

THE SCIENCE OF ELECTRIC GUITARS AND GUITAR ELECTRONICS

Revision July 1, 2014

THE SCIENCE OF ELECTRIC GUITARS AND GUITAR ELECTRONICS

Jarmo Lähdevaara

This is a revised edition of the published first edition (ISBN 978-952-286-420-8).
This revised edition is available only in electronic format without a specific ISBN.
Distribution as a pdf-file only from the website <http://www.guitarscience.net>.

©2012 – 2014 Jarmo Lähdevaara

PRELUDE

This is a book about travelling. It describes the life journey of a ♫ which is set free by plucking a string of an electric guitar. On its fascinating trip towards feeling the freedom of the outside world, our little ♫ travels through pickups, wires, cables, effect devices and amplifiers, all the way down to the loudspeaker. During this trip, ♫ faces a lot of difficult situations; it has to push through resistances, fight between the plates of capacitors and take a ride in the whirls of inductors. Luckily our little ♫ meets a lot of friends along the way. By the help of transistors and electronic tubes, our little ♫ gains a lot of energy and at the end of the signal chain ♫ has grown large enough to set itself free into the surrounding air, feeling the pressure changes and wind in its stem and eventually disappearing into the horizon. And no one will ever hear about it again.

This book is a collection of all the knowledge gained over the years of practising to play bass guitar. Somehow the technical side of things is sometimes more appealing than the artistic side. After several years of studying in the field of technology, the author realised that everything learnt on the way can be thought about the viewpoint of music and especially guitars. This way of thinking has helped the author to learn many topics in engineering in practise by making a concrete reference to the technology of music. Therefore, this book is a documentation of different topics in physics, music and engineering, all connected by that one bass guitar.

All the topics presented in this book are composed together from already known facts, which means that very little new information regarding the topics is produced. The thing that makes this book differ from other similar publications is the view into the whole chain of scientific elements that guitar players are dealing with when playing their instrument. The primary motivation to document all this information in the form of a book was to share the knowledge in one package, instead of having all the bits and pieces scattered around in several books, articles and the Internet. The ultimate goal in the writing process has been that the reader of this book could find answers to most of the fundamental questions which might arise from the technological point of view of electric guitars. The

book tries to answer questions like: What creates the tones that are heard from stringed musical instruments? How scientific properties of a guitar string affect the tone? How can basic components of electronics be used to modify tones? How different pickup types affect the tone? What is inside a guitar amplifier? How to achieve an electroacoustic transformation using loudspeakers? How are guitar effect pedals implemented? How does a guitar tuning device work?

These questions are the ones that have been puzzling the author for a long time. Others possibly share the same questions, but have not yet managed to find satisfying answers to them. This book tries to give an answer to these questions, and tries to form a complete picture of the complex signal chain that is formed by a guitar and all the accessories related to it. In addition to this, the book melts together two completely different branches of science. Hopefully, by the help of this book, a common language between physics and music will be found.

The book is typeset using \LaTeX . All the pictures have been produced with gnuplot or MetaPost. Special thanks to John Hobbes for creating MetaPost and also thanks to Gustavo S. Bustamante Argañaraz for preparing a MetaPost add-on called MakeCirc, which has been used to draw most of the circuit diagrams for this book. In addition, Musi \TeX was used to create the musical notations, and Octave was used as a Matlab replacement for the numerical verification of symbolic results. All the software to support practical projects related to the topics of this book was written in C language and compiled with the gcc compiler. All this artillery is running smoothly on the Linux XUbuntu laptop computer, which gracefully served the author during the whole writing process.

20.10.2012; Oulu, Finland

J.L.

DISCLAIMER

This book contains a few complete circuit diagrams, which are suitable to be built as do-it-yourself projects. While the circuits have been tested to be working in practise and the circuit diagrams have been drawn with care, there may be typographical errors in the schematics, which in the worst case could cause the circuit to malfunction. All the guitar effect devices presented in this book are intended to be powered by a regular 9 volt battery, which is safe to be used as a power source in almost any condition. For your own safety, never use any power source that connects to the mains outlet.

The guitar amplifier circuit diagrams analysed within the scope of this book do not add up into complete amplifiers because the analysis is presented only from a theoretical perspective. Guitar amplifiers would obviously require more than a 9 volt battery as a power source, so it is highly advised not to try to build those as a home-made project unless you are an experienced electrician. When doing wiring work on the tone and volume control section in the guitar, please make sure that the guitar cable is not connected to the guitar. Under the bridge of the guitar there often is a grounding cable, which makes you part of the electric circuit when touching the strings. In some short circuit conditions this might lead to lethal accidents. For your own safety, never use any power source that connects to the mains, and keep the guitar well away from any mains outlet at all times.

Although the text has been read through several times, certainly mistakes and errors, typographical and even conceptual ones, might be lurking within the paragraphs. The accuracy of the information presented within this book has not been inspected by any other person than the author, and therefore the reader is advised to keep a sceptical attitude towards the contents of this book. The author is to be accused for all the remaining mistakes and errors in the text.

All trademarks used in this book, registered or otherwise, are the property of their respective owners. This applies to all manufacturer and product names used when referring to commercially available amplifiers, effect devices, strings, cables and other related accessories.

CONTENTS

PRELUDE	v
DISCLAIMER	vii
1 INTRO	1
1.1 BASIC CONCEPTS OF MUSIC	1
1.1.1 Interpreting notes	2
1.1.2 A connection to physics	4
1.2 BASIC CONCEPTS OF PHYSICS	6
1.2.1 The differential vs. the derivative	6
1.2.2 The differential vs. the integral	9
1.2.3 Continuous time vs. discrete time	11
1.2.4 Fourier analysis vs. an amplitude spectrum	13
1.2.5 The concept of a decibel vs. an amplitude spectrum	29
1.2.6 The Laplace transform vs. a transfer function	32
1.3 BASIC CONCEPTS OF ELECTRONICS	39
1.3.1 Current and voltage	39
1.3.2 Elementary building blocks of electronics	44
1.3.3 Properties of resistors	44
1.3.4 Properties of capacitors	47
1.3.5 Properties of inductors	51
1.3.6 Circuit analysis methods	55
1.3.7 Properties of semiconductors	60
1.3.8 Properties of transistors	67
1.3.9 Transistor circuit DC analysis	70
1.3.10 Transistor circuit AC analysis	75
1.3.11 Input and output impedances	81
1.3.12 Circuit simulations with SPICE	82
2 VIBES	89
2.1 A MATHEMATICAL MODEL OF A GUITAR STRING	89
2.1.1 Setting up the coordinates	89
2.1.2 The ideal string model	91
2.1.3 Setting up the equations	95
2.1.4 Solving the transverse wave equation	99
2.1.5 The significance of A_n : the amplitude spectrum	106
2.1.6 Alternative solutions of the wave equation	109
2.2 NONLINEAR EFFECTS IN STRING VIBRATION	111
2.2.1 The effect of amplitude	111
2.2.2 The effect of friction	114
2.2.3 The effect of stiffness	118
2.2.4 The effect of end supports	123
2.2.5 Coupling between directions of vibration	132
2.3 NONLINEAR EQUATIONS OF MOTION	140
2.4 MEASURING THE STRING VIBRATIONS	151
2.4.1 Properties of the measured string	152

2.4.2	The measurement setup	152
2.4.3	Error limit calculations	155
2.4.4	Measurement results - Amplitude spectrum	157
2.4.5	Measurement results - Decay of upper partials	166
2.4.6	Measurement results - Pitch glide	170
3	GUITAR PICKUPS	175
3.1	THE EFFECT OF PICKUP POSITIONING	175
3.2	MAGNETIC PICKUPS	177
3.2.1	The basic construction and function of a magnetic pickup	178
3.2.2	The magnetic field of a pickup magnet	179
3.2.3	Magnetic interaction between a pickup and strings	184
3.2.4	Electromagnetic induction in a magnetic pickup	191
3.2.5	The magnetic field generated by the coil of a pickup	197
3.2.6	A single-coil pickup vs. a humbucker	199
3.2.7	A magnetic pickup as part of an electric circuit	202
3.2.8	Measurements on a magnetic pickup	209
3.3	OPTICAL PICKUPS	216
3.3.1	An optical transducer as a guitar pickup	217
3.3.2	An optical pickup as a signal source	220
3.3.3	An optical pickup as part of an electric circuit	226
3.3.4	Measurements on an optical pickup	228
3.4	PIEZOELECTRIC PICKUPS	232
3.4.1	The piezoelectric effect	233
3.4.2	A piezoelectric transducer as a signal source	234
3.4.3	A piezoelectric pickup as part of an electric circuit	239
3.4.4	Measurements on a piezoelectric pickup	246
4	GUITAR TONE CONTROL CIRCUITS	251
4.1	A BASIC TONE CONTROL CIRCUIT	251
4.2	MODIFIED TONE CONTROL CIRCUITS	257
4.3	THE 'BIG MUFF π ' TONE CONTROL CIRCUIT	262
4.4	AMPLIFIER TONE CONTROLS FOR GUITAR MOUNTING	266
5	GUITAR CABLES	271
5.1	THE GENERAL TRANSMISSION LINE THEORY	271
5.2	PRACTICAL GUITAR CABLES	276
5.2.1	A shielded instrument cable	276
5.2.2	A parallel wire cable	281
5.3	IMPEDANCE MATCHING	284
5.4	A CABLE AS PART OF A GUITAR TONE CONTROL CIRCUIT	289
6	EFFECT DEVICES	293
6.1	AMPLITUDE EFFECTS	294
6.2	FUZZ BOX VOODOO	295
6.2.1	The DC bias analysis of the fuzz effect	297
6.2.2	The AC analysis of the fuzz effect	300
6.2.3	Side effect: feedback in transistor circuits	303
6.2.4	A SPICE model for the fuzz circuit	325
6.2.5	Results of simulations, calculations and measurements	335
6.3	TREMOLO = AMPLITUDE MODULATION	342
6.3.1	The DC bias analysis of the tremolo effect	346
6.3.2	The AC analysis of the tremolo effect	347

6.3.3	Side effect: the JFET as a resistor	349
6.3.4	Side effect: the RC phase-shift oscillator	355
6.3.5	A SPICE model for the tremolo circuit	367
6.3.6	Results of simulations, calculations and measurements	369
6.4	A COMPRESSOR AND AN EXPANDER	378
6.5	FREQUENCY EFFECTS	384
6.6	THE ORIGINS OF THE WAH-WAH SOUND	385
6.6.1	The DC analysis of the wah effect	386
6.6.2	The AC analysis of the wah effect	391
6.6.3	Side effect: RLC circuit resonance	394
6.6.4	Side effect: the Miller theorem	396
6.6.5	A SPICE model for the wah circuit	399
6.6.6	Results of simulations, calculations and measurements	402
6.7	AN OCTAVE DOUBLER EFFECT	407
6.7.1	The DC analysis of the Green Ringer effect	408
6.7.2	The AC analysis of the Green Ringer effect	410
6.7.3	A SPICE model for the Green Ringer circuit	410
6.7.4	Results of simulations, calculations and measurements	412
6.8	PHASE EFFECTS	415
6.9	A PHASE SHIFTER	415
6.9.1	Side effect: the relaxation oscillator	418
6.9.2	Side effect: an op-amp all-pass filter	425
6.9.3	Side effect: the op-amp in the nodal matrix method	426
6.9.4	The DC analysis of the Phase 45 effect	428
6.9.5	The AC analysis of the Phase 45 effect	429
6.9.6	A SPICE model for the Phase 45 circuit	433
6.9.7	Results of simulations, calculations and measurements	437
6.10	A FLANGER	445
6.11	TIME EFFECTS	448
6.12	THEORETICAL PITCH SHIFT	449
6.13	A THEORETICAL REVERBERATION EFFECT	451
6.14	ANALOGUE DELAY ACTUATORS	453
6.15	A REVERBERATION EFFECT USING MECHANICAL SPRINGS	454
6.15.1	Wave propagation in a helical spring	455
6.15.2	Interfacing electrical and mechanical vibrations	464
6.15.3	A drive and recovery circuit for a reverberation unit	468
6.15.4	Measurements on the reverberation device	472
7	GUITAR AMPLIFIERS	479
7.1	COMMON GUITAR AMPLIFIER BUILDING BLOCKS	480
7.2	OPERATION MODE CLASSES OF AUDIO AMPLIFIERS	481
7.3	DISTORTION IN AUDIO POWER AMPLIFIERS	483
7.4	TUBE AMPLIFIERS	485
7.4.1	The anatomy of different tube types	486
7.4.2	A tube preamplifier	488
7.4.3	A tube amplifier tone control circuit	492
7.4.4	Tube phase splitters	495
7.4.5	A tube power amplifier	503
7.5	TRANSISTOR AMPLIFIERS	508
7.5.1	A solid-state preamplifier	509
7.5.2	A solid-state power amplifier	513
7.6	INTERFACING AMPLIFIERS WITH LOUDSPEAKERS	519

7.7 GUITAR AMPLIFIER SIMULATIONS WITH SPICE	520
8 LOUDSPEAKERS	527
8.1 THE CONSTRUCTION OF A MOVING-COIL LOUDSPEAKER ELEMENT	527
8.2 MODELLING THE VOICE COIL	530
8.3 ELECTROMECHANICAL ANALOGUES REVISITED	532
8.4 AN ELECTRICAL MODEL OF THE DRIVER	538
8.5 MEASURING THE ESSENTIAL DRIVER PARAMETERS	544
9 TUNING DEVICES	549
9.1 STATE VARIABLE BIQUAD FILTERS	550
9.2 A MONOSTABLE MULTIVIBRATOR	558
9.3 AN ANALOGUE GUITAR TUNING DEVICE	561
9.4 MEASUREMENTS ON THE TUNING CIRCUIT	570
9.5 THE STROBOTUNER	576
9.6 AUTOMATIC TUNING SYSTEMS FOR GUITARS	580
10 OUTRO	583
REFERENCES	589
NOMENCLATURE	593
ABBREVIATIONS	598
APPENDIX A	599
APPENDIX B	601
APPENDIX C	605
APPENDIX D	613
APPENDIX E	615

CHAPTER 1

INTRO

This first chapter covers some basic ideas on the languages of music, physics and electronics, including the most relevant theoretical background on the topics discussed in this book. This introductory chapter is intended as a quick reference section extending the rest of the chapters, although it does not form a complete and uniform coverage of all topics. In some cases the theoretical treatment is presented from the viewpoint of the author and should not be taken as the absolutely correct view. It is advisable to read the material with a critical mindset.

1.1 BASIC CONCEPTS OF MUSIC

There are several different theories on the science of music. Different locations of the world have their own view on musical notations, tunings and scales. The so-called mainstream music is considered to be composed in "the West", meaning mainly the countries of Europe and America [1, p. 230]. Therefore, the musical ideologies described in this book follow the theories of Western music. This is justified from the viewpoint of a guitar player, since the guitar is commonly considered a Western instrument. By focusing mainly on Western music, one is restricted to examining only the properties of the twelve-tone equal temperament scale [1, p. 231], a concept that will be explained later.

Musical masterpieces are documented using a distinct language, a notation, which only professional musicians understand. By representing a song by writing notes on a staff, a musician can document every sound and tone to be reproduced with correct timings and distinct tonal embellishments by other musicians. To fully understand the topics covered in this book, some terms and symbols from the musical vocabulary need to be explained.

1.1.1 Interpreting notes

The sound of music is composed from three principal elements. The first of these elements is *pitch*, which indicates the relative highness or lowness of a sound. The second element is *duration*, i.e. the relative length of a sound, and the third principal element is *loudness*, which refers to the intensity or strength of a sound [2, p. 1]. These three basic elements are mapped into musical notations as described in Figure 1.1.



Figure 1.1: Musical notations of pitch and duration

The notes are written into a space limited by five horizontal lines called a *staff*. In Figure 1.1 the staff has been divided to left and right sides. On the left side are the most usual notations that define the pitch of the sound, whereas the right side gives some examples on the notation related to duration. Let's explore the left side first.

The extracted notation starts with a *clef* symbol. There are a few different kinds of clef symbols; the one used here is a *bass clef*, a.k.a. the *F clef* symbol. It automatically assigns fixed note names to the five staff lines. Notes are generally named with capital letters, e.g. *A*, *B*, *C*, *E*, *F* and *G*, although there are national differences in the note naming conventions. Here we will use namings introduced in American music literature; for example, as seen in reference [2]. The different *octaves* of these notes are expressed using subscript numbering, e.g. *A*₁ and *A*₂. This kind of note naming system gives a concrete meaning for the different pitches of notes. The bass clef starts from the note named as *E*₂, representing the sound of the lowest open string of a guitar. Surprisingly, the bass guitar's low E-string actually has the note *E*₁. This small conflict is discussed a little later.

The location of a note on a certain staff line is not enough to accurately depict all possible pitches. In addition there are specific notations for small pitch modifications. The notation of \flat is used to lower the pitch of a note by a semitone and the notation of \sharp is used to raise the pitch of a note by a semitone. The notation of \natural is used to return the normal pitch to a note which has been modified earlier using \flat or \sharp . The concept of semitone is explained in section 1.1.2.

On the right side of Figure 1.1 is a collection of usual notations of duration. First there is a *treble clef* symbol, which is the most common of all clef symbols. This is followed by a *meter signature*, which is used to indicate the rhythmic ordering of notes of different duration in the staff. Then, keeping the pitch constant, notes of different duration are shown. Starting from the whole note, its duration fills one full bar when using the most common $\frac{4}{4}$ meter signature. The following notes are derived from the whole note by dividing the duration of the previous note by a factor of two: half note, quarter note, eighth note, sixteenth note and 32nd note. [2, pp. 2–11]

In addition to these notations of duration, a general *tempo* can be set to fix a certain duration for a specific note such as a quarter note, e.g. $\text{♩} = 80$. This also sets the other notes' durations relative to the tempo setting of the quarter note.

As already mentioned, different clef symbols are used to assign written music to certain instruments. The most common division done with clefs is to use the bass clef for low tone instruments and the treble clef for the rest of the instruments. However, this is not a good method to indicate the actual pitches of notes, since the pitch variation of different instruments is much larger than can be written using normal bass and treble clefs. Of course an additional written notification can be used to indicate a certain instrument, but from a scientific point of view the pitch is not well defined in these cases. Figure 1.2 tries to explain the actual pitch ranges that can be written using the most common clef symbols.

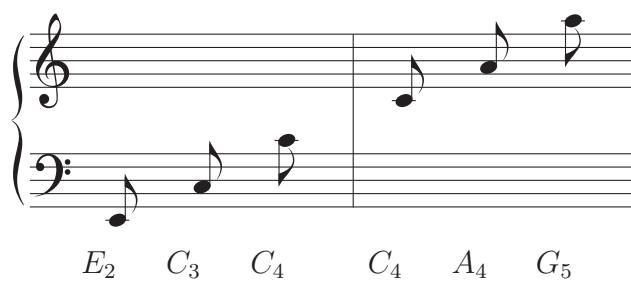


Figure 1.2: The boundary between the bass clef and the treble clef

The use of clefs automatically affects the pitches of the notes written on the staff. Noting from Figure 1.2, the C_4 note acts as a median value between the bass and the treble clefs. Approaching music from a more scientific point of view, one can ask the question: how are different instruments mapped to the bass clef and the treble clef? The thickest string of a bass guitar vibrates as E_1 ... How to write music for the bass guitar since there are not enough staff lines available? The same question can also be assigned to the guitar, which lowest string vibrates as

E_2 . Should guitar notes be written on the bass clef? Sometimes it is just clear from the context that notes to be played by bass are written on the normal bass clef and the instrument name is assigned as bass. But if there is a need to use exact notations, then one would write music for bass as:

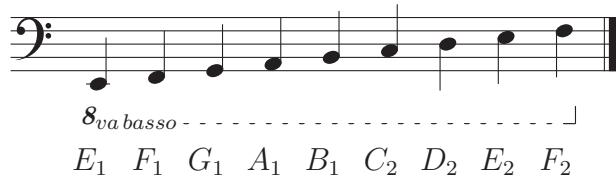


Figure 1.3: Scaling the staff to the actual pitch of the instrument

The same procedure should be applied to guitar notes when using the treble clef. There are similar definitions for raising the actual pitch defined by a certain clef symbol. These definitions and the proper way of using them are described thoroughly in reference [2].

1.1.2 A connection to physics

General scientific definition of sound describes sound waves as pressure fluctuations [3, p. 528]. Also from scientific perspective, pitch is related to *frequency*, duration is related to *time* and loudness relates to *energy* [2, p. 1]. When considering pitch as a scientific quantity, it is a somewhat naive simplification to say that pitch equals frequency, since in the musical world pitch is seen through the *sensation* of a heard sound [4, pp. 1–11].

