which we will call r_e . We can relate the AC base-emitter voltage v_{be} to the AC emitter current i_e with the incremental emitter resistance r_e :

$$i_e = \frac{v_{be}}{r_e} \quad (1241)$$

This is a useful relationship in determining the AC voltage gain and input resistance of various BJT amplifier stages.

Just as the slope of the curve changes with different values of emitter current, so does this resistance, and its value is given by equation 1242:

$$r_e = \frac{V_t}{I_e} \Omega \quad (1242)$$

where V_t is 26mV at room temperature and I_e is the DC emitter current.

This is a very useful and important relationship.

Example

Suppose a BJT is biased at a DC emitter current of 1 mA. If the input AC base-emitter voltage is 10mV peak-peak, what is the emitter AC current?

Solution

First, determine the value of the emitter incremental resistance r_e with equation 1242:

$$\begin{aligned} r_e &= \frac{V_t}{I_e} \\ &= \frac{0.026}{0.001} \\ &= 26 \Omega \end{aligned}$$

Now we can relate the AC base-emitter voltage to the AC emitter current, with equation 1241:

$$\begin{aligned} i_e &= \frac{v_{be}}{r_e} \\ &= \frac{0.010}{26} \\ &= 0.38 \text{ mA peak-peak} \end{aligned}$$

Transconductance

The relationship between the AC base-emitter voltage and the **collector** current (dI_c/dV_{be}) is known as the *transconductance* of the device and given the symbol g_m . The term transconductance (*transfer conductance*, because it relates the output current to the input voltage) originated to characterize vacuum tubes but is used with the BJT, JFET and MOSFET transistors. Transconductance is an important parameter and arises repeatedly in the evaluation of amplifier circuits.

$$g_m = \frac{dI_c}{dV_{be}} \text{ mhos} \quad (1243)$$

Since $I_c = \alpha I_e$ and $dV_{be}/dI_e = r_e$, then

$$\begin{aligned} g_m &= \frac{dI_c}{dV_{be}} \\ &= \frac{\alpha dI_e}{dV_{be}} \\ &= \alpha \frac{1}{r_e} \\ &\approx \frac{1}{r_e} \end{aligned} \quad (1244)$$

Consequently, the emitter incremental resistance r_e and the transconductance g_m are reciprocals of each other.

In modelling the BJT, one can use transconductance or incremental emitter resistance. They are both useful in different circumstances. However, we will focus on the incremental emitter resistance r_e .

29.17 The T Model for the Transistor

The concepts of *emitter incremental resistance* and *transconductance* lead us to a very simple and useful model for the transistor. Figure 735 shows the NPN version of the model.

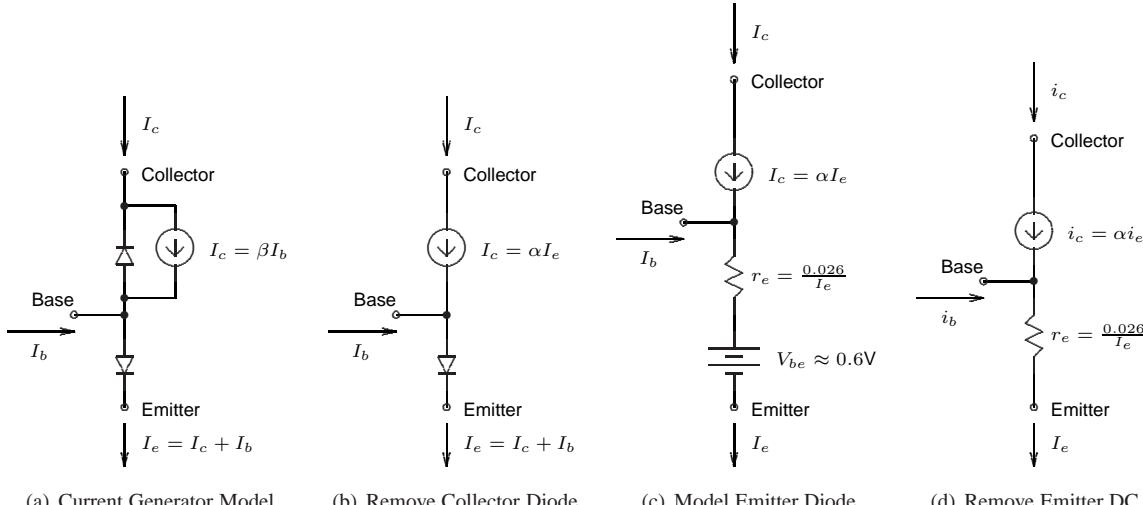


Figure 735: Evolution of the T Model of the BJT

Figure 735(a) shows the Current Generator Model from figure 704 on page 804.

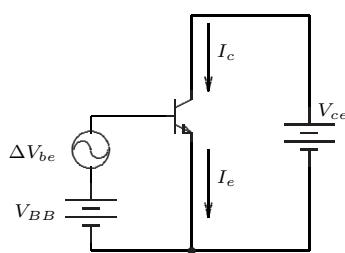
The collector-base diode is reverse biased when the transistor is used as an amplifier, so it appears as an open circuit and can be removed as shown in figure 735(b). As well, the collector current is shown as αI_e to emphasize its relationship to the emitter current.

In figure 735(c) we model the base-emitter diode with a DC voltage source V_{be} in series with the emitter incremental resistance r_e .

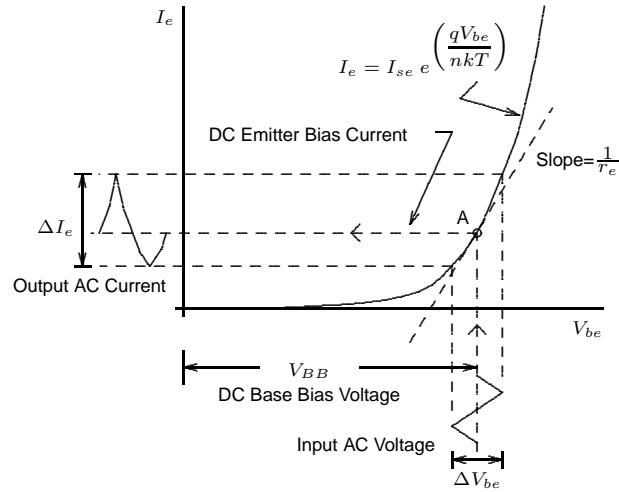
Figure 735(d) shows the final T model for AC voltages and currents. The currents are now indicated as lower case to emphasize that they are AC quantities. The DC component of the base-emitter voltage can be removed, since there is no AC voltage across it. The emitter incremental resistance is indicated as a function of the DC emitter current I_e .

This is a very useful model for AC circuit analysis, and sufficiently accurate for many applications. We'll use it extensively in the material that follows.

29.18 Transconductance and Distortion



(a) Circuit



(b) Collector Current vs Base Voltage

Figure 736: Base Voltage and Collector Current

Figure 736(b) suggests that the nonlinearity of the $V_{be} - I_c$ characteristic will cause distortion between the input and output waveforms. As it turns out, the distortion becomes significant if the peak value of the AC input (which we'll call $V_p = \Delta V_{be}/2$) becomes comparable to V_t , that is, approximately 26mV.

Harmonic distortion occurs when the amplifier generates harmonics of the input sine wave frequency. For example, if the input frequency is at f , then the second harmonic appears at $2f$, the third harmonic at $3f$ and so on. The degree of distortion is related to the magnitude of these harmonics with respect to the fundamental. Since the harmonics usually fall off in magnitude with increasing harmonic number, an approximate idea of the distortion can be obtained by considering the first few harmonics, such as those at $2f$ and $3f$.

In this situation, one can show that the fraction of second harmonic in the output current, relative to the fundamental component is given by

$$\text{HD}_2 \approx \frac{1}{4} \frac{V_p}{V_t} \quad (1245)$$

Similarly, the third harmonic component is:

$$\text{HD}_3 \approx \frac{1}{24} \left(\frac{V_p}{V_t} \right)^2 \quad (1246)$$

Example

If the second harmonic in the output is to be kept less than 5% of the fundamental, what is the maximum allowable magnitude of the peak AC input signal V_p ? Assume V_t is 26mV.

Solution

From equation 1245 we can write:

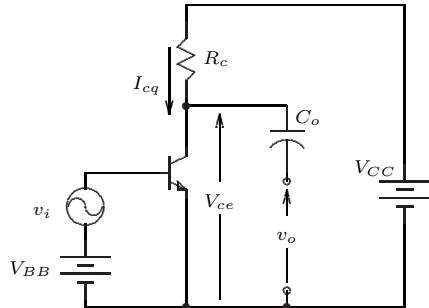
$$\begin{aligned} \text{HD}_2 &= \frac{1}{4} \frac{V_p}{V_t} \\ &= 0.05 \end{aligned}$$

Rearranging and solving for V_p :

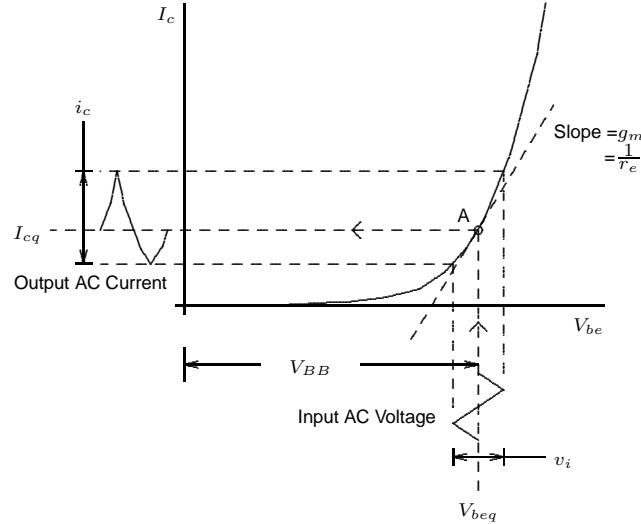
$$\begin{aligned} V_p &= 4 \times 0.026 \times 0.05 \\ &= 5.2\text{mV} \end{aligned}$$

So the AC input signal must be restricted to less than 5.2mV to avoid generating 2nd harmonic content greater than 5% of the fundamental. This is a very small AC signal. To be able to handle larger signals, the amplifier must be *linearized* in some manner or overall negative feedback must be used to reduce the distortion.

29.19 The Common Emitter Amplifier



(a) Circuit



(b) Collector Current vs Base Voltage

Figure 737: A Basic Common Emitter Amplifier

Overview

In the remainder of this section we show how a single BJT can be used as an amplifier. There are three possible configurations for this amplifier stage, defined by the terminal that is common to the input and output signals: common emitter, common collector (emitter follower) and common base.

Of the three possible amplifier configurations, the common-emitter is the most versatile and the most frequently used to provide voltage and current gain. The input resistance is equal to beta times the emitter resistance, so this can be adjusted to meet various circuit requirements. (If very large input resistance is required, a JFET is a

better choice of amplifying device.) The voltage gain is equal to the ratio of collector to emitter resistance, which can be a large value. The current gain²³⁴ is simply beta. Consequently, it is easy to obtain a large power gain with this amplifier configuration.

The output resistance is rather large, being the parallel combination of the external collector resistance and the internal collector resistance r_c . In a current generator, this is desirable. In many amplifier applications a lower output resistance is ultimately required, but that is easily obtained with a subsequent emitter follower stage.

Basic Function

In this section, we'll examine the common-emitter BJT stage as an amplifier for AC signals. To completely characterize this amplifier, we need to determine the voltage gain, input resistance, output resistance and current gain.

Figure 737(a) shows a very simple common-emitter amplifier. This is not yet an entirely practical circuit but it will get us started. Figure 737(b) shows the effect of the transconductance and is similar to the graphs of previous sections. However, the input signal voltage has been relabelled from ΔV_{be} to v_i to indicate that this is an AC voltage. Similarly, the varying collector current is relabelled from ΔI_c to i_c .

First, let us consider the DC conditions. The bias voltage V_{BB} establishes the DC collector current at some value I_{cq} (which might be chosen to optimize the transistor properties.) The collector resistor R_c translates that DC collector current into a DC collector-emitter voltage:

$$V_{ce} = V_{CC} - I_c R_c \quad (1247)$$

We might choose V_{BB} and R_c so that V_{ce} is half the supply voltage, for example. These DC conditions represent the *quiescent* (no AC signal) condition for the amplifier. The transistor is set up so that the collector-emitter voltage can increase **and** decrease, and can thereby trace out an AC signal.

The AC input voltage signal v_i is in series with the bias voltage V_{BB} , so it swings around point A on the graph. Consequently, the output AC current is given by²³⁵

$$i_c = \frac{v_i}{r_e} \quad (1248)$$

where

$$r_e = \frac{V_t}{I_E} \quad (1249)$$

I_E is the DC emitter bias current. (I_E is approximately equal to I_{cq} on the graph.)

$$\begin{aligned} V_t &= \frac{n k T}{q_e} \\ &= 26 \text{ mV at room temperature} \end{aligned} \quad (1250)$$

The AC collector current i_c creates an AC voltage across the collector resistance R_c . The output coupling capacitor passes this AC voltage to the output while blocking the DC component of collector-emitter voltage. Consequently, the output AC voltage is:

$$v_o = -R_c i_c \quad (1251)$$

²³⁴Bear in mind that the voltage gain is obtained with an open-circuited load and the current gain is obtained with a short-circuited load. So the maximum voltage gain and maximum current gain cannot be obtained in a given circuit at the same time.

²³⁵Throughout this section, we could use transconductance g_m or emitter incremental resistance r_e to define the slope of the collector-current base-emitter voltage curve. We'll stick with r_e because it provides a simpler explanation in most cases.

(The negative sign appears because the output voltage decreases as the collector current increases. Taking the differential d/dV_{ce} of equation 1247 yields the same result). Substitute for i_c from equation 1248, and:

$$v_o = -R_c \frac{v_i}{r_e} \quad (1252)$$

The voltage gain of this amplifier is then given by

$$\begin{aligned} A_v &= \frac{v_o}{v_i} \\ &= -\frac{R_c}{r_e} \end{aligned} \quad (1253)$$

The negative sign indicates that there is a polarity inversion between the input and output signals. This is an inverting amplifier.

Although there are a number of details missing from this design, it indicates an important result: *the voltage gain for the common-emitter amplifier is the ratio of the collector resistance R_c to the emitter incremental resistance r_e .*

29.20 The Common Emitter Amplifier, Analysis

While the analysis of the previous section is convincing, it's difficult to extend. A more useful approach is to replace the transistor with its Tee model and then do the analysis.

Figure 738(a) shows the original amplifier circuit. In figure 738(b) the BJT has been replaced with the Tee Model, complete with all its bells and whistles. Figure 738(c) is the result of many simplifications on figure 738(b):

- The DC sources V_{BB} , V_{be} and V_{CC} are all replaced by short circuits, since they have no effect on the alternating currents.
- The base resistor r_b is often small enough to ignore so it is replaced by a short circuit.
- If the collector resistance r_c is large compared to R_c , then it has little effect. In the interest of simplicity, we'll make that assumption, and so it can be replaced with an open circuit.
- The collector current generator is shown as an AC device since we are considering only AC currents.
- The output capacitor C_o is a short-circuit for AC.

Most of the work is in generating the equivalent circuit. Once we have the figure 738(c) it is relatively straightforward to analyse.

Voltage Gain

The definition of voltage gain is

$$A_v = \frac{v_o}{v_i} \quad (1254)$$

where the output voltage is into an open-circuit load.

We begin at the input with the input voltage v_i and work our way to the output voltage v_o :

$$i_e = \frac{v_i}{r_e} \quad (1255)$$

$$i_c = \alpha i_e \quad (1256)$$

$$v_o = -i_c R_c \quad (1257)$$

Collapsing equation 1257 into equation 1256 and then into equation 1255, we have

$$v_o = -\alpha \frac{v_i}{r_e} R_c \quad (1258)$$

Now we can determine the voltage gain:

$$\begin{aligned} A_v &= \frac{v_o}{v_i} \\ &= -\alpha \frac{R_c}{r_e} \\ &\approx -\frac{R_c}{r_e} \end{aligned} \quad (1259)$$

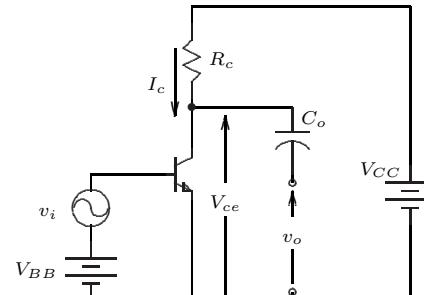
where we have assumed that $\alpha \approx 1$. Equation 1259 is the same as equation 1253, the result we previously obtained by using the diode equation, for the voltage gain of the common emitter amplifier.

Current Gain

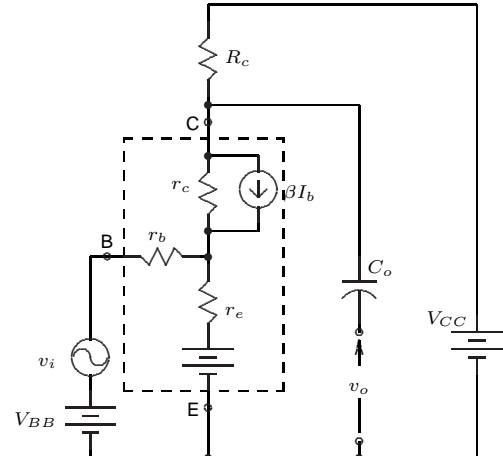
The definition of current gain is

$$A_i = \frac{i_o}{i_i} \quad (1260)$$

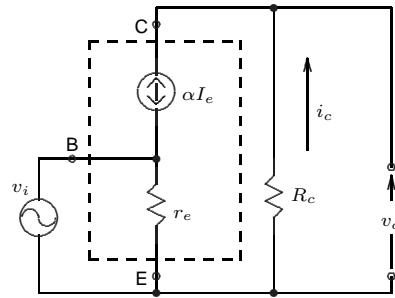
when the load resistance is a short circuit. By inspection, the current gain is simply the ratio of collector current to base current, which is beta.



(a) Amplifier Circuit



(b) Bias Model



(c) AC Model

Figure 738: Common Emitter Amplifier, Model

Input Resistance

The definition of input resistance is:

$$r_i = \frac{v_i}{i_i} \quad (1261)$$

But

$$i_i = i_b \quad (1262)$$

Further, we can relate base current to emitter current as:

$$i_b = \frac{i_e}{\beta + 1} \quad (1263)$$

Collapse these equations backwards and we obtain

$$\begin{aligned} r_i &= \frac{v_i}{i_e / (\beta + 1)} \\ &= (\beta + 1) \frac{v_i}{i_e} \end{aligned} \quad (1264)$$

But

$$\frac{v_i}{i_e} = r_e \quad (1265)$$

Substitute that in equation 1264 and:

$$\begin{aligned} r_i &= (\beta + 1)r_e \\ &\approx \beta r_e \end{aligned} \quad (1266)$$

That is, the emitter resistance multiplied by beta appears as the input resistance.

Incidentally, if we want to take the base resistance into account, it simply adds to the total:

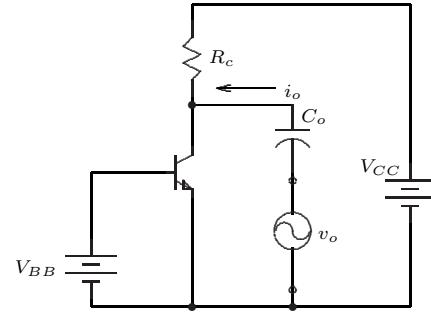
$$r_i = r_b + (\beta + 1)r_e \quad (1267)$$

In practice, the base resistance is usually small compared to the beta-magnified emitter resistance and can be neglected.

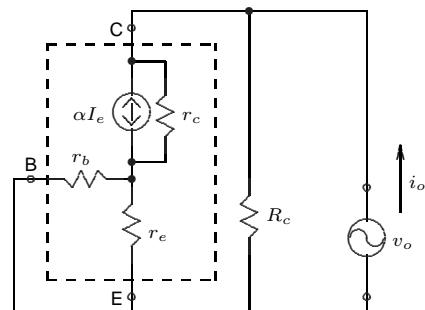
Output Resistance

The conceptual circuit for measuring output resistance is shown in figure 739(a). A voltage source v_o attempts to drive current into the output terminal of the amplifier. The ratio of output voltage to current is the output resistance:

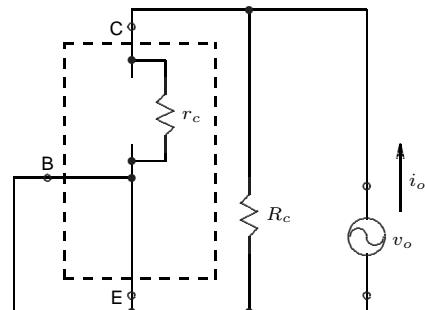
$$\frac{v_o}{i_o} = r_o \quad (1268)$$



(a) Original Circuit



(b) BJT Replaced by Tee Model



(c) AC Equivalent Circuit

Figure 739: Common Emitter Amplifier, Output Resistance

The equivalent circuit is in figure 739(b). It may simplified by:

- Consider that the output capacitor C_o is a short circuit.
- The collector current generator αI_e is an open circuit.
- The parallel combination of base resistance r_b and emitter resistance r_e , compared to the collector resistance r_c , is small enough that r_b and r_e may be ignored and treated as a short circuit.

All the work is in the equivalent circuit. By inspection,

$$r_o = r_c \parallel R_c \quad (1269)$$

For discrete transistors with a physical resistor as R_c , r_c is often in the neighbourhood of $10k\Omega$ and R_c in the vicinity of $5k\Omega$. As we will see, R_c is sometimes replaced with a transistor acting as a constant current generator in which case it is a very large value, 100's of $k\Omega$.

Summary

As shown in figure 740, the approximate relationships for the BJT common emitter amplifier are:

Voltage Gain	A_v	$\frac{R_c}{r_e}$	$r_c \gg R_c$
Current Gain	A_i	β	
Input Resistance	r_i	βr_e	$r_b \ll \beta r_e$
Output Resistance	r_o	R_c	$r_c \gg R_c$

Notice that these values satisfy the generalized amplifier relationship given by equation 580 on page 386. For example, if we substitute values from this summary for voltage gain, input resistance and output resistance into equation 580 to solve for current gain, we obtain:

$$\begin{aligned} A_i &= A_v \frac{r_i}{r_o} \\ &= \left(\frac{R_c}{r_e} \right) \frac{\beta r_e}{R_c} \\ &= \beta \end{aligned}$$

which is the same result we obtained for current gain on page 842.

More exact relationships are:

Voltage Gain	A_v	$\frac{R_c \parallel r_c}{r_e}$
Current Gain	A_i	β
Input Resistance	r_i	$r_b + \beta r_e$
Output Resistance	r_o	$R_c \parallel r_c$

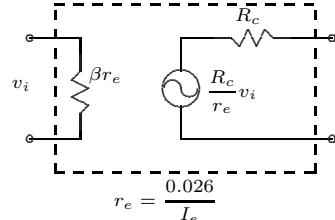


Figure 740: Common Emitter Amplifier Parameters

Example

The transistor shown in figure 741(a) has the following properties:

I_s	$4.95 \times 10^{-13} \text{ A}$
n	1.126
β	100
r_c	$100 \text{ k}\Omega$
r_b	0

The transistor is to be operated at 1mA collector current to optimize β . The supply voltage V_{CC} is 10 volts. Determine:

1. the value of V_{BB} in the bias model of figure 741(b)
2. the value of collector resistance R_c
3. the value of the parameters in the AC model of figure 741(c)

Solution

1. The diode equation 1356 on page 894 relates the base-emitter voltage of the transistor to the collector current. Rearranging that equation to solve for base-emitter voltage, we have

$$\begin{aligned} V_{be} &= nV_t \ln\left(\frac{I_c}{I_s}\right) \\ &= 1.126 \times 0.026 \ln\left(\frac{1 \times 10^{-3}}{4.95 \times 10^{-13}}\right) \\ &= 0.627 \text{ volts} \end{aligned}$$

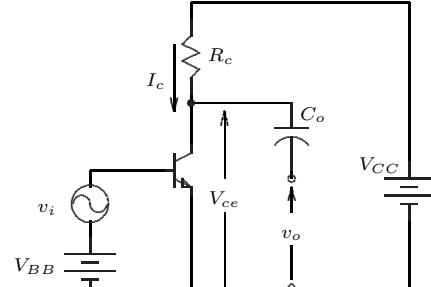
Since the base supply V_{BB} is connected base to emitter in this circuit, it should be set to 0.627 volts.

2. The value of collector resistance R_c affects two things. The bias conditions establish a quiescent value of V_{ce} . The collector resistance also affects the voltage gain – a larger value of R_c increases the voltage gain. In this problem, we'll put the collector-emitter voltage at 5 volts – halfway between V_{CC} and ground. That value provides the maximum room for the collector voltage to swing up and down from its quiescent value. To calculate the value of collector resistance,

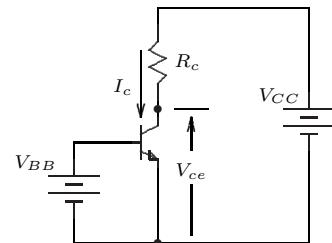
$$V_{ce} = V_{CC} - I_c R_c$$

where I_c is the DC collector bias current. Then:

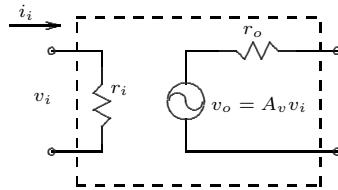
$$R_c = \frac{V_{CC} - V_{ce}}{I_c} = \frac{10 - 5}{1 \times 10^{-3}} = 5 \text{ k}\Omega$$



(a) Amplifier Circuit



(b) DC Bias Model

(c) AC Model
Figure 741: Example Amplifier

3. The first parameter to determine is the emitter incremental resistance r_e :

$$r_e = \frac{0.026}{I_e} = \frac{0.026}{1 \times 10^{-3}} = 26\Omega$$

To determine the parameters of the AC model of figure 741(c), we simply plug numbers into the formulae for the parameters:

$$A_v = \frac{R_c}{r_e} = \frac{5000}{26} = 192\text{v/v}$$

$$r_i = \beta r_e = 100 \times 26 = 2.6k\Omega$$

$$r_o = R_c = 5k\Omega$$

Notice that the value of the internal collector resistance r_c is much larger than the external collector resistance R_c . Then $r_c \parallel R_c \approx R_c$ and this justifies our approximations for voltage gain and output resistance.

Maximum Possible Voltage Gain

What is the maximum possible voltage gain that can be obtained from the amplifier configuration shown in figure 742?

We know that the voltage gain is

$$A_v = \frac{R_c}{r_e} \quad (1270)$$

but

$$r_e = \frac{V_t}{I_e} \quad (1271)$$

If we assume that the transistor is biased so that half the supply voltage appears across R_c , then

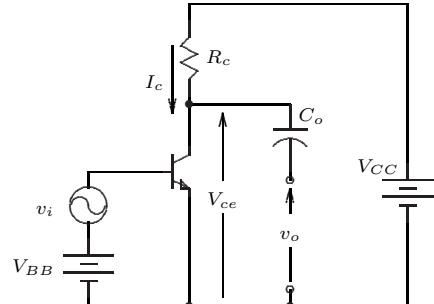


Figure 742: Common Emitter Amplifier

$$\begin{aligned} R_c &= \frac{V_{CC}/2}{I_c} \\ &\approx \frac{V_{CC}/2}{I_e} \end{aligned} \quad (1272)$$

Substitute r_e from equation 1271 and R_c from equation 1272 into equation 1270, and we have:

$$\begin{aligned} A_v &= \frac{V_{CC}/2}{I_e} \div \frac{V_t}{I_e} \\ &= \frac{V_{CC}}{2V_t} \end{aligned} \quad (1273)$$

This is an interesting result. It indicates that the maximum possible voltage gain from a grounded-emitter BJT amplifier, using a resistive load in the collector, is proportional to the collector supply voltage. There is a real limit to the maximum value of the power supply voltage. Many transistors cannot withstand a collector supply voltage of more than 30 volts or so.

Example

What is the maximum possible voltage gain that we could achieve in a circuit of the type shown in figure 742 if the supply voltage V_{CC} is 9 volts?

Solution

Using equation 1273, we have:

$$A_v = \frac{V_{CC}}{2V_t} = \frac{9}{2 \times 0.026} = 173 \text{ volts/volt}$$

There are ways around this limitation. For example, the collector resistor can be replaced by a constant-current source which supplies the required DC collector current while maintaining a very high internal impedance to AC signals. The result is a very large voltage gain, well beyond the limit indicated by equation 1273.

29.21 Common Emitter Amplifier with Emitter Resistance

In section 29.19, we introduced the basic *common-emitter amplifier* circuit. The emitter is directly connected to ground, which results in a significant voltage gain. However, the circuit has several problems. The signal capability is limited to signals below a few millivolts. The bias point is dependent on the base-emitter voltage V_{be} , which changes with temperature. These problems can be addressed by placing a fixed resistor R_e in series with the emitter as shown in figure 743.

Why this works

- **Lower Distortion** The distortion in a common-emitter stage results from the changing slope of the $V_{be} - I_c$ curve (section 29.18). If we make R_e much larger than r_e , then the gain of the stage will remain constant over a complete cycle of the input. This makes the gain more linear and reduces the distortion.

- **Bias Stability** In this circuit, applying KVL to the base-emitter loop provides a relationship for DC emitter current:

$$I_e = \frac{V_{BB} - V_{be}}{R_e} \quad (1274)$$

For a given value of emitter current, the base bias voltage V_{BB} is much larger than it is in a circuit without an emitter resistor. Consequently, temperature-induced changes in V_{be} have much less effect on the emitter (and collector) current, and the bias stability is better.

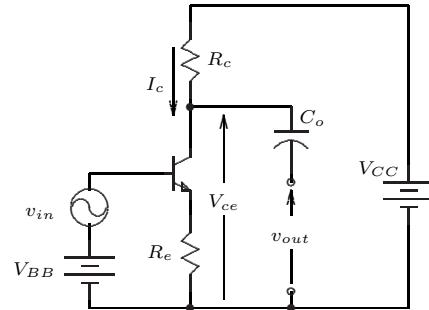


Figure 743: Amplifier with Emitter Resistance

Signal Capability

Figure 744 is helpful in illustrating the behaviour of the signals in the amplifier. The collector voltage swings up and down around its quiescent value V_{cq} as the emitter voltage swings down and up around its quiescent value V_{eq} . There are several constraints on these signals:

- The collector voltage V_c cannot exceed the positive supply voltage
- The emitter voltage V_e cannot go below the negative supply voltage, which in this case is ground (zero volts).
- The two signals approach each other in the second half of the cycle. At the minimum of collector voltage and maximum of emitter voltage, they must not overlap and in fact there must be a small voltage margin V_{cemin} somewhat greater than the saturation voltage of the transistor.
- The peak-peak amplitude of the collector voltage is equal to the voltage gain times the peak-peak amplitude of the emitter voltage.
- The peak-peak amplitude of the emitter voltage is equal to the peak-peak amplitude of the input voltage, V_i .

There are a number of ways we can play with this information. For example, if we know the required output amplitude and voltage gain, we can determine the minimum required value for the supply voltage. Treating the minimum value of V_{ce} as zero for the moment, we have:

$$\begin{aligned} V_{CC} &= V_o + V_i \\ &= V_i(1 + A_v) \end{aligned} \quad (1275)$$

where V_i is the peak-peak value of the input voltage. Anything more than this minimum value of V_{CC} adds to the *headroom* of the amplifier and can be allocated to the minimum value of V_{ce} and the space between the maximum value of V_c .

Similarly, if the supply voltage and voltage gain are known, then the maximum magnitude of the signals can be determined.

In either case, once the supply voltage is known and the signal amplitudes are known, then the quiescent value of collector voltage V_{cq} and emitter voltage V_{eq} can be determined from a diagram like figure 744.

Bias Conditions

When the bias conditions are known, the values of collector resistance R_c , emitter resistance R_e and base bias supply voltage can be determined according to the following equations: The base bias voltage V_{BB} creates a quiescent voltage across the emitter resistor of:

$$V_{eq} = V_{BB} - V_{be} \text{ volts} \quad (1276)$$

This sets up an emitter current

$$I_e = \frac{V_{BB} - V_{be}}{R_e} \quad (1277)$$

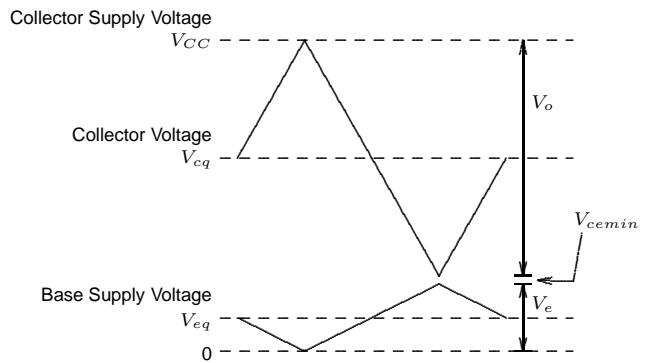


Figure 744: Signal Capability

Then the quiescent collector voltage V_{cq} is:

$$V_{cq} = V_{CC} - I_c R_c \quad (1278)$$

where $I_c \approx I_e$.

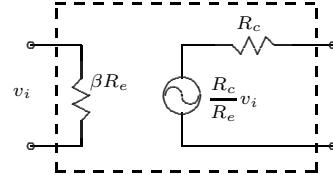


Figure 745: CE Amplifier with R_e , Parameters

AC Parameters

The gain and resistance equations are easily determined: the emitter resistance r_e becomes $r_e + R_e$. If the external is much larger than the intrinsic resistance, then $r_e + R_e \approx R_e$.

Modifying the summary of parameters from page 844, we have the following approximate relationships, summarized in figure 745.

Voltage Gain	A_v	$\frac{R_c}{R_e}$	$r_c \gg R_c$
Current Gain	A_i	β	
Input Resistance	r_i	βR_e	$r_b \ll \beta R_e$
Output Resistance	r_o	R_c	$r_c \gg R_c$

Tradeoffs Around R_e

On the one hand, increasing the emitter resistance R_e reduces distortion, improves the bias stability and increases the input resistance. On the other hand, increasing the emitter resistance reduces the voltage gain and signal-handling capability. If the design should obtain the maximum possible voltage gain without distortion, one would choose R_e to be the minimum size that will satisfy the distortion and stability requirements.

Example

Design an amplifier of the type shown in figure 743, to have a voltage gain A_v of 10V/V and operate at a quiescent collector current I_c of 1mA. The collector voltage supply V_{CC} is 9V.

1. Determine the maximum output voltage swing if the saturation voltage is 0V.
2. Determine values for collector resistance R_c , emitter resistance R_e and base bias voltage supply V_{BB} .

Solution

1. Rearranging equation 1275 to solve for input voltage, we have

$$V_i = \frac{V_{CC}}{(1 + A_v)} = \frac{9}{(1 + 10)} = 0.818 \text{ volts, peak-peak}$$

Then the peak-peak output voltage swing is

$$V_o = A_v V_i = 10 \times 0.818 = 8.18 \text{ volts, peak-peak}$$

2. Based on the values for peak-peak input voltage V_i that we have just determined, the quiescent emitter voltage is,

$$V_{eq} = V_i/2 = 0.818/2 = 0.409\text{V}$$

To create the quiescent emitter current of 1 mA, we require that

$$R_e = \frac{V_{eq}}{1 \times 10^{-3}} = \frac{0.409}{1 \times 10^{-3}} = 409\Omega$$

The value of V_{BB} is above this by one base-emitter voltage drop:

$$V_{BB} = \frac{V_i}{2} + V_{be} = \frac{0.818}{2} + 0.6 = 1.09\text{V}$$

There are various ways to determine the collector resistance R_c at this point. Using $A_v = R_c/R_e$, we have

$$R_c = A_v R_e = 10 \times 409 = 4.09\text{k}\Omega$$

Just for fun, we can calculate the collector quiescent voltage V_{cq} :

$$V_c = V_{CC} - I_c R_c = 9 - (1 \times 10^{-3})(4.09 \times 10^3) = 4.91\text{V}$$

The complete circuit with its parameters is shown in figure 746.

There are a couple of minor adjustments we should make to the design.

First, it should be noted that we calculated the peak-peak output voltage assuming that the collector voltage could swing downwards to meet the emitter voltage at one point in the cycle. In practice, we would have to maintain a minimum voltage across the transistor somewhat greater than its saturation voltage. The saturation voltage of a small-signal silicon transistor is about 0.1 volt, so one might ensure that the collector-emitter voltage never drops below 0.3 volts or so.

Second, when the collector voltage swings upward, we assumed it could go up to the positive power supply, 9 volts. This requires that the collector current go completely to zero, which is not good practice. Various nonlinearities come into play as the collector current approaches zero: the emitter incremental resistance heads toward infinity and the current gain decreases. Both these effects will tend to distort the signal. Consequently, it is a good practice to keep the collector voltage maximum slightly below the supply voltage.

Both these points are accommodated by designing the circuit as we have shown here, and then restricting the amplitude of the signals to slightly below the absolute possible maximum.

This is still not a complete design. We need to design circuitry to get a single-ended signal into the amplifier and remove the collector DC bias from the output signal.

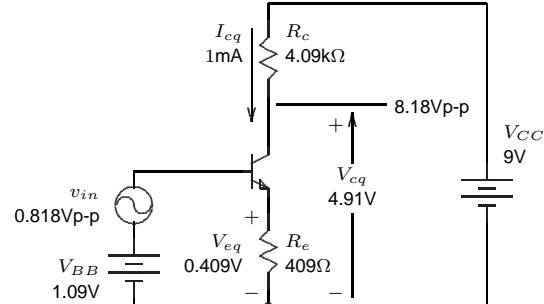


Figure 746: Final Circuit

29.22 Common Emitter Amplifier with Enhancements

In this section, we'll look at various enhancements and modifications of the basic common-emitter amplifier design.

The Input Bias Network

The input bias network must effectively generate the sum of an input AC voltage plus a DC bias voltage as shown in figure 747(a). In practice, it is not convenient to float the input signal or the bias voltage. So some other solution is preferable.

In this section, we'll look at an arrangement that works if the signal is alternating current and the amplifier need only work at frequencies above some lower limit. This is true for audio circuits, for example, which are often designed to operate down to 20Hz. Below that frequency, the amplitude response decreases.

One possible concept is shown in figure 747(b). A coupling capacitor C_{in} serves to prevent the DC base bias supply V_{BB} from being short-circuited via the AC supply v_{in} . The resistor R_b likewise prevents the input voltage V_{in} from being short-circuited via V_{BB} . The DC base current must be supplied by the DC base supply V_{BB} via resistor R_b , so the value of R_b is chosen to be small enough that the DC voltage drop across it is not significant. The AC base current must be supplied via the coupling capacitor C_{in} . The value of the capacitor is chosen so that its impedance is small compared to R_b at the lowest frequency of operation. Put another way, capacitor C_{in} and resistor R_b form a highpass network where the corner frequency is set at or below the lowest frequency of operation.

In summary, the input bias network design equations are:

$$R_b = R_1 \parallel R_2 \quad (1279)$$

$$V_{BB} = V_{CC} \left(\frac{R_2}{R_1 + R_2} \right) \quad (1280)$$

$$f_{min} = \frac{1}{2\pi(R_b \parallel r_{in})C_{in}} \quad (1281)$$

Example

Design a bias network of the type shown in figure 747(c) for the amplifier circuit shown in figure 746 on page 850. Assume $\beta = 100$. The amplifier should work down to a minimum frequency of 20Hz.

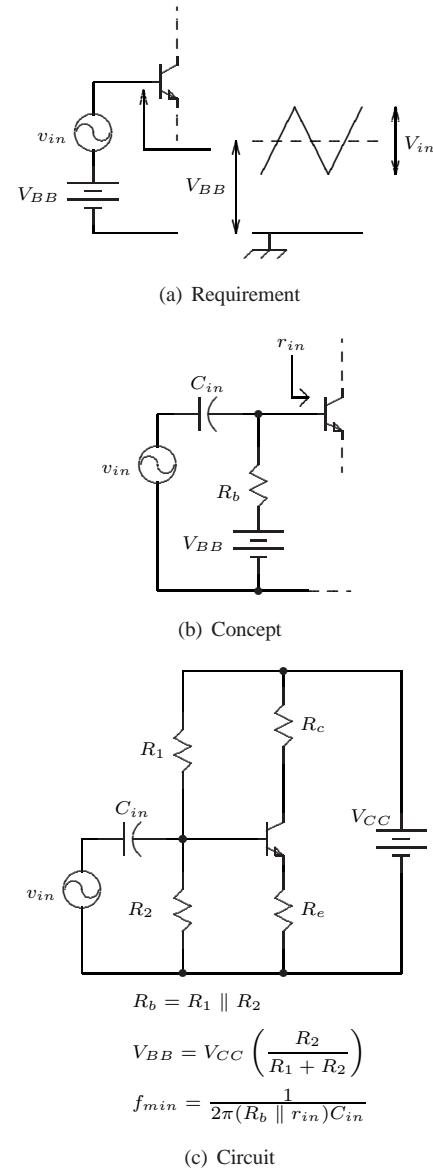


Figure 747: Input Bias Network

Solution

1. **Use the base current to determine R_b** The collector current is 1mA and the value of beta is 100. Consequently, the base current is

$$I_b = \frac{I_c}{\beta} = \frac{1 \times 10^{-3}}{100} = 10\mu\text{A}$$

Somewhat arbitrarily, we might allow a drop of 0.1 volts across the base resistor R_b . Then

$$R_b = \frac{V_{Rb}}{I_b} = \frac{0.1}{10 \times 10^{-6}} = 10k\Omega$$

2. **Use R_b , V_{BB} and V_{CC} to determine R_1 and R_2** This is the straightforward design of a voltage divider based on its open-circuit voltage V_{BB} and its internal resistance R_b . Combining equations 1279 and 1280, we can get an expression for R_1 in terms of the known quantities:

$$R_1 = R_b \left(\frac{V_{CC}}{V_{BB}} \right) = 10 \times 10^3 \left(\frac{9}{1.09} \right) = 82.5k\Omega$$

Now we can rework equation 1280 into a form that solves for R_2 :

$$R_2 = R_1 \left(\frac{V_{BB}}{V_{CC} - V_{BB}} \right) = 82.5 \times 10^3 \left(\frac{1.09}{9 - 1.09} \right) = 11.3k\Omega$$

3. **Determine the input resistance of the amplifier stage** The capacitor sees an input resistance of R_b in parallel with the input resistance of the amplifier, which is beta times the total emitter resistance, $(r_e + R_e)$. First, we'll get the emitter incremental resistance. As usual this is:

$$r_e = \frac{0.026}{I_e} = \frac{0.026}{1 \times 10^{-3}} = 26\Omega$$

Then the total input resistance seen into the base of the transistor is:

$$r_{in} = \beta(r_e + R_e) = 100 \times (26 + 409) = 43.5k\Omega$$

Then the capacitor sees a resistance of R_b in parallel with r_{in} :

$$R_b \parallel r_{in} = 10k\Omega \parallel 43.5k\Omega = 8.13k\Omega$$

Now we can use equation 1281 to determine the capacitor value:

$$C_{in} = \frac{1}{2\pi f_{min}(R_b \parallel r_{in})} = \frac{1}{2 \times 3.14 \times 20 \times 8.13k\Omega} = 0.97\mu\text{F}$$

The final circuit is shown in figure 748.

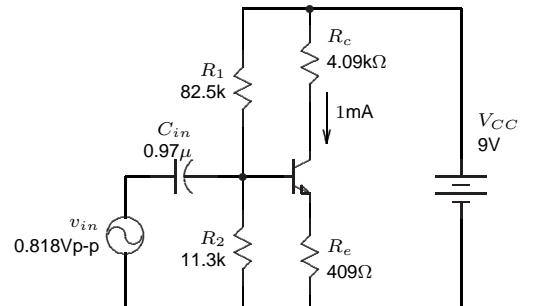


Figure 748: Final Circuit

The Output Coupling Network

Next, we'll look at the calculation of the output capacitor C_o , shown in the abbreviated schematic diagram of figure 749(a). Using a model for the amplifier, we have the equivalent circuit shown in figure 749(b).

We wish to relate the load voltage v_l to the open-circuit output voltage of the amplifier, v_o .

- At low frequencies the capacitor looks like an open circuit, the current through the load resistor will be zero and the load voltage also zero.
- We can redraw the circuit as shown in figure 749(c) without affecting the current through the load resistance, and therefore without affecting the voltage across it. From that figure we can determine the behaviour at high frequencies. At high frequencies the capacitor looks like a short circuit and the output resistance r_o and the load resistance R_l form a voltage divider, in which case:

$$v_l = v_o \left(\frac{R_l}{r_o + R_l} \right) \quad (1282)$$

- The network in figure 749(c) is a highpass RC network where the capacitor is C_o and the resistance $r_o + R_l$. The corner frequency for this network occurs where the capacitive reactance is equal to the resistance. Then the corner frequency is given by

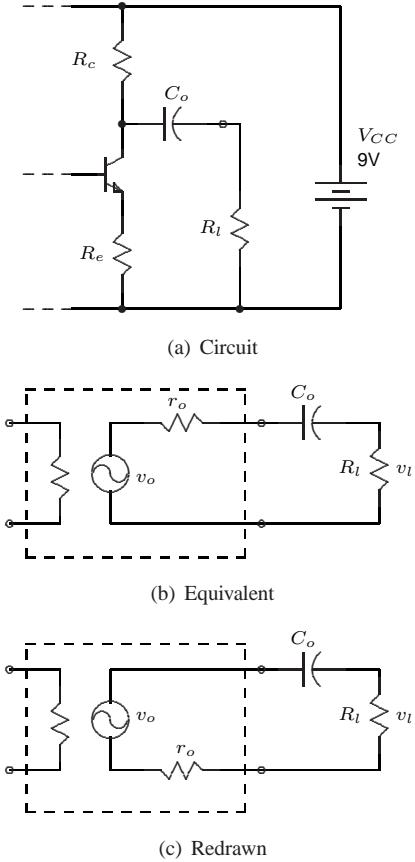


Figure 749: Output Network

$$\begin{aligned} f_c &= \frac{1}{2\pi R C} \\ &= \frac{1}{2\pi(r_o + R_l)C_o} \end{aligned} \quad (1283)$$

If the minimum frequency, output resistance and load resistance are known, then equation 1283 may be used to calculate the value of the output coupling capacitor C_o .

In the common-emitter amplifier, the output resistance r_o is approximately equal to R_c (assuming $R_c \ll r_c$), or more exactly r_o is equal to $R_c \parallel r_c$, where R_c is the external collector resistance and r_c is the incremental collector resistance of the transistor.

Example

Design an output coupling network for the amplifier circuit shown in figure 746 on page 850. Assume the load resistance is $10\text{k}\Omega$. The amplifier should work down to a minimum frequency of 20Hz. The parameters for the transistor are as shown on page 845. The value of incremental collector resistance is $r_c = 100\text{k}\Omega$.

Solution

The value of output resistance r_o is given by:

$$\begin{aligned} r_o &= r_c \parallel R_c \\ &= (100 \times 10^3) \parallel (4.09 \times 10^3) \\ &\approx 4.09 \times 10^3 \end{aligned}$$

Rearranging equation 1283 to solve for output capacitance, we have:

$$\begin{aligned} C_o &= \frac{1}{2\pi(r_o + R_l)f_c} \\ &= \frac{1}{2\pi(4.09 \times 10^3 + 10 \times 10^3)20} \\ &= 0.565\mu\text{F} \end{aligned}$$

Emitter Bypass Capacitor

In the circuits we have been discussing, the gain for AC and DC signals is the same: R_c/R_e . That is, a variation of x volts in v_{in} (at some high frequency) appears multiplied at the output by the same voltage gain as would a variation of x volts in the DC supply V_{BB} . As a consequence, the bias condition and the AC gain are interrelated.

This need not be the case. In figure 750(a), the gain for an AC signal (at a frequency where capacitor C_e appears to be a short circuit) is much larger than the DC gain.

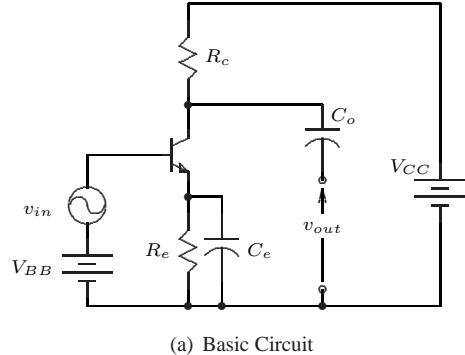
For the DC source, the voltage gain is

$$A_{vdc} = \frac{R_c}{r_e + R_e} \quad (1284)$$

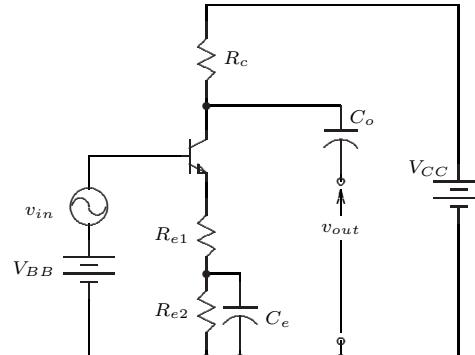
For the AC source, the external emitter resistance R_e is bypassed and the only emitter resistance is the emitter incremental resistance r_e . Then the voltage gain is

$$A_{vac} = \frac{R_c}{r_e} \quad (1285)$$

This might be appealing, but the circuit suffers from the same limitation as the first common emitter circuit we looked at – the input signal must be limited to a few millivolts or the output will be seriously distorted (section 29.18). An alternative circuit is shown in figure 750(b). At low frequencies, the emitter capacitor C_e is an open circuit and the effective emitter resistance is $R_e = R_{e1} + R_{e2}$. At high frequencies, where the capacitor may be considered a short circuit, R_{e2} is eliminated and R_{e1} sets the gain. Consequently, the gain at high frequencies is larger than the gain at DC. If R_{e2} is greater than r_e , the distortion is reduced, or the signal-capability increased for the same level of distortion.



(a) Basic Circuit



(b) Modification

Figure 750: Emitter Bypass Capacitor

Calculating the Emitter Capacitor

Referring to figure 750(a), what should be the size of the emitter capacitor C_e ? It should appear to be a short circuit at the lowest operating frequency of the amplifier. That is, it should appear to be a small reactance compared to some resistance. At first glance, it appears that the resistance in question is R_e , but that's not correct.

To get the true picture, replace the transistor in figure 750(a) with the Tee equivalent circuit for the transistor. The AC equivalent circuit, where we have assumed that the base resistance r_b is a short circuit, is then as shown in figure 751(a) on page 855. As usual, we have obtained the AC equivalent circuit by replacing the V_{BB} and V_{CC} DC sources with short circuits.

The base terminal of the transistor is driven by a voltage source, so the voltage at the base is forced to be v_{in} and is unaffected by the collector current source. (The collector current source does affect the transistor input current, but that's another story.) Then insofar as the voltages across r_e and the rest of the emitter circuit are concerned we can eliminate the collector circuit from consideration. This is shown in figure 751(b).

Finally, if we treat r_e and R_e as a voltage divider, we can Thevenize them into the circuit shown in figure 751(c).

We would like to minimize the voltage across the capacitor so the impedance of the emitter capacitor must be small compared with the internal resistance of the Thevenized source. That is: $X_{Ce} \ll r_e \parallel R_e$. The capacitor and resistor form a lowpass network with a corner frequency given by:

$$f_{max} = \frac{1}{2\pi(r_e \parallel R_e)C_e} \quad (1286)$$

Example

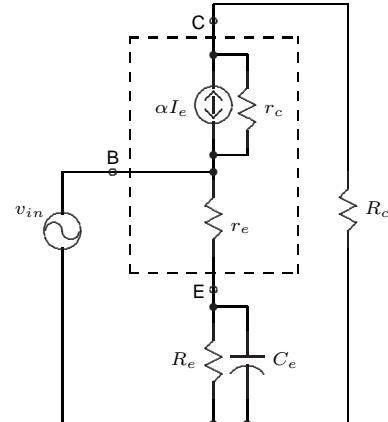
Calculate a suitable emitter resistor bypass capacitor for the amplifier circuit shown in figure 746 on page 850. The capacitor should appear as a low impedance below 20Hz.

Solution

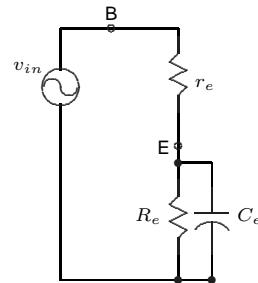
The transistor is operating at an emitter current I_e of 1mA. Then the emitter resistance r_e is 26Ω . The emitter resistance R_e is 409Ω . Rearranging equation 1286 to solve for emitter capacitance C_e , we have:

$$C_e = \frac{1}{2\pi(r_e \parallel R_e)f_{max}} = \frac{1}{2\pi(26 \parallel 409)20} = 325\mu F$$

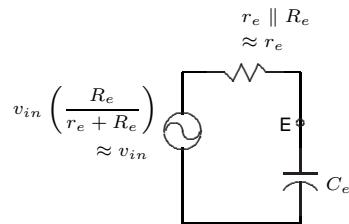
This is a moderately large capacitor, and since it is larger than $1\mu F$, it will probably be a polarized electrolytic.



(a) Equivalent



(b) Simplified



(c) Thevenized

Figure 751: Emitter Capacitor Calculation

Summary

The completed common emitter amplifier with its input bias resistors, coupling and bypass capacitors, is shown in figure 752. In this circuit, the overall voltage gain is approximately

$$\begin{aligned} A_v &= \frac{R_c \parallel R_l}{r_e} \\ &= \frac{4.09k \parallel 10k}{26} \\ &= 111 \text{volts/volt} \end{aligned}$$

As we discussed earlier, the input signal must not exceed a few millivolts or the output signal across the load resistor will be seriously distorted.

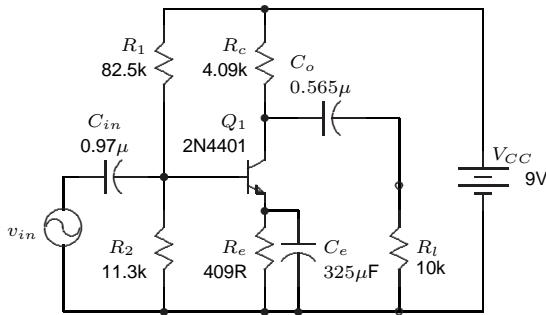


Figure 752: Example Circuit Complete

Dual Supply Configuration

It's a huge advantage in amplifier design to minimize or even eliminate coupling and bypass capacitors. If we split the power supply into two and move the common (ground) point to the junction of the two supplies, this facilitates the design of such an amplifier²³⁶. Here as an introduction we consider a dual-supply configuration with coupling and bypass capacitors. Eventually, we'll move to a completely DC coupled configuration.

The basic configuration is shown in figure 753(a). The bias currents are shown in figure 753(a).

The DC base current must flow through the AC input voltage supply, v_{in} . If that is not acceptable, then some other arrangement must be used.

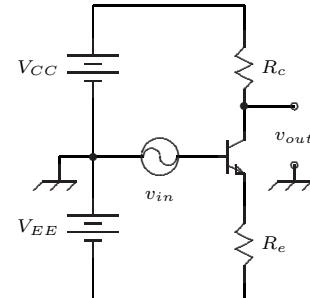
Now consider the AC voltage gain in figure 753(a). Suppose the two supplies are equal in magnitude. We might put the collector voltage at half the positive supply voltage to allow the maximum possible swing at the collector. The current through the two resistors (neglecting base current) is the same since $I_c \approx I_e$. The voltage across the emitter resistor R_e (neglecting the base-emitter voltage drop V_{be}) is equal to the negative supply voltage and also equal to the positive supply voltage. Then the collector resistance has the same current as the emitter resistance but half the voltage across it, so $R_e = 2R_c$. Consequently, this circuit has a voltage gain of

$$A_v = \frac{R_c}{r_e + R_e} \approx \frac{R_c}{R_e} = 0.5 \text{ volts/volt}$$

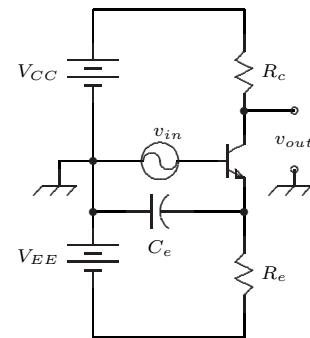
This is not very useful as is. We can obtain a large gain for AC by adding the emitter bypass capacitor C_e shown in figure 753(b). Then, if the capacitor is a short circuit compared to r_e at the frequency of interest, the voltage gain is

$$A_v = \frac{R_c}{r_e} \text{ volts/volt} \quad (1287)$$

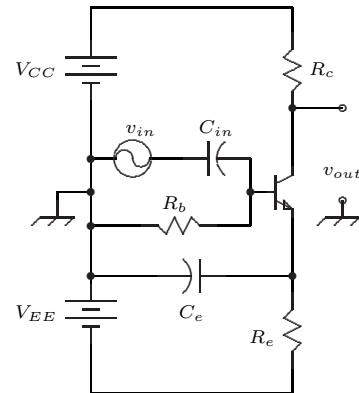
Back to the issue of bias currents: to keep DC base current from flowing through the AC input supply v_{in} , we can modify the circuit as shown in figure 753(c). The input current flows through a capacitor C_{in} , which blocks DC current. The base current is provided via the base resistor R_b . As we showed previously, R_b would be chosen small enough that the base current does not create a significant voltage drop across it, but as large as possible to reduce the size of the input capacitor.



(a) Circuit



(b) Equivalent



(c) Redrawn

Figure 753: Common Emitter Amplifier, Dual Supply

²³⁶The first transistor audio amplifiers were the solid-state version of older tube circuits and used a single power supply. Eventually these designs evolved into the split-supply, DC coupled designs which are used today.

29.23 The Common Emitter Amplifier, Base Current Drive

In previous sections, we showed the amplification of a voltage signal by a common emitter stage. When the AC input signal was connected across the base-emitter junction directly (that is, no emitter resistance R_e), the maximum value of the input signal was limited to a few millivolts. More than that would cause significant distortion in the output signal because of the non-linear V-I characteristic of the base-emitter junction (section 29.18).

When an external emitter resistor R_e is added in series with the emitter, the voltage across the base-emitter diode becomes a small fraction of the input voltage and most of the input voltage appears across R_e . Since R_e is a linear device, this lowers the distortion of the amplifier and allows a much larger input signal.

Now, the signal voltage across the emitter resistor R_e is in antiphase with the voltage across the collector resistor R_c . For large signals, these two voltages tend to collide at one point in the cycle (see figure 744 on page 848), and this limits the maximum voltage across the collector resistance. So the emitter resistance is a nuisance in that respect. Is there another way to linearize the amplifier without an emitter resistance?

In the preceding circuits, the base circuit consisted of a DC source V_{BB} in series with an AC source v_{in} . (See for example figure 747 on page 851). An alternative base circuit is a DC current source I_{bb} in parallel with an AC current source i_{in} , figure 754(a). In this case, the DC base current source I_{bb} sets the collector current at βI_{bb} . Then the input signal current causes the collector current to vary. The collector resistance R_c converts this variation in collector current into an AC output voltage.

(As it is, this circuit will work for an individual transistor but will not work in a production situation because beta varies over a 3:1 range in different units of the same transistor model. We'll deal with that problem later.)

Since the collector current is beta times base current, then the amplifier will have very low distortion.

A practical version of this circuit is shown in figure 754(b). The bias current source I_{bb} is replaced by the supply voltage V_{CC} driving current through R_{bb} . The signal current source i_{in} is replaced by v_{in} driving current through R_b . Each of these networks will function as a current source if the load resistance is much less than the internal resistance. In other words, the resistance looking into the transistor base (βr_e) is required to be much less than either R_{bb} or R_b .

The AC equivalent circuit is shown in figure 754(c) where, as usual, the DC voltage sources have been replaced by short circuits. Since we've already required that R_{bb} be much larger than r_{in} , resistor R_{bb} has little effect and can be removed.

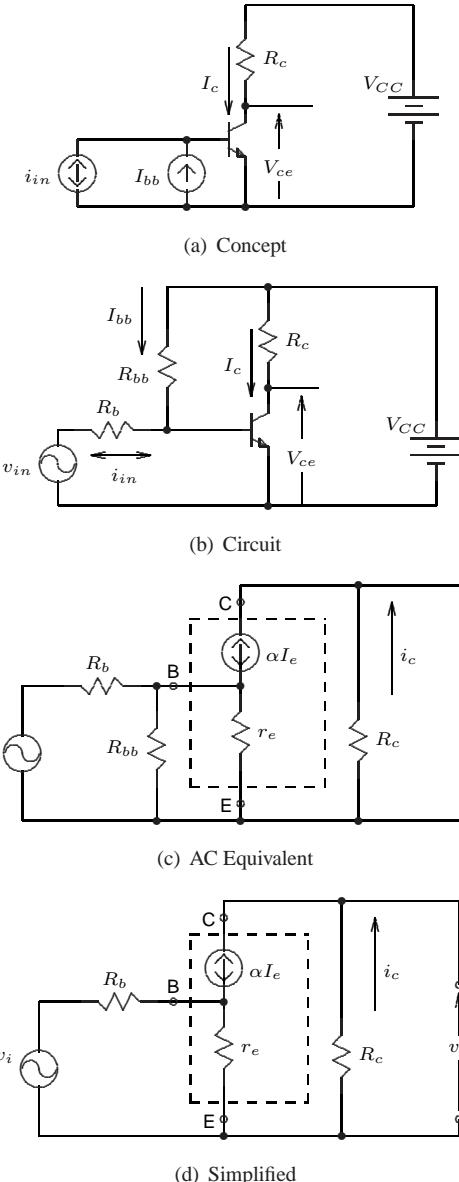


Figure 754: Common Emitter Amplifier, Base Current Drive

Now we can determine the voltage gain of this amplifier stage. The AC base current is given by:

$$i_b = \frac{v_i}{R_b + r_i} \quad (1288)$$

where r_i is the input resistance looking into the base of the transistor, given by:

$$r_i = \beta r_e \quad (1289)$$

The output voltage v_o is given by

$$v_o = i_c R_c \quad (1290)$$

Substitute βi_b for i_c in equation 1290, and then substitute from equation 1288 for i_b and from equation 1289 for r_i :

$$v_o = \beta \frac{v_i}{R_b + \beta r_e} R_c \quad (1291)$$

The voltage gain A_v is the ratio of output to input voltages, so equation 1291 can be manipulated to give:

$$\begin{aligned} A_v &= \frac{v_o}{v_i} \\ &= \frac{\beta R_c}{R_b + \beta r_e} \\ &= \frac{R_c}{\frac{R_b}{\beta} + r_e} \end{aligned} \quad (1292)$$

Compare this to the expression for voltage gain in a common-emitter amplifier with an emitter resistor R_e :

$$A_v = \frac{R_c}{R_e + r_e} \quad (1293)$$

Comparing equations 1292 and 1293, we see that the source resistance, divided by beta, has the same effect on the gain as the emitter resistance. In section 29.10 we saw that moving a resistance from the emitter to the base circuit is equivalent to multiplying it by beta (or beta+1, which is much the same thing). In this section, we see that the effect of the base resistance can be determined by moving it into the emitter and dividing its value by beta.

Conclusions

- For low distortion the value of R_s/β must be much larger than the emitter incremental resistance r_e , which is non-linear.
- If $R_s/\beta \gg r_e$, there is a direct (linear) relationship between output and input voltage, which implies that this is a low-distortion amplifier.
- The voltage gain is proportional to beta, so there will be a large production variation in voltage gain.
- The collector-emitter voltage can swing between the supply voltage V_{CC} and to within V_{cesat} (usually a few tenths of a volt) of ground.

The circuit does lower distortion and permit a larger output swing, but it also has some limitations.

Note: Is $R_{bb} \gg r_i$?

We assumed that resistor R_{bb} is much larger than the input resistance $r_i = \beta r_e$. Let's see if that is true. We'll assume that v_{in} and R_s are disconnected for this calculation.

The base bias resistor R_{bb} is given by:

$$\begin{aligned} R_{bb} &= \frac{V_{CC} - V_{be}}{I_b} \\ &\approx \frac{V_{CC}}{I_b} \end{aligned} \quad (1294)$$

The incremental input resistance r_i is given by:

$$\begin{aligned} r_i &= (\beta + 1)r_e \\ &= (\beta + 1)\frac{V_t}{I_e} \\ &= (\beta + 1)\frac{V_t}{(\beta + 1)I_b} \\ &= \frac{V_t}{I_b} \end{aligned} \quad (1295)$$

Then the ratio of R_{bb} to r_i is:

$$\begin{aligned} \frac{R_{bb}}{r_i} &= \frac{V_{CC}}{I_b} \div \frac{V_t}{I_b} \\ &= \frac{V_{CC}}{V_t} \end{aligned} \quad (1296)$$

V_t is about 26mV at room temperature and V_{CC} is in the order of volts. So this ratio is much larger than 1. For example, for a supply voltage of 9 volts,

$$\begin{aligned} R_{bb} &= \frac{9}{0.026} r_i \\ &= 346 r_i \end{aligned}$$

In general then, it will be true that R_{bb} is much larger than the input resistance r_i .

29.24 Common-Emitter Amplifier with Collector-Base Feedback

One possible method of biasing the common-emitter amplifier is by means of a base resistor (which we will call R_b in this case) to the collector voltage supply, as shown in figure 755(a). Because beta can be expected to vary over a range of 3:1 in a production sample of transistors, the collector bias voltage can be expected to vary over a wide range.

The circuit must be designed in such a way that the transistor does not saturate or cutoff at any point in the output cycle. If the collector bias voltage varies over a wide range then the output signal must be restricted in amplitude to ensure that it doesn't clip at some point in the cycle.

An alternative bias scheme is shown in figure 755(b). In this arrangement, the bias voltage is derived from the collector of the transistor rather than the positive supply. The effect is to provide some negative feedback action to reduce the change in collector voltage for a given change in beta.

For example, suppose that the transistor is biased so that a transistor of average beta²³⁷ causes the collector voltage to be half the supply voltage. Now suppose the beta decreases. For a given base current, this will cause the collector current to decrease. The collector voltage then rises, which drives more current into the base, which (partially) restores the collector voltage to its original value.

A complete amplifier circuit is shown in figure 755(c). This circuit may be regarded as an inverting op-amp circuit where the input resistor is R_s and the feedback resistor is R_b . The 'op-amp' consists of the transistor and its collector resistor R_c . First, we'll examine the bias and then take a look at the AC gain.

Now let's do a brief analysis to see how much of an improvement we obtain with this circuit. Throughout this section, we'll assume that the base-emitter voltage V_{be} is small enough to be ignored. This is a pretty gross assumption, but it simplifies the math and doesn't affect the spirit of the result.

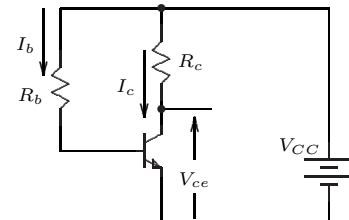
Direct Bias, Analysis

First, consider the direct bias circuit of figure 755(a). Writing the bias equations for the circuit, we have:

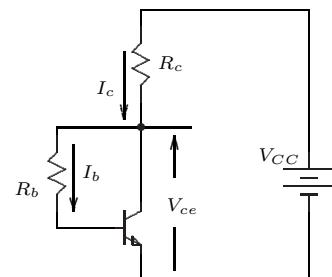
$$V_{ce} = V_{CC} - I_c R_c \quad (1297)$$

$$I_c = \beta I_b \quad (1298)$$

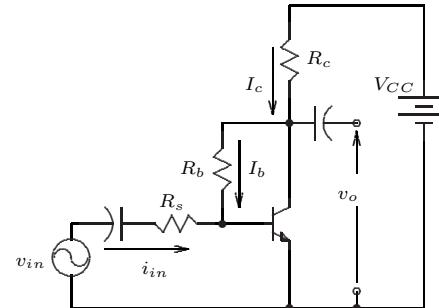
$$I_b = \frac{V_{CC} - V_{be}}{R_b}$$



(a) Direct Bias



(b) Feedback Bias



(c) Complete Amplifier

Figure 755: Collector-Base Feedback Bias

²³⁷Since beta varies in a roughly exponential fashion, the best way to calculate the average is the *geometric mean*: the square root of the product of the maximum and minimum values.

$$\approx \frac{V_{CC}}{R_b} \quad (1299)$$

Combining equations 1297 through 1299 and rearranging, we can get this into the form:

$$\frac{V_{ce}}{V_{CC}} = 1 - \beta \frac{R_c}{R_b} \quad (1300)$$

Let's define the geometric mean of the beta values as

$$\beta_o = \sqrt{\beta_{min} \times \beta_{max}} \quad (1301)$$

We'll design this circuit so that the mean value of beta β_o sets the collector-emitter voltage V_{ce} mid-way between V_{CC} and ground. This isn't essential, but (ignoring the saturation voltage of the transistor, which is usually negligible) it allows the collector-emitter voltage the maximum possible swing. Then:

$$\frac{V_{ce}}{V_{CC}} = 0.5 \quad (1302)$$

Substituting from equation 1302 into equation 1299:

$$0.5 = 1 - \beta_o \frac{R_c}{R_b} \quad (1303)$$

Rearranging equation 1303 to get an expression for R_c/R_b , we have:

$$\frac{R_c}{R_b} = \frac{1}{2\beta_o} \quad (1304)$$

This is a useful relationship. The value of collector resistance R_c is known or can be calculated using supply voltage V_{CC} , collector current I_c and equation 1297. Then the value of base bias resistor R_b can be calculated from equation 1304.

Now let's see how much the output voltage changes when beta changes. Substitute for R_c/R_b from equation 1303 into equation 1300:

$$\frac{V_{ce}}{V_{CC}} = 1 - \frac{\beta}{2\beta_o} \quad (1305)$$

This gives us an expression for the collector-emitter voltage, as a fraction of the supply voltage, for different values of beta. Notice that the resistor values are completely immaterial: the expression is solely a function of the variation of beta.

Example

Suppose we have a transistor for which the beta varies between 50 and 150. Then from equation 1301 the mean beta is:

$$\begin{aligned} \beta_o &= \sqrt{\beta_{min} \times \beta_{max}} \\ &= \sqrt{50 \times 150} \\ &= 87 \end{aligned} \quad (1306)$$

Using equation 1305 the value of V_{ce} for a maximum value of beta is:

$$\begin{aligned} \frac{V_{ce}}{V_{CC}} &= 1 - \frac{\beta}{2\beta_o} \\ &= 1 - \frac{150}{2 \times 87} \\ &= 0.137 \end{aligned} \quad (1307)$$

Similarly, the value of V_{ce} for a minimum value of beta is:

$$\begin{aligned}\frac{V_{ce}}{V_{CC}} &= 1 - \frac{50}{2 \times 87} \\ &= 0.713\end{aligned}\quad (1308)$$

For a value of supply voltage V_{CC} of 9 volts, these would correspond to 1.23 volts and 6.41 volts. At the low end, by the time you allow for a few tenths of a volt for saturation, the allowable peak value of an output AC voltage is limited to something in the order of a volt or so, even though the supply voltage is 9 volts.

Feedback Bias, Analysis

This time, the circuit is shown in figure 755(b). Writing the bias equations we have:

$$\begin{aligned}V_{ce} &= V_{CC} - (I_c + I_b)R_c \\ &\approx V_{CC} - I_c R_c\end{aligned}\quad (1309)$$

$$I_c = \beta I_b \quad (1310)$$

$$\begin{aligned}I_b &= \frac{V_{ce} - V_{be}}{R_b} \\ &\approx \frac{V_{ce}}{R_b}\end{aligned}\quad (1311)$$

Combining equations 1309 through 1311 and rearranging, we can get this into the form:

$$\frac{V_{ce}}{V_{CC}} = \frac{1}{1 + \beta \frac{R_c}{R_b}} \quad (1312)$$

As previously, put V_{ce}/V_{CC} equal to 0.5 and $\beta = \beta_o$. Then we get

$$\frac{R_c}{R_b} = \frac{1}{\beta_o} \quad (1313)$$

Substitute for R_c/R_b from equation 1313 into equation 1312 and we're there:

$$\frac{V_{ce}}{V_{CC}} = \frac{1}{1 + \frac{\beta}{\beta_o}} \quad (1314)$$

This is the equation that characterises the collector-emitter voltage in terms of beta for the amplifier with collector-base feedback. For example, for the same transistor as described above, $\beta_o = 87$, maximum beta is 150 and minimum beta is 50. Plug those numbers into equation 1314, and we get:

$$\begin{aligned}\frac{V_{ce}}{V_{CC}} &= 0.36 \text{ minimum} \\ \frac{V_{ce}}{V_{CC}} &= 0.635 \text{ maximum}\end{aligned}$$

For a 9 volt power supply, these correspond to:

$$\begin{aligned}V_{ce} &= 3.24 \text{ V, minimum} \\V_{ce} &= 5.70 \text{ V, maximum}\end{aligned}$$

Comparing the collector-emitter voltages for direct bias and feedback bias cases, with a 9 volt V_{CC} power supply we have:

	Direct Bias	Feedback Bias
V_{ce} Minimum	1.23V	3.24V
V_{ce} Maximum	6.41V	5.70V
ΔV_{ce}	5.18V	2.46V

As shown in figure 756, with feedback the variation of the collector voltage is about half that of the direct bias case. These results must be taken with a grain of salt, because the effect of the base-emitter voltage V_{be} was neglected and in fact V_{be} does impact these figures.

Voltage Gain

We can regard the amplifier of figure 755(c) as an inverting op-amp where resistor R_b acts as the feedback resistor R_f and the source resistor acts as the input resistor R_i . Providing that the gain of the amplifier is much larger than the closed-loop gain, by analogy with the op-amp case (section 12.5 on page 320) we should expect:

$$G = -\frac{R_b}{R_s} \quad (1315)$$

For this to be true, the open-loop gain A must be much larger than the closed loop gain G . The open-loop gain is the voltage gain of a common emitter stage:

$$A = -\frac{R_c}{r_e} \quad (1316)$$

If the collector-emitter voltage is set to half the supply voltage, then we can get an expression for the collector resistance R_c :

$$R_c = \frac{V_{CC}/2}{I_c} \quad (1317)$$

The expression for incremental emitter resistance r_e is:

$$\begin{aligned}r_e &= \frac{V_t}{I_e} \\&\approx \frac{V_t}{I_c}\end{aligned} \quad (1318)$$

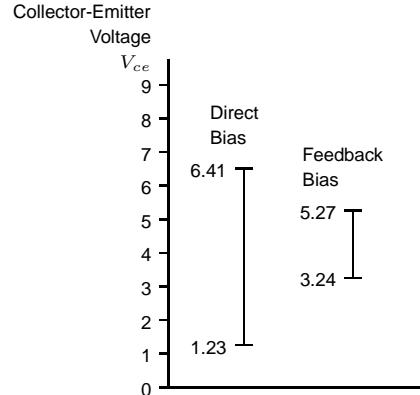


Figure 756: Collector Voltage Variation

Substitute from equations 1317 and 1318 into our equation 1316 for voltage gain:

$$\begin{aligned} A &= -\frac{\frac{V_{CC}/2}{I_c}}{\frac{V_t}{I_c}} \\ &= -\frac{V_{CC}}{2V_t} \end{aligned} \quad (1319)$$

This is useful because it gives the open-loop voltage gain in terms of the supply voltage. If the transistor is operated at room temperature then V_t is 26mV. Suppose the power supply V_{CC} is 9 volts. Then the open-loop voltage gain is:

$$\begin{aligned} A &= -\frac{V_{CC}}{2V_t} \\ &= -\frac{9}{2 \times 0.026} \\ &= -173 \text{ volts/volt} \end{aligned}$$

Consequently, for the closed-loop gain to be much less than the open-loop gain, we might restrict the closed-loop gain to something less than one tenth of this, or about -17.3 volts/volt.

The value of feedback resistor R_b is fixed by the bias requirements. Then the voltage gain can be set by the source resistance R_s .

Conclusions

Suppose that the power supply voltage is 9 volts. Then we have determined that the feedback-bias circuit can be expected to keep the collector supply voltage between 3.2 and 5.7 volts, over a range of beta between 50 and 150. Those collector voltages determine the maximum possible output voltage swing.

To reliably fix the voltage gain at a value determined by the ratio of R_b and R_s , we have also determined that the maximum possible voltage gain is in the order of 17 volts/volt.

There are ways to have the bias stability afforded by the feedback resistance and have a larger voltage gain. See the problem section for an example.

29.25 A Catalogue of Common Emitter Configurations

This section summarizes the circuits we have been discussing and shows some evolutionary relationships between them. This is not an exhaustive collection, but it indicates some of the key concepts for this amplifier.

Base Bias Current and Signal Current

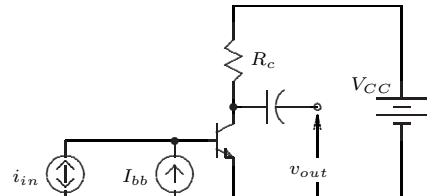
We start with figure 757(a), in which the transistor is biased by a DC current source I_{bb} into the base terminal so that its collector-emitter voltage is somewhere between the supply voltage V_{CC} and ground. The AC signal current i_{in} adds to and subtracts from this DC base bias current, which causes the collector current to vary in a similar fashion but increased in amplitude by the beta of the transistor. The collector resistor R_c turns this variation in collector current into an AC output signal. The output capacitor removes the DC voltage at the collector of the transistor. The AC component of the time-varying collector voltage is the output voltage v_{out} .

Fixed Bias, Practical Circuit

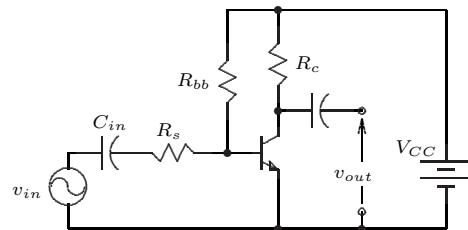
Figure 757(b) shows a (somewhat) practical version of the circuit in figure 757(a). The base DC bias current is provided by V_{CC} driving current through the resistor R_{bb} . The AC signal current is provided by v_{in} driving current through R_s and the input resistance of the transistor, βr_e . Capacitor C_{in} prevents the DC base-emitter voltage from driving a DC current through the AC supply.

Feedback Bias

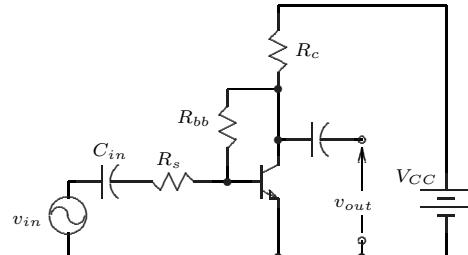
The fixed bias scheme of figure 757(b) does not cope well with the various values of beta that are found in any production transistor. The collector voltage will vary substantially with different values of beta. As described in section 29.24, a more stable arrangement returns the base bias resistor to the collector of the transistor rather than the positive power supply. The circuit is shown in figure 757(c). If the collector bias voltage is centred between V_{CC} and ground, this arrangement allows an output voltage swing that extends from V_{CC} to ground. However, the voltage gain of this circuit is limited unless it is further modified.



(a) Base Current Drive



(b) Fixed Bias, Practical Circuit



(c) Collector-Base Feedback

Figure 757: Common Emitter Circuits

Base Bias Moves to the Emitter

Now let us return to our starting point, figure 757(a), and consider a different evolutionary path. As shown in figure 758(a), the base current source I_{bb} can be relocated into the emitter of the transistor as I_{ee} if it is increased by a factor of $\beta + 1$. This then creates a standing (bias) current in the transistor. The emitter current sink I_{ee} must be bypassed by a capacitor to provide a path for emitter AC current. The base signal current source i_{in} can become voltage source v_{in} , where $v_{in} = i_{in}r_e$.

Emitter Current Sink, Negative Supply

One way of implementing the emitter current sink I_{ee} is shown in figure 758(b). The emitter resistor R_e is returned to a negative power supply V_{EE} so that $I_{ee} \approx V_{EE}/R_e$. In figure 758(a), the base current must flow through the AC input voltage. In figure 758(b) the input capacitor C_{in} blocks base current from flowing through the AC input voltage and it flows through the base resistor R_b instead.

The base resistor R_b is usually chosen so that base current through it does not set up a substantial voltage across it. Then the base of the transistor is approximately at ground potential for DC, and the emitter is one V_{be} drop below ground.

The current sink can also be implemented with a current mirror as shown in figure 777 for the emitter follower amplifier.

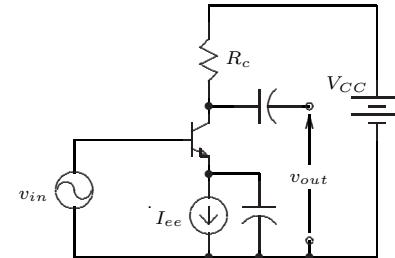
Emitter Current Sink, Single Supply

If a negative voltage is not available, it is possible to return the emitter resistor R_e to ground and raise the base voltage by the amount needed to drive DC current I_{ee} through it. This arrangement is shown in figure 758(c).

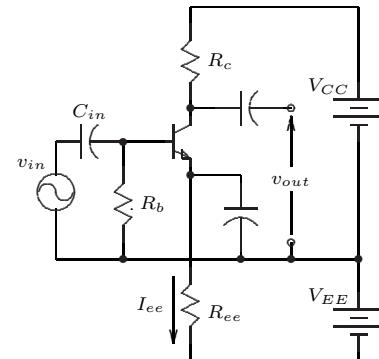
In this case, the voltage to drive the emitter resistor must be taken away from the voltage across the transistor, so the output voltage swing is reduced from the circuit of figure 758(b).

Rather than a negative power supply, the voltage across R_e is now provided by the voltage divider R_1 and R_2 . The value of R_b in figure 758(b) is equal to $R_1 \parallel R_2$.

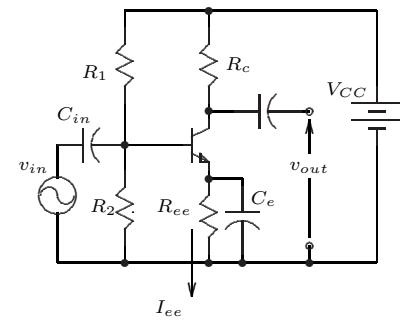
If R_e is replaced by a transistor current sink (see figure 777), then the voltage at the emitter of the amplifier transistor need only be a few tenths of a volt above ground. Then most of the supply voltage can appear across the collector resistor and the transistor, and very little is needed across the emitter current sink. This allows a much larger output signal without clipping.



(a) Emitter Current Bias



(b) Emitter Current Sink, Negative Supply



(c) Emitter Current Sink, Single Supply

Figure 758: Common Emitter Circuits,
cont'd

Distortion and Signal Handling

As we saw in section 29.18, the AC voltage across the base-emitter junction must be limited to a few millivolts. The circuits of figure 757 have a certain amount of base resistance R_s . This source resistance appears in the emitter circuit reduced by beta. So for the same level of distortion the input signal can be increased by an amount approximately proportional to $R_s/\beta r_e$. (The base bias resistor R_b appears in parallel with R_s in calculating the effective input resistance, and may need to be taken into account.) The value of R_s also has an effect on the voltage gain, which must be taken into account.

The circuits of figure 758 have no AC emitter or source resistance, and so have large voltage gain but are very limited in input-signal capability. In all cases, the AC input voltage appears entirely across the emitter incremental resistance r_e . To reduce distortion, we have to add resistance to that circuit loop, either with source resistance or unbypassed emitter resistance.

The Unbypassed Emitter

The circuit of figure 758(c) is modified very simply by removing the emitter resistor bypass capacitor C_e , with the result shown in the circuit of figure 759(a). This does not affect the DC bias conditions of the circuit but it does affect the AC behaviour. The voltage gain goes from R_c/r_e to $R_c/(R_e + r_e)$, and the input resistance increases from βr_e to $\beta(R_e + r_e)$. The DC bias condition and the AC gain are linked and cannot be set independently of each other.

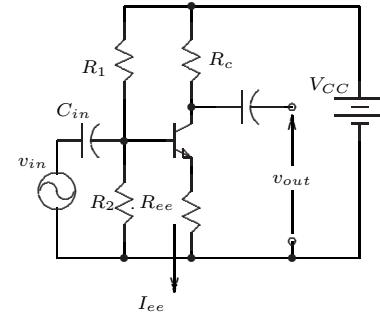
The Partially Bypassed Emitter, Dual Supply

Another approach is to start with the figure 758(a) and put a resistor R_e in series with the emitter bypass capacitor. The circuit is shown in figure 759(b). Emitter resistor R_e appears in series with emitter resistance r_e . Consequently, the input voltage is distributed over $R_e + r_e$ and it can be larger for the same voltage across r_e , and the same level of distortion.

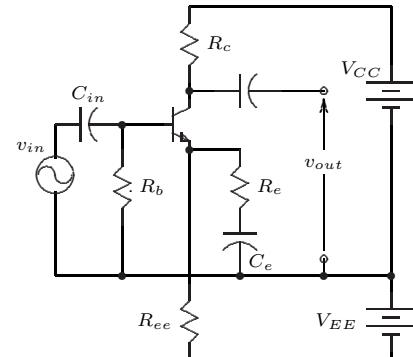
The DC bias point is still primarily established by V_{EE} and R_{ee} , but the AC voltage gain changes from R_c/r_e to $R_c/(R_{ee} \parallel R_e)$.

The Partially Bypassed Emitter, Single Supply

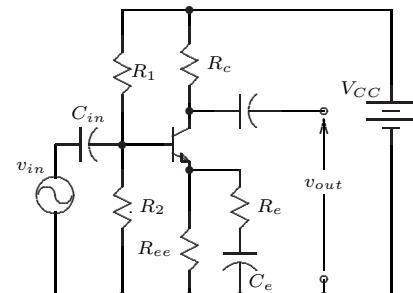
The single-supply version of figure 759(b) is shown in figure 759(c). Again, the voltage gain is $R_c/(R_{ee} \parallel R_e)$. Because there is likely to be much less voltage across R_{ee} in this circuit, the resistance will be less for the same bias current, and R_{ee} and R_e are more likely to be comparable in value.



(a) Unbypassed Emitter



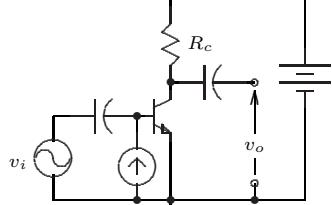
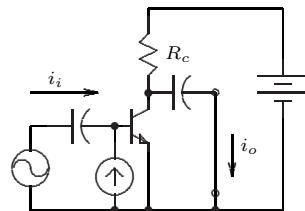
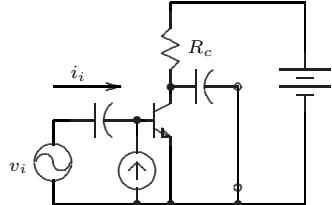
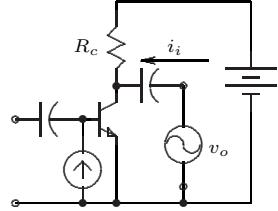
(b) Partially Bypassed Emitter, Dual Supply



(c) Partially Bypassed Emitter, Single Supply

Figure 759: Common Emitter Circuits, cont'd

29.26 Summary: Common-Emitter Equations

Voltage Gain A_v		$\frac{v_o}{v_i} = \frac{R_c}{r_e}$
Current Gain A_i		$\frac{i_o}{i_i} = \beta$
Input Resistance r_i		$\frac{v_i}{i_i} = (\beta + 1)r_e$
Output Resistance r_o		$\frac{v_o}{i_o} = R_c$

Notes

1. If there is an unbypassed emitter resistor R_e between the emitter terminal and ground, replace r_e with $r_e + R_e$.
2. If the source has internal resistance R_s then this acts as an emitter resistance $R_e = R_s/\beta$, which affects the voltage gain and the input resistance.
3. If the collector resistor R_c is more than 1/10 of the collector incremental resistance, replace R_c with $R_c \parallel r_c$.
4. All the capacitors are assumed to be short circuits at the measurement frequency.

29.27 The Common Collector (Emitter Follower) Amplifier

Overview

The transistor amplifier can assume 3 possible configurations: common emitter, common collector, and common base. In this section, we'll study the common collector configuration and its applications.

The common-collector stage is almost universally referred to as an *emitter-follower* stage because the output (the emitter voltage) follows the input voltage at an offset of one base-emitter voltage drop. Consequently, its voltage gain is approximately unity.

The emitter follower is generally used as a buffer amplifier between a source and load resistance. It may be regarded as a *resistance transformation device*. To the source, it makes the load resistance appear beta times as large. To the load, it makes the Thevenin source resistance appear beta times as small. A small current at the input controls a much larger current at the output, and the current gain can be as large as beta.

Basic Function

The basic concept is illustrated in figure 760. The circuit appears similar to the common emitter amplifier, but the behaviour is quite different. The load resistance has moved from the collector to the emitter.

An input voltage V_{in} controls the base terminal of the transistor, and the load resistance R_l receives the emitter current.

The source device supplies a current I_b which appears in the load as emitter current, that is, multiplied by $\beta + 1$. So the current gain of this circuit would be expected to be equal to something like $\beta + 1$.

The output voltage is equal to the input voltage minus the base-emitter voltage drop V_{be} of the transistor.

$$V_l = V_{in} - V_{be} \quad (1320)$$

If the input voltage changes by, say, $+\Delta V_{in}$, then the output changes by the same amount: to $V_l + \Delta V_{in}$. Because the output voltage follows the input voltage (with an offset of V_{be} volts, as shown in figure 760), this implies that the emitter follower amplifier has a gain of around 1 volt/volt.

Now suppose that the input voltage is held fixed and the load resistance decreases. This will initially cause an increase in voltage across the base-emitter junction, which will cause a very large increase in emitter current. In the common emitter amplifier, there is no feedback mechanism of this sort – the collector current is a fixed function of the base or emitter current and does not change with a change in load resistance. Consequently, we can expect that the output resistance of the emitter follower will be lower than the common emitter stage.

The magnitude of the input signal is limited by the following:

- The collector supply voltage V_{CC} must be larger than the maximum input voltage by an amount that will keep the collector-base diode reverse biased.
- The base voltage must be positive with respect to the emitter. In figure 760 that precludes the input voltage going negative.

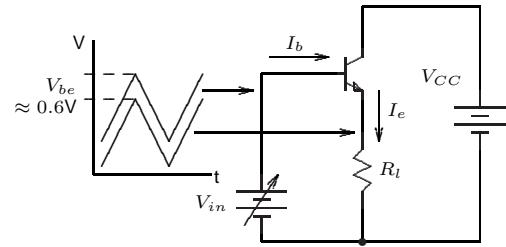


Figure 760: Common Collector Concept

Device Bias Current

Consider the emitter follower circuit shown in figure 761. We'll assume that the input voltage is fixed. The output current I_o can change direction. The emitter bias current source I_{ee} is sometimes referred to as the *tail current* in the circuit.

When the output current is down the page, the emitter current in the transistor is the sum of the tail current and the output current:

$$I_e = I_{ee} + I_o \quad (1321)$$

That is, the transistor increases its current by an amount equal to the output current. There is no hard limit to the amount of current that the transistor can source in this manner.

When the output current reverses (so that it's up the page), the emitter current in the transistor is the difference between the tail current and the output current:

$$I_e = I_{ee} - I_o \quad (1322)$$

That is, the transistor *reduces* its current by the amount equal to the output current.

Current cannot flow into the emitter of the transistor, so the emitter current cannot become negative and, by equation 1322, the *tail current I_{ee} must exceed the output current*. Put another way, in a circuit of the type shown in figure 761 the transistor can source large amounts of current but the current it can sink is limited by the tail current.

Figure 761 shows an NPN follower. For a PNP follower, this rule would be reversed: it can sink large amounts of current but the source current is limited by the tail current.

It is possible to combine the ability of the NPN follower to supply current and the PNP follower to absorb it. Figure 762 shows an example. (Near relatives of this circuit are shown as output stage buffers in section 13.14, page 356). Notice that this configuration has no bias current at all, so the transistor current is entirely equal to the load current. As the current switches from sourcing in the upper transistor to sinking in the lower one, the current in each transistor goes to zero. The result is *crossover distortion* in the output waveform.

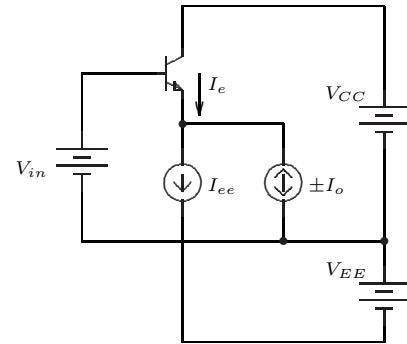


Figure 761: Bias Current

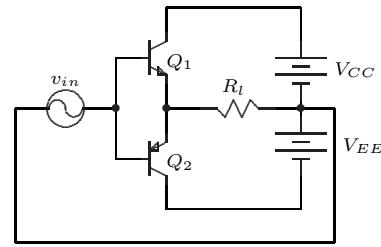


Figure 762: Bidirectional Emitter Follower

Emitter Follower Configurations

A series of emitter follower AC amplifier circuit configurations are shown in figures 763, 764 and 765.

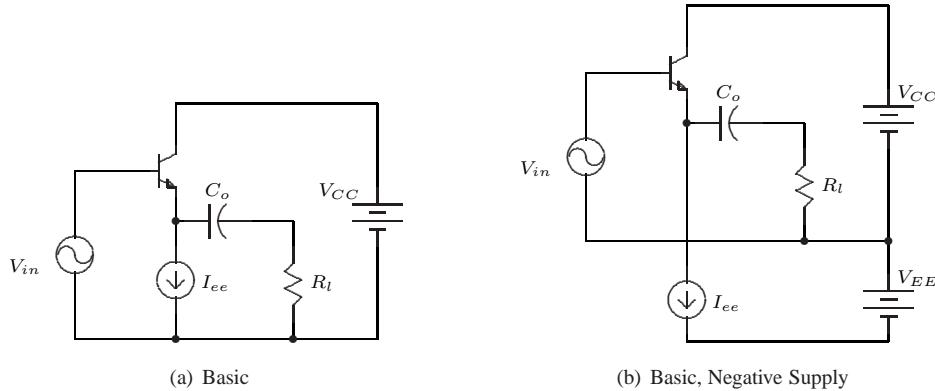


Figure 763: Basic Emitter Follower

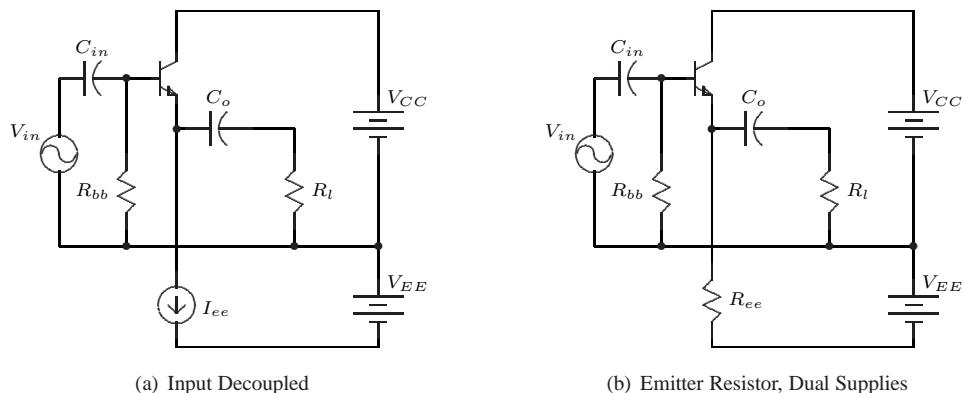


Figure 764: Dual Supply Emitter Follower

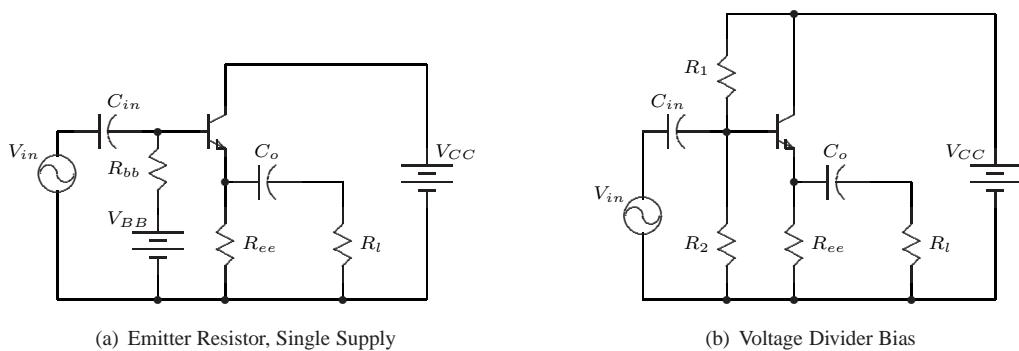


Figure 765: Single Supply Emitter Follower

A basic emitter follower AC coupled amplifier is shown in figure 763(a). The load resistor R_l is coupled to the output by means of a capacitor C_o . The capacitor passes an AC signal to the load but blocks any DC bias at the emitter of the transistor.

This is not a practical circuit because the input voltage swings negative which will put the emitter of the transistor at a negative potential at certain points in the cycle. In practice, the current sink I_{ee} must be biased so that its top end is positive with respect to its bottom end. A more practical circuit is shown in figure 763(b), in which the bottom end of the current sink is connected to negative supply V_{EE} . Then the current source will be correctly biased as long as the emitter of the transistor swings negative by no more than V_{EE} volts. In both these circuits, the tail bias current I_{ee} must be greater than the peak current in the load resistance.

In figure 763, the DC base current flows through the AC source v_{in} . If this is not desirable, then the input circuit can be modified as shown in figure 764(a). (This same scheme was used in the common emitter amplifier, section 29.22 on page 851, and the design considerations are given there.)

The tail current source can be replaced with a resistor as shown in figure 764(b).

When the amplifier must be operated from a single supply, the bias voltage for the emitter resistor may be provided by a DC voltage source in the base circuit, figure 765(a). A practical implementation of that concept is shown in figure 765(b), where the supply V_{BB} and resistance R_b are provided by the R_1, R_2 voltage divider. Again, the design equations are identical to the common emitter case, section 29.22 on page 851.

Modelling the Emitter Follower

In this section, we'll examine the common-collector BJT stage as an amplifier for AC signals.

Our approach to the common emitter amplifier (and common base amplifier in a future section) is determine the voltage gain, input resistance, output resistance and current gain of the amplifier. Then we can attach a source with some internal resistance, and a load with some resistance, and determine the voltages and currents in the circuit.

This approach does not work well with the emitter follower. The characterization of an amplifier as in section 13.24 assumes that the load resistance has no effect on the input resistance, and the source resistance has no effect on the output resistance. That is not true of the emitter follower: there are additional interactions.

The main purpose of an emitter follower is to act as a buffer – making a low impedance load present a higher impedance to the source, and making a high impedance source present a low impedance to the load.

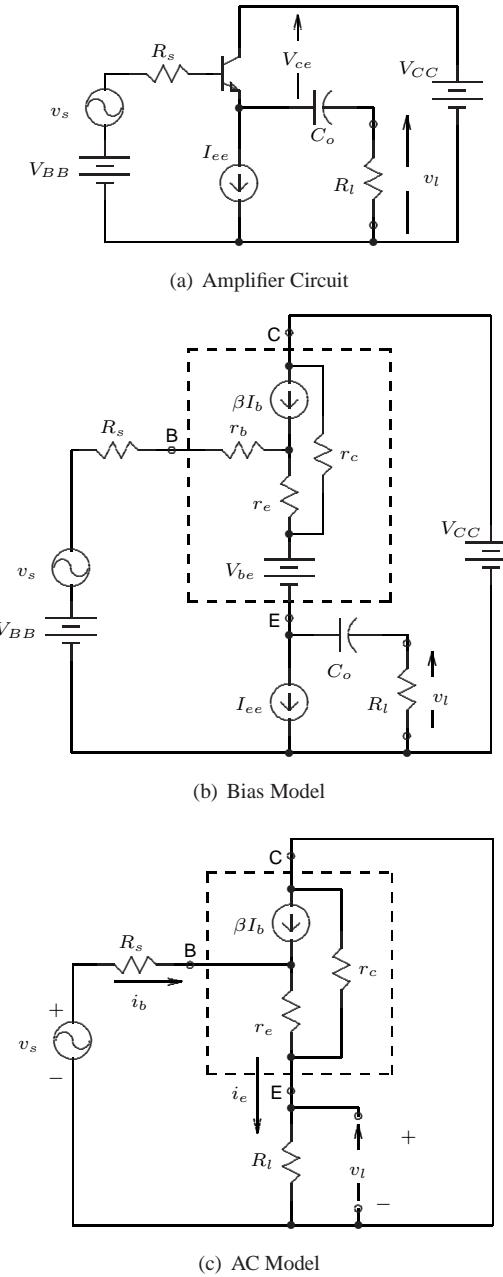


Figure 766: Emitter Follower Model

It's easier to understand the analysis and the results are more useful when we include the effect of source resistance and load resistance from the outset. For this reason, we'll determine the voltage gain between the source voltage v_s and the load voltage v_l with source resistance and load resistance in place.

Figure 766(a) shows the basic emitter follower amplifier circuit. Figure 766(b) shows that circuit with the transistor replaced by the T model. Figure 766(c) shows the AC model of the circuit. In this circuit the tail current sink I_{ee} has become an open circuit. The collector resistance r_c of the transistor appears in parallel with the load resistance. In emitter follower circuits, the load resistance R_l is almost always much smaller than the collector resistance, and so r_c can be ignored. The base resistance r_b is assumed to be negligible and is treated as zero ohms. The output capacitor is assumed to be a short circuit.

Voltage Gain (Source to Load)

We expect on the basis of our understanding of the behaviour of the emitter follower (output follows input) that the AC voltage gain will be close to unity. Let's see if that is the case. Applying KVL to the base-emitter loop in figure 766(c), we have

$$+v_s - i_b R_s - i_e r_e - v_l = 0 \quad (1323)$$

Let's get both currents i_b and i_e in terms of the emitter current i_e :

$$i_b = \frac{i_e}{\beta + 1} \quad (1324)$$

We will also need to involve the load resistance R_l :

$$i_e = \frac{v_l}{R_l} \quad (1325)$$

Putting these three equations together and cranking the algebra machine, we can obtain for the voltage gain:

$$\begin{aligned} A_v &= \frac{v_l}{v_s} \\ &= \frac{R_l}{R_s/(\beta + 1) + r_e + R_l} \end{aligned} \quad (1326)$$

This equation describes the behaviour of the voltage divider shown in figure 767.

By inspection, we can determine the following:

- The gain must always be somewhat less than unity.
- The gain approaches unity as the load resistance R_l becomes larger compared to the source resistance R_s divided by beta.
- A larger value of beta will reduce the effect of the source resistance on the voltage gain.
- If the emitter follower is driven by a voltage source, then $R_s = 0$. Assuming that the emitter incremental resistance r_e is much smaller than the load resistance R_l , then the voltage gain is very close to unity.

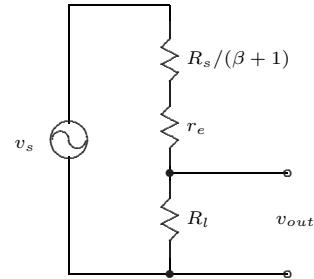


Figure 767: Voltage Gain of Emitter Follower

Current Gain

The current in the load is the emitter current i_e . The current supplied by the source is i_b . Consequently, the current gain is simply

$$\begin{aligned} A_i &= \frac{i_l}{i_i} \\ &= \frac{i_e}{i_b} \\ &= \beta + 1 \\ &\approx \beta \end{aligned} \tag{1327}$$

Input Resistance

By the BJT reflection principle (section 29.10), from the base terminal the resistance seen is equal to the emitter resistance multiplied by $\beta + 1$. Then

$$\begin{aligned} r_i &= (\beta + 1)(r_e + R_l) \\ &\approx \beta R_l \end{aligned} \tag{1328}$$

The approximation is valid when $\beta \gg 1$ and $R_l \gg r_e$, the usual case.

Output Resistance

By the same reflection principle, the resistance in the base circuit divides by $\beta + 1$ to move it into the emitter. Then

$$\begin{aligned} r_o &= \frac{R_s}{\beta + 1} + r_e \\ &\approx \frac{R_s}{\beta} \end{aligned} \tag{1329}$$

The approximation is valid when $\beta \gg 1$ and $R_s/\beta \gg r_e$, usually true.

29.28 Emitter Follower Oscillation

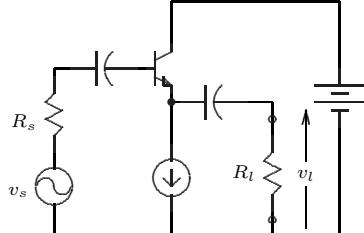
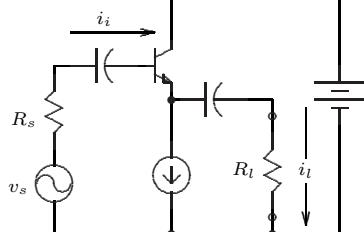
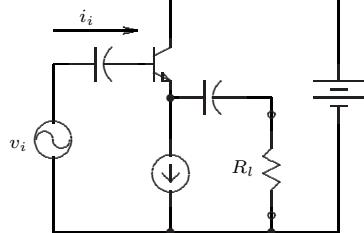
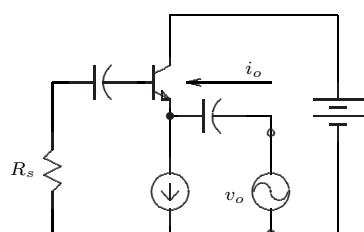
The emitter follower would appear to be an unlikely candidate for spurious oscillation, but given the right conditions, it can and does happen. A high-frequency transistor in the emitter follower configuration can oscillate at a frequency that is well beyond the frequency range of an oscilloscope. The symptom is a DC offset that changes in unpredictable ways. For example, attaching a length of wire to the base terminal may appear to cause a DC shift in the output voltage.

The analysis is complicated [253], [254], [255]. Oscillation is precipitated by inductance in the input circuit and capacitance in the output load. These capacitances can be strays: discrete components are not necessary. Many cases can be cured with the addition of a small resistance (eg, 200Ω) in series with the base lead.

29.29 Summary: Emitter Follower Equations

The equations for the single stage BJT emitter follower amplifier along with some of the usual approximations are shown below. These equations assume a constant current bias. The equations must be modified if a resistor R_e is used to implement the emitter bias current sink. A DC path for base current – such as a resistor between the base terminal and ground – must exist, but for simplicity it's not shown on the schematics.

It's not so important to memorize the equations as it is to remember the model. Then the equations may be determined for this and other similar circuits.

Voltage Gain		$\frac{v_l}{v_s} = \frac{R_l}{\frac{R_s}{\beta + 1} + r_e + R_l}$ $\approx \frac{R_l}{r_e + R_l} \text{ if } \frac{R_s}{\beta + 1} \ll r_e + R_l$
Current Gain		$\frac{i_l}{i_i} = \beta + 1$ $\approx \beta \text{ if } \beta \gg 1$
Input Resistance		$\frac{v_i}{i_i} = (\beta + 1)(r_e + R_l)$ $\approx \beta R_l \text{ if } R_l \gg r_e$
Output Resistance		$\frac{v_o}{i_o} = \frac{R_s}{\beta + 1} + r_e$ $\approx \frac{R_s}{\beta} + r_e \text{ if } \beta \gg 1$

29.30 The Emitter Follower as a Negative Feedback System

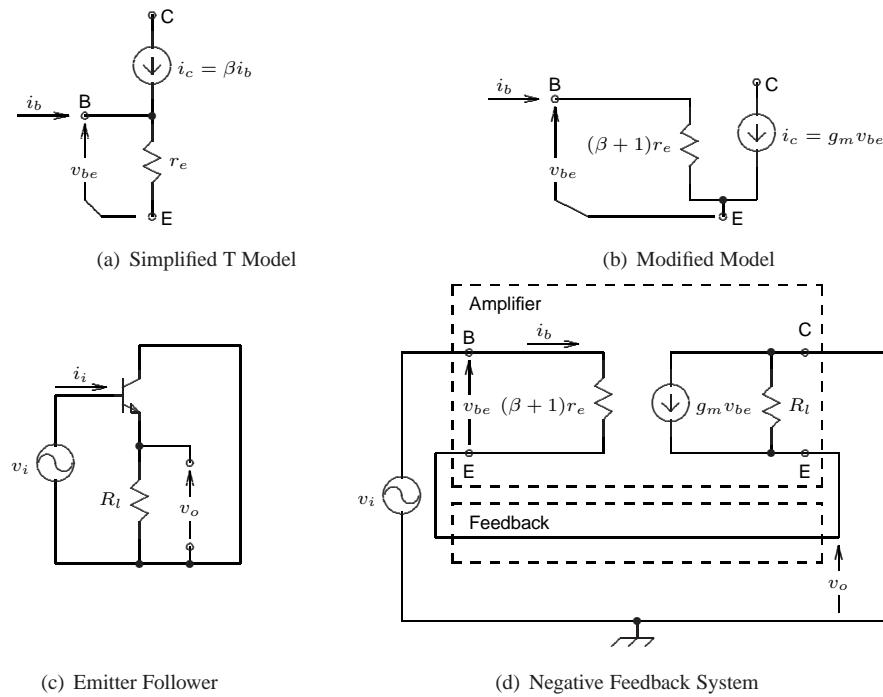


Figure 768: Emitter Follower as NFB System

It's most straightforward to analyse the behaviour of the emitter follower in a frontal attack as in the previous sections. However it's possible to view it as a negative feedback amplifier and this viewpoint leads to some interesting insights.

We begin by modifying a simplified T model of the transistor²³⁸ – as shown in figure 768(a) – into something resembling an operational amplifier, figure 768(b). The emitter resistance r_e can be moved out of the emitter into a separate path for the base current if it is multiplied by $\beta + 1$:

$$\begin{aligned}
 r_i &= \frac{v_i}{i_i} \\
 &= \frac{v_{be}}{i_b} \\
 &= \frac{(i_b + i_c)r_e}{i_b} \\
 &= \frac{(i_b + \beta i_b)r_e}{i_b} \\
 &= (\beta + 1)r_e
 \end{aligned} \tag{1330}$$

²³⁸All the models shown in this section are the AC equivalent circuits, with DC voltage sources short-circuited, DC current sources open-circuited and capacitors treated as short circuits.

Assuming that $\alpha \approx 1$, and using $g_m = 1/r_e$, the collector current generator i_c can be given by

$$\begin{aligned} i_c &= \alpha i_e \\ &= \alpha \frac{v_{be}}{r_e} \\ &= g_m v_{be} \end{aligned} \quad (1331)$$

This represents the collector current generator as a voltage-controlled current source. (This model is not tied to the emitter follower – it can be used to analyse the common-emitter stage as well.)

Figure 768(c) shows the AC equivalent circuit of an emitter follower. Figure 768(d) shows the model of figure 768(b) substituted into the emitter follower circuit of figure 768(c). With some rearrangement of wires, the load resistance can be seen to be in parallel with the collector current generator.

Figure 769 shows an op-amp *buffer* circuit (see section 12.3 on page 318). The output is placed in series with the input error voltage, raising the input resistance seen from the source. The feedback factor is unity, so the closed-loop gain G , which is the reciprocal of the feedback factor, is also equal to unity.

Treating the emitter follower of figure 768(d) as a negative feedback system, we have an open-loop gain A_{ol} of:

$$\begin{aligned} A_{ol} &= \frac{v_{ce}}{v_{be}} \\ &= \frac{g_m v_{be} R_l}{v_{be}} \\ &= g_m R_l \text{ volts/volt} \end{aligned} \quad (1332)$$

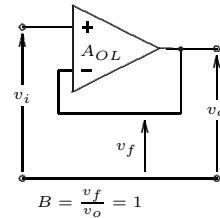


Figure 769: Buffer

and a feedback factor of $B = 1$. Then from the negative feedback equation 424 on page 287, we have that:

$$\begin{aligned} G &= \frac{v_o}{v_i} \\ &= \frac{A_{ol}}{1 + A_{ol}B} \\ &= \frac{g_m R_l}{1 + g_m R_l} \text{ volts/volt} \end{aligned} \quad (1333)$$

Substitute $1/r_e$ for g_m and we have:

$$\begin{aligned} G &= \frac{R_l/r_e}{1 + R_l/r_e} \\ &= \frac{R_l}{r_e + R_l} \text{ volts/volt} \end{aligned} \quad (1334)$$

If $R_l \gg r_e$ (as is usually the case), then $G \approx 1$.

This is the same result we obtain by circuit analysis, so it seems reasonable that the circuit can be analysed as a negative feedback system. Where does this lead? Well, we'd expect that the input resistance $(\beta + 1)r_e$ of the original amplifier is raised by a factor of $1 + A_{ol}B$ by the negative feedback. That is in fact the case. Furthermore, we can expect that disturbances to the output voltage that occur inside the amplifier will be minimized by the same factor $1 + A_{ol}B$.

Relating the Common Emitter and Emitter Follower Amplifiers

The common emitter amplifier and the emitter follower have many similarities. In both cases,

- the controlling voltage is placed between base and emitter.
- the controlling current flows into the base terminal
- the output current is beta (or beta+1) times as large as the input current

In some sense, they are the same device with different feedback arrangements. The common-emitter amplifier has no intrinsic feedback, but an external feedback network can be attached to the amplifier. The emitter follower has unity gain feedback (often described as *100% feedback*) inherent in its circuit. Both circuits have the same amplifier device. This suggests that a common-emitter amplifier with 100% feedback would have the same properties as the emitter follower. Let's see if that is true.

For want of a better name, we'll call this experimental circuit a *collector follower*. The circuit is shown in figure 770(a). The output voltage v_o is coupled to a 1:1 transformer via a capacitor (to block the collector DC voltage). The secondary of the transformer creates a floating voltage source $v_f = v_o$ which is in series with the input signal, and opposite in polarity. The AC equivalent circuit is shown in figure 770(b).

Applying KVL to the input loop, we have

$$+v_i - v_f - v_{be} = 0 \quad (1335)$$

The feedback voltage v_f is equal to the output voltage v_o :

$$v_f = v_o \quad (1336)$$

The amplifier is a conventional common-emitter stage with an open-loop gain of

$$A_{ol} = \frac{R_c}{r_e} \quad (1337)$$

Then the output voltage is:

$$\begin{aligned} v_o &= A_{ol}v_{be} \\ &= \frac{R_c}{r_e}v_{be} \end{aligned} \quad (1338)$$

Rearranging to solve for v_{be} , we have:

$$v_{be} = \frac{r_e}{R_c}v_o \quad (1339)$$

Substitute for v_f from equation 1335 and for v_{be} from equation 1339, into equation 1335.

$$+v_i - v_o - \frac{r_e}{R_c}v_o = 0 \quad (1340)$$

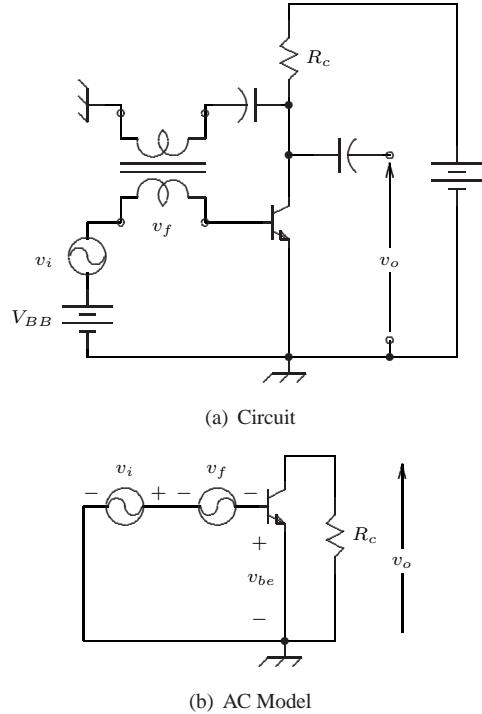


Figure 770: The *Collector Follower*

Rearrange and solve for the closed-loop gain $G = v_o/v_i$:

$$\begin{aligned} G &= \frac{v_o}{v_i} \\ &= \frac{R_c}{r_e + R_c} \end{aligned} \quad (1341)$$

This is exactly the same gain that we obtained for the emitter follower. In a similar manner, one can show that the input resistance seen at v_i is

$$r_i = \beta(r_e + R_c) \quad (1342)$$

Again, this result is very similar to that of the emitter follower, where we found that

$$r_i = (\beta + 1)(r_e + R_e) \quad (1343)$$

This suggests that this collector follower circuit behaves in a very similar manner to the emitter follower as a consequence of the 100% feedback from the output signal in series with the input signal. There are minor differences – the emitter follower output is in phase with the input, the collector follower output is inverted – but overall the two circuits are more alike than different.

29.31 Emitter Follower Design Example

An emitter follower amplifier is to be used between a Thevenin source and a load resistance. The signal open-circuit amplitude is 1 volt p-p at a frequency of 1kHz, with an internal resistance of $5\text{k}\Omega$. The load resistance is $1\text{k}\Omega$. The amplifier is to operate from a single 9 volt DC supply. The transistor to be used is the 2N4401, for which the current gain is approximately 100 over the range of collector current 0.1mA to 10mA.

If the source is directly connected to the load, the voltage across the load (treating R_s and R_l as a voltage divider) is 1/6 of the source voltage. An emitter follower buffer has high input resistance, low output resistance and a voltage gain of approximately unity.

Using it as a buffer between the source and load, the voltage across the load will be approximately equal to the source voltage. A block diagram of the design is shown in figure 771. A diagram of the circuit configuration is shown in figure 772.

1. Design the emitter follower circuit.
2. Determine the amplitude of the signal at the load.

Bias Current I_{ee}

We can begin by determining the bias current. The peak voltage in the load is $v_{lpk} = 0.5\text{V}$, so the peak current i_{lpk} is:

$$\begin{aligned} i_{lpk} &= \frac{v_{lpk}}{R_l} \\ &= \frac{0.5\text{V}}{1\text{k}\Omega} \\ &= 0.5\text{mA} \end{aligned}$$

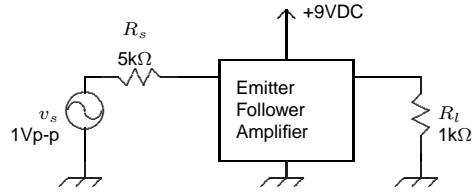


Figure 771: Emitter Follower Design Problem

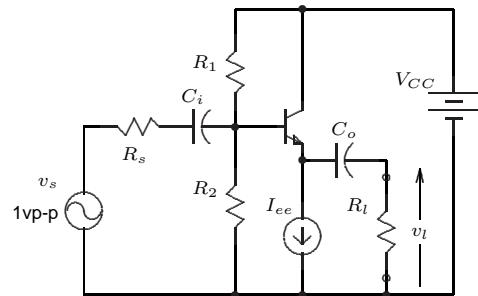


Figure 772: Circuit Configuration

The emitter bias current must be much larger than this, so we might choose 2mA. This provides a safety factor of 4 before the transistor stops conducting on part of the cycle.

Base Bias Network

First, we have to decide where to put the base voltage V_{BB} . Anything more positive than half of the signal plus V_{be} will work. We might put V_{BB} at half the supply voltage so that the divider resistors are equal. (Identical part values make production happy because they can do a larger quantity buy of the common parts and stocking is simplified.) So $R_1 = R_2$ and $V_{BB} = 4.5V$. There is lots of room for the voltage of the divider to drop a bit when supplying base current, so the internal resistance is not critical. Furthermore, we would like the voltage divider to not have a significant loading effect on the source, which means that it should present a large resistance compared the source resistance R_s .

The base bias network and its Thevenin equivalent are shown in figure 773. The voltage divider presents a resistance of $R_1 \parallel R_2$ to the source. The two resistors are equal, call them R , so in parallel they present $R/2$ to the source. Set this equal to 10 times the source resistance and we have:

$$R_{bb} = \frac{R}{2} = 10 \times R_s = 10 \times 5k = 50k\Omega$$

Then $R = R_1 = R_2 = 100k\Omega$ and $R_{bb} = 50k\Omega$.

Let's see how much the base current drops the base voltage from V_{BB} by the time it reaches the base. The DC base current I_b is:

$$I_b = \frac{I_e}{\beta + 1} = \frac{0.002}{101} = 19\mu A$$

Then the DC voltage at the base of the transistor is:

$$V_b = V_{BB} - I_b R_{bb} = 4.5 - (19 \times 10^{-6}) \times (50 \times 10^3) = 3.55 \text{ volts}$$

If the beta of the transistor changes, this will change the base current and consequently the base voltage V_b . In a production design one would have to check the maximum and minimum values to ensure that the transistor does not saturate or cutoff under different values of beta.

Calculation of Voltage Gain

First, we have to Thevenize the input network voltage divider comprised of R_s and R_{bb} . The process is illustrated in figure 774. The open circuit voltage v'_s of the input network is given by the voltage divider equation:

$$v'_s = \frac{R_{bb}}{R_s + R_{bb}} v_s = \frac{50k}{5k + 50k} v_s = 0.909 v_s$$

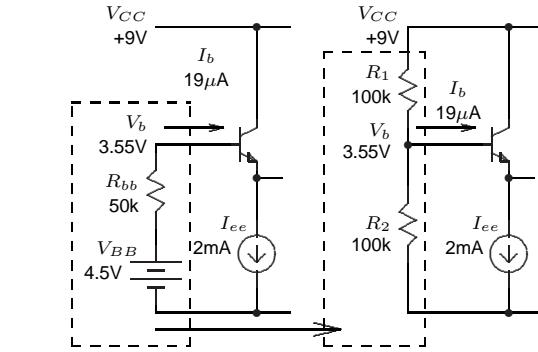


Figure 773: Base Bias Network

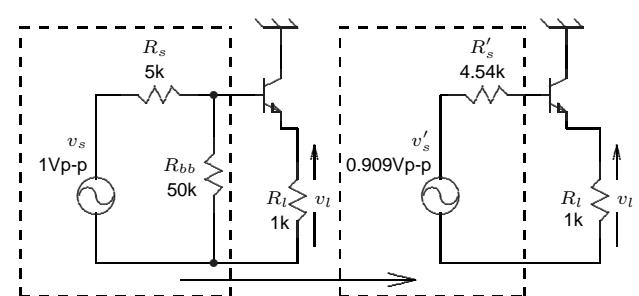


Figure 774: Thevenising the Input Network

Then the gain of the input network is:

$$\frac{v'_s}{v_s} = 0.909 \text{ v/v}$$

The R_s , R_{bb} voltage divider has an internal resistance R'_s which is given by the parallel combination of the two resistors R_s and R_{bb} :

$$R'_s = R_s \parallel R_{bb} = 5000 \parallel 50000 = 4545\Omega$$

Now we can represent the input source as a Thevenin source v'_s with an internal resistance R'_s , as shown in the right half of figure 774.

For the voltage gain calculation we'll need the emitter incremental resistance r_e , so let's calculate that now:

$$r_e = \frac{V_t}{I_e} = \frac{0.026}{0.002} = 13\Omega$$

Now we can determine the gain of the emitter follower (the ratio of v_l/v'_s) with equation 1326 from page 874:

$$A'_v = \frac{v_l}{v'_s} = \frac{R_l}{R'_s/(\beta + 1) + r_e + R_l} = \frac{1000}{4545/(100 + 1) + 13 + 1000} = 0.945 \text{ v/v}$$

The overall gain between source v_s and load v_l is given by the product of the gain of the input network and the gain of the emitter follower:

$$A_v = \frac{v_l}{v_s} = \frac{v'_s}{v_s} \times \frac{v_l}{v'_s} = 0.909 \times 0.944 = 0.858 \text{ v/v}$$

That is, with a source signal v_s of 1 volt peak-peak, the load voltage v_l would be 0.858 volts peak-peak.

Input Capacitor C_i

Figure 775 shows the equivalent circuit for the calculation of input capacitance C_i . Reflecting the resistances r_e and R_l from the emitter to the base, we multiply them by $\beta + 1$. Consequently:

$$r_i = (\beta + 1)(r_e + R_l)$$

The input network is a highpass network RC network. The capacitance is C_i . The resistance R is a combination of the resistors in the input network:

$$\begin{aligned} R &= R_s + (R_{bb} \parallel r_i) \\ &= R_s + (R_{bb} \parallel [(\beta + 1)(r_e + R_l)]) \\ &= 5k + (50k \parallel [(100 + 1)(13 + 1k)]) \\ &= 38.5k\Omega \end{aligned}$$

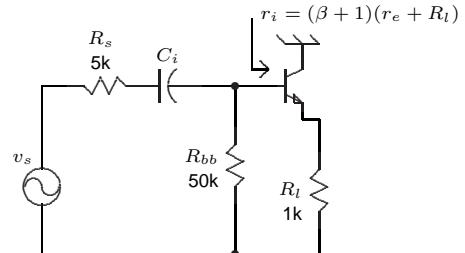


Figure 775: Calculating C_i

To ensure that the operating frequency is well above the corner frequency, we should put the corner frequency much less than 1000Hz. Taking *much less than* to mean *less by a factor of 10*, we would put f_c for the network at 100Hz. The corner frequency for a highpass network is:

$$f_c = \frac{1}{2\pi RC}$$

Solving for the capacitance,

$$C_i = \frac{1}{2\pi R f_c} = \frac{1}{2 \times 3.14 \times 38500 \times 100} = 41.3nF$$

Output Capacitor C_o

Figure 776 shows the equivalent circuit for the calculation of output capacitance C_o . The base resistance is the parallel combination of R_s and R_{bb} . Reflecting this resistance from the base to the emitter, we divide them by $\beta + 1$. This reflected resistance appears in series with the emitter incremental resistance r_e .

(Figure 767 on page 874 shows the equivalent circuit diagram.) Consequently:

$$r_o = r_e + \frac{R_s \parallel R_{bb}}{\beta + 1} = 13 + \frac{5000 \parallel 50000}{100 + 1} = 58\Omega$$

Notice that the output resistance r_o of the emitter follower is much less than the load resistance R_l . This is good news. Any change in the load resistance (and output load current) will then have little effect on the load voltage.

Like the input network, the output network is a highpass RC network. The capacitance is C_o . This time, the resistance R is a combination of the resistors in the output network:

$$R = r_o + R_l = 58 + 1000 = 1058\Omega$$

Taking the same approach as we did with the input network, we can solve for the output capacitance C_o :

$$C_o = \frac{1}{2\pi R f_c} = \frac{1}{2 \times 3.14 \times 1058 \times 100} = 1.5\mu\text{F}$$

The output capacitor is much larger than the input capacitor because the resistances in the output circuit are smaller than those in the input circuit.

Constant Current Sink I_{ee}

Ideally, the current sink should absorb a constant current, independent of the variation in transistor emitter voltage. There are several possible ways to implement the current sink I_{ee} . We'll look at three possibilities: a current mirror, a long-tailed resistor, and a simple resistor to ground.

A very effective constant current sink can be constructed with a current mirror (section 30.3, page 915). The circuit configuration is shown in figure 777.

Resistor R_3 sets up a current equal to I_{ee} in the diode-connected transistor Q_2 . This causes the mirror transistor Q_3 to conduct approximately the same current into its emitter. This tail current is largely independent of the voltage across Q_3 (as long as Q_3 is kept out of saturation), so it's an effective current source.

The calculation of R_3 is straightforward. If V_{be2} is the base-emitter voltage of Q_2 , then:

$$R_3 = \frac{V_{CC} - V_{be2}}{I_{ee}} = \frac{9 - 0.6}{2\text{mA}} = 4.2\text{k}\Omega$$

If the mirror transistors Q_2 and Q_3 will be at exactly the same temperature, then this design will work as shown. If there is some possibility that they will be at slightly different temperatures, then the circuit should include small-value emitter resistors to equalize the currents in Q_2 and Q_3 , as described in section 30.3.

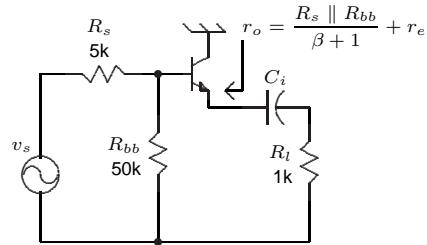


Figure 776: Calculating C_o

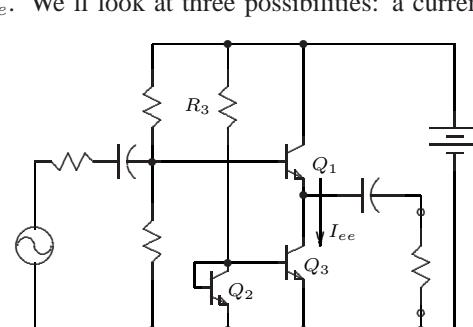


Figure 777: Current Mirror Current Sink

Long-Tailed Resistor Current Sink

If a negative supply is available, then the circuit of figure 778 is possible. This arrangement is referred to as *long-tailing* the emitter resistor, that is, increasing the value of R_{ee} with a large tail voltage.

