

ES222 - Course Notes

Koen Boeckx

March 20, 2023

Contents

1	Introduction	7
2	Circuit Theory	9
2.1	Passive Components	9
2.2	Laws of Kirchhoff	11
2.2.1	Kirchhoff's voltage law	11
2.2.2	Kirchhoff's current law	11
2.2.3	Example	11
2.3	Frequency Representation	12
2.3.1	Series and Parallel Combinations	13
2.3.2	Millman's theorem	14
2.3.3	Example	14
2.4	Circuit Transformations	15
2.4.1	Thévenin's Theorem	15
2.4.2	Norton's Theorem	16
2.4.3	Two Ports	17
2.4.4	Cascaded Circuits	17
I	Components	19
3	Solid State Theory & Semiconductors	21
3.1	Semiconductors	21
3.1.1	Band structure	21
3.2	Electrons and Holes	22
3.2.1	Effective mass	24
3.2.2	Number of carriers	25
3.3	Donor and Acceptor Impurities	26
3.4	Carrier Transport	28
3.4.1	Drift Current	28
3.4.2	Band structure under biasing	28
3.4.3	Diffusion Current	29
3.5	Generation and Recombination	30
3.6	The continuity equations	31

4 The pn-junction	33
4.1 The pn-junction in equilibrium	33
4.1.1 Fermi-levels in equilibrium	34
4.1.2 The built-in potential	35
4.2 The pn-junction under bias	36
4.2.1 Forward bias	36
4.2.2 Reverse - bias	36
4.2.3 Diode Characteristic	37
4.2.4 Practical Diode Characteristic	39
4.3 Avalanche effect	40
4.4 Depletion Capacitance	40
5 Transistors	41
5.1 Bipolar Junction Transistor	41
5.1.1 Operation in Active Mode	41
5.1.2 Currents in Active Mode	42
5.1.3 Carrier distribution in Active Mode	45
5.1.4 Modes of Operation and Ebers-Moll Equations	47
5.1.5 Common-Emitter configuration	48
5.1.6 Early Effect	49
5.2 MOSFET	50
5.2.1 Description and Operation	50
5.2.2 MOS Capacitor	51
5.2.3 I-V characteristic	52
5.2.4 Second Order Effects	54
II Analog Electronics	57
6 Basic Analog Circuits	59
6.1 Non-linear elements in circuits	59
6.1.1 The Diode as a Circuit Element	59
6.1.2 The BJT as a Circuit Element	60
6.1.3 The MOSFET as a Circuit Element	63
6.1.4 Additional remarks	64
6.1.5 A more general circuit	64
6.2 Small-Signal Response	65
6.3 Static and Dynamic Load lines	69
6.3.1 Transistors and Dynamic Load Lines	71
6.4 Biasing	72
6.4.1 BJT Biasing	72
6.4.2 MOSFET Biasing	74
6.5 The Small-Signal Model	76
6.5.1 BJT Small-Signal Model	76
6.5.2 MOSFET Small-Signal Model	77
6.5.3 Orders of magnitude	78

CONTENTS	5
7 Amplifiers	79
7.1 Basic Amplifier	79
7.1.1 Coupling Capacitance	79
7.1.2 The 4-resistor amplifier	81
7.2 Basic Topologies	83
7.2.1 Common Emitter Amplifier (CEA)	83
7.2.2 Common Base Amplifier (CBA)	84
7.2.3 Common Collector Amplifier (CCA)	86
7.2.4 Comparison of Topologies	87
7.3 Difference Amplifier	87
7.3.1 Definition	87
7.3.2 Implementation	89
7.3.3 Common Mode	89
7.3.4 Differential Mode	90
7.3.5 Load-line Analysis	91
7.3.6 Common & Differential Gain	92
7.3.7 Power-Supply Rejection Ratio	94
7.4 Operational Amplifier	95
7.4.1 The Ideal Amplifier	95
7.4.2 The Real Amplifier	97
7.4.3 OPAMP Theory	98
7.4.4 Unity Gain Buffer	102
8 Transistors at High Frequency	105
8.1 Giacoletto Model at High Frequencies	105
8.2 The Miller Capacitor	106
8.3 Miller's Theorem	108
8.4 Conclusion	109
9 Power Amplifiers	111
9.1 Introduction	111
9.2 Class A Amplifier	112
9.2.1 Improvement to the class A amplifier	115
9.3 Class B Amplifier	116
9.4 Push-Pull Amplifiers	119
9.5 Class C Amplifier	121
9.5.1 The RLC Tank	122
9.6 Class D Amplifier	123
9.7 Class S Amplifier	124
9.8 The Selective Amplifier	126
10 Feedback Theory	129
10.1 Accurate Gain Control	129
10.2 Input and output impedance with feedback	131
10.3 Increased bandwidth	131
10.4 Distortion reduction	132
10.5 Stability Issues	133

11 Oscillators	137
11.1 Phase Shift Oscillator	137
11.2 Wien Bridge Oscillator	139
11.3 Colpitts Oscillator	140
11.4 Quartz Oscillator	142
11.5 Relaxation Oscillator	143
11.6 The Fantastron	146
12 DC Voltage Generation	149
III Digital Electronics	151
13 Digital Circuits	153
13.1 Combinatorial Digital Systems	153
13.2 Sequential Digital Systems	153
14 Advanced Digital Circuits	155
15 A/D and D/A Converter Circuits	157

Chapter 1

Introduction

This course is the continuation of the electronics course given by Prof. dr. ir. Patrick MERKEN, who created the course contents and designed most of the figures used in the work. My thanks go out to him.

Chapter 2

Circuit Theory

In this chapter, we reiterate several elements from elementary circuit theory that are necessary to understand basic electronic circuits. These elements include passive components like resistors, capacitors and inductors, the laws of Kirchhoff, the phasor representation and the complex impedance, and several circuit transformations to make analysis easier.

2.1 Passive Components

In the electronic circuits we will study, we encounter 4 kinds of elements:

- Sources: these are relatively complex systems with two terminals that provide either a voltage or a current. Sources can either be constant, i.e. the generated current or voltages doesn't vary with time, or they generate a current or voltage that does vary with time. In the former case we call them DC sources, in the latter case we speak of AC sources, typically with an average value of zero.
- Linear elements: these can be passive, like resistors, capacitors or inductors, or active, like dependent current or voltage sources. The latter are sources that depend linearly on other currents or voltages in the circuit.
- Non-linear elements, such as diodes and transistors. The study of electronics concerns these elements and their usage is circuitry.
- Conductors, which connect the different lumped elements. We suppose that they are ideal: they have no resistance, inductance or capacitance. If any of these non-idealities are present in real conductors, they will be modeled as separate lumped elements.

We always use the quasi-static approximation, where the value of the current in a branch is the same everywhere in that branch. This is valid when the dimensions of the circuit are much smaller than the wavelength of the signal.

The most important linear passive components are:

- A *resistor* is an element that resists the flow of current due to an applied voltage. The current-voltage relation is:

$$v_R = R i_R$$

The resistance of a resistor is measured in ohms (Ω).

- A *capacitor* is a passive electronic component that stores electrical energy in an electric field. It consists of two conductive plates separated by an insulating material, called the dielectric.

When a voltage is applied across the plates of a capacitor, electrical charge Q accumulates on the plates, creating an electric field between them. The current-voltage relation is

$$v_C = \frac{Q}{C} = \frac{1}{C} \int i_C dt$$

or

$$i_C = C \frac{dv_C}{dt}$$

A capacitor resists a sudden change in voltage across its terminals. The capacitance of a capacitor is measured in farad (F).

- An *inductor* is a passive electronic component that stores electrical energy in a magnetic field. It consists of a coil of wire, often wrapped around a core made of a magnetic material, such as iron or ferrite.

When an electric current flows through an inductor, a magnetic field is created around the coil. The strength of the magnetic field is proportional to the amount of current flowing through the coil. When the current changes, the magnetic field changes, inducing a voltage across the coil that opposes the change in current. Thus, an inductor resists a sudden change in current. The current-voltage relation is

$$v_L = L \frac{di_L}{dt}$$

or also

$$i_L = \frac{1}{L} \int v_L dt$$

The inductance of an inductor is measured in henry (H).

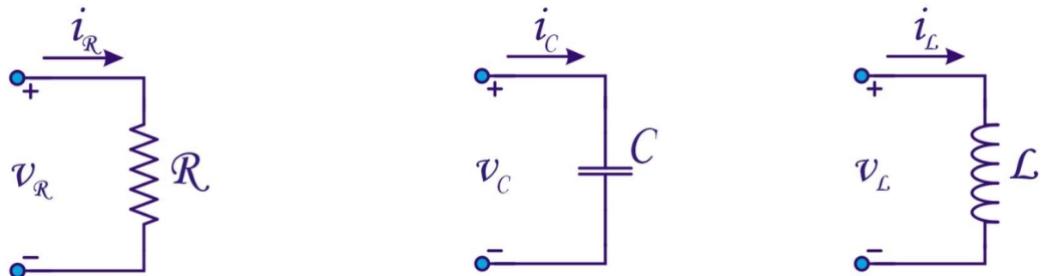


Figure 2.1: Resistor (left), capacitor (middle) and inductor (right)

2.2 Laws of Kirchhoff

2.2.1 Kirchhoff's voltage law

Kirchhoff's voltage law (KVL) states that the directed sum of the potential differences (voltages) around any closed loop is zero:

$$\sum_{k=1}^n v_k = 0 \quad (2.1)$$

KVL is actually a restatement of Faraday's law $\nabla \times \vec{E} = -\frac{\partial B}{\partial t} = 0$ where we assume that that (time-varying) magnetic fields are contained to each component and the field in the region exterior to the circuit is negligible.

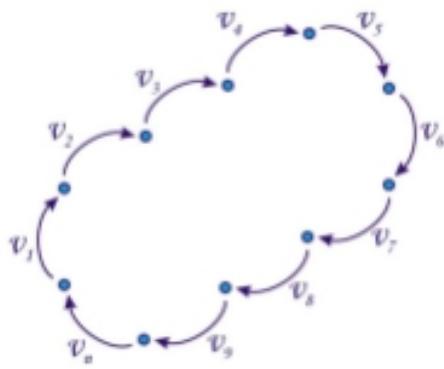


Figure 2.2: Kirchhoff's voltage law

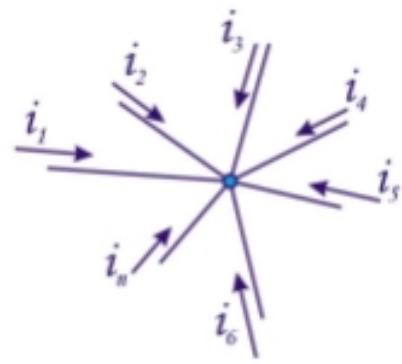


Figure 2.3: Kirchhoff's current law

2.2.2 Kirchhoff's current law

Kirchhoff's current law (KCL) says that in any node of an electric circuit, the algebraic sum of currents is zero:

$$\sum_{k=1}^n i_k = 0 \quad (2.2)$$

or equivalently, that the sum of currents flowing into that node is equal to the sum of currents flowing out of that node. The law relies on the fact that there is no accumulation of charges in any network node.

2.2.3 Example

Consider the circuit in 2.4. With KVL, we can write:

$$v_{in} = v_R + v_C = R i + v_C$$

with $i = C \frac{dv_C}{dt}$. The equation to find v_C becomes:

$$v_{in} = RC \frac{dv_C}{dt} + v_C$$

If v_{in} is suddenly turned on (e.g. by closing a switch at $t = 0$), then $v_{in} = V_0 u(t)$ with $u(t)$ the step function. We can then solve for $v_C(t)$:

$$\begin{aligned} RC \frac{dv_C}{dt} + v_C &= V_0 \\ \frac{dv_C}{dt} &= \frac{1}{RC}(V_0 - v_C) \\ \frac{dv_C}{V_0 - v_C} &= \frac{dt}{RC} \\ \int_0^{v_C} \frac{dv_C}{V_0 - v_C} &= \int_0^t \frac{dt}{RC} \\ -\ln(V_0 - v_C) &= \frac{t}{RC} + K' \\ v_C(t) &= V_0 - K e^{-\frac{t}{RC}} \end{aligned}$$

with $K = V_0$ so that $v_C(t = 0) = 0$:

$$v_C(t) = V_0(1 - e^{-\frac{t}{RC}}) = V_0(1 - e^{-\frac{t}{T}})$$

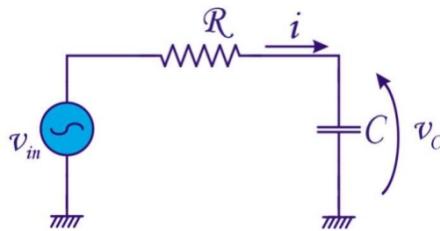


Figure 2.4

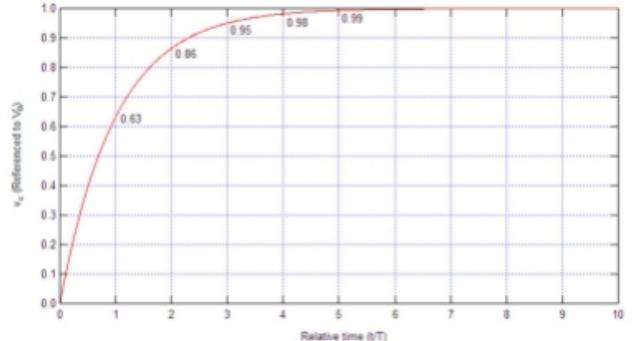


Figure 2.5

This function is represented in figure 2.5. Voltage v_C increases as the step response of a first-order system, and reaches 63% of the final value after 1 characteristic time $T = RC$. It takes $5T$ to reach 99% of the final value.

2.3 Frequency Representation

Assume an sinusoidal signal with frequency $f = \omega/2\pi$ is applied to a circuit that only contains linear elements. Linear system theory learns us that then every current and voltage will also be a sinusoid with the same frequency. Assume the current in an element can be written as $I = I_0 \cos(\omega t) = \operatorname{Re}\{I_0 e^{j\omega t}\}$. In that case, we can write:

- For a resistor: $V_R = R I_R$
- For a capacitor: $V_C = \frac{1}{C} \int I_L dt = \frac{1}{C} \int I_0 e^{j\omega t} dt = \frac{1}{j\omega C} I_0 e^{j\omega t} = \frac{1}{j\omega C} I_C$
- For an inductor: $V_L = L \frac{dI_L}{dt} = L \frac{d(I_0 e^{j\omega t})}{dt} = j\omega L I_0 e^{j\omega t} = j\omega L I_L$

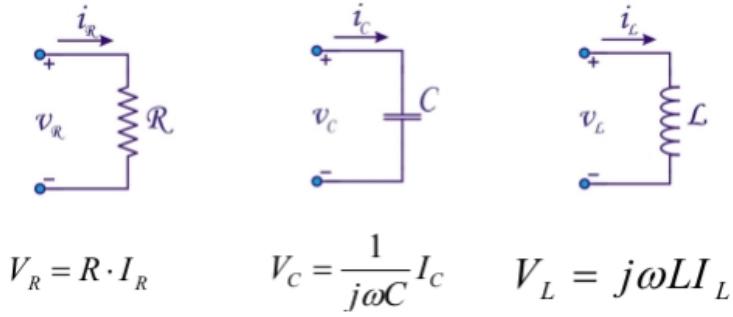


Figure 2.6

These relations are the phasor representation of a sinusoidal signal, or the Steinmetz transformations. They are summarized in figure 2.6.

From this, we conclude that for every element, we can write:

$$V = Z \cdot I$$

where Z is the *complex impedance* of the element: R for a resistor, $\frac{1}{j\omega C}$ for a capacitor and $j\omega L$ for an inductor.

We also define other circuit parameters: $Y = \frac{1}{Z}$ as the *admittance* such that $I = Y \cdot V$, and $G = \frac{1}{R}$ as the *conductance* (measured in siemens or A/V).

2.3.1 Series and Parallel Combinations

Two impedances are in series when the same current I runs through them, as in figure 2.7. The voltage drop over both is thus $Z_1 \cdot I + Z_2 \cdot I = (Z_1 + Z_2) \cdot I = Z \cdot I$ with $Z = Z_1 + Z_2$ the series combination of Z_1 and Z_2 .

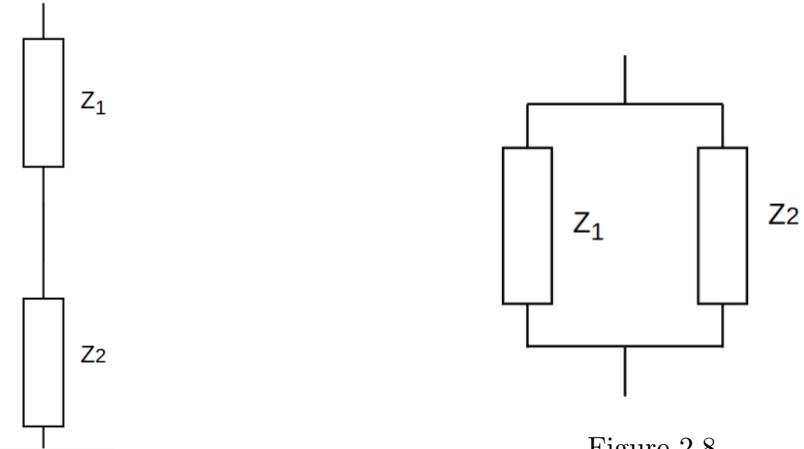


Figure 2.8

Figure 2.7

Two impedances are in parallel when the same voltage V is applied to their terminals, as in figure 2.8. The current through Z_1 is V/Z_1 and the current through Z_2 is V/Z_2 . The total current I through both is thus $V/Z_1 + V/Z_2 = V(\frac{1}{Z_1} + \frac{1}{Z_2}) = \frac{V}{Z}$ with $Z = (\frac{1}{Z_1} + \frac{1}{Z_2})^{-1}$ the parallel combination of Z_1 and Z_2 .

2.3.2 Millman's theorem

Millman's theorem is derived from KCL and allows to calculate the voltage in a node based on the voltages in neighboring nodes (see figure 2.9):

$$v_0 = \frac{\sum_{i=1}^n Y_i v_i + I_{eq}}{\sum_{i=1}^n Y_i} \quad (2.3)$$

with Y_i the conductance of each element.

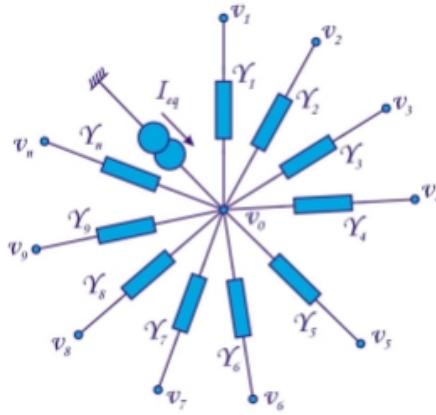


Figure 2.9

We can derive this theorem by stating KCL in node 0:

$$\begin{aligned} \sum_{i=1}^n I_n &= \sum_{i=1}^n Y_i(v_i - v_0) + I_{eq} = 0 \\ \Rightarrow v_0 \sum_{i=1}^n Y_i &= \sum_{i=1}^n Y_i v_i + I_{eq} \\ \Rightarrow v_0 &= \frac{\sum_{i=1}^n Y_i v_i + I_{eq}}{\sum_{i=1}^n Y_i} \end{aligned}$$

The voltage in node 0 is thus the weighted average of the surrounding voltages, with the conductances as weights.

2.3.3 Example

Lets apply the concepts from this section to the problem of figure 2.4. As we have assumed that all signals are sinusoidal, we state that $v_{in} = V_0 e^{j\omega t}$. Consequently:

$$\begin{aligned} V_{in} &= I \cdot R + \frac{1}{j\omega C} I \\ \Rightarrow I &= \frac{j\omega C}{1 + j\omega RC} V_{in} \\ \text{and } V_C &= \frac{1}{1 + j\omega RC} V_{in} \end{aligned}$$

We can thus conclude that:

$$\begin{aligned}\frac{V_C}{V_{in}} &= \frac{1}{1 + j\omega RC} \\ &= \frac{1}{\sqrt{1 + \omega^2 T^2}} e^{-j\phi}\end{aligned}$$

with $\phi = \text{atan}(\omega T)$. Note that we don't find any information on the transient response as before, but we find how the circuit behaves in *permanent harmonic regime* (PHM).

The ratio $\frac{V_C}{V_{in}}$ is called the *transmittance* $H(\omega)$ and it has an amplitude $A = |T(\omega)|$ and a phase ϕ . When we plot $20 \log(A)$ [dB] vs $\log(\omega)$, we find the Bode curve:

$$20 \log(A) = -20 \frac{1}{2} \log_{10}(1 + \omega^2 T^2)$$

From this expression, we deduce that:

- When $\omega T \ll 1$, $20 \log_{10}(A) \approx 20 \log_{10}(1) = 0$.
- On the other hand, when $\omega T \gg 1$, $20 \log_{10}(A) \approx -20 \log_{10}(\omega T)$ and $|H(\omega)|$ decrease by 20 dB for every increase of ω by a factor 10 (-20 dB per decade).

As a general rule, any transmittance $T(\omega)$ can be written as

$$T(\omega) = A_0 \frac{(1 + j\omega T_{n+1}) \dots (1 + j\omega T_p)}{(1 + j\omega T_1) \dots (1 + j\omega T_n)}$$

Every individual term is either equal to 1 if $\omega \ll 1/T_i$ or equal to $j\omega T_i$ if $\omega \gg 1/T_i$. Consequently, a Bode curve can be constructed approximately by following a few simple rule: when we encounter a critical pulsation $\omega = 1/T$ in the numerator (a *zero*) or in the denominator (a *pole*), the curve will:

- decreases by 20 dB per decade for a pole,
- increases by 20 dB per decade for a zero,

TODO: add explanation and rules for the phase.

An example based on figure 2.4 is given in figure 2.10.

For a first-order system, the -3 dB cut-off is reached at $\omega = \omega_0$, the critical pulsation.

2.4 Circuit Transformations

2.4.1 Thévenin's Theorem

Thévenin's theorem states that any linear two-terminal electrical network can be replaced at terminals A–B by an equivalent combination of a voltage source E_{th} in series with a impedance Z_{th} . Two circuits are equivalent if they have the same voltage-current relation at their terminals. This idea is shown in figure 2.11, where the circuit elements in the blue cloud are replaced by the source and impedance on the right.

- The equivalent voltage E_{th} is the voltage obtained at terminals A–B of the network with terminals A–B open circuited.

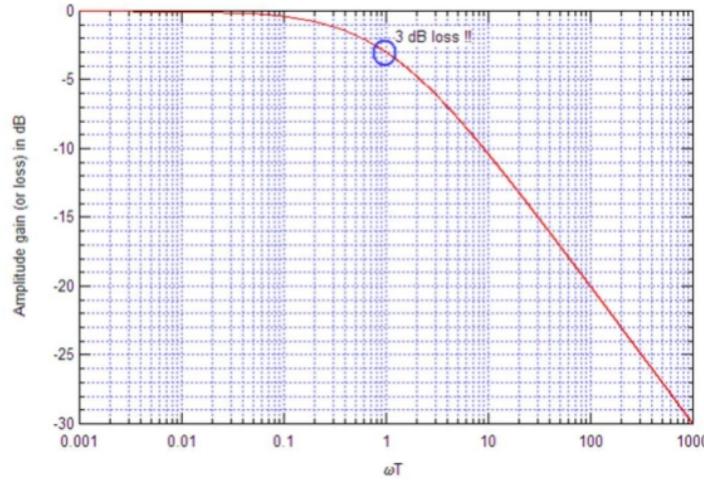


Figure 2.10

- The equivalent impedance Z_{Th} is the impedance that the circuit between terminals A and B would have if all ideal voltage sources in the circuit were replaced by a short circuit and all ideal current sources were replaced by an open circuit.

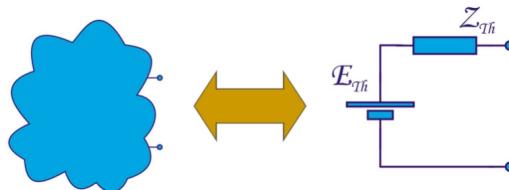


Figure 2.11: Thévenin equivalent

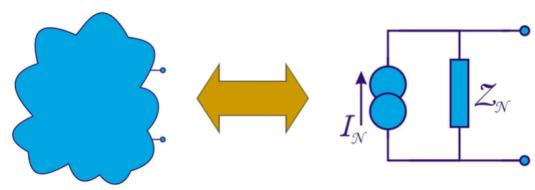


Figure 2.12: Norton equivalent

2.4.2 Norton's Theorem

Norton's theorem tells us how to replace a linear network by a current source I_N in parallel with an impedance Z_N as in figure 2.12. It's the dual of Thevenin's theorem.

- The Norton current I_N is calculated as the current flowing at the terminals into a short circuit (zero resistance between output terminals).
- The impedance Z_N is found by calculating the output voltage produced with no resistance connected at the terminals; equivalently, this is the resistance between the terminals with all (independent) voltage sources short-circuited and independent current sources open-circuited. This is equivalent to calculating the Thevenin resistance: $Z_{Th} = Z_N$.

Furthermore, the output voltage generated by the Norton source with the output terminals open equals the Thevenin voltage: $E_{Th} = Z_N I_N$.

2.4.3 Two Ports

Certain circuits have two access nodes: an input and an output port. We assume that the port is unilaterial: signals transit from input to output and don't come back. We model the input as an impedance Z_i and the output as a Thévenin or Norton equivalent circuit - see figure 2.13. Note how the current and voltage source depend on the input voltage. The

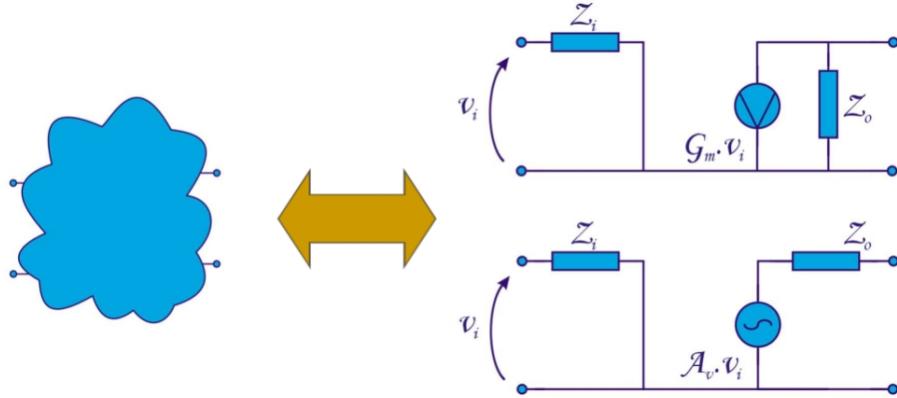


Figure 2.13

different elements have names:

- Z_i : the input impedance,
- Z_o : the output impedance,
- G_m : the transconductance,
- A_v : the voltage gain

They can be calculated in the following way:

- Z_i : apply an input voltage V_i , calculate the input current I_i . Then $Z_i = \frac{V_i}{I_i}$;
- Z_o : eliminate the transconductance by grounding the input: $V_i = 0$, apply a voltage V_o at the output and calculate the output current I_o . Then $Z_o = \frac{V_o}{I_o}$.
- G_m : short-circuit the output so no current flows through Z_o . Apply an input voltage V_i and calculate the output current I_o . Then $G_m = \frac{I_o}{V_i}$.
- A_v : keep the output open. Apply an input voltage V_i and calculate the output voltage V_o . Then $A_v = \frac{V_o}{V_i}$.

2.4.4 Cascaded Circuits

When multiple circuits are connected in a cascade, we speak of cascaded circuits. They typically consist of:

- A source, like an antenna, a signal generator or a microphone, which we model as a current source i_s or voltage source v_s , in parallel or series with an output impedance Z_s .

- An intermediate stage like an amplifier or filter that actually transforms the signal. This stage has input impedance Z_i , output impedance Z_o , a voltage gain A_v or transconductance G_m .
- A load, which can be a PC, a scope, a speaker, ... We model this as a load impedance Z_l .

This configuration is shown in figure 2.14 If we are not carefull, each stage will influence the

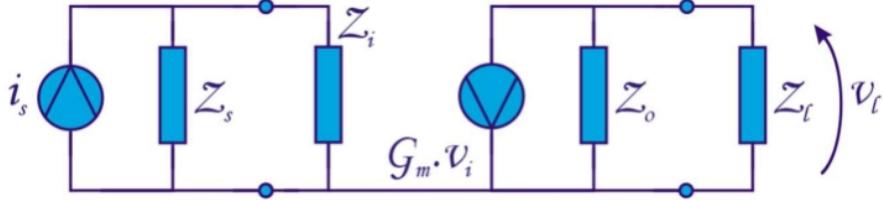


Figure 2.14

previous or next stage. We can compute the relation between source current i_s and load voltage v_l in figure 2.14:

- The voltage v_i across Z_i is:

$$v_i = i_s (Z_s || Z_i) = \frac{Z_s Z_i}{Z_s + Z_i} i_s$$

and not simply $i_s Z_s$. The intermediate stage loads the source stage.

- The load voltage v_l across Z_l is:

$$v_l = -G_m (Z_o || Z_l) v_i = -G_m \frac{Z_o Z_l}{Z_o + Z_l} v_i$$

and not simply $-G_m Z_l v_i$. The ouput impedance of the intermediate stage adds to the load.

- This means that the voltage of the signal available at the output is:

$$v_l = -G_m Z_o \frac{Z_l}{Z_o + Z_l} \frac{Z_i}{Z_s + Z_i} Z_s i_s = \frac{Z_l}{Z_o + Z_l} \frac{Z_i}{Z_s + Z_i} A_v v_s$$

Ideally, this should be $v_l = -A_v v_s$. From this we can conclude that $Z_i \rightarrow \infty$ and that $Z_o \rightarrow 0$ to approach the ideal behavior. We must design to have a high input impedance and low output impedance.

Part I

Components

Chapter 3

Solid State Theory & Semiconductors

In this chapter, we will discuss the concept of a semiconductor. We will describe what kind of material this is, based on the concept of energy bands. In solid form, a semiconductor like silicon forms a crystal, so particular attention is paid to the behavior of electrons in a crystal. The number of charge carriers in a semiconductor (which can be both electrons, like in a metal, or holes, which don't exist in a metal) is calculated. To increase the number of carriers, we will dope the semiconductor with impurities. Next, we will consider the most important transport phenomena in semiconductors, namely drift and diffusion. Finally, we discuss generation and recombination of electron-hole pairs and establish the continuity equations.

3.1 Semiconductors

We refer to elements of the fourth column of the periodic table as semiconductors. Examples are carbon (C), Silicon (Si) and Germanium (Ge). These elements have 4 electrons on their outer shell and tend to form 4 covalent bonds with neighboring atoms to obtain a stable octet structure. As the atoms bond, they form a regular pattern known as a crystal. Silicon, the most commonly used semiconductor atom in electronics, has an external shell ($n = 3$) with a completely filled $3s^2$ orbital, and a partially filled $3p^2$ orbital. In order to form bonds, the orbitals in the exterior shell interact and undergo sp^3 hybridisation. The consequence is that the 4 covalent bonds are uniformly spaced in space, forming bonds of about 109° with each other. The left part of figure 3.1 shows a single cell of the silicon crystal structure. The blue dots represent a single silicon atom, with covalent bonds to 4 neighboring atoms. The figure on the right shows a simplified representation in 2 dimensions of the same crystal.

3.1.1 Band structure

When atoms are very far apart, they don't influence each other and we can find their wave function $\Psi(x, t)$ by solving the Schrodinger equation for a single electron in orbit around a fixed nucleus. This results in the well-known discrete energy spectrum associated with each wave function. However, as atoms come nearer, their electrons in the outer shell start to interact with each other. According to the Pauli exclusion principle, they can no longer occupy

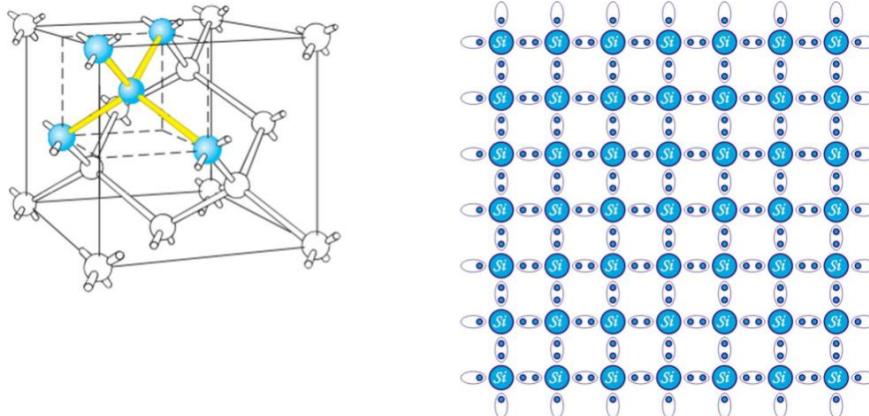


Figure 3.1: Silicon crystal (left) - schematic representation (right)

the same energy levels.

Assume that N Si atoms are initially very far apart, and we bring them closer together with the goal of forming a crystal. We have thus $2N$ electrons at the energy level of the $3s$ orbital, and $2N$ electrons in the $3p$ orbital, where there are $6N$ levels available. When the atoms are close enough, they interact and the energy levels have to shift slightly down or up to be in compliance with the Pauli exclusion principle. The levels are still discrete, but since N , the number of atoms in the crystal, is a huge number ($N \approx 10^{20}$ atoms per cm^3), we can consider the resulting energy band as a continuous spectrum (see figure 3.2). From a certain lattice spacing on, the $3s$ and $3p$ bands merge and form a single band. When the atoms come close enough together to form a crystal as in figure 3.1, we observe that the bands split again. At the interatomic distance in the crystal lattice (5.43\AA for silicon), there is a *valence band* with energies up to an energy E_V and a conduction band with energies higher than $E_C > E_V$. Between bands is a forbidden range where no electrons can exist. This range is called the *bandgap* and the difference between E_C and E_V is the *bandgap energy* E_g .

3.2 Electrons and Holes

At a temperature of 0 K, all $4N$ electrons are in the valence band. This means that they are all part of a covalent bond between two silicon atoms. However, as the temperature increases, electrons obtain more thermal energy and some electrons are able to break the covalent and become free electrons that can roam through the crystal. They have acquired enough energy to transition from the valence band to the conduction band. As a covalent bond gets broken, the electron leaves an unfilled position behind it. This free position in turn can be occupied by another electron from another electron bond. This mechanism is shown in figure ???. An open position in a bond is called a *hole*. Since electrons jumping from one hole to the other cause the hole to move in the crystal, we can consider a hole as another moving particle, with a positive charge $+q$.

There are also other mechanisms beside thermal agitation that can provide enough energy to an electron to jump from valence to conduction band, like an impinging photon with high enough energy E (thus frequency ν , since $E = \hbar\nu$). The bottom of the conduction band E_C

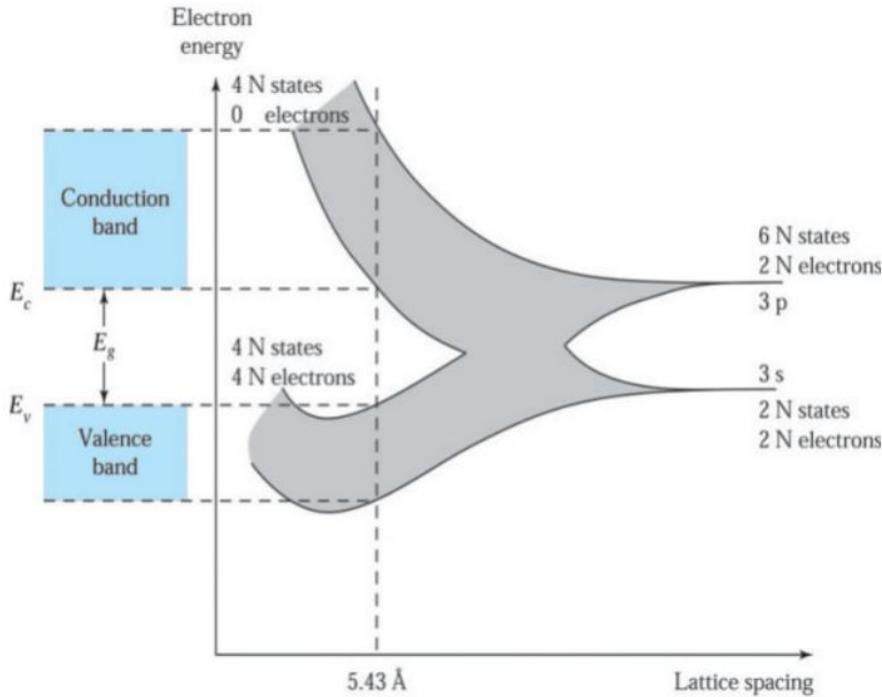


Figure 3.2: Band structure formation

corresponds to the potential energy of an electron, just like the top of the valence band E_V is the potential energy of a hole.

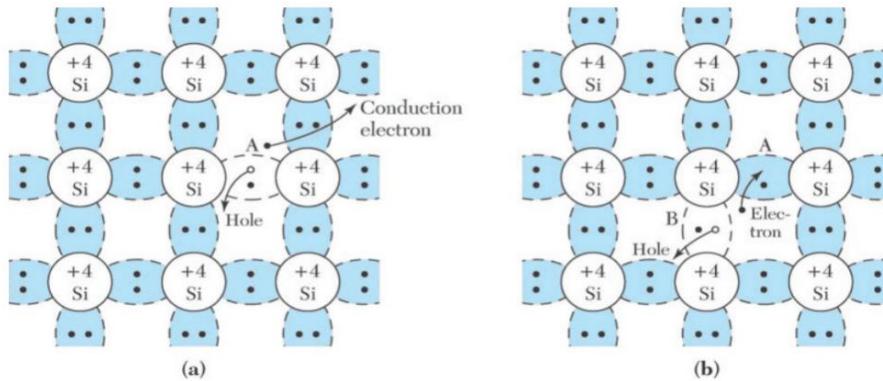


Figure 3.3: Creation of electron hole pair by breaking of covalent bonds

As we consider different directions in the crystal lattice, the band structure differs because the interatomic distance depends on the direction. The direction of propagation of a wave function is determined by its wave vector \vec{k} , and k is directly linked to the momentum p with the relation:

$$\vec{p} = \hbar \vec{k}$$

That's why these *band diagrams* are usually plotted as the energy E as function of momentum p . Figure 3.4 gives the band diagrams for silicon (left) and GaAs (right).

These figures show that for some semiconductors, like GaAs, the minimum of the conduction band is at the same direction as the maximum of the valence band. These semiconductors are called *direct bandgap* semiconductors. Others, like silicon, have different directions for these two points and are thus called *indirect bandgap* semiconductors.

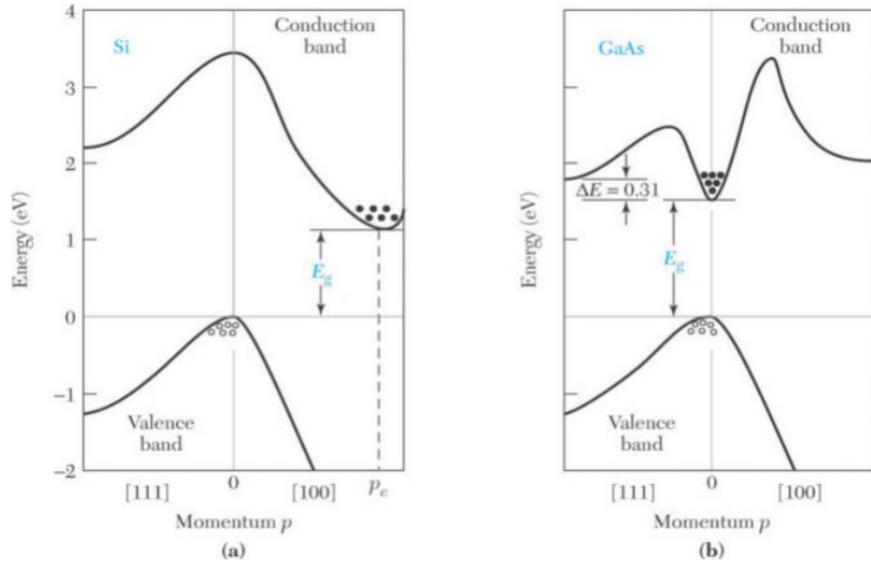


Figure 3.4: Band diagrams: (a) Si and (b) GaAs

3.2.1 Effective mass

A free electron - unencumbered by other particles and their potential functions - have a mass $m_0 = 9.11 \cdot 10^{-31}$ kg. However, within a crystal, electrons in the conduction band are influenced by the nuclei and other electrons. To determine their behavior, we should solve the Schrodinger equation for all these particles. Instead, we introduce the concept of effective mass, which can be considered as the apparent mass of a particle (electron or hole) in a crystal.

Every wave function has a group velocity $v_g = \frac{d\omega}{dk}$ and since $E = \hbar\omega$, we have $v_g = \frac{1}{\hbar} \frac{dE}{dk}$. But since the increase in energy dE of the particle can be considered the result of work dW done by a force F via the relation $dW = Fdl = Fv_g dt$ where dl is the displacement in time dt , we can simplify this relation to $\frac{dk}{dt} = \frac{F}{\hbar}$.

If a particle undergoes a change in group velocity - an acceleration a - we can state that:

$$\begin{aligned} a &= \frac{dv_g}{dt} = \frac{1}{\hbar} \frac{d}{dt} \frac{dW}{dk} \\ &= \frac{1}{\hbar} \frac{d^2W}{dk^2} \frac{dk}{dt} \\ &= \frac{1}{\hbar^2} \frac{d^2W}{dk^2} F \end{aligned} \tag{3.1}$$

This gives a relation between the applied force F and the resulting acceleration a . Hence, according to Newton's second law $F = ma$, we can interpret the proportionality constant as

the effective mass of the particle:

$$m^* = \hbar^2 \left[\frac{d^2 W}{dk^2} \right]^{-1}$$

The effective mass of an electron or hole is thus determined by the local curvature of the $E - k$ relation in the band diagram. From figure 3.4, we see that for silicon, the curvature of the conduction band is higher than that of the valence band. Hence, the effective mass m_e^* of an electron will be smaller than that of a hole m_h^* .

3.2.2 Number of carriers

We can compute the concentration of electrons and holes in pure silicon based on quantum statistics. The distribution of fermions (like electrons) is determined by the Fermi-Dirac distribution:

$$F(E) = \frac{1}{1 - e^{(E-E_F)/kT}}$$

which gives the probability that at temperature T , an electron occupies energy level E . E_F is the Fermi-level, a reference level at which the probability of occupation is exactly one-half and k is the Boltzmann constant (not to be confused with the wavenumber k).

We also consider the number of allowed states at an energy level E . We already know that between E_V and E_C , no states are available because this is the bandgap region. For respectively electrons in the conduction band and holes in the valence band, we have that:

$$\begin{aligned} N_C(E) &= \frac{4\pi}{\hbar} (2m_e^*)^{3/2} (E - E_C)^{1/2} \\ N_V(E) &= \frac{4\pi}{\hbar} (2m_h^*)^{3/2} (E_V - E)^{1/2} \end{aligned} \quad (3.2)$$

where $N_C(E)$ and $N_V(E)$ represent the state density per unit volume in conduction and valence band.

To compute the carrier density, we multiply the density of states with the probability that a state is occupied, and integrate over the relevant energy levels. We write n for the number of electrons in the conduction band, and p for the number of holes in the valence band:

$$\begin{aligned} n &= \int_{E_C}^{\infty} N_C(E) F_e(E) dE \\ p &= \int_{-\infty}^{E_V} N_V(E) F_h(E) dE \end{aligned} \quad (3.3)$$

Since both E_C and E_V are far enough removed from the Fermi level, we can simplify the Fermi-Dirac distribution and obtain a standard Boltzmann distribution:

$$\begin{aligned} F_e(E) &\approx e^{-(E-E_F)/kT} \text{ if } E - E_F \gg kT \\ F_h(E) &\approx e^{(E-E_F)/kT} \text{ if } E - E_F \ll kT \end{aligned} \quad (3.4)$$

Substituting 3.4 in 3.3 gives:

$$\begin{aligned} n &= N_C e^{-(E_C-E_F)/kT} \text{ with } N_C = \dots \\ p &= N_V e^{-(E_F-E_V)/kT} \text{ with } N_V = \dots \end{aligned} \quad (3.5)$$

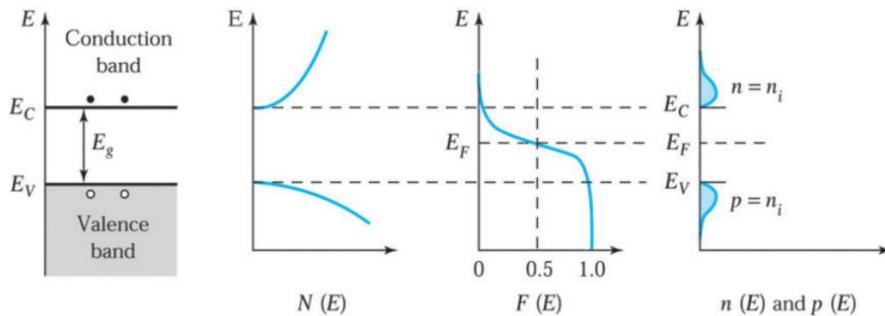


Figure 3.5: Carrier density

Figure 3.5 represents the band diagram, the state density $N(E)$, the carrier distribution $F(E)$ and the final results.

In pure silicon, every electron in the conduction band gives rise to a hole in the valence band - hence $n = p$ and E_F lies almost halfway between E_V and E_C (the small difference is due to a difference in effective mass between holes and electrons). We say that $n = p = n_i$, the intrinsic carrier density. Typically, at room temperature, $n \approx 10^{10}/cm^3$. If we multiply the equations for n and p , the Fermi level disappears:

$$np = n_i^2 = N_C e^{-(E_C - E_F)/kT} N_V e^{-(E_F - E_V)/kT} = N_C N_V e^{-E_g/kT} \quad (3.6)$$

3.3 Donor and Acceptor Impurities

The intrinsic carrier density n_i is quite low, and pure silicon is a poor conductor. However, we can increase the number of carriers (either holes or electrons) with a process called *doping*. Doping is the act of replacing some of the silicon atoms by atoms of group III (acceptor atoms) or group V (donor atoms) of the periodic table. If the replacement atoms are donor atoms (like arsenic or phosphorus), they bind with the neighbouring silicon atoms but still have an additional electron. This electron is only loosely bounded to the nucleus and can easily transition to the conduction band - without creating a hole. Similarly, atoms of group III, like boron, are an electron short to create 4 covalent bonds. If however they could snatch an electron from another bond, they can use it to obtain an octet structure and at the same time have created a hole. Both processes are schematically represented in figure 3.6. Doping also removes the dependence of majority carrier concentration on temperature.