Even when playing the same note, different instruments create a different sensation of sound, which is described to have a certain *timbre*, *tone colour* [2, p. 1] or *tone quality* [5, p. 18]. Physics explains timbre as a sum of several frequencies, where the lowest (fundamental) frequency defines the actual pitch for the sound, but one still cannot equalise the sums of frequencies and the sensation of hearing a sound. The first treatment ever to combine physics and music was published by the great German scientist Hermann Helmholtz back in the year 1862 [5]. This publication also refers to the sensations of tone, as a direct implication from the title of the book. Therefore, it is not all that simple to give an exact definition for sound, tones and music purely based on the theories provided by physics.

On the other hand, it is acceptable to say that every *written* note has a certain frequency (a.k.a. pitch). These frequencies obey a certain logical construction that can be used to calculate specific frequencies for all notes in relation to some

other note for which the frequency is already known.

The frequencies of harmonic upper partials (often misleadingly called overtones) of a note are integer multiples of the fundamental frequency [5, p. 22]. This series of upper partials also covers all octaves of a note because octaves are 2^n ($n = \text{some positive integer}$) multiples of the fundamental frequency. In a typical guitar tuning scheme, the proceeding 12 frets from a certain note cover the whole octave. Moving one fret upwards or downwards from a note equals a change of a half tone (semitone) in the pitch. Hence, it is possible to derive a geometric series equation for the dependencies between different notes: to double the frequency of a note (taking the octave) one needs to take 12 steps (semitones). The multiplying factor to change a note by a half tone would then be

$$\sqrt[12]{2} \cdot \text{♪} = \# \text{♪}, \quad (1.1)$$

where ♪ means any note and #♪ the same note, but raised by a semitone. This kind of formula is naturally different for all the different scales; here we are assuming that the normal twelve-tone equal temperament scale is used.

When mathematically examining musical scales and temperaments, a more accurate unit can be used to measure pitch. This unit is called a *cent* and it divides the semitone into smaller parts. This kind of measure was not yet acknowledged in the times of Helmholtz, but in the English translation a decent definition of a cent is already introduced [5, p. 431]. A cent divides the interval of equal semitones (1.1) into 100 equal intervals, so therefore it can be used to define pitches more accurately. The geometrical factor term of a cent is $\sqrt[1200]{2}$ and this directly implies that an octave is measured as 1200 cents [6, p. 45].

When tunings of different instruments are considered, the equal temperament scale uses the convention that the middle *A*-note (A_4) is fixed to the frequency of 440 Hz. This is the basis for mapping notes to the frequency plane. The frequencies of all other notes can be calculated from the A_4 -note using the equation (1.1). A series of note frequencies is depicted in Figure 1.4.

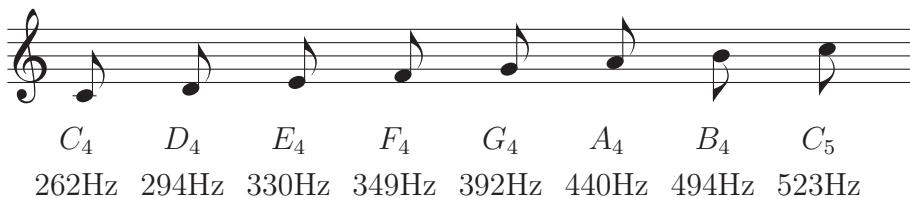


Figure 1.4: The mappings of notes to the physical frequencies of vibration

1.2 BASIC CONCEPTS OF PHYSICS

As the musical notations were already explained in the previous section, some of the most essential mathematical concepts used in physics need to be covered. The presentation is laid out as what is seen by the author to be essential for understanding the physical phenomena introduced in this book. In addition this means that mathematical exactness in this chapter is replaced by practical and intuitive formalism.

1.2.1 The differential vs. the derivative

The idea behind the differential is to express the exact change $\Delta f(x)$ in the value of a function $f(x)$ between two points, x and $x + \Delta x$, i.e.

$$f(x + \Delta x) - f(x) = \Delta f(x), \quad (1.2)$$

using a linear approximation. If the actual increment Δx is relatively small, a good approximation is obtained by using the slope described by the derivative $\frac{df(x)}{dx}$ and the infinitesimal (= very small) increment dx as

$$\Delta f(x) \approx df(x) = \frac{df(x)}{dx} dx = f'(x) dx. \quad (1.3)$$

To understand this property of the differential, one needs to know the practical meaning of the derivative. In connection to slopes and linear approximations, the derivative is based on the mathematical definition of a straight line. For example, a drawing book of a little child contains exercises that require connecting dots together by drawing a line through all the points. The equation that defines all the dots along a straight line is

$$y(x) = k(x - x_0) + y(x_0), \quad (1.4)$$

where k is identified as the slope of the line. This means that if two points, (x_0, y_0) and (x_1, y_1) , exist within the path of the line, the change Δy in the value of $y(x)$ can be calculated from the equation

$$\Delta y = y(x_1) - y(x_0) = y_1 - y_0 = k(x_1 - x_0). \quad (1.5)$$

This linear dependency connecting the slope of the line with some arbitrary increment $x_1 - x_0$ is depicted in Figure 1.5. Related to the derivative where the increment is assumed to be very small, the linear approximation via differentials is purely based on the basic equation of a straight line.

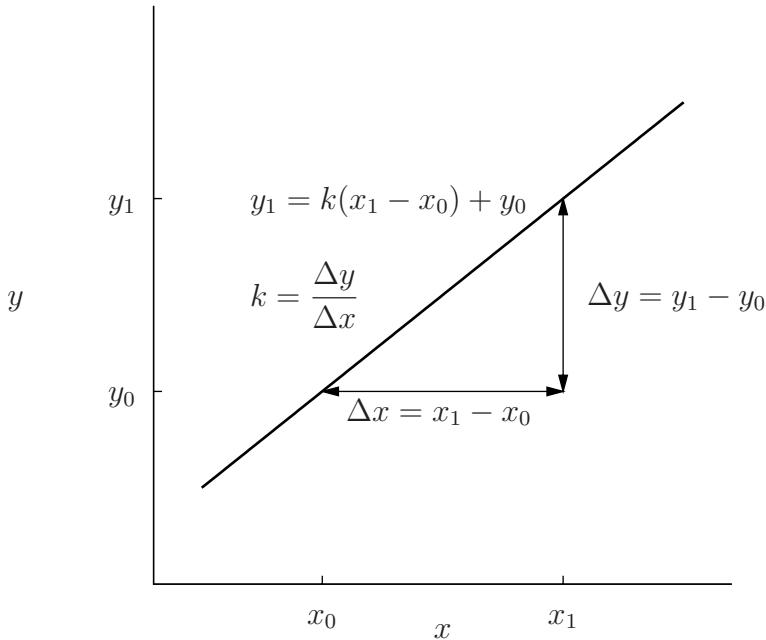


Figure 1.5: Geometry of a straight line

Let's take a closer look at the geometrical properties and the mathematical definition of the derivative. The derivative of a function $f(x)$ calculated at some specific point $Q_0 = f(x_0)$ is geometrically the slope of the tangential line that touches the function $f(x)$ at $f(x_0)$ as depicted in Figure 1.6.

In Figure 1.6 there are also other lines drawn to connect the point Q_0 to points P_1 , P_2 and P_3 respectively. This is to clarify the mathematical definition of the derivative given in equation (1.6); when connecting points of a function $f(x)$ with a straight line, one approaches a tangent line when the two points become closer and closer to each other. When the points are overlapping and on the verge of becoming the one and the same point, the mathematical definition of the derivative

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{x + \Delta x - x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (1.6)$$

is obtained. Hence, the derivative indeed depicts the slope of a straight line connecting points x and $x + \Delta x$ of a function $f(x)$, but only when Δx can be considered to have a value very close to zero.

From the point of view of physics, this purely mathematical definition of the derivative is not very practical. The more important quantity in physics is the differential $df(x)$ of the function $f(x)$. Whereas the derivative only refers to a slope, the differential depicts the actual change of a function or any system in

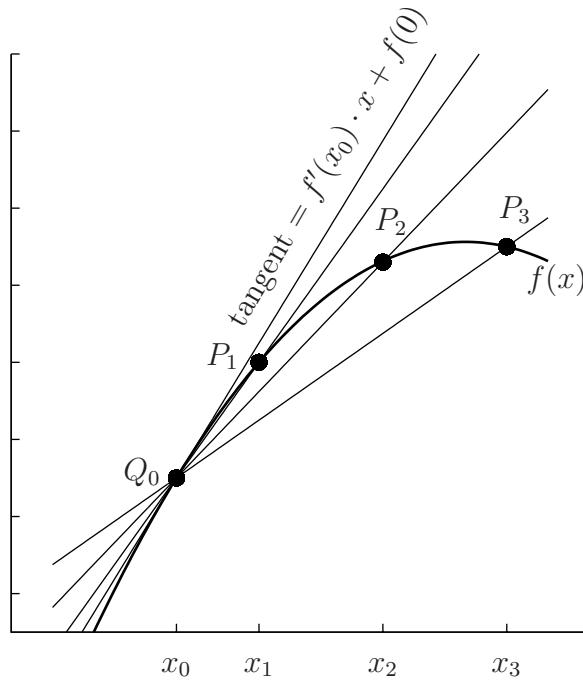


Figure 1.6: The tangential slope at Q_0 vs. lines to points P_x of $f(x)$

some infinitesimal interval. The differential of a function is evaluated as

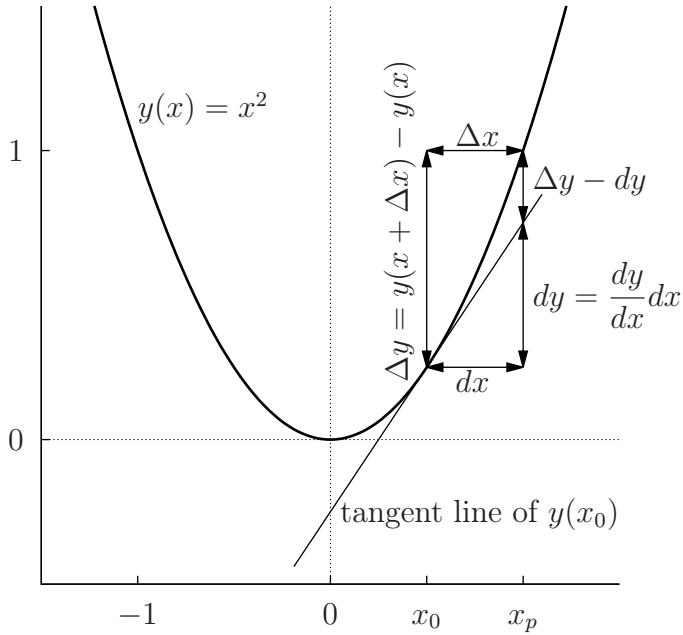
$$dy = y(x + dx) - y(x) = \frac{dy}{dx} dx = y'(x) dx. \quad (1.7)$$

From a mathematical point of view, this definition is not that acceptable, because the definition of infinitesimal values is not very exact. Basically this is only a notational issue, but to make differential equations and the definition of the integral make sense from the practical point of view, one needs to establish the fundamentals of the differential. Unlike the derivative, the differential has physical meaning. The differential is the basis for the mathematical modelling of physical systems. With differential equations one can analyse the behaviour of some real device without making any measurements on the actual device.

In practise the differential means that when moving an amount of dx from x_0 to $x_0 + dx$, the value of the function $f(x)$ has a new value at $f(x_0 + dx)$, which can be estimated by the rate of change (i.e. slope, i.e. derivative) at x_0 and moving along this constant change dx times towards $x_0 + dx$. The exact change is defined simply by equation (1.2) and the estimated approximate change is given by the differential $df(x)$ as depicted in Figure 1.7.

When dividing the equation (1.3) by dx , one obtains the normal derivative to both sides of the equation

$$\frac{df(x)}{dx} = f'(x). \quad (1.8)$$

Figure 1.7: Depiction of the definition of dy

Concluding from Figure 1.7, if one makes the assignment $\Delta x = dx$ then $\Delta y \neq dy$. The smaller dx is, the closer dy is to the exact change Δy in $y(x)$. According to Figure 1.7, it is clear that the differential dy is only a linear approximation of the amount of change in the actual function $y(x)$ in the interval dx . To extend the idea of differentials towards more exact approximations, one could try to fit a polynomial to the original function instead of linear fitting. This idea is known as the Taylor approximation. The Taylor approximation can be also presented as a series of derivatives. If the function $f(x)$ has derivatives of all orders, then

$$f(x_p) = f(x_0) + f'(x_0)(x_p - x_0) + \frac{f''(x_0)}{2!}(x_p - x_0)^2 + \dots = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x_p - x_0)^k, \quad (1.9)$$

where x_0 denotes some constant point on the x -axis, which is chosen as the base point for the approximation of $f(x)$ at some arbitrary point x_p . The practical use of this approximation concerns cases where the differential can easily be calculated for point x_0 , but not at $x = x_p$, which is a point near x_0 . The closer $x = x_p$ is to x_0 , the more accurate results follow from Taylor's formula (1.9).

1.2.2 The differential vs. the integral

Integration and differentiation are opposite mathematical operations. When integrating some function, one should assume that before integration the function

is a differential of some function, i.e.

$$\frac{dy}{dx} = f'(x) \quad \Rightarrow \quad dy = f'(x) dx = d[f(x)] \quad (1.10)$$

To build the original function y from the differential dy , one needs to sum all the infinitesimal differentials together with the integral operator:

$$\int dy = \int f'(x) dx = \int d[f(x)] \quad (1.11)$$

$$y = f(x)$$

Notice how the differential operator d and the integral operator \int cancel each other out. This justifies the use of dx in the integral notation as a real quantity of the differential of $f(x)$, and not just an indicator for the variable to be integrated. The point is that the infinitesimal dx can be used in the equations like a normal constant, which can be divided away with another dx . By taking dx as a real (but very small) physical quantity is the easiest way of understanding differential equations. The equations (1.11) show the essential link between differentiation and integration, which is definitely one of the most fundamental mathematical relations from the view point of practical applications.

Integration is sometimes called "antidifferentiation". The following set of equations (1.12) should explain the reason for this naming to further show the relationship between the differential and the integral.

$$y = f(x)$$

$$dy = \frac{df(x)}{dx} dx = f'(x) dx = d \left[\int_{z=a}^x f'(z) dz \right] \quad (1.12)$$

$$y = \int_{z=a}^x f'(z) dz = \int_{z=a}^x f(z) = f(x) - f(a)$$

$$= f(x) + C.$$

In equation (1.12) a is assumed to be any constant number. Pay careful attention how to raise the integral operator with the dependent variable x as the upper limit of the integral. This yields from the fundamental theorem of calculus, which says that

$$\frac{d}{dx} \int_a^x f'(z) dz = \frac{d}{dx} \int_a^x d[f(z)] = \frac{d}{dx} f(x) - \frac{d}{dx} f(a) = f'(x). \quad (1.13)$$

So what else happened in equation (1.12)? By differentiating a function and then antidifferentiating (cancelling the d operators from both sides), an additional constant C appeared from nowhere. There are two types of integrals, definite and indefinite integrals. This was an example of the indefinite integral, since the limits of integration were not explicitly defined as constant values, but instead contained the dependent variable x . This indefinite form is practical when solving differential equations and normally the integral sign in the indefinite form is written without any limits of integration. The constant of integration C is then fixed by assigning boundary constraints, which arise from a specific practical application. The derivation of the equations of the vibrating string will contain a practical example of this method.

The physical interpretation of the definite integral relates to area calculations within graphs. The idea is that areas in graphs are constrained by two or more variables, and their products lead to physical quantities such as energy, work and flux. Integrals appearing in practical applications are typically definite integrals because the integration limits reflect the constraints due to the physical system under evaluation; the indefinite integrals are used mainly in theoretical derivations.

1.2.3 Continuous time vs. discrete time

The previous sections have covered basic differentiation and integration methods with continuous variables. This means that a variable of a function, typically x , can have any value from the set of real numbers. Usually in practical applications the form of a function is not known, and all data points are measured at certain intervals of time. This creates a discrete time function, since if data points are measured in one second intervals, there is no way of knowing the value at 1.5 seconds, for example. Numerical methods in general have great practical significance when any data from real world measurements is analysed.

Differentials in discrete time: Numerical mathematical methods define the difference $\Delta f(x)$ between two consecutive data points separated by a constant k as

$$\Delta f(x) = f(x + k) - f(x). \quad (1.14)$$

This equation (1.14) is known as the forward difference formula. Naturally there exists a backward difference formula

$$\nabla f(x) = f(x) - f(x - k), \quad (1.15)$$

and the central difference equation

$$\Delta f(x) + \nabla f(x) = f(x+k) - f(x-k). \quad (1.16)$$

These formulae can be extended as numerical derivatives when bringing in the slope property of the derivative. The relative change in some variable, call it x , can be determined via difference formulae.

$$\frac{\Delta f(x)}{\Delta x} = \frac{f(x+k) - f(x)}{\Delta x} = \frac{f(x+k) - f(x)}{x+k-x} = \frac{f(x+k) - f(x)}{k}. \quad (1.17)$$

When combining the forward and backward difference formulae, a derivative based on the central difference is obtained.

$$\begin{aligned} \frac{\Delta f(x) + \nabla f(x)}{\Delta x} &= \frac{f(x+k) - f(x) + f(x) - f(x-k)}{\Delta x} \\ &= \frac{f(x+k) - f(x-k)}{x+k-(x-k)} \\ &= \frac{f(x+k) - f(x-k)}{2k}. \end{aligned} \quad (1.18)$$

All of these numerical derivative equations are officially derived from the Taylor series expansion by taking the first terms of the series, and making the proper substitutions and subtractions. For example, a substitution of $x_p = x_0 + k$ into the equation (1.9) gives $f(x_p) = f(x_0 + k)$ and the first terms are

$$f(x_0 + k) = f(x_0) + f'(x_0)[(x_0 + k) - x_0] \Rightarrow f'(x_0) = \frac{f(x_0 + k) - f(x_0)}{k}.$$

The smaller the constant k is, the more accurate derivative values are obtained from these numerical derivatives.

Integrals in discrete time: Numerical integration methods are obtained by approximating the integrand by functions that can be easily integrated. The simplest numerical integration formula is the *rectangular rule*, where the interval of integration $a \leq x \leq b$ is divided into n subintervals of equal length $h = (b-a)/n$. The value of the integrand is sampled at midpoint x_j of each subinterval j . Then the products $f(x_1)h, \dots, f(x_n)h$ represent the areas of the rectangular boxes depicted in Figure 1.8a. The complete definition of the rectangular rule is

$$J = \int_a^b f(x) dx \approx h[f(x_1) + f(x_2) + f(x_3) + \dots + f(x_n)],$$

where $h = (b-a)/n$.

More accurate numerical approximations for integrals are obtained by using the *trapezoidal rule*. In this method the area under integrand $f(x)$ between points a and b is approximated by n 'trapezoids', which have areas

$$\frac{1}{2}[f(a) + f(x_1)]h, \quad \frac{1}{2}[f(x_1) + f(x_2)]h, \quad \dots, \quad \frac{1}{2}[f(x_{n-1}) + f(b)]h,$$

as shown in Figure 1.8b. When these areas are added together, the definition of the trapezoidal rule is

$$J = \int_a^b f(x) dx \approx h \left[\frac{1}{2}f(a) + f(x_1) + f(x_2) + \dots + f(x_{n-1}) + \frac{1}{2}f(b) \right],$$

where $h = (b - a)/n$.

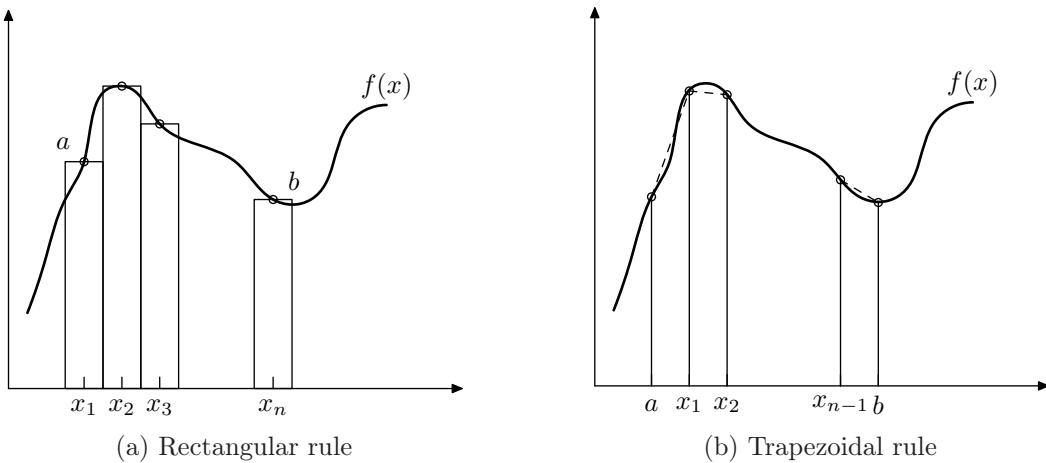


Figure 1.8: Visualisation of simple numerical integration methods

In both methods of numerical integration the accuracy is increased by making the step size h smaller.