Applying KVL to the loop including the base, emitter and negative supply V_{EE} , we have:

$$+V_b - V_{be} - V_{Ree} + V_{EE} = 0$$

Rearranging this to solve for the voltage across R_{ee} , and plugging in the known values, we have:

$$V_{Ree} = +V_b - V_{be} + V_{EE} = 3.55 - 0.6 + 9 = 11.95\text{V}$$

Then for a tail current I_{ee} of 2mA, the tail resistance is:

$$R_{ee} = \frac{V_{Ree}}{I_{ee}} = \frac{11.85}{2 \times 10^{-3}} = 5.93\text{k}\Omega$$

Since the output signal appears at the emitter of the transistor an emitter resistor will conduct a certain level of signal current and this may impact the AC behaviour of the circuit.

With an ideal current sink for I_{ee} , the AC load resistance is simply R_l . With resistor R_{ee} as the current sink, the effective load resistance reduces to $R_l \parallel R_{ee}$. This requires that the DC bias current in R_{ee} be increased and complicates the design. The degree to which this is a problem depends on the amplitude of the output signal: small AC output signals are less of a problem.

In our example design, the effect of R_{ee} is to reduce the load resistance from $1\text{k}\Omega$ to $1\text{k}\parallel 5.93\text{k}=855\Omega$. This will impact the AC gain and input resistance, and may require an increase in the bias current to avoid clipping.

Resistor Current Sink

In a single supply emitter follower circuit, it's tempting to use a single resistor R_{ee} between emitter and ground as the tail current sink²³⁹.

The circuit configuration is shown in figure 779. The value of R_{ee} depends on the base bias voltage V_b :

$$R_{ee} = \frac{V_{BB} - V_{be}}{I_{ee}} = \frac{3.55 - 0.6}{2 \times 10^{-3}} = 1.48\text{k}\Omega$$

In this case, the emitter resistor is comparable in size to the load resistance, so it will have a significant effect on the AC behaviour of the circuit. If the base bias voltage V_b can be moved toward the positive supply (as is possible in this example), then the value of R_{ee} will increase for the same emitter current.

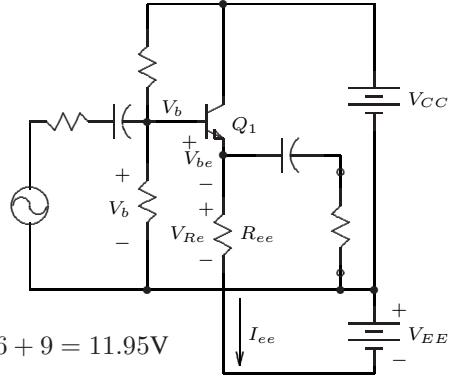


Figure 778: Long Tailed Current Sink

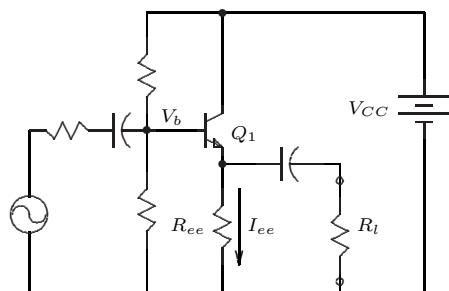


Figure 779: Resistor Current Sink

²³⁹We might refer to this arrangement as the *short-tailed* current sink.

Design Summary

- The overall gain from source to load in this example is only 0.858v/v , well short of unity gain. If this is important, an op-amp unity-gain buffer (section 12.3) or JFET follower (section 31.5) might be a better choice. In this emitter-follower example the shortfall from unity gain is partially a consequence of the base bias divider network that loads down the source. The op-amp buffer and JFET follower have lower input bias current and so the bias divider can be a larger resistance which reduces loading of the source and improves the overall gain.
- The mirror current sink is simple to design and behaves in an ideal fashion. Unfortunately, it requires 1 resistor and 2 transistors to implement.
- The long-tailed resistor current sink requires a negative power supply and may have a small impact on the overall performance of the circuit.
- The short-tailed resistor current sink is economical of parts but significantly complicates the design of the circuit.

A circuit of the completed emitter follower, with component values, is shown in figure 780.

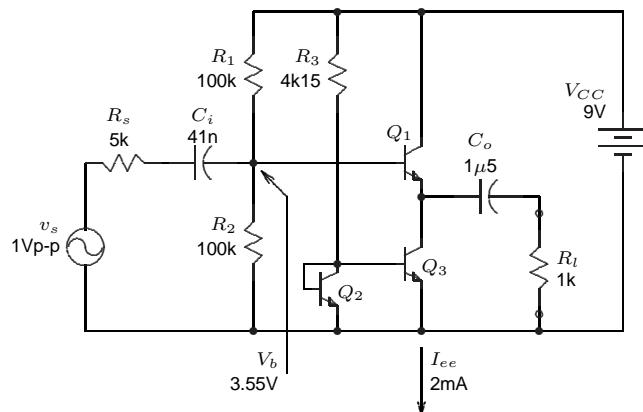


Figure 780: Completed Emitter Follower

29.32 The Common Base Amplifier

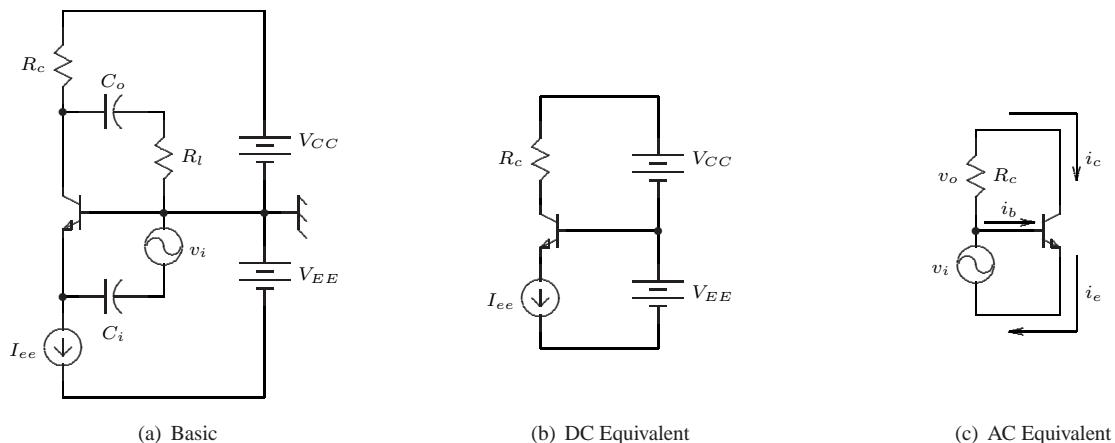


Figure 781: Basic Common Base Amplifier

Overview

The common emitter and emitter follower amplifier configurations are the most used of the three possibilities. The common-base stage is less popular, but it has some important applications.

We will find that a common-base stage has very low input resistance (equal to r_e) and a current gain that is approximately unity. These properties make it unsuitable for many applications. However, it does have a large voltage gain – the same as the common emitter stage, equal to the ratio of collector and emitter resistances. Consequently it does have power gain, which is the definition of a real amplifier.

The common-base stage has the same signal-handling issues as the common emitter stage, so the input signal at the emitter must be kept below a few millivolts to avoid serious distortion of the waveform. It is quite usual for the input source to have a much larger internal resistance R_s than the input resistance r_e of the common-base stage. In that case, the overall voltage gain will be equal to the ratio of the collector to source resistance.

The main use of the common-base stage is in conjunction with the common-emitter stage. This combination of stages is used to create an ultra-high output resistance which is useful in various current source designs. The same combination works well at high frequencies, as we will see when we look at the high-frequency behaviour of various amplifiers.

Since the emitter of the common-base stage is at a very low AC impedance, it's a good point at which to add currents from different sources. For example, a circuit of this type could be used to add together several audio signals in a simple audio mixer.

Basic Function

A dual-supply common-base amplifier is shown in figure 781(a).

The DC equivalent circuit, used to determine the bias conditions, is in figure 781(b). The emitter current generator I_{ee} establishes the DC bias current in the transistor. It might be chosen so that the collector voltage of the transistor is half-way between V_{CC} and the base voltage (the ground point in this circuit²⁴⁰).

²⁴⁰There is nothing magical about the ground point in this circuit – as usual, ground is simply a reference point for the other voltages, like sea level in measuring altitude. However, since the input voltage generator and the output voltage have a common point at the base, it's convenient

The AC equivalent circuit is shown in figure 781(c). The input voltage generator v_i drives current through the emitter incremental resistance r_e . This causes an AC emitter current i_e which appears as collector current. The current gain between emitter and collector is alpha, which is approximately equal to 1. Then the collector resistor R_c converts this AC collector current into an AC collector voltage $v_o = i_c R_c$.

Device Bias Current

The AC emitter current adds to and subtracts from the DC emitter current. When the signal current is out of the emitter, the total emitter current is the sum of the DC bias current and the signal current.

When the signal current is into the emitter, the total emitter current is the difference between the DC bias current and the signal current. If the peak value of the AC emitter current is equal to the DC bias current, the emitter current of the transistor is zero. Any further increase in the signal will drive the transistor further into cutoff. Consequently, the value of the DC emitter bias current I_{ee} must be equal to or larger than the peak value of the input AC current.

Single-Supply Configuration

Various common-base stages can be designed by analogy with the emitter-follower amplifier configurations previously shown in figure 763 through figure 765. One possibility, a single-supply AC-coupled common-base amplifier, is shown in figure 782. In this case, the R_1, R_2 voltage divider establishes a DC voltage V_b at the base of the transistor. This voltage, minus one base-emitter drop, appears at the emitter of the transistor. The emitter voltage V_e divided by R_{ee} establishes the DC bias current in the transistor.

The available signal swing at the collector of the transistor is then between the supply voltage V_{CC} and the DC emitter voltage V_e . The peak-peak output AC voltage must fit within this available space, preferably with a volt or two left over to keep the transistor out of saturation.

Modelling the Common Base Amplifier

The circuit to be modelled is in figure 783(a) and its AC model in figure 783(b). The model assumes that the internal base resistance r_b is zero and that the base terminal is at an AC ground point. (If the base bias source has appreciable resistance, then it must be bypassed to ground by a capacitor).

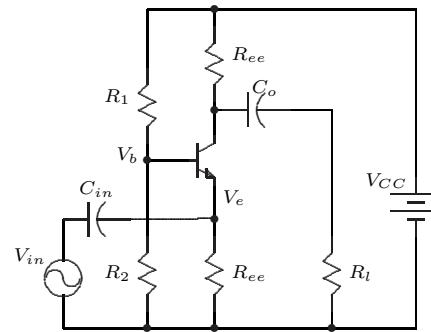


Figure 782: Common Base Single Supply AC Amplifier

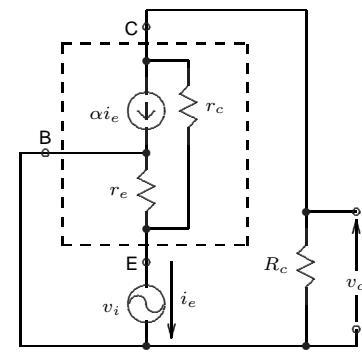
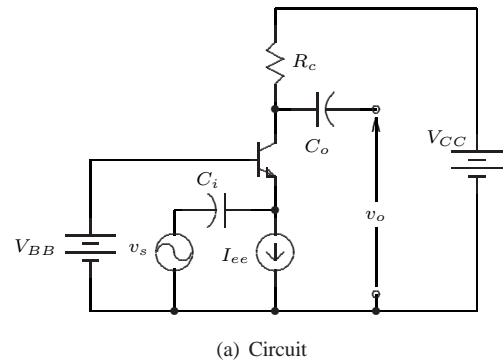


Figure 783: Modelling the Common Base Amplifier

to choose that particular point as ground.

Voltage Gain

The input voltage v_i drives an emitter current i_e through the emitter resistance r_e :

$$i_e = \frac{v_i}{r_e} \quad (1344)$$

The collector current is equal to alpha times the emitter current:

$$i_c = \alpha i_e \quad (1345)$$

The output voltage v_o is the voltage across the collector resistor R_c :

$$v_o = i_c R_c \quad (1346)$$

Equations 1344 through 1346 can be used to find the voltage gain:

$$\begin{aligned} A_v &= \frac{v_o}{v_i} \\ &= \frac{v_i \alpha R_c}{r_e v_i} \\ &= \alpha \frac{R_c}{r_e} \\ &\approx \frac{R_c}{r_e} \end{aligned} \quad (1347)$$

This is the same voltage gain as the common-emitter amplifier.

The common-base amplifier is limited in input signal amplitude in the same way as the common-emitter amplifier (section 29.18). The circuit of figure 784 sacrifices some voltage gain in exchange for an ability to handle larger signals. In this circuit, the source resistance R_s and the emitter resistance r_e form a voltage divider so that the source voltage V_s can be much larger than the voltage across r_e . Then one can show that:

$$\begin{aligned} A_v &= \alpha \frac{R_c}{R_s + r_e} \\ &\approx \frac{R_c}{R_s} \end{aligned} \quad (1348)$$

Current Gain

Current gain is determined with a short-circuit load. Then the current in the load is the collector current i_c . The current supplied by the source is i_e . Consequently, the current gain is simply

$$\begin{aligned} A_i &= \frac{i_l}{i_i} \\ &= \frac{i_c}{i_e} \\ &= \alpha \end{aligned} \quad (1349)$$

Of course, alpha is approximately unity.

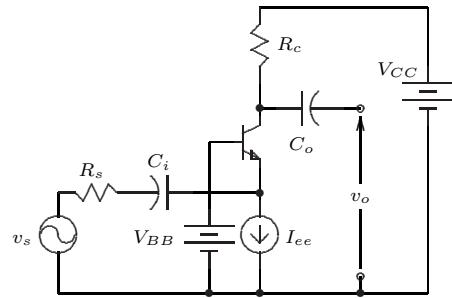


Figure 784: Common Base Amp with Source Resistance

Input Resistance

By inspection of figure 783(b), the input source v_i drives current into the emitter resistance r_e . So:

$$\begin{aligned} r_i &= \frac{v_{be}}{i_e} \\ &= r_e \end{aligned} \quad (1350)$$

This is a very low resistance value, typically in the order of a few tens of ohms.

Output Resistance

The output resistance measurement is made (conceptually) by short-circuiting the input voltage, attaching an AC voltage source between the output terminals, and measuring the resultant current into the output terminals. Then:

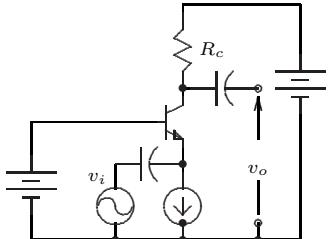
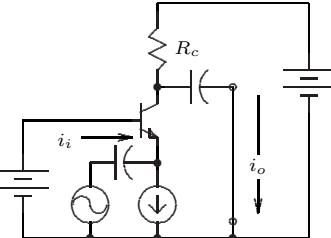
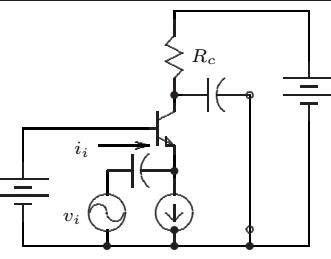
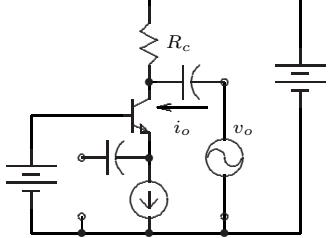
$$\begin{aligned} r_o &= \frac{v_o}{i_o} \\ &= r_c \end{aligned} \quad (1351)$$

We must be careful here. The collector resistance models the Early effect – as the collector voltage increases, so does the collector current. The Early effect is much less pronounced when the transistor is controlled by its emitter current. Consequently, the value of r_c must be the value measured specifically for the common-base amplifier. This is approximately beta times the value for the common emitter amplifier [250], [256].

$$r_{c(cb)} = \beta r_{c(ce)} \quad (1352)$$

where $r_{c(ce)}$ is the collector resistance of the common emitter amplifier and $r_{c(cb)}$ is the collector resistance of the common base amplifier.

29.33 Summary: Common-Base Equations

Voltage Gain A_v		$\frac{v_o}{v_i} = \frac{R_c}{r_e}$ (Notes 1,2)
Current Gain A_i		$\frac{i_o}{i_i} = \alpha \approx 1$
Input Resistance r_i		$\frac{v_i}{i_i} = r_e$
Output Resistance r_o		$\frac{v_o}{i_o} = R_c$ (Note 2)

Notes

1. If the source has internal resistance R_s then replace r_e with $r_e + R_s$
2. If the collector resistor R_c is more than 1/10 of the collector incremental resistance, replace R_c with $R_c \parallel r_c$.
3. All the capacitors are assumed to be short circuits at the measurement frequency.

29.34 The Inverted Transistor

The symmetry of the diode model implies that it might be possible to interchange the collector and emitter terminals. (The equivalent terminals of the JFET transistor, the source and drain, can be interchanged in this way with no effect on the operation.)

In fact, the BJT will function in this configuration, but with different characteristics. The base-collector and base-emitter junctions are constructed differently, and so the BJT is not entirely symmetrical.

- The maximum allowed base-emitter junction voltage is generally much less than the maximum base-collector voltage, (6 volts vs 30 volts) so the supply voltage must be limited.
- The normal configuration of the BJT is referred to as its *forward* condition and its value of beta under this condition, which we have been calling β , is then referred to as the *forward beta* β_f . When the collector and emitter terminals are interchanged, the transistor is said to be in its *reverse* configuration and the current gain is referred to as the *reverse beta* β_r .

The value of reverse beta β_r is *much* less than the normal forward beta. Values of 2 to 5 are typical for β_r compared to values of 20 to 200 for β_f .

This usually shows up when a circuit is functioning but in a marginal fashion. It then turns out that the collector and emitter terminals of a transistor have been inadvertently interchanged. As a consequence, the current gain is much less than it should be.

- The collector-emitter saturation voltage V_{cesat} of a BJT transistor is substantially lower in the inverted mode. There was a time when this characteristic made the inverted transistor attractive for signal chopping applications [77]. However, JFETS and MOSFETs are a much superior choice for signal switching applications (including chopping), so they are the devices of choice in modern designs.

29.35 Noise in the BJT

In a system that must amplify a weak signal, the noise performance of the first stage of the amplifier chain sets a minimum signal-noise ratio for the entire system. A single BJT (or JFET) can have better noise performance than an operational amplifier. Consequently, it may be best to have a discrete transistor in the first stage of the amplifier. It's important to be able to characterise the noise performance of that transistor.

Here we give a simple method of translating data sheet information into noise performance. The minimum of information on a data sheet is the so-called *noise figure* of the transistor when used as an amplifier. Now we show how to convert that into noise voltage.

As shown in figure 785 the noise figure NF of an amplifier describes how much noise the amplifier adds to a signal. It is defined as *the ratio of the amplifier output signal with noise, to the output signal if the amplifier were noiseless*. Both signals are in watts, and the noise figure NF is in decibels.

Power is proportional to voltage squared, and a power ratio is expressed in decibels as $10 \log_{10}$ of the power ratio.

Then the noise figure is given by:

$$\begin{aligned} NF &= 10 \log_{10} \left(\frac{P_{out}}{P_{in}} \right) \\ &= 10 \log_{10} \left(\frac{e_i^2 + e_A^2}{e_i^2} \right) \\ &= 10 \log_{10} \left(1 + \frac{e_A^2}{e_i^2} \right) \end{aligned} \quad (1353)$$

where e_A is the noise voltage of the amplifier referred to the input, e_i is the noise voltage input from the source²⁴¹.

From equation 987 on page 670, noise voltage is related to temperature, bandwidth and resistance by

$$e_n = \sqrt{4KTBR}$$

where

- K is Boltzmann's Constant, 1.38×10^{-23} joules/kelvin
- T is the absolute temperature, kelvins
- B is the measurement bandwidth, Hz.

Substitute this expression of e_n for e_i in equation 1353, and we have:

$$NF = 10 \log_{10} \left(1 + \frac{e_A^2}{4KTBR} \right) \quad (1354)$$

The source resistance for this measurement is critical, since the input thermal noise is proportional to resistance and that means it affects the ratio measurement. For example, with a large source resistance, the contribution of the amplifier becomes insignificant and the noise figure approaches zero.

We can use equation 1354 to relate noise figure NF to noise voltage e_A .

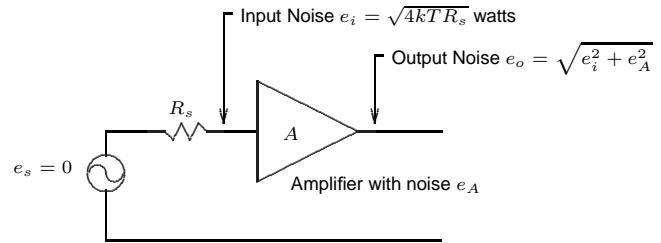


Figure 785: Noisy Amplifier

²⁴¹Notice that the gain of the amplifier is irrelevant.

Example

The BFT82 PNP radio-frequency transistor product specification sheet gives the noise figure as 2.5db (typical). What is the equivalent noise voltage?

Solution

In the absence of any other information, we assume that the noise figure was measured with a source resistance of 50Ω , since that is the standard impedance in radio-frequency designs. At this point, we choose a bandwidth B of 1Hz, which gives us the result in volts per $\sqrt{\text{Hz}}$. Rearrange equation 1354 and substitute values.

$$\begin{aligned}
 \text{NF} &= 10 \log_{10} \left(1 + \frac{e_A^2}{4KTBR} \right) \\
 10^{NF/10} &= 1 + \frac{e_A^2}{4KTBR} \\
 (10^{NF/10} - 1) &= \frac{e_A^2}{4KTBR} \\
 e_A^2 &= (4KTBR) \times (10^{NF/10} - 1) \\
 &= 4 \times (1.38 \times 10^{-23}) \times 290 \times 1 \times 50 \times (10^{2.5/10} - 1) \\
 &= 622 \times 10^{-21} \text{ V}^2/\text{Hz} \\
 e_A &= \sqrt{622 \times 10^{-21}} \\
 &= 789 \times 10^{-12} \text{ V}/\sqrt{\text{Hz}}
 \end{aligned}$$

For example, in a system with a bandwidth of 10MHz, the equivalent noise voltage would be e_A multiplied by the square root of the bandwidth:

$$\begin{aligned}
 e_N &= 789 \times 10^{-12} \times \sqrt{10 \times 10^6} \\
 &= 2.49\mu\text{V}
 \end{aligned}$$

The equivalent circuit, showing the amplifier equivalent noise e_N , is in figure 786.

29.36 BJT Noise Current and Other Considerations

A noise current source contributes to the overall noise of the transistor. This current creates a noise voltage e'_N that appears in series with the voltage noise generator. The magnitude of the noise current depends on the source resistance.

$$e'_N = i_N R_s \quad (1355)$$

Current noise doesn't figure in the previous analysis because the source resistance is low (50Ω). With increasing source resistance, the current noise would be a consideration. As a practical matter it may be preferable to use a low noise JFET when the source resistance is large. JFET current noise is orders of magnitude below BJT current noise.

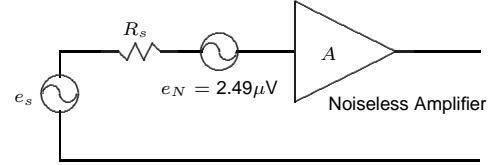


Figure 786: Noisy Amplifier Example

As is the case for many of the properties of a BJT, the noise figure is strongly dependent on the operating conditions such as the collector bias current. References [214] and [100] show how noise in the BJT may be reduced to its absolute minimum.

29.37 Problems

1. Redo the calculations of section 29.24, taking into effect the base-emitter voltage V_{be} .
2. In this development, the mean value of beta was taken as the geometric mean. Repeat the calculations using the arithmetic mean (the average) of the maximum and minimum values of beta.
3. In the circuit of figure 755, there is AC feedback from the collector to the base via the base resistance R_b . To increase the gain, without affecting the bias, it is possible to shunt the AC feedback signal to ground. Resistor R_b is split into two resistors and the centre tap is grounded via a capacitor. When that is done, what is the gain of the stage?
4. Determine the input resistance of the stage, looking into the base of the transistor in figure 755(c). In general, what effect does the feedback have on the input impedance of the stage?
5. Determine the output resistance of the stage, looking back into the collector of the transistor in figure 755(c). In general, what effect does the feedback have on the output impedance of the stage?

29.38 Appendix: Determining the Transistor Parameters

This section brings together all the information necessary to construct the various models of the BJT. These are low-frequency models, which do not take into account the various effects that roll off the gain at higher frequencies.

The classic equation relating the base-emitter voltage to collector current in a BJT transistor is:

$$I_c = I_s e^{\left(\frac{qV_{be}}{nkT}\right)} \quad (1356)$$

where the variables are:

I_c collector current

I_s reverse saturation current. In general, this is a function of temperature $I_s(T)$, but we will treat it as a constant (unless otherwise noted) and write it as I_s .

q electron charge, 1.6×10^{-19} coulomb

V_{be} base-emitter voltage

n the process constant typically ≈ 1

k Boltzmann's constant, 1.38×10^{-23} joule/ $^\circ$ K

T absolute temperature in $^\circ$ Kelvin

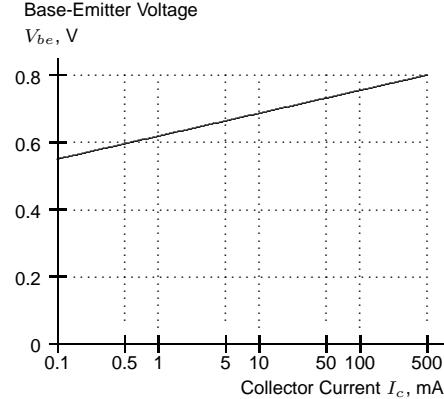


Figure 787: V_{be} vs I_C for the 2N4401

If we have access to values of the base-emitter voltage vs collector-emitter current for a particular transistor, we can solve for the values of n and I_s for that transistor. In this example, we shall use the graph of figure 787, adapted from the Motorola data sheet for the 2N4401. Notice that the relationship between the logarithm of

current on the horizontal axis and base-emitter voltage on the vertical axis is ruler-straight. This indicates that the exponential relationship between I_c and V_{be} holds precisely over a large range of currents.

Calculating the Process Constant n

Taking two points on the $V_{be} - I_c$ characteristic, we can write:

$$\ln \frac{I_{c2}}{I_{c1}} = \frac{q}{nkT} (V_{be2} - V_{be1}) \quad (1357)$$

Two suitable data points from the graph of figure 787 are 0.5 mA and 500.0 mA. Then we have

I_{c1}	0.5 mA	V_{be1}	600 mV
I_{c2}	500.0 mA	V_{be2}	800 mv

Substituting these values in equation 1357, we have:

$$\ln 1000 = \frac{1.6 \times 10^{-19}}{n \times 1.38 \times 10^{-23} \times 298} (0.800 - 0.600)$$

Solving for n ,

$$n = 1.126$$

Solving for the Saturation Current I_s

We are now in a position to substitute a pair of V_{be} , I_c values in equation 1356, along with the newly calculated value of process constant, in order to solve for the saturation current. Choosing 0.6 volts and 0.5mA, we have:

$$\begin{aligned} I_c(T) &= I_s e^{\left(\frac{qV_{be}}{nkT} \right)} \\ 0.5 \times 10^{-3} &= I_s e^{\left(\frac{1.6 \times 10^{-19} \times 0.6}{1.126 \times 1.38 \times 10^{-23} \times 298} \right)} \end{aligned}$$

Solving for I_s , we have

$$I_s = 4.95 \times 10^{-13} \text{ amps}$$

Appendix: The H Parameter Model

Transistor parameters are often quoted by the manufacturer in terms of a particular instance of the *H parameter model* of the transistor²⁴².

Our interest in H Parameters is primarily to extract the parameters for the more useful T model.

The H Parameter Model [256], [257], [250] replaces a BJT amplifier stage with the equivalent circuit shown in figure 788.

²⁴²According to James Early (of the Early effect), who originally suggested this particular model, H parameters were adopted because they provided higher manufacturing yields [248].

The model shows only the behaviour of the alternating currents and voltages that are the signals flowing through the amplifier. It assumes that the BJT is connected to the voltage and current supplies that place it into its active operating region, so that it will function properly as an amplifier for AC signals. These voltage and current supplies are not shown in the model.

The H parameter values for the common-emitter, common-collector and common-base configurations are all different but can be related to each other. Manufacturers usually present the H parameters for the common-emitter configuration, since that is the most popular transistor configuration.

In figure 788, the input terminal **1** is the base of the transistor and the output terminal **3** is the collector. Terminals **2** and **4** are joined together as the emitter of the transistor. Because both the input and output ports have a common connection at the emitter, this is known as the *common emitter* configuration. Each of the four parameters of the model has a following **e** suffix to indicate that the model is for the common emitter configuration.

The *H* in *H Parameter* stands for *Hybrid*, because the model is a mixture: it includes a voltage source in series with a resistance, and a current source in parallel with a resistance that is given as a conductance.

Now we will define and explain each of the four H parameters.

Input Impedance h_{ie}

This the impedance seen between the base and emitter terminals.

$$h_{ie} = \left[\frac{v_{be}}{i_b} \right]_{v_{ce}=0} \quad (1358)$$

Conceptually, the measurement applies an AC voltage between base and emitter with the collector short-circuited to the emitter (for AC signals), and then measures the AC current into the base. The short circuit at the output ensures that the h_{re} generator at the input generates zero volts.

Reverse Transfer Voltage Ratio h_{re}

The reverse transfer voltage ration accounts for the effect that a change in the output voltage of the transistor is slightly coupled back into the input.

$$h_{re} = \left[\frac{v_{be}}{v_{ce}} \right]_{i_b=0} \quad (1359)$$

The measurement applies an AC voltage between collector and emitter with the base open circuited, and then measures the AC voltage between base and emitter. The value of h_{re} is normally very small and so may usually be ignored.

Forward Current Gain h_{fe}

The forward current gain accounts for the collector to base current ratio, which is approximately equal to β .

$$h_{fe} = \left[\frac{i_c}{i_b} \right]_{v_{ce}=0} \quad (1360)$$

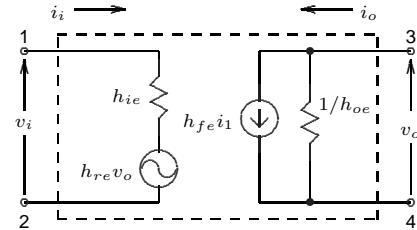


Figure 788: H Parameter Black Box Model

The measurement applies an AC voltage between base and emitter with the collector short-circuited to the emitter for AC signals. The value of h_{fe} is the ratio of the base AC current to collector AC current.

Output Admittance h_{oe}

The forward current gain accounts for the Early effect, whereby there is a slight increase in collector current when the collector-emitter voltage is increased.

$$h_{oe} = \left[\frac{i_c}{v_{ce}} \right]_{i_b=0} \quad (1361)$$

The measurement applies an AC voltage between collector and emitter with the base open circuited, and then measures the resultant AC current into the collector.

Determining the Tee Model Parameters

Now we show how to use H parameter data to determine the T parameters, the value of each component in the equivalent circuit of figure 789.

Current Gain β

The current gain β is the ratio of the total collector current I_c to the total base current I_b . The value of the *forward transfer current* h_{fe} is the ratio of the small signal AC collector current i_c to the small signal AC base current i_b . When the transistor leakage currents are small, as they are in modern transistors, these two are equivalent and:

$$\beta = h_{fe} \quad (1362)$$

Emitter Resistance r_e

As we saw previously, the resistance r_e is created by the slope of the base-emitter diode current-voltage characteristic and is given by:

$$r_e = \frac{nV_t}{I_E} \quad (1363)$$

where

- n is the process constant, previously found to be 1.126 for the 2N4401
- V_t is the thermal voltage, approximately 26mV at 25°C
- I_E is the DC emitter current

so the value of r_e can be obtained without reference to the H parameters, as long as the emitter bias current is known.

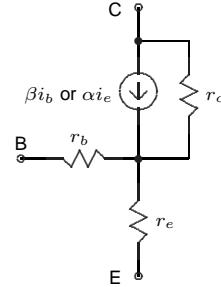


Figure 789: Tee Model

Collector Resistance r_c

Consider first the T Model and that an AC voltage v_{ce} is connected between the collector and emitter terminals. The base terminal is open circuited for this measurement so the collector current generator βi_b is disabled. Then the ratio of collector voltage v_{ce} to collector current i_c is

$$\frac{v_{ce}}{i_c} = r_c + r_e \quad (1364)$$

Now, the value of r_c is at least some 10's of kilohms while r_e is a few 10's of ohms. Then r_e can be neglected and

$$\frac{v_{ce}}{i_c} \approx r_c \quad (1365)$$

For the same measurement on the H parameter model, the h_{fe} current generator is disabled and:

$$\frac{v_{ce}}{i_c} = \frac{1}{h_{oe}} \quad (1366)$$

Consequently we can equate equation 1365 and 1366.

$$r_c \approx \frac{1}{h_{oe}} \quad (1367)$$

Notice that this value of r_c is the collector incremental resistance when the output voltage is between collector and emitter, as it is in the common-emitter case. When the output voltage is between the collector and base, as it is in the common-base amplifier, the collector resistance is beta times as large.

Base Resistance r_b

It's difficult to use H parameter values to calculate the base resistance r_b . In theory it should be possible to do so using the values of r_e , h_{fe} and h_{ie} . In practice, the value of r_b is equal to the difference of two large numbers [258] and a slight inaccuracy in either one results in a major error in the value of r_b . This is an indication that our representation of the behaviour by a fixed resistor is not sufficiently accurate. In fact, the resistance r_b is a function of the base current. An example from [259] is shown in figure 790.

Fortunately, none of this matters in the practice of analysing a low-frequency circuit. The value of base resistance r_b is comparable to the emitter resistance r_e and the current through it is $\beta + 1$ times smaller than the emitter current. For example, at a base current of 0.1mA, figure 790 shows the base resistance as about 100Ω , so the voltage across r_b is $0.1 \text{ mA} \times 100\Omega = 10 \text{ mV}$. If the beta of the transistor is 100, then the emitter current is 10mA and the emitter resistance $0.026/0.010 = 2.6\Omega$. The voltage across this resistance is $0.01\text{A} \times 2.6\Omega = 26\text{mV}$. As a result, the voltage across r_b is small enough to ignore. At high frequencies this resistance does have an important effect but it is also easier to measure.

Here's how it should work using the H parameters:

Referring to the Tee model of the BJT in figure 789, we can write the input voltage v_{be} as:

$$\begin{aligned} v_{be} &= i_b r_b + i_e r_e \\ &= i_b r_b + (\beta + 1) i_b r_e \\ &= i_b [r_b + (\beta + 1) r_e] \end{aligned} \quad (1368)$$

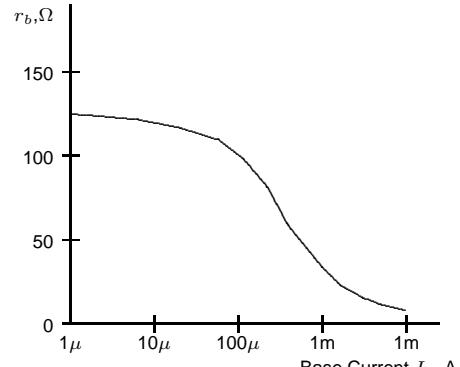


Figure 790: Base Resistance Function

Referring to the H parameter model of figure 788, we can write the input voltage v_{be} as:

$$v_{be} = i_b h_{ie} + v_{ce} h_{oe} \quad (1369)$$

Because h_{oe} is very small (typically 1×10^{-4}), the second term can be neglected. Then

$$v_{be} \approx i_b h_{ie} \quad (1370)$$

Equating equations 1368 and 1370 and solving for h_{ie} , we have:

$$h_{ie} = rb + (\beta + 1)r_e$$

Substitute h_{fe} for β , and rearrange to solve for r_b :

$$r_b = h_{ie} - (h_{fe} + 1)r_e \quad (1371)$$

Consequently, if r_e , h_{fe} and h_{ie} are known values then it should be possible to calculate the value r_b . In practice, a slight error in h_{fe} or h_{ie} makes the result of equation 1371 meaningless.

Example

The 2N4401 transistor H parameters at a temperature of 25°C and collector current of 1mA are: $n = 1.159$, $h_{fe} = 140$, $h_{ie} = 4000\Omega$. Determine the value of r_b for the T model.

Solution

Using equation 1363,

$$r_e = \frac{nV_t}{I_E} = \frac{1.159 \times 0.0256}{0.001} = 29.7\Omega$$

Now apply equation 1371:

$$r_b = h_{ie} - (h_{fe} + 1)r_e = 4000 - (140 + 1)29.7 = -187\Omega$$

Clearly, this approach doesn't work correctly. Many authors assume some nominal resistance around 100Ω , and others ignore it entirely, thereby effectively assuming it is zero.

An alternative calculation for r_e and r_b That Also Does Not Work

There is another way to get to the value of r_b , but to save the overworked reader we'll warn in advance that it doesn't work either. Rather than rely on the DC emitter current to define the emitter resistance r_e (equation 1363) it is possible to calculate r_e from the H parameters, using the definition of the Reverse Transfer Voltage h_{re} . Applying the definition of h_{re} to the T model, we have:

$$\begin{aligned} h_{re} &= \left[\frac{v_{be}}{v_{ce}} \right]_{i_b=0} \\ &= \frac{r_e}{r_e + r_c} \\ &\approx \frac{r_e}{r_c} \end{aligned} \quad (1372)$$

Substitute $1/h_{oe}$ for r_c from equation 1367 and rearrange to solve for r_e :

$$r_e = \frac{h_{re}}{h_{oe}} \quad (1373)$$

This equation generates a value for emitter resistance in the T model, entirely from H parameter values. Does it work any better than our previous method? Let's see.

From the datasheet for the 2N4401 transistor at 25°C and $I_E = 1\text{mA}$, we obtain $h_{re} = 0.7 \times 10^{-4}\text{volts/volt}$, $h_{oe} = 10 \times 10^{-6}\text{V}$, $h_{fe} = 210$ and $h_{ie} = 6000\Omega$. Plugging h_{re} and h_{oe} into equation 1373, we find:

$$r_e = \frac{h_{re}}{h_{oe}} = \frac{0.7 \times 10^{-4}}{10 \times 10^{-6}} = 7\Omega.$$

This is substantially smaller than the value calculated in equation 1363. Nonetheless, we can plug it and our values for h_{fe} and h_{ie} into equation 1371:

$$r_b = h_{ie} - (h_{fe} + 1)r_e = 6000 - (210 + 1) \times 7 = 4530\Omega$$

This value is *way* too large for r_b , which should be under 500Ω , so it doesn't yield correct results either.

Summary: T Model Parameters from H Parameters

This table summarizes the generation of T model Parameters.

Collector Resistance	r_c	$1/h_{oe}$
Current Gain	β	h_{fe}
Base Resistance	r_b	Estimate, 100Ω
Emitter Resistance	r_e	nV_t/I_E

29.39 Appendix: A discussion of Models

All circuit models are necessarily an approximation of reality. This is particularly true for models of the BJT.

There are two possible ways to simplify the diode-with-current generator model of figure 704. These are shown in figure 791.

The alert reader will notice a contradiction in the models of figure 791(a) and 791(b). In figure 791(a) the base current and the collector current flow through the base-emitter diode. In figure 791(b), it is just the base current that flows through the base-emitter diode voltage V_{be} . This is misleading. However, it allows us to show the current generator going directly between collector and emitter terminals, which implies that the voltage between collector and emitter can approach zero volts, which is the case.

The other alternative is to show the base and collector current flowing through V_{be} , which makes it difficult to visualize how the collector-emitter voltage can drop below V_{be} .

In another situation, where we are concerned about the AC gain of the circuit, another model is more appropriate. In that case, as we shall see, it is more accurate to show collector current flowing through the base-emitter diode and less important to consider saturation, so a different model will be appropriate.

29.40 Appendix: Base-Emitter Voltage Function of Temperature

The $V_{be} - I_c$ relationship is usually given by the diode equation²⁴³:

²⁴³More precisely, the current is given by $I_c = I_s(e^{\frac{V_{be}q_e}{nkT}} - 1)$. At any significant value of V_{be} , the -1 term is usually negligible, so we can ignore it.

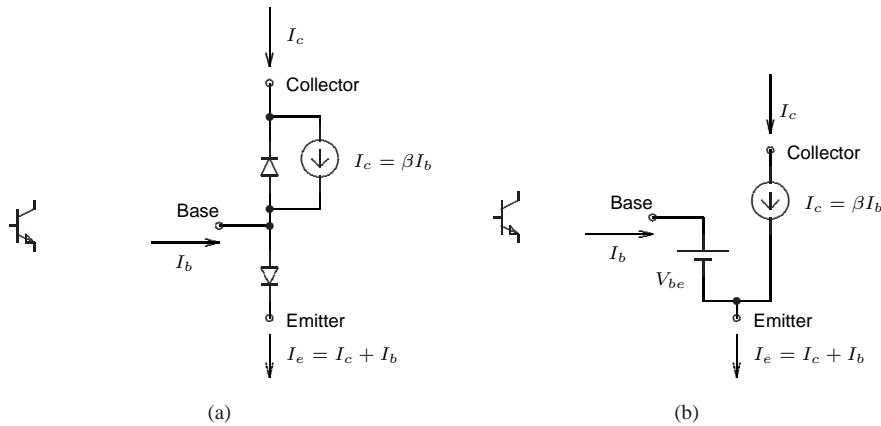


Figure 791: Current Generator Model, Alternate Versions

$$I_c = I_s e \frac{V_{be} q_e}{n k T} \quad (1374)$$

where

- I_s is the *saturation current*, typically equal to 10^{-12} to 10^{-15} amps.
 - n is a constant approximately equal to 1
 - k is Boltzmann's constant, 1.38×10^{-23} joules/kelvin
 - T is the absolute temperature in kelvins
 - q_e is the electron charge, 1.6×10^{-19} coulomb

However, there are a few situations when it is necessary to know the base-emitter voltage collector-current function to greater precision than given by equation 1374. For example, the diode-connected transistor can be used as a temperature sensor the temperature coefficient of base-emitter voltage is known precisely.

Taking the natural logarithm of both sides of the diode equation and solving for V_{be} , we have

$$V_{be}(T) = \frac{nKT}{q} \ln \frac{I_c(T)}{I_s(T)} \quad (1375)$$

In this equation the collector current and saturation current are shown as functions of temperature T to emphasize that they change when the temperature changes. In other words, we cannot treat I_c and I_s as constants and differentiate this equation with respect to temperature T to determine the temperature coefficient.

However, if the base-emitter voltage is known at some reference temperature T_o , the base-emitter voltage at some other temperature T may be calculated by the following [260–262]:

$$V_{be}(T) = V_{G0} \left(1 - \frac{T}{T_c}\right) + V_{beo} \left(\frac{T}{T_c}\right) + \frac{n k T}{a} \ln \left(\frac{T_o}{T}\right) + \frac{k T}{a} \ln \left(\frac{I_c}{I_o}\right) \quad (1376)$$

where the additional terms are

V_{G0} , the bandgap voltage of silicon, typically 1.22V

T_o , a reference temperature in ° Kelvin

V_{beo} , the base-emitter voltage at the reference temperature

I_o , the reference collector current

For example, consider that a diode-connected MTS-102 transistor has a base-emitter voltage of 0.600 volts at a collector current of 0.1 mA and temperature of 20°C. What is the base-emitter voltage at 100°C?

The values of the parameters are:

$$\begin{aligned}
 V_{GO} &= 1.22 \text{ volts} \\
 T &= 273 + 100 \\
 &= 373^\circ\text{K} \\
 T_o &= 273 + 20 \\
 &= 293^\circ\text{K} \\
 V_{beo} &= 0.600 \text{ volts} \\
 n &= 1.014 \\
 k &= 1.38 \times 10^{-23} \\
 q &= 1.6 \times 10^{-19} \\
 I_c &= 0.1 \times 10^{-3} \\
 I_o &= 0.1 \times 10^{-3}
 \end{aligned}$$

Substituting these in equation 1376, we have

$$\begin{aligned}
 V_{be} &= V_{G0} \left(1 - \frac{T}{T_o}\right) + V_{beo} \left(\frac{T}{T_o}\right) + \frac{nkT}{q} \ln \left(\frac{T_o}{T}\right) + \frac{kT}{q} \ln \left(\frac{I_c}{I_o}\right) \\
 &= 1.22 \left(1 - \frac{373}{293}\right) + 0.600 \left(\frac{373}{293}\right) \\
 &\quad + \frac{1.014 \times 1.38 \times 10^{-23} \times 373}{1.6 \times 10^{-19}} \ln \left(\frac{293}{373}\right) \\
 &\quad + \frac{1.38 \times 10^{-23} \times 373}{1.6 \times 10^{-19}} \ln \left(\frac{0.1 \times 10^{-3}}{0.1 \times 10^{-3}}\right) \\
 &= -0.333 + 0.763 - 0.0078 + 0.0 \\
 &= 0.422 \text{ volts}
 \end{aligned}$$

Notice that the third term contributes only a very small amount to the final value. The fourth term has no effect because the collector current is equal to the reference current.

Base-Emitter Voltage Temperature Coefficient

Equation 1376 defines the base-emitter voltage with great precision. We can determine the rate of change of base-emitter voltage (the temperature coefficient of base-emitter voltage) by differentiating with respect to temperature equation 1376.

The first two terms of equation 1376 are linear in T and contribute most of the change in base-emitter voltage with temperature. The third term contributes a small non-linear value, and the fourth term may be ignored if the collector current is constant with temperature.

Consequently, if we assume that the base-emitter junction is driven from a constant current source, the temperature coefficient of the base-emitter voltage may be determined by differentiating the first three terms of equation 1376.

The first two linear coefficient terms of equation 1376 are:

$$V_{be} \simeq V_{G0} - V_{G0} \left(\frac{T}{T_o} \right) + V_{beo} \left(\frac{T}{T_o} \right)$$

then the differential of those terms is:

$$\frac{dV_{be}}{dT} = 0 - \frac{V_{G0}}{T_o} + \frac{V_{beo}}{T_o} \quad (1377)$$

The differential of the third term is a bit more involved: We wish to find

$$\frac{d}{dT} \frac{nkT}{q} \ln \left(\frac{T_o}{T} \right)$$

We begin with the identity

$$\frac{duv}{dx} = u \frac{dv}{dx} + v \frac{du}{dx}$$

and substitute

$$\begin{aligned} u &= \frac{nkT}{q} \\ v &= \ln \left(\frac{T_o}{T} \right) \end{aligned}$$

to obtain

$$\frac{d}{dT} \frac{nkT}{q} \ln \left(\frac{T_o}{T} \right) = \frac{nkT}{q} \times \frac{d}{dT} \ln \frac{T_o}{T} + \ln \frac{T_o}{T} \times \frac{d}{dT} \frac{nkT}{q} \quad (1378)$$

Using the identity

$$\frac{d}{dx} \ln u = \frac{1}{u} \frac{du}{dx}$$

we have that

$$\begin{aligned} \frac{d}{dT} \ln \frac{T_o}{T} &= \frac{1}{T_o/T} \times \frac{d}{dT} \frac{T_o}{T} \\ &= \frac{T}{T_o} \times -\frac{T_o}{T^2} \\ &= -\frac{1}{T} \end{aligned}$$

We also have that

$$\frac{d}{dT} \frac{nkT}{q} = \frac{nk}{q}$$

Finally, substituting in equation 1378, the differential of the third term is given by

$$-\frac{nkT}{q} \times \frac{1}{T} + \ln \frac{T_o}{T} \times \frac{nk}{q} = \frac{nk}{q} \left(\ln \frac{T_o}{T} - 1 \right)$$

Putting this result together with the linear coefficient terms of equation 1377, we have the complete expression for the rate of change of V_{be} with temperature:

$$\frac{dV_{be}}{dT} = -\frac{V_{G0}}{T_o} + \frac{V_{beo}}{T_o} + \frac{nk}{q} \left(\ln \frac{T_o}{T} - 1 \right) \quad (1379)$$

Exercise

The data sheet for the MTS102 temperature sensor transistor gives V_{beo} as 0.595 volts at 25°C. Determine the temperature coefficient of V_{be} over the range of -40°C to +50°C.

Solution

Substituting $V_{G0}=1.22\text{V}$, $V_{beo}=0.595$, $T_o=298^\circ$ and $T=298^\circ$ Kelvin in equation 1379, we have

$$\begin{aligned} \frac{dV_{be}}{dT} &= -\frac{1.22}{298} + \frac{0.595}{298} + \frac{1.014 \times 1.38 \times 10^{-23}}{1.6 \times 10^{-19}} \left(\ln \frac{298}{298} - 1 \right) \\ &= -4.094 \times 10^{-3} + 1.997 \times 10^{-3} - 0.087 \times 10^{-3} \\ &= -2.184 \text{ mV}/^\circ\text{C} \end{aligned}$$

The value from the data sheet is -2.265 mV/°C, which is within 4% of the calculated value.

Plugging temperature values of -40°C and +50°C into equation 1379, we can calculate the transistor temperature coefficients at the extremes of operation:

Temperature	Coefficient of Temperature	Error related to coefficient at 25°C
-40	-2.163 mV/°C	-0.1%
+25	-2.184 mV/°C	+0.0%
+50	-2.191 mV/°C	+0.3%

Evidently the temperature coefficient of base-emitter voltage may be relied upon to be essentially constant over this temperature range.

Temperature and Collector Current

The previous analysis relates base-emitter voltage to changes in temperature if the collector current is maintained constant. Now we'll look at the effect of temperature on collector current if the base-emitter voltage is provided by a constant-voltage source V_{BB} . The test circuit is shown in figure 792.

We start with equation 1376. The left side of the equation is fixed at V_{BB} volts.

$$V_{BB} = \left\{ V_{G0} \left(1 - \frac{T}{T_o} \right) + V_{beo} \left(\frac{T}{T_o} \right) + \frac{nkT}{q} \ln \left(\frac{T_o}{T} \right) \right\} + \frac{kT}{q} \ln \left(\frac{I_c}{I_o} \right) \quad (1380)$$

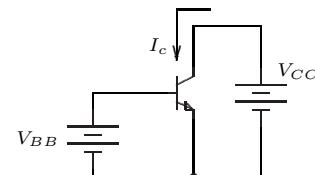


Figure 792: Effect of Temperature on Collector Current

Put

$$\frac{kT}{q} = V_t \quad (1381)$$

The rate of change of V_t with temperature is k/q , which is a small number. Consequently, V_t is only weakly dependent on temperature and to simplify the analysis we'll take it as a constant.

Consider that the temperature changes by 1°C . We previously showed that the term within the braces in equation 1380 will change by -2.2mV . Since the left side of the equation is fixed, the last term on the right side of the equation must change by $+2.2\text{mV}$. Writing this as an equation and substituting for V_t from equation 1381, we have:

$$V_t \ln \left(\frac{I_c}{I_o} \right) = 2.2\text{mV} \quad (1382)$$

where I_c is the collector current at the new temperature and I_o is the original collector current. The value of V_t at 25°C is 26mV . Substitute this for V_t , and take the antilog:

$$\begin{aligned} \frac{I_c}{I_o} &= e^{\frac{0.0022}{0.026}} \\ &= e^{0.0846} \\ &= 1.085 \end{aligned} \quad (1383)$$

That is, the current increases by about 8% with each degree C increase in temperature. In general:

$$I_c = I_o (1.085^{\Delta T}) \quad (1384)$$

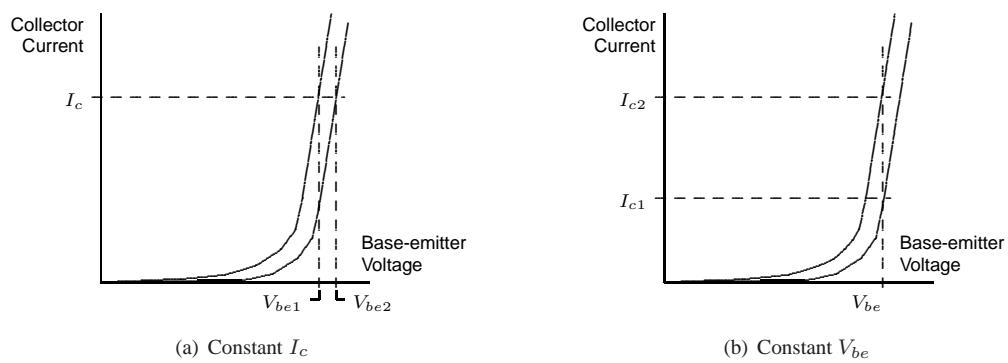
For example, if the temperature increases by 10°C , what is the change in collector current? Using equation 1384,

$$\begin{aligned} I_c &= I_o (1.085^{\Delta T}) \\ &= I_o (1.085^{10}) \\ &= 2.26 I_o \end{aligned}$$

That is, the current approximately doubles with each 10°C increase in temperature.

Summary

The effect of changing the temperature is summarized in figure 793. In figure 793(a), the current is held constant and the effect is to change the base-emitter voltage by $-2.2\text{mV}/^\circ\text{C}$. In figure 793(b), the base-emitter voltage is held constant and the effect is to double the collector current with each 10°C increase in temperature.

Figure 793: Temperature Change, Effect on V_{be} and I_c

29.41 Appendix: BJT Datasheet

This section shows a four-page excerpt from the ON Semiconductor datasheet for the 2N4401. There is a lot of information here, which may or may not be relevant to a given application. We will give an overview of the important specifications, skipping the ones that are used less frequently.

The upper right corner shows the case style and the pinout for the device. This is important. It is risky to make assumptions about the pin connections – there are many different configurations.

The same information indicates the case style, which is basically a Case 29-04, Style 1 in ONsemi parlance. It is similar to the TO-92 case, which is a JEDEC standard. A dimensioned drawing of this particular case is available from the vendor, important when doing the physical layout of the circuit.

The maximum voltage ratings are shown in the upper left corner. (See the table below.) The other ratings list the maximum collector current and power dissipation. (For a discussion of the power dissipation specifications, see section 28 on page 783.)

The maximum collector current and maximum junction temperature must not be exceeded. The *Thermal Characteristics* section shows the values of thermal resistance that are essential for designing a heatsink.




General Purpose Transistors
NPN Silicon

MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector-Emitter Voltage	V_{CEO}	40	Vdc
Collector-Base Voltage	V_{CBO}	60	Vdc
Emitter-Base Voltage	V_{EBO}	6.0	Vdc
Collector Current — Continuous	I_C	600	mAdc
Total Device Dissipation @ $T_A = 25^\circ\text{C}$ Derate above 25°C	P_D	625 5.0	mW mW/ $^\circ\text{C}$
Total Device Dissipation @ $T_C = 25^\circ\text{C}$ Derate above 25°C	P_D	1.5 12	Watts mW/ $^\circ\text{C}$
Operating and Storage Junction Temperature Range	T_J, T_{Stg}	-55 to +150	$^\circ\text{C}$

THERMAL CHARACTERISTICS

Characteristic	Symbol	Max	Unit
Thermal Resistance, Junction to Ambient	R_{JA}	200	$^\circ\text{C/W}$
Thermal Resistance, Junction to Case	R_{JC}	83.3	$^\circ\text{C/W}$

ELECTRICAL CHARACTERISTICS ($T_A = 25^\circ\text{C}$ unless otherwise noted)

Characteristic	Symbol	Min	Max	Unit
OFF CHARACTERISTICS				
Collector-Emitter Breakdown Voltage ⁽¹⁾ ($I_C = 1.0 \text{ mA dc}, I_B = 0$)	$V_{(BR)CEO}$	40	—	Vdc
Collector-Base Breakdown Voltage ($I_C = 0.1 \text{ mA dc}, I_E = 0$)	$V_{(BR)CBO}$	60	—	Vdc
Emitter-Base Breakdown Voltage ($I_E = 0.1 \text{ mA dc}, I_C = 0$)	$V_{(BR)EBO}$	6.0	—	Vdc
Base Cutoff Current ($V_{CE} = 35 \text{ Vdc}, V_{EB} = 0.4 \text{ Vdc}$)	I_{BEV}	—	0.1	$\mu\text{A dc}$
Collector Cutoff Current ($V_{CE} = 35 \text{ Vdc}, V_{EB} = 0.4 \text{ Vdc}$)	I_{CEX}	—	0.1	$\mu\text{A dc}$

1. Pulse Test: Pulse Width $\leq 300 \mu\text{s}$, Duty Cycle $\leq 2.0\%$.

Preferred devices are ON Semiconductor recommended choices for future use and best overall value.

Figure 794: Datasheet Cover Page

$V_{(BR)CEO}$	Breakdown voltage, collector to emitter, base open circuited. This is essentially the maximum supply voltage V_{CC} for the transistor.
$V_{(BR)CBO}$	Breakdown voltage, collector to base, emitter open circuited. This is the maximum reverse voltage across the collector-base diode.
$V_{(BR)EBO}$	Breakdown voltage, emitter to base, collector open circuited. This is the maximum reverse voltage across the emitter-base diode. Notice that it is much less than the other breakdown voltages.

Small Signal Characteristics

The ON characteristics describe the current gain (h_{FE}) in the active region, and the saturation voltages V_{cesat} and V_{besat} in the saturation region. Figure 16 of the datasheet presents a more detailed version of this information.

The datasheet distinguishes between the DC gain h_{FE} and the AC gain h_{fe} of the transistor. In practice both values of current gain are similar so we have treated them as identical and referred to them throughout the section on the BJT as *beta*.

Under *Small Signal Characteristics*, the important specifications are the Current-Gain Bandwidth Product f_T , the collector-base capacitance C_{cb} and emitter base capacitance C_{eb} . These parameters are important in modelling the high frequency behaviour of the transistor.

2N4401

ELECTRICAL CHARACTERISTICS ($T_A = 25^\circ\text{C}$ unless otherwise noted) (Continued)

Characteristic	Symbol	Min	Max	Unit
ON CHARACTERISTICS(1)				
DC Current Gain ($I_C = 0.1 \text{ mA dc}, V_{CE} = 1.0 \text{ Vdc}$) ($I_C = 1.0 \text{ mA dc}, V_{CE} = 1.0 \text{ Vdc}$) ($I_C = 10 \text{ mA dc}, V_{CE} = 1.0 \text{ Vdc}$) ($I_C = 150 \text{ mA dc}, V_{CE} = 1.0 \text{ Vdc}$) ($I_C = 500 \text{ mA dc}, V_{CE} = 2.0 \text{ Vdc}$)	h_{FE}	20 40 80 100 40	— — — 300 —	—
Collector-Emitter Saturation Voltage ($I_C = 150 \text{ mA dc}, I_B = 15 \text{ mA dc}$) ($I_C = 500 \text{ mA dc}, I_B = 50 \text{ mA dc}$)	$V_{CE(\text{sat})}$	— —	0.4 0.75	Vdc
Base-Emitter Saturation Voltage ($I_C = 150 \text{ mA dc}, I_B = 15 \text{ mA dc}$) ($I_C = 500 \text{ mA dc}, I_B = 50 \text{ mA dc}$)	$V_{BE(\text{sat})}$	0.75 —	0.95 1.2	Vdc
SMALL-SIGNAL CHARACTERISTICS				
Current-Gain Bandwidth Product ($I_C = 20 \text{ mA dc}, V_{CE} = 10 \text{ Vdc}, f = 100 \text{ MHz}$)	f_T	250	—	MHz
Collector-Base Capacitance ($V_{CB} = 5.0 \text{ Vdc}, I_E = 0, f = 1.0 \text{ MHz}$)	C_{cb}	—	6.5	pF
Emitter-Base Capacitance ($V_{EB} = 0.5 \text{ Vdc}, I_C = 0, f = 1.0 \text{ MHz}$)	C_{eb}	—	30	pF
Input Impedance ($I_C = 1.0 \text{ mA dc}, V_{CE} = 10 \text{ Vdc}, f = 1.0 \text{ kHz}$)	h_{ie}	1.0	15	kohms
Voltage Feedback Ratio ($I_C = 1.0 \text{ mA dc}, V_{CE} = 10 \text{ Vdc}, f = 1.0 \text{ kHz}$)	h_{re}	0.1	8.0	$\times 10^{-4}$
Small-Signal Current Gain ($I_C = 1.0 \text{ mA dc}, V_{CE} = 10 \text{ Vdc}, f = 1.0 \text{ kHz}$)	h_{fe}	40	500	—
Output Admittance ($I_C = 1.0 \text{ mA dc}, V_{CE} = 10 \text{ Vdc}, f = 1.0 \text{ kHz}$)	h_{oe}	1.0	30	μmhos
SWITCHING CHARACTERISTICS				
Delay Time ($V_{CC} = 30 \text{ Vdc}, V_{BE} = 2.0 \text{ Vdc}, I_C = 150 \text{ mA dc}, I_{B1} = 15 \text{ mA dc}$)	t_d	—	15	ns
Rise Time ($V_{CC} = 30 \text{ Vdc}, I_C = 150 \text{ mA dc}, I_{B1} = I_{B2} = 15 \text{ mA dc}$)	t_r	—	20	ns
Storage Time ($V_{CC} = 30 \text{ Vdc}, I_C = 150 \text{ mA dc}, I_{B1} = I_{B2} = 15 \text{ mA dc}$)	t_s	—	225	ns
Fall Time ($V_{CC} = 30 \text{ Vdc}, I_C = 150 \text{ mA dc}, I_{B1} = I_{B2} = 15 \text{ mA dc}$)	t_f	—	30	ns

1. Pulse Test: Pulse Width $\leq 300 \text{ ns}$, Duty Cycle $\leq 2.0\%$.

SWITCHING TIME EQUIVALENT TEST CIRCUITS

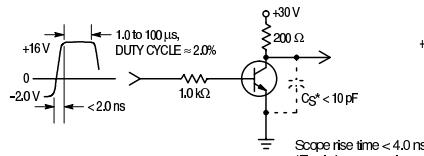


Figure 1. Turn-On Time

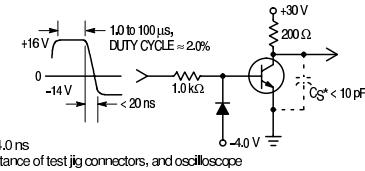


Figure 2. Turn-Off Time

H Parameters

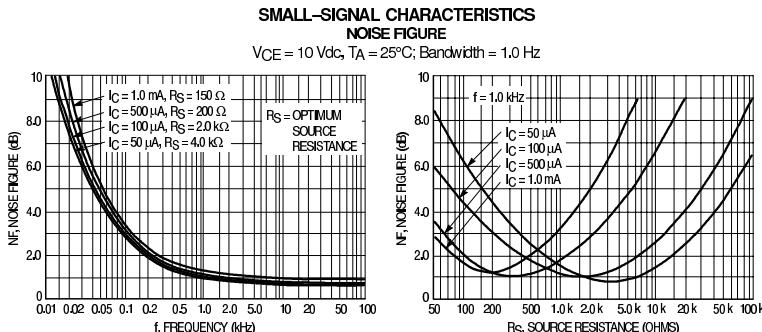
Figures 11 through 14 of the datasheet shows the four H parameters h_{fe} , h_{ie} , h_{re} and h_{oe} plotted against collector current. Earlier in the datasheet under Small Signal Characteristics, a single figure is given for each of these parameters. However, given the substantial variation in the parameters with collector current the plots shown here are much more useful.

Section 29.38 discusses the H parameter model. Section 29.38 shows how the H parameters may be converted to the T parameter model, used throughout Section 29.

DC Current Gain

Figure 15 of the datasheet (next page) shows how the current gain varies with collector current and temperature.

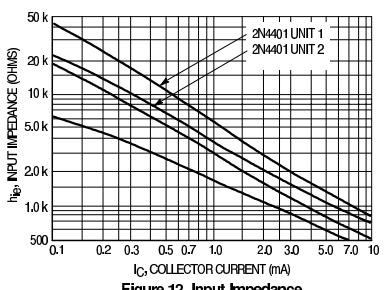
2N4401



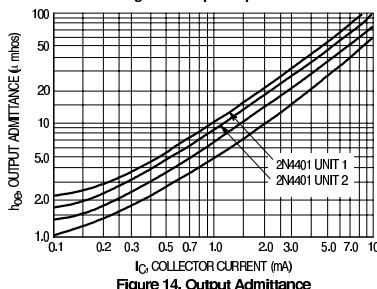
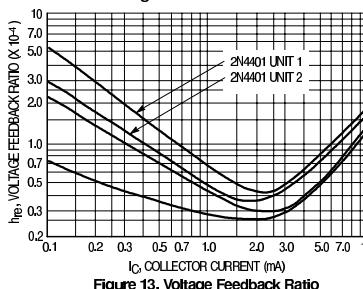
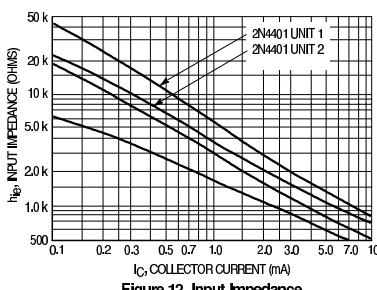
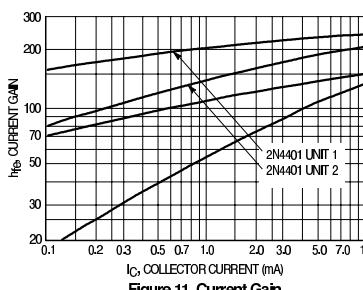
h PARAMETERS

V_{CE} = 10 Vdc, f = 1.0 kHz, T_A = 25°C

This group of graphs illustrates the relationship between h_{fe} and other "h" parameters for this series of transistors. To obtain these curves, a high-gain and a low-gain unit were



selected from the 2N4401 lines, and the same units were used to develop the correspondingly numbered curves on each graph.



Collector Saturation Region

Figure 16 of the datasheet shows how the transistor behaves in its saturation region. For example, taking the left-most trace, the collector-emitter voltage is about 0.1 volts for a collector current of 1mA and a base current of 0.1 mA. These values correspond to a forced beta of 10.

Figure 17 is another view of the same information. The lowest trace shows how the collector saturation voltage V_{cesat} varies with collector current at a forced beta of 10. Over the collector current range 0.1mA to 50mA, the saturation voltage is about 0.1V.

The upper trace in figure 17 shows the base-emitter voltage V_{be} under saturation conditions. At low currents the predominant effect is the exponential characteristic of the base-emitter diode and the trace is linear. Above 50mA collector current (5mA base current), the base-emitter voltage diverges upward from the exponential characteristic, the effect of the intrinsic resistance r_b in the base-emitter junction.

The middle trace shows the base-emitter voltage V_{be} vs collector current. The straight line indicates that it is a very predictable exponential characteristic according to the diode equation. This trace may be used to derive the saturation current I_s and process constant n for the transistor (section 29.38).

Figure 18 shows the temperature coefficients, in millivolts/ $^{\circ}\text{C}$, for the collector-emitter saturation voltage and base-emitter voltage. The saturation voltage is essentially negligible over a wide range of currents. The base-emitter voltage temperature coefficient depends on the collector current. A temperature measurement circuit must ensure that the current through the base-emitter junction does not change with temperature.

2N4401

STATIC CHARACTERISTICS

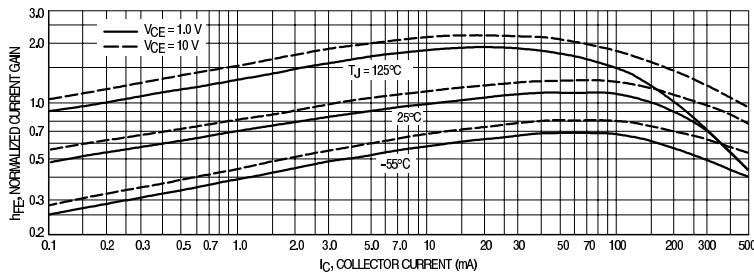


Figure 15. DC Current Gain

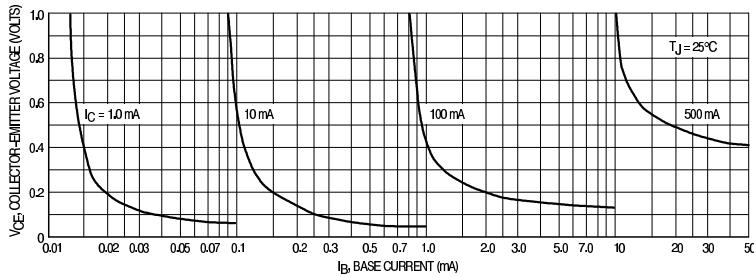


Figure 16. Collector Saturation Region

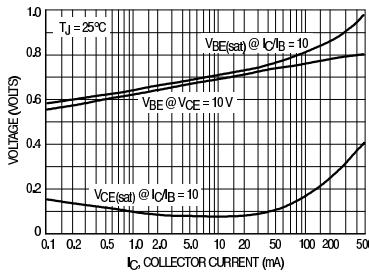


Figure 17. "On" Voltages

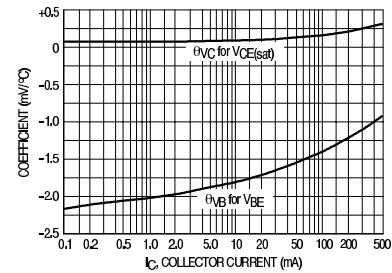


Figure 18. Temperature Coefficients

30 BJT-Compounds

In this section, we look at some common functional blocks using the BJT device. Each functional block consists of an arrangement of a small number of transistors to form a useful subcircuit such as a gain block or constant current sink.

Since this is a book about board-level design, the useful functional blocks are limited to those which can be and are implemented with discrete transistors or transistor arrays. There are other functional blocks that are used only within integrated circuits – because they rely on particular properties of the integrated circuit environment such as the matching of devices. An introductory level description of these circuits is in [263], [252].

For example, many of the same building blocks – darlington pair, differential amplifier, current source – are used in both audio power amplifiers [264] and BJT integrated circuit operational amplifiers [84]. In this work, we focus on the discrete component version of those building blocks.

All of these BJT building blocks have their counterpart in JFET and MOSFET implementations. Once the BJT version is understood, it is usually a simple step to move to the JFET or MOSFET version.

30.1 The Darlington Pair

If the output current of one transistor is used as the input current for a second, the overall current gain is the product of the individual current gains²⁴⁴, that is:

$$\beta = \beta_1 \cdot \beta_2 \quad (1385)$$

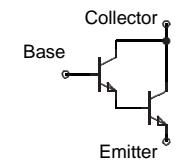
The two transistors can be connected as shown in figure 798 so that 3 terminals are brought to the outside world. Then this device behaves as a transistor with a very large value of current gain, and is known as a *Darlington* pair, after its inventor [265].

The large current gain is attractive, especially when a medium to large current must be controlled, but there are some disadvantages. In general, Darlington devices have a larger saturation voltage, since it must be greater than one V_{be} drop. This increases the power dissipation when the device is saturated (in the ON state). Darlingtons tend to be slower than single transistors. Finally, there is a potential problem in switching off the second transistor when it is hot.

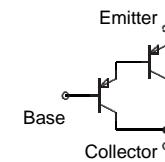
As an example, compare specifications for a conventional power transistor, TIP41A and a Darlington power transistor TIP100.

	Transistor, TIP41A	Darlington, TIP100
Current Gain β (minimum)	30	1000
Saturation Voltage V_{CESat}	0.2V at $I_C = 2A$	1.0V at $I_C = 2A$
F_T	3MHz	4MHz
F_{3db}	$3\text{MHz}/30=100\text{kHz}$	$4\text{MHz}/1000=4\text{kHz}$

The transition frequency F_T (the frequency at which the current gain is unity) is similar in the transistor and Darlington. However, the current gain starts its descent at frequency F_{3db} , and this begins at a much lower frequency for the Darlington transistor.

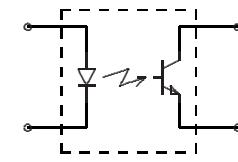


(a) NPN Darlington

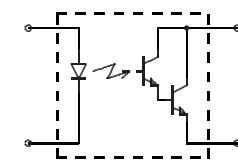


(b) PNP Darlington

Figure 798: Darlington Pair



(a) Transistor



(b) Darlington

Figure 799: Optocouplers

²⁴⁴Assuming the usual approximation that $\beta + 1$ is approximately equal to β .

This effect also shows up in optocouplers and opto-interrupters (section 27.6, page 778), which are available with transistor and Darlington photo-transistors, as shown in figure 799(b) above. The opto-isolators that use a Darlington phototransistor have a much larger current transfer ratio, but a much lower bandwidth:

	Transistor, 4N25	Darlington, 4N32
Current Transfer Ratio CTR	0.2	100 (minimum)
BW (F_{3db})	300kHz	30kHz

The Collector Leakage Current Problem

Darlington transistors are often called upon to dissipate considerable power. This can lead to a problem in shutting off the output device.

Consider the NPN Darlington of figure 800(a), which is composed of a *driver* transistor and an *output* transistor. Recall that the collector-base junction is a diode which is reverse biased under normal operation (not saturated). The leakage current of this diode doubles for each 10°C increase in the junction temperature (section 29.13, page 825).

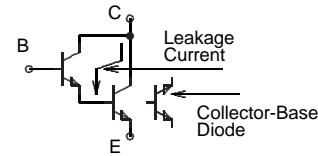
As long as the Darlington is conducting, this is not a major problem. Now suppose that the driver transistor shuts off by reverse biasing its base-emitter diode, and that the output transistor is hot enough to have significant leakage current through the collector-base diode. This leakage current effectively becomes a base current for the output transistor, and it continues to conduct. This is a major problem – the driver transistor has lost control of the output current.

Fortunately, this problem can be fixed by providing a resistive path for the leakage current, such that the leakage current creates a voltage drop across the output base-emitter diode that is insufficient to cause it to conduct. For this reason, commercial Darlington transistors are provided with internal base-emitter resistors, as shown in figure 800(b). (The resistor values given are for the popular TIP100 series of Darlington transistors.)

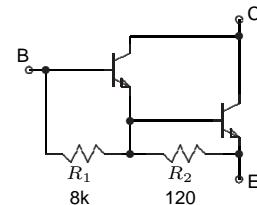
These same resistors (referred to as *flushout* resistors in [266]) improve the frequency response of the Darlington pair. In this case, they provide a relatively low resistance discharge path for charge carriers in the base region, and this improves the speed of the device. The value of F_T given for the Darlington transistor on page 30.1 would be much lower without these resistors.

Applications

One very common application is shown in figure 801. In this circuit, the BJT emitter follower is replaced by a Darlington transistor. The equations for the Darlington follower are obtained by substituting $(\beta + 1) \cdot \beta_2 \approx \beta_1 \cdot \beta_2$ for the value of β in the emitter follower equations of section 29.29, page 876). That is, the Darlington may be treated simply as a transistor with $\beta = \beta_1 \cdot \beta_2$.



(a) Leakage Problem



(b) Solution

Figure 800: Darlington Leakage

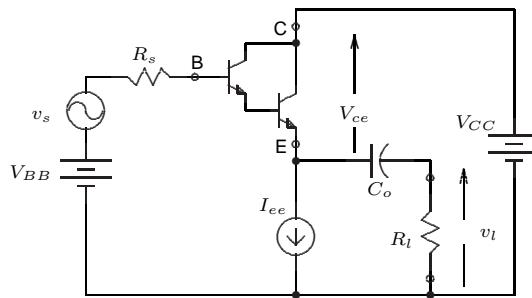


Figure 801: Darlington Follower

The Darlington transistor finds application in conjunction with other compounds, and we'll see examples of that in connection with the differential amplifier and totem-pole output stage²⁴⁵.

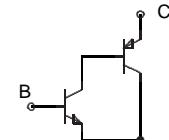
Where the requirement is simply for higher current gain, the power MOSFET transistor may be a better choice. In general, power MOSFETs require less ancillary circuitry to make them function correctly, their current gain is essentially infinite²⁴⁶, they work to a higher frequency, and they are less subject to certain types of failure. For example, the power MOSFET has replaced the Darlington transistor as the preferred device for power control in switching regulators.

Duncan [266] and Self [264] differ on the preferred device for the power output stage – BJT or MOSFET – for audio power amplifiers. However, they agree that Darlington power transistors (like the TIP100) are not suitable for high quality audio power amplifier designs. On the other hand the Darlington transistor configuration, constructed from separate transistors and with tailored flushout resistors, is very common in audio power amplifiers.

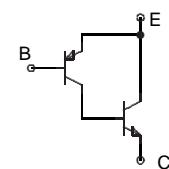
30.2 The Sziklai Pair

An interesting relative of the Darlington pair is the Sziklai pair, also known as the Complementary Feedback Pair (CFP), [264] shown in figure 802.

The Sziklai pair configuration contains an NPN and a PNP transistor connected so that the polarity of the driver transistor determines the overall polarity of the pair. This arrangement can be regarded as a current amplifier with 100% feedback from the output transistor to the emitter of the driver transistor. The driver transistor uses the error voltage across its base-emitter junction to drive the output transistor in such a direction that the error is reduced to zero. Again, a flushout resistor is desirable across the base-emitter junction of the output transistor.

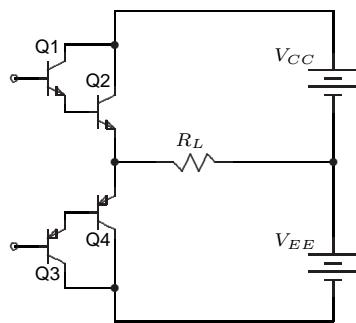


(a) NPN

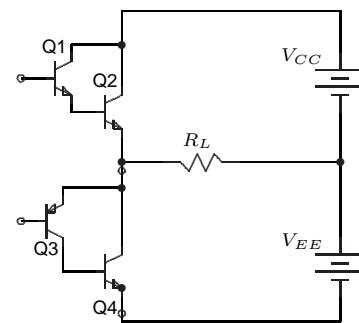


(b) PNP

Figure 802: Sziklai Pair



(a) Complementary Darlington



(b) Synthesized Complement

Figure 803: Complementary Pair

A desirable configuration of output transistors for an audio power amplifier is shown in figure 803(a). The upper and lower Darlingtons alternate conducting half-cycles of AC to the load resistance R_L .

²⁴⁵The term *totem-pole* refers to a configuration of two transistors in series, usually drawn as one above the other. This configuration is often found in the output stage of an amplifier.

²⁴⁶The gate (input) terminal of a power MOSFET is an open circuit at DC, but presents a large capacitance, in the order of hundreds of picofarads, to AC signals. This capacitance must be charged and discharged by the driving signal. Consequently, the average drive signal current is zero, but the peak drive current may be significant.

The Q1-Q2 Darlington transistor has an effective base-emitter voltage of two BJT base-emitter drops (about 1.4 volts), while the Sziklai pair has only one base-emitter voltage drop. This asymmetry contributes some distortion [264], but it can be reduced with the simple addition of a silicon diode in the emitter lead of the Sziklai driver Q3 [264].

Recall that oscillation in a negative feedback system is most likely to occur for 100% feedback (unity closed-loop gain). The Sziklai pair is a unity-closed-loop-gain negative feedback system, and consequently it is more likely to oscillate than a Darlington pair – especially if the output transistor has a lower bandwidth than the driver transistor.

Application

At one point in the history of audio power amplifier design, NPN power transistors were available but their PNP complements were not. Instead, a full complementary NPN-PNP pair was synthesized using two NPN transistors, as shown in figure 803(b). The Sziklai pair Q3-Q4 acts effectively as a PNP power transistor, but requires only a low-power PNP transistor and uses an NPN power transistor. The arrangement of figure 803(b) is known as a *quasi-complementary pair*.

This same technique is used in silicon-based integrated circuits. NPN transistors are relatively easy to fabricate. PNP transistors tend to have low current gain and poor power-handling capability, so they are paired with an NPN transistor in a Sziklai pair to formulate an effective PNP transistor with reasonable specifications.

Gain Equation

Figure 804 shows the circuit of a Sziklai pair used as an emitter follower. As in the case of the Darlington, the current gain of the composite transistor is approximately equal to the product of the two current gains. Consequently, we'd expect that the emitter follower equations of section 29.29, page 876 will apply, where $(\beta + 1) \cdot \beta_2 \approx \beta_1 \cdot \beta_2$ replaces β .

For example, the input resistance of the single-transistor emitter follower is given by

$$\begin{aligned} r_i &= (\beta + 1)(r_e + R_L) \\ &\approx \beta R_L \text{ if } \beta \gg 1, R_L \gg r_e \end{aligned} \quad (1386)$$

That is, the load resistance appears at the input multiplied by beta.

The input resistance of the Sziklai pair is given by

$$\begin{aligned} r_i &= (\beta_1 + 1)(r_{e1}\beta_2\alpha_1 R_L) \\ &\approx \beta_1\beta_2 R_L \text{ if } \beta \gg 1, \alpha_1 \approx 1, R_L \gg r_{e1} \end{aligned} \quad (1387)$$

where the variables are:

- r_i AC input resistance, ohms
- β_1 current gain of the driver transistor Q_1
- β_2 current gain of the output transistor Q_2
- r_{e1} incremental emitter resistance of driver transistor Q_1
- α_1 ratio of collector to emitter current in Q_1
- R_L load resistance, ohms

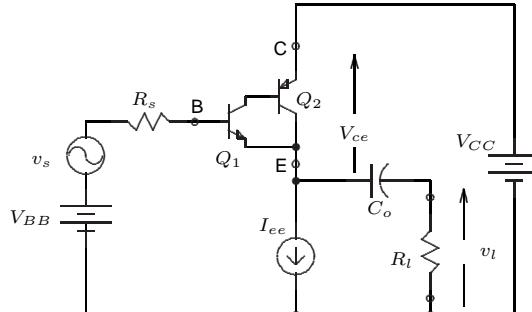


Figure 804: Sziklai Follower

That is, under the usual assumptions, the load resistance appears at the input multiplied by the product of the individual transistor current gains.

30.3 The Current Source

The *current source*²⁴⁷ has a number of important applications in electronic circuits:

- The output current is used to establish the operating point of some other section of the circuit. For example, a current source is used to establish the operating current in a differential pair (section 30.14, page 935).
- Constant current charges a capacitor so that the rate of change of capacitor voltage is a constant. This effect is useful in timing circuits, where there is then a direct relationship between voltage and time.
- The constant current is used to create a constant voltage across a resistor, or a very constant voltage across a zener diode. For example, a constant current can be used in conjunction with a resistor to create a constant-voltage level shift in an amplifier.

Figure 805 shows an example. In this case, the input signal e_i is a 1 volt peak-peak AC voltage sitting on a DC voltage of 3 volts. The resistor-current sink combination creates a fixed voltage shift of 2 volts. The output voltage e_o is the same AC voltage as the input, but shifted to a DC voltage of 1 volt.

- A constant current device can source a direct current while appearing as an AC open circuit. For example, in a common-emitter BJT amplifier circuit, the voltage gain is the ratio of collector resistance to emitter resistance. If the collector resistance is replaced by a constant-current source, then its incremental resistance and the voltage gain become very large. An example is shown in section 30.15 on page 943.

When a BJT is operated in its active region, its collector current is relatively independent of collector voltage. It's useful to regard the collector circuit as a constant-current source in parallel with a large internal resistance, typically $100\text{k}\Omega$. In many circuits, this internal resistance is large enough to ignore, and the transistor may be regarded as a constant-current source. The value of that current is set by other elements in the circuit – the base-emitter voltage or the emitter current for example.

30.4 Discrete Component Sink

A typical discrete component current sink is shown in figure 806.

Starting at the left side, voltage source E_1 and resistor R_1 establish a current through the diode and zener. Both these are constant-voltage devices, so the precise level of current is not critical, but it must be sufficient to bias these devices into constant-voltage operation, and to supply the base current of the transistor.

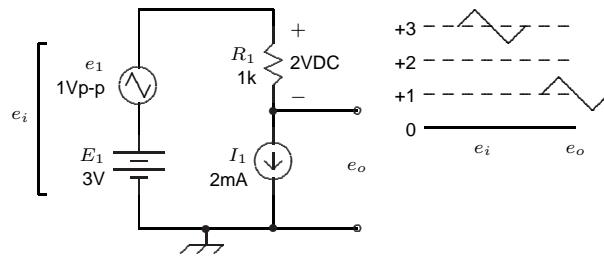


Figure 805: Voltage Shift Circuit

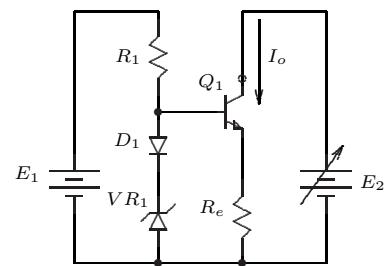


Figure 806: Discrete Current Sink

²⁴⁷A current source produces current from its output terminal. A current sink absorbs current into its output terminal. This distinction is not always made. Source is the more common term, and it is often applied to a sink.

The diode voltage and base-emitter voltage of the transistor cancel, so the voltage across the emitter resistor R_e is approximately equal to the zener voltage. If the transistor is in its active region, then this voltage establishes an emitter current in the transistor. The collector current is alpha times the emitter current.

Now suppose that the collector supply voltage E_2 is increased from zero. The transistor enters its active region when the collector-emitter voltage exceeds the saturation voltage (typically 0.2 volts for a small signal device). Once that happens, the collector current is constant, irrespective of further increases in collector supply voltage. Consequently, the collector supply voltage must be equal to or greater than the emitter (and zener) voltage plus the transistor saturation voltage.

Both the diode voltage and the base-emitter voltage of the transistor change at the same rate (-2.2mV/°C). Consequently, a change in the base-emitter voltage of the transistor is cancelled by the same change in the diode voltage, and this improves the temperature stability of the circuit.

Design Equations

Applying KVL to the centre circuit loop, we have:

$$+V_z + V_d - V_{be} - V_{Re} = 0 \quad (1388)$$

If the diode voltage drop V_d is equal to the transistor base-emitter voltage V_{be} , then

$$V_{Re} = +V_z \quad (1389)$$

Consequently, the emitter current is:

$$\begin{aligned} I_E &= \frac{V_{Re}}{R_e} \\ &= \frac{V_z}{R_e} \end{aligned} \quad (1390)$$

The collector current is

$$\begin{aligned} I_C &= \alpha I_E \\ &= \alpha \frac{V_z}{R_e} \\ &\approx \frac{V_z}{R_e} \end{aligned} \quad (1391)$$

It turns out that the temperature coefficient of a zener diode is nearly zero for zener diodes of about 6.2 volts. If temperature stability is a concern, then 6.2 volts is a suitable choice for V_z . Then the value of emitter resistance is then chosen to suit the output current. Notice that the zener voltage sets the emitter voltage and the collector voltage must be greater than this for the transistor to be in its active region.

Performance

This is rather a complicated circuit, with 4 components in the core current generator. One of them is a zener diode, which requires millamps of current to operate and limits the output voltage range.

An ideal current source should have infinite output resistance. The emitter resistor R_e provides negative feedback that raises the output resistance significantly²⁴⁸. As shown in an appendix (equation 1418, page 928),

²⁴⁸The process of reducing gain to obtain lower distortion or some other desirable property – that is, applying negative feedback – is known as *degeneration*. The emitter resistor is then referred to as an *emitter degeneration* resistor. In this context the negative feedback is usually local to a particular circuit, not around an entire amplifier.

the output resistance is:

$$r_o \approx \frac{R_e}{r_e} r_c \quad (1392)$$

This equation is a useful guide to indicate the effect of emitter degeneration on the output resistance. However, keep in mind that the output incremental resistance r_o is dependent on the operating point of the transistor. The data-sheet value is given at one particular operating point, and this must be modified to suit a different operating point (section 29.11, page 817).

Example

Suppose that the transistor is operated at 1mA, so that the emitter incremental resistance r_e is about 25Ω . The zener diode is 6v2 volts, so that the emitter resistance R_e is $6k2\Omega$. The incremental output resistance r_c of the transistor is $100k\Omega$. What is the output resistance r_o ?

Solution

Using equation 1392 and substituting for the known quantities, we have:

$$\begin{aligned} r_o &= \frac{R_e}{r_e} r_c \\ &= \frac{6200}{25} 100 \times 10^3 \\ &= 24.8 \times 10^6 \Omega \end{aligned}$$

For example, this means that the output current would vary by $1\mu\text{A}$ with every 25 volts change in output voltage.

30.5 The Current Mirror

Widlar [267] recognized that two identical transistors, at the same temperature and carrying the same current, can be expected to have identical base-emitter voltages. He used this concept as the basis for the current generator now known as the *current mirror*, figure 807. The current mirror can be regarded as a development of figure 806, where the zener diode and resistor are eliminated, and the diode replaced by a diode-connected transistor identical to transistor Q_1 .

Some external circuit – in this case, the voltage source E_1 and resistor R_1 – establishes a *reference current* I_{ref} in the diode. This creates a base-emitter voltage in Q_1 , which is parallel with the base-emitter voltage of Q_2 . Then Q_2 carries the same collector current as Q_1 . Well, not quite, because this ignores the effect of base current, but we'll explore that in a moment.

One important subtlety of the circuit: notice that transistor Q_1 is not in saturation: its base-emitter voltage is imposed across the collector-emitter junction. Because saturation voltage (typically 0.2 volts for small signal transistors) is less than base emitter voltage (typically 0.6 volts for silicon transistors), then the transistor must be operating in its active region. As a result, the normal exponential relationship between base-emitter voltage and collector current must apply – not true if the transistor is saturated. If the collector-emitter voltage drops below the base-emitter voltage, then the base current decreases, and beta multiplication causes the collector current to decrease dramatically. This negative-feedback action keeps the collector-emitter voltage V_{ce1} of Q_1 at V_{be1} volts.

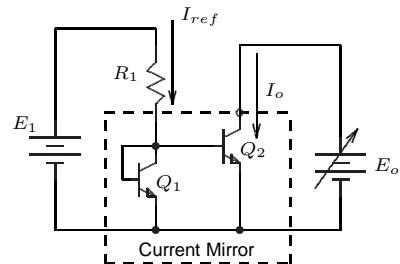


Figure 807: Current Mirror

More formally: taking KVL around the Q_1, Q_2 loop:

$$+ V_{be1} - V_{be2} = 0 \quad (1393)$$

The relationship between base-emitter voltage V_{be} and collector current I_c is given by equation 1157, page 807:

$$I_c = I_s e^{V_{be}/V_t} \quad (1394)$$

Rearrange equation 1394 to solve for V_{be} :

$$V_{be} = V_t \log_e \left(\frac{I_c}{I_s} \right) \quad (1395)$$

Substitute from equation 1395 into equation 1393 for both the V_{be} s:

$$V_{t1} \log_e \left(\frac{I_{c1}}{I_{s1}} \right) = V_{t2} \log_e \left(\frac{I_{c2}}{I_{s2}} \right) \quad (1396)$$

Now let's see what we can eliminate. The value of V_t is given by equation 395 on page 247:

$$V_t = \frac{kT}{q_e} \quad (1397)$$

where k and q_e are constants, and T is the absolute temperature. Consequently, we can substitute for V_t from equation 1397 into equation 1396:

$$\frac{kT_1}{q_e} \log_e \left(\frac{I_{c1}}{I_{s1}} \right) = \frac{kT_2}{q_e} \log_e \left(\frac{I_{c2}}{I_{s2}} \right) \quad (1398)$$

If the two devices are at the same temperature, $T_1 = T_2$, and we can simplify equation 1398 to:

$$\log_e \left(\frac{I_{c1}}{I_{s1}} \right) = \log_e \left(\frac{I_{c2}}{I_{s2}} \right) \quad (1399)$$

Now, if the logs are equal then their arguments are equal, so:

$$\frac{I_{c1}}{I_{s1}} = \frac{I_{c2}}{I_{s2}} \quad (1400)$$

If the two transistors are identical, then their values of I_s will be identical, and

$$I_{c1} = I_{c2} \quad (1401)$$

That is, referring back to figure 807, the output current I_o is equal to the reference current I_{ref} .

Base Current

30.6 Discrete Component Current Mirror

919

The preceding analysis assumes that base currents are negligible. That's not quite true, as figure 808 illustrates. From the figure, we have:

$$\begin{aligned} I_o &= I_{c2} \\ &= I_{c1} \\ &= I_{ref} - I_{c1}/\beta_1 - I_{c2}/\beta_2 \end{aligned} \quad (1402)$$

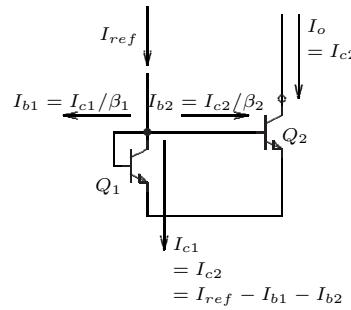


Figure 808: Current Mirror, Analysis

It's reasonable to assume that the current gains are equal, so that $\beta_1 = \beta_2 = \beta$. Then:

$$I_o = I_{ref} - 2I_o/\beta \quad (1403)$$

Rearranging the last line provides the ratio of output to input current:

$$\frac{I_o}{I_{ref}} = \frac{\beta}{\beta + 2} \quad (1404)$$

For $\beta \gg 2$ (the usual case), then the output current is very nearly equal to the reference current.

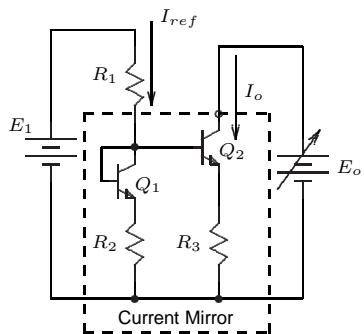
Performance

The current mirror of figure 807 has several advantages over the current source of figure 806:

- it requires fewer parts, and no zener diode
- the operating voltage required is generally less (one V_{cesat} across the output transistor)
- the operating current is easily scaled by adjusting the reference current. For example, in figure 807, the output current can be changed by adjusting resistor R_1 .

30.6 Discrete Component Current Mirror

The Current Mirror was an immediate success and widely adopted as a current source in integrated circuit design. Given its advantages, circuit designers began to consider using the current mirror in discrete component circuits. One possibility is to use an integrated circuit transistor array²⁴⁹ to provide the matched transistors. Arrays are useful when close matching is required, but they tend to be expensive and/or difficult to obtain. The other approach is to use discrete transistors and emitter degeneration resistors to swamp the mismatch between the two.



²⁴⁹eg, National Semiconductor LM394, Intersil CA3046.

Figure 809: Current Mirror, Discrete Version

The circuit of figure 807 will work with discrete PNP or NPN transistors. However, due to mismatch²⁵⁰ between the devices the current ratio will vary. As well, and this is likely to be the main problem, any change in temperature between the devices causes the output current to change. As shown in section 29.40 (page 904), equation 1384 indicates that the output current of the mirror will vary according to:

$$I_o = I_{ref} 1.085^{\Delta T} \quad (1405)$$

where I_o is the mirror output current, I_{ref} the mirror input current and ΔT the difference in temperature between the two transistors. That is, each degree difference creates an 8% difference in output current. The two devices should be mounted in close physical proximity or even forced into contact with each other to minimize the temperature gradient.

Mirror Mismatch, Example

A difference in temperature can arise because the two transistors are dissipating different amounts of power. Consider the example in figure 810, which is constructed from 2N4401 NPN transistors. They are packaged in a TO-92 case, which has a thermal resistance θ_{ja} of $200^{\circ}\text{C}/\text{W}$. The transistors are carrying equal currents of 1mA, but the voltage across Q_1 is 0.6 volts and the voltage across Q_2 is 30 volts, 50 times larger. The power dissipations are:

$$\begin{aligned} P_{D1} &= V_{ce1} I_1 \\ &= 0.6 \times (1 \times 10^{-3}) \\ &= 0.600 \text{ mW} \end{aligned}$$

$$\begin{aligned} P_{D2} &= V_{ce2} I_2 \\ &= 30 \times (1 \times 10^{-3}) \\ &= 30 \text{ mW} \end{aligned}$$

Using equation 1119 from page 787, we have:

$$T_{j1} = T_a + \theta_{ja} P_{D1} \quad (1406)$$

$$T_{j2} = T_a + \theta_{ja} P_{D2} \quad (1407)$$

where

- T_{j1}, T_{j2} Junction temperatures of the two transistors, $^{\circ}\text{C}$
- T_a Ambient temperature, $^{\circ}\text{C}$
- θ_{ja} Thermal resistance, junction to ambient, $200^{\circ}\text{C}/\text{watt}$
- P_{d1}, P_{d2} Power dissipations of the two transistors, watts

The temperature difference between the two transistor junctions is obtained by subtracting equations 1406 and 1407:

$$\Delta T_j = T_{j2} - T_{j1}$$

²⁵⁰This is usually referred to as V_{be} mismatch, since these two devices operated at the same collector currents will have different base-emitter voltages. In fact, the base-emitter voltages are forced to be the same because they are connected in parallel, so the mismatch is reflected in different collector currents [251].

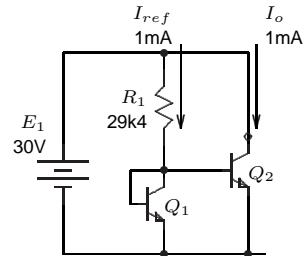


Figure 810: Power Dissipation

$$\begin{aligned}
 &= \theta_{ja} (P_{D2} - P_{D1}) \\
 &= 200 (30 - 0.6) \times 10^{-3} \\
 &= 5.88 \text{ }^{\circ}\text{C}
 \end{aligned}$$

Plugging this into equation 1405, we obtain the output current of the mirror:

$$\begin{aligned}
 I_o &= I_{ref} 1.085^{\Delta T} \\
 &= (1 \times 10^{-3}) \times 1.085^{5.76} \\
 &= 1.6 \text{ mA}
 \end{aligned}$$

As a result of the heating in Q_2 , the output current is 60% larger than the reference current. This is unacceptable. As well, any air currents over the transistors will alter the temperature differential and cause the output current to change. Next, we'll look at the cure for this problem.

Mirror Emitter Degeneration

The mismatch due to differences in the devices and device temperatures can be overcome with emitter degeneration resistors as shown in figure 809. Intuitively, it seems reasonable that a larger voltage across these resistors will reduce the effect of differences in base-emitter voltage. However, in many circuits it's desirable to keep the resistors as small as possible in order to minimize the voltage across them. How large must these resistors be in order to provide the desired degree of temperature stability?

The base-emitter voltage of a transistor changes by $-2.2\text{mV}/^\circ\text{C}$ (section 29.40 (page 902)). A simple design strategy is to

- Using the temperature difference and the base-emitter coefficient of temperature, determine the base-emitter voltage difference between the two transistors.
- This voltage difference, divided by the emitter resistance, determines the current difference in the two mirrors. That is specified by the requirements of the design, so the emitter resistor can be calculated.

This approach is not completely accurate, because it does not account for the change in base-emitter voltage caused by the changing emitter current. In section 30.10 (page 928), we show that the estimate is conservative, and the actual change in output current will be slightly less than predicted by the simple method. The simple method of calculating emitter resistance is illustrated in the following design example.

Design Example

Figure 811 is a discrete-component current mirror. Transistors Q_1 and Q_2 are identical small-signal silicon devices. The reference and output current are nominally $500\mu\text{A}$. A temperature gradient of 2 degrees may exist between the two transistors, under which conditions the output current should deviate from its nominal value by no more than 5%. Calculate the resistor values.

Solution

First, we'll calculate the emitter resistors R_2 , R_3 to meet the temperature stability criterion.

The change in base-emitter voltage of Q_2 will be:

$$\begin{aligned}\Delta V_{be} &= -2.2\text{mV}/^\circ\text{C} \times \Delta T \\ &= (-2.2 \times 10^{-3}) \times 2 \\ &= -4.4 \text{ mV}\end{aligned}$$

This must cause a change in current of:

$$\begin{aligned}\Delta I_o &= 5\% \times 500\mu\text{A} \\ &= 25\mu\text{A}\end{aligned}$$

Consequently, the emitter resistance R_3 must be:

$$R_3 = \frac{\Delta V_{be}}{\Delta I_o}$$

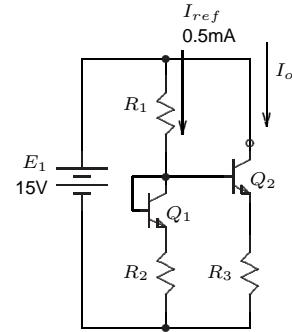


Figure 811: Design Example

$$\begin{aligned}
 &= \frac{4.4 \text{ mV}}{25\mu\text{A}} \\
 &= 176\Omega
 \end{aligned}$$

Resistor R_2 is made the same value to balance the circuit. The voltage dropped across each emitter resistor is:

$$\begin{aligned}
 V_{R3} &= I_o R_3 \\
 &= 500\mu\text{A} \times 176\Omega \\
 &= 88 \text{ mV}
 \end{aligned}$$

Now we can calculate the reference resistor R_1 :

$$\begin{aligned}
 R_1 &= \frac{+E_1 - V_{ce1} - V_{R2}}{I_{ref}} \\
 &= \frac{15 - 0.6 - 0.088}{500 \times 10^{-6}} \\
 &= 28k6\Omega
 \end{aligned}$$

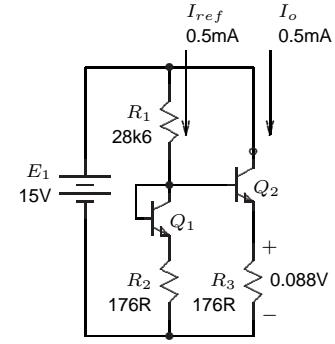


Figure 812: Design Complete

30.7 Multiple-Output Current Mirror

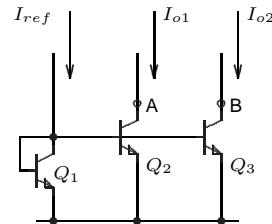
A simple current mirror with multiple outputs²⁵¹ is shown in figure 813(a).

Ignoring base currents, the base-emitter voltage of Q_1 appears at Q_2 and Q_3 , and so Q_2 and Q_3 each produce an output current (approximately) equal to I_{ref} . If points A and B are connected together, then the output currents add so that $I_o = 2I_{ref}$.

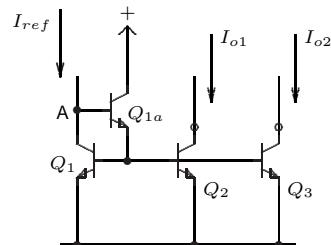
This scenario can be extended to additional outputs. However, the reference current must supply base current for each of the output transistors and this creates an error between reference and output that increases with the number of outputs. The collector current through Q_1 , which sets the output current I_o in each of the other transistors, is less than I_{ref} by $(N + 1)$ base currents, where N is the number of outputs²⁵².

In figure 813(b), an additional transistor Q_{1a} , acts as an emitter follower to supply the various base currents. This substantially reduces the error between I_{ref} and the various output currents²⁵³.

When the circuit is first turned on, transistor Q_1 is in the off state. Then I_{ref} is current driving into an open circuit, so point A rises in voltage. When point A reaches two base-emitter drops above ground, the emitter follower Q_{1a} turns on, dumping emitter current into the base of Q_1 and the output transistors. The circuit settles down with point A sitting at two base-emitter drops above ground. The collector current I_{c1} of Q_1 is equal to the reference current minus the base current in Q_{1a} . The base current of Q_{1a} –



(a) Basic Multiple Output



(b) Buffered Multiple Output

Figure 813: Multiple Output Mirror

²⁵¹For simplicity, this and subsequent current mirrors are shown without the emitter degeneration resistors required in a discrete component implementation.

²⁵²We assume that the current gain β is equal for all the transistors.

²⁵³Gilbert [251] suggests the name *Emitter Follower Augmented (EFA)* mirror for this circuit.

which is the error term between I_{ref} and I_o – is equal to the $(N + 1)$ base currents of the other transistors, divided by β_{1a} . Consequently, the error term is very small and the reference and output currents are nearly equal.

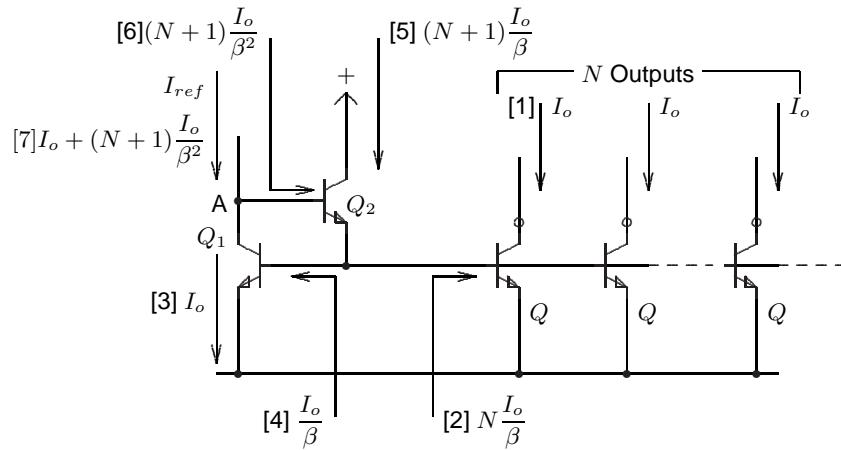


Figure 814: Multiple Output Mirror Analysis

Figure 814 shows a more detailed analysis of the currents in the buffered multiple-output mirror²⁵⁴. The bracketed numbers indicate the sequence of the analysis.

Step	Formula	Notes
1	I_o	The output current from each of the output transistors. Ideally should be equal to I_{ref}
2	$N \frac{I_o}{\beta}$	The total base current of the output transistors is beta times smaller than the total output current, NI_o .
3	I_o	The emitter (and collector) currents in Q_1 must be equal to the output (collector) current in each of the output transistors, since their base-emitter junctions are in parallel.
4	$\frac{I_o}{\beta}$	The base current in Q_1 is equal to its collector current divided by β
5	$(N + 1) \frac{I_o}{\beta}$	The emitter current of Q_2 is the sum of the base current in Q_1 (step 4) plus the total base current in the output transistors (step 2).
6	$(N + 1) \frac{I_o}{\beta^2}$	The base current of Q_2 is $\beta + 1 \approx \beta$ times less than its emitter current (step 5).
7	$I_{ref} = I_o + (N + 1) \frac{I_o}{\beta^2}$	The collector current of Q_1 is obtained by summing the currents at node A, (step 3 + step 6). This is the reference current I_{ref}

Step 7 in the table relates the reference current I_{ref} to the output current I_o , which may be manipulated into the form:

$$\frac{I_o}{I_{ref}} = \frac{1}{1 + \left(\frac{N+1}{\beta^2} \right)} \quad (1408)$$

²⁵⁴All the transistors are assumed to have the same value of current gain β . We assume the collector and emitter currents are approximately equal.

Since β^2 is likely to be much larger than $N + 1$, the right hand side is approximately equal to unity and $I_o \approx I_{ref}$.

30.8 The Wilson Current Mirror

The multiple-output current mirror of figure 813(b) eliminates base-current error and produces one or more output currents that are accurate repetitions of the reference current. However, the output resistance remains unchanged from the basic current mirror of figure 807. The Wilson current mirror shown in figure 815 has only a single output, but it also eliminates base-current error and substantially increases output resistance. For example, a comparison of basic and Wilson current mirrors in [268] shows $r_o = 600\text{k}\Omega$ for the basic mirror and $r_o = 36\text{M}\Omega$ for the Wilson mirror, a factor of 60 improvement. Among other applications, the large output resistance of the Wilson mirror is useful as a dynamic load to increase amplifier gain or as a highly linear constant-current source in a linear ramp circuit (section 29.9, page 811).

Operation

The Wilson mirror obtains its large output resistance by means of negative feedback. To understand the mechanics, consider the following sequence of events:

- The output voltage (the collector voltage of Q_3) increases.
- The output resistance of Q_3 conducts more current.
- This increases the current in the diode-connected transistor Q_2 .
- The other half of the mirror, Q_1 , mimics this increase in current at its collector.
- The reference current I_{ref} is unchanged, so the increase of current at the collector of Q_1 reduces the base current of the output transistor Q_3 .
- Decreasing base current in output transistor Q_3 reduces its collector (output) current.
- Consequently, there is a negative feedback action in which the original increase in collector current in Q_3 is countered by a decrease in its base current. The effect of this negative feedback is to reduce the change in mirror output current for a given change of output voltage, that is, to raise the output resistance.

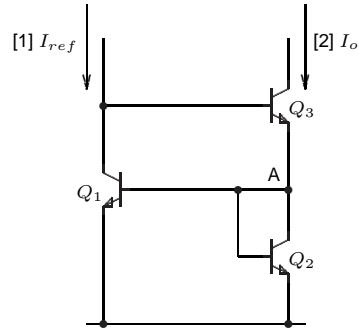


Figure 815: Wilson Current Mirror

Current Flow Analysis

Now we'll show that the output current is very nearly equal to the reference current. The steps in the analysis follow the numbers in figure 816, with the following notes:

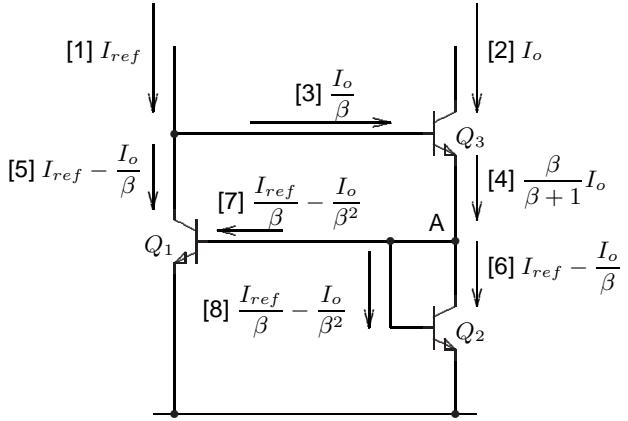


Figure 816: Current Analysis

Step	Formula	Notes
1	I_{ref}	The mirror reference (input) current
2	I_o	The mirror output current. Ideally should be equal to I_{ref}
3	$\frac{I_o}{\beta}$	The base current of Q_3 is beta times smaller than the output current, I_o .
4	$\frac{\beta+1}{\beta}I_o$	The emitter current of Q_3 is the sum of its collector current (step 2) and base current (step 3).
5	$I_{ref} - \frac{I_o}{\beta}$	Using step 1 and step 3, apply KCL at the collector node of Q_1 . This is the collector current of Q_1 .
6	$I_{ref} - \frac{I_o}{\beta}$	Transistors Q_1 and Q_2 form a simple current mirror. Then diode-connected transistor Q_2 must conduct the same collector current as Q_1 (step 5).
7	$\frac{I_{ref}}{\beta} - \frac{I_o}{\beta^2}$	The base current of Q_1 is beta times smaller than its collector current (step 4)
8	$\frac{I_{ref}}{\beta} - \frac{I_o}{\beta^2}$	The base current of Q_2 is beta times smaller than its collector current (step 5)

At node A, we can apply KCL to the currents:

$$+\frac{\beta+1}{\beta}I_o - \left(\frac{I_{ref}}{\beta} - \frac{I_o}{\beta^2}\right) - \left(\frac{I_{ref}}{\beta} - \frac{I_o}{\beta^2}\right) - \left(I_{ref} - \frac{I_o}{\beta}\right) = 0 \quad (1409)$$

After some algebraic manipulation:

$$\begin{aligned} \frac{I_o}{I_{ref}} &= \frac{\beta^2 + 2\beta}{\beta^2 + 2\beta + 2} \\ &= \frac{1}{1 + \frac{2}{\beta^2 + 2\beta}} \end{aligned} \quad (1410)$$

This ratio will be very nearly unity, since the term

$$\frac{2}{\beta^2 + 2\beta}$$

is much smaller than 1. That is, the output current I_o is very nearly equal to the reference current I_r .

Output Resistance

In the appendix on page 929, we show that the output resistance of the Wilson current mirror is given by equation 1429:

$$r_o = \frac{\beta r_{c3}}{2}$$

This is an increase of $\beta/2$ over the output resistance of the basic current mirror (figure 807) and the buffered multiple-output mirror (figure 813(b)).

With equation 1392 (page 917), we showed that the output resistance of a common-emitter current source can be raised with emitter resistance degeneration. This technique can be used to raise the output resistance of the basic current mirror and the buffered multiple-output mirror. However, that requires a voltage drop across an emitter resistor, and this may have an effect on the operating voltage range of the current mirror.

The Wilson current mirror cannot use degeneration resistors to raise its output resistance. However, a discrete component version of the mirror can use emitter resistors to swamp production tolerances and temperature differences in transistors Q_1 and Q_2 .

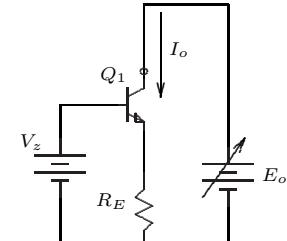
30.9 Emitter Resistor and Output Resistance

In this section, we determine the output resistance of the common-emitter transistor when there is an emitter resistor R_E . The circuit for analysis is based on figure 806. Figure 817(a) shows the circuit simplified by replacing the diode and zener with a constant voltage source V_z ²⁵⁵. Figure 817(b) shows the equivalent circuit for the AC output resistance analysis. The base voltage source becomes a short circuit to ground and the transistor is replaced by its T model.

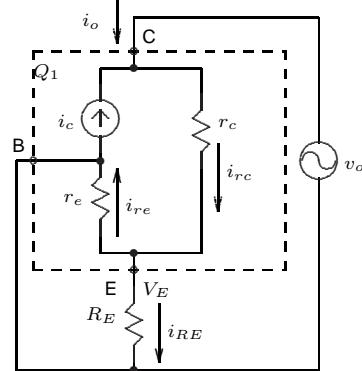
The AC currents and voltages in this circuit reflect the *changes* in the DC bias conditions of the circuit. The emitter and collector currents in figure 817(a) (the bias currents) flow down the page. However, the effect of increasing output voltage is to *decrease* these currents, and so the AC equivalent currents flow in the opposite direction, up the page, and this is shown in figure 817(b). The AC equivalent emitter current i_e is shown as flowing away from the emitter terminal. This is a consequence of the AC model, because the base terminal is grounded and resistors r_e and R_E are in parallel. Then the AC collector current generator i_c must reflect this polarity and it drives current *into* the collector terminal.

Now the mathematical analysis. The basic definition of output resistance is:

$$r_o = \frac{v_o}{i_o} \quad (1411)$$



(a) Circuit



(b) AC Equivalent

Figure 817: Output Resistance Analysis

²⁵⁵In fact, the zener has a small internal resistance, and a more exact analysis would take that into account.

Applying KCL to the collector node, we have

$$i_o = i_{rc} - i_c \quad (1412)$$

The collector current generator is controlled by the emitter current:

$$i_c = \alpha i_{re} \quad (1413)$$

At this point it helps to have some knowledge of typical values. For example, if the transistor was biased at 1mA emitter current, then $r_e = 26$ ohms. If the base bias supply V_z is 2 volts and emitter current 1mA, then the emitter resistor $R_E = 1400$ ohms. Evidently, in many practical cases, R_E is much larger than r_e . Then the current through r_c flows almost entirely through r_e . Then the emitter current through the emitter incremental resistance r_e is:

$$i_{re} \approx \frac{v_o}{r_c + r_e} \quad (1414)$$

We can make another approximation: collector incremental resistance r_c is typically orders of magnitude larger than the emitter incremental resistance r_e (eg, $10k\Omega$ vs 26Ω). So equation 1414 can be further simplified:

$$i_{re} \approx \frac{v_o}{r_c} \quad (1415)$$

With these approximations, we can say that:

$$i_{rc} \approx i_{re} \quad (1416)$$

Now we can back collapse equations 1416 through 1412 into equation 1411. The final result is:

$$r_o = \frac{r_c}{1 - \alpha} \quad (1417)$$

Using the definition of α that $\alpha = \beta/(\beta + 1)$, we have:

$$r_o = \beta r_c \quad (1418)$$

That is, the effect of the emitter resistance R_E is to increase the collector incremental resistance by a factor of beta. This is based on our assumption that $R_E \gg r_e$, which is usually true.

30.10 Appendix: Mirror Emitter Degeneration Resistors

For the purpose of analysis, the discrete-component current mirror of figure 809 can be simplified to the schematic of figure 818. The diode-connected transistor Q_1 and its emitter resistor R_2 are represented by the fixed voltage source V_{BB} . The emitter resistor R_3 of figure 809 is given the generic name R_E .

The base emitter voltage of the transistor is a logarithmic function of the collector current, and the voltage across the emitter resistor is a linear function of the emitter current. Consequently, this circuit does not have a simple analytical solution and an iterative approach is more useful.

Consider the effect of a change in temperature. Suppose that the junction of transistor Q_2 increases in temperature. Then, according to section 29.40 (page 902), the base-emitter voltage will

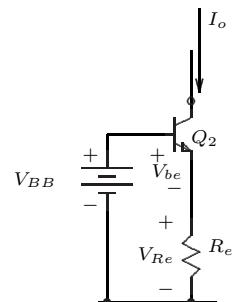


Figure 818: Emitter Degeneration Analysis

change at a rate of about $\Delta V_{be} = -2.2\text{mV}/^\circ\text{C}$. By KVL, this causes the voltage across the emitter resistor to *increase* at the same rate, which in turn causes the collector current to increase by an amount

$$\Delta I_c = \alpha \frac{\Delta V_{be}}{R_e} \quad (1419)$$

(As usual, we'll take $\alpha \approx 1$.) Using equation 1395 again, this increase in collector current moves the base emitter voltage in the opposite direction, an increase this time:

$$V_{be} = V_t \log_e \left(\frac{I_c}{I_s} \right) \quad (1420)$$

This increase opposes the original temperature-caused increase in voltage across the emitter resistor. Iterating around this loop with a spreadsheet confirms that the actual increase in collector current is reduced by the effect of equation 1420.

The table shows typical results for a current mirror of the type shown in figure 818, operating at $100\mu\text{A}$:

Reference Current, μA	100	100	100
Saturation Current I_s , A	3.77×10^{-17}	3.77×10^{-17}	3.77×10^{-17}
V_t , volts	0.026	0.026	0.026
Input Voltage V_{BB} , V	0.625	0.650	0.700
V_{be} for $\Delta T = 0$	0.600	0.600	0.600
ΔT , $^\circ\text{C}$	5	5	5
Emitter resistance R_e	250	500	1000
I_o , simple calculation, μA	144	122	111
I_o , iterated, μA	122	115	109

The main results are in the last two rows of the table, which show the simple and iterated predictions of output current from the mirror. From these results, we determine:

- In all cases, the temperature differential causes the output current to be larger than the reference current, as expected.
- As the emitter resistance increases (with the input voltage adjusted to keep the current the same), the effect of temperature decreases.
- The simple method (ignoring equation 1408) gives a change in current that is somewhat higher than the iterated value.
- The disparity between the simple and iterated methods decreases with larger emitter resistors.

Consequently, *the simple method may be used with confidence that it gives a result that is conservative – somewhat larger than the actual output current*.

30.11 Appendix: Wilson Mirror Output Resistance

In this section, we'll develop an expression for the output resistance of the Wilson current mirror. To begin with, some remarks on the strategy of solving this type of problem. One starts with the equivalent circuit – in this case, figure 819(a).

From there, one method is write the 7 loop and 4 node equations and then solve them (manually or using a computer equation solver) for the desired result, which is a large equation with many terms. Then one applies

various approximations and the equation collapses into something manageable. This is the *approximations-at-the-end* approach.

Approximations-at-the-end has the advantage that it requires very few assumptions on the circuit operation and much of the work is mechanical. However, the resulting equations can be cumbersome. An example of this approach for the analysis of this circuit is in [252].

Alternatively, one can look at the circuit for various simplifications and, once those have been applied, generate the equations. In some cases – and this particular case is an example – the equations are much simpler to deal with. This is the *approximations-at-the-beginning* approach.

This approach requires some knowledge of the circuit operation in order to spot components that can be combined or removed altogether. The mechanics are simpler, but prone to error if one makes a mistaken approximation.

The full-blown equivalent circuit for the determination of output resistance in the Miller current mirror is shown in figure 819(a). Notice that the reference current source has been treated as an open-circuit. The ratio of the applied AC output voltage v_o to the resulting current i_o gives the output r_o , the output resistance of the circuit.

Careful examination of figure 819(a) reveals that three simplifications are possible:

- Current source i_{c2} is short-circuited, so it may be removed.
- Collector resistance r_{c2} is in parallel with r_{e2} . Since collector incremental resistance is several orders of magnitude larger than emitter incremental resistance, the emitter resistance predominates and r_{c2} may be removed.
- By similar reasoning, r_{c1} appears in parallel with the series combination of r_{e3} and r_{e2} , so it may be removed.

The result is shown in figure 819(b). Notice also in figure 819(b) that the current source i_{c3} acts to reduce the input current i_i and therefore points up the page. This then correctly reflects the operation of the circuit.

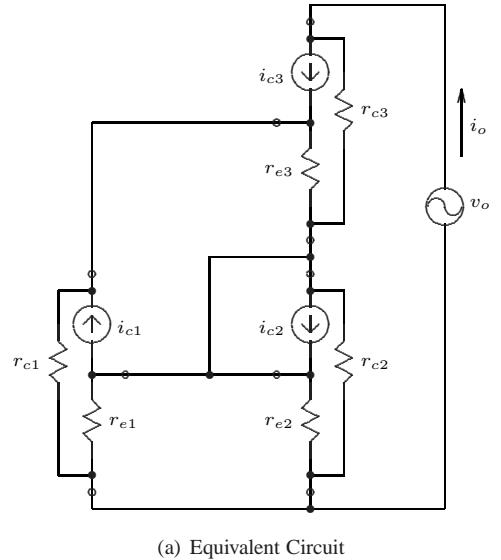
As usual, we'll assume that all the current gains are equal to β , that $\alpha = 1$, and the emitter incremental resistances r_e are equal. Now we can assail this circuit with equations. The definition of the output resistance is:

$$r_o = \frac{v_o}{i_o} \quad (1421)$$

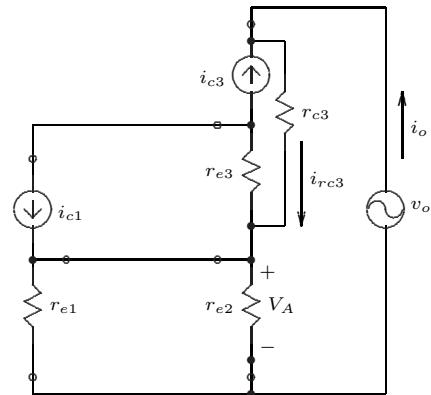
By KCL at the output node of the circuit,

$$i_o = i_{r3} - i_{c3} \quad (1422)$$

Now we can make another approximation. It's a reasonable assumption that the voltage at point A is a small fraction of the test voltage v_i . Why? Consider the current flowing through r_{c3} : the parallel combination of resistors r_{e1} and r_{e2} must be much less than the collector resistance r_{c3} , so the voltage developed by



(a) Equivalent Circuit



(b) Simplified

Figure 819: Wilson Mirror, Output Resistance

current i_{r3} must be small. What about current i_{c3} ? That flows in the opposite direction so it lowers the voltage across r_{e1} and r_{e2} . So: the voltage across r_{c3} is approximately equal to v_i , and²⁵⁶:

$$i_{r3} = \frac{v_i}{r_{c3}} \quad (1423)$$

The collector current of Q_3 is beta times its base current:

$$i_{c3} = \beta i_{c1} \quad (1424)$$

Now we need to trace the collector current of Q_1 to the right side of the circuit:

$$i_{c1} = \alpha i_{e1} \quad (1425)$$

$$i_{e1} = \frac{v_A}{r_{e1}} \quad (1426)$$

where v_A is as defined on figure 819(b).

And finally, the definition of v_A :

$$v_A = (i_{r3} - i_{c3})(r_{e1} \parallel r_{e2}) \quad (1427)$$

Since the emitter resistances are all equal to r_e , we can rewrite equation 1427 as:

$$v_A = (i_{r3} - i_{c3}) \frac{r_e}{2} \quad (1428)$$

Now back-substitute these equations with the objective of obtaining i_i in terms of v_i for equation 1421. One obtains:

$$\begin{aligned} r_o &= r_{c3} \left(1 + \frac{\beta}{2} \right) \\ &\approx \frac{\beta r_{c3}}{2} \end{aligned} \quad (1429)$$

30.12 The Cascode Pair

The *cascode pair* is an amplifier stage that serves three useful purposes:

- It has very large output resistance, which is useful in a current source or to create high voltage gain in an amplifier.
- It distributes the supply voltage so that an amplifier stage can include a series of lower-voltage devices.
- It minimizes the effect of collector-base feedback capacitance, thereby extending the high-frequency response.

A simple NPN cascode is shown in figure 820. Common-emitter stage Q_1 drives a common-base stage Q_2 .

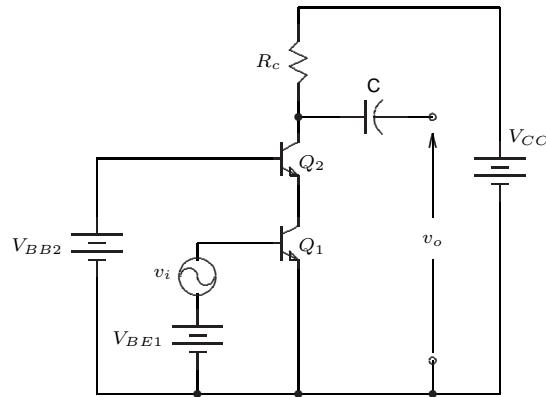


Figure 820: Cascode Pair

²⁵⁶An examination of the original circuit (figure 815) also suggests that the AC voltage across the diode connected transistor Q_2 will be much smaller than the voltage across the output transistor Q_3 .

Biassing

Essentially any bias scheme that will work with a common-emitter amplifier stage will work in this situation to bias Q_1 . In this circuit, we've chosen to use the same method as shown in figure 734 (page 834). Voltage source V_{BE1} biases the base of Q_1 so that the AC voltage v_i creates a collector current that includes an AC current superimposed on a DC bias current.

Voltage source V_{BB2} establishes a DC voltage at the base of Q_2 that is sufficient to ensure that Q_1 is not saturated.

$$+ V_{ce1} = V_{BB2} - V_{be2} \quad (1430)$$

For example, if V_{BB2} is set at 2 volts, then the collector-emitter voltage of Q_1 is about 1.4 volts. This voltage is essentially static – it doesn't change with the AC signal.

The DC current in Q_1 appears at the collector of Q_2 (diminished very slightly by the current gain α) and then establishes a DC voltage across the collector resistor R_c . The varying collector current in Q_2 creates an output voltage at the collector of Q_2 . The coupling capacitor removes the DC component of the voltage, leaving the AC component v_o at the output terminals.

There are many variations on this circuit. The cascode is often incorporated into other circuitry, and DC coupling into the input and output terminals is common. However, the circuit in figure 820 is a useful place to begin.

Voltage Gain

The voltage gain of the cascode amplifier is identically the same as the common-emitter amplifier: the ratio of collector to emitter resistance. In this circuit, the collector resistance is R_c , the emitter resistance r_{e1} . Transistor Q_2 is a grounded base stage, so its AC output (collector) current is equal to its input (emitter) current. Consequently, it has very little effect on the voltage gain of the stage.

Now in more detail: The AC emitter current in Q_1 is given by:

$$i_{e1} = \frac{v_i}{r_{e1}} \quad (1431)$$

The collector current of Q_1 is alpha times its emitter current:

$$i_{c1} = \alpha_1 i_{e1} \quad (1432)$$

The collector current of Q_1 becomes the emitter current of Q_2 :

$$i_{e2} = i_{c1} \quad (1433)$$

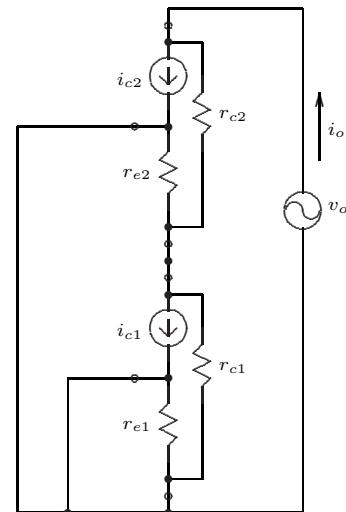
The AC collector current of Q_2 is alpha times its emitter current.

$$i_{c2} = \alpha_2 i_{e2} \quad (1434)$$

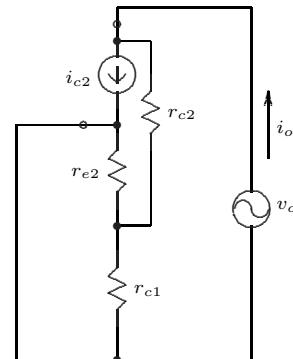
The collector resistance R_c converts the AC output current into an AC voltage²⁵⁷:

$$v_o = R_c i_{c2} \quad (1435)$$

²⁵⁷More exactly, the collector resistance is the parallel combination of R_c and the incremental output resistance r_{o2} of Q_2 . As we'll see, r_{o2} is very large, so we'll assume it can be ignored.