Conventionally, we use N_d for the donor concentration and N_a for the acceptor density. They have typical values of $10^{16}/cm^3$. This is much lower than the number of atoms ($\sim 10^{20}/cm^3$) but much higher than intrinsic carrier concentration n_i ($\sim 10^{10}/cm^3$). We will refer to *n-type* or *p-type* semiconductors when we are talking about doping with donors or acceptors, respectively.

Expression 3.5 remains valid, also in a doped semiconductor. Hence the majority charge carriers (electrons in n-type, holes in p-type semiconductor) largely outnumber the minority carriers (holes in n-type, electrons in p-type). In the case of an n-type semiconductor for example, we assume that all donor atoms lose their spare electron and become ionized. Hence $n_n = N_d \approx 10^{16}/cm^3$. Consequently, $p_n = n_i^2/n_n \approx 10^4/cm^3$. Mind the subscript n to indicate that it is an n-type semiconductor.

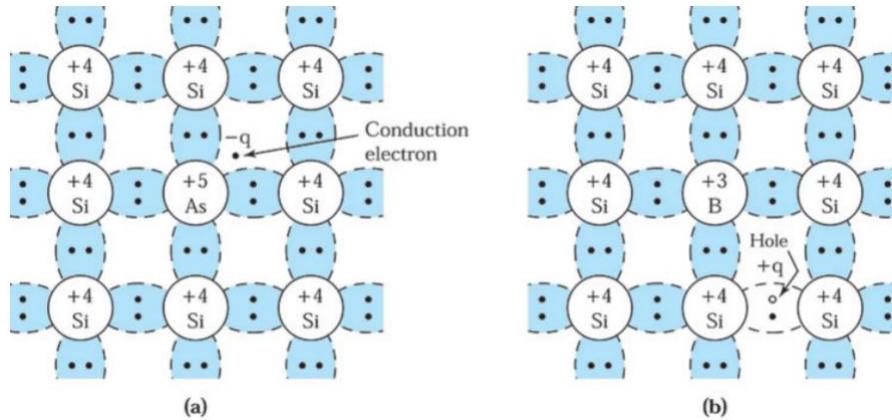


Figure 3.6: Doping with (a) donor atoms and (b) acceptor atoms

The Fermi-level for doped (or *extrinsic*) semiconductors is no longer located halfway between valence and conduction band. Since:

$$n_n = N_d = N_C e^{-(E_C - E_{Fn})/kT} \Rightarrow E_{Fn} = E_C - kT \ln \frac{N_d}{N_C}$$

and we also know that $E_i = E_C - kT \ln \frac{n_i}{N_C}$ for an intrinsic semiconductor. Thus:

$$E_{Fn} = E_i + kT \ln \frac{N_d}{n_i}$$

and

$$E_{Fp} = E_i - kT \ln \frac{N_a}{n_i}$$

This means that in an n-type semiconductor, the Fermi level lies above the intrinsic Fermi level, while in a p-type, it lies below. Even more, the Fermi level lies closer to the conduction (valence) band if N_d (N_a) is higher. Figure 3.7 shows how the charge distribution changes in an n-type semiconductor, due to a shift of the Fermi level E_{Fn} . Also mind the surplus of electrons compared to holes, since almost all of the electrons in the conduction band come from ionization of the donor atoms.

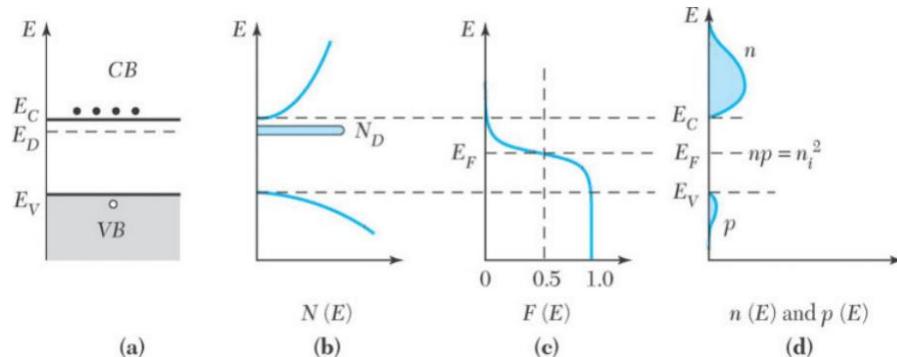


Figure 3.7: Carrier density due to donor doping in n-type semiconductor

If we transform the energies in potentials via the relation $E = -qV$ and define V as the potential difference between the intrinsic and actual Fermi level: $V = V_{Fi} - V_F$, we can rewrite these equations and obtain the *Boltzmann equations*:

$$\begin{aligned} n &= n_i e^{qV/kT} = n_i e^{V/v_{th}} \\ p &= n_i e^{-qV/kT} = n_i e^{-V/v_{th}} \end{aligned} \quad (3.7)$$

with $v_{th} = kT/q$ the thermal voltage (≈ 26 mV at $T = 300$ K). These equations are only valid in thermal equilibrium, i.e. when there is no flow of charges.

3.4 Carrier Transport

Several carrier transport mechanisms exists in a semiconductor. We discuss the 2 most common ones: *drift current*, due to the presence of an electric field, and *diffusion current*, due to a concentration gradient.

3.4.1 Drift Current

In the absence of an electric field, the charges move around in a random manner because they have thermal energy. However, since movement is random, here exists no preferential direction of flow. This changes when an external electric field is applied. In the presence of an electric field \mathcal{E} , a electron with charge $-q$ is accelerated by a force $F_1 = -q\mathcal{E}$. However, at the same time the charge interacts with the lattice, and is slowed down through to collisions with atoms and impurities. This damping force F_2 is proportional to the velocity of the particle: $F_2 = -\alpha v_d$ with α the damping factor. Thus:

$$m_e^* \frac{dv_d}{dt} = -q\mathcal{E} - \alpha v_d$$

The particle will reach an equilibrium velocity v_{d0} when $\frac{dv_d}{dt} = 0$, thus $v_{d0} = \frac{-q\mathcal{E}}{\alpha}$. We can solve the first-order equation for a particle that initially is at velocity $v_d(0) = 0$ and obtain $v_d(t) = v_{d0}e^{-t/\tau_e}$ where $\tau_e = \frac{m_e^*}{\alpha}$ a characteristic time called the *relaxation time*. It represents the time to reach v_{d0} and is of the order of 1ps.

From a macroscopic point of view, in order to compute the total electron current, we have to average over all electrons available for conduction, i.e. the electrons present in the conduction band. After some calculations, we obtain for the electron current density $J_n = q\mu_n n \mathcal{E}$ with the *electron mobility* $\mu_n = \frac{q\tau_e}{m_e^*}$. A similar expression can be found from the hole current density J_p . The total current is the sum of both:

$$J = J_n + J_p = q(\mu_n n + \mu_p p)\mathcal{E} = \sigma \mathcal{E} \quad (3.8)$$

which is the formulation of Ohm's law for a semiconductor.

3.4.2 Band structure under biasing

We consider conduction in a homogeneous semiconductor material due to an electric field. Figure 3.8 shows an n-type semiconductor and its band diagram at thermal equilibrium (left) and the band diagram when a positive biasing voltage is applied to the right-hand terminal

(right). When an electric field E is applied to a semiconductor, each electron will experience a force $-q\mathcal{E}$ from the field. The force is equal to the negative gradient of potential energy:

$$-q\mathcal{E} = \frac{dE_C}{dx}$$

Since we are interested in the gradient of the potential energy, we can use any part of the band diagram that is parallel to E_C . It is convenient to use the intrinsic Fermi level E_i because we shall use E_i when we consider p-n junctions. Thus:

$$\mathcal{E} = -\frac{1}{q} \frac{dE_C}{dx} = -\frac{1}{q} \frac{dE_i}{dx}$$

We can define a related quantity V as the electrostatic potential whose negative gradient equals the electric field: $\mathcal{E} = -\frac{dV}{dx}$. Comparing both equations, we get $V = -\frac{E_i}{q}$, which provides a relationship between the electrostatic potential and the potential energy of an electron. For the homogeneous semiconductor shown in figure 3.8, the potential energy and E_i , decrease linearly with distance; thus, the electric field is a constant in the negative x-direction. Its magnitude is the applied voltage divided by the sample length.

The electrons in the conduction band move to the right side. The kinetic energy corresponds

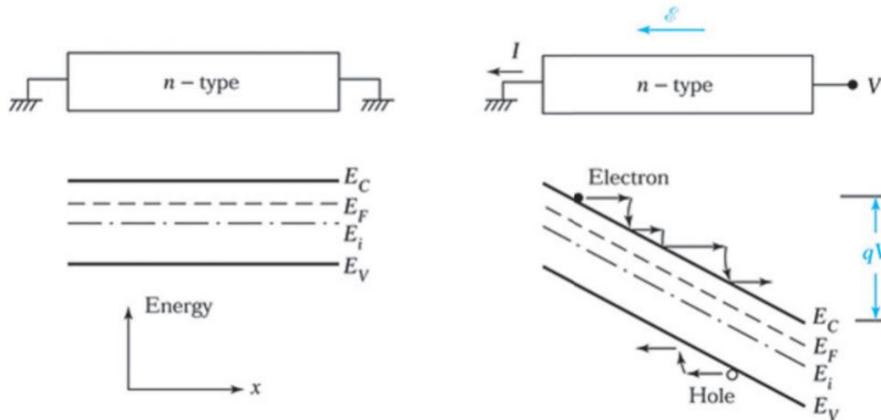


Figure 3.8: Band structure under biasing

to the distance from the band edge (i.e., E_C for electrons). When an electron undergoes a collision, it loses some or all of its kinetic energy to the lattice and drops toward its thermal equilibrium position. This is the origin of attenuation factor α of the previous section. After the electron has lost some or all its kinetic energy, it will again begin to move toward the right and the same process will be repeated many times. Conduction by holes can be visualized in a similar manner but in the opposite direction.

3.4.3 Diffusion Current

Another transport mechanism is diffusion where carriers move from one location to another due to spatial variation of carrier concentration. As carriers move randomly due to thermal agitation, more carriers will move from the higher to the lower carrier concentration than the other way around, effectively leading to a net movement of electrons from left to right as in

figure 3.9. This current is described by a standard diffusion equation:

$$J_n = qD_n \frac{dn}{dx}$$

with D_n the diffusion coefficient for electrons. An equivalent relation exists for holes:

$$J_p = -qD_p \frac{dp}{dx}$$

However, this displacement of carriers will induce an electric field \mathcal{E} and hence also a drift

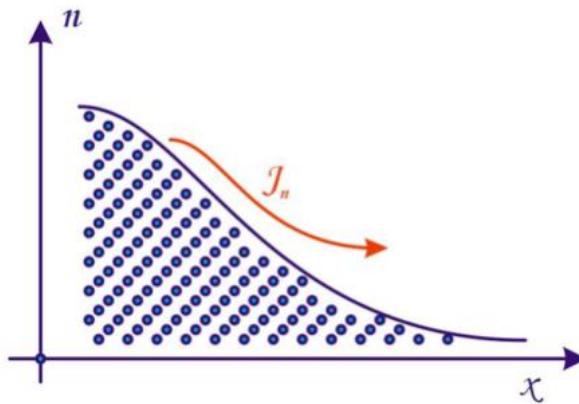


Figure 3.9: Current J_n due to diffusion

current $J = q\mu_n n \mathcal{E}$. After a while, both currents will be equal and opposite and a dynamic equilibrium will set in:

$$J_n = q\mu_n n \mathcal{E} + qD_n \frac{dn}{dx} = 0$$

Using the Boltzmann equations (3.7) we can derive a relation between D_n and μ_n :

$$D_n = \frac{kT}{q} \mu_n$$

and

$$D_p = \frac{kT}{q} \mu_p$$

These relations are known as the Einstein equations.

3.5 Generation and Recombination

In thermal equilibrium, the relation $pn = n_i^2$ is valid. This is a dynamic equilibrium: the thermal generation of electron-hole pairs at rate G_{th} is counteracted by electrons that fall back from conduction to valence band (i.e. a free electron that combines with a hole to form a covalent bond). This process is called recombination.

If excess carriers are introduced, we are no longer in equilibrium and $pn > n_i^2$. The creation of excess carriers is called *carrier injection* and can be done by optical excitation or forward-biasing a pn-junction (see chapter 4). The mechanism that restores equilibrium is the

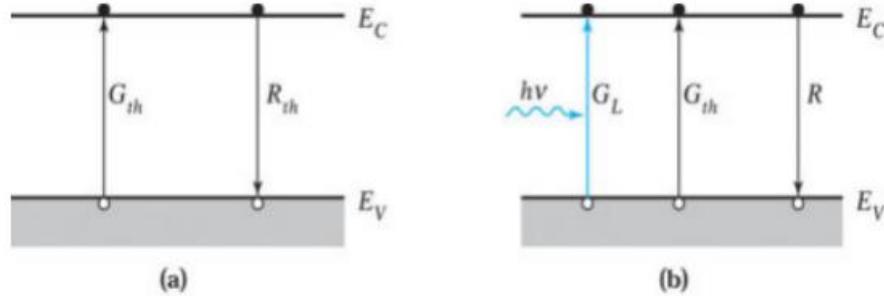


Figure 3.10: Generation and recombination under (a) equilibrium and (b) optical excitation

recombination of the injected minority carriers with the majority carriers present in the semiconductor. Both thermal and excess carrier generation and recombination are represented in We assume that excess carriers lead to recombination at rate $R(n, p, T)$. In equilibrium is $G_{th}(n_0, p_0, T) = R(n_0, p_0, T) = R_{th}$. When we create excess carriers, we can write that $n = n_0 + \delta n$ and $p = p_0 + \delta p$ and we will assume that the injection is small compared to equilibrium conditions. This means that $\delta n \ll n_0$ and $\delta p \ll p_0$. Under these conditions, we can approximate the recombination rate by the first-order terms:

$$\begin{aligned} R(n, p, T) &= R(n_0, p_0, T) + (n - n_0) \frac{\partial R}{\partial n} + (p - p_0) \frac{\partial R}{\partial p} \\ &= G_{th} + \frac{(n - n_0)}{\tau_n} + \frac{(p - p_0)}{\tau_p} \end{aligned} \quad (3.9)$$

In an n-type material, the recombination rate is determined by the excess holes, since holes are much more likely to find a candidate for recombination (i.e. an electron) than by the excess electrons. The rate at which the carriers recombine is thus determined by the number of minority carriers:

- In n-type: $R - G_{th} = \frac{(p - p_0)}{\tau_p}$
- In p-type: $R - G_{th} = \frac{(n - n_0)}{\tau_n}$

The lifetime of the excess minority carriers τ_p and τ_n are a lot larger than the relaxation time τ_e .

3.6 The continuity equations

The change of carrier density in a volume V within a semiconductor can due to three causes:

1. generation of carriers,
2. recombination,
3. flow of carriers in or out of the volume the surrounding surface S

Assume we want to study the change of electron density in a p-type material. We can write:

$$\begin{aligned} -q \iiint_V \frac{\partial n}{\partial t} dV &= -q \iiint_V (G - R) dV - \iint_S \vec{J}_n \vec{n} dS \\ &= -q \iiint_V (G - R) dV - \iiint_V \nabla \vec{J}_n dV \end{aligned} \quad (3.10)$$

where we used the divergence theorem. Since this is valid for any volume, the terms within the integrals must be equal:

$$\begin{aligned} -q \frac{\partial n}{\partial t} &= -q(G - R) - \nabla \vec{J}_n \\ &= -q(G_{th} + g - R) - \nabla \vec{J}_n \\ \Rightarrow \frac{\partial n}{\partial t} &= \left(\frac{n - n_0}{\tau_n} \right) + \frac{1}{q} \nabla \vec{J}_n + g \end{aligned} \quad (3.11)$$

where we replace $G_{th} - R$ with $\frac{(n-n_0)}{\tau_n}$ since we are in a p-type material. In one dimension, this becomes:

$$\frac{\partial n}{\partial t} = \left(\frac{n - n_0}{\tau_n} \right) + \frac{1}{q} \frac{\partial J_n}{\partial x} + g \quad (3.12)$$

where J_n in general has a drift and a diffusion contribution: $J_n = q\mu_n n \mathcal{E} + qD_n \frac{dn}{dx}$. Substituting this in 3.12 - with both n and \mathcal{E} depending on x - gives:

$$\frac{\partial n}{\partial t} = \left(\frac{n - n_0}{\tau_n} \right) - n\mu_n \frac{\partial \mathcal{E}}{\partial x} + \mu_n \mathcal{E} \frac{\partial n}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2} + g \quad (3.13)$$

This equation is called the *continuity equation* for n-type carriers in a p-type material. Similar expressions can be found for the other cases.

Chapter 4

The pn-junction

In this section we demonstrate that when a junction is formed between a sample of *p*-type and one of *n*-type semiconductor, this combination possesses the properties of a rectifier. This two-terminal device is called a *junction diode*. The physical structure and symbol of a pn-junction is shown in figure 4.1

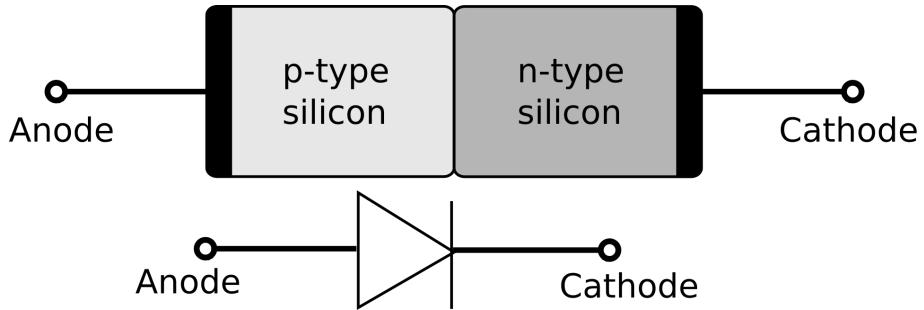


Figure 4.1: Structure (above) and symbol (below) of a pn-junction

We will discuss the junction in equilibrium and under forward and reverse bias. The I-V characteristic will be derived and the concepts of avalanche breakdown and depletion capacitance will be discussed.

4.1 The pn-junction in equilibrium

If donor impurities are introduced into one side and acceptors in the other side of a single crystal of a semiconductor, a *pn*-junction is formed¹. At the interface, there is by construction a concentration gradient and hence electrons and holes will diffuse to either side. Electrons from the n-type will recombine with the majority holes in the p-type and the holes from the p-type will recombine with the majority electrons in the n-type. As a consequence, the doping ions will be exposed and the n-side will have a fixed positive charge density (N_d) while the p-side will have a fixed negative charge density (N_a) (figure 4.2 (a)). This charge buildup will lead to an internal electric field pointing from n-type to p-type and it will thus act against any further diffusion current (4.2 (b)). We assume that no external bias is applied, so that

¹This is called an *abrupt junction*

no net current can flow. The zone where the majority carriers have diffused and recombined is the *space-charge region*². To preserve charge neutrality, we require that:

$$N_A x_p = N_d x_n$$

with x_p and x_n the depth of the space-charge region in p- and n-type (W_{Dp} and W_{Dn} in figure 4.2). As can be deduced from this expression, the space-charge region extends further in the lightly-doped material.

Because $\frac{d\mathcal{E}}{dx} = \rho/\epsilon$, the electrical field that results from this charge buildup is computed as $\mathcal{E}(x) = \int \rho(x)/\epsilon dx$ with $\rho(x)$ the local charge profile:

$$\begin{aligned} \rho(x) &= -N_a \text{ if } x < 0 \\ &= N_d \text{ if } x > 0 \end{aligned} \quad (4.1)$$

The maximum electric field corresponds to the total charge buildup and is negative since the p-type is placed left: $|\mathcal{E}_m| = \epsilon N_a x_p = \epsilon N_d x_n$. An electric field gives rise to a potential difference since $\mathcal{E} = -\frac{d\psi}{dx}$ as in figure 4.2(c) where the so-called built-in potential ψ_{bi} is equal to the surface under $\mathcal{E}(x)$:

$$\psi_{bi} = \frac{1}{2}(x_n + x_p)|\mathcal{E}_m| = \frac{1}{2}\epsilon(x_n + x_p)N_a x_p$$

This built-in potential is also visible in 4.2 (d) where the energy bands are shown. Due to electric field, the bands will bend (in opposite direction as for the potential, since $E = -qV$) at the junction. This last figure is important and can also be constructed by reasoning with the Fermi levels.

The total width $W = x_n + x_p$ of the space charge region can be computed as function of the doping levels and the built-in voltage. The result is:

$$W = \sqrt{\frac{2\epsilon}{q} \left(\frac{N_a + N_d}{N_a N_d} \right) V_{bi}} \quad (4.2)$$

4.1.1 Fermi-levels in equilibrium

Since no net current can flow in the junction with no external bias, the drift current must be equal to the diffusion current. For the holes, this condition gives:

$$\begin{aligned} J_p &= J_{p,drift} + J_{p,diffusion} \\ &= q\mu_p p\mathcal{E} - qD_p \frac{dp}{dx} \\ &= q\mu_p p \left(\frac{1}{q} \frac{dE_i}{dx} \right) - kT\mu_p \frac{dp}{dx} = 0 \end{aligned} \quad (4.3)$$

In equilibrium we can apply the Boltzmann equations:

$$p = n_i e^{(E_i - E_F)/kT} \quad (4.4)$$

²Also called the depletion region

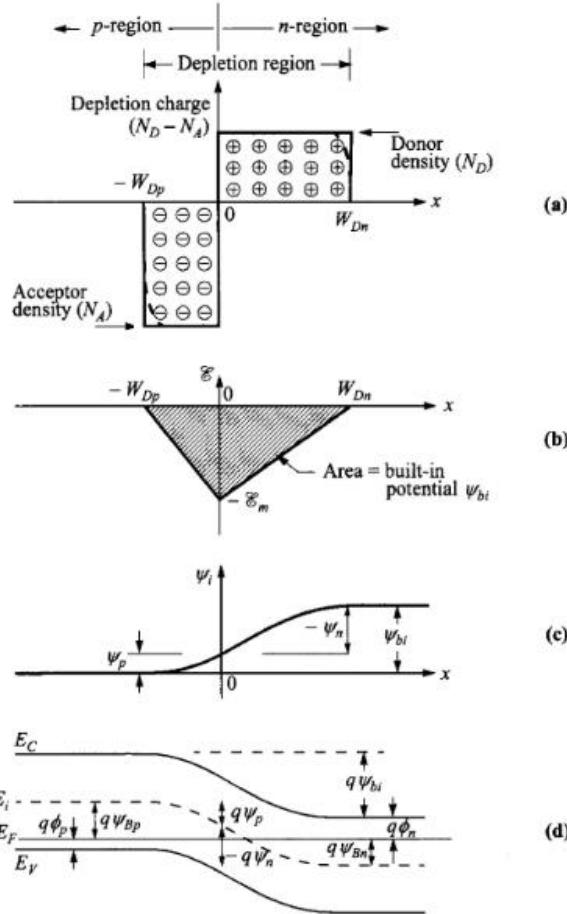


Figure 4.2: pn-junction in thermal equilibrium

and compute $\frac{dp}{dx}$:

$$\frac{dp}{dx} = \frac{p}{kT} \left(\frac{dE_i}{dx} - \frac{dE_F}{dx} \right)$$

Substituting this in equation 4.3 gives:

$$J_p = \mu_p p \frac{dE_F}{dx} = 0$$

or $\frac{dE_F}{dx} = 0$. A similar result is valid for the electrons. **For there to be zero net electron and hole currents, the Fermi level must be constant.**

As can be seen in figure 4.2(d), the Fermi level E_F remains constant along the junction because there is no net current. Since E_F is close to E_V in the p-type and close the E_C in the n-type, the bands must bend like they do in the figure to keep a constant E_F .

4.1.2 The built-in potential

We will now compute $V_{bi} = \psi_n - \psi_p$ (previously denoted as ψ_{bi}). We know that far away from the junction, no net charges are present, so the potential must comply with the Laplace

equation:

$$\frac{d^2\psi}{dx^2} = 0$$

and $N_d - N_a + p - n = 0$ to preserve charge neutrality. In a p-type material, we assume that $N_d = 0$ and $p \gg n$. This results in $p = N_a$. Inserting this in equation 4.4 gives:

$$\psi_p = -\frac{1}{q}(E_i - E_f)|_{x < x_p} = \frac{kT}{q} \ln \left(\frac{N_a}{n_i} \right)$$

Similarly, the electrostatic potential for the n-type material is

$$\psi_n = -\frac{1}{q}(E_i - E_f)|_{x > x_n} = \frac{kT}{q} \ln \left(\frac{N_d}{n_i} \right)$$

The built-in potential V_{bi} is the difference of electrostatic potential between p-side and n-side at thermal equilibrium:

$$V_{bi} = \psi_n - \psi_p = \frac{kT}{q} \ln \left(\frac{N_a N_d}{n_i^2} \right) \quad (4.5)$$

4.2 The pn-junction under bias

4.2.1 Forward bias

Assume we apply a positive voltage V_F to the anode (p-side) while keeping the cathode (n-side) at ground. The effect is that we reduce the barrier of the built-in potential from V_{bi} to $V_{bi} - V_F$, as shown on the left side of figure 4.3. This has a couple of consequences:

- The width of the depletion region will reduce, because equation 4.2 can be rewritten as:

$$W = \sqrt{\frac{2\epsilon}{q} \left(\frac{N_a + N_d}{N_a N_d} \right) (V_{bi} - V_F)} \quad (4.6)$$

- The value of the maximum \mathcal{E} is reduced. This means that there is an disequilibrium between diffusion and drift current, and there will be a net (diffusion) flow of holes from p-type to n-type and a flow of electrons from n-side to p-side. The result is a net current from p-side to n-side.³

There is thus a diffusion current of majority carriers through the depletion region to the other side, where they become the minority carriers. Since there are a lot of majority carriers available, this current can become very large when the potential barrier is lowered enough. Also notice that the Fermi levels in figure 4.3 are no longer constant, because there is a current and we are no longer in equilibrium.

4.2.2 Reverse - bias

When we apply a positive voltage V_R to the cathode (n-side) while keeping the anode at ground, we increase the barrier of the built-in potential to $V_{bi} + V_R$. This situation is sketched on the right of figure 4.3. According to equation 4.6, the width of the space-charge

³Notice that electrons can diffuse from low to higher energy, while they always drift from high to low.

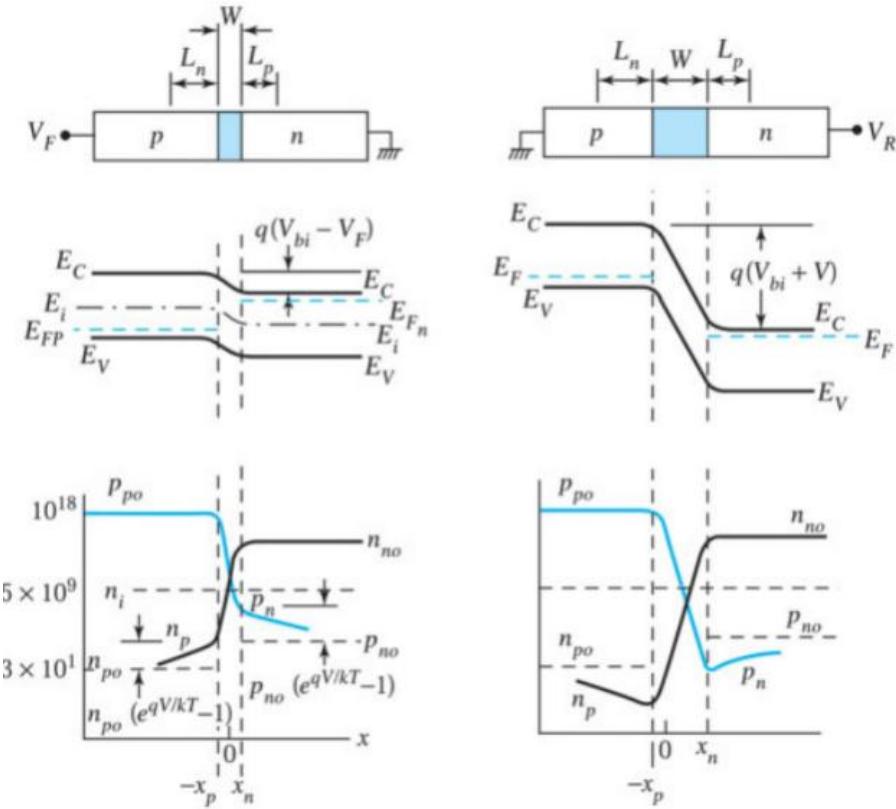


Figure 4.3: pn-junction under forward (left) and reverse (right) bias

region will increase (because we can replace V_F with $-V_R$). The internal electric field at the junction becomes larger than before, so there can be a net drift current. However, the only available carriers to be swept across the space-charge region by the electric field are the minority carriers on both sides of the junction. Consequently, the resulting current (the *saturation current*) will be very low.

4.2.3 Diode Characteristic

Let's compute the currents. We'll make the following assumptions:

1. the depletion region has abrupt boundaries and, outside the boundaries, the semiconductor is assumed to be neutral;
2. the carrier densities at the boundaries are related by the electrostatic potential difference across the junction;
3. the injected minority carrier densities are small compared to the majority carrier densities;
4. neither generation nor recombination current exists in the depletion region and the electron and hole currents are constant throughout the depletion region.

At equilibrium, we get $p_{p0} = N_a$ and $n_{n0} = N_d$. Together with the mass-action law $p_{p0}n_{n0} = n_i^2$, we can rewrite equation 4.5 as:

$$V_{bi} = \frac{kT}{q} \ln \frac{p_{p0}n_{n0}}{n_i^2} = \frac{kT}{q} \ln \frac{n_{n0}}{n_{p0}}$$

Rearranging this gives:

$$n_{n0} = n_{p0} e^{qV_{bi}/kT} \quad (4.7)$$

and

$$n_{n0} = n_{p0} e^{qV_{bi}/kT} \quad (4.8)$$

Because of the second assumption, these equations remain valid when we change the net potential. Thus:

$$n_n = n_p e^{q(V_{bi} - V_F)/kT} \quad (4.9)$$

with n_n and n_p the non-equilibrium electron densities at the boundaries of the space-charge region at n- and p-sides, respectively. Substituting 4.7 in 4.9 yields the electron density at the boundary of the depletion region on the p-side ($x = -x_p$):

$$n_p = n_{p0} e^{qV/kT} \quad (4.10)$$

and similarly:

$$p_n = p_{n0} e^{qV/kT} \quad (4.11)$$

where V can be both V_F or V_R , namely the externally applied voltage across the junction. We can also write this as:

$$\begin{aligned} n_p - n_{p0} &= n_{p0}(e^{qV/kT} - 1) \\ p_n - p_{n0} &= p_{n0}(e^{qV/kT} - 1) \end{aligned} \quad (4.12)$$

Note that the minority carriers at the boundaries of the space-charge region increase substantially above their equilibrium under forward bias. Hence, there is an injection of minority carriers at the depletion region.

In the neutral n-region, there is no electric field, so the steady-state continuity equation 3.13 reduces to:

$$\frac{\partial p_n}{\partial t} = D_p \frac{d^2 p}{dx^2} - \frac{p_n - p_{n0}}{\tau_p} = 0 \quad (4.13)$$

Solving this equation with boundary conditions of eq. 4.12 and $p_n(x = \infty) = p_{n0}$ gives:

$$p_n - p_{n0} = p_{n0}(e^{qV/kT} - 1)e^{-(x-x_n)/L_p} \quad (4.14)$$

with $L_p = \sqrt{D_p \tau_p}$ the diffusion length of holes. This graph is shown in the lower left part of figure 4.3. At the boundary $x = x_n$:

$$J_p(x_n) = -qD_p \frac{dp_n}{dx} \Big|_{x=x_n} = \frac{qD_p p_{n0}}{L_p} (e^{qV/kT} - 1) \quad (4.15)$$

By applying the same reasoning for the n-region, we obtain a similar relation for J_n . Since the total current is the sum of both, we finally find:

$$J = J_p(x_n) + J_n(-x_p) = J_S(e^{qV/kT} - 1) \quad (4.16)$$

with J_S the saturation current:

$$J_S = \frac{qD_n p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n} \quad (4.17)$$

Equation 4.16 is the diode equation. Its graph is shown in figure 4.4. It is important to notice that the current increases exponentially when $V > 0$ because the potential barrier is removed. The junction will act as a conductor. On the other hand, when $V < 0$, there is only a small saturation current that is not impacted by the value of V . The junction is an open circuit (with a loss current).

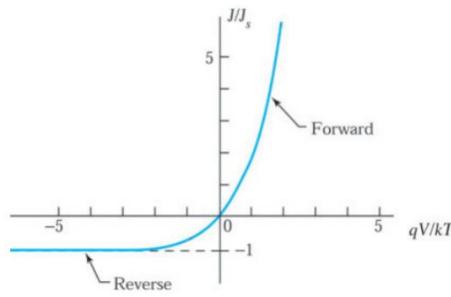


Figure 4.4: Current characteristic of equation 4.16

4.2.4 Practical Diode Characteristic

Equation 4.16 gives a current density. By multiplying with the surface A of the cross-section of the junction, we obtain the I-V curve:

$$i_D = I_S(e^{v_D/v_{th}} - 1) \quad (4.18)$$

with $v_{th} = \frac{kT}{q} \approx 26mV$ (at $T = 300K$) the thermal voltage (see figure 4.5). When $v_D \gg v_{th}$:

$$i_D \approx I_S e^{v_D/v_{th}} \quad (4.19)$$

Furthermore, as v_D increases by $\sim 60mV$, the current i_D is multiplied by a factor 10. As can be seen in figure 4.5, we can consider $v_D = 0.6V$ as a threshold voltage:

- If $v_D > 0.6V$, the diode will conduct,
- If $v_D < 0.6V$, the diode will not conduct.

Hence the diode works as a rectifier: only current from p- to n-side can pass, while the other direction is blocked. Of course, because of the saturation current and the avalanche effect (see 4.3) this is not completely true.

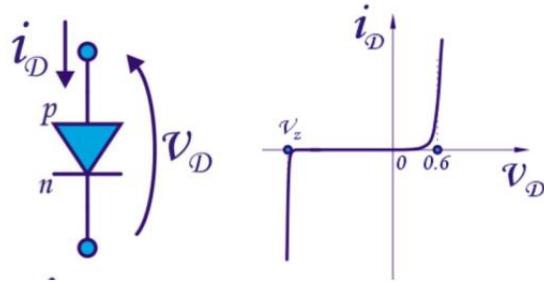


Figure 4.5: Symbol and I-V curve

4.3 Avalanche effect

TODO

4.4 Depletion Capacitance

When we reverse-bias a pn-junction, more positive charges appear on the n-side and more negative charge on the p-side. Thus, the device basically operates as a *capacitor*. In essence, we can view the conductive n and p sections as the two plates of the capacitor. We also assume the charge in the depletion region equivalently resides on each plate.

But there is more: as V_R increases, so does the width of the depletion region. That is, the capacitance of the structure decreases as the two plates move away from each other. The junction therefore displays a voltage-dependent capacitance. We can show that the junction capacitance C_j is given by:

$$C_j = \frac{C_{j0}}{\sqrt{1 - \frac{V_R}{V_{bi}}}} \quad (4.20)$$

with C_{j0} the capacitance under zero external bias ($V_R = 0$).

A pn-junction under reverse bias is thus a voltage-controllable capacitor. This kind of device has many uses, like e.g. in frequency-tunable circuits.

Chapter 5

Transistors

A transistor is a semiconductor device used to amplify or switch electrical signals and power. It is a component with (at least) three terminals. A voltage or current applied to one of the terminals controls the current through another pair of terminals. Because the controlled (output) power can be higher than the controlling (input) power, a transistor can amplify a signal.

We will discuss the two most prevalent transistors: the Bipolar Junction Transistor (BJT) and the Metal-Oxide-Semiconductor Field Effect transistor (MOSFET).

5.1 Bipolar Junction Transistor

A bipolar junction transistor (BJT) is formed by placing a n-type semiconductor between two p-type semiconductors (pnp) - or vice versa (npn), as shown in figure 5.1. In the case of the pnp transistor, the heavily doped p^+ -region is called the *emitter* (E). The narrow central *n*-region is the *base* (B). Its width is small compared with the diffusion length of the minority carriers. The lightly doped *p*-region is the *collector* (C). Figure 5.1 also shows the circuit symbols for the two BJT transistors, together with the conventional directions for voltages and currents.

We will discuss the pnp-transistor in some detail; the reasoning for the npn-transistor is similar and left as an exercise.

5.1.1 Operation in Active Mode

First consider the pnp-transistor with the three terminals (emitter, base and collector) connected to ground as in figure 5.2 (a). Under these circumstances we can construct the charge profile like we did for the pn-junction. Notice that the space-charge regions at both junctions extend deeper in the lower doped regions (base in E-B junction, collector in B-C junction).

As for the pn-junction, the electric field that emerges in the space-charge region is such that it is in equilibrium with the diffusion current due to the concentration gradients at the junctions. Figure 5.2(d) shows the band diagram in equilibrium. Since there is no current, the Fermi-level E_F is constant in the three regions.

We will polarize the transistor in the so-called active mode. This is the most commonly used mode in practice. We do this by applying a positive voltage $V_{EB} > 0$ between emitter and base, and a negative voltage $V_{CB} < 0$ between collector and base. The pn-junction between

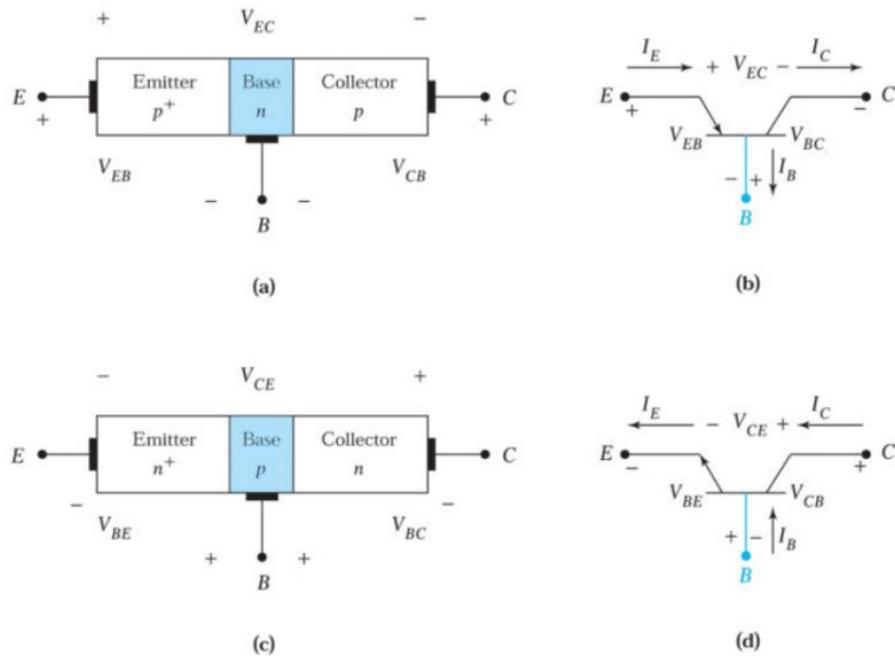


Figure 5.1: (a) Idealized one-dimensional schematic of a p-n-p bipolar transistor and (b) its circuit symbol. (c) Idealized one-dimensional schematic of an n-p-n bipolar transistor and (d) its circuit symbol.

emitter and base is thus forward-biased, while the junction between collector and base is reversed biased.

Since we apply a positive voltage to the emitter and a negative voltage to the collector (while keeping the base grounded; this is a *common-base* configuration), we lower the energy bands in the emitter and raise them in the collector (just as in the pn-junction) - see figure 5.3(d). As a consequence, the potential barrier between emitter and base is lowered so majority carriers of the emitter side (i.e. holes) can diffuse through the space-charge region to the base. If the base would be a lot larger than the diffusion length, almost all injected holes would recombine with electrons in the base and generate a base current, just like in an ordinary pn-junction. However, since the base is small, most injected holes do not recombine but diffuse into the space-charge region between base and collector. There they are swept by the electrical field to the collector. As a consequence, most holes leaving the emitter end up in the collector. This phenomenon is called the *transistor action*. It is important to realize that the collector depends on the emitter current and thus on the height of the emitter-base barrier and V_{EB} , but not on the base-collector voltage V_{CB} (as long as the base-collector junction is reversed biased).

5.1.2 Currents in Active Mode

For each terminal, we can identify the carrier flows that contribute to the currents I_E , I_C and I_B .

- ### 1. At the emitter:

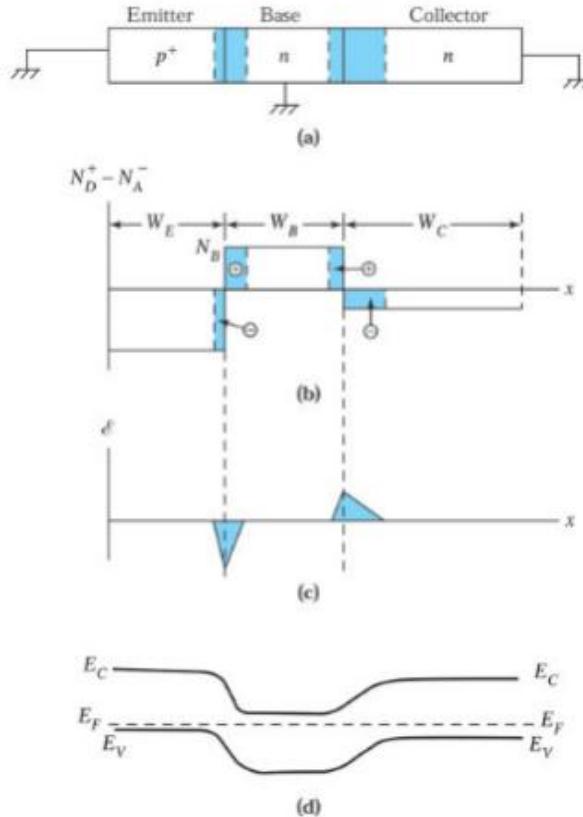


Figure 5.2: (a) pnp transistor with all leads grounded (b) Charge profile in equilibrium (c) Electric field (d) Energy band diagram.

- a hole current from emitter to the base I_{Ep}
- an electron flow from base to emitter I_{En}

Thus $I_E = I_{Ep} + I_{En}$

2. At the base:

- a current I_{BB} due to recombination of injected holes with electrons (which have to be resupplied by the battery). This current equals the difference between I_{Ep} and I_{Cp} : $I_{BB} = I_{Ep} - I_{Cp}$
- an electron flow from base to emitter I_{En}
- an electron flow from collector to base: I_{Cn} (i.e. the leakage current from the reversed biased B-C junction)

Thus $I_B = I_{En} + (I_{Ep} - I_{Cp}) - I_{Cn}$

3. At the collector:

- a current I_{Cp} that is what remains of I_{Ep} after transition through the base.

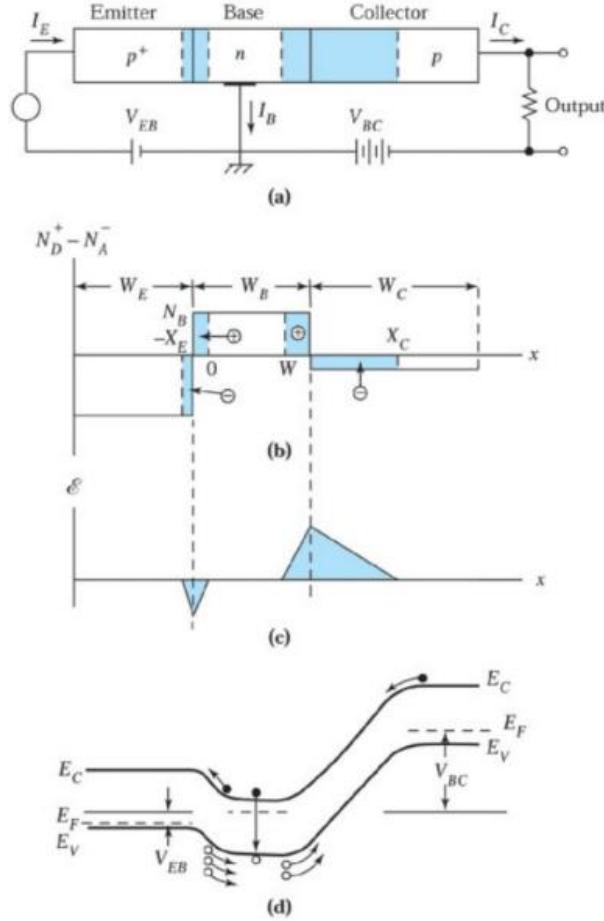


Figure 5.3: pnp transistor in active mode

- an electron flow from collector to base through the reversed biased B-C junction: I_{Cn}

$$\text{Thus } I_C = I_{Cp} + I_{Cn}$$

These currents are represented in figure 5.4. As the figure implies, I_{Ep} is the major current in the device.

We characterize the transistor by the *common-base current gain* α_0 , which is the ratio between I_{Cp} and the emitter current:

$$\alpha_0 \equiv \frac{I_{Cp}}{I_E}$$

We can rewrite this as:

$$\alpha_0 = \frac{I_{Cp}}{I_{En} + I_{Ep}} = \left(\frac{I_{Ep}}{I_{En} + I_{Ep}} \right) \left(\frac{I_{Cp}}{I_{Ep}} \right) = \alpha_T \gamma$$

with α_T the *base transfer factor* and γ the *emitter efficiency*. For proper operation, we require that $I_{Ep} \gg I_{En}$, thus that $\alpha_T \approx 1$. This can be accomplished by requiring that the

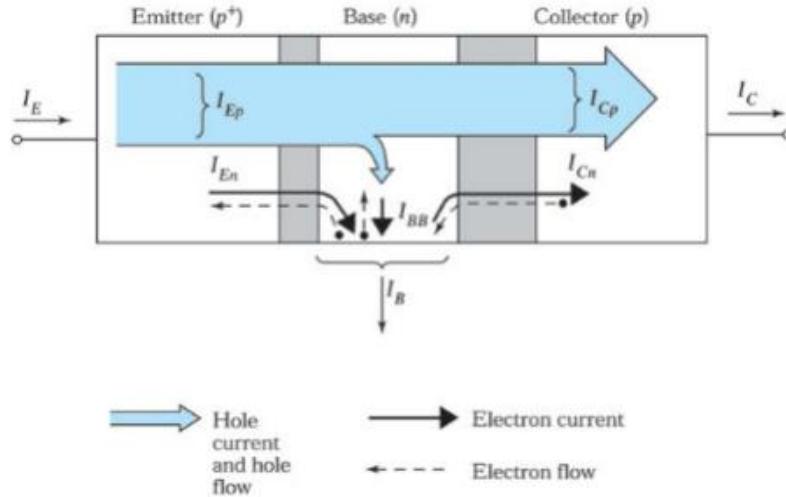


Figure 5.4: Currents in active mode

emitter doping level be much greater than that of the base. The factor γ can be increased by decreasing the length of the base.