These simple ideas of numerical differentiation and integration can be extended further to methods for solving partial differential equations in numerical form. The basic idea of advancing the evaluation process by a certain fixed step size is a common procedure in all fields of numerical analysis. As computers become more and more faster, the accuracy of numerical methods is approaching exact values as the step size can be reduced closer and closer towards zero.

1.2.4 Fourier analysis vs. an amplitude spectrum

When solving differential equations it often happens that a simple answer does not exist. Sometimes it is possible to derive an exact answer in the form of an

infinite series expansion. A sum of all the terms of the series gives an exact answer and partial sums of the series give approximations of different accuracy. There are several kinds of infinite series, and clearly the most common of them is called a *Fourier series*. The Fourier series is one of the most important mathematical tools for physicists and engineers because its relation to practical applications is so strong. By using the Fourier series it is possible to reconstruct *periodic* functions of almost any shape with sums of sinusoidal functions of different frequencies. In the frequency domain, the components of the Fourier series represent the *spectrum* of the reconstructed function or series of measurement points. This spectral representation is especially usable in applications concerning vibrations, where periodicity is almost always inevitable.

The theoretical background for the analysis was developed by Jean Baptiste Joseph Fourier in his publication 'The Analytical Theory of Heat' [7], where the idea of the series is derived in a clear and detailed manner from first principles. The result of Fourier's derivation was a trigonometric series that estimates the form of periodic signals by sums of sines and cosines, i.e. one can reconstruct the original signal by summing the components of the trigonometric series

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx. \quad (1.19)$$

The factors a_n and b_n are commonly known as *Fourier coefficients* and those can be physically interpreted as amplitudes of the trigonometric components of the series. The term a_0 is different from other coefficients, and is therefore taken outside of the general sum. The term a_0 depicts the constant offset of the whole function, and it can be interpreted also as the average value of the function $f(x)$.

The sum (1.19) builds up from sinusoidal functions of integer multiples of index n in frequency. As an example, Figure 1.9 presents two trigonometric functions and their sum function. Firstly, the sine function has the amplitude value 1 ($1 \cdot \sin$) and integer multiplicity $n = 1$ ($1 \cdot x$). Secondly, the cosine function has the amplitude value 2 ($2 \cdot \cos$) and integer multiplicity $n = 2$ ($2 \cdot x$). The period T_1 of the sine function is 2π ; the cosine function has period $T_2 = \pi$. Because the cosine function is the second harmonic frequency with index $n = 2$, it vibrates two full periods within the same time where sine function does only one full vibration sequence. The first harmonic frequency is commonly denoted as the *fundamental frequency* and the higher harmonic frequencies are *upper partials* of the fundamental frequency [5]. The term *harmonic components* or other similar expressions are also very often used to denote the upper partials.

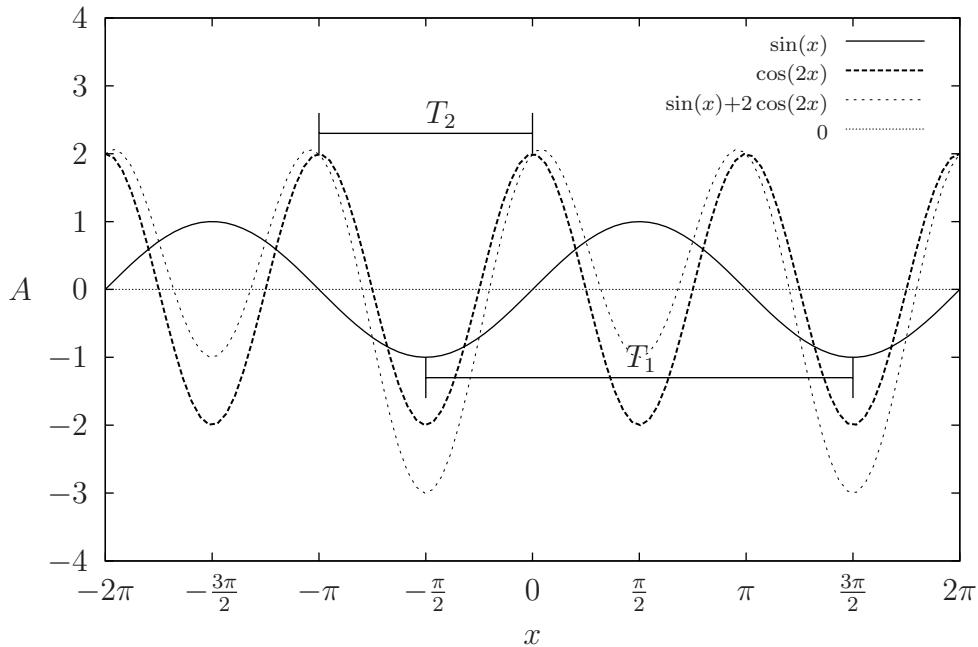


Figure 1.9: Examples of periodic trigonometric functions

The idea of the Fourier series is now decoded into the sum function of the sine and cosine functions. With these two simple sinusoids, it is possible to generate a slightly more complex periodic function, which repeats itself in the same period (T_1) as the largest period of the added sinusoids (fundamental frequency). Here Fourier's idea is constructed in a reverse engineering kind of way. The process of reconstructing an arbitrary signal from a sum of sinusoidal signals is commonly referred to as Fourier synthesis.

Normally there is just an arbitrary periodic waveform $f(x)$, from where the Fourier series representation is derived. To determine the Fourier series for some function $f(x)$, the period T of $f(x)$ and coefficients a_0, a_n and b_n need to be evaluated. If one sets $k = 1$ in equation (1.19), the period of the fundamental frequency will be simply 2π . In this case, the Fourier coefficients are calculated as integrals over one full period:

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx ; \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx ; \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx.$$

To scale the basic trigonometric period 2π in equation (1.19) so that it would be suitable for arbitrary periods $2L$, one needs to set $x = \frac{2\pi}{2L}x$ and the period

$2\pi \Rightarrow 2L$, so that

$$a_0 = \frac{1}{2L} \int_{-L}^L f(x) dx \quad (1.20)$$

$$a_n = \frac{1}{L} \int_{-L}^L f(x) \cos\left(n \frac{\pi}{L} x\right) dx \quad (1.21)$$

$$b_n = \frac{1}{L} \int_{-L}^L f(x) \sin\left(n \frac{\pi}{L} x\right) dx. \quad (1.22)$$

Clearly the calculation process of the Fourier coefficients resembles a correlation calculation of an arbitrary periodic function $f(x)$ with the trigonometric base functions sin and cos. If $f(x)$ is an even or odd function, then either the sin or cos terms will be zero, and these can be handled more efficiently as special cases using the *Fourier Cosine series* for even functions and the *Fourier Sine series* for odd functions. Equation (1.19) is the one-dimensional form of the series, but there also exist Fourier series representations for several dimensions. The process of calculating the series representation is called Fourier analysis and it is an opposite operation compared to the Fourier synthesis. [8, pp 103–104]

The sine and cosine components of the basic Fourier series (1.19) can be combined to a single sinusoid with a certain amplitude A and phase ϕ . By setting up an equation,

$$\begin{aligned} a_n \sin nx + b_n \cos nx &= A_n \sin(nx + \phi) \\ &= A_n (\sin nx \cos \phi + \cos nx \sin \phi) \\ &= (A_n \cos \phi) \sin nx + (A_n \sin \phi) \cos nx, \end{aligned}$$

one gets an equation pair,

$$a_n = (A_n \cos \phi) \quad \Rightarrow \quad a_n^2 = A_n^2 \cos^2 \phi \quad (1.23)$$

$$b_n = (A_n \sin \phi) \quad \Rightarrow \quad b_n^2 = A_n^2 \sin^2 \phi. \quad (1.24)$$

If the equation pair is summed together, the result

$$a_n^2 + b_n^2 = A_n^2 (\cos^2 \phi + \sin^2 \phi) \quad \Rightarrow \quad \boxed{A_n = \sqrt{a_n^2 + b_n^2}} \quad (1.25)$$

is obtained, indicating that the common amplitude for a single frequency component equals a squared summation over the individual amplitudes of sine and

cosine. The phase ϕ can be determined from the equation pair (1.23) by a division

$$\frac{b_n}{a_n} = \frac{A_n \sin \phi}{A_n \cos \phi} = \tan \phi \quad \Rightarrow \quad \boxed{\phi = \arctan \frac{b_n}{a_n}}. \quad (1.26)$$

Because of the properties of the tangent function, the phase angle resulting from equation (1.25) might need to be reduced to the first quadrant. With the results obtained in equations (1.25) and (1.26), it is possible to represent the total amplitude of each harmonic frequency f_n by one single amplitude factor A_n . This way of presenting the Fourier series leads to an amplitude spectrum representation of a signal.

In addition to the Fourier series, the concept of Fourier analysis also extends to cover Fourier integrals, which in turn lead to integral transforms. The Fourier integral transform is used to map a function of x to a domain of the complex frequency $i\omega$ and leads to a spectrum described by a function of continuous frequency instead of the discrete set of the harmonic frequencies of the Fourier series. The Fourier integrals do not require the analysed function to be periodic, but instead the period L needs to cover the whole x -axis, i.e. $L = \infty$. Because of their analytical nature, the Fourier integrals are primarily used as integral transform pairs to solve differential equations and only secondly for spectrum analysis. The analytic Fourier transform is not very suitable for practical applications of spectrum analysis. In some very rare cases the Fourier transform might be used as a special case of the Laplace transform to calculate transfer functions of electric circuits.

The basic Fourier series (1.19) is only applicable to functions of continuous time, but spectrum analysis is more often needed for measured signals. A measured signal consist of sequential samples of the original continuous function, and the time scale of the sequence of samples is defined by the frequency of taking the samples. The counterpart of the Fourier series for discrete time functions is the Discrete Fourier Transform (DFT). Similarly to the basic Fourier series analysis, the DFT leads to a set of sines and cosines of harmonic frequencies, where the Fourier coefficients multiplying the trigonometric base functions represent the amplitude spectrum of the analysed signal.

By definition, the Discrete Fourier Transform coefficient X_k of data series x_n is calculated as

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi k \frac{n}{N}}, \quad (k = 0, 1, \dots, N - 1) \quad (1.27)$$

where N is the total amount of data points within one period of the sampled data sequence x_n . In a general case, both X_k and x_n are complex numbers. The index k indicates the ordinal number of the frequency component so that $k = 0$ is the average value and $k = 1$ is the fundamental frequency. The period T_1 of the fundamental frequency is defined by N in connection with the sampling frequency f_S of x_n so that

$$T_1 = \frac{N}{f_S}.$$

The DFT therefore assumes periodicity of x_n as N samples. If the DFT is defined as in equation (1.27), the physical interpretation of the coefficients X_k is not directly the amplitude of the frequency component k . The details of coefficients X_k are derived after some more background information has been covered.

The expression that uses the imaginary exponential is not very practical if one wants to understand the true nature of the DFT. Based on the Euler identities,

$$\begin{aligned} e^{inx} &= \cos nx + i \sin nx \\ e^{-inx} &= \cos nx - i \sin nx, \end{aligned} \tag{1.28}$$

the imaginary exponent is depicted as a periodic sum of sines and cosines in the complex plane, and the magnitude of equations (1.28) is rotating on the unit circle of the complex plane. This unit circle is illustrated in Figure 1.10.

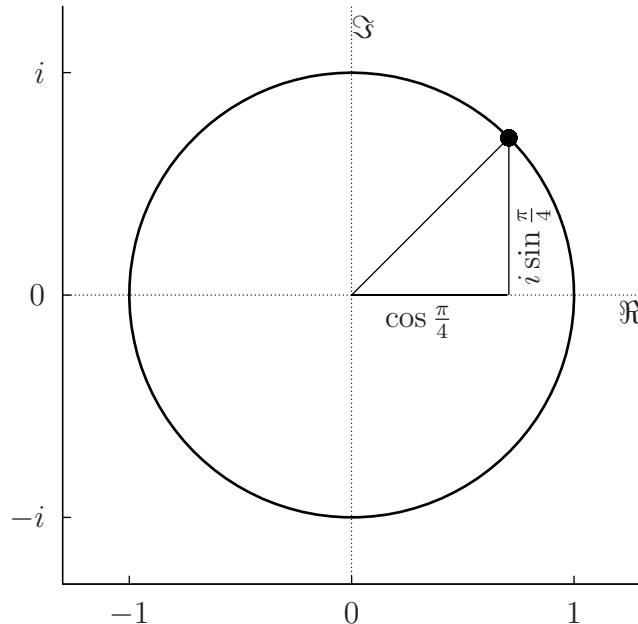


Figure 1.10: The unit circle in the complex plane

With the help of the Euler equations (1.28), the DFT can be written as

$$X_k = \sum_{n=0}^{N-1} x_n \cdot \left[\cos\left(2\pi k \frac{n}{N}\right) - i \sin\left(2\pi k \frac{n}{N}\right) \right], \quad (1.29)$$

and (1.29) can be separated into the real and imaginary components of

$$\begin{aligned} \Re(X_k) &= \sum_{n=0}^{N-1} x_n \cdot \cos\left(2\pi k \frac{n}{N}\right) \\ \Im(X_k) &= \sum_{n=0}^{N-1} -i \cdot x_n \cdot \sin\left(2\pi k \frac{n}{N}\right). \end{aligned} \quad (1.30)$$

From equation (1.29) one can see that when the index n goes through N sampled values, the measured values x_n are correlated with a complex sinusoidal function of the fundamental frequency and all the k numbers of its harmonic frequencies. The range of values of the integer variable $k = 0, 1, \dots, N - 1$.

The basic idea behind the DFT is therefore to define a correlation factor between the data series values x_n and sinusoidal harmonic frequencies f_k , which are integer multiples of the fundamental frequency f_1 . The fundamental frequency is determined by the sample size N , which represents the amount of samples in one full period of x_n . The DFT assumes that x_n repeats itself in N sample intervals. The possible phase difference is covered by calculating the complex coefficients X_k for both the sine and cosine correlation and the result is a squared sum of these two components. This seems to be equivalent to the process of determining the coefficients of the Fourier series for continuous functions!

To cover the Fourier synthesis part, it is possible to recover the original time domain data sequence from a set of Fourier transformed values using the Inverse Discrete Fourier Transform. The IDFT is defined as:

$$x_k = \frac{1}{N} \sum_{n=0}^{N-1} X_n \cdot e^{i2\pi k \frac{n}{N}}, \quad (1.31)$$

and from here one can notice that the Fourier transformed coefficients X_n in the inverse transform relate to the amplitudes of the complex form of the continuous Fourier series (1.19). The complex form of the Fourier series is derived using the Euler identities (1.28)

$$e^{inx} + e^{-inx} = 2 \cdot \cos nx \Rightarrow \cos nx = \frac{1}{2} (e^{inx} + e^{-inx}) \quad (1.32)$$

$$e^{inx} - e^{-inx} = 2 \cdot i \sin nx \Rightarrow \sin nx = \frac{1}{2i} (e^{inx} - e^{-inx}), \quad (1.33)$$

and a substitution of these equalities into the trigonometric Fourier series (1.19) yields

$$\begin{aligned}
 f(x) &= a_0 + \sum_{n=1}^{\infty} \left(\frac{1}{2} a_n \cdot [e^{inx} + e^{-inx}] - \frac{1}{2} i b_n \cdot [e^{inx} - e^{-inx}] \right) \\
 &= a_0 + \sum_{n=1}^{\infty} \left(\frac{1}{2} [a_n - i b_n] \cdot e^{inx} + \frac{1}{2} [a_n + i b_n] \cdot e^{-inx} \right) \\
 &= \sum_{n=-\infty}^{\infty} c_n \cdot e^{inx}.
 \end{aligned} \tag{1.34}$$

The last form in equation (1.34) is explained with the notations

$$\begin{aligned}
 c_n &= \frac{1}{2} (a_n - i b_n) (= X_n) \\
 c_{-n} &= \frac{1}{2} (a_n + i b_n) (= X_{-n}) \\
 c_0 &= a_0 = X_0.
 \end{aligned} \tag{1.35}$$

This gives a relation to the amplitudes of the complex (c_n) and real (a_n, b_n) forms of the Fourier series. The equality of c_n in equation (1.34) and X_n in equation (1.31) is obvious, and the relation of X_n to the real amplitudes a_n and b_n is the same as for c_n .

The notations c_n and c_{-n} in (1.35) mean that in the complex representation of the Fourier series, the amplitude components of the real-valued Fourier series have been split into two parts between the positive $i\omega$ and the negative $-i\omega$ frequencies. What are the negative frequencies in the case of X_n then, since the summation starts from 0 in equations (1.27) and (1.31)?

The theory for the DFT coefficients says that

$$X_k = X_{N-k}^*, \tag{1.36}$$

where $k = 0, 1, \dots, \frac{N}{2} - 1$ if N is even and $k = 0, 1, \dots, \frac{N-1}{2}$ if N is odd, and

$$X_k = X_{N+k}, \tag{1.37}$$

on the basis of periodicity. A substitution of the periodicity condition $k = N + k$ from equation (1.37) into the imaginary component in equation (1.30) gives

$$\Im(X_k) = \sum_{n=0}^{N-1} -i \cdot x_n \cdot \sin \left(2\pi n + 2\pi k \frac{n}{N} \right) = \sum_{n=0}^{N-1} -i \cdot x_n \cdot \sin \left(2\pi k \frac{n}{N} \right)$$

because of the rule $\sin(x + y) = \sin x \cos y + \cos x \sin y$. This obviously keeps the reference sine wave rotating to the positive direction. If x_n is a sine wave of the same frequency and phase, the correlation value of the two sine waves is positive and this will result in a negative imaginary DFT component $-iX_k$. This seems to match with $a_k - ib_k$.

A substitution of the complex conjugate condition $k = N - n$ from (1.36) into the imaginary component in equation (1.30) gives

$$\Im(X_k) = \sum_{n=0}^{N-1} -i \cdot x_n \cdot \sin\left(2\pi n - 2\pi k \frac{n}{N}\right) = \sum_{n=0}^{N-1} i \cdot x_n \cdot \sin\left(2\pi k \frac{n}{N}\right),$$

and this also keeps the reference sine wave rotating to the positive direction, but changes the sign in front of the sine function. By the same argument as above, this will result in a positive imaginary DFT component $+iX_k$ that seems to match with $a_k + ib_k$.

This weak argument is trying to prove that

$$c_n = X_n \quad \text{and} \quad c_{-n} = X_{N-n},$$

where n extends to a range $[0, N/2]$ as already defined more robustly for the even and odd values of N separately. The amplitudes of the positive frequency components are obtained in a growing order corresponding directly to the subindex until the value of the subindex reaches the limit $\frac{N}{2}$. After this, the DFT splits out the negative frequencies in reversed order. This is visualised by plotting the amplitude spectrum from the waveform presented in Figure 1.11. The sampled waveform, drawn with a solid line in Figure 1.11, is constructed from two sine waves $f(x)$ and $g(x)$, which are drawn with dashed lines to the same plot. The analytic expressions for the single sine waves are

$$f(x) = 2.5 \sin(x) - 0.25 \quad \text{and} \quad g(x) = 1.25 \sin\left(2x - \frac{\pi}{4}\right),$$

and the direct sum of these waves is sampled equidistantly at intervals of $\frac{\pi}{4}$ to create the input data sequence x_n for the DFT calculation.

Using equation (1.27), the data sequence x_n with $N = 16$ sampled data points is transformed into a set of Fourier coefficients, resulting in 16 complex values X_k . The previous theoretical discussion suggests that X_0 is the average value, or offset, of the data sequence x_n , $X_1 \dots X_7$ are the complex amplitudes X_+ of the positive frequencies and $X_9 \dots X_{15}$ are the negative frequency components with complex conjugate amplitudes $X_- = X_+^*$. The components $X_9 \dots X_{15}$ are more

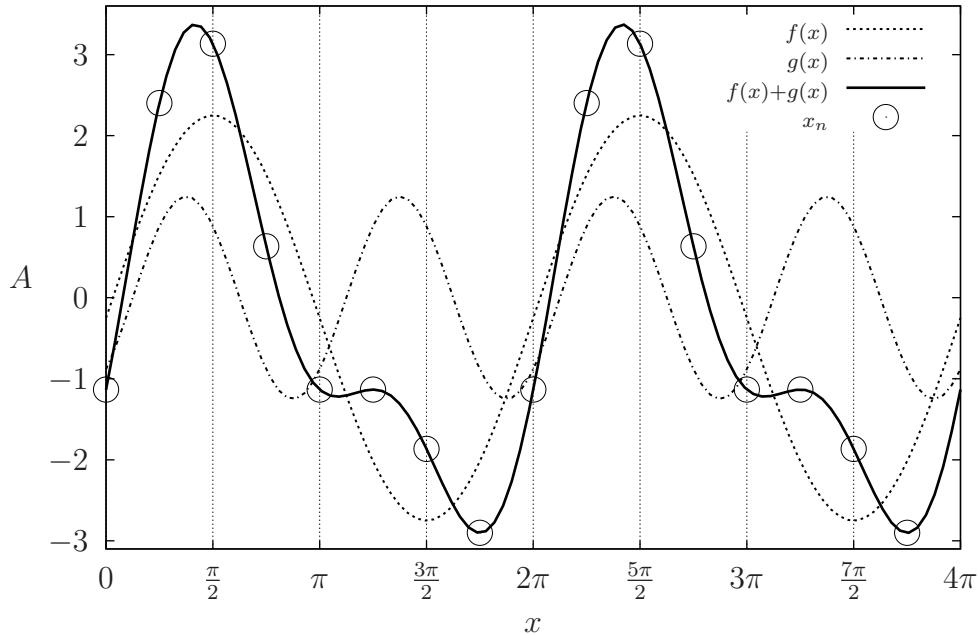


Figure 1.11: Sampled sum of two sine waves

intuitively referenced as $X_{N-7} \dots X_{N-1}$ to emphasise these as negative frequency components with an orderly connection to the positive frequency components $X_7 \dots X_1$.