(a) Equivalent Circuit



(b) Simplified

Figure 821: Cascode Output Resistance

Back collapsing equations 1435 through 1431, and making the usual approximation that $\alpha \approx 1$, we have

$$\begin{aligned} A_v &= \frac{v_o}{v_i} \\ &= \alpha_2 \alpha_1 \frac{R_c}{r_{e1}} \\ &\approx \frac{R_c}{r_{e1}} \end{aligned} \quad (1436)$$

This is the same result as for the common-emitter amplifier (section 29.26, page 869).

Output Resistance

The AC equivalent circuit for determining output resistance is shown in figure 821(a). For this calculation, both the bases are connected to an AC ground. We can simplify this circuit:

- Emitter resistance r_{e1} is short-circuited, so it may be removed.
- Current generator i_{c1} has no input base current, so its AC current is zero and it appears as an open circuit in the circuit and may be removed.

The simplified circuit is in figure 821(b).

In section 30.9 (page 927) we developed equation 1418, which shows the effect of emitter resistance on output resistance in the common-emitter stage:

$$r_o \approx \left(\frac{R_e}{r_e} \right) r_c$$

That is, the output resistance r_o is equal to the collector incremental resistance r_c , increased by the ratio of external emitter resistance R_e to emitter incremental resistance r_e . In the cascode stage, the external emitter resistance of the output stage Q_2 is the collector incremental resistance of the first stage, r_{c1} . By analogy with this case, we can write:

$$r_{o2} \approx \left(\frac{r_{c1}}{r_{e2}} \right) r_{c2}$$

Since collector incremental resistance r_c is typically in the order of $100\text{k}\Omega$ and emitter incremental resistance r_e is in the order of tens of ohms, the ratio r_{c1}/r_{e2} is a very large number, and the output resistance of the cascode amplifier is much larger than the intrinsic collector incremental resistance r_{c2} of Q_2 .

Frequency Response

The cascode amplifier performs better than the common-emitter amplifier at high frequencies. With some judicious approximations, it's relatively simple to understand why.

In section 13.1 on page 335, we showed that feedback capacitance C_f around a voltage amplifier appears as a capacitor $1 + A_v$ times as large across the input terminals of the amplifier (figure 268).

A common-emitter amplifier is shown in figure 822. The transistor has an unavoidable parasitic capacitance C_{be} between base and emitter, and capacitance C_{cb} between collector and base.

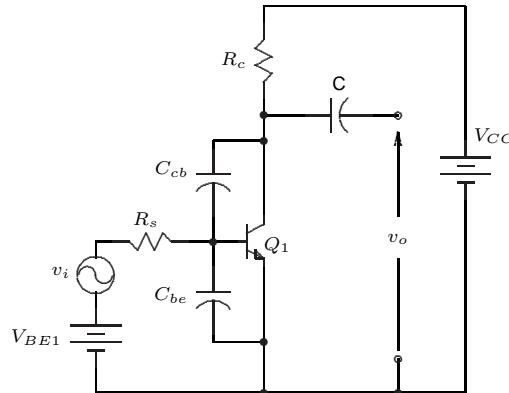


Figure 822: Common-Emitter Amplifier

If the transistor has a voltage gain of A_v volts/volt between the base and collector terminals, then the feedback capacitance C_{cb} will appear as a capacitance in parallel with C_{be} , but multiplied by the factor $1 + A_v$.

This capacitance then interacts with the source resistance R_s to form a lowpass filter, restricting the frequency response of the amplifier.

Example: Common-Emitter Amplifier Frequency Response

The transistor of figure 822 is biased at 1mA and the collector resistance is 7k5Ω. The device capacitances are $C_{be} = 30\text{pF}$, $C_{cb} = 6\text{pF}$, and the source resistance R_s is 250Ω. What is the corner frequency of the input lowpass filter?

Solution

The emitter incremental resistance r_e (equation 1242 on page 836) is:

$$\begin{aligned} r_e &= \frac{0.026}{I_e} \\ &= \frac{0.026}{1 \times 10^{-3}} \\ &= 26\Omega \end{aligned}$$

The voltage gain (assuming $r_c \gg R_c$) is:

$$\begin{aligned} A_v &= \frac{R_c}{r_e} \\ &= \frac{7500}{26} \\ &= 288 \text{ volts/volt} \end{aligned}$$

Then the feedback capacitance C_{cb} appears multiplied by $1 + A_v$ in parallel with C_{be} , so that the total input capacitance is:

$$\begin{aligned} C_{in} &= C_{be} + (1 + A_v)C_{cb} \\ &= 30\text{pF} + (1 + 288) \times 6\text{pF} \\ &= 1766 \text{ pF} \end{aligned}$$

This capacitance forms a lowpass filter with source resistance R_s at a corner frequency:

$$\begin{aligned} f_c &= \frac{1}{2\pi R_s C_{in}} \\ &= \frac{1}{2 \times 3.14 \times 250 \times (1766 \times 10^{-12})} \\ &= 360 \text{ kHz} \end{aligned}$$

This lowpass network will reduce the input voltage beyond f_c , and consequently the output voltage as well²⁵⁸.

Notice that the feedback capacitance is smaller than the base-emitter capacitance but has a larger effect on the frequency response because of the Miller multiplication effect.

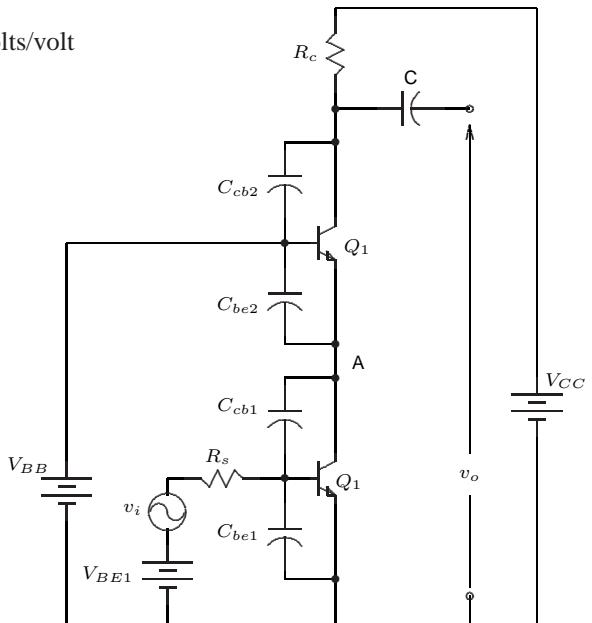


Figure 823: Cascode Frequency Response

²⁵⁸This analysis simplifies the situation, inasmuch as it ignores the effect of the output resistance R_c . However, the basic idea still holds.

Cascode Frequency Response

We'll now see that the cascode circuit has a much-extended frequency response. Consider figure 823. We'll show that both the Q_1 and Q_2 stages defeat the Miller effect.

The collector of the first stage (point A) has very little voltage swing, because the effective collector load resistance for Q_1 is the emitter incremental resistance of Q_2 , which is very small. Consequently, the voltage gain of the Q_1 stage and the Miller multiplication effect are also small.

For transistor Q_2 , the base terminal is an AC ground, so the collector-base capacitance is not between the output and input terminals and the Miller effect does not apply to this capacitor either²⁵⁹.

Consequently, the input capacitance C_{in} seen into the base of the first transistor is simply the sum of the collector-base and base-emitter capacitances. Using the figures for the previous example, this moves the input lowpass filter corner frequency from 346kHz to 17MHz.

The other lowpass filter is formed by the collector resistance R_C with the collector-base capacitance C_{cb2} of the second transistor. The corner frequency for that circuit is 3.5MHz. Consequently, it will dominate the frequency response, and the cascode amplifier high-frequency cutoff is ten times that of the simple common-emitter amplifier stage.

30.13 Folded Cascode Pair

It is possible to construct a *folded* cascode pair using an NPN transistor followed by a PNP transistor, as shown in figure 824. The bias current in Q_2 is determined by the current source I_1 minus the bias current in Q_1 . An AC signal current out of the collector of Q_1 will see the current source I_1 as an open circuit and flow into the emitter of Q_2 .

The voltage across the current source I_1 need only be a fraction of a volt to keep the current source transistors from saturating. As a result, the output of Q_2 can swing close to the supply rails and there is wide latitude for locating the quiescent output voltage of Q_2 . For example, if the collector resistor R_c is run to a negative supply voltage, then the collector voltage of Q_2 can be placed at zero volts.

The current source I_1 can be replaced by a resistor. To be effective, it must have a resistance that is large compared to the input resistance r_{e2} seen at the emitter of Q_2 .

30.14 The Differential Pair

The *differential pair* was invented in 1938 by the prolific British engineer Alan Blumlein [269]. The basic BJT version is shown in figure 825(a).

The emitters of the two transistors are tied together and to a constant current sink, I_{EE} , which is known as the *tail current*.

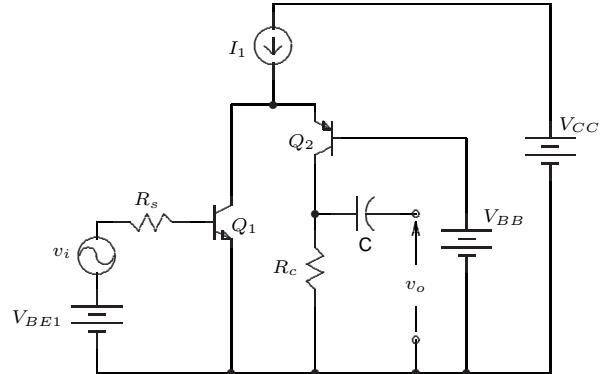


Figure 824: Folded Cascode

²⁵⁹What about the collector-emitter capacitance C_{ce} ? It's between the output and input terminals, and the common-base transistor has a large voltage gain. However, the effect of C_{ce} capacitance is mitigated by two factors: (a) the capacitance is very small and (b) the effective source resistance of the lowpass filter is r_{e2} , the very low input resistance of the second transistor. Consequently, collector-emitter capacitance doesn't have much effect.

If the two base voltages V_{B1} and V_{B2} are equal²⁶⁰, then the tail current splits equally between the two transistors so that each has a collector current equal to half the tail current.

$$I_{C1} = I_{C2} = \frac{I_{EE}}{2} \quad (1437)$$

Now suppose the base voltage V_{B1} of transistor Q_1 increases slightly with respect to the other base voltage V_{B2} . Then transistor Q_1 conducts more heavily – its collector current I_{C1} increases. Since the emitter current is constant, as the collector current of transistor Q_1 increases, the collector current I_{C2} of the other transistor decreases. As we shall see, a small difference voltage between the two bases, in the order of 80mV, completely switches the tail current from one side of the pair to the other.

It's useful to redraw the input voltages V_{B1} and V_{B2} as a *common-mode* voltage V_{CM} and a *difference* voltage V_d . Then:

$$V_{CM} = \frac{V_{B1} + V_{B2}}{2} \quad (1438)$$

$$V_d = V_{B1} - V_{B2} \quad (1439)$$

If the transistors operate in their active region and are not saturated or cutoff, the differential pair largely responds to the difference voltage V_d and ignores the common-mode voltage V_{CM} . This property is very useful.

- It forms the basis for the input stage of the operational amplifier.
- It allows direct coupling between differential pair stages, without regard to biasing level.

As well, since the switching of current from one side to another involves minimal change of voltage, stray capacitance need not be charged or discharged. Consequently, the differential pair can be used as the basis for a high-speed logic family, Emitter Coupled Logic (ECL).

30.14.1 BJT Differential Pair, Transfer Characteristic

The *transfer function* of the differential pair is shown in figure 826. It shows the two collector currents as a function of the differential input voltage V_d .

The collector currents are shown as a ratio of the emitter tail current I_{EE} . The difference voltage is shown in units of V_t , approximately 26mV at room temperature.

²⁶⁰We also assume that the transistors have identical characteristics and the transistor alphas are close enough to unity that the collector and emitter current are essentially equal.

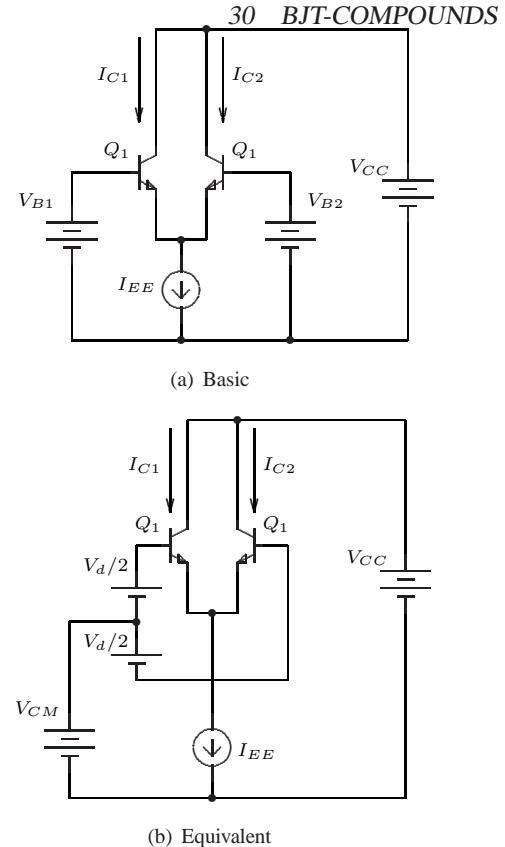


Figure 825: Differential Pair

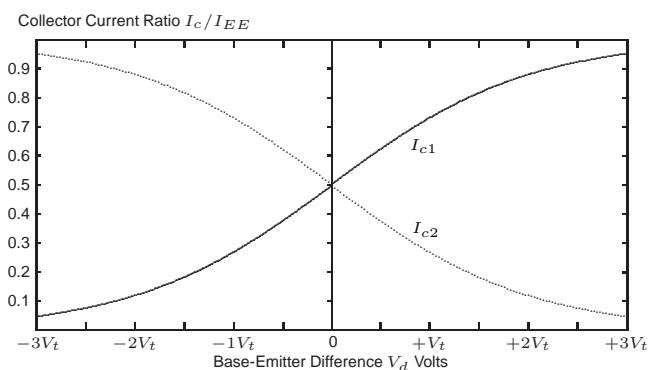


Figure 826: Diff Amp Transfer Characteristic

Notice:

- When the differential input voltage is zero (that is, the two base terminals are at the same potential), both collector currents are equal to one-half the emitter tail current I_{EE} .
- As the difference voltage is made positive, the collector current of Q_1 increases at the expense of the collector current in Q_2 .
- When the difference voltage is $3V_t$ (approximately 78mV), 95% of the tail current is flowing in one transistor and 5% in the other. In other words, this relatively small differential input voltage has transferred the tail current almost entirely to one side of the differential pair.
- In circuits where linearity is important (low distortion audio amplifiers, for example), it's important to note that the transfer characteristic becomes progressively less linear as the difference voltage increases.

Example Design: Biasing

In this section, we look at an example design – first, from the standpoint of biasing, and then to determine the gain equations. This will demonstrate some of the issues around the discrete BJT differential amplifier.

The bias circuitry is the surrounding circuitry to make the differential amplifier fully functional. There are a number of issues:

- Establishment of the quiescent voltage of the transistor base terminals with respect to the power supply or supplies.
- The emitter current sink.
- Connection of the input signal.
- Extraction of the output signal.

Figure 827 provides a starting point. This is a very simple differential amplifier, which produces an output signal that is the difference of two input signals, multiplied by some gain G . The non-inverting input is at terminals 3, 4, the inverting input at terminals 1, 2, and the output at terminals 5, 6.

$$v_{56} = G(v_{34} - v_{12}) \quad (1440)$$

We'll assume that the input and output signals are coupled via a series capacitor, so that a DC level at the input terminal has no effect on the input signal, and a DC level at the output terminal has no effect on the output signal.

Resistors R_{B1} and R_{B2} provide a path for the transistor base currents. The emitter resistance R_{EE} with negative supply V_{EE} form the tail current source²⁶¹. Collector resistance R_{C1} converts the variation of collector current into a variation in collector voltage.

Now we will calculate the values of the tail resistance R_{EE} , collector resistance R_{C1} , and base resistances R_{B1} , R_{B2} . We assume the following conditions:

$$V_{CC}=12\text{V}, V_{EE}=12\text{V}, I_{C1} = I_{C2} = 1\text{mA}, V_{ce1} = 6\text{V}.$$

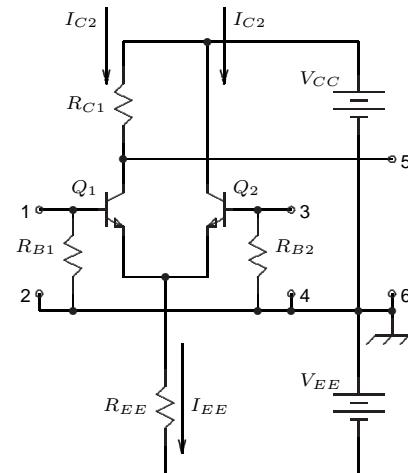


Figure 827: Example Differential Amplifier

²⁶¹As we'll see, this arrangement provides rather poor common-mode rejection ratio and is only useful in non-critical applications.

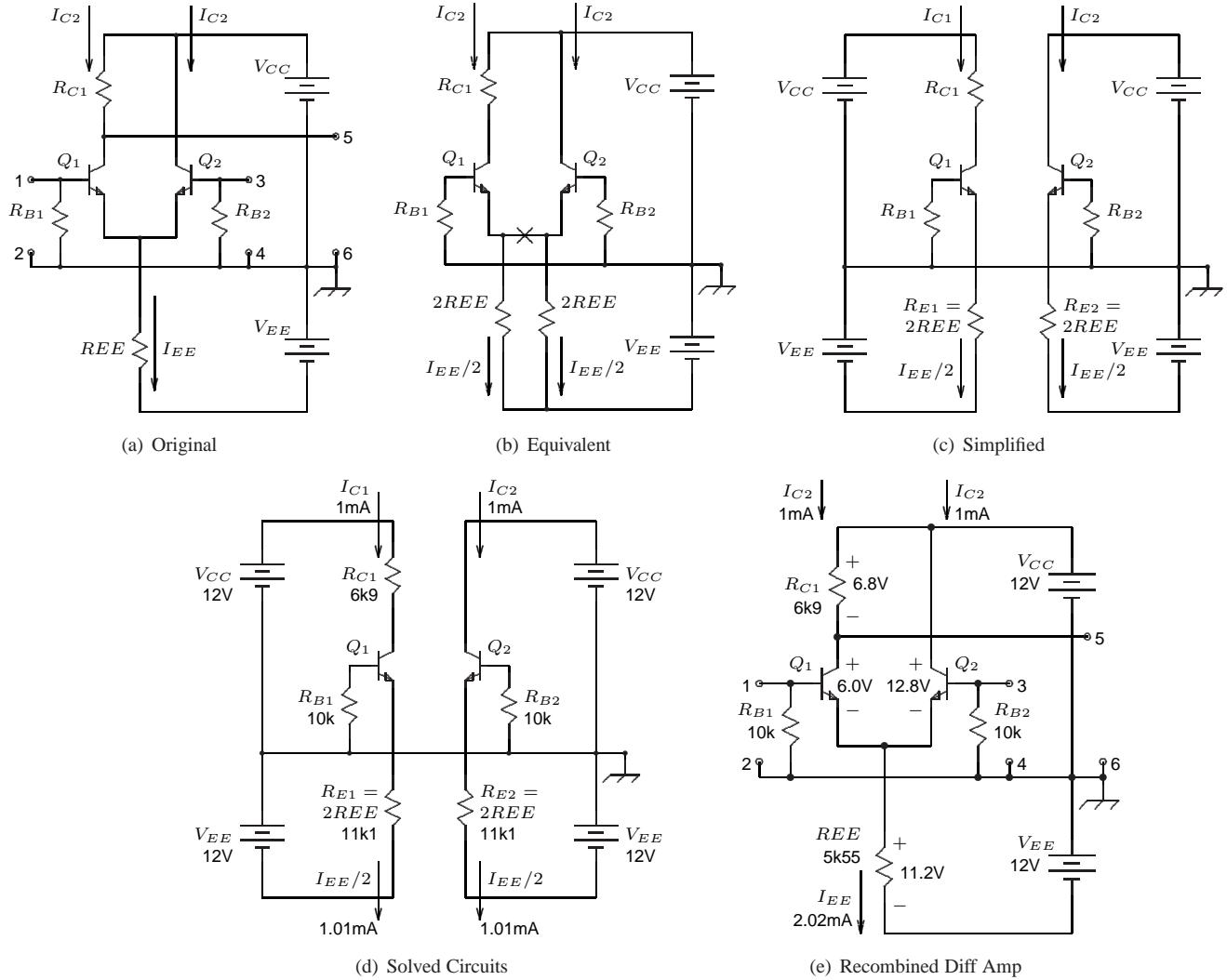


Figure 828: Differential Pair DC Analysis

The differential amplifier circuit can be simplified by the process shown in figure 828. Figure 828(a) shows the original circuit. In figure 828(b), the signal input and output terminals have been removed for clarity. As well, the emitter tail resistor R_{EE} is split into two parallel resistors each of value $2R_{EE}$. The connection at X may be removed since it has no effect on the currents or voltages of the circuit. Then in figure 828(c) the supply voltages are replicated, again without affecting the circuit.

Figure 828(c) consists of two BJT bias circuits. The left circuit was analysed as a BJT bias example in section 29.14 on page 831, and suitable resistance values are shown in figure 733 on page 832.

A collector resistance is missing from the right circuit. However, the collector and emitter current are essentially unaffected by collector resistance, so the values of R_e and R_b are the same as the left circuit. Figure 828(d) shows the separate circuits, with the values copied in from figure 733 on page 832.

They are recombined into a differential amplifier by combining the two tail currents and replacing the two 11k1 resistors in parallel with one 5k55 resistor. Figure 828(e) shows these two circuits recombined into one

differential amplifier.

Because there is no resistance in the collector lead of Q_2 , the collector-emitter voltage for Q_2 is larger than that for Q_1 . This has no material effect on the bias conditions for the two transistors²⁶².

Example Design: Differential Gain

There are two voltage gains of interest in the differential of figure 827: the *differential gain* and the *common-mode gain*. In an ideal differential amplifier the differential gain will be some large value (usually the larger the better) and the common-mode gain will be zero. Neither of these is the case for the circuit of figure 827, but we'll show improvements momentarily.

The differential gain is defined as:

$$A_d = \frac{e_o}{e_2 - e_1} \quad (1441)$$

where e_o is the output voltage, e_1 is one of the input voltages and e_2 is the other input voltage.

We can determine the gain using the superposition theorem, together with the gain behaviour of single stage BJT amplifier stages.

The total output is the sum of the output signal e_{o1} due to input signal e_{i1} (which we will write as $[e_{o1}]_{e_{i1}}$), and output e_{o2} due to input signal e_{i2} (which we will write as $[e_{o1}]_{e_{i2}}$). When that's known, we can determine the differential gain A_d from equation 1441.

It's possible to do this analysis almost entirely by redrawing the circuit in the sequence shown in figure 829, and with reference to the gain relationships we previously developed for the single stage BJT amplifier.

- The original circuit is shown in figure 829(a), together with the AC input sources e_1 and e_2 . These are coupled to the input terminals via capacitors so that the input sources do not disturb the bias conditions.
- In figure 829(b), the power supplies are replaced by their internal impedances (short circuits), which puts the power supply rails at an AC ground point. The capacitors are assumed to be short circuits at the frequency of interest. The base-bias resistors are in parallel with the input voltage sources, so they have no effect on those voltages and are removed from the circuit.
- In figure 829(c), we perform the first stage of the superposition. The input e_{i2} is replaced by a short circuit, which grounds the base of Q_2 . Resistor R_{EE} is effectively in parallel with the emitter incremental resistance r_{e2} . Since the emitter current is about 1mA, the incremental resistance is in the order of 25Ω . Consequently resistor R_{EE} is large compared to r_{e2} and it can be ignored.
- Figure 829(d) shows the final simplified equivalent circuit when the input is e_{i1} . From the analysis of the common-emitter amplifier in section 29.20, the voltage gain is equal to the ratio of the collector resistance to the total emitter resistance. (As usual, this assumes that $\alpha \approx 1$.) Consequently,

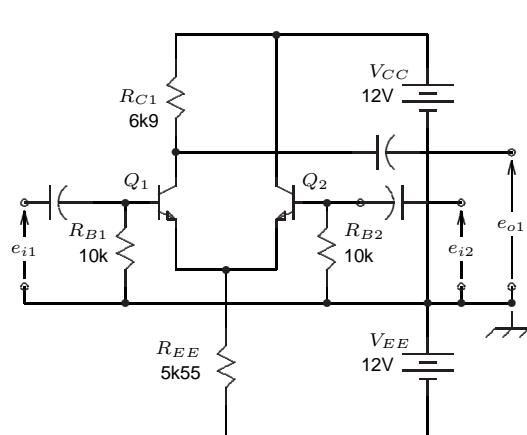
$$[e_{o1}]_{e_{i1}} = -e_{i1} \frac{R_{c1}}{r_{e1} + r_{e2}} \quad (1442)$$

Assuming that the emitter currents are approximately equal and the transistors similar, then we can assume that the incremental emitter resistances are equal:

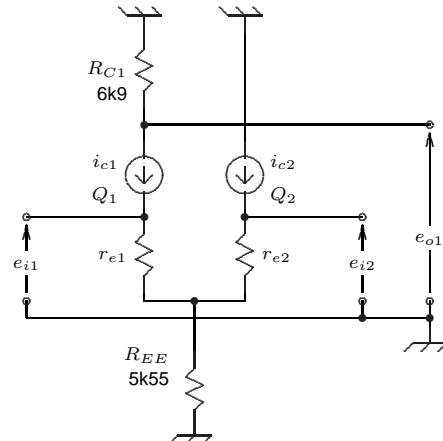
$$[e_{o1}]_{e_{i1}} = -\frac{R_{c1}}{2r_e} e_{i1} \quad (1443)$$

The minus sign is important: it indicates that the output is inverted with respect to the input.

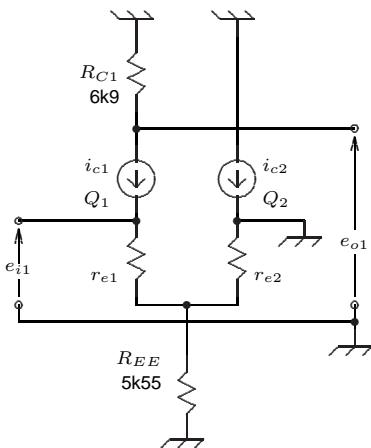
²⁶²The Early effect will increase the collector current in Q_2 , but in a first-cut analysis the effect is small enough to be ignored.



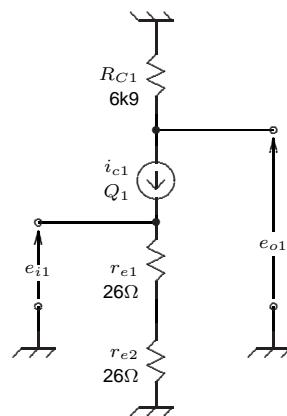
(a) Original



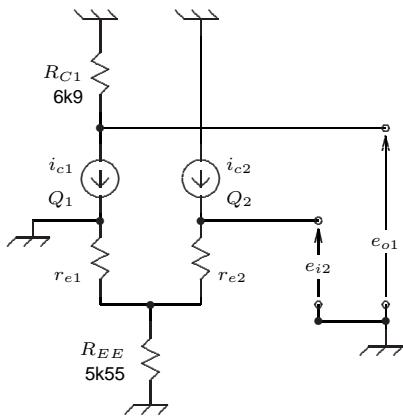
(b) AC Equivalent



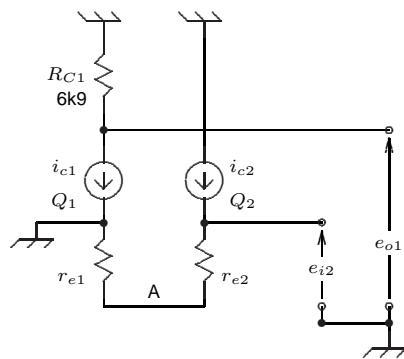
(c) AC equivalent, Input 1



(d) Input 1 Simplified



(e) AC equivalent, Input 2



(f) Input 2 Simplified

Figure 829: Differential Amplifier AC Analysis

- Now we can determine the output due to the input e_{i2} . The AC equivalent source, with e_{i1} short-circuited to ground, is shown in figure 829(e). Again, R_{EE} is large enough compared to r_e to be ignored.

Input voltage e_{i2} appears across the two emitter resistances r_{e1} and r_{e2} . Assuming again that they are equal to r_e , then the AC emitter current in Q_1 is

$$i_{e1} = \frac{e_{i2}}{2r_e} \quad (1444)$$

This current appears as collector current ($\alpha \approx 1$), and then translates into output voltage $[e_{o1}]_{e_{i2}}$ by the collector resistance R_{c1} . Consequently, the output voltage $[e_{o1}]_{e_{i2}}$ is given by:

$$[e_{o1}]_{e_{i2}} = \frac{R_{c1}}{2r_e} e_{i2} \quad (1445)$$

- In the final step, we add the two output voltages: e_{o1} from equation 1443, e_{o2} from equation 1445, to get the total output voltage:

$$\begin{aligned} e_{o1} &= [e_{o1}]_{e_{i1}} + [e_{o1}]_{e_{i2}} \\ &= -\frac{R_{c1}}{2r_e} e_{i1} + \frac{R_{c1}}{2r_e} e_{i2} \\ &= \frac{R_{c1}}{2r_e} (e_{i2} - e_{i1}) \end{aligned} \quad (1446)$$

Rearranging this equation gives the differential gain:

$$\begin{aligned} A_d &= \frac{e_{o1}}{e_{i2} - e_{i1}} \\ &= \frac{R_{c1}}{2r_e} \end{aligned} \quad (1447)$$

Because the output is taken from one collector of the amplifier and referred to ground, we refer to this configuration as the *single-ended* output configuration of the differential amplifier. (This is in distinction to the *differential* output configuration, where the output is taken between the two collector terminals.) In words, *the voltage gain of the single-ended configuration is half the ratio of collector to emitter resistance*.

In the case of this particular amplifier the transistors are biased at 1mA emitter current, so the value of r_e is 26Ω (section 29.16 on page 834). Then the single-ended voltage gain is:

$$\begin{aligned} A_d &= \frac{R_{c1}}{2r_e} \\ &= \frac{6900}{2 \times 25} \\ &= 138 \text{ volts/volt} \end{aligned} \quad (1448)$$

Example Design: Common-Mode Gain

Continuing with the previous design example, we now calculate the common-mode gain. For the common-mode gain calculation the two input terminals are connected together and driven from the same input. The AC equivalent circuit is shown in figure 830. Resistors r_{e1} and r_{e2} are in parallel, and in series with the much larger resistance R_{EE} . This time, the emitter resistances r_{e1} and r_{e2} are insignificant compared to R_{EE} . The input AC voltage appears across R_{EE} , creating an AC emitter current i_e . Because the circuit is symmetrical, with equal voltages across r_{e1} and r_{e2} , the emitter current divides equally between Q_1 and Q_2 . Then the output common-mode voltage is:

$$e_{ocm} = i_{c1} R_{c1} = \frac{e_{icm}}{2R_{EE}} R_{c1} \quad (1449)$$

Then the common-mode gain is given by:

$$A_{cm} = \frac{e_{ocm}}{e_{icm}} = \frac{R_{c1}}{2R_{EE}} \quad (1450)$$

In this case, the common-mode gain is:

$$A_{cm} = \frac{R_{c1}}{2R_{EE}} = \frac{6900}{2 \times 5500} = 0.627 \text{ volts/volt} \quad (1451)$$

Ideally, this value should be as close to zero as possible, but for a non-demanding application, this value might be acceptable.

The common-mode rejection ratio (CMRR) is the ratio of differential gain to common-mode gain, a useful indicator of the effectiveness of the differential action²⁶³. Using equations 1447 and 1450, we have the CMRR for this amplifier as:

$$CMRR = \frac{A_d}{A_{cm}} = \frac{R_{c1}/2r_e}{R_{c1}/2R_{EE}} = \frac{R_{EE}}{r_e} \text{ volts/volt} \quad (1452)$$

For the circuit of figure 829, the CMRR is:

$$CMRR = \frac{R_{EE}}{r_e} = \frac{6900}{25} = 276$$

In decibels:

$$CMRR_{db} = 20 \log_{10}(CMRR) = 20 \log_{10}(276) = 49 \text{ db}$$

A typical integrated circuit op-amp has a CMRR in the order of 80 db.

Input Resistance

The input resistance seen at the base of Q_1 can be determined from an examination of figure 829(c). By the reflection principle, the resistance in the emitter of Q_1 appears in the base circuit as if multiplied by $\beta + 1$. Then:

$$r_i = (\beta + 1)(r_{e1} + R_{EE} \parallel r_{e2}) \quad (1453)$$

²⁶³For CMRR in op-amps, see section 21.5 on page 649

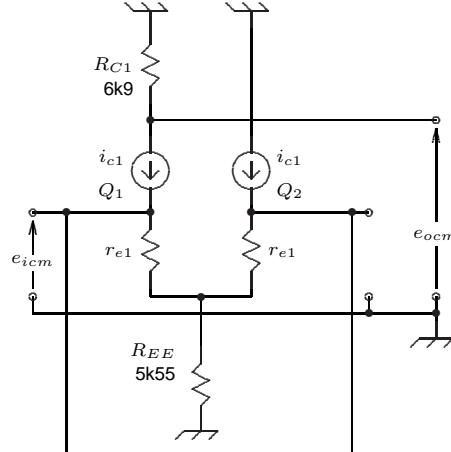


Figure 830: Common-Mode AC Analysis

We can make some assumptions and approximations to simplify this:

Assumption	Approximation
$\beta \gg 1$	$\beta + 1 \approx \beta$
$R_{EE} \gg r_e$	$R_{EE} \parallel r_{e2} \approx r_{e2}$
$I_{E1} \approx I_{E2}$	$r_{e1} \approx r_{e2} = r_e$

Then the input resistance is simply double the single, grounded emitter BJT case:

$$r_i \approx 2\beta r_e \quad (1454)$$

For example, with $\beta = 100$ and $r_e = 25\Omega$, $r_i \approx 2500\Omega$.

30.15 Enhanced Differential Pair

The arrangement shown in figure 831 brings several benefits to the differential pair. It improves the common-mode rejection and generates a single-ended current output²⁶⁴. The current output can be converted into a voltage by means of a load resistance or amplified in a subsequent stage. As we'll show, the overall transfer function is given by

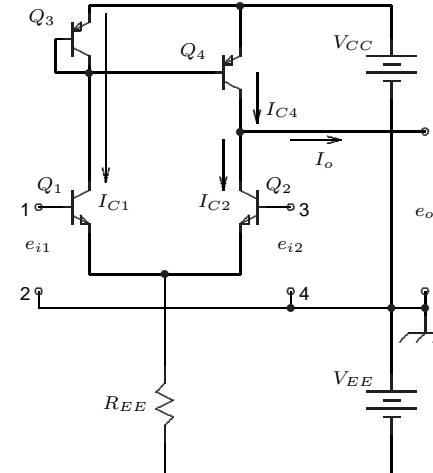
$$i_o = \frac{(e_{i1} - e_{i2})}{r_e} \quad (1455)$$

Since this stage is usefully regarded as a voltage-current converter, it's frequently referred to as a *transconductance* stage.

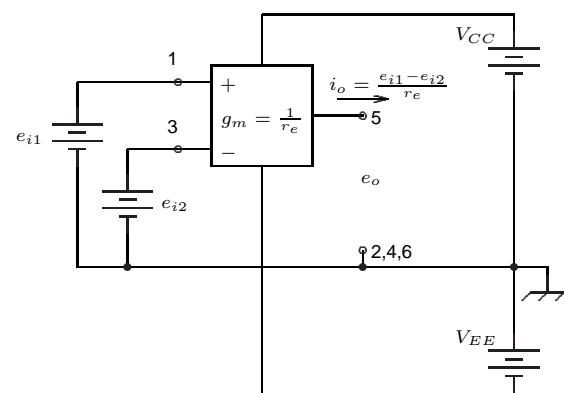
Circuit Operation

Diode-connected transistor Q_3 and transistor Q_4 form a current mirror. As a result, the collector current of differential pair transistor Q_2 is subtracted from the collector current of the differential pair transistor Q_1 at the output node 5. This achieves two results:

- The two DC collector bias currents I_{C1} and I_{C2} subtract at the output node, making the output current independent of the bias current and common-mode input voltage.
- The two AC collector currents i_{C1} and i_{C2} have opposite sense. That is, in response to some input voltage one current increases while the other decreases. The net effect is that the two AC collector currents add to form the output AC current i_o .



(a) Circuit



(b) Block Diagram

Figure 831: Transconductance Stage

²⁶⁴In figure 831 this current is pictured as heading off into empty space, which is obviously problematic. In practice, the effective load circuit is a constant voltage source (such as the base-emitter junction of a transistor) that behaves as an AC short circuit and ensures that Q_2 and Q_4 are not saturated or cutoff.

Common-Mode Operation

The DC output current is:

$$I_o = I_{C4} - I_{C2} \quad (1456)$$

But the mirror makes

$$I_{C4} = I_{C1} \quad (1457)$$

Also, since the stage is symmetrical:

$$I_{C2} = I_{C1} \quad (1458)$$

Combining these three equations, we have that

$$\begin{aligned} I_o &= I_{C4} - I_{C2} \\ &= I_{C1} - I_{C2} \\ &= I_{C1} - I_{C1} \\ &= 0 \end{aligned} \quad (1459)$$

That is, if the differential amplifier is balanced, then its collector currents cancel at the output terminal. Varying the common-mode voltage changes the tail current I_{EE} , but the collector currents remain equal, the stage remains balanced, and this has no effect on the output current.

However, for this to be true the input and output currents of the mirror must be equal. That's not quite true for the simple mirror shown in figure 831(a), but this mirror can be replaced by other circuits that more closely approach the ideal.

Differential Operation

This circuit can be analysed using small-signal AC quantities, but there is an alternate analysis that is more illuminating.

Consider that the stage is initially balanced with voltage E_i into both inputs. Then consider that the left input increases to $E_i + \Delta E$ while the right input is unchanged. The common-mode input increases (by $\Delta E/2$) but that has no effect on the output.

Then the collector current I_{C1} changes by the amount:

$$\Delta I_{C1} = \Delta E/r_e \quad (1460)$$

This change in collector current is reflected by the mirror and shows up as an increase of the same amount in the output current I_{C4} of the mirror.

The collector current of Q_2 is unchanged, so the entire change in mirror output current must flow into the output terminal:

$$I_o = \Delta E/r_e \quad (1461)$$

The change in input voltage ΔE is equal to the difference voltage between the two input terminals, so:

$$I_o = \frac{e_{i1} - e_{i2}}{r_e} \quad (1462)$$

Constant Tail Current

To this point, the differential amplifier has been shown with a tail current supply created by resistor R_{EE} and negative power supply V_{EE} . This arrangement has the virtue of simplicity, but two significant shortcomings:

- In circuits where the output is taken from a collector resistor, the finite value of the tail resistor leads to a modest value of common-mode rejection ratio. Increasing R_{EE} does not help because equation 1452 shows that CMRR is the ratio of R_{EE} to emitter incremental resistance r_e . Increasing R_{EE} decreases the emitter tail current, which causes r_e to increase in like amount.
- As the common-mode input voltage changes over a large range, this changes the voltage across the tail resistance R_{EE} , which changes tail current I_{EE} and the corresponding DC bias currents I_{C1} and I_{C2} in the transistors. This alters the value of r_e , which affects voltage gain, and the transistor operating point, which affects the current gain β .

Both these problems may be addressed by placing a current generator in the tail circuit, as shown in figure 832. A wide variety of current generators can be used for this task, but the current mirror of figure 832 is convenient. In this case, resistor R_1 sets the input current of the mirror in Q_4 . This is reflected by Q_3 to become constant current I_{EE} . By KVL:

$$+V_{EE} + V_{CC} - V_{R1} - V_{BE4} = 0 \quad (1463)$$

Rearranging and using $V_{R1} = I_{EE}R_1$, we can solve for the tail current:

$$I_{EE} = \frac{V_{CC} + V_{EE} - V_{BE4}}{R_1} \quad (1464)$$

For example, to obtain a tail current of 1mA with 12 volt supplies, we would make R_1 :

$$\begin{aligned} R_1 &= \frac{V_{CC} + V_{EE} - V_{BE4}}{I_{EE}} \\ &= \frac{12 + 12 - 0.6}{1 \times 10^{-3}} \\ &= 23k4\Omega \end{aligned}$$

Notice that the common-mode input voltage has no effect on the tail current until the emitters of Q_1 and Q_2 are lowered to the point where the tail current generator Q_3 becomes saturated.

Emitter Degeneration Resistors

In section 29.21, we showed the effect of adding a fixed emitter resistance R_e to a common-emitter amplifier stage. This resistance provides local negative feedback to the stage, so it is referred to as *emitter degeneration* resistance. Now, the incremental resistance r_e is non-linear – it varies with emitter current. This variation introduces distortion into the amplifier stage. However, if r_e is in series with a fixed resistor and the fixed resistor is significantly larger than the incremental resistance, then the fixed resistance dominates, reducing the distortion. The price to be paid for this is lower gain²⁶⁵.

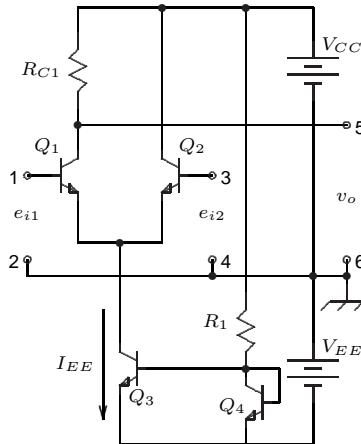


Figure 832: Constant Tail Current

²⁶⁵Douglas Self suggests another approach in [264]. Increase the bias current I_{EE} so that r_e is reduced by some amount. Then add a fixed resistance R_e to bring the total emitter resistance back to its original value. In effect, the gain remains at its original value while the bias

The effect of a fixed emitter resistor on the gain of the differential amplifier is same as the effect of a fixed emitter resistor on the gain single-stage BJT common-emitter amplifier: in the gain equation, replace the emitter incremental resistance r_e with $r_e + R_E$.

For example, for the circuit of figure 829(a) on page 940 we found that the single-ended voltage gain is:

$$A_d = \frac{R_{c1}}{2r_e} \quad (1465)$$

When degeneration resistors $R_{E1} = R_{E2} = R_E$ are installed as in figure 833, the single-ended differential voltage gain becomes:

$$A_d = \frac{R_{c1}}{2(r_e + R_E)} \quad (1466)$$

Since emitter incremental resistance r_e is usually in the order of a few tens of ohms, the degeneration resistors R_E can be in the order of tens or hundreds of ohms to be effective. In general, this resistance is sufficiently small that it has no material effect on the biasing of the stage and the DC voltage across R_E can be ignored.

Single Gain-Setting Resistor

The configuration of figure 825 uses two equal resistors R_{E1}, R_{E2} to set the gain of the differential pair. Figure 834 shows how this can be accomplished with one resistor R_E . This is particularly useful when the gain must be adjusted – this circuit requires only one adjustable resistor.

There are now two constant current sinks Q_{3A} and Q_{3B} , each associated with one of the differential pair transistors. In figure 834 the current sources are a two-output current mirror where the output current is set by the current through resistor R_1 . The AC differential gain of the circuit is given by:

$$A_d = \frac{R_{c1}}{2r_e + R_E} \quad (1467)$$

This arrangement surfaces again in the predistortion circuitry of the Gilbert transconductance multiplier, section 36.4, page 1100.

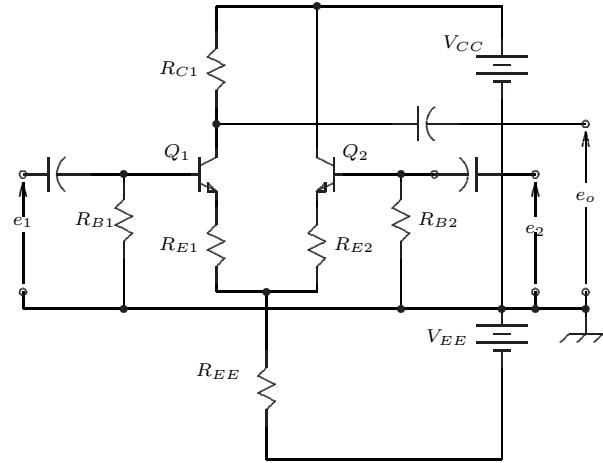


Figure 833: Emitter Degeneration Resistors R_E

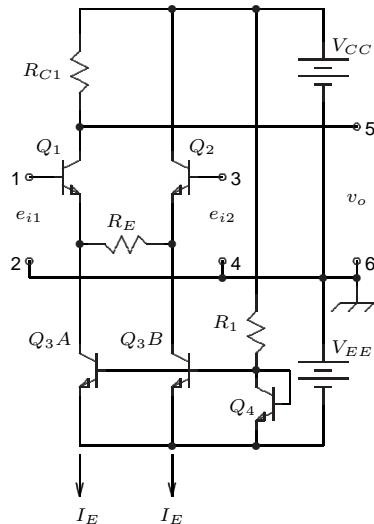


Figure 834: Single Gain-Setting Resistor

current increases. Self measured a reduction of distortion by a factor of 10 using this technique. The increased bias current in the differential stage increases the available drive current to subsequent stages, which may improve the slew rate of the amplifier. If the differential amplifier is the input stage of an amplifier, then it also increases the bias current, which may be an issue.

30.15.1 Darlington-Differential

The differential amplifier can be combined with other building-blocks in various interesting ways. For example, figure 835 shows the input stage of a single-supply operational amplifier, the LM124 [270].

This amplifier uses PNP darlington pair transistors Q_1 , Q_2 and Q_3 , Q_4 in place of the single NPN transistors we've seen in previous differential amplifier circuits. The differential output current I_o is generated by the NPN current mirror Q_5 , Q_6 .

It's a challenge to construct an operational amplifier where the common-mode input voltage can extend to the negative supply – in this case zero volts, the ground potential. As indicated in the figure by the boxed voltages, this amplifier can function correctly under that condition. We can verify that as follows:

1. The minimum voltage across the diode-connected transistor Q_5 is the base-emitter voltage of Q_6 , +0.6 volts.
2. The minimum voltage across the collector-emitter of Q_2 is its saturation voltage, which we'll take as +0.2 volts. Then point 2 is at +0.8 volts.
3. The voltage at the base of Q_2 , which is the collector-emitter voltage of Q_1 , is one base-emitter drop below the emitter of Q_2 , that is, +0.2 volts. This is just sufficient to keep Q_1 out of saturation.
4. The base voltage of Q_1 , which is the input voltage, is another base-emitter drop below its emitter voltage, that is, -0.4 volts. Consequently, the amplifier should be able to function correctly with either or both input terminals at or slightly below zero volts.

This is an integrated circuit, and so the transistors can be matched quite closely to each other. If this circuit were to be constructed in a discrete form, one would add resistors in the current mirror to swamp any mismatch in base-emitter voltages. As well, one might add gain-setting resistors in the differential pair, and flushout resistors in each Darlington pair.

30.15.2 Cascode-Differential

As explained in section 30.12, a cascode pair consists of a common-emitter stage driving a common-base stage. The cascode pair minimizes the effect of collector-base capacitance, thereby improving the frequency response of the stage. It's also useful where large supply voltage is required – perhaps to generate a large output voltage swing.

The cascode pair may be incorporated into the differential amplifier, as shown in the example of figure 836.

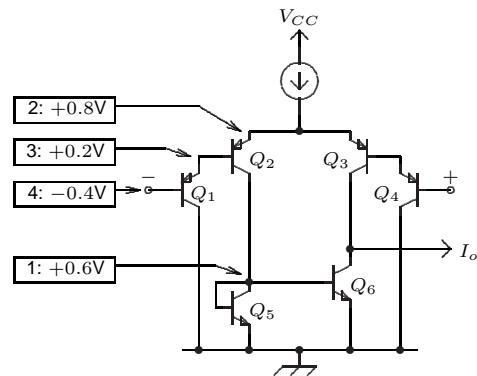


Figure 835: LM124 Input Stage

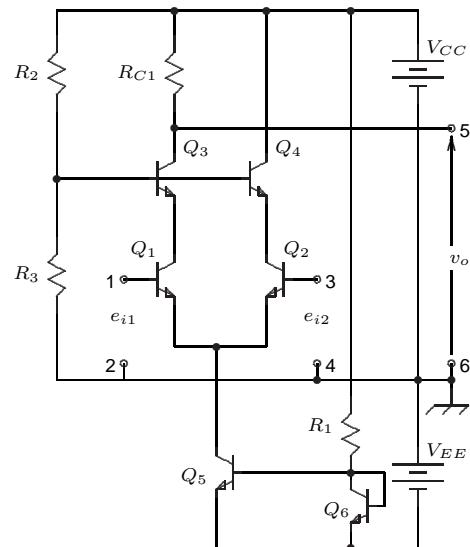


Figure 836: Cascode Differential

The cascode pairs Q_1, Q_3 and Q_2, Q_4 effectively function as single BJT devices with enhanced bandwidth capability. Resistors R_2 and R_3 set the base voltage of the common-base stages Q_3, Q_4 such that there is sufficient voltage to prevent saturation in the common emitter stages Q_1, Q_2 .

For example, suppose the maximum input voltage at the base terminal of Q_1, Q_2 is 2 volts. Then, assuming a transistor saturation voltage of 0.2 volts, the base voltage of the common-base stage must be at least 2.2 volts.

The common-base stages simply relay their input (emitter) current to their (collector) output. Consequently, the single-ended gain is unchanged from figure 827, and is given by:

$$A_d = \frac{R_{c1}}{2r_e} \quad (1468)$$

30.15.3 Differential Stage Coupling

An individual BJT common-emitter amplifier stage creates a significant level shift between its input and output terminals. In order to accommodate a voltage swing at the collector of the transistor, there must be a bias voltage across the transistor equal to or greater than the peak value of the output AC voltage swing. In general, this output bias voltage is not convenient. The DC bias can be removed with a coupling capacitor, but this restricts the operation of the amplifier to frequencies where the capacitance is a low impedance. When the amplifier is DC coupled, the signal must be *level shifted* to an appropriate bias level for the subsequent stage.

However, because a differential amplifier stage is not sensitive to common-mode input voltage²⁶⁶, the differential stage solves this problem. Consider for example the two-stage DC coupled differential amplifier of figure 837.

The first stage of this amplifier is the differential pair Q_1, Q_2 . The collectors of Q_1, Q_2 are probably biased to some voltage between zero and V_{CC} volts. This bias voltage forms the common-mode input voltage for the following differential amplifier stage Q_3, Q_4 so it is largely ignored by the second stage. The operating current for the second stage is determined by the tail current generator I_{EE2} . The value of collector resistor R_3 then determines the quiescent output voltage.

This technique is common in the design of discrete component preamplifiers of analog oscilloscopes.

30.15.4 Appendix: Differential Amplifier Transfer Characteristic

Here we develop the equations that relate the differential input voltage V_d to the relative magnitudes of the two collector currents, leading to the transfer characteristic shown in figure 826.

We start with the equation relating the BJT base-emitter voltage to its collector current (equation 1158 on page 807):

$$I_c = I_s e^{V_{be}/V_t} \quad (1469)$$

²⁶⁶Within the usual limits that the transistors are not saturated or cutoff.

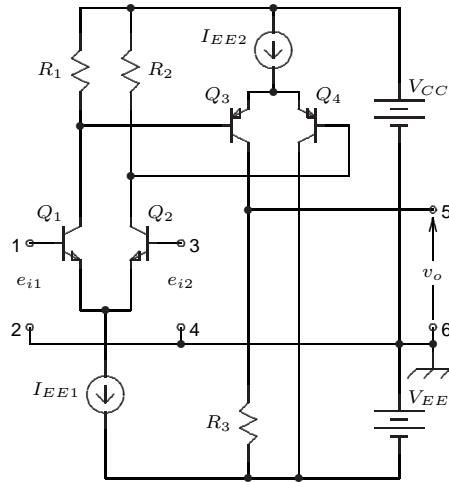


Figure 837: Two Stage Diff Amp

where

- V_{be} Base-emitter voltage, volts
- V_t Transistor thermal voltage, 26mV at room temperature
- I_c Collector current, amperes
- I_s Saturation current for the transistor, typically 10^{-12} to 10^{-15} amps.

Then the base emitter voltages of the two transistors in the differential pair may be written as:

$$I_{c1} = I_{s1} e^{V_{be1}/V_t} \quad (1470)$$

$$I_{c2} = I_{s2} e^{V_{be2}/V_t} \quad (1471)$$

For identical transistors, the two saturation currents are equal and

$$I_{s1} = I_{s2} = I_s \quad (1472)$$

Taking the ratio of equations 1470 and 1471, the saturation currents cancel and the ratio of the two collector currents is:

$$\begin{aligned} \frac{I_{c1}}{I_{c2}} &= \frac{e^{V_{be1}/V_t}}{e^{V_{be2}/V_t}} \\ &= e^{(V_{be1}-V_{be2})/V_t} \end{aligned} \quad (1473)$$

Taking KVL around the loop shown in figure 838:

$$+V_{B1} - V_{be1} + V_{be2} - V_{B2} = 0$$

Rearranging:

$$\begin{aligned} V_{be1} - V_{be2} &= +V_{B1} - V_{B2} \\ &= V_d \end{aligned}$$

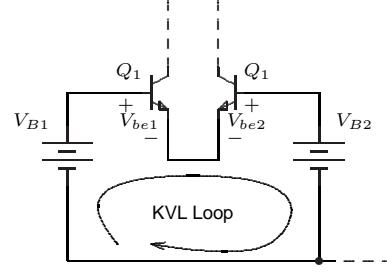


Figure 838: V_{be} Loop

That is, the quantity $V_{be1} - V_{be2}$ is the voltage difference between the two base terminals, V_d . Substituting V_d for $V_{be1} - V_{be2}$ in equation 1473, we have:

$$\frac{I_{c1}}{I_{c2}} = e^{\frac{V_d}{V_t}} \quad (1474)$$

A more useful version of this relationship would show the collector current as a ratio of the tail current, as a function of the differential input voltage. By KCL, the collector currents must sum to alpha times the tail current:

$$\begin{aligned} I_{c1} + I_{c2} &= \alpha I_{EE} \\ &\approx I_{EE} \text{ (since } \alpha \approx 1) \end{aligned} \quad (1475)$$

Rearrange equation 1475 to solve for I_{c2} and substitute for I_{c2} in equation 1474:

$$\frac{I_{c1}}{I_{EE} - I_{c1}} = e^{\frac{V_d}{V_t}} \quad (1476)$$

With some algebraic manipulation, this can be rearranged into the desired form:

$$\begin{aligned}\frac{I_{c1}}{I_{EE}} &= \frac{e^{V_d/V_t}}{1 + e^{V_d/V_t}} \\ &= \frac{1}{1 + e^{-V_d/V_t}}\end{aligned}\quad (1477)$$

As a quick check on this result:

- For the differential input voltage V_d equal to $3V_t$, $I_{c1} = 0.95I_{EE}$, that is, the left side is conducting most of the tail current.
- For the differential input voltage V_d equal to 0, both base terminals are at the same potential. The collector current $I_{c1} = 0.5I_{EE}$, that is, both sides are conducting equally.
- For the differential input voltage V_d equal to $-3V_t$, $I_{c1} = 0.05I_{EE}$, that is, the left side is conducting lightly, and most of the tail current has shifted to the right side of the pair.

In a similar manner, it may be shown that the collector current in the right hand transistor Q_2 is given by

$$\begin{aligned}\frac{I_{c2}}{I_{EE}} &= \frac{e^{V_d/V_t}}{1 + e^{V_d/V_t}} \\ &= \frac{1}{1 + e^{-V_d/V_t}}\end{aligned}\quad (1478)$$

Equations 1477 and 1478 give the transfer characteristic of figure 826.

30.15.5 The \tanh Function

When the output is the difference between the two collector currents, as in figure 831, then we have the following for the output current:

$$I_o = I_{C1} - I_{C2} \quad (1479)$$

Substituting for I_{C2} and I_{C1} from equations 1477 and 1478, we have:

$$I_o = I_{EE} \left(\frac{1}{1 + e^{-V_d/V_t}} \right) - I_{EE} \left(\frac{1}{1 + e^{V_d/V_t}} \right) \quad (1480)$$

For notational convenience, put $V_d/V_t = x$. Then:

$$\begin{aligned}I_o &= I_{EE} \left(\frac{1}{1 + e^{-x}} \right) - I_{EE} \left(\frac{1}{1 + e^{+x}} \right) \\ &= I_{EE} \left(\frac{1 + e^{+x} - 1 - e^{-x}}{(1 + e^{-x})(1 + e^{+x})} \right) \\ &= I_{EE} \left(\frac{e^{+x} - e^{-x}}{1 + e^{+x} + e^{-x} + e^0} \right) \\ &= I_{EE} \left(\frac{e^{+x} - e^{-x}}{e^{+x} + e^{-x} + 2} \right)\end{aligned}\quad (1481)$$

Use the relationship $a^2 - b^2 = (a - b)(a + b)$ on the numerator. Use the relationship $a^2 + 2ab + b^2 = (a + b)^2$ on the denominator. Then we have:

$$\begin{aligned} I_o &= I_{EE} \left(\frac{(e^{+x/2} - e^{-x/2})(e^{+x/2} + e^{-x/2})}{(e^{+x/2} + e^{-x/2})^2} \right) \\ &= I_{EE} \left(\frac{e^{+x/2} - e^{-x/2}}{e^{+x/2} + e^{-x/2}} \right) \end{aligned} \quad (1482)$$

The term in the brackets is the *hyperbolic tangent* of $x/2$. Back substituting for x , we have:

$$I_o = I_{EE} \tanh \left(\frac{V_d}{2V_t} \right) \quad (1483)$$

Consequently, the output current is said to be a tanh function of the differential input voltage.

30.15.6 Appendix: Dual Stage Differential Amplifier Bias and Gain

Here we show one plausible development for a two-stage differential amplifier. To start, we'll assume that the transistors have been chosen. They have a satisfactory current gain at a current of 1mA and acceptable voltage ratings. The amplifier should produce as large as possible a voltage swing at the output. The supply voltages V_{CC} and V_{EE} are 15 volts.

For the transistors used in this circuit assume that the operating point is $I_c = 1\text{mA}$. At that operating point, the current gain $\beta = 100$, emitter incremental resistance $r_e = 25\Omega$, and collector incremental resistance $r_c = 100k\Omega$.

Bias

- Put the quiescent output voltage at zero, midway between the two supplies. The current through collector resistor R_3 is 1mA and it has 15 volts across it, so $R_3 = 15k\Omega$.
- We would like the output to be able to swing as close as possible to the positive supply rail, which requires that the common-mode input voltage to the second stage be set at the largest possible voltage. Allow one saturation drop across the constant current source I_{EE2} and a further base-emitter drop of 0.6 volts at Q_3, Q_4 , and the common-mode voltage is a maximum of +14.2 volts. Put it at 14 volts.
- The voltages across R_1, R_2 are equal to the difference between the positive supply and the base voltage of the Q_3, Q_4 pair: one volt. The collector current in the first differential pair is 1mA. So these resistors $R_1 = R_2 = 1k\Omega$.

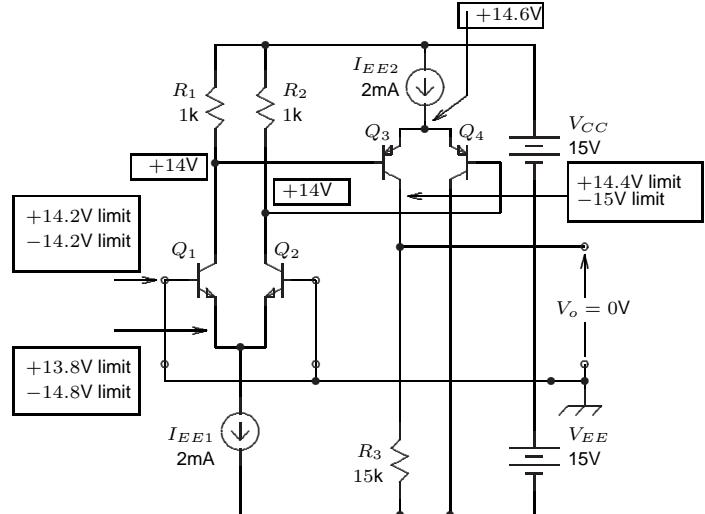


Figure 839: Two Stage Diff Amp: Bias

That's it. The maximum output positive voltage swing is equal to the positive supply voltage minus two saturation voltages, one across the constant current source I_{EE2} , and one across the differential transistor Q_3 : about 14.6 volts. The maximum output negative voltage swing is equal to the negative supply voltage: -15 volts.

The maximum positive common-mode input is determined by the collector voltage (14 volts), and the maximum negative common-mode input by saturation of the current source I_{EE1} .

Voltage Gain

A good place to start is the input resistance of the second stage. This is beta times the emitter incremental resistance:

$$r_{i3} = \beta_3 r_{e3} = 100 \times 25 = 2k5\Omega$$

Now we can calculate the load resistance seen at the collector of the first stage.

$$r_{l1} = R_1 \parallel r_{c1} \parallel r_{i3} = 1k \parallel 100k \parallel 2k5 = 709\Omega$$

The first stage is differential in, differential out, so its voltage gain is given by the ratio of collector load resistance to emitter resistance.

$$A_{d1} = \frac{e_{od1}}{e_{id1}} = \frac{r_{l1}}{r_{e1}} = \frac{709\Omega}{25\Omega} = 28 \text{ volts-volt}$$

The second stage voltage gain is differential in, single ended out, so its voltage gain is the ratio of its collector resistance $R_3 \parallel r_{c3}$ to twice the emitter incremental resistance r_{e3} .

$$A_2 = \frac{v_o}{e_{id2}} = \frac{R_3 \parallel r_{c3}}{2r_{e3}} = \frac{15k \parallel 100k}{25} = 522 \text{ volts-volt}$$

The overall voltage gain is the product of the two voltage gains:

$$A_v = \frac{v_o}{e_{id1}} = \frac{e_{od1}}{e_{id1}} \frac{v_o}{e_{id2}} = 28 \times 522 = 14616 \text{ volts-volt}$$

For reference, the symbols used in this development are:

- r_{i3} Input resistance looking into the base of transistor Q_3 , equal to β_3 times the incremental resistance r_{e3}
- β_3 Current gain of transistor Q_3
- r_{l1} Total resistance seen at the collector of transistor Q_1
- r_{e3} Emitter incremental resistance of transistor Q_3 , approximately equal to $0.026/I_{E3}$. In this case, I_{E3} is 1mA, so $r_{e3} = 25\Omega$
- I_{E3} DC bias current in transistor Q_3 , 1mA
- e_{od1} Differential output voltage of first stage, measured between the collector terminals of transistors Q_1 and Q_2
- e_{id1} Differential input voltage of first stage, measured between the base terminals of transistors Q_1 and Q_2
- R_1 Value of the collector resistor for transistor Q_1 , $1k\Omega$
- r_{c1} Collector incremental resistance of transistor Q_1 , a property of the transistor, $100k\Omega$
- r_{e1} Emitter incremental resistance of transistor Q_1 , $r_{e1} = 25\Omega$
- R_3 Value of the collector resistor for transistor Q_3 , $15k\Omega$
- r_{c3} Collector incremental resistance of transistor Q_3 , $100k\Omega$
- v_o Voltage swing at the output terminal of the amplifier

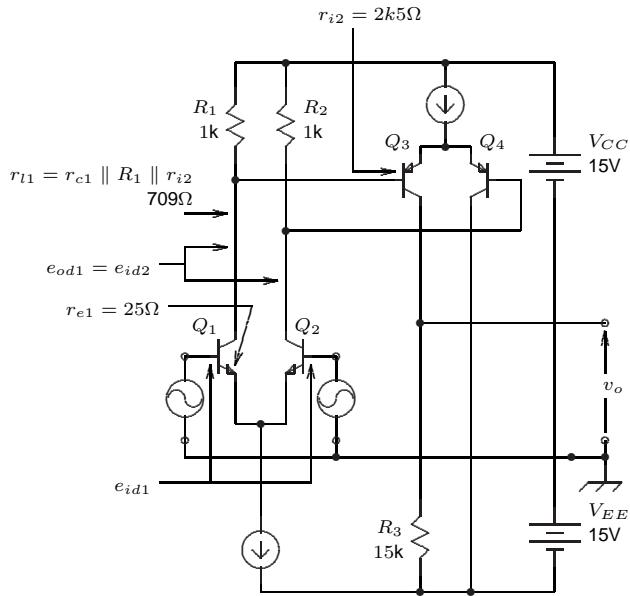


Figure 840: Two Stage Diff Amp: Gain

Final Circuit

Figure 841 shows the completed two-stage amplifier, with implementation of the two constant current sources. Source I_{EE1} is created by the current mirror Q_7, Q_8 . Source I_{EE2} is created by the current mirror Q_5, Q_6 . Resistor R_4 establishes the reference current of 1mA for both mirrors.

Is this a useful circuit? As a general purpose op-amp, the bias current (the base current of the Q_1, Q_2 differential pair) is $10\mu\text{A}$, which is very large by modern standards. Constructed with discrete transistors, there is likely to be significant difference between the base-emitter voltages of the two input transistors, and this appears as offset voltage in the op-amp.

An integrated circuit, with both transistors on the same substrate, will have much better match and smaller offset voltage. The output resistance, seen looking back into the output terminal of the op-amp, is in the order of $15\text{k}\Omega$, which is unacceptably large for most applications. At a minimum, an emitter follower would be required to lower the output resistance. (It could be biased by a current sink similar to Q_7 .)

Circuits similar to this are found as the vertical amplifier of an analog oscilloscope. However, those differential amplifiers are heavily degenerated with emitter resistance in order to increase the frequency bandwidth, reduce the distortion and establish the voltage gain as a function of fixed resistor ratios.

30.16 Series-Shunt Feedback Pair

The *series-shunt pair* is an arrangement that has extended high-frequency response [271]. It is commonly found in the vertical and horizontal amplifiers of analog oscilloscopes [272], [273]. The basic circuit is shown in figure 842. Our main interest here is the AC performance of the circuit, so we will ignore the issue of biasing this amplifier, and assume that some other circuitry establishes bias currents in the amplifiers. (Figure 863 on page 972 shows a complete integrated circuit, the $\mu\text{A}733$ differential amplifier, which contains suitable bias circuitry. The values in figure 842 are from this circuit.)

This circuit is not particularly common, but it is a useful example to illustrate how the basic gain equations of the single stage BJT amplifier can be applied to a multi-stage amplifier.

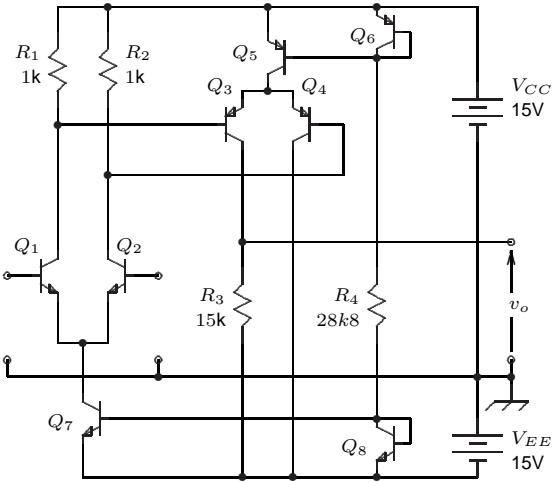


Figure 841: Two Stage Diff Amp: Final

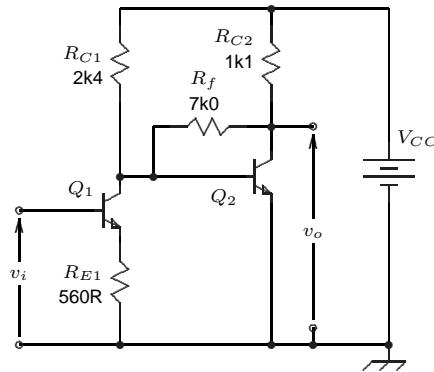


Figure 842: Series-Shunt Pair

Voltage Gain

With the second stage disconnected as shown in figure 843(a), the voltage gain of the first stage is the ratio of collector to emitter resistance:

$$\begin{aligned} A_{v1} &= -\frac{R_{C1}}{r_{e1} + R_{E1}} \\ &\approx -\frac{R_{C1}}{R_{E1}} \end{aligned} \quad (1484)$$

The output resistance of this stage is:

$$\begin{aligned} r_o &= R_{C1} \parallel r_{c1} \\ &\approx R_{C1} \end{aligned} \quad (1485)$$

Now consider the second stage, figure 843(b). It may be considered as an inverting operational amplifier comprised of transistor Q_2 and collector resistor R_{C2} . The input resistance R_i is the output resistance of the previous stage, and the feedback resistance is R_f . If the open-loop gain is much smaller than the closed-loop gain, and if the output resistance of this stage is much smaller than the feedback resistance, then the overall voltage gain of the second stage is:

$$\begin{aligned} A_{v2} &= -\frac{R_f}{R_i} \\ &= -\frac{R_f}{R_{C1}} \end{aligned} \quad (1486)$$

where the collector resistance R_{C1} in figure 843(a) is equivalent to the input resistance R_i as shown in figure 843(b).

The total gain is the product of equations 1484 and 1486:

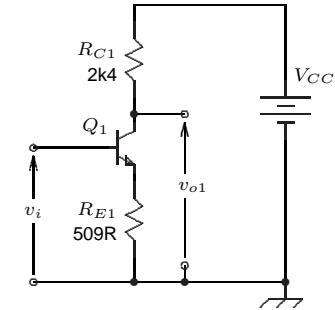
$$\begin{aligned} A_v &= A_{v1} \cdot A_{v2} \\ &= \left(-\frac{R_{C1}}{R_{E1}} \right) \cdot \left(-\frac{R_f}{R_{C1}} \right) \\ &= \frac{R_f}{R_{E1}} \end{aligned} \quad (1487)$$

Alternative Formulation of Voltage Gain

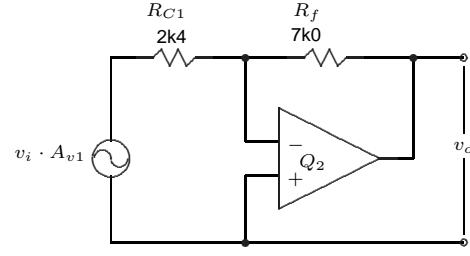
Another way of looking at this amplifier is the following:

The input voltage v_i appears across the emitter resistor of the first stage, and creates a collector current in the transistor of:

$$\begin{aligned} i_{c1} &= \alpha \frac{v_i}{R_{E1}} \\ &\approx \frac{v_i}{R_{E1}} \end{aligned} \quad (1488)$$



(a) First Stage



(b) Second Stage

Figure 843: Series-Shunt Voltage Gain

The collector of the first transistor is connected to the virtual earth point of the second-stage operational amplifier. Consequently, that stage is fixed in voltage, so all of the first stage AC current flows into that node. The op-amp forces this current to flow through the feedback resistor R_f . Using equation 1488 for i_{c1} , the output voltage is:

$$\begin{aligned} v_o &= R_f i_{c1} \\ &= R_f \frac{v_i}{R_{E1}} \end{aligned} \quad (1489)$$

The overall gain of the amplifier is the ratio of output to input voltage:

$$\begin{aligned} A_v &= \frac{v_o}{v_i} \\ &= \frac{R_f}{R_{E1}} \end{aligned} \quad (1490)$$

High Frequency Performance

This amplifier extends frequency response by reducing the Miller Effect capacitance in both stages.

In the first stage, there is very little collector voltage swing because the collector is connected to the virtual earth point of the operational-amplifier second stage. Consequently, the collector-base capacitance appears between the base terminal and ground, essentially unmultiplied.

In the second stage, there is a large increase from the voltage swing at the base terminal to the voltage swing at the collector terminal and so this stage has a large voltage gain. As a result, the collector-base capacitance of the second transistor appears between the base terminal and ground, multiplied by the voltage gain, a large number. However, the feedback resistance R_f is also subject to the Miller Effect, and it appears between the base terminal and ground divided by the voltage gain. The result is to lower the source resistance, and this mitigates the effect of the Miller capacitance.

Assumptions and Approximations

The preceding analysis makes a number of assumptions and approximations. This is legitimate when doing a first-cut analysis or design. However, the assumptions should be checked carefully by a more rigorous analysis and/or by simulating the circuit. This is particularly true if the high-frequency performance of the circuit is important.

The important assumptions in the foregoing development are as follows:

- The discrete emitter resistance R_{E1} is large enough compared to the emitter incremental resistance r_{e1} that the latter may be ignored.
- The voltage gain of the second stage is large enough that it may be considered to be an operational amplifier, that is, the open-loop gain is large compared to the closed-loop gain.

The open-loop gain of the second stage is the ratio of its collector to emitter resistance:

$$A_{v2OL} = \frac{R_{C2}}{r_{e2}} \quad (1491)$$

The closed-loop gain is given by equation 1486 above.

- The output resistance of the operational amplifier must be low compared to its load resistance. The output resistance of the op-amp stage is $R_{C2} \parallel r_{c2} \approx R_{C2}$. The load resistance is the feedback resistance R_f .

- The input resistance of the operational amplifier $\beta_2 r_{e2}$ is large compared to the Millerized feedback resistance $R_f/(1 + A_{v2OL})$.
- The output resistance of the first stage is large compared to the Millerized feedback resistance $R_f/(1 + A_{v2OL})$.

The first assumption is easy to check. The other assumptions are not entirely obvious, so it's a good idea to do a complete analysis of that stage. Here it is.

Appendix: Gain of the Shunt-Feedback BJT Amplifier

The circuit to be analysed is shown in figure 844(a). (The component labels have been simplified to facilitate decoding the formulae.) The circuit values are from the μ A733 Differential Amplifier integrated circuit. The value of r_e is based on emitter current of 1.75mA. We'll take beta = 100.

We will determine how closely this circuit approximates the behaviour of an inverting op-amp circuit, for which the voltage gain would be $-R_f/R_i$.

The equivalent circuit is shown in figure 844(b). The transistor has been replaced by its equivalent circuit and the supply replaced by a short-circuit to the circuit common. Certain things are known about this circuit, so they have been incorporated into the figure. For example, the direction of currents is known. As well, this is an inverting amplifier, so we know that the polarity of v_o is as shown in the figure. Defining v_o in this manner aids in the physical understanding of the circuit. For example, it can be seen from the diagram that the total voltage across R_f is the sum of v_A and v_o .

If we defined v_o in the opposite polarity, our gain equation would contain a leading minus sign.

Now we can begin to write equations for this circuit. First, we'll obtain the ratio of output to base voltage, v_o/v_A . Then we'll determine the input resistance R_A looking into the virtual earth point. Finally, we'll determine the ratio of output to input voltage v_o/v_i , and compare that with an operational amplifier.

Step 1: Open-Loop Gain

The first step is to determine the open-loop gain v_o/v_A of the op-amp.

$$v_o = i_o R_c \quad (1492)$$

$$i_o = i_c - i_f \quad (1493)$$

$$\begin{aligned} i_c &= \alpha i_e \\ &\approx i_e \end{aligned} \quad (1494)$$

$$v_A = i_e r_e \quad (1495)$$

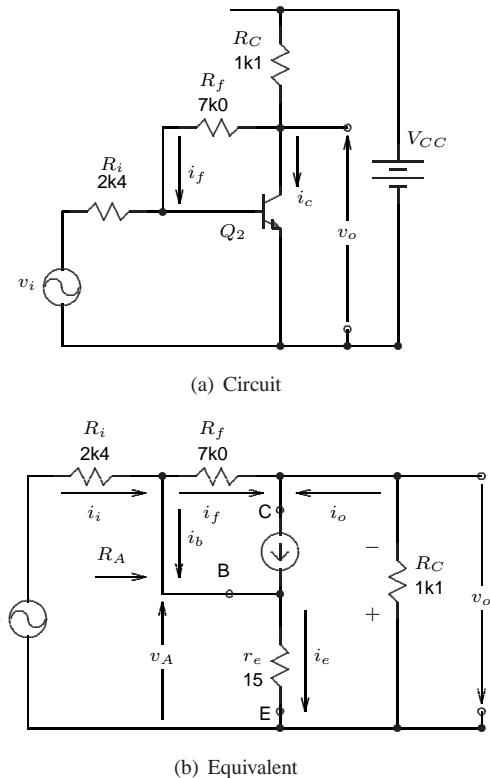


Figure 844: Series-Shunt Gain Analysis

$$i_f = \frac{v_A + v_o}{R_f} \quad (1496)$$

Substitute from equation 1495 into equation 1494 for i_e :

$$i_c = \frac{v_A}{r_e} \quad (1497)$$

Substitute from equation 1497 for i_c and from equation 1496 for i_f , into equation 1493:

$$i_o = \frac{v_A}{r_e} - \frac{v_A + v_o}{R_f} \quad (1498)$$

Substitute from equation 1498 for i_o into equation 1492, and after some manipulation:

$$v_o = R_c \left(\frac{v_A}{r_e} - \frac{v_A}{R_f} - \frac{v_o}{R_f} \right) \quad (1499)$$

Solve for the gain v_o/v_A :

$$\begin{aligned} \frac{v_o}{v_A} &= \left(\frac{R_f R_c}{R_f + R_c} \times \frac{1}{r_e} \right) - \left(\frac{R_c}{R_f + R_c} \right) \\ &= \frac{R_f \parallel R_c}{r_e} - \frac{R_c}{R_f + R_c} \end{aligned} \quad (1500)$$

Now we should see what we can throw out. Using the typical values from figure 844, the first term is *much* larger than the second term. So (introducing a new variable A to simplify the notation) we can write:

$$\begin{aligned} A &= \frac{v_o}{v_A} \\ &\approx \frac{R_f \parallel R_c}{r_e} \end{aligned} \quad (1501)$$

This looks reasonable. The gain between the virtual earth node and the output is the usual ratio of collector to emitter resistance, where the collector resistance is R_c and R_f in parallel. The quantity A is in effect the open-loop gain of the operational amplifier in this circuit.

Step 2: Input Current

There are two components to the input current: the feedback current i_f and the base current i_b . First we'll consider the feedback current:

From equation 1502:

$$i_f = \frac{v_A + v_o}{R_f} \quad (1502)$$

But by equation 1501, we defined $v_o = Av_A$. Substituting for v_o in equation 1502, and rearranging we have

$$\begin{aligned} \frac{v_A}{i_f} &= \frac{R_f}{1 + A} \\ &\approx \frac{R_f}{A} \end{aligned} \quad (1503)$$

That is, the current flowing into the feedback resistor sees a resistance that is apparently $1 + A$ times smaller than the physical resistance R_f . This is the Millerized feedback resistance.

Now for the base current. The incremental emitter resistance r_e reflects into the base circuit, multiplied by $\beta + 1$. So the resistance seen by the base current is $(\beta + 1)r_e$.

The resistance seen into the feedback resistor and the resistance seen into the base terminal are in parallel. For the typical values of our circuit, the base resistance is *much* larger than the Millerized feedback resistance, and so it may be neglected.

Step 3: Closed-Loop Gain

Referring to figure 845, the Millerized feedback resistance and the input resistance R_i form a voltage divider, so that the virtual earth voltage is:

$$\begin{aligned} v_A &= \left(\frac{R_f/A}{R_f/A + R_i} \right) v_i \\ &= \left(\frac{R_f}{R_f + AR_i} \right) v_i \end{aligned} \quad (1504)$$

The output voltage v_o is A times the virtual earth voltage:

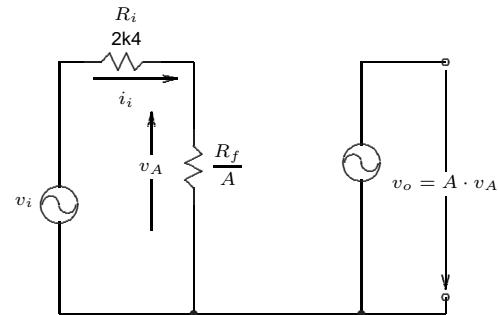


Figure 845: Closed-Loop Gain

$$\begin{aligned} v_o &= A \cdot v_A \\ &= A \left(\frac{R_f}{R_f + AR_i} \right) v_i \\ &= \left(\frac{R_f}{R_f/A + R_i} \right) v_i \end{aligned} \quad (1505)$$

The Millerized feedback resistance R_f/A is much smaller than the input resistance R_i , and so the closed-loop gain G of this amplifier is:

$$\begin{aligned} G &= \frac{v_o}{v_i} \\ &\approx \frac{R_f}{R_i} \end{aligned}$$

which corresponds to the original approximation²⁶⁷.

²⁶⁷The minus sign, which indicates inversion of the signal, is missing in this equation because of our original assumption in figure 844(b) of the polarity of v_o .

30.17 An Operational Amplifier Design

Discrete component operational amplifiers are primarily of tutorial and historical interest [274], but there are some niche applications that continue to require discrete amplifiers. For example, amplifiers with high-power or high-voltage output [275], [276], and amplifiers with very low noise [277] may require discrete component design.

In this section, we'll look at a simple discrete component operational amplifier circuit and then consider some of the tradeoffs in its design.

Along the way, we'll relate the specifications of the integrated circuit op-amp (gain-bandwidth, slew rate, bias current and so on) to the internal construction of this particular operational amplifier.

Although the focus of this material is on the discrete-component operational amplifier, many of the ideas in this section are transferable to the integrated circuit amplifier.

Much of this presentation reflects work by Solomon [84].

The operational amplifier circuit is shown in figure 846(a). There are three stages:

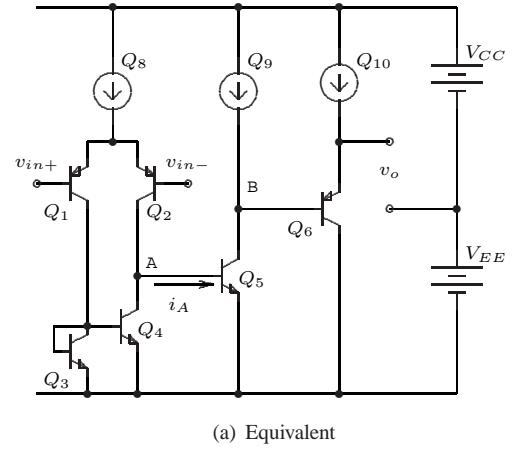
- The input stage, transistors Q_1 through Q_4 , is a PNP differential amplifier with an active current mirror load in the collector. This stage produces an output current at point A according to equation 1462 (page 944) which is:

$$i_A = \frac{v_{in+} - v_{in-}}{r_e} \quad (1506)$$

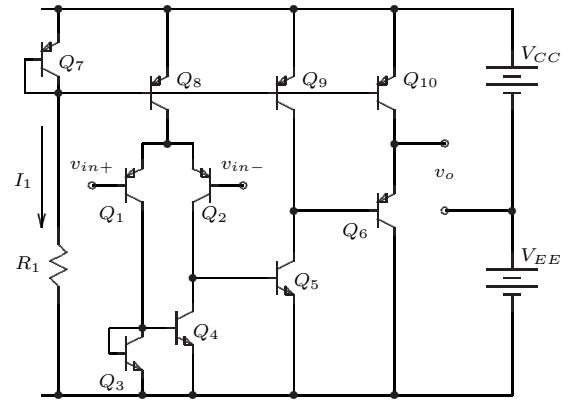
- The second stage, Q_5 , is a current amplifier: its collector current is simply beta times its base current. The output current is translated into a voltage by its collector resistance, the current source Q_9 . The effective value of the resistance at this point will be the incremental collector resistance of Q_5 in parallel with the resistance looking into the output terminal of the current source Q_9 . If we call this point B, then the voltage swing at this point (ignoring any load resistance) is

$$v_B = i_A \beta_5 r_{c5} \parallel r_{o9} \quad (1507)$$

- Point B could be used as an output terminal, but this is a high-impedance point in the circuit. As a source, it would have very high internal resistance ($r_{c5} \parallel r_{o9}$), and any load resistance would change the output voltage. Consequently, some sort of large-input-resistance, small-output-resistance buffer is required. The third stage, an emitter follower Q_6 , provides the buffer function. The emitter follower has a voltage gain of approximately unity and a current gain of beta. Consequently, the load resistance seen at the base of Q_6 is β_6 times the load resistance, and the internal resistance seen at the output terminal of the op-amp is the



(a) Equivalent



(b) Operational Amplifier

Figure 846: Operational Amplifier

resistance at point B, divided by β_6 . Without a buffer, the load resistance not only loads down the output voltage but appears between point B and ground, and thereby affects the voltage gain of the Q_5 stage.

Voltage Gain

A block diagram of the amplifier is shown in figure 847.

We can obtain an expression for the differential voltage gain of the amplifier by combining equations 1506 and 1507:

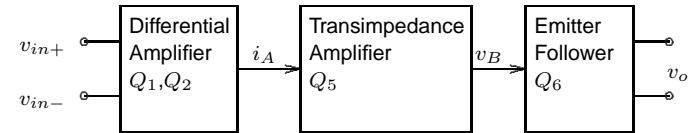


Figure 847: Block Diagram

$$\begin{aligned}
 A_D &= \frac{v_o}{v_{in+} - v_{in-}} \\
 &= \frac{v_B}{i_A r_{e1}} \\
 &= \frac{\beta_5(r_{c5} \parallel r_{o9})}{r_{e1}}
 \end{aligned} \tag{1508}$$

Example

Calculate the differential voltage gain of the amplifier with the following assumptions²⁶⁸:

The op-amp drives an open-circuit load. The transistors are biased at an emitter current of 1mA so that the emitter incremental resistance is 25Ω . The current gain β is 100 and the output resistance $r_{c5} = r_{o9} = 100\text{k}\Omega$.

Solution

Substituting values in equation 1508, we have

$$A_D = \frac{\beta_5(r_{c5} \parallel r_{o9})}{r_{e1}} = \frac{100(10^5 \parallel 10^5)}{25} = 2 \times 10^5 \text{ volts/volt} = 106 \text{ db}$$

Bias Network

The bias network implements the current sources I_8 , I_9 and I_{10} . In figure 846(b), diode-connected transistor Q_7 carries a current I_1 established by the lone resistor R_1 in the circuit:

$$I_1 = \frac{V_{CC} + V_{EE} - V_{be7}}{R_1} \quad (1509)$$

This current then reflects through the current generators Q_8 , Q_9 and Q_{10} . As shown in figure 846(b), all these currents are equal.

A more flexible approach would set the bias current for each stage according to its requirements. The first stage, the input differential amplifier, is likely to be biased at a low tail current in order to decrease the amplifier bias current (the base currents of Q_1 and Q_2). A low tail current also raises the incremental emitter resistance r_{e1} and r_{e2} , which raises the differential input resistance.

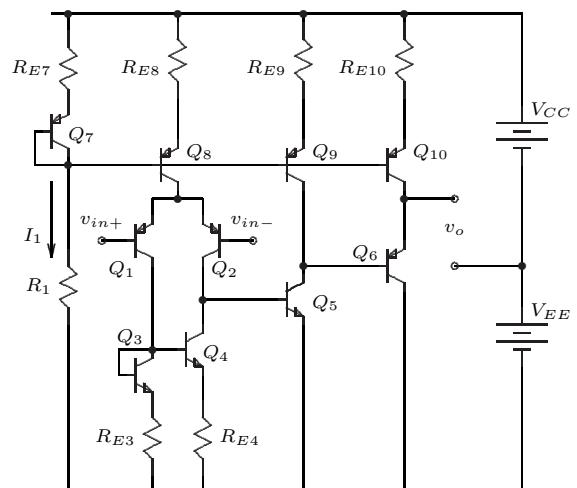


Figure 848: Emitter Resistors

²⁶⁸ Solomon points out in [84] that the voltage gain of an integrated circuit operational amplifier is much less than the same circuit in discrete component form, as a consequence of thermal feedback between the output transistors and the input differential pair.

The output stage bias current must be sufficient to drive the minimum load resistance to full output voltage.

A final design of this amplifier circuit would include emitter resistors in Q_3, Q_4, Q_7 through Q_{10} as shown in figure 848 in order to swamp any device or temperature differences. Resistors R_{E7} through R_{E10} could be scaled to adjust the bias currents in the different stages.

Emitter resistance raises the output resistance of a current mirror (page 927), which enhances the performance of the operational amplifier. Increased output resistance in Q_8 improves the common-mode rejection ratio, and increased output resistance in Q_9 increases the open-loop voltage gain.

Stability

The open-loop frequency response of the op-amp in figure 846 is defined by the circuit impedances and the transistor parasitic capacitances. At medium and high frequencies, each stage contributes a phase lag, and so the frequency response breaks at several points. The resultant equivalent circuit and frequency response are shown in figure 601 (page 709). As discussed in detail in that section, a negative feedback system using this amplifier is likely to oscillate. If the closed-loop gain intersects with the open-loop frequency response where the latter is decreasing at 40db/decade (or greater) then the circuit will be unstable.

Early discrete component solid state operational amplifiers used three gain stages with a system of RC and LC *compensation networks* to shape the open-loop frequency response of the op-amp so that it would be stable with negative feedback. Some of these networks were quite complex [274]. Early integrated-circuit operational amplifiers continued with this tradition. The compensation components were external to the integrated circuit and applications engineers were required to design the compensation network for each application.

In 1967, National Semiconductor released the *LM101* integrated circuit operational amplifier. The LM101 used current-source collector loads to generate high gain in a simplified two-stage circuit. As a result, this circuit could be compensated by a single capacitor, and this circuit topology became a standard of analog op-amp design.

Fairchild Semiconductor then released the *741* op-amp with a compensation capacitor integrated onto the chip, making the op-amp very convenient to use.

Subsequently, the integrated circuit LM101 topology influenced discrete component amplifier design, specifically the design of audio power amplifiers [243].

Dominant Pole Compensation

An operational amplifier will be stable in many circuits if the op-amp open-loop frequency response decreases at a fixed $-20\text{db}/\text{decade}$ until the open-loop gain is unity. It also ensures that the circuit will be unity-gain stable, that is, under the worst-case condition of maximum loop gain.

This is illustrated in figure 849. The original open-loop frequency response (dashed line) contains two break frequencies, each corresponding to an RC lag network in the amplifier circuit. Each lag adds another 90° of phase shift. A closed-loop gain of unity, for example, would be unstable because it intersects with the open-loop response at a slope of $-40\text{db}/\text{decade}$.

The compensated frequency response (solid line) is arranged to break at frequency f_o , run down at $-20\text{db}/\text{decade}$, and pass through the unity gain (0db) axis at a lower frequency than the open-loop response. Any closed loop gain

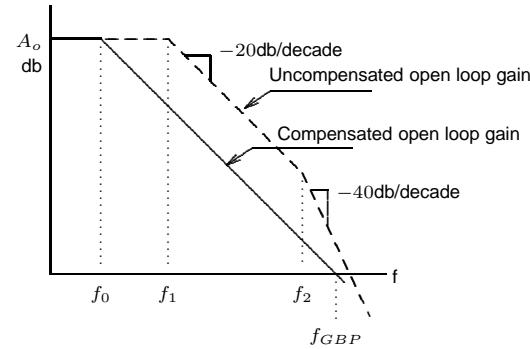


Figure 849: Compensated Response

will intersect with the open-loop gain where it has a slope of -20db/decade , and consequently any closed-loop gain condition will be stable²⁶⁹.

This is referred to as *dominant pole compensation*, because the pole frequency f_o determines the overall open-loop frequency response. (Frequencies f_1 and f_2 are referred to as *non-dominant poles*.)

As shown in figure 850(a), the compensated frequency response of figure 849 may be obtained by adding a single capacitor C_c between the input and output of the voltage gain stage Q_5 . The voltage amplifier stage then has the equivalent circuit of figure 850(b).

The differential amplifier stage Q_1 to Q_4 produces a current v_i/r_{e1} , where v_i is the differential input voltage and r_{e1} the emitter incremental resistance of either of the differential transistors Q_1, Q_2 . The output voltage v_{o5} of the voltage amplifier stage Q_5 is buffered by a unity-gain voltage follower, so this is also the output voltage v_o of the amplifier.

Treating Q_5 as an inverting operational amplifier, we have:

$$v_o = v_{o5} = -Z_f i_i = -\frac{1}{sC_c r_{e1}} \frac{v_i}{r_{e1}} = -\frac{1}{sr_e C_c} = -\frac{1}{j\omega r_e C_c}$$

The magnitude of the output voltage is then given by:

$$|v_o| = \frac{1}{\omega r_e C_c} \quad (1510)$$

At the point where the compensated amplitude response crosses the frequency axis, the frequency is the *gain-bandwidth product* and the gain is unity. Then the gain-bandwidth product ω_{GBP} radians/second is given by:

$$\omega_{GBP} = \frac{1}{r_e C_c} \text{ radians/sec} \quad (1511)$$

and

$$f_{GBP} = \frac{1}{2\pi r_e C_c} \text{ Hz} \quad (1512)$$

A typical gain-bandwidth product for integrated circuit op-amps is 1MHz.

Referring to equation 1512, the gain-bandwidth product is dependent very simply on the emitter incremental resistance r_{e1} of the differential amplifier and the compensation capacitance C_c – both of which are known quantities. Consequently, the gain-bandwidth product can be specified with some confidence for any given amplifier design.

In contrast, the dominant pole frequency f_o (figure 849) cannot be tightly specified. This is the case because (as equation 1508 shows) the open-loop gain A_o is dependent on current gain beta and collector incremental resistance r_o , both of which are much less predictable. Referring to figure 849, it can be seen that variations in A_o will be reflected in variations in dominant pole frequency f_o . Mathematically, the product of gain and frequency are constant anywhere along the -20db/decade slope of the open-loop frequency response. This also applies at frequency f_o , so :

$$f_{GBP} = A_o f_o \quad (1513)$$

The gain-bandwidth product f_{GBP} at f_o , so variations in open-loop gain A_o must be reflected by opposite variations in the dominant pole frequency f_o .

²⁶⁹This requires that the characteristic of the closed-loop gain is flat with frequency.

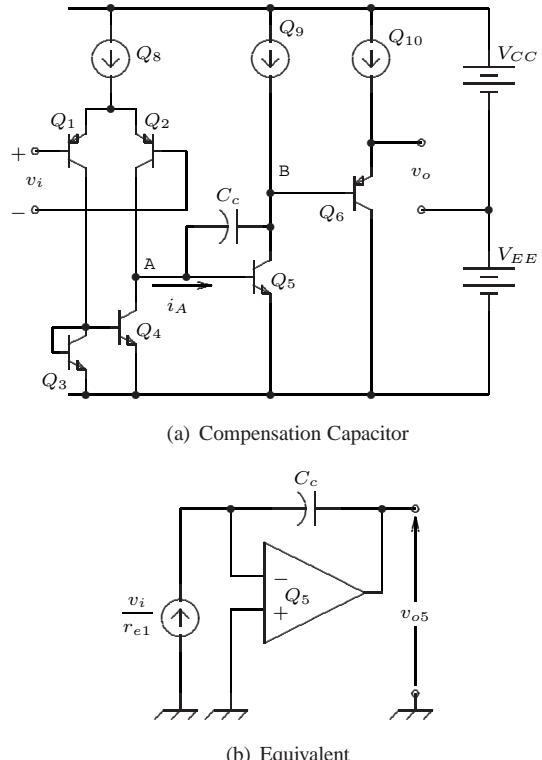


Figure 850: Compensation Technique

Slew Rate

For small signal amplitudes, the amplifier operates in a linear fashion, and no stages are saturated or cutoff. A small amplitude step input causes the output of the amplifier to rise in an exponential fashion at a rate determined by the loop gain and f_o (section 23.5, page 698).

A large amplitude step causes a large error signal in the differential stage. This drives subsequent stages into saturation or cutoff and the rate of rise of the output is determined by the rate at which an available current can charge an internal capacitance. Under this condition, the output signal is a linear ramp that changes at the *slew rate* of the operational amplifier (section 21.8, page 655).

In the amplifier circuit of figure 850, capacitance C_c and its charging current determine the slew rate. In most op-amps, the bias currents increase from stage to stage so that the available current from source Q_9 is large compared to the output current of the differential amplifier stage. Consequently, the maximum value of the current i_A determines the slew rate. The maximum value of i_A is equal to the tail current of the differential amplifier. For example, when Q_1 is fully OFF and Q_2 fully ON, the current I_8 from Q_8 flows entirely into the compensation capacitance. Then the slew rate is:

$$\frac{dV}{dt} = \frac{I_8}{C_c} \text{ volts/sec} \quad (1514)$$

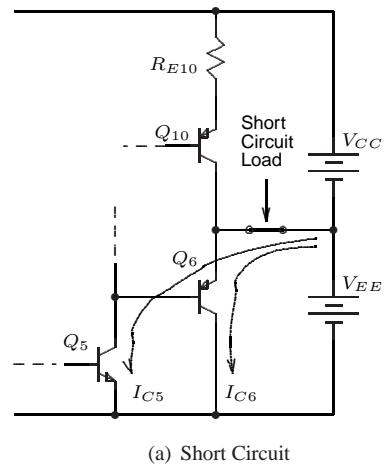
Output Current Limit

Although the amplifier shown in figure 850 is functional, it is not completely protected against a short circuited load.

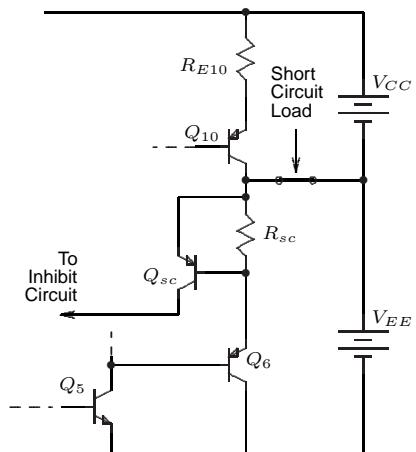
Refer to figure 851 and consider that the output is connected directly to ground. If the output is a positive voltage, then current flows from the positive supply out of the op-amp to ground. The output current will be limited to the collector current Q_{10} . The power dissipation in Q_{10} will be its collector-emitter voltage V_{CC} times its collector current I_{C10} . Providing the transistor does not overheat, this is fine.

Now consider that the output is a negative voltage. Output current flows from ground, into the output terminal of the op-amp, through emitter follower Q_6 , to the negative power supply. If driver transistor Q_6 is fully ON (which is likely), the base current of Q_6 can be extremely large – and the collector current of Q_6 will be beta times larger. Consequently, Q_6 will fail because of excessive current or power dissipation. Driver transistor Q_5 is also likely to be destroyed.

A cure for this problem, which limits the output current that the op-amp can sink, is shown in figure 851(b). A resistor R_{sc} is located in the path of the output current and sized so that the voltage across it is about 0.6 volts at the maximum allowable output current. Then if the maximum allowable output current flows through that resistor it turns on Q_{sc} . The output current from Q_{sc} is fed to an *inhibit* circuit which removes the drive current from the output transistor Q_7 . (One possibility is to dump the collector current from Q_{sc} into the collector of diode-connected transistor Q_3 . This causes Q_4 to conduct heavily and removes base current from transimpedance amplifier Q_5 , which shuts off Q_7 .) The net effect is to limit the current through Q_7 . Notice that Q_7 still has approximately V_{EE} volts across it and is conducting the maximum allowable output current, so it must be able to dissipate the corresponding power.



(a) Short Circuit



(b) Protection

Figure 851: Short Circuit Protection

Output Resistance

Looking back into the output terminal of the op-amp, we see the output of an emitter follower Q_6 . By the reflection principle (section 29.27, page 870), the resistance seen is approximately equal to the source resistance for the Q_6 stage divided by the current gain β_6 .

At the input terminal to Q_6 at low frequency (so that the compensation capacitor may be treated as an open circuit), the source resistance consists of the output resistance of current source Q_9 in parallel with the output resistance of the transimpedance amplifier Q_5 . These resistances might be in the order of $10^5\Omega$ each and the current gain current gain β_6 in the order of 10^2 , so the output resistance is in the order of 500Ω . When the amplifier is operated closed-loop, this resistance is further decreased by the loop gain $A_{ol}B$.

The output resistance can be reduced, and the output current capability increased, with the addition of the circuitry shown in figure 852.

Transistor Q_{11} acts as an emitter follower for sourcing current into a load resistance. Transistor Q_{12} acts as an emitter follower for sinking. The two diodes D_1 , D_2 function as a floating voltage which compensates for the base-emitter voltages of Q_{11} and Q_{12} . The bias source Q_{10} ensures that forward current always flows through these diodes.

This output stage brings three benefits:

- Transistors Q_{11} and Q_{12} can be made larger than the other transistors in the op-amp, so that they can conduct more current. This increases the output current capability of the amplifier, *but only if* the transistors are heatsinked sufficiently to dissipate the resultant power.
- Looking into the output terminal of the op-amp, this output stage lowers the internal resistance of the op-amp by an additional factor equal to the beta of Q_{11} (or Q_{12} , depending on which transistor is conducting at the time).
- The new output stage raises the resistance seen at the collector of Q_5 (point B in figure 852), which has the effect of raising the overall open-loop voltage gain of the amplifier.

Figure 862 (exercise 7 on page 971) shows a circuit that limits the output current of this type of stage.

Offset Voltage

The *offset voltage* of this amplifier is caused by a mismatch between the transistors of the differential pair, Q_1 , Q_2 . From equation 1158 on page 807, the collector current I_c in each transistor of the differential pair is given by:

$$I_c = I_s e^{V_{be}/V_t}$$

If the bases of Q_1 and Q_2 are tied together, then this forces the base-emitter voltages to be equal. The collector currents will be slightly different because the saturation current I_s is slightly different for each device, due to minor processing differences.

This difference in collector current will translate into a non-zero voltage at the output of the operational amplifier.

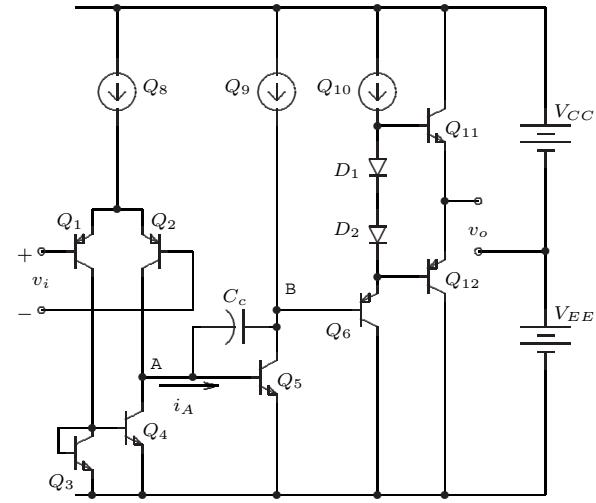


Figure 852: Enhanced Output Stage

A small voltage can be applied between the bases of Q_1 and Q_2 and adjusted so that the output of the op-amp is zero. Then this voltage is the offset voltage of the amplifier. The offset voltage is the difference between the two base-emitter voltages when the collector currents are equal.

Offset voltage is particularly problematic in high-gain DC amplifiers, where it causes a static output error. It is also important in integrators because it causes a dynamic output error.

It is possible to add circuitry to the op-amp that will adjust the offset voltage to zero. An example, the scheme used in the 741 integrated circuit amplifier, is shown in figure 853. When the potentiometer R_{offset} is centred, it has no effect on the offset voltage of the amplifier. Moving the wiper to one side effectively decreases the emitter resistance on that side. For example, if I_{C1} is larger than I_{C2} , the wiper is moved to the left to decrease the emitter resistance of Q_3 , thereby reducing the effect of that larger current.

In general potentiometer adjustments should be avoided in a production design. Potentiometers are expensive, require adjustment, have the potential for misadjustment, and can become electrically noisy. For all these reasons, it is best to match the differential amplifier transistors as closely as required to obtain the necessary offset voltage specification. In a one-off discrete component design, that can be done by manual selection of the transistors. In production, the differential pair can be purchased as a matched BJT pair.

Bias Current

The operational amplifier input bias current (section 21.2, page 641) is simply the average of the base currents of the differential amplifier transistors Q_1 Q_2 . These are PNP transistors, so the bias current flows out of the input terminals.

There are op-amp applications where a small bias current is a requirement. In this design, the bias current can be reduced by the following measures:

- Reduce the tail current of the differential pair, thereby reducing the base current of transistors Q_1 Q_2 .
- Convert the input transistors into Darlington or Sziklai pairs, which have larger current gain.
- Replace the BJT devices Q_1 Q_2 by JFET or MOSFET devices. An op-amp with these devices in the input stage will have an input bias current orders of magnitude smaller than a conventional BJT input. However, it is more difficult to match the transfer characteristics of these devices, so the offset voltage is likely to be larger than a BJT stage. Figure 854 shows the input BJT transistors replaced by P-Channel JFET devices.
- Replace the BJT devices Q_1 Q_2 by *superbeta transistors*. These are BJTs that have a current gain β in the order of 5000, and consequently small base current [278]. However, this technique is only available in integrated circuit designs.

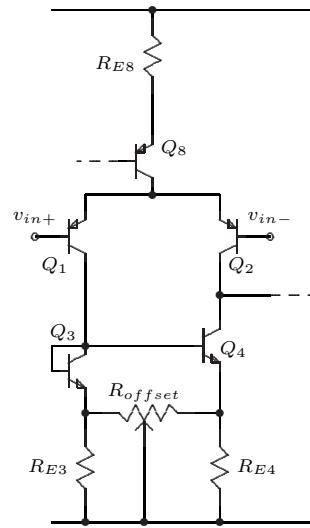


Figure 853: Offset Adjustment

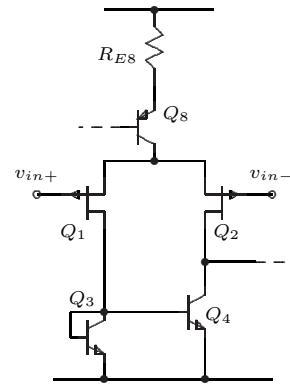


Figure 854: P-Channel JFET Inputs

30.18 Exercises

1. A logic level translator circuit is shown in figure 855.

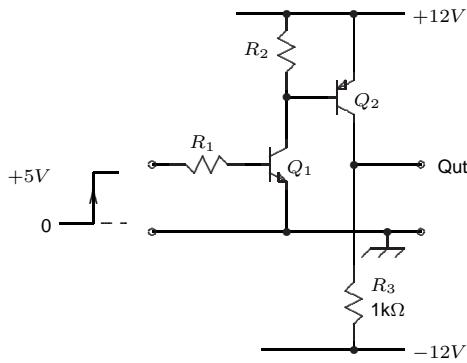


Figure 855: Logic Level Translator

The input to this circuit is a logic signal that switches between 0 and +5 volts. The level translator converts this signal to one that varies between +12V and -12V. The transistors are silicon and have a minimum beta of 50.

- (a) To what level does a zero volt input correspond at the output? Is this an inverting or non-inverting circuit?
 - (b) Resistor R_2 should be chosen so that most of the collector current flows into Q_2 . Calculate a suitable value for R_2 .
 - (c) Calculate a suitable value for R_1 .
 - (d) The *input threshold value* for the circuit of figure 855 – the input voltage at which the circuit begins to switch – is about one V_{BE} . Consequently, noise pulses of that magnitude can trigger the circuit incorrectly. Redesign the circuit so that the input threshold is 2.5 volts.
 - (e) Redesign the circuit so that it inverts the logic: +5 volts input corresponds to -12 volts output. You can change the polarity of either or both of the transistors. Two transistors are still sufficient.
 - (f) Modify your circuit so that its input threshold is 2.5 volts.
 - (g) Estimate the *switching gain* (the ratio of the change in output voltage to input voltage) of the circuit shown in figure 855. What change in voltage at the input is required to cause the output to switch fully from one supply rail to the other?
2. A current sink design is shown in figure 856(a). A circuit designer has suggested the (possibly) improved version shown in figure 856(b). The transistor parameters are: $V_{BE} = 0.6\text{V}$, $\beta = 50$, $r_c = 50\text{k}\Omega$. The diode forward voltage is $V_F = 0.6\text{V}$.
 - (a) The circuit operates from a positive supply of +10 volts. The diode current should be 10 times the base current of the transistor. The output current I_o of the sink is $500\mu\text{A}$. For circuit 856(a), calculate the values of R_1 and R_2 .
 - (b) Determine the output resistance of the current sink in figure 856(a).
 - (c) The circuit of figure 856(b) is claimed to be an improved version of 856(a). Is this the case? Explain.

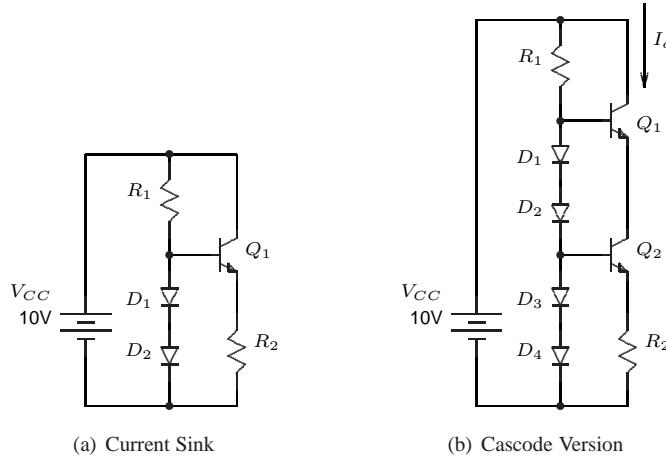


Figure 856: Cascode Current Sink

- (d) What is the collector-emitter voltage of transistor Q_1 in figure 856(b)? Can D_1 be eliminated? Explain.
- (e) Determine the output resistance of the current sink in figure 856(b).
3. A simple current sink, composed of transistor Q_1 and resistors R_1 and R_2 , is shown in figure 857. In this circuit, the current sink is controlling the tail current of the differential pair. (The differential pair don't do anything useful in this circuit.)

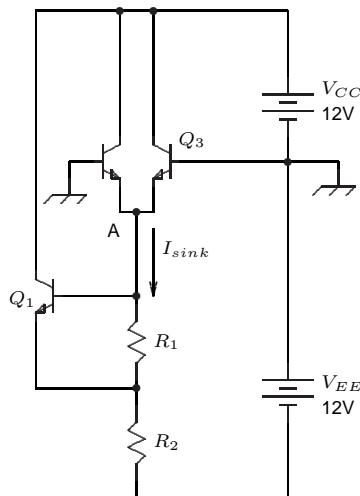


Figure 857: Current Sink

The operation of the current sink can be viewed as follows:

- Resistor R_1 is selected to drop a voltage of V_{BE} volts at the sink current.

- Resistor R_2 is made as large as possible, consistent with the requirements of the rest of the circuit. In figure 857, point A must be at $-V_{BE}$ volts in order for the differential pair to operate properly, so for silicon transistors where $V_{BE} = 0.6$ volts the maximum voltage across R_2 is $V_{EE} - 2V_{BE} = 10.8$ volts.

The current through R_2 is set to the sink current I_{sink} plus whatever emitter current I_{E1} is desired in the transistor.

- The transistor senses the voltage across R_1 and regulates that current by dumping more or less current into resistor R_2 . This dumping action corrects the voltage across R_1 in such a direction as to oppose any change.

For example, if the sink current I_{sink} decreases for some reason, then the voltage drop across R_1 decreases. The transistor responds by conducting less current, which reduces the voltage across R_2 . This moves the voltage at the emitter of the transistor more negative, which increases the current through R_1 .

- If the sink current in figure 857 is to be 1mA, and the transistor emitter current 2mA, calculate suitable values for R_1 and R_2 .
- Comment on the temperature stability of this current sink.
- Compare the common-mode input range of this circuit against the circuit shown in figure 832 on page 945. Would this circuit be more suitable for an op-amp in the inverting or non-inverting configuration?
- (Advanced) Determine the input resistance of this current sink, seen at point A, looking into the current sink circuit, in terms of the parameters of Q_1 and circuit resistances R_1 , R_2 .

4. Figure 858 shows a *complementary buffer amplifier*.

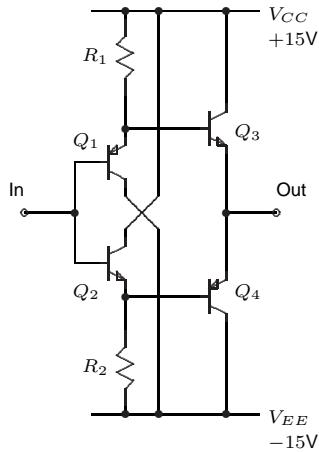


Figure 858: Complementary Buffer Amplifier

- The amplifier is entirely composed of what type of single stage amplifier?
- For large signals, the currents through the resistors vary substantially, causing the incremental emitter resistance r_e in Q_1 and Q_2 to vary, which causes distortion. Show how the resistors may be replaced with current sources. Hint: The additional devices required are two transistors, two diode-connected transistors, and one resistor R_{cc} .

- (c) Calculate a value for the resistor R_{cc} such that the bias current in transistors Q_1 and Q_2 is 1mA.
- (d) When the bias current in transistors Q_1 and Q_2 is 1mA, what is the bias current in transistors Q_3 and Q_4 ?
- (e) Modify the circuit so that the output current is limited to a maximum of 10mA. Hint: For one possible solution to this circuit, the additional devices are two resistors and two transistors.
- (f) (Advanced) Design a short-circuit current limit circuit that under normal operation has no effect on the bias current in the output stage.
- (g) Determine the power dissipation in the amplifier when it is delivering a sine wave, 10mA peak current, into a short circuit.

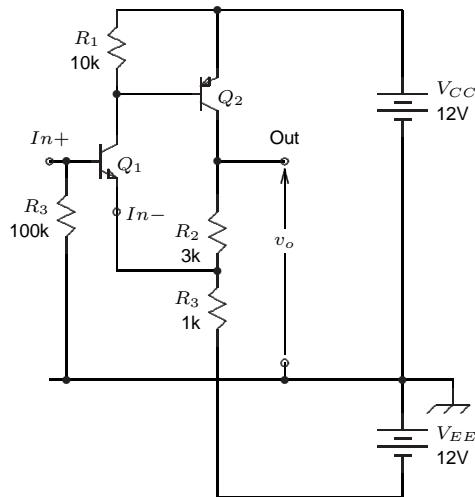
5. In this exercise, we investigate the design of a simple transistor amplifier.

Throughout this question, the transistors have $\beta = 100$, $V_{BE} = 0.6$ volts, $r_c = 50k\Omega$.

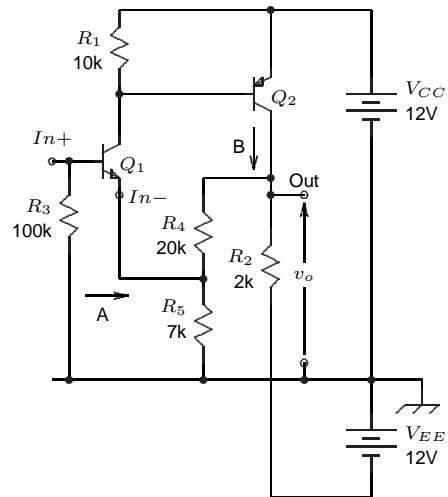
(a) **First Attempt**

A simple discrete component operational amplifier circuit is shown in figure 859(a). The amplifier itself is transistors Q_1 , Q_2 and resistors R_1 , R_2 . The inverting and non-inverting inputs, and the output of the op-amp, are marked on the diagram.

The maximum output voltage swing of the amplifier before clipping should be at least ± 5 volts.



(a) Attempt 1



(b) Attempt 2

Figure 859: Amplifier Design

- Verify that the inverting and non-inverting terminals of this ‘operational amplifier’ are correct.
- Resistors R_2 and R_3 are the feedback network. What is the intended closed-loop gain of the amplifier?
- This amplifier will not work properly, because the bias conditions are incorrect. Show that an input voltage of zero volts saturates Q_2 .

(b) **Second Attempt**

A second attempt at the design is shown in figure 859(b).

- i. Assuming that the base current of Q_1 is negligible (so that the base voltage of Q_1 is at zero volts), determine by inspection the DC output voltage of the operational amplifier.
- ii. Calculate the currents in the two transistors under this same condition. Are they saturated or cutoff?
- iii. Assuming that the open-loop gain of the amplifier is large compared to the closed-loop gain, determine by inspection the closed-loop voltage gain G for this circuit.
- iv. The resistor R_1 may be ignored if it is large compared to the resistance seen into the base of transistor Q_2 . Show that this is indeed the case.
- v. Calculate the open-loop gain of the amplifier. For this calculation, assume that the negative feedback network is disconnected and an input signal applied to the inverting input terminal. The open-loop gain is the ratio of this signal to the output signal.
When the feedback signal is open-circuited, the amplifier AC and DC conditions must remain unchanged. Consequently, the network at the emitter of Q_1 must be replaced with the Thevenin equivalent at point A. The network seen by the collector of Q_2 must be replaced by the Thevenin equivalent seen at point B.
- vi. Is this open-loop gain adequate compared to the closed-loop gain? Determine the true value of the closed-loop gain.
- vii. What effect does a temperature change have on the output voltage, in volts per $^{\circ}\text{C}$?

(c) **Third Attempt**

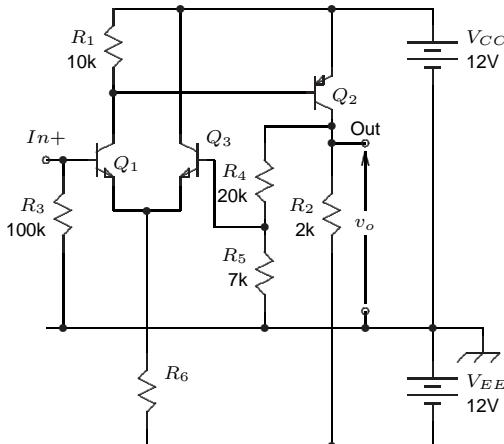


Figure 860: Attempt 3

A further improvement of the design is shown in figure 860.

- i. The inverting input terminal has moved. Where is it?
- ii. This modification improves the stability of the output voltage with temperature. Why?
- iii. This modification increases the open-loop gain of the amplifier. Explain in general terms why this occurs.
- iv. Determine the open-loop gain of this version of the amplifier.
- v. Each transistor Q_1, Q_3 is to operate at a bias current of $500\mu\text{A}$. Calculate a suitable value for R_6 .

6. For the circuit shown in figure 861, the transistors have a current gain $\beta = 100$, collector resistance $r_c = 50k\Omega$, and base-emitter voltage $V_{BE} = 0.6V$. The value of alpha may be taken as unity. Assume a load resistance R_l of $2k\Omega$.

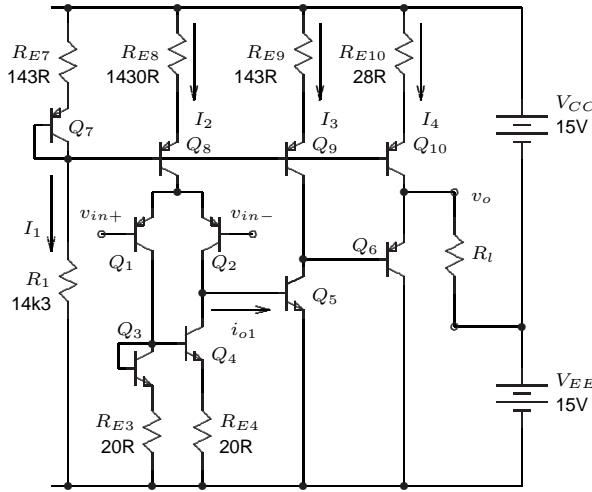


Figure 861: Operational Amplifier

Answer the following questions about this circuit.

- What is the voltage across resistor R_1 ?
 - What is the voltage across resistor R_{E7} ?
 - What is the reference current I_1 ?
 - Calculate the differential amplifier stage bias current I_2 .
 - Calculate the transimpedance stage collector bias current I_3 .
 - Calculate the output stage bias current I_4 .
 - Determine the first stage differential transconductance gain $i_{o1}/(V_{in+} - V_{in-})$.
 - What is the internal resistance of the current source seen looking into the collector terminal of Q_9 ?
 - Determine the effective load resistance seen at the collector of transistor Q_5 .
 - What effect does the load resistance have on the open-loop gain of the amplifier? Explain.
 - Calculate the overall differential voltage gain.
 - Calculate the input bias current of the amplifier.
7. Figure 862 shows a circuit for short-circuit protection of the op-amp discussed in section 30.17.
- Explain how transistors Q_{13} and Q_{14} limit the output current.
 - If the maximum output current is to be limited to 100mA and the transistors are all silicon type, calculate the value of resistors R_{sc1} and R_{sc2}

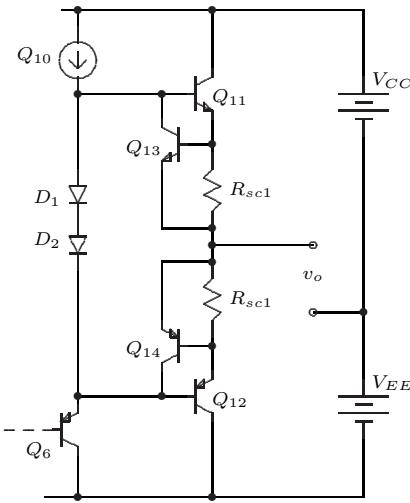


Figure 862: Op-Amp Short Circuit Protection

8. Figure 863 shows the μ A733, a differential amplifier optimized for high speed operation.

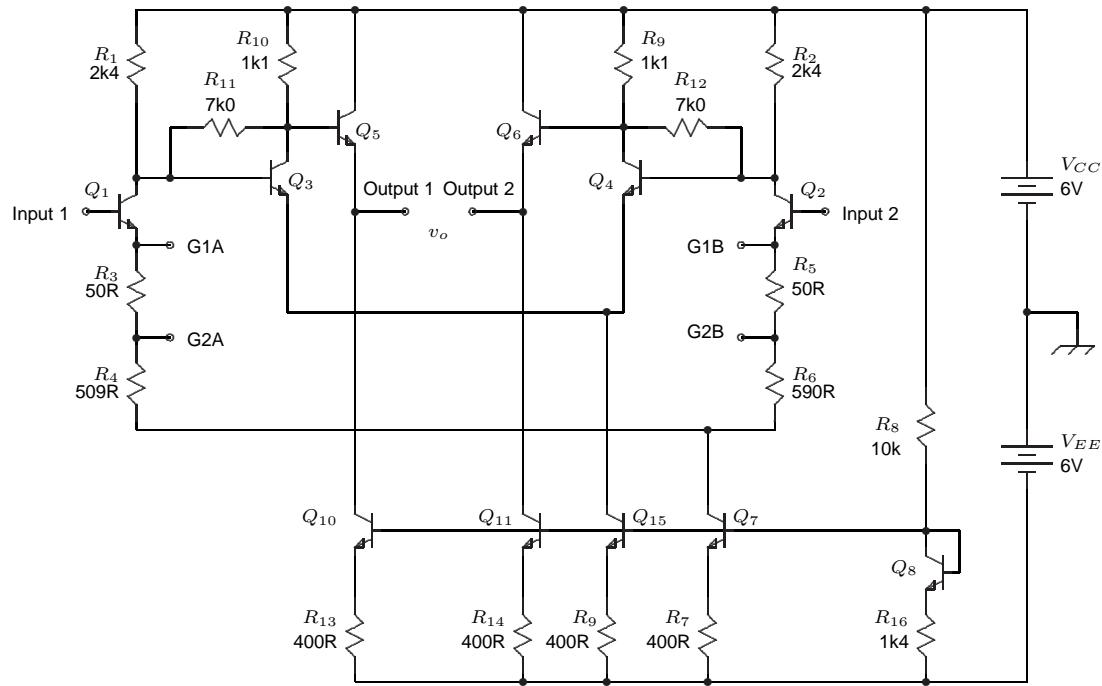
The transistors have a current gain $\beta = 100$, collector resistance $r_c = 50k\Omega$, and base-emitter voltage $V_{BE} = 0.6V$. The value of alpha may be taken as unity.

(a) DC Analysis

- i. Identify the bias system for the amplifier: the reference current generator and the slave current sources.
- ii. Identify the input stage differential amplifier.
- iii. Identify the second stage differential amplifier.
- iv. Identify the output stage emitter followers.
- v. Determine the tail current for each of these stages.
- vi. Assume that both input terminals are connected to the ground point. Determine the collector DC voltage to ground at the output of the first stage.
- vii. Ignoring the feedback resistors R_{11} and R_{12} , determine the collector DC voltage to ground at the output of the second stage.
- viii. Determine the DC voltage to ground at the output of the amplifier.
- ix. What is the maximum available output current that can be supplied or absorbed at the output terminal?

(b) AC Analysis

- i. Amplifiers reduce distortion with negative feedback. In some amplifiers, all the feedback is *global*, that is, applied external to the amplifier. In other amplifiers, the feedback is *local*, that is, applied at each stage of the amplifier. In a third case, a combination of local and global feedback is used. In general, this amplifier is used without global feedback. Identify the local feedback mechanism in each stage.
- ii. Using the values shown on the schematic of figure 863, and the bias conditions calculated above, determine the voltage gain of the amplifier.

Figure 863: μ A733 Differential Amplifier

- iii. The voltage gain may be adjusted by connecting terminals G2A and G2B. What effect does this have on the voltage gain? Explain.
9. The Fairchild μ A723 was one of the first integrated circuit regulators. It's now obsolescent, but the schematic makes a useful study of compound BJT devices. A simplified drawing of the amplifier in the μ A723 integrated circuit regulator, adapted from the datasheet [279], is shown in figure 864.
- The value of β is typically 150 for the small signal NPN transistors used in the circuit, and 50 for the power transistor Q9. The Early voltage is 200 volts for the NPN transistors, 80 for the PNP devices.
- Each of the transistors in this schematic has a function associated with the bias system, the differential amplifier, and the buffer amplifier. Which transistors belong in each of these groups?
 - Starting with the Zener diode D1, estimate the zero signal (bias) currents in the various transistors.
 - Determine which of Input 1 and Input 2 are the inverting and non-inverting terminals of the amplifier.
 - Estimate the voltage gain of the amplifier under the conditions shown on the schematic.
 - The thermal resistance Θ_{ja} junction to ambient of the plastic DIP package is $150^\circ\text{C}/\text{Watt}$ worst case. Estimate the temperature of the chip above ambient under the conditions shown in the figure. The maximum allowable operating temperature of the chip is 70°C . Is the power dissipation within acceptable limits?
10. A simple discrete-component AC-coupled amplifier is shown in figure 865. It should have a gain e_o/e_i of 10 volts/volt over the audio frequency range 20Hz to 20kHz.

You may assume:

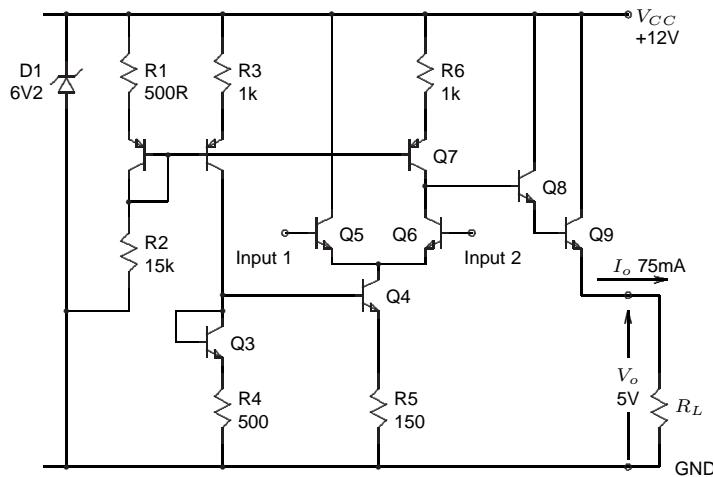
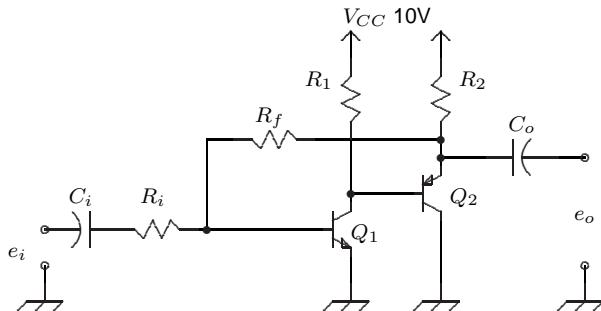
Figure 864: μ A723 Regulator

Figure 865: AC Coupled Amplifier

- The DC current through Q_1 is 1mA. Through Q_2 it is 10mA.
 - The DC voltage across Q_2 is half the supply voltage. (Why might this be a good choice?)
 - The current gain β of the transistors is 50.
- (a) Calculate the values of R_1 and R_2 .
 - (b) Calculate the value of R_f . You can assume the current through R_f is much smaller than the current through R_2 , but then confirm that this is the case.
 - (c) Calculate the value of R_i to give a gain of 10 V/V. You can assume the ‘open-loop’ gain of the amplifier is much larger than 10.
 - (d) Now calculate the open-loop gain of the amplifier, to confirm the previous assumption.
 - (e) Calculate the value of C_i , assuming a source impedance of zero ohms.
 - (f) Calculate the value of C_o , assuming a load impedance of 30Ω .
 - (g) What effect, if any, will attaching a 32Ω load to the output have on the magnitude of the output signal? Explain.