We can rewrite the collector current I_C as:

$$\begin{aligned} I_C &= I_{Cp} + I_{Cn} = \alpha_T I_{Ep} = \alpha_0 I_E + I_{Cn} \\ &= \alpha_0 I_E + I_{CBO} \end{aligned} \quad (5.1)$$

where I_{CBO} is the leakage current between collector and base with the emitter-base junction open. This expresses that the collector current is a fraction ($0 < \alpha_0 < 1$) of the emitter current, plus a leakage current.

5.1.3 Carrier distribution in Active Mode

In order to compute the different currents, we first need to determine the carrier distribution in each region. We will assume the following:

1. Uniform doping in each region
2. No hole drift current in base
3. Collector saturation current is negligible
4. Only low-level injection
5. No generation-recombination in the depletion zone
6. No series resistance in the device

We will assume that the $x = 0$ corresponds to the end of the emitter-base depletion zone, and $x = W$ to the start of the base-collector depletion zone. Starting from the continuity equation for minority carriers in the base region:

$$\frac{dp_n}{dt} = -\frac{(p_n - p_{n0})}{t_p} - \frac{1}{q} \frac{dJ_p}{dx} + g \text{ with } J_p = q\mu_p pE + qD_p \frac{dp_n}{dx}$$

If we assume steady-state, no electric field and no other generation mechanisms, the equation reduces to:

$$D_p \frac{d^2 p_n}{dx^2} = \frac{(p_n - p_{n0})}{t_p}$$

This equation has as general solution

$$p_n(x) = p_{n0} + C_1 e^{x/L_p} + C_2 e^{-x/L_p}$$

with $L_p = \sqrt{D_p t_p}$ the so-called *diffusion length* of the holes in the n-type semiconductor. Constants C_1 and C_2 can be determined by the boundary conditions:

- $p_n(0) = p_{n0} e^{qV_{EB}/kT}$ because this is the standard concentration at the boundary of the depletion region, as in equation 4.11 of chapter 4,
- $p_n(W) = 0$ because all holes at $x = W$ will be swept through the base-collector space-charge region by the induced electric field.

Solving for C_1 and C_2 leads to a rather complex distribution. However, if $W/L_p \ll 1$ we can simplify and obtain:

$$p_n(x) = p_{n0} e^{qV_{EB}/kT} \left(1 - \frac{x}{W}\right)$$

The minority carrier concentration decreases thus linearly in the base. In a similar manner, we can find an expression for the minority carriers (electrons) in emitter and collector. The results are:

$$\begin{aligned} n_E(x) &= n_{E0} + n_{E0}(e^{qV_{EB}/kT} - 1)e^{(x+x_E)/L_E} \\ n_C(x) &= n_{C0} - n_{C0}e^{(x-x_C)/L_C} \end{aligned} \tag{5.2}$$

These distributions are also represented in figure 5.5. Once the carrier concentrations are known, the currents are easily computed as they are proportional to the concentration gradient:

- $I_{Ep} = A \left(-q D_p \frac{dp_n}{dx} \Big|_{x=0} \right) = \frac{qAD_p p_{n0}}{W} e^{qV_{EB}/kT}$
- $I_{En} = A \left(-q D_E \frac{dn_E}{dx} \Big|_{x=-x_E} \right) = \frac{qAD_E n_{E0}}{L_E} (e^{qV_{EB}/kT} - 1)$
- ...

By combining these results, we obtain expressions for the three terminal currents:

- $I_E = a_{11}(e^{qV_{EB}/kT} - 1) + a_{12}$
- $I_C = a_{12}(e^{qV_{EB}/kT} - 1) + a_{22}$
- $I_B = I_E - I_C$

where all the a_{ij} depend on the transistor characteristics like doping concentrations and dimensions.

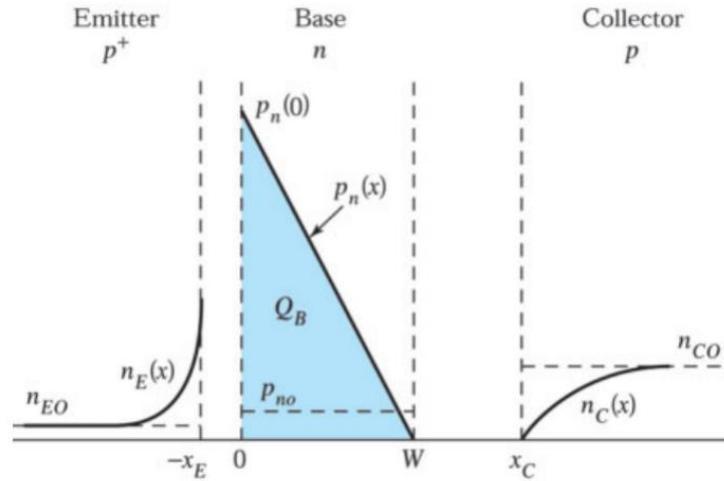


Figure 5.5: Minority carrier distribution in active mode

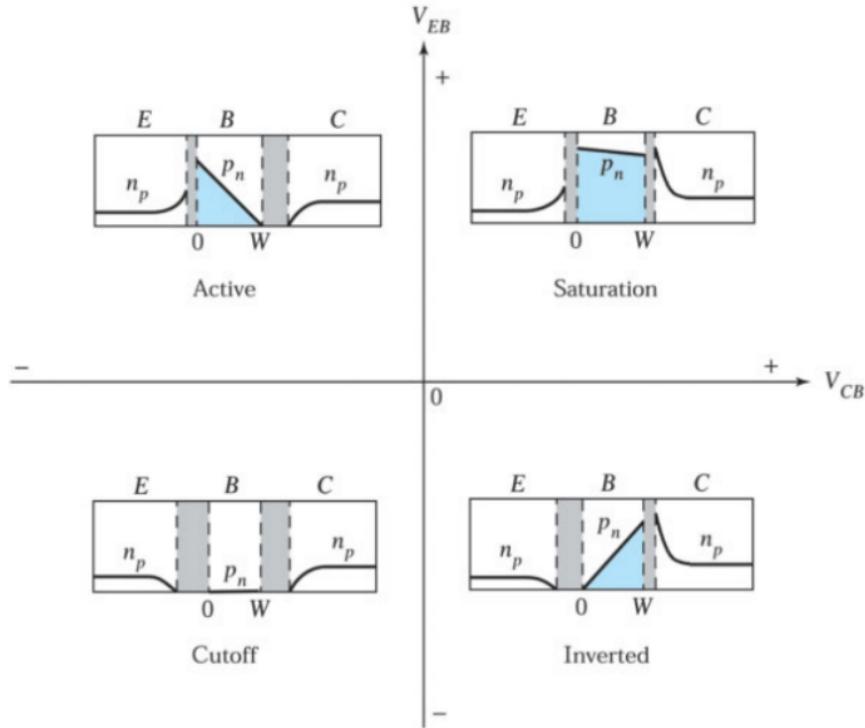


Figure 5.6: Modes of operation

5.1.4 Modes of Operation and Ebers-Moll Equations

Based on the signs of V_{EB} and V_{CB} , we can distinguish 4 different modes of operation. We already studied the case where $V_{EB} > 0$ and $V_{CB} < 0$, namely the active mode. The other modes are represented in figure 5.6, together with the minority carrier distribution in emitter, base and collector. In saturation mode, both junctions are forward biased. The minority-carrier distribution at the edge of each depletion region is not zero, as in the active mode.

Small biasing voltages lead to a large output current. The transistor acts a closed switch. In cutoff mode, both junctions are reversed biased. Only a small current will flow. The transistor is an open switch.

In inverted mode, emitter and collector are reversed. The behavior is however not exactly like in the active mode, because the doping levels in emitter and collector are different.

All these modes can be analyzed in a similar way as we have done for the active mode. This analysis leads to the *Ebers-Moll equations* which are used to model the BJT in simulators like SPICE.

TODO: Add Ebers-Moll equations and equivalent circuits

5.1.5 Common-Emitter configuration

Up until now, we have kept the base at a fixed voltage. However, the most common use of the bipolar transistor is with the emitter at a fixed voltage as in figure 5.7(a). This configuration has V_{EB} and I_B as inputs, and I_C and V_{EC} as outputs.

We can express I_C as function of I_B by substituting $I_E = I_B + I_C$ in equation 5.1:

$$I_C = \alpha_0(I_B + I_C) + I_{CBO}$$

Solving for I_C gives:

$$\begin{aligned} I_C &= \frac{\alpha_0}{1 - \alpha_0} I_B + \frac{I_{CBO}}{1 - \alpha_0} \\ &= \beta_0 I_B + I_{CEO} \end{aligned} \tag{5.3}$$

with $\beta_0 = \frac{\alpha_0}{1 - \alpha_0}$ the *common-emitter current gain* and $I_{CEO} = \frac{I_{CBO}}{1 - \alpha_0}$ the collector-emitter leakage current. This current is always present, even if $I_B = 0$ and thus makes the bipolar transistor a very bad switch because it will leak current when closed.

In the previous, common-base configuration, and with $\alpha_0 \approx 1$, a change in emitter current I_E produces a changes of approximately the same amount in the collector current I_C and a much smaller change in the base current (factor of $1 - \alpha_0$). To achieve current amplification, the change is initiated in the base current rather than in the emitter current. This causes the collector current to change by a factor of $\frac{\alpha_0}{1 - \alpha_0} = \beta_0$.

When α_0 is close to one, β_0 becomes very large. This is a good thing because it allows us to amplify the base current I_B . However, β_0 can change a lot from transistor to transistor. This is why we will explore polarisation techniques that mitigate the variability of β_0 .

The defining I-V characteristic of a common-emitter configuration is shown in 5.7(b). If V_{EC} is too low, V_{CB} is positive and the base-collector junction is not reversed biased. The transistor in saturation mode. Thus V_{EC} must be higher than some threshold to put the transistor in the active mode. This value is called $V_{EC,Sat}$ and is about 0.2V.

When $V_{EC} > V_{EC,Sat}$, we can say based on equation 5.3 that $I_C \approx \beta_0 I_B$ and hence independent of V_{EC} . This is the ideal behavior, because we now control output current I_C with input current I_B and not with the output voltage V_{EB} . However, as can be seen from the figure, the I-V characteristics are not entirely flat and thus still depend on V_{EB} . This is due to the *Early-effect* and will be addressed in the next section.

If V_{EB} is lower than 0.6V, the emitter-base junction is not forward biased and I_B is very small. We say that the transistor is in cut-off. The only current that still flows from emitter to collector is the leakage current I_{CEO} .

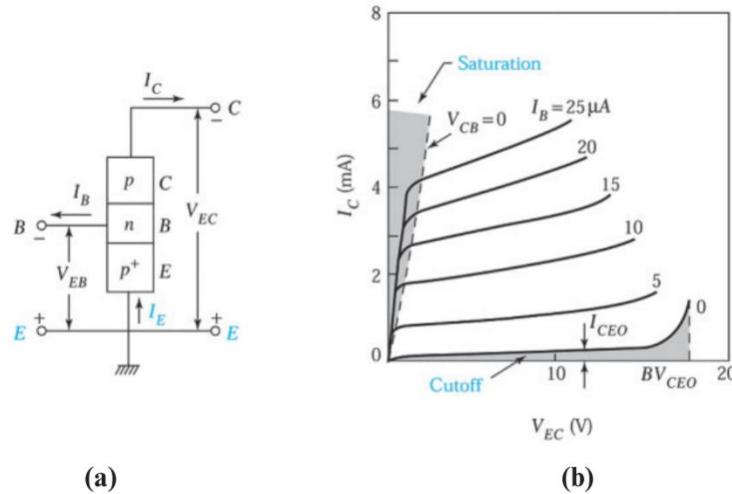


Figure 5.7: BJT in common-emitter configuration (a) and I-V characteristic (b)

5.1.6 Early Effect

In principle, I_C should be independent of the base-collector junction voltage V_{BC} . However, as V_{BC} increases, the width of the space-charge region between base and collector increases, as predicted by equation 4.6 where $V_F = -V_{BC}$. This means that the effective length of the base decreases and this in turn has two effects:

1. More holes coming from the emitter will reach the collector because there is less room for recombination. Effectively, the base transfer factor α_T increases.
2. The hole current is determined by the slope of the hole concentration in the base, as seen previously. As W decreases - while the boundary conditions remain the same - the diffusion hole current I_{Ep} increases. The increase of slope (in absolute terms) is shown in figure 5.8(a).

Both effects contribute to an increase in I_C when V_{CB} (or V_{EV} in common-emitter) increases and I_B remains constant. This phenomenon is called the *Early effect*. We can show that all those non-flat current curves in the I-V characteristic (see 5.8(b)) pass through the same point on the V-axis when extended. The corresponding voltage is the Early voltage V_A .

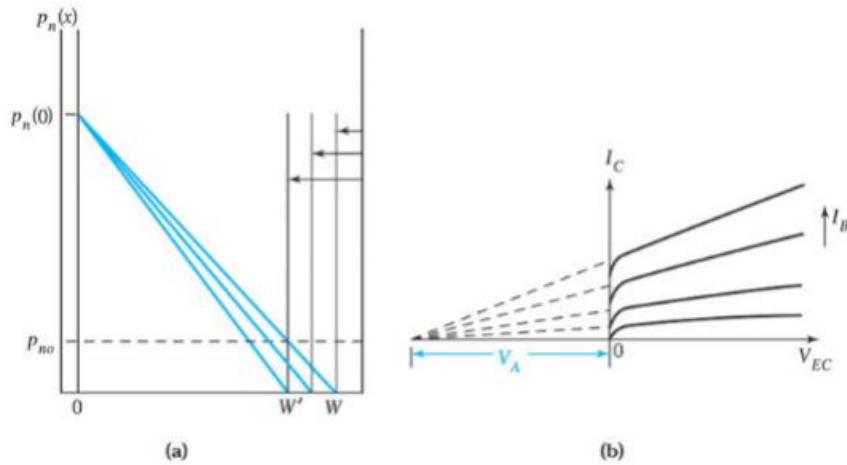


Figure 5.8: Schematic diagram of (a) the Early effect and (b) Early voltage V_A

5.2 MOSFET

The metal–oxide–semiconductor field-effect transistor or *MOSFET* is a transistor based on the field-effect. It has an insulated gate, made of a conductor like a metal or poly-silicon. This terminal is isolated from the rest of the device by a narrow layer of SiO_2 which serves as dielectric. The voltage applied at the gate determines the conductivity of the device and hence can control the current between the two other terminals named source and drain. This ability to change conductivity with the amount of applied voltage can be used for amplifying or switching electronic signals. Figure Figure 5.9(a) shows the structure of a MOSFET, while Figure 5.9(c) shows its symbols together with the three terminals.

5.2.1 Description and Operation

Figure 5.9(b) shows the cross-section of an n-channel MOSFET. Source (S) and drain (D) are both n^+ contacts and are isolated from each other by the p-type substrate. As the voltage on the gate increases, the holes in the p-type material are expelled and the region below the gate becomes depleted of charge carriers. If the gate voltage increases even further, electrons which are present in the p-type bulk as minority carriers are attracted to the region below the gate. A narrow layer of electrons between source and drain is formed and conduction of electrons between source and drain becomes possible. If the electron density below the gate is equal to the hole concentration in the bulk, inversion has happened and we say that a channel has formed. The gate voltage at which this inversion from p to n happens, is called the threshold voltage V_T .

To have a current flow from source to drain, we don't only need a channel but also a positive voltage difference between drain and source. As we increase the voltage at the drain, the current increases, but at the same time the depth of the channel at the drain decreases because $V_{GD} < V_{GS}$. At a certain moment the voltage difference between gate and drain is no longer enough to sustain an n-type channel ($V_{GD} = V_T$). We call this *pinch-off* and say that the transistor is saturated. The current between drain and source no longer increases as the drain voltage increases but remains constant.

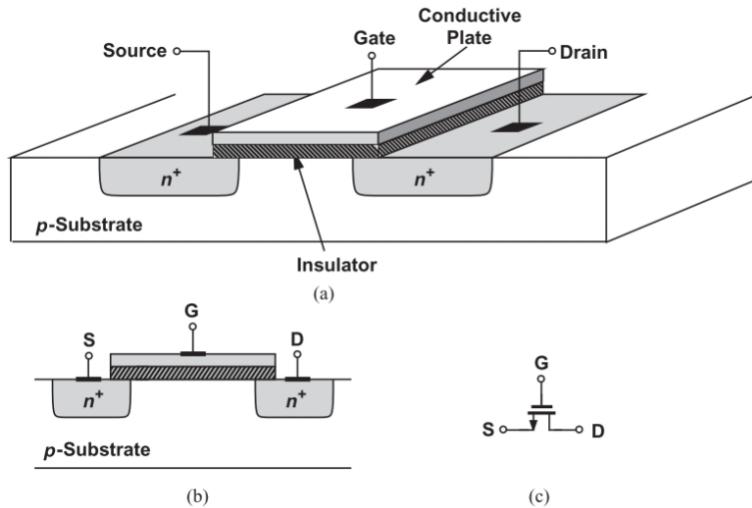


Figure 5.9: Schematic diagram of an n-channel MOSFET: (a) structure, (b) side view, (c) symbol

The dimensions of a MOSFET can become very small; a typical gate oxide thickness t_{ox} is about 15\AA and shrinks with every new generation. The channel length L is typically multiple tens of nanometers.

5.2.2 MOS Capacitor

The region below the gate is called a MOS capacitor. In integrated circuits, it can store charges and forms the basis building block for charge-coupled devices (CCD). A MOSFET can thus be seen as a MOS capacitor and two pn-junctions placed immediately adjacent to it. We will briefly see how charge builds up in the semiconductor as a voltage is applied to the metal gate.

Figure 5.10 represents a MOS capacitor with the gate on the left and the p-type semiconductor on the right of the oxide. If a negative voltage $V < 0$ is applied to the metal plate of the gate, as in 5.10(a), positive carriers are attracted to the $\text{SiO}_2 - \text{Si}$ interface. No current flows in the device, so the Fermi level remains constant. The carrier distribution depends on the difference $E_i - E_F$: $p_p = n_i e^{(E_i - E_F)/kT}$ so the conduction and valence bands at the interface have to bend upward to increase $E_i - E_F$ because E_F is constant. The holes "float" up to the maximum in the valence band and accumulate at the interface.

If $V > 0$, as in 5.10(b), holes are repelled from the interface and the bands bend up. Initially, all acceptor donors are exposed and a charge layer of depth W is created inside the semiconductor. The induced charge density per unit area is $Q_d = qN_A W$. We call this process *depletion*. If V increases further and the bands bend even more, the intrinsic Fermi level will fall below the true Fermi level as in 5.10(c) and electrons will swarm to the interface because the exponent in expression $n_p = n_i e^{(E_F - E_i)/kT}$ becomes positive. This process is called *inversion* and is the condition needed to form a channel in a MOSFET. The channel is a highly charged region just right of the interface (see the charge distribution), separated from the p-type semiconductor by a relatively wide depletion region. The voltage where the Fermi levels cross is the threshold voltage V_T and we denote the associated charge Q_n .

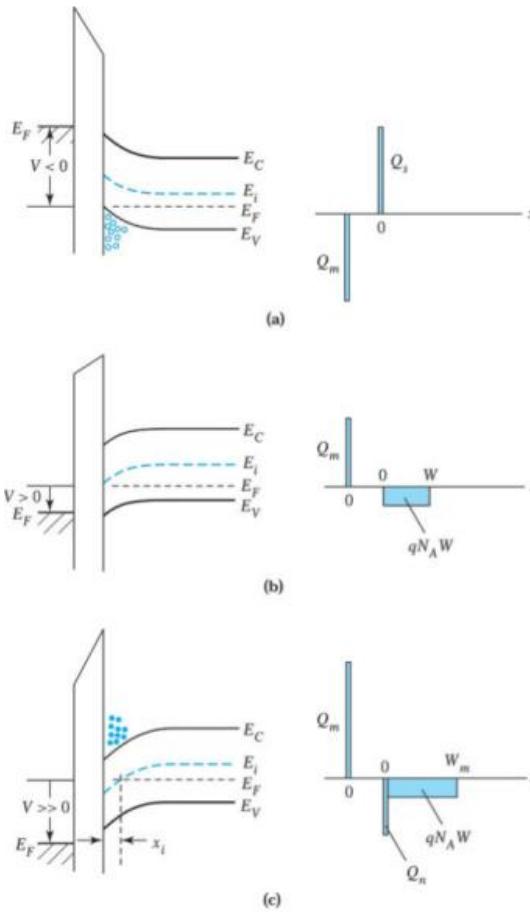


Figure 5.10: Band diagrams and charge distribution of a MOS capacitor in (a) accumulation, (b) depletion, and (c) inversion

5.2.3 I-V characteristic

If V is the applied voltage difference to the plates of a capacitor, and C is the capacitance, then the induced charge is $Q = CV$. In the case of the MOSFET of figure 5.9(b), we are only interested in the mobile charges under the gate, namely the electrons Q_n attracted to the interface in figure 5.10(c), and not the depletion charge $Q_d = qN_A W$. This means that $V = V_{GS} - V_T$ because no mobile charges exist for $V_{GS} < V_T$ ¹. If we assume that the MOS capacitor has a gate capacitance C_{ox} per unit area, this relation becomes

$$Q_n = WC_{ox}(V_{GS} - V_T)$$

with W the width of the transistor. Mind that Q is a charge density per unit length. As drain and source voltage are not the same, the channel voltage varies along the length of the channel. If we denote the channel potential by $V(x)$, we can rewrite the equation above as:

$$Q_n(x) = WC_{ox}(V_{GS} - V(x) - V_T)$$

¹This assumption will be revised in section 5.2.4

where $V(x)$ goes from zero to V_D if the channel is not pinched off (figure 5.11(a)).

We know that the current is the charge density times the velocity of the charges: $I_D = Q_n(x) \cdot v$. The current is a drift current as we apply an electric field \mathcal{E} across the channel. Thus:

$$v = -\mu_n \mathcal{E} = \mu_n \frac{dV}{dx}$$

with μ_n the electron mobility. Substituting this in the expression for I_D , we find:

$$I_D = WC_{ox}(V_{GS} - V(x) - V_T)\mu_n \frac{dV(x)}{dx}$$

Multiplying both sides by dx and integrating from $x = 0$ to the channel length L :

$$\int_{x=0}^{x=L} I_D dx = \int_{V(x)=0}^{V(x)=V_{DS}} \mu_n C_{ox} W (V_{GS} - V(x) - V_T) dV$$

Because I_D is constant along the channel, we can solve both integrals and express I_D as

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (2(V_{GS} - V_T)V_{DS} - V_{DS}^2) \quad (5.4)$$

This is a parabolic function that reaches a maximum for $V_{DS} = V_{GS} - V_T$:

$$I_{D,max} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$

We already established that $V_{DS} = V_{GS} - V_T$ is the condition for saturation: from this point on, the current will no longer increase as V_{DS} increases. This is the situation in figure 5.11(b) and (c). As V_{DS} increases, the pinch-off point P moves closer to the source. The voltage difference in the shrinking channel at P remains $V_{GS} - V_T$. The transistor is in saturation² and the drain current is - in a first approximation - equal to:

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 \quad (5.5)$$

Note that contrary to the BJT, there is no DC current through the gate.³

If V_{DS} is relatively small, we can approximate I_D as:

$$I_D \approx \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)V_{DS}$$

This is the expression of a resistor with value:

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)}$$

We say that the transistor is in the *linear* or *triode* region. Since the resistance is a function of V_{GS} , the MOSFET in this region can be seen as a programmable resistance.

Figure 5.12 gives the overall output characteristic of an n-channel MOSFET. Notice how the transition point from linear to saturation depends on V_{GS} : for the transistor to be in saturation V_{DS} must be larger than the *overdrive voltage* $V_{ov} = V_{GS} - V_T$. This is not the case for the BJT, where we used a fixed cutoff $V_{CE,sat}$ with a value of 0.2V.

We have studied the n-channel MOSFET or *NMOS*. The study of the p-channel MOSFET or *PMOS* is left as an exercise for the reader.

²Mind that saturation for a MOSFET is not the same concept as saturation for a BJT

³A note on notation: we will often use K for the product $\mu_n C_{ox}$.

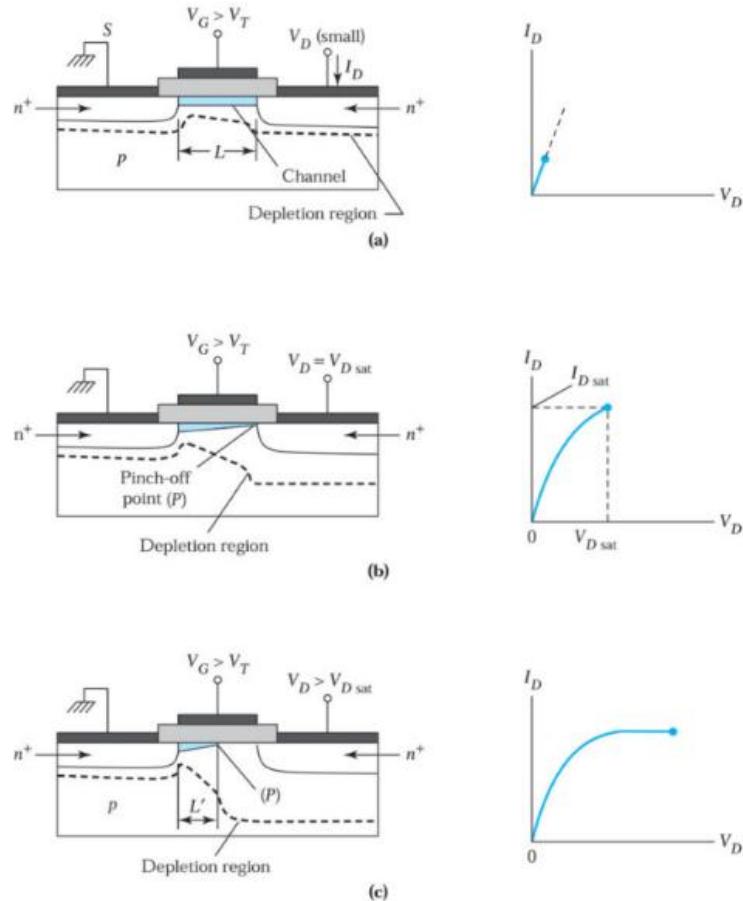


Figure 5.11: Operations of the MOSFET and output I-V characteristics. (a) Low drain voltage. (b) Onset of saturation. (c) Beyond saturation.

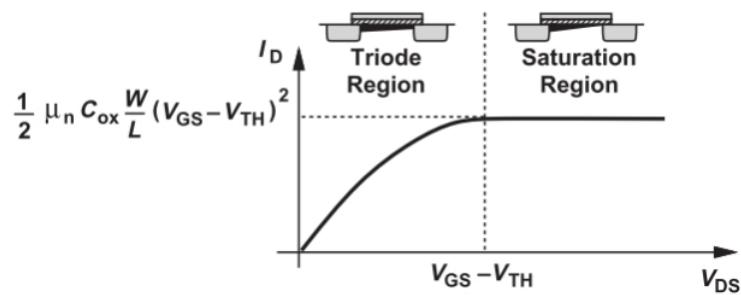


Figure 5.12: Overall MOSFET $I_D - V_{DS}$ characteristic

5.2.4 Second Order Effects

We will discuss several second-order effects that will cause the MOSFET to behave differently than the ideal behavior of figure 5.12.

Channel-length Modulation

Notice that in figure 5.12(c) the effective length of the channel decreases as the drain voltage increase (the pinch-off point moves to the left). We established equation 5.5 with the implicit assumption that L is constant. If however the effective channel length decreases, current I_D will increase with increasing V_{DS} and the I-V characteristic of figure 5.12 is not flat. This is similar to the Early effect in the BJT.

To model this, we assume that L doesn't change, but do include an explicit dependence on V_{DS} in equation 5.5:

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \quad (5.6)$$

The factor λ is the *channel length modulation coefficient*. To decrease λ , the designer can increase the length of the transistor, because this makes the relative impact of a change in L is smaller.

Body Effect

Until now, we have supposed that both the p-type substrate and the n-type source are connected to a common ground. This is however not always the case in real circuits, where the source can be tied to voltages higher than the substrate. Even in that case, the pn-junction between source and substrate is still reversed biased and the device still works properly.

However, something does change as the source voltage increases with respect to the bulk: as the source becomes more positive with respect to the substrate, the threshold voltage V_T increases. Called "body effect," this phenomenon is formulated as

$$V_T = V_{T0} + \gamma(|2\phi_F + V_{SB} - |2\phi_F|)|$$

where V_{SB} is the voltage difference between source and substrate (bulk), V_{T0} the threshold voltage when $V_{SB} = 0$ and γ and ϕ_F technology-dependent parameters.

Subthreshold Conduction

We have assumed that the MOSFET turns on abruptly when the gate voltage exceeds the threshold voltage. In practice however, the device turns on gradually and there is already a source-drain current before V_T is reached. This current depends exponentially on V_{GS} , similarly as in a BJT. Called the *subthreshold conduction*, this effect has become a critical issue in modern MOS devices.

Part II

Analog Electronics

Chapter 6

Basic Analog Circuits

In this chapter, we will see how non-linear elements are used in electrical circuits. We discuss the concept of a load line, both static and dynamic, and study the small-signal response of a non-linear element, which essentially requires a linearization around an operating point. Finally, we provide small-signal models for both diodes and transistors at high and low frequencies.

6.1 Non-linear elements in circuits

In this section, we will see how non-linear elements like diodes and transistors are used in electrical circuits. In principle, adding a non-linear element makes the analysis of the circuit a lot harder, compared to circuits with only linear elements (resistors, capacitors, inductors, ...). In practice however, we will rely on a graphical solution method, which simplifies the analysis while still being rigorous.

6.1.1 The Diode as a Circuit Element

Let's use a diode as a lumped element in an electronic circuit, as in the figure below.

By applying KVL, we obtain the equation of the **load-line**:

$$E - v_D = R i_D$$

This equation has to be combined with the diode I-V characteristic:

$$i_D = \phi(v_D) = I_S(e^{v_D/v_{th}} - 1)$$

From a formal point of view, we have two unknowns, v_D and i_d , and two equations, so in principle we can solve for both unknowns. However, there is no analytical solution to our problem, so we prefer a graphical method.

In figure 6.2 we combine $i_D = \phi(v_D)$ (the red curve) with the expression of the load line (green line).

We apply a simplification and assume that $v_D = V_{DQ} \approx 0.6$ V. The operating point¹ Q lies at the intersection of both lines. In order for the diode to conduct, we can immediately

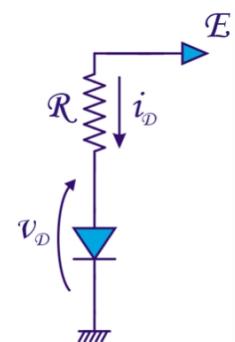


Figure 6.1

¹The letter Q stands for *quiescent*

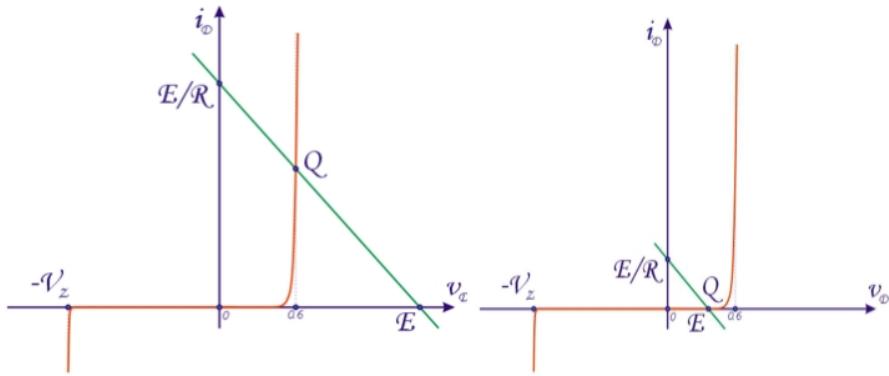


Figure 6.2: I-V with load line (a) with $E > 0.6V$ and (b) $E < 0.6 V$

conclude by comparing both figures that it is necessary that $E > 0.6V = V_{DQ}$. The diode current is easily calculated as

$$i_D \approx I_{DQ} = \frac{E - V_{DQ}}{R}$$

We can conclude that the diode will always conduct as long as $E > 0.6 V$. Variations in E only mean that the load line will move parallel. The operating point (V_{DQ}, I_{DQ}) can only move vertically because of the nature of $\phi(v_D)$ when $v_D > 0.6 V$. Only when E becomes smaller than $0.6V$ will conduction stop until we reach the Zener region. Remember that the $0.6 V$ is specific for silicon.

As an example, consider the circuit in figure 6.3. It is not obvious when the diode will conduct, so let's replace the current source and the resistor by the Thevenin equivalent, namely a resistor $R_{th} = R$ and a voltage source $V_{th} = RI_0$ in **series** with this resistor. This circuit is identical to the one in figure 6.1 with a load line through the points $(R I_0, 0)$ and $(0, I_0)$. We can conclude that the transistor will conduct when $V_{th} = R I_0 > 0.6V$. Furthermore, when $R \rightarrow \infty$ and the resistor become an open circuit, the load line will become a horizontal line through I_0 .

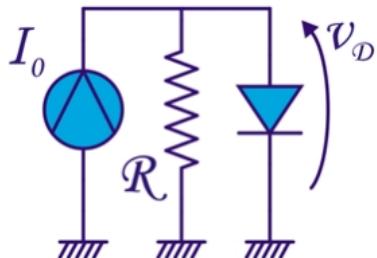


Figure 6.3: With current source

6.1.2 The BJT as a Circuit Element

A typical bipolar transistor circuit is shown in figure 6.4(a). To analyze this circuit, we cut at the entrance of the base and replace the left loop of resistors R_1 and R_2 and the supply voltage E by the Thevenin equivalent circuit in figure 6.4(a).

The Thevenin voltage E_B and impedance R_B are easily computed by seeing that (a) the voltage at the base is the result of applying a voltage divider to the supply voltage E , and (b) that when we ground E (as required by the Thevenin rules) R_1 and R_2 are in par-

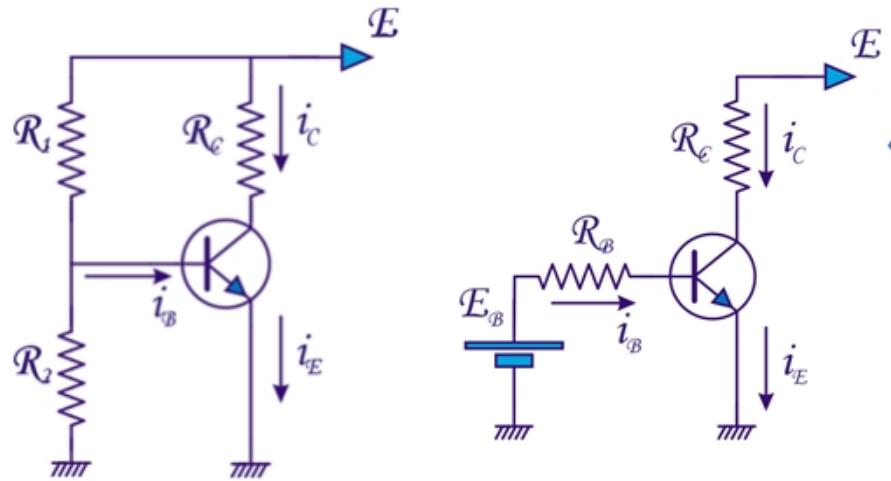


Figure 6.4: (a) Transistor circuit and (b) simplified with Thevenin

allel:

$$\begin{aligned} E_B &= \frac{R_2}{R_1 + R_2} E \\ R_B &= R_1 || R_2 = \frac{R_1 R_2}{R_1 + R_2} \end{aligned} \quad (6.1)$$

In this left loop, we can write:

$$E_B - v_{BE} = R_B i_B \quad (6.2)$$

Consider the right loop, which consists of resistor R_C and the voltage v_{CE} . In this loop, we can write:

$$E - v_{CE} = R_C i_C \quad (6.3)$$

Let's assume that all the currents are constant and we have biased the transistor in its operating point. We indicate this by adding a Q to the currents and voltages, just as for the diode. With this convention, we can rewrite these equations to obtain:

$$\begin{aligned} I_{BQ} &= \frac{E_B - V_{BEQ}}{R_B} \\ V_{CEQ} &= E - R_C I_{CQ} \end{aligned} \quad (6.4)$$

with $V_{BEQ} \approx 0.6$ V because we want to bias the base-emitter junction in the forward (conducting) region. From the diode analysis, we know this will be the case when $E_B > 0.6$ V. We also know the relation between I_{BQ} and I_{CQ} (neglecting any leak currents):

$$I_{CQ} = \beta I_{BQ} \quad (6.5)$$

where β is given by the manufacturer and is typically very high, but it can vary a lot from one transistor to the next. With this equation, we have enough information to compute all currents and voltages:

1. We know R_B and E_B , and we assume $V_{BEQ} = 0.6V$, so we can compute I_{BQ} .
2. We know β , so we can compute I_{CQ} .
3. We can now compute V_{CE} .

All these calculations can all be represented graphically; see the (idealized) BJT characteristic in figure 6.5. The figure on the left represents the left loop, with the load line $E_B - v_{BE} =$

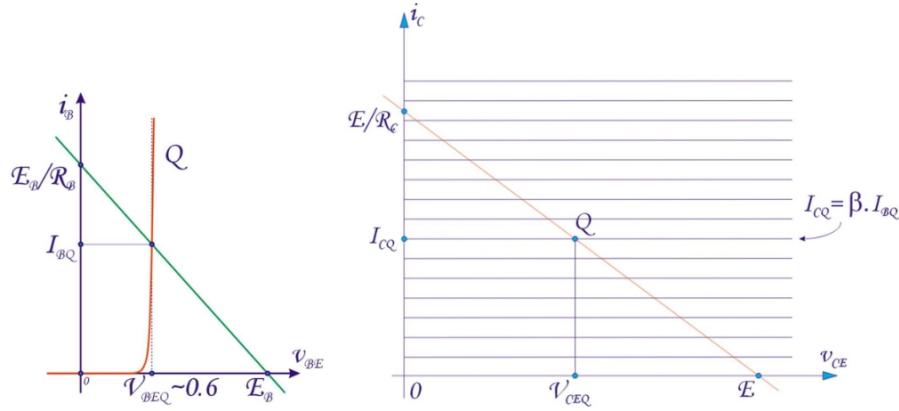


Figure 6.5: (a) Transistor circuit and (b) simplified with Thevenin

R_B i_B is green and the v_{BE} junction characteristic in red. The intersection between both functions is the operating point $Q = (V_{BEQ}, I_{BQ})$.

The right loop is shown in the graph on the right; the horizontal line on which the transistor operates is given by $I_C = \beta I_B$ and the load line is given by $V_{CE} = E - R_C I_C$. Once again, the operating point is there where both lines intersect.

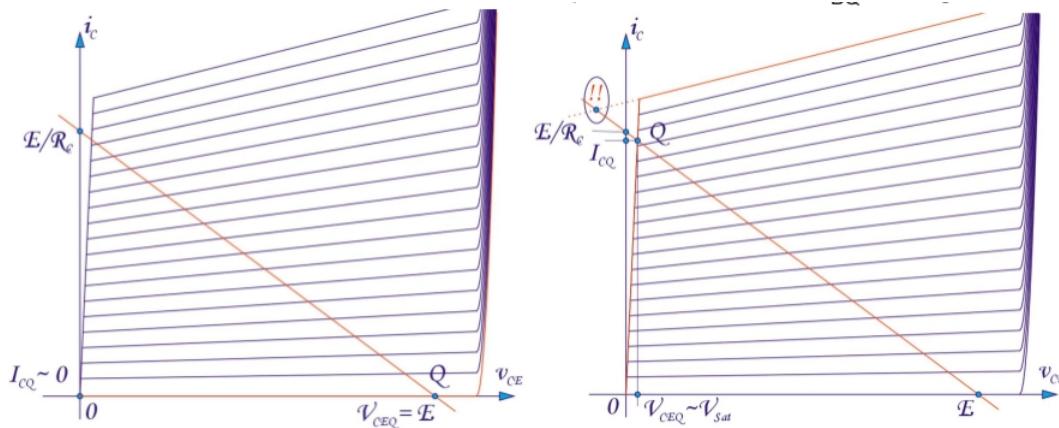
We can reason about the operation of the circuit by thinking about these figures. For example, if R_B would decrease, the slope of the $I_B - V_{BE}$ load-line would increase, so that the operating point Q will move up and I_{BQ} will increase. This increase will cause in increase of I_{CQ} in the graph on the right, and the operating point will move along the load line to a higher I_C and a lower V_{CE} .

We can conclude that:

- The left loop determines I_{BQ} .
- By assumption, we are in the normal working domain and $I_{CQ} = \beta I_{BQ}$.
- The right loop gives V_{CEQ} .
- Given $Q = (V_{CEQ}, I_{CQ})$, we verify we are in the normal working domain.

Let's discuss a couple of edge cases:

- If $E_B < 0.6V$ then $V_{BEQ} < 0.6V$. Then $I_{BQ} \approx 0$ and $I_{CQ} \approx 0$ (neglecting leakage current). See figure 6.6(a). If that's the case, the operating point is $Q = (V_{CEQ} = E, 0)$ and the transistor is blocked (*cut-off mode*).
- As V_{BEQ} and I_{BQ} keep increasing, I_{CQ} will become larger and eventually V_{CEQ} will become too small to keep the transistor in active mode. Then $I_{CQ} \neq \beta I_{BQ}$ and $V_{CEQ} \approx V_{CE,Sat}$. So, in that case, $I_{CQ} = \frac{E - V_{CE,Sat}}{R_C}$. The transistor is *saturated*.

Figure 6.6: (a) $V_{BEQ} < 0.6V$ and (b) $V_{CEQ} \approx V_{CE,Sat}$

6.1.3 The MOSFET as a Circuit Element

Just as for the BJT, we consider a biasing circuit for the n-channel MOSFET transistor as in figure 6.7(a) and we simplify the circuit with a Thevenin equivalent circuit - just as we did for the BJT - as in figure 6.7(b). with

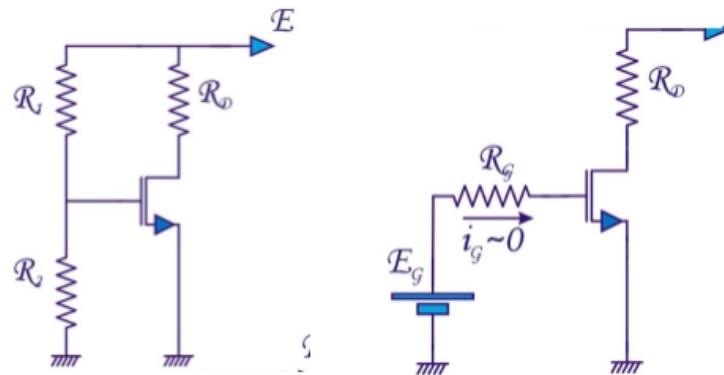


Figure 6.7: (a) MOS Transistor circuit and (b) simplified with Thevenin

$$E_B - G = \frac{R_2}{R_1 + R_2} E \quad (6.6)$$

$$R_G = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2}$$

Note that the gate current $I_G = 0$ as the gate is a capacitor where the DC current is zero. And just as for the BJT, we can express an equation for the left and right loop:

$$V_{GSQ} = E_G \text{ because } I_G = 0$$

$$I_{DSQ} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{GSQ} - V_T)^2 \text{ if } V_{GSQ} > V_T \quad (6.7)$$

$$V_{DSQ} = E - R_D I_{DSQ} \text{ if } V_{DSQ} > V_{GS} - V_T$$

The last condition is required for the transistor to be saturated, which is similar to the active mode for a BJT.

We can also represent these equations graphically. Figure 6.8(a) represents the quadratic relation between I_{DS} and V_{GS} to determine V_{GSQ} . As long as $V_{DS} > V_{GSQ} - V_T$, figure 6.8(b) gives the value of V_{DSQ} .

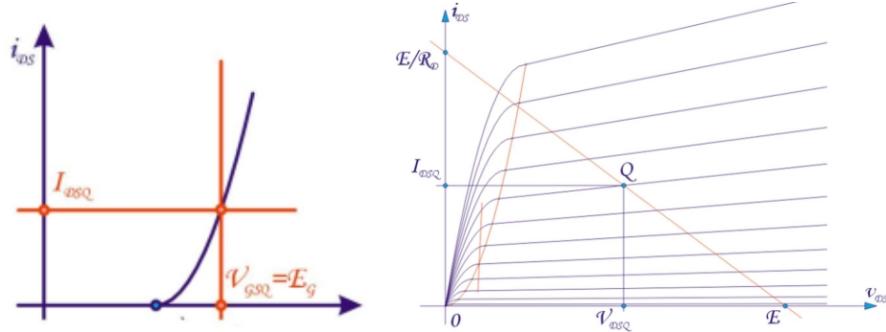


Figure 6.8: (a) $i_{DS} = f(v_{DS})$ and (b) $I_{DS} = f(V_{DS})$

6.1.4 Additional remarks

For both type of transistors, there are three working domains, as in figure 6.9:

- MOSFET: blocked, saturated, linear
 - BJT: blocked, normal (or active), saturated
- For each transistor there exist three working domains :
 - FET : Blocked, Saturated, Linear
 - BJT : Blocked, saturated, Normal

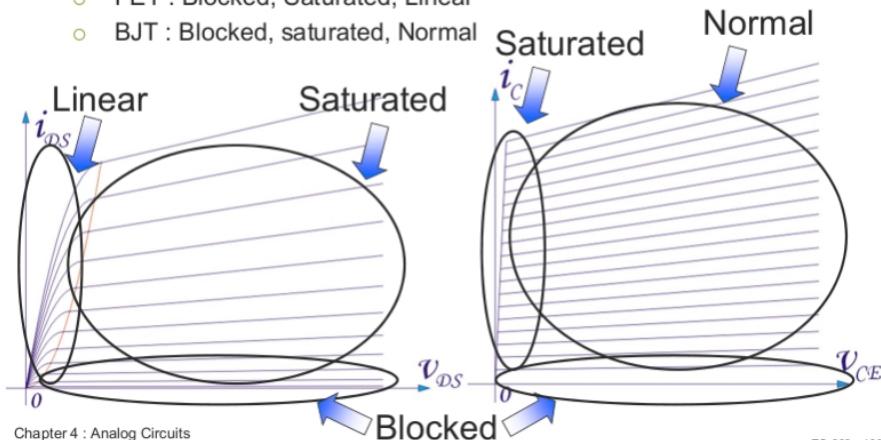


Figure 6.9: (a) $i_{DS} = f(v_{DS})$ and (b) $I_{DS} = f(V_{DS})$

6.1.5 A more general circuit

To improve the linearity (see the chapter on feedback) and biasing of the circuit, an emitter resistance R_E is often added, as in figure 6.10(a). Just as before, the right side is simplified

with the Thevenin equivalent (see 6.10(b)). Once again, we can write the KVL in left and

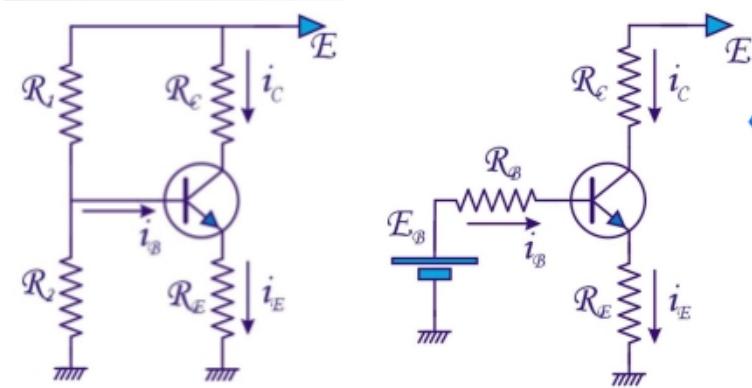


Figure 6.10: (a) General circuit and (b) Thevenin simplification

right loop:

$$\begin{aligned} E_B - V_{BEQ} &= R_B I_{BQ} + R_E I_{EQ} = R_B I_{BQ} + R_E(\beta + 1) I_{BQ} \\ E_B - V_{BEQ} &= R_B \frac{I_{CQ}}{\beta} + R_E(\beta + 1) \frac{I_{CQ}}{\beta} \approx \frac{R_B}{\beta} I_{CQ} + R_E I_{EQ} \end{aligned} \quad (6.8)$$

where we have used $I_{CQ} = \beta I_{BQ}$ because we suppose we're working in the normal operating region. These equations lead to expressions for current I_{CQ} and voltage V_{CEQ} :

$$\begin{aligned} I_{CQ} &= \frac{E_B - V_{BEQ}}{\frac{R_B}{\beta} + R_E} \\ V_{CEQ} &= E - (R_C + R_E) I_{CQ} \end{aligned} \quad (6.9)$$

These equations can be plotted on the different I-V characteristics as in figure 6.11. Notice that the figure on the left gives i_C as function of v_{BE} . Since the relation between i_B and v_{BE} is the exponential diode characteristic, the same goes for the relation between $i_C = \beta i_B$ and v_{BE} .