Figure 1.12 illustrates the spectrum representation of these 16 amplitude coefficients X_k . The absolute value of the coefficients has been used to make X_k real-valued $|X_k|$, and all the values $|X_k|$ have been divided by N to reach the amplitude interpretation. In addition, the components have been divided into groups of positive and negative frequencies around the X_0 component to emphasise mirroring. As all X_k are plotted as magnitudes $|X_k|$, the nasty side effect is that the real sign of X_0 is lost. In reality, the DFT had given $\frac{X_0}{N}$ a value -0.25 , which is exactly the same value what was given to $f(x)$ as offset. Generally Figure 1.12 shows that with the periodicity of $N = 16$ samples, a total of six upper partial coefficients are obtained in addition to the fundamental component X_1 . The eighth coefficient is shared between the negative and the positive frequencies, since it is exactly at the boundary point of these frequencies. If the total number of samples would have been an odd number, then there would not be any shared components.

The magnitudes of the amplitude coefficients are split in half between the positive and negative frequency components. In practical analysis, when the input samples x_n are always real-valued, it is common to calculate the DFT for $N/2$

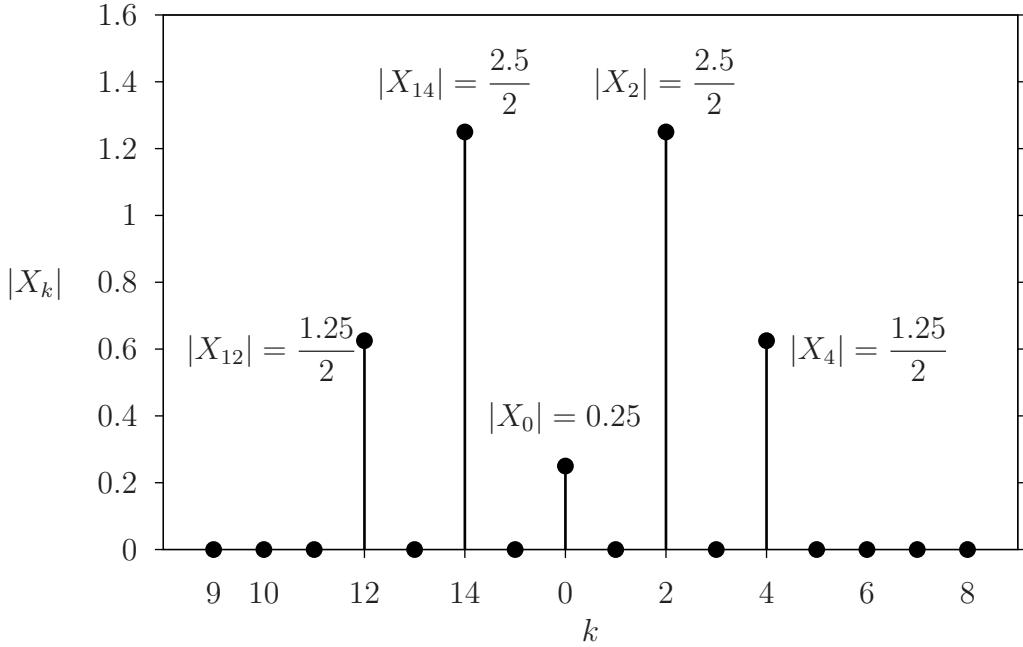


Figure 1.12: The DFT coefficients for the sampled sum of sines

components and multiply the magnitudes of the amplitude coefficients by 2. The negative frequency components have no physical significance in real world applications and they can be skipped when calculating the DFT. This leads to the conclusion that if and only if the input data sequence x_n is real, then the true amplitude A_k of a particular frequency component k can be calculated using the formula

$$A_k = 2 \cdot \sqrt{\Re(X_k)^2 + \Im(X_k)^2} = \sqrt{a_k^2 + b_k^2}, \quad (1.38)$$

where $k = 0, 1, \dots, N/2$, X_k are the complex coefficients resulting from the DFT and a_k, b_k are the amplitudes of the cosine and sine functions in the trigonometric Fourier series (1.19). This result can also be concluded from equation (1.35). It should be mentioned that in a general case where the input signal can have complex values, the amplitudes of the negative frequencies should be summed with the positive frequencies to obtain the true spectrum of the signal.

The phase information of the frequency components is also included in the complex amplitude coefficients X_k . The general theories of complex numbers define that the phase

$$\phi_k = \arctan \frac{\Im(X_k)}{\Re(X_k)} \quad (1.39)$$

for a certain harmonic frequency of order k .

The relation between the real and imaginary Fourier series states that the DFT

results in amplitude components which are half in magnitude compared to the real amplitudes of the trigonometric frequency components. This can also be analysed from another point of view. Suppose that we have only the fundamental frequency of a pure sine wave with amplitude A in our system to be measured. Figure 1.13 depicts one full period of this equidistantly sampled sine wave.

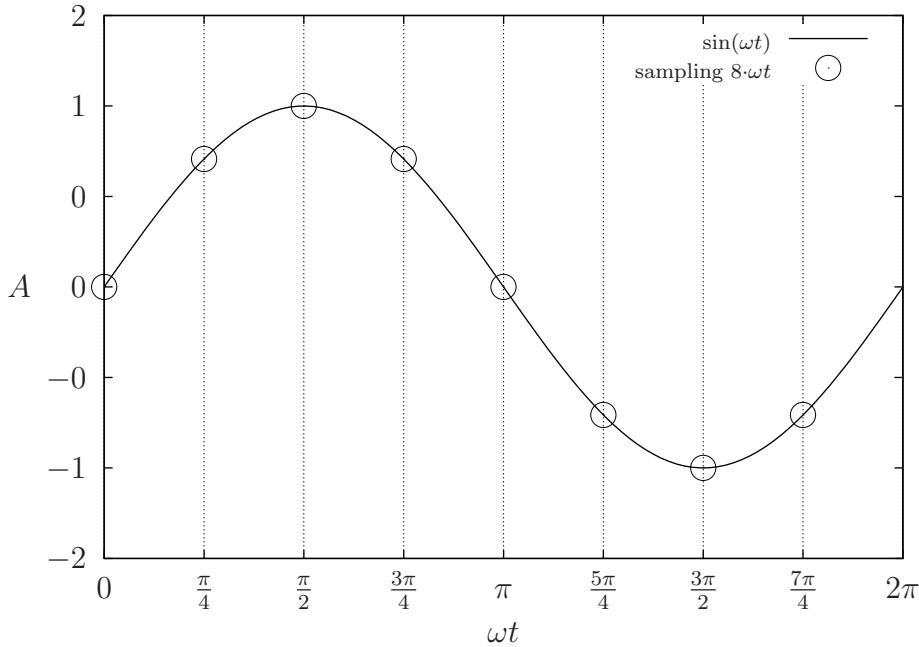


Figure 1.13: Equidistant sampling of a sine wave

Then the DFT is applied for the fundamental frequency of the signal of Figure 1.13 leading to a derivation sequence:

$$\begin{aligned}
 \sum_{n=0}^{N-1} A \sin\left(2\pi \frac{n}{N}\right) \sin\left(2\pi \frac{n}{N}\right) &= A \sum_{n=0}^{N-1} \left[\sin\left(2\pi \frac{n}{N}\right) \right]^2 = \\
 A \sum_{n=0}^{N-1} \left[\frac{1}{2i} \left(e^{i2\pi \frac{n}{N}} - e^{-i2\pi \frac{n}{N}} \right) \right]^2 &= A \sum_{n=0}^{N-1} -\frac{1}{4} \left(e^{i4\pi \frac{n}{N}} - 2 + e^{-i4\pi \frac{n}{N}} \right) = \\
 A \sum_{n=0}^{N-1} \frac{1}{2} - \frac{1}{2} \cos\left(4\pi \frac{n}{N}\right) &= \frac{A}{2}N - \sum_{n=0}^{N-1} \frac{A}{2} \cos\left(4\pi \frac{n}{N}\right) = \boxed{\frac{A}{2}N}. \quad (1.40)
 \end{aligned}$$

The result is N times the half value of the real amplitude of the sampled sine wave. The cosine term in the equation is always zero when the sampling frequency is an *exact multiple* of the measured sine wave frequency. This way one can get the original amplitude of the signal by dividing it by N and multiplying

it with 2. In the more thorough formulation of DFT analysis, this can be thought of as using a rectangular window function w_n with amplitude 1 to multiply the original signal. A brief discussion about window functions follows later.

In basic DFT analysis, it is extremely important to fill the requirement for the relation

$$\frac{f_S}{f_n} = k_n \quad (1.41)$$

between the sampling frequency f_S and the measured signal frequency f , where k_n is an integer number and the subscript n refers to the upper partial, i.e. $n = 1$ refers to the fundamental frequency of a periodic signal.

If the requirement for relation (1.41) is not fulfilled, spectral leaking will be added to the analysis results because of asymmetric sampling or a wrong truncation point of the measured signal. This phenomenon deserves to be treated with care, since it is quite typical to have erroneous results because of spectral leaking in DFT analysis. Figure 1.14 describes a situation where the analysis frame is deliberately chosen to be larger than the actual period of the sinusoid.

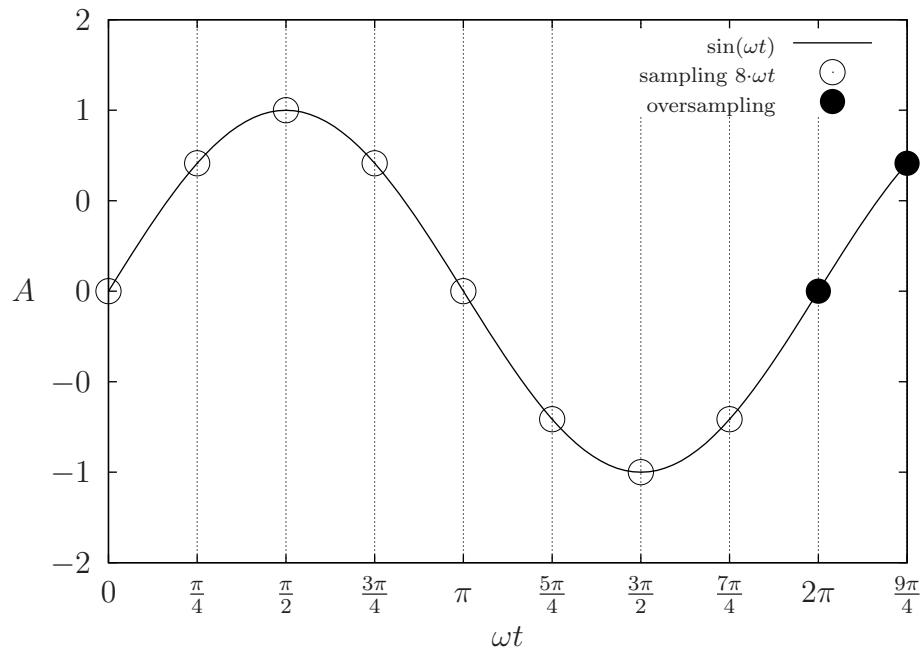


Figure 1.14: Nonperiodic truncation when sampling a sine wave

When the sampling frequency is not an exact multiple of the analysed signal, the DFT analysis still gives a good approximation of the amplitude of the fundamental frequency, but some of the magnitude gets chopped off to the other harmonic frequencies. This is not that clear at first thought, but below is an attempt to

prove this effect in a calculation where the sampled sine wave is transformed with respect to the imaginary sine component of the DFT.

$$\begin{aligned} \frac{1}{N+\epsilon} \sum_{n=0}^{(N+\epsilon)-1} \sin\left(2\pi \frac{n}{N}\right) \cdot \sin\left(2\pi \frac{n}{N+\epsilon}\right) = \\ \frac{1}{N+\epsilon} \sum_{n=0}^{(N+\epsilon)-1} -\frac{1}{4} \left(e^{i2\pi \frac{n}{N}} - e^{-i2\pi \frac{n}{N}} \right) \left(e^{i2\pi \frac{n}{N+\epsilon}} - e^{-i2\pi \frac{n}{N+\epsilon}} \right) = \\ \frac{1}{N+\epsilon} \sum_{n=0}^{(N+\epsilon)-1} \frac{1}{2} \left[\cos\left[\frac{2\pi n}{N} \left(1 - \frac{1}{1 + \frac{\epsilon}{N}}\right)\right] - \cos\left[\frac{2\pi n}{N} \left(1 + \frac{1}{1 + \frac{\epsilon}{N}}\right)\right] \right] \end{aligned} \quad (1.42)$$

When $N \gg \epsilon$, this equation reduces to the equivalent of (1.40). In this case the cosine correlation also gives some contribution to the result, so the calculation needs to be repeated for the cosine component as well. The calculation is similar as in equation (1.40), so it is not repeated here. From these calculations one can derive a formula to estimate the error \mathcal{E}_k in the amplitudes of the frequency components due to the unharmonic truncation point of the analysed data series. Only the result of a very long calculation is shown here.

$$\mathcal{E}_k = \frac{\sqrt{2}}{N+\epsilon} \sqrt{\sum_{m=0}^{(N+\epsilon)-1} \sum_{n=0}^{(N+\epsilon)-1} \cos\left(k \cdot 2\pi \frac{n-m}{N+\epsilon}\right) \left[\cos\left(\frac{N}{T} \cdot 2\pi \frac{m-n}{N}\right) - \cos\left(\frac{N}{T} \cdot 2\pi \frac{m+n}{N}\right) \right]}$$

This equation results in a relative error term for the fundamental ($k = 1$) frequency. For other harmonic components of index k this gives a leakage addition term that is chopped off from the amplitude of the fundamental frequency. When spectral leakage happens, the amplitude of the fundamental is decreased and the amplitudes of the following harmonic frequencies are increased. The multiplier T in the equation refers to the number of periods of the fundamental frequency taken inside the frame of analysis. The derived equation is just a theoretical approximation formula for simple sinusoids, but it depicts the idea behind the possible error coming from the DFT calculations. It should not be used for official error estimation of DFT results.

If only simple periodic sinusoidal signals are analysed, it is more than sufficient to use basic DFT for spectrum reconstruction. Especially in this book there is no need to go further into Fourier theorems than already presented. There has still been some personal interest in getting a more advanced level of understanding

of the DFT, especially knowledge about the Fast Fourier Transform and the effects of window functions.

Usually Fourier transforms are used to reveal the *relative* amplitude differences between the fundamental frequency and the harmonic overtones of periodic signals. It would still be nice to be able to use the DFT as a measurement instrument, to reveal the true amplitudes of each harmonic in the analysed signal. This was proven to be quite simple as a result of calculations made in (1.40) when using periodic signals and a rectangular window function, which simply scales every sampled point of the signal with unity.

To go one step further, the following question is: what if one uses some other window function to weigh the measured sine wave? Let us name the window function as w_n , which has the same N number of data points as our sampled sine wave does. Then, as calculated in (1.40):

$$\begin{aligned} \sum_{n=0}^{N-1} w_n A \sin\left(2\pi \frac{n}{N}\right) \sin\left(2\pi \frac{n}{N}\right) &= A \sum_{n=0}^{N-1} w_n \left[\sin\left(2\pi \frac{n}{N}\right) \right]^2 = \\ A \sum_{n=0}^{N-1} w_n \left[\frac{1}{2i} \left(e^{i2\pi \frac{n}{N}} - e^{-i2\pi \frac{n}{N}} \right) \right]^2 &= A \sum_{n=0}^{N-1} -w_n \frac{1}{4} \left(e^{i4\pi \frac{n}{N}} - 2 + e^{-i4\pi \frac{n}{N}} \right) = \\ A \sum_{n=0}^{N-1} \frac{w_n}{2} - \frac{w_n}{2} \cos\left(4\pi \frac{n}{N}\right) &= \boxed{\sum_{n=0}^{N-1} \frac{Aw_n}{2} - \sum_{n=0}^{N-1} \frac{Aw_n}{2} \cos\left(4\pi \frac{n}{N}\right)}. \quad (1.43) \end{aligned}$$

Now there is a situation where it is not possible to state that the cosine term would always be zero. In fact, in this case it is never zero. From the transform against the cosine component of the DFT there will be a sine term similar to the cosine term to make things a bit harder. The leading term is still similar to what was found in equation (1.40), but now the term N has changed to the sum of the discrete values of the window function.

The scaling factor to normalise the amplitude of windowed Fourier transforms is not clear from this context. More accurate analysis needs to be made numerically with the help of mathematical software. Numerical analysis of the scaling factors from the most common window functions shows that quite often the effects of the cosine and sine terms are very small compared to the leading term. It is therefore common to deduce that the scaling factors of window functions are calculated by adding up all the terms in the discrete window function. This scaling factor is more commonly known as a Coherent Scaling Factor, and these factors are

tabulated for the most common window functions.

Window functions are used to reduce the effect of spectral leaking in situations where there is no possibility to determine the periodicity of a signal. The most common window functions are all symmetric and they have a damping effect on both end points of the analysed signal. Since the DFT assumes continuity at the end points of the signal (periodicity), by using special windowing the effect of discontinuity at the end points can be smoothed out a bit. The window functions aim their full weighting factors at the middle of the analysed signal.

If real-valued input signals are used, then to obtain the real amplitude from a window-weighted DFT calculation, one needs to divide the magnitude $|X_k|$ with the sum of the values of the window function and multiply it with 2. Here we only analysed the simplest possible situation, but it should work somehow for more complicated systems as well. At least this way one gets a good *estimate* for the actual amplitudes of the frequency components of the measured system.

Sometimes there is more interest in the power spectrum instead of the amplitude spectrum. For the power spectrum one just squares the amplitude, such as $|X_k|^2$, or multiplies the complex conjugates $X_k \cdot X_k^*$. The second method is more analytic and a more correct way to obtain the power spectrum. The assumption that power is directly related to the square of the amplitude is not a very robust definition, at least in the sense of exact values of power.

The Fast Fourier Transform (FFT) is a special case of the DFT. It is a kind of a short-cut way to calculate the transform, and as its name suggests, it is a much faster method to calculate the FFT. There are of course restrictions, since the calculational method of the FFT relies on symmetries. Therefore, the amount of samples taken to the FFT calculation must be a power of 2 (e.g. 2^{12}). This was the requirement for the traditional FFT, but nowadays there are FFT methods for other amounts of samples as well.

Zero padding is a quite commonly used method when using the FFT or when extending the range of the DFT calculation to interpolate the amplitudes of the frequency components that are not integer multiples of the fundamental frequency determined by the truncation point of the data series. Zero padding is unfortunately commonly misused, because the amount of zeros in the padding should again be an integer multiple of the fundamental period of the data sequence. If it is not, spectral leakage or distortion will appear in the analysis results.

1.2.5 The concept of a decibel vs. an amplitude spectrum

The decibel is a common term when spectrums and frequency response graphs are concerned. Actually *bel* [B] is a unit of a physical quantity that measures logarithmic power ratio with respect to some reference power value. In a mathematical formulation bel is defined as

$$\log_{10} \left(\frac{P}{P_0} \right) \text{B}. \quad (1.44)$$

In practical applications bel behaves like farads or henrys. It is a quite large unit and therefore the most commonly used unit is the decibel, which is derived from bel as follows:

$$\log_{10} \left(\frac{P}{P_0} \right) \text{B} = \log_{10} \left(\frac{P}{P_0} \right) \frac{1 \text{ B}}{10 \text{ dB}} 10 \text{ dB} = 10 \log_{10} \left(\frac{P}{P_0} \right) \text{dB}. \quad (1.45)$$

There are two ways of using the decibel as a measure of the power ratio: the absolute measure and the relative measure. In the absolute measure, the reference power value is some standard unit of power such as 1 mW, and then the decibels are officially abbreviated as dBm. With similar reasoning the abbreviation for 1 μ W is dB μ . As an example for 4 dBm, the absolute power value P can be retrieved as

$$4 \text{ dBm} = 10 \log_{10} \left(\frac{P}{1 \text{ mW}} \right) \text{ dBm}$$

$$10^{\frac{4}{10}} = \frac{P}{1 \text{ mW}}$$

$$P = 10^{\frac{4}{10}} \text{ mW}.$$

Please note that the notation of dBm is *only a notation!* This way the absolute values are distinguished from the relative values.