- (h) What effect will it have on this circuit if the current gain of Q_1 changes over the range 40 to 100?
 (i) We assumed a bias current of 1mA in Q_1 and 10mA in Q_2 . Why might we assume those particular values?

Optional Construct the amplifier, using the NPN transistor 2N4401 and PNP transistor 2N4403. Verify the operation of the circuit.

11. Figure 866 shows an audio power amplifier.

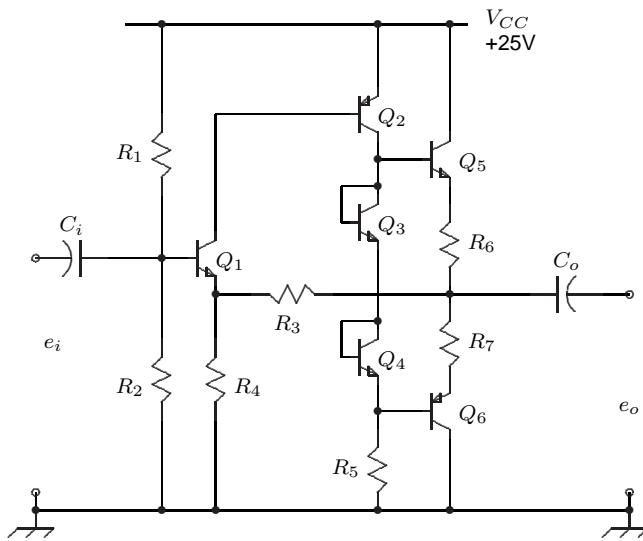


Figure 866: Audio Power Amplifier

- (a) The circuit may be treated as an op-amp with resistive feedback. Identify the inverting and non-inverting input terminals of the op-amp.
 (b) Identify the resistors that determine the closed-loop gain of this op-amp.
 (c) What is the function of transistors Q_3 and Q_4 ?
 (d) Which stage of this amplifier (transistor and resistor) determines the open-loop gain of the op-amp?
 (e) Approximately at what voltage would you expect to find the base terminal of Q_1 ? Explain.
 (f) This design could be greatly improved by using a differential input pair for the input stage. Give 2 reasons why this would be an improvement.
 (g) Redraw the circuit showing a differential amplifier input stage and a constant-current collector load in the second stage to increase open-loop gain.
 (h) If the amplifier drives an 8Ω load (a loudspeaker), calculate a suitable value for the output capacitor C_o .

31 JFET Characteristics and Applications

In section 8.2 (page 268), we introduced the Junction Field Effect Transistor, or JFET. In this section, we add some details and show how the JFET can be used as a switch and as an amplifier.

31.1 Constant Current Diode

We previously discussed the zener diode in section 7. The zener is a *constant voltage* two-terminal device. The JFET may be used to construct its dual: a *constant current* two-terminal device [280].

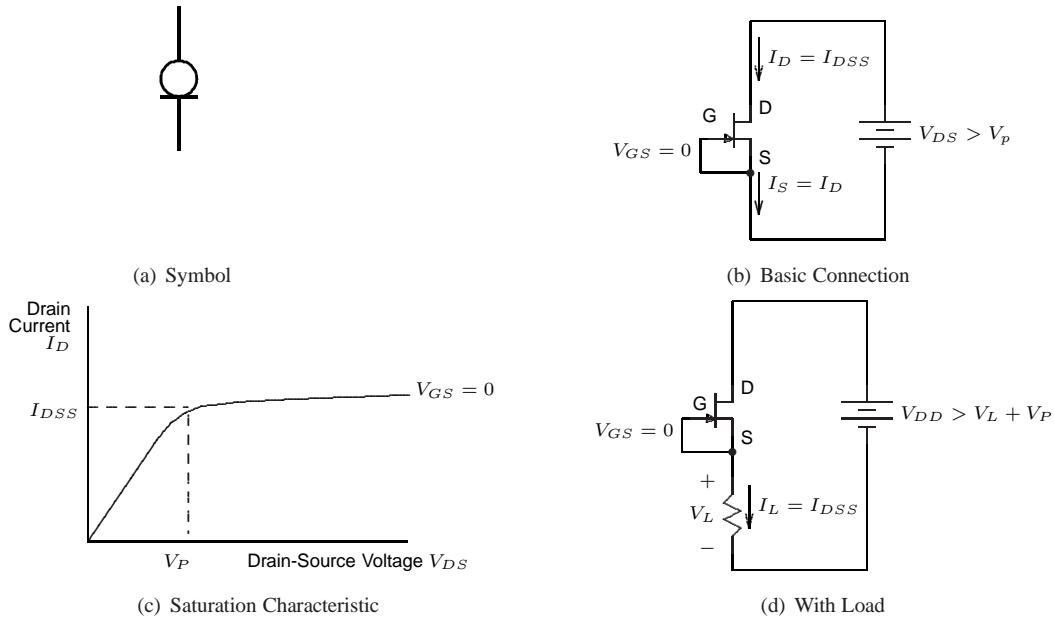


Figure 867: JFET Constant Current Diode

A symbol for the constant-current diode is shown in figure 867(a). The basic JFET diode circuit is shown in figure 867(b), and the behaviour in figure 867(c).

As the drain-source voltage V_{DS} is increased, the drain current I_D initially increases in a like manner. However, once the drain voltage exceeds the saturation voltage of the device (equal to V_P , the pinch-off voltage), the drain current becomes essentially constant, independent of the drain-source voltage. The gate-source voltage V_{GS} is zero, so there is no gate current and the source current I_S is equal to the drain current I_D .

A more practical circuit, in which the constant current diode supplies a load resistor, is shown in figure 867(d). The supply voltage V_{DD} must be sufficient to create the load voltage plus the minimum drain-source voltage for the JFET.

The constant-current diode is a convenient device in some applications. However, it does have some limitations:

- The production spread of the drain saturation current I_{DSS} is quite large. For example, the value of I_{DSS} in the MPF102 N-Channel JFET may be anywhere from 2mA to 20mA, a range of 10:1.

- Even in the pinch-off region, where V_{DS} is greater than the saturation voltage, the drain current increases slightly with drain voltage. This may be modelled by a drain-source resistor in parallel with the current source. In the case of the MPF-102, this resistance is in the order of $5\text{k}\Omega$. This is not a particularly large output resistance for a constant current source, and it should be checked against the application requirements.

Adjustable Constant Current Diode

The output current of the JFET current source may be adjusted by the *self-bias* technique, shown in figure 868.

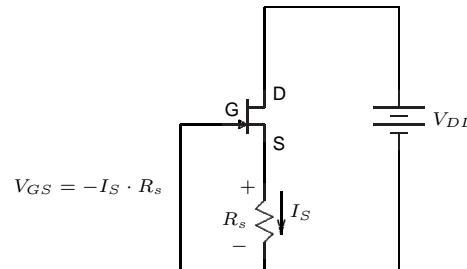
Figure 868(a) shows the basic configuration. For the moment, we ignore the load resistance, or, if you prefer, you can think of the load as a short circuit. Current sources are quite happy with a short-circuit load. Source current flows through the source resistor R_s and creates a positive voltage at the source terminal with respect to the gate. Consequently, this reverse-biases the gate-source junction, which throttles the drain current to some value below its saturation current.

The situation is illustrated on the JFET transfer characteristic of figure 868(b). The behaviour of the JFET is characterized by the curved trace running between I_{DSS} and V_P . As the gate-source voltage V_{GS} becomes more negative, the drain current I_D decreases.

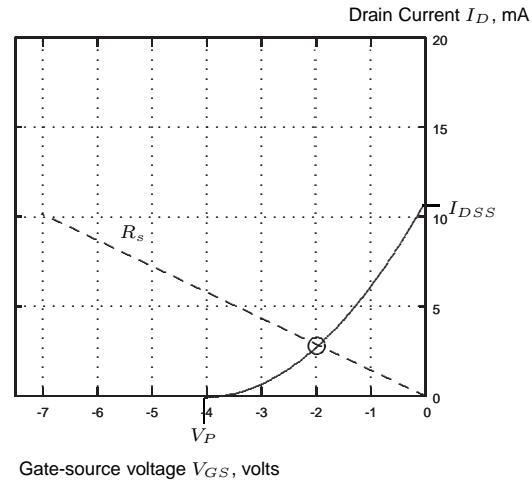
The voltage-current characteristic of the source resistor R_s is shown as a dashed line on figure 868(b). The circuit stabilizes at the point where the transfer characteristic of the JFET and the characteristic of the source resistor intersect – indicated by a circle on the diagram. This point is called the *Q Point*, *Q* for *Quiescent* (resting) – the circuit comes to rest at this particular combination of current and voltage.

In figure 868(b), the Q point corresponds to a drain current of 2.5mA and a gate-source voltage of -2 volts.

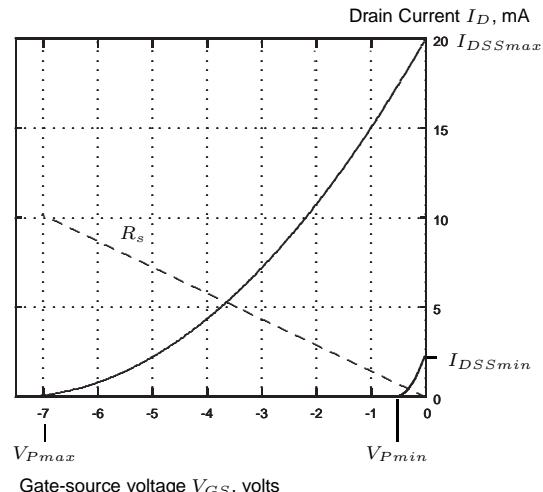
This graphical technique is useful in visualizing the spread in operating voltage and current expected for various values of the JFET drain saturation current and gate cutoff voltage. Figure 868(c) shows the maximum and minimum transfer curves for the JFET. The drain current can be expected to vary over the range approximately 0.5mA to 5mA.



(a) Basic Connection



(b) Typical Q Point



(c) Worst Case Q Points

Figure 868: Adjustable Current Source

Long-Tailing

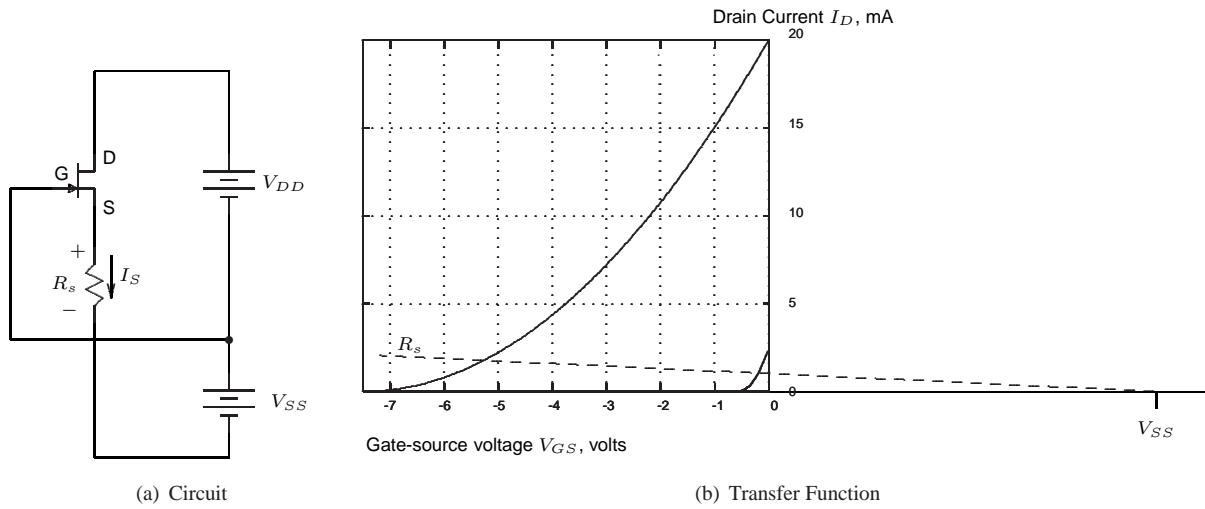


Figure 869: Long-Tailed Current Source

By long-tailing the source circuit, it is possible to establish a drain current that is less variable with production variations in the JFET. The circuit is shown in figure 869(a) and the effect on the transfer characteristic diagram in figure 869(b).

When the source resistor R_s is connected to a negative supply, the magnitude of the source resistance can be increased, flattening the slope. The effect is to reduce the variation in drain current: in figure 869(b) the drain current would be expected to vary between 1 to 2mA. The range of corresponding gate-source voltages is increased, but that is often not a major concern.

In effect, the combination of the increased source resistance R_s and negative supply V_{SS} are approaching a current source. This isn't a useful technique in constructing a current source, but it is useful in biasing a JFET amplifier, as we'll see.

31.2 Transfer Characteristic

The transfer characteristic of figures 868 and 869 has an analytical expression [281]:

$$I_D = I_{DSS} \left(1 - \frac{V_{GS}}{V_P} \right)^2 \quad (1515)$$

where the quantities are:

- | | |
|-----------|--|
| I_D | Drain Current at some value of gate-source voltage |
| I_{DSS} | Drain Saturation Current: the drain current when the gate-source voltage is zero |
| V_{GS} | Gate to Source Voltage |
| V_P | Pinch-Off Voltage: the gate voltage that reduces the drain current to zero. |

This is the equation of a parabola and consequently the JFET is referred to as *a square-law device*. This is useful when an application requires a current or voltage that increases as the square of some other voltage.

The values of drain saturation current and pinch-off voltage are available from the JFET datasheet (section 31.9) on page 993. Alternatively, it's possible to measure these parameters. Connect the gate terminal to the source terminal and apply an adjustable drain-source voltage V_{DS} . Increase the voltage from zero. As the voltage increases, the drain current will increase up to a point and then stabilize at a constant value as shown in figure 867(c) on page 976. As illustrated in figure 867, the pinch-off voltage V_P is the voltage at which the drain current becomes constant. For voltages at or above this, the drain current is the drain saturation current I_{DSS} .

Example

A certain JFET has a drain saturation current I_{DSS} of 20mA and a pinch-off voltage of -7 volts. What is the drain current when the gate-source voltage is -4 volts?

Solution

Substitute in equation 1515:

$$\begin{aligned} I_D &= I_{DSS} \left(1 - \frac{V_{GS}}{V_P}\right)^2 \\ &= 20 \left(1 - \frac{-4}{-7}\right)^2 \\ &= 3.67 \text{ mA} \end{aligned}$$

JFET Low Voltage Region

In figure 870 the JFET characteristic of figure 867(c) is expanded and plotted for various values of gate voltage. There are two regions of operation.

In the *ohmic region*, the drain current increases with increasing drain-source voltage. The channel of the JFET is behaving as a (somewhat nonlinear) resistor.

In the *pinch-off region*²⁷⁰, the JFET assumes constant-current operation.

The transition between the two regions is defined by the dashed line. At any point on the dashed line, the drain-source voltage V_{DS} is equal to the the pinch-off voltage plus the gate-source voltage. For example, if the pinch-off voltage V_P of the JFET is 5 volts and gate-source voltage is -3 volts, the transition occurs at:

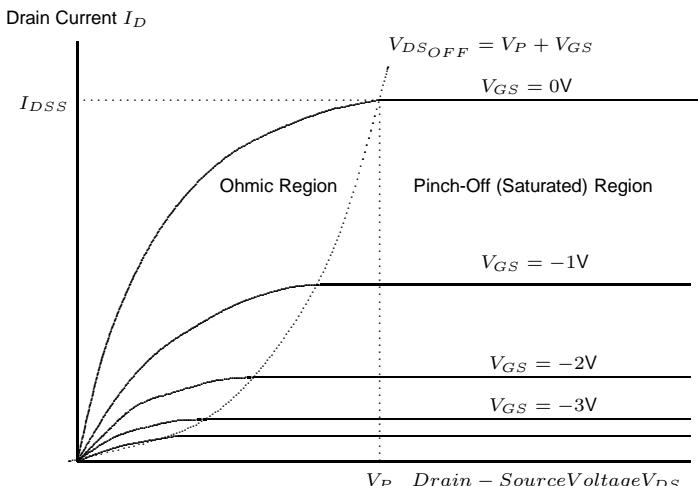


Figure 870: Low Voltage Region

$$\begin{aligned} V_{DS,OFF} &= V_P + V_{GS} \\ &= 5 - 3 \\ &= 2 \text{ V} \end{aligned}$$

²⁷⁰Confusingly, this region is often referred to as *saturation*. However, saturation in the JFET is precisely the opposite of saturation in the BJT, so we will avoid that term.

In other words, the JFET will behave as a constant-current device when its drain-source voltage is equal to or greater than 2 volts.

When the JFET is operated in its ohmic region, it behaves essentially as a voltage-controlled resistor. A more negative gate-source voltage increases the resistance. It's not shown in figure 870, but the behaviour continues into the third quadrant – that is, the drain-source voltage can reverse and this behaviour can apply for alternating voltages.

This is a significant difference from the behaviour of a BJT, which has diode junctions in its collector-emitter path and a finite minimum saturation voltage. The BJT is not symmetrical: when the emitter and collector are interchanged the current gain is very small.

As a consequence of these properties, early in the history of the JFET there was a flurry of interest in using it in voltage-controlled variable gain amplifiers [282] and as a frequency control element in active filters [283]. However, as the curvature of the traces indicates in figure 870, the drain-source resistance is non-linear. This effectively limits the maximum allowable signal across the JFET. As a result, this technique is unsuitable for high quality audio applications and other techniques (such as the Operational Transconductance Amplifier, section 35.5, page 1069) have proven to be more useful for gain control and tuning active-filters.

Acoustic Radar Compressor

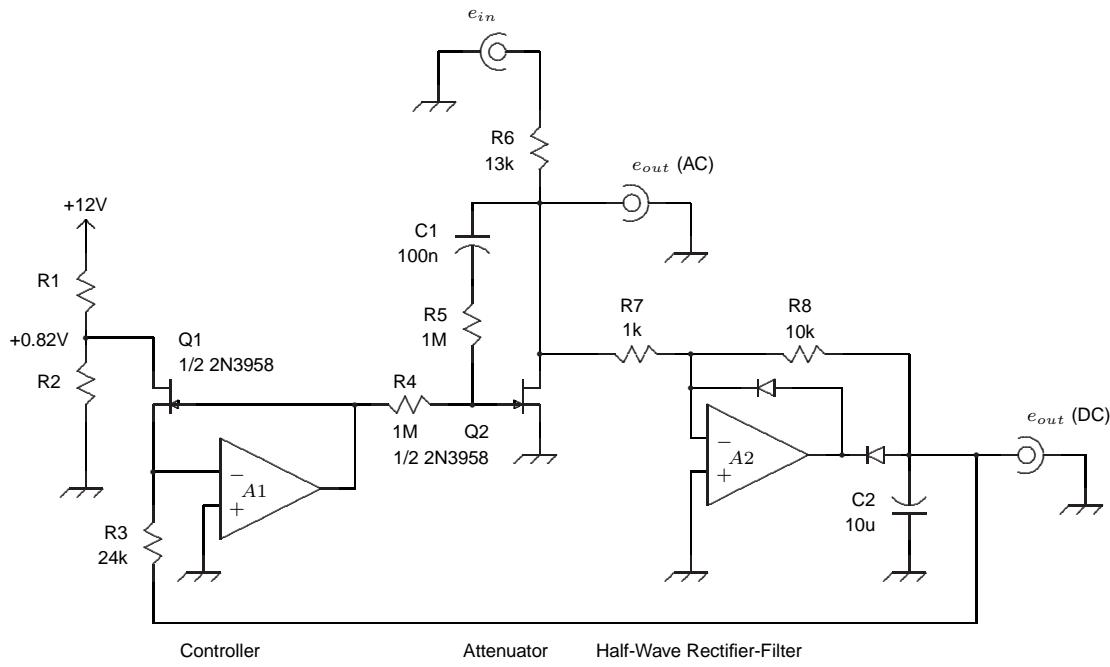


Figure 871: Compressor

These ideas are illustrated in the circuit of figure 871.

A *compressor* is a circuit that reduces the dynamic range between the largest and smallest signals. It is used in audio systems and, in this case, an acoustic radar system. Radar transmits a pulse and then receives and displays the echoes. There is a large range in amplitude between the weakest and strongest received signals. Large signals are liable to clip and small signals to be buried in noise. A compressor reduces the dynamic range so that it is more compatible with other electronic circuits.

A compressor circuit is sometimes referred to as *automatic gain control* (AGC). It acts as if a human operator is adjusting a volume control to keep the signal at a roughly constant level.

The circuit of figure 871 has three sections. In the centre, the JFET Q2 acts as a variable resistor. Then the input signal e_{in} passes through a voltage divider comprised of R6 and Q2 to the output point $e_{out}(\text{AC})$. As Q2 assumes a lower resistance, the input signal is attenuated.

The object is to accomplish this automatically. Op-amp A2 acts as an ideal half-wave rectifier which charges the filter capacitor C2 to the peak value of the output signal. This quasi-DC signal then controls the gate of the attenuator JFET Q2 via the controller A1.

The controller linearizes the control characteristic of the attenuator JFET. To understand the operation, first recognize that there is a negative feedback loop around the controller op-amp A1. As a consequence, the inverting terminal of A1 is at zero volts. Then whatever current flows through R3 must also flow through the controller JFET Q1. The negative feedback action of op-amp A1 adjusts the resistance of Q1 to ensure that this is the case. Therefore *the resistance of the controller JFET Q1 is a linear function of the rectified signal voltage*.

Now, the op-amp output voltage will be at whatever value is necessary to adjust Q1 to the correct resistance. That same value is connected to the gate of the gain control JFET Q2. The JFETs are matched, so the resistance of the attenuator JFET Q2 is equal to the resistance of the controller JFET Q1, which is in turn proportional to the amplitude of the output signal. As the output signal becomes larger, the control system increases the attenuation, which partially cancels that increase.

The drain of the controller JFET is connected to a supply of 0.82 volts, which is less than the pinch-off voltage of this JFET. That ensures that the JFET stays in its ohmic region and therefore behaves as a variable resistor.

The capacitor-resistor network connected to the gate of the attenuator JFET causes some of the AC signal to be superimposed on the DC control signal. The effect is to reduce distortion in the attenuator. As the drain voltage rises, the curve in the resistance characteristic decreases current through the JFET. However, increasing drain voltage slightly increases the gate voltage, which tends to increase current through the JFET – and these two effects cancel – at least, to some extent.

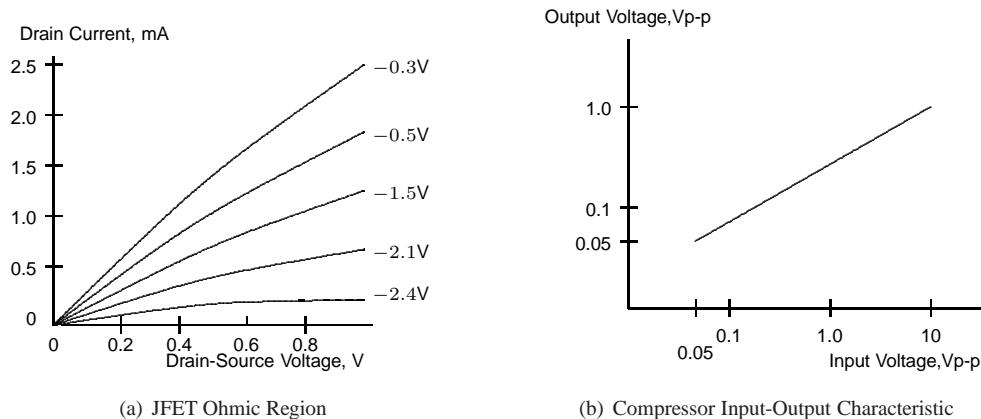


Figure 872: Compressor Details

31.3 The JFET Switch

There are many applications for an electronic switch, one that can block or pass an AC signal. For example, such a switch can replace a mechanical switch or relay with a substantial saving in size and expense. An electronic switch can turn on and off at high frequencies, thereby becoming a type of modulator or multiplier.

Unlike the BJT, a JFET can conduct equally well in either direction, so it's a promising device to use as the basis of an electronic switch. In practice, the JFET analog switch has largely been replaced by the CMOS integrated circuit analog switch (section 33, page 1009) which is easier to apply and requires fewer parts. However, JFET switches appear in legacy circuits and make sense in a few applications, so we'll study them briefly here.

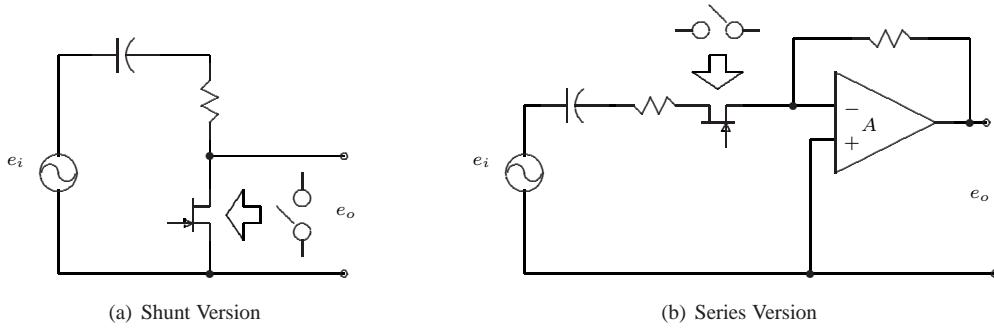


Figure 873: JFET Analog Switch

Basic JFET switch configurations are shown in figure 873.

The shunt switch in figure 873(a) routes the input signal to ground when it is conducting. Strictly speaking, the input capacitor is not required, but it emphasizes that the input is an AC signal.

The series switch in figure 873(b) has the opposite effect: it allows the signal to pass when it is conducting. The op-amp establishes a virtual earth point (essentially zero volts) at the output of the JFET switch, which simplifies the design of the switch.

Figure 874 shows the drive circuitry for the shunt switch, with the addition of resistor R_g , diode D and switching voltage V_g (g for gate). The drive circuit is identical for the series JFET switch version.

When the gate drive V_g is positive, the diode is reverse biased and appears as an open circuit. Then the gate is connected to the drain terminal and the JFET is ON.

When the gate drive V_g is negative, the diode conducts. The gate terminal is forced to a negative voltage, and the JFET is OFF.

This will work, regardless of the polarity of the input signal, as long as the gate voltage V_g is greater than the voltage appearing across the drain-source terminals. The JFET is a symmetrical device, and when the input voltage reverses in polarity, the drain and source terminals exchange.

In some applications, a single JFET switch does not attenuate the signal sufficiently. Then *both* a series and shunt JFET switch may be used in combination.

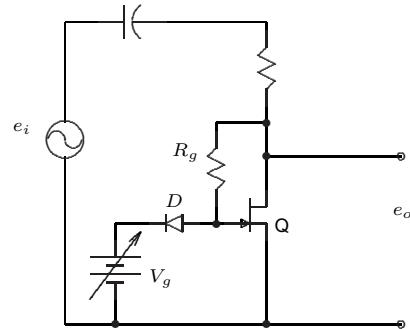


Figure 874: Shunt Switch with Drive

31.3.1 Opto-Isolated JFET Switch

The JFET switch shown previously must share a common terminal between the signal and the switching voltages. There are applications where it is desirable for reasons of noise isolation to keep these two signals on entirely separate ground systems.

An opto-isolated JFET controls the conduction of the drain-source path from an integral light-emitting diode source, so the grounds can be entirely separate. Furthermore, this eliminates the requirement of the previous switch design that the switching voltage exceed the signal voltage. For the opto-isolated JFET, it is sufficient to ensure that the LED conducts sufficiently.

As in the case of a normal JFET, the drain-source terminals of the opto-isolated JFET may be interchanged, so the channel conducts equally well in either direction and appears as a low-value resistance when conducting.

A commercial product is the H11L1,2,3 *Photo-FET Optocoupler*, [284], shown in figure 875. There are some limitations to this device. As a switch, the datasheet shows the LED requires 16mA of forward current for the JFET to be in the ON state. As a variable resistor, the voltage across the JFET must be limited to something in the order of 100mV to keep the JFET in its ohmic region. With a 25 μ sec switching speed, it's slow compared to other solid-state switches. However, it is bi-directional and the speed is an improvement over an isolated resistance such as an LED-CdS photoresistor pair [285].

31.4 The JFET Amplifier

In section 29, page 798, we showed how the BJT could be used to amplify bipolar (AC) signals.

We also discussed three BJT amplifier configurations: common-emitter, common-collector and common-base.

The JFET may be used in circuits that are remarkably similar to the BJT amplifier circuits. However, there are some differences.

The primary difference between the BJT and JFET is the variability in bias conditions. In an amplifier configuration, the base-emitter voltage of a BJT is in the order of 0.65 volts, give or take a few millivolts and the collector bias current is quite predictable – even when BJT's are changed.

Figure 869 shows that this is not the case for the JFET. There is an enormous unit-unit variation in transfer characteristic. For one-off circuits, the JFETs can be selected to meet certain requirements or the bias circuit adjusted to suit the JFET. A production can only use a JFET if the requirement is tolerant of this large variation. For example, in amplifying small signals clipping is not an issue so a large variation in bias point is acceptable. Alternatively, the design requires a circuit configuration of matched devices that cancels out this variation. Unfortunately, matched JFETs are expensive.

31.4.1 Playing to the Strengths of the JFET

The gate current of a JFET is orders of magnitude smaller than the base current of a BJT. As a result, large-value resistors can be used in the gate circuit of a JFET without loading the source circuit or affecting the bias point²⁷¹.

The BJT and JFET have different noise behaviour. The BJT has low voltage noise and high current noise, so best noise behaviour is obtained in a circuit with low impedances. The JFET is the reverse: high voltage noise and low current noise.

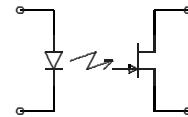


Figure 875: Opto-Isolated JFET

²⁷¹JFET gate current increases rapidly with temperature so if the JFET must operate at an elevated temperature the circuit must be checked for correct operation under that condition. Also, keep in mind that noise increases with the square root of resistance (section 22.1 on page 670), so a large value gate resistor may contribute significant noise to the circuit.

31.4.2 BJT and JFET Op-Amps

Operational amplifiers are available with BJT inputs and with JFET inputs. Using an op-amp with JFETs in the input is a simple way to take advantage of the strengths of the JFET without having to deal with bias variation. The input circuitry JFETs are inherently matched in the integrated-circuit process. The first JFET op-amps had significantly higher offset voltage than BJT op-amps, but that difference has disappeared in modern devices.

Transconductance of the BJT

The BJT may be regarded as a device in which the output (collector) current is controlled by a much smaller input (base) current, in which case the constant of proportionality is *beta*.

Alternatively, we can regard the BJT as a device in which its output current is controlled by its base-emitter voltage. Increasing base-emitter voltage increases the collector current in an exponential fashion.

The slope of this exponential transfer characteristic is the *transconductance* of the BJT (section 29.16). Since the transfer characteristic is a curve, the slope is highly dependent on the operating bias point. It is given (approximately) by

$$g_m = \frac{I_E}{0.026} \text{ amps/volt} \quad (1516)$$

where I_E is the DC current establishing the operating point.

For example, at an emitter current of 1mA, the transconductance of a BJT is:

$$g_m = \frac{I_E}{0.026} = \frac{0.001}{0.026} = 38 \text{ mA/volt} \quad (1517)$$

That is, if the BJT is biased by a 1mA DC collector current, then applying an AC voltage of 10mV p-p to the base would cause the collector current to vary by 0.38mA p-p.

As well, we showed in section 29.16 that this transconductance could be modelled as an *emitter incremental resistance* r_e , which is the reciprocal of the transconductance.

$$r_e = \frac{0.026}{I_E} \text{ volts/amp (ohms)} \quad (1518)$$

where I_E is the DC current establishing the operating point.

Again, at an emitter current of 1mA, the emitter incremental resistance is

$$r_e = \frac{0.026}{I_E} = \frac{0.026}{0.001} = 26 \text{ volts/amp (ohms)}$$

31.4.3 Transconductance of the JFET

The JFET input terminal is connected to a diode that must be reverse biased. Consequently, there is no significant input current. The JFET is controlled by its gate-source voltage: a more negative voltage reduces the drain current.

The transconductance of a JFET is a measure of the gain of the device. It is the ratio of the change in drain current to the change in gate-source voltage ²⁷².

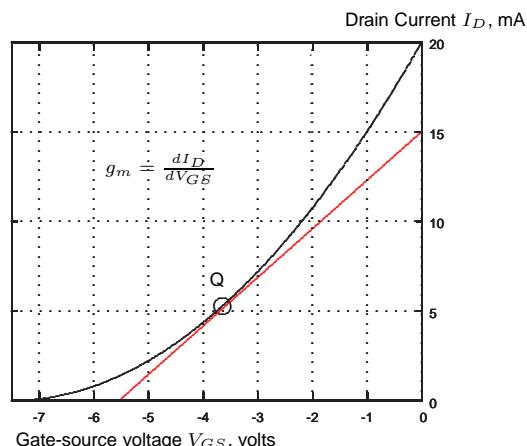


Figure 876: Transconductance of the JFET

²⁷²Compare this to the BJT case (section 29.16 on page 836, where g_m is the ratio of the change in collector current to the change in base-emitter voltage. In both cases, the transconductance is the ratio of change in output current (I_d or I_c) to change in input voltage (V_{gs} or V_{be}).

$$g_m = \frac{dI_d}{dV_{gs}} \text{ amps/volt (mhos)} \quad (1519)$$

It turns out that the transconductance of the JFET is significantly smaller and less predictable than that of the BJT.

Graphically, the transconductance is simply the slope of the transfer characteristic, a plot of drain-current vs gain-source voltage (figure 876).

We obtain the transconductance by differentiating equation 1515 (page 978) for the transfer characteristic:

$$\begin{aligned} g_m &= \frac{d(I_D)}{d(V_{gs})} \\ &= \frac{d}{d(V_{gs})} \left[I_{DSS} \left(1 - \frac{V_{GS}}{V_P} \right)^2 \right] \\ &= \frac{2I_{DSS}}{V_P} \left(\frac{V_{GS}}{V_P} - 1 \right) \end{aligned} \quad (1520)$$

Example

Calculate the transconductance of a JFET which has a drain saturation current I_{DSS} of 20mA and pinch-off voltage V_P of -7 volts. The JFET is being operated at a quiescent point of $V_{GS} = -3.5$ volts (figure 876).

Solution

Substitute in equation 1520:

$$g_m = \frac{2I_{DSS}}{V_P} \left(\frac{V_{GS}}{V_P} - 1 \right) = 2 \times \frac{20}{-7} \left(\frac{-3.5}{-7} - 1 \right) = 2.85 \text{ mA/Volt}$$

Notice that this is an order of magnitude less than the BJT transconductance calculated above: 38 mA/V for the BJT compared to 2.85 mA/V for the JFET.

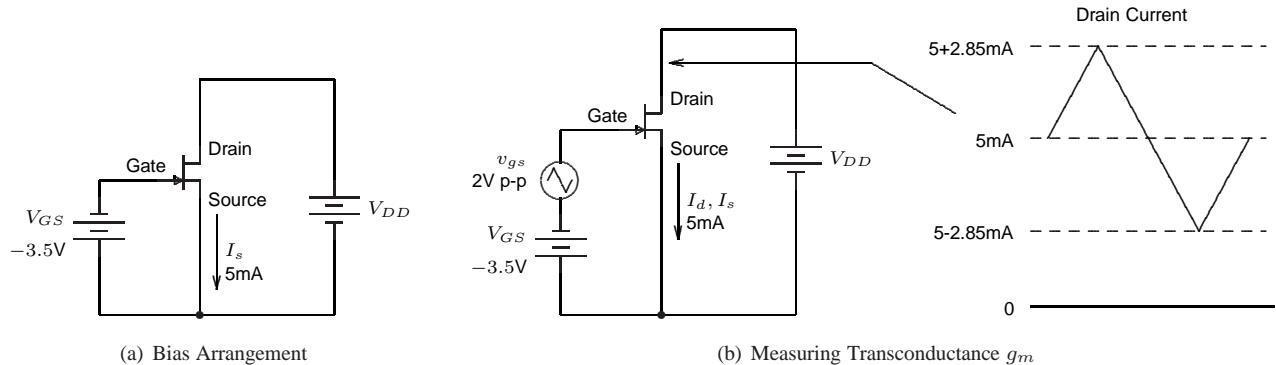


Figure 877: Measuring JFET Transconductance

The measurement circuit is shown in figure 877. Figure 877(a) shows the JFET biased by a $-3.5V$ V_{GS} constant-voltage source. Referring to figure 876, this establishes the DC drain current (and source current, since they are equal in the JFET) as 5mA.

In figure 877(b) we add an alternating voltage source v_{gs} of 2 volts peak-peak, in series with the gate bias. This has the effect of varying the drain current by an amount

$$i_d = g_m v_{gs} = 2.85 \text{ mA/Volt} \times 2 \text{ Volts p-p} = 5.7 \text{ mA p-p} = \pm 2.85 \text{ mA} \quad (1521)$$

31.4.4 Transresistance of the JFET

We can also express the JFET transconductance as a transresistance, which is analogous to the BJT emitter incremental resistance r_e . We'll call it the incremental source resistance r_s :

$$r_s = \frac{1}{g_m} = \frac{1}{2.85 \times 10^{-3}} = 350 \text{ volts/amp (ohms)}$$

This resistance is an order of magnitude larger than the BJT incremental emitter resistance. We will see that the resultant single stage voltage gain from a JFET stage is smaller than that of a BJT.

31.5 The Source Follower

The BJT is inherently a low input impedance device although the input impedance can be raised by various feedback techniques. The JFET input is inherently a high input impedance device. It appears as an open circuit with a few picofarads of stray capacitance to the drain and source terminals. Consequently the JFET can be used as a very simple device to present a source with a high impedance and a load with a low impedance.

The BJT emitter follower was discussed in section 29.27, page 870. The JFET equivalent, the *source follower*, is shown in figure 878. There are many possible variations on this circuit. This one is a simple AC amplifier or *buffer*²⁷³.

Like the BJT emitter follower, the JFET source follower has a low output impedance. In figure 878, the input resistor R_i can be very large compared to the load resistor R_l while the overall voltage gain v_l/e_i remains close to unity.

31.5.1 Notes on the Source Follower Circuit

- The source follower of figure 878 uses *self-bias* to establish a DC bias current in the JFET.

Then the input voltage modulates that DC current, causing it to increase or decrease – but never to reverse. This variation in current causes a varying voltage to appear across the source resistor R_s . The output capacitance C_o extracts the AC component and passes it on to the load resistor R_l . The coupling capacitor limits the minimum frequency of operation. The circuit will only operate at frequencies high enough that the capacitor appears as a low impedance.

- Current through resistor R_s establishes a positive voltage at the source terminal. The gate terminal is grounded via the gate resistor R_g . A positive voltage on the source reverse-biases the gate-source diode so

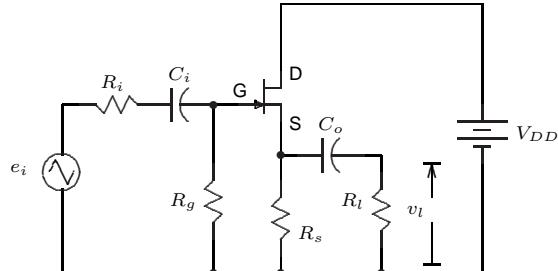


Figure 878: JFET Source Follower

²⁷³A *buffer amplifier* usually describes an amplifier with high input impedance, low output impedance, and voltage gain near unity. It has no voltage gain but it does have power gain.

there is very little current in the gate terminal and its voltage is essentially zero. The whole system settles out at a gate-source voltage determined by the intersection of the source resistance R_s and the transfer characteristic, as shown in figure 868 on page 977.

- The input capacitor C_i removes any DC component in the input signal, which may or may not be a requirement. If the DC component is small or non-existent, then C_i is not required.
- The gate resistor R_g provides a path for gate leakage current. It's required if there is no other DC path for current to ground. For example, if the source is a piezoelectric crystal, which is an open-circuit for DC, then the gate resistor is required. On the other hand, if the source is a moving coil sensor, then it appears as a (near) short-circuit to ground. Then remove input capacitor C_i and the gate resistor R_g is not necessary.
- In this circuit, the source terminal is sitting at some positive voltage with respect to ground. Capacitor C_o prevents that voltage from setting up a DC current in the load resistor R_l . For example, if the load is a moving-coil earphone, the DC current would bias the diaphragm to one side, increasing distortion.
- The designer must take into account the variability of JFET characteristics. Refer again to the transfer characteristic of figure 868(c) on page 977. In a production run, the Q point can vary substantially. The drain current could be anywhere between some fraction of 1mA up to 5mA, and the corresponding gate-source voltage 0.3V up to 3.5V. The maximum allowable signal before clipping is limited by the lower of these numbers. For example, if the gate-source bias voltage is 0.3 volts, then the voltage swing at the source is limited to something less than 0.15 volts peak.

31.5.2 The JFET Small Signal Model

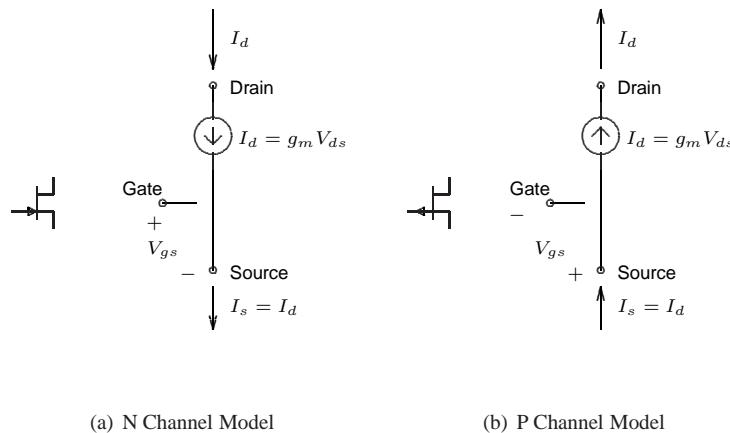


Figure 879: JFET Small Signal Models

When the JFET is biased so that it is operating in its constant-current region, then the small signal model is as shown in figure 879. The gate terminal is simply an open circuit. The source current and drain current are equal, and the drain current is equal to the transconductance times the gate-source voltage.

This is a low frequency model: it ignores stray capacitances and other high-frequency effects. It also assumes small currents, so that the resistance of the drain-source channel can be neglected.

31.5.3 Voltage Gain of the Source Follower

We'll illustrate the application of the model by calculating the voltage gain of the source follower. The model for the small-signal analysis of the source follower of figure 878 (page 986) is shown in figure 880.

The model replaces capacitors by short circuits. It ignores the input resistance R_i since it's connected to an open circuit, conducts no current and therefore has no voltage drop. It combines the load resistance R_L and source resistance R_s into one resistor. And it replaces the drain supply by a short circuit, since a pure voltage source has zero internal resistance.

There are three equations and then some algebra. By KVL around the gate-input loop,

$$+ v_o + v_{gs} - e_i = 0 \quad (1522)$$

$$+ v_o = i_s R_s = i_d R_s \quad (1523)$$

$$+ i_d = g_m v_{gs} \quad (1524)$$

Combine and solve for the voltage gain v_o/e_i and we obtain:

$$\frac{v_o}{e_i} = \frac{R_s}{1/g_m + R_s} \quad (1525)$$

This is where the concept of *transresistance* r_m becomes useful. Replace $1/g_m$ by r_m :

$$\frac{v_o}{e_i} = \frac{R_s}{r_m + R_s} \quad (1526)$$

If we now refer to the BJT *Summary of Emitter Follower Equations* in section 29.29 (page 876), we find that the equation for voltage gain of the emitter follower is very similar to that of the JFET. In the BJT emitter follower, the voltage gain is somewhat less than unity and depends on the ratio of the emitter incremental resistance r_e to the load resistance. In the JFET source follower, the gain is somewhat less than unity and depends on the ratio of the transresistance r_m to the load resistance. And because the BJT incremental emitter resistance is less than the JFET transresistance, we can predict that the JFET follower voltage gain is likely to be less than the BJT emitter follower voltage gain.

This is a *Happy Moment in Electronics*, because it means that the concepts which we struggled to understand about the BJT can largely translate to the JFET.

31.6 JFET Follower Oscillation

As in the case of the emitter follower (section 29.28 on page 875), a JFET follower can oscillate at very high frequency. As in that case, the symptom is a DC offset that changes in unpredictable ways that do not seem to be related to the DC conditions of the circuit²⁷⁴.

The analysis is shown in [79] and [286]. (In this respect, the MOSFET and JFET are very similar, so the analysis is comparable.) The oscillation may be cured with the addition of a resistance (eg, 3000Ω) in series with

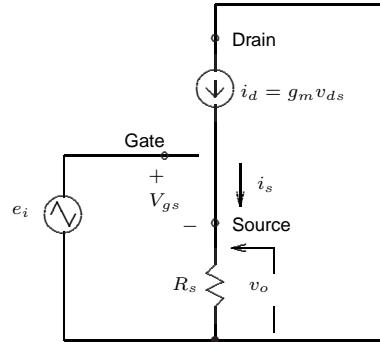


Figure 880: JFET Source Follower Model

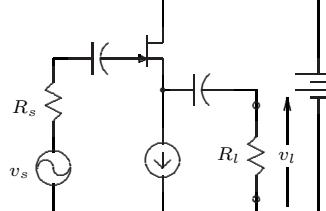
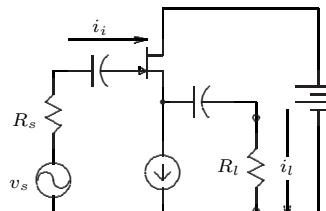
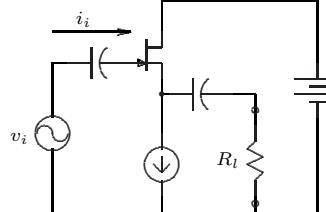
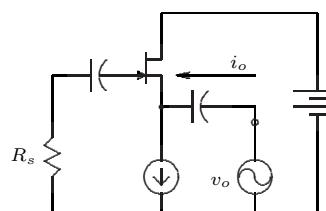
²⁷⁴Morgan [286] describes a case where attaching an inductance to the input of an oscilloscope caused its preamplifier to oscillate!

the base lead. For low frequency applications, this is sufficient. Unfortunately, for high frequency applications, this resistance may interact with the device capacitances to limit the frequency response. Morgan [286] shows a somewhat more sophisticated approach, with a shunt RC network attached to the gate terminal. This is sufficiently complex that it should be verified with simulation.

31.6.1 Summary: Source Follower Equations

The equations for the JFET single stage source follower amplifier, along with some of the usual approximations, are shown below. These equations assume a constant current bias. The equations must be modified if a resistor R_s is used to implement the source bias current sink. A DC path for gate current – such as a resistor between the base terminal and ground – must exist, but for simplicity it's not shown on the schematics.

It's not so important to memorize the equations as it is to memorize the model. Then the equations may be determined for this and other similar circuits.

Voltage Gain		$A_v = \frac{v_l}{v_s} = \frac{R_l}{r_m + R_l}$
Current Gain		$A_i = \frac{i_l}{i_i} = \infty$
Input Resistance		$r_i = \frac{v_i}{i_i} = \infty$
Output Resistance		$r_o = \frac{v_o}{i_o} = r_m$

31.7 JFET Common Source Amplifier and Common Gate Amplifier

The BJT common-emitter amplifier and common base amplifier were described and analysed in sections 29.19 through section 29.33 (pages 839 through 890). We leave the equation derivations to the curious reader (and the beleaguered student solver of the section exercises), for three reasons:

- With the introduction to the JFET in this section, is straightforward to extrapolate to the Common Source and Common Gate JFET amplifiers. Change the bias circuit to suit the JFET, recognize that the gate terminal is an open circuit, and replace r_e by r_m .
- Because of the unit to unit variation in JFET drain-saturation current I_{DSS} and pinch-off voltage V_p , these are not particularly useful circuits in a production circuit.
- The advantages of a JFET amplifier can be obtained as a JFET input operational amplifier integrated circuit, where the parameter variations are taken care of by matching and IC fabrication techniques.

31.8 JFET Compound Amplifiers

In this section, we show amplifiers using two JFET devices.

31.8.1 JFET Differential Amplifier

A simple high-impedance voltmeter circuit, adapted from [33], is shown in figure 881. This circuit is a JFET adaptation of a vacuum tube voltmeter measurement circuit known as a VTVM (vacuum tube volt meter), [32] in which two JFETs replace the original dual-triode vacuum tube. The reading on the $200\mu\text{A}$ moving-coil meter is proportional to the input voltage V_i .

The *Balance* control adjusts the currents in the two sides of the differential amplifier so that they are equal when the input voltage is zero. Ideally, the JFETs should be matched, but the *Balance* control compensates for any remaining difference between the transfer characteristics of the two devices. The *Scale* control sets the deflection of the ammeter so that the scale factor of the voltmeter is correct.

This circuit illustrates a discrete component JFET amplifier and it has a certain retrograde charm, but there are better approaches with modern components. Question 4 on page 995 shows one possibility.

31.8.2 Oscilloscope Buffer Amplifier

Figure 882 shows a clever circuit that is frequently used as the input impedance buffer in an oscilloscope vertical amplifier.

A simple JFET follower (figure 880 on page 988 for example) has a source terminal voltage equal to the gate-source voltage V_{gs} , when the gate is at zero volts. In figure 880 this DC offset is removed by an output coupling capacitor C_o . This DC offset is a problem in DC coupled amplifiers. We'd prefer the output to be zero volts when the input is zero volts. The circuit of figure 882 solves that problem.

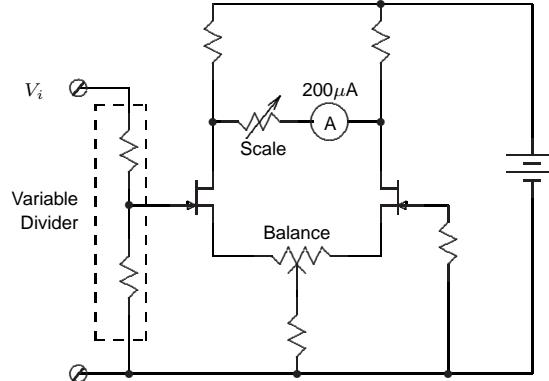


Figure 881: JFET Differential Amplifier

Assume that the two JFETs Q_1 and Q_2 are matched, so they have the same gate-source voltage when their drain currents are equal. Also consider that R_1 and R_2 are equal.

The lower JFET Q_2 is operating at self-bias condition with some value of drain current and gate-source voltage V_{gs} . The gate of Q_1 is grounded through R_g , so its potential is zero volts. Then the source voltage of Q_1 is sitting at $+V_{gs}$ volts.

The resistors are equal, and both are conducting the same current, so both have V_{gs} volts across them. Then the voltage drop across the upper resistor R_1 exactly cancels the voltage at the source of Q_1 , and the output voltage is zero volts.

In practice, one or other of the resistors is made adjustable to compensate for any difference between the JFETs.

Is this a practical circuit in modern applications? On the one hand, one JFET or MOSFET op-amp buffer can replace the entire circuit. On the other hand, modern op-amps tend to have low voltage supplies and are therefore limited in their signal-handling capability. A typical high-speed mosfet op-amp might use supplies of $\pm 2.5V$, for example. In contrast, discrete component JFETs can typically withstand 30 volts, so the circuit could operate from, say $\pm 12V$ supplies.

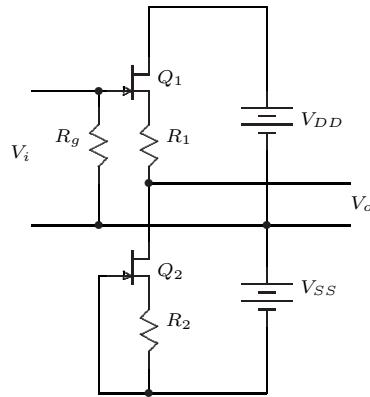


Figure 882: JFET Buffer Amplifier

31.9 Appendix: Interpreting the JFET Datasheet

The MPF102 is an inexpensive, general purpose, N-Channel JFET. The first page of its datasheet, which contains the electrical specifications, is shown in figure 883 on page 994.

Here is a brief description of each of the quantities on the datasheet.

Absolute Maximum Ratings	The device is seriously damaged or destroyed if these ratings are exceeded.
Electrical Characteristics	
$V_{(BR)GSS}, I_{GSS}$	Not particularly important, but these figures identify the reverse leakage current of the gate-source diode at the maximum allowed reverse voltage.
$V_{gs(off)}$	The gate-source voltage required to reduce the drain current to 2nA.
V_{gs}	The gate-source voltage that corresponds to a small value of drain current, 200 μ A. Notice that this can vary over a substantial range, between -0.5V and -7.5V.
I_{DSS}	The drain current corresponding to a gate-source voltage of zero volts. Like the previous value for V_{gs} , this varies substantially. The saturation current for a given MPF102 can be anywhere between 2.0mA and 20mA. The forward transconductance, di_d/dv_{gs} . This is the primary gain specification of the device. For example, the maximum value is 7500 μ S, 7500×10^{-6} Siemens (amps/volt), which corresponds to 7.5mA/V.
g_{fs}	
Small Signal Characteristics	
C_{iss}	The parasitic capacitance appearing between the input (gate) terminal and the source terminal.
C_{rss}	The parasitic capacitance appearing between the output (drain) terminal and the input (gate) terminal.
Thermal Characteristics	Both these capacitances are relevant to the high-frequency model of the JFET.
	These specifications are not terribly important, since the JFET is not normally required to dissipate significant power. See section 28 on page 783.

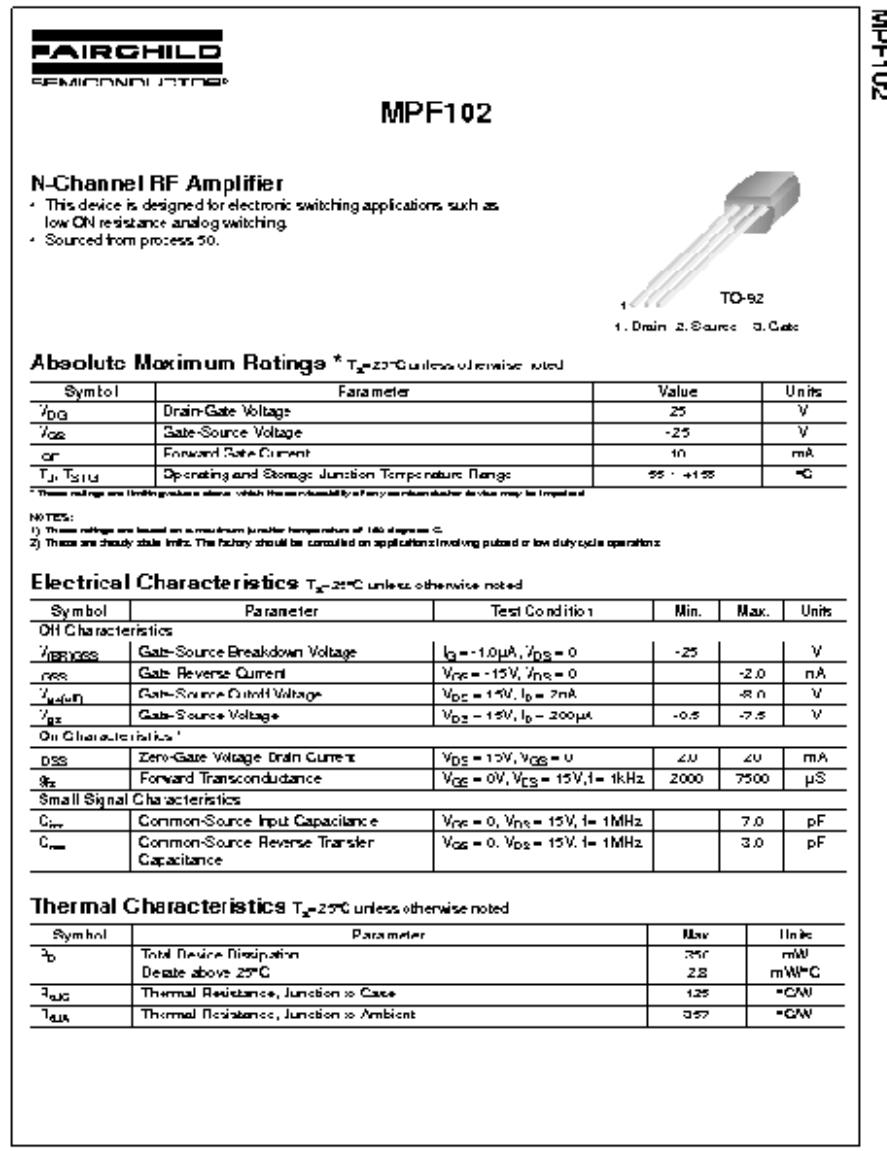


Figure 883: MPF102 Datasheet

31.10 Exercises

1. The PN4392 JFET is specified to have the following characteristics:

Pinch-Off Voltage V_P (V_{GSOFF})	-2V (min)	-5V (max)
Drain Saturation Current I_{DSS}	25mA	75mA (max)

- (a) Sketch the transfer characteristic of the JFET.
 - (b) Determine the JFET transconductance at drain current of 5mA.
 - (c) This JFET is biased by a fixed gate-source voltage of -1V. Draw the bias line on the transfer characteristic. Over a production run, what circuit conditions would vary and by how much?
 - (d) This JFET is biased by a fixed source current of 10mA. Draw the bias line on the transfer characteristic. Over a production run, what circuit conditions would vary and by how much?
2. An audio amplifier company is using JFETs in its circuits and decides to match the devices by hand.
- Design a circuit to measure the drain saturation current I_{DSS} and pinch-off voltage V_P for an N-Channel JFET. The circuit should use a multimeter on its current and voltage ranges as the readout device.
 - Modify the circuit so it can be switched to accommodate both N-Channel and P-Channel JFETs.
3. (a) In the compressor circuit of figure 872 on page 981 determine the relationship between the DC output voltage e_{oDC} and the resistance of the JFET Q1. Hint: it is not necessary to know any characteristics of the JFET except that it is operating in its ohmic region.
- (b) The compressor circuit uses a dual JFET, 2N3958. That device is obsolescent. Redraw the circuit, replacing the dual JFET with two opto-isolated JFETs H11F1 (See datasheet at [284]).
4. The circuit shown in figure 884 below is a modern version of the voltmeter previously shown in figure 881 on page 991. The op-amp has JFET or MOSFET inputs.

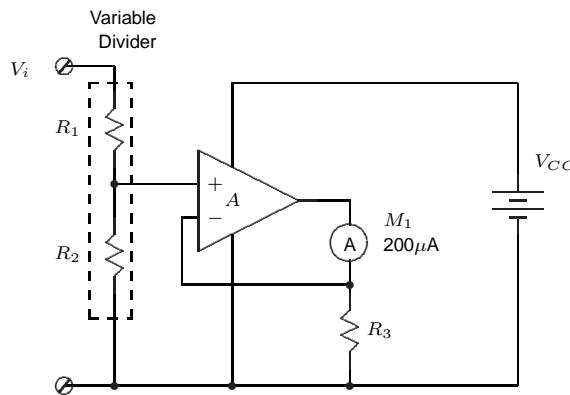


Figure 884: Modern JFET Voltmeter

- (a) Relate the meter current to the input voltage and the values of the resistors in the circuit.
- (b) The *Balance* control is no longer necessary. Why?
- (c) What output voltage range is required of the operational amplifier (minimum and maximum)?

- (d) What range of common-mode input voltage is required of the operational amplifier (minimum and maximum)?
- (e) It might be necessary to use two batteries in a split-supply arrangement for this circuit. Why?
- (f) The circuit can be modified to measure rectified average voltage by putting a diode bridge around the ammeter. Draw the modified circuit.
- (g) What effect will the diode voltage drop have on the meter reading? Why?

32 The MOSFET Family

32.1 Linear Applications of the MOSFET Transistor

Section 8.3 served as an introduction to the MOSFET transistor. Used as a switch, the MOSFET is a simple device and the information in that section is sufficient to use power MOSFETs in common switching applications. In this section, we discuss the use of MOSFETs in linear applications.

The MOSFET *transfer characteristic* relates the input control signal of the MOSFET (the gate-source voltage) against the output (the drain current). It gives us a useful entry into understanding MOSFET devices and applications, so we'll start there.

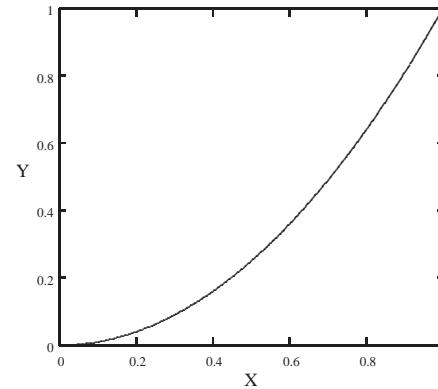


Figure 885: Square Law Function

32.2 MOSFET Transfer Characteristic

All field effect transistors – the JFET and MOSFET – are basically square law devices. That is, a plot of the drain current against gate-source voltage (the *transfer characteristic*) shows that drain current is proportional to the *square* of the gate-source voltage²⁷⁵.

Consider the square-law characteristic shown in figure 885 and equation 1527. This is the basis for the transfer function of the MOSFET.

$$Y = X^2 \quad (1527)$$

The MOSFET form of the square-law equation looks like this:

$$\frac{I_D}{I_{DSS}} = \left(\frac{V_{gs}}{V_p} \right)^2 \quad (1528)$$

As in equation 1527, in equation 1528 both the LHS and the RHS run from 0 to 1. The dependent variable (Y axis) is the ratio of drain current I_D to some maximum value of drain current I_{DSS} . The independent variable (X axis) is the ratio of gate-source voltage V_{gs} to some maximum value of gate-source voltage V_p . Equation 1528 is usually written as:

$$I_D = I_{DSS} \left(\frac{V_{gs}}{V_p} \right)^2 \quad (1529)$$

For example, if the value of I_{DSS} is 20mA and V_p is 3 volts, then the transfer characteristic is as shown in figure 886.

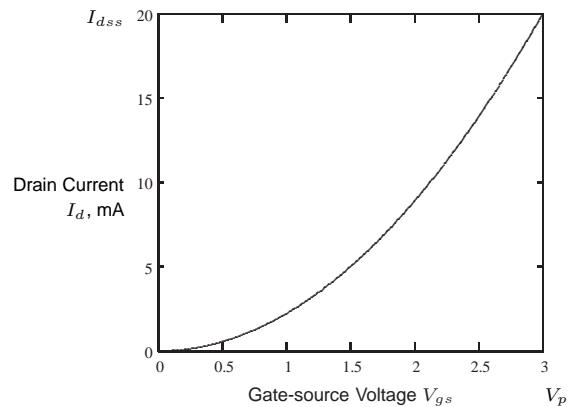


Figure 886: MOSFET Transfer Curve

²⁷⁵Such a non-linear relationship is a poor basis for linear amplifier design. The behavior can and must be linearized with various forms of negative feedback.

32.3 The Enhancement Mode MOSFET

The curve of figure 1529 describes an N-channel *enhancement mode* MOSFET. The FET is normally off and is caused to conduct by making the gate positive with respect to the source. In some enhancement mode MOSFETs, the characteristic is offset to the right by a *threshold voltage* V_t , which can range between 0.7V and 4V. Then the equation of the transfer characteristic becomes:

$$I_D = I_{DSS} \left(\frac{V_{gs} - V_t}{V_p} \right)^2 \quad (1530)$$

As illustrated in figure 888, the P-channel enhancement mode MOSFET differs only by the polarity of the voltage and current. Drain current is zero for zero gate-source voltage. For this device, the drain current flows into the source and out of the drain. The drain current increases as the gate-source voltage is made more negative. In effect, the transfer characteristic is identical to figure 887, but rotated through 90°.

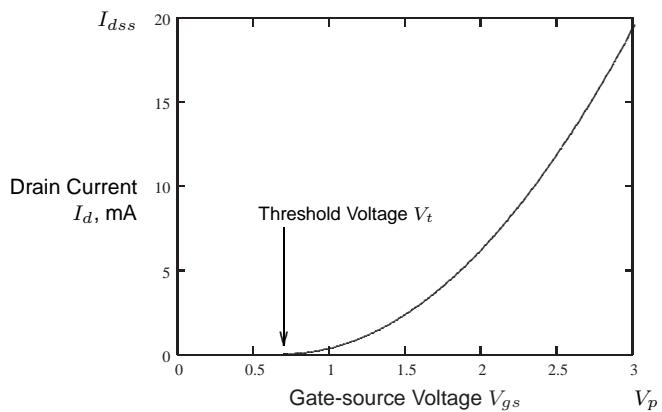


Figure 887: Threshold Voltage V_t

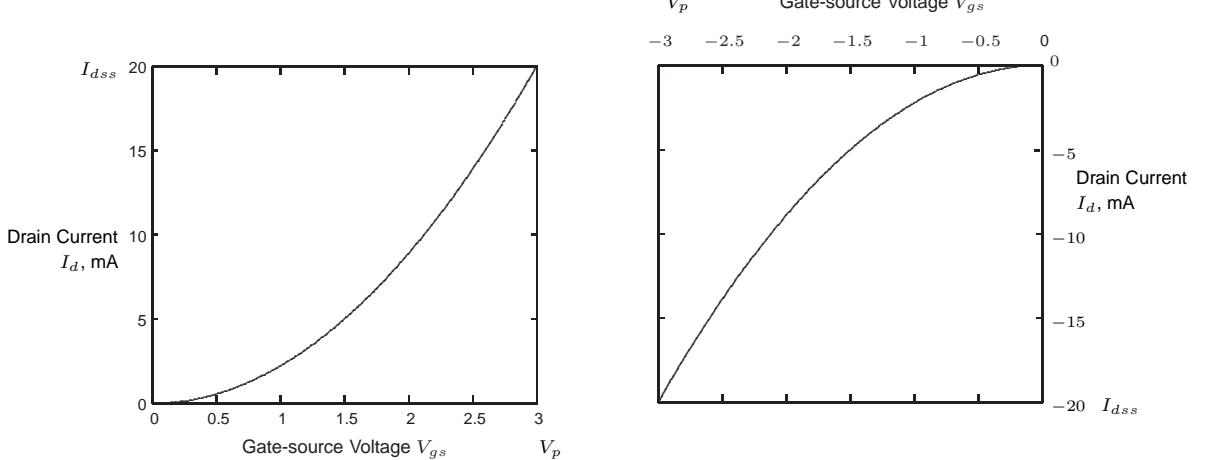


Figure 888: N and P Channel Enhancement Mode MOSFETs

In practice, the device manufacturer can manipulate the construction of the MOSFET to produce a transfer characteristic that is essentially linear above the threshold voltage. Consequently, the best design approach is graphical, using the transfer characteristic from the device datasheet.

32.4 The Depletion Mode MOSFET

We've already met an example of a depletion mode device: the JFET introduced in section 8.2 (page 268) and elaborated in section 31 (page 976). The enhancement-mode MOSFET is like a Normally Off switch, enabled by increasing gate-source voltage. The depletion-mode MOSFET is like a Normally On switch, disabled by increasing gate-source voltage.

The transfer characteristic of a depletion mode MOSFET is shown in figure 889.

Many of the concepts of the enhancement mode MOSFET apply to a depletion mode device. We focus on enhancement mode devices because they are more readily available.

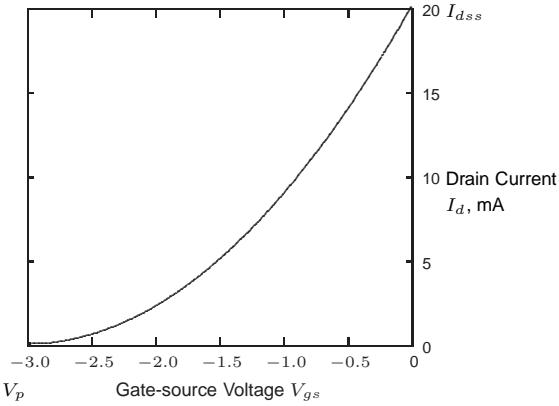


Figure 889: Depletion Mode MOSFET Transfer Curve

32.5 DC Coupled Source Follower

Although MOSFETs can be used in common-source and common-gate configurations, their most useful application is in the source follower (common drain) configuration. The source follower is an effective impedance converter or buffer. It presents a very large DC load resistance to the source and much lower internal resistance to the load. In this section we'll first look at biasing an enhancement-mode source follower. Then we'll examine the performance of the follower.

32.5.1 Biasing the Follower

Figure 890 shows a DC coupled source follower. Resistor R_g is necessary to conduct any gate leakage current and although this is very small, some sort of DC path to ground is required. Since R_g conducts almost no DC current (since the gate of the MOSFET is an open circuit for DC), the gate voltage is exactly equal to v_i . For the purpose of bias analysis, we assume v_i is zero.

The negative supply V_{ss} pulls the source terminal of the MOSFET negative, which creates a positive gate-source voltage V_{gs} . As the gate-source voltage increases, more drain current (and source current, since they are equal in a MOSFET) flows, according to the transfer characteristic. This current creates a voltage across the source resistor R_s that makes the

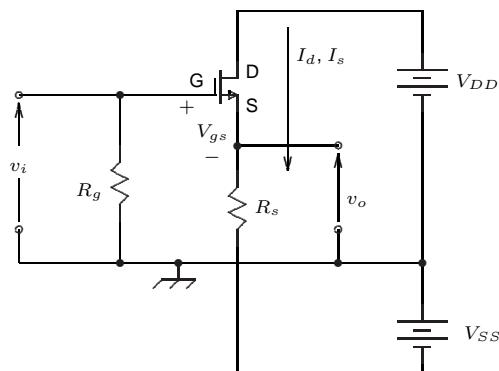


Figure 890: N Channel, Enhancement Mode, Source Follower, DC Coupled

MOSFET source terminal more positive. The circuit reaches equilibrium at the Q point, which we'll determine by a graphical construction, figure 891.

To develop this graphical construction, consider KVL around the loop including gate voltage:

$$-V_{gs} - I_d R_s + V_{ss} = 0 \quad (1531)$$

We superimpose this on the transfer characteristic in order to show us the Q point. Equation 1531 is a straight line, with X axis intercept at $I_d = 0$ and y intercept at $V_{gs} = 0$. Plugging those values into equation 1531 to obtain the X and Y intercepts, we obtain a straight line that runs between a current of V_{ss}/R_s on the current axis to V_{ss} on the voltage axis.

The result is shown on figure 891, using the transfer characteristic for the MOSFET of figure 886, negative supply V_{ss} of 12 volts, and source resistor R_s of $2k4\Omega$. The MOSFET Q point is at $V_{gsq} \approx 1.4V$, $I_{dq} \approx 4mA$. Consequently, in figure 890 we'd expect to find the source terminal sitting at -1.4 volts with respect to ground.

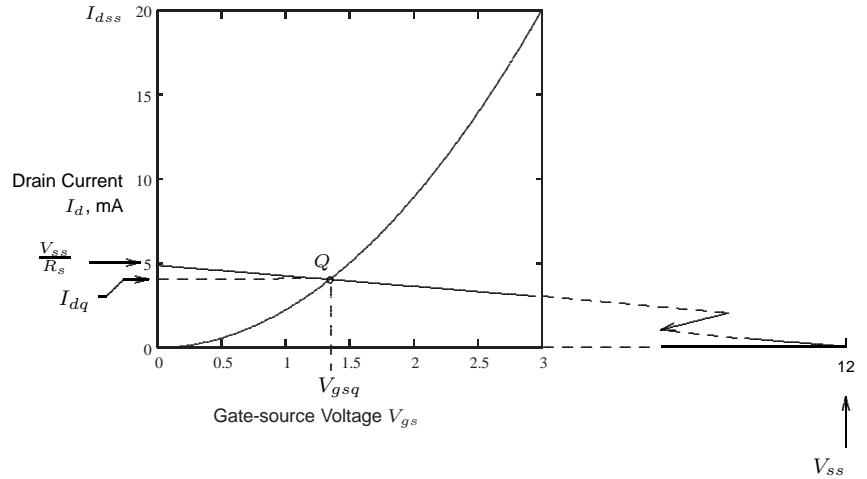


Figure 891: Bias Condition for DC Follower

32.5.2 Voltage Gain of the Follower

First, we will determine the AC gain of the follower circuit of figure 890. The AC equivalent circuit for figure 890 is in figure 892.

The equations to determine the AC gain are:

$$v_o = i_d R_s \quad (1532)$$

$$i_d = g_m v_{gs} \quad (1533)$$

$$v_{gs} = e_i - v_o \quad (1534)$$

Collapsing these three equations and solving for voltage gain $A_v = v_o/e_i$, we obtain:

$$A_v = \frac{v_o}{e_i} = \frac{g_m R_s}{1 + g_m R_s} \quad (1535)$$

This is fine as far as it goes, but we can get a bit more insight. In section 29.16 (page 29.16) on the BJT, we showed that there is a relationship between transconductance g_m and incremental emitter resistance r_e :

$$g_m = \frac{1}{r_e} \quad (1536)$$

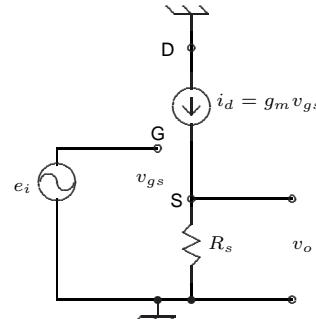


Figure 892: Small Signal AC model of Source Follower

Unlike the BJT, the MOSFET does not have any physical quantity that can be recognized as an incremental resistance. Nonetheless, it is useful to create an analogous incremental resistance which we'll call the *transresistance* r_m . For the MOSFET:

$$r_m = \frac{1}{g_m} \quad (1537)$$

Now substitute this in equation 1535 and simplify. The result is:

$$A_v = \frac{v_o}{e_i} = \frac{R_s}{r_m + R_s} \quad (1538)$$

This equation describes a voltage divider, so the MOSFET AC equivalent circuit can be redrawn as figure 893.

It's clear from this diagram that the gain of the circuit is dependent on the ratio of r_m to R_s . If the source resistance R_s is much larger than r_m the gain will approach unity.

32.5.3 Output Resistance of the Follower

The test setup for calculating the output resistance of the source follower is shown in figure 894. The relevant equations are:

$$r_o = \frac{v_o}{i_o} \quad (1539)$$

$$i_o = i_s - i_d \quad (1540)$$

$$i_s = \frac{v_o}{R_s} \quad (1541)$$

$$i_d = g_m v_{gs} \quad (1542)$$

$$v_{gs} = -v_o \quad (1543)$$

Notice that v_{gs} is negative with respect to v_o because a more positive value of v_o makes the gate-source voltage less positive (more negative). Back collapse these equations and put $g_m = 1/r_m$ and we obtain:

$$r_o = r_m \parallel R_s \quad (1544)$$

This is the same expression that would be obtained by looking into the output terminal of the voltage divider in figure 893. If the source resistance R_s is much larger than r_m , the internal impedance will be approximately equal to r_m .

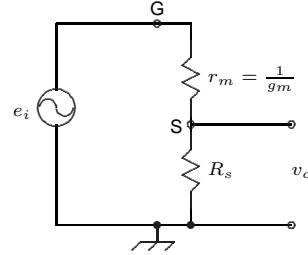


Figure 893: Simplified Model of Source Follower

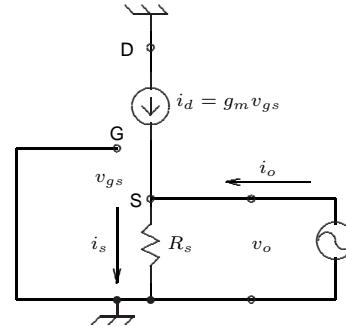


Figure 894: Output Resistance Model of Source Follower

Determining the Transconductance g_m

The MOSFET transconductance is generally available from the device datasheet. The transconductance quoted on the datasheet depends on the bias conditions. If the value is important, you can determine the transconductance from the transfer characteristic, at the selected bias point.

The transconductance is the slope of the transfer characteristic:

$$g_m = \frac{\Delta I_d}{\Delta V_{gs}} \quad (1545)$$

This is illustrated in figure 895 for the MOSFET with bias point of figure 891. Plugging in the numbers from figure 895 we obtain:

$$g_m = \frac{\Delta I_d}{\Delta V_{gs}} = \frac{16 \times 10^{-3}}{3 - 0.6} = 5.83 \times 10^{-3} \text{ amps/volt} \quad (1546)$$

$$r_m = \frac{1}{g_m} = \frac{1}{6.66 \times 10^{-3}} = 171\Omega \quad (1547)$$

Notice the large value of MOSFET transresistance compared to the BJT, or, stated alternately, the low value of transconductance of the MOSFET compared to the BJT.

Example: Voltage Gain and Internal Resistance

Now we can determine the voltage gain of the follower in figure 890. Plug $r_m = 150\Omega$ and $R_s = 2400\Omega$ into figure 893 and equation 1538:

$$A_v = \frac{R_s}{r_m + R_s} = \frac{2400}{171 + 2400} = 0.93 \text{ volts/volt} \quad (1548)$$

As predicted, the voltage gain is close to unity. We can also determine the internal resistance of the source follower, seen into the output terminals, with equation 1544:

$$r_{int} = r_m \parallel R_s = 171 \parallel 2400 = 160\Omega \quad (1549)$$

As predicted, the internal resistance is approximately equal to r_m .

Signal Handling

To this point, we've assumed that the transfer characteristic of figure 886 applies. However, for that to be correct, the drain-source voltage V_{ds} must be greater than the gate-source voltage V_{gs} ²⁷⁶.

Now, the source voltage is V_{gs} volts below the gate. The drain must be at least V_{gs} volts above the source. Consequently they cancel out and the drain terminal voltage must be more positive than the largest positive gate voltage.

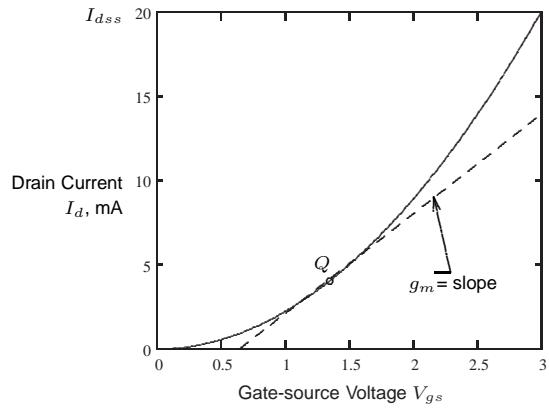


Figure 895: Graphical Determination of Transconductance g_m

²⁷⁶This assumes a zero threshold voltage V_t . If there is a significant threshold voltage, then $V_{ds} \geq V_{gs} - V_t$.

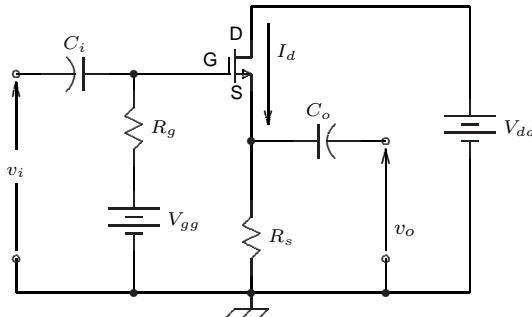
An AC input voltage to figure 891 appears across the source resistor R_s . This will cause a variation in current through R_s , and a consequent AC variation in the Q point. Changing the Q point changes the transconductance, which affects the voltage gain. Changing gain with signal level distorts the signal. Whether this level of distortion is acceptable depends on the application.

To avoid this effect, source resistor R_s would be replaced with a current sink, conducting the bias current I_d .

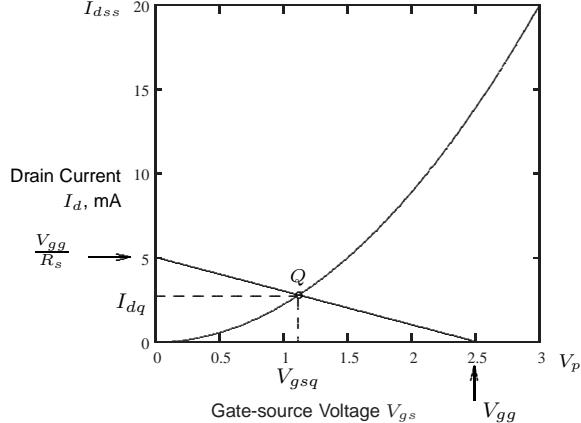
32.6 Biasing the Enhancement Mode MOSFET: AC Coupled Follower

Figure 896(a) is another follower circuit, AC coupled this time. If the input signal is AC, then the MOSFET must conduct some DC current. Then the input voltage signal modulates the DC current through the MOSFET. The variation in this DC current appears as a voltage across the source resistor R_s .

A positive gate voltage is supplied by source V_{gg} . Resistor R_g conducts no DC current (since the gate of the MOSFET and capacitor C_i are both open-circuits for DC), so the DC gate voltage is exactly equal to V_{gg} . The resistor R_g prevents input AC voltage from being shorted to ground via V_{gg} . (In practice, the gate supply V_{gg} and resistance R_g could be created by a voltage divider between the supply V_{dd} and ground.).



(a) Circuit



(b) Bias Condition

Figure 896: N Channel, Enhancement Mode, Source Follower

Making the gate positive (an amount we assume to be greater than the threshold voltage) causes the MOSFET to conduct drain current. The drain current creates a voltage across the source resistor R_s . This voltage rises until the MOSFET is at equilibrium, at some point (the Q point) on the transfer characteristic.

To develop a graphical method for designing the bias system, consider KVL around the loop including gate voltage:

$$+V_{gg} - I_g R_g - V_{gs} - I_d R_s = 0 \quad (1550)$$

But the gate current I_g is zero, so this simplifies to:

$$+V_{gg} - V_{gs} - I_d R_s = 0 \quad (1551)$$

We superimpose this on the transfer characteristic in order to show us the Q point. Equation 1551 is a straight line, with X axis intercept at $I_d = 0$ and y intercept at $V_{gs} = 0$. Plugging those values into equation 1551, we obtain a straight line that runs between a current of V_{gg}/R_s on the current axis to V_{gg} on the voltage axis.

The result is shown in figure 896(b), using the transfer characteristic for the MOSFET of figure 886, gate bias voltage V_{gg} of 2.5 volts, and source resistor R_s of 500Ω . The MOSFET Q point is at $V_{gsq} \approx 1.1V$, $I_{dq} \approx 2.6mA$.

32.7 Applying MOSFETs in Linear Applications

Many different power MOSFETs are available. By far the majority, the so-called *vertical mode* devices, are used in power switching applications such as power supplies.

Millions of small-signal MOSFETs are found in integrated circuits, digital and analog. However, it's relatively difficult to find them as discrete devices that can be purchased over the counter.²⁷⁷

The small-signal MOSFETs have two primary applications: in high-input impedance amplifiers, and as analog switches. Both these applications are readily available as low-cost integrated circuits. Integrated circuit MOSFETs can be enhanced in various ways and matched to each other, so there is little demand for the discrete devices.

For those who wish to experiment with MOSFETs in linear applications, these are the options:

- **CMOS Integrated Circuit** The CD4007 integrated circuit contains 3 independent P channel MOSFETs, 3 independent N channel MOSFETs, and a complementary pair driver. Reference [287] shows some simple amplifiers and oscillators using this part.
- **Matched MOSFET pairs** MOSFET transistor arrays of two to 4 devices, P and N channel, enhancement and depletion mode, are available from Advanced Linear Devices at modest cost. Particularly notable: these devices are matched to within a few millivolts, which makes them suitable for the differential stage front end of an amplifier.
- **Power MOSFETs** As discussed in detail below, power MOSFETs can be used in the output stage of an audio power amplifier.

32.8 Vertical and Lateral Format MOSFETs

Power MOSFETs can be constructed in *vertical* or *lateral* format. Both types of devices are enhancement mode devices, but there are some important differences.

In a vertical format MOSFET, the current flow is vertical through the chip and the drain-source channel is short. Very low drain-source resistance is obtained, in the order of 0.04 ohms. This is useful in an application that switches a large current, since it minimizes the power loss in the switch.

Vertical construction MOSFETs are primarily intended for switching applications. However, they are inexpensive and readily available so there is interest in using them the power output stage of an audio amplifier. A web search turns up many examples. The circuit shown in [288] uses the IRF532 and IRF9532 vertical format power MOSFETs.

The lateral format MOSFET is less common: it consists essentially of a small-signal MOSFET scaled to higher voltages and currents. One way of achieving this is to parallel a large number of small transistors. In this arrangement the drain-source channel is relatively long and therefore higher resistance, in the order of an ohm or so.

Lateral construction MOSFETs are more expensive, but are specifically designed for audio power amplifiers and have some advantages over vertical devices. The example circuit shown in [289] uses the Hitachi 2SK1058 and 2SJ162.

The threshold voltage of the audio devices – such as the Hitachi 2SK1056 mentioned earlier – is much lower: essentially zero.

32.9 MOSFET Power Amplifier Output Stage

²⁷⁷Of the 16 devices listed in [100] (published in 1980), only one (the VN2222L) was readily available in 2009.

This is a subject of great interest to the audio hardware community, so we provide a brief introduction and pointers to additional information.

An audio power amplifier is essentially a high power op-amp. The supply voltages are large (± 60 V is typical) and the amplifier must deliver substantial current into a 4 or 8 ohm load²⁷⁸. The audio power amplifier must be stable, have a frequency response that is flat over the audio frequency band and not distort the signal.

Figure 897 shows an audio power amplifier. This is an operational amplifier with an N-Channel and P-Channel power MOSFET output stage, where the MOSFETs act as followers with voltage gain less than unity and very large current gain. Resistors R_1 and R_2 provide negative feedback and establish the voltage gain of the amplifier. Notice the similarity to the booster amplifier stage shown in section 13.14, page 356. The transistors in figure 297(a) have been replaced by MOSFETs.

At first blush, MOSFETs appear to be particularly attractive for this application. However, with a thorough investigation, some of those advantages are not so evident.

32.9.1 Drive Current and Voltage

A MOSFET is a voltage-controlled device. In steady state, the input appears as an open circuit. This is in contrast to the power BJT, which requires considerable current to operate. Thus it appears there might be some simplification of the circuitry that drives the output transistors.

Alas, while the steady state load is an open circuit, the dynamic load (ie, for an AC signal) is a large capacitance. Figure 898 shows a MOSFET follower with its parasitic capacitances C_{gd} , C_{gs} and C_{ds} .

For example, for the IRF740 the gate-source capacitance C_{gs} is 1600 pF and the gate-drain capacitance C_{gd} is 450 pF. These are the ones that usually matter, because they affect the frequency response and the input impedance.

In section 13.1, the *Miller Effect*, page 333, we showed that the effective input capacitance is $C_i = (1 + A_v)C_f$, where A is the inverting (negative) gain and C_f is the feedback capacitance. That is, the effect of the feedback capacitance is increased by the voltage gain.

In the case of a source follower, the voltage gain is a positive number somewhat less than 1. That is, the feedback capacitance is *reduced* in effect when the amplifier is non-inverting, gain less than 1.

This makes physical sense. Suppose the non-inverting gain is unity. Then both ends of the feedback capacitance are driven by the same AC voltage. Then the net voltage across the feedback capacitor is zero, and there is no current through it. Then it has no effect and can be removed.

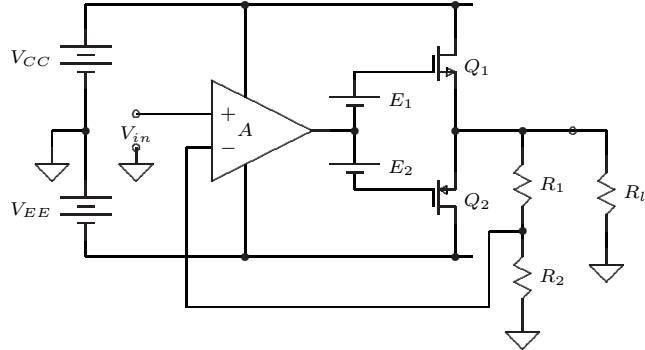


Figure 897: Audio Amp, MOSFET Outputs

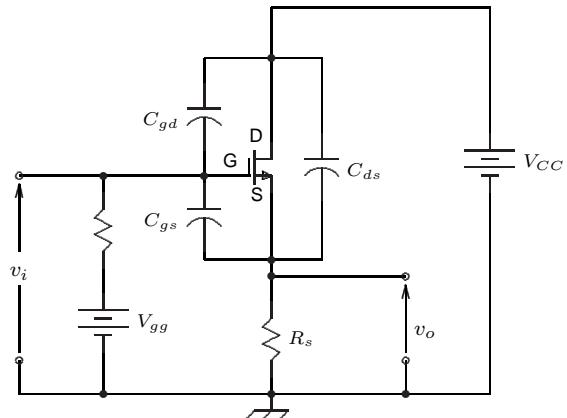


Figure 898: Source Follower with Parasitic Capacitances

²⁷⁸A loudspeaker is not a purely resistive load. It shows substantial changes in impedance over the audio band. This complicates the stability issue.

For a non-inverting gain somewhat less than 1, there is some current through the feedback capacitance. Its effective value at the input of the amplifier is *less* than the original feedback capacitance.

We can use the same Miller effect formula, but the voltage gain A_v changes sign. In this case, $C_i = (1 - A_v)C_f$, where A_v might be something like 0.8.

For example, a source follower using the IRF740 has an effective input capacitance according to the following:

$$\begin{aligned} C_i &= C_{gd} + (1 - A_v)C_{gs} \\ &= 450 + (1 - 0.8)1600 \\ &= 770 \text{ pF} \end{aligned}$$

At 100kHz, this appears as an impedance of about $2\text{k}\Omega$. A peak drive of 50 volts must supply a peak current of 25mA. This is not outrageous, but neither is it zero. The drive current increases if the output transistors are connected in parallel to increase output power.

32.9.2 Thermal Stability and Crossover Distortion

The power BJT suffers from the phenomenon of *second breakdown*. At high current levels, a crowding effect occurs. In effect, the voltage drop decreases across the junction as the junction temperature increases. This tends to concentrate current into a small area of the chip, which causes a further voltage drop in that area. The result is a hot-spot on the chip which fails. In effect, the BJT has a negative coefficient of resistance: the resistance decreases as the temperature increases.

In contrast, power MOSFETs have a positive coefficient of resistance. Consequently, an area of the chip that increases in temperature has a decrease in current, which reduces the current in that area. As a consequence, MOSFETs are less likely to experience second-breakdown. However, MOSFETs have a different problem, the change of threshold voltage with temperature. To understand that, we refer again to figure 897.

Without the voltage bias sources E_1 and E_2 , there is a large *dead zone* in the output. When the output of the op-amp is less than the threshold voltage of a transistor, then the output is zero. If MOSFETs have a threshold voltage of 3 volts each, then the dead zone occurs when the op-amp output voltage is between ± 3 volts. The action of negative feedback forces the op-amp to switch very quickly through this deadzone. For a low output frequency (such that the slew rate of the op-amp moves the output quickly between -3 to $+3$ volts for example) and an application where distortion is not important (eg, driving an AC motor) this may not matter. For an audio power amplifier application, the dead zone creates so-called *crossover distortion*.

The bias voltage E_1 and E_2 establishes a quiescent current in the MOSFETs. To minimize crossover distortion, the bias voltage is set to such a value that the conduction of the two MOSFETs overlaps in the crossover region – over a range of output voltage both MOSFETs are conducting. The threshold voltage of power MOSFETs changes with temperature, in the order of $-1\text{mV}/^\circ\text{C}$ [290]. As a consequence, when the MOSFETs heat up (as they will in a power amplifier application), then the quiescent current changes. At a minimum, this changes the bias condition of the transistors and may affect the distortion. This effect is less evident in lateral format MOSFETs, which have a zero threshold voltage. If this effect is important for the selected MOSFETs and power level, then additional circuitry is required to stabilize the bias current against temperature change.

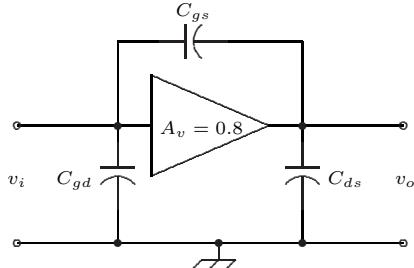


Figure 899: Source Follower Equivalent Circuit

Notice that the change in threshold voltage is not an issue when the MOSFET is used as a switch. A small change in threshold voltage is unimportant when, as is the usual case, the input voltage is much larger than the threshold voltage.

32.9.3 Output Voltage Swing

Compare the BJT power amplifier of figure 297(a) on page 357 with the MOSFET amplifier of figure 897 on page 1005. The bias voltage in the BJT amplifier, provided by a forward-biased diode, is about 0.6 volts. The bias voltage in the MOSFET amplifier is equal to the threshold voltage plus whatever voltage is required to create the bias current. The threshold voltage (for vertical format MOSFETs) is in the order of volts. This voltage subtracts from the power supply voltage to obtain the available peak voltage swing at the input of the MOSFETs. Consequently, for a given power supply voltage a BJT amplifier will have a larger available output voltage swing.

32.9.4 Bandwidth

In general, power MOSFETs have much wider bandwidth than power BJTs. As a consequence, power MOSFETs are used almost exclusively in switching power supplies where high switching frequency allows correspondingly smaller inductors and capacitors for energy storage.

In power amplifiers, the advantages are not so clear. If the amplifier must drive a piezoelectric transducer at several hundred kHz, then a MOSFET output stage is the obvious choice. An audio power amplifier need only function without distortion at a maximum of 20kHz, so the choice of output device – MOSFET or BJT – is more debatable. Furthermore, if the BJT output stage generates less inherent distortion [243], then that may be more of a consideration than bandwidth.

On the one hand, a negative feedback amplifier with a fast driver stage and slow output stage is likely to oscillate [225]. The driver stage overshoots while waiting for the output stage to respond. This may be a problem with relatively slow BJTs in an output stage. It's not likely to be a problem with power MOSFETs.

On the other hand, the wide bandwidth of a MOSFET can lead to parasitic oscillations at a radio frequency. From Douglas Self in [243]: *The extended FET frequency response ... means that rigorous care must be taken to avoid parasitic oscillation, as this is promptly followed by an explosion of disconcerting violence.* To make this even more exciting, loudspeakers present wide variations in impedance over the audio frequency range, and those variations can trigger instability.

Evidently a MOSFET-based audio amplifier design must be constructed with parasitic components in mind. It would be prudent to test with a current-limited power supply and a wide bandwidth oscilloscope.

32.10 Exercises

1. There are many audio power amplifier designs that use power MOSFETs in the output stage.
 - (a) Do a web search to identify a power amplifier that uses vertical format MOSFETs. Identify the threshold bias mechanism. Does it compensate for temperature? Explain.
 - (b) Find the circuit of a single-supply audio power amplifier. Redraw figure 897 to show operation from a single power supply voltage. Hints: you will need an input bias network, input and output capacitors, and a capacitor in the feedback network. What effect do these components have on the frequency response of this amplifier?
 - (c) Based on what you now know of power BJTs and power MOSFETs, which do you think is more suitable for the output amplifier stage of an audio amplifier? Justify your answer.
2. Suppose that a current sink is used to replace the source resistance in figure 890. (A suitable mirror circuit is in figure 809, page 919).

- (a) What effect would this have on the voltage gain of the follower?
- (b) What effect would it have on the output resistance?
- (c) What effect would it have on distortion? Explain.

33 CMOS Analog Switch and Applications

An *analog switch* is an integrated circuit that can enable and disable the flow of current. The current can flow in either direction through the switch, and the voltage at the switch terminals can vary over some range. The analog switch can be viewed as a solid-state version of the relay (section 2.31 on page 108). The schematic symbols for the two devices are shown in figure 900. Their function is similar but there are some important differences.

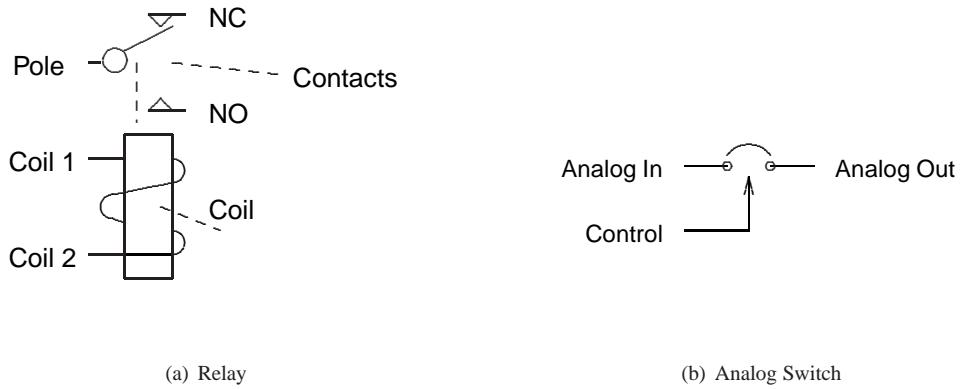


Figure 900: Relay and Analog Switch

- **Isolation** In a relay, the switch is electrically independent from the control circuit (the *coil* or *armature*). In an analog switch, the circuits must share a common power supply terminal.
- **Current and Voltage Capability** The switch contacts of a relay can carry much larger currents than an integrated-circuit analog switch. Relay switch currents range from one ampere up to hundreds of amps. An IC analog switch typically conducts a few tens of milliamperes²⁷⁹.
- Similarly, the switch contacts of a relay can withstand much larger voltages than an integrated-circuit analog switch.
- **Switching Speed** The switching speed of an electro-mechanical relay is quite slow: tens or hundreds of milliseconds. An analog switch can open or close in tens of nanoseconds, so the switching frequency can be in the megahertz range.
- **Switch Resistance** The resistance of a relay switch is extremely low: in the milliohms. An analog switch has much larger resistance and this resistance varies with current. This variation of resistance is a potential source of signal distortion, although circuits can be designed to minimize the effect.
- **Switch Configuration** Like the relay, the analog switch is available in various configurations such as SPST (single pole, single throw), SPDT (single pole, double throw), DPTD (double pole, double throw) and so on.

The analog switch is also available as a *many to one* or *multiplexer*. Since a signal can flow in either direction through the switch, it can also function as a *one to many* or *demultiplexer* (sometimes also known as a *distributor*). The mechanical analog of a multiplexer is a rotary switch. For example, a 5 position rotary

²⁷⁹There is an electronic version of the relay known as a *solid state switch* or *SSR*. The SSR control circuit is optically isolated from the switch section and the switch section can carry currents of many amperes. SSR's are relatively large and expensive. They are used in industrial power control applications.

switch is described as a SP5T switch. Equivalent devices (SP8T, DP4T) are available as integrated circuit analog switches. Figure 901 shows a SP4T rotary switch or multiplexer.

The position of a rotary switch is determined by rotating a mechanical shaft. The position of an analog switch multiplexer is determined by an electronic binary logic code. For example, the position of the switch in figure 901 is specified by a two bit binary code.

- **Power Consumption** A relay coil consumes many orders of magnitude more electrical power than an analog switch. The power required to actuate an analog switch is essentially zero.

However, the analog switch requires power supplies. If the signal swings above and below ground, then there must be a negative power supply. The supply voltages must exceed the largest signal through the switch.

- **Size, weight and cost** Even the smallest relays are large and expensive compared to an integrated circuit.
- **Reliability** The switch contacts of a relay are subject to wear and failure from contaminants, and have a finite life. The integrated-circuit analog switch shows no such wear-and-tear: its life is essentially infinite.

Wherever possible, switching applications that use relays are being replaced by integrated circuit analog switches (or other solid-state devices.) The result is an electronic design that is smaller, lighter, less expensive and more reliable.

However, there are some applications where the currents and voltages are large or the switch must be electrically isolated from the control circuit. In those cases, a relay may still be necessary²⁸⁰.

33.1 Applications

In this section, we'll look at some typical applications of the analog switch.

33.1.1 Mixer Input Select

Figure 902 shows a typical application for an analog switch. When the three analog switches are closed, the three input signals (which could be audio signals) are added together²⁸¹ and appear at the output of the op-amp (section 13.2 on page 336).

Any one of these inputs can be disabled by opening the corresponding switch. In a mixing console, this circuit could be the method of selecting various tracks to be combined in the final recording.

The 74HC4316 analog switch is one possible device for this application. The switch would be powered by +/-5 volt supplies. The control signals would be 0 volts for OFF and +5 volts for ON.

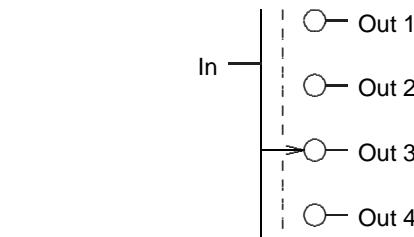


Figure 901: Rotary Switch/Multiplexer. The arrow moves vertically to connect the input to one of the 4 outputs.

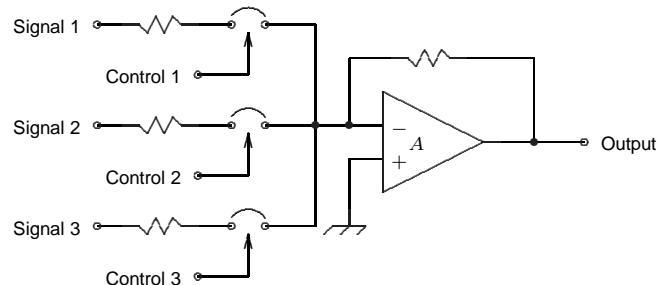


Figure 902: Switchable Mixer

²⁸⁰The optically isolated JFET can provide low voltage and current switching with isolation: section 31.3.1 on page 983

²⁸¹Confusingly, an audio circuit that combines signals is referred to as a *mixer*. In RF and signal processing, a device which multiplies signals is also known as a mixer. So it is essential to know the context of the word.

According to the data sheet for the 74HC4316, the ON resistance of a switch is in the order of 100Ω and varies slightly with current. This variation of resistance with current has the potential of distorting an audio signal. However, if the fixed input resistors are much larger ($10k\Omega$, for example) then the switch resistance is effectively swamped by the fixed resistance, and a variation in switch resistance will have negligible effect on the gain and on distortion.

33.1.2 Switched Gain Amplifier

It is a common requirement in signal processing applications to be able to change the gain of an amplifier somewhere in the signal processing chain. Then the amplifier can match the signal amplitude to, say, the input of an A/D converter. For small signals, select the high gain position. For large signals, which would overload a high gain amplifier, select the low gain.

The circuit of figure 903 shows a simple example. When the switch is closed, resistor R_2 is grounded and the circuit is a straightforward non-inverting amplifier, section 12.4 on page 319. The gain is

$$\frac{v_o}{v_i} = 1 + \frac{R_1}{R_2}$$

If the switch is opened, then (because there is no current through R_1 and therefore no voltage drop across it) the circuit has 100% feedback. The feedback signal to the inverting output is exactly equal to the output signal. Then the circuit becomes a unity-gain follower. Alternatively, when the switch is open the value of R_2 in the previous expression becomes infinite, $R_1/R_2 = 0$ and the gain becomes unity.

There is, however, a secondary effect at work here, one that appears at high frequencies. The analog switch has a certain amount of capacitance between its input terminal and ground: 35pF for the 74HC4316. At high frequencies, this capacitance provides a path to ground even when the switch is open. As a result, at a high enough frequency that the reactance of this capacitance becomes comparable to the resistance R_2 , the gain of the circuit increases. This high frequency peaking will show up as overshoot on the transitions of a square wave, which is probably undesirable.

Many other configurations are possible for a switched gain amplifier. If the amplifier must operate at high frequency, the circuit model should be checked in a simulator for the effect of the analog switch capacitance.

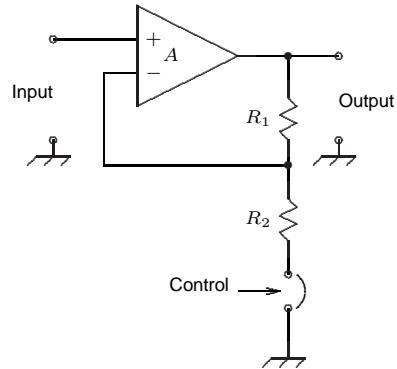


Figure 903: Switched-Gain Amplifier

33.1.3 Logic Circuits

Here we illustrate digital logic circuits²⁸² constructed with analog switches. This is sometimes useful for a simple logic requirement.

Figure 904(a) shows a positive logic AND gate. If we take a logical 1 as a positive voltage and a logical 0 as zero volts, then the output from the gate is a logic 1 when control inputs A AND B are both logic 1.

Similarly, figure 904(b) shows a positive logic OR gate. The output from the gate is a logic 1 when control either input A or input B are logic 1.

A *complete* logic family (one that can implement *any* logic function) requires an inverter: a device that can converts a logic 0 at the input to a logic 1 at the output, and vice versa. Figure 904(c) shows a simple inverter. When the input control signal is a logic 0, the switch is open and resistor R pulls up the output terminal to the supply voltage, a logic 1. When the control signal is a logic 1, the switch is closed and that shorts the output terminal to ground. This circuit has the virtue of simplicity, but it has two shortcomings. First, current flows through resistor R when the output is in the logic 0 state: this represents wasted power. Second, when the load has a capacitive component C , the output rise time will be determined by the time constant RC . It's usually not practical to reduce the capacitance, so the rise time (and the speed of the logic) can only be increased by reducing the value of R , which increases wasted power.

Figure 904(d) shows an improved inverter. The output circuit now consists of two switches in a *totem-pole* configuration. When the output is a logic 1, the upper switch closes and the lower switch opens. When the output is a logic 0, the upper switch opens and the lower switch closes. Ideally, they are never *both* closed: that would place a short circuit between the power supply and ground. With this arrangement, no waste current flows in the totem-pole transistors. Furthermore, a load capacitance can be charged and discharged rapidly through the low-resistance path of the analog switches.

The lower switch is driven by the input signal, the upper switch is driven by the simple inverter of figure 904(c). Doesn't the simple inverter waste power and slow down the circuit? It need not, because it is only driving the control input of one analog gate. Consequently the load capacitance is small so resistor R can be relatively large without a speed penalty. A large resistor minimizes wasted power.

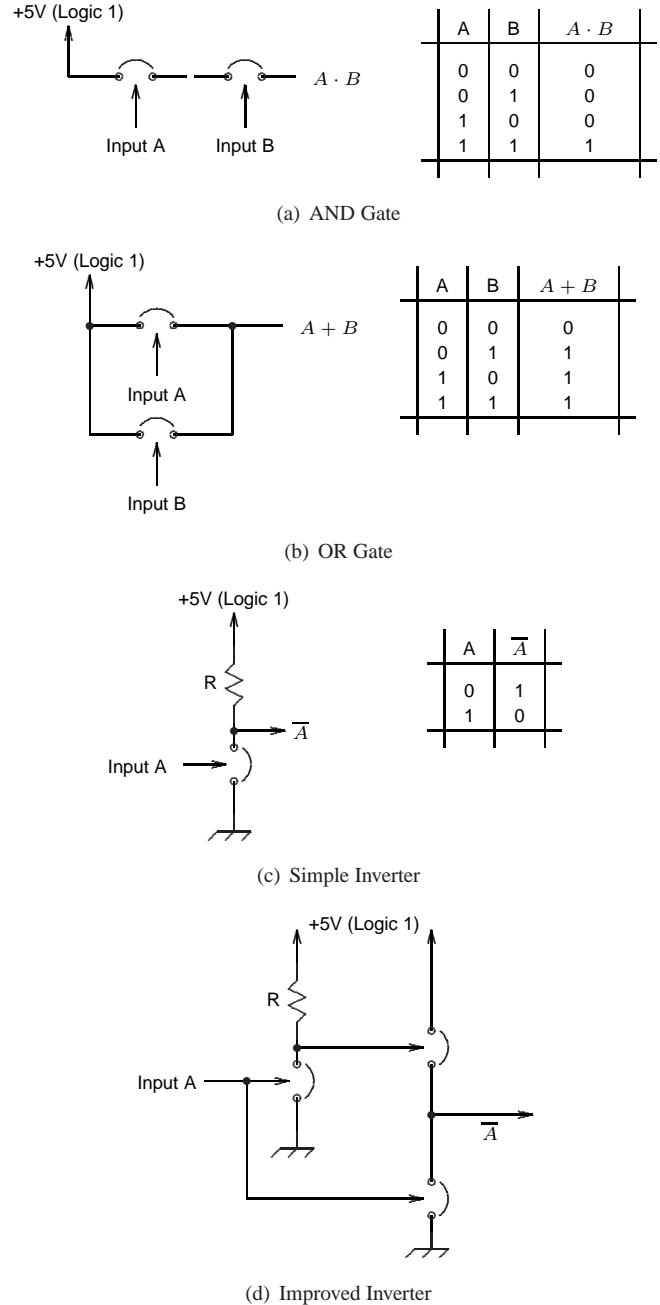


Figure 904: Analog Switch Logic

²⁸²This section assumes some familiarity with digital logic. Any introductory level text in Digital Logic will provide the necessary background.

34 Interference

Consider the following case histories:

- A certain car engine control system quits while close to a radio transmission tower.
- A development engineer finds that her signal acquisition system is swamped in noise from the on-board microprocessor.
- An airborne computer control system works fine on the bench when the modules are spread out. When the modules are installed in the case, noise from the switching power supply appears on the video output.
- An industrial automation control system is being produced as a commercial product. When tested for compliance with emission regulations, it fails the test. The unit must be redesigned to incorporate filters on the connecting cables.
- An aerial photography camera is installed in an aircraft, powered from the 28VDC bus. Whenever the camera triggers, the aircraft navigation computer resets.
- A film projection system is equipped with a magnetic head to read the sound track on magnetic tape. A nearby motor induces a power line interference signal in the magnetic head. The noise signal is stronger than the desired signal.

These are examples of *interference* – an electronic signal is somehow corrupted by another electrical signal. The interfering signal may be known as *EMI: ElectroMagnetic Interference*. We say that the *signal integrity* is compromised by an interfering signal. The interference may show up as a degradation of some signal-noise ratio, such as background television sync buzz on an audio signal. Or it may be severe enough to cause the system to fail.

As well, the allowable level of emissions from an electronic device are regulated, and must meet the requirements of *EMC: ElectroMagnetic Compliance*.

In other words, it may be necessary to protect an electronic circuit from some external interference signal, or it may be necessary to reduce the interfering signal that the electronic circuit is producing.

In all of the cases cited above, some signal interferes with the proper operation of the device. The precise mechanism is different in each case. However, it is possible to approach each problem with the same general understanding. Simple tests and measurements identify the specific interference mechanism and then specific remedies can be applied to the design. In some cases, there may be more than one mechanism of interference at work, and so it is important to be systematic.

34.1 Scope of the Problem

Interference problems are growing more common and severe, for the following reasons:

- Devices progress to smaller sizes, so different sections of the system are in closer proximity. For example, in a cellular telephone the RF transmitter is in close proximity to a microprocessor. Their signals must be kept separate.
- In the days of large equipment and vacuum tubes, interference problems could be solved by using shielding²⁸³ in the form of enclosed metal boxes to separate the interfering circuits. This approach is less useful because of modern demands in size, weight and cost. Shielding, if it is used, must be part of the printed circuit board. The outer case is likely to be plastic, and so is no help as a shield.

²⁸³In this context, a *shield* is a physical barrier to some interfering signal. A shield is a metal plate or enclosure that isolates two electric, magnetic or electromagnetic areas in space.

- Many circuits include digital and analog circuitry in the same unit. For example, a data acquisition system will typically include an analog front end that conditions a sensor signal and presents it to an A/D converter. The output from the A/D converter is directed to a microprocessor for further processing. The digital signals of the microprocessor must be kept separate from low-level signals in the analog front end.
- Electronic devices are ubiquitous, and so there are more opportunities for interference between different products.
- The clock speeds of digital circuits are extremely high – in the hundreds of MHz. Digital waveforms at these frequencies trigger parasitic component effects, such as stray inductance. Short connecting leads and small patches of a circuit board may become antennas.

Curing these problems can be a challenge, because:

1. **Very small amounts of coupling between a source and receiver of interference may introduce unacceptable levels of interference.** For example, an audio system may require a signal-noise ratio of 80db which is a factor of 10,000 volts/volt. If the reference level is 1 volt, then interfering signals must be kept below $100\mu\text{V}$. In a metropolitan area, a short length of wire will pick up radio and television signals that exceed this by a factor of 1000. Consequently, the defences must attenuate these undesired signals by the same factor.

Another example: depending on the frequency, an unbalanced current of only $5\mu\text{A}$ in a cable may be sufficient to cause unacceptable radiated electromagnetic interference [291].

2. **Schematic symbols of components do not accurately reflect their true behaviour.** Schematic symbols have a dual meaning: they represent the ideal component and simultaneously represent the real device. For example, an ideal voltage source maintains its output voltage regardless of the output current while a real voltage source has internal resistance. When the output current of the voltage source is small, the voltage drop across the internal resistance can be ignored. However, that same internal resistance can cause coupling between a source and receiver of interference.

Similarly, capacitors have internal inductance and resistance. These unmarked, extra behaviours are modelled by *parasitic* components: resistors and inductors that appear in series and parallel with an idealized capacitor.

In the interest of simplification, we can often ignore these parasitics. However, when interference is involved – and particularly at high frequency where inductive reactance due to wire length and stray capacitance becomes significant – the parasitics must be taken into account.

3. **A schematic diagram does not reflect the wiring layout.** Many interference mechanisms are a consequence of the routing of wiring and the placement of components. It is necessary to combine an understanding of the schematic and the wiring layout to identify the important parasitic components and coupling mechanisms.

To troubleshoot interference and EMI problems, the engineer must understand the schematic as it is drawn, and must also understand the effect of the parasitic components that are not shown but may have an important effect. The schematic with its parasitic components has been termed *the hidden schematic* [292].

34.2 The Mechanism of Interference

A generalized view of interference is shown in figure 905.

There is a *source* for the interfering signal. The signal is conveyed by some *medium* of transmission to the *receiver* circuit²⁸⁴.

For example, in the car engine control problem described on page 1013:

- The source of interference is a strong radio transmitter.
- The medium of transmission is an airborne radio signal which couples to car wiring. The car wiring – or possibly the metal structure of the car – behaves as a receiving antenna.
- The receiver is the engine control system electronics.

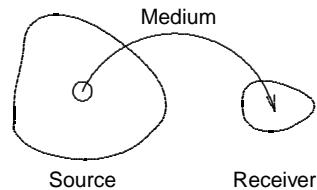


Figure 905: Interference in General

In this generalized model, the interference may be minimized by **eliminating the source, blocking the medium, or by making the receiver resistant to the interference**. In difficult cases, it may be necessary to use some combination or all three.

In this particular example, it is not feasible to eliminate the source, so the focus must be on blocking the medium and hardening the receiver.

34.3 The Source

In dealing with a problem of interference, it is critical to identify the components in figure 905.

The identification of the source is usually based on the concept of *correlation*. For example, every time the soldering iron switches the radio produces a popping noise. These two events are then said to be correlated. Correlation does not necessarily indicate causality, but it's an important clue to investigate further. For example, if the soldering iron is moved away from the radio and the popping noise decreases in severity, that is a second correlation that points at the same suspect.

As another example, suppose a switching power supply is suspected as being the source of interference in an analog amplifier. Synchronize an oscilloscope on the power supply switching waveform and see if the amplifier noise display is stable. If that's the case, then the two signals are correlated. If there is still some question, operate the circuit from a separate lab power supply and see if the problem persists.

34.4 The Coupling Mechanism

Having identified the source, the next step is to determine the coupling mechanism or *medium of transmission*. The usual suspects for coupling are:

- *conductive* (via a power supply or ground path, for example)
- *electric-field*
- *magnetic-field*
- *electro-magnetic field* (radio wave)

Now we'll look at the details of each of these mechanisms.

²⁸⁴Some authors refer to the source and the receiver as the *aggressor* and *victim* circuits.

34.5 Conductive Coupling

In this case, the source and the receiver share a common impedance. An example audio circuit is shown in figure 906. A preamplifier stage is represented by amplifier A . A power amplifier stage is represented by the switch and load resistance R_L .

In figure 906(a), both the amplifier and preamplifier are connected back to the power supply E by the same wiring. The wiring has resistance, represented by R_{1A} through R_{2B} .

The load resistance R_L is such that significant current flows through it when the switch is closed. This current is supplied via resistor R_{1A} and returned via resistor R_{1B} . As a consequence, the voltage V_{xy} changes when the amplifier draws more or less current.

This change in voltage is communicated to the preamplifier supply terminals. If the preamp has insufficient power supply rejection ratio (section 21.4), then this variation in voltage will find its way into the output signal of the preamp.

A particularly insidious result of this arrangement is the possibility of low-frequency oscillation, called *motorboating*. The preamplifier, power amplifier, and common supply impedance form a loop. The coupling and bypass capacitances provide various phase shifts which result in positive feedback at some low frequency, and the entire arrangement becomes unstable. In this case, we could view the interfering signal as being generated at the power amplifier and coupled by the common power supply impedance, with the preamplifier as the receiver of the interference. Additional capacitance across the supply at xy will not cure this problem – it simply moves the oscillation to a lower frequency.

Identification of this coupling mechanism may simply be a matter of recognizing that the same signal appears at different points in the circuit, for example on the power supply line. (The measurement technique must be such as to isolate conductive effects from other effects.)

In this particular case, the cure is shown in figure 906(b). The preamplifier power supply lines are moved from the power amplifier connection to the power supply, where the source impedance is low. This eliminates power supply fluctuations caused by the interaction of the power amplifier current and power supply wiring. The power supply resistance wiring of the preamp R_{2A} , R_{2B} has little effect because it is not carrying significant current. In effect, this change blocks the interference coupling mechanism.

Alternatively, the preamplifier could be equipped with a power supply regulator, thereby making it more resistant to variations in power supply voltage.

In this particular example, the coupling impedance is a resistance. It is quite common for the coupling impedance to be an inductance, in which case coupling occurs at a frequency where the impedance is significant. The previous example is for an audio circuit, but the concept applies equally well to circuits which contain digital switching circuitry and sensitive analog circuitry.

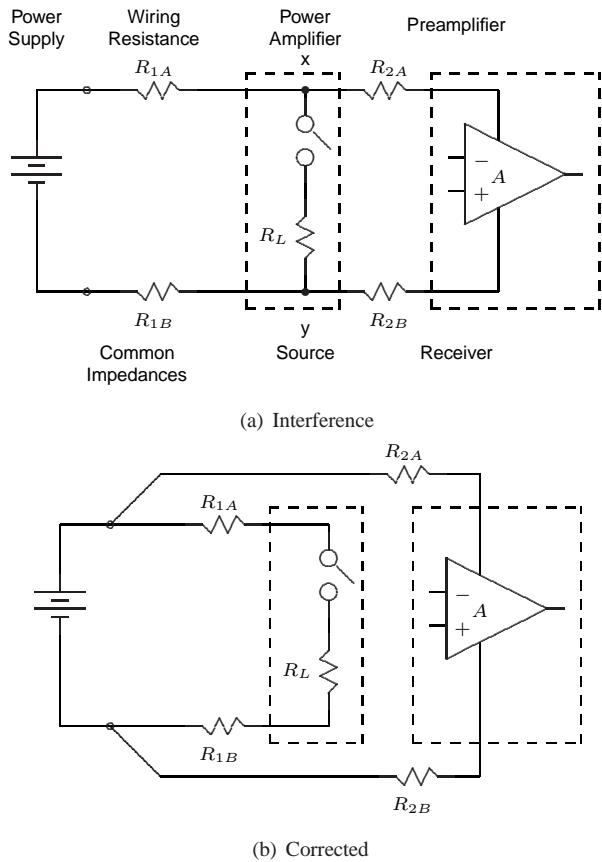


Figure 906: Example, Common Impedance

Reduce conductive coupling between circuits by providing separate power supply and ground conductors back to the power supply. In the case of a mixed digital-analog circuit, provide separate analog and digital supply and ground lines. The analog and digital supply meet at the power supply output. The analog and digital ground meet at the power supply common point.

For this to be effective, there must be no other connection points between the analog and digital systems.

34.5.1 Implications for PC Board Layout

When laying out a circuit board, if the analog power line and digital power line connect together at the power supply, then a PC board autorouter program will think that they're the same net. It will then mindlessly intersperse digital and analog power supply connections, exactly what is not wanted. The analog supply line and digital supply line must be kept separated and given distinct names, such as +5A, +5D. The autorouter then brings the leads back to the power supply area, and a removable jumper wire connects them together. Similarly, the grounds need to be separated and given distinct names, such as AGND, DGND, and brought back to jumpers²⁸⁵ (figure 907).

This particular strategy is suitable for circuits where the digital signal edge rates are modest and electromagnetic radiation is not an issue. However, it's not a cure-all. A different strategy, using a ground plane, may be more effective in other situations.

34.6 Electric Field Coupling

A time-varying voltage creates a time-varying electric field that will couple into a receiver circuit via any stray capacitance, as indicated in figure 908(a). The degree of coupling increases with the capacitance, which is a function of the surface areas of the source and receiver and the spacing between them. For example, when the two plates of the capacitor have area A and spacing S , then the capacitance is

$$C = \frac{\epsilon A}{S} \quad (1552)$$

where ϵ is a constant, the permittivity of free space. Then the capacitance and the coupling can be reduced by reducing the areas of the two plates and by increasing the spacing between them.

The coupling circuit is effectively a highpass RC circuit, for which the Bode plot is given in figure 908(b). As shown in this figure, the coupling from the interfering source e into the receiver circuit is reduced by increasing the highpass corner frequency. In the figure, the original corner frequency is ω_1 and the signal at frequency ω is attenuated to point A . If the corner frequency is moved up to ω_2 , then the attenuation is moved down to point B , that is, there is less coupling between the source and receiver circuit. The corner frequency is given by

$$\omega_c = \frac{1}{R_{\text{circuit}} C_{\text{stray}}} \quad (1553)$$

so the corner frequency may be increased by decreasing either the capacitance or resistance. That is, **low impedance circuits are less susceptible to electric-field interference.**

²⁸⁵This has other advantages: at a later date, you can separate analog and digital power in the event that they seem to be interfering and must be mutually filtered. The jumper can be used as a test point for board power supply voltage.

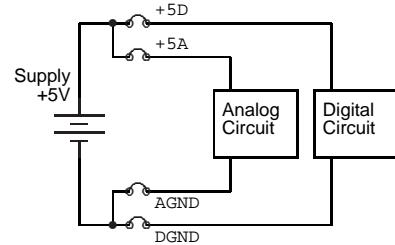


Figure 907: Common Point Grounding

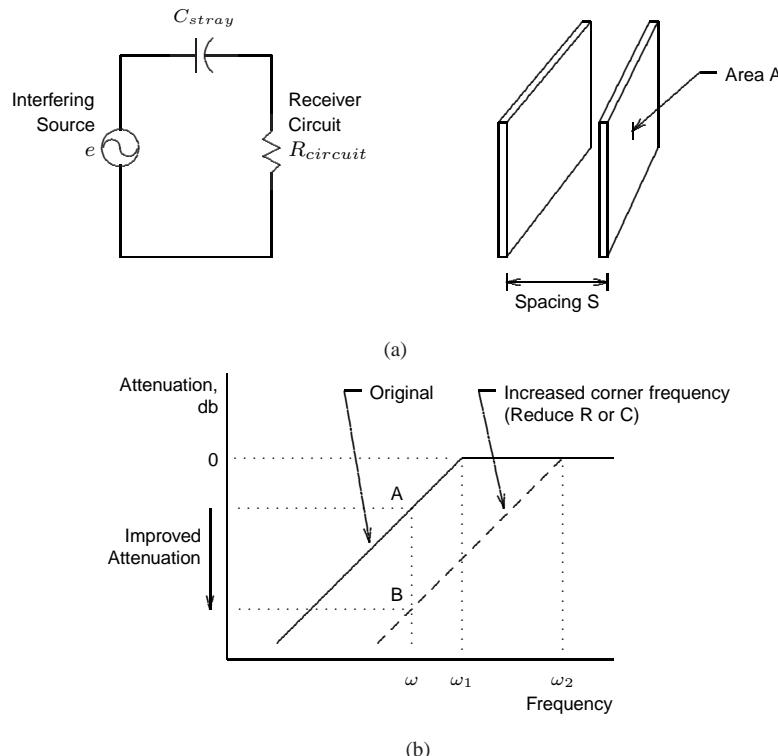


Figure 908: Electric Field Coupling

34.6.1 Electric-Field Detection

Electric-field detection is straightforward. The probe is a length of wire, or a flat plate that is coupled to the display device – oscilloscope or spectrum analyser. The source and the probe form the two terminals of the coupling capacitance. The strength of the detected signal depends on the spacing, but not the orientation, of the probe.

34.6.2 Electric-Field Shielding

If as shown in figure 909 a metallic plate is interposed between the two plates of a capacitor and this plate is connected to the circuit common, then the electric field stops at the plate. In effect, the shield becomes the second plate of the capacitor and coupled current is short-circuited to ground.

Moreover, an electric field cannot penetrate inside a metallic enclosure. Consequently, an electric-field shield is easily constructed – a thin metal foil is sufficient.

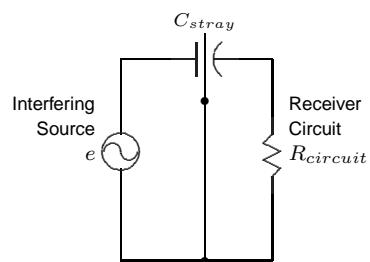


Figure 909: Electric Field Shield

34.6.3 Demonstration: Electric Field Pickup

In this case, as shown in figure 910(a), the function generator drives a conductive aluminum plate, approximately 19cm by 19cm. The electric field probe is a stiff piece of wire, about 5cm total length, spaced about 0.5 cm above the aluminum plate. The probe plugs into the end of a BNC cable. The cable leads to an oscilloscope with an input resistance-capacitance of $1M\Omega$ in parallel with $47pF$. The generator output is 8 volts p-p sine wave at a frequency of 1MHz.

The measured signal on the oscilloscope is 100mVp-p. The plate and probe are two electrodes in a 'stray' capacitance. The signal couples through this stray capacitance to the oscilloscope.

We'll assume that the 50Ω internal resistance of the signal generator is much smaller than the impedance of C_{stray} , so it can be ignored. The $1M\Omega$ input resistance of the oscilloscope is much larger than the impedance of C_{in} of the oscilloscope, so it too may be ignored. Then the equivalent circuit is as shown in figure 910(b). The stray capacitance and the input capacitance of the oscilloscope form a capacitive voltage divider.

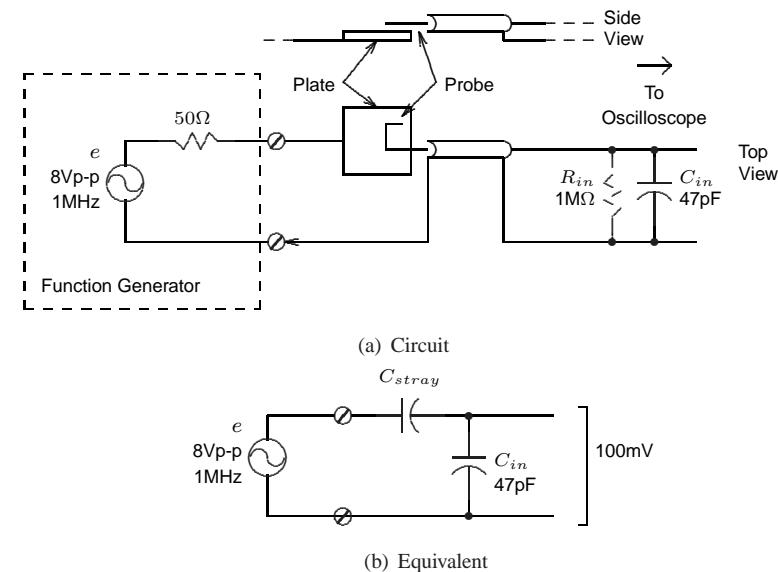


Figure 910: Electric Field Pickup

$$\frac{e_o}{e} = \frac{Xc_{in}}{Xc_{stray} + Xc_{in}} \quad (1554)$$

$$= \frac{1/\omega C_{in}}{1/\omega C_{stray} + 1/\omega C_{in}} \quad (1555)$$

$$= \frac{1}{C_{in}} \frac{C_{in} C_{stray}}{C_{in} + C_{stray}} \quad (1556)$$

Since C_{in} is much greater than C_{stray} , we can simplify this to

$$\frac{e_o}{e} \approx \frac{C_{stray}}{C_{in}} \quad (1557)$$

Plugging values for e , e_o and C_{in} into equation 1557 puts the stray capacitance C_{stray} at $0.59pF$.