TODO: expand + explain for MOSFET.

6.2 Small-Signal Response

In this section, we will introduce the concept of a small-signal response, namely how do voltage and currents in a circuit change when we apply only a small change to the input values. The general idea is that we design the circuit such that it operates at an operating point Q , and we linearize the circuit around this operating point to study only small deviations. We will use the diode as an example. in section 6.5 we apply the same reasoning to transistors. First, we introduce some notation to distinguish large-signal from small signal quantities.

- x_A : measure of a specific variable,
- X_A : the average value of this specific variable,
- x_a : variation of the specific variable around average X_A .

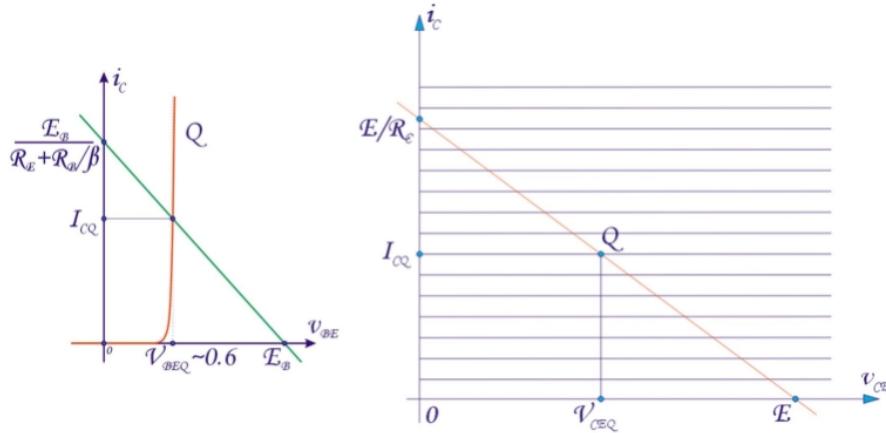


Figure 6.11: (a) Load line: $I_{CQ} = \frac{E_B - V_{BEQ}}{\frac{R_E}{\beta} + R_E}$ (b) $V_{CEQ} = E - (R_C + R_E)I_{CQ}$

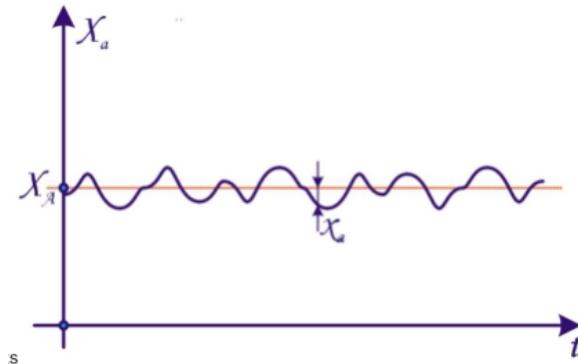


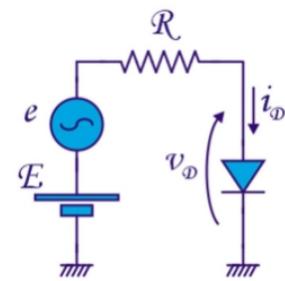
Figure 6.12: Signal quantities

Refer to figure 6.12 for a visual representation of these variables. Note that only x_A and x_a vary with time, and that the average value of x_a is zero: $\mathbb{E}[x_a] = 0$.

Let's apply this to the simple diode circuit. In the figure to the right, the supply has two components: a fixed voltage E , and a varying voltage e with average value $\mathbb{E}[e(t)] = 0$. The quantities we're looking for, v_D and i_D , can be split in two components: an average value and a variation around this average.

$$\begin{aligned} v_D &= V_D + v_d \\ i_D &= I_D + i_d \end{aligned} \quad (6.10)$$

Assume that $e = 0$, i.e. we study the system with no variations. If $E > 0.6V$, we can write - like we've done before - that:



$$\begin{aligned} V_{DQ} &= 0.6V \\ I_{DQ} &= \frac{E - V_{DQ}}{R} \end{aligned} \quad (6.11)$$

In this way, we determine the operating point as the intersection between the load line and the diode characteristic, as in figure 6.13.

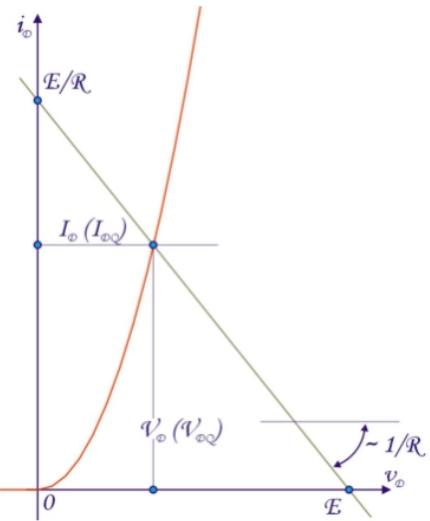
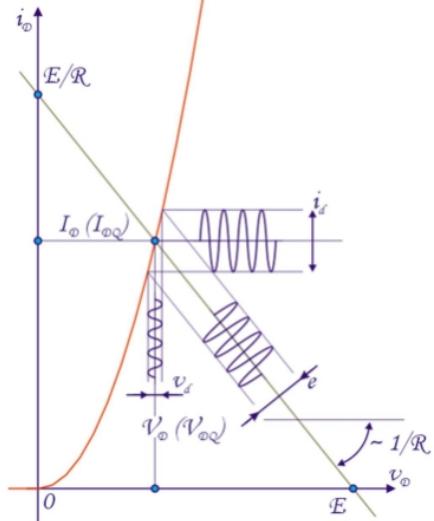


Figure 6.13: Diode equation and load line

Figure 6.14: Variations around Q

Now assume $e \neq 0$. This changes the equation for the load line:

$$E + e - v_D = R i_D \quad (6.12)$$

This equation can be rewritten as:

$$(E + e) - (V_{DQ} + v_d) = R(I_{DQ} + i_d)$$

or, since $E - V_{DQ} = R I_{DQ}$:

$$e - v_d = R i_d \quad (6.13)$$

This is the equation of the small-signal load line, where the center of the coordinate system is translated to the operating point $Q = (V_{DQ}, I_{DQ})$. Figure 6.14 shows what happens: small variations of e move the load line parallel to the original load line $E - V_{BEQ} = R I_{DQ}$. As this moving load line intersects with the diode characteristic, small voltage variations v_d and current variations i_d appear across the diode. We can determine the relation between v_d and i_d :

$$\begin{aligned} i_D &= \phi(v_D) \approx I_S e^{v_D/v_{th}} \\ di_D &= \frac{I_S}{v_{th}} e^{v_D/v_{th}} dv_D \\ \Rightarrow i_d &= \frac{i_D}{v_{th}} v_d \end{aligned} \quad (6.14)$$

and finally

$$v_d = \rho_d i_d \text{ with } \rho_d = \frac{v_{th}}{I_{DQ}} \quad (6.15)$$

We have effectively linearized the diode characteristic around the operating point. Locally, for small variations, the diode operates as a resistor with resistance ρ_d .

By doing this, we have transformed the original problem into two subproblems:

1. Determine the DC solution by solving (graphically) the equations on the left of figure 6.15. This solution determines Q and the so-called small signal parameters, like ρ_d .
2. Solve a linear circuit where the nonlinear element has been replaced by the small-signal equivalent - a resistor in the case of our diode. See the right part of figure 6.15.

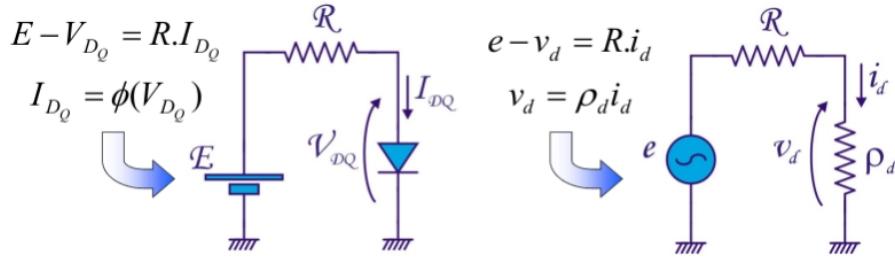


Figure 6.15: Signal quantities

However, to be correct, the small-signal equivalent model of a diode is not just a resistor. A pn-junction creates a space-charge region on its interface. As the voltage across the junction changes, charges (both n and p) have to be transported to and from the junction to increase or decrease the SCR. This means that a diode - or any pn-junction - is also capacitive. This phenomenon of depletion capacitance was already explained in section 4.4.

To model this behavior, we replace a diode in as small-signal equivalent circuit by:

1. A dynamic resistance ρ_d , in parallel with
2. A junction capacitance C_j , as in figure 6.16 (with $\alpha = 1/2$). This capacitance can be neglected for small frequencies.

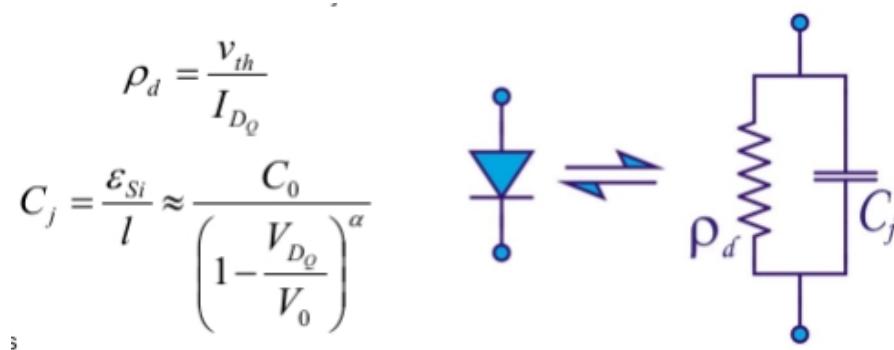


Figure 6.16: Small-signal model of a diode

Finally, to establish the small-signal equivalent circuit:

- We replace all nonlinear devices by their small-signal model (like the one in 6.16 for the diode).
- We replace all independent voltage source by a short-circuit (i.e. $E = 0$) because we assume they don't vary with time.
- For the same reason, we replace all independent current source by an open circuit (i.e. $I = 0$).

6.3 Static and Dynamic Load lines

In section 6.1, we saw the concept of a load line. However, there is more to it than we've seen up to now. The reason is the fundamental difference between the operating point Q and the small-signal response. The former is fundamentally a DC concept, because there are no time-varying quantities involved. The small-signal response at the other hand deals with AC signals: signals that vary in time and thus have a non-zero frequency. Let's thus study a circuit that contains frequency-dependent components like capacitors or inductors.

Consider the circuit in figure 6.17, where a load charge R_L is connected to the original diode circuit through a capacitor C . To compute the operating point Q , we assume $e = 0$. Since there are no variations, the capacitor C is an open circuit and the circuit is reduced to the one in figure 6.18. This is the same circuit as before, so we conclude that $V_{DQ} = 0.6$ V and the load line is $E - V_{DQ} = R I_{DQ}$. The operating point allows us to compute the small-signal resistance ρ_d of the diode.

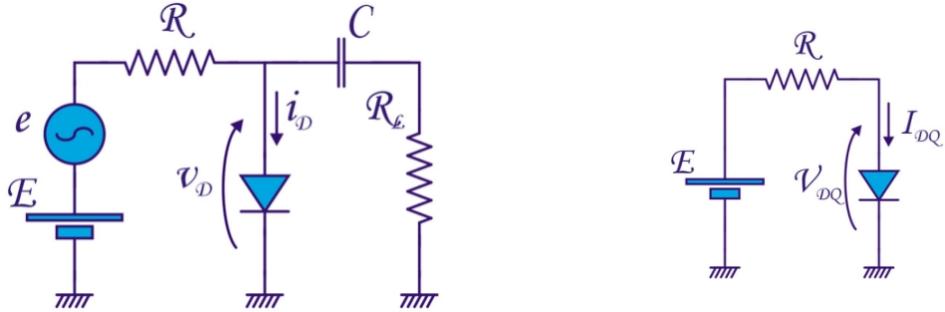


Figure 6.18: Circuit to determine Q

Figure 6.17: Diode circuit with capacitor

For the small-signal response, we can't neglect C . We do replace the diode by a resistance ρ_d (and neglect - for simplicity - the capacitance C_j) and obtain the circuit in figure 6.19 with $E = 0$. This circuit can be simplified with Thevenin's theorem, which gives the circuit in figure 6.20 where we have made a cut just above ρ_d . Z_{th} and e_{th} can be computed as follows:

- For e_{th} , we assume an open circuit, so no current through ρ_d . As such, we have a voltage divider consisting of resistors R and R_L and capacitor C . Let Z_C be the series combination of R_L and C , namely:

$$Z_C = R_L + \frac{1}{j\omega C} = \frac{1 + j\omega R_L C}{j\omega C} \quad (6.16)$$

Then, we apply the expression for a voltage divider:

$$e_{th} = \frac{Z_C}{R + Z_C} e = \frac{\frac{1 + j\omega R_L C}{j\omega C}}{R + \frac{1 + j\omega R_L C}{j\omega C}} e = \frac{1 + j\omega R_L C}{1 + j\omega(R + R_L)C} e \quad (6.17)$$

- For Z_{th} , we replace e by a short-circuit. Z_{th} is then the parallel combination of R with Z_C :

$$Z_{th} = \frac{Z_C R}{Z_C + R} = R \frac{1 + j\omega R_L C}{1 + j\omega(R + R_L)C} \quad (6.18)$$

Obviously, the load line is different: at DC, it is $E - V_{DQ} = R I_{DQ}$, but at AC it is $e_{th} - v_d = i_d Z_{th}$. For high frequencies, we can simplify $Z_{th}|_{\omega \rightarrow \infty} = R \frac{R_L}{R + R_L} = R || R_L$ and the small-signal

load line becomes $e_{th} - v_d = i_d(R||R_L)$ which has a different slope than the DC load line (keep in mind that the AC load line is centered at the operating point Q).

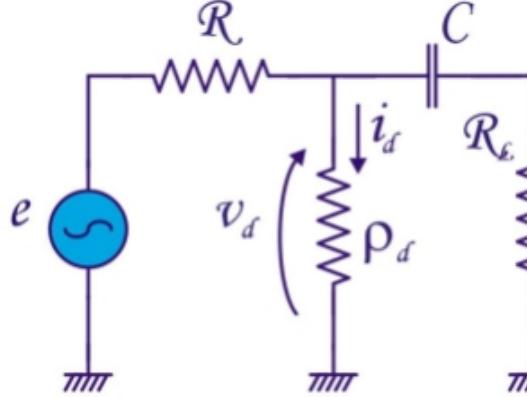


Figure 6.19: Diode circuit with capacitor

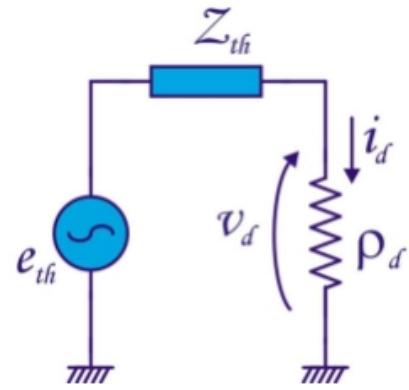


Figure 6.20: Circuit to determine Q

Both load lines are show in figure 6.21, with in green the static load line (slope = $\frac{-1}{R_{stat}}$ where $R_{stat} = R$) and in blue the dynamic load line (slope = $\frac{-1}{R_{dyn}}$ where $R_{dyn} = R||R_L$).

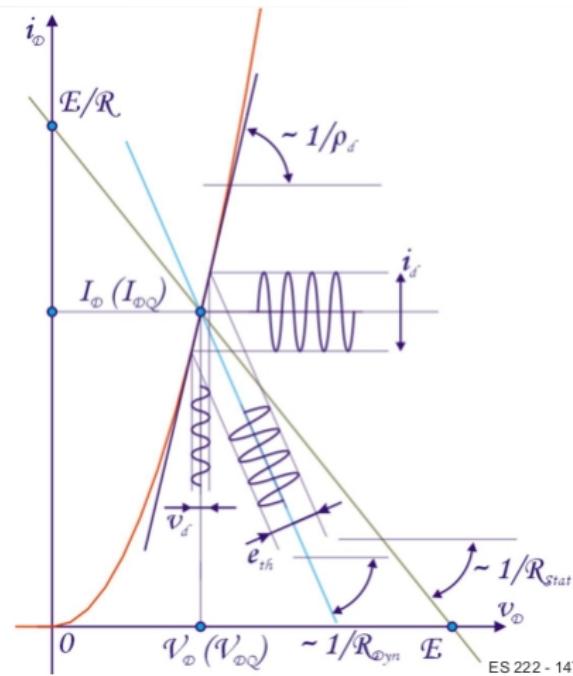


Figure 6.21: Static (green) and dynamic (blue) load lines

We conclude that there are two load lines:

1. The *static* load line, determined at zero frequency (DC) and used to place the operating point
2. The *dynamic* load line, at the frequency of interest, which typically is high enough so that we can simplify the impedance. It determines how the operating will move (the small-signal response)

The later remark implies that there is a critical frequency from which the capacitor C can be neglected. From equation 6.16, we see that this impedance has a pole in $\omega = 0$ and a zero in $\omega = 1/(R_L C)$. For frequencies higher than $\frac{2\pi}{R_L C}$, the impedance becomes frequency-independent and is equal² to R_L . Thus the circuit reduces to the one in figure 6.22. It is this circuit that determines the dynamic load line.

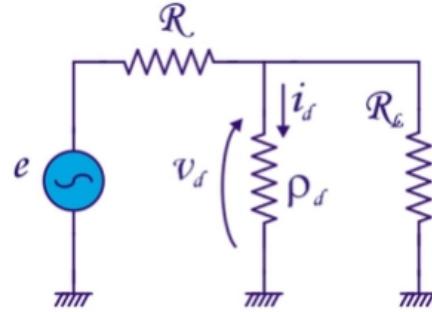


Figure 6.22: Circuit to determine the dynamic load line

6.3.1 Transistors and Dynamic Load Lines

Consider the circuit in figure 6.23. This is the same circuit as we saw in figure 6.10, but with a capacitor C_E in parallel with the emitter resistance R_E . After simplifying the left part with

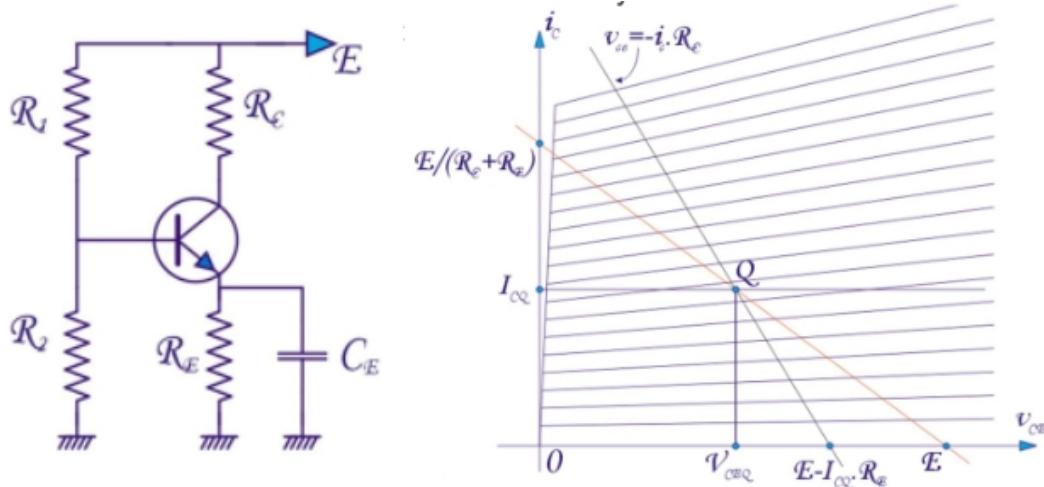


Figure 6.23: BJT circuit with emitter bypass capacitor C_E

the Thevenin theorem, we can establish the equation in the left loop, which hasn't changed from section 6.1.5:

$$E_B - V_{BEQ} = \left(\frac{R_B}{\beta} + R_E \right) I_{CQ} \quad (6.19)$$

The equations in the right loop change, with $Z_E = R_E || C_E = \frac{R_E}{1+j\omega C_E R_E}$, becomes

$$v_{CE} = E - (R_C + Z_E) i_C \quad (6.20)$$

²Convince yourself by sketching the Bode graph

which can be split into two equations:

1. The DC operating point:

$$V_{CEQ} = E - (R_C + R_E) I_{CQ}$$

,

2. A small signal equation, valid for $\omega \gg \omega_0$:

$$v_{ce} = -R_C i_c$$

These static and dynamic load lines are represented in figure 6.23 (red and green lines, respectively). The intersection of the dynamic load line with the horizontal axis is given by $E - R_E I_{CQ}$ because the voltage at the emitter is fixed (if ω is high enough) and equals $R_E I_{CQ}$. Hence, on the dynamic load line, the maximum swing of v_{ce} is between 0 and $E - R_E I_{CQ}$.

6.4 Biasing

In the previous section, we developed a way to determine the operating point of a transistor if the resistors are given:

- Determine the current via the left loop: $E_B - V_{BEQ} = \left(\frac{R_B}{\beta} + R_E \right) I_{CQ}$ or $E_G - V_{GSQ} = R_S I_{DSQ}$.
- Determine V_{CEQ} or V_{DSQ} based on the right loop, and check whether we are in the normal (saturation) region of the transistor (i.e. $V_{CEQ} > 0.6$ V or $V_{DSQ} > V_{GSQ} - V_T$).

However, many values are not exactly known:

- V_{BEQ} varies over a specific production lot,
- β depends on i_C , and varies (from -50% to +200 %) over a specific lot,
- V_T is only specified with a certain precision,
- $K = \mu_n C_{ox}$ (or $\mu_p C_{ox}$) is only specified with a certain precision,
- All these parameters vary with temperature.

This section will describe a method to choose the biasing resistors (R_1 , R_2 , R_C or R_D and R_E or R_S) such that variation in the parameters above has minimal impact on the quiescent currents and voltages for which the circuit is designed.

6.4.1 BJT Biasing

The goal of BJT biasing is to choose R_1 , R_2 and R_E to reduce the impact of variations on V_{BEQ} and β . Furthermore, R_C will be chosen such that the operating point Q lies in the middle of the normal operating region.

Consider the general four-resistor BJT circuit in 6.10(b). We reproduce the Thevenin simplification here for convenience.

As a reminder, the Thevenin voltage E_B and impedance R_B are derived from the biasing resistors R_1 and R_2 and the supply voltage E : $E_B = \frac{R_2}{R_1+R_2} E$ and $R_B = R_1 \parallel R_2$.

From this circuit, we see that:

$$E_B - V_{BEQ} = \left(\frac{R_B}{\beta} + R_E \right) I_{CQ}$$

which means that:

$$I_{CQ} = \frac{E_B - V_{BEQ}}{\frac{R_B}{\beta} + R_E}$$

To make I_{CQ} independent from β , we should choose

$$R_B \ll \beta R_E \quad (6.21)$$

such that:

$$I_{CQ} \approx \frac{E_B - V_{BEQ}}{R_E} \quad (6.22)$$

In this equation, we suppose that E_B and R_E are constant and can be produced with high accuracy (there is still a temperature dependence, but this is relatively low). Suppose that V_{BEQ} can vary, which has an impact on I_{CQ} . This can be quantized as:

$$\Delta I_{CQ} = \frac{\Delta V_{BEQ}}{R_E} \quad (6.23)$$

A typical problem then goes as follows:

- The limits of V_{BEQ} are known, so we know ΔV_{BEQ}
- The limits of β are known: $(\beta_{min}, \beta_{max})$
- The value of I_{CQ} is given, with a certain precision ΔI_{CQ}

This problem can be solved by following these steps:

- With ΔV_{BEQ} and ΔI_{CQ} , determine $R_E = \frac{\Delta V_{BEQ}}{\Delta I_{CQ}}$,
- With R_E and the equation of the left loop, determine $E_B = R_E I_{CQ} + V_{BEQ}$,
- With β_{min} , choose a R_B such that $R_B \approx \beta_{min} R_E / 10$. This guarantees that condition 6.21 is satisfied.
- With R_B and E_B , determine R_1 and R_2 based on the Thevenin equations.

This procedure allows us to determine R_1 , R_2 and R_E . The only remaining unknown is R_C . We will determine this resistor by requiring that the operating point Q lies in the middle of the normal operating region. Refer to figure 6.23 of the biasing circuit with bypass capacitor. We assume that $C_E \rightarrow \infty$, such that it is a short circuit for small signals³. In this way, the static and dynamic load lines correspond to those of figure 6.23(b).

The slope of the dynamic load line is given by R_C . To determine this resistor, we need to set two points of the load line. We choose to limit the current I_C between 0 and $2I_{CQ}$.

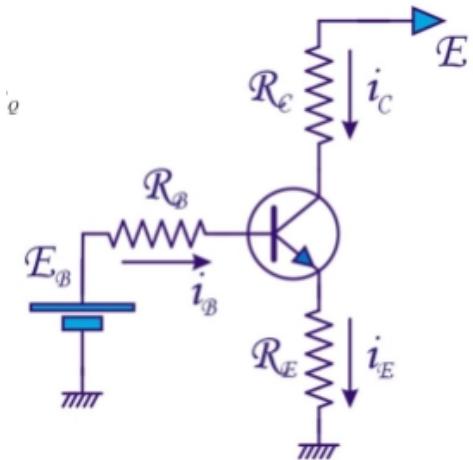


Figure 6.24: Circuit for biasing a BJT

³Or we assume that the only frequencies of interest are much higher than $\omega_0/2\pi$.

This corresponds to a maximum attainable current swing. The minimum voltage is $V_{CE,Sat}$, because below this value the transistor goes into saturation. The maximum voltage is $E - R_E I_{CQ}$, as can be derived from the figure. Note that in AC, the emitter is grounded, so the voltage at the emitter is fixed. Hence we compute R_E as the slope between these two points:

$$R_C = \frac{E - R_E I_{CQ} - V_{CE,Sat}}{2I_{CQ}} \quad (6.24)$$

This reasoning is graphically represented in figure 6.25.

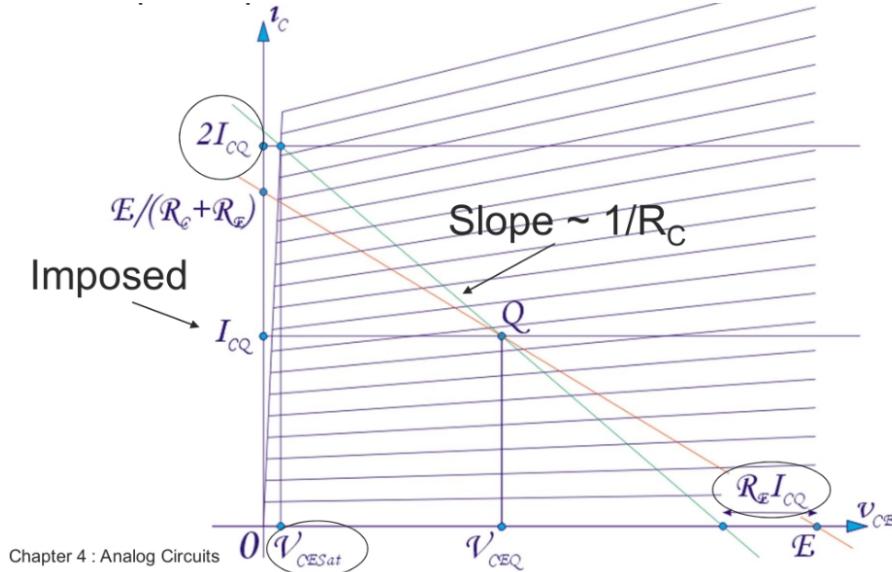


Figure 6.25: Determining R_C to place in Q in the middle of normal operating region

6.4.2 MOSFET Biasing

The goal of MOSFET biasing is to choose R_1 , R_2 and R_S to reduce the impact of variations on V_T and K . Furthermore, R_D will be chosen such that the operating point Q lies in the middle of the normal operating region.

The circuit we consider is the one in figure 6.26. As always, we replace the maze on the left by the Thevenin equivalent with E_G and R_G . We don't know the exact position of the i_{DS} vs v_{GS} curve because:

- The required value of I_{DSQ} is only given within certain limits: $I_{DSQ,min} < I_{DSQ} < I_{DSQ,max}$
- The manufacturer gives the value of $K = \mu C_{ox}$ within limits: $K_{min} < K < K_{max}$
- The same goes for the threshold voltage V_T : $V_{Tmin} < V_T < V_{Tmax}$.

This allows us to draw a minimum (based on V_{Tmax} and K_{min}) and maximum curve (based on V_{Tmin} and K_{max}), as in figure 6.27. The intersection of these curves with resp. $I_{DSQ,min}$ and $I_{DSQ,max}$ gives two points on which the load line $E_G = v_{GS} + R_S i_{DS}$. This is the only way to ensure that the intersection between load line and the real curve gives

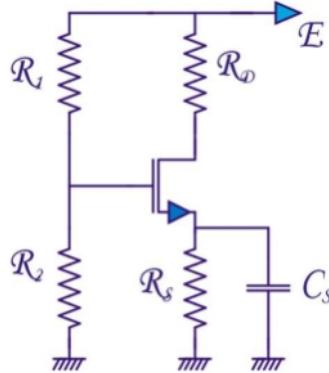


Figure 6.26

a current between $I_{DSQ,min}$ and $I_{DSQ,max}$. The slope of the line between these two points determines thus R_S , while the intersection with the x -axis sets E_G . As there is no DC current through R_G , its value doesn't really matter for setting the operating point. We can choose R_G freely and have an additional degree of freedom to determine R_1 and R_2 from the Thevenin equations⁴. To determine R_D , we apply the same reasoning as for the

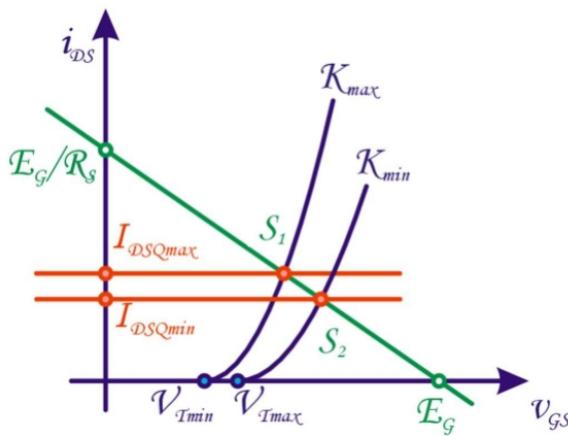


Figure 6.27

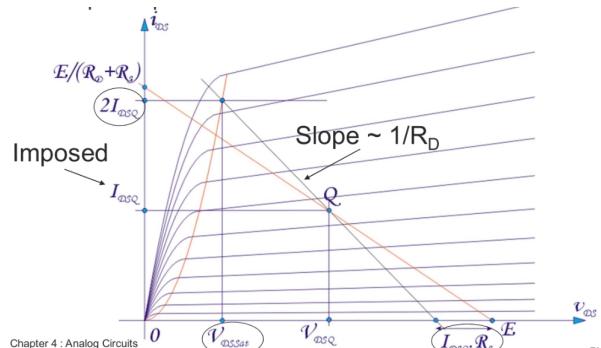


Figure 6.28

BJT: we want to place Q in the middle of the saturation region. The minimum and maximum current is 0 and $2 I_{DSQ}$; the corresponding voltage range is $V_{DS,Sat} < V_{DS} < E - R_S I_{DSQ}$. The value of $V_{DS,Sat} = V_{GSQ} - V_T$, the minimum V_{DS} to stay in saturation, has to be determined by solving $2I_{DSQ} = \frac{K}{2} \frac{W}{L} V_{DS,Sat}^2$ because at the edge of saturation, the required current is $2I_{DSQ}$. We then determine R_D as:

$$R_D = \frac{E - R_S I_{DSQ} - V_{DS,Sat}}{2 I_{DSQ}} \quad (6.25)$$

⁴In the exercises, R_G will be given.

6.5 The Small-Signal Model

Basically, the small-signal model of a (non-linear) component is a representation of the component that can be used as a substitute when we consider small signals. It should only contain linear elements (resistors, inductors, capacitors, linearly depended current or voltage sources, ...) because it is obtained by linearizing the behavior of the component around an operating point Q .

In section 6.2, we established the small-signal model for diode. This was a resistance ρ_d in parallel with a capacitor C_j , as shown in figure 6.16. The values of both elements are set by the operating point (V_{DQ} , I_{DQ}). In this section, we will develop the small-signal model for a bipolar junction transistor and for a MOSFET.

6.5.1 BJT Small-Signal Model

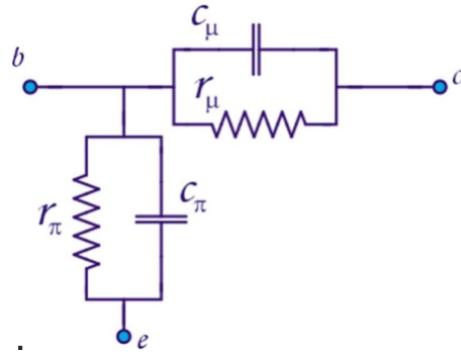


Figure 6.29: Just the pn-junctions

A bipolar junction transistor is nothing else than two pn-junctions put against each other. So we could just put the small-signal model of diode between base and emitter and between base and collector, as in figure 6.29 (for a npn transistor). However, in doing so, we don't have any component that mimics the transistor action, namely the dependence of i_c on i_b . To do this, we define a current gain h_{fe}

$$h_{fe} = \frac{di_C}{di_B} = \frac{i_c}{i_b} = \frac{d\beta i_b}{di_b} = \beta + \frac{d\beta}{di_b} i_b \approx \beta \quad (6.26)$$

and place a dependent current source $h_{fe} i_b$ between collector and emitter. Furthermore, since the output current between collector and emitter also depends on v_{ce} because of the Early effect, we add an "Early" resistor r_c between both terminals. Figure 6.30 gives the entire small-signal model for a npn BJT. The different parameters depend on the biasing conditions:

- Resistance r_π is the diode resistance between base-emitter junction, thus: $r_\pi = \frac{v_{th}}{I_{BQ}}$.
- The base-collector junction is reversed biased, so $r_\mu \approx 0$.
- The Early resistance depends on the Early voltage V_E , which is about 40V. Hence $r_c \approx \frac{V_E}{I_{CQ}}$.

- Often, we express i_c as function of v_{be} . The ratio between both is the *transconductance* g :

$$g = \frac{i_c}{v_{be}} = \frac{i_c}{r_\pi i_b} = \frac{h_{fe}}{r_\pi} \approx \frac{\beta I_{BQ}}{v_{th}} = \frac{I_{CQ}}{v_{th}}$$

By introducing the transconductance, we replace the current-dependent current source $h_{fe} i_b$ with a voltage-dependent current source $g v_{be}$. This is much better, because we can control I_{CQ} by choosing R_E and we can thus set g with high precision. This is not the case for β , as explained previously. The result is shown in the model in figure 6.31, which is also called *Giacoleto's model*.

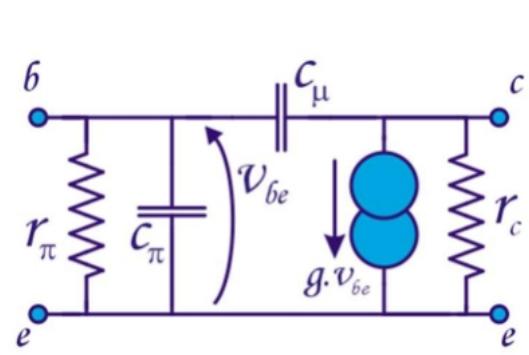
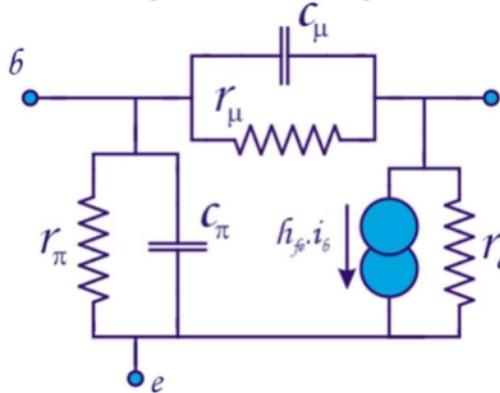


Figure 6.31: Giacoleto's model

Figure 6.30: BJT small signal model

Note that $C_\pi \gg C_\mu$ because the width of the depletion zone is much smaller between emitter and base than between base and collector (the latter is reversed biased while the former is forward biased) and $C \sim \epsilon/t$ with t the thickness of the depletion region. When working at low frequencies, the capacitors are omitted and replaced by open circuits to obtain the low-frequency model.

6.5.2 MOSFET Small-Signal Model

For the MOSFET transistor, we observe that:

- A voltage v_{gs} causes a current between drain and source. We model this by a transconductance g_m .
- Because of channel-length modulation, v_{ds} also has an impact on the current between drain and source. We model this with a resistor r_{ds} .
- The connections between gate and source and gate and drain are capacitors: C_{gs} and C_{gd} .

All these elements are represented in figure 6.32. We can compute g_m , based on the $i_{DS} - v_{GS}$ characteristic (when $v_{GS} > V_T$):

$$\begin{aligned} i_{DS} &= \frac{K}{2} \frac{W}{L} (v_{GS} - V_T)^2 \\ \Rightarrow \frac{di_{DS}}{dv_{GS}} &= K \frac{W}{L} (v_{GS} - V_T) = \frac{2i_{DS}}{v_{GS} - V_T} \\ \Rightarrow g_m &= \frac{di_{DS}}{dv_{GS}} = \frac{i_{ds}}{v_{gs}} = \frac{2I_{DSQ}}{V_{DSQ} - V_T} \end{aligned} \quad (6.27)$$

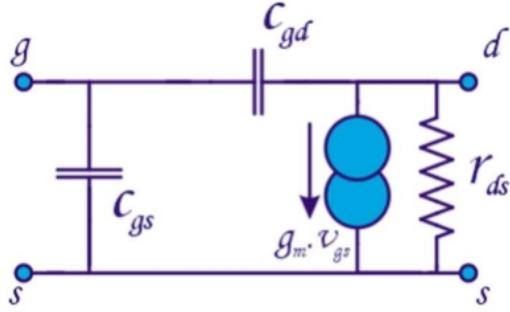


Figure 6.32: MOSFET Small signal model

As for r_{ds} , this quantity is related to the channel-length modulation factor λ from equation 5.6:

$$i_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (v_{GS} - V_T)^2 (1 + \lambda v_{DS}) \quad (6.28)$$

This allows us to compute the change in i_{DS} for small variations of v_{DS} :

$$\begin{aligned} \frac{\partial i_{DS}}{\partial v_{DS}} &= \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (v_{GS} - V_T)^2 \lambda \\ &\approx \lambda I_{DSQ} \end{aligned} \quad (6.29)$$

This expression is the conductivity g_{ds} and thus $r_{ds} = \frac{1}{g_{ds}} \approx \frac{1}{\lambda I_{DSQ}}$.

6.5.3 Orders of magnitude

To estimate the values of the transistor parameters, we will assume that (a) a good value for the Early voltage ≈ 40 V and that (b) the designer should choose a small $V_{DS,Sat}$ to maximize g_m , typically ≈ 200 mV.

For the bipolar transistor, we observe that:

- $g_\pi = \frac{1}{r_\pi} = \frac{I_{BQ}}{v_{th}} = \frac{I_{CQ}}{\beta v_{th}} = \frac{g}{\beta}$
- $g_c = \frac{1}{r_c} = \frac{I_{CQ}}{V_E} = \frac{v_{th}}{V_E} \frac{I_{CQ}}{v_{th}} \approx \frac{g}{1600}$

and thus: $g \gg g_\pi \gg g_c$.

For the MOSFET, we find that:

$$\bullet \quad g_{ds} = \frac{1}{r_{ds}} = \frac{I_{DSQ}}{V_E} = \frac{v_{DS,Sat}}{2V_E} \frac{2I_{DSQ}}{v_{DS,Sat}} \approx \frac{g_m}{400}$$

and thus: $g_m \gg g_{ds}$.

Chapter 7

Amplifiers

In this section, we will study amplifiers: circuits that take a signal as input and produce an identical but magnified version at the output. We'll see the basic amplifier and the more stable four-resistor version. Next, we study the most common amplifier topologies - common emitter, common base and common collector - and what are their advantages and drawbacks. Finally, we study the differential amplifier and the operational amplifier or OPAMP.

7.1 Basic Amplifier

In this section, we will develop and improve a basic amplifier. The elementary circuit we will be using is shown in figure 7.1. Bias currents I_{BQ} and I_{CQ} are generated by the DC voltage source E_B . The time-varying voltage source v_i is the input signal, and is applied at the base of the transistor¹. The output is measured at the collector.

As v_i increases, so does i_B , just as in figure 6.14. If the transistor is biased in the normal operating region, then $i_C = \beta i_B$ will also increase. As we move along the load line with increasing i_C , the voltage drop along R_C increases and the output voltage v_o decreases. We want to compute the voltage gain A_v . But before we do that, we first see how the input part of the circuit in figure 7.1 can be constructed.

7.1.1 Coupling Capacitance

Let's compute the corresponding Thevenin equivalent of the circuit in figure 7.2. To find e_0 , apply Millman's theorem:

$$\begin{aligned} e_0 &= \frac{E/R_1 + e_i j\omega C_B}{1/R_1 + 1/R_2 + j\omega C_B} \\ &= \frac{R_2 E + e_i j\omega C_B R_2}{R_1 + R_2 + j\omega C_B R_1 R_2} \\ &= E \frac{R_2}{R_1 + R_2} \frac{1}{1 + j\omega C_B R_B} + e_i \frac{j\omega C_B R_B}{1 + j\omega C_B R_B} \end{aligned} \tag{7.1}$$

with $R_B = \frac{R_1 R_2}{R_1 + R_2}$. If ω is much smaller than the critical pulsation $\omega_c = \frac{1}{R_B C_B}$, then $e_0 \approx E \frac{R_2}{R_1 + R_2}$. If $\omega \gg \omega_c$, then $e_0 \approx e_i$. The latter condition corresponds to assuming C_B is a

¹Note that in reality this is the result of applying Thevenin's theorem to resistors R_1 and R_2

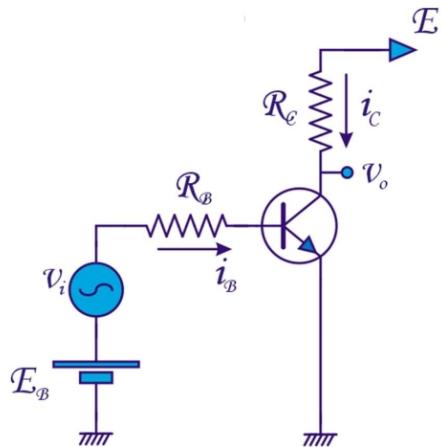


Figure 7.1: Simple amplifier

short-circuit. Another way to find equation 7.1, is to use the superposition principle: first, consider only E with $e_i = 0$, then consider e_i with $E = 0$, and add both results.

The circuit is thus equivalent to a DC source $E_B = E \frac{R_2}{R_1 + R_2}$ in series with a small-signal, high-frequency source e_i , just as in figure 7.3. This means we can use this circuit to couple e_i to the input of the amplifier, while keeping the DC biasing, just as in figure 7.1. Capacitor C_B is a *coupling capacitor* because it "couples" v_i into the circuit.

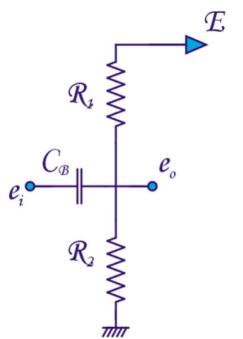


Figure 7.2

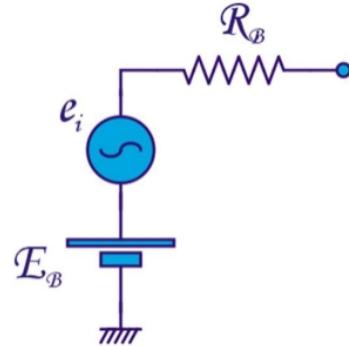


Figure 7.3

Voltage gain A_v

To calculate A_v , we draw the small-signal equivalent circuit, by replacing the npn-transistor by its small-signal model, and by grounding the DC voltage source E . We also assume that the frequency we consider is higher than $\frac{1}{R_B C_B}$, so we can consider C_B as a short-circuit. The small-signal circuit is shown in figure 7.4. The parameters of the model are set by the operating currents and voltages:

- $r_\pi = \frac{v_{th}}{I_{BQ}}$,

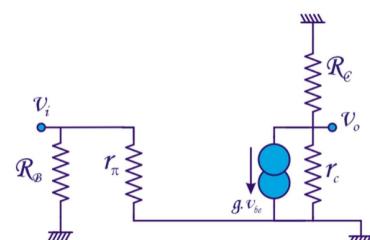


Figure 7.4: Simple amplifier - small-signal model

- $r_c = \frac{V_E}{I_{CQ}}$,

- $g = \frac{I_{CQ}}{v_{th}}$

In the equivalent circuit, we apply Millman in v_o :

$$v_o = \frac{-g v_{be}}{g_c + G_C}$$

and we see that $v_{be} = v_i$. As a consequence:

$$A_v = \frac{v_o}{v_i} = -g (r_c || R_C) \quad (7.2)$$

For a MOSFET, we would have found a similar expression: $A_v = -g_m(r_{ds} || R_D)$. Note that the gain is negative, because an increase in i_b and thus in i_c will lead to a larger drop across R_C and will decrease v_o , as explained previously.