In the relative measure the absolute value of the reference power is not known. This type of measure is commonly used in amplifier specifications, where the gain of the amplifier is measured as the relative power ratio of the output and input signals (output in the numerator, input in the denominator). In this case, 0 dB means that the output and input signals have the same power, +3 dB means that the output power is two times the input power, +6 dB means that the output power is four times the input power, +9 dB means that the output power is eight times the input power and -3 dB means that the output power is half of the

input power. These results can be directly calculated from equation (1.45), e.g.

$$3 \text{ dB} = 10 \log_{10} \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right) \text{ dB}$$

.....

$$P_{\text{out}} = 10^{\frac{3}{10}} P_{\text{in}} \approx 2P_{\text{in}}.$$

The most important thing to understand here is that decibels always depict power quantities. Every time when one sees a graph using decibels as a measure of voltage amplitude, it is actually measuring power.

It is anyhow possible to retrieve voltage or pressure amplitudes from decibels via the relation

$$P = \frac{V^2}{R},$$

where V refers to the voltage magnitude and R is resistance. According to the mathematical laws of logarithms, the exponent is taken down as a multiplier so that the decibel-formatted voltage relation is written as

$$10 \log_{10} \left(\frac{P}{P_0} \right) \text{ dB} = 10 \log_{10} \left(\frac{V}{V_0} \right)^2 \text{ dB} = 20 \log_{10} \left(\frac{V}{V_0} \right) \text{ dB.} \quad (1.46)$$

Because of the factor of 2 in the voltage conversion, +6 dB means that the output voltage is two times the input voltage, and so on.

But wait a minute ... Something is not right here ...? The resistances R cancelled each other out from the voltage equation, meaning that the resistance between the measurement points of V and V_0 should be the same! From this it follows that if some voltage ratio is first converted to decibels and then this decibel value is converted to a power ratio, the only case that gives correct powers is when the resistances of the input and output are the same. From this it follows that a strict way to define the power ratio from voltages is

$$\frac{P}{P_0} = \frac{V^2}{V_0^2} \frac{R_0}{R},$$

and the corresponding conversion to decibels is written as

$$10 \log_{10} \left(\frac{P}{P_0} \right) \text{ dB} = 20 \log_{10} \left(\frac{V}{V_0} \right) \text{ dB} + 10 \log_{10} \left(\frac{R_0}{R} \right) \text{ dB.}$$

The rules above are often broken in electronics, where the transfer functions showing the relative gain are derived directly from the amplitude ratio of the

input and output voltages, neglecting the impedances. This is possibly just a common agreement to misuse the decibels when plotting transfer functions, but it is definitely against the definition of the decibel. Unfortunately this misuse will be continued in this book as well.

For voltages there also exists an absolute decibel measure, where the reference voltage is 1 V. The decibels using this reference voltage are officially abbreviated as dBV. As an example for 4 dBV, the true voltage value V_x can be retrieved as

$$4 \text{ dBV} = 20 \log_{10} \left(\frac{V_x}{1 \text{ V}} \right) \text{ dBV}$$

$$10^{\frac{4}{20}} = \frac{V_x}{1 \text{ V}}$$

$$V_x = 10^{\frac{4}{20}} \text{ V.}$$

Additionally, it is a common habit that when calculating the power level in dBm, the reference voltage amplitude V_0 is set to 0.77459 V and R_0 to 600 Ω. The terms V_0 and R_0 can be chosen freely, as long as they result in a power level of 0.001 W. [9, pp. 33-37]

The acoustical equivalent of voltage is pressure p , and in the acoustical world there exists a similar relation,

$$I = \frac{p^2}{S},$$

where I = intensity and S = area where the pressure p is affecting. Decibels are commonly used as a measure of intensity as well.

Another unit of measure rarely used instead of the decibel is the *neper* [Np]. It is defined as a natural logarithm of a voltage ratio

$$\frac{V}{V_0} = e^x,$$

and after taking a natural logarithm from both sides,

$$\ln \left(\frac{V}{V_0} \right) \text{ Np} = x \ln e \text{ Np} = x \text{ Np.}$$

The relationship between the neper and the decibel is

$$10 \log_{10} \left(\frac{P}{P_0} \right) \text{ dB} = 10 \log_{10} e^2 \text{ Np} = 8.686 \frac{\text{dB}}{\text{Np}}.$$

This correlation between nepers and decibels is good to know, although this piece of information is not needed very often.

1.2.6 The Laplace transform vs. a transfer function

Both the Laplace and Fourier transform belong to a widespread group of integral transforms. These mathematical transforms have the general form of

$$u(z) = \int_C K(z, t)v(t) dt,$$

where $K(z, t)$ is called a *kernel* of a certain transform, and the integral is evaluated around some contour C . [10, p. 433]

Integral transforms are used to map a function of some real variable, like time t for instance, to another domain of some other variable, which is linked to the original variable via the transform used. After transforming the domain of a variable, it is possible to return to use the variable in the original domain via an inverse transform. Inverse operations in mathematics are typically more complex than 'direct' operations. As an example, integration is an inverse operation of differentiation, since it returns the differentiated function to its original form before differentiation.

Mathematical analysis of electric circuits leads to a set of differential equations that are often difficult to solve. The analysis of circuits can be made simpler by mapping the time dependent variables to a domain where the differential equations are handled as basic linear equations. This is why the Laplace transform is frequently used in circuit analysis to solve a group of differential equations without integration. The idea of the Laplace transform is to change a differential equation into a polynomial function $P(s)$, where the powers of s relate to the order of the substituted differential. After transformation, mathematical operations can be made in the s -domain, and when the answer is found in the s -domain, the expression is transformed back (if necessary) to the time domain via the inverse transform.

The Laplace transform of a function $f(t)$ is defined as

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty e^{-st} f(t) dt, \quad (1.47)$$

where the factor e^{-st} is identified as the kernel $K(s, t)$ of the Laplace transform [11, p. 221]. The limits of the transform can be extended to cover the interval $[-\infty, \infty]$, but normally it is unnecessary since the variable of integration is normally time t and negative time is an obscure concept.

When the integration in equation (1.47) starts from zero, this automatically gives

the differential equation initial condition(s) at time $t = 0$. If the initial conditions need to be given at some other time than $t = 0$, a simple change of variables $t = t' + t_0$ is needed so that $t = t_0$ gives $t' = 0$. This follows directly from the lower limit of the Laplace transform integral (1.47).

Despite presenting the general equation (1.47) from where all the transforms can be calculated, in practise most of the transforms are done directly from tables of precalculated transform functions. These tables of Laplace transforms can be found from any book that covers engineering mathematics because these transforms are so often used in practical applications in all fields of technology. The most common Laplace transform pairs are listed in appendix E.

The Laplace variable s is a variable of a complex type, where $s = \sigma + j\omega$. In connection with the exponential function

$$e^{-(\sigma+j\omega)t} = e^{-\sigma t} [\cos(\omega t) - j \sin(\omega t)],$$

according to the properties of the complex exponential function. This suggests that σ refers to the damping coefficient and ω to the angular frequency coefficient in the basic solution formulae of differential equations.

In the context of differential equations, the Laplace transform is used to establish a so-called *transfer function* $H(s)$, which is just a Laplace s -domain equivalent of the original time domain differential equation. For a linear differential equation

$$H(s) = \frac{Y(s)}{X(s)},$$

where $Y(s)$ is the Laplace transform of the output function $y(t)$, and $X(s)$ is the transform of the input function $x(t)$. Often $y(t)$ is referred to as a response function and $x(t)$ as a driving function. [12, p. 60]

The Laplace transform is used primarily in two different ways in the design process of electrical circuits. The first way is to use the transform with the inverse transform to find a solution to the differential equation of a circuit in the time domain. This is called *transient analysis*, and it shows the real output waveform $y(t)$ of the circuit with respect to some input function $x(t)$.

Secondly, it is tremendously popular to examine the frequency response of electric circuits (and also other systems) directly using the Laplace transform function of the circuit, without the need for an inverse transform. The mathematical foundations for this method are inaccurately explained in many contexts. This is why a detailed walkthrough is in order.

The most common goal in the field of electronics is to find out the output of an electric circuit when the input signal is a sine wave of the form

$$x(t) = A_{\text{in}} \sin(\omega t), \quad (1.48)$$

where A_{in} is the amplitude of the sinusoidal input signal and ω refers to the angular frequency $2\pi f$. Assuming that the transfer function $H(s)$ can be expressed as a ratio of general polynomials,

$$H(s) = \frac{p(s)}{q(s)} = \frac{p(s)}{(s + s_1)(s + s_2) \cdots (s + s_n)},$$

which usually is the case, the Laplace transform of the output function $y(t)$ is

$$Y(s) = H(s)X(s) = \frac{p(s)}{q(s)}X(s) = \frac{p(s)}{q(s)} \frac{\omega A_{\text{in}}}{s^2 + \omega^2},$$

where $X(s)$ is the Laplace transform of the input function $x(t)$ given in equation (1.48). The transform has been made using the transform pair of function $\sin(\omega)$ from the table of Laplace transforms in Appendix E.

Using a partial fraction expansion,

$$Y(s) = \frac{a}{s + j\omega} + \frac{\bar{a}}{s - j\omega} + \frac{b_1}{s + s_1} + \frac{b_2}{s + s_2} + \cdots + \frac{b_n}{s + s_n}, \quad (1.49)$$

where a , \bar{a} and b_i are undetermined constants. The complex conjugate pair of a follows from the input sine wave, and the b_i terms appear when the function $q(s)$ is written open. Using the standard table of Laplace transforms for inverse transform of equation (1.49), the inverse transform of $Y(s)$ has a general form

$$y(t) = ae^{-j\omega t} + \bar{a}e^{j\omega t} + b_1e^{-s_1t} + b_2e^{-s_2t} + \cdots + b_ne^{-s_nt}. \quad (1.50)$$

If the system is stable, the coefficients $-s_1, -s_2, \dots, -s_n$ have negative real parts, and the transient terms with these constants will eventually fade away. Therefore, only the steady state part

$$y(t) = ae^{-j\omega t} + \bar{a}e^{j\omega t} \quad (1.51)$$

from equation (1.50) remains. This result is still not the complete solution because the constants a and \bar{a} are not yet evaluated. Those constants can be determined from the partial fraction expansion (1.49). To solve for a , according to the theory of partial fractions as shown in e.g. [13, p. 332], one can multiply $Y(s)$ by $(s + j\omega)$ to cancel the term from the denominator, so that

$$Y(s)(s + j\omega) = H(s) \frac{\omega A_{\text{in}}}{s^2 + \omega^2} (s + j\omega) = H(s) \frac{\omega A_{\text{in}}}{s - j\omega},$$

and evaluate this at the root of $(s + j\omega)$, which is $s = -j\omega$. This gives

$$a = -H(-j\omega) \frac{A_{\text{in}}}{2j}.$$

Similarly for \bar{a} ,

$$\bar{a} = H(j\omega) \frac{A_{\text{in}}}{2j}.$$

Because $H(j\omega)$ is complex-valued, it can be written in two parts as

$$H(j\omega) = |H(j\omega)| e^{j\phi},$$

where $|H(j\omega)|$ is the magnitude and $e^{j\phi}$ carries the angle information. The angle ϕ is calculated as

$$\phi = \angle H(j\omega) = \arctan \left[\frac{\Im\{H(j\omega)\}}{\Re\{H(j\omega)\}} \right]. \quad (1.52)$$

The same reasoning applies to $H(-j\omega)$:

$$H(-j\omega) = |H(-j\omega)| e^{-j\phi} = |H(j\omega)| e^{-j\phi}$$

because the magnitude does not change with $-j\omega$, it just points in the opposite direction than in the case of $j\omega$. The minus sign on the phase term comes directly from equation (1.52) defining the angle because $\arctan(-x) = -\arctan(x)$ is the negative angle equivalence for the arcus tangent.

Now the constants have been defined, and their values are

$$a = -\frac{A_{\text{in}}}{2j} |H(j\omega)| e^{-j\phi} \quad \text{and} \quad \bar{a} = \frac{A_{\text{in}}}{2j} |H(j\omega)| e^{j\phi}.$$

Substitution into equation (1.51) yields an expression for the steady state output as

$$y(t) = A_{\text{in}} |H(j\omega)| \frac{e^{j(\omega t + \phi)} - e^{-j(\omega t + \phi)}}{2j} = \boxed{A_{\text{in}} |H(j\omega)| \sin(\omega t + \phi)},$$

which means that the output signal is also a sine wave with the same frequency ω as the input signal, but the phase has shifted and the amplitude of the output sine wave is related to the input amplitude by the equation

$$A_{\text{out}} = A_{\text{in}} |H(j\omega)| \quad \Rightarrow \quad |H(j\omega)| = \boxed{\frac{A_{\text{out}}}{A_{\text{in}}}}.$$

This is an extremely significant result! The amplitude of the output signal is proportional to the amplitude of the input sine wave with the factor of magnitude of the transfer function evaluated at some angular frequency $\omega = 2\pi f$. This way it is possible to scan through a set of frequencies $[f_{\text{low}}, f_{\text{high}}]$, to see how the amplitude of the sinusoidal output signal changes as a function of the frequency.

Due to the derivations provided, for sinusoidal input signals the characteristics of any system can be directly modelled by a transfer function,

$$H(j\omega) = \frac{Y(j\omega)}{X(j\omega)},$$

using the complex angular frequency $j\omega$ as the argument instead of the general Laplace variable s . This also leads to definitions of magnitude

$$|H(j\omega)| = \frac{|Y(j\omega)|}{|X(j\omega)|}$$

and phase

$$\angle H(j\omega) = \angle \frac{Y(j\omega)}{X(j\omega)}$$

having the $j\omega$ term directly as the argument. All this information is available already from the transfer function without the need for inverse transform! [12, pp. 269 – 273]

Apart from the transfer functions of general circuit analysis, Laplace transforms are used extensively to solve all kinds of differential equations. From the general definition (1.47), it is possible to derive the formulae (1.53) – (1.55) to transform integrals and differentials.

$$\mathcal{L} \left[\int_0^t f(\tau) d\tau \right] = \frac{1}{s} \mathcal{L}[f(\tau)] \quad (1.53)$$

$$\mathcal{L} \left[\frac{df(t)}{dt} \right] = s \mathcal{L}[f(t)] - f(0) \quad (1.54)$$

$$\mathcal{L} \left[\frac{d^2 f(t)}{dt^2} \right] = s^2 \mathcal{L}[f(t)] - s f(0) - \frac{df(0)}{dt}. \quad (1.55)$$

These formulae reveal the use of the Laplace variable s as a replacement of the differential operator in the Laplace domain, where factor $\frac{1}{s}$ represents an integrator and factor s a differentiator. With these transforms it is possible to linearise a differential equation in the s domain, so that in many cases a complicated differential equation is solved with less effort using Laplace transforms. When a differential equation is transformed to the s domain, it is represented in the form of a transfer function.

The Laplace transformed transfer function gives direct information about the stability of the system. The roots of the Laplace variable s in the denominator of the transfer function are called *poles*. These roots can be used to indicate if

the analysed system is stable with respect to a certain input function. The basic idea of the analysis method is drawn in Figure 1.15, where the green colour on the left side refers to a stable system and the red colour on the right side to an unstable system. The marker 'X' is used to indicate the location of a pole.

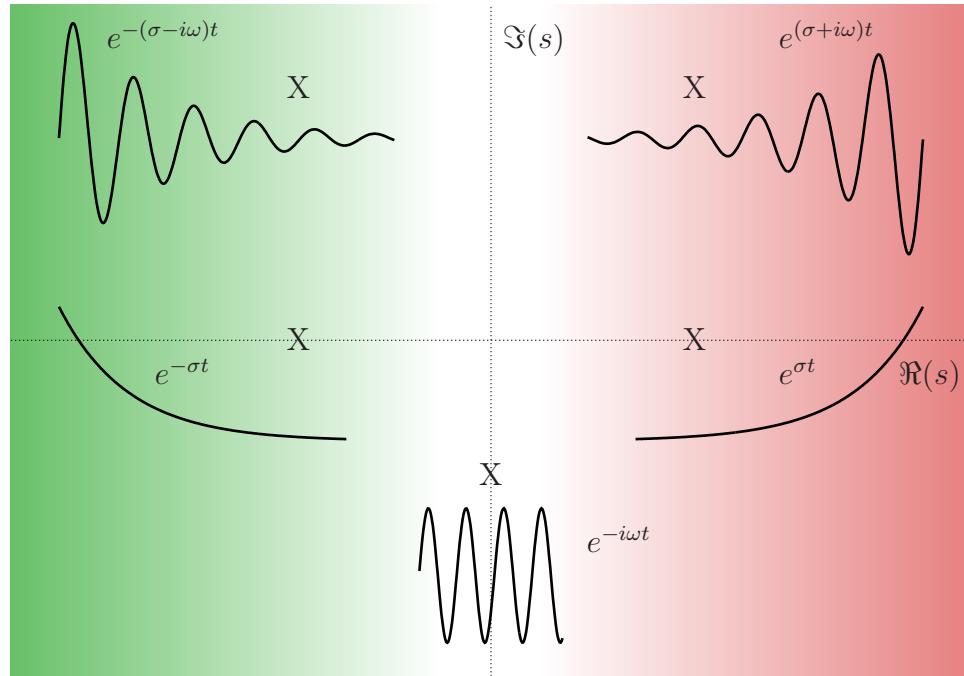


Figure 1.15: The Laplace s domain quadrants of stability

The general rules of stability state that

- If all poles of a transfer function are located on the left side of the imaginary axis in the complex plane of the Laplace variable s , then the system described by the transfer function is stable.
- If any of the poles is located exactly on the imaginary axis, the system is undeterministic.
- If any of the poles is located in the right side of the imaginary axis in the complex plane, the system is unstable.

All of these cases are related to the solutions of a differential equation by the characteristic equation that determines the argument for the exponential function

$$y(t) = \sum_{i=1}^n C_i e^{-r_i t}$$

acting as the basis of the solution of function $y(t)$. The arguments r_i are the roots of a characteristic equation for the n th order differential equation and by definition of the s domain transfer function $H(s)$, the roots r_i correspond to the roots s_i . This is why the different functions depicted in Figure 1.15 are equivalent to the three different cases of solutions of the n th order differential equation.

The direct substitution of the Laplace variable s with $j\omega$ in transfer functions causes confusion in several contexts, because of the similarity of the Laplace transform with the Fourier integral transform

$$F(\omega) = \mathcal{F}[f(x)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{-j\omega x} dx, \quad (1.56)$$

with kernel $K(j\omega, x)$. The Fourier integral transform yields a spectral representation of the original function $f(x)$ in the frequency domain. In practical applications, Fourier transforms can be equally used with functions of space $f(x, y, z)$ and functions of time $f(t)$ unlike Laplace transforms, which are mainly applied in the time domain.

Although the integral definitions of Laplace and Fourier transforms are relatively similar, and one could be considered as a special case of the other, the Laplace transform variable substitution $s = j\omega$ presented here is a special case of the general complex variable $s = \sigma + j\omega$ and it applies only as a 'short-cut' to the sinusoidal frequency response modelling as explained above. In general engineering practises, Laplace transforms are used to study the stability of any linear system by examining the complex roots of the denominator in the transfer function. The Fourier transform, on the other hand, can be used to find analytic functions that describe how energy in some system with some arbitrary function $f(x)$ is divided between different frequencies in the frequency plane. The Fourier integrals lead to continuous spectrum presentation instead of the discrete point spectrum given by Fourier's series. Being analytic functions, the Fourier integrals do not have that many practical applications where they could be used.

To clear the possible confusion concerning basic engineering applications, there is no link between the Fourier $j\omega$ and the Laplace $j\omega$ whatsoever. Both transforms can be used as purely mathematical tools to make a set of linear equations from relatively complicated differential equations to simplify the solution process, which can be considered as the only similarity between these two transforms.

1.3 BASIC CONCEPTS OF ELECTRONICS

The current density \vec{J} in a conductor has a dependency to the electric field \vec{E} and to the properties of the conductive material itself. Especially in metals \vec{J} is nearly directly proportional to \vec{E} . This phenomenon was first discovered by German physicist Georg Ohm [3, p. 850]. When properly machined into symmetric geometrical shapes, conducting and insulating materials can be used to construct building blocks, which connect together to perform mathematical transformations to the basic physical concept of current flow.

To lay a steady foundation to basic circuit analysis, it is sufficient to introduce one single equation

$$\boxed{V = IR}. \quad (1.57)$$

This equation is generally known as Ohm's law, where the symbol V denotes voltage, I refers to current and R is the resistance. This law is only an idealised model for certain metallic materials, but it is sufficient and accurate enough for basic electronics calculations. The following sections give more detailed background information on the variables V , I and R mentioned in Ohm's law and explain other concepts in the field of analogue electronics as well.