Notice that there is no frequency term in equation 1557, so this divider affects all frequencies equally. A complex waveform such as a square wave will be undistorted passing through this divider. That is confirmed by measurement: when the generator is switched to square and triangle waveforms, the oscilloscope shows the same waveform shape as the generator.

34.6.4 Demonstration: Reducing Electric Field Pickup by Lowering Impedance

If the oscilloscope input is shunted by a 50Ω resistor, the effect is dramatic: the detected signal drops from 100mV to about 3mV, an attenuation of

$$\begin{aligned} 20 \log_{10} \frac{V_1}{V_2} &= 20 \log_{10} \frac{100}{3} \\ &= 30.4 \text{ db} \end{aligned}$$

Furthermore, the detected waveform is a differentiated version of the signal-generator waveform. A triangle signal becomes a square wave, a square wave becomes alternating positive and negative spikes.

The new equivalent circuit is shown in figure 911. The input capacitance of the oscilloscope is 47pF, which has an impedance of about $3.3\text{k}\Omega$ at 1MHz. Consequently, it's large enough to be ignored compared to 50Ω . This circuit is an RC highpass network, with a corner frequency around 6GHz. Frequencies below this will be attenuated by a greater or lesser extent. For waveforms well below the corner frequency, the network will act as a differentiator.

34.6.5 Demonstration: Reducing Electric Field Pickup by Shielding

If a grounded metallic sheet is placed between the plate and probe in figure 910, then we obtain the equivalent circuit of figure 909. The electric-field coupling into the probe should drop significantly. For a flat shield approximately the same size as the plate in figure 910, the signal dropped by a factor of 16, some 24db. Enclosing the probe in a sealed metal container would undoubtably improve the shielding effect.

Electric field pickup may be reduced by lowering the impedance of the receiver circuit or placing a grounded metallic shield between the source and the receiver.

34.7 Magnetic Field Coupling

When a wiring *loop* is driven by an alternating current, it creates a time-varying magnetic field. If that AC magnetic field passes through a second loop, a time-varying voltage will be induced in the second loop. For example, the magnetic field might originate in the filter inductor of a switching power supply and be received in a circuit loop that is nearby wiring.

A little theoretical investigation will give us some insight into this effect.

The situation is shown in figure 912. A transmitter coil (shown as one turn here, which could be multiple turns) is driven by an AC source and carries an AC current i .

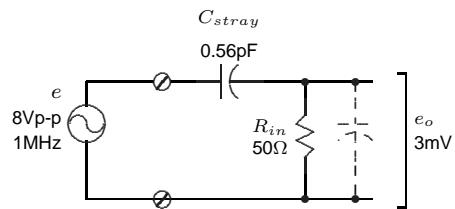


Figure 911: Equivalent Circuit

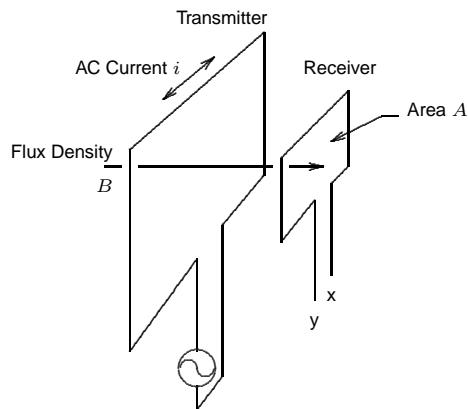


Figure 912: Magnetic Coupling

This current generates a magnetic field:

$$B = \mu_o H \quad (1558)$$

where the variables are:

- B Flux Density, webers/metre², from the transmitting coil
- μ_o magnetic permeability of space, a constant
- H magnetic field intensity, amp-turns.

As well, we can write:

$$H = Ni(t) \quad (1559)$$

where

- N is the number of turns on the coil, (in this case, $N=1$)
- $i(t)$ is the current in the coil

Putting $N = 1$ and combining these two equations, we have:

$$B = \mu_o i(t) \quad (1560)$$

as the time-varying magnetic field intensity in the region of the transmitting coil.

This same field passes through the smaller receiving coil. According to Lenz's law:

$$e_{xy} = N \frac{d\phi}{dt} \quad (1561)$$

where

- e_{xy} is the voltage created at the terminals of the coil
- ϕ is the total flux flowing through the receiver coil
- N is the number of turns on the receiver coil (again, one turn)

The flux in the receiver coil is equal to the transmitter magnetic field times the receiver coil area:

$$\phi = BA \quad (1562)$$

Substituting for B from equation 1560 into equation 1562, and then for ϕ into equation 1561, we have that the receiver terminal voltage is:

$$\begin{aligned} e_{xy} &= A \frac{d}{dt} \mu_o i(t) \\ &= A \mu_o \frac{di(t)}{dt} \end{aligned} \quad (1563)$$

For example, if $i(t)$ is a sine wave of radian frequency ω and peak value I :

$$\begin{aligned} e_{xy} &= A \mu_o \frac{dI \sin \omega t}{dt} \\ &= A \mu_o \omega (I \cos \omega t) \end{aligned} \quad (1564)$$

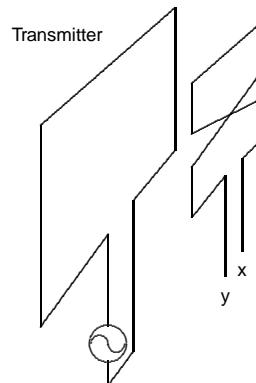


Figure 913: Coil Cancellation

Most real-world configurations are much more complicated and so numerical results and predictions are not as simple as this formula would indicate. However, there are a number of implications of the mathematics that are useful guidelines:

- The generated voltage is proportional to the area of the receiver coil. Pickup can be minimized by reducing this area.
- In the example shown, the receiver coil is smaller than the transmitter coil, so the receiver coil is in a uniform magnetic field and the voltage is proportional to the area of the receiver coil. If the coil sizes are reversed so that the transmitter is smaller than the receiver, then this will reduce the flux flowing through the receiver coil just as if the receiver coil area had been reduced. Consequently, reducing either the transmitter or receiver coil area reduces the induced voltage.
- The induced voltage is proportional to frequency. This coupling is likely to be more of a problem at higher frequencies.
- In the diagram, the coils are parallel. If the coils are at an angle to each other, the induced voltage becomes proportional to the cosine of the relative angle. At 90° , where the transmitter coil and receiver coil are at right angles, the induced voltage is zero. This is a convenient test to determine whether pickup is from an electric or magnetic field. Rotation of the probe will have no effect on electric field pickup.
- If the receiver coil is rotated through 180° with respect to the transmitter coil, the cosine effect continues to hold and the voltage across the terminals reverses polarity. For example, if the two-loop arrangement in figure 913 is illuminated with a uniform magnetic field, the generated voltages in the two loops will cancel and V_{xy} will be zero.
- This analysis assumes that there is close proximity of the two coils. The field of the transmitter coil falls off with distance, so the pickup can be reduced by separating the two coils.
- The coupling between the two coils will be reduced if the magnetic field is somehow redirected away from the receiver coil. That is the function of magnetic *shielding*, as discussed below.

34.7.1 Magnetic Field Identification

A small coil of wire, connected to an oscilloscope or spectrum analyser, may be used as a *search coil* (probe) to detect magnetic fields. Ideally, the probe is surrounded by an electrostatic shield (with a slot in it) so that electric fields do not generate a signal in the probe [293] The concept is shown in figure 915(a).

However, that is not essential: Smith in [294] describes an unshielded magnetic field probe using a bent paper clip, figure 915(b). The coil is usually terminated through a series 50Ω resistor at the coil or a 50Ω resistive input at the oscilloscope or spectrum analyser to avoid reflections in the connecting coaxial cable.

34.7.2 Voltage in the Conductor

If the magnetic probe senses the flux due to current in a conductor, then the shape of detected voltage waveform will be the same as the voltage across that segment of the conductor. This follows from equation 1563, where the

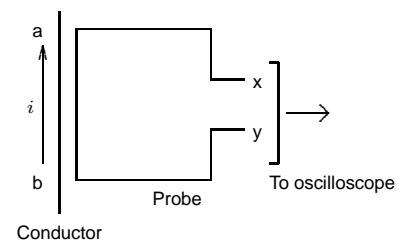


Figure 914: Measuring Voltage in a Conductor

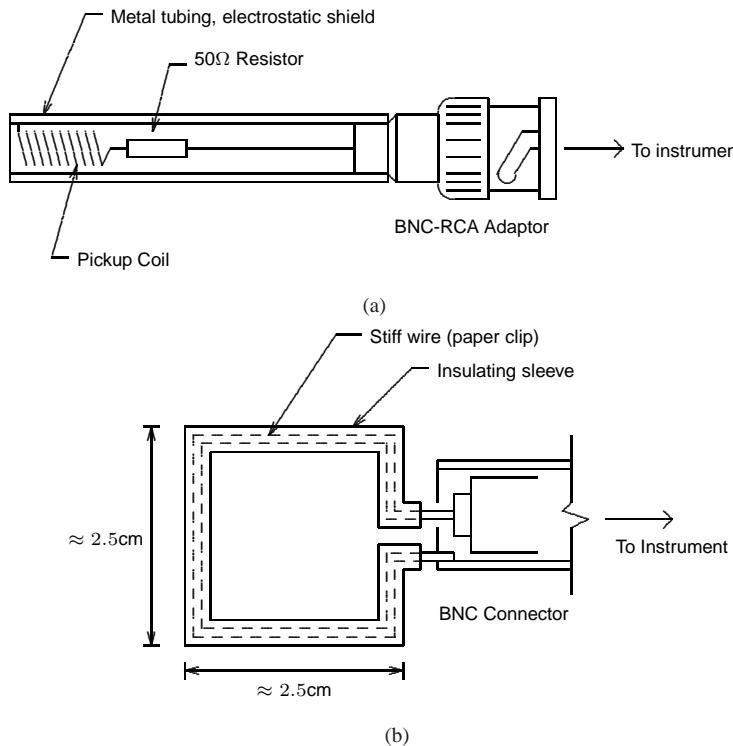


Figure 915: Magnetic Field Probes

probe voltage e_{xy} is shown to be proportional to the rate of change of the current $di(t)/dt$ in the probed conductor. If that section ab of the conductor has an inductance L_{ab} , then the voltage across it will be

$$e_{ab} = L_{ab} \frac{di(t)}{dt} \propto e_{xy} \quad (1565)$$

where

e_{ab} is the voltage between a and b on the current-carrying conductor

L_{ab} is the inductance of the conductor between a and b

e_{xy} is the open-circuit output voltage of the magnetic probe

If there is perfect magnetic coupling between the conducting wire and the probe, then the magnitude of the two voltages is the same: $e_{ab} = e_{xy}$. This is not likely to be the case, so the value of the probe voltage sets an upper limit on the voltage across the conductor.

The current-carrying wire and the probe constitute a transformer. If the flux of the wire couples perfectly to the probe, then they form a 1:1 transformer. The voltage across the two parallel wires are experiencing the same rate of change of flux and the same voltage. In general, there will be some flux leakage and the voltage across the secondary (the probe) will be somewhat less than the voltage in the current-carrying conductor.

34.7.3 Demonstration: Magnetic Field Pickup

Figure 916(a) shows an arrangement to demonstrate magnetic field pickup. A stiff wire is bent into a square-shaped probe and then plugged into one end of a coaxial cable. The other end goes to the input of an oscilloscope. A signal generator drives current into a separate length of wire to create a magnetic field.

In this example, the probe is approximately 3.5cm on a side. The length of wire is about 35cm long. As the probe is rotated, the detected voltage decreases, going through a null when the probe and wire are at right angles. Since the detected voltage changes with the orientation of the probe, the pickup is magnetic, rather than electric.

The voltage across the probe terminals will be equal to or less than the voltage in a 3.5cm segment of the conductor [295]. In the measurement setup of figure 916(a), the current-carrying conductor was 35cm. The 3.5cm probe produced 20mVp-p, implying a total voltage across the conductor of 200mV. The voltage across the conductor actually measured some 300mV because the prediction sets a lower bound on the voltage.

The electrical equivalent circuit for the probe is shown in figure 916(b). The inductance of the transformer secondary forms a lowpass filter with the input resistance R_i of the oscilloscope, corner frequency $\omega_o = R/L$ (section 5.2.4 on page 203). When the oscilloscope cable is unterminated, then the effective load resistance R_i for the transformer is $1M\Omega$. If the 1MHz operating frequency is below the cutoff frequency of the highpass filter then the detected waveform has the same shape as the signal generator waveform.

If the cable is terminated with 50Ω , the cutoff frequency of the lowpass filter is moved downward. If the operating frequency is above this new cutoff frequency, the detected waveform is a lowpass filtered version of the signal generator waveform – the high frequency edges are removed.

34.7.4 Magnetic Field Shielding

Electric-field shielding was discussed in section 34.6.2 on page 1018. A thin conductive sheet is effective in blocking an electric field.

Magnetic field shielding is not so straightforward. A ferrous material such as iron functions as a magnetic field shield by providing a high-permeability path for the field lines. Non-ferrous materials such as copper and aluminum function as a shield for an AC magnetic field by inducing circulating currents in the material that create an opposing magnetic field. The shielding effectiveness increases with frequency (see reference [296], Part 5). However, non-ferrous materials are not an effective shield for DC or low-frequency magnetic fields. For example, shielding for a 60Hz line-current generated field often requires a special magnetic alloy, commonly referred to as *mu-metal*. This approach tends to be complicated and expensive, and it's usually more cost-effective to shape the

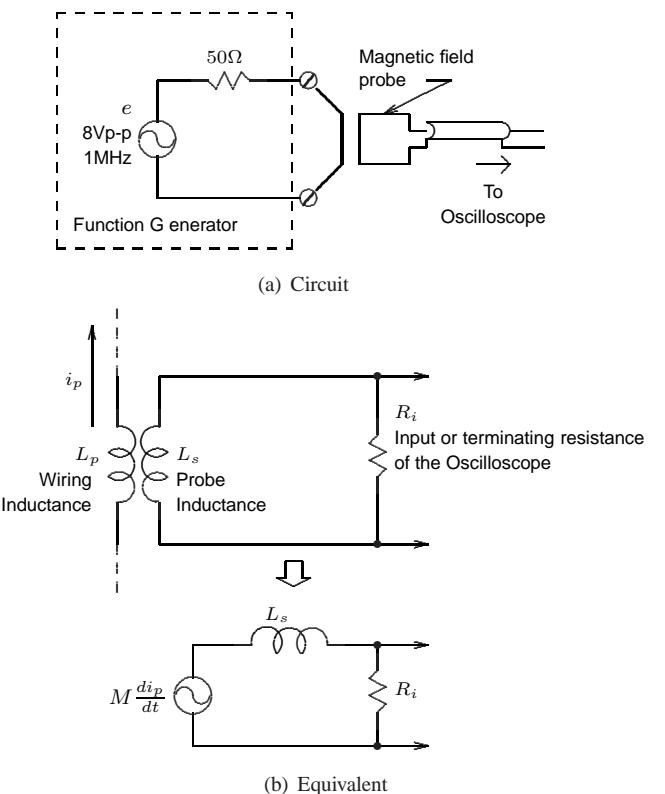


Figure 916: Magnetic Field Pickup

magnetic field. For example, changing from an E-core transformer to a toroidal power transformer may reduce the offending field, since a magnetic toroid has very little external magnetic field.

It can be shown that the external magnetic field decreases inside the shielding material to $1/e$ of its original value in a distance equal to one *skin depth*, where skin depth is given by [297]:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \quad (1566)$$

where the variables are

- δ Skin depth, metres
- ω Frequency of the magnetic field, radians/sec
- μ Magnetic permeability of the shielding material, given by $\mu = \mu_r\mu_o$
- μ_r Relative permeability, a dimensionless constant, the ratio of the material permeability to free space
- μ_o Magnetic permeability of free space, a constant, $4\pi \times 10^{-7}$
- σ Conductivity of the shield material, V/metre given by $\sigma = 1/\rho$, Siemens/metre
- ρ Resistivity of the shield material, $\Omega\text{-metre}$

Example

A switching power supply, operating at 100kHz, is generating a magnetic field that couples into nearby wiring. What would be the shielding effectiveness of 0.050 inch thick aluminum in blocking this field?

The resistivity of aluminum is found in a table of Properties of Materials to be $2.6 \mu\Omega\text{-cm}$.

Solution

First, convert the resistivity to MKS units:

$$\rho = 2.6 \mu\Omega\text{-cm} = 2.6 \times 10^{-6} \Omega\text{-cm} \times \frac{1 \text{ M}}{100 \text{ cm}} = 2.6 \times 10^{-8} \Omega\text{-M}$$

The conductivity is the reciprocal of the resistivity:

$$\sigma = \frac{1}{\rho} = \frac{1}{2.6 \times 10^{-8} \Omega\text{-M}} = 38 \times 10^6 \text{ Siemens/metre}$$

Since aluminum is not a ferromagnetic material (ie, cannot be magnetized), its relative permeability is close to unity. Then the permeability is simply equal to the permeability of space:

$$\mu = \mu_r\mu_o = 1 \times 4\pi \times 10^{-7} = 4\pi \times 10^{-7}$$

Now we can use equation 1566 to calculate the skin-depth of the magnetic field:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} = \sqrt{\frac{2}{(2\pi \times 100 \times 10^3) \times (4\pi \times 10^{-7}) \times (38 \times 10^6)}} = 2.58 \times 10^{-4} \text{ metres}$$

Now we can relate the skin depth to the attenuation. We convert the material thickness to metres, determine the number of skin-depths in the material, and then calculate the attenuation. The material thickness T is:

$$T = 0.050 \text{ inches} = 0.050 \text{ inches} \times \frac{2.54 \text{ cm}}{1 \text{ inch}} \times \frac{1 \text{ metre}}{100 \text{ cm}} = 1.27 \times 10^{-3} \text{ metres}$$

Each skin depth contributes a factor of $1/e$ attenuation, and the overall thickness of the material contains T/δ skin depths so the overall attenuation factor K is:

$$K = \left(\frac{1}{e}\right)^{T/\delta} = \left(\frac{1}{2.71}\right)^{1.27 \times 10^{-3}/2.58 \times 10^{-4}} = 7.28 \times 10^{-3}$$

That is, magnetic field inside the shield is approximately 100 times less than the field outside the shield. The voltage induced in a coil would reduce by the same factor.

In decibels, this factor is:

$$K_{db} = 20 \log_{10} K = 20 \log_{10} (7.28 \times 10^{-3}) = 42.7 \text{ db}$$

34.8 Spectrum of Pulse Waveform

Digital signals, such as those present in a switching power supply or microprocessor, are a frequent source of interference with other circuits. The spectrum of such a signal gives some important clues on minimizing this interference. Once the spectrum is known, then it is possible to design suitable filters to contain the digital noise.

34.8.1 Square Wave Spectrum

A square wave is shown in figure 917(a). The formula for the spectrum of this type of square wave (one with zero rise and fall time) is given in [298] as:

$$C_n = 2A_{AV} \left| \frac{\sin(n\pi t_o/T)}{n\pi t_o/T} \right| \quad (1567)$$

where the variables are:

- C_n amplitude of the n th component of the spectrum, volts
- A_{AV} average value of the waveform, volts
- n harmonic number (multiple of the fundamental)
- t_o time for half the period, seconds
- T period, seconds

The average value of the waveform is:

$$A_{AV} = A \left(\frac{t_o + t_r}{T} \right) \quad (1568)$$

An example spectrum given by equations 1567 and 1568 for a 1MHz square wave ($T=1\mu\text{Sec}$) of amplitude 5 volts ($A=5$), is shown in figure 917(b). The vertical axis is expressed in decibels relative to $1\mu\text{V}$, that is:

$$\text{Amplitude, db} = 20 \log_{10} \left(\frac{C_n}{1 \times 10^{-6}} \right) \quad (1569)$$

Notice that the harmonics fall at the expected frequencies for a square wave: 1MHz, 3MHz, 5MHz and so on. The amplitude decreases at a rate of 20db/decade.

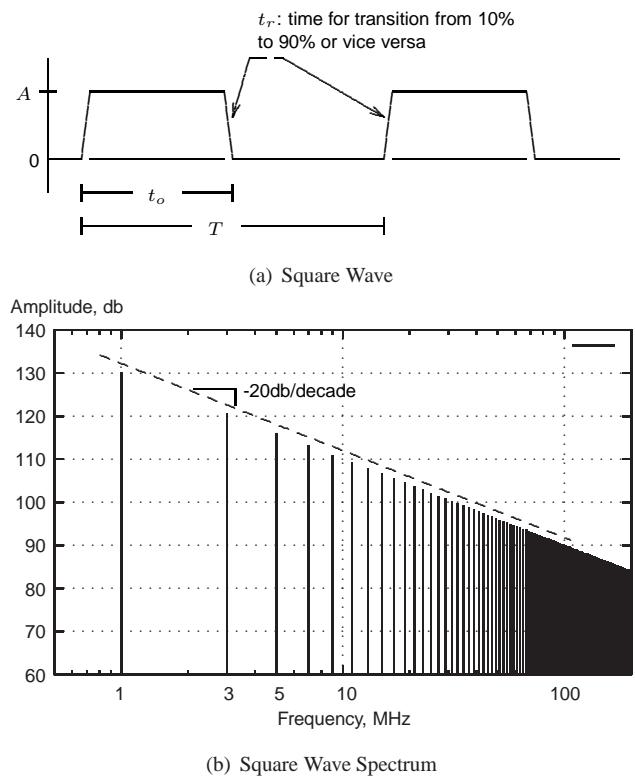


Figure 917: Square Wave Spectrum

34.8.2 Trapezoidal Wave Spectrum

Now consider the trapezoidal waveform shown in figure 918(a). This has significant rise time t_r and fall time t_f . Assuming that the rise and fall time are equal, the expression for the spectrum [298] is similar to equation 1567, but with an additional term:

$$C_n = 2A_{AV} \left| \frac{\sin(n\pi t_r/T)}{n\pi t_r/T} \frac{\sin(n\pi(t_o + t_r)/T)}{n\pi(t_o + t_r)/T} \right| \quad (1570)$$

where the additional variable t_r is the rise (and fall) time. The rise time is usually much smaller than the duration of the pulse, so equation 1570 may be simplified to:

$$C_n = 2A_{AV} \left| \frac{\sin(n\pi t_o/T)}{n\pi t_o/T} \frac{\sin(n\pi t_r/T)}{n\pi t_r/T} \right| \quad (1571)$$

where the average value of the waveform is given by equation 1568. Using equation 1571 and the same parameters as the previous square wave example, with a rise time t_r of 20nSec, the spectrum for the trapezoidal waveform is shown in figure 918(b).

Notice that the amplitude decreases at a rate of 20db/decade up to a certain frequency, and then more rapidly thereafter, at 40db/decade. Finite rise time in the time domain decreases the spectral energy at high frequencies in the frequency domain.

34.8.3 The Bode Approximation

In this section we'll look at a method for predicting the amplitude spectrum for a trapezoid waveform.

The spectrum of figure 918(b) can be approximated as shown in figure 918(c).

There are two break frequencies, which we'll call f_1 and f_2 . The lower break frequency is determined by the pulse width t_o ; the upper break frequency f_2 is determined by the pulse rise (and fall) time t_r .

$$f_1 = \frac{1}{\pi t_o}, \quad f_2 = \frac{1}{\pi t_r} \quad (1572)$$

Above f_2 , the amplitude of the spectrum decreases rapidly, and so f_2 is usually taken as the 'upper frequency limit' of the signal.

Filtering of noise and containment of electromagnetic radiation is simpler at lower frequencies, so it is desirable to limit this high frequency content. Consequently, **to minimize noise and interference problems digital circuitry should use logic and pulse waveforms with the slowest transition times that are acceptable.**

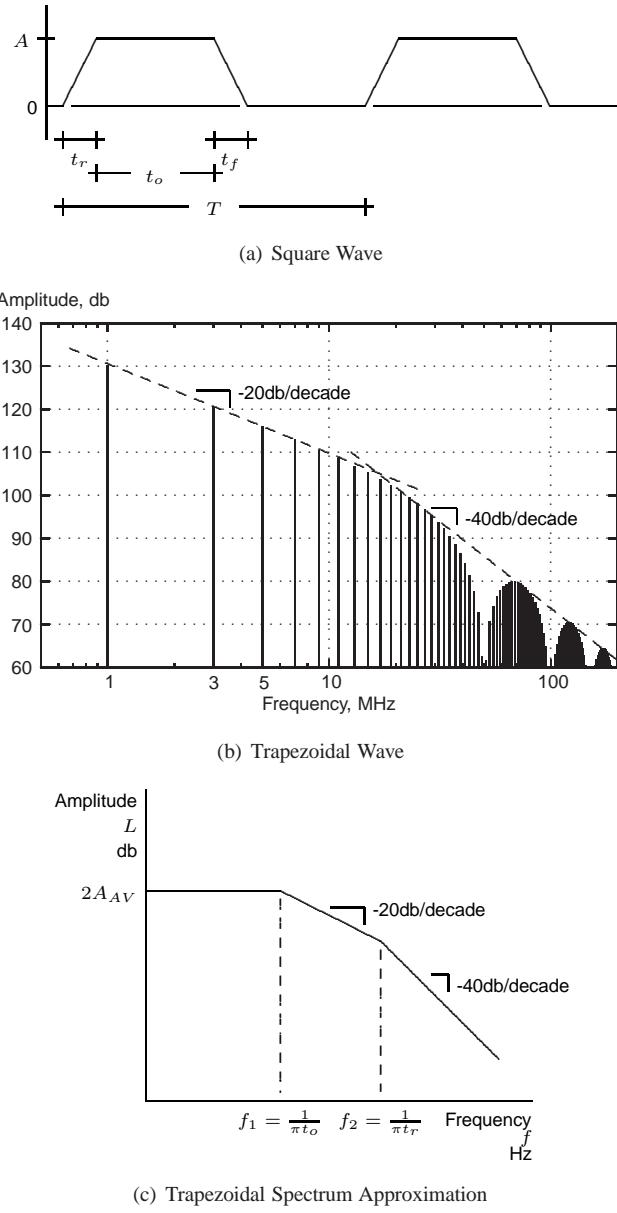


Figure 918: Trapezoid Wave Spectra

Example

A certain pulse waveform is a square wave with a peak-peak amplitude of 5 volts, a frequency of 1MHz and rise/fall time of 20nSec. Calculate the spectrum amplitude and the two break frequencies.

Solution

The amplitude L is twice the average value A_{AV} . The duty cycle t_o/T is 50% or 0.5. Then using equation 1568:

$$L = 2A_{AV} = 2A \frac{t_o}{T} = 2 \times 5 \times 0.5 = 5 \text{ volts} \quad (1573)$$

On a Bode plot this must be expressed in decibels relative to some reference. For example, for a $1\mu\text{V}$ reference, we have:

$$L_{db\mu\text{V}} = 20 \log \frac{L}{1\mu\text{V}} = 20 \log \frac{5}{1 \times 10^{-6}} = 133.9 \text{ db}\mu\text{V} \quad (1574)$$

The break frequencies are given by equation 1572:

$$f_1 = \frac{1}{\pi t_o} = \frac{1}{\pi \times (0.5 \times 10^{-9})} = 637 \text{ kHz} \quad (1575)$$

$$f_2 = \frac{1}{\pi t_r} = \frac{1}{\pi \times (20 \times 10^{-9})} = 15.9 \text{ MHz} \quad (1576)$$

Then the amplitude spectrum decreases rapidly above 15.9 MHz

34.8.4 Derivation: Spectrum of Trapezoidal Waveform

In this section, we'll derive the break frequency expressions for f_1 and f_2 given by equation 1572.

The amplitude spectrum for a trapezoid digital wave is given by equation 1571. From that starting point, the envelope of the trapezoid spectrum is determined by the expression:

$$C_f = 2A_{AV} |sinc(\pi t_o f) sinc(\pi t_r f)| \quad (1577)$$

where $sinc(x) = \frac{\sin(x)}{x}$ and the discrete harmonics at frequencies n/T have been replaced by a continuously varying frequency f .

The sinc function is shown in figure 919, when both axes are plotted to logarithmic scales. For values of x less than 1, the value of the function is 1. For values of x more than 1, the function decreases by a factor of 10 when the value of x increases by a factor of 10. Consequently, the function is proportional to $1/x$. At the corner frequency, these two are equal, that is:

$$\begin{aligned} 1/x &= 1 \\ x &= 1 \end{aligned} \quad (1578)$$

The Bode approximation of a sinc function is thus a constant value of 1 over the range $0 < x < 1$, decreasing at 20db/decade for $x > 1$. The corner occurs at $x = 1$.

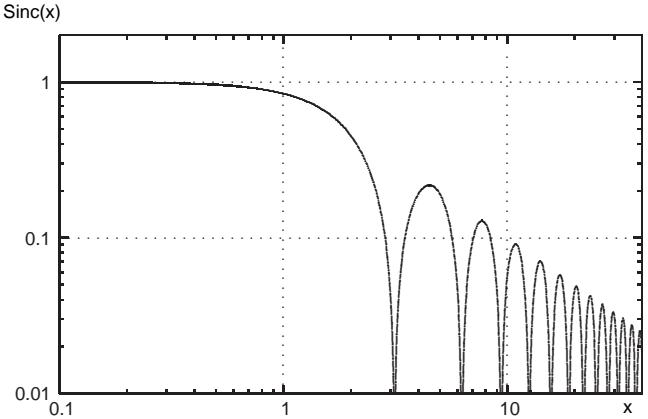


Figure 919: Sinc Function

Now, in order to convert equation 1577 into a suitable format for a Bode plot, we take 20 times the \log_{10} of C_f :

$$\begin{aligned} 20 \log_{10} C_f &= 20 \log_{10} (2A_{AV} |\text{sinc}(\pi t_o f) \text{sinc}(\pi t_r f)|) \\ &= 20 \log_{10} (2A_{AV}) + 20 \log_{10} \text{sinc}(\pi t_o f) + 20 \log_{10} \text{sinc}(\pi t_r f) \end{aligned} \quad (1579)$$

This function has three parts: a constant value, followed by two sinc functions. Each of the sinc functions is a lowpass filter.

In the case of the first sinc function, the corner frequency f_1 occurs when:

$$\begin{aligned} \pi t_o f_1 &= 1 \\ f_1 &= \frac{1}{\pi t_o} \end{aligned} \quad (1580)$$

Similarly

$$\begin{aligned} \pi t_r f_2 &= 1 \\ f_2 &= \frac{1}{\pi t_r} \end{aligned} \quad (1581)$$

Consequently, the frequency spectrum is as shown in figure 918 on page 1027.

34.9 Components of The Hidden Schematic

In this section, we identify several components of the *hidden schematic*. These are electrical properties of components that are not normally indicated on a schematic diagram, but may be important in the functioning of the circuit.

For example, wires are generally assumed to have zero resistance. However, there are situations where wiring resistance is critical and must be considered. Here we identify some examples of those components and situations.

34.9.1 Wiring Resistance

Connecting wire has resistance, which can be significant under certain circumstances. For example, the resistance of #22AWG solid wire is 52.9Ω per Km. An example where this is important is shown in figure 920(a).

In this example, a 2 amp capacitor-filter full-wave rectified power supply (section 15.2) is found to have 100 millivolt noise 'bumps' superimposed on the DC output voltage. These pulses find their way into the load circuitry as noise.

The noise pulses occur at a rate of 120 per second, twice the line frequency. Increasing the filter capacitance *increases* the amplitude of the noise pulses. What is the cause?

The schematic doesn't tell the whole story. The wiring, shown in figure 920(b), is more informative.

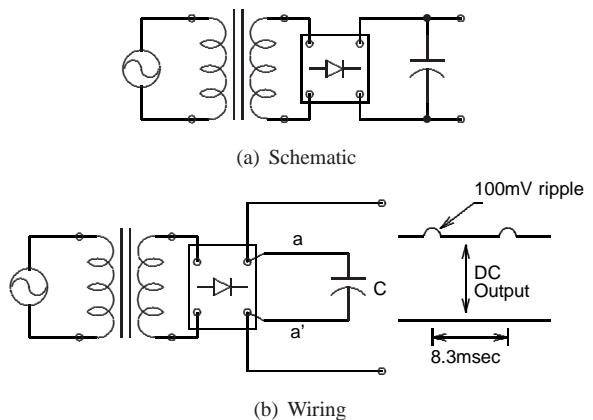


Figure 920: Power Supply Capacitive Filter

The power supply has been wired so that the output is taken from the full-wave bridge, with separate connecting wires to the filter capacitor C . The wires a-a' connecting the capacitor to the bridge rectifier are #22AWG, 20cm in length, so each connecting wire has a resistance of:

$$\frac{52.9 \text{ ohms/km}}{1000 \text{ metres/km}} \times 0.20 \text{ metres} = 10.6 \times 10^{-3} \text{ ohms}$$

In a capacitor filtered power supply, the bridge rectifier supplies current to the filter capacitor in short pulses. (Figure 367(c) on page 426 shows an example of the sawtooth-shaped current waveform for a half-wave rectifier.) The average value of these current pulses is equal to the average output current. If the duration of the pulses is short, then the peak charging current into the capacitor is substantially larger than the output DC current. In this particular case, the peak current is 5 amps²⁸⁶, in which case the voltage developed across the capacitor connecting wires is

$$\begin{aligned} V &= 10.6 \times 10^{-3} \times 5 \\ &= 0.053 \text{ volts} \end{aligned}$$

There are two connecting wires, so the total voltage developed by the pulse of charging current is 100mV.

Increasing the filter capacitance C worsens the noise because it reduces the duty cycle of the charging pulses, increases the peak amplitude of the charging current pulse, and thereby increases the magnitude of the noise pulses.

The solution is simply to move the output connections to the terminals of the power supply. Assuming that the filter capacitor has low internal resistance, then the voltage *at its terminals* will be constant, and this is what should be taken to the load circuit.

Notice the steps in solving this particular puzzle:

- Be aware of the actual *current* waveform in the circuit. An oscilloscope shows the voltage waveform and in this case the shape of the current waveform is quite different from the voltage.
- Treat connecting wires as having resistance.
- Consider the actual wiring arrangement of the circuit.

Incidentally, here is a trick to remembering approximate values of wiring resistance.

First, in English units:

- The resistance of #10AWG wire is about 1Ω per 1000 feet.
- Resistance increases by a factor of 10 with each increase of 10 wire gauges. So #20 would be about 10Ω per 1000 feet.

You can interpolate between gauges with the relationship

$$\Delta \text{AWG} = 10 \log_{10} \frac{R_2}{R_1} \quad (1582)$$

where ΔAWG is the change in wire gauge between resistances R_1 and R_2 . This formula is similar to the decibel relationship of powers, so it's not unreasonable to remember.

²⁸⁶The power supply example of page 426 has a peak to average current ratio of 28, so 5 amps peak is not at all unlikely in a power supply with 2 amperes output current.

Example

What is the approximate resistance of #26AWG copper wire?

Solution

We know that #10AWG is 1Ω per 1000 feet. Then #20 is about 10Ω per 1000 feet and we'll choose that as R_1 , our starting point. The value of ΔAWG is $26 - 20 = 6$. Plug those values into equation 1582 and we obtain $R_2 = 39.8\Omega$ per 1000 feet. (According to a wire table, it's actually 40.82Ω per 1000 feet, so this approach is fine for estimates.)

If you prefer metric units, #5AWG is about 1Ω per kilometer, and the same method applies.

34.9.2 Resistance of Printed Wiring Traces

The resistance of printed wiring traces depends on their length and cross-sectional area, according to

$$R = \rho \frac{L}{A} \quad (1583)$$

where in CGS units the quantities are:

- R resistance in Ω
- ρ resistivity of copper, $1.724\mu\Omega\text{-cm}$
- L length of the conductor in centimetres
- A cross-sectional area of the conductor in square centimetres

In the construction of a printed wiring board, the copper thickness is specified in ounces, which is the weight of copper plated over a one-foot square area. Boards are typically plated with 1, 1.5 and 2 ounce copper.

Using the density of copper (8.92 grams/cm^3) and working through the various conversion factors, a 1 ounce copper plating is equivalent to a thickness of 0.00135 inches (0.00343 centimetres).

The width of printed circuit traces is specified in 'mils', or thousandths of an inch.

Example

A 30 mil trace on a printed wiring board is 9 inches long. The board is plated with 1oz copper. If the trace conducts a peak current of 2 amps, what is the end-end voltage drop across this trace?

Solution

First, we need to use equation 1583 to calculate the resistance of the trace. The trace width is 0.030 inches, or 0.0762 centimetres. The trace height for 1oz copper is 0.00343 centimetres. The length is 9 inches, or 22.8 centimetres. Plugging these figures into equation 1583, we have

$$\begin{aligned} R &= \rho \frac{L}{A} \\ &= 1.724 \times 10^{-6} \times \frac{22.8}{0.0762 \times 0.00343} \\ &= 0.15\Omega \end{aligned}$$

The voltage drop is

$$\begin{aligned} V &= IR \\ &= 2 \times 0.15 \\ &= 0.3 \text{ volts} \end{aligned}$$

Let us now put this in perspective. For an 8 bit, 5 volt A/D converter, each step is about 19mV, so this voltage drop represents a possible error source of 15 digital counts, or a 6% error. There are many systems where this error would be unacceptable. If this voltage drop is in the form of an AC voltage, then it would show up as jitter in the A/D reading. The wiring of the circuit board must be arranged so that this voltage does not become a source of error – by adding to the signal voltage, for example.

Where traces must have low resistance, they should be made as wide as possible. It is may be possible to design a printed wiring board so that conducting areas occupy most of the board, and the insulating gaps are narrow. This substantially reduces the resistance of the conducting traces.

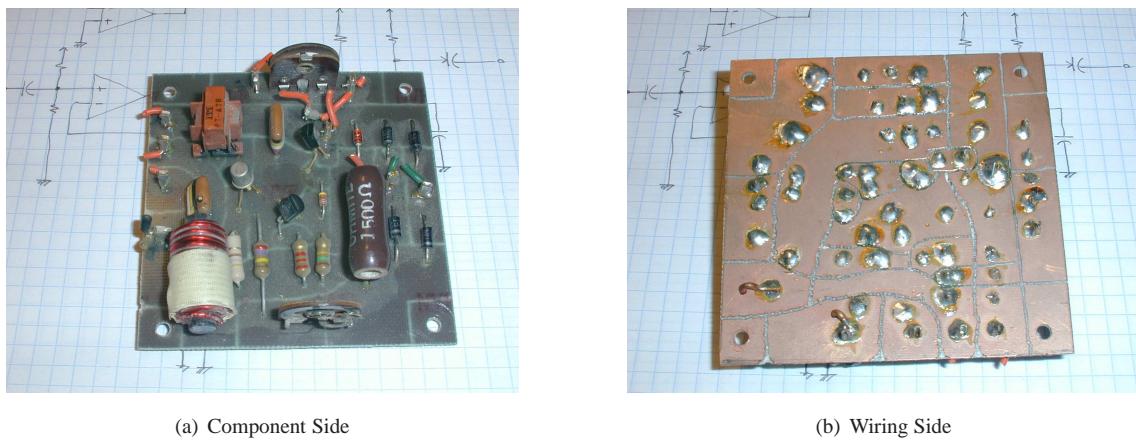


Figure 921: Reducing Trace Resistance

Figure 921 shows a circuit board that was laid out in this fashion. The conductors are islands of copper that are isolated by narrow channels.

34.9.3 Wiring Inductance

Modern digital circuits operate at fast transition rates, in the order of tens of nanoseconds through to fractions of a nanosecond. Even if the associated currents are small, the rapid **change** creates large values of di/dt .

According to the equation

$$e = L \frac{di}{dt} \quad (1584)$$

this rapidly changing current can create significant voltage across a very small inductance.

Even a short length of wire can have sufficient inductance to generate significant voltage under these circumstances. This is somewhat counterintuitive, so consider the following demonstration.

As shown in figure 922, an #18AWG (Pomona 1440) test lead is connected across the terminals of a signal generator. At DC and low frequency, this shorting strap has a very low impedance and so the measured voltage on the oscilloscope is quite small. However, at a frequency of 1MHz, a short length of wire has an appreciable inductance and consequently a significant voltage appears between the ends of the wire.

The open-circuit voltage of the generator is set to 8Vp-p, 1MHz. The internal 50Ω resistance of the gen-

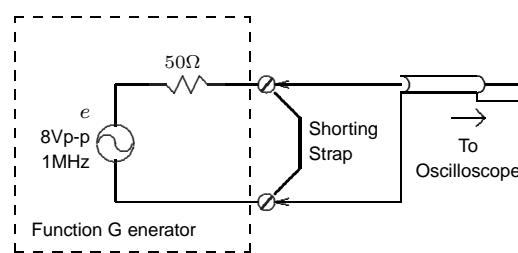


Figure 922: Wiring Inductance

erator and the inductance of the wire form a voltage divider, with the results shown under *Measured Inductance*.

Wire Length Inches	Developed Voltage mVp-p	Reactance Ω	Measured Inductance, μH	Predicted Inductance μH
12	300	2.05	0.326	0.382
24	600	3.6	0.644	0.856
50	1200	8.8	1.41	1.97

Notice that the developed voltage and the measured inductance of the shorting wire increase with the length of the wire. If the frequency is dropped to 100kHz, the developed voltage drops by a factor of ten, confirming that the inductance is proportional to frequency.

A wire has significant inductance at high frequency.

Can this inductance be predicted? Reference [299] gives the following formula for the inductance of a straight wire:

$$L = 2 \times 10^{-4} \left(\ln \frac{4l}{d} - 0.75 \right) \mu\text{H} \quad (1585)$$

where the variables are:

- L inductance in μH
- l length of the conductor, mm
- d diameter of the conductor, mm

The diameter of #18AWG, according to a wire table, is 1.024mm. Plugging the wire lengths (in mm) from the measurement into equation 1585, we obtain the figures under *Predicted Inductance*. The predictions are in the ball-park, which is usually sufficient.

34.9.4 Inductance of a Printed Wiring Trace

The inductance of a printed wiring trace is given in [299] as:

$$L = 2 \times 10^{-4} \left[\ln \left(\frac{2l}{w+h} \right) + 0.2235 \left(\frac{w+h}{l} \right) + 0.5 \right] \mu\text{H} \quad (1586)$$

where the variables are:

- L inductance in μH
- l length of the conductor, mm
- w width of the conductor, mm
- h height of the conductor, mm

Example

What is the inductance of a 20cm length of 10mil wide copper trace, printed with 1 oz copper?

Solution

First, we need to calculate the dimensions of the conductor in millimetres.

Length l :

$$\begin{aligned} l &= 20 \text{ cm} \times \frac{10 \text{ mm}}{1 \text{ cm}} \\ &= 200 \text{ mm} \end{aligned}$$

Width w : One mil is 1/1000 inch. An inch is 25.4mm. Then 10 mils is:

$$\begin{aligned} w &= 10 \text{ mil} \times \frac{1}{1000} \frac{\text{inch}}{\text{mil}} \times \frac{25.4 \text{ mm}}{1 \text{ inch}} \\ &= 0.254 \text{ mm} \end{aligned}$$

Height h : On page 1031 we established that 1oz copper plate has a thickness of 0.00343 centimetres, or 0.0343 millimetres.

Now we can calculate the inductance, using equation 1586:

$$\begin{aligned} L &= 2 \times 10^{-4} \left[\ln \left(\frac{2l}{w+h} \right) + 0.2235 \left(\frac{w+h}{l} \right) + 0.5 \right] \mu\text{H} \\ &= 2 \times 10^{-4} \left[\ln \left(\frac{2 \times 200}{0.254 + 0.0343} \right) + 0.2235 \left(\frac{0.254 + 0.0343}{200} \right) + 0.5 \right] \mu\text{H} \\ &= 1.546 \times 10^{-3} \mu\text{H} \\ &= 1.546 \text{ nH} \end{aligned}$$

Conductor Dimensions and Inductance

Equation 1586 indicates that trace inductance

- increases directly with trace length.
- decreases slowly with trace width, since the log term dominates
- is largely unaffected by trace thickness

Consequently **trace inductance is most effectively reduced by shortening the length of a conductor.**

34.9.5 Power Supply Inductance and The Bypass Capacitor

Many systems include both digital and analog circuitry. The high frequency components of a digital signal (figure 917) tend to propagate via power supply lines into other sections of the circuit. It is a major challenge to keep these digital signals from interfering with analog sections of the system.

One of the main lines of defence is the *bypass capacitor*. A bypass capacitor is connected between a power supply line and ground. Ideally, it conducts all frequencies of alternating currents, thereby preventing any variation in power supply voltage. Alternatively, the bypass capacitor can be regarded as a reservoir of charge that can be provided to keep the power supply line at a constant voltage.

An example of a mixed analog-digital system power supply is shown in figure 923(a).

The system consists of digital gates U_1 and U_2 . The same power supply is used by the analog section, represented by the op-amp U_3 .

The digital section is redrawn in figure 923(b) to indicate its essential features. The power supply connections create inductances L_1 and L_2 . The input of logic gate U_2 can be represented by the load capacitance C_{in2} .

In figure 923(c) the logic gate is represented by a SPDT switch, which connects the output pin to the supply V_{CC} or to ground.

Suppose that this switch is originally in the lower position, so that the capacitor C_{in2} is discharged. Then the switch is moved to the upper position. The capacitor must charge through the power supply inductance L_1 . Since inductor current cannot change instantaneously, inductor L_1 appears briefly as an open circuit, and the voltage at A drops to ground level. Then the capacitor charges to the supply voltage and point A returns to V_{CC} . Inductance L_1 and capacitance V_{CC} form a resonant circuit. Depending on the circuit resistance – which affects the damping – the transient may result in ringing at the resonant frequency of the inductor-capacitor. In any case, an ugly voltage transient appears on the power supply line.

Now, back to figure 923(a). This same transient, created by the digital circuitry, will appear at the V_{CC} terminal of the op-amp. Since the op-amp is unlikely to have significant power supply rejection at the frequency of this transient, the transient will appear at the output of the op-amp and contaminate that signal.

In some cases, the transient voltage at A may communicate itself sufficiently to the subsequent logic gate U_2 to cause it to switch incorrectly.

Now consider figure 923(d), in which a *bypass capacitor* C_{bypass} is provided near the digital device. This capacitor is chosen to be large compared to C_{in2} . It acts as a reservoir of charge that can be delivered rapidly into the load capacitance, without materially affecting the supply voltage. The charge on C_{bypass} is then restored via L_1 and the power supply over a longer period of time. The net effect is to substantially reduce the magnitude of switching transients on the power supply line. (It would help significantly to provide the analog circuitry with its own power supply lines, independent of the digital power supply lines. However, since the power supply undoubtably has some finite internal inductance, separate leads are not a perfect solution and bypass capacitors must be provided at the digital circuitry.)

For this to work, the bypass capacitor itself must not contain significant inductance, and that is the subject of the next section.

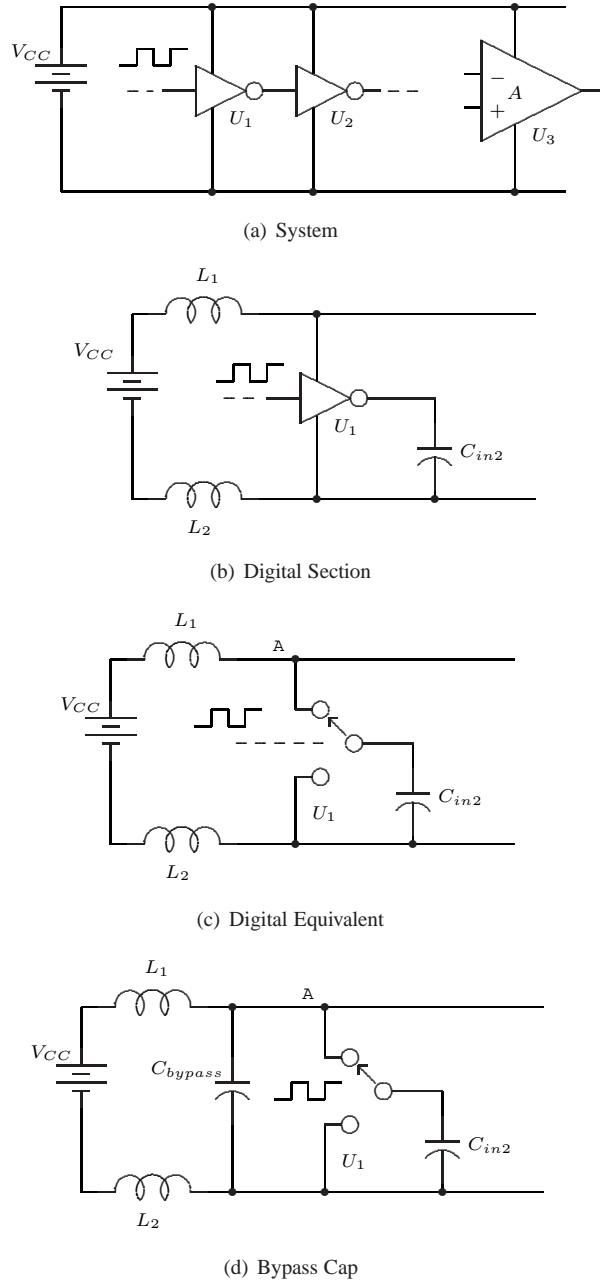


Figure 923: Analog/Digital Power Supply System

34.9.6 Parasitic Inductance in Capacitors

The reactance of an ideal capacitor decreases with frequency, without limit. Unfortunately, the connecting leads of a real-world capacitor form an inductance in series with the capacitor. The reactance of this inductance rises with frequency until it dominates, and the capacitor ceases to function as such.

Figure 924 illustrates this effect in the time domain.

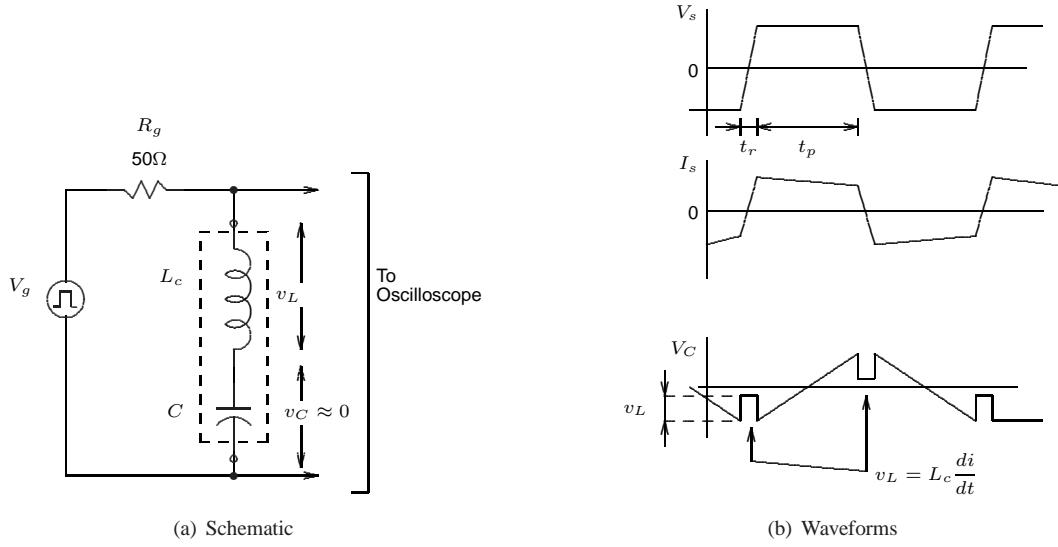


Figure 924: Capacitor Parasitic Inductance

A square wave generator with internal resistance R_g connects to a capacitor C . At low frequencies, this would result in the usual exponential charging waveform of an RC lowpass filter. However, the frequency of the generator is set to a large enough value that the capacitor cannot change substantially in voltage over one half of the pulse cycle. Consequently, its a reasonable approximation to take the capacitor voltage as zero.

The capacitor includes parasitic inductance L_p . When the generator voltage transitions from its negative value to its positive value, there is an abrupt change of capacitor current. This change in current induces a voltage in the parasitic inductance during the transition time t_r . This inductive spike is shown in idealized form as a 'notch' in the capacitor waveform V_c .

(In practice, the situation is more complex. The lead inductance and the capacitance form a resonant circuit. The generated pulse is a damped sinusoid with a decay time that depends on the damping in the circuit.)

To minimize the parasitic capacitance of bypass capacitors, their lead length and the length of the connecting traces to the device being bypassed should be as short as possible. Surface mount capacitors, with their zero lead length, are most effective.

Example

In the circuit diagram of figure 924, an 11 volt peak-peak, 1MHz square wave generator voltage is driven through a 50Ω resistance into a 100nF capacitor. The rise (and fall) time of the square wave is 17nSec . The capacitor lead lengths are 2.7cm each. The inductive spikes across the capacitor measure 200mV peak-peak. Determine the parasitic inductance of the capacitor.

Solution

Assume that the voltage across the capacitor is approximately zero volts. Then the capacitor current changes by an amount equal to the step change in voltage divided by the generator resistance.

$$\Delta i = \frac{\Delta V_g}{R_g} = \frac{11 \text{ volts}}{50\Omega} = 0.22\text{A} \quad (1587)$$

This change in current occurs during the rise time of the generator. Using

$$v_L = L_p \frac{\Delta i}{\Delta t}$$

we can solve this equation for the inductance as:

$$L_p = v_L \frac{\Delta t}{\Delta i} = (200 \times 10^{-3}) \frac{17 \times 10^{-9}}{0.22} = 15.4 \times 10^{-9} \text{ H} \quad (1588)$$

34.9.7 Capacitor Resonance

The capacitance and parasitic inductance form a series resonant circuit, along with a very small series resistance. In the frequency domain, the overall impedance is as shown with the solid line in figure 925. At low frequencies, the total impedance is essentially capacitive, since the inductance is small. The total impedance decreases with frequency. At the resonant frequency, the capacitive and inductive reactances are equal and cancel, leaving only the series resistance. Above the resonant frequency, the total impedance is inductive, since the capacitance is small. The total impedance increases with frequency. A **bypass capacitor** is usable in the region below resonance. The connecting leads of a bypass capacitor should be as short as feasible to minimize parasitic inductance.

34.9.8 Broadbanding the Bypass Capacitor

A digital waveform spectrum contains a wide range of frequencies, as seen on page 1026. Ideally, then, the bypassing capacitor should have a broadband response. This can be obtained by placing two capacitors of different values in parallel. Then the resonant frequencies are staggered and the lower impedance capacitor will define the total impedance.

Figure 926 shows the effect of a 100nF capacitor in parallel with a 1nF capacitor. The lower of the individual impedances (dotted lines) defines the total impedance (solid line), which is now maintained at a low value over a wider range of frequencies.

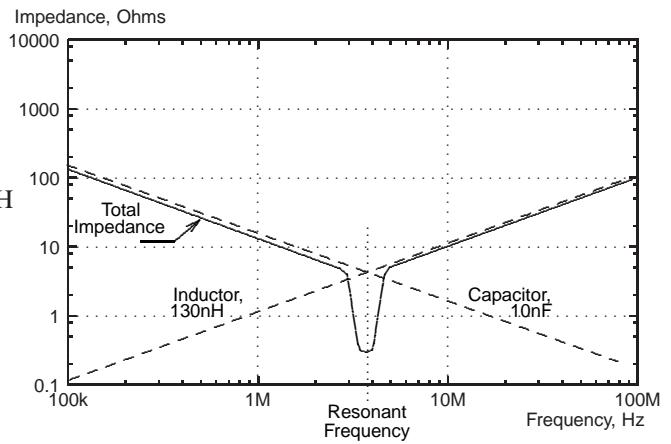


Figure 925: Bypass Capacitor Resonance

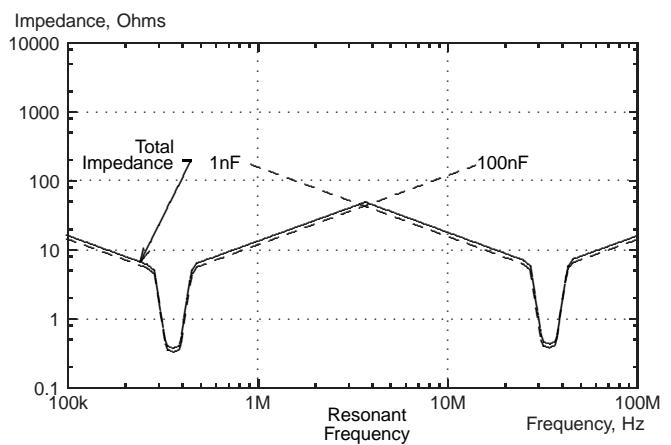


Figure 926: Broadbanding the Bypass Response

34.10 Common Mode and Differential Signals

The concept of *differential* and *common-mode* signals (section 21.5) is important in the control of interference. In this section, we'll show an example where the input circuitry of an amplifier can reject noise by (a) setting up a differential input and (b) rejecting the common-mode input signal.

34.10.1 Rejection of Ground Conductor Noise

There is nothing magical about a *ground* connection. It connects between various points of a circuit and, for convenience, is taken as the reference point for certain voltage measurements. However, currents flow through the various ground connections and thereby create voltage gradients that put different so-called ground points at different potentials.

It is an illustrative exercise to measure the potential between the ground terminals of two AC outlets that are separated in space. There can be millivolts or even volts of potential difference between them.

This potential difference between ground points is an issue in the situation shown in figure 927(a). A sensor of some type produces a signal e_s and is connected to ground at point A. The signal is connected to an amplifier which is connected to ground at point B. By virtue of circulating currents in the ground path, points A and B are separated by a potential e_n .

Redrawing figure 927(a) as figure 927(b), it can be seen that the signal and noise voltages are in series. The input voltage is the sum of e_s and e_n . The signal-noise ratio is entirely dependent on the relative amplitudes of the two voltages: there is no rejection of the noise signal.

Figure 927(c) shows way to reduce the effect of the noise. Providing that the signal is AC (as in an audio system, for example) a transformer at the input of the amplifier will reject the noise voltage. By superposition, the noise voltage appears equally at both transformer primary leads, so it cancels. The signal voltage appears differentially across the transformer primary, so it creates a voltage in the secondary.

Stray capacitance may exist between the transformer primary and secondary windings, and this electrostatic coupling will insinuate a noise signal into the input of the amplifier. High quality audio transformers include an electrostatic shield between the windings that can be grounded to prevent this coupling.

In effect, the sensor voltage e_s is a differential signal and the noise voltage e_n is the common mode signal. The system rejects the common-mode signal.

Figure 927(d) shows an approach similar to figure 927(c). In this case, a differential amplifier or instrumentation amplifier rejects the common-mode noise signal e_n . Since the input is direct coupled (no transformer) this approach will work down to zero frequency (DC).

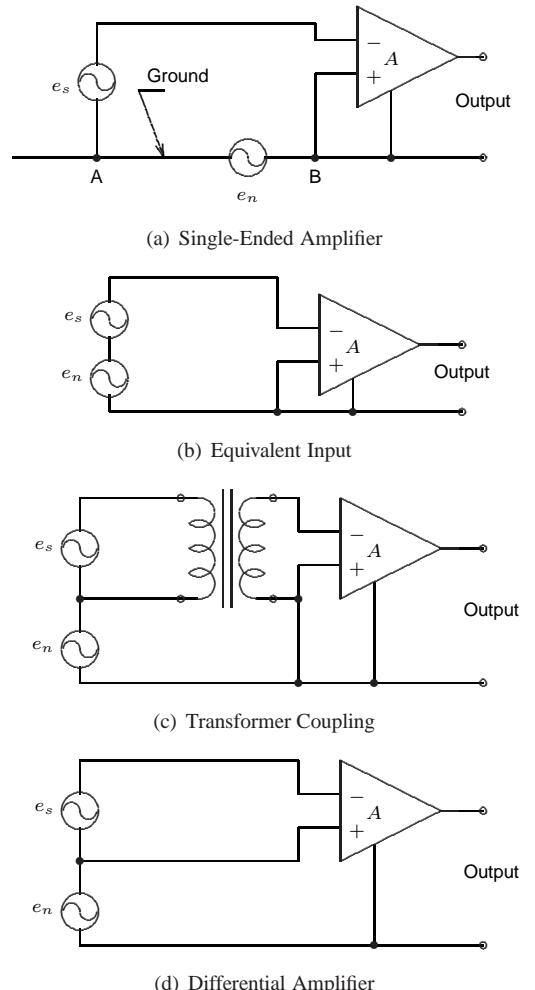


Figure 927: Differential Amplifier Application

34.10.2 Differential Amplifier Enhancements

To further reduce noise pickup, there are additional improvements that can be made to the differential input circuit.

The differential arrangements of figures 927(c) and 927(d) allow the signal conductors to be equal in length and physically close together. This minimizes the loop area in the input circuit of the amplifier, and therefore reduces pickup of any AC magnetic field. For even better magnetic field rejection, the two signal leads can be twisted together as shown in figure 928(a). Any field passing through the two input conductors will generate equal and opposite voltages in two adjacent loops of the conductor, so that magnetically induced voltages cancel.

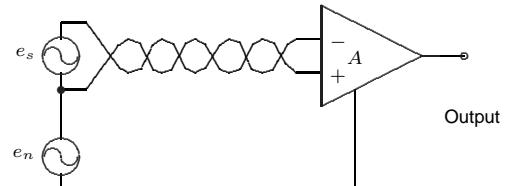
When Shakespeare wrote *When sorrows come, they come not single spies but in battalions*²⁸⁷, he could well have had interference problems in mind. Not only is magnetic field pickup an issue, but there may be an interfering electric field, and there may be interfering radio signals.

Electric field interference couples via stray capacitance into the input leads. If these capacitances are precisely equal and at a frequency where the differential amplifier has sufficient common-mode rejection, then the differential amplifier will eliminate the interfering signal. However, cable is readily available with a twisted pair and an outer electrostatic shield, and this provides an additional level of defence. A twisted-shielded pair is shown in figure 928(b). As shown in that diagram, a capacitively coupled electric field signal e_{n2} will be directed to ground by the shield.

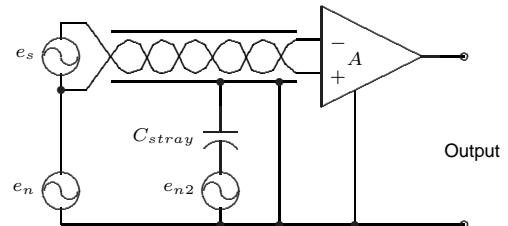
This same shield acts as a defence against radio-frequency interference. This is important, because differential and instrumentation amplifiers have little or no common-mode rejection at radio frequencies. Radio signals cannot penetrate a sealed metal box. The cable shield is connected to the metallic enclosure in such a way that it becomes an extension of the enclosure and prevents radio signal access into the internal circuitry.

Notice that the shield is connected to the enclosure – and ultimately, the reference potential – **at one point only**. If it is connected to ground at two points, then it and the ground path become a complete loop. This is undesirable: AC magnetic fields may induce currents in this loop, or the shield may provide an additional path for ground currents. In either case, these currents may develop a voltage across the shield that creates interference.

The scheme shown in figure 928(b) is widely used in audio systems for microphone signals. A dynamic (moving coil) microphone generates signals in the millivolt range and requires a signal-noise ratio of at least 60db. A microphone cable is often long and therefore acts as an excellent antenna, and may pass through regions around lighting equipment with large magnetic interfering fields. Consequently, stringent measures are required to minimize noise pickup. Twisted, shielded pair cable is standard, with connection of the shield to the equipment enclosure and the microphone casing. The input circuitry is differential, either using a transformer or an instrumentation amplifier.



(a) Twisted Pair Wiring



(b) Twisted-Shielded Pair Wiring

Figure 928: Differential Amplifier Enhancements

²⁸⁷Words uttered by King Claudius in Hamlet, Act 4, Scene 5.

34.10.3 Guarding

There is one further enhancement that may be useful under some circumstances: the *guard* [141], [300]. According to [301]:

Guarding..involves driving a shield, at low impedance, with a potential essentially equal to the common-mode voltage on the signal wire contained within the shield. Guarding has many useful purposes: it reduces common-mode capacitance, improves common-mode rejection, and eliminates leakage currents in high-impedance measurement circuits.

There is a certain amount of stray capacitance between the shield and the signal-carrying conductors of a balanced-shielded cable. If these stray capacitances are not equal, then the common-mode voltage will drive more current through one than the other. This creates a differential voltage on the internal signal wires.

If the shield is driven to the same potential as the common-mode voltage, then regardless of the magnitude of the stray capacitances there will no current through them, thereby improving rejection of the common-mode signal. Figure 929 shows a method of accomplishing this [141].

The resistors R_{3A}, R_{3B} establish a voltage equal to the common mode input voltage E_{CM} at the midpoint of the divider. Amplifier A3 buffers this voltage and provides the necessary drive to the cable. This technique is limited in effectiveness by the frequency response of the op-amps, so it should be regarded as an option only at low frequencies.

34.10.4 Bifilar Choke

Most differential amplifiers are limited to the frequency regime below 100kHz. It is common for one or more radio-frequency signals to appear as a common-mode noise signal at the input of an amplifier. Amplitude modulated RF signals are detected in an amplifier nonlinearity (such as the base-emitter junctions of a BJT differential amplifier pair) and then appear as noise in the amplified output signal.

The *bifilar choke* is often effective against this type of high-frequency common-mode signal. The choke is inserted in the input leads as shown in figure 930(a).

The winding of a bifilar choke is shown in figure 930(b). The two signal-carrying conductors are wound in the same direction around a core. The differential current (the desired signal due

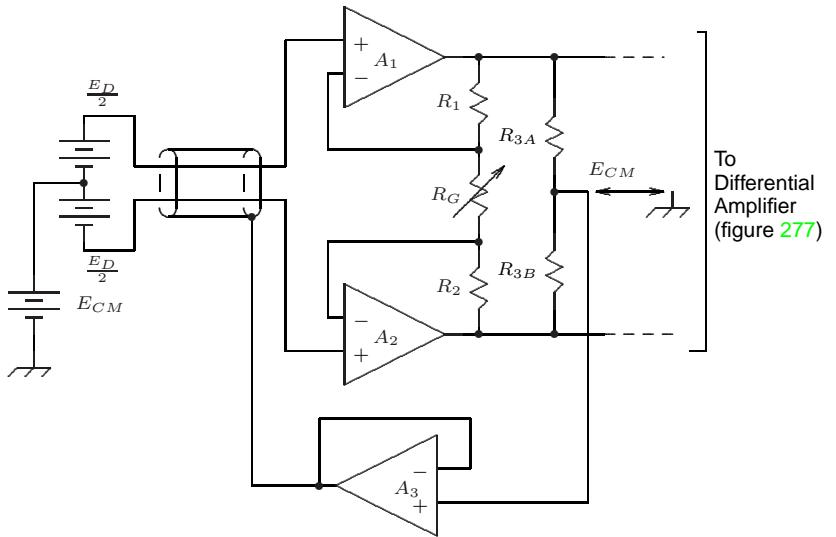
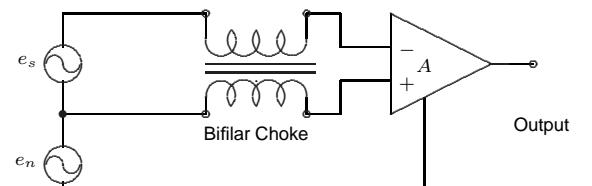
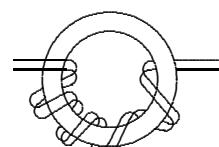


Figure 929: Guarding



(a) Circuit



(b) Winding

Figure 930: Bifilar Choke

to e_s) flows in opposite directions in the two conductors so their magnetic fluxes cancel and the choke has no inductance and no effect.

The common-mode current (the undesired signal due to e_n) is flowing in the same direction in both conductors. It generates a magnetic flux in the core and so the core has significant inductance and opposes the flow of this current.

In the presence of strong radio fields, as in the downtown of a large city, the bifilar choke may be useful in the *output* wiring of an amplifier. Many amplifier systems include a negative feedback connection between the output of the amplifier and the input. If there are long leads attached to the output of the amplifier (the speaker leads of an audio amplifier, for example), they behave as an antenna, bringing the RF signal to the negative feedback path and then to the input stage of the amplifier. The unwanted RF signal then becomes part of the signal path of the system.

If the amplifier behaved as an ideal voltage source at all frequencies, the output circuit of the amplifier would absorb the unwanted signal. However, the amplifier gain and the effectiveness of negative feedback are usually absent at radio frequencies.

34.11 EMC: Electromagnetic Compliance

Alternating currents in a conductor generate alternating electric and magnetic fields that launch radio waves – an *electromagnetic field* – into space, that is, the current and conductor behave as a *transmitting antenna*. The wiring of nearby electronic devices acts in a reverse fashion: these conductors intercept the electromagnetic field, this creates an alternating current in the conductor, and these conductors act as a *receiving antenna*.

If the received electromagnetic signal is sufficiently large compared to the normal currents in the receiving device, they can interfere with its operation. For example, personal computers and other digital devices can radiate electromagnetic signals of significant magnitude over a wide range of the frequency spectrum. To prevent interference with radio receivers and other electronic apparatus, the emitted signals must be reduced and contained or the receiving equipment must be made less susceptible to interfering signals²⁸⁸.

To minimize the social disruption that is caused by mutual interference between electronic devices, governments regulate the maximum allowable field strength of emitted electromagnetic waves. For example, in the United States of America, the allowable maximum emissions are specified in the Federal Communications Commission *Part 15 Rules* [302], [303]. In Europe, the corresponding standard is *EMC Directive 89/336/EEC*. In addition to specifying allowable emissions, the European standard specifies the required immunity of devices to external electromagnetic fields [304].

Other industries and organizations may have their own requirements. For example, the allowable emission levels for certain automotive manufacturers (eg, General Motors, with standard GMW3097) are substantially stricter than the FCC Part 15 Rules²⁸⁹.

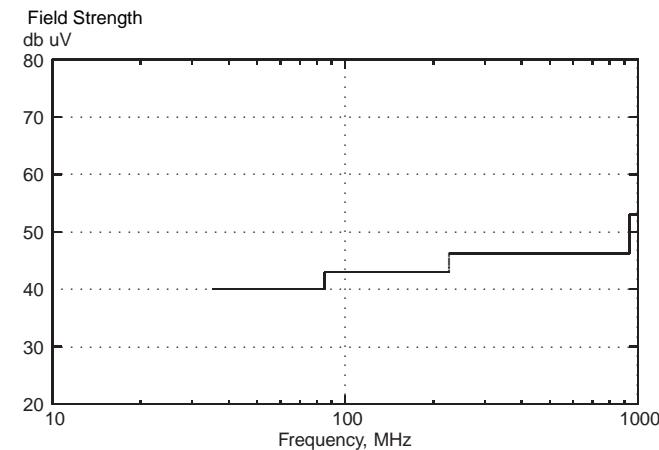


Figure 931: FCC Part 15 Class B Emission Limits

²⁸⁸In one case, a student using a pacemaker inadvertently placed himself in the beam of a laboratory microwave transmitter. Interference from the microwave signal stopped the pacemaker and the patient collapsed. Fortunately, this moved him out of the radio field, the pacemaker restarted again, and the person survived.

²⁸⁹Electronic devices in automobiles exist in close proximity to each other, and automotive radios are more sensitive than domestic radios.

The allowable emission limits for FCC Regulation Part 15 are shown in figure 931. There are two limits: the higher one, Class A, for industrial equipment. The lower limit, Class B, shown in the figure, is required for domestic equipment since television and radio receivers are likely to be in close proximity to electronic equipment in a home environment. In both cases, the equipment emission in microvolts/metre must not exceed the limits shown.

A calibrated spectrum analysis from the equipment under test can be overlaid on figure 931 to determine whether the equipment meets the compliance requirement.

34.11.1 EMI from Differential and Common Mode Currents

Both differential and common mode currents cause electromagnetic interference.

A differential mode current is a current that flows in a loop on the printed circuit board. The source current is balanced by the return current.

A common mode current is a current that is not balanced by a return current. Typically this occurs in a cable that is attached to the board, in which the outgoing and return currents are not completely balanced. In this situation, the common-mode current flows through stray capacitances associated with the cable, the chassis, and earth grounds.

The magnitude of the field due to differential current in a loop and common-mode current in a conductor may be shown to be:

$$E = \frac{\mu_o I A \pi f^2}{2rC} \quad (1589)$$

$$E = \frac{\mu_o I l f}{2r} \quad (1590)$$

where the variables are

- E Field strength, volts per metre
- μ_o Magnetic permeability of space, a constant, $4\pi \times 10^{-7}$
- I Current in the loop, amps.
- A Area of the circuit loop, square metres
- f Frequency, Hz
- r Distance of measurement from coil centre, metres
- C Speed of light, 3×10^8 metres/second
- l Length of the radiating conductor, metres

Example

The FCC Part 15 limit on emissions for a 'Part B' device, used in a residential location, are not to exceed $150\mu\text{V}/\text{metre}$ in the frequency band 88 to 216MHz, measured at a distance of 3 metres.

Determine the (a) maximum allowable (differential) current in a circuit loop 1cm x 1cm, and (b) the maximum allowable common-mode current in a cable 1 metre long.

Solution

(a) **Differential current:** The area of the circuit loop is 0.01 metres x 0.01 metres, or 10^{-4} square metres. According to equation 1589, the radiated field increases with frequency, so the worst case is at a frequency of 216 MHz. Rearranging equation 1589 to solve for differential current, we have:

This forces lower radiation limits.

$$I = E \left[\frac{2rC}{\mu_0 A \pi f^2} \right] = 150 \times 10^{-6} \left[\frac{2 \times 3 \times (3 \times 10^8)}{(4\pi \times 10^{-7}) \times (10^{-4}) \times 3.141 \times (216 \times 10^6)^2} \right] = 14 \text{mA} \quad (1591)$$

(b) **Common-Mode Current:** Equation 1590 indicates that the common-mode current increases with frequency, so the worst case frequency is again 216 MHz. Rearranging equation 1590 to solve for common-mode current, we have:

$$I = E \left[\frac{2r}{\mu_0 l f} \right] = 150 \times 10^{-6} \left[\frac{2 \times 3}{(4\pi \times 10^{-7}) \times 1 \times (216 \times 10^6)} \right] = 3.3 \mu\text{A} \quad (1592)$$

Notice that the same field strength is created with a much smaller common-mode current than differential current. This is an indication that *a cable attached to an electronic circuit has the potential of creating a substantial interfering electromagnetic field, even with very small amounts of common-mode current.* Common-mode currents should be measured with a current probe and if necessary reduced with a common-mode choke.

34.11.2 Common Mode Current, EMI, and the Bifilar Choke

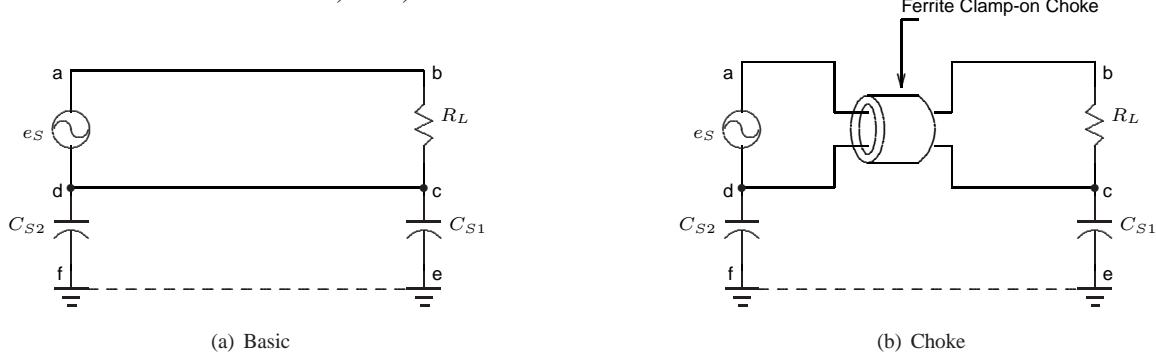


Figure 932: Common-Mode Output Signal

Common mode current, generated by digital circuitry, is a major source of EMI (electromagnetic interference). As an example, consider the circuit shown in figure 932(a), in which signal source e_S drives AC current through a load resistor R_L .

In an ideal world, all the current would flow around the smaller loop abcd. Since the conductors ab and cd can be located close together, the area of loop abcd and the radiated magnetic field are both minimized. However, the return conductor cd has a certain amount of resistance and inductance. As a consequence, some of the return current will flow via the stray capacitances C_{S1} and C_{S2} and the ground conductor ef. This creates a much larger loop abef which then radiates a significant interfering signal.

Since the ground current appears in the signal lead ab but not in the return lead cd, it appears as a common-mode current in the cable abcd. The magnitude of the common-mode current can be detected with a clamp-on current probe that encloses both conductors ab and cd.

Unfortunately, a very small common-mode current can cause unacceptable electromagnetic radiation. Ott [291] cites $5 \mu\text{A}$ as the maximum allowable common-mode current in a one-metre cable at a frequency of 50MHz, to meet EMI requirements.

One possible cure is shown in figure 932(b). A common-mode ferrite choke²⁹⁰ is clamped around the cable [305]. The differential currents in the conductor cancel any magnetic field, so the choke has no effect on those

²⁹⁰Also known as a *balun*, for *balanced to unbalanced* transformer. In this application, the operation is the reverse: to balance an unbalanced current.

currents. Any common-mode current does create a net magnetic field, and so the choke opposes that current. In effect, the choke forces the currents ab and cd to be equal, so that the common mode current in the ground path ef and the radiated interference is zero.

34.11.3 Differential Drive and Common-Mode EMI Current

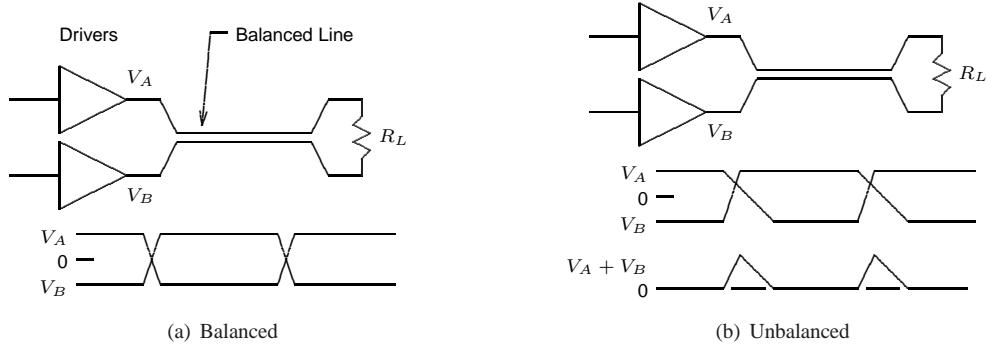


Figure 933: Unbalanced Differential Drive

Figure 933(a) shows another possible example of common-mode output current. The two amplifiers drive a balanced line in antiphase. The peak-peak voltage developed across the load resistance is twice the peak value of each amplifier output. The voltages are opposite and equal at all times, so the average voltage across the load is zero.

Now consider figure 932(b) where the amplifiers have a longer fall time than rise time. The voltages are not equal during the transition times. Triangular pulses of common-mode voltage appear at the load resistance. As in figure 932(a), this common mode voltage will drive current through stray capacitances and the ground path, creating an interfering electromagnetic signal.

Again, a coaxial choke will reduce this common-mode current.

34.11.4 Evaluating EMC

For a complete circuit board with many digital signals, it is not simple to predict the level of interfering electromagnetic field caused by differential currents in the various circuit loops. Digital signals radiate at a variety of frequencies and the resultant field is a vector sum of a large number of individual fields. The usual approach is to

- Construct the board in a manner that minimizes differential radiation, for example by keeping the area of circuit loops to a minimum [296], [306].
- Measure the overall radiated field, and if it exceeds the allowable level, then
- Determine the magnetic and electric fields at various points on the board to determine the *hot spots* that must be remediated.

On the other hand, a current probe can be used to measure the common-mode current in a cable, and thereby predict its contribution to the radiated electromagnetic field.

34.12 Electromagnetic Field Coupling

An electronic circuit may interact with an electromagnetic field as a source or as a receiver.

As a source, the circuit must not radiate a signal with sufficient field strength to interfere with other victim circuits. This is the EMC (electromagnetic compliance) requirement discussed in section 34.11.

As a receiver, an external electromagnetic field couples into the connecting wires of the circuit which act as receiving antennas. This field becomes an electrical signal in the circuit which disturbs the circuit operation. This phenomenon of an external radio signal disturbing the operation of a circuit is generally known as RFI for *radio frequency interference*.

For example, it is very common in metropolitan areas for television radio-frequency signals to be received by sound reinforcement audio equipment. The radio signal is demodulated by the audio circuitry and is heard as *sync buzz* in the output. An examination of the audio signal shows a demodulated video signal riding on the audio.

As another example, radio frequency interference appears in low-frequency measurement circuits as a variable DC output voltage. If a measurement circuit shows a DC output voltage that varies with the placement of the circuit or the presence of a human body, RFI should be investigated as a cause²⁹¹.

The susceptibility of an electronic circuit to external radio signals may be reduced by three measures:

- **Choose components that are less susceptible to radio-frequency signals.** The translation of a high-frequency RF signal to a DC signal occurs as the result of non-linearity in the active devices. There is some evidence that JFET's and MOSFET's are less susceptible to this RFI demodulation process. All other things being equal, a JFET or MOSFET op-amp is a better choice where RFI may be a problem.
- **Filter all leads connected to the circuit.** Radio-frequency signals tend to be common-mode in nature, so common-mode chokes are most effective. For example, in audio circuitry, radio signals tend to ride on the shield conductor of input signal cables, and thereby enter the sensitive preamplifier circuitry.
- **Enclose the circuit and its cables in a metallic shield.** An unbroken metal shield can attenuate an external radio signal by more than 100db, so this measure is very effective when it can be applied. *Shielded cables* [307] are widely used to reduce electrostatic and RF coupling into cables.

Shielding is compromised by running one or more cables into the enclosure – these must be filtered at the point of entry or shielded in such a way that the cable shield is an extension of the enclosure, as shown in figure 934(a).

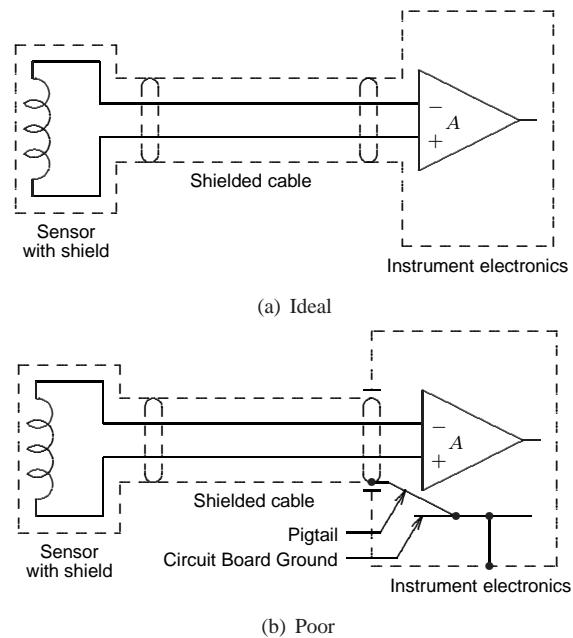


Figure 934: Shielded Enclosure

²⁹¹The other candidate for this behaviour is an unintended high-frequency oscillation in the circuit. A high frequency oscilloscope, properly connected to the circuit through $\times 10$ probes (section 34.13.4) should show whether oscillation is occurring. In some pathological cases, attaching a scope probe puts an additional capacitance into the circuit that stops the oscillation, and this adds to the diagnostic challenge.

An inferior arrangement is shown in figure 934(b). In this case, the cable shield is connected by a length of wire (the *pigtail*) to a ground on the circuit board and then by another length of wire to the enclosure. A radio-frequency signal will travel in common-mode form on the shield of the cable, onto the pigtail, and then via various stray capacitances into the instrument electronics. In figure 934(b) the cable shield should be directly connected to the enclosure shield. When the shield cannot be grounded in this fashion (because it violates the single-point grounding rule, for example), then the pigtail should be fitted with a radio frequency ferrite choke and bypassed to the enclosure by a capacitor. The capacitor is chosen so that it is a short circuit to RF interference and an open circuit to the desired (lower frequency) signals.

Any openings in the shield – to allow for cooling air or a display unit, for example – may also allow radio energy into the enclosure. To maintain the shielding effect, openings should be small compared to the lowest radio frequency wavelength. An approximate formula for the shielding effectiveness of an aperture [296] is:

$$K_{db} = 20 \log_{10} \left(\frac{\lambda}{2d} \right) \quad (1593)$$

where the variables are

- | | |
|-----------|--|
| K_{db} | Shielding effectiveness, db |
| λ | Wavelength of the interfering radio signal, metres |
| d | Largest dimension of the aperture, metres |

The wavelength is given by:

$$c = f\lambda \quad (1594)$$

where:

- | | |
|-----------|---|
| c | Velocity of an electromagnetic wave, 3×10^8 metres/sec |
| λ | Wavelength, metres |
| f | Frequency, Hz |

34.13 Measurement Technique

Ideally, our measurement instruments should show an accurate rendition of what is taking place in the circuit under test. However – and this is particularly true when troubleshooting noise problems – the observed measurement may be totally misleading and incorrect. There are two principal reasons:

- *The measuring instrument affects the operation of the circuit.* For example, the capacitance of the measurement probe cable may reduce the observed noise level in the circuit, or may cause the circuit to burst into oscillation.
- *Noise signals couple into the test leads of the measuring instrument.* For example, when you short-circuit the oscilloscope test lead to its ground, the measured signal should go to zero. In many situations it shows a substantial amplitude because one or more noise signals are coupling into the leads of the oscilloscope.

Great care must be taken to validate measurements – make sure they are real results and not the result of some error in the measurement technique.

This section shows various methods of connecting an oscilloscope into a circuit. The oscilloscope is the primary measurement tool for identifying noise problems, and the connection into the circuit is critical in making accurate measurements. The equivalent circuit of each of the following circuits, showing components in the hidden schematic, helps to identify strengths and weaknesses of the measurement technique.

34.13.1 Single Lead Connection

An example of poor measurement technique is shown in figure 935. This arrangement is often seen in a university electrical engineering lab, where test leads are in short supply. A signal source e_s is connected by a single test lead to an oscilloscope. Completion of the measurement loop is via the AC line ground connection of the source (via a power supply, for example) and the AC line ground connection of the oscilloscope. This setup will work after a fashion, but it has several problems:

- If there are significant currents i_{AC} flowing through the Power Line Ground connection, then there will be a 60Hz voltage signal v_{AC} between the two ground pins. This AC noise voltage appears in series with the signal, so it reduces the signal-noise ratio.
- The measurement circuit and the Power Line Ground form a loop with a large area. A magnetic field in the vicinity will generate a voltage in the loop and again, noise voltage appears in series with the signal and reduces the signal-noise ratio.
- The unshielded test lead functions as one electrode of a stray capacitance C_{stray} , coupling electric field signal e_n into the signal path. The degree to which this is a problem depends on the magnitude of the noise voltage, the stray capacitance, and the impedances of the source and load.
- The return path for signal current is extremely long, and so it will contain significant inductance. If the signal is a pulse waveform, then it will exhibit significant overshoot and ringing at the pulse edges. For pulse waveforms, this is a completely unacceptable arrangement.

34.13.2 Twin Lead Connection

Another common measuring setup is shown in figure 936(a). The oscilloscope is connected by a signal and ground lead to the circuit under test.

The test leads are connected to the oscilloscope via a BNC adaptor, figure 936(b).

The merits and problems of this circuit are:

- Figure 936(a) shows the circuit under test is isolated from ground, which is preferred since it prevents current in the Power Line Ground conductor from inducing noise into the circuit. Most lab power supplies

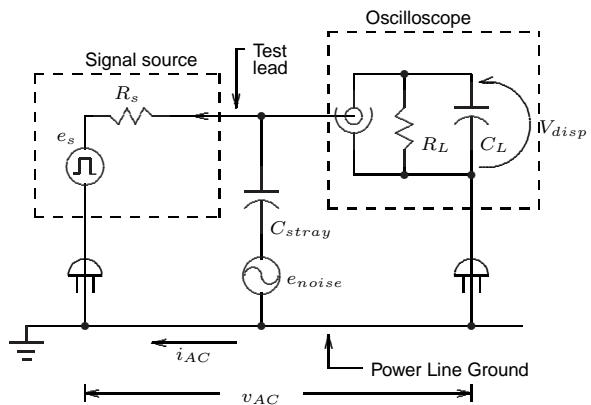
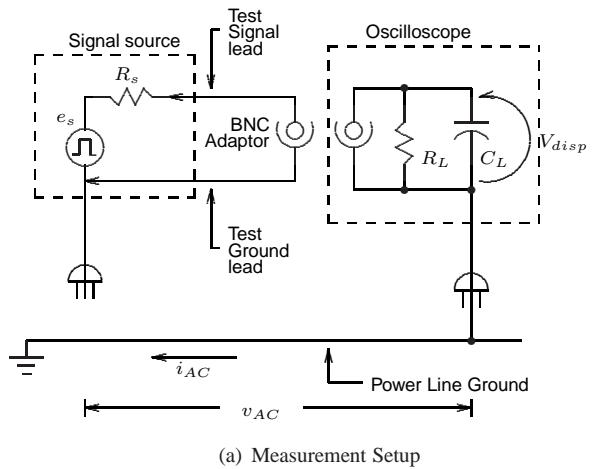
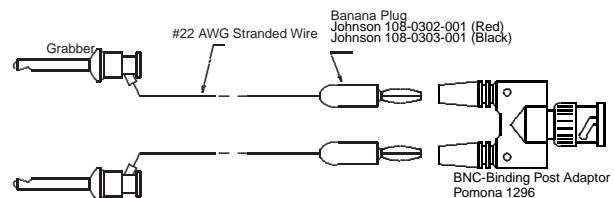


Figure 935: Measurement Using AC Ground



(a) Measurement Setup



(b) BNC Adaptor

Figure 936: Twin Unshielded Leads

are isolated from ground unless the positive (red) or negative (black) terminal is explicitly jumpered to the earth ground (green) terminal.

For this reason, the setup of figure 936(a) is an improvement over that of figure 935.

- To avoid clutter in the diagram, both the noise source e_{noise} and coupling capacitance C_{stray} are omitted from figure 936(a), but they are still present and potentially problematic.
- The inductance of the return path is reduced, but not enough to meet the requirement for accurate rendition of digital pulse waveforms.
- This is a very poor technique for assessing the noise level in a circuit. In addition to the possible electrostatic field pickup via stray capacitance into the unshielded signal lead, the signal and ground test leads form a large loop. A time-varying magnetic field will induce voltage into this loop.
- The input capacitance of the oscilloscope loads the circuit under test. Connecting this capacitance into the circuit under test can have a significant effect on the measurement. The internal resistance R_s of the source on the one hand and the cable capacitance plus the oscilloscope input capacitance on the other, form an RC lowpass filter that will shunt high frequencies to ground. This may show up as a degraded rise time in pulses.

In the frequency domain, the input impedance of the oscilloscope decreases at high frequencies, as illustrated by the following example:

Example

The equivalent input circuit of a certain oscilloscope is shown as $1M\Omega \parallel 50pF$. What is the input impedance of the oscilloscope at 1MHz?

Solution

$$X_c = \frac{1}{2\pi f C} = \frac{1}{2 \times 3.14 \times (1 \times 10^6) \times (50 \times 10^{-12})} = 3k2\Omega \quad (1595)$$

At 1MHz, the oscilloscope input impedance is only $3k2\Omega$. The $1M\Omega$ resistive component is irrelevant by comparison.

The input impedance of an oscilloscope decreases significantly at high frequencies.

The twin-lead oscilloscope connection is physically convenient and suitable only for non-critical measurement of analog and low-edge-rate digital signals where the signals are large compared to noise.

34.13.3 Coaxial Cable Connection

When noise pickup is a potential problem, the signal may be conveyed to the oscilloscope by means of a coaxial cable as shown in figure 937. This is also the arrangement when the oscilloscope is equipped with a so-called $\times 1$ oscilloscope probe.

- The coaxial cable shields against electrostatic noise pickup, so this source of error is reduced.

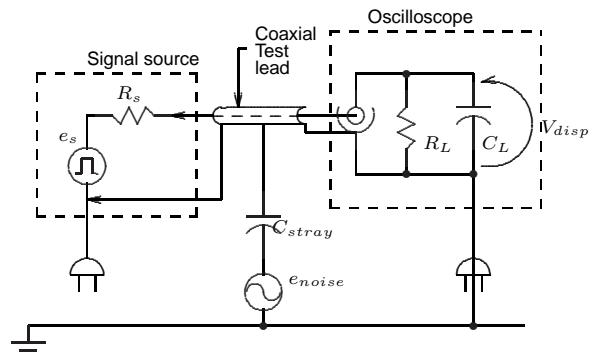


Figure 937: Measurement Using Coaxial Cable

- The ground return path is shorter than that of figure 935 and (probably) figure 936 so the transient response to a pulse is less likely to show overshoot and ringing.
- However, the coaxial cable has a certain amount of capacitance that appears in parallel with the input capacitance C_L of the oscilloscope. This further decreases the rise time of pulses and lowers the oscilloscope impedance at high frequencies.

34.13.4 $\times 10$ Probe Connection

An oscilloscope $\times 10$ probe reduces the loading on the circuit under test at the expense of the signal magnitude. Resistors R_P and R_L form a voltage divider, so that the signal is reduced by a factor of 10 and the input resistance increased from $1M\Omega$ to $10M\Omega$. The probe capacitance C_P together with the cable capacitance and oscilloscope input capacitance form a capacitive voltage divider. For the attenuation ratio to be independent of frequency, the impedances of the capacitors must have the same ratio as the resistances (section 4.15). The cable capacitance C_T is assumed to be $55pF$. It and input capacitance C_L are in parallel, so that section of the capacitive divider is $80pF$. Then the probe capacitance C_P is one ninth that value, $8.9pF$.

Without the $\times 10$ probe, the input capacitance would be $80pF$. With the $\times 10$ probe, the input capacitance is $8.9pF$ in series with $80pF$, or about $8pF$. The $\times 10$ probe reduces the input capacitance by a factor of 10. This reduces the loading on the measurement circuit, and as a result a $\times 10$ probe is less likely to affect the risetime of pulse waveforms.

Notice that the measurement ground point is connected to the oscilloscope probe, not to the oscilloscope. This is important, because it reduces the length of the ground return and the inductance in the measurement ground lead. Furthermore, the area in the loop formed between the probe (point A) and its ground (point B) is susceptible to pickup of any AC magnetic field. Consequently, this area should be minimized by keeping points A and B close together, and minimizing the length of the ground lead.

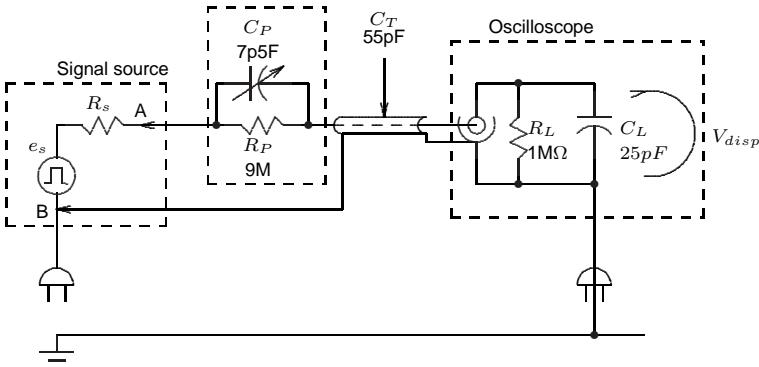


Figure 938: Measurement Using $\times 10$ Probe

34.13.5 Active Probe Connection

The $\times 10$ probe is useful for many routine analog and pulse measurements. However, where even the resistance and capacitance of a $\times 10$ probe cannot be tolerated, there is the *active probe*.

Typically, an active probe consists of a discrete component unity-gain amplifier, using a JFET or mosfet input device. In [113], Pease shows an active probe circuit with an input resistance of $10^{11}\Omega$ in parallel with a capacitance of only $0.29pF$.

On the downside, an active probe must be supplied with power, there is a requirement for additional connections to the probe and the maximum signal amplitude is limited to something in the order of $\pm 10V$.

34.14 Printed Wiring Board Layout

Complete books [29] have been written on printed wiring board layout, and so this section will necessarily mention only some of the highlights. We will assume that the circuit includes both analog and digital circuitry, and the

primary objective is to keep digital noise from corrupting analog signals. Recall that the rapid transitions of digital currents create noise spikes across wiring inductance, and these noise spikes may couple into the analog circuitry.

As an illustration of the difficulty involved, consider that the system being designed uses an 8 bit A/D converter with a 3 volt span. Then one step on the A/D converter is:

$$V_{step} = \frac{V_{span}}{2^N} = \frac{3}{2^8} = 11.7\text{mV} \quad (1596)$$

As a rough indicator, noise induced in the analog system previous to the A/D converter must be less than this amount.

Now consider that the digital system contains pulses of current equal to 20mA which transition in 10nSec. Using

$$e = L \frac{di}{dt}$$

an 11mV signal will be created by as little as 5.5nH, approximately the inductance of a 5 inch printed wiring trace. The problem grows more severe as the number of A/D bits increase and the edge rate of the digital logic increases. To prevent this problem:

- Separate the digital circuitry and analog circuitry into two distinct areas of the printed wiring board.
- Provide each digital integrated circuit with its own decoupling capacitor, located as close as possible to the device. The decoupling capacitors should have the lowest possible lead inductance.
- If a four-layer (or greater) circuit board is allowable, provide the digital logic with one plane dedicated to ground and a second plane dedicated to power supply positive voltage. The power and ground planes are on the outer layers, the signal lines run on the two inner layers. This has two benefits:
 - The inductance of the power supply and ground lines is at an absolute minimum. Rapidly transitioning currents in these layers will run in such a path that minimizes the inductance (just as current tends to choose the path of lowest resistance). This will then generate the smallest possible noise spikes. If allowed, the currents will mirror each other, with the source current running directly above the return current and the loop area at a minimum. Notice that any slot in the copper plane, that forces the current to take a circuitous path away from the mating current in the other plane, will raise the inductance of that path and the noise voltage.
 - Noise voltages generated by loops in the signal path are shielded by the ground and power plane, and this minimizes electromagnetic radiation interference.

- If a two-layer board is required, then provide separate power and ground leads for the analog and digital circuits. The power leads meet only at the power supply positive, and the ground leads only at the power supply ground. Ensure that there is adequate decoupling, so that noise on the digital supply and ground cannot couple into the analog supply and ground.

Low-frequency decoupling can be obtained by providing the digital and analog circuitry each with its own three-terminal regulator. However, integrated-circuit regulators do not function well at the frequency of digital noise pulses. High frequency decoupling requires ceramic capacitors, possibly in an RC or LC lowpass filter configuration.

- Assuming that power and ground lines must be run as separate traces on a two-layer board, route the traces to minimize the enclosed loop area [308], [309].

An example of a poor power trace layout is shown in figure 939(a). (The circuit is shown with 16-pin dual inline packages, for which pin 8 is GND , pin 16 is V_{CC} . The same principles apply to other package

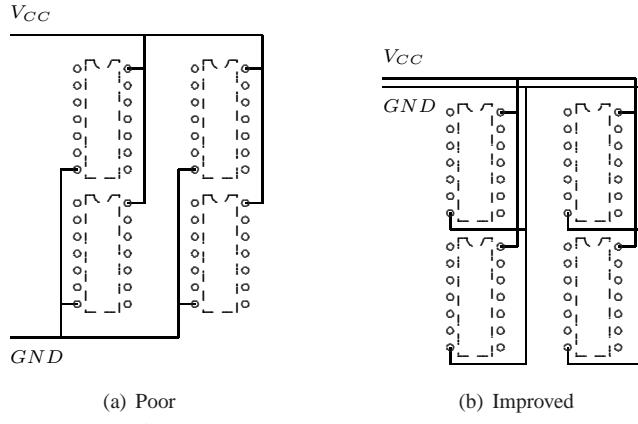


Figure 939: PCB Power Trace Layout

shapes.) The V_{CC} and GND lines are some distance from each other and enclose a large loop area. Consequently, the inductance of the leads will be higher than necessary, and this will create electrical noise.

Figure 939(b) shows an improved arrangement. This arrangement requires a double-sided board or wire jumpers where the traces cross. However, the supply and ground traces are much closer together and the enclosed area is a minimum, reducing electrical noise.

34.14.1 A/D Converter Layout and Grounding

Many instrumentation circuits consist of an analog section that conditions an input signal, followed by a digital system that provides various digital functions such as storage and display. Linking these two is an analog-digital converter.

The analog and digital sections must have a common ground. How should the analog-digital converter be connected to these circuits and the circuit common, in order to minimize the injection of digital noise into the analog circuit?

Some possible measures are:

- The noise problem increases in difficulty as the step size of the A/D converter decreases, so a 12-bit A/D converter is more likely than an 8-bit converter to be accompanied by noise issues. Use the least precise A/D converter that will satisfy the requirements. (Lower precision A/Ds are also significantly less expensive.)
- Integrating type A/D converters (single and dual slope, for example) are slow but inherently noise resistant. If speed is not a requirement, then these are a preferred choice.
- If an integrating A/D cannot be used, then limit the analog bandwidth to match the signal. Any excess bandwidth in the analog circuitry provides a window for noise.
- High-frequency clocks and high speed edges in the digital circuitry are more likely to cause noise injection into the analog circuitry. Use the slowest logic and lowest acceptable clock speed.
- If the budget permits, use a four-layer circuit board with power supply and ground dedicated to two of the layers. This reduces the impedance of the power and ground conductors and helps contain electromagnetic interference.

In addition to these measures, it is important to understand how the A/D converter should be connected to the supply and ground lines.

An equivalent diagram for a high-speed A/D converter [310] is shown in figure 940.

The analog section contains buffer amplifiers and comparator switches. The digital section may be pictured as a n-way collection of SPDT switches that connect the digital output lines to the digital supply DV_{DD} or DV_{SS} ²⁹².

Any transition on the digital supply line or digital ground line will couple through the stray capacitances C_{S1} and C_{S2} into the analog circuitry. Consequently, the DV_{DD} and DV_{SS} lines must be kept stable – or track the analog supply and ground.

Furthermore, when an A/D output logic level changes, the output voltage must charge or discharge the input capacitance of the subsequent logic load, shown as C_L in figure 940. These pulses of current must be supplied via the DV_{DD} or DV_{SS} power supply and ground lines, and may cause voltage excursions that couple through the stray capacitances C_{S1} and C_{S2} .

One possible arrangement of power wiring is shown in figure 941(a). This arrangement uses two separate supplies – one for analog circuits and one for digital circuits – and ties the two of them together at a single point. (In practice, the analog supply and digital supply would each be a 3-terminal regulator that derives power from some common power supply.) Unfortunately, spikes of current in the digital circuit loop, in the inductance of the ground line at A will generate spikes of voltage that will couple through stray capacitance C_{S2} into the analog circuitry. Furthermore, digital noise on the digital supply line at B will also couple through stray capacitance C_{S1} into the analog circuitry.

A better arrangement is shown in figure 941(b). The digital circuit and supply share no conductors with the analog section, so voltage spikes induced on the supply and ground lines are less likely to couple into the A/D converter. The analog and digital grounds of the A/D converter are tied together so voltage cannot be created across stray capacitance C_{S2} . One possible problem remains: the digital section of the A/D converter must source spikes of current into the load capacitance C_L , and this will tend to generate spikes across the analog supply line. The filter capacitor C_F provides a low-impedance source for these pulses of current and inductor L_F prevents transients at the DV_{DD} terminal from coupling into the analog section of the A/D.

Notice that the noise problem worsens with increasing load capacitance C_L . If many logic devices must be driven by the A/D logic output, logic buffers should be used to minimize the load capacitance seen by logic outputs of the A/D converter.

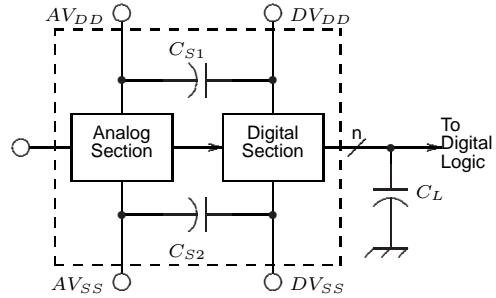


Figure 940: A/D Converter Internal Wiring

²⁹²The terms DV_{DD} and DV_{SS} are used when the A/D converter uses CMOS construction. DV_{DD} is the positive supply (+5 volts, for example) and DV_{SS} is the negative supply (usually the digital ground).

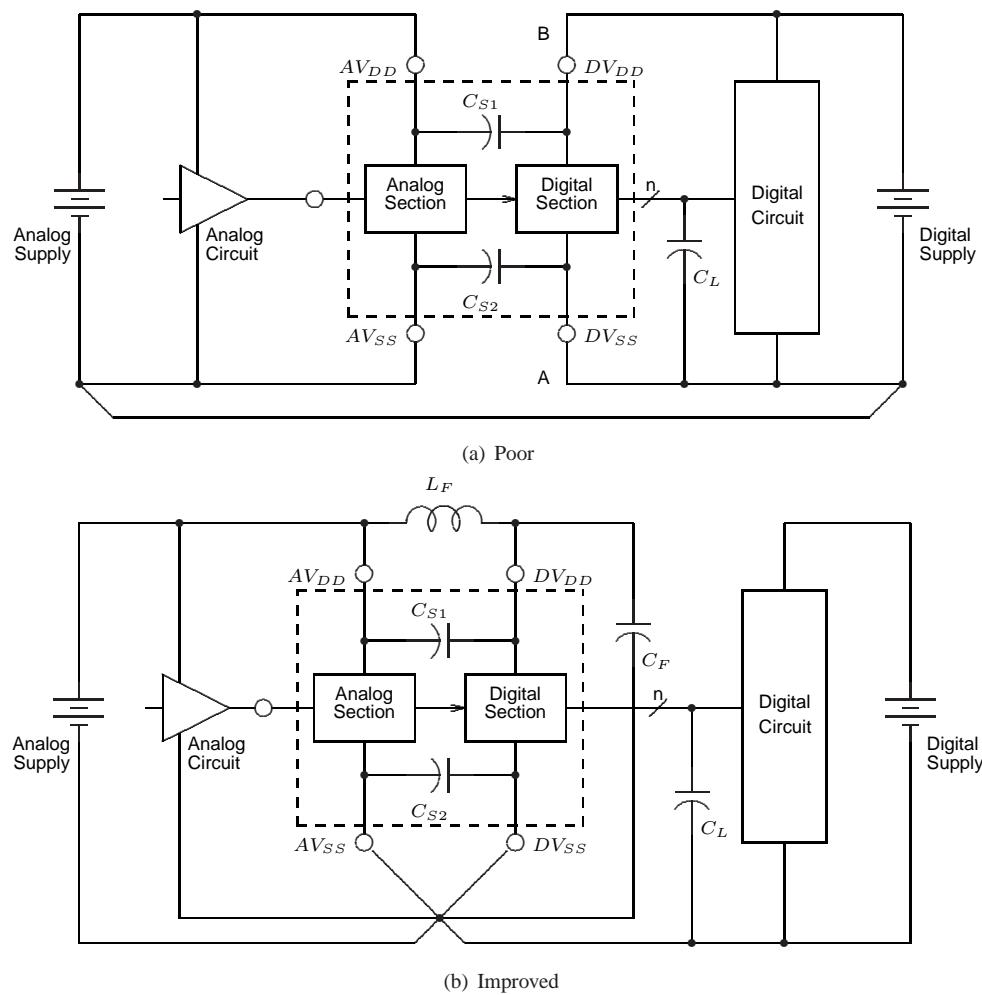


Figure 941: A/D Converter System Wiring

34.14.2 Ground Plane and the A/D Converter

We will use the term *ground plane*, as a generic phrase for *area of copper on the PC board*. Large conductive areas for ground and power supply are attractive because they lower the impedance of the supply and ground. Under the right circumstances, a ground plane can reduce electromagnetic emissions from the circuit board.

There are two schools of thought on the design of ground planes.

On the one hand, authors such as Banyai and Gerke [292] recommend splitting the ground plane into areas that isolate signals – between the analog and digital sections, for example. This technique is often recommended for the design of ground planes around an A/D converter [311], [312]. The analog circuitry and digital circuitry each have their own separate ground plane, joined by a narrow conductor at the A/D mecca ground point. This narrow bridge (figure 942) has a high impedance for high frequency noise, preventing it from transferring from the digital ground plane to the analog ground plane. Care must be taken that traces from the digital circuit do not overlap into the analog circuit area.

Ott [313] maintains that better control of emissions can be achieved with a single ground plane. The analog and digital circuits are each kept in separate areas, and care is taken that traces do not cross from one area into the other. Separate power planes (or traces) are used for the analog and digital supplies.

For high-resolution A/D conversion using multiple A/D converters, the situation is more complicated. All A/D converter grounds should attach to the mecca ground point, but this is physically impossible with multiple chips and a segmented ground plane. Ott cites this as a situation that requires a single continuous ground plane with careful component placement. Johnson [311] recommends segmentation and low-impedance ground, using external metal if necessary.

Simulation will eventually take the uncertainty out of this design issue. In the meantime, it is necessary to prototype a circuit, measure its performance, and be prepared to modify the design.

34.15 Case Histories: Solved

Here are the cures that were applied in the case histories of section 34 at the beginning of the section.

- *A certain car engine control system quits while close to a radio transmission tower.*

The engine control system was susceptible to external electromagnetic radiation. External radio signals were picked up on cables attached to the system, and then carried into the electronics. Operational amplifiers rectified the RF signal and created DC output levels that caused the system to malfunction.

It was a surprise to find that the metal body of the car was an insufficient barrier to these signals. However, there was an opening at the front of the car, filled by a plastic grille. The opening was many wavelengths at various strong RF signals, and so it provided an opportunity for entry of the offending RF signal. The cure: Place the engine control system in a metal box and ensure that every line going into the box has an RF filter.

- *A development engineer finds that her signal acquisition system is swamped in noise from the on-board microprocessor.*

The board layout had intermingled the analog and digital circuitry, and the grounds were co-mingled. The cure: re-lay out the board, moving analog and digital circuitry into separate sections, well away from each

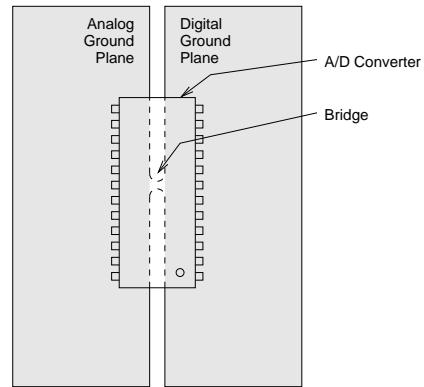


Figure 942: Segmented Ground

other. Organize the central ground point to be at the system A/D converter. Provide separate analog and digital grounds.

- *An airborne computer control system works fine on the bench when the modules are spread out. When the modules are installed in the case, noise from the switching power supply appears on the video output.*

Some exploration with electric field and magnetic field probes identified that the offending field was primarily electric in nature. (The switching power supply used a toroidal inductor, so the magnetic field was contained, as expected.) The cure: Place an electrostatic shield between the power supply and video circuitry. (As a temporary measure, the shield was simply a sheet of aluminum cooking foil glued to cardboard, then connected to the system ground by a wire lead.)

- *An industrial automation control system is being produced as a commercial product. When tested for compliance with emission regulations, it fails the test.*

Probing with a current probe, oscilloscope and spectrum analyser indicates a hot spot on the circuit board associated with the microprocessor clock signal. The clock lines are long and unterminated, and show evidence of ringing at high frequency. Cure #1: A terminating network is patched into the clock line to *damp down* the oscillation. This reduces the offending noise signal.

A measurement of the common-mode current on an external cable indicates that it is a major source of emissions. Cure #2: A clamp-on ferrite core is placed on the cable so that all the lines pass through the core of the cable. This reduces the noise emission to the point where the product is compliant with regulations.

- *An aerial photography camera is installed in an aircraft, powered from the 28VDC bus. Whenever the camera triggers, the aircraft navigation computer resets.*

Inspection of the aircraft wiring showed that the camera and nav computer were on the same electrical circuit. A bench test of the camera revealed that the starting current of the camera DC motor was in the order of 20 amps. This current surge created a large voltage drop across the aircraft wiring, a case of conducted interference. The cure: The camera connection was moved from the circuit that supplied the navigation computer and provided with separate wiring and circuit breaker back to the main 28VDC bus. A line filter was installed into the 28VDC power supply line at the camera to further contain transients in the area of the camera.

- *A film projection system is equipped with a magnetic head to read the sound track on magnetic tape. A nearby motor induces a power line interference signal in the magnetic head. The noise signal is stronger than the desired signal.*

Examination of the noise waveform with an oscilloscope shows a strong component at 60Hz and a smaller component at 120Hz. Probing the area with a search coil shows conclusively that the field is magnetic. The cure: Ideally, one would construct a metal shield (steel or mu-metal, to be effective at these frequencies) and enclose either the motor or pickup coil. Unfortunately, the physical arrangement and time constraints made this approach very difficult. Instead, two tunable notch filters were constructed, one for each of the offending frequencies, similar to the circuits shown in section 17.15.1 on page 553. These were inserted in the audio signal path and tuned to give maximum attenuation of the noise signals. There was a minor effect on the low-frequency component of the program material, but overall the result was quite satisfactory.

34.16 Exercises

1. A 20MHz square-wave clock pulse waveform has an amplitude of 3.3 volts with transition times of 5nSec. Calculate the spectrum amplitude and the break frequencies.

2. By what factor is the resistance of a trace reduced if its width is increased by a factor of 16?
3. The figure shows a driver circuit of the type described in section 34.11.3 (page 1044), coupled to the load resistance via a transformer.

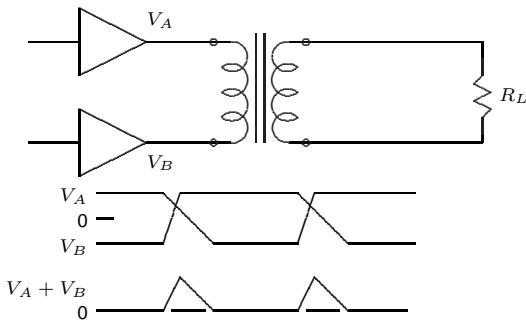


Figure 943: Unbalanced Differential Drive

When the common-mode current is measured in the cable driving R_L , it is found to be excessive.

- (a) Given that the transformer has a certain capacitance between its primary and secondary windings, use this and any other stray capacitances to explain why this common-mode current occurs.
- (b) It's suggested that the common-mode current could be reduced by using a transformer which has an electrostatic shield between primary and secondary. Would this work? Explain.
4. A certain oscilloscope has an input resistance of $1M\Omega$ in parallel with capacitance of $20pF$. A 2 foot long piece of RG-174 coaxial cable ($30.8pF$ per foot) connects the oscilloscope input into a test circuit.
 - (a) What is the effective capacitive load seen by the test circuit?
 - (b) What is the impedance of the load seen by the test circuit at a frequency of $100kHz$?
 - (c) If the internal resistance of the test circuit is $10k\Omega$, what is the rise time of the measurement circuit?
 - (d) The oscilloscope used for this measurement has a 3db bandwidth of $60MHz$. Assuming a $10k\Omega$ source resistance for the circuit, what is the 3db bandwidth of the measurement of this circuit?
 - (e) If a $\times 10$ probe circuit is constructed at the test circuit end of the cable, what will be the effective capacitive load on the test circuit?
5. Draw up a table showing the allowable amounts of common-mode current for the different frequency bands of the FCC Part 15 rules, Part B.
6. The example of section 34.11.1 on page 1042 calculated an allowable level of common-mode current in a cable. Assuming that this common-mode current is created by a 5 volt RMS AC signal driving current through a stray capacitance, calculate the magnitude of this capacitance. How does this value compare with the value of capacitor components?
7. A certain vendor of shielding products provides a copper strip with fingers that mount into corresponding holes in a printed circuit board. The shielding effect is specified as '70 db attenuation of a magnetic field at 200KHz'. The material thickness is 0.005 inches.
Could this level of attenuation be expected from ordinary copper sheeting, resistivity $\rho = 1.589\mu\Omega\cdot cm$, permeability $\mu_r \approx 1$? Explain.

8. In a laboratory setting, the temperature of a small oven is sensed by means of a thermistor. The resultant voltage signal is converted to digital format and recorded by a computer.

The temperature of the oven is controlled by a simple on-off thermostatic control. When the heater switches, the temperature recording system shows a large transient error.

- (a) Identify the possible sources of interference in this situation.
- (b) How would you determine which of these possibilities is the offending problem?
- (c) For each possible source of the interference, give the appropriate solution.

9. In pursuit of the lowest possible part cost in a switching power supply, the project supervisor has pointed out that a ferrite rod inductor is one-third the cost of an equivalent toroidal wound inductor. Why is a toroidal inductor a better choice, in spite of its higher part cost?

10. An airborne data acquisition system is being designed. It consists of inductive sensors which produce a low-level voltage signal, amplifier circuitry, analog-digital conversion, followed by a data-logging microprocessor system.

The system is being custom-built for a one-off experiment. The project deadline is very important. Cost is a secondary consideration.

The system circuit board can be built on one, two or four layers. Which would you recommend? Explain your choice.

The project leader has expressed a preference for a plastic enclosure in order to minimize the instrument package weight. What is your recommendation, and why?

35 Beyond the Op-Amp

35.1 Introduction

In this section we will examine alternative amplifiers, that is, amplifiers that function differently from the common-garden variety of operational amplifier.

35.1.1 Voltage-Mode and Current Mode Circuits

Electronic circuitry has evolved from the past. The first amplifying devices were electron tubes, which have a high input impedance and moderate output impedance. In an amplifier using tubes as the amplifying devices, the signals were predominately voltage-based, what we now refer to as *voltage mode*.

BJT's are a current-in, current-out device. When first invented in the 1950's, BJT's were substituted in tube-type circuits, with minimal modifications. After some time, designers created circuits that used BJT's more effectively, using currents rather than voltages to represent signals. These circuits are known as *current mode* [314].

Small signal MOSFET's are now widely used as the active device in electronic circuits. On the one hand, a MOSFET has high input impedance and moderate output impedance, so it is well suited to support voltage mode circuitry. On the other hand, the advantages of current mode circuits are now well known, so current mode techniques are also in use.

35.1.2 Why is the Op-Amp so Popular?

The op-amp – a voltage amplifier, with high input impedance and low output impedance – is far more popular than any of the alternative devices, and it's useful to consider why that is so.

- In *small-signal* electronic circuits, the focus of this text, it is most common to represent an electronic signal (the representation of information) as a voltage, rather than a current. Both voltage and current are present, but the resistances involved are such that the voltage is relatively large compared to the current. For example, a typical signal might consist of volts of potential accompanied by milliamperes of current. Furthermore, the voltage signal remains relatively undistorted in its passage through the circuitry and the current adjusts – in some cases in a non-linear fashion – to accommodate this undistorted voltage.
- A voltage measurement is non-invasive – it may be measured by attaching an instrument between two points in the circuit. For example, we typically measure voltages in a circuit with a high-impedance AC or DC voltmeter or a high-input impedance oscilloscope.

In contrast, measuring a current is invasive: it requires breaking the circuit and inserting the measurement instrument in series with the current or clamping a magnetic core around the current-carrying conductor ²⁹³.

Where signals appear in the form of a voltage, a voltage amplifier presents a high impedance to an input signal to reduce loading. Similarly, a voltage amplifier produces a signal from a low impedance, so it is not affected by a load resistance. In this arrangement, both the input and output signals can be measured with a voltage-sensing device, and so the circuit is straightforward to monitor and debug.

²⁹³Current may be measured non-invasively by sensing the magnetic field that accompanies the current, by means of a current transformer or hall-effect device. A current transformer measurement requires that the signal be alternating current. A hall-effect can measure both DC and AC currents and is an extremely useful device in applications where the current waveform is a different shape than the voltage waveform. However, hall-effect probes tend to be expensive and have limited sensitivity. Both the current transformers and hall-effect probe must be clamped around the current-carrying conductor, and that limits their usefulness in miniature circuits.

35.1.3 Why Consider Other Amplifier Architectures?

The operational amplifier is readily available and adaptable to a huge range of applications. Why consider any other architecture?

- The operational amplifier is a *voltage mode* device and the voltage swings inside the amplifier must charge and discharge any stray capacitances. As a consequence, it is necessary to trade off operating current for operating speed: high-speed voltage mode operational amplifiers with low power consumption are difficult to design.

Current mode circuits operate with low voltage swings and as a consequence parasitic stray capacitance has less effect. As a result, current mode circuits are inherently capable of higher frequency response²⁹⁴.

- In linear circuit applications, the op-amp is placed in a negative feedback loop. The open-loop gain exceeds the closed-loop gain, and then a number of benefits accrue. However, for stability the open-loop gain is generally rolled off at 6db/octave from a very low corner frequency, typically in the tens of hertz. As a consequence, the gain-bandwidth product is a constant at medium and high frequencies and there is a tradeoff between the loop gain and frequency of operation.

For example, a 1MHz GBP op-amp operating at 100kHz has a loop gain of 10, so the closed-loop gain (which must be much less than the loop gain) is very severely constrained. Some other amplifier architectures (the CFA, current-feedback amplifier) do not require this gain-bandwidth tradeoff – they have a flat amplitude, irrespective of the closed-loop gain, up to the maximum operating frequency.

- Some amplifier architectures have useful features. For example, the OTA (operational transconductance amplifier) can function as a gain-controlled amplifier or two-quadrant multiplier in audio applications.
- In active filter designs, non-traditional amplifier designs may reduce the number of passive components and/or support electronic adjustment of the filter resonant frequency [315] [316] [317].

35.1.4 Other Types of Amplifiers

Although the voltage amplifier is most common, it's a useful exercise to tabulate other combinations of input and output resistance, and consider their applications.

Input Impedance	Output Impedance	Amplifier Name
Low	Low	Transresistance (Norton) amplifier
Low	High	Current conveyor
High	Low	Voltage amplifier (Op-Amp)
High	High	Transconductance amplifier (OTA)

Figure 944: Amplifier Classification

It is quite possible to synthesize these devices from one or more operational amplifiers. For example, a transresistance amplifier (also known as a *current-voltage converter*) can be created with an operational amplifier

²⁹⁴Another way of stating this: in a voltage mode circuit, the circuit impedances are large and even small stray capacitances tend to adversely affect high frequencies as a result. In a current mode circuit, the circuit impedances are lower and so the same order of magnitude of stray capacitance has less effect on frequency response. One might expect that stray inductance would have an analogous effect in a current mode circuit to the stray capacitance in a voltage mode circuit, but that turns out not to be the case.

and a single feedback resistor (section 13.9 on page 345)²⁹⁵. However, these circuits are then restricted by the limitations of the operational amplifier.

In this section, we describe devices that operate in the described condition as their native configuration and may therefore obtain certain advantages.

35.2 The Current Conveyor

Of the four amplifiers listed in figure 944 (page 1059), the current-conveyor is the odd one out. It's never been implemented as a commercial integrated circuit, and its main application is as the subject material of academic papers. However, the ideas have been incorporated into other circuits, such as the current feedback amplifier²⁹⁶.

The basic theory is in [318]. Some applications are in [319] and [320].

The circuit of a current-conveyor, redrawn from [320], is shown in figure 945(a).

The current conveyor is essentially a unity gain current-in current-out device. Input Y is a high impedance input that establishes a voltage at input X . Terminal X is a low impedance node that produces or accepts a current. This current reappears at the output current source Z .

35.2.1 Example

A simple application is shown in figure 945(b). Input voltage v_i transfers across from the high impedance input port In_Y to the low impedance input port In_X . This voltage drives a current $i_a = v_i/R_1$ through R_1 .

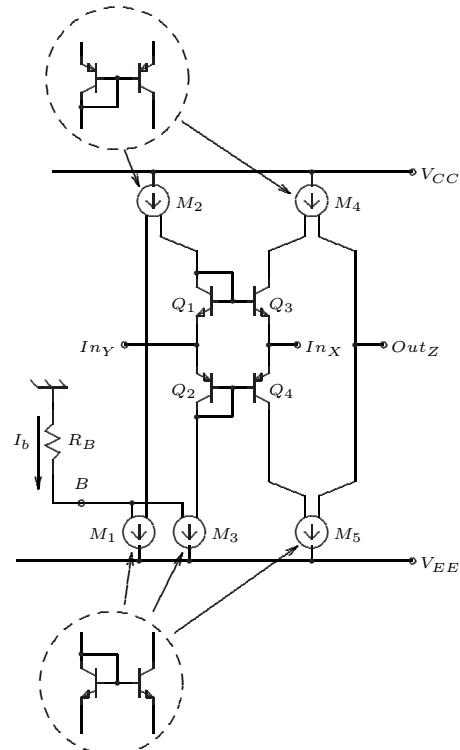
The conveyor mechanism transfers current $i_o = i_a$ to the output, where it flows through resistor R_2 and creates an output voltage v_o .

Working backwards from the output:

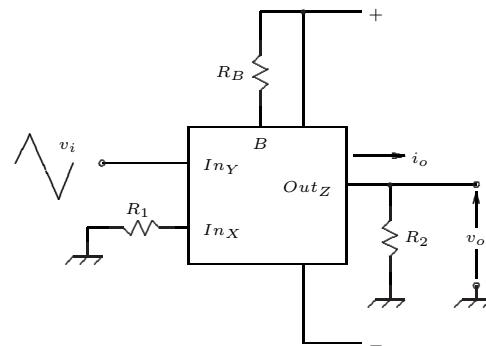
$$\begin{aligned} v_o &= i_o R_2 \\ i_o &= \frac{v_i}{R_1} \\ \frac{v_o}{v_i} &= \frac{R_2}{R_1} \end{aligned} \quad (1597)$$

That is, the voltage gain is equal to the ratio of the two resistors.

Since the output port Z is a current source, the output voltage is completely dependent on whatever resistance is attached to it.



(a) Internal Schematic



(b) Application

Figure 945: Current Conveyor

²⁹⁵In fact, it may be shown that the four amplifier types of the previous table may be combined with four types of feedback to yield sixteen different amplifier-feedback configurations [107].

²⁹⁶The schematic diagram of figure 952 (page 1066) shows the input circuitry of a current feedback amplifier, which is similar to the current conveyor of figure 945 above.

As a practical matter, the circuit would probably require an impedance buffer (such as an op-amp follower) to drive any subsequent circuitry.

35.2.2 Internal Circuit Operation

For simplicity in studying the internal circuitry, assume the current conveyor is configured as in the example of figure 945(b).

Now referring to figure 945(a), the device bias current is set by an external resistor R_B . This current then reflects through a series of current mirrors M_1 through M_3 so that transistors Q_1 and Q_2 carry the same bias current. Diode Q_1 and transistor Q_3 form a current mirror, so that the collector current of Q_3 is also equal to the bias current. Diode Q_2 and transistor Q_4 act in a similar fashion.

Mirrors M_4 and M_5 then reflect these currents.

Diode-connected transistors Q_1 and Q_2 cancel out the base-emitter voltages of transistors Q_3 and Q_4 , so that any voltage appearing at input terminal Y appears also at terminal X . Assuming the transistor base currents are negligible, no current flows into node Y .

Now consider that some current ΔI flows out of terminal X through R_1 . This current has no effect on the current through Q_4 , but it increases the current in Q_3 . Mirrors M_4 and M_5 then reflect these currents. At node Y , the current into the node is $I_B + \Delta I$. The current out of the node is I_B . So current ΔI must flow out of node Y through R_2 . The resistance at node Y must be such that the generated voltage does not cause mirror M_4 to saturate.

A positive voltage at input port Y causes Q_3 to drive current through resistor R_1 at the port X . A negative voltage causes Q_4 to drive current through R_1 .

Is this an amplifier? In other words, does it have power gain? If there is a large input impedance looking into terminal Y , then the input current is near zero and the input power P_i is zero. The output power is the product of the voltage and current in resistor R_2 , which is non zero. So yes, the device has power gain and is an amplifier.

35.2.3 Application

A negative-feedback op-amp based differentiator circuit was shown in section 13.12 (page 351). The closed loop gain rises at 20 db/decade. If this characteristic is allowed to intersect the open-loop gain, then the circuit may be unstable. (section 24).