The gain can be increased by:

1. Increasing g by setting a higher I_{CQ} . This will also decrease r_c , but this is usually not a problem since most of the times $r_c \gg R_C$ and hence $R_C || r_c \approx R_C$. In that case, $A_v \approx -g R_C$.
2. Increase R_C . The drawback is that this decrease the potential swing of v_o .

The maximum voltage gain we can obtain for a BJT is found when $R_C \rightarrow \infty$. Then is

$$A_{v,max} = -gr_c = -\frac{I_{CQ}}{v_{th}} \frac{V_E}{I_{CQ}} \approx -40 \times \frac{1}{0.026} \approx -1600$$

For a MOSFET, $A_{v,max}$ is about -400 .

7.1.2 The 4-resistor amplifier

We will study a more general circuit, namely the amplifier with 4 biasing resistors that we saw before and is reproduced in figure 7.5(left). As before, we assume that the input frequency of interest is such that we can consider C_B as a short circuit. Note that the small-signal circuit would be the same for an n-channel MOSFET, if $r_\pi \rightarrow \infty$.

To compute A_v , apply Millmnn at both the emitter and collector (output) node:

- At collector: $v_o = \frac{g_c v_e - g(v_i - v_e)}{G_c + g_c}$ because $v_{be} = v_i - v_e$. Note that we used conductivities. For example, $G_c = 1/R_c$.
- At emitter: $v_e = \frac{g_\pi v_i g_c v_o + g(v_i - v_e)}{G_E + g_c + g_\pi}$

After eliminating v_e , we obtain:

$$\begin{aligned} A_v &= \frac{-g R_C r_c r_\pi + R_C R_E}{(r_c + R_C)(R_E + r_\pi) + r_\pi R_E (1 + gr_c)} \\ &= -g \frac{r_c R_C}{r_c + R_C} \frac{r_\pi - \frac{R_E}{gr_c}}{R_E + r_\pi (1 + R_E \frac{1+gr_c}{r_c+R_C})} \\ &\approx -\frac{R_C}{R_E} \end{aligned} \quad (7.3)$$

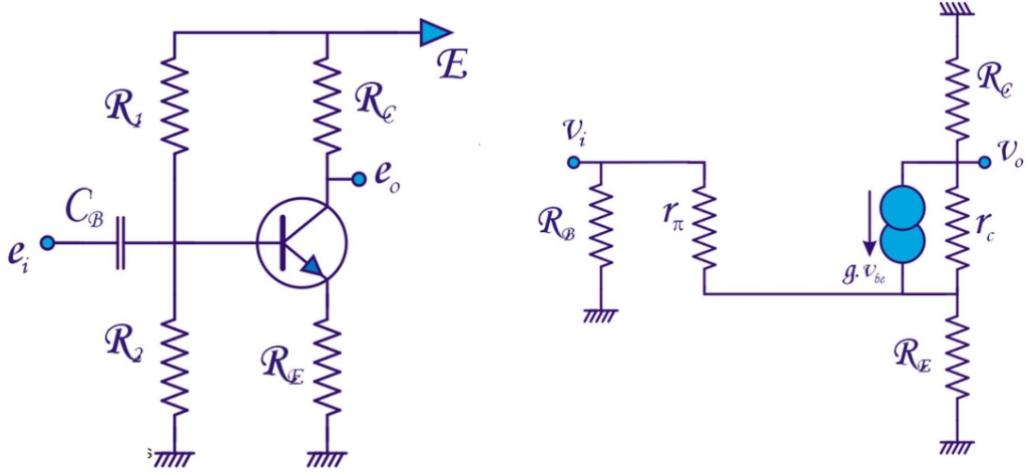


Figure 7.5: Four-resistor amplifier (left) and small-signal equivalent circuit (right)

where we assumed that $r_c \gg R_C$ and $gr_c \gg 1$.

It is logical that $A_v \approx -R_C/R_E$: in a first approximation, as the base voltage increases with v_i , the emitter voltage will follow because they are tied together by the drop across base-emitter junction, which is about 0.6 V. If the emitter voltage increases by v_i , the emitter current will increase by $\frac{v_i}{R_E}$. This is also approximately the increase in collector current, thus the voltage drop increase at R_C is equal to $v_o \approx -R_C \frac{v_i}{R_E}$.

Comparing this result to the version with no emitter resistance, for which $A_v \approx -g R_C$. For $R_C = 1k\Omega$, $R_E = 500\Omega$ and $g = 40mA/V$. Without R_E , we find a gain of 40, while with R_E , the gain is 2. Thus, while R_E is necessary to obtain a stable bias point, its presence significantly reduces the gain. Hence we use a bypass capacitor as in figure 6.23.

Frequency analysis

We will study how the amplifier gain $A_v = \frac{v_o}{v_i}$ depends on frequency. To do this, we establish the small-signal circuit for the four-resistor amplifier with bypass capacitance C_E and coupling capacitance C_B , as in figure 7.6. The parallel combination of R_E and C_E gives:

$$Z_E = \frac{R_E}{1 + j\omega R_E C_E} \quad (7.4)$$

and we substitute this expression for R_E in equation 7.3. At the same time, we multiply by $\frac{j\omega R_B C_B}{1 + j\omega R_B C_B}$, just as in equation 7.1.

$$A_v = \frac{j\omega R_B C_B}{1 + j\omega R_B C_B} \frac{-g R_C r_c r_\pi + R_C Z_E}{(r_c + R_C)(Z_E + r_\pi) + r_\pi Z_E (1 + gr_c)} \quad (7.5)$$

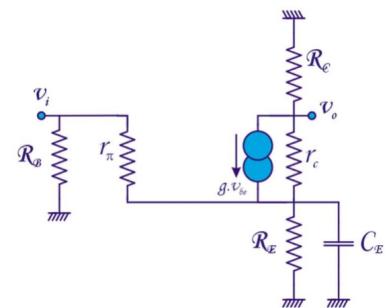


Figure 7.6

If we set $T_B = 1/(R_B C_B)$ and $T_E = 1/(R_E C_E)$, this equation can be simplified (see appendix ??) to:

$$A_v \approx -A_{vE} \frac{j\omega T_B}{1 + j\omega T_B} \frac{1 + j\omega T_E}{1 + j\omega T_E \frac{A_{vE}}{A_{v0}}} \quad (7.6)$$

with $A_{vE} = \frac{R_C}{R_E}$ and $A_{v0} = g \frac{R_C r_c}{R_C + r_c}$. This transmittance has two poles in $\omega = 1/T_B$ and $\omega = \frac{A_{v0}}{T_E A_{vE}}$ and zeros in $\omega = 0$ and $\omega = 1/T_E$. With $A_{v0} \gg A_{vE}$ and a correct choice for $R_B C_B$ and $R_E C_E$, we have:

$$1/T_B < 1/T_E < \frac{A_{v0}}{T_E A_{vE}}$$

Figure 7.7 shows the bode plot of $A_v(\omega)$. We find four different domains based on the

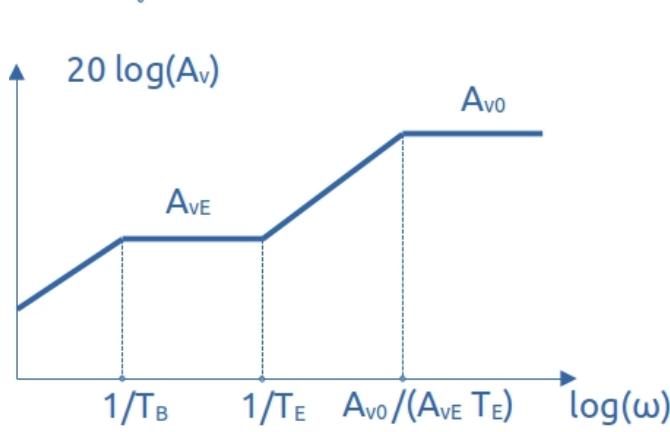


Figure 7.7: 4R A_v Bode Curve

frequency of the input signal:

1. $\omega < 1/T_B$: the signal v_i is not yet coupled into the base of the amplifier through capacitance C_B .
2. $1/T_B < \omega < 1/T_E$: capacitance C_B can be considered as a short circuit. However, ω is too small to short-circuit C_E and bypass R_E . The gain is thus about $-R_C/R_E$.
3. $1/T_E < \omega < A_{v0}/(A_{vE} T_E)$: as ω increases, R_E starts to get bypassed.
4. $\omega > A_{v0}/(A_{vE} T_E)$: the maximum gain (with bypassed R_E) is reached: $A_{v0} = -g(R_C||r_c)$. This is the domain in which we want to use the amplifier.

It is important to note that the critical pulsation is not $1/T_E$ but rather $\omega_{crit} = \frac{A_{v0}}{A_{vE} T_E}$. Note also that $A_{vE} \approx -\frac{R_C}{R_E}$ and $A_{v0} \approx -g R_C$, so the critical pulsation is $\frac{1}{T_E} \frac{A_{v0}}{A_{vE}} \approx \frac{1}{R_E C_E} \frac{g R_C}{\frac{R_C}{R_E}} = \frac{g}{C_E}$.

7.2 Basic Topologies

7.2.1 Common Emitter Amplifier (CEA)

The amplifier configuration of the previous section is the *common-emitter* configuration: input is at the transistor base (gate) and the output is at the collector (drain). For the frequencies

of interest, the emitter (source) is bound to ground and thus at a "common" voltage. Other configurations are the common base and common collector. We will study these configurations in the next sections.

Our goal is to establish the voltage gain, and input- and output impedances for the common-emitter configuration. For this purpose, we once again draw the small-signal equivalent circuit as in figure 7.8.

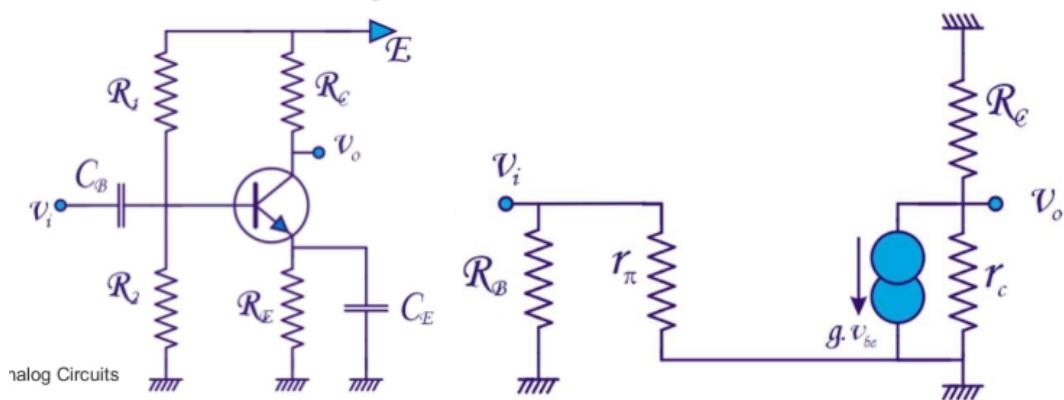


Figure 7.8: CEA: circuit (left) and small-signal equivalent (right)

- Voltage gain A_v : For an input voltage v_i , $v_{be} = v_i$. Thus $v_o = -g v_i (R_C || r_c)$ and $A_v = -g(R_C || r_c) = -\frac{g}{g_c + G_C}$ as we've seen before.
- Input impedance $Z_i = v_i / i_i = \frac{1}{g_\pi + G_B} = r_\pi || R_B$.
- Output impedance: to compute Z_o :

- Shorten v_i to ground,
- Apply v_o (or i_o) to the output,
- Compute i_o (or v_o),
- The output impedance $Z_o = \frac{v_o}{i_o}$.

The effect of shortening the input to ground is that $v_{be} = 0$ is in figure 7.8. Thus the current source with transconductance g can be omitted. We see that then the output impedance is the parallel combination of R_C and r_c : $Z_o = \frac{1}{G_C + g_c}$.

7.2.2 Common Base Amplifier (CBA)

In a common-base configuration, the base of the transistor is kept at a constant voltage (i.e. an AC ground). The input signal is applied to the emitter and the output voltage is measured at the collector. The circuit is shown in the left part of figure 7.9, with the small-signal circuit on the right. Obviously, a bypass capacitance is not added.

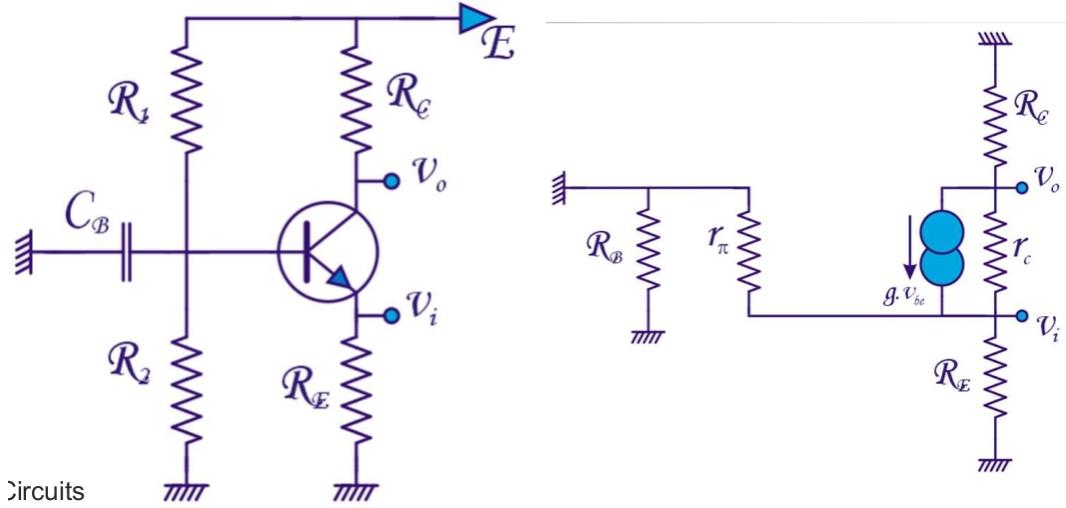


Figure 7.9: CBA: circuit (left) and small-signal equivalent (right)

- Voltage gain A_v : we see that $v_{be} = -v_i$. In the output node, we can write:

$$\begin{aligned} v_o &= \frac{g_c v_i - g(-v_i)}{G_C + g_c} \\ \rightarrow A_v &= \frac{v_o}{v_i} = \frac{g + g_c}{G_C + g_c} \\ &\approx \frac{g}{G_C + g_c} \end{aligned} \quad (7.7)$$

because $g \gg g_c$. Notice how the gain is positive.

- Input impedance Z_i : We compute the current drawn at the input node:

$$i_i = G_E v_i + g_\pi v_i + g_c(v_i - v_o) - g v_i \quad (7.8)$$

Substituting the expression for v_o : $v_o = \frac{g+g_c}{G_C+g_c} v_i$ into this equation gives:

$$\begin{aligned} Z_i &= \frac{g_c + G_c}{(g + g_c)G_C + (g_c + G_C)(g_\pi + G_E)} \\ &\approx \frac{1}{G_E + g} \end{aligned} \quad (7.9)$$

This last expression is the parallel combination of R_E with a resistance $\frac{1}{g}$; if we look in the emitter (or source), we see an impedance $1/g$ (or $1/g_m$).

- Output impedance Z_o : by shorting the input, $v_{be} = 0$ thus there is no current through the transconductance. We see R_C in parallel with r_c :

$$\Rightarrow Z_o = \frac{1}{G_C + g_c},$$

just as for the CEA.

7.2.3 Common Collector Amplifier (CCA)

In a common collector amplifier, the input is applied to the base, and the output is measured at the emitter, as in figure 7.10. The common node - the AC ground - is the collector, thus no collector resistance R_C is needed. Furthermore, we don't use a decoupling capacitor C_E . In a first approximation, we can say that $v_I - v_O = v_{BE} \approx 0.6$ V and remains constant. That's why $\frac{v_o}{v_i} \approx 1$ and we also call this topology a *follower* because the output follows the input.

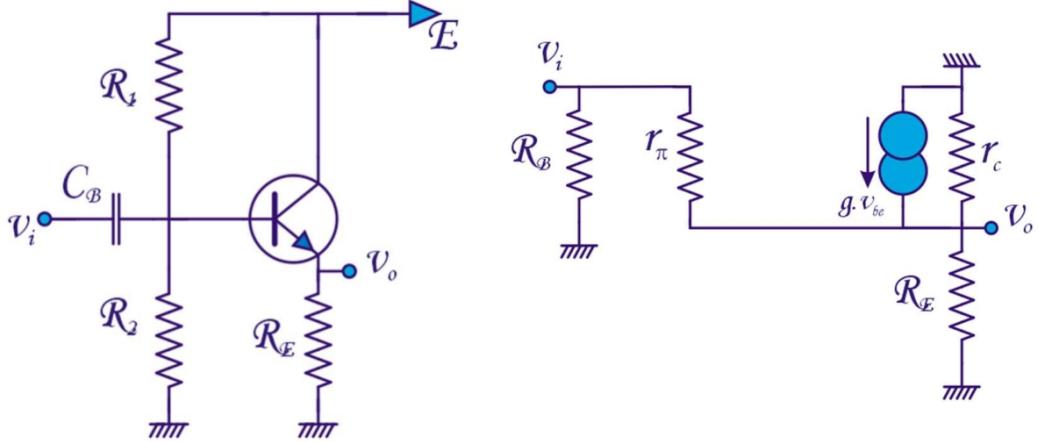


Figure 7.10: CCA: circuit (left) and small-signal equivalent (right)

- Voltage gain A_v : we see that $v_{be} = v_i - v_o$. In the output node, we can write:

$$\begin{aligned}
 v_o &= \frac{-g v_{be} + g_\pi v_i}{g_\pi + g_c + G_E} \\
 &= \frac{-g (v_o - v_i) + g_\pi v_i}{g_\pi + g_c + G_E} \\
 \rightarrow (g + g_\pi + g_c + G_E) v_o &= (g + g_\pi) v_i \\
 \rightarrow A_v = \frac{v_o}{v_i} &= \frac{g + g_\pi}{g + g_\pi + g_c + G_E} \approx \frac{g}{g + G_E} \approx 1
 \end{aligned} \tag{7.10}$$

- Input impedance Z_i .

Consider the circuit initially without R_B . Then

$$\begin{aligned}
 i_o &= g_\pi(v_i - v_o) \\
 &= g_\pi\left(1 - \frac{g + g_\pi}{g + g_\pi + g_c + G_E}\right)v_i \\
 &= g_\pi\left(\frac{g_c + G_E}{g + g_\pi + g_c + G_E}\right)v_i \\
 &\approx g_\pi \frac{G_E}{g} \\
 \rightarrow Z_i &= R_E \frac{g}{g_\pi} = \beta R_E
 \end{aligned} \tag{7.11}$$

and thus $Z_i = \beta R_E \parallel R_B$.

- Output impedance Z_o :

$$i_o = (G_E + g_c + g_\pi + g) v_o$$

and thus

$$Z_o = \frac{1}{G_E + g_c + g_\pi + g} \approx \frac{1}{g + G_E} \approx \frac{1}{g}$$

7.2.4 Comparison of Topologies

The characteristics of the different topologies - both for BJT as for the MOSFET (Common source, gate, and drain configurations) - are summarized in table 7.2.4. We only use intrinsic parameters of the transistors.

	Z_i	Z_o	$ A_v $
CEA	r_π	r_c	$g r_c$
CBA	$1/g$	r_c	$g r_c$
CCA	βR_E	$1/g$	1
CSA	∞	r_{ds}	$g_m r_{ds}$
CGA	$1/g_m$	r_{ds}	$g_m r_{ds}$
CDA	∞	$1/g_m$	1

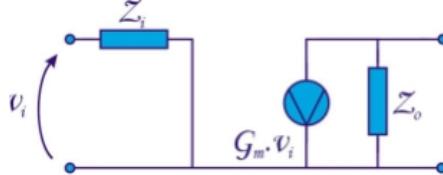


Figure 7.11

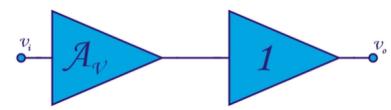


Figure 7.12

Figure 7.11, the schematic representation of an amplifier as seen in chapter 1, shows that as a general rule $|A_v| = g Z_o$. This can be verified in table 7.2.4. From this figure, we can also deduce that a good amplifier needs (a) a high input impedance to avoid drawing a large current from the previous stage and (b) a low output impedance to avoid making the output voltage depended on the impedance of the next stage. The only suitable configuration is the common collector (or CDA), but this amplifier has a gain of ≈ 1 . To implement a good amplifier, we need a cascade of:

- An amplifier with high gain, medium Z_i and high Z_o ,
- A buffer stage with gain ≈ 1 , high Z_i and low Z_o , as in figure 7.12.

Consequently, the gain will be high, and in- and output impedances will be as required.

7.3 Difference Amplifier

7.3.1 Definition

Until now, the amplifiers we studied only had a single input terminal and a single output terminal. We would like take create an amplifier that amplifies the *difference* between two voltage. This can be useful to compare two signals as in an operational amplifier, or to remove

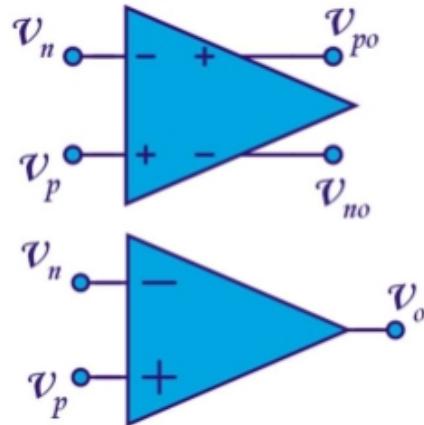


Figure 7.13: Differential amplifier with two (top) or one (bottom) output.

the noise on two signals when this noise is significantly correlated (e.g. because it was created by the same noise source). There are 2 inputs v_p and v_n (the positive and negative terminal). The goal is to amplify the difference $v_p - v_n$. This can be done with one or two outputs. In the latter case - also called the *differential case* - the voltage between outputs v_{po} and v_{no} is:

$$v_{os} = v_{po} - v_{no} = A_v (v_p - v_n)$$

as in the top of figure 7.13. The other option is the single-output, asymmetrical amplifier:

$$v_o = A_v (v_p - v_n)$$

To make the analysis easier, we split the signal in two components:

- A *differential mode*: $v_d = v_p - v_n$,
- A *common mode* $v_c = \frac{v_p + v_n}{2}$

This means that we can write $v_p = v_c + \frac{v_d}{2}$ and $v_n = v_c - \frac{v_d}{2}$. In a typical scenario, the common-mode signal can be a lot larger (order of magnitude several volts) than the differential mode (several milivolts).

With a single output, the output voltage is thus $v_o = A_d v_d + A_c v_c$:

- The differential gain A_d , which is actually what we want, so it has to be as high as possible.
- The common mode gain A_c , which is a gain that we want to keep as low as possible - we want to reject the common mode. Ideally, it should be zero, but we will see that this is not possible (at least for a single-output amplifier).

We also define two so-called rejection ratios: the common-mode rejection ratio (CMRR) which is the ratio between A_d and A_c : $CMRR = A_d/A_c$, and the power supply rejection ratio PSRR, which expresses how variations in the power supply have an impact on the output. Both rejection ratios should be as high as possible, to remove unwanted, parasitic components in the output signal. They are typically expressed in decibel and in a good amplifier, are about 100 – 120 dB.

7.3.2 Implementation

The circuit to accomplish the requirements from the previous section, is shown in figure 7.14. We use two branches with identical npn transistors and identical collector resistances R . Both emitters are tied to a common node with voltage v_e . The bases of both transistors are the inputs. The outputs are measured at or between the collectors. The resistance R_E carries a current I_{RE} .

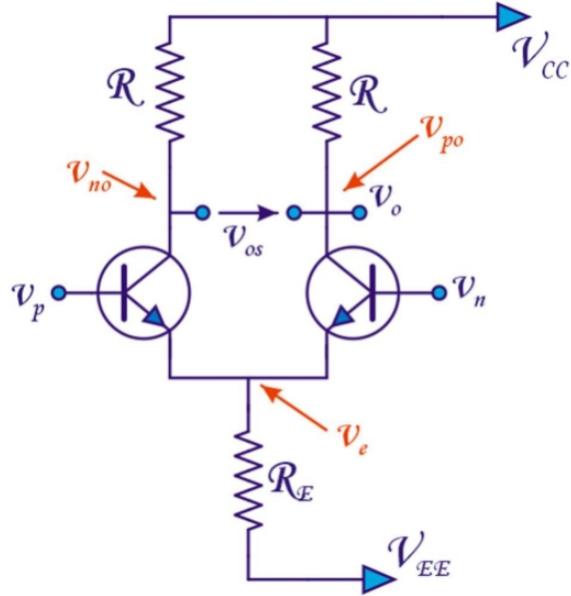


Figure 7.14: Structure of the differential amplifier

7.3.3 Common Mode

If the input nodes v_n and v_p are at the same voltage, $v_n = v_p = v_c$, the two branches carry an identical current $I_{RE}/2$ and the outputs v_{no} and v_{po} are both equal to $V_{CC} - R I_{RE}/2$. The differential output v_{os} is than zero. Because of the symmetry in the circuit, this output is very robust and depends only on the differential signals. However, if there is some mismatch between both branches, e.g. the transistors or resistances are not completely identical, this is no longer the case.

When $v_n = v_p = v_c$, then $v_e = v_c - V_{BEQ} \approx v_c - 0.6$ V. The current in both branches is equal to $i_c = \frac{I_{RE}}{2} = \frac{v_e - V_{EE}}{2R_E}$ and $v_{op} = v_{on} = v_o = V_{CC} - R i_c$. The symmetrical output $v_{os} = v_{po} - v_{no}$ is still zero, as it should be. The single output v_o however, varies with v_c . This variation should be as small as possible.

To study the common-mode response, we can split the circuit into two parts. In a first stage, we split resistance R_E into two parallel resistors of value $2R_E$, as in figure 7.15. Because of the symmetry and because both inputs are the same, there can be no current in the wire connecting both emitters, so we remove it. So, in effect, we can study both halves independently which means we only have to analyze the circuit in figure 7.16.

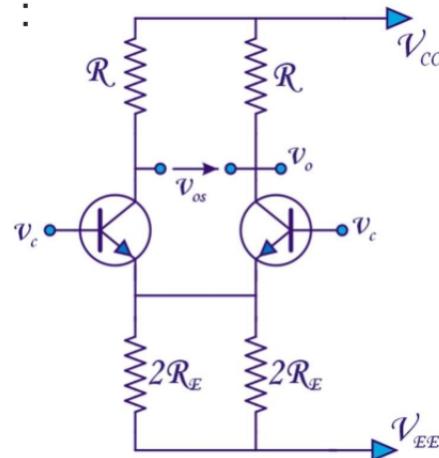


Figure 7.15

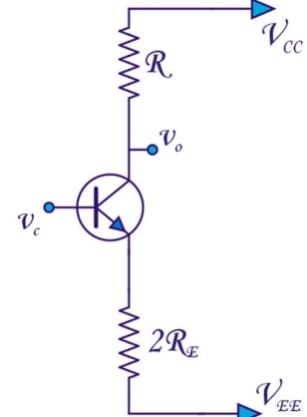


Figure 7.16

7.3.4 Differential Mode

With the common mode v_c is zero, $v_p = +v_d/2$ and $v_n = -v_d/2$. This is a purely differential signal: if v_p increases, v_n would decrease with the same amount. To demonstrate that in this case $v_e = 0$, we draw the small-signal equivalent circuit as in figure 7.17 and use Millman's theorem to compute the voltage in v_{on} , v_{op} and v_e . Note that v_{be} in the left branch is $\frac{v_d}{2} - v_e$, and $-\frac{v_d}{2} - v_e$ in the right branch.

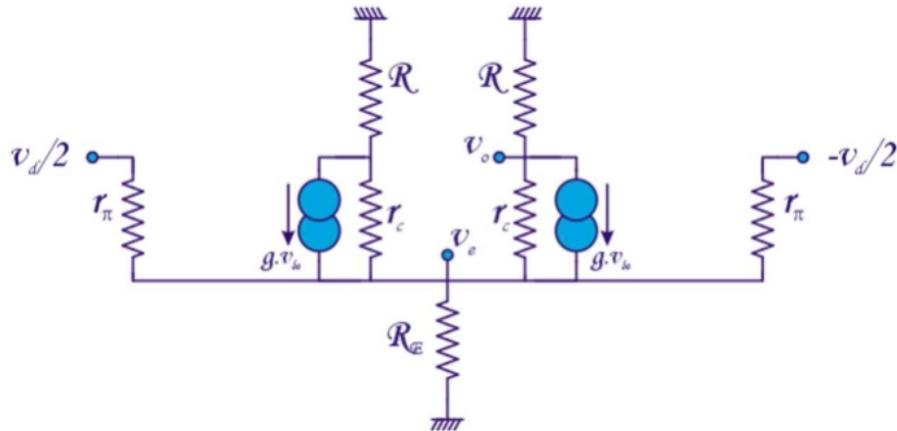


Figure 7.17: Differential mode: small-signal model

$$\bullet \quad v_{po} = \frac{g_c v_e - g(-\frac{v_d}{2} - v_e)}{G + g_c},$$

$$\bullet \quad v_{no} = \frac{g_c v_e - g(\frac{v_d}{2} - v_e)}{G + g_c},$$

$$\bullet \quad v_e = \frac{g_c v_{po} + g_c v_{no} + g(-\frac{v_d}{2} - v_e) + g(\frac{v_d}{2} - v_e) - g_\pi \frac{v_d}{2} + g_\pi \frac{v_d}{2}}{G_E + 2g_c + 2g_\pi}$$

Substituting the expressions for v_{on} and v_{op} in the one for v_e gives:

$$v_e = \frac{g_c \frac{g_c v_e + g v_e}{g_c + G} + g_c \frac{g_c v_e + g v_e}{g_c + G} - g v_e - g v_e}{G_E + 2g_c + 2g_\pi}$$

Because of the symmetries, all mentions of v_d have disappeared in this expression. The solution is $v_e = 0$.

Building on the knowledge that $v_e = 0$ in a purely differential input signal (i.e. v_E doesn't change when both v_{op} and v_{on} change with equal magnitude but opposite sign), we can draw the following circuit for the differential input of figure 7.18. In this circuit, the current through resistance R_E is constant because v_E is constant for differential input signals. The collector current I_{CQ} is constant and set by the common mode: $I_{CQ} = I_{RE}/2$.

Since we apply $+v_d/2$ to the left input, the current in the left loop increases by $i_c \approx g v_{be} = g v_d/2$ (remember: $v_e = 0$). The current in the right loop decreases by the same amount: $i_c \approx -g v_d/2$. So any additional current in the left loop flows in the right loop, keeping the current through R_E constant, as established previously. The output voltages change as well: $v_{on} = -R g v_d/2$ and $v_{op} = +R g v_d/2$ and thus $v_{os} = g R v_d$.

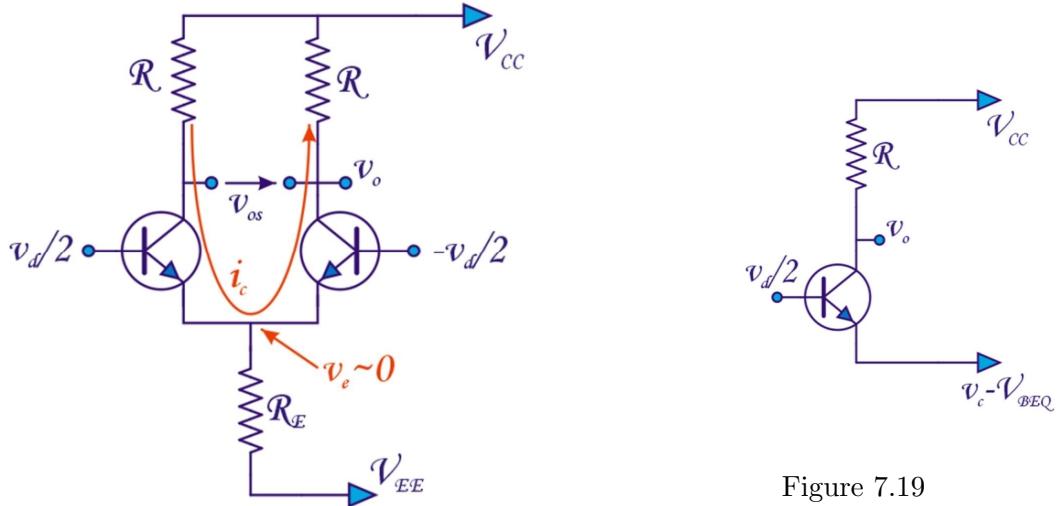


Figure 7.19

Figure 7.18

We can thus summarize that when we apply a differential signal, the emitter voltage v_E stays fixed and equal to $\approx v_c - 0.6$ V and the circuit can be studied by considering only one branch, namely the one in figure 7.19.

This means we have an equivalent circuit to study the common mode in figure 7.16 and a circuit to study the differential mode, in figure 7.19.

7.3.5 Load-line Analysis

The current i_C is determined by the common mode, from figure 7.16:

$$i_C = \frac{v_c - V_{BEQ} - V_{EE}}{2R_E} \quad (7.12)$$

The load line equation is given by:

$$V_{CC} - (v_c - V_{BEQ}) = Ri_C + v_{CE} \quad (7.13)$$

from figure 7.19.

The operating point Q is found by the intersection of these two lines, as in figure 7.20. Note how v_c impacts both lines: the current shifts up and down, and the load line moves parallel with varying v_c . This observation makes it very difficult to control the operating point. Commonly, the outer limits of v_c are given, so an estimate of the range of Q can be made. As v_d changes, one transistor moves up along the load line, while the other one moves down because the load line equation was derived from the circuit where we supposed the input was symmetrical around v_c .

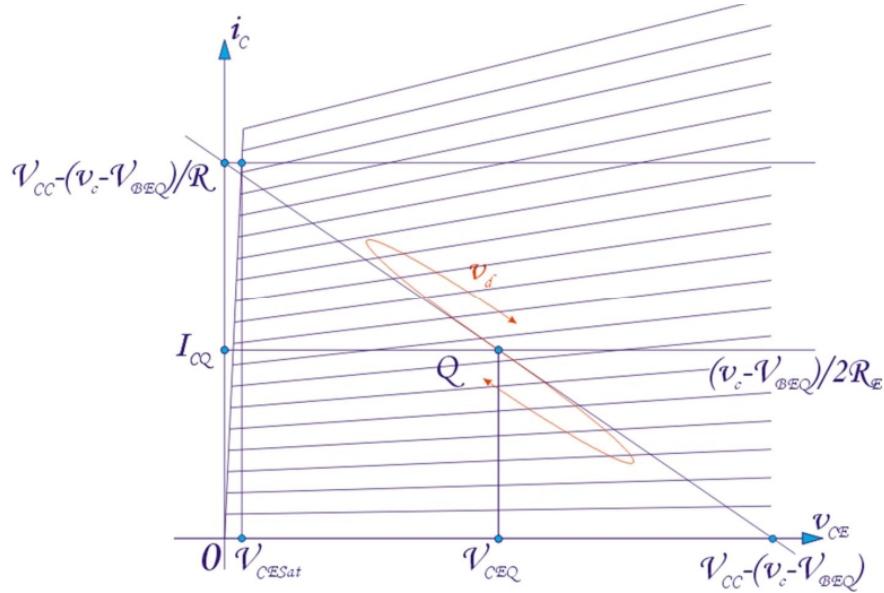


Figure 7.20

7.3.6 Common & Differential Gain

The *differential gain* A_d can easily be computed by transforming the circuit in figure 7.19 to its small-signal equivalent. The emitter voltage is an AC ground because this voltage doesn't change with v_d . The circuit is in fact a common-emitter amplifier. With r_c in parallel with R and $v_{be} = v_d/2$, we find:

$$A_d = \frac{v_o}{v_d} = -\frac{g(R||r_c)}{2} \quad (7.14)$$

The *common-mode gain* A_c can be calculated by using the small-signal equivalent circuit for the circuit in figure 7.16. The result is found in figure 7.21.

This circuit is the same as the one in figure 7.5, but with $2R_E$ in stead of R_E . So we can reuse the result from equation 7.3:

$$\begin{aligned} A_c &= \frac{v_o}{v_c} = \frac{-gRr_cr_\pi + 2RR_E}{(r_c + R)(2R_E + r_\pi) + 2r_\pi R_E(1 + gr_c)} \\ &\approx -\frac{R}{2R_E} \end{aligned} \quad (7.15)$$

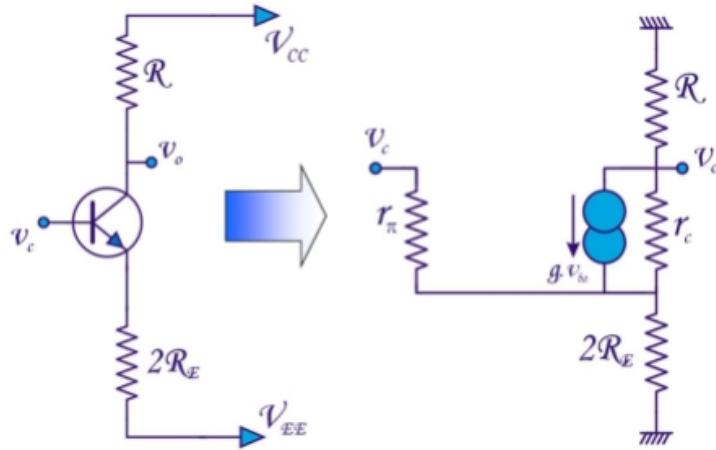


Figure 7.21

for a single output. For the differential output, $A_c = 0$.

From the results in equations 7.14 and 7.15, we find that:

$$\begin{aligned} CMRR &= \frac{A_d}{A_c} \approx \frac{-\frac{gR}{2}}{-\frac{R}{2R_E}} \\ &\approx g R_E \end{aligned} \quad (7.16)$$

This means that if we want to increase CMRR, we have to increase R_E . However, if we increase R_E , and since the voltage at the emitters is equal to $v_c - V_{BEQ}$, the current I_{RE} through R_E decreases. For each transistor, we have $I_{CQ} = I_{RE}/2$, and $g = \frac{I_{CQ}}{v_{th}}$. So each increase in R_E will lead to a decrease in I_{CQ} and thus also in a proportional decrease in g . Hence the CMRR will remain about the same.

A better way to increase R_E is by using the circuit in figure 7.22, where we add an additional transistor between R_E and v_E . We can explain the functioning as follows: with a fixed V_{BB} , the voltage at the emitter of the added transistor is $V_{BB} - V_{BEQ} = V_{BB} - 0.6$ V. This voltage is (relatively) fixed, and controls the current I_{CQ} :

$$I_{CQ} = \frac{1}{2} \frac{V_{BB} - V_{BEQ} - V_{EE}}{R_E}$$

This current is fixed and doesn't depend (in a first approximation) on the collector voltage of the transistor (as long as this voltage is high enough). As the collector voltage can vary and the current remains constant, the collector sees a very large resistance.

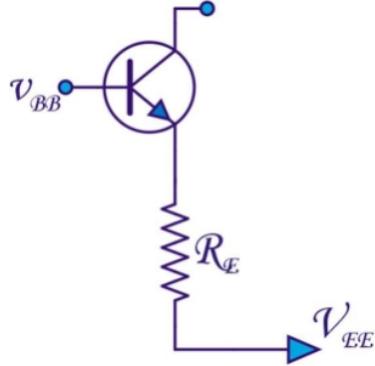


Figure 7.22

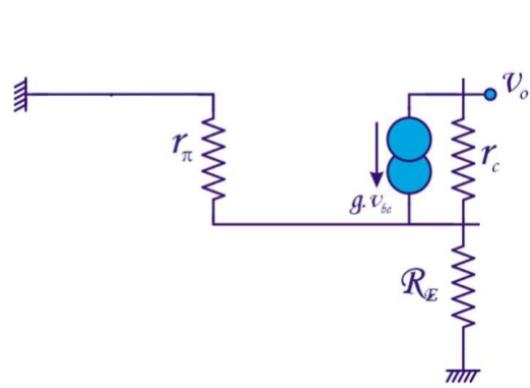


Figure 7.23

More formally, we can sketch the small-signal circuit as in figure 7.23 and compute the output impedance as seen in node v_o :

$$i_o = g_c(v_o - (-v_{be})) + gv_{be}$$

Since i_o also passes through the parallel combination of R_E and r_π , we can state that:

$$i_o = -(G_E + g_\pi)v_{be}$$

and thus:

$$\begin{aligned} i_o &= g_c v_o - (g + g_c) \frac{i_o}{G_E + g_\pi} \\ (G_E + g_\pi)i_o &= (G_E + g_\pi)g_c v_o - (g + g_c)i_o \\ (G_E + g_\pi + g + g_c)i_o &= (G_E + g_\pi)g_c v_o \\ \Rightarrow Z_o &= \frac{v_o}{i_o} = \frac{G_E + g_\pi + g + g_c}{(G_E + g_\pi)g_c} \\ &\approx \frac{g}{g_c G_E} \\ \Rightarrow Z_o &\approx (gr_c) R_E \end{aligned} \tag{7.17}$$

The resistance R_E is thus multiplied by the intrinsic transistor gain $gr_c \approx 1600$. The CMRR is thus $g(gr_c)R_E = g^2r_cR_E$.

7.3.7 Power-Supply Rejection Ratio

To compute the power supply gain A_{cc} , we put the two inputs to AC ground and compute how v_o varies if the power supply undergoes a voltage change v_{cc} . Using the small-signal equivalent circuit of figure 7.19 with $v_d = 0$, we find that $v_o = \frac{r_c}{r_c + R_C}v_{cc} \approx v_{cc}$ so the supply voltage variation is almost completely transferred to the output: $A_{cc} \approx 1$. The power supply rejection ratio PSRR = $\left| \frac{A_d}{A_{cc}} \right| \approx |A_d| \approx \frac{gR}{2}$.

7.4 Operational Amplifier

An operational amplifier or *OPAMP* is basically a differential amplifier with very high gain. It is typically constructed with a differential amplifier as input stage and an additional common-emitter stage to boost the gain of the differential amplifier. Schematically, we can see the internal structure of an OPAMP as in figure 7.24 where there are 4 stages:

1. A single-output differential amplifier as input stage,
2. A buffer stage to avoid that the next stage loads the output of the differential amplifier. This buffer stage has a high input impedance and low output impedance and a gain $A_v \approx 1$, like a common-collector amplifier.
3. A high-gain stage, like a common-emitter amplifier,
4. Another buffer stage to isolate the load circuitry from the OPAMP.

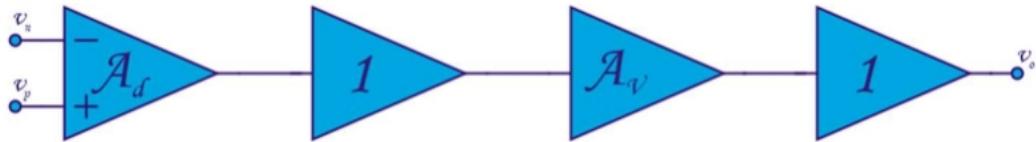


Figure 7.24

An OPAMP is very close to an ideal amplifier, which we will see in the next section. After that, we study the real OPAMP and see how non-idealities will impact its behavior.

7.4.1 The Ideal Amplifier

An OPAMP is represented by the symbol below. Just as the differential amplifier, it has two inputs, and the goal is to amplify the difference $v_i = v_p - v_n$, so that $v_o = A_v v_i$. Often, we define v_i in the other sense ($v_i = v_n - v_p$) and assume A_v is negative. Ideally, the amplifier has these characteristics:

1. The input impedance is infinite: $Z_i = \infty$. This means that currents i^- and i^+ at the inputs are always zero.
2. The output impedance is zero: $Z_o = 0$.
3. The voltage gain is infinite: $A_v = \infty$.²

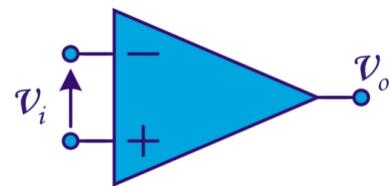


Figure 7.25

An OPAMP is never used in isolation, because with such a high gain, the output will almost certainly be at the supply voltage. Most often an OPAMP is used with other elements that apply feedback from output to the input. The concept of feedback will be studied in more detail in chapter 10.

²In practice, the gain will really be very high: ~ 100000 .

Consider the topology in figure 7.26. The output of the OPAMP is fed back to the negative input terminal through resistance R_2 - this is called *negative feedback*. The input v_i is connected through resistance R_1 to the negative terminal of the OPAMP.

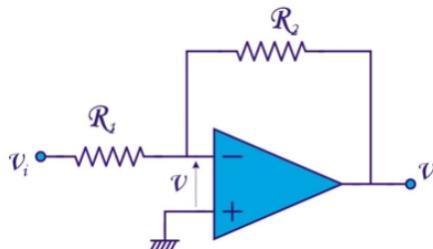


Figure 7.26

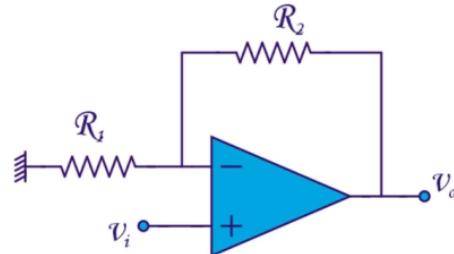


Figure 7.27

Voltage v is found by considering that all current through R_1 also goes through R_2 because the input impedance of the OPAMP is infinite:

$$\begin{aligned} \frac{v_i - v}{R_1} &= \frac{v - v_o}{R_2} \\ \Rightarrow (R_1 + R_2)v &= R_2v_i + R_1v_o \\ \Rightarrow v &= \frac{R_2v_i + R_1v_o}{R_1 + R_2} \end{aligned} \quad (7.18)$$

Output v_o is equal to $A_v v$, so:

$$\begin{aligned} v_o &= -A_v v \\ &= -A_v \frac{R_2v_i + R_1v_o}{R_1 + R_2} \\ \Rightarrow (R_1 + R_2)v_o &= -A_v R_1 v_o - A_v R_2 v_i \\ \Rightarrow v_o &= \frac{-A_v R_2}{R_1 + R_2 + A_v R_1} v_i \\ &\approx -\frac{R_2}{R_1} v_i \text{ if } A_v \rightarrow \infty \end{aligned} \quad (7.19)$$

Thus if A_v is very high, the gain of this topology does not depend on the OPAMP gain and is set by the ratio of resistors R_1 and R_2 : $A = -\frac{R_2}{R_1}$. Because the gain is negative, this topology is called the *inverting amplifier*. The ratio $\frac{R_1}{R_2}$ can be set very precisely, and any temperature dependence disappears because both resistors vary in the same way with temperature. The gain $-\frac{R_2}{R_1}$ is the nominal gain A_n . Expression 7.19 can be rewritten as:

$$\begin{aligned} \frac{v_o}{v_i} &= -\frac{A_n}{1 + \frac{A_n}{A_v} + \frac{1}{A_v}} = \frac{A_v A_n}{A_n + A_v + 1} \\ &\approx -\frac{A_n}{1 + \frac{A_n}{A_v}} \end{aligned} \quad (7.20)$$

This expression is valid for all feedback topologies.