1.3.1 Current and voltage

Voltage V carries a unit named *volt* [V] and the unit for current I is called *ampere* [A]. The elementary element of electricity is electric charge q , which constructs of the basic SI units [A · s]. This combination of units is named as the *coulomb*, named after French physicist Charles Augustin de Coulomb. Charge can be either positive or negative; opposite charges attract each other and similarly polarised charges repel each other. This result yields from a fundamental physical law about the interaction of two charges, and this law is also named after Mr. de Coulomb. Coulomb's law says that between charges q_1 and q_2

$$\vec{F}_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{|\vec{r}_{12}|^2} \frac{\vec{r}_{12}}{|\vec{r}_{12}|}, \quad (1.58)$$

where the convention with the subscripts is that with this notation of \vec{F}_{12} , charge 1 exerts a force on charge 2 along the vector \vec{r}_{12} joining the two charges. The notation $|\vec{r}_{12}|$ refers to the length of the vector \vec{r}_{12} , so that the last term in equation (1.58) is just a unit vector giving a direction to the force between the charges. Of course the absolute value of the force does not change if the subscripts are changed, so it is irrelevant to get stuck with the conventions. The force is acting between the two charges, and that is all one needs to know. The factor $(4\pi\epsilon_0)^{-1}$

in front of the equation is only there to compensate for the differences of the unit systems used. In this book only SI units are relevant, so that the factor is used in related expressions throughout the text.

Figure 1.16 shows a more general situation where several charges are interacting with each other and the total force \vec{F} affecting charge q is a superposition (a sum) of forces exerted by the individual charges q_1 and q_2 .

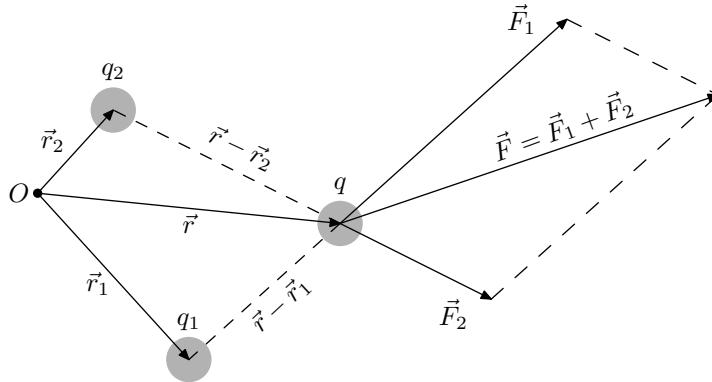


Figure 1.16: Forces acting on charge q

In this case the formula to describe the force acting on charge q (with reference to the notations used in Figure 1.16) is written as

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \sum_i \frac{qq_i}{|\vec{r} - \vec{r}_i|^2} \frac{\vec{r} - \vec{r}_i}{|\vec{r} - \vec{r}_i|}. \quad (1.59)$$

The vector \vec{r} in equation (1.59) holds the coordinates of charge q , which is under the influence of force \vec{F} due to charges q_i . Basically vector \vec{r} is drawn from the origin of the chosen coordinate system to the location of charge q . From here it is simple to define the concept of an electric field \vec{E} , which is defined as the quotient

$$\vec{E} = \frac{\vec{F}}{q} = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{|\vec{r} - \vec{r}_i|^2} \frac{\vec{r} - \vec{r}_i}{|\vec{r} - \vec{r}_i|}. \quad (1.60)$$

In relation to Figure 1.16, now the charge q is divided away and replaced with a point P (with a position vector \vec{r}) in space, to which the electric fields \vec{E}_i of all surrounding charges are affecting. Equation (1.60) holds in the case of simple point charges, and in a more general case the sum is replaced by an integral and the point charges by a continuous charge density ρ , according to

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \int_{\text{all space}} \frac{\rho(\vec{r}_\tau)}{|\vec{r} - \vec{r}_\tau|^2} \frac{\vec{r} - \vec{r}_\tau}{|\vec{r} - \vec{r}_\tau|} d\tau.$$

Here \vec{E} is the total electric field at point P , which is connected to the coordinate origin with the position vector \vec{r} , and vector \vec{r}_τ expands from the origin to the infinitesimal piece of matter $d\tau$. As a general summary, charged particles generate an electric field around them, and certain forces exist between charged particles that give mechanical properties to the basic laws of electrics.

Figure 1.17 shows the electric field lines of charged particles with positive and negative polarities. The dashed lines depict a constant value of a potential function ϕ that exists at a certain distance away from the centre of the charge. The

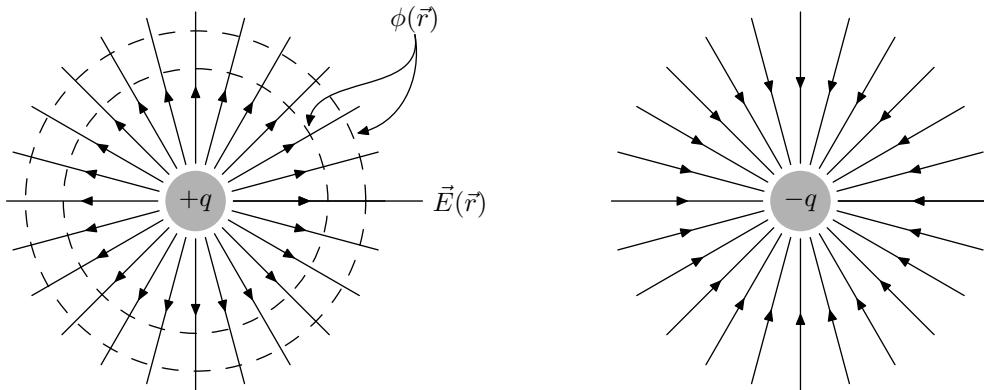


Figure 1.17: Electric field lines (\vec{E}) and equipotential lines (ϕ) of charge q

potential function relates to familiar potential energy U by a relation,

$$U = q\phi, \quad (1.61)$$

where U has the units of joule just like the mechanical equivalent measure of energy. The concept of a potential ϕ and potential energy U in the context of electric fields is explained by defining the concept of work in electricity. External physical work is needed to move a charged particle within the influence of an electric field. As an example, in Figure 1.18 a test charge is moved in the electric field generated by charge q . If work is done to push the test charge against the force of the electric field, the test charge gains potential energy. On the other hand, work done by the electrostatic force on the charge represents a loss of potential energy of the test charge. Setting \vec{r}_A and \vec{r}_B as the position vectors of the points A and B of Figure 1.18, the potential energy difference

$$q\phi(\vec{r}_B) - q\phi(\vec{r}_A) = -q \int_A^B \vec{E} \bullet d\vec{l}. \quad (1.62)$$

The charge q can be divided out from equation (1.18), and now it is possible to

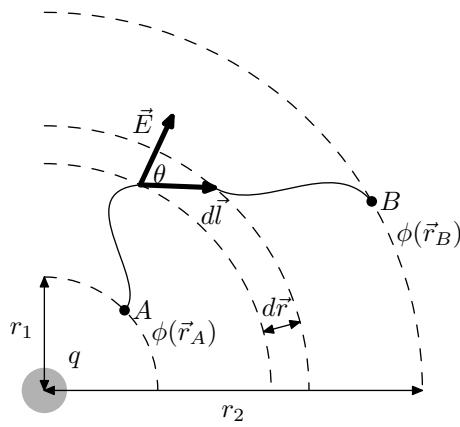


Figure 1.18: Moving a test charge in an electric field of charge q

define voltage as the potential difference

$$V = \phi(\vec{r}_B) - \phi(\vec{r}_A) = - \int_A^B \vec{E} \bullet d\vec{l}.$$

So eventually, the unit of volt is defined as the potential difference $\phi_B - \phi_A$. From equation (1.61) it also follows that dividing U by electric charge q yields again the complicated definition of volt:

$$V = \frac{U}{q} = \frac{U}{[A \cdot s]} = \frac{[\text{kg} \cdot \text{m}^2]}{[A \cdot s^3]}. \quad (1.63)$$

In other words, when one coulomb of charge is moved through a potential difference of one volt, one joule of energy is used [14, p. 2].

The electric field has many practically significant properties. When

$$\oint_C \vec{E} \bullet d\vec{l} = 0, \quad (1.64)$$

the electric field is said to be conservative. This is an important concept and it is needed to justify phenomena in the theory of electromagnetic induction. For example, inside the two conducting plates of Figure 1.19 the electric field is zero, but between the plates there exists a potential difference caused by the electric field. This phenomenon is used to create capacitors, which can store energy and act as a temporary voltage source or a storage of charge.

Electric current is much simpler to define than voltage. Again proceeding from the elementary charge, current is established as the derivative

$$I = \frac{dq}{dt},$$

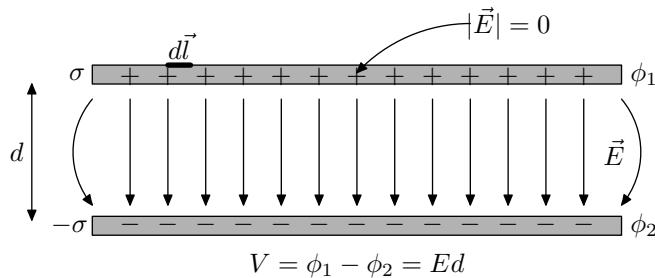


Figure 1.19: The electric field between two conducting plates

meaning that current measures the rate of change of charge in some system. Alternatively, a flow of electrons might be a more descriptive term to describe current, and current is often referred to as an analogy to fluid flow.

Electric currents can exist as alternating current (AC), referring to a continuously changing current value, or direct current (DC), which holds a constant value at all times. Ohm's law (1.57) binds the waveforms of current and voltage to be similar, and somehow justifies the use of notations AC or DC also in the context of voltages. With a small lack of conceptual rigour, sometimes it is common to make a reference to alternating voltage as AC, although the C-letter comes from current.

Before getting deeper in to the subject, let's introduce some more notations. In equation (1.57) capital letters are used to write the voltage and current symbols. Ohm's law could just as well be written using i to denote current and v to denote voltage. There is a reason for this and the reasoning is based on the information provided in Table 1.1.

Table 1.1: Notation preferences for symbols and subscripts

I_B	DC only	capital letter symbol + capital subscript
i_b	AC only	small letter symbol + small subscript
i_B	AC + DC	small letter symbol + capital subscript
I_b	AC related	capital letter symbol + small subscript

With reference to Table 1.1, it is more common to write currents and voltages with small letters when it is not directly specified if the value is pure DC or not. The notation preferences defined in Table 1.1 are used throughout this book, but based on personal experience, this notation principle is difficult to master.

1.3.2 Elementary building blocks of electronics

The most basic components of analogue electronics are a resistor (R), a capacitor (C) and an inductor (L). It is common to think that these basic components are also the simplest ones. Wrong. It is relatively simple to construct a circuit from a readily available set of integrated circuits, but creating a design from a mixture of R , C and L from scratch is definitely not simple. These components have interesting physical properties, which can be accurately modelled with mathematical equations. The fact that has been blurred by integrated circuits of digital electronics is that the basic analogue components make use of the natural properties of materials to interact with electric currents and voltages. In that sense, analogue electronics is a branch of true natural science, whereas digital electronics in principle is only a simplified application of analogue electronics.

In addition to the notations already explained, there exist several different standards that define the schematic symbols for components in circuit diagram drawings. Figure 1.20 presents the schematic symbols used for a resistor, a capacitor and an inductor in this book.

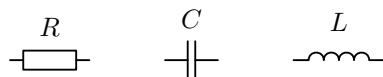


Figure 1.20: The schematic symbols for a resistor, a capacitor and an inductor

At least the resistor symbol is a bit different from the one that is mostly used in electronics literature, such as [15]. Different standards have also had several versions of the inductor symbol, but the one used in Figure 1.20 is the most common one. Sometimes a black-painted resistor has also been used to depict an inductor, but this symbol was commonly in use only for a few years at the beginning of the 21st century and definitely very rarely encountered in general.

A common practise is to designate resistors with the letter R , to which a subscript can be added to identify a certain resistor in a circuit diagram containing several resistors. A similar reasoning applies to capacitors identified with the letter C and inductors with the letter L .

1.3.3 Properties of resistors

The resistor is the simplest of all components, since its behaviour does not depend on frequency. Resistors' job in circuits is to create voltage differences in

specified places in a circuit or to set current levels to needed values according to Ohm's law (1.57). Resistor is a component that wastes energy from the circuit with a mechanism where a voltage drop over a resistor dissipates heat as a sign of lost power. As real components, resistors are also categorised by their ability to dissipate power. The most common resistor types are defined as 1/4 watts and the bigger the physical size of the component is, the more it can dissipate power.

The basic equations for resistance follow directly from Ohm's law. The unit for resistance is called *ohm* and it is marked with a capital omega symbol Ω . In relation to voltage, resistance also constructs from a quite peculiar combination of basic SI units.

The resistances of commonly available resistors vary from 1Ω to about $100\text{ M}\Omega$. Discrete resistor components are fabricated according to different accuracy/ tolerance standards, which also define the basic resistance values of resistors. The most common series of possible resistance values is the E12 series, which is also applicable for capacitors. Here the number 12 means that a decade interval of values has been divided into 12 logarithmic steps, which define the possible component values in the series. In this E12 series the tolerance (maximum deviation from a given value) is 10 %. Other series such as E24, E48, E96 and E192 have smaller tolerances and offer more standard resistance values to choose from. The safest choice in circuit design is to use values defined by the E12 series, and any critical components in a design that depend on a high accuracy component value should be avoided. Table 1.2 presents the most common E-series values and also enlightens the idea behind the division among different series of resistor values.

Table 1.2: Selected series of resistance/capacitance values

	tol %	$10^n \Omega$ ($n = 0, 1, 2, \dots, 8$)											
E6	20	10		15		22		33		47		68	
E12	10	10	12	15	18	22	27	33	39	47	56	68	82
E24	5	10	11	12	13	15	16	18	20	22	24	27	30
		33	36	39	43	47	51	56	62	68	75	82	91

When resistors are connected in series, the total resistance is calculated by a direct sum of all resistance values. Things are not that simple when resistors are connected in parallel. In this case, resistance values of individual components are added reciprocally. Figure 1.21 depicts the combinations of resistors and the equations to calculate the combined resistance R in the case of series and parallel

connected resistors.

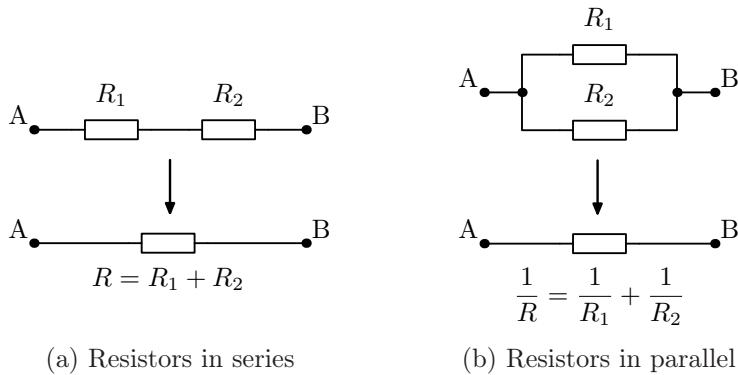


Figure 1.21: Resistors connected in series and in parallel

From the perspective of materials science, every material can be measured to have some amount of resistivity

$$\rho = \frac{E}{J}.$$

By placing a slab of material under the influence of an electric field E and measuring the value of current density J , the resistivity of the material is defined. Resistivities of different materials are tabulated in standard tables of physics. Some resistivity values of selected materials are gathered to Table 1.3. The division to conductors and insulators is quite clear, but as a reference the semiconductor substances of germanium and silicon are also added to the end of the table.

Table 1.3: Resistivities ρ [$\Omega \cdot m$] of selected materials at room temperature

conductors		insulators	
silver	$1.5 \cdot 10^{-8}$	amber	10^{14}
copper	$1.7 \cdot 10^{-8}$	glass	$10^{10} - 10^{14}$
gold	$2.4 \cdot 10^{-8}$	lucite	$> 10^{13}$
aluminium	$2.8 \cdot 10^{-8}$	mica	$10^{11} - 10^{15}$
tungsten	$5.3 \cdot 10^{-8}$	quartz	10^{16}
steel	$20 \cdot 10^{-8}$	sulfur	10^{15}
lead	$22 \cdot 10^{-8}$	teflon	$> 10^{13}$
mercury	$95 \cdot 10^{-8}$	wood	$10^6 - 10^{11}$
germanium	0.60	silicon	2300

As resistivity defines only a property of a material, there needs to be a way to

define a single value of resistance for a resistor. Clearly resistivity and resistance are related via the dimensions of the component, more accurately expressed as

$$R = \frac{\rho L}{A},$$

where L refers to the dimension of length and A to the dimension of a cross-sectional area.

In a laboratory environment where the ambient temperature is practically constant, one can just solve R from equation (1.57) and use that in calculations. In reality resistors have some dependency on temperature changes. This can be taken into account in calculations by using equation (1.65), which defines the temperature dependency of resistance as

$$R(T) = R_0 [1 + \alpha (T - T_0)]. \quad (1.65)$$

The temperature coefficient α is again material dependent. Table 1.4 shows some temperature coefficient values for materials of importance.

Table 1.4: Temperature coefficients α of selected materials at room temperature

material	α	material	α
copper	0.00393	silver	0.00380
aluminium	0.00390	tungsten	0.00450
iron	0.00500	lead	0.00430

1.3.4 Properties of capacitors

Along with resistors, capacitors are the most commonly used discrete components in electric circuits. A capacitor stores energy into itself, so that it can be used as a short time power source. With alternating currents a capacitor behaves like a resistor, whose ohmic value depends on frequency. Therefore capacitors are commonly used to filter some unwanted range of frequencies out from an AC signal. A capacitor acts like a break in the circuit for DC signals, so that is why capacitors are used to isolate DC signals from certain parts of the circuit.

Just like resistors have resistance, capacitors have capacitance. The official definition states that

$$C = \frac{q}{V}, \quad (1.66)$$

meaning that capacitance C holds some amount of charge q within potential difference (voltage) V . [3, p. 816]

Capacitance is closely connected to electric fields and dielectrics. Figure 1.19 from section 1.3.1 is a textbook example of a plate capacitor, where the capacitance is calculated as

$$C = \frac{q}{V} = \frac{\epsilon_0 A}{d},$$

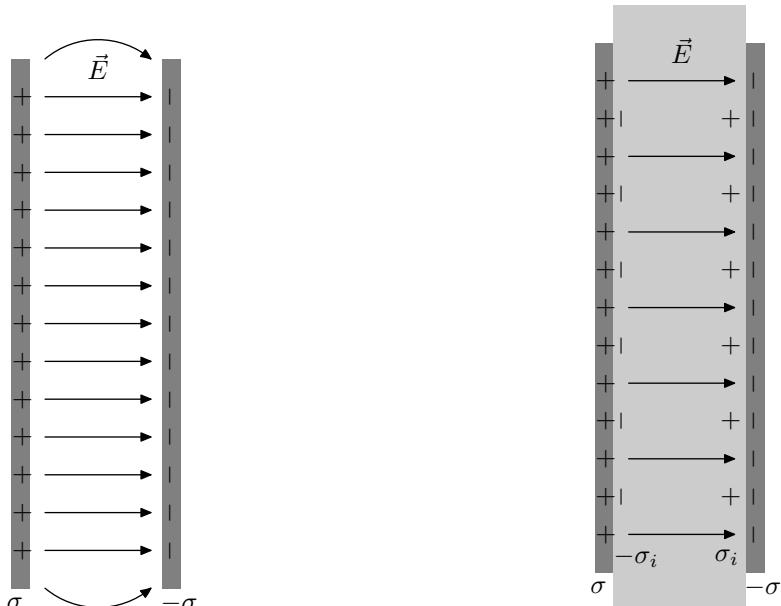
because the uniform charge density inside the plate is

$$\sigma = \frac{q}{A} = \epsilon_0 E = \epsilon_0 \frac{V}{d},$$

as was defined in section 1.3.1. The connection to dielectrics (induced charges in insulating material) arises when the space between the plates of the capacitor is filled with insulating material. In that case the capacitance of the parallel plate capacitor

$$C = \epsilon C_0 = \frac{\epsilon \epsilon_0 A}{d},$$

where ϵ is the relative permittivity of the dielectric (insulating material). Dielectric substances can be used to increase the capacitance value of capacitors without increasing the physical size of the components. Adding a dielectric into a capacitor reduces the potential difference across the capacitor. Figure 1.22 explains the effect of an added dielectric between the parallel plate capacitor.



(a) A parallel plate capacitor with a vacuum dielectric

(b) A parallel plate capacitor with an insulating dielectric

Figure 1.22: The effect of an insulating dielectric between capacitor plates

Values of relative permittivity for some selected materials are collected in Table 1.5.