The current-conveyor differentiator shown in figure 946 does not have this problem, because it operates open-loop, without negative feedback.

If the voltage at the Y input terminal is v_i , then this appears at the X terminal, across the capacitor. The current in the capacitor is

$$i_c = C \frac{dv_i}{dt} \quad (1598)$$

This current then appears at the output terminal Z as i_o . Resistor R translates this current into a voltage $v_o = i_o R$. Putting all this together we have:

$$v_o = RC \frac{dv_i}{dt} \quad (1599)$$

The output voltage is the differential of the input.

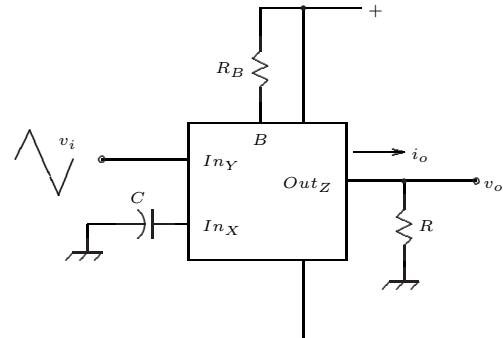


Figure 946: Current Conveyor Differentiator

35.3 The Current Feedback Amplifier

For many years of their history, op-amps were low-frequency devices, with a typical slew rate of 1 volt/ μ Sec and bandwidth of 1MHz. As a consequence, their operation was limited to frequencies between DC and some tens of kilohertz.

With the advent of digital video, there arose a need for higher frequency amplifiers. Semiconductor companies such as Comlinear and Elantec then developed the CFA (Current Feedback Amplifier) to address that market.

Op-amp speed had been limited by the slow speed of integrated circuit PNP transistors. With the invention of *dielectric isolation* [321], PNP transistors became as fast as NPN devices and an IC op-amp could contain both NPN and PNP devices without a speed penalty.

This enabled new, faster op-amp architectures such as the Current Feedback Amplifier described here.

The National Semiconductor LM6181, a typical CFA, has a slew rate of 2000 volts/ μ Sec and a bandwidth of 60MHz. Unlike the VOA, the CFA is usable up to a large fraction of its bandwidth.

As usual in this industry, CFA's were initially quite expensive but prices subsequently dropped to those of a commodity op-amp.

The slew rate and bandwidth of the CFA are a significant advantage in many applications. However, they are not a direct replacement for op-amps.

- The configuration of resistors around a current feedback amplifier appears to be identical to that of a traditional op-amp circuit. The non-inverting op-amp of figure 250 on page 319 is apparently identical to the non-inverting CFA amplifier in figure 947 above.

As well, the voltage gain equation is the same:

$$G = \frac{e_o}{e_i} = 1 + \frac{R_1}{R_2} \quad (1600)$$

For the op-amp, the voltage gain of the op-amp circuit depends on the ratio of R_1 and R_2 and their absolute values are not critical.

For the CFA this is not the case. The voltage gain is given by the ratio of these resistors as before. However, the value of the feedback resistor R_1 affects the stability and transient response, and must be chosen within a narrow range.

- Feedback capacitance usually helps stabilize an op-amp circuit. It has the opposite effect in a CFA circuit.
- The bias current, noise voltage and noise current of a CFA tend to be larger than the same parameter for an op-amp. For example, the bias current of the LM6181 is in the order of $5\mu A$, an order of magnitude larger than many op-amps.
- The two input terminals of an op-amp are identical, with similar input resistance and bias current. The two input terminals of the CFA are *very* different: the non-inverting input is a high-impedance input for voltage signals, the inverting terminal is a low-impedance input for currents.

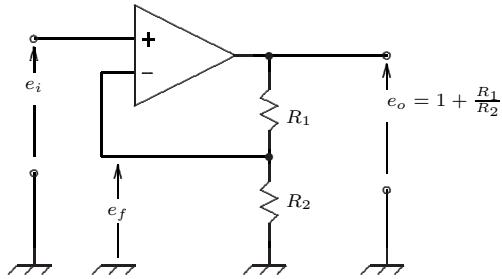


Figure 947: CFA Non Inverting Amplifier

35.3.1 Basic Operation

The basic operation of the CFA is illustrated in figure 948.

Internally, the CFA amplifies the current i_N in the non-inverting terminal by a factor T to produce the output voltage e_o .

$$e_o = T i_N \quad (1601)$$

The dimensions of the gain T are *volts/amp*. To indicate that this is a measure of gain – and not a common-garden variety resistance – the gain is known as *transresistance* (*transfer resistance*), measured in ohms.

The current i_N is the current equivalent of the error voltage that appears between the input terminals of an op-amp. A properly functioning negative feedback system around an op-amp reduces this error voltage to a very small value, which may be regarded as zero for purposes of calculation²⁹⁷.

In the case of the CFA, the value of transresistance gain – the open-loop gain – is very large. As a result, negative feedback minimizes the value of the error current i_N .

Assuming that i_N is small enough to be ignored, the circuit analysis is very simple. The input voltage e_i appears across resistor R_2 . This establishes a constant current in i_2 in R_2 . If the error current is small, then the current in R_1 is the same.

$$i_2 = \frac{e_i}{R_2} \quad (1602)$$

$$i_1 \approx i_2 \quad (1603)$$

$$\begin{aligned} e_o &= i_1 R_1 + i_2 R_2 \\ &= i_2 (R_1 + R_2) \end{aligned} \quad (1604)$$

Use these three equations to solve for e_o/e_i and the output voltage is as given in equation 1600 above.

Now suppose some disturbance increases the value of the output voltage. That drives more current through R_1 . Because the current in R_2 is fixed by e_i , it cannot change. Consequently, an increase in current must flow into the non-inverting input terminal. This new current is amplified by the transresistance gain and drives the output voltage back downward. The net result is to maintain the output voltage of equation 1600.

In fact, the non-inverting and inverting terminals are not connected together by the wire shown in figure 948. There is a unity-gain buffer between them such that the non-inverting input terminal sees a high impedance input. The buffer forces the two input terminals to be at the same voltage, and ensures that the inverting input terminal sees a low input impedance.

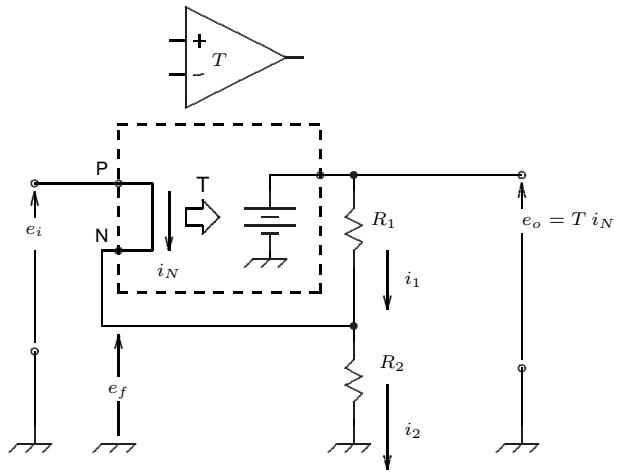


Figure 948: CFA Equivalent Circuit

²⁹⁷It's not quite zero: there must be some error voltage to produce an output voltage.

35.3.2 Gain Analysis

A more detailed analysis of the gain yields further insight. Again referring to figure 948, we can write the following equations:

The input buffer forces the voltages to be equal at the non-inverting and inverting inputs.

$$V_N = V_P = e_i \quad (1605)$$

Summing currents at the inverting input, we have:

$$i_N = \frac{V_N}{R_2} - \left(\frac{e_o - V_N}{R_1} \right) \quad (1606)$$

The transimpedance gain T multiplies the error current to produce the output voltage.

$$e_o = T i_N \quad (1607)$$

Substitute for i_N from equation 1606 in equation 1607 and after some manipulation we can obtain:

$$e_o \left(1 + \frac{T}{R_1} \right) = T V_N \left(\frac{1}{R_2} + \frac{1}{R_1} \right) \quad (1608)$$

This simplifies considerably for the condition that $T \gg R_1$. This condition is analogous to the op-amp negative feedback requirement that the open-loop gain be much larger than the closed-loop gain.

Further substitute e_i for V_N from equation 1605 into equation 1608, and simplify to obtain:

$$\begin{aligned} \frac{e_o}{e_i} &= R_1 \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \\ &= \left(1 + \frac{R_1}{R_2} \right) \end{aligned} \quad (1609)$$

as we obtained previously.

35.3.3 Loop Gain

The loop gain is obtained by tracing around the loop from the output terminal, through the feedback device (R_1 in this case), through the forward transresistance gain T to the output terminal again.

The sensor gain B is equal to the feedback current i_N per unit of output voltage e_o , or $1/R_1$. Then the loop gain is

$$\text{Loop Gain} = \text{Forward Gain} \times \text{Feedback Gain} = \frac{T}{R_1} \quad (1610)$$

Notice that Loop Gain is dimensionless.

Figure 949 shows the characteristic of Transresistance vs Frequency for the LM6181 current feedback amplifier. This is analogous to the frequency plot of open-loop gain A_{ol} of an op-amp. As in section 23.4, page 696, we can plot the value of the closed-loop gain $1/B = R_1$ on this graph to identify the loop gain and useful frequency range of operation.

Notice that transresistance is given in db. For example, at low frequency, the transresistance gain is about 130db Ω . That is

$$\begin{aligned} 20 \log_{10} T &= 130 \\ T &= 10^{130/20} \\ &= 3.16 \times 10^6 \Omega \end{aligned}$$

35.3.4 Maximum Frequency of Operation

Suppose we require a non-inverting amplifier circuit with operation up to 10MHz. What is the maximum allowable value of R_1 in figure 948?

According to figure 949, a frequency of 10MHz corresponds to a transresistance of 70db or so. Then the corresponding resistance is $T = 10^{70/20} = 3162\Omega$

The feedback resistance R_1 must be less than or equal to this value. Put another way, the loop gain is unity at a frequency of 10MHz when $R_1 = 3162\Omega$.

Notice that figure 949 shows the CFA transresistance. The feedback factor R_1 can be plotted on the same graph, and the intersection of the transresistance and feedback factor shows the maximum frequency of operation. However, figure 949 is *not* a plot of voltage gain vs frequency.

To summarize, *the voltage gain of the non-inverting CFA amplifier can be set independently of the bandwidth*. Referring to figure 947 on page 1062, the bandwidth is established by the value of R_1 . The voltage gain is established by the ratio of R_1 to R_2 , according to equation 1600.

35.3.5 Transresistance Equivalent Circuit

Figure 950 adds some more details to the equivalent circuit. There are two buffer amplifiers, each of gain $\times 1$. These present a high impedance (open circuit) at their input and low impedance (voltage source) at their output, and have a voltage gain of unity.

The transresistance gain is now shown as a resistor R_T which is force-fed the current i_N . Capacitance C_T rolls off the transresistance gain above the corner frequency

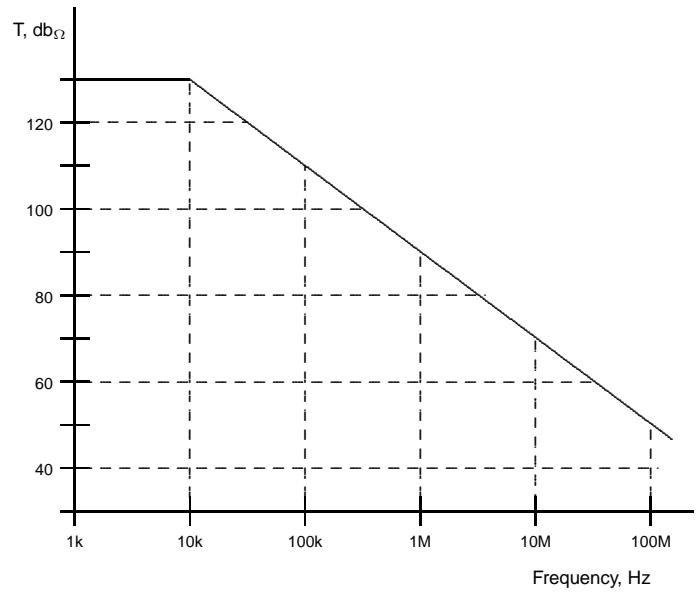


Figure 949: CFA Transresistance Gain

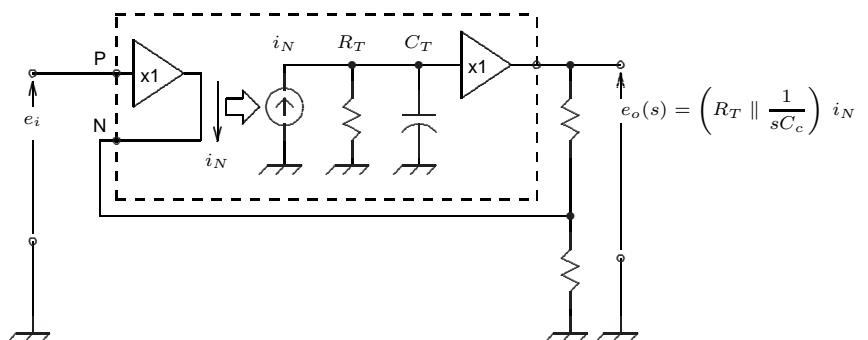


Figure 950: CFA Equivalent

$$\omega_o = 2\pi f_o = \frac{1}{R_T C_T} \quad (1611)$$

For the CFA characteristic of figure 949 the values of these parameters are $R_T = 3.16 \times 10^6 \Omega$ and $f_o = 10\text{kHz}$. Then equation 1611 gives $C_T = 5.04\text{pF}$.

35.3.6 Slew Rate

Why is the CFA so much faster in slew rate than an operational amplifier?

The slew-rate equivalent circuit of an op-amp is shown in figure 951. The input signal (aka *error voltage*, in a negative feedback system) appears as a voltage between the bases of the two transistors in the differential pair. The effect of this voltage is to direct the tail current I_{EE} to one side or the other of the pair (figure 826 on page 936). As a consequence, the maximum current into the capacitor is I_{EE} and the maximum slew rate

$$SR = \frac{I_{EE}}{C_c} \text{ volts/second} \quad (1612)$$

The tail current should be low to minimize the differential transistor base currents and this limits the maximum slew rate in this architecture.

The slew rate equivalent circuit of a current feedback amplifier is shown in figure 952. The input voltage appears between the V_P and V_N terminals. For example, a large positive voltage appears across the base-emitter junction of Q_2 , causing a large current i_1 . At the same time, the base-emitter junction of Q_4 is reverse biased, so it shuts off, reducing i_2 to zero.

The current i_1 is then reflected by the current mirror and charges the transimpedance capacitance C_T . There are limits to the maximum value of this current, but it can be many orders of magnitude larger than the capacitor charging current in a conventional op-amp. As a result, the slew rate is much higher.

The amplifier transresistance is the input impedance of the unity-voltage gain output buffer amplifier. The buffer amplifier is a series of NPN and PNP voltage followers similar to the input stage. As a result, the value of R_T may be somewhat dependent on the load resistance seen at the output terminal.

35.3.7 References

Further information on the theory and application of the current feedback amplifier is in references [139], [322], [323], [324], [325].

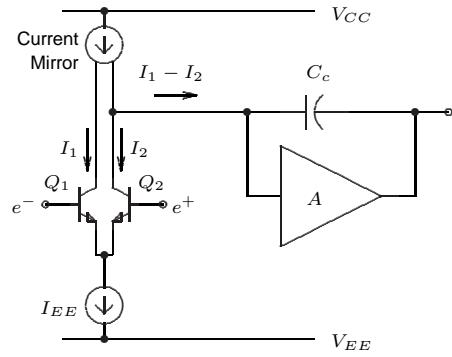


Figure 951: Op-amp slew rate model

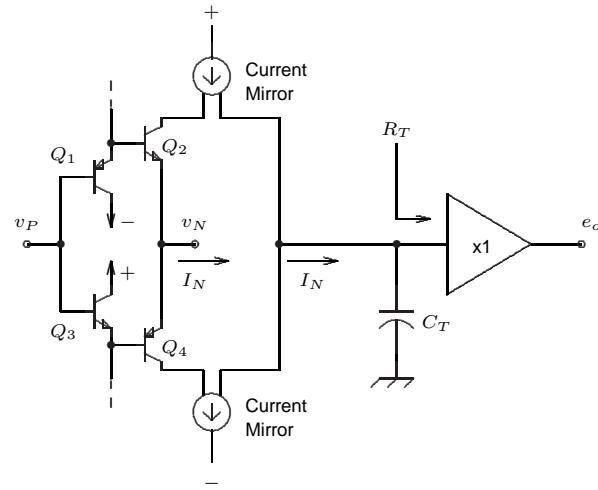


Figure 952: CFA slew rate model

35.4 The Norton Amplifier

In 1972, circuit designers at National Semiconductor invented an alternate form of the operational amplifier – one that depended on input *currents* rather than voltages. For that reason, it was referred to as a *Norton* amplifier.²⁹⁸

At that time, most operational amplifiers required a positive and negative voltage supply. The input common-mode voltage range and output voltage range were required to be significantly less than the total available range between power supplies. This was a significant problem at low supply voltages.

In contrast, the Norton amplifier design operated from a single supply and the output voltage could approach the power supply rails. As well, the simple circuitry of the amplifier permitted four amplifiers to be included on one low-cost integrated circuit.

The National LM3900, a quad Norton amplifier [326], was popular for a period of time. However, the Norton amplifier requires additional understanding to use in circuit designs and it is not well suited for DC-coupled applications. Quad single-supply op-amps such as the LM324 became available at about the same time [327] and eventually exceeded the Norton amplifier in use. Nonetheless, it is a useful exercise to understand the operation of this device.

35.4.1 Norton Internals

A simplified version of the Norton amplifier internal circuitry is shown in figure 953(a).

Transistor Q_1 is the amplifying device, a grounded emitter stage with a constant-current load I_1 to create a large voltage gain. Transistor Q_2 is an emitter follower so that the output of the amplifier is a low impedance source. Current load I_2 biases the emitter follower²⁹⁹. Consequently, the base terminal of Q_1 is the inverting input of the amplifier.

Transistor Q_3 and diode D_1 function as a current mirror, to siphon away current from the inverting input. A positive input voltage at this terminal causes the output voltage to rise, so this is the non-inverting input terminal.

35.4.2 Equivalent Circuit

The Norton amplifier equivalent circuit is shown in figure 953(b).

The input circuitry subtracts the inverting and non-inverting input currents I^- and I^+ , generating an *error current* I_e . The remaining circuitry amplifies this current and converts it to an output voltage, with a transresistance gain R_m volts/amp. Notice the similarity between the operational amplifier of figure 242 on page 315, and the Norton amplifier of figure 953(b). Like the op-amp, the Norton amplifier can be used as the basis for a negative feedback system, where the reference and feedback signals are currents.

²⁹⁸There was no suggestion that the conventional op-amp should be referred to as the Thevenin version, although that would have been consistent nomenclature.

²⁹⁹The actual circuit includes an additional stage of current gain.

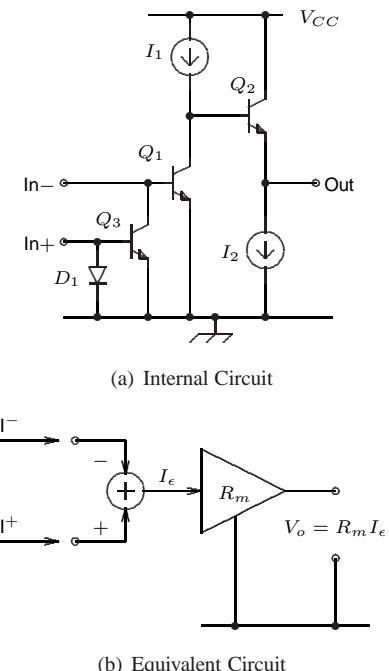


Figure 953: Norton Amplifier

In the op-amp inverting amplifier negative feedback configuration, we found it useful to recognize that the output of the amplifier goes to whatever voltage is necessary to equalize the two input voltages.

In the analogous Norton amplifier circuit, *the output rises to whatever voltage is necessary to equalize the two input currents*.

35.4.3 Bias for AC Amplification

Figure 954(a) shows another version of the equivalent circuit. The amplifier is set up for AC amplification with the output biased at half the supply voltage.

Assume that the base-emitter voltages are small enough compared to the supply voltage to be ignored. Then, referring back to figure 953(a), the non-inverting input accepts a current of $V_{CC}/2R$ amps. This turns on Q_3 which starves Q_1 of base current, and it begins to turn off. As Q_3 turns off, the output voltage rises. This drives more current through the feedback resistor R . The output voltage continues to rise until the currents are equal in both inputs.

Because the bias resistor $2R$ is double the value of the feedback resistor R , the output voltage will rise to half the supply voltage in order to equalize the two input currents. That is, the output voltage is biased to half the supply voltage and it can produce an AC output voltage of $V_{CC}/2$ volts peak before clipping.

35.4.4 AC Amplifier

Figure 954(b) shows the previous circuit, modified to amplify an AC signal e_i . Capacitor C_1 prevents the DC voltage at the inverting input terminal from driving current through the input voltage e_i . Capacitor C_2 prevents the DC voltage at the output of the amplifier from flowing through a load resistance. Both capacitors are chosen to be a low impedance at the lowest frequency of operation³⁰⁰. As in the case of the non-inverting op-amp circuit (section 12.4) the overall voltage gain of the circuit may be shown to be

$$A_v = \frac{e_o}{e_i} = -\frac{R_f}{R_i} = -\frac{R_2}{R_1} \quad (1613)$$

35.4.5 Discussion

The Norton amplifier can be used to implement a very wide variety of circuit building blocks [326]. However, the gain of the current mirror (D_1, Q_3 in figure 953(a)) can vary between 0.9 to 1.1 amps/amp, [328] and this limits the precision of DC amplifier circuits. As in the case of all single-supply amplifiers, there is no power supply rejection on the negative (ground) power supply terminal.

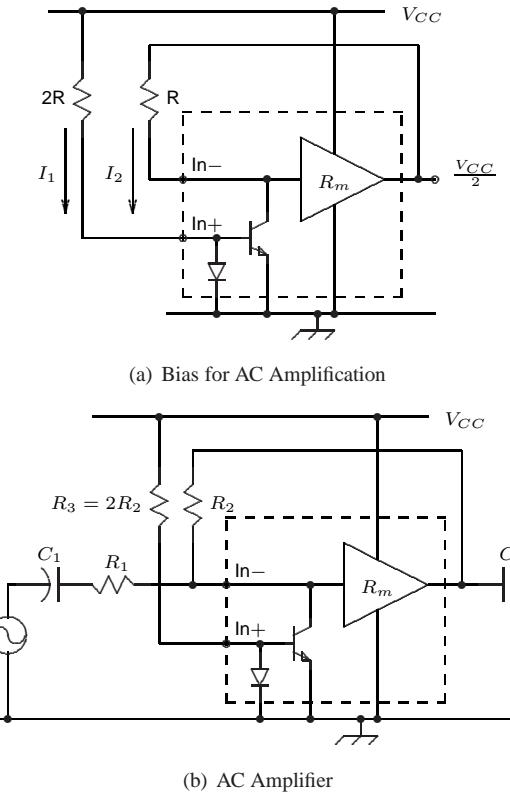


Figure 954: Norton Amplifier Application

³⁰⁰The capacitor value must be low compared to the Thevenin resistance at that point. Some writers designate such a capacitor as C^∞ , to cover all eventualities. Such infinite capacitors are a useful abstraction but not readily available for production work.

35.5 Operational Transconductance Amplifier

The operational transconductance amplifier was originally developed by RCA in 1969 [111].

Functionally, the OTA is a differential input, variable-gain, transconductance amplifier, in which the transconductance is proportional to a control current I_g ³⁰¹.

$$I_o = g_m \cdot e_{id} \text{ Amps} \quad (1614)$$

where

I_o Output current, amps

e_{id} Voltage between the inverting and non-inverting input terminals, $= e^+ - e^-$

g_m Transconductance (from the datasheet), amps per volt (mhos, \mathcal{G}), $= 19.2I_g$

I_g Gain control current, amps

As we would expect in a BJT-based circuit (section 29.11), the output impedance decreases with increasing control current. Again, from the datasheet:

$$r_o = \frac{7500}{I_g} \Omega \quad (1615)$$

35.5.1 Internal Circuitry

The internal circuitry of the original OTA, the RCA CA3080, is shown in figure 956. Each of the circles is a current mirror, in which the output current is equal to the input current.

Consider the input differential pair. A clever arrangement of current mirrors creates an output current that is the difference of the currents in the diff-amp transistors.

This difference current is proportional to the input voltage difference, divided by the incremental emitter resistance (section 30.14), that is:

$$\begin{aligned} I_o &= I_2 - I_1 \\ &= \frac{e^+ - e^-}{2r_e} \\ &= \frac{e_{id}}{2r_e} \end{aligned} \quad (1616)$$

Now consider the tail current of the differential pair. The total tail current I_{EE} is equal to the control current I_g . Assuming small signals then this current splits approximately equally between the two transistors and each transistor conducts $I_g/2$ amps.

Then the emitter incremental resistance of each transistor is:

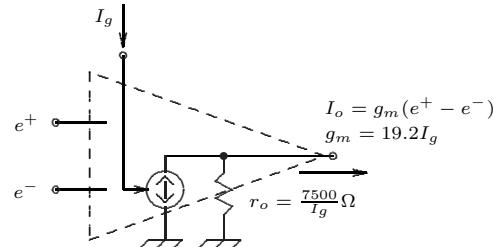


Figure 955: OTA Function

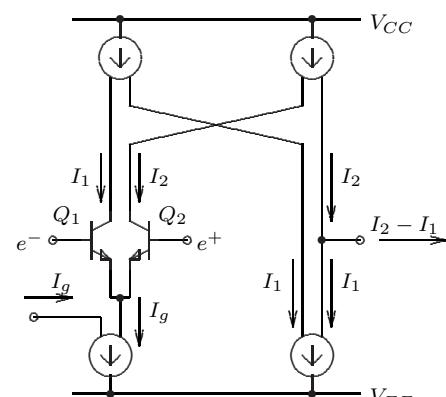


Figure 956: OTA Internal Circuit

³⁰¹In the original datasheets and application notes the gain control current is known as I_{ABC}

$$r_e = \frac{0.026}{I_e} = \frac{0.026}{I_g/2} = \frac{0.052}{I_g} \quad (1617)$$

Substitute for r_e from equation 1617 into equation 1616:

$$I_o = e_{id} 9.6 I_g \quad (1618)$$

This calculation generates a gain constant of 9.6 while the datasheet (quoted in figure 955 and equation 1614) shows a value of 19.2, a discrepancy that can be attributed to the type of transistors actually used in the device.

Some additional notes on the operation of the CA3080 OTA:

- The OTA may be operated open-loop, in which case there is no negative feedback around the device and there is no mechanism to reduce distortion introduced by the OTA.

We saw in section 30.14 that a differential pair can accommodate only a few tens of millivolts before being driven into distortion. As a result, the voltage signal into the OTA must be attenuated into an acceptable range or risk being distorted.

- The data sheet for the CA3080 states *gm linearity: 3 decades*. The performance curves show the gain control current I_g (called I_{ABC} in the datasheet) varying over a range of $0.1\mu\text{A}$ to 1mA . This wide range of operation is an advantage in many applications.
- The current gain of this OTA design is *linear in control current*. In many cases, the control current is generated by a control voltage across a resistor, in which case a control range of 1000:1 requires a voltage range of 1000:1. In audio applications, it's often more useful to have a control characteristic that is a logarithmic or exponential function of the control signal.

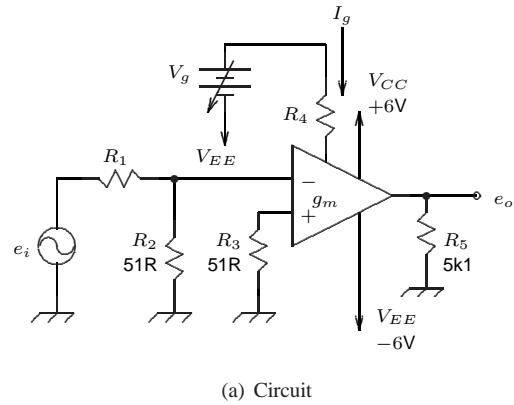
35.5.2 OTA Application: Voltage Controlled Amplifier

Figure 957(a) shows the circuit of a simple OTA-based variable gain amplifier (VGA). The input signal e_i is conveyed to the output at an amplitude determined by the control signal V_g . Figure 957(b) shows the signal that would result if the input e_i is a sine wave and the control signal V_g is adjusted from zero up to some positive value.

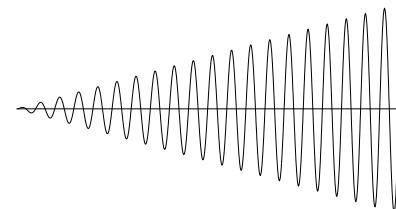
This circuit may also be regarded as a 2-quadrant multiplier. The input signal (multiplicand) may assume positive and negative values. The control signal (multiplier) may only assume positive values (or zero).

Notice that one terminal of the control signal must be returned to the negative supply rail V_{EE} in order to be able to reduce the control voltage to zero. The magnitude of the control voltage V_g and resistor R_4 are proportioned to set the maximum value of the control current.

Resistors R_2 and R_3 supply bias current to the input of the OTA. Resistor R_1 reduces the amplitude of the input voltage to some tens of millivolts peak. Resistor R_5 converts the OTA output current into a signal voltage. Alternatively, an op-amp current-voltage converter (section 13.9) could be used for this purpose.



(a) Circuit



(b) Output Waveform

Figure 957: Voltage-Controlled Amplifier

35.6 The Comparator Device

Although any operational amplifier can be used as a comparator, there is a specific comparator integrated circuit which is optimized for this type of operation. For example, the LM339 device from National Semiconductor is a set of four comparators in one IC package. (see *A Quad of Independently Functioning Comparators* in [199]). This comparator can operate from a single supply and has an *uncommitted collector* at the output. The comparator can be visualized as a single-supply op-amp driving a transistor, as shown in the timer circuit of figure 958.

This circuit illuminates the LED some time after the switch S_1 is pushed and then released. The operation is as follows:

- When the pushbutton S_1 is operated, it discharges the capacitor C_1 to zero volts.
- The non-inverting input of the comparator is established at +4 volts by the resistive divider $R_2 - R_3$.
- Since the voltage at the non-inverting input is higher than the voltage at the inverting input, the output of the comparator will be *high*, that is, the output transistor will be off, and the LED will be extinguished.
- When the pushbutton is released, capacitor C_1 begins to charge toward 5 volts with a time constant $R_1 C_1$. When this voltage passes through +4 volts, the voltage at the inverting comparator input, the output of the comparator changes state.
- Now the output of the op-amp is *high*, the transistor is ON, and the LED illuminates.

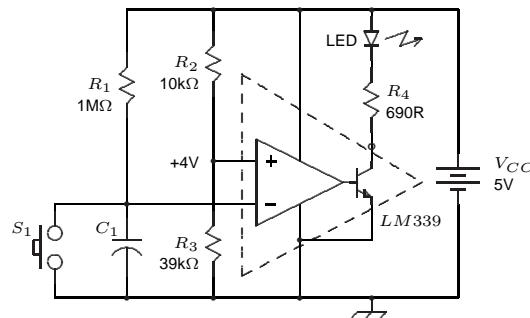


Figure 958: Time Delay

Comparing the Comparators

If an op-amp can be used as a comparator, why is a special-purpose comparator necessary [329]?

Op-amps often come several in a package. For example, the LM324 contains four op-amps in a single 14 pin package. If both op-amps and comparators are needed in a design, it's tempting to use an op-amp as a comparator. Will this work? The answer is a qualified *yes*. The comparator has several features that make it preferable but not mandatory:

Input Voltage The op-amp usually operates in a negative feedback system where the input terminals are at essentially the same voltage. A comparator is expected to operate with very different voltages on the input terminals. Some op-amps misbehave with large voltages between their input terminals. If you are using an op-amp as a comparator, check that it can handle this condition.

Output Circuitry The output of the op-amp is a current-limited voltage supply that swings in some range between the positive and negative supply terminals. (The negative terminal is ground for a single supply op-amp.)

The output of a comparator is an open-collector transistor that can be connected, usually via a resistor, to some voltage different from the power supply voltage. For example, consider the case where a comparator must be able to accept inputs from 0V to +12V and produce a logic signal that varies between 0V and +5V. A comparator could be operated from +12V and ground, with the output connected to +5V via a resistor. Then the output would swing between +5V and ground.

Alternatively, the output voltage could be much *larger* than the comparator supply voltage.

Finally, two or more comparators can have their outputs tied together through a common pullup resistor. Then the combined output is a positive logic AND of the individual comparator outputs. The combined output is HIGH only when the individual comparator outputs are both HIGH.

Speed and Stability The op-amp is usually embedded in a negative feedback circuit. This arrangement has the potential for oscillation (section 24). Most op-amps contain a compensation capacitor that rolls off the open-loop gain at high frequency to prevent this from happening. However, the compensation capacitor limits the maximum rate of change of voltage at the output of the op-amp (the slew rate, section 21.8).

Since comparators operate with the output at the positive or negative limit, oscillation is not as much of a problem. The compensation capacitor is not required and so the comparator has a much faster slew rate than an op-amp.

In conclusion, an op-amp makes a satisfactory comparator as long as its limitations are recognized.

35.7 Instrumentation Amplifier

The *instrumentation amplifier* (IA) is an operational amplifier which has been optimized for critical low frequency operations. In particular, an IA has small offset voltage and large common-mode rejection ratio. As a consequence, it is useful in applications which require sensing a small differential voltage in the presence of a large common-mode voltage. The IA then ignores the common-mode voltage and amplifies the differential signal by a factor typically of several hundred volts/volt.

Section 13.7, page 341, showed that an instrumentation amplifier can be constructed from three operational amplifiers, as repeated in figure 959 above.

Section 37.7, figure 1020, page 1135 shows an instrumentation amplifier of this type in an electronic barometer circuit. This approach is low cost and has the advantage that it can be tailored for a specific application.

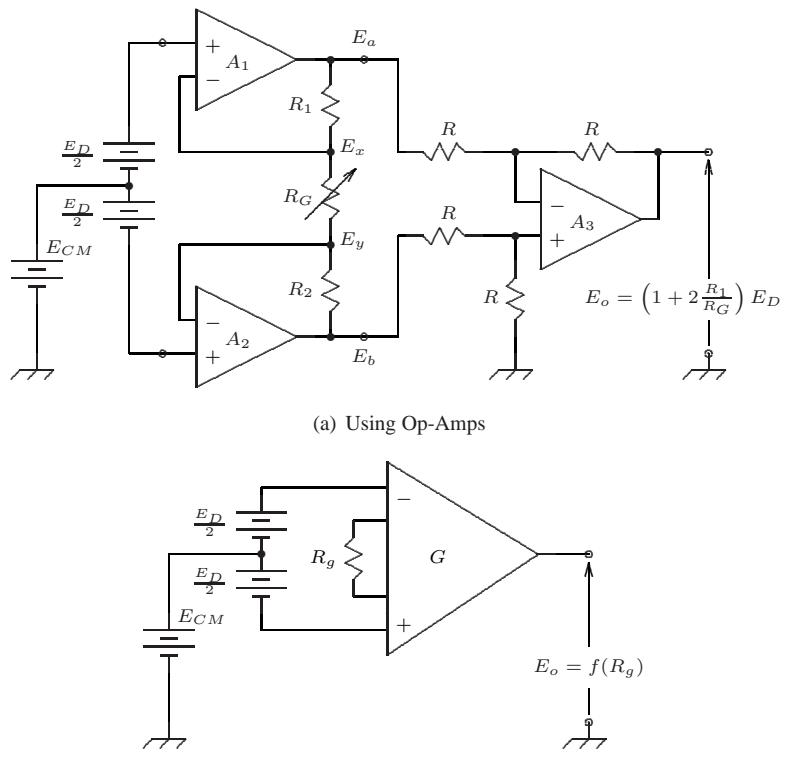


Figure 959: Instrumentation Amplifier

However, the common-mode rejection ratio is limited by the matching of the 4 resistors in the differential amplifier. A typical CMRR for this type of circuit is about 40 to 50db.

A comparison of a conventional operational amplifier with two representative instrumentation amplifiers is shown in the following table³⁰².

	General Purpose TL081	Instrumentation AD620A	Instrumentation INA114
Gain Error at 1000 V/V	1%	0.35%	0.5%
Offset Voltage	5mV	30 μ V	50 μ V
Offset Voltage TC	10 μ V/ $^{\circ}$ C	0.3 μ V/ $^{\circ}$ C	0.25 μ V/ $^{\circ}$ C
CMRR, db	100 (50)	130	120
GBP, Hz	4MHz	12MHz	1MHz
Manufacturer	National Semiconductor	Analog Devices	Texas Instruments
Cost, Qty 1, USD	\$0.48	\$6.62	\$9.50

The general purpose op-amp is clearly much less expensive than a typical instrumentation amplifier. However, for critical applications the instrumentation amplifier has compelling advantages.

- **Gain Error** The figures shown are the gain error contributed by finite open-loop gain in the amplifier. An instrumentation amplifier is routinely operated at a gain of 1000 volts/volt.
- **Offset Voltage and Drift** The offset voltage of the instrumentation amplifier is two orders of magnitude less than a typical ordinary op-amp. Similarly, the drift in offset voltage, which determines the maximum useful DC gain of the amplifier, is much less in the IA.
- **Common Mode Rejection Ratio** For the ordinary op-amp in an instrumentation amplifier circuit, the CMRR is primarily determined by the matching of the resistors R in figure 959(a). Consequently, the achievable CMRR using ordinary op-amps is in the order of 50db.

The IA includes in its differential stage resistors matched to far better tolerance than can be achieved with discrete resistors (they may be matched by *laser trimming*), and the common-mode rejection ratio is therefore much larger.

- **Gain Bandwidth Product** For an IA with a GBP of 1MHz which is operated at a gain of 1000 volts/volt, the closed-loop gain intersects the open-loop gain at 1kHz. Consequently, the gain error is large at and above that frequency. Depending on the required gain accuracy, and therefore the required loop gain, the maximum usable frequency might be 10 or 100Hz. These are not high-speed devices.

³⁰²These figures should not be used for design purposes. Check the data sheets on the manufacturer's web sites for detailed information and device selection information.

35.8 Fully Differential Amplifier

Most signal paths are *single-ended* and use ground as a return. However, there are advantages to a *differential* signal path, where both the signal and its return have conductors that are separate from ground.

35.8.1 Differential and Common-Mode Signals

First, let's briefly review differential and common-mode signals.

Consider the lower pair of waveforms in figure 962(b) on page 1077. Each triangular waveform represents a voltage measured between that conductor and ground, and is known as the *single-ended* signal. In this particular case, each single-ended signal has a DC component of 0.9 volts and a single-ended AC component of 0.7 volts peak.

These single-ended signals, each on its own conductor, are inverted versions of each other. Consequently, the *differential* signal (between the two conductors) is twice the single-ended signal: 1.4 volts peak, or 2.8 volts peak-peak.

Mathematically, let's refer to the two single-ended signals in figure 962(b) as e_A and e_B . The DC component of the signal is E_{DC} and the two AC components are e_1 and e_2 .

Then:

$$\begin{aligned} e_A &= E_{DC} + e_1 \\ e_B &= E_{DC} + e_2 \end{aligned}$$

But AC signal e_2 is the inverted version of e_1 , that is, $e_2 = -e_1$.

$$\begin{aligned} e_A &= E_{DC} + e_1 \\ e_B &= E_{DC} - e_1 \end{aligned}$$

The difference signal between these two conductors is:

$$\begin{aligned} e_D &= e_A - e_B \\ &= E_{DC} + e_1 - (E_{DC} - e_1) \\ &= 2e_1 \end{aligned}$$

That is, the differential signal is twice the AC component of the single-ended signal.

The common-mode signal between the conductors is the average of the single-ended voltages:

$$\begin{aligned} e_{CM} &= \frac{e_A + e_B}{2} \\ &= \frac{E_{DC} + e_1 + E_{DC} - e_1}{2} \\ &= E_{DC} \end{aligned}$$

That is, in this case the common-mode signal is simply the DC component.

In general, we are interested in the AC component of the signal, so we can extract that by finding the difference between the two single-ended voltages.

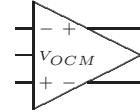


Figure 960: Fully Differential Amplifier

35.8.2 Differential Signalling

Section 13.6 on the subtractor and section 13.7 on the instrumentation amplifier showed circuits that could find the difference between two voltages and generate a single-ended result. The subtractor and the instrumentation amplifier could be described as *partially differential* amplifiers. In this section we describe the *fully differential* amplifier, which has differential inputs and outputs.

Differential signalling has been around for a long time in low-level instrumentation circuits, audio, and oscilloscope vertical amplifiers. It has several attractions:

- Differential signalling prevents noise currents flowing in the ground from mingling with the signal.
- Noise that is coupled into the two conductors of a differential signal path appears as a common-mode signal. A differential amplifier rejects this common-mode noise³⁰³.
- A fully differential amplifier can produce a peak-peak differential voltage that is twice the magnitude of the power supply voltage, which is useful in low voltage systems. This is increasingly important as supply voltages approach ± 3 volts, or even less.
- A fully differential amplifier chain can be *much* easier to design than one that uses single-ended signals. Each amplifier stage ignores the common-mode voltage of the previous stage, which simplifies coupling between stages.

Fully differential amplifiers are not entirely new³⁰⁴. However, recent integrated circuit designs differ by using external components to define the gain and have an adjustable common-mode output voltage. These new designs have become available primarily to drive the differential inputs of high-resolution, high-speed A/D converters.

The symbol for a fully differential amplifier is shown in figure 960. The key features are these:

- The common-mode input voltage (the average of the two input voltages) is ignored. Providing that it's within the allowable range at the input, it has no effect on the output voltage.
- The common-mode output voltage may be established at any useful value within the allowable range at the output. The default value is typically half the supply voltage.

As an added bonus, fully differential amplifiers are available with very wide bandwidth, approaching 1GHz. On the detriment side, fully differential amplifiers are low impedance devices: they require low-value resistors to operate at high frequency, so the input impedance is low. The input bias current is high, tens of μA is typical. Driving a fully differential amplifier from a source with high internal resistance will require a buffer of some sort, such as an op-amp follower.

These features are illustrated in the example that follows.

³⁰³For this to work, the common-mode rejection ratio (CMRR) must be sufficient at the frequency of the noise. The CMRR decreases at high frequencies. For example, in audio amplifiers, the CMRR may not be effective at radio frequencies.

³⁰⁴Discrete component versions were introduced into oscilloscope vertical amplifiers just after WWII [330], [273]. The $\mu\text{A}733$ amplifier is an early integrated circuit fully differential amplifier [331].

35.8.3 Internal Circuit

As in the treatment of the operational amplifier, it is possible to understand the operation of the fully-differential amplifier without reference to the internal circuitry. For those who are interested, we provide figure 961. The diagram shows the internal circuit of a fully-differential amplifier (adapted from [332]).

Transistors Q1 through Q4 are a *folded cascode differential amplifier*. The differential amplifier (section 30.14 on page 935) outputs the two signals of a differential pair and rejects any common-mode input voltage. Cascode transistors Q3-Q4 (section 30.12 on page 931) inhibit Miller-effect capacitance around the Q1-Q2 pair. Consequently there is very little voltage swing at the collectors of the differential pair, ensuring wide bandwidth. The current output of the cascode pair, driving into the high impedance of Q5-Q6, generates the differential voltage swing at the output terminals of the amplifier.

The common-mode output signal is sensed by the two-resistor divider. The amplifier drives the Q5-Q6 pair to shift the common-mode output to be equal to V_{ocm} . As shown, this circuit cannot supply significant current into a load, and the output terminals would be driven by buffer amplifiers of some sort.

35.8.4 Example Design: Transient Capture

Consider that we wish to capture the behaviour of an aerodynamic shock wave, using a strain gauge sensor³⁰⁵. A conceptual circuit is shown in figure 962(a).

The pressure sensor consists of a four-resistor bridge that is bonded to a diaphragm. When a shock wave impinges on the pressure sensor, it deforms a diaphragm on which these four resistors are mounted. This stretches or contracts the resistors, changing their resistance.

The bridge is normally balanced, so points A and B are both at half the supply voltage, +2.5V. Under pressure, the resistors change value in such a way as to raise the voltage at A by 100mV and lower the voltage at B by 100mV. The signal is then composed of a common-mode component of 2.5V, upon which appears a differential component of 200mV peak-peak.

The A/D converter digitizes this signal and transfers it to a host computer for further display and analysis³⁰⁶. The A/D converter has a differential input. The common-mode input signal must be 0.9V with a maximum differential amplitude of 1.4V. That is, the two input connections to the A/D carry a single-ended signal of $0.9V \pm 0.7V$.

The fully differential amplifier serves two functions: it shifts the common-mode voltage from +2.5V at the output of the sensor bridge to +0.9V at the input of the A/D converter. It also provides the required differential gain, from 200mVp-p to 1.4Vp-p. These functions can be provided with conventional op-amps or discrete transistor circuits, but the discrete component circuits are much more complicated.

³⁰⁵The numerical values shown in this example are purely fictitious and provided for the purpose of illustrating concepts. An exact duplicate of this circuit will probably not work.

³⁰⁶Actually, it's unlikely that the host PC could absorb the data quickly enough over a serial or USB connection. Instead, the A/D would put the sample values in a digital memory. Then the host computer would read out the samples for further display and analysis.

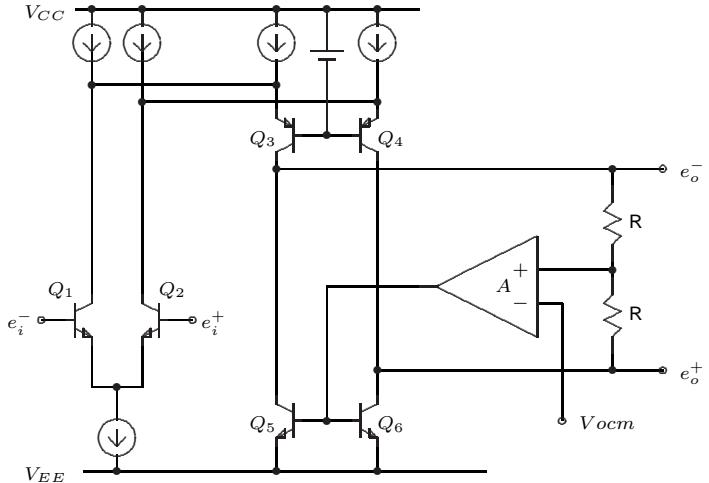


Figure 961: Internal Circuit

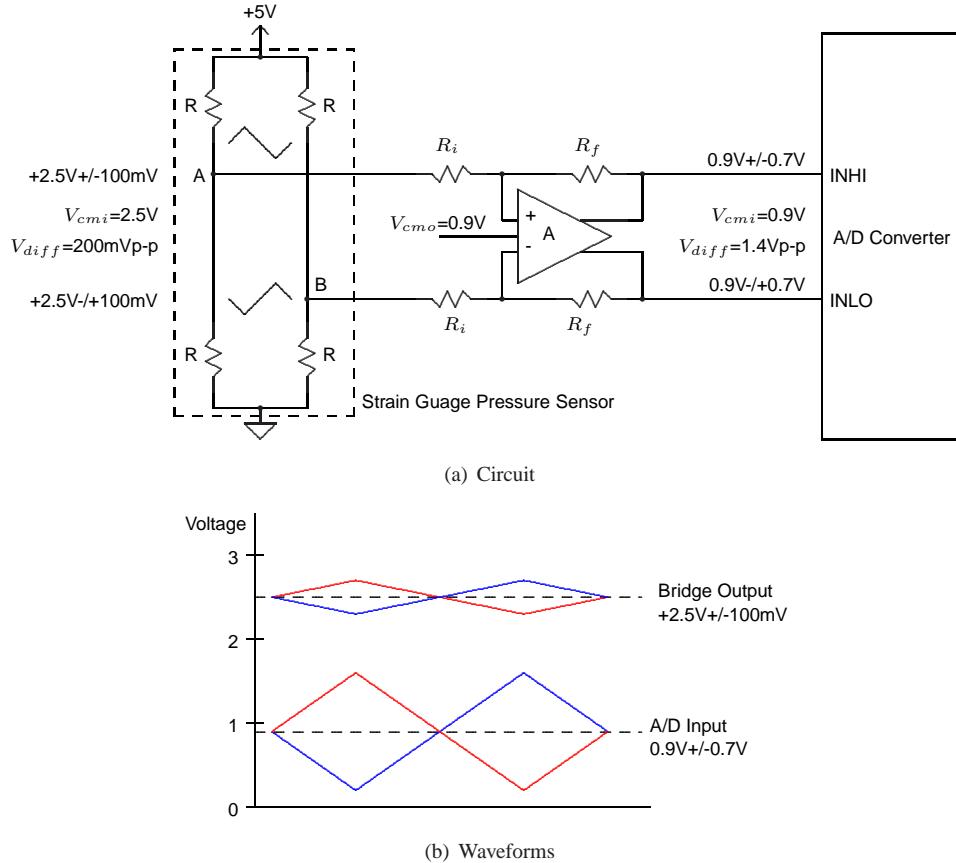


Figure 962: Fully-Differential Amplifier Application

The output common-mode voltage is set to 0.9V by the V_{cmo} input signal. This could be provided by a voltage divider from the +5V supply to ground, with the divider current much larger than the bias current of this input³⁰⁷.

Now we need to determine values for R_f and R_i . The next section equips us with the necessary analytical tools to make that calculation.

35.8.5 Rules of Ideal Behaviour

Here are the rules of ideal behaviour, useful in the analysis of a circuit with a fully-differential amplifier:

- Negative feedback drives the differential input terminals to the same voltage. That is, when negative feedback is operating properly – as in the case of a conventional op-amp – there is zero voltage between the input terminals.
- No signal current flows into the amplifier input terminals.

³⁰⁷It's good practice to bypass this constant-voltage terminal to ground with a capacitor. That ensures constant voltage at the V_{cmo} terminal, even if there are spikes of current demand.

- The plus and minus symbols indicate the relative polarity of the differential input and output.
- The output at the + terminal is equal to the common-mode output voltage plus half the total differential output voltage.
- The output at the - terminal is equal to the common-mode output voltage minus half the total differential output voltage.

35.8.6 Differential Gain Analysis

To begin with, we make the assumption that the feedback network is symmetrical. As shown in figure 963, the input resistors R_i are equal and the feedback resistors R_f are equal. This need not be the case³⁰⁸. However, this symmetry is used in the primary applications of the fully-differential amplifier.

We'll take a visual and intuitive approach. Assume the common-mode input and output voltages are zero. Then, using symmetry, figure 963 shows how the fully-differential amplifier in figure 963(a) can be treated as two inverting op-amps in figure 963(b). The dashed line represents zero potential. There is one conventional inverting op-amp above the dashed line and one below it.

For the upper op-amp in figure 963(b),

$$V_c = -\frac{R_f}{R_i} V_a \quad (1619)$$

In other words:

$$\frac{V_{od}}{2} = -\frac{R_f}{R_i} \frac{E_{id}}{2} \quad (1620)$$

or,

$$V_{od} = -\frac{R_f}{R_i} E_{id} \quad (1621)$$

That is, for a positive input voltage to the upper op-amp, the output voltage is negative. The lower half of the fully differential amplifier works exactly the same way. Its input voltage is negative and its output voltage is positive.

Overall, then, the gain is given in equation 1621. The arrangement of polarity symbols on the fully-differential amplifier shows that the input and output differential voltages are of opposite polarity: $+$ at the input results in $-$ at the output, and vice versa.

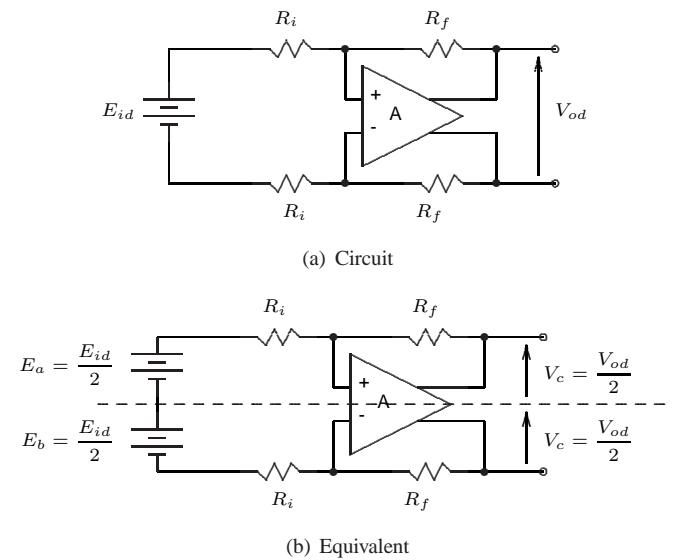


Figure 963: Differential Gain Analysis

³⁰⁸Reference [333] shows some examples of non-symmetric resistor configurations.

A more analytical approach is shown in figure 964. This follows the same approach we used with the conventional operational amplifier: (a) the voltage between the input terminals of the amplifier is zero and (b) signal current cannot flow into the input terminals of the amplifier.

Apply KVL around Loop 1, noting that the input terminals of the amplifier have zero voltage between them. Then the voltage E_{id} is distributed equally across the two resistors R_i , and the current I_i is given by:

$$I_i = \frac{E_{id}}{2R_i} \quad (1622)$$

Since current cannot flow into the input terminals of the amplifier, this current must be equal to the feedback current I_f ³⁰⁹. Now apply KVL around Loop 2, remembering that the voltage across the amplifier input is zero:

$$0 - I_f R_f + V_{od} - I_f R_f = 0 \quad (1623)$$

Substitute I_i from equation 1622 for I_f in equation 1623 and solve for the voltage gain:

$$\frac{V_{od}}{E_{id}} = \frac{R_f}{R_i} \quad (1624)$$

We selected the polarity of the output voltage to reflect physical reality (the output voltage is inverted with respect to the input). Consequently the minus sign is missing in the resultant equation. Either you have to remember the relative polarities or assume them the same and include a minus sign in the equation.

35.8.7 Common-Mode Analysis

The common-mode input voltage and output voltage have no effect on the differential output voltage or the differential gain. However, they do affect the common-mode voltage at the input to the amplifier. This common-mode voltage must not exceed the maximum allowable common-mode input, which is specified on the data sheet for the amplifier. Notice that these common-mode signals may have an AC component, in which case the maximum and minimum values are of interest. This is easiest to picture with a lever diagram, which follows in section 35.8.8 below.

A differential input signal causes no voltage swing at the input terminals of the amplifier.

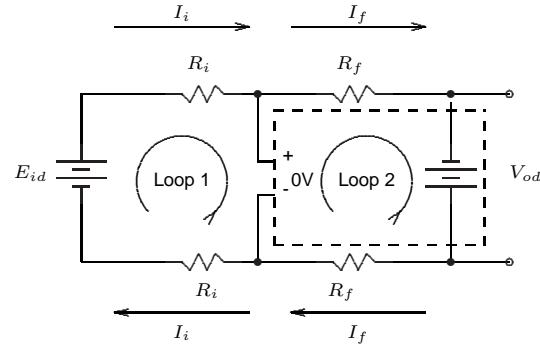


Figure 964: Differential Gain Analysis

³⁰⁹More correctly, we should say that the operation of negative feedback is to adjust the output voltage in order that the voltage between the amplifier input terminals is near zero. To accomplish that, the amplifier adjusts the output voltage so that all the input current flows through the feedback resistor.

35.8.8 Design Example: Single-Ended to Differential Converter

The lever diagram³¹⁰ is useful in designing a circuit that uses the fully-differential-amplifier. An example, the single-ended to differential converter circuit, is shown in figure 965. As previously, we assume that the two ratios of R_f/R_i are equal.

The design challenge: calculate the ratio of $R_f : R_i$ so that the differential output voltage v_o is twice the single-ended input voltage e_i .

To begin with, we'll set the output common-mode voltage to zero.

Now we can construct the lever diagram, shown in figure 966. This construction is a little more complicated than the conventional op-amp lever diagram, so we'll do it for a snapshot of the voltages at one instant in time, and in several steps. (To help, vectors in the diagrams are referred to in the text with square brackets).

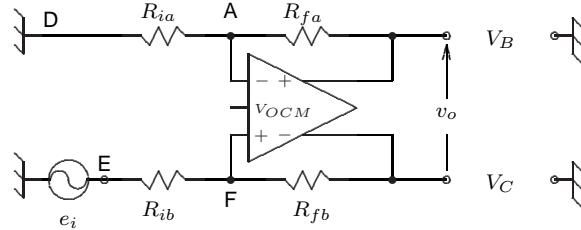


Figure 965: Single to Differential Signal Converter

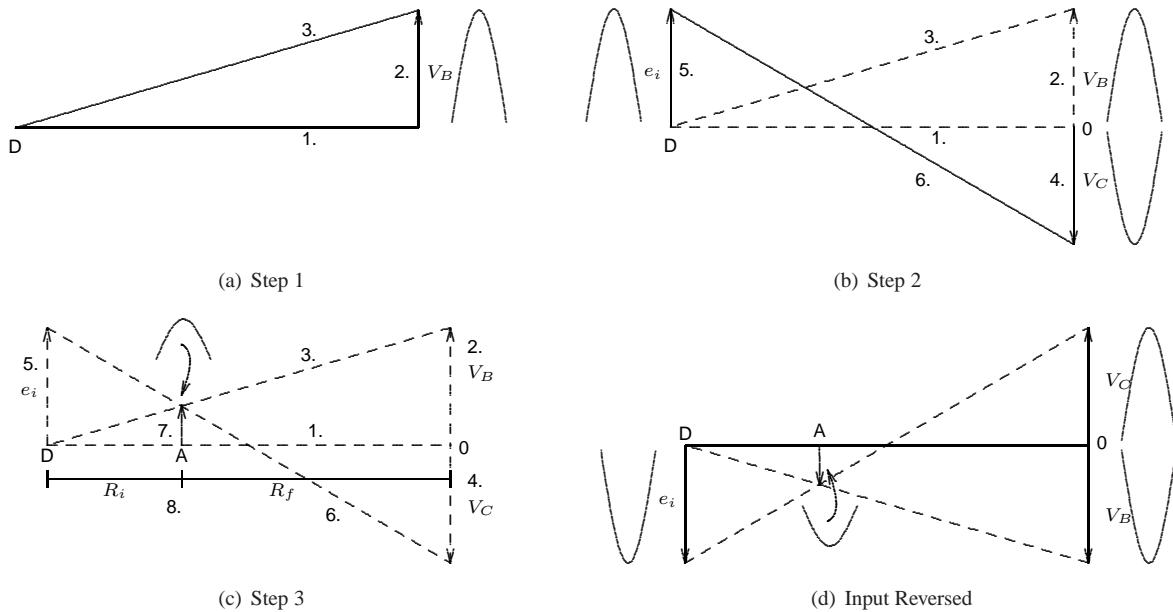


Figure 966: Lever Diagram, Single to Differential Signal Converter

- Passive Side.** As in many op-amp analysis problems, it helps to start with the network that does not have a signal input, R_{ia} and R_{fa} .

- (a) Figure 966(a): Draw the zero-voltage baseline [1], some arbitrary length.
- (b) Draw the output voltage V_B , some arbitrary height [2].

³¹⁰The use of the lever diagram in an op-amp circuit is shown in section 12.5 on page 322 and section 13.8 on page 343.

- (c) Now draw the lever diagram [3] for the voltage divider R_{ia}, R_{fa} . The left end of this divider connects to zero volts at D ; the right end connects to V_B volts, so the lever ramps upward from left to right, from zero to V_B volts.

2. **Active side.** Now we can work on the side of the amplifier that has a signal input R_{ib}, R_{fb} .

- (a) As figure 966(b) shows, the amplifier outputs V_B and V_C move in opposite directions around the common-mode output voltage (zero in this case), so draw the voltage V_C as [4], equal and opposite to V_B .
- (b) Now we must draw the input voltage vector e_i to suit our requirement that the differential output v_o be twice the single-ended input e_i . The differential output voltage v_o is twice the single-ended output voltage, so the magnitude of e_i is equal to the magnitude of V_B or V_C . The circuit indicates that output V_C is opposite in polarity from e_i , so draw e_i as [5], upward with the same length as V_C .
- (c) Connect the tips of the vectors for e_i and V_C . This line [6] is the distribution of voltage (the lever diagram) for the R_{ib} and R_{fb} voltage divider.

3. **Results.** Now this diagram can provide useful information. Read off the resistor ratio and the voltage at the amplifier input terminals.

- (a) Figure 966(c): The intersection point A of the two levers defines horizontally the ratio of the input and feedback resistors. In this case a measurement from the diagram indicates that R_f is twice R_i . If we alter the output/input ratio, point A slides to the left or right, altering the ratio of feedback to input resistance.
- (b) As a matter of interest, a vertical vector [7] at the intersection of the two sloping lines defines the input voltage at the terminals of the amplifier when the input voltage is positive. This is the voltage at *both* input terminals, since negative feedback forces them to be equal.

4. Input Voltage Reverses.

- (a) Figure 966(d) shows what happens when the input voltage reverses polarity: the entire diagram flips upside-down.

The voltage at the input terminals of the amplifier reverses with the input signal e_i . With an AC input for e_i , we would expect to find an AC voltage at the input terminals of the amplifier.

Adding a Common-Mode Output Offset

What is the effect on this circuit if we dial in a positive common-mode output voltage?

On the lever diagram, it's very simple: the two outputs V_B and V_C shift upward by the common mode voltage, figure 967(a).

The ratio of e_i to v_o – the voltage gain – does not change. The differential output voltage is unaffected by the common-mode output voltage.

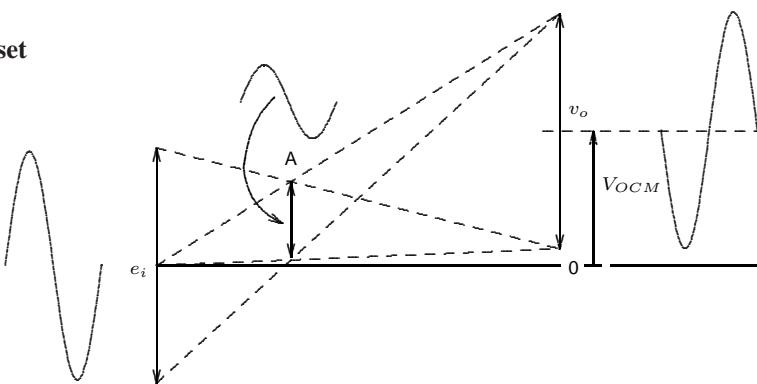


Figure 967: Adding Common Mode Output

Alternative Analysis

The lever diagram is helpful in picturing the behaviour of the circuit. It's also possible to get to an answer using our previous result.

In section 35.8.6 above, we showed that

$$\frac{v_{od}}{e_{id}} = -\frac{R_f}{R_i} \quad (1625)$$

where v_{od} is the differential output voltage and e_{id} is the differential input voltage.

In the single-ended to differential converter circuit of figure 965, the differential input voltage is simply e_i . Because the upper terminal of this source is held at zero volts, the average value of the input voltage is $e_i/2$ and this appears as a common-mode AC voltage at the input. This common-mode input voltage shows up on the lever diagram of the previous section and should be taken into account to ensure the common-mode input range of the amplifier is not exceeded. However, this common-mode input voltage has no effect on the output differential voltage.

Then the overall gain between differential input and differential output is given by equation 1625, so for a differential gain of 2 V/V we require that $R_f = 2R_i$.

35.8.9 Selecting Resistor Value

The voltage gain of the fully-differential amplifier is established by the ratio of two resistors. To complete the design, we need to choose one of the resistors. The choice is not entirely arbitrary. A lower value results in higher frequency response, larger current consumption and lower input resistance, so engineering tradeoffs are required.

Some indication of the effect of feedback resistance can be obtained from figure 968, from the datasheet for the Texas Instruments THS4521 fully differential amplifier [334].

These curves are specific to that device, so they should not be applied to other amplifiers. However, the curves do show how frequency response decreases as the resistance increases³¹¹.

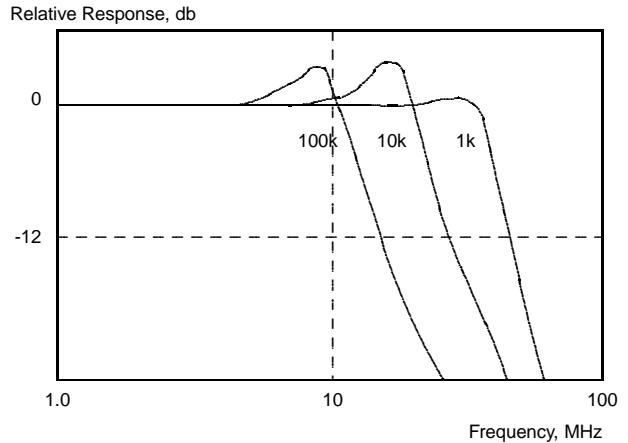


Figure 968: Feedback Resistor Value and Bandwidth

³¹¹The curves also show some degree of peaking as the feedback resistance increases. This would show up as overshoot on square waves, so it's undesirable. It might be possible to reduce the peaking by adding a capacitor in parallel with the feedback resistance to reduce this effect.

35.9 Autozero Amplifier

There are certain applications where a low-level DC signal³¹² must be amplified to a more substantial voltage.

Some examples:

- **Strain Gauge** The *strain gauge* is a printed-film resistor that is bonded to a surface. When the surface is deformed by mechanical strain, the resistor stretches in length which increases its resistance. Strain gauges are used in mechanical engineering to measure strain (and by implication, stress) in a structure such as an airplane wing. They are also the basis of other transducers where a physical quantity can be converted to strain, such as a pressure gauge or weigh scale.

It's usual to group several strain gauges in a bridge configuration, which minimizes temperature effects and increases the output voltage. Nonetheless, the output signal from a strain gauge is in the order of hundreds of nanovolts.

- **Thermocouple** When two dissimilar metals are placed in contact, they generate a small voltage according to the *Seebeck effect*. For example for an iron-constantan pair the generated voltage is proportional to absolute temperature at a rate of about $50\mu\text{V}/^\circ\text{C}$ [141].

A thermocouple can withstand a very high temperature. In the case of iron-constantan, the maximum is 750°C . Other thermocouple materials can be used up to in excess of 2000°C , so thermocouples are useful in measuring the temperature of industrial ovens and aircraft jet engine exhaust.

For accurate temperature measurement, the measuring system must be capable of accurate amplification of a thermocouple signal which is in the order of a few tens of millivolts. Consequently, the errors must be less than a fraction of a millivolt.

- **Precision Voltage Reference** A *standard cell* is used to provide an accurate voltage reference for the calibration of voltmeters and other measuring equipment. This device cannot supply significant current into a load, and so a buffer amplifier is required. Any offset voltage in the buffer amplifier affects the accuracy of the reference voltage, so this must be kept very small.

- **Precision Current Measurement** The usual approach to measuring current in a load is to place a small value resistance (often known as a *shunt*) in series with the load, and measure the voltage across the resistance. To minimize disturbance to the circuit, the sensing resistance should be as small as possible. However at low currents this results in a low value of sensed voltage. For an accurate measurement of current, this sensed voltage must be measured accurately, and that requires an amplifier with a low offset voltage.

The bias current, offset voltage, common-mode input voltage and finite open-loop gain all contribute DC errors in an op-amp amplifier. Of these, the most troublesome is offset voltage and its change with temperature.

The Autozero Concept

Offset voltage may be characterised as a small voltage in series with one of the op-amp inputs. If it can be measured and an equal voltage applied in series with the other op-amp input, then the voltages cancel and the amplifier is free of offset voltage error. If this can be done on a continuing basis, then a change in the offset voltage due to changing temperature will be mirrored in the cancelling voltage, and the amplifier will be free of any drift at the output due to changing offset.

³¹²Referring to a signal as *DC* implies that it is completely static. This is not correct because the signal does change with time or it would not be of any interest. A better description would be *ultra-low frequency*, since the signal does change, but very slowly. Calling a signal DC implies that its signal components are of such a low frequency that AC coupling would require capacitors and inductors that are unreasonably large.

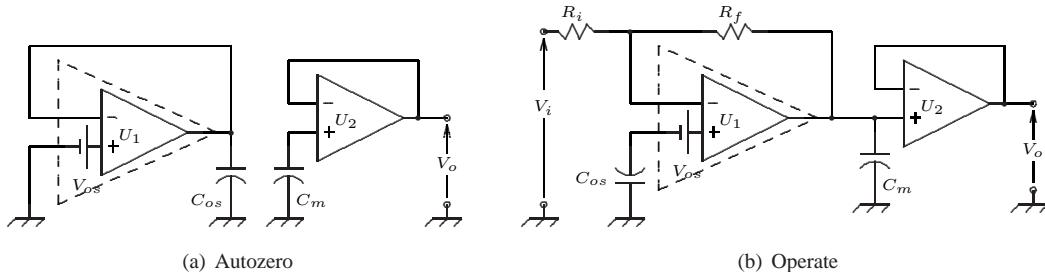


Figure 969: Autozero Operation

Because many of the applications of an autozero amplifier are low frequency, then it is allowable to divide the operation into two phases. Referring to figure 969, the operation is as follows:

- Autozero

- The output voltage of the amplifier is stored in a sample-hold analog memory. The output of the sample-hold provides an output voltage to the load for the system..
 - The amplifier is disabled and temporarily configured as a voltage follower. Its output voltage is equal to its offset voltage.
 - The offset voltage of U_1 is stored on a capacitor.

- Operate

- The capacitor is moved so that it is reversed and placed in series with one of the inputs of the op-amp, in order to cancel the offset voltage of U_1 .
 - The amplifier is reconfigured with an input and feedback resistance so that it functions normally.
 - The sample-hold circuit switches to *sample* mode and stores the current output of U_1 in the memory capacitor.

This circuit is incomplete inasmuch as it requires a number of analog switches to make the circuit change from 969(a) to 969(b), and it does not indicate how to compensate for any offset voltage in U_2 . It does however indicate how the offset voltage is captured on a capacitor, and then moved to cancel that same offset voltage.

Pros and Cons

- This amplifier is capable of an offset voltage of a few nanovolts. Consequently, the amplifier may be designed for an extremely large gain, in the order of 1000V/V. However, at that point, other error sources such as unintended thermocouple generators or finite open-loop gain, may become important. Autozero operation is not a panacea for a high gain design.
 - This is a *sampled-data system*. As a result the frequency of operation is limited to some fraction of the rate at which the system switches between *autozero* and *operate*.
 - The switching activities inevitably inject some noise into the signal path. In a low-frequency operation, this switching noise can be removed with a lowpass filter.
 - It is likely (but not inevitable, as evidenced by some recent designs) that external capacitors are required to act as analog memory for the offset voltage and output voltage.

35.10 Chopper Amplifier

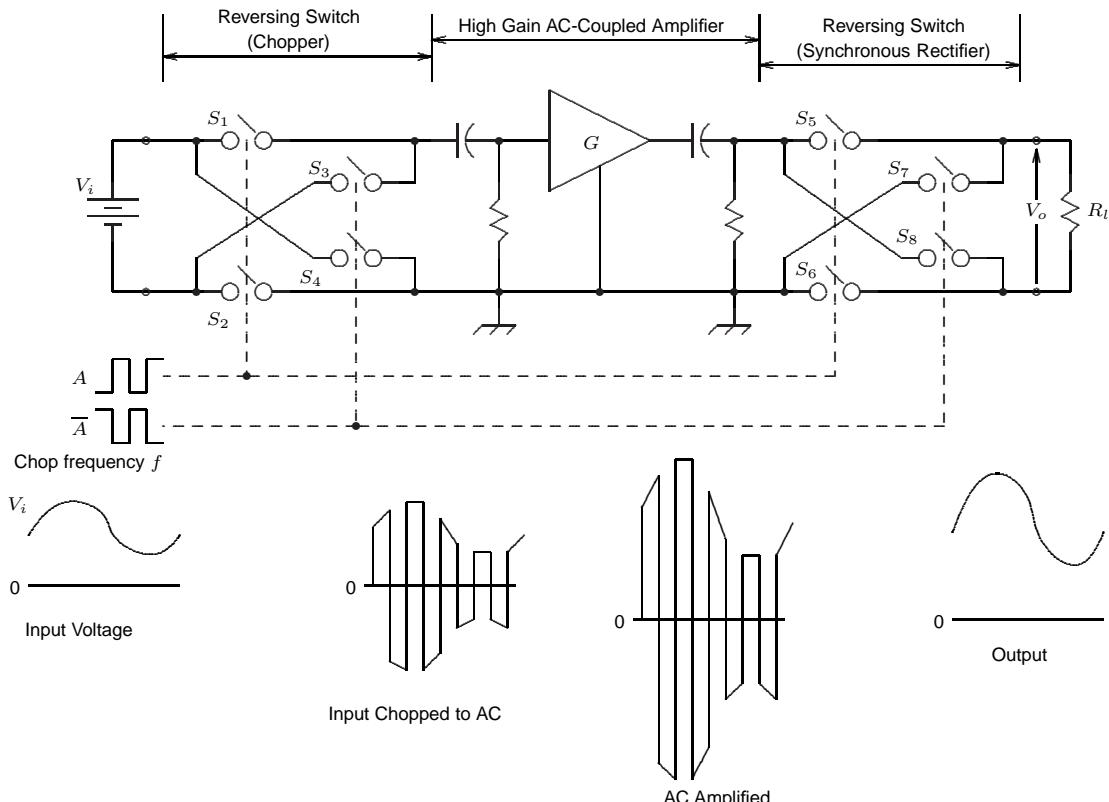


Figure 970: Chopper Amplifier Concept

As mentioned in section 13.22, offset voltage has no effect on the output of an AC-coupled amplifier. Consequently, one useful strategy in building a stable DC coupled amplifier is to

- Chop the input DC voltage into a square wave of peak amplitude equal to the DC input
- Amplify this AC waveform in a high-gain AC-coupled amplifier
- Synchronously rectify the AC waveform, returning it to a DC voltage. (A *synchronous rectifier* is a switch that can pass through or invert the incoming waveform. The switch is operated in synchronism with the AC waveform. It passes through the positive half cycles and inverts the negative half cycles of the incoming waveform, so that the output is full-wave rectified.)

An example circuit is shown in figure 970 above. The four input switches chop the input DC voltage V_i at some frequency f . They are driven by two antiphase control signals A and \bar{A} , acting as a reversing switch act to alternate the input voltage to the amplifier between $+V_i$ and $-V_i$. The resultant AC waveform has a peak amplitude equal to the input voltage.

This AC waveform is amplified by the gain factor G and then fed into the synchronous rectifier section. The effect of the synchronous rectifier is to flip the negative portions of the AC waveform so that they fill in the gaps between the positive pulses. The result is the original DC input voltage, amplified by the gain G .

Figure 970 is a simple example that requires both the input voltage and load to float off ground. However, that is not a general requirement of chopper stabilization.

Chopping Switches

The chopper amplifier technique was widely used in the heyday of the vacuum tube. A low-drift direct-coupled tube amplifier is a challenge to design (though it could be done). However, it was simpler to design an AC-coupled amplifier with a chopper at the input and a synchronous rectifier at the output.

Originally, the chopping device was a mechanical vibrating reed that operated at a few 10's of Hz. Other devices that have been used as choppers include photoresistors that were strobed with a flashing light source, BJT's, JFETs and MOSFETs. At this time, the analog switch, similar to the CD4016, is widely used. The 4016 includes four SPST analog switches and the necessary driver circuitry to work from a digital clock signal. This device would be suitable for the switches shown in figure 970.

Single-Ended Chopper Amplifier

A single-ended chopping amplifier, described by Jim Williams in [335], is shown in figure 971.

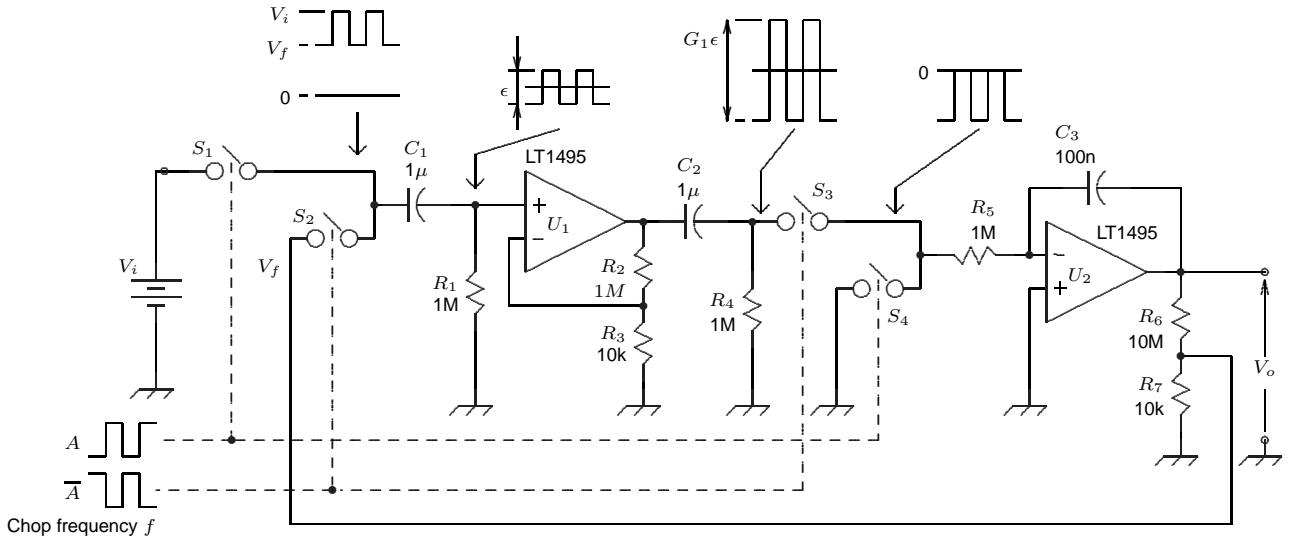


Figure 971: Single-Ended Chopper Amplifier

Basically, this is a non-inverting amplifier with the closed-loop gain defined by R_6 and R_7 :

$$\begin{aligned} G &= \frac{V_o}{V_i} \\ &= 1 + \frac{R_6}{R_7} \end{aligned} \quad (1626)$$

For the values shown, $G = 1001\text{V/V}$. At this magnitude of voltage gain, even a fraction of a millivolt offset would produce a substantial output. This amplifier eliminates the effect of offset voltage altogether. Here is how it works:

- The feedback voltage V_f is of the same polarity as the input voltage V_i .
- Chopper switches S_1 and S_2 alternate between V_i and V_f so that the voltage at their junction is a square wave with amplitude proportional to the difference between them. This signal is an AC version of the error

signal ϵ in a negative feedback system. If the system is completely in balance, there is no error signal and this AC signal has zero amplitude.

- The highpass network R_1, C_1 ensures that no DC voltage reaches the input of amplifier U_1 . Because U_1 is configured with very high gain, any DC voltage at this point would drive it to saturate against a supply rail.
- Resistors R_2 and R_3 establish the closed-loop gain G_1 of U_1 , at 101V/V . The gain of this stage is a compromise. On the one hand, it should be as large as possible compared to the closed-loop gain of the system. On the other hand, it can't be so large that its own offset voltage causes a significant offset at the output. The remainder of the required open-loop gain is made up by the integrator, U_2 .
- The error signal is amplified by the gain G_1 and then passed through another highpass filter, C_2, R_4 .
- Switches S_3 and S_4 are the synchronous rectifier in this system. Switch S_3 operates to select the negative-going half of the amplified error waveform.
- If the error waveform is non-zero amplitude, the synchronous rectifier produces negative-going pulses that drive the integrator output in a positive direction.
- This drives the error voltage in the direction of the input voltage, reducing the error. This is the behaviour we require of a negative feedback system.
- The steady-state (DC) gain of an ideal integrator is infinite. In this case, the steady-state gain of the integrator stage is limited by the open-loop gain of op-amp U_2 . The *total* open-loop gain is the gain G_1 of the first op-amp stage times the open-loop gain of amplifier U_2 , which must be much larger than the closed-loop gain G defined in equation 1626.

Chopper Stabilization Generalized

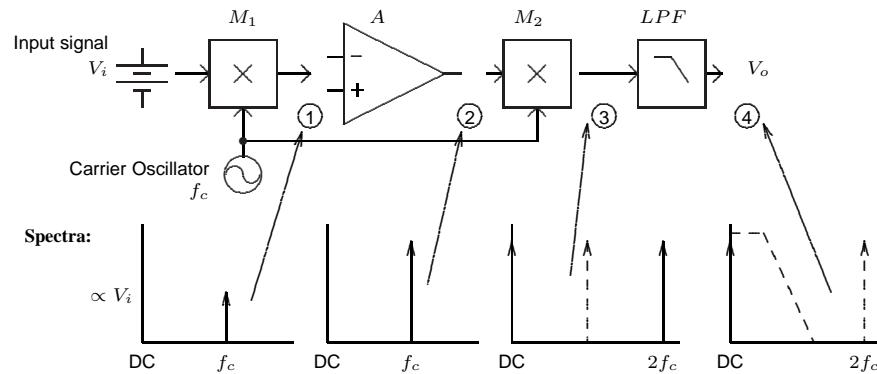


Figure 972: Chopper Amplifier Generalized

The operation of a chopper-stabilized amplifier may be generalized as shown in figure 972. This concept turns out to be useful in a number of applications where a small signal must be recovered in the presence of large amounts of noise.

The two multipliers correspond to the chopper and synchronous rectifier of the chopper stabilized amplifier circuits we have examined earlier.

The first multiplier M_1 accepts the DC input signal V_i and a carrier signal at the chopping or *carrier* frequency from the carrier oscillator. The chopping signals we've been looking at in previous circuits have all been square

waves. In order to simplify the explanation, in the first instance we'll assume that the carrier wave is a sine wave. We'll return to square waves in a moment.

The output of M_1 is a signal at the carrier frequency, the amplitude of which is proportional to V_i . This is represented in the spectral diagram #1 which shows a single frequency at the carrier frequency f_c .

This carrier signal is then amplified in the AC amplifier A and shows up on the spectrum diagram at #3 as a larger signal than previously, but still at frequency f_c .

At the output of the amplifier, the carrier signal is multiplied by the original carrier oscillator signal. When two frequencies are multiplied together, the resulting output spectrum contains two new frequencies: one at the *sum* of the input frequencies and one at the *difference* of the input frequencies. In this case, the two input signals are at exactly the same frequency f_c Hz. The sum output is at $2f_c$ Hz, and the difference frequency at 0 Hz (DC), as shown in the spectral diagram #3.

The $2f_c$ Hz sum frequency is removed by a lowpass filter. The zero Hz difference frequency is the amplified version of the input signal, which becomes the output of the system, as indicated in the spectral diagram #4.

If the bandwidth of the lowpass filter is made very narrow, it effectively rejects any noise or spurious AC signals introduced in the amplification process.

Another Example

Another example of a chopper-stabilized amplifier is shown in problem 1 on page 1089.

35.11 Composite Amplifier

In the design of an operational amplifier, it turns out that certain specifications can be enhanced only at the expense of other specifications. For example, it is difficult to make an amplifier that simultaneously has high bandwidth and low offset voltage.

When such an amplifier is required, it is sometimes possible to combine two amplifiers to achieve both specifications: high bandwidth and low offset, for example. (It is very common to combine discrete transistors with one or more operational amplifiers. For the purposes of this section, we define a composite amplifier as one that includes only op-amps.)

Some application domains where a composite amplifier is useful are:

Application Requirement	Concept	Reference
High speed, low offset voltage	A low-offset voltage op-amp senses the offset voltage of the high-speed amplifier and drives the offset in the high speed amplifier to zero.	[336], [337], [325]
Low noise	An array of N op-amps are paralleled, which reduces the input voltage noise by a factor of \sqrt{N} .	[336]
High output current	An array of N op-amps are paralleled in order to increase the available output current by the factor N . Care must be taken to ensure that the current load is shared among the op-amps.	[336], [338]
Open loop gain in excess of one op-amp	Two op-amps are cascaded to create an open-loop gain equal to the product of the individual open-loop gains. Some care must be taken to ensure stability.	[139]

The first of these applications – combining high-speed and low-offset voltage op-amps – is the most common. An example from [336] is shown in figure 973.

In this circuit, A_1 is the high-speed amplifier and A_2 is a low-offset device. (For the ultimate in low offset voltage, A_2 would be an auto-zero or chopper-stabilized amplifier). Amplifier A_1 is configured as an inverting unity-gain amplifier by resistors R_1 and R_2 . Amplifier A_2 is configured as an integrator by resistor R_3 and capacitor C_1 . The integrator senses the voltage at the virtual earth point of the high-speed amplifier. Suppose that this point is slightly positive due to the offset voltage of the amplifier. Then the output of the integrator will ramp in a negative direction until the inverting terminal of the high-speed amplifier is back to zero volts, thereby cancelling the offset voltage in the high-speed amplifier. Because the offset correction amplifier A_2 is an integrator, it responds only to low-frequency signals.

Since the correction signal at the output of A_2 is very small (a few millivolts) the large gain of the integrator is unnecessary and the output of the integrator is reduced by resistor network R_4 and R_5 . This improves the stability of the system.

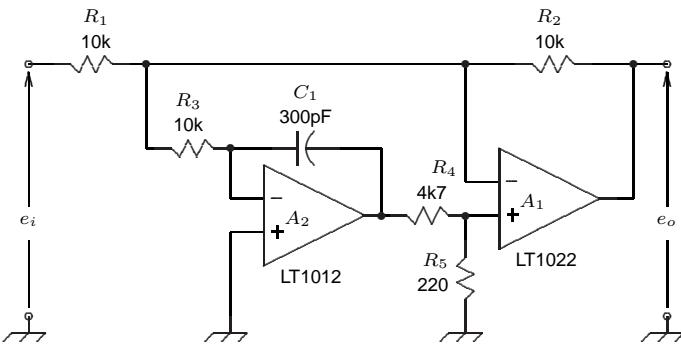


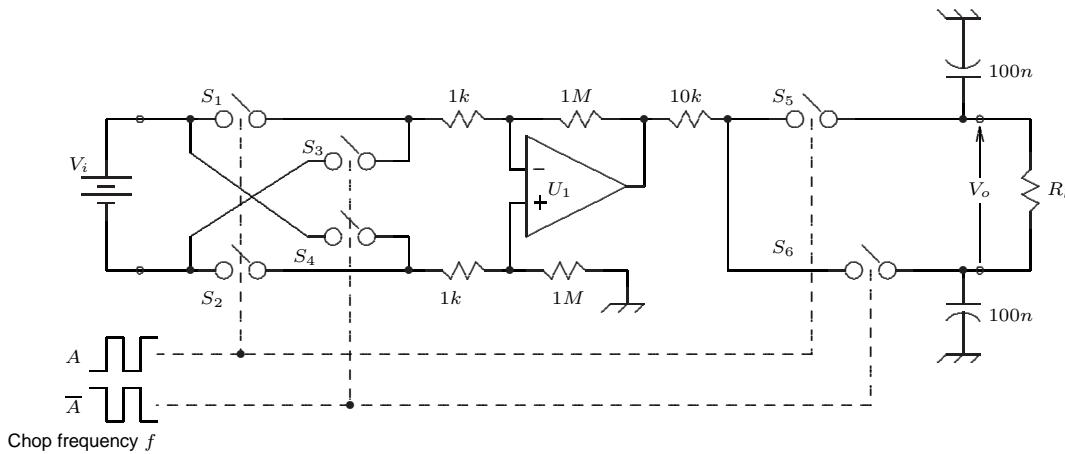
Figure 973: Composite Amplifier

35.12 Exercises

1. **Comparator** Referring to figure 958, section 35.6 on page 1071:

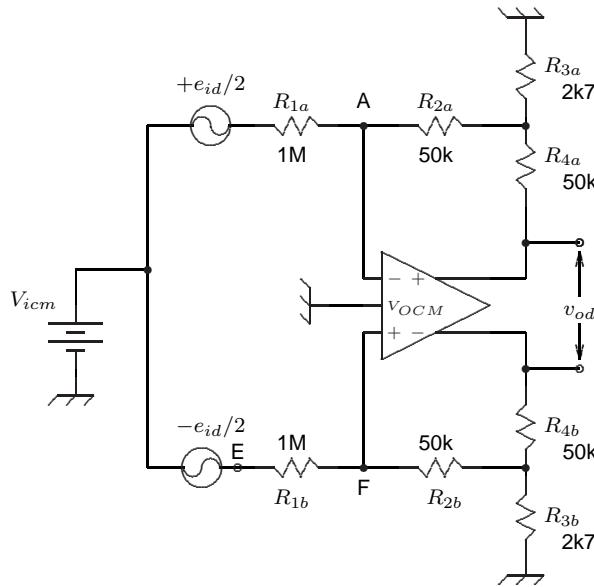
- (a) Calculate a value for R_4 so that the LED current is 10mA when it is illuminated. You can assume that the LED voltage is 2 volts when conducting and the saturation voltage of the comparator output transistor is zero.
- (b) Calculate value for C_1 to give a 4 second delay from the time the switch is released until the LED illuminates.

2. A simple differential, chopper-stabilized amplifier, based on [339] is shown in the figure below.



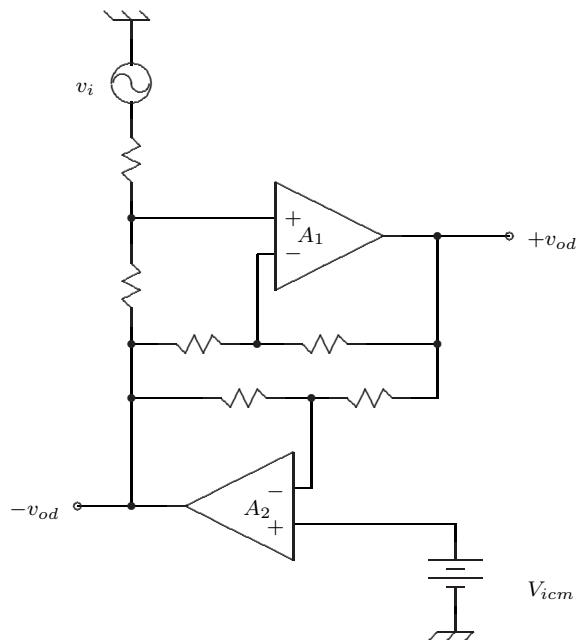
- (a) Determine the overall voltage gain of this circuit V_o/V_i .
- (b) Show that the offset voltage of amplifier U_1 has no effect on the output.
3. Regarding the composite amplifier:
- (a) Using figure 216 of a negative feedback system on page 284 as a guide, draw a block diagram of the composite amplifier shown in figure 973 on page 1089. The integrator in that diagram may be regarded as having a transfer function of
- $$\frac{1}{R_3 C_1 s}$$
- Amplifier A_1 has a finite open-loop gain of $A_1 \text{V/V}$ and may be regarded as having infinite bandwidth.
- (b) Using the approach shown in section 10 on page 284, determine the overall transfer function e_o/e_i of this composite amplifier.
- (c) Apply reasonable approximations to this formula to get an approximate result, and explain that result.
- (d) Because of the finite open-loop gain of amplifier A_2 in figure 973, the expression
- $$\frac{1}{R_3 C_1 s}$$
- for the integrator is not completely accurate. Assuming the open-loop gain of the amplifier is 10^5V/V and the integrator components R_3 and C_1 have the values shown in figure 973, write a more accurate transfer function for the integrator. (For the purposes of this calculation, you may assume that the bandwidth of amplifier A_2 is large enough to ignore its effect.)
- (e) Using this more accurate characterization of the transfer function of the integrator, determine the overall transfer function e_o/e_i of the composite amplifier. Make reasonable approximations and explain the result.
4. A *thermocouple* is a device used to measure high temperatures in ovens that are used to melt metals and other high-temperature chemical processes. It consists of two dissimilar metals welded together. The output voltage from a thermocouple is approximately linear with temperature. In this case, we will assume a type S thermocouple, using platinum and rhodium metals, which generates an output of $6\mu\text{volts}/^\circ\text{C}$. This is a relatively small signal, so it usually must be amplified in a signal-conditioning amplifier before being read by a computer.
- Your assignment is to read a thermocouple signal and record it by means of a microcomputer. The A/D converter on the computer has a 5 volt span and 10 bit resolution.
- (a) If the temperature ranges up from 25 to 1200 degrees C, what voltage gain is required from the signal conditioning amplifier?
- (b) Why is a conventional operational amplifier (such as the LM324) not suitable for this application?
- (c) What general type of amplifier would be suitable? Briefly explain why.
- (d) Assuming that the signal from the thermocouple is amplified to occupy the full span of the A/D converter, what is the resolution of the system in degrees centigrade?
5. The current conveyor of figure 945 (page 1060) has a single high-impedance (current) output.

- (a) Redraw the circuit, expanding the circuit mirrors into their component transistors.
- (b) Add circuitry to create a *complementary* current output. As one output increases in current, the other one decreases.
6. Calculate the input resistance seen by the source E_{id} in figure 963. Assume that the common-mode output voltage is zero.
7. The noise specifications for the LMH6551 fully differential amplifier are $e_n = 6.0\text{nV}/\sqrt{\text{Hz}}$, $i_n = 1.5\text{pA}/\sqrt{\text{Hz}}$. The input and feedback resistors are 500Ω . Assume the bandwidth is 10MHz and the circuit operates at 20°C . Hint: A model of the noise circuit is in [333].
- Determine the differential output noise.
 - What noise sources may be ignored in this calculation, and why?
8. The fully differential amplifier circuit shown below is suggested as a method of increasing the common-mode input range independently of the differential gain.



- (a) What is the common-mode output voltage?
- (b) Assuming that the allowable common-mode voltage range at the input terminals of the amplifier is $\pm 2\text{V}$, what is the allowable range of the common-mode voltage V_{icm} ? You can assume that the differential mode voltage e_{id} is zero.
- (c) Resistors R_3 and R_4 form a voltage divider which attenuates the output signal which is fed back to the input. What is the *gain* of this resistor network?
- (d) Show that the internal resistance of the voltage divider R_3 and R_4 is much smaller than feedback resistance R_2 .
- (e) Replace voltage divider R_3 and R_4 by its Thevenin equivalent and calculate the differential voltage gain v_{od}/e_{id} .
- (f) If you had to adjust the differential voltage gain without materially affecting the common-mode input range, what would you adjust?

- (g) Redesign the circuit to double the input common mode voltage range, keeping the differential voltage gain constant.
- (h) Consider that there is a noise voltage source e_n volts in series with one of the amplifier inputs. What is the magnitude of this noise signal at the output of the amplifier?
9. The circuit shown below is suggested [340] as a method of implementing a fully differential amplifier using conventional operational amplifiers. The circuit converts a single-ended input v_i into differential outputs $+v_{od}$ and $-v_{od}$. The common-mode output voltage is established at a separate input terminal, v_{icm} . All the resistors are equal value.



- (a) Confirm the behaviour of this circuit.

Hints:

- Treat the differential input voltage and the common mode voltage separately.
- It's easiest to start at the output of the circuit and work back to the input. Assume some output voltage, relate that to the other output voltage and then to the input voltage.

- (b) The schematic shows a complete signal loop around the A_1 , A_2 amplifiers. Is this likely to create instability? Explain.

10. When attempting to measure a small differential voltage riding on a large common-mode voltage, the standard 4-resistor differential amplifier is one possibility (section 13.6 on page 339):

As shown in figure 974(a), the input resistors are made much larger than the feedback resistors. As a consequence, the common-mode voltage is reduced – by a factor of 20 for the circuit shown. So, for example, if the op-amp could cope with an input common-mode voltage range of ± 5 volts, the circuit could cope with a common-mode voltage range of ± 100 volts.

Unfortunately, attenuating the common-mode voltage attenuates the differential voltage by the same factor. Then a small differential signal may be compromised by noise.

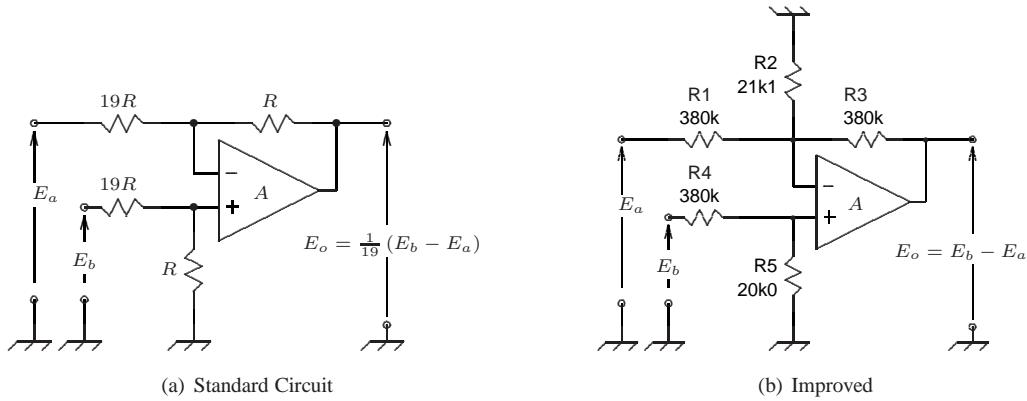


Figure 974: Difference Amplifiers

What we really need is a circuit that will attenuate the common-mode voltage without attenuating the differential voltage. This is the function of the Burr-Brown INA117 difference amplifier [341], shown in figure 974(b). The resistors in this chip are laser-trimmed to ensure close matching, which allows very large common-mode rejection.

- Using superposition, confirm that the common-mode output of the 4-resistor circuit in figure 974(a) is attenuated by a factor of 19.
- To analyse the circuit of figure 974(b), use Thevenin's theorem to reduce R1 and R2 to an equivalent voltage and internal resistance. Repeat for R4 and R5. Now apply superposition to determine an expression for the output voltage in terms of the inputs.

36 Multipliers

36.1 Introduction

There are several applications that require the product of two electronic signals. The general symbol for a multiplier is shown here:

Instrumentation. Many instrument applications require multiplying two analog signals. For example, multiply e_1 , which is proportional to some voltage $e(t)$ by e_2 , which is proportional to some current $i(t)$. Then the output of the multiplier is proportional to the instantaneous power $p(t) = e(t) \cdot i(t)$.

Amplitude Control Electronic control of a signal amplitude, such as used in an audio mixing desk, is accomplished using a VCA, *voltage controlled amplifier*. A VCA is in effect a two-quadrant multiplier, where one input is an AC signal and the other input is a control voltage.

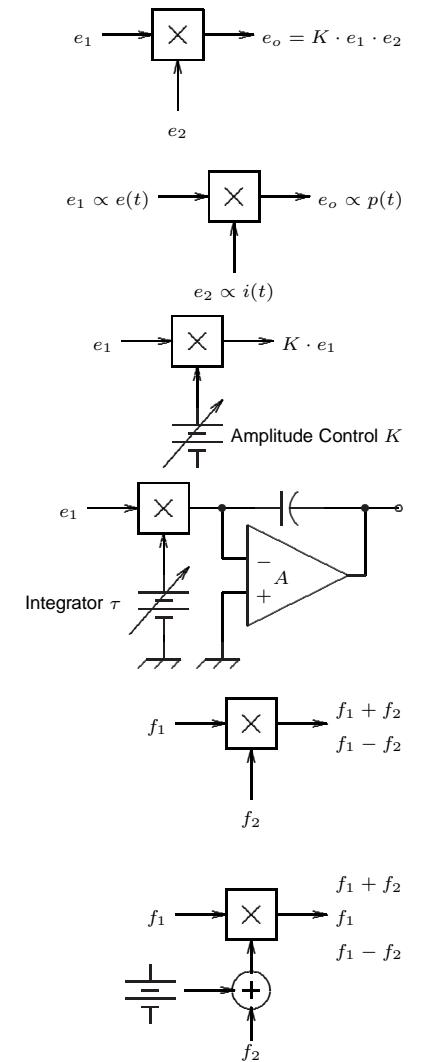
Filter Frequency Control The state variable bandpass filter (section 17.12, page 546) contains two integrators. If the integrator input resistors are replaced by multipliers, then a DC control signal can be used to sweep the centre frequency of the filter. In effect, the multiplier control voltage adjusts the time constant τ of the integrator.

Modulation, DSB The multiplication of two signals is fundamental in signal processing. For example, multiplying two signal frequencies f_1 and f_2 generates new frequencies at the sum and the difference frequencies of the two inputs [342], [343]. In radio parlance this is known as a DSB (double sideband) generator. This technique also forms the basis for the heterodyne method of spectrum analysis [147].

Modulation, AM Multiplication is also the basis of *amplitude modulation*. One input to the multiplier, the so-called *carrier*, is at a radio frequency. The other input is the *signal*, which is the information to be transmitted, such as a voice signal from a microphone. The signal is biased so that it has a DC component and never goes negative. The multiplier output is an AM (amplitude modulated) radio-frequency signal, such as used on the AM broadcast band.

The reverse process, demodulation, occurs in a radio receiver. The incoming radio frequency signal is multiplied by a *local oscillator* signal. A true multiplier can be used as the modulator. However, in practice the multiplier is often a specialized circuit using switching techniques.

Audio Signal Processing Multiplication is used in the production of electronic music. An interesting application in sound reinforcement is the *frequency shifter* [190] which reduces acoustic feedback (section 19.3 on page 617). This is an example of *single sideband modulation, SSB*.



(see figure 524, page 616)

Analog or Digital Multiplication?

It's possible to obtain the product of two signals by analog or digital means. The analog approach uses special-purpose circuitry, described below. The digital approach uses two A/D converters to change the analog signals to a stream of digital numbers which are then passed to a microprocessor. Software then multiplies these two data streams. What are the relative merits of the two approaches?

- **Frequency Range** The analog approach is much faster, extending to signals of hundreds of MHz. The digital approach requires that the signals be sampled at least twice per cycle, which limits the maximum frequency to tens of kHz in commodity market microprocessors. On the other hand, special-purpose DSP (digital signal processors) are approaching radio frequencies.
- **Overhead** The analog method can be accomplished with one or two ICs, and produces an analog result. If the requirement is analog-in-analog-out, then this is attractive. A digital method has the additional overhead of the A-D converters. It is however attractive if the result must be presented in digital form, on a display for example: an analog-in-digital-out system.
- **Accuracy** Analog methods are accurate to 1% or 0.1% accuracy, depending on the method. Digital methods depend entirely on the digitization process. For example, the least significant bit of 12 bit data represents about 0.02% of the full scale signal. (This requires an accurate A/D reference and noise-free, linear A/D conversion, which is not trivial.)
- **Flexibility** Changing a digital process requires modifying the software, which is relatively simple³¹³. Changing an analog circuit is not so straightforward.

The digital method is gradually replacing analog circuitry where multiplications are concerned, but there are still applications that can benefit from an analog solution. A competent engineer will be able to design analog and digital circuitry, and thereby choose an optimal solution for the system.

Multiplier Quadrant

In section 25.5 (page 740), in connection with the MDAC (multiplying digital-analog converter) we introduced the concept of multiplier *quadrant of operation*. The quadrant of operation describes the allowable inputs to a multiplier. These are summarized in the table below.

Quadrants	Description	Example
One	Both signals must be uni-polar. For example, both signals could be positive.	Current mode digital-analog converter
Two	One signal is uni-polar, the other can be both polarities (AC)	VCA, Voltage Controlled Amplifier, signal is AC, amplitude control is uni-polar DC
Four	Both signals can be either polarity	DSB (Double sideband modulator) in communications system.

Now we are in a position to describe various multipliers.

³¹³Providing that the development tools and documentation (such as the source code) are available.

36.2 MDAC: Multiplying Digital to Analog Converter

In section 25.5 (page 739) we introduced the MDAC. This is a bona-fide multiplying device, where one input is an analog signal, the other is digital. Section 25.6 (page 741) showed example applications, where the MDAC is used to control amplifier gain and function generator frequency.

36.3 Duty Cycle (Pulse Width-Height) Multiplier

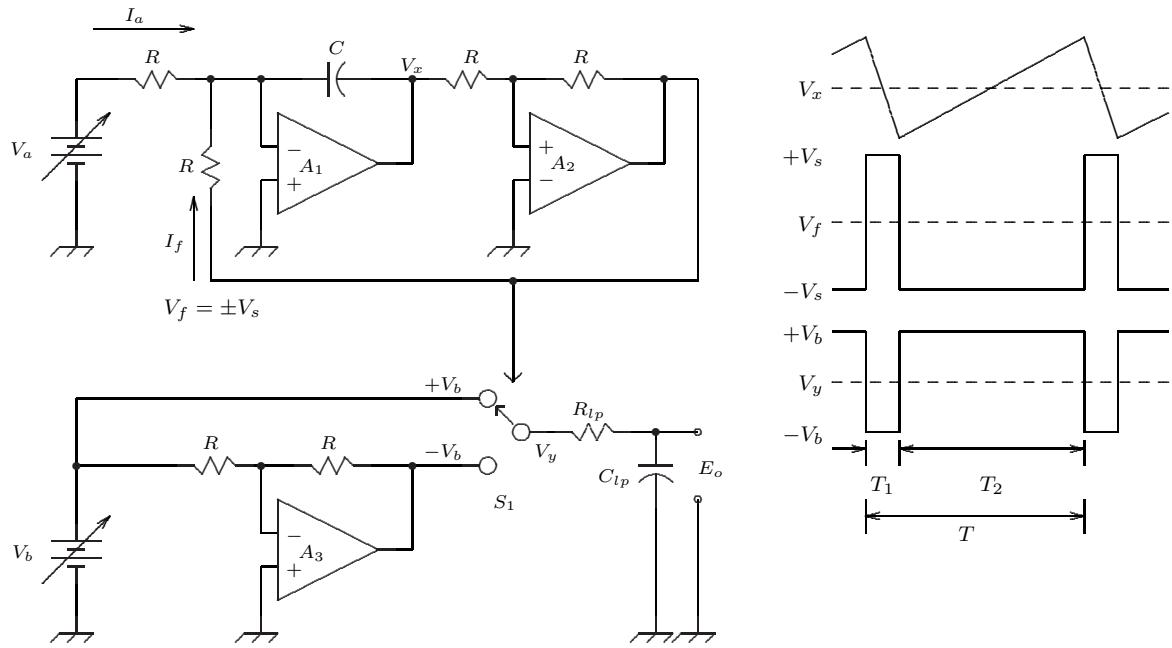


Figure 975: Duty Cycle Multiplier

This multiplier exploits the fact that the area of a rectangle is the product of its width and height. The circuit shown in figure 975 shows an example [344].

The upper part of the circuit generates a pulse waveform with a duty cycle proportional to input voltage V_a . This pulse waveform controls a SPDT switch³¹⁴ so that the output of the switch is a pulse waveform with duty cycle proportional to V_a and a height equal to V_b . The RC lowpass network R_pC_p then extracts the average of this pulse waveform, which is proportional to the product of V_a and V_b .

Analysis

This is a 2-quadrant multiplier. Input V_a must be positive. Input V_b can be either polarity.

Amplifier A_1 is an integrator, amplifier A_2 is a Schmitt trigger. Together, they form a function generator type oscillator (section 20.7 on page 633). If input V_a is zero, then the integrator produces a triangle wave and the Schmitt spends equal time with a *high* output and *low* output, that is, the duty cycle of the Schmitt output is 50%.

Now consider what happens when voltage V_a increases. The integrator output ramps downward more quickly, and the Schmitt output duty cycle changes. We could work through a detailed analysis of the voltages and currents, but there is an insight that simplifies the understanding and analysis of this circuit.

³¹⁴In practice, this is an electronic switch as described in section 33 on page 1009. The CD4066 was used in the original circuit.

For the circuit to operate properly, the average current into the integrator must be zero. If that were not true, the integrator output would gradually increase indefinitely in one direction or the other. That's not what happens: the Schmitt trigger output adjusts its duty cycle to bring the system back into equilibrium. The two amplifiers constitute a negative feedback system. In equation form:

$$\overline{I_a} + \overline{I_f} = 0 \quad (1627)$$

where the overline indicates *average*, I_a is the current due to the input voltage V_a and I_f is the current due to the feedback voltage V_f .

Recognize that the summing junction for these two currents is the input of a negative feedback system, so it is at a virtual earth point. Then we can relate these currents to the voltages:

$$I_a = \frac{V_a}{R} \quad (1628)$$

$$I_f = \frac{V_f}{R} \quad (1629)$$

Back substitute in equation 1627 and we have:

$$\overline{V_a} + \overline{V_f} = 0 \quad (1630)$$

Input voltage V_a is fixed, so $\overline{V_a} = V_a$. For the average value of the feedback voltage, refer to figure 975.

The feedback signal has an amplitude³¹⁵ of $\pm V_s$. Then the average value of the feedback signal V_f is:

$$\begin{aligned} \overline{V_f} &= \frac{1}{T} \int_0^T v_f(t) dt \\ &= \frac{1}{T} (V_s T_1 - V_s T_2) \end{aligned} \quad (1631)$$

$$= V_s \left[\frac{T_1 - T_2}{T} \right] \quad (1632)$$

Substitute in equation 1630, and we have:

$$V_a = -V_s \left[\frac{T_1 - T_2}{T} \right] \quad (1633)$$

Now refer to the lower part of the circuit, which generates the output voltage. The output voltage E_o is the average of the switch voltage V_y :

$$\begin{aligned} E_o &= \overline{V_y} \\ &= \frac{1}{T} \int_0^T v_y(t) dt \\ &= \frac{1}{T} (-V_b T_1 + V_b T_2) \end{aligned} \quad (1634)$$

$$= -V_b \left[\frac{T_1 - T_2}{T} \right] \quad (1635)$$

³¹⁵This can be achieved by using an op-amp that produces a rail-rail output (that is, an output that goes to the two supply rails). Such op-amps did not exist at the time this circuit was invented, and it used a second analog switch like S_1 to select the two voltages.

We can rearrange equation 1633 as follows:

$$\frac{V_a}{V_s} = - \left[\frac{T_1 - T_2}{T} \right] \quad (1636)$$

Substitute for the quantity on the RHS of equation 1636 into equation 1635, and we have finally:

$$E_o = V_b \frac{V_a}{V_s} \quad (1637)$$

That is, the output is equal to the product of the two input voltages, scaled by V_s . For simplicity, V_s might be chosen to be 10 volts, in which case $E_o = 0.1V_aV_b$.

36.4 Transconductance Multiplier

To find the product of two numbers, one may add their logarithms and take the antilogarithm of the result³¹⁶.

The transconductance multiplier uses the logarithmic and exponential behaviour of the junction transistor to implement multiplication [79], [252]. The result is a multiplier with an accuracy in the order of 1% and a frequency response in excess of 100MHz.

Rather than proceeding directly to the transconductance multiplier, we first introduce the *2-quadrant differential amplifier multiplier*. We then show how *pre-distortion* can improve the signal handling capability of the multiplier.

The 2-quadrant Differential Amplifier Multiplier

In section 35.5.2 (page 1070), we showed how the operational transconductance amplifier (OTA) can operate as a 2-quadrant multiplier in applications such as a voltage controlled amplifier.

The multiplication action occurs because

- The gain of the amplifier depends on the transconductance g_m (or equivalently, the incremental emitter resistance r_e) of the input differential pair.
- The transconductance is related to the bias current (the tail current of the differential pair).
- Consequently, increasing the bias current increases the gain and consequently the magnitude of the output signal.
- The bias current can be made proportional to a so-called *control voltage* e_c .

The net result is:

$$e_o = k e_c e_i$$

Where e_i and e_o are AC signals, e_c is a DC control signal that varies between zero and some maximum value, and k is a constant.

³¹⁶Before the invention of the pocket scientific calculator, multiplication was widely accomplished by this method, aided by lengthy tables of logarithms and anti-logarithms. For moderate precision (in the order of 1%) the slide rule was faster and worked well.

The Problem of Limited Input Amplitude

While this circuit can act as a multiplier, the magnitude of the input signal e_i is very limited. Figure 826 (page 936) shows how the current in a differential pair is related to the differential voltage. The relationship is only approximately linear for voltages in the order of V_t (26mV). The schematic of an OTA 2-quadrant multiplier in figure 957 on page 1070 shows one way of coping with this limited range of amplitudes. It is the function of resistor divider R_1 and R_2 to ensure that the signal is reduced to an acceptable level, a few tens of millivolts, at the input of the OTA. Larger than that level will distort the signal significantly.

Gilbert showed in [136] and [345] that this problem may be dealt with by *predistortion diodes*, as shown in figure 976. The input signal is converted to complementary current drive currents $(1 + x)I_1$ and $(1 - x)I_1$. The input signal causes the value of x to vary between 0 and 1. (We'll refer to x as the *modulation index* and show the circuit that accomplishes this magic in a moment.) Diode-connected transistors Q_1 and Q_4 generate a collector-emitter voltage that is proportional to the logarithm of these currents. The currents in the differential pair Q_2 and Q_3 respond exponentially to these voltages. As a net effect, the logarithmic operation followed by an exponential operation makes the current in the differential pair directly proportional to the drive currents, without distortion. Here's the analysis:

Analysis

Take KVL around the base-emitter loop of the four transistors:

$$+ V_{be1} - V_{be2} + V_{be3} - V_{be4} = 0 \quad (1638)$$

Collector current is related to the base-emitter voltage by the Ebers-Moll equation (equation 1157 on page 807).

$$I_c = I_s e^{V_{be}/V_t} \quad (1639)$$

Equation 1639 can be rearranged in the logarithmic form:

$$V_{be} = V_t \ln \frac{I_c}{I_s} \quad (1640)$$

Substitute for the V_{be} s in equation 1638, assuming that the transistors are identical, that is, V_t and I_s are the same for each of the transistors.

$$+ V_t \ln \frac{I_{c1}}{I_s} - V_t \ln \frac{I_{c2}}{I_s} + V_t \ln \frac{I_{c3}}{I_s} - V_t \ln \frac{I_{c4}}{I_s} = 0 \quad (1641)$$

Simplify to:

$$\ln I_{c1} - \ln I_{c2} + \ln I_{c3} - \ln I_{c4} = 0 \quad (1642)$$

Adding logarithms is equivalent to multiplication, subtraction to division, so:

$$\ln \left[\frac{I_{c1}}{I_{c2}} \frac{I_{c3}}{I_{c4}} \right] = 0 \quad (1643)$$

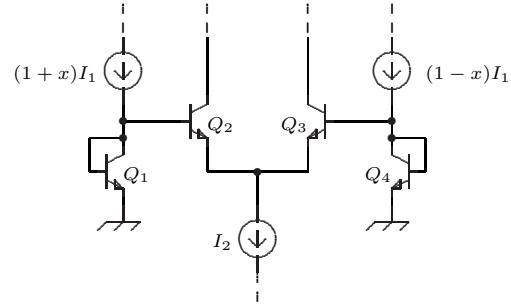


Figure 976: Predistortion Diodes

Take the anti-logarithm, and we have:

$$\frac{I_{c1}}{I_{c2}} \frac{I_{c3}}{I_{c4}} = 1 \quad (1644)$$

Substitute $I_{c1} = (1+x)I_1$ and $I_{c4} = (1-x)I_1$, and after some algebra:

$$I_{c3} - I_{c2} = x(I_{c3} + I_{c2}). \quad (1645)$$

The LHS $I_{c3} - I_{c2}$ is the differential current output, which can be extracted using some sort of analog current-voltage and subtraction circuit. The current $I_{c3} + I_{c2}$ is simply equal to the diff amp tail current I_2 . Then equation 1645 can be written:

$$I_{c3} - I_{c2} = xI_2. \quad (1646)$$

This is a 2-quadrant multiplier where the x factor is proportional to the input AC signal and I_2 is proportional to the control voltage. The output differential current $I_{c3} - I_{c2}$ is therefore proportional to the product of the input AC signal and the control voltage. Notice that there are no small signal approximations involved: the modulation index can vary between 0 and 1. For an AC input, it would have an average value of 0.5 with a variation of ± 0.5 maximum.

Generating the Drive Currents

As first step, we'll revise figure 976 to change the position of the pre-distortion diodes Q_1 and Q_4 as shown in figure 977(a). It turns out that this has no effect on the basic operation of the circuit, and the second version is more convenient to drive with NPN transistors³¹⁷.

The complementary current drive circuit is in figure 977(b). We met this circuit earlier as a differential amplifier with a single gain setting resistor, section 30.15, page 946.

Voltage V_x is an input to the multiplier. The base-emitter voltages cancel, so V_x appears across resistor R_x causing a current through it of

$$I_x = \frac{V_x}{R_x} \quad (1647)$$

Notice that the voltage V_x can be quite large. It is not restricted to a few millivolts as required in the pure differential amplifier.

The current I_x through R_x cannot flow through either of the I_1 current sinks. Therefore, by KCL, the emitter currents of the transistors Q_5 and Q_6 must change to suit. If the transistor betas are large, then their emitter and collector currents are approximately equal. Then the collector currents are:

$$I_{c5} = I_1 + I_x \quad (1648)$$

$$I_{c6} = I_1 - I_x \quad (1649)$$

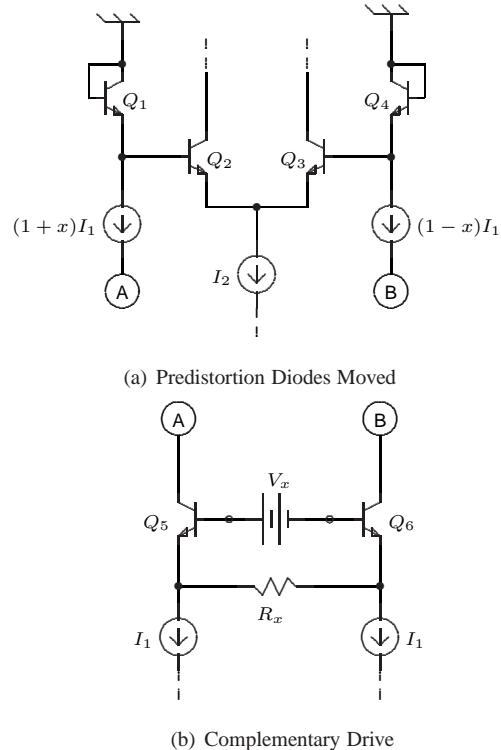


Figure 977: Generating Drive Currents

³¹⁷In general, NPN transistors are easier to find as matched pairs and they have a wider bandwidth than equivalent PNP devices.

The magnitude of I_x must be less than I_1 in order not to shut off Q_5 or Q_6 . For notational convenience, let's create another variable x such that:

$$I_x = xI_1 \quad (1650)$$

where x is some number between 0 and 1. Now equations 1649 and 1649 can be rewritten as

$$I_{c5} = I_1(1 + x) \quad (1651)$$

$$I_{c6} = I_1(1 - x) \quad (1652)$$

These are the complementary currents needed to drive the circuit of figure 977(a). Notice that V_x can reverse polarity. Then x may be positive or negative, so that the circuit of figure 977(a) is a two-quadrant multiplier.

Completing 4 Quadrant Operation

The final building block for a 4 quadrant multiplier is the *Gilbert Quad*, figure 978. This consists of two differential amplifiers: the Q_7, Q_{10} pair with its tail current generator $(1 + y)I_2$, and the Q_8, Q_9 pair with its tail current generator $(1 - y)I_2$.

The base inputs are driven by pre-distortion diodes as in figure 977(a), where the simple differential amplifier is replaced by the Gilbert Quad. The complementary currents are controlled by one of the multiplier input voltages, V_x .

The tail currents in figure 978 are generated by a copy of the complementary current generator of figure 977(b), which is controlled by the other multiplier input voltage V_y .

The output of the quad is the 4 quadrant product, available as a differential signal between the collectors of the transistors.

Analysis

We previously showed (equation 1646) for pre-distorted complementary input currents that the differential output of the differential pair is x times the tail current. Consequently, we can apply this twice to the Gilbert Quad in figure 978.

For the outer differential pair:

$$I_{c7} - I_{c10} = x(1 + y)I_2 \quad (1653)$$

Similarly, for the inner differential pair:

$$I_{c8} - I_{c9} = x(1 - y)I_2 \quad (1654)$$

Rearrange equations 1653 and 1654 to solve for I_{c7} and I_{c9} :

$$I_{c7} = x(1 + y)I_2 + I_{c10} \quad (1655)$$

$$I_{c9} = -x(1 - y)I_2 + I_{c8} \quad (1656)$$

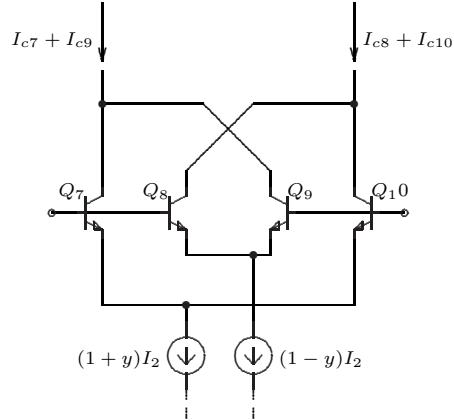


Figure 978: Gilbert Quad

The differential current output I_d from the Gilbert Quad is:

$$I_d = (I_{c7} + I_{c10}) - (I_{c8} + I_{c10}) \quad (1657)$$

Substitute for I_{c7} and I_{c10} from equations 1656 and 1656:

$$\begin{aligned} I_d &= x(1+y)I_2 + I_{c10} - x(1-y)I_2 + I_{c8} - I_{c8} - I_{c10} \\ &= 2xy I_2 \end{aligned} \quad (1658)$$

From equations 1647 and 1650:

$$x = \frac{V_x}{R_x} \frac{1}{I_1} \quad (1659)$$

Similarly for y :

$$y = \frac{V_y}{R_y} \frac{1}{I_2} \quad (1660)$$

Substitute for x and y from equations 1659 and 1660 into equation 1658, with the result:

$$I_d = V_x V_y \left[\frac{2}{I_1} \frac{1}{R_x} \frac{1}{R_y} \right] \quad (1661)$$

Since voltages V_x and V_y can be positive or negative, this defines a 4-quadrant multiplier. The scale factor is set by the quantities inside the square brackets of equation 1661.

The complete multiplier circuit is in figure 979.

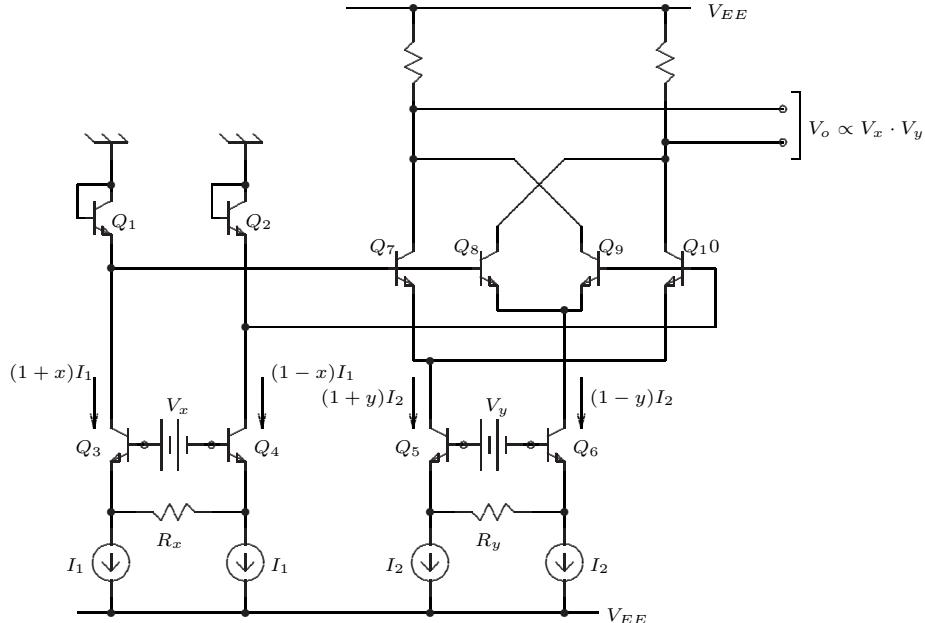


Figure 979: Complete 4-Quadrant Multiplier

The output currents are converted into voltages by 2 collector resistors. The difference could then be extracted by an op-amp differential amplifier. (Modern multiplier circuits often include this circuitry.)

36.5 Exercises

1. For the multiplier of figure 975 on page 1096:
 - (a) What are the maximum allowable input voltages?
 - (b) As the input voltage V_a changes, is there any effect on the period T of the switching waveform? How would that affect the design of the lowpass filter?
2. Figure 980 shows an alternate design for a pulse width-height multiplier. The switch is closed during the time T_{on} .

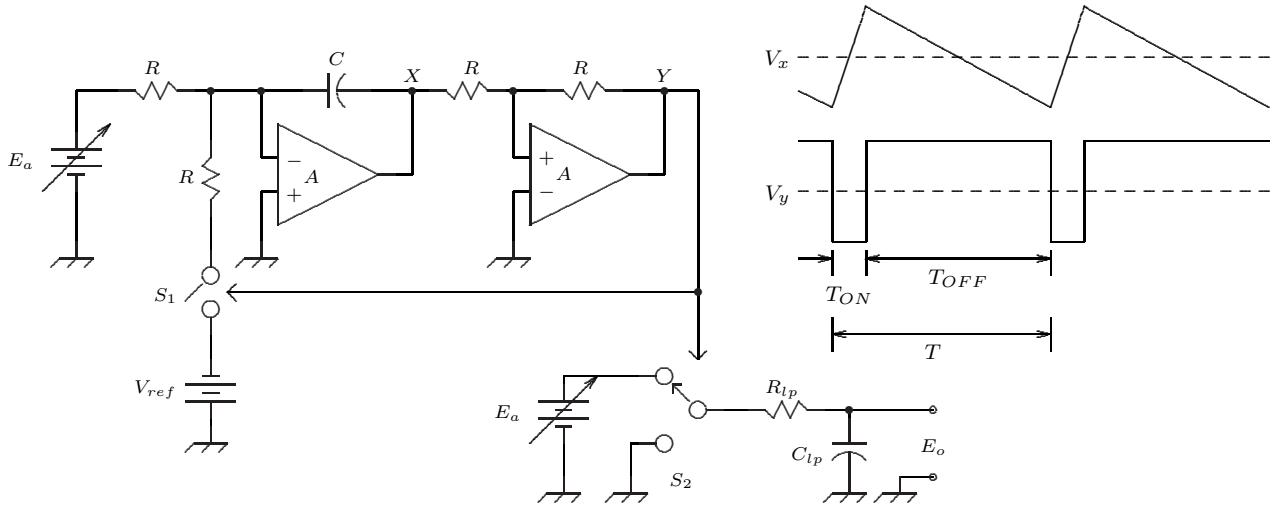


Figure 980: Alternate Duty Cycle Multiplier

- (a) Analyse the operation of this multiplier to verify that it functions correctly. You can assume that the input voltage E_a is less than the bias voltage V_{ref} .
- (b) As the input voltage E_a changes, is there any effect on the period T of the switching waveform? How would that affect the design of the lowpass filter?
3. Figure 981 shows a multiplier using a lamp and two photoresistors [346]. Resistors RP_1 and RP_2 are a matched pair of photoresistive cells (section 27.4) that are equally illuminated by the incandescent lamp PL_1 . The resistance of the cells is a non-linear function of their illumination. The op-amps may be considered to be ideal. The resistor-lamp system around A_1 is a negative feedback system.
 - (a) Show that the output is given by:

$$v_o = K(E_x \cdot v_y)$$
 - (b) Derive an expression for RP_1 in terms of the other circuit parameters.
 - (c) Why could this circuit be described as a *linearization* circuit?

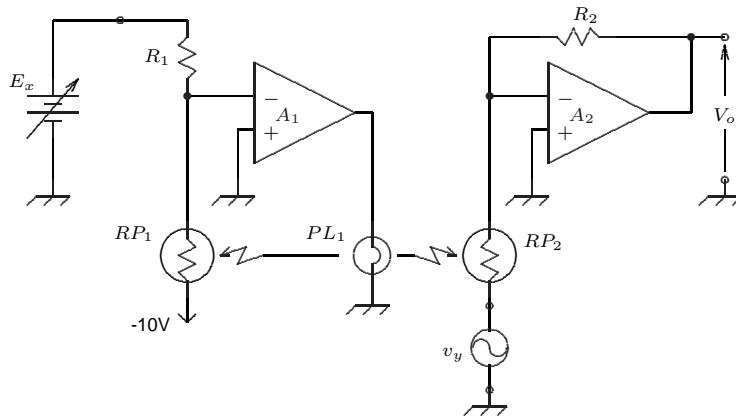


Figure 981: Photo-Resistive Multiplier

- (d) Using the previous result for RP_1 and taking into account that $RP_1 = RP_2$, derive an expression for v_o that shows 2-quadrant multiplication.
- (e) What limitation in frequency response would apply to the E_x input? To the v_y input?
- (f) This is a negative feedback system in which the input terminals of amplifier A_1 can be reversed *without affecting the operation of the system*. Explain why this is so.
- (g) The system could be modified by replacing the lamp-photocell devices with LEDs and phototransistors, or a device like the H1N11 optically coupled JFET. What effect would this have on the performance of the multiplier? Does the previous statement remain true?
4. Figure 979 shows the 4-quadrant Gilbert multiplier with four current sinks. Redraw this circuit showing BJT transistors used as current sinks in these locations.
5. A *frequency doubler* may be constructed using a multiplier as a squaring device. An AC signal at frequency f_1 is fed into both inputs of a multiplier. The output is an AC signal at frequency $2f_1$. Show the trigonometric identity that describes this operation, and sketch the input and output waveforms.