We can also compute $\frac{v}{v_i}$:

$$\begin{aligned}\frac{v}{v_i} &= \frac{v}{v_o} \frac{v_o}{v_i} \\ &= \frac{1}{A_v} \frac{-A_v R_2}{R_1 + R_2 + A_v R_1} \\ &\approx 0 \text{ if } A_v \rightarrow \infty\end{aligned}\tag{7.21}$$

If $A_v \rightarrow \infty$, the voltage $v \rightarrow 0$. This does not mean we can short-circuit both input terminals, but this realization often simplifies analysis. This is why the negative input is called a *virtual ground*.

For example, consider the circuit in 7.27, which is the same circuit as in 7.26, but with v_i applied at the positive terminal. If $v \rightarrow 0$, the voltage at the negative input node is also v_i and the current through R_1 is then equal to $\frac{v_i}{R_1}$. The same current flows through R_2 :

$$v_o - v_i = R_2 \frac{v_i}{R_1}$$

and consequently:

$$v_o = \left(1 + \frac{R_2}{R_1}\right) v_i$$

In figure 7.26, we had an amplifier with a negative gain; here, we obtain an amplifier with a positive gain (the *non-inverting* amplifier).

7.4.2 The Real Amplifier

A real OPAMP, like the 741 OPAMP in figure 7.29 is far from ideal. The most important non-idealities are:

- The input impedance is not infinite, hence $i^-, i^+ \neq 0$. For example, the base currents when bipolar transistors are the inputs of the differential amplifier as first stage.
- There is an offset current $i_d = i^+ - i^-$, when $v_o = 0$. (typically $\sim 20nA$).
- Similarly, there is an offset voltage e_d that's required to make $v_o = 0$. (typically $\sim mV$) as in figure 7.28. This also means that when $v_d \approx 0$, then $v_o \approx E_{supply}$.
- The voltage gain is not infinite: $A_v \neq \infty$. (typically around 100 dB or 100000)
- The bandwidth ω_0 is limited.
- The slew rate and settling time. An OPAMP behaves as a two-pole system, so the output can not immediately follow the input. If we apply a step at the input, the output will increase but not immediately and will take some time to stabilize around the final value. These two effects are characterized by the slew rate and settling time, respectively.

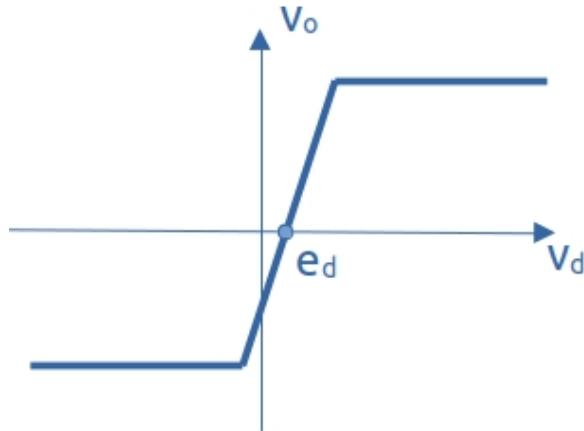


Figure 7.28

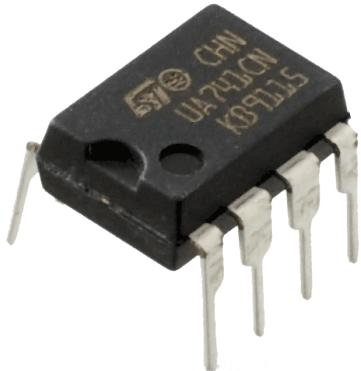


Figure 7.29

Furthermore, a real OPAMP requires a supply voltage ($+E, -E$) and has a maximum allowed power consumption. The supply voltage can be symmetrical ($\pm 1.5V, \pm 5V, \pm 15V$) or with the ground as reference ($+3.3V, +5V, +15V$).

Figure 7.30 shows the typical pin configuration of a LM741 OPAMP, originally developed by Texas Instruments. Note the in -and output voltage v_p, v_n and v_o , the supply voltages $+E$ and $-E$ and two offset pins to adjust offset currents and voltage.

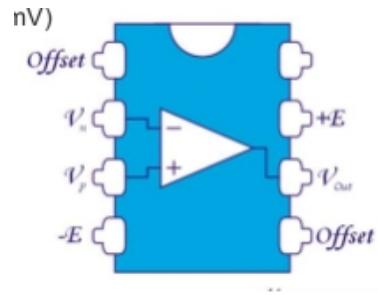


Figure 7.30

7.4.3 OPAMP Theory

To model these non-ideal effects, we will use the OPAMP model in figure 7.31, including a non-zero output impedance Z_o . To demonstrate this, we study the effect of input bias currents and offset voltages when the OPAMP is used in the simple circuit of figure 7.26. Replacing the OPAMP by its model (with $Z_o = 0$) gives the equivalent circuit in figure 7.32. The idea to solve this problem is to consider all sources individually (with all other sources = 0), and adding the results at the end. This is an application of the superposition principle. When we suppose $A_v \rightarrow \infty$:

- $v_i : v_0 = -\frac{R_2}{R_1} v_i$,
- $e_d : v_o = (1 + \frac{R_2}{R_1}) e_d$, because this situation is as in figure 7.27 with voltage e_d at the

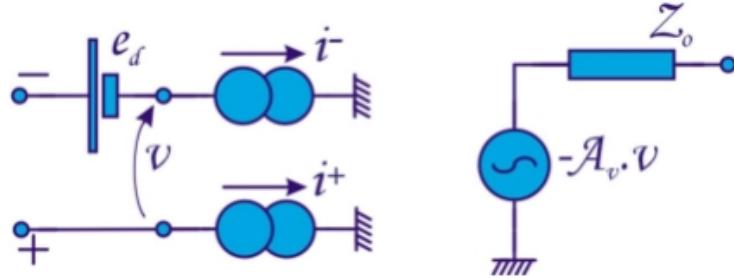


Figure 7.31

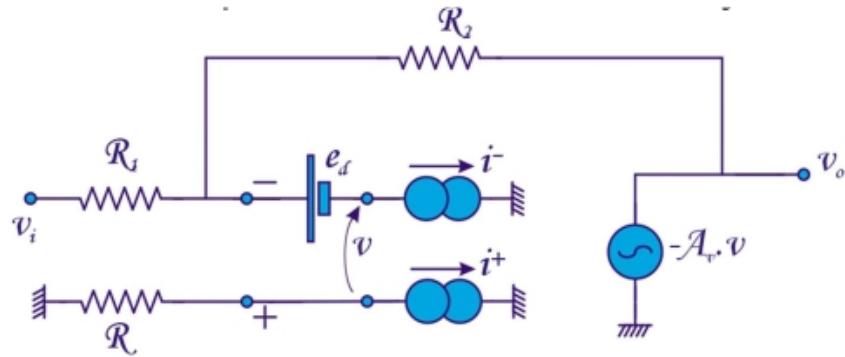


Figure 7.32

negative input instead of v_i .

- $i^- : v_o = R_2 i^-$ because with Millman:

$$v = \frac{G_2 v_o - i^-}{G_1 + G_2}$$

and since $v_o = -A_v v$, we rearrange and obtain:

$$v_o = \frac{A_v}{G_1 + G_2 + A_v G_2} i^- \approx R_2 i^-$$

Another way to see this, is to realize there is no current through R_1 : because $v = 0$, both terminals of R_1 are at zero voltage. All current drawn by i^- goes through R_2 , thus $v_o - v = R_2 i^-$.

- $i^+ : v_o = -R\left(1 + \frac{R_2}{R_1}\right) i^+$ because a current i^+ generates a voltage $-R i^+$ at the positive terminal, replicating the situation of figure 7.26;

The final expression is thus:

$$v_o = -\frac{R_2}{R_1} v_i + \left(1 + \frac{R_2}{R_1}\right) e_d + R_2 i^- - R\left(1 + \frac{R_2}{R_1}\right) i^+$$

We can draw these conclusions:

- If e_d is of the same order of magnitude as v_i , it will have a large impact on the output. The best way to take e_d into account, is to first measure the output voltage without applying v_i . In this way, you measure $\left(1 + \frac{R_2}{R_1}\right) e_d$ and you subtract this value when you measure v_o with v_i at the input.
- The impact of i^- can be reduced by choosing a small R_2 . This is feasible because A_n is the ratio of resistors; their actual value is of less importance.
- If you choose $R = R_1 || R_2$, then $R_2 i^- - R \left(1 + \frac{R_2}{R_1}\right) i^+ = R_2(i^- - i^+)$ and you significantly reduce the impact of the input bias currents.

Since the gain is not infinite, a relative gain error ϵ will always be present. If we call the true gain of the OPAMP with feedback circuitry A , we know from equation 7.20 that:

$$A = \frac{v_o}{v_i} \approx -\frac{A_n}{1 + \frac{A_n}{A_v}} \approx -A_n \quad (7.22)$$

The relative gain error ϵ is thus:

$$\begin{aligned} \epsilon &= \frac{A_n - A}{A_n} = \frac{A_n - \frac{A_n}{1 + \frac{A_n}{A_v}}}{A_n} \\ &= 1 - \frac{1}{1 + \frac{A_n}{A_v}} = 1 - \frac{A_v}{A_v + A_n} \\ &= \frac{A_n}{A_v + A_n} \approx \frac{A_n}{A_v} \end{aligned} \quad (7.23)$$

The gain error thus increases with increasing A_n and with decreasing OPAMP gain (which happens at higher frequencies).

The effect of feedback on the bandwidth

In reality, the OPAMP has a limited bandwidth. We will model the OPAMP as a first-order system with a pole in $\omega = \omega_0$:

$$A_v = \frac{A_{v0}}{1 + j \frac{\omega}{\omega_0}}$$

This system has a cut-off frequency $f_0 = \omega_0/2\pi$.

Substituting this expression in equation 7.22, gives:

$$\begin{aligned} A &\approx -\frac{A_n}{1 + \frac{A_n}{A_v}} && \text{equation 7.22} \\ &= -\frac{A_n}{1 + \frac{A_n}{A_{v0}} \left(1 + j \frac{\omega}{\omega_0}\right)} && \text{substitution} \\ &= -\frac{A_n}{\left(1 + \frac{A_n}{A_{v0}}\right) + j \frac{\omega}{\omega_0} \left(\frac{A_n}{A_{v0}}\right)} && \text{rearrange real - imaginary parts} \\ &\approx -\frac{A_n}{\left(1 + \frac{A_n}{A_{v0}}\right)} \frac{1}{1 + j \frac{\omega}{\omega_0} \left(\frac{A_n}{A_{v0}}\right)} && \text{because } \left(\frac{A_n}{A_{v0}}\right)^2 \text{ is very small} \\ &\approx -\frac{A_n}{1 + j \frac{\omega}{\omega_n}} \end{aligned}$$

where ω_n is a new cut-off frequency:

$$\omega_n = \frac{A_{v0} \omega_0}{A_n}$$

This means that by using the OPAMP with the feedback circuitry, the gain has decreased and becomes A_n , but the bandwidth (the cut-off frequency) has increased. What remains constant is the *gain-bandwidth product GBW*:

$$A_n \times \omega_n = A_{v0} \times \omega_0 = GBW$$

Figure 7.33 represents the Bode curve of the OPAMP self (in blue), with a cut-off at ω_0 and DC-gain A_{v0} , and the OPAMP with feedback circuitry (in green). The latter curve has a lower DC-gain A_n , but a larger bandwidth ω_n . Another issue is also highlighted in the graph: if the frequency increases, the gain decreases and the relative gain error increases. If a maximum ϵ is imposed, the signal frequency cannot go beyond ω_{max} .

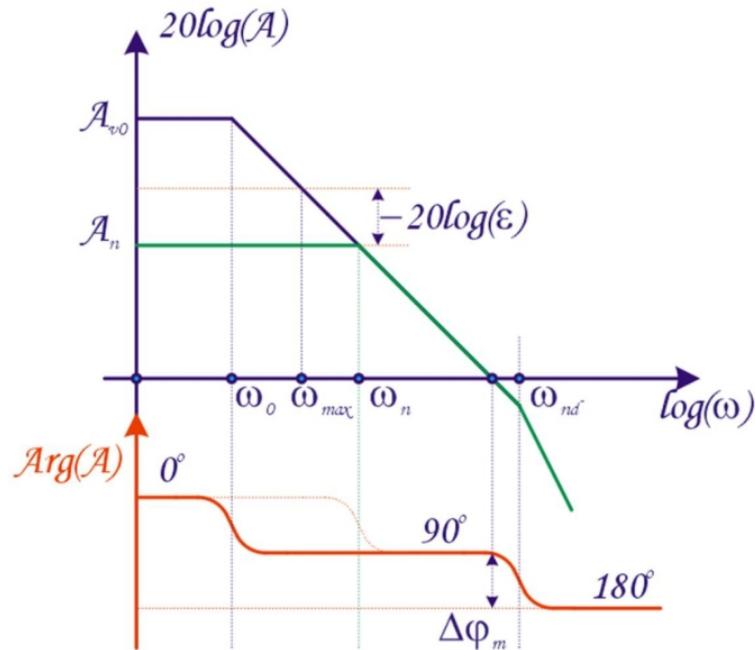


Figure 7.33

Summary

When an OPAMP is used for data acquisition (which is often the case), lots of errors will be introduced:

- due to an offset voltage e_d ,
- due to a differential input current i_d ,
- due to the limited gain $A_v < \infty$, which results in a relative gain error ϵ ,

- due to the limited bandwidth ω_0

You'll have to take all these errors into account, or avoid them by e.g. using OPAMPS with a low offset. In any case, because an OPAMP is inherently a second-order system, we have to live with a delay during acquisition, due to a non-zero settling time.

7.4.4 Unity Gain Buffer

A commonly used OPAMP configuration is the one in figure 7.34, where the output node is directly connected to the negative input terminal. With v equal to zero, we immediately see that $v_o = v_i$, or that $A_n = 1$. We use the equivalent model in figure 7.35 to study the different parameters more formally.

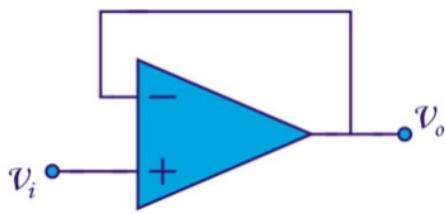


Figure 7.34

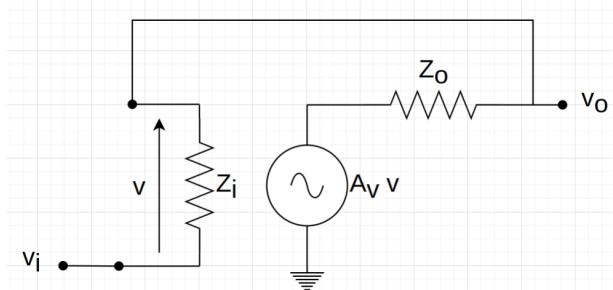


Figure 7.35

1. Voltage gain A_v :

$$\begin{aligned} v_o &= \frac{G_o(-A_v v) + G_i v_i}{G_o + G_i} \\ &= \frac{G_o A_v (v_i - v_o) + G_i v_i}{G_o + G_i} \\ \rightarrow \frac{v_o}{v_i} &= \frac{G_i + A_v G_o}{G_o + G_i + A_v G_o} \\ &\approx \frac{A_v G_o}{A_v G_o} = 1 \end{aligned}$$

2. Input impedance Z_i :

$$\begin{aligned} i_i &= G_i(v_i - v_o) \\ &= G_i\left(1 - \frac{G_i + A_v G_o}{G_o + G_i + A_v G_o}\right)v_i \\ &= G_i \frac{G_o}{G_o + G_i + A_v G_o} v_i \\ &\approx \frac{G_i}{A_v} v_i \\ \rightarrow Z_i &= \frac{v_i}{i_i} = A_v Z_i \end{aligned}$$

3. Output impedance Z_o :

$$\begin{aligned} i_o &= G_i v_o + G_o(v_o - (-A_v v)) \\ &= (G_i + (1 + A_v)G_o)v_o \\ &\approx A_v G_o v_o \\ \rightarrow Z_o &= \frac{v_o}{i_o} = \frac{Z_o}{A_v} \end{aligned}$$

So this circuit has unity gain, a very high input impedance $A_v Z_i$ and a very low output impedance $\frac{Z_o}{A_v}$. In summary, it is the ideal buffer to isolate one stage from the next. Appendix ?? contains more OPAMP examples.

Chapter 8

Transistors at High Frequency

In the BJT small-signal model of figure 6.31, there are two capacitors, C_π and C_μ . These capacitors model the pn-junctions between base and emitter and base and collector. Because the emitter-base junction is forward biased and the base-collector junction is reversed biased, the depletion zone in the former junction is a lot larger than in the latter, and thus $C_\pi \gg C_\mu$. We first analyze the impact of C_π on the behavior of the BJT transistor at high frequencies. Next, we study the impact of C_μ . To do this, we replace C_μ by a Miller capacitor C_M to simplify the analysis.

8.1 Giacoletto Model at High Frequencies

In this section, we assume that $C_\pi \gg C_\mu$ and that $C_\mu \approx 0$. The small-signal model under these assumptions is shown in figure 8.1. With an input current i_b , the base-emitter voltage is:

$$v_{be} = \frac{r_\pi}{1 + j\omega r_\pi C_\pi} i_b$$

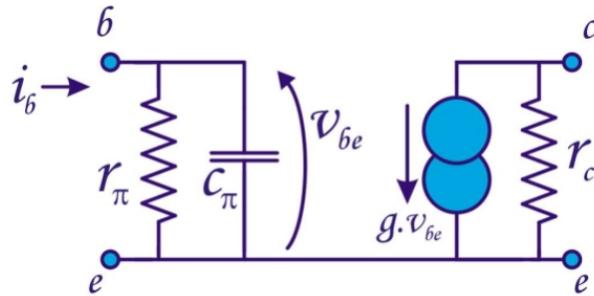


Figure 8.1

with $Z_\pi = \frac{r_\pi}{1 + j\omega r_\pi C_\pi}$ the parallel combination of r_π and C_π . This impedance is equal to r_π for low frequencies, but begins to decrease at a frequency f_β :

$$2\pi f_\beta = \omega_\beta = \frac{1}{r_\pi C_\pi}$$

Beyond this frequency the base starts to degrade and the transistor capacity to amplify the input current decreases.

However, even beyond f_β the transistor can still be used. He stops working when there is no current amplification, i.e. when $\frac{i_c}{i_b} = 1$, where i_c is the current between collector and emitter when we short-circuit them (in the AC equivalent circuit, obviously). This short-circuit current gain $A_{i,sc}$ is the current generated by the dependent current source:

$$A_{i,sc} = \frac{i_c}{i_b} = g v_{be} = g \frac{r_\pi}{1 + j\omega r_\pi C_\pi}$$

We compute the pulsation ω_T when $|A_{i,sc}| = 1$:

$$\begin{aligned} \left| g \frac{r_\pi}{1 + j\omega r_\pi C_\pi} \right| &= 1 \\ &= \frac{gr_\pi}{\sqrt{1 + \frac{\omega^2}{\omega_\beta^2}}} = \frac{\beta}{\sqrt{1 + \frac{\omega^2}{\omega_\beta^2}}} \end{aligned}$$

because $gr_\pi = \frac{I_{CQ}}{v_{th}} \frac{v_{th}}{I_{BQ}} = \frac{I_{CQ}}{I_{BQ}} = \beta$. Consequently:

$$\begin{aligned} \omega_\beta^2 \beta^2 &= \omega_\beta^2 + \omega^2 \\ \Rightarrow \omega &= \omega_\beta \sqrt{\beta^2 - 1} \approx \omega_\beta \beta \end{aligned}$$

This pulsation $\omega_T = \beta \omega_\beta$ is the pulsation beyond which the BJT no longer amplifies the current and thus becomes useless.

8.2 The Miller Capacitor

We study the common-emitter amplifier from figure 7.8 when the input signal v_i has high-frequency components. We assume that both C_B and C_E are short circuits, i.e. we are in the correct working domain (the right part of the Bode curve in figure 7.7). With both C_π and C_μ present, and an output impedance R_S of the signal source, we obtain the AC circuit from figure 8.2.

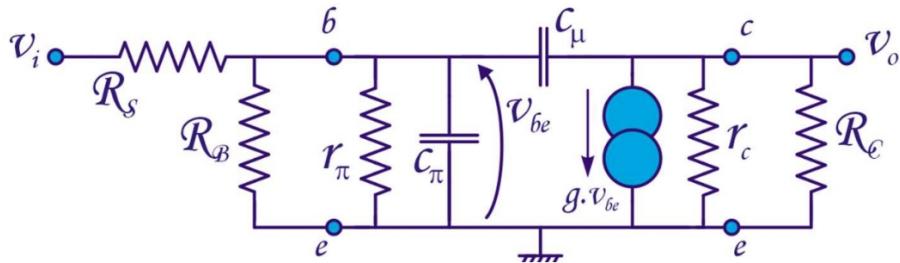


Figure 8.2

The difficulty in this circuit is the presence of C_μ : this capacitor couples the base to the collector, making the analysis hard. To proceed, we want to remove C_μ and replace it by something else, like an element in parallel with C_π . This is where the Miller capacitor comes in.

The key insight is that we can replace the series capacitor C in figure 8.3, which induces a

current $I = Z_C(v_2 - v_1)$, by the two loops in figure 8.4. From an electrical point of view, these circuits are identical because the currents and voltages at the in- and output terminals are identical.

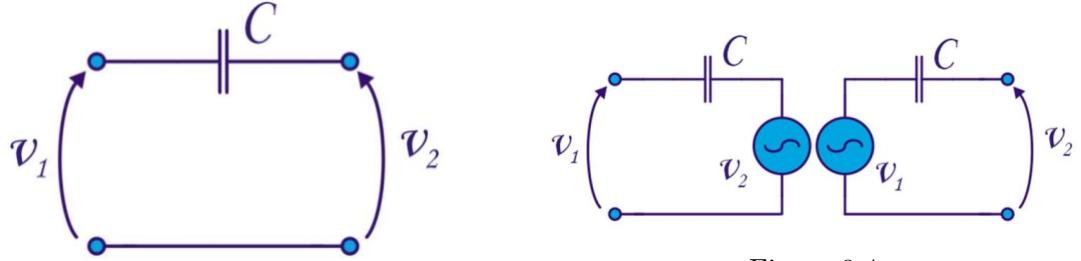


Figure 8.4

Figure 8.3

In the right loop, we replace the voltage source with its Norton equivalent $I_N = j\omega C v_1$, as in figure 8.5.

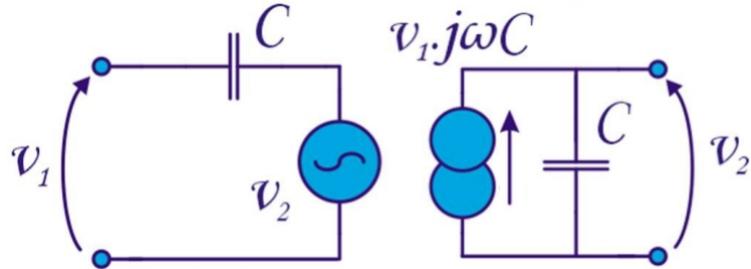


Figure 8.5

If we replace C_μ in this way in the AC-circuit of figure 8.2, we obtain the circuit in figure 8.6. Note that Z_π is $r_\pi || C_\pi || R_B$ and $R_{eq} = r_c || R_C$. In this circuit, we will now:

- Neglect current source $v_{be}j\omega C_\mu$ when it is dominated by $g v_{be}$. This is valid when:

$$\begin{aligned} v_{be}j\omega C_\mu &\ll g v_{be} \\ |j\omega C_\mu| &\ll |g| \\ \rightarrow \omega &\ll \omega_1 = \frac{g}{C_\mu} \end{aligned}$$

This is the first criterion: $\omega \ll \frac{g}{C_\mu}$

- Neglect C_μ in the output loop because its impedance is lot larger then R_{eq} . This is valid if:

$$\begin{aligned} \frac{1}{j\omega C_\mu} &\gg R_{eq} = r_c || R_C \\ \rightarrow \omega &\ll \omega_2 = \frac{1}{R_{eq}C_\mu} \end{aligned}$$

This is the second criterion: $\omega \ll \frac{1}{R_{eq}C_\mu}$

Note that in a typical scenario, e.g. with $I_{CQ} = 1 \text{ mA}$ and $R_C = 1 \text{ k}\Omega$, $g \gg G_C$ and $g \gg g_c$ (this is always valid) and thus mostly $g \gg G_c + g_c = \frac{1}{R_{eq}}$. So if the second criterion is valid,

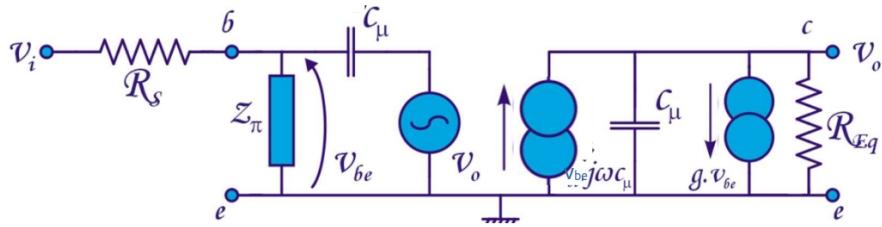


Figure 8.6

the first will valid as well. In general, the Miller conditions are satisfied.

After neglecting $j\omega C_\mu v_{be}$ and C_μ , the small-signal circuit becomes the one in figure 8.7. In this circuit, we know that $v_o = -g R_{eq} v_{be}$. Hence, the current through C_μ in the left loop is equal to:

$$\begin{aligned} i_{C_\mu} &= j\omega C_\mu (v_{be} - v_o) \\ &= j\omega C_\mu (1 + g R_{eq}) v_{be} \end{aligned}$$

So we can replace C_μ and the source that provides v_o by a single capacitor with capacitance $C_\mu(1 + g R_{eq})$. This is the Miller capacitor:

$$C_M = C_\mu(1 + g R_{eq}) \quad (8.1)$$

as in figure 8.8.

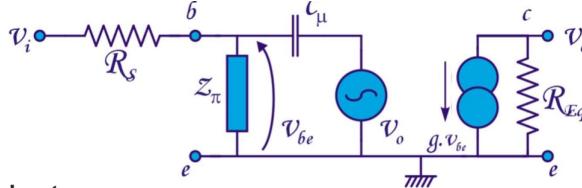


Figure 8.7

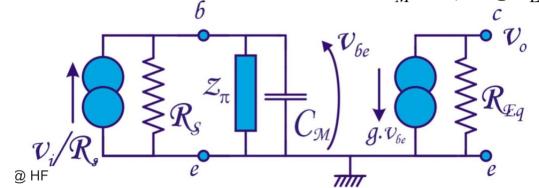


Figure 8.8

We can in the left loop of figure 8.8 group all impedance in a single $Z_{eq} = \frac{R}{1+j\omega R(C_\pi + C_M)}$ with $R = r_\pi || R_B || R_S$ the parallel combination of all resistors in the input part of the circuit, and $C_\pi + C_M$ the parallel combination of C_π and C_M . The base-emitter voltage is then equal to:

$$v_{be} = \frac{1}{R_S} \frac{R}{1 + j\omega R(C_\pi + C_M)} v_i$$

and the voltage gain is:

$$A_v = \frac{v_o}{v_i} = -g R_{eq} \frac{v_{be}}{v_i} = -g R_{eq} \frac{1}{R_S} \frac{R}{1 + j\omega R(C_\pi + C_M)}$$

When R_S is small:

$$A_v \approx -g \frac{R_{eq}}{1 + j\omega R(C_\pi + C_M)} \quad (8.2)$$

8.3 Miller's Theorem

Equation 8.1 is an instance of Miller's theorem, which aims to replace a floating impedance with two grounded impedances as we did in figure 8.3. For the derivation, refer to figure 8.9.

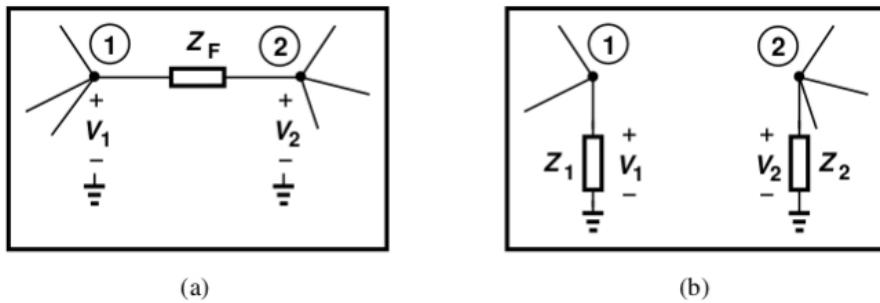


Figure 8.9

We wish to transform Z_F to two grounded impedances as depicted in figure 8.9(b), while ensuring all of the currents and voltages in the circuit remain unchanged. This means:

$$\frac{V_2 - V_1}{Z_F} = \frac{V_1}{Z_1}$$

$$\frac{V_1 - V_2}{Z_F} = -\frac{V_2}{Z_2}$$

If we denote the voltage gain from node 1 to node 2 as $A_v = \frac{V_2}{V_1}$, we find:

$$Z_1 = \frac{V_1}{V_2 - V_1} Z_F$$

$$= \frac{1}{1 - A_v} Z_F$$

for Z_1 and

$$Z_2 = \frac{V_2}{V_1 - V_2} Z_F$$

$$= \frac{1}{1 - A_v^{-1}} Z_F$$

for Z_2 .

8.4 Conclusion

At high frequencies, we must take the parasitic capacitances of the transistor into account. For floating capacitances like C_μ , we use Miller's theorem and replace it by C_M . Note that we can only do this for the if the Miller conditions are satisfied:

1. $\omega \ll \frac{g}{C_\mu}$
 2. $\omega \ll \frac{1}{R_{eq} C}$

From equation 8.2, we see that there is a cut-off frequency for the gain:

$$\omega_H = \frac{1}{c_{\pi} + C_M}$$

Above this frequency, the gain A_v of the common-emitter amplifier begins to decrease, and we can extend figure 7.7 to include this cut-off frequency, as in figure 8.10. We also know that $\omega_H < \frac{1}{r_\pi C_\pi} = \omega_\beta$. This means that performance degradation is initially due to a direct signal path from base to collector through C_μ , and that base degradation happens later (i.e. at higher frequencies).

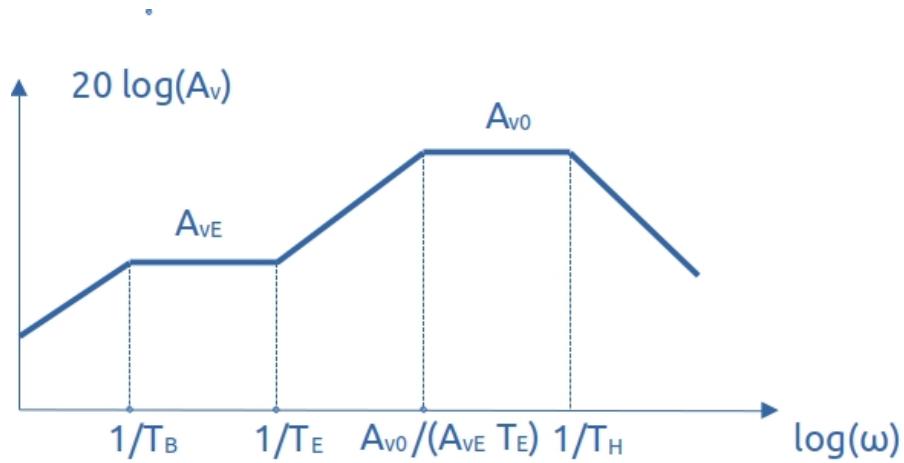


Figure 8.10

Chapter 9

Power Amplifiers

In the previous chapter, we studied all kinds of amplifiers. However, we never looked at power amplification and how power was provided to the load resistance, but we only were interested in amplification of the input voltage. In this chapter, we're mostly interested in power consumption and efficiency, and we'll look at amplifiers of class A, B, C and S. We'll also study current amplifiers like push-pull amplifiers, and selective amplifiers that only amplify signals around a central frequency.

9.1 Introduction

Until now, we were not concerned with power consumption or efficiency of the amplifiers we have studied. We only cared about achieving high gain. In this chapter, we look at *power amplifiers*, amplifiers made to deliver power - and do this as efficiently as possible.

First, we will consider a simple common emitter amplifier in figure 9.1. Notice that we don't consider an emitter resistance R_E . This is because typically, R_E is quite low; if not, the possible voltage swing along the dynamic load line would become too small.

With proper biasing and neglecting $V_{CE,Sat}$, we see that $V_{CEQ} = \frac{E}{2}$ and $I_{CQ} = \frac{E}{2R_C}$ so that Q is nicely in the middle of the operating domain, as in figure 9.2. The load line is given by $E = R_C i_c + v_{CE}$. Furthermore, we assume that the collector resistor R_C is the load to which we want to deliver power.

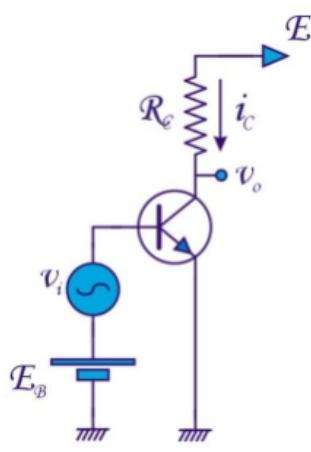


Figure 9.1

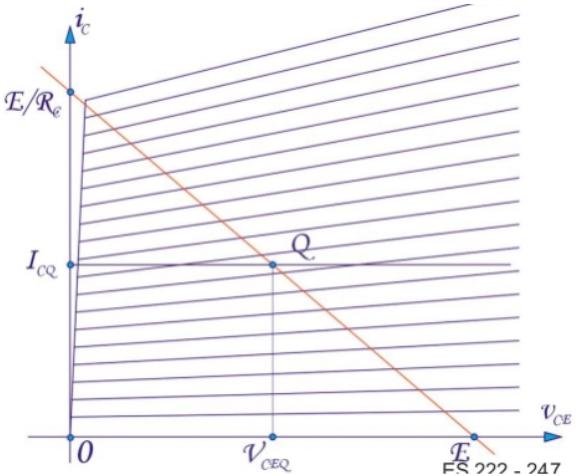


Figure 9.2

We assume that the applied input signal v_i is sinusoidal. If E_B is large enough, the collector current i_C will also be a sinusoid and the transistor is always conducting, as in the top of figure 9.3. But if we decrease the input bias source E_B , I_{CQ} can be reduced, and from a certain point on (namely when $E_B + v_i = V_{BEQ}$), i_C will become equal to zero (figure 9.4 bottom), and can't become negative because the circuit can't conduct in the other direction. If E_B decreases even further, the transistor will conduct less than half of the time (figure 9.4). The *conduction angle* θ is defined as:

$$\theta = \frac{T_{conduct}}{T} \times 180^\circ \quad (9.1)$$

Based on the angle of conduction, 4 different classes of amplifiers are defined:

1. Class A: $\theta = 180^\circ$
2. Class AB: $90^\circ < \theta < 180^\circ$
3. Class B: $\theta = 90^\circ$
4. Class C: $\theta < 90^\circ$

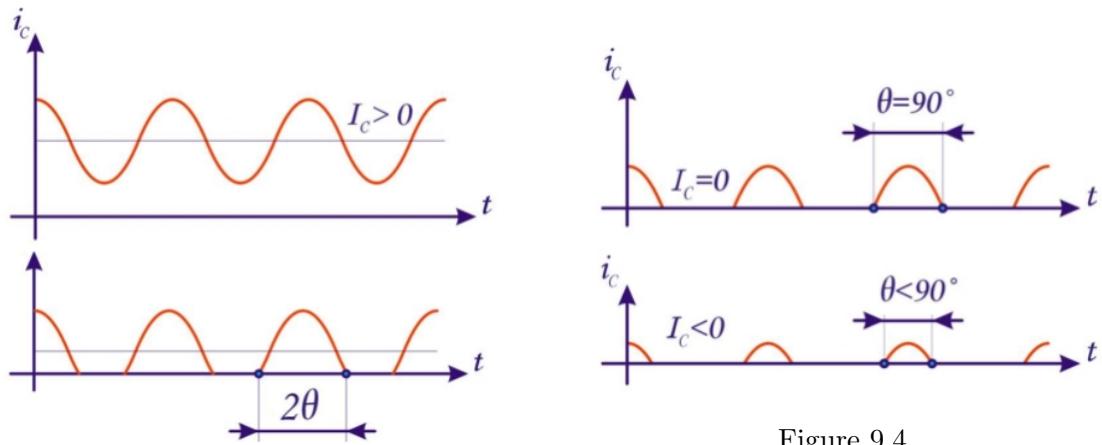


Figure 9.3

In the rest of this chapter, we study a circuit for each type of conduction. We will compute the power efficiencies, by considering only the circuit on the collector side. All other power consumptions, like base currents or power consumed in the emitter (source) resistance will be neglected.

The instantaneous power consumed by an element in a circuit is $p(t) = v(t)i(t)$ with v and i the instantaneous voltage and current through the element. When the signals are sinusoidal, we define the average power dissipated during one period T :

$$P = \frac{1}{T} \int_{t_0}^{t_0+T} v(t) i(t) dt$$

9.2 Class A Amplifier

We assume that the applied input signal v_i is sinusoidal and thus:

$$\begin{aligned} v_{CE} &= V_{CEQ} + V_{cem} \sin(\omega t) \\ i_C &= I_{CQ} - I_{cm} \sin(\omega t) \end{aligned}$$

with $V_{cem} = R_C I_{cm}$.

Because Q is located in the middle of the operating region, we also have $V_{CEQ} = E/2$ and $I_{CQ} = \frac{E}{2R_C}$. Consequently, the voltage and current amplitudes are limited to: $V_{cem} \leq E/2$ and $I_{cm} \leq \frac{E}{2R_C}$.

We calculate¹:

- P_D , the power delivered to the circuit by the supply E :

$$P_D = \frac{1}{T} \int_{t_0}^{t_0+T} E i_C(t) dt = E I_{CQ} = \frac{E^2}{2R_C}$$

- P_C , the power dissipated by the transistor:

$$\begin{aligned} P_C &= \frac{1}{T} \int_{t_0}^{t_0+T} v_{CE}(t) i_C(t) dt \\ &= \frac{1}{T} \int_{t_0}^{t_0+T} (V_{CEQ} + V_{cem} \sin(\omega t)) (I_{CQ} - I_{cm} \sin(\omega t)) dt \\ &= \frac{1}{T} \int_{t_0}^{t_0+T} (V_{CEQ} I_{CQ} - V_{cem} I_{cm} \sin(\omega t)^2) dt \\ &= V_{CEQ} I_{CQ} - \frac{1}{2} V_{cem} I_{cm} = \frac{E^2}{4R_C} - \frac{1}{2} \frac{V_{cem}^2}{R_C} \\ &= \frac{E^2}{4R_C} - \frac{1}{2} \frac{E^2}{4R_C} = \frac{E^2}{8R_C} \text{ at } V_{cem} = E/2 \text{ and } I_{cm} = \frac{E}{2R_C} \end{aligned}$$

- P_L , the AC power delivered to the load R_C because the only useful power delivered to the load is the variation of the signal around the average value:

$$\begin{aligned} P_L &= \frac{1}{T} \int_{t_0}^{t_0+T} v_{ce}(t) i_c(t) dt \\ &= \frac{1}{2} V_{cem} I_{cm} \\ &= \frac{1}{2} \frac{E^2}{4R_C} = \frac{E^2}{8R_C} \text{ at } V_{cem} = E/2 \text{ and } I_{cm} = \frac{E}{2R_C} \end{aligned}$$

- P_J , the rest:

$$P_J = P_D - P_C - P_L = \frac{E}{4R_C}$$

The efficiency of the amplifier η is defined as the ratio between power delivered to the load over the total power delivered by the supply:

$$\eta = \frac{P_L}{P_D} = \frac{V_{cem}^2}{E^2} \quad (9.2)$$

With $V_{cem,max} = \frac{E}{2}$ we find the maximum efficiency: $\eta_{max} = \frac{1}{4}$. The quality factor F is the ratio between the maximum power consumed by the transistor $P_{C,max}$ and the maximum

¹Note that $\int_0^T \sin(\omega t) dt = 0$ and $\int_0^T \sin^2(\omega t) dt = \frac{T}{2}$

power consumed by the load $P_{L,max}$. The former happens when $V_{cem} = 0$ and is equal to: $P_{C,max} = \frac{E^2}{4R_C}$, the latter occurs at full swing $V_{cem} = E/2$ and is equal to $P_{L,max} = \frac{E^2}{8R_C}$. Thus

$$F = \frac{P_{C,max}}{P_{L,max}} = 2$$

These different power contributions are shown in figure 9.5 as a function of the voltage swing

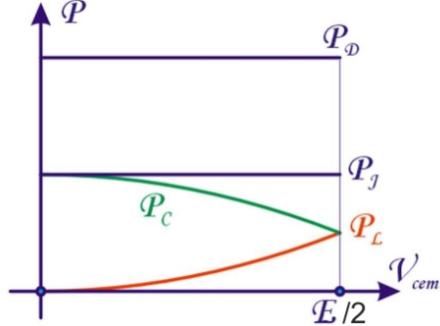


Figure 9.5: Power distribution for increasing V_{cem}

V_{cem} . Note how the maximum efficiency is only 25%, achieved at the maximal value of V_{cem} . Half the power $P_J = \frac{E^2}{8R_C} = R_C I_{CQ}^2$ is used to generate the bias current for the transistor, which also runs through the load resistor R_C but doesn't provide anything useful. This power is useless and should be eliminated to increase the efficiency η .

To improve this, we propose the circuit in figure 9.6. If L is very large, the inductor will be a short-circuit for DC signals. And if C_L is very large, the load resistor R_L is isolated from the bias currents. Thus I_{CQ} doesn't flow through the load, in contrast with the previous situation. Consequently, $V_{CEQ} = E$ and $I_{CQ} = \frac{E}{R_C}$ and the operating point Q is uniquely determined.

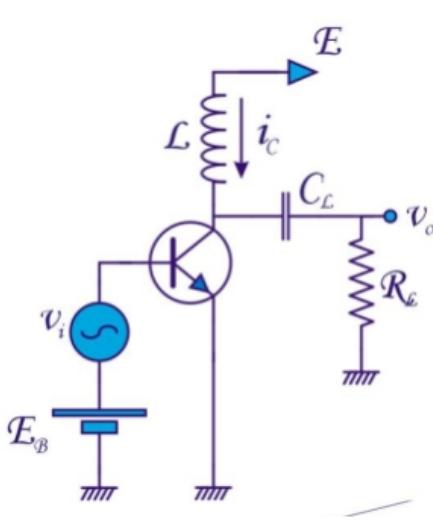


Figure 9.6: Class A Amplifier

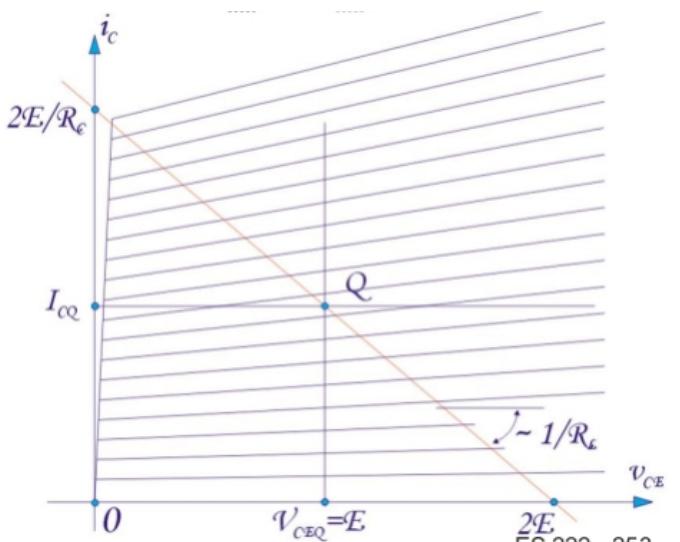


Figure 9.7: Load lines and Q-point

For AC signals however, the inductor is an open circuit and C_L is a short circuit, so we can write the dynamic load line: $v_{ce} = -R_L i_c$. The load lines (static in blue, dynamic in red)

are sketched in figure 9.7. Note how v_{CE} can become higher than the power supply E . This is because the inductance will create an electromotive force (EMF) at the frequency we are working at to keep the current through it constant. This EMF can increase the voltage at the collector to $2E$. This also means that $V_{cem,max} = E$ and $I_{cem,max} = E/R_L$.

We can recompute the different powers:

- $P_D = \frac{1}{T} \int_{t_0}^{t_0+T} E i_C dt = E I_{CQ} = \frac{E^2}{R_L}$
- $P_C = \frac{1}{T} \int_{t_0}^{t_0+T} v_{CE} i_C dt = V_{CEQ} I_{CQ} - \frac{V_{cem}^2}{2R_L} = \frac{E^2}{2R_L}$
- $P_L = \frac{1}{T} \int_{t_0}^{t_0+T} v_{ce}(t) i_c(t) dt = \frac{V_{cem}^2}{2R_L} = \frac{E^2}{2R_L}$

Note how $P_C + P_L = P_D$ and thus $P_J = 0$. This is because there runs no DC (and hence useless) power through the load resistance. The different powers are shown in figure 9.8.

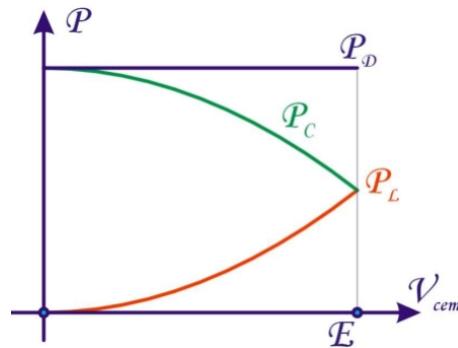


Figure 9.8: Power distribution for class A amplifier

The efficiency $\eta = \frac{P_L}{P_D} = \frac{V_{cem}^2}{2E^2}$ so that $\eta_{max} = \frac{1}{2}$, and the quality factor $F = 2$. The maximum efficiency (reached at maximum amplitude) is thus 50%, instead of 25% as before.