Table 1.5: Values of relative permittivity ϵ of selected materials

material	ϵ	material	ϵ
vacuum	1.0	air	1
mylar	3.1	mica	6
neoprene	6.7	glass	7
Polyethylene	2.3	germanium	16

The definition of capacitance (1.66) does not give any meaningful information with direct currents. Writing equation (1.66) as

$$q = Cv,$$

and differentiating both sides with respect to time,

$$\frac{d}{dt}q(t)dt = C \frac{d}{dt}v(t)dt,$$

reveals the equation

$$i(t) = C \frac{d}{dt}v(t) \quad (1.67)$$

relating current and voltage of a capacitor. If the charge of the capacitor stays constant, then the derivative $dq(t)/dt$ is zero, and no current is flowing anywhere. For this reason a capacitor is kind of a blocking issue when operated with DC voltages (although the initial charging of the capacitor happens also with DC voltage).

For sinusoidal input voltage, equation (1.67) says

$$i = C \frac{d}{dt}V_0 \sin(\omega t) = \omega CV_0 \cos(\omega t) = \omega CV_0 \sin\left(\omega t + \frac{\pi}{2}\right),$$

which means that the phase of the current in a capacitor is leading the input voltage by 90 degrees. If this current-voltage relation is transferred to the complex plane where the angle difference between the real and imaginary axes is 90 degrees,

$$i = j\omega CV_0 \quad \Rightarrow \quad i = \frac{V_0}{j\omega C}.$$

This expression has the same form as Ohm's law, the frequency dependent impedance of the capacitor being $\frac{1}{j\omega C}$.

The same complex impedance expression is obtained using the Laplace transform. An application of the transform rule (1.54) for the derivative in equation (1.67) gives

$$\mathcal{L}[i(t)] = Cs\mathcal{L}[v(t)] \quad \Rightarrow \quad \frac{\mathcal{L}[v(t)]}{\mathcal{L}[i(t)]} = \frac{1}{sC}. \quad (1.68)$$

When the input signal is a sine wave, the Laplace variable $s = j\omega$ as explained in section 1.2.6, and the capacitor impedance expression is in agreement with the earlier result. The reactance of a capacitor is $X_C = \frac{1}{\omega C}$, and it is simply derived from the complex impedance expression by taking out the imaginary factor j . [14, pp. 25 – 27]

Capacitors have similar properties as resistors when connected in series or in parallel. As Figure 1.23 indicates, the rules of capacitance addition are the opposite to the rules of resistance addition. When capacitors are connected in parallel, the combined capacitance is calculated as a direct sum of individual capacitance values, whereas series-connected capacitances are summed reciprocally.

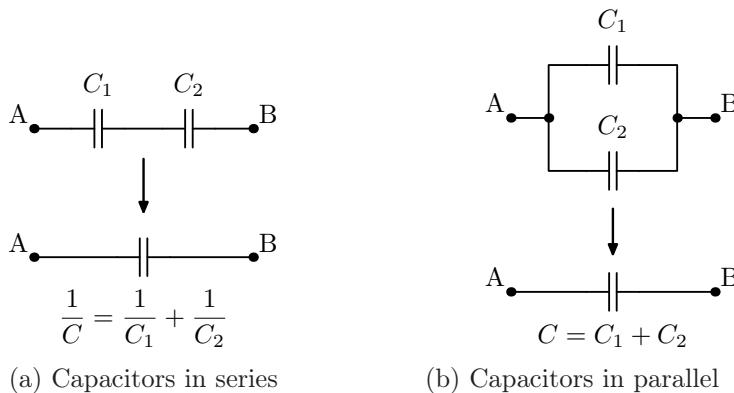


Figure 1.23: Capacitors connected in series and in parallel

The practical capacitor components that can be bought from a local electronics store have a variety of different types to choose from. Table 1.6 lists a few common types of capacitors and their properties. The table is adapted directly from reference [14, p. 19]. The most common capacitor types sold by component dealers are ceramic, mylar (polyester) and electrolytic capacitors. The quality difference between ceramic and mylar capacitors is subtle, although some might favour mylar capacitors over ceramic ones in audio applications. Electrolytic capacitors are commonly used to give large capacitance values with a single com-

ponent. Unfortunately electrolytic capacitors are the most problematic because of poor quality and relatively short life span.

Table 1.6: Capacitor types and their properties

type	value range	accuracy	leakage	description
mica	1 pF – 10 nF	good	good	excellent quality, good for RF applications
ceramic	10 pF – 1 μ F	-	-	small, inexpensive, very popular, poor temperature stability
mylar	1 pF – 10 μ F	good	good	inexpensive, good, very popular, poor temperature stability, also known as polyester capacitors
polystyrene	10 pF – 10 nF	good	very good	high quality, large in size, suitable for signal filters
polycarbonate	100 pF – 10 μ F	good	good	high quality, suitable for integrators, good temperature stability
tantalum	0.1 μ F – 500 μ F	poor	medium	popular, small, polarised, poor temperature stability, recommended to replace electrolytic caps
electrolytic	0.1 μ F – 0.2 F	terrible	awful	recommended only for power supply filters, polarised, awful temperature stability

1.3.5 Properties of inductors

Inductors inherit their properties from the theory of electromagnetic induction, which consists of many different laws and internal relations between those laws. A practical inductor component is just a piece of wire that is bent into a loop or a helical structure. Usually inductors consist of several turns of wire stacked side by side or even in many layers on top of each other. As a commonly used special case, an inductor consisting of several sequential loops of wire is called a *solenoid*. The solenoid therefore looks like a tube made of wire as shown in Figure 1.24.

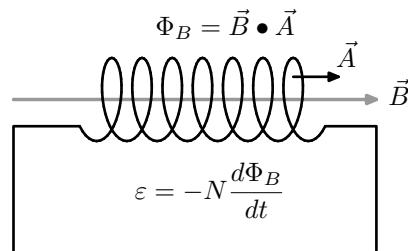


Figure 1.24: Solenoid made of N loops of conducting wire

Just like resistors have resistance and capacitors have capacitance ... inductors

have inductance. For a single inductor the definition of inductance arises from the ability of a wire loop (or any piece of wire in general) to generate a magnetic field when current is applied to the wire. Mathematically the definition of inductance derives from the flux of a magnetic field Φ , the relation being

$$L = \frac{N\Phi_B}{i}, \quad (1.69)$$

where N is the number of loops in the inductor and i is the current flowing in the wire. Inductance is officially called *self-inductance* because the inductance properties of any electrical component appear internally when a flowing current generates a related magnetic flux around the component.

Multiplying equation (1.69) by current i and differentiating with respect to time,

$$L \frac{di}{dt} = N \frac{d\Phi_B}{dt},$$

and furthermore there exists a relation to electromotive force,

$$\varepsilon = -L \frac{di}{dt}. \quad (1.70)$$

This e.m.f. is generally named as *self-induced e.m.f.*, since the voltage ε over the inductor arises internally due to excitation with alternating current. Because of the differentials in equation (1.70), only alternating currents are able to generate the e.m.f.; static direct current only initiates a constant magnetic field. The minus sign in equation (1.70) indicates that the polarity of the self-induced e.m.f. opposes the changes in the flowing current.

The deeper theoretical background of inductors is based on two types of electric fields, electrostatic fields \vec{E}_c and nonelectrostatic fields \vec{E}_n . Earlier when treating electric fields in the context of voltage in section 1.3.1, it was stated that a line integral over a closed path enclosing an electric field equals zero. This always holds, but only when the field is an electrostatic field \vec{E}_c . The nonelectrostatic field \vec{E}_n refers especially to situations where the electromotive force ε is a result of electromagnetic induction. In this case the line integral of equation (1.64) should be written

$$\oint_C \vec{E}_n \bullet d\vec{l} = \varepsilon = -\frac{d\Phi_B}{dt}, \quad (1.71)$$

meaning that the total work done by the field \vec{E}_n equals the induced electromotive force.

Additionally, Gauss's law for an electric field states that a surface integral of \vec{E} over a closed surface with a differential surface normal vector $d\vec{A}$

$$\oint_C \vec{E}_c \bullet d\vec{A} = \frac{Q}{\epsilon_0},$$

where Q is the total charge enclosed within the surface. Note that the electrostatic field \vec{E}_c is used here. For a nonelectrostatic field,

$$\oint_C \vec{E}_n \bullet d\vec{A} = 0.$$

The electrostatic field \vec{E}_c and the nonelectrostatic field \vec{E}_n sum up linearly to a total field

$$\vec{E} = \vec{E}_c + \vec{E}_n.$$

Especially when connecting inductors as part of a closed circuit, the total field $\vec{E}_c + \vec{E}_n$ within the inductor must be zero to keep the total field conservative. When combining equations (1.70) and (1.71)

$$\oint_C \vec{E}_n \bullet d\vec{l} = -L \frac{di}{dt}, \quad (1.72)$$

and in the case of a closed electric circuit, the line integral of equation (1.72) is calculated around the whole circuit. But because the \vec{E}_n field is nonzero only inside the inductor, equation (1.72) can be written as

$$\int_a^b \vec{E}_n \bullet d\vec{l} = -L \frac{di}{dt}, \quad (1.73)$$

where points a and b are the connecting terminals of the inductor. In this case it is clear that the electrostatic field has value $\vec{E}_c = -\vec{E}_n$ to make the total field zero. This leads to the equation

$$\int_a^b \vec{E}_c \bullet d\vec{l} = L \frac{di}{dt},$$

from where follows an expression for the voltage over the inductor terminals,

$$V_{ab} = V_a - V_b = L \frac{di}{dt}. \quad (1.74)$$

[3, pp. 1151 – 1153]

The inductance equations presented so far concern only air core inductors. Similarly as capacitance can be increased by adding insulating materials between plates, inductance can be increased by adding magnetically sensitive material inside a solenoid structure. The magnetic field generated by the solenoid also magnetises the core material, which increases the total magnetic flux inside the solenoid. According to equation (1.69), this also increases the inductance of the

inductor. The magnetic equivalent quantity of dielectric relative permittivity ϵ is magnetic permeability μ .

When evaluating expressions for the magnetic field density \vec{B} for air core inductors, Ampère's law and the Biot-Savart law automatically add the permeability of free space μ_0 to the equations. The relative permeability is used as a multiplier in the form $\mu\mu_0$ to scale the value of free space permeability. Useful core materials to increase the inductance of a solenoid are materials with ferromagnetic and diamagnetic tendencies. As a side effect, magnetic core materials introduce nonlinearities to an inductor so that the dependency between the generated magnetic flux and the applied current is not linear anymore.

With alternating currents, inductors have similar resistive properties as capacitors have. The basic inductor equation (1.74) in the case of alternating signals is written as

$$v(t) = L \frac{d}{dt} i(t) \quad (1.75)$$

relating the voltage and current of an inductor. For sinusoidal input current, equation (1.75) says

$$v = L \frac{d}{dt} I_0 \sin(\omega t) = \omega L I_0 \cos(\omega t) = \omega L I_0 \sin\left(\omega t + \frac{\pi}{2}\right),$$

which is the same as in the case of capacitance, but now the roles of voltage and current have changed places.

If this current/voltage relation is transferred to the complex plane where the angle difference between the real and imaginary axes is 90 degrees,

$$v = j\omega L I_0.$$

This expression has the same form as Ohm's law, the frequency dependent impedance of an inductor being $j\omega L$. The same complex impedance expression can be obtained with Laplace transforms. When applying the derivative transform rule (1.54) to equation (1.75),

$$\mathcal{L}[v(t)] = L s \mathcal{L}[i(t)] \quad \Rightarrow \quad \frac{\mathcal{L}[v(t)]}{\mathcal{L}[i(t)]} = sL. \quad (1.76)$$

When the input signal is a sine wave, the Laplace variable $s = j\omega$ as explained in section 1.2.6, and the inductor impedance expression is in agreement with the earlier result. The reactance of an inductor is $X_L = \omega L$, and it is simply derived from the complex impedance expression by taking out the imaginary factor j .

When connected in series or in parallel, inductance values are added with the same rules as pure resistances. The addition rules for inductances are summarised in Figure 1.25. One word of warning concerning practical circuit de-

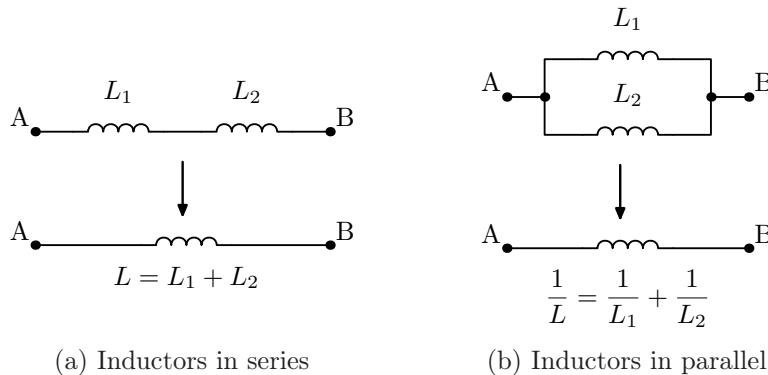


Figure 1.25: Inductors connected in series and parallel

sign: closely connected inductors also generate mutual inductance, which alters the value of both inductors. Usually the effect of mutual inductance is unwanted, so just keep in mind to keep inductors separated when building circuits.

Inductors as real world components are the most difficult ones to acquire. Some standard high-frequency inductors are sold directly from the shelf, but if some special inductance values are desired, it is better to wind the inductor as a do-it-yourself project because combinations of several discrete closely connected inductors easily result in the mutual inductance effect. If a large inductor is wound from a wire several meters long, then the pure resistance R_L of the wire has to be taken into account in calculations.

1.3.6 Circuit analysis methods

Circuit analysis refers to a process of solving unknown currents and voltages in circuit networks. From the designing perspective, different circuit analysis methods enable the modelling of the designed circuits mathematically, without the need for excessive prototype builds. The most basic circuit analysis techniques rely on the so-called *Kirchhoff's laws*: Kirchhoff's Voltage Law (KVL) and Kirchhoff's Current Law (KCL). More advanced methods include *mesh* analysis and *nodal* analysis, both of which are systematic matrix methods built on top of Kirchhoff's laws. The simple circuit in Figure 1.26 is used as a basis for explaining how to use Kirchhoff's laws in practise. In Figure 1.26 the black dots are called *nodes*, the elements between the nodes are called *branches* and the closed path formed by the branches is called a *mesh* or just simply a *loop*.

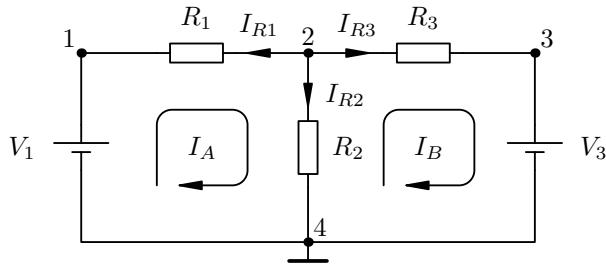


Figure 1.26: An example circuit to explain Kirchhoff's laws

Kirchhoff's Current Law: *The sum of all currents leaving a specific node is equal to the sum of currents entering that node.* By a general agreement, currents leaving a node are marked as positive and currents entering a node are marked as negative. On a deeper level KCL is based on the law of conservation of charge. Systematic use of KCL also leads to nodal analysis, because from Figure 1.26:

$$I_{R1} = \frac{V_{24} - V_{14}}{R_1} \quad I_{R2} = \frac{V_{24}}{R_2} \quad I_{R3} = \frac{V_{24} - V_{34}}{R_3}.$$

Since the actual values of the voltage sources are unknown in this case, it is common practise to make the initial assumption that all the currents are leaving the specific node for which the equations are written (V_{24} in this case). Then as a consequence of KCL, the sum of these leaving currents must equal zero, i.e.

$$\frac{V_{24} - V_{14}}{R_1} + \frac{V_{24}}{R_2} + \frac{V_{24} - V_{34}}{R_3} = 0.$$

The actual resulting node voltage values for all the unknown voltages will correct the direction of the currents automatically, as long as the equations are written with consistency for the chosen signs for leaving and entering currents. By gathering terms with respect to nodal voltages,

$$-V_{14} \frac{1}{R_1} + V_{24} \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right) - V_{34} \frac{1}{R_3} = 0.$$

This equation can be written in matrix form,

$$\begin{bmatrix} -\frac{1}{R_1} & \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right) & -\frac{1}{R_3} \end{bmatrix} \times \begin{bmatrix} V_{14} \\ V_{24} \\ V_{34} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}.$$

Unfortunately this is not that good example of nodal analysis because there is only one node in this simple circuit, and that one node creates that one row to the admittance matrix. Normally the circuit is written directly into a matrix equation, and the unknown node voltages are solved using Cramer's rule, leading

to a determinant division. There will be a lot of practical examples of nodal analysis throughout the text. To give short step-by-step instructions at this phase, the rules for nodal analysis are:

1. Convert all voltage sources to current sources and redraw the circuit using these current sources
2. Select the reference node (usually the ground node) and indicate all other nodes with numbers 1, 2, 3, ..., N
3. From KCL it follows that the matrix representation will have the general form of

$$\begin{bmatrix} Y_{11} & -Y_{12} & \cdots & -Y_{1N} \\ -Y_{21} & Y_{22} & \cdots & -Y_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -Y_{N1} & -Y_{N2} & \cdots & Y_{NN} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}$$

where

- Y_{ii} = sum of all admittances connected to node i
 - Y_{ij} = sum of all admittances connected between nodes i and j
 - I_i = sum of all current sources entering node i , leaving current sources are prefixed with a minus sign
4. From the matrix representation, all needed node voltages can be solved using Cramer's rule

When denoting matrix rows with index i and columns with index j , it follows that the admittance matrix has all entries Y_{ij} prefixed with a minus sign except when $i = j$. This comes from the initial assumption that all currents are leaving from the specific node i . I_i is the sum of converted current source terms entering or leaving node i . Leaving current source terms have a minus sign and entering current sources have a plus sign. This is exactly the opposite of the common agreement stated earlier, but here the terms are only moved to the other side of the equality sign, so that is why the signs are opposite.

In the case of controlled current sources (such as $g_m v_\pi$ in a transistor equivalent circuit), those should be written in terms of node voltage variables (V_1, V_2, \dots) at

the admittance matrix side before solving the matrix equation. There will be lots of examples of this method presented in the following chapters.

Kirchhoff's Voltage Law: *The sum of all voltage drops around a closed loop is zero at all times.* In practise one just goes around a closed path in a circuit (either clockwise or counterclockwise) and sums up all voltages on the way. To be more specific about the voltage drops, there are two rules:

1. a positive voltage drop is in the direction of the current in a resistor (or more generally impedance)
2. a battery has a positive drop in the direction from the + terminal to the – terminal, independent of the direction of the current

With these rules in mind it is possible to solve the branch currents I_{R1} , I_{R2} and I_{R3} in Figure 1.26. Going around the loop specified as I_A in the clockwise direction yields,

$$-V_{14} - I_{R1}R_1 + I_{R2}R_2 = 0,$$

and the second loop in the clockwise direction equates to

$$-I_{R2}R_2 + V_{34} + I_{R3}R_3 = 0.$$

These two equations form a pair of equations with three unknowns. A third equation is needed to solve these equations. That third equation comes from KCL:

$$I_{R1} + I_{R2} + I_{R3} = 0.$$

From the set of these three equations it is possible to solve the unknown branch currents I_{R1} , I_{R2} and I_{R3} . The initial directions of the currents are just thrown in based on a guess. The final result indicates whether the guess had been a successful one.

The KVL method leads to a systematic mesh analysis where mesh currents are solved and the individual branch currents are obtained as sums or differences of the mesh currents. The good thing is that there is no need to make initial guesses on the directions of the branch currents anymore. The following matrix equation can be written from the circuit in Figure 1.26:

$$\begin{bmatrix} R_1 + R_2 & -R_2 \\ -R_2 & R_2 + R_3 \end{bmatrix} \times \begin{bmatrix} I_A \\ I_B \end{bmatrix} = \begin{bmatrix} V_{14} \\ -V_{34} \end{bmatrix}.$$

For basic use of the mesh analysis, the following list gives short step-by-step instructions to create the mesh matrix:

1. Convert all current sources to voltage sources and redraw the circuit using these voltage sources
2. Select a mesh current variable for each loop, e.g. $I_1, I_2, I_3, \dots, I_N$
3. From KVL it follows that the matrix representation will have the general form of

$$\begin{bmatrix} Z_{11} & -Z_{12} & \cdots & -Z_{1N} \\ -Z_{21} & Z_{22} & \cdots & -Z_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -Z_{N1} & -Z_{N2} & \cdots & Z_{NN} \end{bmatrix} \times \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}$$

where

- Z_{ii} = sum of all impedances that the mesh current I_i passes through
 - Z_{ij} = sum of all common impedances for mesh currents I_i and I_j . It is extremely important to note that if the mesh currents pass through the common impedances in opposite directions, then Z_{ij} appears as negative to the matrix. If the mesh currents pass through the common impedances in the same direction, then Z_{ij} appears as positive to the matrix.
 - V_i = sum of all voltage sources in the path of mesh current i . If the voltage source gives a potential rise to the direction of the mesh current, it is taken to the sum as positive. If the voltage source causes a potential drop to the direction of the mesh current, it is taken to the sum as negative.
4. From the matrix representation, all needed mesh currents can be solved using Cramer's rule

In the case of controlled voltage sources, those should be written in terms of mesh current variables (I_1, I_2, \dots) to the impedance matrix before solving the matrix equation.