37 Complete Systems

Introduction

This section shows some examples of complete analog electronic systems.

Although digital electronics and microprocessors have taken over many electronic systems, there is still a place for systems that are entirely analog. Most signals originate as an analog voltage or current, proportional to some physical variable. And some output signals (audio, for example) have to be in analog form to drive the final transducer.

The story is expressed in the graph of figure 37.

If a microprocessor (or digital signal processor) is used to implement some processing function P_D , and the source signal is analog, then an A/D converter must be used to convert the signal into digital form for processing. Similarly, if the output is analog, a D/A converter is required to convert the microprocessor output into analog form. The D/A and A/D conversion represent additional costs and complexity for the design. Consequently, the microprocessor-based system starts from a certain *cost threshold*.

A microprocessor-based system may also incur certain unique NRE (Non Recurring Engineering) costs, such as the hardware cost of a development and emulator system, and the cost of training engineers in the software and the quirks of a microprocessor.

This cost threshold has decreased over time as much of the required additional microprocessor hardware has been incorporated into a complete microcontroller chip, available at very low cost.

However, when the signal processing is relatively simple (and/or has to be performed at high speed), then it may be more economical to build a completely analog system. This avoids the cost of A/D, D/A converters and may have other advantages, such as lower system noise or higher processing bandwidth.

Each system design has to be evaluated carefully to determine whether a microprocessor is an asset. However, a competent design engineer must be able to evaluate the requirements and design a microprocessor-based system (which may have analog processing elements in it) or a purely analog system.

For some fascinating discussions on the process of analog circuit design, see references [347] and [348].

37.1 Equation Solver

Before the advent of the modern digital computer, much simulation and the solution of differential equations was done on the *analog computer*. The analog computer is a collection of operational amplifiers that can be configured into a circuit that simulates some physical system.

The analog computer took some skill to use effectively. The problem had to be *scaled* in time and amplitude to fit the capabilities of the op-amps.

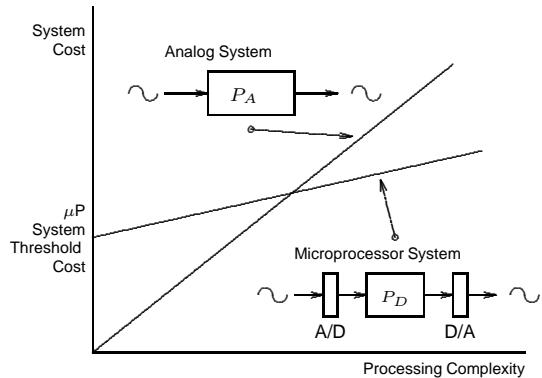


Figure 982: Analog and Microprocessor System Cost

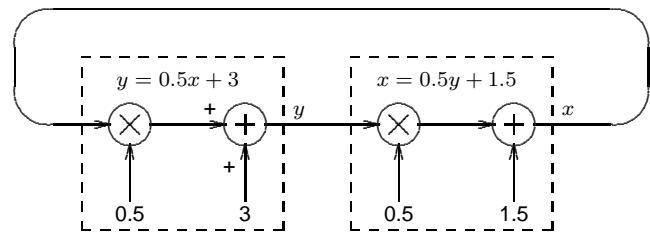


Figure 983: Equation Solver Block Diagram

Analog simulation has fallen into disuse³¹⁸, but the technology is still occasionally useful in electronic systems. The circuit shown here will simultaneously solve two linear equations, and illustrates how an analog computation is structured.

We'll start with specific equations and then generalize. Let's say that the equations to be solved are:

$$y = \frac{1}{2}x + 3 \quad (1662)$$

$$y = 2x - 3 \quad (1663)$$

To implement this as an analog circuit, we first rearrange equation 1663 in terms of x :

$$x = \frac{y}{2} + 1.5 \quad (1664)$$

Assuming x is known, equation 1663 can be used to generate a value for y . Assuming y is known, equation 1664 can be used to generate a value of x . To solve these equations simultaneously, we simply close the loop so that each result can be used by the other equation. (It's an animal that feeds on its own tail). A block diagram of the circuit is shown in figure 983.

At first blush, it may seem unlikely that you can use one unknown to solve for another in this fashion. However, we are building a device that will converge on a stable state that represents the solution to the equation. If the equations do not have a solution (if their lines are parallel, for example), then the circuit will not settle down correctly. Now we'll turn this into a circuit. The adders can be implemented by an operational amplifier adder. The multipliers can be potentiometers: multiplying a signal by a constant 0.5 is equivalent to feeding it through an equal-valued voltage divider. But there are a couple of minor details to watch: the adders normally invert, so there must be a re-inverter to produce a positive sum. And the potentiometers must drive a very high impedance if their coefficient setting is to be accurate. Keeping these points in mind, we have the circuit shown in figure 984.

To use this circuit:

- Adjust the constant potentiometers P_3 to produce 3 volts and P_4 to produce 1.5 volts. (The analog computers of yesteryear used 10 turn potentiometers with a mechanical digital readout. Such dials and pots are still available commercially.)

These pots are loaded with the input resistance of the following summer, so the dial setting on the pot may not be entirely reliable.

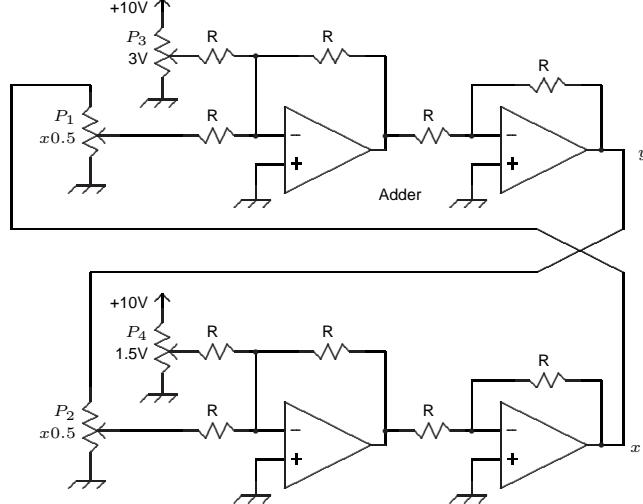


Figure 984: Equation Solver Circuit

³¹⁸It is especially ironic that analog computing fell into disuse about the same time that their hundred-dollar operational amplifiers fell to a cost in the tens of cents.

However, if the summing resistance is much larger than the pot resistance, then the loading effect will be minimal. For example, with a 10k pot and $1M\Omega$ summing resistance R , the error is in the order of 1%.

- Adjust the coefficient potentiometers P_1 and P_2 to exactly mid-point to set the coefficient to 0.5.
- The output voltages x and y , which can be read with a voltmeter, are the solution of the equations. (The results should be $x = 4V$, $y = 5V$.)

By varying the coefficient and constant potentiometers, various other equations may be solved.

Construction Notes

The tradition is to use $1M\Omega$ resistors for the input and feedback resistors R . To avoid bias current errors, use a JFET or MOSFET input operational amplifier, such as the TL074. With supply voltages of ± 15 volts, the computing signals can range over ± 10 volts without saturating the op-amps.

Notice that it is not sufficient to ensure that the output signals stay within that 10 volt range. *None* of the op-amp signals can exceed that range. (Some analog computers had *overload* lamps to indicate when this occurred).

37.2 DC Motor Controller

In this section, we show a motor speed controller which senses the load on the motor and automatically compensates for this load, thereby keeping the motor speed essentially independent of mechanical load. The effect is to allow very precise control over motor speed, which is especially useful at low motor speeds.

37.2.1 Real DC Motor Behaviour

Recall the two fundamental relationships of an ideal DC motor:

- The motor speed is proportional to the voltage across the armature.
- The motor torque is proportional to the current through the armature.

The armature resistance of any real motor is not zero. As a result, when the motor is loaded and the motor current must increase to output mechanical torque, the voltage across the armature resistance increases, leaving less voltage for the armature itself, and the speed decreases. This corresponds to our experience of a DC motor: increasing the mechanical load causes the motor current to increase and the speed to decrease.

This effect may be modelled as shown in figure 985. A real DC motor consists of an ideal motor (in which the speed is exactly proportional to the voltage across it) in series with an armature resistance R_A .

Here is the main idea for this control concept: if the voltage across the *ideal motor* could be maintained constant in spite of changes in current through the motor, the motor speed would be independent of the load torque. For this to occur the supply voltage must increase by an amount equal to the voltage drop across the armature resistance.

If we had a controller that would increase the motor supply voltage automatically in response to increased mechanical load, the *speed regulation* would be very accurate: the motor speed would not change substantially with increasing loads.

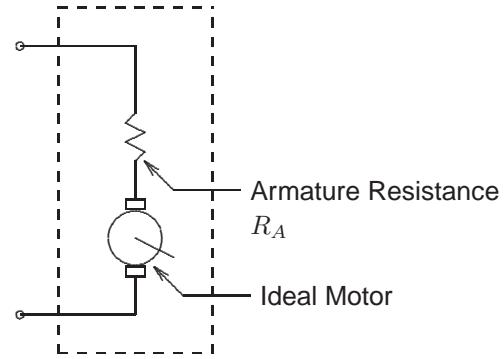


Figure 985: Real Motor, Equivalent Circuit

37.2.2 The Control System

To see how such a control system could be implemented, consider the motor control circuit shown in figure 986.

In this circuit, the amplifier controls the power supply current into the motor such that the power amplifier output voltage is the same as its input control voltage E_s . The amplifier has a voltage gain of unity and the output of the amplifier is a constant voltage characteristic. (The output voltage is unaffected by the output current.)

In the next step, we add the control circuitry for regulating the speed of the DC motor, shown in block diagram form in figure 987.

There are a number of additions to this circuit.

- A *current sensing resistor* R_s is wired in series with the motor. Without additional circuitry, this resistor would worsen the speed regulation of the motor, but we will be adding additional control circuitry to compensate for the voltage drop across this resistor. The voltage developed across this resistor is proportional to the motor torque. This same voltage is therefore also proportional to the motor torque, so the resistor creates a voltage that we can use in our control system to signify the magnitude of the motor torque.

e

- A feedback amplifier with gain $\times 2$ multiplies the voltage across R_s by 2 and delivers it back to the adder.
- The input voltage to the power amplifier is now equal to the control voltage E_s plus the voltage at the output of the feedback amplifier.

If we make the sensing resistor R_s equal to the armature resistance R_a , then the feedback system will always add on a voltage to the control voltage E_s , by an amount equal to *twice* the voltage across the armature resistance. This voltage appears at the output of the power amplifier and drives the DC motor. The net result is that the armature voltage is exactly equal to E_s , regardless of the armature current, and so the motor speed is regulated exactly.

37.2.3 Control System Equations

We can prove this last paragraph by writing the equations for the control system, as follows:

The voltage across the ideal motor armature is E_m . By Kirchhoff's voltage law around the control loop including the motor:

$$+ V_{pa} - I_m R_a - E_m - I_m R_s = 0 \quad (1665)$$

where I_m is the motor current.

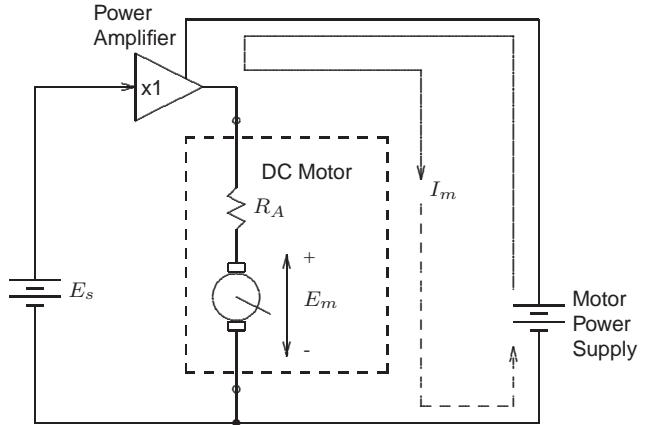


Figure 986: Motor Power Circuit

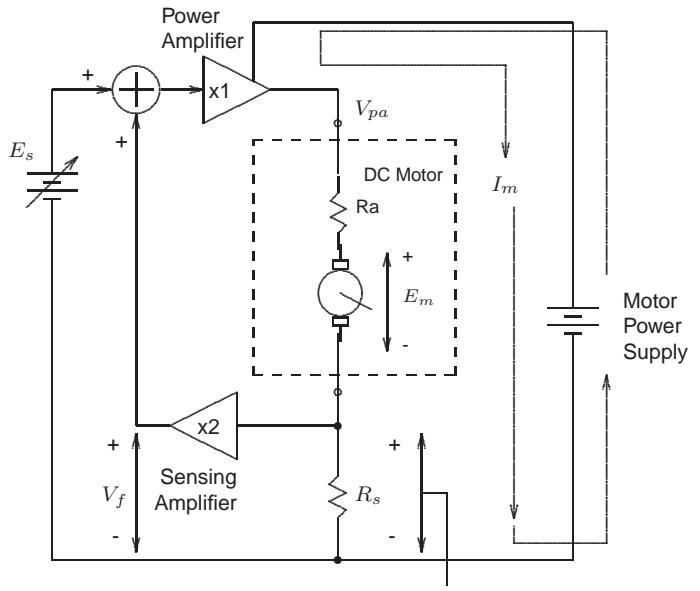


Figure 987: Motor Speed Controller Block Diagram

The power amplifier output voltage is equal to the power amplifier input voltage, which is in turn equal to the sum of the control and feedback voltages:

$$V_{pa} = E_s + V_f \quad (1666)$$

The feedback signal V_f is equal to the voltage V_s across the current sensing resistor times the gain of the feedback amplifier. In figure 987 the gain was shown as $\times 2$, but we will give it the symbol K_f , so that:

$$V_f = V_s K_f \quad (1667)$$

The voltage V_s across the sensing resistor R_s is

$$V_s = I_m R_s \quad (1668)$$

Substituting in equation 1665 from equations 1666, 1667 and 1668 we have

$$\begin{aligned} +V_{pa} - I_m R_a - E_m - I_m R_s &= (E_s + V_f) - I_m R_a - E_m - I_m R_s \\ &= E_s + (V_s K_f) - I_m R_a - E_m - I_m R_s \\ &= E_s + (I_m R_s) K_f - I_m R_a - E_m - I_m R_s \\ &= 0 \end{aligned}$$

We want the value of the motor voltage E_m to be equal to the control voltage E_s , so that we can substitute zero for $E_s - E_m$ in this equation. Substitute zero for $E_s - E_m$ in the above equation to obtain:

$$I_m R_s K_f - I_m R_a - I_m R_s = 0$$

Cancelling I_m from this equation, we have

$$R_s K_f - R_a - R_s = 0 \quad (1669)$$

One possible configuration to satisfy this equation would be to make $R_s = R_a$ and $K_f = 2$, that is:

$$2R_a - R_a - R_a = 0$$

This condition will ensure that the motor voltage E_m will be equal to the control voltage E_s , and was the setup shown in figure 987. However, we can do better than this.

The motor current is significant, so that we will waste considerable power in the sensing resistor R_s if we make it equal to the armature resistance R_a . It would be better to make it as small as possible. We can increase the gain of the feedback amplifier to compensate for the smaller value of the sensing resistance. For example, we could choose

$$R_s = \frac{1}{5} R_a$$

Then from equation 1669, substituting $R_a/5$ for R_s we have:

$$\frac{R_s}{5} K_f - R_a - \frac{R_s}{5} = 0 \quad (1670)$$

$$(1671)$$

Solving for the value of K_f , we obtain that

$$K_f = 6$$

That is, the gain of the feedback amplifier must be 6 volts/volt to obtain the same control system as before.

37.2.4 Electronic Circuit

The circuit diagram of the complete motor speed controller is shown in figure 988.

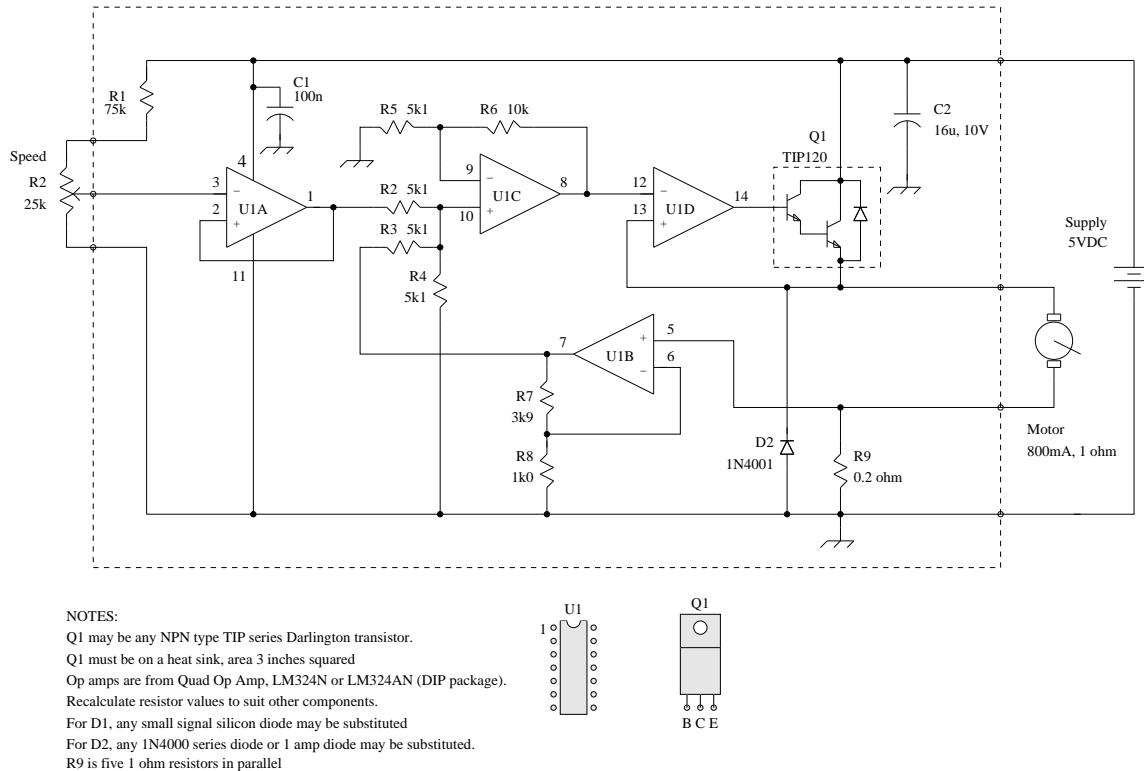


Figure 988: Motor Speed Controller Circuit Diagram

- Resistor R2 is the *speed* control. Notice that its voltage can go right down to zero, signifying that the motor speed is controllable right down to zero speed.
- Opamp U1A is configured as a voltage follower, and buffers the signal from the speed potentiometer.
- Amplifier U1C is a non-inverting amplifier (section 13.3, page 337). An inverting amplifier will not work in this circuit because the circuit is powered by a single positive power supply, so all signals must be positive with respect to ground. Resistors R2, R3 and R4 function as a lossy adder circuit: the output, which may be verified using the superposition theorem, is

$$\frac{e_1 + e_2}{3}$$

This is followed by a non-inverting amplifier of gain 3 so the output of the amplifier is simply the sum of e_1 and e_2 .

- Opamp U1D and transistor Q1 form the power amplifier, configured as a voltage follower, so that the output (motor) voltage is the same as the input voltage.

- Resistor R9 is the current sensing resistor R_s . The motor resistance R_a for this type of motor is typically about 1Ω , so that $R_s = R_a/5$, as in the previous example.
- The feedback amplifier U1B is a non-inverting amplifier with gain

$$\begin{aligned} A_v &= 1 + \frac{R7}{R8} \\ &= 1 + \frac{3900}{1000} \\ &\approx 5 \text{ volts/volt} \end{aligned}$$

According to our previous theory, the voltage gain of the feedback amplifier should be 6 volts/volt for perfect compensation. However, when the circuit was constructed and tested with a feedback gain of 6, it was found that the motor had a slight tendency to *creep* when it was supposed to stop completely. This was fixed by reducing the gain of the feedback amplifier, thereby undercompensating the system slightly, so that the motor is slightly affected by load torque.

37.2.5 Results

The circuit does maintain constant motor speed in spite of different mechanical loads. To accomplish this, the control system has to supply enough current to generate the required mechanical torque. The main control transistor Q1 must dissipate a power equal to the motor current times the difference in supply and motor terminal voltage. This causes considerable dissipation in the transistor, which must be removed by a heatsink, and the efficiency is low compared to a duty-cycle switching scheme (section 37.3.3).

The circuit operates correctly for a particular value of motor armature resistance. If the motor type is changed, then the sensing circuit must be modified to suit.

Since this compensation is a form of *positive* feedback, care must be taken that the loop gain does not exceed unity.

An alternative possibility for speed control is a system of tachometer feedback, in which the motor speed is sensed and used as a negative feedback signal. This eliminates the dependence on motor resistance, and acquires the advantages of a negative feedback system.

37.3 Sumo Competition Robot

Designing and constructing a mobile robot is an excellent project for a course in electrical engineering or technology. This mobile robot platform is equipped with circuitry that enables it to participate in a Sumo robot wrestling competition (lightweight autonomous class).

Two robots are placed in a circular raised plywood ring area. They move about in the ring, detecting the edge of the ring and backing away from it. If they sense another robot, they move toward that robot and try to push it out of the ring, without falling out themselves.

Most robot designs use analog sensor circuits and one or more microcomputers for control and guidance. This design, the *Mechbot* robot (because it was originally used in a course for mechanical engineers) uses analog circuits entirely³¹⁹.

Figure 989 shows a photograph of the robot platform. Figure 990 on page 1113 shows a three-view of the robot.

37.3.1 Robot Behaviour

The Sumo robot has the following behaviours:

- The motors normally operate at a low speed (*cruise mode*). They operate open-loop but with individual speed adjustment. Then when both motors are operating, the robot speed adjustment ensures that the robot drives approximately in a straight line.
- Two object detectors face forward at the front of the robot. When either object detector senses an object, the robot assumes that object is another robot. It moves the opposite motor into high speed (*engage mode*). The effect is to rotate the robot so that the other detector and/or whiskers can detect the object. When the second detector detects the robot, both motors are then in engage mode and the robot drives into the obstacle and tries to push it out of the ring.
- Two microswitches (*whiskers*) are located at the front of the robot. When a whisker is triggered (ie, the switch is closed), the corresponding motor goes into high speed. The effect is to ensure that the motors stay in high speed mode as long as the robot is pushing the opponent.
- There are two edge detectors in the front corners of the robot. When either of these detects the black line that defines the edge of the ring, it throws both of the motors into reverse. A time delay circuit runs one motor in reverse longer than the other, so the robot does a *back and turn* manoeuvre away from the edge. The direction of the back-and-turn depends on which edge detector was triggered. The edge detector circuitry works regardless of the motor speed, so even if the motor is in high-speed mode, it will back away from edges.

A circuit prototyping area is mounted on a foamcore plastic base. Two DC servomotors are mounted on L brackets under the plastic base and attached to model aircraft wheels. A castoring wheel at the rear allows the robot to swivel in response to differential steering from the driven wheels.

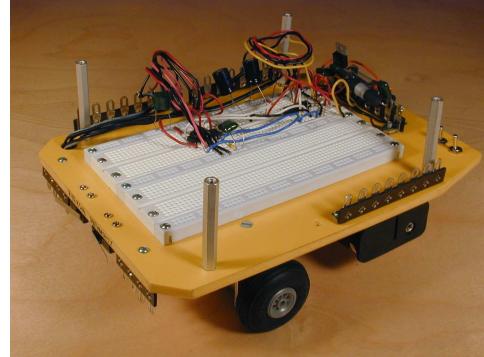


Figure 989: Mechbot

³¹⁹ The design of the robot presented here was inspired by *Basey*, an analog-guided robot by Andrew O'Malley [349]. The detail design and concept verification was done by Devin Ostrom of Ryerson University.

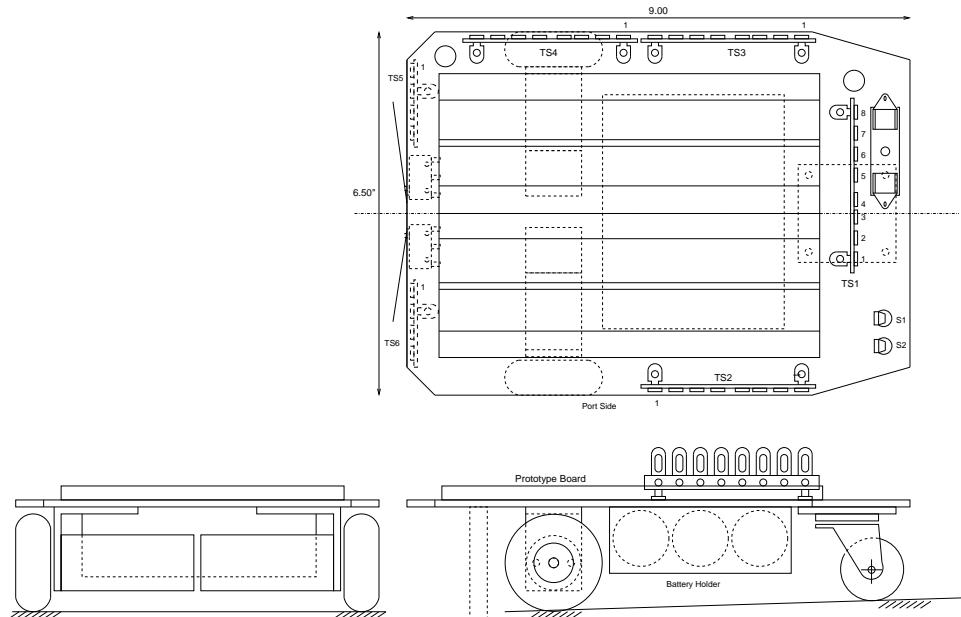


Figure 990: Mechbot

37.3.2 Robot Power Supply

The design of the power supply is heavily dependent on the specifications of the DC motors. The gearmotor chosen was a 12VDC unit. At 9VDC, the output shaft turns at about one revolution per second, which results in a suitable robot speed. The current draw is a modest 20mA per motor at 9VDC.

The battery type and voltage is a very important decision. A rechargeable 9 volt nicad battery was considered but not chosen for two reasons: the battery and charger are expensive, and a nicad battery is capable of very large short-circuit current, which is a hazard in this application.

1.5 volt cells are used as the basis for the battery because they are readily available and inexpensive. The required voltage determines the number of cells in series. The required battery life (amp-hour capacity) determines the physical size of the cells.

Battery capacity is a complex issue. Batteries of different types discharge at different rates, so the capacity depends strongly on the end-point voltage. Battery capacity also depends on the current being drawn from the battery, and whether the battery is given a chance at intervals to recover.

Battery Type	Eveready Number	Capacity, Amp-hours	Current mA
AAA	912	0.2	300
AA	1215	1.09	375
C	935	1.56	375
D	950	4.06	375
9V	1222	0.28	60

Figure 991: Battery Capacity

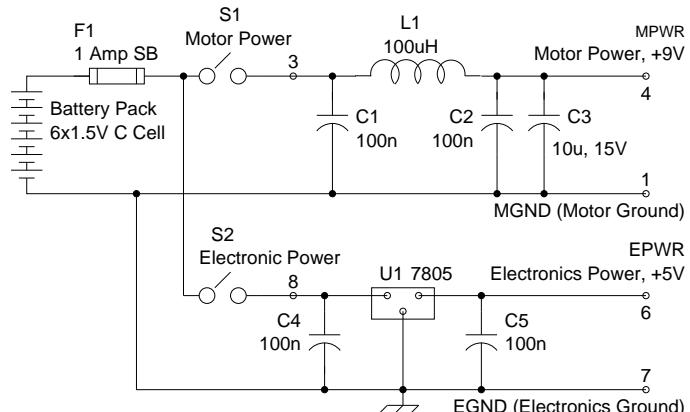


Figure 992: Power Supply Wiring

The capacity figures in figure 991 were determined from the Eveready Battery Engineering Handbook [350], but should be regarded only as a very rough guide. (The last column shows the battery load current that was used to determine the battery capacity. Ideally, this should be in the same order-of-magnitude as the actual load current. A current smaller than this will result in somewhat larger capacity.) The robot battery current was estimated to be about 100mA worst case, so a set of C sized batteries provide an endurance of 15.6 hours.

Notice that the 9 volt battery (the so-called *transistor radio* battery) would operate the robot for about 2.8 hours.

The circuit of the robot power supply is shown in figure 992. The power supply provides a high-current 9 volt supply (6, 1.5 volt C cells in series) for the motors and other high current loads. A 5 volt logic supply is derived from the 9 volt supply by means of a three-terminal regulator. This will operate correctly until the battery voltage drops below 7 volts.

DC motors generate significant *hash* electrical noise on the power supply line. The filter comprised of C1-3, L1 – and the 5 volt regulator U1, prevent this noise from finding its way into the analog circuits.

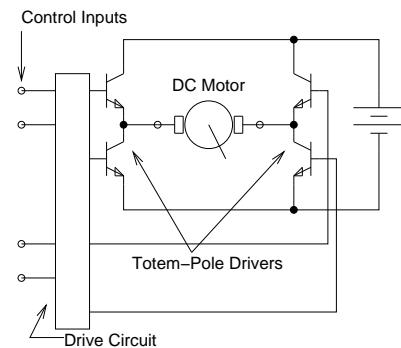
Notice how the motor ground and electronics ground are brought separately to the power supply and then joined at one point. This prevents conducted noise on the motor ground line from corrupting the analog ground.

37.3.3 Motor Driver Circuit

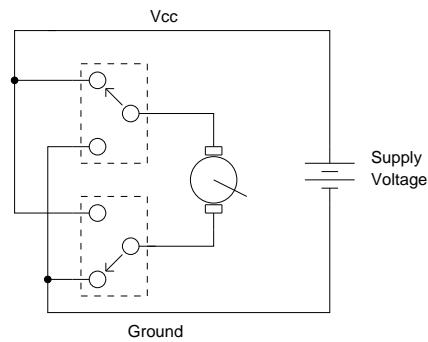
The motor control circuit accepts 5 volt logic-level signals and reverses voltage to the motor in order to reverse its rotation. Fortunately, there is an integrated circuit, the Texas Instruments L293³²⁰, which contains most of the circuitry to drive two motors.

The L293 contains four BJT totem-pole output circuits. Two outputs can implement the motor control circuit shown in figure 993(a). As shown in figure 993(b), each totem-pole output is equivalent to an SPDT switch. The polarity of the motor voltage reverses when the switches reverse position.

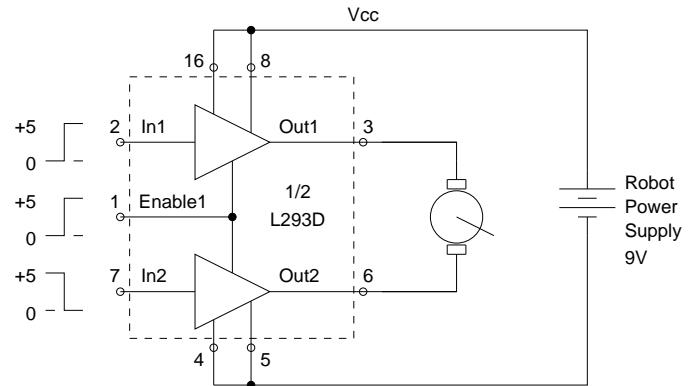
A more complete diagram is shown in figure 993(c). The 5 volt inputs to the two driver sections must be operated in antiphase in order to simulate the operation of the SPDT switches in figure 993(b).



(a) Totem-Pole Drivers



(b) Equivalent Circuit



(c) Driver Circuit

Figure 993: Motor Driver Circuit

³²⁰The L293 is widely second-sourced from other vendors. It is capable of one ampere output current, includes flyback diodes on the outputs and includes a temperature sensing circuit which disables the outputs at high temperature.

If the motor is enabled at all times, the L293 *Enable* input is simply connected to a +5 volt input. To disable both drive outputs simultaneously, the *Enable* input is grounded. (Disabling an output is accomplished in the chip by turning off *both* transistors in the totem-pole output.)

If the *Enable* input is duty cycle modulated at some frequency, then the ratio of ON time to OFF time controls the average motor voltage and consequently the motor speed. Since the totem-pole transistors are either ON or OFF, they dissipate very little power. This digital speed control scheme is much more efficient than continuous adjustment of the voltage across the motor.

37.3.4 Edge Detection Subsystem

A block diagram of the edge detection and back-and-turn concept is shown in figure 994.

The two photoresistive cells PR1 and PR2 are located at the front corners of the robot. The photoresistive cells are normally over a light-coloured playing area. Then the voltage into comparators U1 and U2 is below the voltage set by the potentiometer P1, and the output of the comparators is low.

When a photoresistive cell extends into the void beyond the playing area, it receives little reflected light so its resistance increases and the output voltage from its voltage divider also increases. If this voltage is above the voltage set at the threshold pot P1, then the comparator produces a high voltage.

Consequently, the outputs of the comparators U1 and U2 are low for *inside* playing area and high for *outside* playing area.

The transition of a comparator output voltage to indicate *outside* the playing area triggers two time delays, shown on the diagram as 0.2 seconds and 0.5 seconds. These delays are the backup times for the two motors when the robot detects an edge.

A pair of diodes for each motor act as a logical OR gate. For example, when the Port detector is triggered, this causes the port motor to back for 0.2 seconds and the starboard motor for 0.5 seconds. The net effect is to back the robot away from the edge, but longer on the side that did not detect the edge. This tends to rotate the robot more toward the centre of the field.

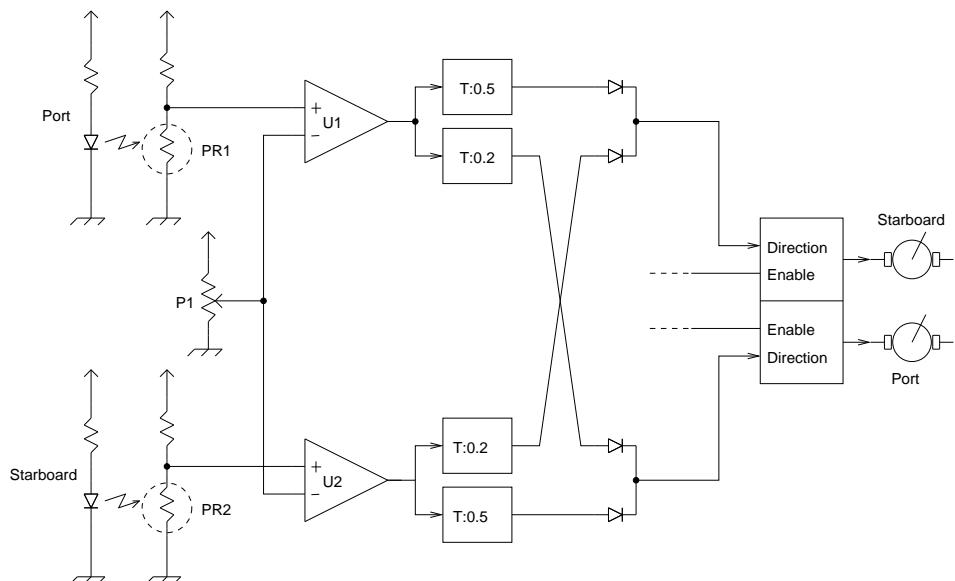


Figure 994: Edge Detection, Back and Turn Concept

The circuitry for the edge detector is shown in figure 995.

The robot includes a *DIP switch*³²¹ which may be used to enable and disable robot functions. If these switches are opened, then the edge detector is disabled, which is useful when debugging other functions.

Each of the four delay networks consists of a leftmost diode, resistor, capacitor and rightmost diode. When the output of a comparator goes high, it rapidly charges the capacitor. When the robot starts backing up, the edge detection signal will disappear, and the output of the comparator will immediately go low. The leftmost diode prevents the capacitor from discharging back through the output of the comparator.

The rightmost diode creates a logical OR gate so that, for example, either the Port or Starboard edge detector can trigger the Starboard motor.

The outputs from the time delay circuits drive Schmitt trigger inverters that produce the signals to operate the L293 motor driver. The *enable* pins on the L293 are operated from a different circuit, the *Object Detector and Speed Control*, described on page 1116.

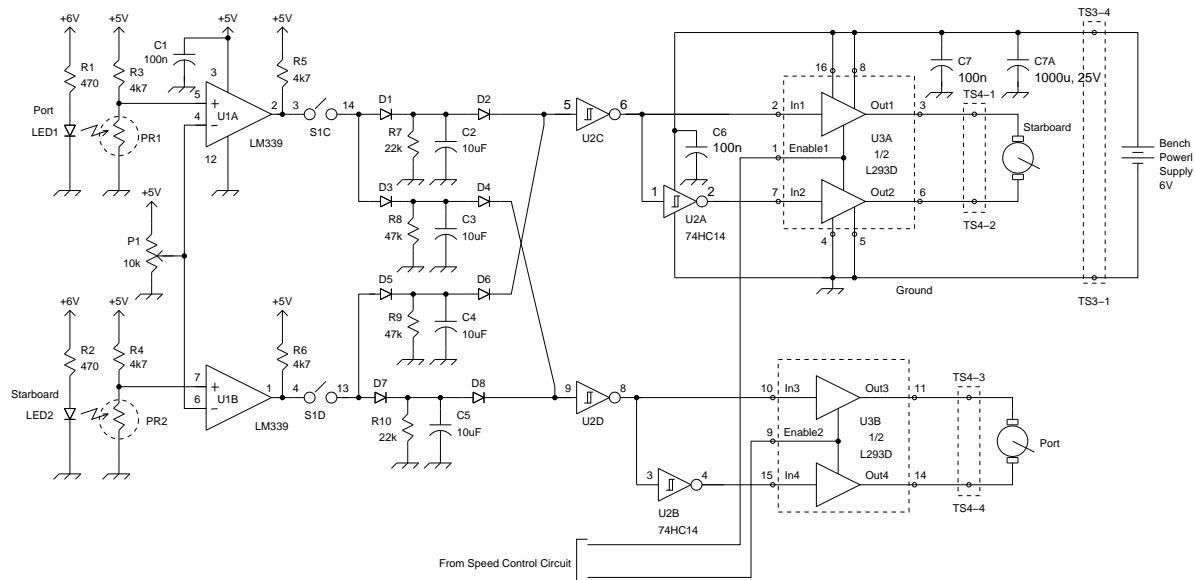


Figure 995: Edge Detection Circuit

37.3.5 Object Detector and Speed Control

The circuit concept for detecting another robot, the *object detector*, is shown in figure 996.

Two infrared LED emitters, one on each side at the front of the robot, are driven from a 4KHz oscillator so that their light beams (which are invisible because they are infrared radiation) are chopped at a frequency of 4KHz. The detector is designed to respond to this 4KHz signal and ignore other illumination such as sunlight or 120Hz from fluorescent lamps.

There are two identical detector channels, so we'll just describe one of them. The reflected infrared illumination is detected by an infrared photodiode, and amplified in a *transresistance* amplifier. This is an op-amp circuit that converts the current from the photodiode into an AC voltage at 4KHz.

³²¹DIP stands for *dual-inline package*, so this is a switch with the same footprint as a 16-pin integrated circuit.

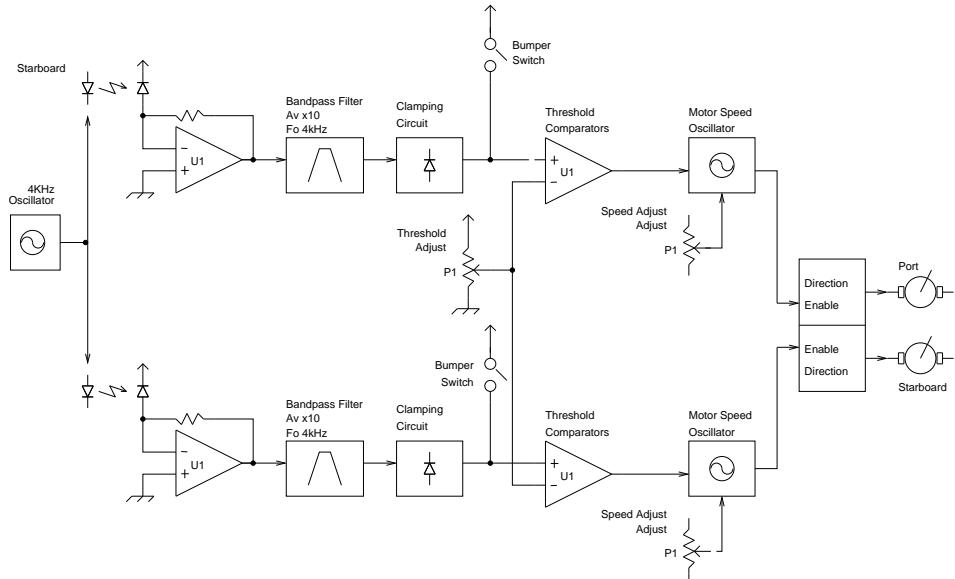


Figure 996: Object Detector Concept

This signal, along with any other spurious signals, is fed into a 4KHz *bandpass filter*. The filter amplifies a signal with a frequency at around 4KHz and rejects ambient light and frequencies of 120Hz.

The output of the bandpass filter is an AC signal of unknown amplitude riding on a DC level. The next function is a *clamping circuit*, which fixes the lower level of the 4KHz waveform at around zero volts. Then as the signal changes in strength, it grows upward from the zero volt level.

The clamped signal is fed into a *threshold detector*, which compares the level of the signal with a fixed level from the *threshold adjust* potentiometer. If the signal exceeds the threshold level, the comparator will produce a 4KHz train of pulses at 5 volts amplitude.

Each motor has a *speed adjust* circuit so that the two motors can be adjusted to run at the same, low speed. When an object is detected, the signal into the speed adjust circuits puts them into full speed mode, so the robot closes at maximum speed and pushes the other robot with maximum torque.

When this robot touches the other robot, it should actuate either or both the bumper switches. As long as the bumper switch is closed, that simulates an object detected signal and keeps the robot in high speed mode. The switch artificially pulls up the comparator input to create this signal.

37.3.6 Object Detector Circuit

The complete object detector circuit is shown in figure 997 on page 1118.

- U4A is the 4KHz oscillator which drives the infrared emitters LED3 and LED4.
- U5A and U5C are the transresistance amplifiers that convert the photodiode current to a voltage.
- U5B is the bandpass filter for one channel, U5D for the other channel.
- The clamp circuit on the upper channel is C13 and D9. These devices behave as a half wave rectifier that charges up the right plate of C13 to the peak value of the signal voltage. The result is that AC voltages rest on the zero volt baseline.

- U6A is an op-amp schmitt trigger with positive feedback via R21 and R20. The threshold voltage is set at the inverting input pin.
 - Switch S1A allows the object detector to be disabled during debugging sessions.
 - Schmitt trigger U4A, P3, diodes D11 and 12, and C17 form an oscillator which generates the motor drive waveform. The duty cycle can be adjusted by changing P3.
 - When an object signal is generated, that turns on Q2, which then forces the input of the schmitt trigger U4B to ground. This forces the output of U4B to a high state, which puts the motor into full output.
 - The other channel operation is identical.

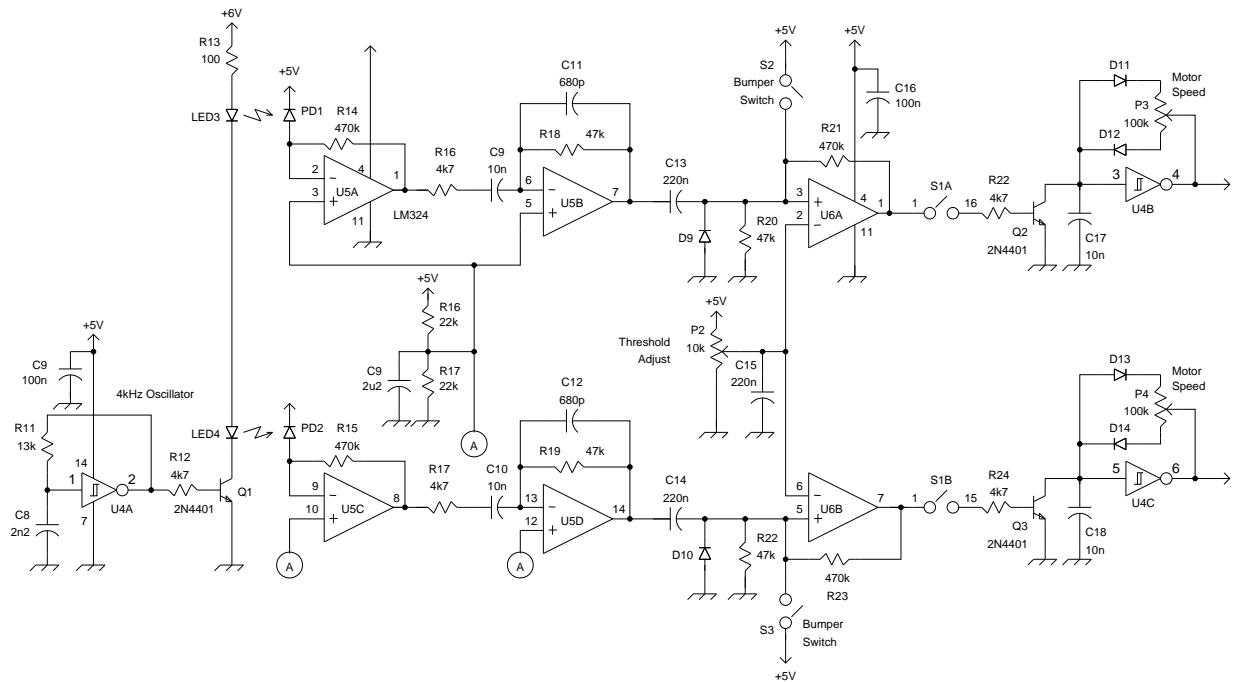


Figure 997: Object Detector Circuit

37.4 The Synchro Resolver

In this section, we'll look at the technique of *carrier coding and recovery*, that has a wide variety of possible applications. This same technique appears in slightly different form in connection with correlation, network analysis, lock-in measurement technique and spectrum analysis of signals. We'll start with a specific application and then generalize to these other techniques.

The *resolver* is a electro-mechanical component widely used in commercial and military avionics equipment for the measurement of shaft angle [351], [352], [353]. It looks much like a small motor, but is actually more closely related to transformers.

The resolver is capable of *absolute* measurement of shaft angle³²². It is possible to obtain considerable precision from a resolver: 12 bit accuracy, (one part in 4096) is routine. Some resolvers are brushless, so there are no moving contacts, resulting in excellent reliability.

The resolver consists of three coils, a *rotor* attached to the shaft of the resolver and two orthogonal *stators* attached to the frame. The terminals of each coil are brought to the outside world. An AC voltage supply, typically 117 volts RMS at 400Hz, is connected to the rotor winding. (Some models are designed to operate at 26 volts.) As the shaft rotates, the coupling between the rotor and the two stators changes, changing the magnitude of the output voltage of the stator windings. The voltage induced in a stator winding is at a maximum when the rotor and stator windings are aligned, and zero when they are at right angles to each other, figure 998.

The equations for the two stator voltages are:

$$e_{s1} = e_r K_r \sin \theta \quad (1672)$$

$$e_{s2} = e_r K_r \cos \theta \quad (1673)$$

where

- e_r is the rotor AC supply voltage
- K_r is a constant for the resolver, typically 0.3
- e_{s1}, e_{s2} are the voltages induced in the two stator windings
- θ is the angle between the rotor and stator S1

The magnitude of the AC stator voltage follows the sine function through a full 360° of shaft rotation³²³, figure 999.

For the first 180° of rotation, the sine function and stator voltage are positive, so the AC voltage induced in the stator is in phase with the rotor voltage. For the second 180° of rotation, the sine function and stator voltage are negative, implying that the AC voltage induced in the stator is *inverted* with respect to the phase of the rotor AC voltage.

For any given voltage on this waveform (except for those at a maximum or minimum point) there are two possible angles. To unambiguously determine the shaft angle, we need the cosine output of the resolver as well, figure 1000.

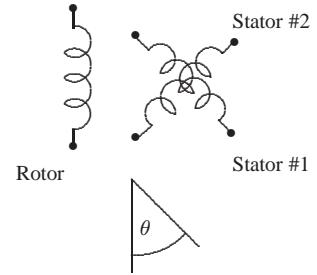


Figure 998: Resolver

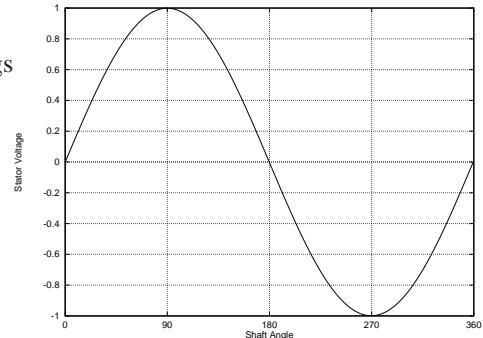


Figure 999: Resolver SINE Output

³²²In contrast, for example, to the perforated wheel encoders found inside a computer mouse, which are *incremental* devices.

³²³Notice carefully the labels on the axes of this graph. The function is the magnitude of the AC voltage VS the shaft angle. The output voltage is a sine wave that varies in magnitude.

The shaft angle is then

$$\tan \theta = \frac{e_r K_r \sin \theta}{e_r K_r \cos \theta}$$

so

$$\theta = \tan^{-1} \frac{\sin \theta}{\cos \theta} \quad (1674)$$

(Some care must be taken in calculating θ at the angles 90° and 270° , where the value of the cosine approaches zero and the tangent function approaches infinity.)

Figure 1001 shows what the stator voltage waveform would look like if the rotor were to be rotated so that one rotation corresponded to 20 cycles of the AC excitation voltage³²⁴.

A careful examination of that figure shows that the stator voltages in the $0-180^\circ$ segment and $180-360^\circ$ segment are equal in magnitude but out of phase. A suitable electronic circuit can decode the magnitude and phase to determine the angle of the rotor.

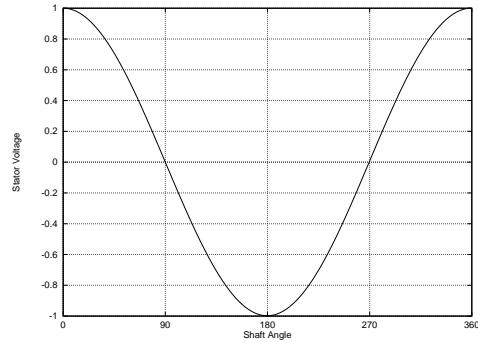


Figure 1000: Resolver COSINE Output

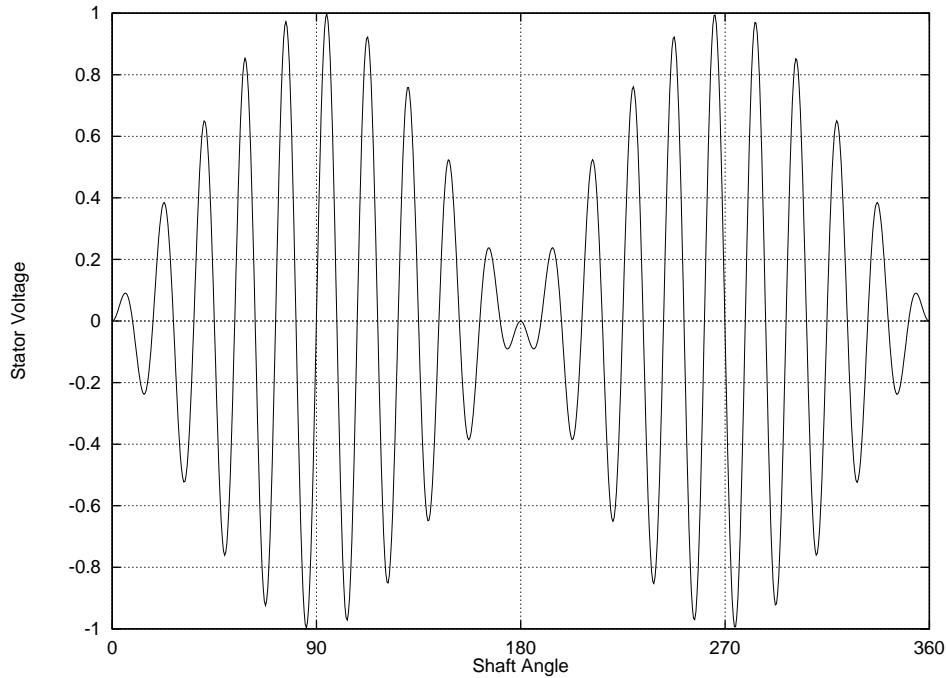


Figure 1001: Resolver Stator Voltage Waveform vs Shaft Angle

Now we need a method of converting the resolver signals into a useful form. We'll assume that the desired output is two DC voltages, proportional respectively to the sine and cosine winding outputs. These signals could then be digitized by an analog-digital converter and further processed in microprocessor software.

³²⁴To obtain the waveform shown in figure 1001, the resolver shaft would have to be rotated at *exactly* the correct frequency, which is all but impossible to do in practice.

A suitable circuit for the resolver is described in the next section, and then generalized in section 37.6 (page 1125).

37.4.1 Resolver Interface, Design

A block diagram of the resolver interface is shown in figure 1002.

A 400Hz sine wave is generated by the oscillator and drives the resolver stator winding. The magnetic field generates 400Hz AC voltages in the two secondaries, where the magnitude of these AC voltages depends on the angle of rotation of the resolver. (Where the frequency of the power distribution grid is 60Hz in North America, the frequency 400Hz is typically used in aircraft. As the frequency of operation increases, the size and weight decreases for magnetic components, such as transformers. In this application, the exact frequency is not critical, and any frequency approximately equal to 400Hz would work.)

This same sine wave is squared up in a comparator and used as the reference phase in the two phase-sensitive rectifiers (section 16.5).

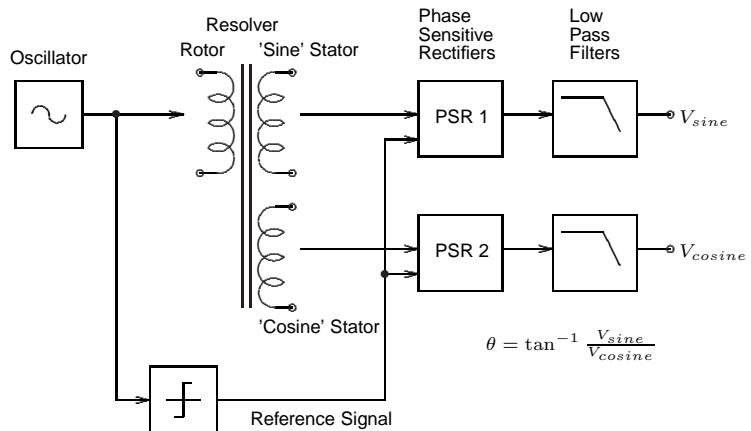


Figure 1002: Resolver Interface Block Diagram

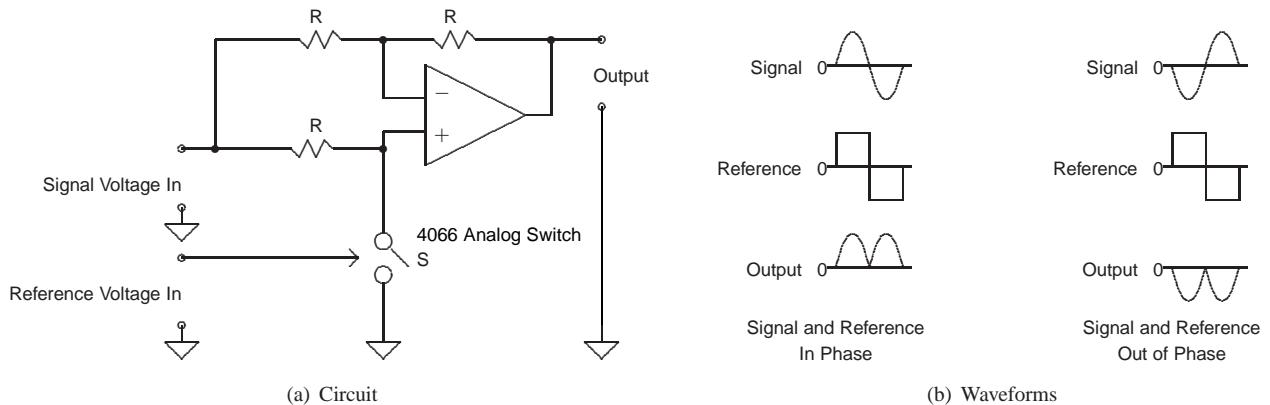


Figure 1003: Phase Sensitive Rectifier

Referring to the circuit diagram of figure 1003(a), the phase sensitive rectifier has a gain of plus or minus 1, depending on the position of the reference switch, open or closed. The waveforms of figure 1003(b) show the effect of the phase relationship between the input and reference signals.

The effect of phase sensitive rectification on the signal from the resolver may be seen by comparing figure 1001 on page 1120 with the waveform at the output of the phase-sensitive rectifier, figure 1004 on page 1122.

Figure 1004 shows the output of one of the phase sensitive rectifiers as the resolver is rotated through 360°. Notice how the 400Hz excitation sine wave is full-wave rectified to a positive or negative average value, depending on the resolver shaft angle. The average value of this waveform can be extracted by a lowpass filter to produce a DC voltage that varies with the angle of the resolver.

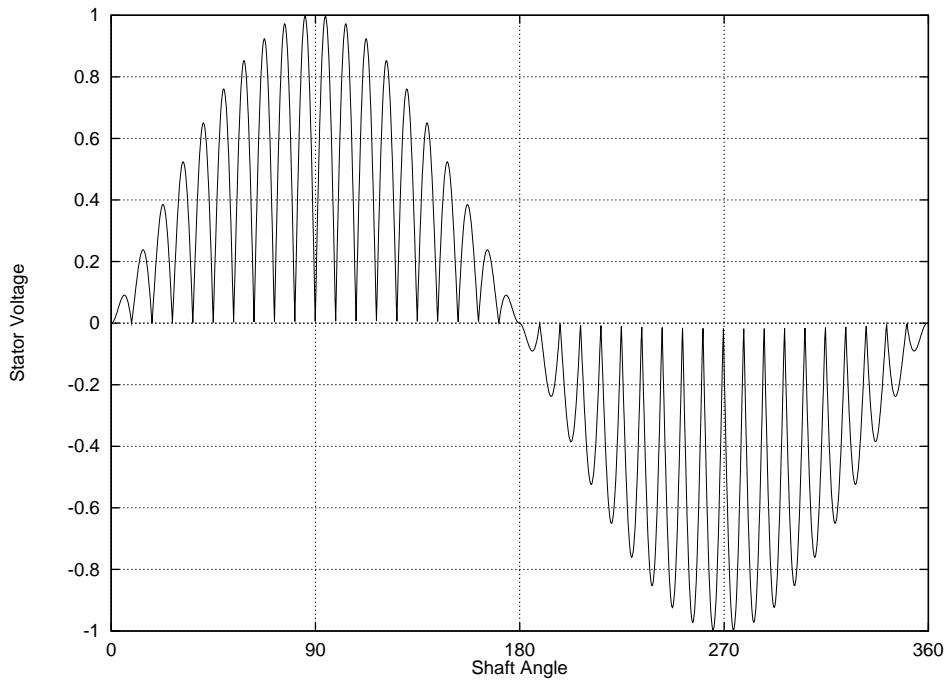


Figure 1004: Phase-Sensitive Rectifier Output

37.4.2 Resolver Interface, Circuit Diagram

The complete circuit diagram of the resolver interface is shown in figure 1005.

- The available power supply for this circuit is +5 volts. Amplifier U3 is an audio power amplifier that contains circuitry to place its output at half the supply voltage. Its output current capability is in excess of 150mA. This output creates a *pseudo-ground* half way between 5 volts and ground. Then the circuit can be designed as if it had dual supplies of $\pm 2.5V$ around the pseudo-ground.

Many op-amps require several volts of headroom between the power supply voltages and the maximum input and output signals. There is very little room between the signal and supply voltages in this circuit, and so it requires the use of an RRIO (Rail-Rail, Input and Output) op amp. In RRIO operation, the input common-mode voltage range and output voltage range include the power supply rails.

- The carrier oscillator is op-amp U1A and its associated components. It's a Wien Bridge oscillator (section 20.2, page 622) where the network R1,C1,R2,C2 determines the frequency of oscillation. Resistors R3 and R4 define the closed-loop forward gain of the op-amp circuit and are proportioned to provide a gain of slightly more than 3 to ensure that the oscillator will start reliably.
- The comparator is U1B. It squares up the sine wave to generate a reference square wave for the phase sensitive rectifiers. The network R11,R12,C3 generates a small amount of phase lead so that the reference waveform is exactly aligned with the carrier sine wave output from the resolver.
- Op-amps U1C, U1D are the sine and cosine channel phase sensitive rectifiers. The switches U3A, U3B are two sections of a four-channel analog switch, the CD4066. The output of each PSR stage is a full-wave rectified waveform with amplitude dependent on the rotary position of the resolver.

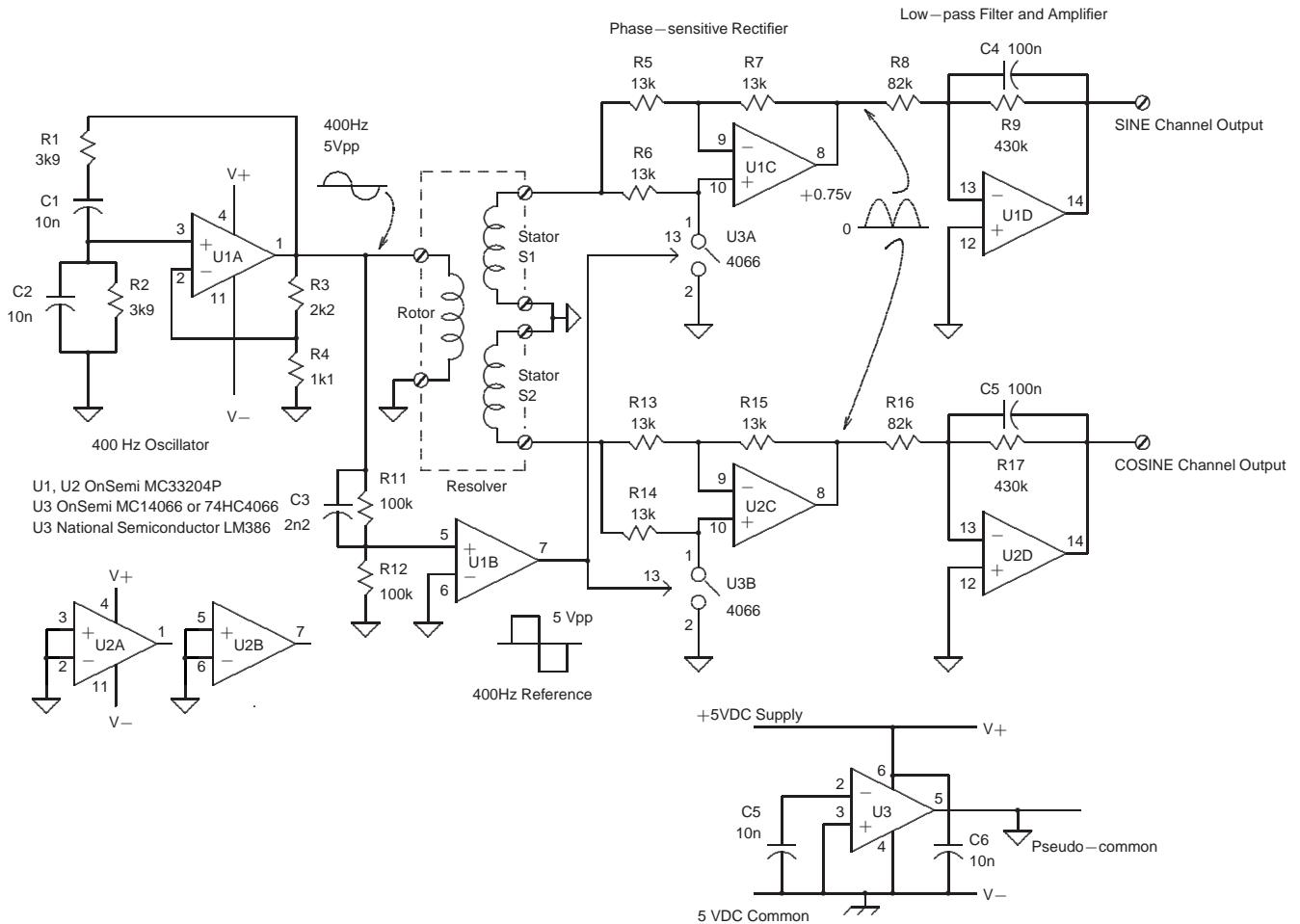


Figure 1005: Resolver Interface Schematic

- The AC component of the PSR output is reduced in the lowpass filter stages U1D, U2D. These stages have a low-frequency gain of 5.25 volts/volt to compensate for losses in the system and generate a full scale output in the order of 4 volts DC. The filter is a first-order lowpass filter with a cutoff frequency defined by the feedback resistance and capacitance (section 17.2), approximately 4Hz for the values shown in the schematic.

37.5 The Linear Variable Differential Transformer

The Linear Variable Differential Transformer (LVDT) [354] is a close relative to the Synchro Resolver described in section 37.4. It is a useful device for measuring linear displacement or sensing a null reference position in linear movement. It is quite feasible to measure displacement with resolution in the order of 0.001 inches. Because there are no brushes or sliding parts (unlike, say, a linear potentiometer), the LVDT is reliable and suitable for use in an industrial environment.

Like the resolver, the LVDT is a variant on the transformer. It includes one primary and two secondary windings arranged in a linear fashion, as indicated in the cross-sectional diagram of figure 1006(a).

The primary winding is driven from an AC source. If the movable core piece is equidistant from the two secondary windings, then the voltage induced in them is equal. When the core moves toward secondary #2, for example, the voltage in secondary #1 decreases, the voltage in secondary #2 increases.

The two secondaries are wired in series in such a way that the two induced voltages oppose each other. When the core is at the centre position the voltages in the secondaries cancel, and this null can be used as a reference position. As well, the phase of the total-secondary-voltage changes as the core passes through the mid-point, and so the phase of this voltage can be used to indicate the direction of the core displacement. For example, if secondary voltage in phase with the primary indicates that the core is left of the centre position then secondary voltage antiphase with the primary indicates that the core is to the right of centre. The magnitude of the secondary voltage indicates the magnitude of the displacement.

In figure 1006(a) the three windings are shown as equal size, but this need not be the case. Various relative sizes of windings can be used to proportion the output voltage relative to core displacement as required.

A block diagram of the LVDT electronic interface is shown in figure 1006(b).

This circuit is very similar to the resolver interface of figure 1002. One phase sensitive rectifier detects the LVDT secondary voltage and produces an average positive or negative output depending on the phase of the secondary output voltage. The lowpass filter removes the AC component at the output of the phase sensitive rectifier. The resultant output from the LVDT interface is shown in figure 1006(c).

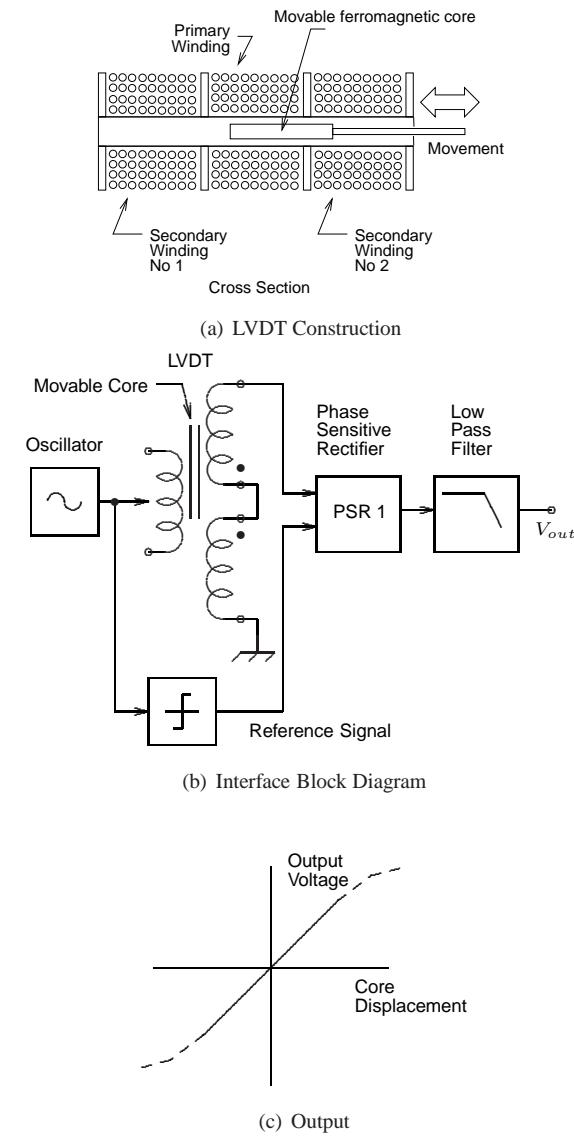


Figure 1006: LVDT Interface

37.6 Analog Correlation Techniques

It is possible to generalize the electronic systems shown in section 37.4.1 for the resolver and in section 37.5 for the linear variable differential transformer. This leads to some important insights and connections with other analog systems such as the lock-in amplifier, spectrum analyser, network analyser and correlation detector³²⁵.

37.6.1 Correlation Overview

The *correlation* of two electronic signals is a measure of their similarity [355] [356]. If two signals are identical, then they have a *correlation coefficient* C_{12} which is some positive value. If two signals are completely different, then their correlation coefficient is zero. If the two signals are related in such a way that one is an inverted version of the other, then the correlation coefficient is some negative value.

The correlation of two signals $f_1(t)$ and $f_2(t)$ is given by:

$$C_{12} = \frac{1}{T} \int_0^T f_1(t) \cdot f_2(t + \tau) dt \quad (1675)$$

To begin with, assume the quantity τ is zero. Translating this equation into a block diagram requires multiplying the two signals together and then averaging the result over some period of time T .

The signals must have zero mean value, that is, the DC value must have been removed. As well, the signals must be *stationary* during the averaging interval – their statistics must not change. However, the signals need not be periodic and in practice correlation is often performed on random noise signals.

The architecture of this correlator then consists of an analog multiplier followed by an integrator. The process is started by resetting the integrator to zero, applying the two signals, and integrating for some period of time T .

In practice, the output of an integrator tends to drift to one of the op-amp bounds because an integrator has extremely large gain at zero frequency, so any offset voltage eventually becomes an extremely large error signal in the output. For this reason, the integrator is usually replaced with a lowpass filter, as shown in figure 1007. For frequencies well beyond the corner frequency of the filter, a lowpass filter approximates the behaviour of an integrator: the amplitude decreases at 20db/decade and the phase shift is 90° . A lowpass filter can be constructed with modest gain at low frequencies, so that an offset voltage does not create a large error signal.

37.6.2 The Correlation Time Delay

The τ term in equation 1675 represents a time delay that is applied to one of the two signals. In practice, many signals show a strong correlation when one of the signals is delayed with respect to the other, figure 1008.

For example [357], consider the situation where an engineer is measuring the sound absorption of a wall panel, figure 1009. A test signal (which could be random noise, bandlimited to the frequencies of interest) is generated by a loudspeaker. A microphone is placed near the panel. The microphone signal is correlated against the test signal. As the correlation time delay τ is varied,

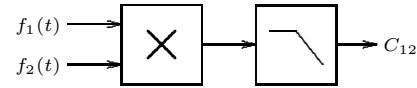


Figure 1007: Correlator

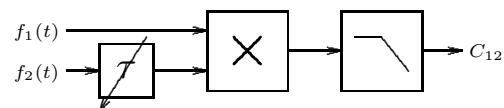


Figure 1008: Correlator With Delay

³²⁵All of these systems may be implemented with digital signal processing methods. However, understanding the analog approach is useful, even when the implementation uses digital techniques.

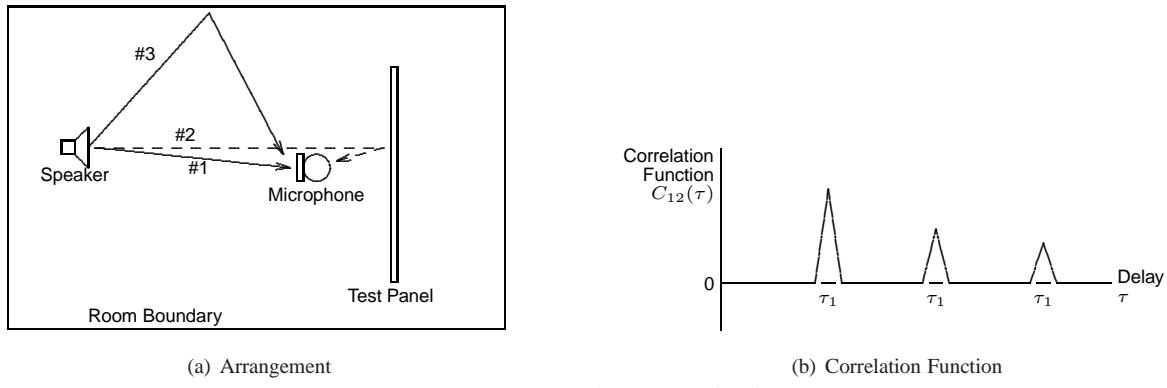


Figure 1009: Correlation Application

there will be three correlation peaks: one that corresponds to the time of propagation of the sound wave directly to the microphone (the *direct path*), a second peak at a greater time delay that corresponds to the time of propagation of the sound wave from the panel (the *reflected path*), and then subsequent peaks that correspond to reflections from the room boundaries. The ratio of the first to second peaks is a measure of the absorption of the sound panel³²⁶.

The basic correlation structure – multiplier followed by lowpass filter, without time delay – can be implemented entirely in analog circuitry. It proves to be useful in a variety of applications, and it is worthy of some investigation.

37.6.3 Correlation and Signal Processing

We can gain some insight into correlation by considering it as an operation of *signal processing*. Consider the measurement system shown in figure 10.10. In this situation there exists some *experiment (black box)* for which we wish to know the transfer function exactly. A straightforward measurement circuit is shown in figure 10.10(a).

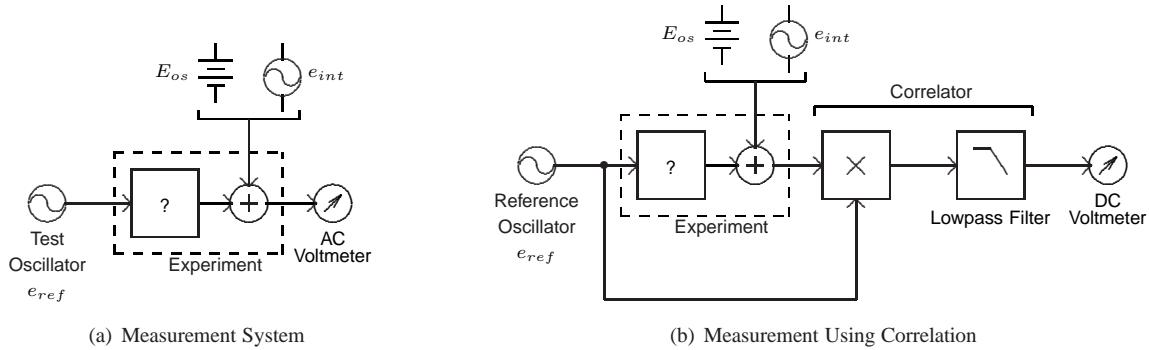


Figure 1010: Signal Processing with Correlation

A test signal is applied to the input and the magnitude of the output is measured with an AC voltmeter³²⁷. The magnitude of the transfer function of the experiment is simply the ratio of the input signal to the output signal. However, DC offsets and noise signals (either intrinsic or the result of interference) are effectively added to the output signal, and this may obscure the measurement.

³²⁶ Interested readers are referred to Schomer [358] and Broch [359] for details and limitations of this type of measurement.

³²⁷We assume that the phase is not of interest in this case.

An alternative system is shown in figure 1010(b). In effect, as we shall show, the correlator creates a narrow-band filter in the output that eliminates signals except the signal at the input. The effect is dramatic – it is possible to retrieve a signal which is *smaller* than the noise signal.

To understand the operation, suppose that the reference oscillator operates at a frequency of f_r (for *reference* frequency). Also suppose that the experiment affects the amplitude, but not the phase, of this signal. Then the signal appearing at the output of the experiment will be at frequency f_r as well.

When two frequencies f_1 and f_2 are applied to the input of a multiplier, the output consists of two new frequencies: the *sum* frequency at $f_1 + f_2$, and the *difference* frequency at $f_1 - f_2$.

The output frequencies of the multiplier will occur in the manner shown in the following table. Each entry line shows a combination of frequencies, followed by an example. In the example, the reference frequency is assumed to be 1000Hz, the interference frequency f_i is assumed to be 60Hz, and the lowpass filter cutoff f_c is at 10Hz.

Type	Sum	Difference
Signal	$f_r + f_r = 2f_r$ 1000 + 1000 = 2000	$f_r - f_r = 0 \text{ (DC)}$ 1000 - 1000 = 0
DC Offset	$f_r + 0 = f_r$ 1000 + 0 = 1000	$f_r - 0 = f_r$ 1000 - 0 = 1000
Interference	$f_r + f_i$ 1000 + 60 = 1060	$f_r - f_i$ 1000 - 60 = 940

The lowpass filter is chosen to have a cutoff frequency that is close to zero Hz (DC). All the outputs, except the signal that is at the same frequency of the reference, are translated to some non-zero frequency and so are rejected by the lowpass filter. The magnitude of that component reflects the effect of the experiment on the reference signal. Notice that the instrument in figure 1010(a) uses an AC voltmeter. The instrument in figure 1010(b) uses a DC voltmeter.

What about broadband noise? Noise can be modelled as a signal e_n which includes a range of frequencies and is added to the output of the experiment in the same manner as the offset and interference signals. At the output of the multiplier, only noise frequencies in the range of 0 to 10Hz are passed by the lowpass filter. At the input to the multiplier, these noise frequencies correspond to the frequency range $f_r \pm f_c$. Consequently, the measurement bandwidth of this system is $2f_c$ Hz.

The same results – rejecting offset and interference, and reducing the noise bandwidth – could be achieved with a narrow bandpass filter at the output of the experiment, shown in figure 1011. However, and this is a crucial point, *the system of figure 1010(b) can function with an arbitrarily narrow lowpass filter, so the signal-noise ratio can be very substantially improved*. For example, we could set the cutoff frequency f_c of the lowpass filter to 0.01Hz. The signal would still pass through the lowpass filter, as desired, but most of the noise would be outside the bandwidth $f_r \pm f_c$.

In theory, the bandpass filter can function in the same manner. However, there is a practical problem: the reference frequency must be centered in the bandpass, and that's difficult to accomplish with a filter of bandwidth 0.02Hz. In essence, the correlator structure ensures that the centre of the effective bandpass filter tracks the reference frequency.

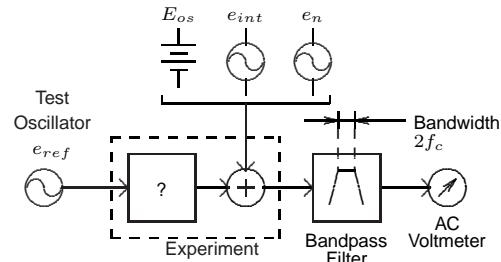


Figure 1011: Experiment with Bandpass Filter

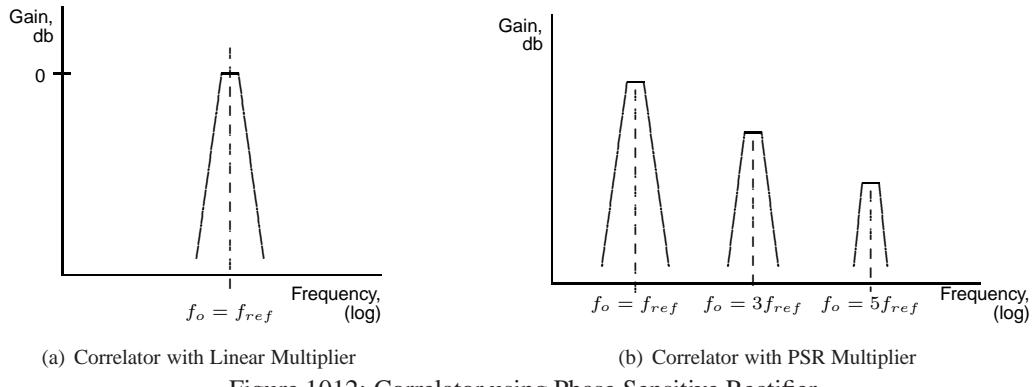
37.6.4 Correlation and the Phase Sensitive Rectifier

It is entirely possible to use a linear analog multiplier (section 36.1) in the previous correlator circuits. However, such a multiplier is relatively complicated and expensive. With a small loss in effectiveness it is possible to use a phase sensitive rectifier instead, which is a much simpler circuit (section 16.5).

The phase-sensitive-rectifier may be regarded as a multiplier: the multiplicand is a linear signal, the multiplier is ± 1 , and the result is equal to the input voltage or the input voltage inverted.

A correlator with a perfectly linear multiplier creates one bandpass filter centred at the reference frequency. The bandwidth of this filter is twice the cutoff frequency of the correlator lowpass filter.

Multiplying by ± 1 is equivalent to multiplying by a square wave of magnitude ± 1 . In spectral terms, this square wave will contain harmonics at the fundamental, 3rd harmonic, 5th harmonic and so on. Each of these harmonic frequencies effectively creates a new bandpass filter (*a harmonic filter*) at the centre frequency of that harmonic.



(a) Correlator with Linear Multiplier (b) Correlator with PSR Multiplier

Figure 1012: Correlator using Phase Sensitive Rectifier

This has two effects:

- The output measurement includes signals not only within the frequency range $f_{ref} \pm f_c$, but also at harmonics of f_{ref} .
- The cumulative total bandwidth for noise is increased to some extent by the effective multiple bandpass filters.

These effects are mitigated to some extent by the decreasing amplitude of higher order harmonics. Then the effective gain of each bandpass filter is reduced proportionally, and this reduces the contribution of each harmonic filter.

In practice, there are many applications where these effects are quite acceptable. The synchro resolver circuit of section 37.4 is one such example. Each phase sensitive rectifier–lowpass filter functions as a correlator that substantially improves the resistance of the interface circuit to noise and offsets. However, in that particular circuit, the cutoff frequency of the lowpass filter must be a compromise: lower cutoff frequency improves the noise immunity but lengthens the transient response. Consequently, the cutoff frequency of the lowpass filter must be large enough that the filter can respond quickly to changes in input signal amplitude caused by changing resolver shaft angle.

37.7 Barometer Interface

It is a common scenario in electronic technology that some critical component becomes available at reasonable cost and that creates an entirely new opportunity for a product. Pressure sensors have been available in the past, but at a cost that makes them uneconomical to use in a consumer product. However, measurement of pressure is a requirement in car engine environmental controllers, and that has resulted in low-cost pressure sensors becoming available on the market.

Although these pressure sensors are inexpensive, the signal of interest is very small and the sensor produces a temperature-sensitive offset voltage which must be taken into account. It's proposed in this example to design a circuit which will interface the pressure sensor to the A/D converter of a microprocessor. The microprocessor can also measure temperature and microprocessor software used to compensate for the offset from the pressure sensor.

The design objective is a low-cost barometer module that can be used in consumer-grade weather stations³²⁸.

37.7.1 The Pressure Sensor

Air pressure typically varies between 95 and 105 kilo-Pascals (kPa). The design shown here is based on the Motorola MPX100AP sensor (figure 1013), a sensor for absolute pressures between 0 and 100 kPa. The heart of the sensor is a thin silicon diaphragm into which has been diffused a network of four resistors in a bridge configuration. The resistors function as sensitive *strain gauges*, changing resistance as atmospheric pressure deforms the diaphragm.

The resistance of a conductor is given by

$$R = \rho \frac{l}{A} \quad (1676)$$

where

ρ is the resistivity of the conductor

l is the length

A is the cross-sectional area

When a resistor is stretched, its length increases and its cross-sectional area decreases, both increasing the resistance. In most conductors, this effect is very slight. In the pressure sensors, the resistors are constructed of semiconductor material that shows large changes with small deformations.

Four resistors are configured as a strain gauge bridge, figure 1014. The resistors are located so that diagonally opposite resistors in the bridge change resistance in the same direction, either $R(1 + \Delta)$ or $R(1 - \Delta)$.

The differential output voltage is then simply $V_{24} = V_{CC} \times \Delta$.

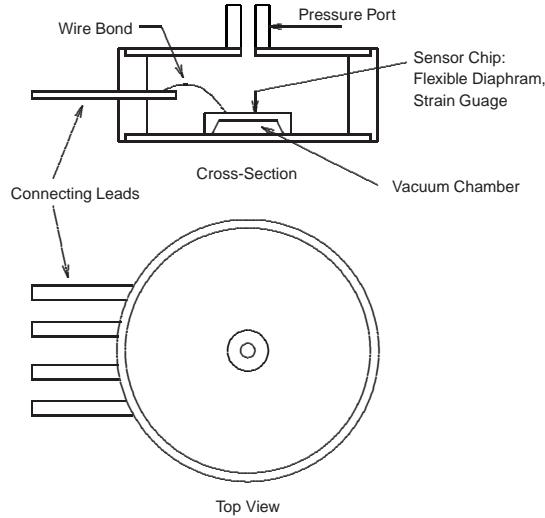


Figure 1013: Pressure Sensor

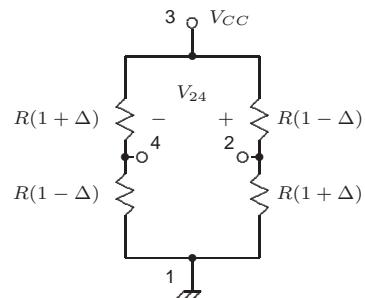


Figure 1014: Resistor Bridge

³²⁸This same pressure sensor and interface circuit can be used as the basis for an electronic altimeter: see [360].

37.7.2 Sensor Specifications

To design the interface, we need to know the signal produced by the sensor and the corresponding signal required by the A/D converter. Then we can determine the required amplifier configuration and gain.

The key specifications for the MPX100AP pressure sensor are as follows:

	Minimum	Typical	Maximum	Unit
Burst Pressure	-	-	200	kPa
Pressure Range	0	-	100	kPa
Supply Voltage	-	3.0	6.0	Volts
Supply Current	-	6.0	-	mA
Full Scale Span (3V Supply)	45	60	90	mV
Offset	0	20	35	mV
Offset TC	-	± 15	-	$\mu\text{V}/^\circ\text{C}$
Sensitivity	0.45	0.60	0.90	mV/kPa

Notes:

Pressure Range The pressure range is given as 0 to 100kPa. We will exceed the maximum pressure slightly to 105kPa. This is still well below the burst pressure.

Supply Voltage If the three volt supply is obtained by dropping a 5 volt supply by 2 volts across resistors in series with the sensor, it turns out that the temperature drift of the sensor is substantially reduced [361].

Supply Current This figure enables us to determine that the bridge resistors are nominally 500Ω . From these figures, we can determine that the sensor gain varies from 0.45 to 0.9mV/kPa .

Full Scale Span This voltage is caused by mismatch of the bridge resistors and appears as a fixed voltage at the output of the sensor and varies with temperature. Notice that the value given for the temperature coefficient of offset voltage is a *typical* value, and no maximum value is given. Consequently, there is no obligation on the part of the sensor supplier to ensure that the temperature coefficient is below a certain limit. There is a serious degree of risk involved in using this sensor if this is a critical specification.

Offset Sensitivity K_T This is a somewhat redundant statement of the transducer gain, which can also be determined from the full scale span specification. Notice that the transducer sensitivity varies over a range of 2:1. The circuit must compensate for this effect.

37.7.3 Pressure Sensor Equivalent Circuit

Putting together all the specifications of the pressure sensor, we have the sensor equivalent circuit shown in figure 1015. From bottom to top, the elements of this equivalent circuit are:

- V_{CM} is the common-mode voltage that appears at both output terminals of the bridge. For a 5 volt supply dropped symmetrically to 3 volts, V_{CM} is 2.5 volts. This voltage must be removed from the signal path by a differential amplifier.

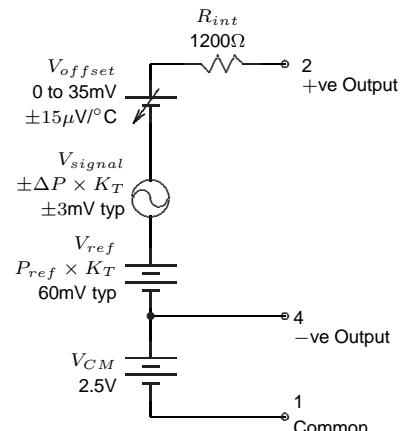


Figure 1015: Sensor Equivalent Circuit

- V_{ref} is the signal voltage that appears at the output terminals due to the average standing air pressure of 100kPa on the sensor. This appears at the sensor terminals but is not of interest, so it must be subtracted from the signal.
- V_{signal} is the change in voltage at the sensor terminals due to the variation in air pressure. This is the signal of interest, about $\pm 3\text{mV}$ for a total pressure variation of $\pm 5\text{kPa}$.
- V_{offset} is a DC output that appears at the output of the sensor. It varies with temperature at a rate of $\pm 15\mu\text{V}/^\circ\text{C}$. If we assume that the sensor will be used in an environment where the temperature varies by 20°C , the offset voltage can be expected to vary by $\pm 0.3\text{mV}$. This represents a potential 10% error compared to the air pressure signal. Some measure will be required to compensate for this effect.

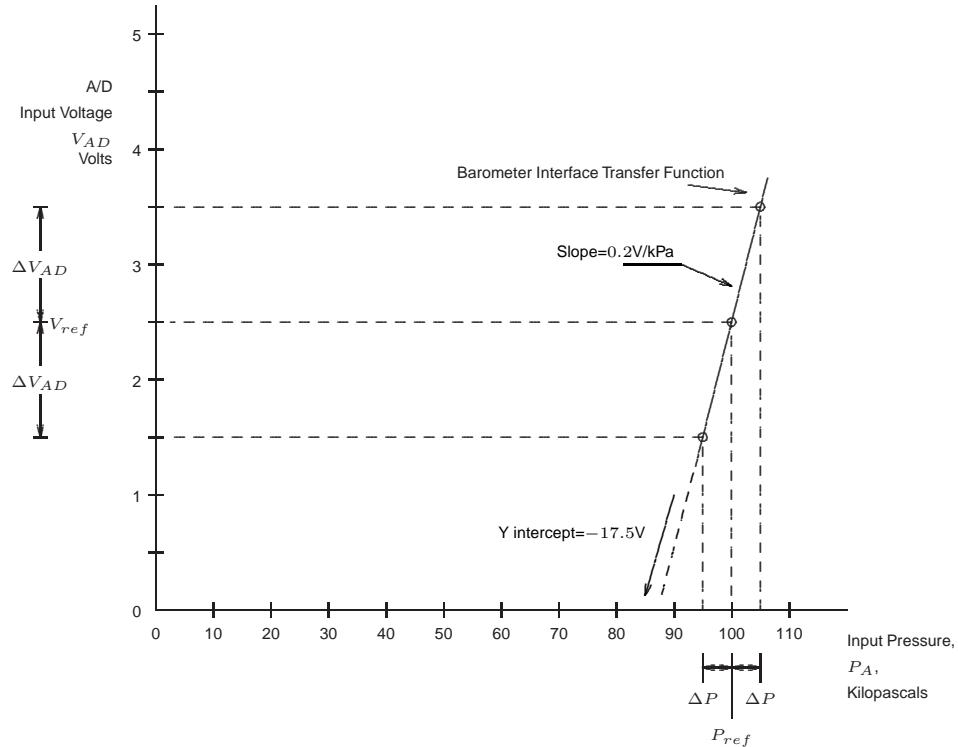


Figure 1016: Barometer Interface Transfer Function

37.7.4 Barometer Resolution and Dynamic Range

Early in the design, we must decide the *resolution* of the barometer, in units of A/D counts per kilopascal of pressure. We'd like a large resolution in order to detect small changes in air pressure. However, higher resolution requires higher voltage gain from the interface and consequent greater sensitivity to a variety of nasty drift signals. Our philosophy should therefore be to make the resolution no higher than necessary.

The face of an aneroid barometer is typically divided into 60 divisions and weather broadcasts are typically given to the nearest tenth of a kilopascal. This would imply 100 steps over the 10kPa variation in air pressure.

Therefore, we might fix on 1 part in 100 as a suitable target for resolution. A suitable *dynamic range* might be 95 to 105 kPa.

37.7.5 Transfer Function

It is useful to characterize the fixed component of air pressure, 100kPa, as P_{ref} , which creates a fixed component of voltage V_{ref} at the input to the microcomputer A/D converter. The variation in air pressure is $\pm\Delta P_a$ around P_{ref} , creating a variation in A/D voltage of $\pm\Delta V_{AD}$ around V_{ref} .

The value of ΔV_{AD} is the product of the resolution, previously fixed at 100 steps, and the voltage per step, 19.5 mV/step, for a 5 volt, 8 bit A/D converter.

Then

$$\begin{aligned}\Delta V_{AD} &= 100 \times 19.5 \times 10^{-3}/2 \\ &= 1.95/2 \\ &= 0.975 \text{ volts}\end{aligned}$$

We'll round this off to ± 1.0 volts.

Now we can fix V_{ref} . It must be large enough that the amplifier doesn't exceed its maximum or minimum output voltages. A good choice is 2.5 volts, halfway between 0 and 5 volts.

With this information, we can draw the transfer function, shown in figure 1016.

Substituting two point coordinates in the straight line equation $y = mx + b$, we can solve for m and b , determining that the transfer function is

$$V_{AD} = 0.2P_A - 17.5 \quad (1677)$$

where

V_{AD} is the input voltage to the A/D converter

P_A is the air pressure in kilopascals

The slope or m value we refer to as the *signal gain*. The intercept or b value we refer to as the *bias*. The bias signal removes the effect of the average air pressure, leaving only the variation in pressure. The signal gain amplifies this variation in pressure sufficiently to drive an A/D converter.

Now we can translate this transfer function into a block diagram, keeping in mind the power supply constraint. This interface works with a microprocessor that is powered from a single +5 volt supply. Ideally, the interface should use this same supply. Then all signals must be constrained to operate in the region between 0 and +5 volts, and the op-amps used must operate correctly in that region.

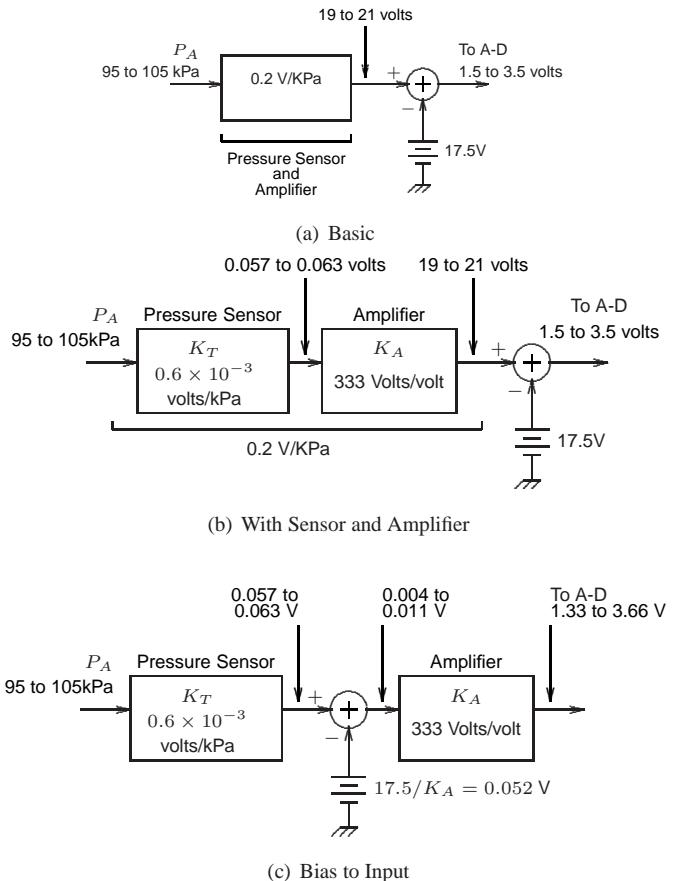


Figure 1017: Barometer Interface Block Diagram

The output voltage of a single supply bipolar op-amp such as the National LM324 or Motorola MC34074 is very limited: 0.5 to 3.5 volts when operated from a +5 volt power supply. On the other hand, the data sheet for the LMC660CN shows 0.2V to 4.7V for a +5 volt supply and load greater than $2\text{K}\Omega$, so this is a suitable amplifier for the pressure sensor interface.

As a starting point for the block diagram, we have figure 1017(a), which is a direct translation of the transfer function, equation 1677.

The typical sensor gain K_T (see the table on page 1130) is 0.6×10^{-3} volts/kPa, so to create an overall gain of 0.2 volts/kPa an amplifier is required, and the amplifier gain must be $0.2/(0.6 \times 10^{-3}) = 333$ volts/volt. The block diagram for this arrangement is shown in figure 1017(b).

There are two practical problems with figure 1017(b).

- The 100kPa pressure P_{ref} tries to generate a 19 to 21 volt signal at the output of the amplifier K_A . This will saturate the amplifier, since it is operated from a +5 volt supply.
- The 17.5 bias volt signal is difficult to generate in a +5 volt system.

Alternatively, we could move the bias signal to the *input* of the amplifier. We need to reduce the magnitude of the bias by the gain of the amplifier, as shown in figure 1017(c). This solves the overload problem of the previous block diagram and this one can be translated into a functional circuit. However, there are some disadvantages:

- The bias signal is now in the millivolt range, so it may be susceptible to interference from other voltages. It's generally preferable to work with large adjustment voltages.
- As we'll see, the input stage to the K_A amplifier must be differential to subtract out a common-mode voltage across the sensor bridge. Furthermore, this same stage must have adjustable gain to compensate for the variation in sensor gain. Those two functions are easy to provide in a differential amplifier, but bias subtraction is a significant additional complication. In the days when op-amps were tens of dollars each, it might have been worthwhile to expend some effort to design a suitable stage. In these days of ten cent quad op-amps, it's often preferable to have separate op-amps performing each function. Then each function can be optimized, interactions are less likely to occur, and access to signals is simplified – which helps with debugging.

A better arrangement is to divide the gain K_A into two roughly equal stages, K_{A1} and K_{A2} , as shown in figure 1018.

In this case, the offset voltage V_{bias} is +1.0 volts, easily generated from +5 volts. Neither amplifier is driven into saturation or cutoff: the signals are well inside the zero to +5V region. Each of these stages can be optimized for its function.

37.7.6 Designing the Circuit

There are essentially two stages to this circuit.

- **Differential Conversion Stage** The output of the pressure sensor bridge is a differential signal sitting on a half-supply common mode signal. The common mode signal must be ignored, so the sensor amplifier must be differential and have a satisfactory common mode rejection ratio. It must also have a voltage gain of 19 Volts/Volt.

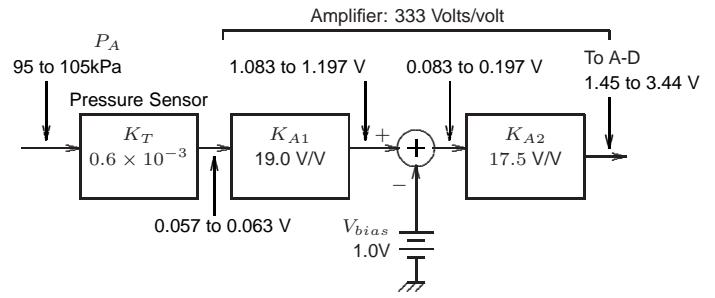


Figure 1018: Splitting the Gain

The instrumentation amplifier can be constructed from three operational amplifiers, as shown in section 13.7 page 341. Alternatively, it is possible to purchase an instrumentation amplifier as a single integrated circuit, section 35.7. A careful engineering design requires comparison of the relative performance and cost of these two approaches.

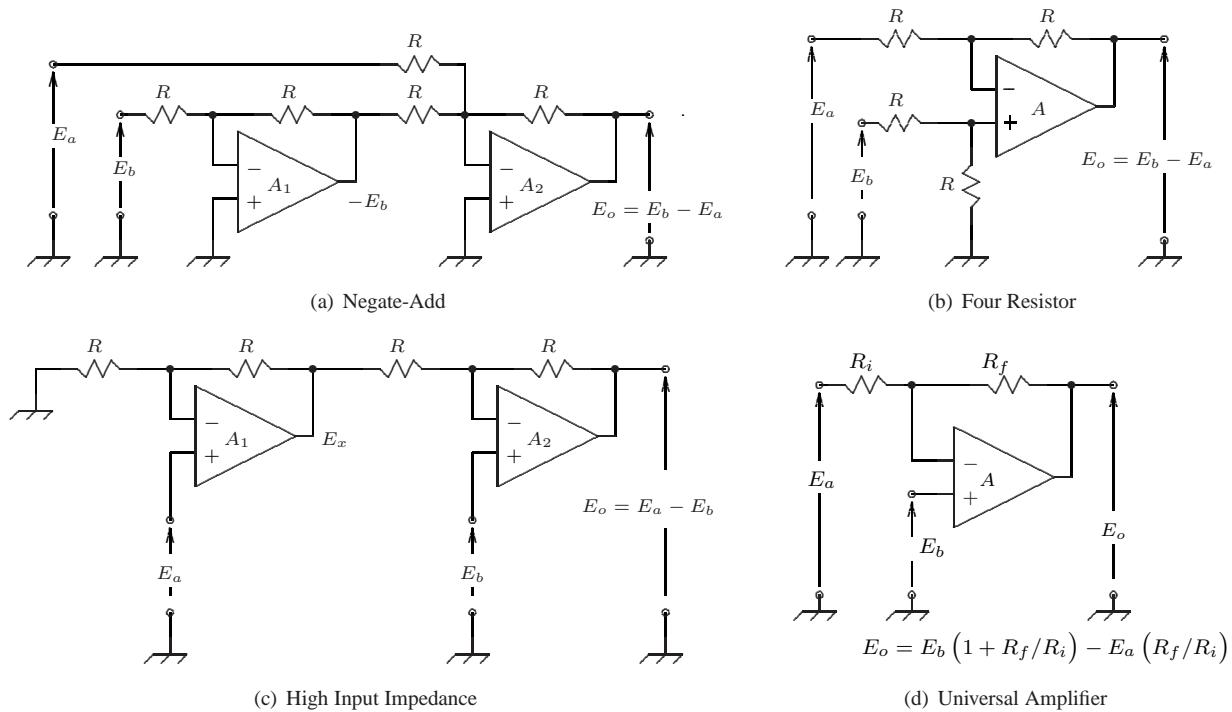


Figure 1019: Subtractors

- **Subtractor Stage** A suitable circuit is not immediately obvious. There are four choices – the subtractors shown in section 13.6, and the universal amplifier of section 13.8 – as summarized in figure 1019. Now we'll consider each of these circuits on its merits:
- The barometer circuit must operate from a single positive source, and so all signals must be positive. Consequently the *negate-and-add* subtractor of figure 1019(a) can be eliminated because it generates an intermediate result that is a negative voltage.
- The *difference amplifier* of figure 1019(b) will work correctly, providing that E_b is always greater than E_a . However, this circuit requires careful matching of resistors to get good common mode rejection, and it does not inherently provide gain³²⁹.
- The *high impedance input difference amplifier* of figure 1019(c) will also work, again providing that E_b is always greater than E_a . However, it requires two op-amps and matched resistors.
- The *universal amplifier* of figure 1019(d) can function as a subtractor, providing that we recognize that the non-inverting gain is slightly larger than the inverting gain. This circuit is economical of resistors, provides a

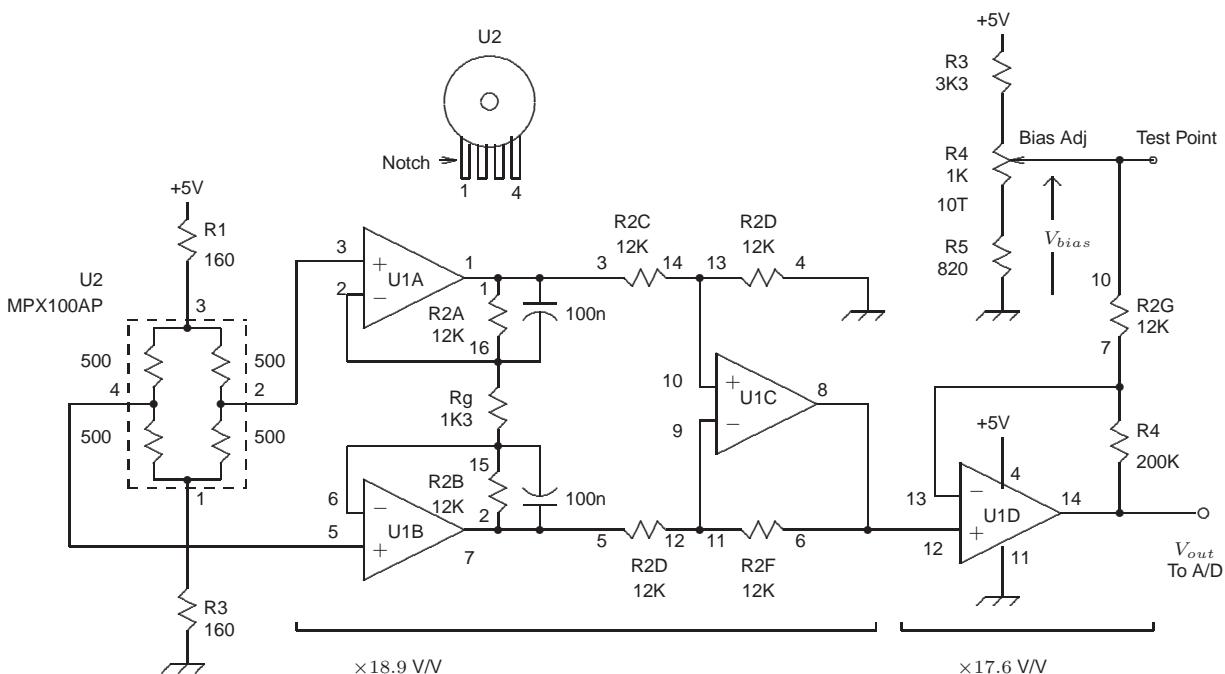
³²⁹The subtractors of figures 1019(b) and 1019(c) can be made to have voltage gain, but to achieve gain and good common-mode rejection requires very careful matching of resistor ratios. In general, it's better to separate the *subtraction* and *gain* function.

high-impedance input at the non-inverting terminal, inherently provides gain, and does not require matched resistors.

The universal amplifier can be used as a subtractor, in which the non-inverting gain – for the signal – is 17.5 Volts/Volt, and the inverting gain – for the bias – is 16.5 Volts/Volt. This slightly reduced bias gain can be compensated by a slightly larger bias voltage.

37.7.7 Complete Circuit

A complete circuit is shown in figure 1020.



U1 National Semiconductor LMC660CN

U2 Motorola MPX100AP, absolute pressure, hose port

U3 Bourns 4116R-001-123 Resistor Array

All 12K resistors are part of U3. Small numbers are DIP package pin numbers.

Discrete resistors must be low temperature coefficient, Philips MRS 25F series or equivalent.

R4 Cermet 10 turn pot, Bourns 3299-1-102 or equivalent.

Change Rg to alter gain.

R1,R3,Rg 160Ω, 1/4 watt, 1%

Figure 1020: Barometer Interface Schematic

- Sensor

According to the pressure sensor data sheet, the resistors in the sensor U2 are 500Ω each and 3 volts should appear across the pressure sensor. Then resistors R_1 and R_2 are 160Ω each.

- **Instrumentation Amplifier** The instrumentation amplifier stage U1A, U1B and U1C removes the 2.5 volt common mode sensor voltage and amplifies the differential sensor voltage. The voltage gain of this stage is given by

$$K_{A1} = 1 + \frac{2R_2}{R_g} \quad (1678)$$

and is set to 18.9 volts/volt. Somewhat arbitrarily, we chose R_2 as $12\text{K}\Omega$, which makes R_g

$$\begin{aligned} R_g &= \frac{2R_2}{K_{A1} - 1} \\ &= \frac{2 \times 12000}{18.9 - 1} \\ &= 1341\Omega \end{aligned}$$

The gain is adjusted by changing R_g . (But see *Calibration*, below.)

For good common mode rejection, the resistors of the differential stage U1C are from a resistor array. All other resistors must be low temperature film, $\pm 50\text{ppm}$ temperature coefficient.

- **Subtractor**

The second stage, U1D, subtracts the bias and provides a final gain to the signal of 17.6 Volts/Volt, and to the bias of 16.6 Volts/Volt. The bias pot R4 should be *cermet* for low drift.

Now we have a provisional circuit design. However, there are some details to take care of.

37.7.8 Reliability of the Design

Now that we have a circuit design, we must ensure that the circuit will work reliably, allowing for component tolerances and the effect of temperature induced drift of the components. To some extent, the temperature sensor and microprocessor can compensate for temperature drift and component tolerances, but there are two potential problems:

- Will the circuit function, or will some voltage or current run into saturation or cutoff?
- Can the circuit calibration procedure compensate for circuit tolerances, or do we lose measurement accuracy under some conditions?

For example, the sensor constant K_T can vary over a range of 2:1. Resistors have a tolerance ($\pm 5\%$ or $\pm 1\%$). The operational amplifiers have offset voltages which can vary by $\pm 7\text{mV}$.

The sensor has an offset voltage drift of $\pm 15\mu\text{V}/^\circ\text{C}$, the resistors change by 250ppm (parts per million) per degree C, and the amplifier offset voltages may change by as much as $\pm 30\mu\text{volts}$ per degree C. With a voltage gain in excess of 300V/V, any drift in offset voltage, bias, power supply or resistance values has the potential for being amplified to appear as drift in the output voltage. What effect will these drifts have on the operation of the barometer?

We should build and test one or more prototypes. However, the correct functioning of a prototype is a *necessary but not sufficient* condition to determine a reliable design. The fact that a prototype works merely means that at least one version of the circuit will function. It's no guarantee that *all* circuits will function.

To ensure the reliable operation of the circuit, the correct strategy is to perform an engineering analysis, checking circuit operation by calculation and simulation. This will provide the necessary confidence to build the

circuit in quantity, and be assured that it will function under all specified conditions. Where possible, to ensure that some massive blunder has not occurred, the calculations should be checked against the prototype. If the calculations and simulation accurately predict the behaviour of the prototype, then we can have some confidence in the predictions.

37.7.9 Circuit Analysis for Component Tolerance and Temperature Drift

This is potentially an unwieldy problem, because of the combinatorial explosion of tolerance variables. Consequently, we need some sort of systematic method of investigating the effect of parameter tolerances and temperature drift. A spreadsheet model is a useful technique for this type of problem. The parameters and equations for the interface are shown in figure 1021. A typical spreadsheet printout is shown in figure 1022.

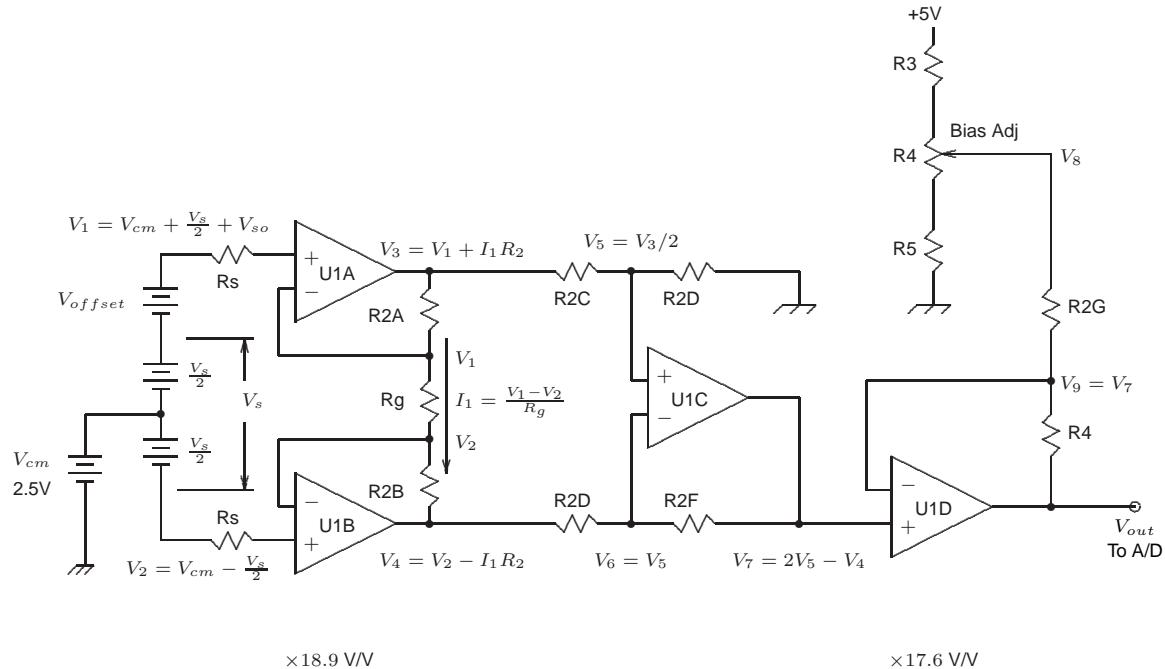


Figure 1021: Circuit Equations

The tinkering with spreadsheet model turned up the following results:

- The sensor offset V_{offset} does not cause the amplifier to saturate but does have a dramatic effect on the output voltage V_{AD} .

(Notice that *sensor offset* V_{offset} is a property of the sensor: do not confuse it with the offset voltage V_{bias} of the interface.)

- The bias voltage V_{bias} may be adjusted to compensate for the effect of sensor offset voltage. The output voltage V_{AD} is very sensitive to the setting of V_{OS} , so the pot R_4 should be a multi-turn unit.
- For low sensor gain, the change in A/D reading over the full range of air pressures (95 to 105 kPa) is over 60 counts, so the resolution is satisfactory even for low sensor gain.

Sensor Constant	K_T	(data)	0.6×10^{-3}	V/kPa
Air Pressure	P_A	(data)	100	kPa
Sensor Offset	V_{so}	(data)	10×10^{-3}	V
Gain Resistor	R_g	(data)	1300	Ω
Resistor	R_2	(data)	12000	Ω
Sensor Output	V_s	$K_T P_A$	0.06	V
	V_1	$V_{cm} + \frac{V_s}{2} + V_{so}$	2.53	V
	V_2	$V_{cm} - \frac{V_s}{2}$	2.47	V
	I_1	$(V_1 - V_2)/R_g$	48	μAmp
	V_3	$V_1 + R_2 I_1$	3.084	V
	V_4	$V_1 - R_2 I_1$	1.916	V
	V_5	$V_3/2$	1.542	V
	V_6	V_5	1.524	V
	V_7	$2V_5 - V_4$	1.167	V
K_{os}	V_8	(data)	1.68	V
	V_9	V_7	1.167	V
Resistor	R_4	(data)	200000	Ω
Output Voltage	V_{AD}	$V_7(1 + \frac{R_4}{R_2}) - V_8 \frac{R_4}{R_2}$	2.00	V
A/D Reading	N_{AD}	$\frac{V_{AD}}{5} 256$	102	counts

Figure 1022: Amplifier Spreadsheet Model and Results

These results could have been predicted from an analysis of the circuit, but the spreadsheet model makes it easy to explore the effect of a variety of options and combinations of parameters.

A circuit simulation program such as *SPICE* could also be used to analyse the circuit, and is a better choice where an accurate op-amp model is required. The spreadsheet model assumes ideal op-amps. On the other hand, spreadsheet programs are readily available and easy to use.

37.7.10 Calibration

The Pressure Sensor Specifications shown on page 1130 show that the sensor gain (K_T in figure 1018) can vary over a 2:1 range and the sensor offset voltage can vary from 0 to 35mV, so some sort of calibration procedure will be required. Variations in sensor gain change the slope of the transfer characteristic. Variations in sensor offset voltage move the Y intercept along the Y axis. Calibration adjusts the gain and offset of the amplifier so that the slope and Y intercept of the complete transfer function match figure 1016.

A suitable strategy is to adjust the gain so that the slope of the transfer function is equal to 0.2V/kPa, and then adjust the offset so that the reference pressure P_{ref} (100kPa) corresponds to the reference voltage V_{ref} (+2.5V).

Gain To adjust the transfer function gain, we inject an AC signal into the input of the system and adjust the amplifier gain so that the magnitude of the AC output is correct. In this case, we cause the air pressure to vary over some range δP and then adjust the amplifier gain so that the output voltage varies over a corresponding range

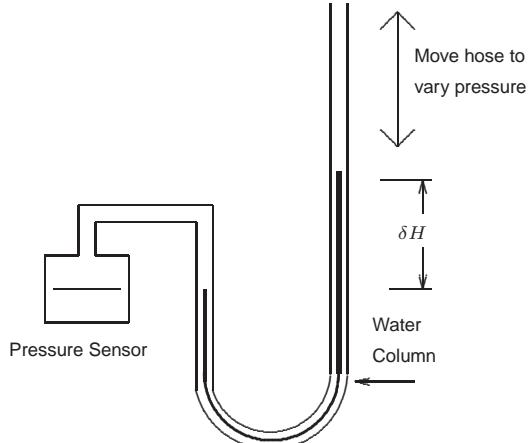


Figure 1023: Water Manometer

δV , where

$$\begin{aligned}\delta V &= K \cdot \delta P \\ &= 0.2 \cdot \delta P\end{aligned}$$

A suitable apparatus for generating a known change in pressure is shown in figure 1023. The liquid is water, laced with red food colouring to make it visible. The tubing is flexible plastic hose available from the local hardware store. The hose is filled with water so that it forms a U shape³³⁰.

The right side of the manometer is raised or lowered to create a height differential of ΔH . The resulting pressure may be determined from a table of physical constants, in which we find that one atmosphere corresponds to 101.3kPa and 160.2 centimetres of water. For example, a pressure differential of 2 kilopascals may be created by a height differential of:

$$\begin{aligned}\delta H &= 2 \text{ kPa} \times \frac{160.2 \text{ cm}}{101.3 \text{ kPa}} \\ &= 3.16 \text{ cm}\end{aligned}$$

The amplifier gain is adjusted until the variation in output voltage is

$$\begin{aligned}\delta V &= 2 \text{ kPa} \times K \frac{\text{volts}}{\text{kPa}} \\ &= 2 \text{ kPa} \times 0.2 \frac{\text{volts}}{\text{kPa}} \\ &= 0.4 \text{ volts}\end{aligned}$$

Offset Now the transfer function must be moved vertically by adjusting the *bias* control in the amplifier interface. The manometer is disconnected so that the sensor is exposed to ambient air pressure. Then the bias control is adjusted until the amplifier output voltage corresponds to the current ambient air pressure (available from the local weather broadcast).

For example, if the current air pressure is 102.4kPa, the output voltage is given by the transfer function:

$$\begin{aligned}V_{out} &= 0.2P_{ambient} - 17.5 \\ &= 0.2 \times 102.4 - 17.5 \\ &= 2.98 \text{ volts}\end{aligned}$$

The *bias* control is adjusted until the output voltage is set to this value. The transfer function should now correspond to figure 1016 and the barometer is calibrated.

³³⁰Apparatus suggested by Jim Koch of Ryerson University.

37.8 Exercises

1. Equation Solver

- (a) How would the computing circuit of figure 984 on page 1106 be modified to solve 3 equations in 3 unknowns?
- (b) Assuming that the amplifiers have a gain-bandwidth product of 1MHz and a slew rate of 0.5 volts/ μ Sec, estimate the settling time of this computing circuit. You can assume that the op-amps are initially producing an output voltage of zero volts and the answer requires some op-amps to produce an output voltage of 10 volts.

2. Motor Speed Control

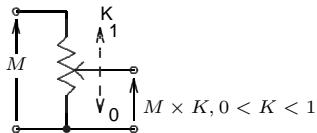
- (a) Verify that the motor control system of figure 988 uses positive feedback.
- (b) The resistance of copper wire increases as its temperature increases. In this motor control circuit, the armature winding will tend to increase in temperature as it is being operated. What effect will this have on the operation of the circuit?

3. Steering Computer

A mobile robot is driven by two DC motors. The speed is increased by collectively increasing the speed of both motors. The direction is controlled by reducing the speed of one motor while increasing the speed of the other. To make the operator interface as simple as possible, the operator is to be provided with a speed control S and a direction control D . The speed value varies between 0 and 10.

Direction varies between 0 and 1, such that 0 indicates full left turn, 1 indicates full right turn, and 0.5 indicates straight ahead.

- (a) Show that the motor signals should be SD to one motor $S(1 - D)$ to the other.
- (b) Design an analog computing circuit to generate these signals from potentiometer inputs that generate signals S and D . As shown in the diagram below, multiplication of some signal M by factor K , where $0 < K < 1$, can be accomplished by applying signal M to the track of a potentiometer and extracting signal $M \times K$ at the wiper of the potentiometer. The value of K varies between 0 and 1 according to its position on the track.



4. Sumo Robot

The Sumo Robot of section 37.3 (page 1112) uses the object detector circuit described in section 37.3.5 (page 1116) to locate its opponent. It's a competitive advantage for a robot locating device to have the longest possible range, and here we consider such a modification.

The object detector described in section 37.3.5 modulates the light emitters with an AC signal. The detector then bandpass filters the received signal, and then it is level detected against a *decision threshold* to determine if another robot is in view.

Based on the ideas of section 37.6, a correlation type detector would allow a much narrower receiver bandpass, which should result in a better signal-noise ratio on the received signal and therefore greater detection range.

- (a) Using the method of section 20.6 (page 631) determine the actual frequency of the astable oscillator in figure 996. The 74HC14 Schmitt trigger may be considered to have trigger thresholds at $1/3V_{CC}$ and $2/3V_{CC}$.
- (b) Determine the 3db frequencies of the bandpass filter in figure 996.
- (c) As a result of the inverse-square law, the detected signal decreases as the square of distance. Assuming that the noise level is unchanged, and the threshold of detection signal-noise ratio requirement is the same, how much of an improvement in detection distance can be expected when the bandwidth is decreased from its original value to 1Hz?
- (d) Redesign the sumo robot detector circuit to use a synchronous detector correlation technique with a 1Hz detection bandwidth.

5. Automating Barometer Calibration

In section 37.7.10, we showed how the barometer interface could be adjusted to compensate for variations in sensor gain and offset. However, if this device were to be produced as a consumer product, it would be highly desirable to eliminate the requirement for human intervention, which is expensive.

- (a) Referring to figure 1020, suppose that resistor R_g is a fixed resistor (to suit the typical gain of the sensor) and the gain is not adjustable. What affect does the tolerance in sensor gain K_T have on the overall gain ($\delta V_{out} / \delta P$)?
- (b) What is the minimum expected resolution in kPa if the amplifier gain is fixed?
- (c) When the sensor gain is at its maximum value and the amplifier gain is fixed (to suit the typical gain of the sensor), show that the bias *must* be adjusted to keep the amplifier from saturating.
- (d) It is proposed that the production barometer have fixed gain and that the microprocessor be responsible for adjusting the bias voltage. Draw a block diagram of the system.
- (e) Verify that the amplifier will work correctly under the circumstances of fixed gain, adjustable bias.
- (f) Assuming that the production facility includes a manometer which can apply various pressures to the sensor, draw the flow chart for a computer-controlled calibration procedure.

37.9 Design Exercises

Each of these analog circuit design exercises requires creative design thinking for low-complexity analog circuits. For each circuit, produce a *circuit concept* that includes

- a block diagram
- a schematic diagram
- identification of any design issues that must be clarified before moving to a final design.

Electronic product catalogues – from Digikey, Electrosonic, Future Electronics or Jameco, for example – will be a useful resource.

1. Light meter using photoresistive cell

A Cadmium Sulphide (CdS) photoresistor is to be used to construct a low-cost analog light meter that shows light level on an analog milliammeter movement (1mA full scale, 64Ω) or an electronic thermometer display (National Semiconductor LM3914, LM3915).

In the light levels expected, a CdS sensor will vary from $10k\Omega$ to $100k\Omega$ in a fashion that is logarithmic with light level. The display may also be logarithmic, that is, you need not linearize the sensor response. The circuit should operate from a 9VDC battery.

Optionally, you may use some other photosensitive device.

2. Electronic door chime

A *chime* sound such as *bong* or *ding* can be created by pulsing an underdamped resonant circuit. The resonant circuit can be a passive LC circuit or an active filter with high Q (low damping factor). This circuit is to be used as the basis for an electronic door chime that can be tuned in frequency and damping by the end user. The underdamped pulse waveform feeds an audio amplifier that drives a loudspeaker.

3. Analog thermometer

A *thermistor* is a temperature sensitive resistor that may be used to construct a simple temperature sensing circuit. The thermistor resistance decreases with temperature and is given by:

$$R_T = Ae^{B/T}$$

where

- | | |
|-------|---|
| R_T | is the thermistor resistance |
| A | is the <i>linear factor</i> constant |
| B | is the <i>exponential</i> factor constant |
| T | is the temperature in Kelvins |

For design purposes, assume $B = 3892$ and the resistance of the thermistor is $10k\Omega$ at 25°C . A common arrangement is to put the thermistor in a voltage divider with a fixed resistance. The value of the fixed resistance is equal to the thermistor resistance at the mid-point of the temperature range.

Design a thermistor-based temperature measuring circuit that will operate over the range 0°C to $+50^\circ\text{C}$ and generates a voltage between 0 and 5 volts that can be displayed on a digital voltmeter. The circuit should operate from 9VDC.

4. LED flashlight with variable intensity

An LED flashlight is very economical of current and as a result, the batteries last a long time. Design a circuit to drive 4 high-intensity white LEDs from a 9VDC supply, with variable intensity.

The variable intensity control should be a duty-cycle modulated square wave. This can be obtained by creating a triangle wave and then comparing the triangle wave with a DC voltage derived from a potentiometer. The output of the comparator is a pulse waveform. As the pot is adjusted, the duty cycle of this pulse waveform changes. (Other circuits, for example using the 555 timer, are also possible).

The maximum current through any LED should be 50mA. At that current, the voltage across the LED is 1.7 volts.

5. Lamp Control Circuit This circuit ensures that a household lamp is turned off after an interval, thereby saving the lamp life and electrical power costs.

When a pushbutton is actuated, the lamp illuminates immediately. After a period of time (which is adjustable), the lamp extinguishes.

A suitable lamp control device is the *triac*.

6. High impedance analog voltmeter This voltmeter should present a high impedance (at least $10M\Omega$) to the measurement circuit and display the result on an analog milliammeter movement (1mA full scale, 64Ω) or an electronic thermometer display (National Semiconductor LM3914, LM3915.) It should operate from one or two 9VDC batteries and, with a range switch, be capable of measuring DC voltages 0-1 volt, 0-10 volts, 0-100 volts, full scale display.**7. Barking dog triggers sprinkler system** This product detects the noise of a barking dog and closes a relay. The relay contact closure can be used to actuate a solenoid water valve that turns on a sprinkler for a fixed period of time, such as 20 seconds.

The barking dog signal is picked up by an electret microphone. The output of the microphone may be assumed to be a 100mV pulse of 500 mSec duration for each dog bark. The solenoid water valve requires 1 Amp, 117VAC to actuate. The circuit should operate from a 9VDC battery.

8. Moisture monitor As soil dries out, its resistance increases. Consequently, the moisture of soil may be monitored by measuring the resistance. To avoid ion migration effects, the resistance measurement must be done with an AC voltage and current. A typical range of soil resistance is from 60Ω to $35k\Omega$. One possible approach to this measurement is to use a Wheatstone Bridge that is driven by an AC voltage. Then the voltage at the output of the bridge is a null when the bridge is balanced. If three of the bridge resistances are known, then the fourth (the soil resistance) may be calculated. The resistances can be calibrated on their dial faces and the reading read off a dial setting when the bridge is balanced. The bridge can be excited by an AC oscillator such as a Wien Bridge or phase shift oscillator.**9. Analog circuit for line-following motorized robot**

Design an analog circuit that will use two photosensitive resistors to guide a motorized robot, similar to the *Sumo* robot shown previously. The result of two photosensitive sensors is fed into two comparators, which then drive the motor circuit.