9.2.1 Improvement to the class A amplifier

This issue with the previous circuit is this: typically, the power supply E is given, and when you buy a speaker, the internal resistance R_L is also given. This means that the maximum power you can deliver to the load, namely $\frac{E^2}{2R_L}$ is also fixed, even though the speaker may have a higher $P_{L,max}$.

An alternative circuit to the one in figure 9.6 is the one in figure 9.9 where a transformer is used to transfer the AC power to the load and the conversion factor n can be set. This additional degree of freedom will allow to increase the maximum power.

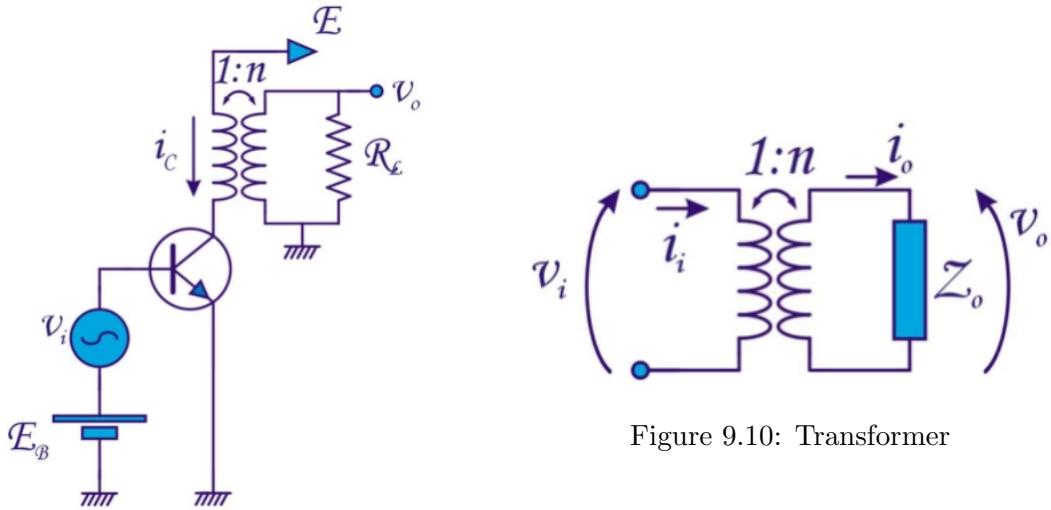


Figure 9.10: Transformer

Figure 9.9: Improved Class A Amplifier

In this circuit, we have the same DC load line as before: $v_{CE} = E$, but the AC load line is now $v_{ce} = -R'_L i_e$ where R'_L is the load as seen by the biasing circuit.

To understand this, consider the transformer in figure 9.10. A transformer works by generating a changing magnetic field $\Phi_B(t)$ with an inductor and this flux flows then through another inductor which generates an EMF due to Faraday's law of induction: $\mathcal{E} = -\frac{d\Phi_B}{dt}$. Because a transformer only works for AC signals, the load seen by the DC circuit is zero - the transformer acts as a short-circuit. For AC signals however, as in figure 9.10, the impedance Z_i seen by the circuit is different. Because the relation between currents and voltages is:

$$\begin{aligned} n I_o &= I_i \\ V_o &= n V_i \end{aligned}$$

we find that $Z_i = \frac{V_i}{I_i} = \frac{V_o}{n^2 I_o} = \frac{Z_o}{n^2}$. This means that the load R_L is reflected into the circuit and the circuit sees R_L/n^2 . For a given E , R_L and $P_{L,max}$, we know for the reflected load (the apparent resistor) that $P_{L,max} = \frac{E^2}{2R'_L}$ and thus:

$$R'_L = \frac{E^2}{2P_{L,max}}$$

and from R'_L and the true R_L , we find the turns ratio n :

$$n = \sqrt{\frac{R_L}{R'_L}}$$

In an audio amplifier, this turn ratio can often be set by turning a screw.

9.3 Class B Amplifier

In the class A amplifier from the previous section, there was always a current I_{CQ} and a current variation i_c around this value because there is a bias source E_B that lifts the base voltage at the input transistor to a level that allows conduction in the right branch. This also

means that the transistor dissipates power even when there is no signal ($V_{cem} = 0$) - it will even dissipate all power delivered to the circuit, which is constant.

We could remove this bias source, as in figure 9.11. However, in that case there will only be a non-zero voltage of R_L during half of the period, namely when $v_i > 0$, as in figure 9.12. Note that the voltage is negative because the circuit has a negative gain.

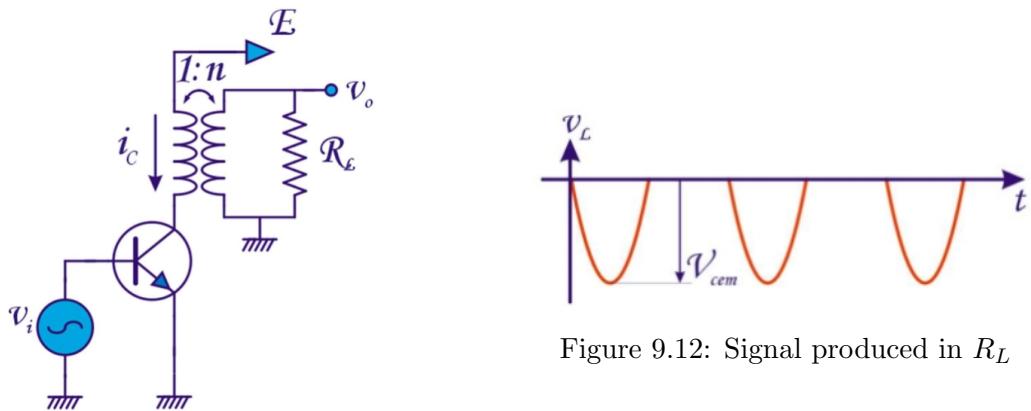


Figure 9.11: Class A Amplifier without E_B

Figure 9.12: Signal produced in R_L

This is off course not an acceptable situation, but we can create a signal during the entire period of the input signal by effectively putting two of these circuits on top of each other, as in figure 9.13. Note that we use both v_i and $-v_i$ and two npn-transistors with interconnected emitters and opposite base-emitter voltages. On the load side, note that the dots signify whether the turns of the inductors turn in the same direction or not.

Transistor ① will conduct only when $v_i > 0$ and generates a signal of the form of figure 9.12 in R_L . Transistor ② conducts only when $v_i < 0$ and generates the opposite signal of figure 9.12. When both signals are added together, the complete wave form is restored. Note that we have neglected that $V_{BEQ} \approx 0.6$ V - we will come back to this later.

The DC load line for transistor ① has the same expression as behavior: $v_{CE} = E$, but $I_{CQ} = 0$ because the transistor is blocked when $v_i = 0$. Thus the operating point lies on the x-axis of figure 9.14 and the dynamic load line with slope $-\frac{1}{R'_L}$ passes through this point. As v_i increases, v_{CE} of transistor ① decreases because we move up the load line, as shown in figure 9.14, until the maximum current E/R'_L is reached. At the same time, transistor ② is blocked but v_{CE} of ② increases from E to $2E$ because the transformer induces an EMF from the right inductor to the inductor in the loop with the emitter-collector of transistor ②. The v_{CE} of ② moves along the horizontal line from Q to $2E$. When v_i becomes negative, the situation is reversed for both transistors.

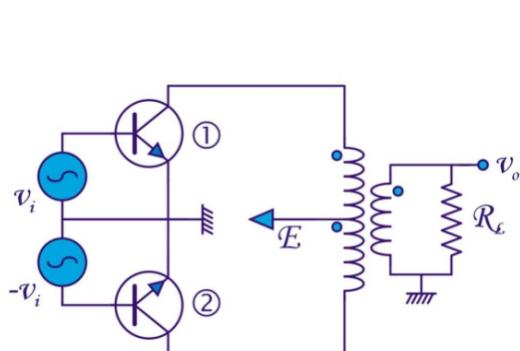


Figure 9.13: Class B Amplifier

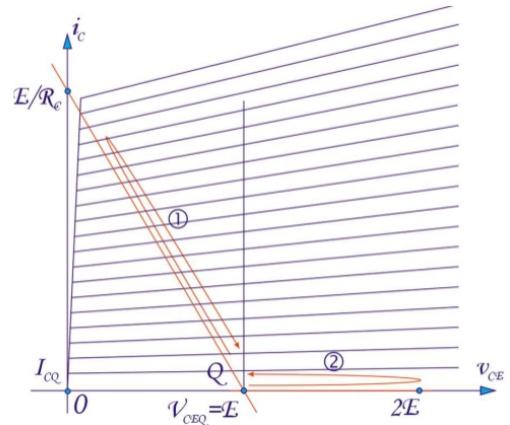


Figure 9.14: Load lines of figure 9.13

Let's compute the different powers, just as before. However, keep in mind that every transistor is only active half of the time, so we integrate over $T/2^2$: (**TODO:** check results)

- $P_D = \frac{2}{T} \int_{t_0}^{t_0+T/2} E I_{cm} \sin(\omega t) dt = \frac{2}{\pi} E I_{cm} = \frac{2}{\pi} \frac{E V_{cem}}{R'_L} = \frac{2}{\pi} \frac{E^2}{R'_L}$
- $P_L = \frac{2}{T} \int_{t_0}^{t_0+T/2} V_{cem} I_{cm} \sin^2(\omega t) dt = \frac{V_{cem}^2}{2R'_L} = \frac{E^2}{2R'_L}$
- And consequently, since there is no power dissipated for biasing:
 $P_C = P_D - P_L = \frac{2}{\pi} \frac{E V_{cem}}{R'_L} - \frac{V_{cem}^2}{2R'_L}$

We see that P_D increases monotonically with the signal amplitude V_{cem} (there was no dependence in the class A amplifier) - see figure 9.15.

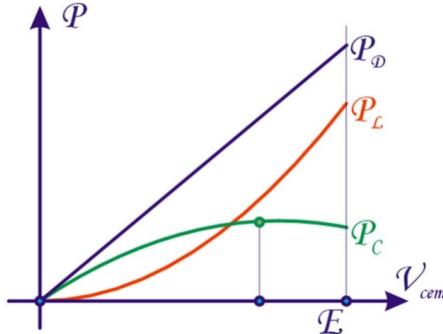


Figure 9.15: Power distribution for class B amplifier

P_C has a quadratic behavior and reaches a maximum when:

$$\frac{dP_C}{dV_{cem}} = \frac{2}{\pi} \frac{E}{R'_L} - \frac{V_{cem}}{R'_L} = 0$$

this is at $V_{cem} = \frac{2}{\pi} E$ and thus

$$P_{C,max} = P_C \Big|_{V_{cem} = \frac{2E}{\pi}} = \frac{2E^2}{\pi^2 R'_L}$$

²Remember: $\int_0^\pi \sin(x) dx = 2$ and $\int_0^\pi \sin^2(x) dx = \pi/2$

The efficiency η is equal to:

$$\eta = \frac{P_L}{P_D} = \frac{\pi}{4} \frac{V_{cem}}{E}$$

and $\eta_{max} = \frac{\pi}{4}$. In other words, the maximum efficiency, reached when $V_{cem} = V_{cem,max} = E$, is about 78%. The quality factor F is:

$$F = \frac{P_{C,max}}{P_{L,max}} = \frac{2E^2}{\pi^2 R'_L} / \frac{E^2}{2R'_L} = \frac{4}{\pi^2}$$

and for each transistor $F_{Tr} = \frac{2}{\pi^2} \approx 0.2$. So for every 5 W delivered to the load, the transistor consumes only 1 W. This is a lot better than before.

How to generate $-v_i$?

TODO: Given v_i , there are basically two ways to generate $-v_i$, as required by the class B amplifier:

1. Use a phase splitter, as in figure ??.
2. Use a transformer, as in figure ??.

What if $V_{BEQ} \neq 0$?

Until now, we have neglected the threshold voltage V_{BEQ} needed to generate a current in the transistor. But in reality, $v_i > 0.6$ V before the transistor conducts. If we take this into account, the resulting current waveform is not a perfect sinusoid, but rather a cascade of half-periods interrupted by regions of zero current, when v_i is too low to generate a forward bias in the base-emitter junctions, as in figure 9.16. This phenomenon is called *cross distortion*.

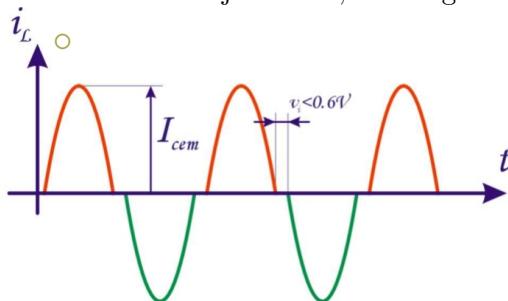


Figure 9.16

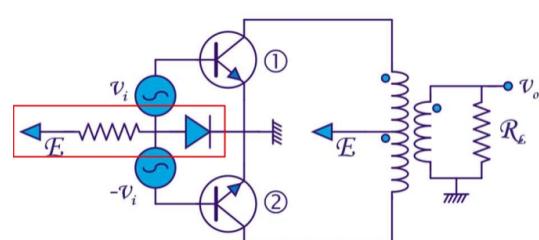


Figure 9.17

A way to solve this is to pre-bias the loop with the base-emitter junction, by adding one threshold voltage via a diode, as in figure 9.17 - the part in red is the pre-bias circuit. This circuit adds one V_{BEQ} to v_i such that a positive v_i is enough to generate a collector current in transistor ① because the voltage at its base will be $v_i + 0.6$ V.

9.4 Push-Pull Amplifiers

A similar idea of using two transistors circuits on top of each other, is the so-called *push-pull* topology of figure 9.18. The goal is not to amplify a voltage, but to deliver current to and from a load R_L .

The circuit consists of an input voltage v_i that will be coupled in through capacitors C_B , two biasing circuits with resistors R_B and R that put the base voltage at $\frac{R_B R}{R_B + R} E$, and two transistors in common-collector configuration (i.e. as emitter followers). Note that transistor ① is an npn, but ② is a pnp transistor. The load R_L is thus connected to the emitters of both transistors.

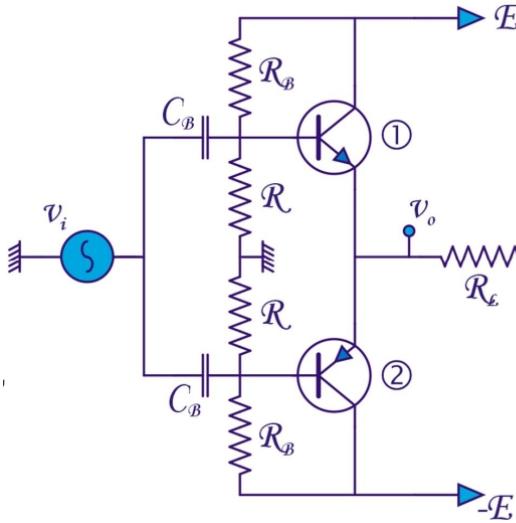


Figure 9.18

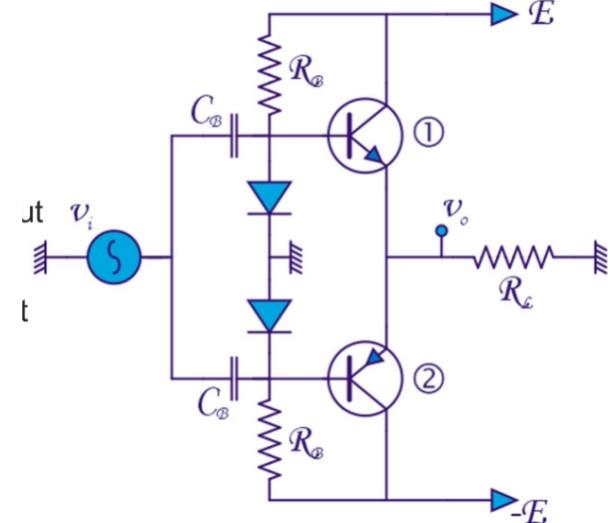


Figure 9.19

If v_i is zero, both transistors carry the same current because of the symmetry, and hence there is no current in the load: $i_{C1} = i_{C2}$ and $i_{RL} = i_{C1} - i_{C2} = 0$. If v_i increases, i_{C1} will increase and i_{C2} will decrease. Consequently, $i_{RL} = i_{C1} - i_{C2} > 0$ and transistor ① will "push" current into the load. On the other hand, if v_i decreases, i_{C2} increases and i_{C1} decreases, and transistor ② will "pull" current out of the load. Because both transistors are emitter followers, $v_o \approx v_i$ and $A_v \approx 1$. To reiterate: the goal is to deliver current, not to increase the gain.

The configuration in figure 9.18 is a class A amplifier because it has a bias circuit with two resistors. If we replace resistors R with diodes as in figure 9.19, we have the same pre-biasing circuitry as the class B amplifier in figure 9.17. In this case, resistor R_B biases the diode to prevent cross distortion.

A way to provide both voltage amplification and current, is by using this circuit as the output trap after an OPAMP as in figure 9.20.

The voltage gain is set by the OPAMP: $A_v = 1 + \frac{R_2}{R_1}$. Note that the feedback via R_2 is applied to the output of the circuit, and not to the output of the OPAMP. Because of its high gain, the OPAMP will try to keep the voltage v between its two terminals equal to zero, so that we can write at the negative terminal:

$$v_i = \frac{0/R_1 + v_o/R_2}{1/R_1 + 1/R_2} = \frac{R_1}{R_1 + R_2} v_o$$

and this equation will set the output voltage. In this way, we avoid voltage loss through the push-pull stage.

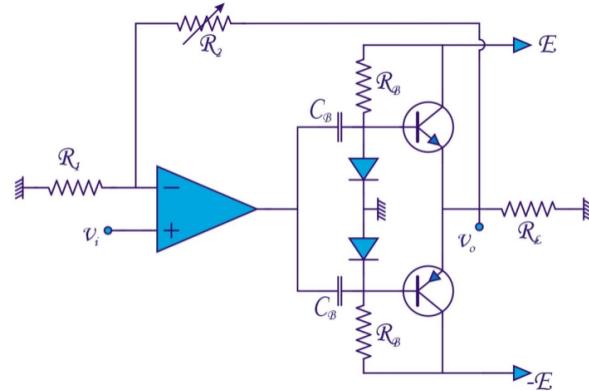


Figure 9.20: Push-Pull amplifier with OPAMP

9.5 Class C Amplifier

For the class C amplifier, we return to the circuit of the class A amplifier in figure 9.6, but we now apply a negative bias (notice the orientation of \$E_B\$). This means that the transistor will conduct for only a fraction \$2\tau\$ of the period \$T\$, as the waveform of \$i_C\$ in figure 9.22 shows.

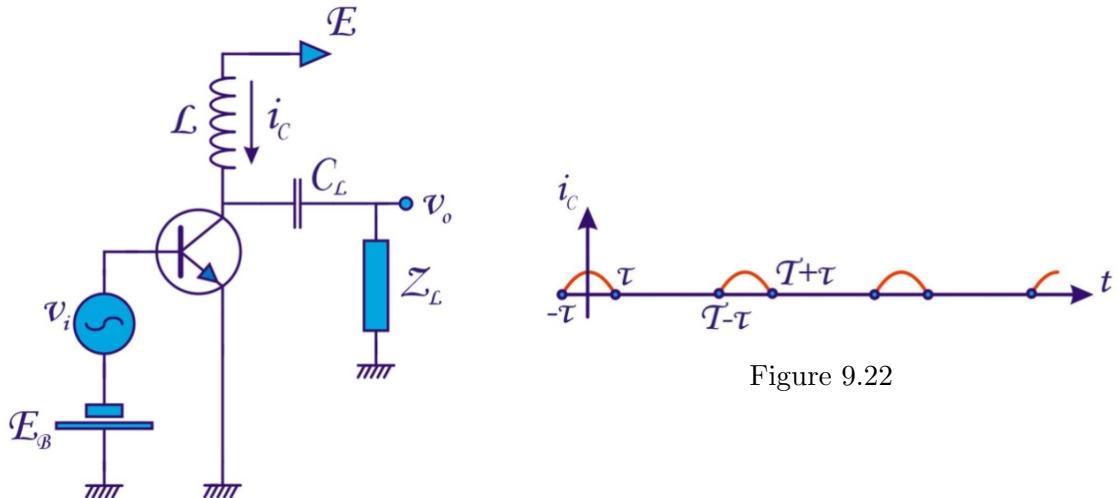


Figure 9.22

Figure 9.21

The current \$i_C\$ in figure 9.22 can be expressed as a cosine from which we subtract the lower \$\cos(\omega_0 t)\$ part:

$$i_c = I_{cm} \frac{\cos(\omega_0 t) - \cos(\omega_0 \tau)}{1 - \cos(\omega_0 t)} \text{ if } kT - \tau < t < kT + \tau$$

and 0 elsewhere. Because this function is periodic, we can write it as a Fourier series i.e. the spectrum of the function:

$$i_c(t) = I_0 + \sum_{n=1}^{\infty} I_n \cos(n \omega_0 t) \text{ with } I_n = \frac{2}{T} \int_{-T/2}^{T/2} i_c(t) \cos(n \omega_0 t) dt$$

This spectrum is shown in figure 9.23. From this spectrum, we only need the first harmonic, centered on \$\omega = \omega_0\$.

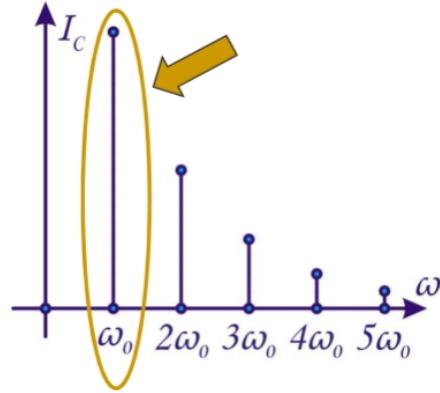


Figure 9.23: Spectrum of figure 9.22

The output voltage v_o is equal to $v_o = Z_L i_c$. To just keep the first harmonic, we need a Z_L that is maximum at frequency $\omega_0/2\pi$, and zero at all other frequencies. This can be achieved by using an RLC tank as in figure 9.24 for Z_L . The impedance is:

$$Z(\omega) = \frac{R}{1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)}$$

with resonance pulsation $\omega_0 = \sqrt{1/LC}$ and quality factor $Q = \omega_0 RC = \omega_0/BW$ as in figure 9.25. So a proper choice of L and C determines ω_0 , and the choice of R and C sets the bandwidth. The bandwidth should be set by considering the bandwidth of the signals of interest, and by taking into account that non-linear distortion creates other harmonics. How the expressions for the RLC circuit are found is explained in section 9.5.1

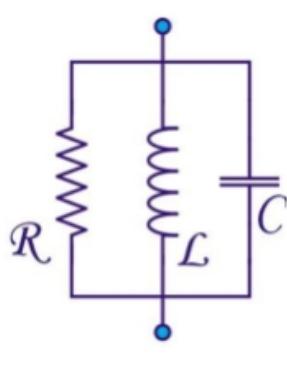


Figure 9.24

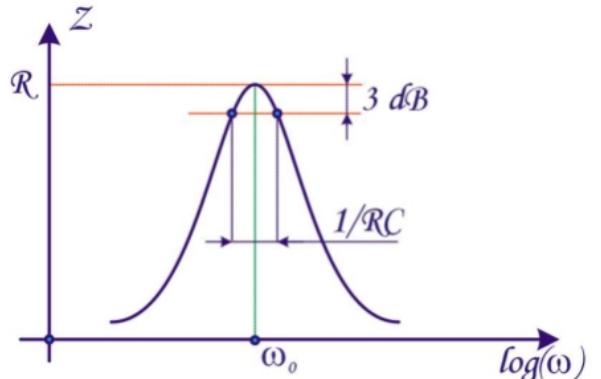


Figure 9.25

The class C amplifier has less dissipation in the transistor and higher power transfer. It is mostly used in telecommunication applications and high power transmission.

9.5.1 The RLC Tank

Consider the parallel combination of a resistance R , a capacitance C and an inductor L , as in figure 9.24. Qualitatively, we see that if $\omega \ll \omega_0$, the inductance will act as a short circuit, and $Z \approx 0$. Similarly, if $\omega \gg \omega_0$, C becomes a short and again $Z \approx 0$. Now let's analyze this

circuit more formally.

Working with the admittances, we can write:

$$\begin{aligned} Y &= \frac{1}{R} + j\omega C + \frac{1}{j\omega L} \\ &= \frac{j\omega L + R - \omega^2 RLC}{j\omega RL} \end{aligned}$$

and thus, for the impedance Z , where we use the resonance frequency $\omega_0^2 = \frac{1}{LC}$:

$$\begin{aligned} Z(\omega) &= \frac{j\omega LR}{j\omega L + R - \omega^2 RLC} = \frac{R}{1 - j\frac{R}{\omega L} + j\omega RC} \\ &= \frac{R}{1 + j(\omega RC - \frac{R}{\omega L})} = \frac{R}{1 + jRC(\omega - \frac{1}{\omega LC})} \\ &= \frac{R}{1 + jRC(\omega - \frac{\omega_0^2}{\omega})} = \frac{R}{1 + j\omega_0 RC\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)} \end{aligned}$$

If $\omega \ll \omega_0$ or if $\omega \gg \omega_0$, $Z \rightarrow 0$. If $\omega = \omega_0$, $Z(\omega_0) = R$. The impedance as function of frequency is plotted in figure 9.25. If we define the width where $Z = \frac{R}{\sqrt{2}}$ (i.e. a decrease of -3 dB) as the impedance bandwidth $\Delta\omega$, we can show that $\Delta\omega = \frac{1}{RC}$. This means that the resonance pulsation ω_0 is controlled by L and C , and the bandwidth is set by R and C .

The quality factor is defined as the ratio between ω_0 and the bandwidth:

$$Q = \frac{\omega_0}{\Delta\omega} = \omega_0 RC$$

With this expression, we can rewrite the expression for Z :

$$Z(\omega) = \frac{R}{1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)} \quad (9.3)$$

9.6 Class D Amplifier

Figure 9.26 shows the circuit of a class D amplifier. This amplifier uses switches: when the top switch is open, the one in the bottom is closed, and vice versa. This can be implemented with two MOS transistors, as in figure 9.27: a PMOS at the top and an NMOS at the bottom. When the input is high, the v_{GS} of the bottom transistor is high and it will conduct and pull the output to the ground. When the input is low, the top transistor will conduct and pull the output to the supply E . These circuits will be studied in more detail in chapter 13.

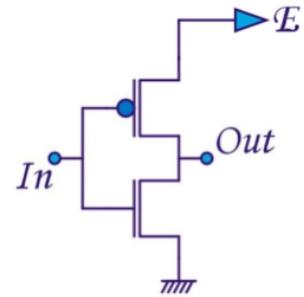
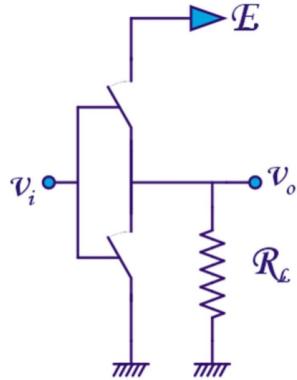


Figure 9.27

Figure 9.26

The advantage of a switch is that it never dissipates power: one of the switches will always be open, so no bias current can flow. In reality however, each of these switches (transistors) will have a parasitic resistor: each transistor will consume - when closed - a power P_C because it will operate in the linear region (see 5.2.3) and we can model it as a (small) resistor r_{ds} in series with R_L :

$$P_C = i_{ds} v_{ds} = \frac{E}{r_{ds} + R_L} \frac{r_{ds} E}{r_{ds} + R_L} = E^2 \frac{r_{ds}}{(r_{ds} + R_L)^2} \approx E^2 \frac{r_{ds}}{R_L^2}$$

The issue with this amplifier is that the amplitude is not preserved, so we can only transmit signals where the information is not encoded in the amplitude, but for example in the pulse width (pulse width modulation - PWM) or in the frequency (frequency modulation - FM). If the load is not a resistor but a pure capacitor, i.e. if we replace R_L by C_L , we can compute the associated power dissipation. When the top switch closes and the bottom one is open, a charge $Q = C_L E$ is transferred from the supply to the capacitance C_L . When the bottom switch is closed, the capacitor will discharge and the same amount of charge is transferred to ground. If this happens with a frequency f , we generate an effective current i_{cap} equal to $fC_L E$. The average power consumption over one period is thus:

$$P_{cap} = i_{cap} E = fC_L E^2$$

9.7 Class S Amplifier

A class D amplifier can not be used when the signal amplitude is important, as in AM modulation. However, we can modify a class D amplifier to obtain a class S amplifier that is capable of amplifying AM signals.

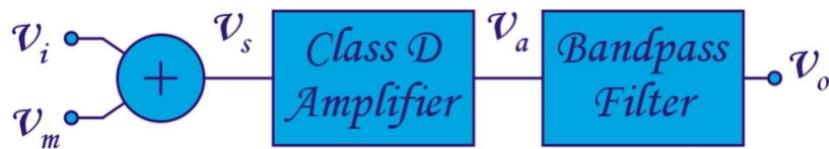


Figure 9.28: Block diagram of a class S amplifier

To do this, we first take the sum of the input signal v_i , which we assume has the form $v_i = A(t) \cos(\omega_0 t)$, with a signal $v_m = \cos(\frac{\omega_0}{2}t)$. We assume that $|A(t)| \ll 1$, so that $|v_m| \gg |v_i|$. By representing these signals as Fresnel vectors as in figure 9.21, with v_m in red, v_i in green and $v_s = v_i + v_m$ in yellow, we see the small green arrow v_i rotates fast around the slowly rotating red arrow v_m . By tracing out the end-points of the green arrow v_i , we see that the amplitude information $A(t)$ of v_i is now encoded in the phase of v_s . We can now use a class D amplifier to amplify the signal, losing all amplitude information but keeping the phase information intact, and extract the amplitude information from the output signal with a bandpass filter, as in figure 9.28.

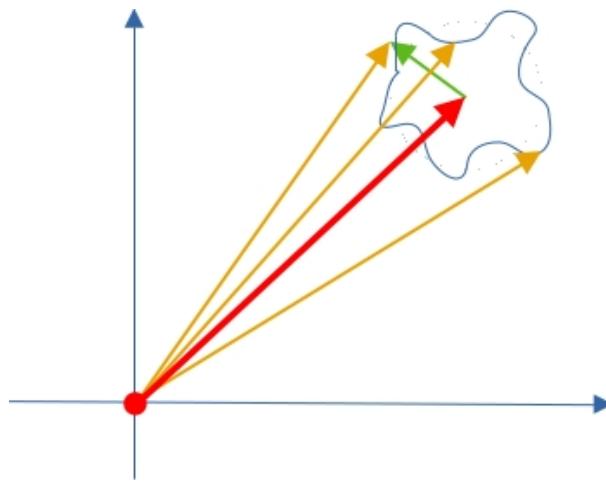


Figure 9.29

Mathematically, we can write:

$$\begin{aligned} v_s &= A(t) \cos(\omega_0 t) + \cos\left(\frac{\omega_0}{2}t\right) = A(t) \cos\left(\frac{\omega_0}{2}t\right) \cos\left(\frac{\omega_0}{2}t\right) - A(t) \sin\left(\frac{\omega_0}{2}t\right) \sin\left(\frac{\omega_0}{2}t\right) + \cos\left(\frac{\omega_0}{2}t\right) \\ &= [1 + A(t) \cos\left(\frac{\omega_0}{2}t\right)] \cos\left(\frac{\omega_0}{2}t\right) - [A(t) \sin\left(\frac{\omega_0}{2}t\right)] \sin\left(\frac{\omega_0}{2}t\right) \\ &= \sqrt{[1 + A(t) \cos\left(\frac{\omega_0}{2}t\right)]^2 + [A(t) \sin\left(\frac{\omega_0}{2}t\right)]^2} \cos\left[\frac{\omega_0}{2}t + \arctan \frac{A(t) \sin\left(\frac{\omega_0}{2}t\right)}{1 + A(t) \cos\left(\frac{\omega_0}{2}t\right)}\right] \end{aligned}$$

Because we lose the amplitude, the output of the class D amplifier v_a is equal to:

$$v_a = K \cos\left[\frac{\omega_0}{2}t + \arctan \frac{A(t) \sin\left(\frac{\omega_0}{2}t\right)}{1 + A(t) \cos\left(\frac{\omega_0}{2}t\right)}\right]$$

with K the gain of the amplifier. When we take into account that $|A(t)| \ll 1$, we can simplify this expression:

$$\begin{aligned} v_a &= K \cos\left[\frac{\omega_0}{2}t + \arctan \frac{A(t) \sin\left(\frac{\omega_0}{2}t\right)}{1 + A(t) \cos\left(\frac{\omega_0}{2}t\right)}\right] = K \cos\left[\frac{\omega_0}{2}t + \arctan A(t) \sin\left(\frac{\omega_0}{2}t\right)\right] \\ &\approx K \cos\left[\frac{\omega_0}{2}t + A(t) \sin\left(\frac{\omega_0}{2}t\right)\right] \approx K \cos\left(\frac{\omega_0}{2}t\right) \cos[A(t) \sin\left(\frac{\omega_0}{2}t\right)] - K \sin\left(\frac{\omega_0}{2}t\right) \sin[A(t) \sin\left(\frac{\omega_0}{2}t\right)] \\ &\approx K \cos\left(\frac{\omega_0}{2}t\right) - KA(t) \sin^2\left(\frac{\omega_0}{2}t\right) \\ &\approx K \cos\left(\frac{\omega_0}{2}t\right) - \frac{K}{2}A(t) + \frac{K}{2}A(t) \cos(\omega_0 t) \end{aligned}$$

After the bandpass filter, centered on ω_0 , the output signal becomes:

$$v_o \approx \frac{K}{2} A(t) \cos(\omega_0 t)$$

We have thus amplified with high efficiency, and conserved the amplitude $|A(t)|$.

9.8 The Selective Amplifier

Until now, we have amplified a relatively large bandwidth, and we have made no effort to restrict the amplification only to a specific frequency. In most telecommunications applications however, we are often only interested in a specific segment of the spectrum. This means we want to amplify a narrow bandwidth around the center frequency ω_0 , and discard all other frequencies.

A perfect solution for this would be to use the RLC-circuit from section 9.5.1, which has as impedance the resistor R at resonance and zero impedance elsewhere as in figure 9.30. However, in reality every inductance L also has a small series resistance r_s . We want to replace this series combination of an ideal L with a parasitic r_s with a parallel combination of L with a resistance R_p as indicated in figure 9.31.

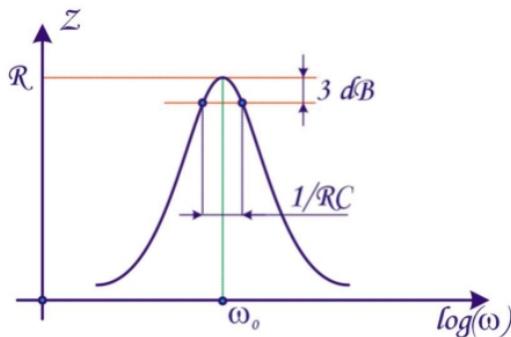


Figure 9.30

The impedance of an inductor L in series with a resistor r_s is:

$$\begin{aligned} j\omega L \parallel r_s &= j\omega L + r_s \\ \frac{1}{j\omega L + r_s} &= \frac{r_s - j\omega L}{r_s^2 + \omega^2 L^2} \\ &\approx \frac{r_s}{\omega^2 L^2} + \frac{1}{j\omega L} \end{aligned}$$

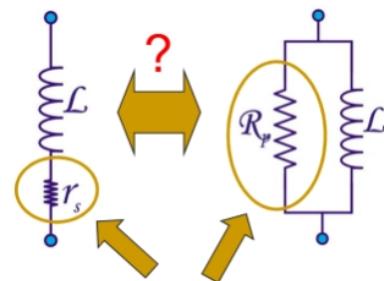


Figure 9.31

This last simplification is only valid if $\omega L \gg r_s$, or in other words, if the quality factor of the coil (the inductor) $Q_C = \frac{\omega L}{r_s}$ is a lot larger than 1. This means that we have an equivalent impedance consisting of a resistor $R_p = \frac{\omega^2 L^2}{r_s}$ in parallel with an inductance $j\omega L$.

When we add a capacitor in parallel with R_p and L , we get a true RLC-tank as in figure 9.24 - but in reality we have a non-ideal coil in parallel with C . The quality factor of the whole is $Q = \omega R_p C = \frac{R_p}{\omega_0 L} = Q_C$. So any LC-combination has in reality an impedance as in figure 9.30, but with a maximum of $R_p = \frac{\omega^2 L^2}{r_s}$ and a bandwidth of $\frac{1}{R_p C}$.

In a real RLC-circuit, we can push the resistance r_s to become a resistance R_p in parallel with

the true resistance R . The quality factor if the entire circuit is $Q = \omega_0(R_p||R)C < \omega_0 R_p C = Q_C$. So the use of an external resistor decreases the quality factor.

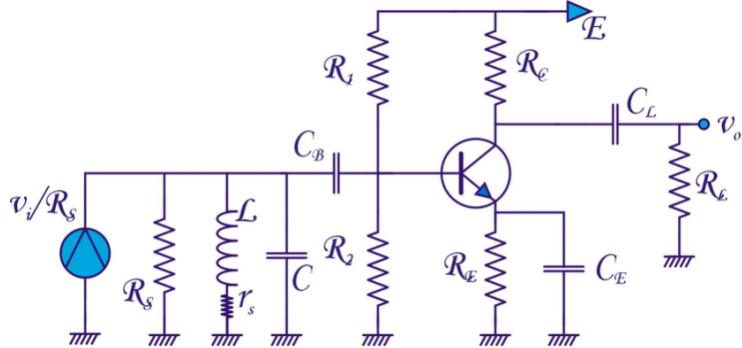


Figure 9.32

To construct a selective amplifier, we add the LC-circuit to the input stage of a common-emitter amplifier, as in figure 9.32. Note how the input has been replaced by the Norton equivalent, namely $Z_N = R_S$ and $I_N = v_i/R_S$. Because the parallel combination of L and C forms an RLC tank, there is an impedance Z that pulls the input to zero at almost all frequencies, except for the narrow band around ω_0 , where $Z = R_p$.

To compute the gain, we introduce the small-signal equivalent circuit in figure 9.33, where we suppose that both C_B and C_E are short circuits, i.e. we are in the normal operating domain, and C_L is also a short circuit. **TODO: gv_{be} arrow in wrong direction**

We can simplify the circuit in figure 9.33 by replacing C_μ by the Miller capacitance $C_M = (1 + gR_{eq})C_\mu$ with $R_{eq} = r_c||R_C||R_L$ and by grouping all resistors in the left side together:

$$R_\pi = R_p||R_s||R_1||R_2||r_\pi$$

. To compute the gain, we observe that $v_{be} = \frac{v_i}{R_S} R_\pi$ - if R_S is very small, this reduces to $v_{be} \approx v_i$ - and the output voltage is:

$$v_o = -R_{eq} g \frac{v_i}{R_S} R_\pi$$

at $\omega = \omega_0$. At frequencies below or above the central frequency plus or minus the bandwidth, the gain is (almost) zero.

In summary we can say that (1) OPAMPS have multiple (2 or more) poles and (2) this will have an effect on circuit stability. This stability depends on the poles (and zeros) of the loop gain $T(j\omega)$ and thus implicitly on $A(j\omega)$ en $H(j\omega)$, of which the latter term is application dependent. In essence, there are two types of OPAMPS:

1. Compensated OPAMPS: for these devices, as long as $|H(j\omega)| < 1$, stability is guaranteed. These circuits have a large phase margin, but are inherently slow because their bandwidth is artificially reduced.
2. Uncompensated OPAMPS: no measures were taken to assure stability; the user is responsible for this. These devices do have maximal bandwidth.

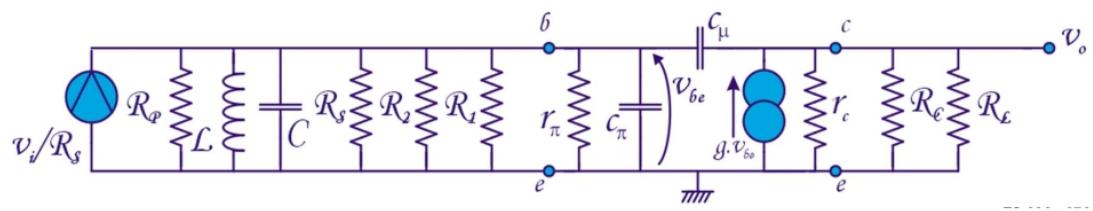


Figure 9.33: The selective amplifier

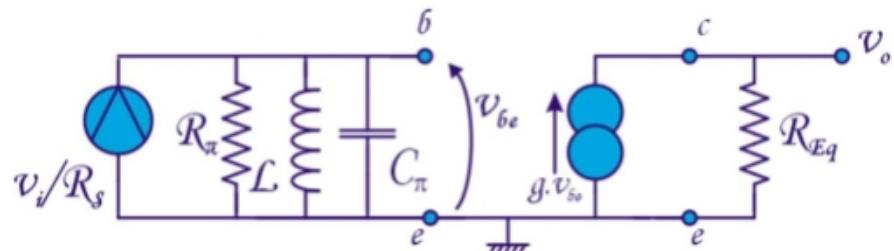


Figure 9.34

Chapter 10

Feedback Theory

When we talk about feedback, we mean that part of the output signal is fed back to the input. What is applied to the input of the circuit is not the input signal, but rather the difference between input signal and the returned output signal.

We will see in this chapter that feedback offers many advantages:

- It allows for accurate gain control,
- You can control the input and output impedance,
- You can extend the bandwidth,
- The distortions are reduced

We discuss these topics in turn, and afterwards have a look at some stability issues that may arise by applying feedback.

10.1 Accurate Gain Control

In figure 10.1, the output voltage v_o , generated by an amplifier with transmittance $A(j\omega)$, is returned to the input through a system with transmittance $H(j\omega)$. We can write:

$$\begin{aligned} v_f &= H(j\omega) v_o \\ v_c &= v_i - v_f = v_i - H(j\omega) v_o \\ v_o &= A(j\omega) v_c = A(j\omega) (v_i - H(j\omega) v_o) \\ \Rightarrow A_{CL} &= \frac{v_o}{v_i} = \frac{A(j\omega)}{1 + A(j\omega) H(j\omega)} \\ &= \frac{A}{1 + AH} = \frac{A}{1 + T} \end{aligned} \tag{10.1}$$

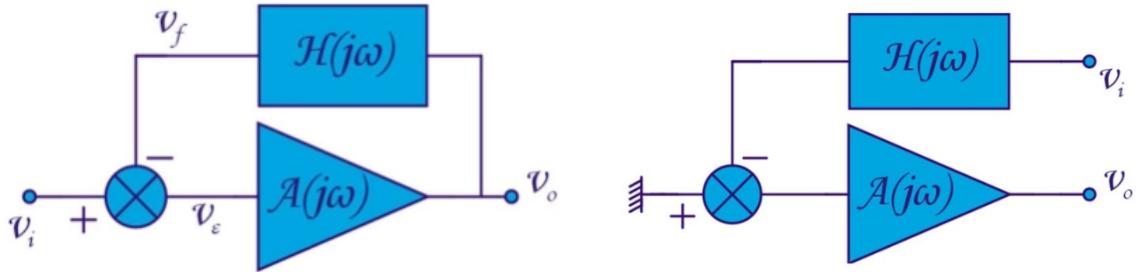


Figure 10.1

Figure 10.2

Some definitions: A is the open-loop gain and H the feedback gain, A_{CL} is called the closed-loop gain and $T = AH$ is the *loop gain*. The loop gain T can be obtained by this procedure:

1. Open the feedback loop at the amplifier output,
2. Ground the input,
3. Apply v_i at the open end of the loop,
4. Obtain T by computing v_o/v_i .

This method is visualized in figure 10.2, where we applied it to the circuit in 10.1. The signal at the input of the amplifier is $H v_i$, and $v_o = AH v_i$, so we find the loop gain.

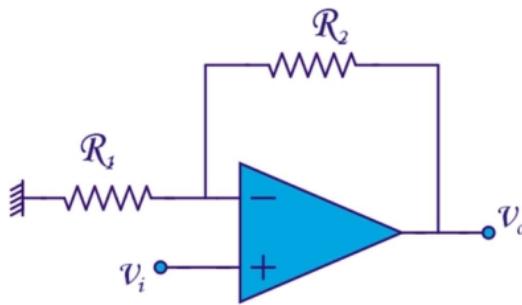


Figure 10.3

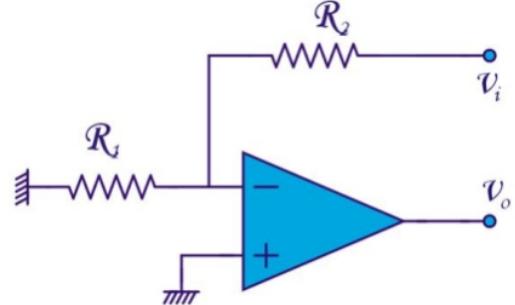


Figure 10.4

When we apply this to the non-inverting OPAMP topology that we saw in section 7.4, which is pictured in 10.3, we get the circuit in figure 10.4 with the feedback loop opened and the input grounded. The voltage at the negative terminal is $v^- = \frac{R_1}{R_1+R_2} v_i$, and the output is then $v_o = -A \frac{R_1}{R_1+R_2} v_i$. The loop gain T is thus :

$$T = -A \frac{R_1}{R_1 + R_2}$$

Typically, the amplifier gain A will be very high. In that case, the closed-loop gain doesn't depend on A :

$$\begin{aligned} A_{CL} &= \frac{A}{1 + AH} = \frac{1}{H} \frac{A}{1/H + A} \\ &= \frac{1}{H} \frac{1}{1 + T^{-1}} = \frac{1}{H} \frac{T}{T + 1} \\ &\approx \frac{1}{H} \end{aligned}$$

Usually, A is not only very large, but also difficult to control. It also depends on the small-signal parameters like g , g_m , r_c , r_{ds} , ... On the other hand, $\frac{1}{H}$ is not very high, but typically only depends on passive elements like resistors and capacitors, which can be very well controlled.

Applying this to the example of the non-inverting amplifier, we had $T = AH = -A \frac{R_1}{R_1+R_2}$, so $H = \frac{R_1}{R_1+R_2}$ or $H^{-1} = \frac{R_1+R_2}{R_1} = 1 + \frac{R_2}{R_1}$, just as we found in section 7.4.

10.2 Input and output impedance with feedback

Consider the circuit in figure 10.5. We compute the in- and output impedances Z_i and Z_o taking into account that no current enters the OPAMP:

$$Z_i = \frac{v_i}{i_i} = \frac{v_i}{(v_i - v_o)/Z} = \frac{v_i Z}{v_i + Av_i}$$

$$Z_o = \frac{v_o}{i_o} = \frac{v_o}{(v_o - v_i)/Z} = \frac{v_o Z}{v_o + v_o/A}$$

or, in other words:

$$Z_i = \frac{Z}{1 + A}$$

$$Z_o = \frac{Z}{1 + A^{-1}}$$

This is the same result as we found for the Miller capacitor in section 8.3.

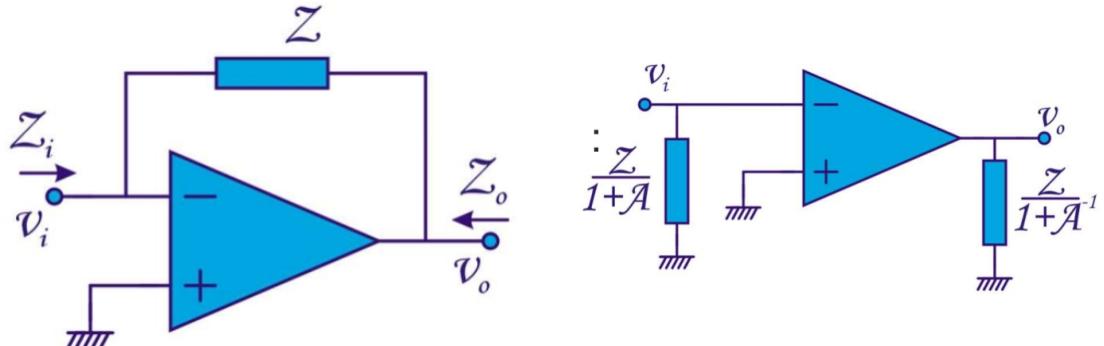


Figure 10.6

Figure 10.5

By using these relations, we can redraw the circuit as the one in figure 10.6, where we explicitly represented the in- and output impedances.

10.3 Increased bandwidth

Assume that the amplifier $A(j\omega)$ can be modeled as a system with a single pole in ω_0 :

$$A(j\omega) = \frac{A_0}{1 + j\frac{\omega}{\omega_0}}$$

Substituting this in the expression for the closed-loop gain A_{CL} gives:

$$\begin{aligned} A_{CL} &= \frac{A(j\omega)}{1 + A(j\omega)H} = \frac{\frac{A_0}{1+j\frac{\omega}{\omega_0}}}{1 + \frac{A_0}{1+j\frac{\omega}{\omega_0}}H} \\ &= \frac{\omega_0 A_0}{(\omega_0 + j\omega) + \omega_0 A_0 H} = \frac{\omega_0 A_0}{(1 + T_0)\omega_0 + j\omega} \\ &= \frac{\frac{A_0}{1+T_0}}{1 + j\frac{\omega}{\omega_0(1+T_0)}} \end{aligned}$$

So the closed-loop system is also a first-order system, but now:

- The gain A_0 is divided by the loop gain plus one.
- The bandwidth is increased by the same factor because the pole lies in $\omega_0(1 + T_0)$.
- The gain-bandwidth product GBW remains constant.

This is the same conclusion as in figure 7.33.

10.4 Distortion reduction

We model a distortion as an unwanted signal somewhere downstream in our circuit. Take for example the circuit in figure 10.7, where the first amplifier is an OPAMP, and the second is a class B push-pull amplifier, like in figure 9.20, including the feedback path through R_1 and R_2 , modeled with $H(j\omega)$. The second amplifier suffers from a large distortion, modeled as a voltage v_d applied at its input. Without feedback, this distortion is directly visible at the

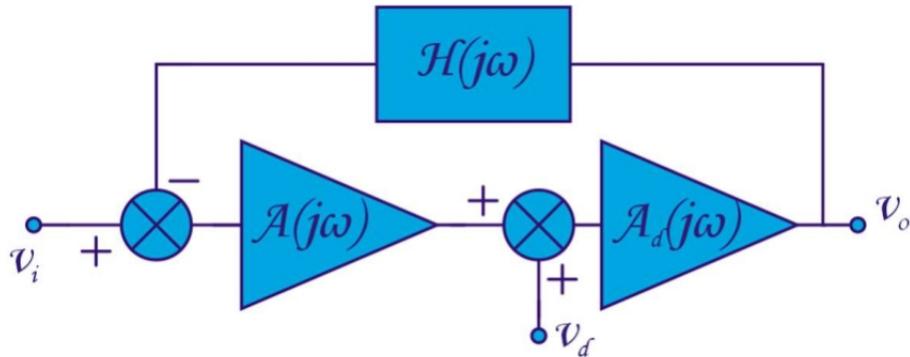


Figure 10.7

output, magnified by the gain A_d : $v_o = A(j\omega)A_d(j\omega)v_i + A_d(j\omega)v_d$. However, with feedback this becomes:

$$\begin{aligned} v_o &= A_d(v_d + A(v_i - Hv_o)) \\ (1 + A_dAH) v_o &= A_dv_d + A_dAv_i \\ v_o &= \frac{AA_d}{1+T}v_i + \frac{A_d}{1+T}v_d \\ &\approx \frac{1}{H}v_i + \frac{1}{AH}v_d \end{aligned}$$

with the loop gain $T = AA_dH$. This means that the distortion is reduced by a factor A compared to the input signal. Knowing that a distortion downstream will be reduced by the gain of the previous stages, it is important to place a distortion-less amplifier as the first stage of the loop.

In the circuit in figure 9.20, we needed diodes as a pre-bias to avoid cross-distortion. But thanks to the feedback, if you would omit the diodes in the second stage of figure 10.7, there would be no distortion either, because every distortion in the second stage will be divided by A .

10.5 Stability Issues

To keep the operating point of the OPAMP with feedback, is essential that we apply the feedback to the negative terminal. Otherwise, the system becomes unstable and the output will go to infinity (will explode) - in reality, off course, the output will be set to the highest voltage available in the circuit, typically the OPAMP supply $\pm E$.

In principle, we can analyze the frequency response of the feedback topology in figure 10.1 by expressing both the Laplace transforms¹ on $A(s)$ and $H(s)$ as the ratio of two polynomials:

$$A(s) = \frac{N_A(s)}{D_A(s)} \text{ and } H(s) = \frac{N_H(s)}{D_H(s)}$$

and substituting in the expression for the closed-loop 10.1:

$$\begin{aligned} A_{CL}(s) &= \frac{A(s)}{1 + A(s)H(s)} = \frac{\frac{N_A(s)}{D_A(s)}}{1 + \frac{N_A(s)}{D_A(s)} \frac{N_H(s)}{D_H(s)}} \\ &= \frac{N_A(s)D_H(s)}{D_A(s)D_H(s) + N_A(s)N_H(s)} \end{aligned}$$

To verify the stability, we must calculate the poles and check whether their real part is negative (i.e. do they lie in the left half of the s-plane). If this is the case for all poles, the system is asymptotically stable. However, to find the poles, we must solve $D_A(s)D_H(s) + N_A(s)N_H(s) = 0$. This method is usually very complex because it requires the factoring of high order polynomials, and it gives no indication of the stability margins, i.e. how much can the different parameters vary before the system becomes unstable.

Let's take another approach. We will assume that the open loop gain $A(j\omega)$ is low-pass: i.e. it has a pole in $\omega = \omega_d$ and for all frequencies $\omega < \omega_d$, $A(j\omega) = A_0$ is constant. For all $\omega > \omega_d$ the gain decreases. This means we model $A(j\omega)$ with following expression, where ω_d is the dominant pole and ω_{nd} is the non-dominant pole²:

$$A(j\omega) = \frac{A_0}{(1 + \frac{\omega}{\omega_d})(1 + \frac{\omega}{\omega_{nd}})}$$

With $T(j\omega) = A(j\omega)H(j\omega)$, we know that the closed-loop gain is

$$A_{CL} = \frac{A(j\omega)}{1 + T(j\omega)}$$

¹Note that $s = \sigma + j\omega$. We usually set $\sigma = 0$, i.e. we analyze the frequency response of the system.

²In general, there can be many of those non-dominating poles; what is important is that the ω_d determines the low-frequency behavior.

This means that if $T(j\omega) > 0$ the system is always stable. The problem arises when $T(j\omega) = -1$, i.e. if for a certain ω_ϕ :

- $|T(\omega_\phi)| = 1$, and
- $\angle T(\omega_\phi)) = 180^\circ$

So even when we apply feedback to the negative terminal, if the frequency response is such that the signal undergoes an additional phase shift of 180° , it can create an unstable circuit. The magnitude at the frequency ω where this phase shift happens must be ≥ 1 .

To demonstrate this, consider the topology in figure 10.8. We established earlier that $T = \frac{R_1}{R_1+R_2} A(j\omega)$. Notice that the feedback is applied to the negative terminal. A small perturbation that occurs at the negative input, as shown by the up arrow, will be amplified and becomes a larger perturbation at the output, but in the other direction. This output perturbation will be transmitted to the input by the feedback path but, as long as $T > 0$, **will act against the original perturbation**. The perturbation can not be sustained and will die out. The circuit is unconditionally stable.

Now assume the feedback is done on the positive terminal as in figure 10.9. The perturbation will travel around the circuit and will arrive back with the same direction at the input node. This means that the perturbation will not only be sustained but magnified, and the system becomes unstable. Note that this situation would be identical to the one in figure 10.8 if T were negative.

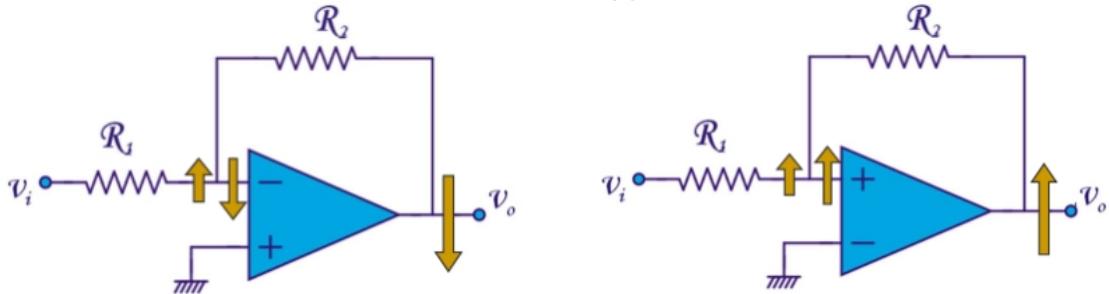


Figure 10.8

Figure 10.9

If H has no frequency dependence and $A(j\omega)$ has a single pole ω_d , we can solve the problem analytically. With $A(j\omega) = \frac{A_0}{1+j\frac{\omega}{\omega_d}}$, we find for the (closed) loop gain:

$$T(j\omega) = \frac{A_0 H}{1 + j\frac{\omega}{\omega_d}} \text{ and } A_{CL} = \frac{A_0}{1 + A_0 H} \frac{1}{1 + j\frac{\omega}{\omega_d(1+A_0 H)}} = \frac{A_{CL0}}{1 + j\frac{\omega}{\omega_{CL}}}$$

which is a result we already found when we discussed the increased bandwidth. Interpreting this as the movement of pole in the s -plane, we see that the existing pole ω_d has moved to the left:

$$\omega_{CL} = (1 + A_0 H) \omega_d$$

This movement is shown in figure 10.10. This is the *root-locus*, because it traces how the root(s) (poles) of the closed-loop response move as function of H . The closed-loop always remains stable. This can also be learned from the Bode curve, because a first-order system can only have a maximal phase shift of 90° , and not 180° as needed for an instability.

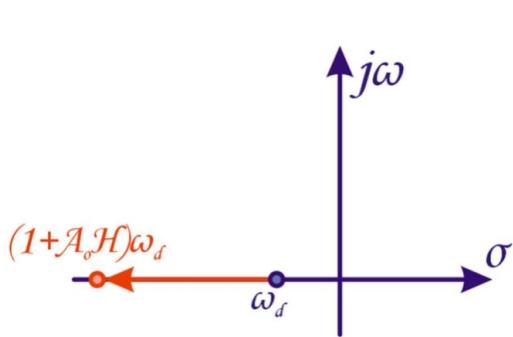


Figure 10.10

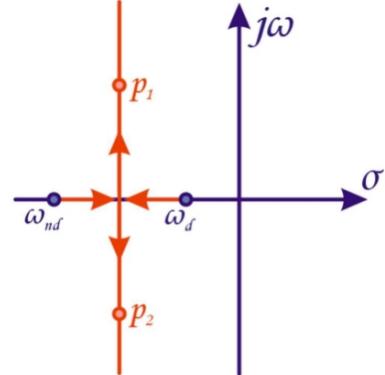


Figure 10.11

We can do the same analysis for a system with two poles, where $A(j\omega)$ is given by:

$$A(j\omega) = \frac{A_0}{(1 + \frac{\omega}{\omega_d})(1 + \frac{\omega}{\omega_{nd}})}$$

with ω_d the dominant and ω_{nd} the non-dominant pole, as before. Then:

$$T(j\omega) = A(j\omega)H(j\omega) = \frac{A_0 H}{(1 + \frac{\omega}{\omega_d})(1 + \frac{\omega}{\omega_{nd}})}$$

and

$$A_{CL} = \frac{A(j\omega)}{1 + T(j\omega)} = \frac{A_0 \omega_d \omega_{nd}}{(1 + A_0 H) \omega_d \omega_{nd} + j\omega(\omega_d + \omega_{nd}) - \omega^2}$$

The poles of A_{CL} are found by solving for the roots of the denominator. The result is:

$$\begin{aligned} p_{1,2} &= -\frac{1}{2}(\omega_d + \omega_{nd}) \pm \frac{1}{2}\sqrt{(\omega_d + \omega_{nd})^2 - 4(1 + A_0 H)\omega_d \omega_{nd}} \\ &\approx -\frac{1}{2}\omega_{nd} \pm \frac{1}{2}\sqrt{\omega_{nd}^2 - 4\omega_{CLd}\omega_{nd}} \end{aligned}$$

This means that if $1 + A_0 H$ is small, the poles are real and located close to $-\omega_d$ and $-\omega_{nd}$. As $A_0 H$ increases, they move closer together. At some point (i.e. when the part under the square root is zero) they collide and are both equal to $-\frac{1}{2}(\omega_d + \omega_{nd})$. If $A_0 H$ increases further, the term under the square root becomes negative and the poles acquire an imaginary part, i.e. they become complex when

$$A_0 H > \frac{(\omega_d - \omega_{nd})^2}{4\omega_d \omega_{nd}} \approx \frac{\omega_{nd}}{4\omega_d}$$

This situation is represented in the root-locus of figure 10.11. Note how even for a second-order system, the poles always remain in the left half-plane and so the system stays stable. The imaginary component of the poles is reflected in the fact that the circuit may oscillate temporarily and may appear unstable.

The maximal phase-shift of a second order system is 180° , but this shift only occurs when $\omega \rightarrow \infty$ so the system stays stable. This however does not mean that there are no issues. Even stable systems can have a large overshoot (i.e. how much does the step response goes beyond its final value) in the step response, as in figure 10.12. This figure shows the step

response of a second order system, where the damping ratio ζ decreases. The damping ratio ζ is related to the ratio of the imaginary and real component of the complex pole: if the poles are both real, $\zeta > 1$ and there are no oscillations; if $\zeta = 0$, the system is undamped and even a small perturbation is enough to generate a continuous oscillation. If ζ is small, the overshoot increases. Ideally, $\zeta \approx 0.46$ to have a reasonable fast step response without too much overshoot.

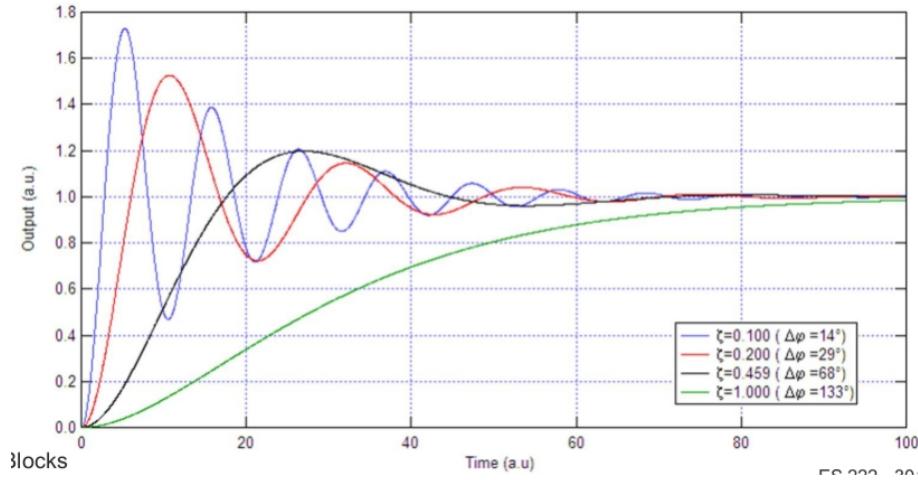


Figure 10.12

The damping ratio is related to the *phase margin* $\Delta\phi$. This is the difference between the phase at the frequency where the gain is 0 dB (i.e. = 1) and a phase of 180° . Figure 10.13 shows the (normalized) bode curves and phase margin for different values of ζ . The ideal ζ corresponds to a phase margin $\Delta\phi = 68^\circ$. If $\Delta\phi = 0$, the system is unstable. So in a sense the phase margin gives an indication how far we are from instability.

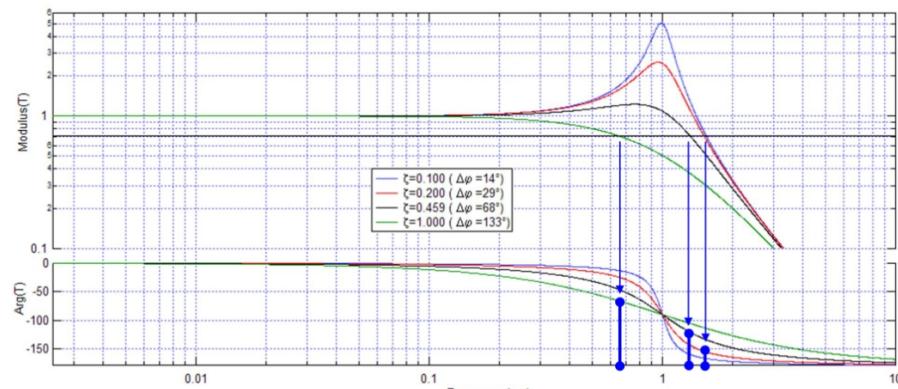


Figure 10.13

Chapter 11

Oscillators

Usually, what we want to do when applying feedback is to control the loop gain $T(j\omega)$ such that the amplifier becomes and stays stable. However, in this chapter we try to do the opposite: for a certain frequency ω_0 , we design the circuit such that $T(j\omega_0) = H(j\omega_0)A(j\omega_0) = 1$. This means that the circuit will be unstable for one single frequency ω_0 . These circuits generate sinusoidal output signals at ω_0 , even when there is no signal at the input¹. We call them *oscillators*.

11.1 Phase Shift Oscillator

This oscillator consists of a single common-emitter amplifier that (a) provides the gain and (b) provides a 180° phase shift. To generate the other 180° phase shift, 3 RC blocks are placed in series at the output of the amplifier, as in figure 11.1. So we have an amplifier, followed by feedback loop of 3 RC blocks, that couple the output signal back to the input. If each RC circuit could function independent from the others, we could just tune R and C to provide a 60° shift for a required frequency ω_0 . However, because the each next block loads the previous one, the calculation is more complicated.

To determine ω_0 , we establish the AC equivalent circuit as in figure 11.2 - with $R_{eq} = r_c || R_C$ and $R_\pi = r_\pi || R_1 || R_2$ - and compute the loop gain $T(j\omega)$. Then we verify the oscillation conditions:

- $|T(j\omega_0)| = 1$
- $\angle T(j\omega_0) = 0$

so for one frequency, the signal travels around the loop and arrives in phase and with the same magnitude back at the input² This condition also implies that $\Im\{T(j\omega)\} = 0$ and $\Re\{T(j\omega)\} = 1$.

To compute the loop gain, we cut the link between the input and output of the transistor, and compute which v'_{be} at R_π will be generated by current source $g v_{be}$. The loop gain is then $T = \frac{v'_{be}}{v_{be}}$. The resistance R' is chosen such that $R_\pi + R' = R$.

¹There is always *noise* present, and white noise contains every frequency. This will be studied in other courses

²If $|T(j\omega_0)| > 1$, we wont get a nice sinusoid but a block signal at the output.

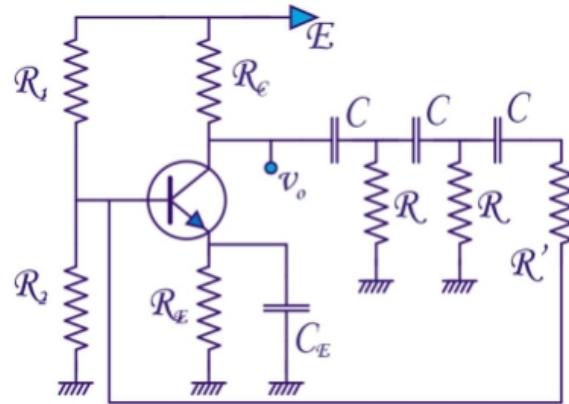


Figure 11.1

From the AC equivalent circuit, we find:

$$-g v_{be} = -g R_\pi i_b \approx i_b \left(3 + \frac{4}{j\omega RC} - \frac{1}{\omega^2 R^2 C^2} + \frac{R}{R_{eq}} + \frac{6}{j\omega R_{eq} C} - \frac{5}{\omega^2 C^2 R R_{eq}} - \frac{1}{j\omega^3 R^2 R_{eq} C^3} \right)$$

The imaginary part gives us ω_0 :

$$\begin{aligned} \frac{4}{RC} + \frac{6}{R_{eq}C} - \frac{1}{\omega_0^2 R^2 R_{eq} C^3} &= 0 \\ \Rightarrow \frac{4}{R} + \frac{6}{R_{eq}} &= \frac{1}{\omega_0^2 R^2 R_{eq} C^2} \\ \Rightarrow 4R_{eq} + 6R &= \frac{1}{\omega_0^2 R C^2} \\ \Rightarrow \omega_0^2 &= \frac{1}{(4R_{eq} + 6R) R C^2} \end{aligned}$$

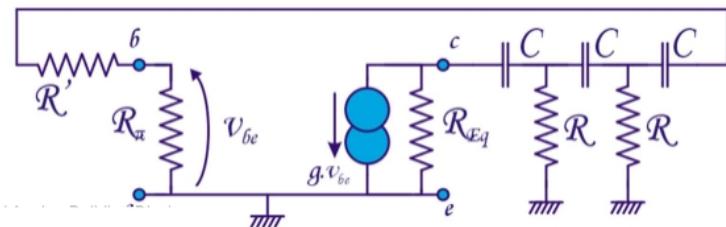


Figure 11.2

The real part provides the criterion for oscillation:

$$\begin{aligned} -gR_\pi &= 3 + \frac{1}{\omega^2 R^2 C^2} + \frac{R}{R_{eq}} - \frac{5}{\omega^2 C^2 R R_{eq}} \\ \Rightarrow -gR_\pi &= 3 - \frac{6R + 4R_{eq}}{R} + \frac{R}{R_{eq}} - \frac{5(6R + 4R_{eq})}{R_{eq}} \\ (gR_\pi - 23) + 4\frac{R_{eq}}{R} - 29\frac{R}{R_{eq}} &= 0 \end{aligned}$$

This equation can be solved for the ratio $\frac{R}{R_{eq}}$:

$$\frac{R}{R_{eq}} = \frac{gR_\pi - 23}{58} + \sqrt{(gR_\pi - 23)^2 - 46458}$$

This equation only has a solution only if $gR_\pi > 23 + \sqrt{464} \approx 44.6$, which puts a lower limit on the transconductance g and so on the bias current I_{CQ} of the amplifier. If g is higher, the output will be distorted; if it is lower, there will be no oscillation.

11.2 Wien Bridge Oscillator

The Wien bridge oscillator is a popular oscillator that uses an OPAMP with two feedback loops:

- One stable loop through resistances R_S and R_F
- An unstable loop (on the positive terminal) through C in series with R ($= Z_F$), and C in parallel with R ($= Z_S$):

$$\begin{aligned} Z_F &= R + \frac{1}{j\omega C} = \frac{1 + jRC\omega}{j\omega C} \\ Z_S &= R \parallel \frac{1}{j\omega C} = \frac{R}{1 + jRC\omega} \end{aligned}$$

The loop gain can be found by computing the input of the OPAMP $v_i = v_p - v_n$ and by setting $v_o = Av_i$:

$$\begin{aligned} v_n &= \frac{R_S}{R_S + R_F} v_o \\ v_p &= \frac{Z_S}{Z_S + Z_F} v_o = \frac{j\omega RC}{1 + j3RC\omega - R^2 C^2 \omega^2} v_o \\ v_o &= A v_i = A (v_p - v_n) \\ &= A \left(\frac{j\omega RC}{1 + j3RC\omega - R^2 C^2 \omega^2} - \frac{R_S}{R_S + R_F} \right) v_o \end{aligned}$$

With $A \gg 1$, this becomes:

$$\begin{aligned} \frac{j\omega RC}{1 + j3RC\omega - R^2 C^2 \omega^2} &\approx \frac{R_S}{R_S + R_F} = k \\ \Rightarrow j\omega RC &= k + j\omega 3kRC - \omega^2 kR^2 C^2 \end{aligned}$$

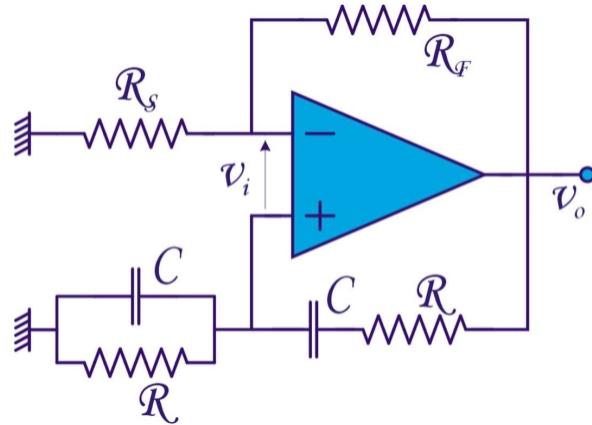


Figure 11.3

In this case, the real part gives ω_0 :

$$\begin{aligned} k - \omega_0^2 k R^2 C^2 &= 0 \\ \Rightarrow \omega_0 &= \frac{1}{RC} \end{aligned}$$

and the oscillation condition is given by the imaginary part:

$$\begin{aligned} j\omega RC &= j\omega 3kRC \\ \Rightarrow k &= \frac{R_S}{R_S + R_F} = \frac{1}{3} \end{aligned}$$

11.3 Colpitts Oscillator

A Colpitts oscillator is an oscillator that's very popular in radio application. It consists of a common-emitter amplifier followed by an inductor and two capacitors in the feedback path.

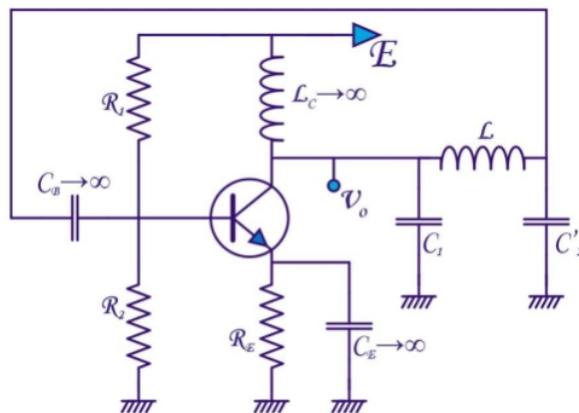


Figure 11.4

Inductor L_C and capacitor C_B are used for biasing and coupling in the signal respectively, as seen previously. The same goes for capacitor C_E , which short-circuits R_E . We can thus

omit them from the AC equivalent circuit in figure 11.5, where we do consider the parasitic transistor capacitances c_π and c_μ and replace the latter by the Miller capacitance C_M (i.e. we suppose that the Miller conditions are satisfied).

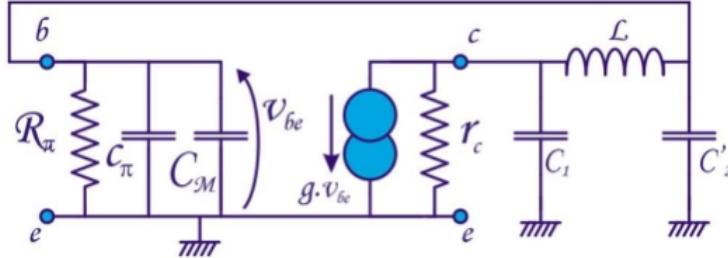


Figure 11.5

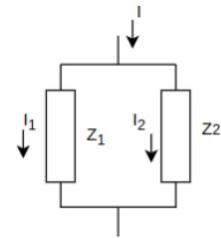


Figure 11.6

We can simplify further by grouping all capacitors between base and ground: $C_2 = c_\pi + C_M + C'_2$. Let $Z_1 = r_c||C_1 = \frac{r_c}{1+j\omega r_c C_1}$ and $Z_2 = j\omega L + (\frac{1}{j\omega C_2}||R_\pi) = j\omega L + \frac{R_\pi}{1+j\omega R_\pi C_2}$. As such, Z_1 and Z_2 are two impedances in parallel and they operate as a voltage divider (see figure 11.6):

$$I_2 = \frac{Z_1}{Z_1 + Z_2} I$$

Like for the phase-shift, we assume that the base-emitter voltage v_{be} is generated with the current source $-g v_{be}$. The current through the parallel combination of R_π and C_2 is then:

$$i_{R_\pi||C_2} = -g v_{be} \frac{Z_1}{Z_1 + Z_2}$$

and v_{be} is equal to:

$$\begin{aligned} v_{be} &= \frac{R_\pi}{1 + j\omega R_\pi C_2} i_{R_\pi||C_2} = -g v_{be} \frac{Z_1}{Z_1 + Z_2} \frac{R_\pi}{1 + j\omega R_\pi C_2} \\ &= -g v_{be} \frac{\frac{r_c}{1 + j\omega r_c C_1}}{\frac{r_c}{1 + j\omega r_c C_1} + j\omega L + \frac{R_\pi}{1 + j\omega R_\pi C_2}} \frac{R_\pi}{1 + j\omega R_\pi C_2} \\ &= -v_{be} \frac{gr_c R_\pi}{j\omega L(1 + j\omega R_\pi C_2)(1 + j\omega r_c C_1) + R_\pi(1 + j\omega r_c C_1) + r_c(1 + j\omega R_\pi C_2)} \end{aligned}$$

This last expression results in the condition:

$$j\omega L - \omega^2(R_\pi C_2 + r_c C_1)L - j\omega^3 r_c R_\pi C_1 C_2 L + R_\pi + j\omega r_c R_\pi C_1 + r_c + j\omega R_\pi r_c C_2 = -gr_c R_\pi$$

By examining the real parts of this equation, we find ω_0 :

$$\omega_0^2 = \frac{L + r_c R_\pi (C_1 + C_2)}{r_c R_\pi C_1 C_2 L} = \frac{1}{r_c R_\pi C_1 C_2} + \frac{C_1 + C_2}{C_1 C_2 L} \approx \frac{C_1 + C_2}{C_1 C_2 L}$$

From the imaginary part, the oscillation condition on the transconductance g can be found:

$$g + \frac{R_\pi + r_c}{r_c R_\pi} \approx \frac{C_1 + C_2}{C_1 C_2 L} (R_\pi C_2 + r_c C_1)$$

11.4 Quartz Oscillator

Regular oscillators are not very good time references: they drift have a precision of only $\sim 1\%$. This means that daily they have an error of about 15 minutes. A *quartz oscillator* - i.e. an oscillator based on the vibrations of quartz crystal - has a much higher precision: from $\sim 10^{-6}$ (about 30 seconds per year) to $\sim 10^{-7}$ if they are kept at a stabilized temperature. Quartz is a mineral composed of silicon and oxygen atoms arranged in a specific crystal structure. It exhibits a *piezoelectric effect*, which means that it can generate an electric charge when subjected to mechanical stress, such as pressure or vibration. This effect also works in the other way: quartz changes shape when a voltage is applied.

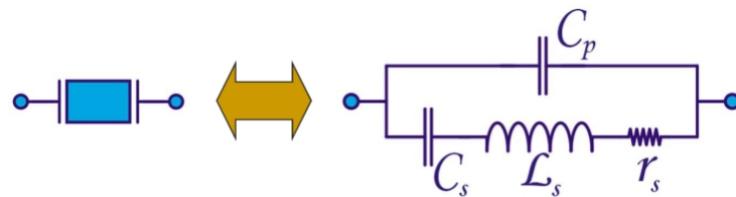


Figure 11.7

The quartz crystal with the external clamps is represented by the symbol in the left part of figure 11.7. From an electrical point of view, a quartz crystal can be modeled as in the circuit in the right part of the figure. The different components are justified because the moving crystal has several mechanical characteristics for which we use an electrical equivalent:

- Stiffness, because you have to provide energy before anything moves: modeled by a capacitor $C_s \sim 10fF$
- Mass, because there is inertia when you have to displace mass: modeled by an inductor $L_s \sim 1000H$
- Friction, because heat is generated when the crystal moves: modeled by a resistance $r_s \sim 100\Omega$
- The clamps (due to the external connectors): modeled by a capacitor $C_p \sim 10pF$

If we ignore r_s , we can compute the admittance $Y(j\omega)$ of the quartz crystal:

$$\begin{aligned} Y(j\omega) &= j\omega C_p + \frac{1}{j\omega L_s + \frac{1}{j\omega C_s}} = j\omega C_p + \frac{j\omega C_s}{1 - \frac{\omega^2}{\omega_s^2}} \\ &= j\omega C_p \frac{(1 + \frac{C_s}{C_p})\omega_s^2 - \omega^2}{\omega_s^2 - \omega^2} = j\omega C_p \frac{\omega_p^2 - \omega^2}{\omega_s^2 - \omega^2} \end{aligned}$$

with important frequencies $\omega_s^2 = \frac{1}{L_s C_s}$ and $\omega_p^2 = (1 + \frac{C_s}{C_p})\omega_s^2$.

This admittance is purely imaginary. When we plot the imaginary part of the impedance $Z(j\omega) = \frac{1}{Y(j\omega)}$ as function of ω , we find the curve in blue in figure 11.8, with an asymptote in ω_p . The curve in red is the same result, but now with the friction resistance r_c . If $\omega < \omega_s$ or $\omega > \omega_p$, the imaginary part of Z is negative, so Z is capacitive. When $\omega_s < \omega < \omega_p$, Z becomes inductive.

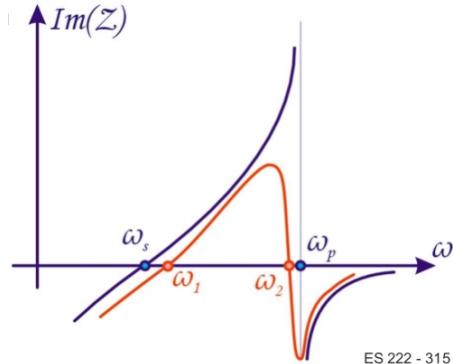


Figure 11.8

We can construct a Colpitts oscillator with a quartz crystal instead of an inductor L , as in figure 11.9. This oscillator has an oscillation frequency $\omega_0 = \frac{1}{LC}$ with C the parallel combination of C_1 and C_2 . This frequency necessarily lies between ω_s and ω_p because Z is only inductive in this region, and the corresponding inductance is determined by the red curve in figure 11.8. If ω_0 lies outside this region, Z would be capacitive and the oscillator would not work. If the frequency ω_0 would try to drift, e.g. because of a variation in C , the inductance L would also change, and the relatively high slope of Z will adjust L so that ω_0 remains relatively constant:

$$C \uparrow \Rightarrow \omega_0 = \sqrt{\frac{1}{LC}} \downarrow \Rightarrow L = Z(j\omega_0) \downarrow \Rightarrow \omega_0 \uparrow$$

This means the system is self-corrective and it is also why the quartz oscillator is very stable.

11.5 Relaxation Oscillator

The oscillators we saw so far were all based on the principle that for a certain frequency ω_0 , we want to create a loop gain $T(j\omega_0) = 1$. Another type of oscillator is the *relaxation oscillator*, which is based on the instability of an amplifier. We configure the circuit such that the output of an amplifier keeps switching between high and low, and a capacitor elsewhere in the circuit will eternally be charged and discharged (the *relaxation* of a capacitor). The output waveform will no longer be a sinusoid, but a block or triangle signal.

Whether a circuit is stable or not can be verified by Lyapounov's theorem. To do this:

- First find all the fixed points of the circuit, i.e. the solutions of the governing equations with all time derivatives and inputs set to zero.
- Linearize the system around these solutions.
- Identify how the operating point moves as time t increases (typically, the solutions have an exponential behavior).
- If the operating point returns to the solution, the system is stable. If not, the system might be unstable.

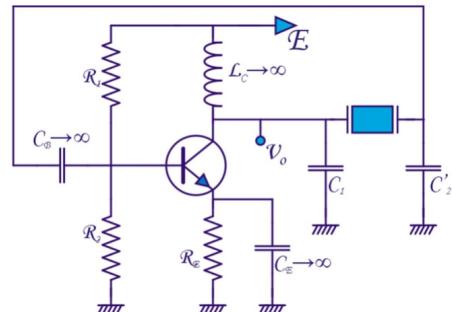


Figure 11.9

As an example, consider the circuit in figure 11.10, with its operating equations:

$$v_o = \phi(v) \text{ (the OPAMP characteristic)}$$

$$i = \frac{v_i + v}{R} = C \frac{dv_c}{dt}$$

$$v_o = -v + v_c$$

If $\frac{dv_c}{dt} = 0$, we find the fixed point: $i = 0, v_i = 0, v = 0, v_o = 0$.

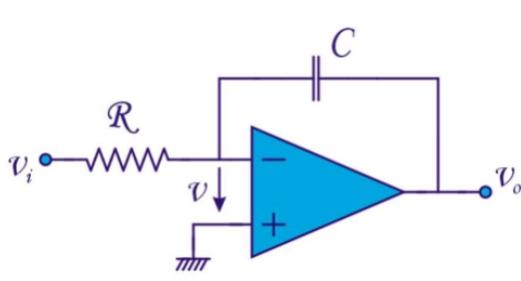


Figure 11.10

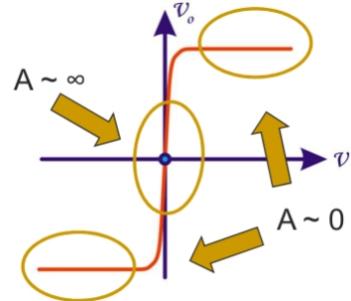


Figure 11.11

We linearize the operating equations around this fixed point. To do this, we approximate $v_o = \phi(v)$ by three different linear regions as in figure 11.11: two with gain $A \approx 0$, and one with $A \approx \infty$. The fixed point $v = v_o = 0$ is associated with $A \approx \infty$.

$$\begin{aligned} v_o &= A v = -v + v_c \Rightarrow v = \frac{v_c}{A+1} \\ (A+1)v_i &= (A+1)RC \frac{dv_c}{dt} - v_c \\ \Rightarrow v_c(t) &= (1+A)v_i(e^{t/(1+A)T} - 1) \end{aligned}$$

So this system is unstable because $v_c \rightarrow \infty$ if $t \rightarrow \infty$ (except when $v_i = 0$). This is evident from the exponent in $e^{\alpha t}$: if $\alpha > 0$, the behavior is unstable, as we find here for $\alpha = \frac{1}{(1+A)T}$. Because $e^x \approx 1 + x$ when x is small, we find that for small values of t , $v_c(t) \approx \frac{t}{T}v_i$, i.e. we find the expression for an integrator, as expected.

Let's apply this to the circuit in figure 11.12, which is a *relaxation oscillator*.

The operating equations are:

$$v_o = \phi(v) \approx A v$$

$$i = C \frac{dv_c}{dt}$$

$$v_o - v_c = R i$$

$$v + v_c = \frac{R_1}{R_1 + R_2} v_o = k v_o$$

These equations have a fixed point around $v_o = v_c = v = 0$ and $i = 0$. When we linearize

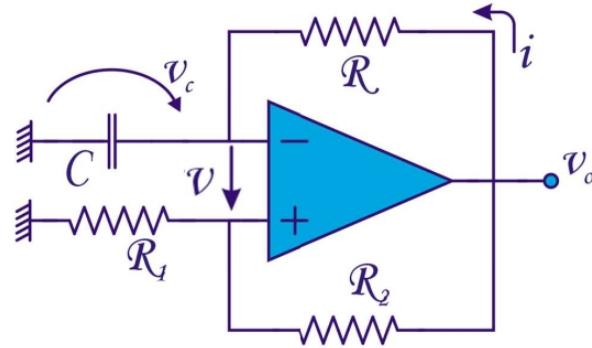


Figure 11.12

around this point ($v_o = A v$), we find:

$$\begin{aligned} v_c &= v_o - R i = A v - R C \frac{dv_c}{dt} \\ v + v_c &= k v_o = A k v \\ \rightarrow v_c &= (A k - 1) v \\ \rightarrow v_c &= \frac{A}{A k - 1} v_c - R C \frac{dv_c}{dt} \end{aligned}$$

and so:

$$\begin{aligned} v_c &= \frac{A k - 1}{A k - A - 1} R C \frac{dv_c}{dt} \\ &= \frac{k - 1/A}{1 - k + 1/A} R C \frac{dv_c}{dt} \end{aligned}$$

The stability criterion is thus:

$$\frac{k - 1/A}{1 - k + 1/A} > 0$$

or in other words: $A > \frac{1}{k}$.

To understand how the circuit works, assume that $v_o = E$. Then capacitor C is charging and the voltage at the negative input increases. At some point in time, the voltage at v^- becomes larger than $v^+ = \frac{R_1}{R_1 + R_2} E$ and because of the large gain, the output voltage switches to $v_o = -E$. The capacitor C has still a charge of $C \frac{R_1}{R_1 + R_2} E$ on it, so v^- changes abruptly to $-E + \frac{R_1}{R_1 + R_2} E$. Now, the process reverses: the capacitor starts to discharge, the voltage at v^- decreases and suddenly v becomes positive and the output switches again. The process is shown in figure 11.13.

When C is charging or discharging, we can write:

$$v_c(t) = v_c(\infty) + (v_c(0) - v_c(\infty)) e^{-t/T}$$

With this equation, we can compute the frequency of oscillation: the switching criterion is when $v = 0$, i.e. at time $t = T$ when

$$kE = E + (-kE - E) e^{-\frac{T}{2RC}}$$

and thus $T = 2RC \ln \frac{1+k}{1-k}$.

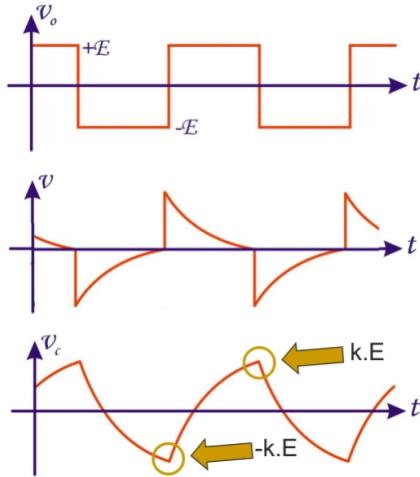


Figure 11.13

11.6 The Fantastron

The *fantastron* (figure 11.14) is a type of relaxation oscillator that consists of

1. An unstable comparator (also called a *Schmidt trigger*): if v_p switches sign, the output voltage will switch from $+E$ to $-E$ because the feedback happens on v^+ .
2. An integrator with time constant RC .

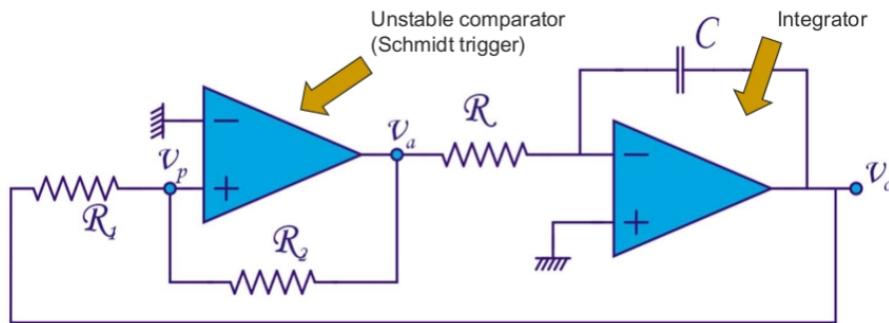


Figure 11.14

With v_a equal to $\pm E$, we can write the voltage v_p as (Millman):

$$v_p = \frac{R_2 v_o \pm R_1 E}{R_1 + R_2}$$

If $v_p > 0$, $v_a = E$, and the integrator will start to integrate down with slope $-\frac{E}{RC}$, until the output reaches $v_o = \pm E \frac{R_1}{R_2}$. At that time, v_p will become negative, v_a switches suddenly to $v_a = -E$, and the process starts all over again in the other direction. The waveforms for v_a and v_o are shown in figure 11.15. In this circuit, we have simultaneously access to a block signal as to a triangle waveform.

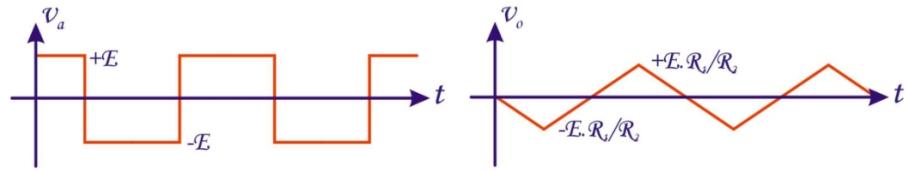


Figure 11.15

The period is thus determined by when v_p switches sign, i.e. one half period $\frac{T}{2}$ corresponds to:

$$-\frac{R_1}{R_2}E + \frac{E}{RC} \frac{T}{2} = \frac{R_1}{R_2}E$$

and so:

$$T = 4 \frac{R_1}{R_2} RC$$

Chapter 12

DC Voltage Generation

The problem we address in this chapter is the generation of a fixed DC voltage of arbitrary value, without too much variation (*ripple*). The supply we have at our disposal is the local AC voltage, which can have a value of 110 V, 210 V or 220 V, and a frequency of 50 Hz or 60 Hz, depending on where you are in the world.

Part III

Digital Electronics

Chapter 13

Digital Circuits

13.1 Combinatorial Digital Systems

13.2 Sequential Digital Systems

Chapter 14

Advanced Digital Circuits

Chapter 15

A/D and D/A Converter Circuits

Bibliography

- [1] Sze, Simon M., Yiming Li, and Kwok K. Ng. Physics of semiconductor devices. John wiley & sons, 2021.
- [2] Razavi, Behzad. Fundamentals of microelectronics. John Wiley & Sons, 2021.
- [3] Schilling, Donald L., Charles Below. Electronic circuits: discrete and integrated. New York: McGraw-Hill, 1989.