Ideally, both motors should run continuously. When a sensor strays over the line, it slows down one of the two motors so that the robot corrects its steering to stay on the line. That is, the correction signals should add to or subtract from the basic speed signal such that the basic speed can be adjusted independently of steering. The circuit must operate from a single 12VDC supply.

- 10. Magnetic Suspension Control System** In this system, a steel sphere is suspended some distance below an electromagnet. A negative feedback system senses the vertical position of the sphere and adjusts the current in the electromagnet to keep the sphere in the same position.

The position of the sphere is detected by a partially obscured light beam which shines on a photoresistor (or other photodector device). The signal from this circuit is then passed through a PID controller (11.7 on pageref 309) for adjustment of the loop stability. The output from the PID controller operates a 2 amp current source that drives the electromagnet.

- 11. Cloudy-Clear Sky Detector** Clouds tend to block infrared radiation from the surface of the earth. Consequently, a clear sky allows heat to radiate quickly into the sky and there is a large temperature differential between the ground and a patch of sky. When the sky is cloudy, the temperature differential is less.

This effect can be used as the basis for a cloud detector , using a *Peltier Cell*³³¹. The Peltier Cell is a large series of PN junctions that produces a voltage in response to a temperature differential between its two surfaces. A typical Peltier Cell is 4cm x 4cm in area. In this application, the cell is mounted in a tube so that one side faces the sky and the other the ground.

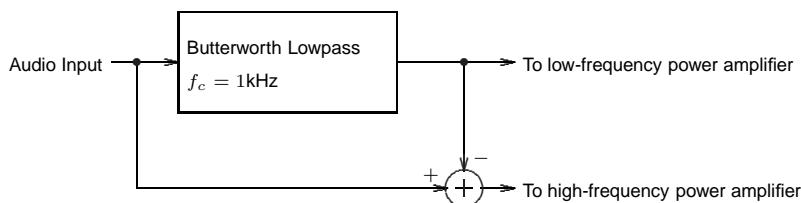
The Peltier Cell may be modelled as a voltage source of up to 40mV in series with an internal resistance of 2Ω . Design a circuit that will detect the cell voltage and close a relay when the sky is cloudy. The circuit should operate from a 9VDC battery.

- 12. Audio Crossover Unit** A *crossover unit* separates high and low frequencies and routes them to the corresponding tweeter and woofer loudspeaker driver units.

A *passive* crossover is placed between the audio power amplifier and the driver units, and consists of inductors and non-polarized capacitors.

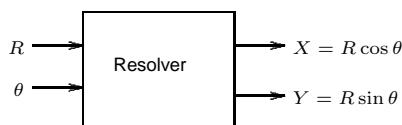
An *active* crossover system is placed before the audio power amplifier stage. The high and low frequencies are separated and routed to two power amplifiers, each of which operates a driver unit.

One possible design for a two-channel active crossover system is shown below. The lowpass signal is subtracted from the input signal to generate the highpass output.



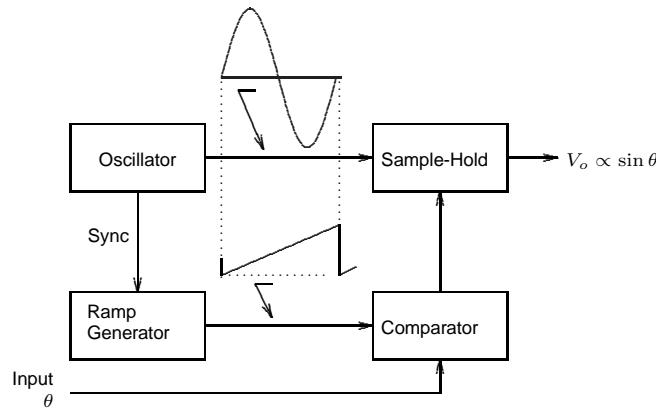
- (a) If a second-order lowpass filter is used, draw the diagram for the complete circuit.
- (b) Draw the frequency response for the high and lowpass outputs. (Note: The highpass and lowpass amplitude characteristics are not *mirror images* because of phase shift through the lowpass filter.)

- 13. Electronic Resolver** A *resolver* is a device which accepts the polar form of a vector $R\angle\theta$ and converts it to the rectangular form X, Y .



³³¹The Peltier Cell can be used in a reverse fashion. A current passed through the cell causes one plate to become hot and the other cold. This effect is used in solid-state coolers.

This system requires a circuit that will convert a voltage (representing the angle θ) into a voltage proportional to the sine of θ . One possible scheme is shown below. An oscillator generates a sine wave. The sine wave is then sampled at some point in time which is proportional to the angle θ . The combination of the ramp generator and comparator functions as a voltage-to-time converter that generates a sample signal at the correct time, proportional to angle.



- (a) How would you modify this circuit so that it will also generate a voltage proportional to the cosine of the angle θ ? Draw a suitable oscillator circuit.
 - (b) Draw a circuit for the ramp generator, showing how it is synchronized to the sine wave.
 - (c) Draw a circuit for the comparator and sample-hold circuits.
14. **Angle Measurement System** Section 37.4 described one method of driving the resolver, and the resultant output signals. There is an alternative [362]: the two stators are driven by AC signals that are in phase quadrature. The output voltage from the rotor is a sine wave that varies in phase angle with respect to the stator voltage. Taking one of the stator winding voltages as the zero degree phase reference, the rotor voltage remains constant in amplitude but varies smoothly from 0° to 360° as the rotor turns through a full rotation.
- Design an electronic interface for a resolver using this scheme, including the electronics to drive the resolver.
15. **Hand-On-Throttle Detector** Normally, the driver of a locomotive is expected to have his/her hand on the throttle lever at all times. As a safety feature, a temperature detector is used in the throttle handle to ensure that a live human hand is attached.
- The temperature sensor is a thermistor that varies in resistance between 3910Ω at room temperature and 3240Ω when heated by a human hand. The detector circuit is to be powered from a single 9 volt DC power supply that can vary in output by $\pm 0.5V$. The design should be such that a power failure causes the alarm to sound.
- Design a suitable circuit to operate a 100mA, 9 volt relay.

37.10 Extended Design Exercise: Telescope Heater

This section walks through the details of a design exercise for a possible electronic product.

As the earth makes the transition from day into night, the atmosphere cools. As it does so, moisture in the form of *dew* forms on surfaces at ground level. This is a major nuisance for amateur astronomers. Dew condenses on the lenses and mirrors of a telescope, rendering it unusable. However, if the critical surfaces are kept a few degrees about the ambient temperature they will remain moisture free.

At first blush, this sounds very simple. However, there are two complicating factors:

- Excessive heat creates convection currents which disturb seeing, so the heat must be kept to a minimum.
- Amateur telescopes are often operated at some distance from AC power, so the heaters must be powered from a 12 volt battery. In the interest of long battery life, the heater should operate at the minimum acceptable power level that prevents dew from forming. If the ambient temperature changes, then the heating system must respond which implies some kind of feedback.

A review of information on the internet indicates that dew prevention can be accomplished by maintaining the telescope a fixed amount ΔT above the ambient temperature – typically 5 degrees centigrade. For a telescope aperture of 8 inches diameter, the maximum power required is in the order of 10 watts. The controller switches the heater on and off to maintain ΔT at the setpoint value, so the average power requirement is less than this. As well as extending battery life, this minimal level of heating does not affect seeing.

In this design exercise, you will design the dew prevention heater and its controller.

Specifications

Here is a summary of the specifications:

Supply Voltage	12 VDC (nominal), 14 to 11V
Temperature Range	0 to 40°C
Control Range ΔT	0 to 5°C
LED Indicators	Battery Power Heater Power

In a commercial product, the part and assembly cost would also be important specifications.

Temperature Measurement

The suggested temperature sensor is a *thermistor*, which was described briefly in section 10.5. The thermistor is a resistor with a strong negative coefficient of resistance with temperature. The thermistor resistance is a highly non-linear function of temperature, with a characteristic given by:

$$R_T = R_{25} e^{\beta \left(\frac{1}{T} - \frac{1}{T_{25}} \right)} \quad (1679)$$

where the variables are:

R_T	Resistance of the thermistor at some temperature T
R_{25}	Resistance of the thermistor at 25°C
β	A constant that is the property of the device
T	Temperature of the thermistor in Kelvins (°C plus 273)
T_{25}	25° in Kelvins (298k)

Fortunately, it is possible to generate a voltage that is a more linear function of temperature by placing the thermistor in a voltage divider. The voltage divider fixed resistance is made equal to the thermistor resistance at the mid-point of its temperature range. The output of the voltage divider is then approximately a linear function of temperature. For example, if the thermistor is used to measure between 10°C and 50°C, the fixed resistance would be made equal to the thermistor resistance for 30°C.

Block Diagram

Now we are in a position to sketch out a block diagram, figure 1024.

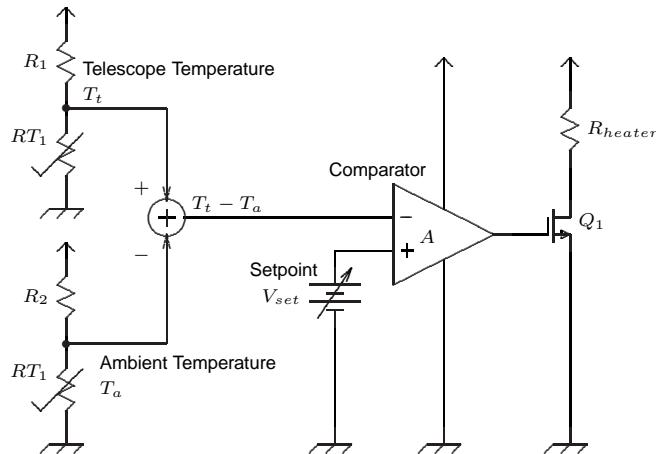


Figure 1024: Telescope Dew Heater, Block Diagram

Two voltage dividers generate voltages proportional to the telescope temperature and the ambient temperature. A subtractor calculates the difference voltage V_{diff} between them. This voltage is applied to a comparator. If the difference voltage V_{diff} drops below the setpoint voltage V_{set} , then the output of the comparator swings high, turning on the power FET Q_1 , which then applies power to the heater.

The difference calculation and comparator are shown as two different stages. It's possible to combine that circuitry into one op-amp, but

- This arrangement is much easier to troubleshoot. A voltmeter at the output of the subtractor immediately indicates the relative temperature of the two thermistors, and
- Op-amps come in packages of 4, so using an extra op-amp is not extravagant.

Detailed Electronic Design

Here are some hints and notes on the electronic design of this circuit.

- **Operating Temperature**

A reasonable range of operating temperatures³³² is 0°C to 40°.

- **Battery Voltage**

The battery voltage is nominally 12 volts. The variation in actual output voltage depends on the battery type and current load. Assuming an SLA (sealed lead-acid) battery, you may assume it will vary between 14 volts when the battery is fully charged to 11 volts at full discharge.

- **Battery Endurance**

Assuming that the average output power of the heater is 5 watts over the course of an observing session, how long can a 6 amp-hour battery provide power for this device? You can assume that the controller electronic circuits require negligible power.

- **Heater Resistance**

The heater operates from the full battery supply voltage, 12V. What should be the heater resistance if it can develop 10 watts from 12 volts?

- **Single Supply Operation**

The circuit must operate from a single supply voltage, so a single-supply op-amp is appropriate. By using a single-supply quad op-amp like the LM324, one op-amp can function as the subtractor and a second op-amp as the comparator, with two spares for other functions. For example, the subtractor can use two op-amps. With a single supply, the outputs from the op-amps cannot produce a negative voltage, so all signal voltages must be positive.

- **Voltage Regulation**

To guarantee reliable operation, certain internal voltages of the circuit should be regulated. A good choice for the regulated voltage would be 5 volts. Five volt regulators are the most common type and consequently readily available and inexpensive.

- **Thermistor Specifications**

A wide variety of thermistors could be used in this circuit. One inexpensive device is:

Type	Panasonic PNT113-ND
R_{25}	1000Ω
R_{25}/R_{50}	2.55
Diameter	5mm
Supplier	Digikey
Price	\$1.49 US, Qty 1

The R_{25}/R_{50} gives the resistance ratio between the resistance at 25°C and 50°C. Using the specifications for this device (or some other thermistor of your choosing), determine the resistance of the thermistor at the temperature endpoints and midpoint.

- **Thermistor Voltage Dividers**

Design the thermistor voltage dividers. Based on a divider supply voltage of 5 volts, what is the output voltage range of the voltage divider at the extremes and mid-point of temperature?

Plot divider output voltage vs temperature to confirm that the relationship is approximately linear.

³³²This depends entirely on geographic location, of course. As this is being written in Toronto, Canada, it is -16°C.

What is the measurement sensitivity in volts/ $^{\circ}\text{C}$? Assuming that ΔT can vary between 0°C and 5°C , what range of voltages would you expect at the output of the subtractor?

- **Op-amp Input Voltage Range**

An operational amplifier has certain limits to the voltage range at its input terminals. For example, the LM324 can accept voltages between zero and $V_{CC} - 1.5$ volts, where V_{CC} is the positive supply voltage.

The subtractor circuit and the comparator circuit must be checked to ensure that the signal voltages do not exceed this range. For example, the output of the subtractor can range down to zero volts, so the comparator must be able to accept this input voltage.

- **Setpoint Network**

The setpoint voltage is generated by an adjustable voltage divider. The control is a front panel potentiometer that adjusts ΔT between 0°C and 5°C . Within wide limits, this divider design can be driven by the availability of a low-cost linear panel-mount potentiometer. The adjustment range should use the full 270° rotation range of the pot.

- **Comparator Hysteresis**

To avoid oscillation of the comparator output when the two input voltages are equal, the comparator should include a small amount of hysteresis. This hysteresis is created by positive feedback from the output back to the input, as shown in figure 1025, so that the comparator operates as a Schmitt Trigger. A suitable value of hysteresis might be a voltage equivalent to 0.2°C . Then, if the value of ΔT is set to 3°C , it would actually ramp up and down between 3.1°C and 2.9°C . The precise value of the hysteresis is not critical as long as it's not excessively large, nor is it important that the hysteresis be symmetrical about the setpoint.

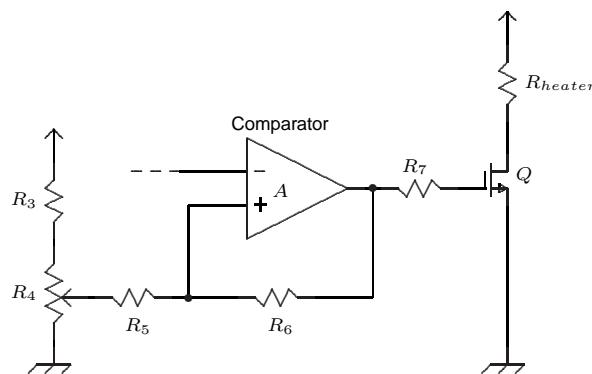


Figure 1025: Comparator Detail

A possible design process for this network would be:

- Choose the pot R_4 , a fairly low value resistance.
- Choose R_5 to be large compared to R_4 so that it has no effect on the wiper voltage of R_4 .
- Based on the output voltage swing of the op-amp, choose R_6 to provide the necessary hysteresis.

- **Power Switch Requirements**

The choices for the power switch device are:

Relay A relay is a relatively bulky and expensive device, and the contacts may eventually become intermittent. Its main advantage is that the switched circuit is totally independent from the drive circuit, but that's not a requirement here.

Power BJT A collector current of 1 amp and beta of about 20 or so would require a base current of 50mA. This is beyond the capability of an op-amp, so an additional driver stage would be required.

Power BJT Darlington A Darlington BJT helps with the base current problem, but introduces another one. The saturation voltage is in the order of 2 volts, so the power dissipation with 1 amp collector current is in the order of 2 watts. This is likely to require a heatsink.

Power MOSFET A MOSFET requires no current drive, which makes it an improvement over the power BJT. Its power dissipation is proportional to $I^2 R_{DS(on)}$, which is an improvement over the power BJT Darlington. It is also competitive in price with a power BJT or Darlington BJT.

Consequently, the device of choice for this application is the Power MOSFET.

The heater resistance is specified at its operating temperature. When the heater is cold, its resistance is much lower. At first switch-on, there is a significant surge of current through the resistance. This has two important consequences:

- For reliability, the current rating of the MOSFET should be well in excess of the requirement. (The extra cost of providing this extra current capability is very modest.)
- The surge current will cause a sudden increase in voltage drop across the internal resistance and an abrupt momentary drop in battery voltage. This could be a problem for other equipment connected to the same battery. (There are reports of telescope microprocessors crashing under these circumstances.)

• Power MOSFET Specifications

A search of distributor catalogues turns up the IRL2703 as a suitable MOSFET device. From the datasheet:

Type	International Rectifier IRL2703-ND
Drain Current $I_{DS(max)}$	25A
On Resistance $R_{DS(on)}$	0.06Ω
Drain-Source Voltage $V_{DS(max)}$	30V
Gate-Source Voltage Maximum $V_{DS(max)}$	±16V
Case	TO-220
Operating junction temperature range	–55 to +175°C
Thermal Resistance, Junction to Case θ_{jc}	3.3°C/W
Thermal Resistance, Case to Sink (greased) θ_{cs}	0.5°C/W
Thermal Resistance, Junction to Ambient θ_{ja}	62°C/W
Supplier	Digikey
Price	\$0.94 US, Qty 1

The MOSFET maximum drain current $I_{DS(max)}$ and drain-source voltage $V_{DS(max)}$ both meet the requirements of this application.

We should also check that the MOSFET cannot be destroyed by excessive gate-source voltage. Since the op-amp comparator stage is powered from +12 volts, it cannot exceed the MOSFET Gate-Source Voltage Maximum $V_{DS(max)}$ of 16 volts.

• Heatsink

Determine whether a heatsink will be required, and if so, the required heatsink area. A suitable choice for maximum allowable junction temperature is 100°C, which provides a significant safety margin over the specification. The transistor must not be allowed to overheat even when the heater is continuously ON.

- **MOSFET Drive**

Referring to figure 3 of the datasheet, *Typical Transfer Characteristics*, it can be seen that the gate-source voltage must be at least 3.6 volts for the MOSFET to conduct a drain current of 10 amps. An op-amp like the quad, single-supply LM324 is a low-cost choice but its maximum output voltage is at least 1.2 volts below the positive power supply rail. An LM324 operated from 5 volts would be limited to a maximum output voltage of 3.8 volts. This is rather marginal.

There are two alternatives:

- Use an LM324 type device, but operate its V_{CC} supply from +12 volts. Then the maximum output voltage is +10.8 volts, sufficient to drive the gate of the MOSFET.
- Use a so-called rail-rail op-amp such as the LMC660, which can output a full +5 volts when operated from a 5 volt power supply.

In a commercial design, the first alternative is to be preferred since the LM324 is much less expensive than the LM660.

- **Noise Control**

When the heater switches on, there is an abrupt change of current level in the heater. The coupling into other circuits is proportional to the rate of change of the current. To prevent the transmission of noise into other circuits that may be using the same battery, it is prudent to provide a noise filter at the point where 12 volt power enters the circuit. Commercial noise filter modules are available, but they are expensive. In practice, a small inductor called a *hash choke* of about $100\mu\text{H}$ at a suitable current rating in series with the electrical current, accompanied by 100nF ceramic capacitors to ground, is effective in containing the noise.

- **Indicators**

For the cost of a few LEDs, it is possible to equip the circuit with useful indicators. For example, it would be useful to know

- When battery power is present and in the correct range
- When the heater is operating correctly (ie, switching on and off.)

Design these indicators into your circuit.

Mechanical Design

The mechanical design of the circuit impacts

- Cost of the parts
- Cost of manufacturing and assembly
- Ease of use
- Reliability

It is therefore worthwhile to put some careful thought into the mechanical design.

In a large company, there may be product designers and mechanical engineers who handle this aspect of the design. However, in many smaller firms and certainly for individual practitioners, the designer of the electronics is responsible for the packaging and mechanical design.

- Using a circuit layout CAD program, design the printed circuit board.

- Based on the size of the resultant circuit, choose a suitable container for the circuit and the arrangement of the circuit card in the enclosure.
- Determine how the cable connections will attach to the circuit board. For example, here are two possible configurations:
 - Figure 1026. There are two screw-type terminal wiring blocks. The user opens the box, wires the cables into the screw terminals, and replaces the box cover. The screw terminal blocks mount on the printed circuit board and are soldered directly into it. The customer is provided with cables that must be cut to the correct length, stripped, and then wired into the terminal blocks.

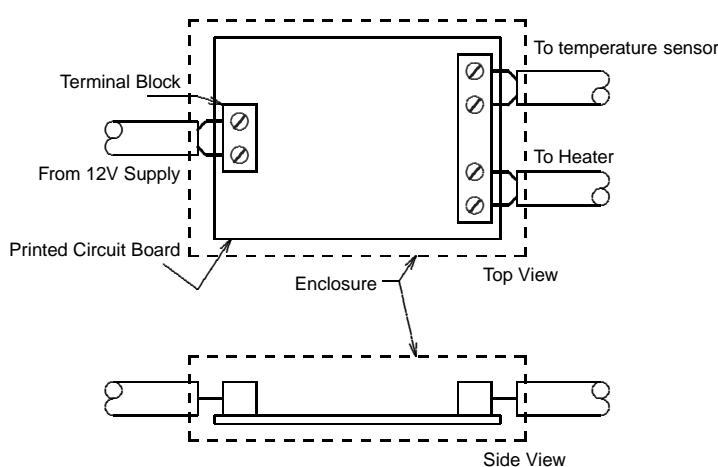


Figure 1026: Enclosure Concept

- There are two connectors, which mount on the box. One connector is for 12 volt input power. The other connector is for the telescope temperature sensor and the heater. The connectors are wired back to the printed circuit board. It should be possible to remove the box without unwiring the connectors. The customer is provided with cables with the connectors attached.

In both cases, the cables must be mechanically secure (*strain relieved*). What are the pros and cons of these two schemes?

- How will the temperature adjustment control mount so that it minimizes cost and wiring?
- How would the indicator LEDs be mounted?

Summary Documentation

Prepare a complete package of the intellectual property, including

- Final specifications
- Schematic diagram
- Wiring diagram, showing terminal connections to external cables
- Mechanical assembly diagram
- Parts list, showing quantity, description, manufacturer, supplier and cost for each part.

References

- [1] Frank Goodenough. *The Magic of Analog Design.* *Electronic Design*, pages 75–80, November 25 1992.
- [2] National Magnetics Group Inc. *Effective Permeability of Ferrite Rods.* <http://www.magneticsgroup.com/pdf/erods.pdf>.
- [3] Editor Robert Pease. *Analog Circuits, World Class Designs.* Newnes, Oxford, UK, 2008.
- [4] Murray R. Spiegel. *Mathematical Handbook of Formulas and Tables.* McGraw-Hill, New York, N.Y., 1968. A useful reference text for mathematical formulae.
- [5] Syscomp Electronic Design. <http://www.syscompdesign.com>
- [6] Eric Weisstein. *Franklin, Benjamin (1706-1790)*, 2008. <http://scienceworld.wolfram.com/biography/FranklinBenjamin.html>.
- [7] Tom Tucker. *Bolt of Fate: Benjamin Franklin and his Fabulous Kite.* Public Affairs, New York, 2003.
- [8] American Institute of Physics. *A Look Inside the Atom*, 2008. <http://www.aip.org/history/electron/jjhome.htm>
- [9] *Scientific Notation.* http://en.wikipedia.org/wiki/Scientific_notation
- [10] *Millikan Oil Drop Experiment.* http://en.wikipedia.org/wiki/Millikan_oil_drop_experiment.
- [11] John Linsley Hood. *Valve and Transistor Audio Amplifiers.* Newnes, 1997.
John Linsley Hood is an expert of long standing in the design of audio amplifiers. This book includes interesting examples of various vacuum tube and solid state amplifiers.
- [12] Arie van Rhijn. *Integrated Circuits for High Performance Electret Microphones.* *Audio Engineering Society Convention Paper, 114th Convention*, March 22-25, 2003. <http://www.national.com/appinfo/amps/files/aes2003.pdf>.
- [13] Donnie Curington. *Isolation, Is Your Measurement System Safe.* <http://archives.sensorsmag.com/articles/0900/60/main.shtml>.
- [14] *Lightning.* <http://en.wikipedia.org/wiki/Lightning>.
- [15] Powerhouse Museum Collection: Battery, *Standard Cell, 1924.* <http://www.powerhousemuseum.com/collection/database/?irn=319067>.
- [16] *A Practical Josephson Voltage Standard at One Volt.* <http://nvl.nist.gov/pub/nistpubs/sp958-lide/315-318.pdf>.
- [17] *History of the ampere.* <http://www.sizes.com/units/ampHist.htm>.

- [18] *Absolute Determination of the Ampere.*
<http://nvl.nist.gov/pub/nistpubs/sp958-lide/056-058.pdf>.
- [19] *Residual-current device.*
http://en.wikipedia.org/wiki/Ground_Fault_Interrupter.
- [20] C.R. Nave. *Microscopic View of Ohm's Law*, 2006.
<http://hyperphysics.phy-astr.gsu.edu/hbase/electric/ohmmic.html>.
- [21] *Georg Ohm.*
http://en.wikipedia.org/wiki/Georg_Ohm.
- [22] Gerald Recktenwald. *Conversion of Thermocouple Voltage to Temperature*, April 16, 2006.
web.cecs.pdx.edu/~gerry/epub/pdf/thermocouple.pdf.
- [23] *Ohm's law.*
http://en.wikipedia.org/wiki/Ohm%27s_Law.
- [24] *Quantum Hall Effect as Resistance Standard.*
[http://hrma.physics.sjtu.edu/PhysicsHorizon/QHE-as-R-S_A01.2\(01455\).pdf](http://hrma.physics.sjtu.edu/PhysicsHorizon/QHE-as-R-S_A01.2(01455).pdf).
- [25] *Precision Resistors.*
<http://www.riedon.com>.
- [26] *Preferred Number.*
http://en.wikipedia.org/wiki/Preferred_number#Capacitors_and_resistors.
- [27] *Resistivity.*
<http://en.wikipedia.org/wiki/Resistivity>.
- [28] *Nichrome.*
<http://en.wikipedia.org/wiki/Nichrome>.
- [29] Howard Johnson, Martin Graham. *High-Speed Digital Design: A Handbook of Black Magic*. Prentice Hall, 1993.
- [30] John Vandersleen. *Printed Wiring Board Manufacturing Advances. Power Electronics Technology*, pages 40–43, September 2004.
<http://powerelectronics.com/mag/409pet23.pdf>.
- [31] *Galvanometer.*
<http://en.wikipedia.org/wiki/Galvanometer>.
- [32] *Electronic Volt-Ohm Meter*. In *RCA Receiving Tube Manual*. RCA Victor Company, 1966.
- [33] National Semiconductor. *FET Circuit Applications*, February 1970.
<http://www.national.com/an/AN/AN-32.pdf>.
- [34] Tony R. Kuphalldt et al. *High-Impedance Voltmeter*.
www.allaboutcircuits.com/vol_6/chpt_6/5.html.
- [35] *Shunt (electrical)*.
[http://en.wikipedia.org/wiki/Shunt_\(electrical\)](http://en.wikipedia.org/wiki/Shunt_(electrical)).

- [36] William David Cooper. *Electronic Instrumentation and Measurement Techniques, 2nd Edition*. Prentice Hall, Englewood Cliffs, New Jersey, 1978.
- [37] Auto Batteries.com. *Frequently Asked Questions (on car batteries)*, 2008.
<http://www.autobatteries.com/faq/index.asp>.
- [38] *Leyden Jar*.
http://en.wikipedia.org/wiki/Leyden_jar.
- [39] *Wimshurst Machine*.
http://en.wikipedia.org/wiki/Wimshurst_machine.
- [40] *Battery energy*. <http://www.allaboutbatteries.com/Battery-Energy.html>.
- [41] *Electric double-layer capacitor*.
<http://en.wikipedia.org/wiki/Supercapacitor>.
- [42] *Energy Density*.
http://en.wikipedia.org/wiki/Energy_density.
- [43] *Ignition System*.
http://en.wikipedia.org/wiki/Ignition_system.
- [44] Dr. Richard Vawter. *Energy Stored in an Inductor*.
<http://www.ac.wvu.edu/~vawter/PhysicsNet/Topics/Inductance/EnergyStorage.html>.
- [45] *The ARRL Handbook*. American Radio Relay League.
Published annually, always contains one chapter on construction techniques. Reasonable price, a must-have.
- [46] Vannevar Bush. *The Differential Analyser, a new machine for solving differential equations*. *Journal of the Franklin Institute*, 212:447 fol, 1931.
- [47] Peter D. Hiscocks. *A Relay Counter*, 2000.
<http://www.ee.ryerson.ca/~phiscock/papers/relay-counter.pdf>.
- [48] *Marx generator*.
http://en.wikipedia.org/wiki/Marx_generator.
- [49] Lucid Science. *Rock Disaggregator*.
<http://www.lucidscience.com/gal-rock%20disaggregator-1.aspx>.
- [50] *COS/MOS Rate Multipliers: Versatile Circuits for Synthesizing Digital Functions*. RCA Application Note ICAN-6739, 1980.
- [51] *Complex Number*.
http://en.wikipedia.org/wiki/Complex_number.
- [52] *Paul's Online Math Notes*.
<http://tutorial.math.lamar.edu/Extras/ComplexPrimer/Forms.aspx>.
- [53] Complex number arithmetic.
http://www.allaboutcircuits.com/vol_2/chpt_2/6.html.

- [54] *Euler's formula.*
http://en.wikipedia.org/wiki/Euler%27s_formula.
- [55] *Reactance Slide Rule.*
http://www.hpmemory.org/pict/technics/applet/slide_rule/reactance.htm
A marvellous web-based version of the HP sliderule reactance calculator.
- [56] *Shure Reactance Slide Rule.*
<http://500kc.com/downloads/1957%20Shure%20Reactance%20Slide%20Rule>
High resolution copies of a reactance slide rule.
- [57] *LC circuit.*
http://en.wikipedia.org/wiki/LC_circuit.
- [58] *Simple parallel (tank circuit) resonance. All About Circuits*, 2008.
http://www.allaboutcircuits.com/vol_2/chpt_6/2.html.
- [59] Hiscocks et al. *Electrical Concepts and Applications*. West, Minneapolis-St.Paul, 1997.
- [60] Roy Bishop. *Tides and the Earth-Moon System*. In Rajiv Gupta, editor, *Observer's Handbook 2003*. The Royal Astronomical Society of Canada, 2003.
- [61] J.David Irwin. *Basic Engineering Circuit Analysis, Fifth Edition*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.
- [62] *Tuned mass damper.*
http://en.wikipedia.org/wiki/Tuned_mass_damper.
- [63] *KollÁgumi Jegyzetszervere.*
<http://jegyzet.sth.sze.hu/ftp/!BSc/Szabalyozastechnika/BodePl.pdf>.
- [64] Peter Hiscocks and James Gaston. *A Software-Based Network Analyzer*. June 2006.
<http://www.syscompdesign.com/AppNote.html>.
- [65] *Computer Algebra System*. http://en.wikipedia.org/wiki/Computer_algebra_system.
- [66] *Comparison of Computer Algebra Systems*.
http://en.wikipedia.org/wiki/Comparison_of_computer_algebra_systems.
- [67] *Maxima home page*.
<http://maxima.sourceforge.net/>.
- [68] Donald P. Cambell Gordon S. Brown. *Principles of Servomechanisms*. John Wiley and Sons Inc, London, 1948.
- [69] Keith E. Holbert. *Laplace Transform Solutions of Transient Circuits*.
www.eas.asu.edu/~holbert/eee202/EEE202_Lect13_LaplaceSolutionTransientCircuits.ppt
Slides on Laplace Transform of Initially Charged Capacitor.
- [70] Robert Miller. *Algebraic Transient Analysis*. Rinehart Press, San Francisco, 1971.
- [71] Artice M. Davis. *Linear Circuit Analysis*. PWS Publishing Company, 1998.
- [72] Kurt Bryan.
<http://www.rose-hulman.edu/~rickert/Classes/ma222/Wint0102/dirac.pdf>.

- [73] Kobayashi Ichiro, Sato Katsuhiko, Tsuyuki Shigeru, and Fukushima Kazuhiro. *International Education Software*.
<http://www.ies.co.jp/math/java/calc/lopi/rule.html>.
- [74] Steven W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing, Chapter 32, The Laplace Transform*.
<http://www.dspguide.com/CH32.PDF>.
- [75] Richard G. Lyons. *Understanding Digital Signal Processing, 2nd Edition, Section 6.2, The Laplace Transform*. Pearson Education, 2004.
- [76] Academic Book Exchange, AbeBooks.com.
<http://www.abebooks.com/>.
- [77] Millman & Taub. *Pulse, Digital and Switching Waveforms*. McGraw-Hill, New York, 1965.
- [78] Adel S. Sedra and Kenneth C. Smith. *Microelectronic Circuits*. Oxford University Press, 1991.
- [79] David A. Johns and Ken Martin. *Analog Integrated Circuit Design*. Wiley, 1997.
- [80] Geoff Phillips. *Power Mosfets*. *Electronics Today International (ETI)*, pages 16–19, November 1986.
- [81] Bill Davies. *Practical Robotics*. WERD Technology Inc, Toronto, Ontario, 1997.
- [82] Bill Davies. *Electric Motors and Mechanical Devices for Hobbyists and Engineers*. WERD Technology Inc, Toronto, Ontario, 1997.
- [83] Harold S. Black. *Inventing the Negative Feedback Amplifier*. *IEEE Spectrum*, pages 55–60, December 1977.
- [84] James E. Solomon. *The Monolithic Operational Amplifier: A Tutorial Study*. *IEEE Journal of Solid State Circuits*, SC-9, No. 6:314 to 332, December 1974. www.national.com/an/AN-AN-A.pdf.
- [85] Jiří Dostál. *Operational Amplifiers*. Elsevier Scientific Publishing Company, New York, 1981. An excellent source of material on testing operational amplifier characteristics, and on compensation and stability.
- [86] Michael Street. *Simplify ac circuit analysis with Reactance-Bode charts*. *EDN*, pages 83–90, May 5 1976. The definitive article on reactance charts.
- [87] C.J. Savant. *Control System Design, 2nd Edition*. McGraw-Hill, New York, 1964.
- [88] Donald G. Schultz Charles E. Rhors, James L. Melsa. *Linear Control Systems*. McGraw-Hill, New York, 1993.
- [89] Katsuhiko Ogata. *Modern Control Engineering, 2nd Edition*. Prentice-Hall, Englewood Cliffs, 1990.
- [90] J.G. Ziegler and N.B. Nichols. *Optimal settings for automatic controllers*. *Trans. ASME*, 64:759–768, 1942.
- [91] David Marsh. *Op amps take the next step*. *EDN*, pages 75–84, September 26 2002. A review of current technology in op-amp devices.
- [92] Ron Mancini. *Single-Supply Op Amp Design Techniques*. *Texas Instruments Application Report SLOA030A*, pages 1–18, March 2001.

- [93] Bruce Carter. *A Single-Supply Op-Amp Collection*. Texas Instruments Application Report SLOA058, pages 1–26, November 2000.
- [94] John M. Miller. *Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit*. *Scientific Papers of the Bureau of Standards*, 15(351):367–385, 1919-20. This is the original paper on what is now known as the *Miller effect*. Kent Lundberg has kindly posted a copy of this paper on his web page at:
<http://web.mit.edu/klund/www/papers/jmiller.pdf>.
- [95] B.H.List R.C.McMaster, R.L.Merrill. *Analogous System in Engineering Design*. In H.M.Paynter, editor, *A Palimpsest on the Electronic Analog Art*. George A. Philbrick Researches, 1955.
- [96] George Fox Lang. ‘*Analog*’ was not a Computer Trademark. *Sound and Vibration*, pages 16–24, August 2000.
- [97] AN272: *Op Amp Booster Designs*. In *Linear Applications Handbook*. National Semiconductor, 1994.
- [98] Jim Williams. *Application Note 18: Power Gain Stages for Monolithic Amplifiers*. In *Linear Applications Handbook Volume 1*. Linear Technology, 1990.
- [99] Boon-Chau Tho Jong-Lick Lin. *Analysis and μ -Based Controller Design for an Electromagnetic Suspension System*. *IEEE Transactions on Education*, 41, No 2:116–129, May 1998.
- [100] P.Horowitz and W.Hill. *The Art of Electronics*, 2nd Ed. Cambridge University Press, Cambridge, 1989.
- [101] Peter Carroll Dunn. *Gateways into Electronics*. Wiley-Interscience, New York, 2000. An interesting and somewhat different approach to an electronics textbook.
- [102] Wikipedia. *Phantom Power*.
http://en.wikipedia.org/wiki/Phantom_power.
- [103] Jim Williams. *Max Wien, Mr Hewlett, and a Rainy Sunday afternoon*. In Jim Williams, editor, *Analog Circuit Design: Art, Science and Personalities*. Butterworth-Heinemann, 1991. Jim Williams brings his inimitable style to an analysis of the Wein Bridge oscillator, a historical description of the original HP200A oscillator, and a modern approach to the design. Also included, a list of important references: the original paper by Max Wien, and William Hewlett’s thesis. One of the definitive works on the circuit.
- [104] Bruce Carter. *An audio circuit collection, Part 1*. *Analog Applications Journal, Texas Instruments Incorporated*, pages 39–43, November 2000.
- [105] Bruce Carter. *An audio circuit collection, Part 2*. *Analog Applications Journal, Texas Instruments Incorporated*, pages 41–47, February 2001.
- [106] Bruce Carter. *An audio circuit collection, Part 3*. *Analog Applications Journal, Texas Instruments Incorporated*, pages 34–38, July 2001.
- [107] Bernard Sklar. *The Sixteen Faces of Feedback: A Catalogue of the Basic Negative Feedback Amplifier Configurations*. *IEEE Transactions on Education*, E-10:161–166, September 1967. In a very clear presentation, Sklar shows how the units of a negative feedback sensor network (volts/volt, amps/volt, etc) determine the overall transfer function. The four different types of sensors may be used with the four different types of amplifiers, giving 16 possible amplifier-feedback systems.

- [108] *Applications Manual for Computing Amplifiers for Modelling, Measuring, Manipulating & Much Else.* Philbrick Researches Inc, 1966. A somewhat antique but still educational collection of operational amplifier circuits, written with an endearing style.
- [109] D.E.Pippenger and E.J.Tobaben. *Linear and Interface Circuits Applications, Volume 1: Amplifiers, Comparators, Timers, Voltage Regulators.* Texas Instruments, 1985. TI Publication TI-SLYA001.
- [110] Wikipedia. *AC power plugs and sockets.*
http://en.wikipedia.org/wiki/AC_power_plugs_and_sockets.
- [111] O.H.Schade. *Analysis of Rectifier Operation.* Proc IRE, 31:341–361, July 1943. A classic paper on the design of half and full-wave rectifier-capacitor power supplies.
- [112] On Semiconductor. *Linear & Switching Voltage Regulator Handbook.* Publication HB206/D, Rev 4, Feb-2002. A comprehensive introduction to power supply design. Includes design curves of reference [111].
<http://www.onsemi.com/pub/Collateral/HB206-D.PDF>.
- [113] Robert.A.Pease. *Troubleshooting Analog Circuits.* Butterworth-Heinemann, 1991. Anyone who designs and builds analog circuits should read this book. A wealth of practical information on circuit pitfalls and troubleshooting.
- [114] Wikipedia. *Fourier series.*
http://en.wikipedia.org/wiki/Fourier_series.
- [115] C.Fouret C.Basso and P.Kadanka. *Get the best from your low-dropout regulator.* EDN, pages 115–126, February 19 1999.
- [116] Chester Simpson. *Linear Regulators: Theory of Operation and Compensation.* National Semiconductor Application Note 1148, May 2000. Notes on the stabilization of the Low Drop Out regulator.
- [117] Jerrold Foutz. *Switching-Mode Power Supply Design.* 2003.
<http://www.smpstech.com/tutorial>.
- [118] David Duff. *034 Electronic Circuits II, Supplementary Notes.* Ryerson University, 1992. Notes on the switching regulator.
- [119] *Aluminum Electrolytic Capacitors, Catalog H7.* United Chemi-Con, Rosemount, Illinois, USA, 1986. A manufacturer's catalog of data for electrolytic capacitors.
- [120] Daewoo Corp. *Mini Glossary, Summary of Electrolytic Capacitors,*.
http://www.dwecc.com/data/content/document/221/MINI_GLOSSARY.pdf
A very useful summary of the properties of electrolytic capacitors.
- [121] Brian Huffman. *Application Note 91: Efficiency and Power Characteristics of Switching Regulator Circuits.* In *Linear Applications Handbook Volume 2.* Linear Technology, 1993. An excellent collection of circuits and information on the design of switching regulators.
- [122] Maxim Semiconductor. *Charge pumps shine in portable designs.* Application Note 669, 15 March 2001.
- [123] John P. Skurla. *Analog switch acts as dc/dc converter.* EDN, pages 118,119, January 20 2000.
- [124] J.S.Bugler. *Theoretical Performance of Voltage Multiplier Circuits.* IEEE Journal of Solid State Circuits, pages 132–135, June 1971.

- [125] Eric Lowdon. *Practical Transformer Design Handbook*. Howard W. Sams & Co, Indianapolis, Indiana, 1980. A useful source for practical information on power transformer design. Unfortunately, the author does not show development of the equations in the text.
- [126] Marty Brown. *Power Supply Cookbook*. Butterworth-Heinemann, 1994. An excellent introduction to switching power supply design by an obviously experienced designer.
- [127] Magnetics Inc. *Powder Cores Magnetics*. Design information on the application of ferrite cores.
<http://www.mag-inc.com/design/technical-documents>.
- [128] Wikipedia. *Circular mil*.
http://en.wikipedia.org/wiki/Circular_mil.
- [129] Douglas Brooks. *Fusing Current: When Traces Melt Without a Trace*. Originally printed in 'Printed Circuit Design, a Miller Freeman publication, December, 1998
<http://www.ultracad.com/articles/fusing.pdf>.
- [130] Abraham I. Pressman. *Switching Power Supplies, Second Edition*. Abraham I. Pressman, 1998. A clear, extensive and thorough treatise on switching power supply design.
- [131] Dan L. Vogler. *Bi-FET devices improve absolute-value amplifier*. *Electronics*, page 105, May 12 1977.
- [132] Dan Sheingold. *Nonlinear circuits handbook*. Analog Devices, Cambridge, Massachusetts. A very useful collection of non-linear op-amp and analog module applications.
- [133] Gerald G.Graeme. *Applications of Operational Amplifiers – Third Generation techniques*. McGraw-Hill, New York, 1973.
- [134] Peter D. Hiscocks. *A simple and accurate voltage squaring circuit*. *Canadian Electronics Engineering*, page 29, Oct 1976. A simple squaring circuit using a JFET.
- [135] Barrie Gilbert. *Novel Technique for RMS-DC Conversion Based on the Difference of Squares*. *Electronics Letters*, 11:181–182, April 17 1975 1975.
- [136] Barrie Gilbert. *Current Mode Circuits from a Translinear Viewpoint: A Tutorial*. In D.G.Haigh C. Toumzou, F.J.Lidgey, editor, *Analogue IC Design: the Current Mode Approach*, chapter 2. Peter Peregrinus, 1990.
- [137] Tom Cate Howard Handler. *True RMS voltage conversion*. *Electronic Design*, pages 66–72, February 15 1974. A design discussion and circuit for the implicit method of RMS conversion.
- [138] Jim Williams. *A Monolithic IC for 100MHz RMS-DC Conversion*. In James M. Williams, editor, *1990 Linear Applications Handbook*, chapter 22. Linear Technology, 1987.
- [139] S.Franco. *Design with Operational Amplifiers and Analog Integrated Circuits*. McGraw-Hill, New York, 1998.
- [140] *LF198 Monolithic Sample and Hold Circuits*. In *National Semiconductor Linear Applications Handbook*. National Semiconductor Corporation, 1994.
- [141] E.J.Kennedy. *Operational Amplifier Circuits, Theory and Applications*. Holt, Rinehart and Winston, 1998.
- [142] Steve Guinta. *Ask the Applications Engineer 21: Capacitance and Capacitors*. *Analog Dialogue*, 30-2, 1996. A very useful summary of the properties of different types of capacitors.

- [143] *M68HC11 Reference Manual*. Motorola Semiconductor, Phoenix, Arizona, 1991. Document number M68HC11RM/AD.
- [144] National Semiconductor. *Specification and Architectures of Sample-Hold Amplifiers. Application Note 775*, July 1992.
- [145] Dan L. Vogler. *Precision peak-to-peak ac-dc converter built with single-supply op-amp circuit*. *Electronic Design*, page 160, November 22 1977.
- [146] Martin Giles. *Audio/Radio Handbook*. National Semiconductor Corporation, 1980. A classic source of information on audio circuit designs. Out of print from the original source, but currently available from audioexpress.com.
- [147] Peter Hiscocks. *Audio Spectrum Analyser*. *Wireless World Magazine*, pages 55–60, May 1980. An audio-spectrum analyser that is suitable for construction by hobbyists.
- [148] Don Lancaster. *Active Filter Cookbook*. Howard W. Sams, 1975.
- [149] Lawrence P.Huelsman. *Active and Passive Analog Filter Design: an introduction*. McGraw-Hill, New York, 1993.
- [150] *IEEE Std 315-1975, IEEE Standard American National Standard Canadian Standard graphic symbols for electrical and electronics diagrams*. 1975.
- [151] M. Bronzite. *Simple Active Filter: Design Procedure*. *Wireless World*, pages 117–119, March 1970. A simple two-amplifier lowpass or highpass second-order circuit design.
- [152] Michael Wyatt. *Active filter makes component selection easier*. *EDN*, pages 42,43, July 22 1993. A brief description of a two-amplifier lowpass or highpass second-order circuit design.
- [153] Russell Kincaid. *RC filter design by the numbers*. *The Electronic Engineer*, pages 57–64, October 1968. A very useful tabular method for designing Sallen-Key lowpass and highpass filters, for the Bessel, Butterworth and Chebyschev characteristics up to order 10. If you can live with a unity gain Sallen-Key design, this is the most convenient way to get there.
- [154] E.L.Key R.P.Sallen. *A Practical Method of Designing RC Active Filters*. *IRE Transactions on Circuit Theory*, CT-2:74–85, March 1985. The Sallen-Key filter configuration is revealed to the world.
- [155] Robert Paz. *Analog Computing Technique*.
http://courses.engr.illinois.edu/ece486/labs/lab1/analog_computer_manual.pdf.
- [156] R.W. Newcomb W.J.Kerwin, L.P.Huelsman. *State-Variable Synthesis for Insensitive Integrated Circuit Transfer Functions*. *IEEE Journal of Solid State Circuits*, SC-2:87–92, September 1967.
- [157] Adel S. Sedra Robert G. Sparkes. *Programmable Active Filters*. *IEEE Journal of Solid State Circuits*, pages 93–95, February 1973. The use of multipliers to control the cutoff frequency and bandwidth of a state-variable filter.
- [158] Yishay Nezer. *Active notch filters*. *Wireless World*, pages 307–311, July 1975. An excellent tutorial article on the design of active notch filters.
- [159] Ralph Glasgal. *Tuneable RC Null Networks*. *EEE*, pages 70–74, October 1969. Shows and explains various passive RC notch filter networks. A letter in the November issue (Bill Neeland) shows another useful tuneable notch filter design.

- [160] Pen-Min Lin Raymond A. DeCarlo. *Linear Circuit Analysis*. Oxford University Press, New York, 2001. Basic techniques of linear circuit analysis.
- [161] S. Butterworth. *On the Theory of Filter Amplifiers*. *Wireless Engineer*, 7:536–541, 1930. The Butterworth filter characteristic.
- [162] John Derbyshire. *Prime Obsession*. Plume, New York, N.Y., 2003. An entertaining account of the mathematics and history of the Reimann Hypothesis, which relates to the distribution of prime numbers.
- [163] M.Tepian. *Fundamentals of EEG Measurement. Measurement Science Review*, 2, 2002 .
<http://www.measurement.sk/2002/S2/Tepian.pdf>.
- [164] W.E.Thompson. *Delay Networks Having Maximally Flat Frequency Characteristics. Proceedings of the IEE, Part 3*, 96:487–490, November 1949. Presentation of the filter characteristic based on Bessel polynomials.
- [165] David Johnson. *Introduction to Filter Theory*. Prentice-Hall, Englewood Cliffs, New Jersey, 1976.
- [166] Bernard M.Oliver and John M.Cage. *Electronic Measurements and Instrumentation*. McGraw-Hill, New York, 1971. A classic reference for electronic measurement techniques and equipment.
- [167] Stefan Niewiadomski. *Filter Handbook: A practical design guide*. CRC Press, Boca Raton, Florida, USA, 1989. Much useful design information, especially on passive filters useable at radio frequencies. Includes filter characteristic tables in the Q- ω format.
- [168] Kerry Lakanette. *Switched-Capacitor Filter Handbook*. National Semiconductor Corporation, National Semiconductor Corporation, 2900 Semiconductor Drive, P.O.Box 58090, Santa Clara, California, 95052-8090, USA, 1985. Long out of print, but useful for it's collection of filter tables.
- [169] Richard C. Dorf. *Modern Control Systems*. Addison-Wesley, New York, N.Y., 1967. A standard text on control systems.
- [170] Brian Williams. *Pole-Zero Visualization Tool*, 2005 .
<http://web.mit.edu/6.302/www/pz/>.
- [171] Michael Rettinger. *Acoustical Design and Noise Control*. Chemical Publishing Co., New York, N.Y., 1977. A definitive reference on acoustical noise control techniques.
- [172] *Application Note 100: Acoustics Handbook*. Hewlett-Packard GMBH, 703 Boblingen, Herrenberger Str. 110, West Germany, 1968. A handbook of acoustical measurement techniques.
- [173] Shawn Carlson. *Detecting Extraterrestrial Gravity*. *Scientific American*, pages 94–96, January 2000. A very clever design for a gravimeter-seismometer. Note - some connections in the schematic may be incorrect.
- [174] Mohammed S. Ghausi. *Analog Active Filters*. *IEEE Transactions on Circuits and Systems*, CAS-31:13–31, January 1984. A survey article on the historical development of active filters.
- [175] Alexander Belousov. *Negatron yeilds real natural frequency*. *EDN*, page 96, August 19 1993.
- [176] Robert A. Pease. *Improve circuit performance with a 1-op-amp current pump*. In Ian Hickman, editor, *Electronic Circuits, Systems and Standards*. Butterworth-Heinemann, 1991. A detailed article on applying the Howland current source.

- [177] Tim Henry. *Analysis and Design of the Op Amp Current Source*. Motorola Applications Engineering, AN-587, 1972.
- [178] Application Note AN-435: *Designing with the LMC835 Digital-Controlled Graphic Equalizer*. In *Linear Applications Handbook*. National Semiconductor, 1994. The Gyrator is used to simulate an inductance in an audio graphic-equalizer.
- [179] Thomas H. Lynch. *The right gyrator trims the fat off active filters*. *Electronics*, pages 115–118, July 21 1977. LC filter design with gyrators and FDNR's.
- [180] John Dent. *Negative resistance in AF filters*. *Electronics World + Wireless World*, pages 1203–1207, December 1989. High Q band-stop filter design using gyrators.
- [181] R.H.Troughton. *Gyrators in band-pass filter design*. *Electronics & Wireless World*, pages 1227–1229.
- [182] Ian Hickman. *Filters using negative resistance*. *Electronics World*, pages 217–221, March 1997. A very lucid introduction to the FDNR and its application.
- [183] Philips Semiconductors. AN98035: *Circulators and Isolators, unique passive devices*, March 23, 1998.
<http://www.glsle.com/files/downloads/circulator.pdf>.
- [184] P.E.Greenaway J.M.Rollett. *Directly Coupled Active Circulators*. *Electronic Letters*, 4, No 26:579–580, 27 December 1968.
- [185] Charles Wenzel. *Low Frequency Circulator/Isolator. Uses No Ferrite or Magnet*, 1991.
www.wenzel.com/pdffiles1/pdfs/RFDesign3.pdf.
- [186] Ian Hickman. *Analog circuits cookbook*, 2nd Ed. Newnes, Boston, 1999.
- [187] *Negative Impedance Converter*, 2008.
http://en.wikipedia.org/wiki/Negative_impedance_converter.
- [188] Russell Kincaid. *AC amplifier eliminates use of coupling capacitor*. *Electronic Design*, 26:54, December 20 1970. A clever capacitor-synthesis circuit.
- [189] Kevin Schmitt. *Phase-Shift Network Analysis and Optimization*. *QEX*, August 1994. Also on the WEB at
<http://fermi.la.asu.edu/w9cf>.
- [190] M.Hartley Jones. *Frequency Shifter for Howl Suppression*. *Wireless World*, pages 317–322, July 1973. An audio circuit based on the single-sideband phasing method, for reducing audio feedback in a sound reinforcement system. Interesting for the phase shift networks and the use of analog multipliers.
- [191] *Audio Feedback Eliminator*. *ETI Canada*, pages 38–43, May 1978.
- [192] Trevor Brook. *Frequency Shifter*. *Studio Sound*, pages 44–47, October 1974.
- [193] J.Donald Tillman. *Moog Patents*.
<http://www.till.com/articles/moog/patents.html>
Refers to US Patent: 3,800,088: *Apparatus For Producing Special Audio Effects Utilizing Phase Shift Techniques*.
- [194] *Barkhausen stability criterion*, 2008.
http://en.wikipedia.org/wiki/Barkhausen_stability_criterion.

- [195] Marc Thompson. *Review of Feedback Systems*. In Robert A. Pease, editor, *Linear Applications Handbook*. Newnes, 2008. The Routh Stability Criterion explained.
- [196] Alexander Belousov. *Negative impedance improves capacitive sensors*. EDN, March 30, 1995.
<http://www.edn.com/archives/1995/033095/07DI7.htm>
- [197] Alexander Belousov. *Improved frequency modulator uses Negatron*. EDN, July 11, 2002.
<http://www.edn.com/article/CA231573.html>
- [198] Ron Mancini. *Design of op amp sinewave oscillators*. Analog Applications Journal, pages 33–37, August 2000. An excellent examination of various op-amp sine wave oscillators.
- [199] *National Semiconductor Linear Applications Handbook*. National Semiconductor Corporation, National Semiconductor Corporation, 2900 Semiconductor Drive, P.O.Box 58090, Santa Clara, California, 95052-8090, USA, 1994.
- [200] John Vanderkooy and Cameron J. Koch. *Generate low-distortion sinewaves*. Electronic Design, pages 70–73, July 5 1973. A Wien bridge oscillator with high-power output, using a discrete component op-amp. Includes an analysis of the distortion contributed by incandescent lamp stabilization.
- [201] R.D.Crawford. *One-IC Audio Generator*. Radio-Electronics, pages 37,38, February 1971. A simple amplitude-stabilized Wein bridge oscillator.
- [202] Walter G. Jung. *IC Op-Amp Cookbook*. Howard W. Sams & Co., Indianapolis, 1974.
- [203] Tobey, Graeme & Huelsman. *Operational Amplifiers: Design and Applications*. McGraw-Hill Kogakusha, Tokyo, 1971.
- [204] *A Precision Waveform Generator and Voltage Controlled Oscillator*. Intersil Application Note, October 1972. Detailed information on the internals and applications of the 8038 integrated circuit function generator.
- [205] Michael Lampton. *A Wide Range Audio Sweep Oscillator*. Audio, pages 30–33, July 1971. Description of a function generator that can be swept over the audio range.
- [206] William E. Peterson. *Field effect transistor converts triangle to sines*. Electronics, pages 69,70, August 31 1970. An example of a JFET being used for triangle-sine wave shaping.
- [207] J.I.Smith. *Modern Operational Circuit Design*. Wiley, 1971.
- [208] *AN-263: Sine Wave Generation Techniques*. 1999. An excellent overview of oscillator and sine wave shaping techniques.
<http://www.national.com/an/AN/AN-263.pdf>.
- [209] *Operational Amplifier Databook, 1993 Edition*. National Semiconductor Corporation, National Semiconductor Corporation, 2900 Semiconductor Drive, P.O.Box 58090, Santa Clara, California, 95052-8090, USA, 1993.
- [210] Texas Instruments. *Datasheet, TLC07x Operational Amplifier*, 2000
<http://focus.ti.com/docs/prod/folders/print/tlc074.html>.
- [211] *LM118 Op Amp Sweeps 70V/μsec*. In Dan Sheingold, editor, *National Semiconductor Linear Applications Handbook*, chapter LB-17. National Semiconductor Corporation, 1994.

- [212] Melvin C. Sprinkle. *The Ultimate Noise. db*, pages 23–27, June 1969.
- [213] Gary Franklin and Troy Hatley. *Don't Eyeball Noise. Electronic Design*, 24:184–187, November 22 1973.
The tangential method of measuring noise RMS amplitude, using an oscilloscope.
- [214] C.D.Motchenbacher and J.A.Connely. *Low Noise Electronic System Design*. John Wiley and Sons, New York, 1993. Another classic reference in noise control.
- [215] Henry W. Ott. *Noise Reduction Techniques in Electronic Systems*. John Wiley and Sons, New York, 1976.
A classic reference in noise control.
- [216] G.B.Clayton. *Operational Amplifiers*. Butterworths, London, 1971. A general-purpose book on the op-amp. Some useful material on op-amp computing circuits.
- [217] Texas Instruments. *Noise Analysis in Operational Amplifier Circuits. Application Note SLVA043A*, November 22 1999. A general overview of op-amp noise.
- [218] D.Abbott, B.R.Davis,N.J.Philips,K. Eshraghian. *Simple Derivation of the Thermal Noise Formula Using Window-Limited Fourier Transforms and other Conundrums. IEEE Transactions on Education*, 39:1 – 13, February 1996.
- [219] Jiří Dostál. *Operational Amplifiers, Second Edition*. Butterworth-Heinemann, Stoneham, MA, 1993. An excellent general text on operational amplifiers.
- [220] C Gullett. *The Hidden Schematic. Common Ground (Intel Automotive Newsletter)*, 1, February 1993.
Possibly the first mention of the term *hidden schematic*.
- [221] T.D.Towers. *Hanging on Cs at High Frequencies. Wireless World*, pages 619–623, December 1966.
- [222] Peter D. Hiscocks. *Construction Techniques*, 2000.
<http://www.ee.ryerson.ca/~phiscock/papers/construction.pdf>.
- [223] Walt Kester. *Prototyping techniques help verify analog-circuit performance. EDN*, Feb 15 1996.
<http://www.edn.com/archives/1996/021596/04df3.htm>.
- [224] Bruce Carter. *How (Not) to Decouple High-Speed Operational Amplifiers. Texas Instruments Application Report SLOA069*, pages 1–13, September 2001.
- [225] Jim Williams. *AN-18, Power Gain Stages for Monolithic Amplifiers*. In James M. Williams, editor, *1990 Linear Applications Handbook*. Linear Technology, March 1986.
- [226] Gerald Graeme. *Phase compensation exetnd op amp stability and speed. EDN*, pages 181–192, September 16 1991. A very detailed presentation of various methods of op-amp compensation.
- [227] *Accuracy and Precision*.
http://en.wikipedia.org/wiki/Accuracy_and_precision.
- [228] Walt Kester. *MT-001 Tutorial: Taking the Mystery out of the Infamous Formula SNR= 6.02N + 1.76dB, and Why You Should Care*, 2009.
<http://www.analog.com/static/imported-files/tutorials/MT-001.pdf>.
- [229] *The Integrating A/D Converter (ICL7135). Intersil Application Note AN017*, pages 1–5, 1999. An examination of the dual-slope A/D converter and its errors.

- [230] Dave Pryce. *Audio DACs push CD players to higher performance.* *EDN*, pages 112–121, December 7 1989. A survey article on audio D/A conversion, just before the sigma-delta converter revolution.
- [231] Analog Devices. *Sigma-Delta ADCs and DACs. Application Note AN-283.* A very clear tutorial on Sigma-Delta oversampling conversion.
- [232] Gary J. Summers. *Miles/gallon measurement made easy with CMOS rate multipliers.* *EDN*, pages 61–63, January 20 1980.
- [233] Max W. Hauser. *Principles of Oversampling A/D Conversion.* *Journal of the Audio Engineering Society*, 39:3–26, January/February 1991. A very comprehensive tutorial on oversampling A/D conversion.
- [234] Analog Devices. *Using Sigma-Delta Converters-Parts 1 and 2. Application Notes AN-388, AN-389.* Useful supplementary information on the oversampling A/D converter.
- [235] Ralph Weir. *Sigma-delta conversion.* *Electronics World+Wireless World*, pages 1–32, December 1989. A tutorial on oversampling A/D ($\Sigma - \Delta$) conversion.
- [236] John Bliss. *Applications of PhotoTransistors in Electro-Optic Systems.* *Motorola Application Note*, AN-508, 1970.
- [237] *CdS Light Sensor.*
www.extremenxt.com/lego.htm
- [238] *CdS Photoresistor Data.*
<http://ronja.twibright.com/datasheets/cds-resistor-pgm.pdf>.
- [239] *Liteon Optoelectronics Databook.* Taiwan Liton Electronic Co. Ltd, Taipei, Taiwan, ROC, 1994-95.
- [240] *QT Optoelectronics Product Data Book.* QT Optoelectronics, Sunnyvale, California, 1995/96.
- [241] *The Cray-1*, 2008.
<http://en.wikipedia.org/wiki/Cray-1>.
- [242] *AN16: Selecting Commercial Power Transistor Heat Sinks.* In *The Semiconductor Data Book*. Motorola Semiconductor Products, 1969.
- [243] Douglas Self. *Audio power amplifier design handbook.* Newnes, 3rd edition, 2002.
- [244] Douglas Self. *Self on audio.* Newnes, 2000.
- [245] Douglas Self. *Various notes on audio system design*, 2004.
<http://www.douglas-self.com/>.
- [246] William W. Troutman William F. Brinkman, Douglas E Haggan. *A History of the Invention of the Transistor and Where It Will Lead Us.* *IEEE Journal of Solid State Circuits*, pages 1858–1864, December 1997.
- [247] James M. Early. *Effects of Space-Charge Layer Widening in Junction Transistors.* *Proceedings of the IRE*, pages 1401–1406, November 1952. The original paper describing what is now known as the Early effect.
- [248] James M. Early. *Out to Murray Hill to Play: An Early History of Transistors.* *IEEE Transactions on Electron Devices*, 48:2468–2472, November 2001. Historical notes by the discoverer of the Early effect.

- [249] Bryan L. Hart. *Modelling the Early Effect in Bipolar Transistors*. *IEEE Journal of Solid State Circuits*, SC-18:139–140, February 1983. Establishes that the effective value of output resistance in the BJT common-base configuration is increased by a factor of beta over the common-emitter configuration.
- [250] AN8: *RF Small-Signal Design using 2-Port Parameters*. In *The Semiconductor Data Book, 4th Edition*. Motorola Semiconductor, 1969. Appendix A of this application note relates transistor h-parameters in the common emitter, common base and common collector (emitter follower) configurations.
- [251] Barrie Gilbert. *Bipolar Current Mirrors*. In F.J.Lidgey & D.G.Haigh C.Toumazou, editor, *Analog IC Design: The Current Mode Approach*, chapter 6. Peter Peregrinus, 1990.
- [252] Paul R. Gray, Paul J. Hurst, Stephen H. Lewis, Robert G. Meyer. *Analysis and Design of Analog Electronic Circuits*. John Wiley and Sons, New York, 2001.
- [253] Glenn B. DeBella. *Stability of Capacitively-Loaded Emitter Followers*. *Hewlett-Packard Journal*, pages 16,17, April 1966.
- [254] Saverio Trotta et al. *An 84 GHz Bandwidth and 20dB Gain Broadband Amplifier in SiGe Bipolar Technology*. *IEEE Journal of Solid State Circuits*, 42:2099–2106, October 2007.
- [255] Paul J. Beneteau. *Stable Wideband Emitter Followers*. *Solid State Design*, pages 29–33, February 1962.
- [256] Gordon J. Deboo and Clifford N. Burrous. *Integrated Circuits and Semiconductor Devices: Theory and Application*. McGraw-Hill, New York, 1971. Appendix B contains a clear description of the H-Parameter and Hybrid- π transistor models.
- [257] Evor L. Hines. *Course Notes, ES261, School of Engineering, University of Warwick*, 2000 .
www.eng.warwick.ac.uk/staff/elh/elh_es261_notes/notes2_m_2000.htm
- [258] Paul E. Gray and Campbell L. Searle. *Electronic Principles: Physics, Models and Circuits*. Wiley, New York, 1969. Section 12 contains a comprehensive presentation on the determination of BJT model parameters.
- [259] Martin Linder. *DC Parameter Extraction and Modelling of Bipolar Transistors*, 2001. Doctoral Dissertation, KTH, Royal Institute of Technology, Department of Microelectronics and Information Technology, Sweden.
- [260] Robert J. Widlar. *An Exact Expression for the Thermal Variation of the Emitter Base Voltage of Bi-Polar Transistors*. *IEEE Proceedings*, pages 96,97, January 1967.
- [261] Yannis P. Tsividis. *Accurate Analysis of Temperature Effects in Ic-Vbe Characteristics with Application to Bandgap Reference Sources*. *IEEE Journal of Solid State Circuits*, Vol SC-15, No. 6:1076–1084, December 1980.
- [262] AN460: *LM34/LM35 Precision Monolithic Temperature Sensors*. In *Linear Applications Handbook*. National Semiconductor, 1994.
- [263] Ken Martin David A. Johns. *Analog Integrated Circuit Design*. Wiley, 1997. An introductory level text on analog IC design.
- [264] Douglas Self. *Audio Power Amplifier Design Handbook, 3rd Edition*. Newnes, Oxford, UK, 2003. A definitive textbook on audio power amplifier design.

- [265] Sidney Darlington. *Semiconductor Signal Translating Device*. US Patent 2,663.806, May 1952. Invention of the Darlington pair.
- [266] Ben Duncan. *High Performance Audio Power Amplifiers*. Newnes, Oxford, UK, 1996. A comprehensive description of audio power amplifier design technique. This reference and [264] are a wealth of information on distortion mechanisms, power output stages, and the topology of amplifiers.
- [267] Robert J. Widlar. *Some Circuit Design Techniques for Linear Integrated Circuits*. IEEE Transactions on Circuit Theory, CT-12, No 4:585 to 590, December 1965. The current mirror and other techniques.
- [268] Hans R. Camenzind. *700 Series Bipolar Array Design Manual*. pages 5–1 to 5–76, 2003. A collection of integrated circuit building blocks, including current mirrors
www.arraydesign.com/_count/countdown.cgi?7MANUAL.PDF.
- [269] David Clayden. *The 'Alan Blumlein' Home Page*. 2005.
www.alanturing.net/turing_archive/pages/blumlein/#3.
- [270] *LM124/224/324/2902 Low Power Quad Op-Amp*. In *National Semiconductor Operational Amplifier Handbook*. National Semiconductor Corporation, 1993.
- [271] D.E. Hooper E.M. Cherry. *The design of wide-band transistor feedback amplifiers*. Australian Proceedings of the IEE, 110:375–388, February 1963. A description of the principles and design technique for series-shunt feedback amplifiers, especially to maximize bandwidth.
- [272] S. Sytstra. *Application Note 319: A vertical deflection amplifier for 150 MHz oscilloscopes*. In *Philips Application Information*. Philips Electronic Components and Materials Division, 1967. Detailed description of the design of a high-bandwidth oscilloscope amplifier.
- [273] Bob Orwiler. *Vertical Amplifier Concepts*. Tektronix, Beaverton, Oregon, 1969. A rather dated but nonetheless interesting description of oscilloscope vertical amplifier circuits.
- [274] Walter G. Jung. *Op Amp Applications Handbook*. Newnes, 2005. An excellent, comprehensive text on operational amplifiers and their applications. Also contains a detailed history of op-amp technology. Available as a .pdf document at: www.analog.com/library/analogDialogue/archives/39-05/op_amp_applications_handbook.html.
- [275] Holger Muller. *Fast high-voltage amplifiers for driving electro-optic modulators*. Review of Scientific Instruments, 76:084701, August 2005. Discrete component BJT (and electron tube) amplifier designs.
- [276] M. S. Colclough. *A fast high-voltage amplifier for driving piezoelectric positioners*. Review of Scientific Instruments, 71:4323, November 2000. A high-speed, high-voltage amplifier using an IC op-amp and discrete transistors.
- [277] Douglas Self. *Various notes on audio system design*, 2004.
<http://douglas-self.com/amps/closed.htm>.
- [278] Robert J. Widlar. *Application Note AN-29, IC Op Amp Beats FETs on Input Current*. In *Linear Applications Handbook*. National Semiconductor, 1994. Description of the National Semiconductor LM-108 Op-Amp, which uses superbeta transistors to obtain very low bias current and offset voltage.
www.national.com/an/AN/AN-29.pdf.
- [279] National Semiconductor. *LM723/LM723C Voltage Regulator*, June 1999.
<http://www.national.com/ds/LM/LM723.pdf>.

- [280] *Duality (electrical circuits)*, 2008.
http://en.wikipedia.org/wiki/Dual_%28electronics%29.
- [281] Leonce J. Sevin. *Field Effect Transistors*. Texas Instruments Electronics Series, McGraw-Hill, New York, 1965. An early but useful text on junction field effect transistors.
- [282] James Sherwin. *Program multichannel audio gain*. *Electronic Design*, 3:68–72, February 1 1975.
- [283] Arthur D. Delagrange. *FETs in RC network tune active filter*. *Electronics*, page 76, December 7 1970.
- [284] Fairchild Semiconductor. *H11F1M, H11F2M, H11F3M Photo FET Optocouplers*, May, 2007.
<http://www.fairchildsemi.com/ds/H1%2FH11F1M.pdf>.
- [285] Perkin-Elmer. *Vactrol Analog Opto-Isolator*, 2001.
<http://www.datasheetcatalog.org/datasheet/perkinelmer/VT500.pdf>.
- [286] Erik Morgan. *Amplifier Instability*. *Electronics World*, pages 311,312, April 1998.
- [287] Tony van Roon Ray Marston. *Transistor Tutorial: Part 10, Metal Oxide Semiconductor Field Effect Transistors*.
<http://www.uoguelph.ca/~antoon/tutorial/xtor/xtor10/xtor10.html>.
- [288] Application Note AN-948
Linear Power Amplifier Using Complementary HEXFET Power MOSFETs. Undated.
<http://www.irf.com/technical-info/appnotes/an-948.pdf>.
- [289] Rod Elliott. *High Power, High Fidelity MOSFET Power Amplifier*. 2004.
<http://sound.westhost.com/project101.htm>
- [290] Mitch Hodges and Rod Elliott. *Using HEXFETs in High Fidelity Audio*. 2006.
<http://sound.westhost.com/articles/hexfet.htm>
- [291] Henry Ott. *Measuring Common-Mode Currents on Cables*. 2000.
<http://www.hottconsultants.com/techtips/tips-cm.html>.
- [292] Daryl Gerke Chris Banyai. *EMI Design Techniques for Microcontrollers in Automotive Applications*. Intel Application Note AP-711, pages 1–28, February 1996. Electromagnetic Interference requirements for automotive applications are very stringent. This very useful application note provides guidelines and a case history of interference reduction. Guidance on the design and segmentation of ground planes for minimizing EMI. The authors coined the term 'hidden schematic.'
ece-www.colorado.edu/~mcclure1/iap711.pdf.
- [293] Bruce Carsten. *Sniffer probe locates sources of EMI*. *EDN*, pages 155–162, June 4 1998. Construction details and applications for a probe to sense magnetic field interference.
- [294] Douglas C. Smith. *Signal and Noise Measurement Techniques Using Magnetic Field Probes*. IEEE International Symposium on Electromagnetic Compatibility, Symposium Record, pages 559–563, 1999. Also available at <http://www.emcesd.com/tt080699.htm>.
- [295] Douglas C. Smith. *Current probes, more useful than you think*. IEEE International Symposium on Electromagnetic Compatibility, Symposium Record, pages 284–289, 1998. A variety of applications for magnetic field probes are described. Also available at
<http://www.emcesd.com/pdf/iprobe98.pdf>.

- [296] Keith Armstrong. *Advanced PCB design and layout for EMC. Parts 1-6. EMC & Compliance Journal*, Issues 51-56, 2004. <http://www.compliance-club.com/archive/index.asp> A set of six articles giving extensive and detailed information on the design of electronic systems to control EMI and signal integrity.
- [297] Isador Straus. *Shielding Review: Tips and Tricks. Conformity*, pages 26–32, April 2004. An overview of shielding techniques.
<http://wenku.baidu.com/view/6308a03383c4bb4cf7ecd1a2.html>.
- [298] Editor H.P. Westman. *Reference Data for Radio Engineers*. Howard W. Sams & Co., Inc., Indianapolis, 1974.
- [299] Analog Devices. *Mixed Signal Circuit Techniques*. Application Note AN-280, 2005.
<http://www.analog.com> A very comprehensive discussion of noise, grounding, shielding and other practical design issues.
- [300] James C. Huhta and John G. Webster. *60Hz Interference in Electrocardiography*. *IEEE Transactions on Biomedical Engineering*, BME-20:91–101, March 1973. A clear exposition on the use of differential amplifier techniques in a biomedical application.
- [301] Alan Rich. *Shielding and guarding: How to exclude interference-type noise*. *Analog Dialog*, 17-1:8–13, 1983. Practical information on shielding techniques. Also published as Analog Devices Application Note AN-347.
- [302] Roland WW. Gubisch. *Inside FCC Part 15 and Canada's Corresponding Standards. Compliance Engineering*, 1999. A tutorial introduction on FCC Part 15 rules on Electromagnetic Compliance.
<http://www.ce-mag.com/99ARG/Gubish31.html>.
- [303] Federal Communications Commission. *Part 15 – Radio Frequency Devices*. pages 1–156, 2005. The rules of the FCC in the United State of America, specifying measurement methods and compliance rules for electromagnetic radiation.
<http://www.fcc.gov/oet/info/rules/>.
- [304] Dag Björklöf. *Guidelines for Application of the EMC Directive. Compliance Engineering*, 1999. Electromagnetic compliance rules in Europe.
<http://www.autoemc.net/Standards/StandardsMain.htm>.
- [305] Gilbert P. Condon. *Baluns cut ground noise*. *Electronic Design*, pages 62–63, July 22 1971. The use of common-mode chokes to reduce ground path noise.
- [306] Thurman J. (Bill) Ritenour. *Stopping Noise Before It Starts*. *Medical Electronics Manufacturing*, Fall, 1998. Common-mode radiation related to 'board bounce', and mechanisms for minimizing common-mode noise.
<http://www.devicelink.com/mem/archive/98/10/007.html>.
- [307] Belden. *Sending All the Right Signals*.
<http://www.belden.com/pdfs/Cable101/Shielding.pdf>.
- [308] D.Walton. *PCB layout for high speed Schottky ttl*. *Wireless World*, pages 63–65, February 1978. Design of printed circuit boards for digital circuits.

- [309] Motorola Semiconductor. *Reducing A/D Errors in Microcontroller Applications*. Motorola Application Note AN1058, 2001. Analysis and case history regarding minimizing noise in a microprocessor A/D application.
www.freescale.com/files/microcontrollers/doc/app_note/AN1058.pdf.
- [310] Walt Kester. *Ask The Applications Engineer-12: Grounding (Again)*. Analog Dialog, 26-2:26,27, 1992. Proper grounding method for A/D converters.
- [311] Howard Johnson. *Multiple ADC Grounding*. EDN, page 30, February 1 2001.
- [312] Nicholas C. Gray. *Attack The Noise Gremlins that Plague High-Speed ADCs*. Electronic Design, pages 107–110, December 1999. Guidance on minimizing noise problems in A/D converters.
www.electcdesign.com/Articles/ArticleID/7487/7487.html.
- [313] Henry W. Ott. *Partitioning and Layout of a Mixed Signal PCB*. Printed Circuit Design, pages 62–63, June 2001. Minimizing digital noise contamination of analog circuitry. Advantages and disadvantages of split ground planes. <http://hottconsultants.com/papers.html>.
- [314] F.J.Lidgey & D.G.Haigh C.Toumazou. *Analog IC Design: The Current Mode Approach*. Peter Peregrinus, 1990.
- [315] Randal L. Geiger and Edgar Sanchez-Sinencio. *Active Filter Design using Operational Transconductance Amplifiers: A Tutorial*. IEEE Circuits and Devices Magazine, 1:20–32, March 1985.
- [316] Henrique S. Malvar. *OTAs and op-amp form voltage-controlled equalizer*. Electronics Magazine, page 181, May 19 1982.
- [317] Iqbal A. Khan and Muslim T. Ahmed. *Wide-Range Electronically Tunable Multifunction OTA-C Filter for Instrumentation Applications*. IEEE Transactions on Instrumentation and Measurement, IM-36:13–17, March 1987.
- [318] F.Gohh A.S.Sedra, G.W.Roberts. *The current conveyor: history, progress and new results*. IEE Proceedings, 137, Pt G, No 2:446–454, April 1990.
- [319] L. Grigorescu. *Amplifiers Built With Current Conveyors*. Rom. Journ. Phys., Vol. 53, Nos. 1&2, pages 109–113, 2008. http://www.nipne.ro/rjp/2008_53_1-2/0109_0114.pdf.
- [320] Abhirup Lahiri. *Sinusoid generator uses dual-output current-controlled conveyors*. EDN, pages 54,56, January 22, 2009. <http://www.edn.com/contents/images/6629476.pdf>.
- [321] Elantec. *Dielectric Isolation*.
<http://www.nalanda.nitc.ac.in/industry/appnotes/Elantec/d40940.pdf>.
- [322] Sergio Franco. *Current-feedback amplifiers benefit high-speed designs*. EDN, pages 161–170, Jan 1989. Tutorial article on current feedback amplifiers.
- [323] Ian Hickman. *CFBOs: delivering speed at any gain?* Electronics World+Wireless World, pages 78–80, Jan 1993. Test measurements on the LM6181 current feedback amplifier.
- [324] Debbie Brandenburg. *OA-30: Current vs Voltage Feedback Amplifiers*, 1998.
<http://www.national.com/an/oa/oa-30.pdf>.
- [325] David Potson. *OA-07: Current Feedback Op Amp Applications Circuit Guide*, 1988.
<http://www.national.com/an/oa/oa-07.pdf>.

- [326] *Application Note AN-72, The LM3900: A New Current-Differencing Quad of +/- Input Amplifiers.* In *Linear Applications Handbook*. National Semiconductor, 1994. Description of the National Semiconductor LM3900 Norton Amplifier, and its applications.
www.national.com/an/AN/AN-72.pdf.
- [327] Thomas Frederiksen Ronald Russell. *Automotive and Industrial Electronic Building Blocks. IEEE Journal of Solid State Circuits*, SC-7:446–454, December 1972.
- [328] *LM2900/LM3900/LM3301 Quad Amplifiers.* In *National Semiconductor Operational Amplifiers Databook*. National Semiconductor Corporation, 1993.
- [329] Bruce Carter. *Op Amp and Comparators – Don't Confuse Them! Texas Instruments Application Report SLOA067*, pages 1–12, September 2001.
- [330] Peter Hiscocks. *Oscilloscope Development, 1943-57*. April 2009.
<http://www.syscompdesign.com/AppNotes/scope-history.pdf>.
- [331] Jim Williams. *AN47: High Speed Amplifier Techniques*. 1991.
- [332] Jim Karki. *Fully differential amplifiers*. Texas Instruments, 2000.
www-s.ti.com/sc/techlit/slyt165.pdf.
- [333] Jim Karki. *Analysis of fully differential amplifiers*. Texas Instruments, 2000.
focus.ti.com/lit/an/slyt157/slyt157.pdf.
- [334] *Very Low Power, Negative Rail Input, Rail-to-Rail Output, Fully Differential Amp (Rev. B)*. Texas Instruments, 2008.
<http://www.ti.com/lit/gpn/ths4521>.
- [335] Jim Williams. *Application Note 87: 0.05µV/°C Chopped Amplifier Requires Only 5µA Supply Current*. In <http://cds.linear.com/docs/Application%20Note/an87f.pdf>. Linear Technology, 1992.
- [336] Jim Williams. *Composite Amplifiers*. In James M. Williams, editor, *1990 Linear Applications Handbook*, chapter 21. Linear Technology, March 1986.
- [337] S.Raymond Kushnir P.Kent Hardage and Thomas J. Zamborelli. *Optimizing the Design of a High-Performance Oscilloscope. Hewlett-Packard Journal*. A high-speed, low drift composite amplifier design for the front-end of an oscilloscope.
- [338] Robert.A.Pease. *Linear Brief LB-44, Get More Power Out of Dual or Quad Op Amps*. In *Linear Applications Handbook*. National Semiconductor, 1994. How to parallel op-amps to increase the output power.
- [339] G.C.Hammond. *Chopper stabilised amplifier for dc voltmeter*. *Wireless World*, page 91, June 1979. A very simple, fully differential chopper-stabilised amplifier design.
- [340] Bruce Carter. *A Differential Op-Amp Collection*. *Texas Instruments Application Report SLOA064*, July, 2001.
focus.ti.com/jp/lit/an/sloa064/sloa064.pdf.
- [341] Burr-Brown products from Texas Instruments. *INA117 High Common-Mode Voltage Difference Amplifier*, 16-Apr-2009. <http://focus.ti.com/lit/ds/symlink/ina117.pdf>.

- [342] Wikipedia. *Heterodyne*.
<http://en.wikipedia.org/wiki/Heterodyne>.
- [343] Wikipedia. *Frequency mixer*.
http://en.wikipedia.org/wiki/Frequency_mixer.
- [344] Harold Anderson, Peter Hiscocks. *Switching multiplier is accurate at low frequencies*. *Electronics*, February 2, 1978, pp 114-117. A low-cost pulse width-height multiplier with 0.5% accuracy.
- [345] Barrie Gilbert. *A Precise Four-Quadrant Multiplier with Subnanosecond Response*. *IEEE Journal of Solid-State Circuits*, SC-3, No 4:365–373, December 1968.
- [346] Mark E. Connelly. *An analog photoresistive multiplier*. *Simulation*, November 1966, pp 247-259.
- [347] Jim Williams. *Analog Circuit Design: Art, Science and Personalities*. Butterworth-Heinemann, Newton, MA, 1991.
- [348] Jim Williams. *The Art and Science of Analog Circuit Design*. Butterworth-Heinemann, Newton, MA, 1995.
- [349] Andrew O'Malley. *All Your Base Are Autonomous*. December 2002. An autotomous analog sumo robot.
<http://www.aomalley.org/robots/basey/>.
- [350] Union Carbide F.B.Pipal. *Eveready Battery Engineering Data*. New York, 1976.
- [351] J.H.Batchelor A.R.Upson. *Synchro Engineering Handbook*. Muirhead & Co Limited, New York, 1965. An introduction to rotating components such as the resolver and *synchro*, useful in transmitting and receiving a rotation control signal.
- [352] Wikipedia. *Synchro*.
<http://en.wikipedia.org/wiki/Synchro>.
- [353] Wikipedia. *Resolver*.
[http://en.wikipedia.org/wiki/Resolver_\(electrical\)](http://en.wikipedia.org/wiki/Resolver_(electrical)).
- [354] Mike Powell. *Mike's Flight Deck: LVDT's*. Construction of an LVDT plus circuit of an interface.
<http://www.mikesflightdeck.com/lvdt.htm>.
- [355] F.H.Lange. *Correlation Techniques*. Van Nostrand Reinhold Company, New York, 1967. Although the details of the electronic circuits implementation details are dated, this is a comprehensive introduction to the concept of correlation.
- [356] Robert Cushman. *Make the most of noise: Correlate it, Parts 1 and 2*. *EDN Magazine*, Part 1: March 1, pp 29-35. Part 2, April 15, pp 39-45 1971. A readable introduction to correlation techniques, including the construction of a simple analog correlator.
- [357] Gordon T. Roberts Richard L. Rex. *Correlation, Signal Averaging and Probability Analysis*. *Hewlett-Packard Journal*, 21, No 3:2–20, November 1971. A tutorial article on correlation which includes a number of example applications.
- [358] P.D.Schomer. *Measurement of Sound Transmission Loss by Combining Correlation and Fourier Techniques*. *Journal of the Acoustical Society of America*, 51, No. 4, Part 1:1127–1141, 1972. Correlation applied to the measurement of sound transmission through panels.

- [359] J.T.Broch. *On the Applicability and Limitations of the Cross-Correlation and Cross-Spectral Density Techniques.* *Bruel & Kjaer Technical Review*, No 4:3–27, 1970. Some basic limitations of the cross-correlation technique.
- [360] Frank Ogden. *Electronic Altimeter.* *Wireless World*, pages 37–41, May 1985.
- [361] J. Cragg C. Swartz, C. Derrington. *Temperature Compensation Methods for the Motorola X-ducer Pressure Sensor Element.* *Motorola Application Note AN-840*, 1982
www.freescale.com/files/sensors/doc/app_note/AN840.pdf.
- [362] John Gasking. *Resolver to Digital Conversion – A simple and Cost Effective Alternative to Optical Shaft Encoders.* Analog Devices Application Note AN-263.
<http://www.analog.com/en/rList/0,2881,760%255F791%255F874,00.html>.
- [363] Peter Hiscocks. *Schematic Symbols for Xfig.*
<http://www.ee.ryerson.ca/~phiscock/>.

Index

A

- AC adaptor, 416, 435
- AC power distribution, 416
 - ground, 417
 - ground fault interrupter (GFI), 418
 - neutral, 417
 - safety, 417
- AC, alternating current, 154
- accuracy and precision, 732
- active devices, 266
- active filter
 - 2nd order transient response, 567
 - all pass, 613
 - applications, 513
 - Bessel lowpass characteristic, 564
 - biquad, 573, 590
 - Bronzite circuit, 537
 - Butterworth highpass characteristic, 557
 - Butterworth lowpass characteristic, 554
 - centre frequency, 514
 - Chebyschev lowpass characteristic, 557
 - cutoff frequency, 514
 - design method overview, 572
 - FDNR, frequency dependent negative resistance, 605
 - filter order, 515
 - filter shape, 515
 - first order active filters, summary table, 520
 - first order cascade, 524
 - function types, 514
 - higher order, 553
 - higher order transient response, 569
 - implementation, 517
 - introduction, 513
 - multiple-feedback-bandpass, 573
 - notch, 550
 - pass band, 514
 - roadmap, 517
 - Sallen-Key, 517
 - Sallen-Key highpass, 545
 - Sallen-Key lowpass, 539
 - second order bandpass, 529
 - second order highpass, 531
 - second order lowpass, 524
 - second order notch, 533
 - second order summary, 536
 - sensitivity, 516
 - simple bandpass, 521
 - state variable, 550, 573
 - stop band, 514
 - tables, 570
 - transition band, 514
- adc, 749
 - aliasing, 752, 761
 - dual slope, 758
 - feedback type, 754
 - quantization, 749
 - quantization noise, 749
 - sigma-delta oversampling, 761
 - signal to noise ratio, 750
 - single slope, 756
 - successive approximation converter, 755
 - up counting type, 754
 - up-down counter converter, 754
 - voltage to frequency, 764
- ADC, Analog to Digital Conversion, *see* adc adder
 - non-inverting, 337
 - op-amp, 336
 - passive, 137
- aliasing, 570, 752
- all pass network, 613
 - analysis, 613
 - Bode plots, 614
- ammeter
 - moving coil, 70
- amp-turns, 101
- amp-turns per meter, 101
- amplifier, 266
 - ac coupled, non-inverting, 370
 - ac-coupled, inverting, 364
 - autozero, 1083
 - bjt common emitter, 841, 847
 - bjt common-base, 886
 - buffer, 318
 - charge, 360
 - charge, design of, 362
 - chopper stabilized, 1085
 - classification, 387

- common gate, 991
- common source, 991
- common-collector, 870
- composite, 1088
- concept, 106
- configurations, 839
- current feedback, 1062
- differential, 1039
- fully differential, 1074
- instrumentation, 341, 1072
- inverting, 320
- j fet compound, 991
- j fet differential, 991
- lever diagram, 344
- mechanical, 107
- non-inverting, 288, 319
- Norton, 1067
- oscilloscope buffer, 991
- parameters, 383
- pneumatic, 108
- power output stage, 1004
- switched gain, 1011
- transconductance, 1069
- transimpedance, 346
- transresistance, 384, 385
- types, 387
- universal, 343
- voltage, 315
 - voltage controlled gain, 1070
- analog, 43
- analog computer, 1105
- analog switch, 1009
 - applications, 1010
 - logic circuits, 1012
- analog to digital converter, *see* adc
- analog variable, 732
- analogy, hydraulic, 50
- analysis, 44
- astable multivibrator, 631
- attenuator
 - compensated, 184
- autozero amplifier
 - concept, 1083
 - pros and cons, 1084
- B**
- bandwidth, 181
- Q factor, 182
- Barkhausen Criterion, 619, 622, 626, 631
- barometer
 - calibration, 1138
 - interface, 1129
 - interface circuit, 1133
 - reliability, 1136
 - resolution and dynamic range, 1131
- base-emitter voltage, V_{be} , 267
- battery
 - compared to capacitor, 87
 - tester, 84
- Bessel filter characteristic, 564
- Beyond the op-amp, 1060
- bias
 - direct, 861
 - bias current, 639
 - biquad, 573, 590
 - bjt
 - V_{be} and temperature, 822
 - β , 802
 - active device, 266
 - alpha, 806, 807, 810
 - base current bias, 866
 - base current control, 802
 - base current drive, 858
 - base voltage control, 805
 - base-emitter diode equation, 806
 - base-emitter voltage and temperature, 900
 - base-emitter voltage temperature coefficient, 902
 - bias, 851
 - bias example, 831
 - bias stability and beta, 833
 - bipolar junction transistor, 798, 799
 - cascode pair, 931
 - catalogue of common-emitter amplifiers, 866
 - collector resistance, 817
 - collector-base feedback, 861
 - common-base amplifier, 886
 - common-collector amplifier, 870
 - common-emitter amplifier, 841, 847, 851, 858, 861, 866
 - common-emitter equations, 869
 - current amplification, 802
 - current gain model, 803
 - current generator model, 801
 - current or voltage control, 808
 - datasheet, 907

- differential amplifier, 935
- differential pair, 935
- diode model, 799
- direct bias, 861
- discussion of models, 900
- distortion, 868
- dual supply, 857
- Early voltage, 820
- Ebers-Moll equation, 807
- emitter current control, 810
- emitter current sink, 867
- emitter follower, 870
- emitter incremental resistance r_e , 834
- emitter-follower model, 873
- emitter-follower voltage gain, 874
- fixed bias, 866
- h-parameter model, 895
- input coupling, 851
- inverted, 891
- leakage current, 825
- logic device, 276
- model, 894, 895, 897, 900
- nomenclature, 800
- output coupling, 853
- parameters, 894
- polarity, NPN and PNP, 267
- process constant, 895
- ramp generator, 808
- reflecting resistance, 815, 875
- saturation, 267
- saturation current I_s , 807, 895
- summary of emitter follower equations, 876
- summary, circuits and equations, 828
- switch, 274
- t-parameter model, 897
- tee-model, 837
- temperature and collector current, 904
- terminals, 800
- transconductance, 836, 984
- unbypassed emitter, 868
- variation in current gain, 805
- voltage controlled current, 805
- BJT, bipolar junction transistor, *see bjt*
- Bode plot, 300, 304
 - poles and zeros, 305
 - constant factor, 305
 - differentiator factor, 305
 - first order lag, 307
- first order lead, 307
- integrator factor, 305
- limitations, 300
- quadratic factor, 312
- why useful, 300
- Boltzmann's constant, 247, 670, 892
- bridge, 138
- Bronzite filter
 - highpass, 538
 - lowpass, 537
- buffer, 318
 - amplifier, JFET, 986
 - JFET, 986
 - op-amp, 318
 - stabilizing, 716
- buffer amplifier, 318
- buildout resistance, 724
 - enclosed, 725
- Butterworth filter characteristic, 554

C

- capacitance
 - equation, 86
 - multiplier, 595
- capacitor
 - basics, 85
 - bypass, 1034
 - ceramic, 508
 - compared to battery, 87
 - construction, 86
 - decoupling, 708
 - defective circuit, 90
 - demonstration, 85
 - dielectric absorption, 508
 - differential equation, 89
 - energy storage, 87, 88, 90
 - ESR, equivalent series resistance, 444
 - film dielectric, 511
 - high voltage, 94
 - hydraulic analogy, 86
 - hysteresis, 508
 - in time-domain, 88
 - inductance, 1036
 - integral equation, 89
 - polypropylene, 508
 - polystyrene, 508
 - resonance, 707, 1037

- rule of thumb, 89
- series and parallel, 92
- soakage, 508
- capacitors
 - in parallel, 92
 - in series, 92
- carrier coding, 1119
- cascode pair, 931
 - folded, 935
- catastrophic failure, 290
- CFA, current feedback amplifier, 1062
- charge
 - balancing-converter, 764
- charge amplifier, 360
- charge balancing, 764
- charge pump, 455
- charge sensor, 361
- chatter, 401
- Chebyschev filter characteristic, 557
- chopper amplifier, 1085
 - chopping switches, 1086
 - circuit, 1085
 - generalized, 1087
 - single ended, 1086
- circular mil, 482
- circulator, 607
- cloud detector, 1144
- CMRR, 942
- CMRR, common mode rejection ratio, 649
- collector follower, 879
- common-base amplifier, 886
 - bias, 887
 - current gain, 888
 - input resistance, 889
 - model, 887
 - output resistance, 889
 - summary of equations, 890
 - supply configuration, 887
 - voltage gain, 888
- common-mode rejection ratio, 942
- common-mode signal, 341, 1074
- common-mode voltage, 649
- comparator, 401, 1071
 - comparison with op-amp, 1071
 - op-amp, 317
 - zero crossing detector, 317
- compensation
 - buildout resistance, 724, 725
- by reduced noise gain, 713
- compensated attenuator, 726
- dominant pole, 706
- lag-lead network, 719, 720
- complete systems
 - exercises, 1140
- complex frequency plane, 305, 573
 - amplitude and phase response, 578
 - membrane model, 577
 - visualization, 577
- complex notation
 - addition and subtraction, 159
 - introduction, 158
 - multiplication and division, 160
 - polar notation, 160
 - rectangular notation, 159
- composite amplifier, 1088
- computer algebra
 - Maxima, 212
- condensor, 98
- conductance, 56
- connection
 - parallel, 67
 - series, 66
- constant current generator
 - floating load, 358
 - grounded load, 359
 - Howland circuit, 597
- conventional current flow, 51
- converter
 - voltage to frequency, 764
- converter, current to voltage, 345
- cooling
 - derating curve, 785
 - fan, 791
 - fan design, 792
 - Fourier law, 784
 - heat sink, 783
 - heat sink design, 786
 - killer problem, 783
 - thermal circuit, 784
 - thermal resistance, 785
- correlation, 1119, 1125
 - analog, 1125
 - overview, 1125
 - phase sensitive rectifier, 1128
 - signal processing, 1126
 - time delay, 1125

coulomb, 51
 crossover distortion, 871, 1006
 CTR, *see* current transfer ratio
 current, 53
 alternating, basics, 154
 amplification, 802
 conventional current flow, 51
 electron flow, 51
 example, 54
 FSD, full scale deflection, 68
 fusing, 482
 leakage, in bjt, 825
 limit, 432
 meter, 54
 precision measurement, 1083
 short-circuit, 73
 shunt, 71
 sink, discrete version, 915
 standard, 54
 current conveyor, 1060
 application, 1061
 internal circuit, 1061
 current divider, 133
 application, 133
 current feedback amplifier
 basic operation, 1063
 gain analysis, 1064
 loop gain, 1064
 maximum frequency, 1065
 references, 1066
 slew rate, 1066
 transresistance equivalent circuit, 1065
 current flow
 conventional, 51
 electron, 51
 current gain β , 266
 current generator
 Howland circuit, 597
 model of bjt, 801
 current mirror, 917
 degeneration, 922
 design example, 922
 discrete version, 919
 effect of mismatch, 920
 example, 819
 mirror degeneration, 928
 multiple output, 923
 Wilson mirror, 925, 929

current source, 915
 floating load, 358
 grounded load, 359
 ideal, 62
 non-ideal, 77
 current to voltage converter, 345
 current transfer ratio, 778
 cutoff frequency, 527

D

D element, 605
 d'Arsonval meter movement, 68
 dac, 732
 aliasing, 752
 current steering, 737
 glitches, 735
 inexpensive, for CMOS logic, 739
 introduction, 732
 monotonicity, 734
 multiplying, applications, 741
 multiplying, four quadrant, 741
 multiplying, gain control, 740
 multiplying, introduction, 739
 multiplying, single quadrant, 740
 multiplying, two quadrant, 740
 non-linearity, 734
 pulse width modulation, 744
 quantization, 749
 quantization noise, 749
 R-2R network, 736
 reference stability, 733
 resolution, 733
 settling time, 734
 signal to noise ratio, 750
 specifications, 733
 voltage switching, 737
 weighted resistor, 735

DAC, Digital to Analog Conversion, *see* dac
 damping, 178
 critically damped, 178
 heavily damped, 178
 lightly damped, 178
 overdamped, 178
 underdamped, 178
 damping coefficient d , 178, 570
 Damping factor, 312
 definition, 178

- damping factor, 710
- Darlington
- complementary, 913
 - differential, 947
 - follower, 912
 - pair, 432, 911
 - Sziklai pair, 913
- datasheet
- jfet, 993
 - dc motor, 1107
 - dead zone, 1006
 - decoupling, 707
 - degeneration, 916
 - demultiplexer, 1009
 - derating curve, 785
 - dielectric absorption, 508
 - dielectric isolation, 1062
 - differential amplifier, 935, 1039
 - bias and gain, 951
 - cascode, 947
 - darlington, 947
 - emitter degeneration, 945
 - enhanced, 943
 - gain, 939
 - input resistance, 942
 - inter-stage coupling, 948
 - single gain-setting resistor, 946
 - tanh function, 950
 - transfer-function-equations, 948
- differential pair, 935
- transfer characteristic, 936
- differential signal, 341, 1074
- differentiator, 351, 1020
- improved, 353
 - in frequency domain, 352
 - in time domain, 351
 - stability, 711
- digital to analog converter, *see* dac
- digital variable, 732
- diode, 246
- AND gate, 248
 - anode, cathode, 246
 - bridge, 251
 - clamp, 252
 - clipping circuit, 252
 - constant current, 976
 - equation, 246
 - free-wheeling, 270, 427
 - limiter, 252
 - OR gate, 247
 - photodiode, 774
 - precision, 495, 497
 - snubber, 254, 270, 427
 - types, 254
 - zener, 259
- distortion
- bjt, 838
 - crossover, 871
 - transconductance, 838
- distributor, 1009
- disturbance, 288
- documentation, 44
- Dostal, Jiri, 291
- double sideband, 616
- drift
- temperature, 1137
- duplex receptacle, 417

E

- Early
- effect, 820
 - voltage, 820
- earth, *see* ground
- Ebers-Moll equation, 807
- Ebers-Moll model, 894
- effective value, 483
- pulse waveform, 486
 - triangle wave, 485
- electret, 360
- electroencephalography, 559
- electron
- charge, 55
 - conduction, 54
 - mass, 55
 - velocity, 54
 - velocity in vacuum, 55
- electron charge, 247
- electron flow, 51
- electronic charge, 51
- electronic system, 1105
- EMI, Electromagnetic Interference, 1013
- emitter-follower, 870
- and negative feedback, 877
 - bias, 871
 - configurations, 872

current gain, 875
 design example, 880
 design summary, 885
 input resistance, 875
 model, 873
 oscillation, 875
 output resistance, 875
 summary of equations, 876
 voltage gain, 874
 energy storage
 capacitor vs battery, 87
 in capacitor, 87, 88, 90
 in inductor, 100
 equation solver, 1105
 equivalent circuit
 bjt cascode, 931
 bjt differential pair, 939
 bjt emitter follower, 877
 bjt output resistance, 927
 bjt ramp generator, 819
 bjt reflecting resistance, 816
 bjt series-shunt feedback, 956
 current feedback amplifier, 1063
 current to voltage converter, 345
 Thevenin, 119
 voltage divider design, 126
 zener diode regulator, 260
 ESR, equivalent series resistance, 444
 Euler's Formula, 160
 European Style Notation, 61

F

fan cooling, 791
 Faraday's law, 96
 Faraday, Michael, 95
 FDNR, frequency dependent negative resistance, 605
 feedback
 bjt collector-base, 861
 collector-base, 866
 emitter-follower as negative feedback, 877
 negative, 284
 feedback factor, 694
 filter, 513
 active, 513
 anti-alias, 753
 Bessel lowpass characteristic, 564
 brick wall, 673

Butterworth lowpass characteristic, 554
 Chebyschev lowpass characteristic, 557
 first order passive, summary table, 518
 first order transient response, 567
 frequency normalization, 574
 higher order, 553
 highpass, 545
 lowpass, 539
 notch, 550, 551
 plotting high-order characteristic, 565
 pulse width modulation, 744
 second order bandpass RLC, 529
 second order highpass RLC, 531
 second order notch RLC, 533
 second order summary, 536
 second-order RLC lowpass, 525
 state variable, 546
 summary of common pole-zero configurations, 582
 tables, 570
 transient response, 567
 tuneable notch, 553
 twin tee, 551
 first order active filters, summary table, 520
 first order passive filter, summary table, 518
 flushout resistor, 912
 flux
 density, 101
 magnetic, 101
 folded cascode pair, 935
 follower
 BJT emitter, 870
 instability, 727
 JFET source, 986
 op-amp, 318
 force
 magnetomotive magnetizing, 101
 Franklin, Benjamin, 50
 frequency dependent negative resistance, 605
 application, 605
 circuit, 606
 frequency response
 introduction, 691
 large signal, 658
 maximum usable frequency, 696
 op-amp gain bandwidth, 692
 peaking, 710
 quadratic response, 701

single pole, 698
 fully differential amplifier, 1074
 applications, 1075
 common-mode output offset, 1081
 example design, 1076
 internal circuit, 1076
 lever diagram, 1080
 rules of behaviour, 1077
 single-ended to differential converter, 1080
 function generator, 633
 fusing current, 482

G

gain
 current, 385, 802, 896
 loop, 694
 maximum voltage gain, 846
 noise, 694
 open-loop, 694
 relating voltage and current gain, 386
 sensor, 694
 voltage, 864
 galvanometer, 68
 Gauss, 101
 generator, 282
 constant current, floating load, 358
 constant current, grounded load, 359
 ramp voltage, 808
 square wave, 633
 triangle wave, 633
 graceful degradation, 290
 ground, 65
 chassis, 65
 conductor noise, 1038
 earth, 65, 416, 417
 high quality, 65
 safety, 417
 guarding, 1040
 gyrator
 analysis, 600
 assessment, 602
 improved, 603
 simple, 599
 transient analysis: a reality check, 600

H

h-parameter model, 895
 Happy Moment, 988
 Hartley-Jones, 618
 heat sink, 454, 783
 area and thermal resistance, 788
 derating curve, 785
 design, 786
 thermal capacitance, 789

Henries, 101
 hidden schematic, 1014, 1029
 hit-and-miss engine, 438
 hobby course, 43
 Howland current generator, 597
 analysis, 597
 differential input, 598
 opinions on usefulness, 599
 Howland integrator
 quadrature oscillator, 628
 hysteresis, 401
 capacitor, 508

I

IA, instrumentation amplifier, *see* instrumentation amplifier
 ignition coil, 98
 impedance, 190, 191
 in-phase component, 616
 inductance, 98, 101
 air core coil calculation, 104
 bifilar choke, 1040, 1043
 capacitor, 1036
 differential equation, 98
 effect of conductor length and width, 1034
 energy storage, 100
 equation, 103
 Faraday's law, 98
 parasitic, 708
 power supply, 1034
 printed wiring trace, 1033
 rule of thumb, 99
 snubber, 98
 wiring, 708, 1032
 induction, 95
 magnetic law of, 95
 moving magnet, 95
 self induction, 95, 96
 inductor, 95

instrumentation amplifier, 1072
 performance, 1073
 using op-amp, 341
integrator, 346
 Howland non inverting, 350
 in frequency domain, 348
 in time domain, 346
 non-inverting, 349
interference, 1013
 A/D converter, 1051, 1054
 bifilar choke, 1040, 1043
 case histories, 1013, 1054
 case histories, solved, 1054
 common-mode current, 1043
 common-mode interference, 1044
 component parasitics, 1014
 conductive coupling, 1016
 coupling mechanism, 1015
 coupling, conductive, 1016
 coupling, electric field, 1017
 coupling, electromagnetic, 1045
 coupling, magnetic field, 1020
 differential amplifier, 1039
 electric field, 1019
 electric field coupling, 1017
 electric field shielding, 1018
 electromagnetic compliance, 1013
EMC, electro-magnetic compliance, 1013, 1041, 1044
EMI, electro-magnetic interference, 1043
ground conductor noise, 1038
ground plane, 1054
grounding, 1051
guarding, 1040
hidden schematic, 1014
ideal and real component behaviour, 1014
magnetic field, 1022, 1024
magnetic field coupling, 1020
measurement technique, 1046
measurement, active probe, 1049
measurement, coax cable connection, 1048
measurement, single lead connection, 1047
measurement, twin lead connection, 1047
mechanism, 1015
medium, 1015
PC board layout, 1049
receiver, 1015
scope of the problem, 1013

shielding, 1013
shielding, magnetic field, 1024
source, 1015
invisible schematic, 718

J

JFET
 buffer, 986
 source follower, 986
jfet, 268, 976
 active device, 266
 amplifier, 983
 applications, 976
 common-gate amplifier, 991
 common-source amplifier, 991
 compound amplifier, 991
 compressor, 980
 constant current diode, 976
 datasheet, 993
 differential amplifier, 991
 drain saturation current I_{dss} , 268
 gate, drain, 268
 long-tailing, 978
 low voltage region, 979
 ohmic region, 979
 opto-isolated switch, 983
 oscillation, 988
 oscilloscope buffer amplifier, 991
 pinch-off voltage, 979
 pinch-off voltage V_p , 268
 saturation, 269, 979
 small signal model, 987
 source follower equations, 990
 strengths of, 983
 switch, 981
 transconductance, 984
 transconductance g_m , 268
 transfer characteristic, 978
 transresistance, 986

JFET, junction field effect transistor, *see* jfet
Joseph Henry, 99

K

KCL, 64
Kettering ignition system, 98
Kirchhoff's Current Law, 64

Kirchhoff's Voltage Law, 63

KVL, 63

- and KCL, in circuit analysis, 65
- and the capacitor, 88

L

Lag

- first order, 307

lag-lead network, 720

Laplace transform, 195

- arbitrary waveform, 238
- as shorthand, 197
- capacitive reactance, 199
- capacitor with initial charge, 226
- computer algebra, 212
- further reading, 244
- inductor differential equation, 198
- LC resonant circuit, 228
- linearity property, 238
- LR lowpass filter, 203
- overview, 195
- RC lowpass filter, 199
- RLC resonant circuit, 232
- second order lowpass, 236
- table of quadratics, 214
- table of transforms, 214
- time shift property, 239
- transfer function and frequency response, 199
- two stage RC lowpass filter, 210

unit impulse, 240

unit impulse and RC lowpass response, 217

unit impulse forcing function, 215

unit ramp and RC lowpass, 223

unit ramp forcing function, 216

unit step, 242

unit step and RC lowpass, 220

unit step forcing function, 216

large signal frequency response, 658

lateral format mosfet, 1004

Lead

- first order, 307

LED, *see* light emitting diode

Lenz's law, 96, 280

level shift, 798, 915, 948

level translator, 966

lever diagram, 320, 344

fully differential amplifier, 1080

inverting amplifier, 322

non-inverting amplifier, 320

light emitting diode, 85, 773

colour, 773

driving multiple, 773

flasher, 773

infrared, 773

ultraviolet, 773

wavelength, 773

white, 773

load line, 79

design example, 139

loading effect, 53

logic

gate, 139

logic device

BJT, 276

flip-flop, 278

LSTTL, 142

model, 140

output characteristic, 141

logic level translator, 966

long-tailing, 978

loop gain, 694

LVDT, linear variable differential transformer, 1124

M

magnetic

circuit, 101

circuit analog, 101

force, 280

permeability, 101

saturation, 104

magnetism, 101

B-H curve, 104

electric circuit analogy, 101

magnetizing force, 101

Marx generator, 110

mdac

frequency control, 742

gain control, 741

membrane model, 577

meter, 68

characterization, 68

current, 54

full wave driver circuit, 502

moving coil, 68

- voltage, 53
 - mil, 1034
 - circular, 482
 - Miller effect, 345
 - Minas Basin, 176
 - mirror
 - current, 917
 - mixer, 1010
 - MMF, magnetomotive force, 101
 - mobile robot, 1112
 - model, *see* equivalent circuit
 - bjt current generator, 801
 - bjt h-parameter, 895
 - bjt t-parameter, 897
 - bjt, as diode, 799
 - discussion of bjt models, 900
 - jfet small signal, 987
 - logic device, 140
 - mosfet, 269, 272, 997
 - active device, 266
 - bias, 999, 1003
 - determining transconductance, 1002
 - enhancement mode, 998
 - family, 997
 - follower bias, AC coupled, 1003
 - follower bias, DC coupled, 999
 - follower output resistance, 1002
 - follower voltage gain, 1002
 - linear applications, 1004
 - power amplifier, 1004
 - source follower voltage gain, 1000
 - switch, 272, 273
 - threshold voltage, 998, 1006
 - transfer characteristic, 997
 - transresistance, 1000
 - vertical and lateral format, 1004
 - motor, 280
 - behaviour, 1107
 - control system, 1108
 - controller, 1107
 - driver circuit, 1114
 - equations, 281
 - ideal, 281
 - non-ideal, 282
 - speed-torque curve, 281
 - motorboating, 1016
 - moving coil
 - ammeter, 70
 - meter, 68
 - ohmmeter, 71
 - voltmeter, 69
 - MSB, 735
 - mu-metal, 1024
 - multiplexer, 1009
 - multiplier, 338, 1094, 1128
 - 2 quadrant, 1070, 1098
 - duty-cycle, 1096
 - mdac, 1096
 - ota, 1069
 - photoresistive cell, 1103
 - pulse width-height, 1096, 1103
 - quadrant of operation, 740
 - transconductance, 1098
 - multiplying digital-analog converter, *see* mdac
- N**
- negative feedback
 - adder, 286
 - and amplifier, 286
 - block diagram, 284
 - concept, 284
 - emitter follower, 877
 - error signal, 286
 - frequency response, 691, 702
 - goes bad, 290
 - hunting, 291
 - in current feedback amplifier, 1064
 - minimizing disturbance, 288
 - operational amplifier, 287
 - oscillation, 290
 - output resistance, 654
 - plant, 286, 309
 - preventing oscillation, 291
 - saturation, 290
 - sensor gain, 286
 - setpoint, 309
 - steering system example, 285
 - temperature controller, 291
 - voltage regulator, 431
 - negative impedance converter, 593
 - negative resistance, 594
 - network, RC highpass, 172
 - network, RC lowpass, 169
 - NIC, *see* negative impedance converter
 - noise, 667

- adding noise signals, 669
- amplitude, 667
- effective bandwidth, 673
- excess, 673
- figure, 892
- formulae, 684
- gain, 678
- in op-amp inverter, 681
- in voltage follower, 679
- introduction, 667
- measuring amplitude, 668
- op-amp, 674
- pink, 672, 684
- popcorn, 672
- references, 670
- resistor, 670
- signal to noise ratio, 668
- spectrum, 671
- symbols, 686
- tangential measurement, 668
- transistor, 892
- voltage density, 684
- white, 672, 684
- noise gain, 645, 683, 694, 702
- stability, 713
- non-ideal op-amp, 639
- Norton
 - current source, 121
 - equivalent circuit, 121
 - internal resistance, 121
 - theorem, 121
- Norton amplifier, 1067
 - AC amplifier, 1068
 - bias, 1068
 - equivalent circuit, 1067
 - internals, 1067
- Norton's Theorem, 121
- Nyquist frequency, 570
- O**
- Ørsted, Hans Christian, 95
- offset, 336
- offset voltage, 1085
- Ohm's Law, 56
- Ohm, Georg, 56
- ohmmeter
 - moving coil, 71
- op-amp
 - adder, 336
 - bias current, 639, 641
 - booster stage, 356
 - buffer, 318
 - charge amplifier, 360
 - CMRR, common mode rejection ratio, 639, 649
 - comparator, 317, 1071
 - compensation, 706
 - composite, 716
 - difference amplifier, 650
 - differentiator, 351
 - drift, 639
 - finite gain, 640
 - finite open-loop gain, 639
 - follower, 318, 727
 - formal analysis, 323
 - frequency response, 692
 - fully differential amplifier, 1074
 - GBP, gain bandwidth product, 639, 692
 - high frequency, 708
 - improved booster stage, 356
 - input resistance, 639
 - input resistance, effect of negative feedback, 377
 - input signals, 316
 - instrumentation amplifier, 1072
 - integrator, 346
 - introduction, 315
 - inverting amplifier, 320
 - lever diagram, 320, 322
 - limits on output voltage and current, 652
 - noise, 640, 674
 - noise gain, 645
 - noise specifications, 675
 - non-ideal, 639
 - non-inverting adder, 337
 - non-inverting amplifier, 319
 - offset voltage, 643
 - offset voltage correction, 645
 - offset voltage drift, 644
 - offsets, 639
 - output resistance, 653
 - output resistance, effect of negative feedback, 377
 - phase splitter, 615
 - power, 315
 - power dissipation, 793
 - pseudo ground, 327
 - PSRR, power supply rejection ratio, 639, 646

- rules of behaviour, 325
- sign changer, 337
- single supply operation, 326
- sitcheable sign changer, 338
- slew rate, 640, 655
- stability, 640, 706
- subtracter, 338, 650
- universal amplifier, 343
- virtual earth, 322
- Operational Amplifier, *see* op-amp
- opto-electronics
 - exercises, 779
 - introduction, 772
 - isolated jfet switch, 983
 - light emitting diode, 773
 - opto-interrupter, 779
 - opto-isolator, 778
 - optocoupler, 778
 - photodiode, 774
 - photoresistive cell, 776
 - phototransistor, 777
 - standard observer, 772
 - visible spectrum, 772
- opto-interrupter, 779
- optocoupler, 778
 - current transfer ratio, 778
- oscillator, 619
 - as negative feedback system, 619
 - comparator driven, 619
 - introduction, 619
 - negative resistance, 594, 619
 - phase shift, 620
 - quadrature, 626
 - simulation, 619
 - square wave, 633
 - three phase, 629
 - triangle wave, 633
 - wien bridge, 622
- OTA, operational transconductance amplifier, 1069
 - application, 1070
 - internal circuit, 1069
- overshoot, 567
- P**
- parasitic component, 1036
- pc board
 - wiring inductance, 1032
- wiring resistance, 1031
- peak detector, 507
 - improved, 508
 - references, 509
- Pease, Robert, 425
- Peltier Cell, 1144
- permeability
 - magnetic, 101
- phantom power, 362
- phase
 - shifter, 613
 - splitter, 615
- phase margin
 - closed-loop gain, 711
 - peaking, 710
 - Q factor, 709
 - transient response, 710
- phase sensitive rectifier, 503
- phase shifter
 - active, 615
 - all pass network, 613
 - improved, 613
 - oscillator, 620
 - passive, 613
 - wide band, 616
 - wideband, audio application, 617
- photodiode, 774
 - amplifier circuits, 775
 - photoconductive mode amplifier, 775
 - photovoltaic mode amplifier, 775
- photoresistive cell, 776
- photoresistor, 776, 1103
 - CdS cell, 776
 - photo relay, 776
 - spectral response, 776
 - transient response, 776
- phototransistor, 772, 777
 - photodarlington, 777
 - sensitivity, 777
 - spectral response, 777
 - switching speed, 777
- PID controller, 309
 - Bode plot, 311
 - example design, 310
 - Ziegler-Nichols rules, 310
- piezo-electric crystal, 360
- piezoelectric material, 86
- plotting

amplitude and phase, 242
 polarity
 voltage, 62
 poles and zeros, 305, 573
 complex, 583
 high order filters, 584
 on the real axis, 582
 potentiometer, 124
 power
 definition, 83
 dissipation, 83
 equation by analogy, 83
 op-amp dissipation, 793
 phantom, 362
 power gain, 266
 power supply, 416, 1034
 capacitor filter and linear regulator, tradeoffs, 423
 capacitor filter, idealized, 422
 charge pump, 455
 components, 418
 current practice, 428
 design summary, 474
 flaming debris, 419
 inductor filter, 427
 isolated, 468
 line regulation, 428
 load regulation, 428
 measurement checklist, 475
 rectification, 419
 references, 487
 regulation, 419
 regulator, 428
 remote sensing, 433
 robot, 1112
 transformer resistance, 424
 voltage inverter, 456
 voltage multiplier, 459
 pre-distortion, 1098
 precision and accuracy, 732
 precision diode, 495
 precision rectifier, 495
 full wave, 500
 half wave, 499
 pressure sensor, 1129
 primary winding, 98
 process constant, 894, 895
 PSR, *see* phase sensitive rectifier
 PSRR, power supply rejection ratio, 646

Q

Q factor, 175, 177, 178, 527, 702
 bandwidth, 180, 182
 damping factor, 178
 definition, 178
 energy dissipation, 179
 phase margin, 709
 transient response, 711
 voltage magnification, 180
 Q point, 79
 quadrature component, 616
 quadrature oscillator, 626
 applications, 628
 based on state variable filter, 628
 basic, 627
 using Howland integrator, 628
 quantization, 749
 quantization noise, 749

R

rabbits and wolves, 354
 ramp generator, 89, 91
 non-ideal, 91
 rc highpass
 frequency response, 172
 RC highpass, time domain, 172
 rc lowpass
 frequency response, 170
 rc lowpass, time domain, 169
 RC network, highpass, 172
 RC network, lowpass, 169
 reactance, 157
 calculators, 166
 capacitive, 161
 capacitive in vector notation, 162
 capacitive summary, 163
 capacitive, effect of frequency, 163
 chart, 166
 inductive, 164
 inductive, effect of frequency, 165
 inductive, summary, 165
 inductive, vector notation, 165
 plot, 300, 301
 plotting routine, 168
 slide rule, 168
 rectifier

- bridge, 251
- full wave, 249
- full wave bridge, dual polarity, 422
- full wave meter driver, 502
- full wave, bridge, 422
- full wave, centre tapped, 421
- half wave, 249, 421
- half wave, dual polarity, 421
- phase sensitive, 503, 1128
- precision, 495
- precision full wave, 500, 501
- precision half wave, 499
- precision rectifier, references, 502
- switchable amplifier, 500
- reference
 - precision voltage, 1083
- reflecting resistance, 815
- reflection principle, 875
- regulation
 - distributed vs local, 419
- regulator
 - adjustable voltage, 434
 - buck, 416
 - fixed voltage, 434
 - linear, 416
 - low dropout, 434
 - series, 429
 - shunt, 263, 429
 - step down, 416
 - switching, 416, 437
 - three terminal, 433
- relay, 110
- reluctance, 101
 - variable, 102
- resistance, 56
 - bjt emitter incremental, 834
 - bleeder, 121
 - buildout, 724
 - collector, 817
 - dropper, 121
 - example, 57
 - incremental, 57
 - input, 639
 - internal, 72, 73
 - internal, R_{int} , 77
 - internal, by Test Voltage method, 78
 - negative, 593, 619
 - negative incremental, 595
- non-linear, 57
- printed wiring trace, 1031
- reflecting, 815
- standard, 58
- Thevenin, 119
- transresistance, 986
- wiring, 1029
- resistivity, 58
- resistor
 - carbon composition, 673
 - design, 59
 - flushout, 912
 - metal film, 673
 - noise, 670
 - photoresistor, 776
 - precision, 58
 - standard values, 58
- resistors
 - parallel, 67
 - series, 66
- resolver, 1119
 - interface, 1121
- resonance
 - capacitor, 1037
 - half-power frequencies, 182
 - introduction, 173
 - parallel, 176
 - parallel, frequency response, 177
 - series, 174
 - series, frequency response, 175
 - tank circuit, 173
 - voltage magnification, 176, 180
- resonant frequency, 527
- reverse saturation current, 894
- right hand rule, 95
- ringing, 567
- RMS conversion, 505
 - implicit method, 506
 - obvious implementation, 505
 - thermal method, 507
- RMS value of waveform, 483
- robot
 - analog, 1112
 - edge detection, 1115
 - object detector, 1116
 - power supply, 1112
 - speed control, 1116
 - sumo, 1112

rotor, 68

S

Sallen-Key filter, 539

- highpass, 545

- lowpass, 539

- references, 543

sample-hold, 509

- application, A/D converter, 510

- reference, 511

- specifications, 511

- track-hold, 510

saturation current, I_s , 267

saturation voltage V_{cesat} , 266

Schmitt trigger, 401

- astable, 631

- design equations, 406

- design equations, inverting version, 407

- design overview, 402

- fan thermostat example, 410

- hysteresis, 401

- inverting, 403

- lever diagram, 405, 413

- non-inverting, 405

- non-inverting design equations, 408

- summary of design equations, 410

- transfer function, 402

second breakdown, 1006

secondary winding, 98

self-bias, 986

sensor

- air pressure, 1129

- linear displacement, 1124

- rotation, 1119

- variable reluctance, 102

sensor gain, 694

shunt, 71

shunt regulator, 263

sign changer, 337, 505

- switchable, 338

signal, 43

- analog, 732

- average, 495

- common mode, 1038

- common-mode, 341, 1074

- differential, 341, 1038, 1074

- digital, 732

peak, 496

peak to peak, 496

properties, 495

signal integrity, 1013

simulation, 44

- spring-mass system, 354

- wrong answers quickly, 425

sine wave shaper, 634

single sideband, 616

slew rate, 498, 655

- increasing, 656

- sine wave signals, 657

snubber, 98

soakage, 508

source

- controlled, 266

- current, 915

- current controlled current, 266

- floating, 65, 188

- generalized, 81

- grounded, 65

- ideal current, 62

- ideal voltage, 61

- non-ideal, *see source*, practical

- practical current, 77

- practical voltage, 72

- voltage and current, compared, 81

- voltage controlled current, 266, 272

source follower, 986

- equations, 990

spectrum, 772

Spice, 65

spreadsheet

- calculation of tolerance, 1137

stability

- buffer, 716

- buildout resistance, 724, 725

- by reduced noise gain, 713

- capacitive loading, 722

- compensated attenuator, 726

- decoupling, 707

- differentiator, 711

- dominant pole compensation, 706

- exercises, 729

- hidden schematic, 707

- input capacitance, 718

- intercept rule, 711

- introduction, 706

lag-lead network, 719
 loop gain, 709
 op-amp, 706
 output buffer stage, 716
 phase margin, 708
 standard
 current, 54
 primary voltage, 53
 resistance, 58
 voltage, 53
 standard observer, 772
 state variable filter, 546
 stator, 68
 Steinmetz, Charles, 157
 step response, 701
 strain gauge, 1083
 subtractor, 338
 difference amplifier, 339
 high-impedance inputs, 339
 negate and add, 338
 superposition, 134
 and power, 136
 applications, 134
 passive adder, 137
 switch
 analog CMOS, 1009
 bjt, 274
 electromechanical, 108
 mosfet, 272, 273
 whisker, 1112
 switching regulator, 437
 circuit layout, 455
 concept, 437
 design example, 449
 discontinuous operation, 446
 energy storage inductor, 442
 filter capacitor, 442
 heat sink, 454
 mechanical analog, 438
 operation, 438
 power losses, 448
 rationale, 449
 synchro resolver, 1119
 synchronous demodulator, 628
 system
 analog vs digital, 1105
 electronic, 1105
 Sziklai pair, 913

T

t-parameter model, 897
 tanh function, 950
 tank, 173
 tee model of bjt, 837
 Teslas, 101
 thermal capacitance, 789
 thermal circuit, 784
 thermal inertia, 624, 789
 thermal resistance, 784, 785
 thermocouple, 56, 1083
 Thevenin
 and AC circuits, 120
 and Norton, 121
 equivalent circuit, 119
 puzzle, 152
 resistance, 119
 theorem, 119
 thevenizing, 119
 voltage, 119
 Thompson, J.J., 50
 three phase oscillator, 629
 threshold voltage
 mosfet, 998
 time constant, 169, 172, 200, 307, 567, 745
 tolerance, 58, 145, 1137
 buildup, 145
 cumulative, 145
 error budget, 149
 nominal value, 145
 spreadsheet analysis, 147
 totem-pole, 913, 1012, 1114
 track-hold, 510
 transconductance, 272, 984, 1000, 1069
 amplifier, 385
 bjt, 836
 jfet, 268
 mosfet, 1000
 multiplier, 1098
 transfer characteristic
 mosfet, 997
 transformer
 core materials, 465
 design, 460
 design example, 468
 dotted terminal, 189
 equation, 476

ideal, 186
 phase, 189
 polarity, 189
 reflection concept, 188
 resistance, 425
 symbols, 476
 transistor
 BJT, *see* bjt
 bjt switch, 274
 JFET, *see* jfet
 mosfet, 997
 mosfet switch, 272
 phototransistor, 777
 translinear, 506
 transresistance, 986, 1000, 1065
 triboelectricity, 51

U

units, converting, 48

V

vacuum tube, 55
 vertical format mosfet, 1004
 virtual earth, 322, 325
 virtual ground, 325
 volt, 52
 volt-seconds, 96
 voltage, 51
 compliance, 62
 divider, 119, 121
 Early, 820
 example, 52
 induction, 95
 line voltage equation, 161
 open-source, E_{oc} , 73
 polarity, 62
 reference polarity, 62
 standard, 53
 Thevenin, 119
 to frequency converter, 764
 voltage divider
 compensated, 184
 design nomograph, 132
 design of, 125
 design, exact method, 125
 design, graphical method, 130

design, large ratio shortcut, 131
 design, simple method, 127
 evolution of, 121
 generalized, 122
 thevenin equivalent, 124
 think Thevenin, 124
 two resistor, 123
 voltage gain, 864
 voltage multiplier, 459
 voltage source
 ideal, 61
 non-ideal, 72
 voltmeter, 53
 moving coil, 69
 sensitivity, 70

W

wall wart, 416
 watt, 83
 watt-hour, 83
 watt-second, 83
 wave filter, 513
 waveform, 154
 periodic, 155
 sine wave, 155
 sine wave generation, 155
 sinusoidal, 155
 spectrum of pulse, 1026
 spectrum of square wave, 1026
 spectrum of trapezoid, 1027
 vector notation and sine waves, 157

Webers, 101
 wheatstone bridge, 138
 whisker switch, 1112
 Widlar, Robert, 433
 wien bridge oscillator, 622
 amplitude stabilization, 624
 analysis, 623
 comparison, 626
 diode limiter, 625
 integrator comparator, 625
 references, 624
 Williams, Jim, 716
 wire gauge, 482
 wire table, 482

X

X-ray detector, 361

Z

zener

- diode, 259
 - impedance, 261
 - regulator, 259
 - with constant-current drive, 263
 - with current amplifier, 262
- Ziegler-Nichols rules, 310
- zoo
- unusual circuits, 593

Colophon

The source files for this text were created on a computer running the Suse 10.3 variant of the Linux operating system. The text was edited using the `joe` editor and processed in the L^AT_EX typesetting program.

Diagrams were created using the `xfig` drawing program and exported in `pictex` format. This format is processed as L^AT_EXtext, and has the advantage that it can incorporate mathematical formulae.

The electronic symbols were created using `xfig`. With each new drawing, the complete collection of symbols was loaded into `xfig` and the necessary symbols copied from the collection. The collection of symbols is available as a download [363].

Graphs were created with `gnuplot` and exported in `fig` format. They were then processed in `xfig` and exported in `pictex` format for inclusion in the text.

Images were scanned on a Canon Lide 30 flatbed scanner hosted by a computer running the Microsoft XP operating system. They were then transferred to the Linux machine and formatted using the `xv` image processing program.

The final text has approximately 1100 body text pages and 1000 diagrams, for a diagram:page ratio of 91%. It meets the design objective that most pages should have at least one diagram.

The L^AT_EX packages `wrapfig` and `subfig` were used extensively throughout the manuscript. The files were organized into 380 subject files that were then input by the master document file. The `comment` package was used to select sections of the manuscript for debugging.

The index was prepared with `Makeindex` and the bibliography with `BIBTEX`.

With the exception of the scanner program (which was bundled with the scanner) all these programs were Open Source. The project was made possible by these amazing tools. Special thanks to the programmers that created them.