1.3.7 Properties of semiconductors

Different materials conduct electrical current differently and for that reason materials can be classified into resistors, semiconductors and conductors. To understand some of the practical topics presented in other chapters of this book, it is necessary to briefly explain the scientific background of semiconductors.

Current flow on the atomic level is directly related to the mobility of electrons. A typical structure of a single atom is commonly described with a *Bohr model*, where the electrons of an atom are placed in circular shells around the core of the atom, as shown in Figure 1.27. The physical properties of an atom are mainly dependent on the amount of electrons in the outermost shell. This outer shell is named as the *valence shell*, and the electrons having the highest energy with respect to the core of the atom are called valence electrons. Suitable materials for semiconductor fabrication are silicon and germanium, which both have four valence electrons in their valence shell according to the Bohr model.

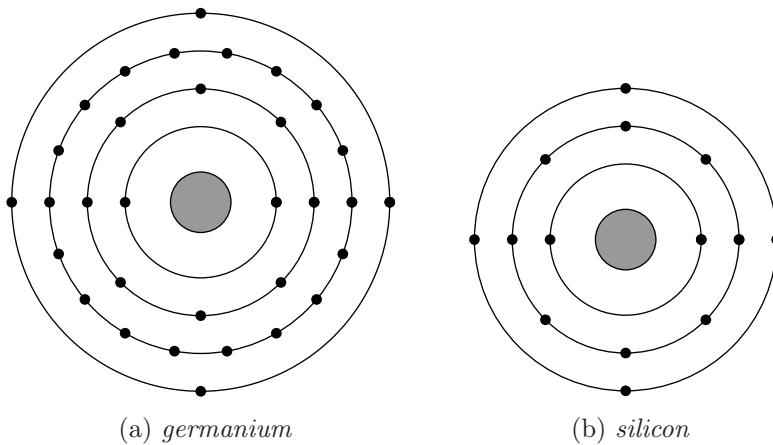


Figure 1.27: Bohr models of germanium and silicon atoms

Especially in metals the atoms of the material align themselves in systematic structures. These structures are called lattices, which most often appear in three basic types; simple cubic, body centred cubic and face centred cubic. For the most common semiconducting materials such as silicon and germanium, the lattice structure is called a diamond lattice, which can be constructed by joining face centred lattices suitably together. The unit cell of a diamond lattice structure is depicted in Figure 1.28. In this lattice the atoms are located in every corner, in the centre of every face and four atoms are inside the cube somewhat symmetrically. [16, pp. 8–11]

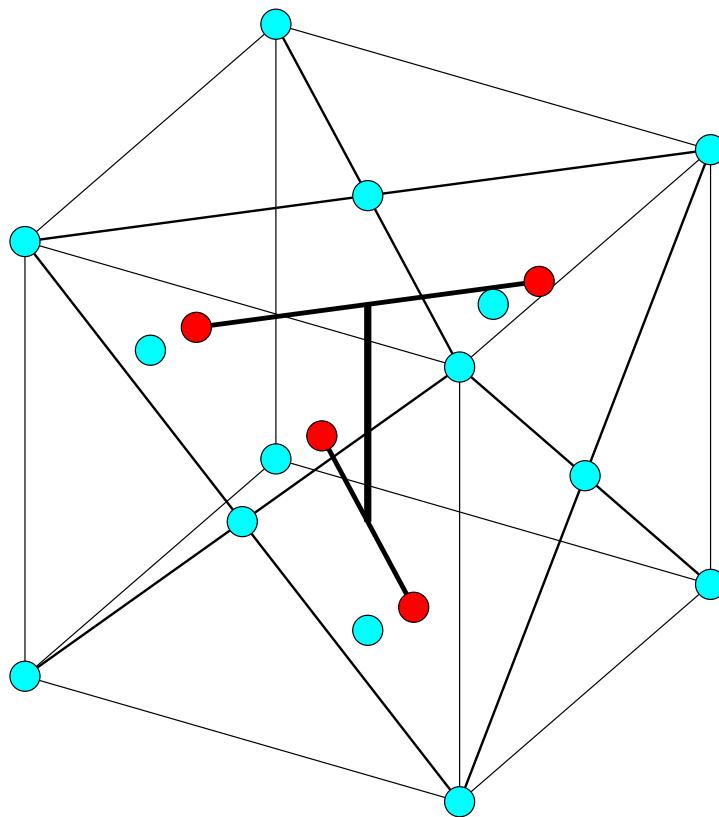


Figure 1.28: A diamond lattice

The Bohr model is a simple model for the electron band (or orbital) structure, but does not reveal how the electrons create the current flow inside a slab of material. In the so-called *Drude model*, the valence electrons are taken to be the conduction electrons by forming an electron gas inside the material. The valence electrons are free to move inside the material, continuously colliding with other electrons. This brings in a stochastic element to the motion of the electrons inside a metal conductor [17, pp. 2–6]. Figure 1.29 shows the idea of the electron gas of valence electrons inside a conductor.

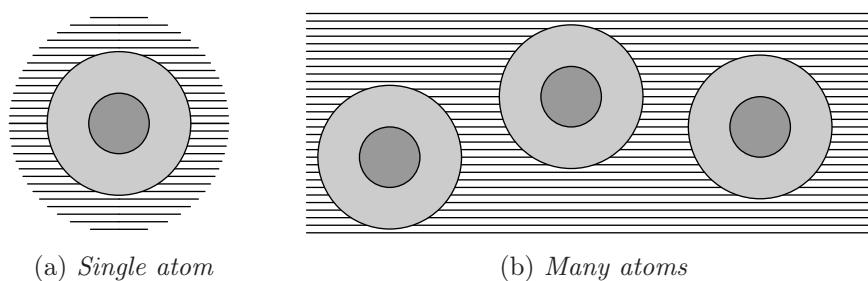


Figure 1.29: Drude theory of electron gas

The Drude model describes pure conductors better than semiconductors, for

which the Bohr model is adequate enough for creating a realistic image of the current flow in the atomic level. When several silicon or germanium atoms are interconnected as a solid piece of material, the four valence electrons form covalent bonds with each other to keep the atom structure firmly together. When the valence electrons are given a certain amount of energy (heat, voltage, work), they are able to dislocate from the valence band of the atom. Outside the valence band the electrons are considered to be in a *conduction band*, where they are free to move around toward other atoms. After travelling randomly for some time in the conduction band, the free electrons relocate back to a valence band of another atom, which has also lost an electron, leaving a dismounting hole for the free electrons. A rapid exchange between free electrons and holes creates the current flow mechanism in a semiconductor.

There exists an energy gap that the valence electrons need to pass from the energy level E_V of the valence band to the energy level E_C of the conduction band. Mathematically the energy gap is simply defined as the difference $E_G = E_C - E_V$. The magnitudes of these energy gaps are material-dependent, being 0.66 eV for germanium and 1.12 eV for silicon [18, p. 72]. Figure 1.30 indicates the differences in energy gaps between insulators, semiconductors and conductors. The

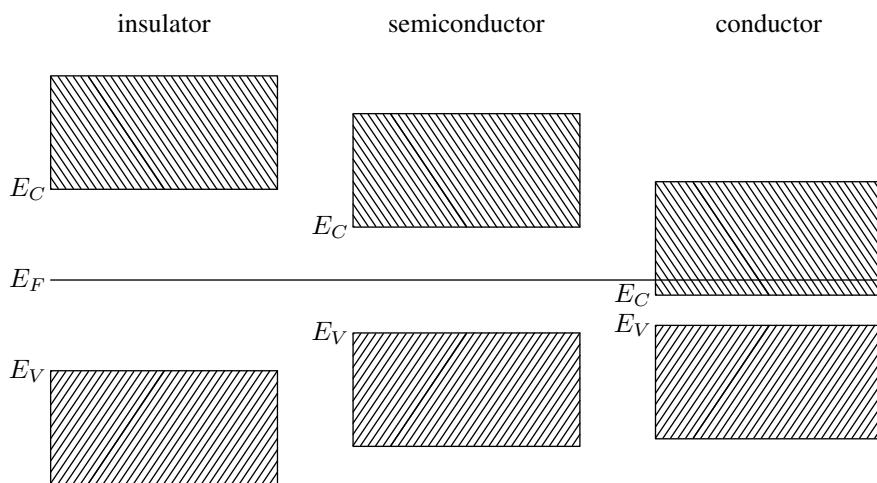


Figure 1.30: Energy gaps for an insulator, a semiconductor and a conductor

Fermi level E_F indicates the highest energy level, which is occupied by electrons with a 50 % probability at any temperature. When temperature is increased from absolute zero (0 kelvins), the probability to have electrons above the Fermi level increases slightly. The further away the Fermi level is from the conduction band, the more unlikely it is that there will be electron exchange between the valence band and the conduction band. For conductors the Fermi level is inside

the conduction band so that current will certainly flow, but for pure (or *intrinsic*) semiconductors the probability is quite low.

As also seen in Figure 1.30, in semiconductors the Fermi level is in the middle of the valence and conduction bands. This is due to the fact that in an electrically neutral semiconductor crystal the free electron concentration and hole concentration are equal. Naturally all the dislocated free electrons occupy the conduction band and all the holes are the empty spaces left in the valence band. Therefore, the valence band concentration N_V indicates holes (positive charge) and the conduction band concentration N_C refers to the free electrons (negative charge).

To increase the conductivity of a semiconductor, impurity atoms are added to the pure semiconductor crystal. The impurity atoms will bring along either one extra electron (atoms with five valence electrons) or one extra hole (atoms with three valence electrons). These add up to the intrinsic concentrations as *donor* N_D or *acceptor* N_A concentrations as shown in Figure 1.31. The impurity atoms will make the pure semiconductor either a negatively charged *n-type* or a positively charged *p-type* semiconductor. The added energy due to the impurity atoms will bring the energy levels closer to the Fermi level as indicated in Figure 1.31.

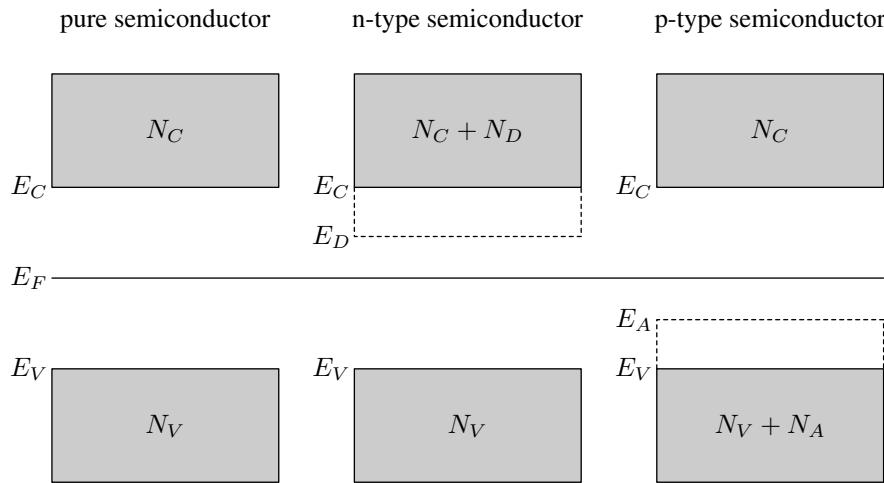


Figure 1.31: Energy gaps for intrinsic, *n*-type and *p*-type semiconductors

Related to the intrinsic conduction and valence band concentrations, the quantities of

$$n = N_C e^{-(E_C - E_F)/kT} \quad \text{and} \quad p = N_V e^{-(E_F - E_V)/kT} \quad (1.77)$$

can be defined to indicate the temperature dependency of the concentrations N_C and N_V . In equation (1.77), k is the Boltzmann's constant and T is the tem-

perature. Additionally, n refers to the negatively charged electrons and p to the positively charged holes. The product

$$np = N_C N_V e^{-E_G/kT} \quad \text{and also} \quad np = n_i^2, \quad (1.78)$$

where n_i refers to the concentration of the intrinsic semiconductor. The concentrations n and p can be assigned with subscripts to clarify the n -type and p -type semiconductors, and still the products

$$n_n p_n = n_p p_p = n_i^2$$

are all equal to the square of the total concentration of the pure semiconductor.

In an n -type material the free electron concentration is approximately equal to the density of the donor atoms. Also, since in the n -type semiconductor $N_A = 0$, the electron and hole concentrations can be evaluated against the concentration of the donor atoms, namely

$$n_n \approx N_D \quad \text{and} \quad p_n = \frac{n_i^2}{N_D}. \quad (1.79)$$

Similarly, in the p -type semiconductor

$$p_p \approx N_A \quad \text{and} \quad n_p = \frac{n_i^2}{N_A}. \quad (1.80)$$

The n - or p -type semiconductors are of no use when treated as separate elements. The magic happens to the energy gap when the combined pn junction is examined. Figure 1.32 explains what happens at the junction when the p -type and n -type substances are brought in direct contact. The most interesting part of Figure 1.32 is the energy term E_0 at the middle of the junction. The mathematical expression for this energy can be evaluated qualitatively from the figure. When the energy terms

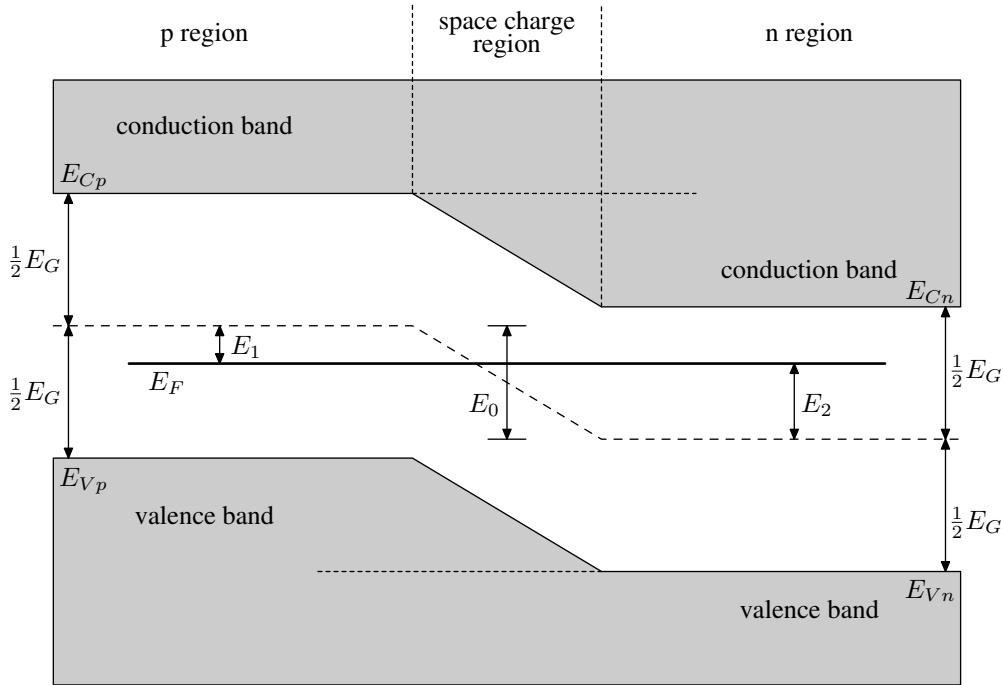
$$E_1 = E_{Vp} - E_F - \frac{1}{2}E_G \quad \text{and} \quad E_2 = -E_{Cn} + E_F - \frac{1}{2}E_G$$

are added together, the energy at the junction

$$E_0 = E_1 + E_2 = E_G - (E_{Cn} - E_F) - (E_F - E_{Vp}).$$

From equation (1.78) one can solve the expression for the energy gap:

$$E_G = kT \ln \frac{N_C N_V}{n_i^2},$$

Figure 1.32: The energy gap of a *pn* junction

and when substituting $n = N_D$ and $p = N_A$ into the equations (1.77)

$$E_{Cn} - E_F = kT \ln \frac{N_C}{N_D} \quad \text{and} \quad E_F - E_{Vp} = kT \ln \frac{N_V}{N_A}.$$

With these intermediate results the expression for the junction energy becomes

$$E_0 = kT \ln \frac{N_D N_A}{n_i^2}.$$

Using the results of equations (1.79) and (1.80), the combined concentrations of the *pn* junction can be used to define E_0 as

$$E_0 = kT \ln \frac{p_{p0}}{p_{n0}} = kT \ln \frac{n_{n0}}{n_{p0}}.$$

Solving for p_{p0} yields

$$p_{p0} = p_{n0} e^{E_0/kT} = p_{n0} e^{V_0/V_T}, \quad (1.81)$$

where V_0 is the voltage over the *pn* junction in the equilibrium state. In this state no external voltages have been applied to the *pn* junction.

If a forward bias voltage is applied over the *pn* junction, holes are injected from the *p* region to the *n* region. The concentration p_n of holes in the *n* region is increased according to equation

$$p_n(x) = p_{n0} + P_n(0) e^{-x/L_p}, \quad (1.82)$$

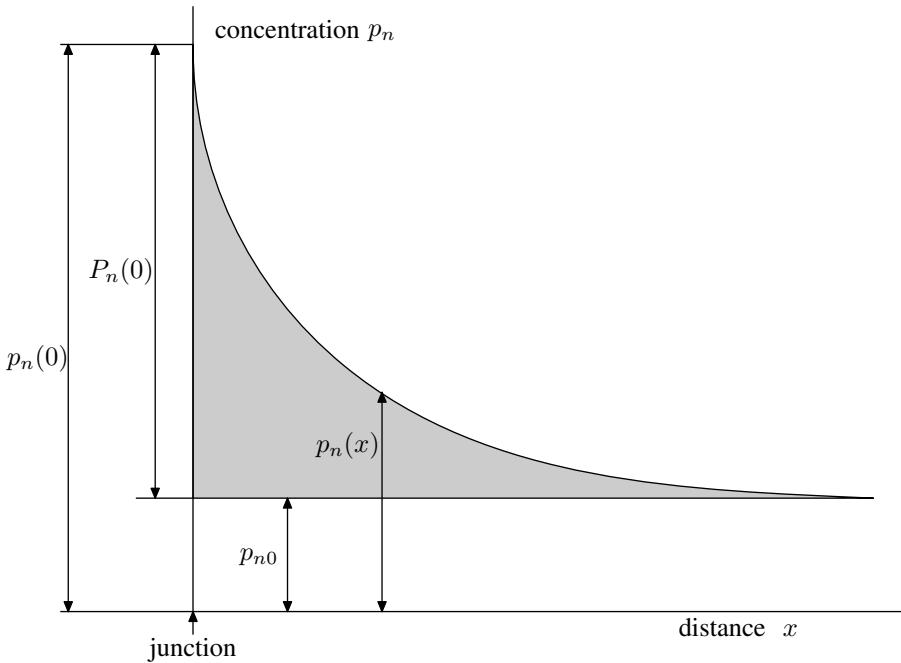


Figure 1.33: Hole concentrations at the pn junction

where the parameter L_p is the hole diffusion length in the n region. Figure 1.33 illustrates the exponential decrease of the concentration density $p_n(x)$ with distance x into the n region. Related to the concentration rates of holes in the n region, the diffusion hole current

$$I_{pn} = -AeD_p \frac{dp_n}{dx} = \frac{AeD_p P_n(0)}{L_p} e^{-x/L_p}. \quad (1.83)$$

If the hole concentrations at the edges of the space-charge region are p_p and p_n in the p and n regions respectively, a total barrier potential V_B existing over the pn junction obeys the law

$$p_p = p_n e^{V_B/V_T},$$

which is the Boltzmann equation of kinetic gas theory. If the junction is biased in the forward direction by applying voltage V , the barrier potential is decreased from the equilibrium value V_0 as $V_B = V_0 - V$. The hole concentration in the p region is in constant equilibrium p_{p0} , but in the n region the hole concentration varies with distance. At the junction $x = 0$, and the Boltzmann relation is

$$p_{p0} = p_n(0) e^{(V_0-V)/V_T}.$$

When this equation is combined with equation (1.81),

$$p_n(0) = p_{n0} e^{V/V_T}. \quad (1.84)$$

This equation is referred to as the law of the junction. It indicates that for a positive forward bias voltage, the hole concentration $p_n(0)$ at the junction is greater than the equilibrium value p_{n0} . The hole concentration $P_n(0)$ injected into the n region is obtained from equations (1.84) and (1.82) as

$$P_n(0) = p_{n0} (e^{V/V_T} - 1). \quad (1.85)$$

The hole current $I_{pn}(0)$ crossing the junction into the n region is given by equation (1.83). Since the current is evaluated directly at the pn junction, the distance x according to Figure 1.33 is $x = 0$, and using equation (1.85) as $P_n(0)$, the currents

$$I_{pn}(0) = \frac{AeD_p p_{n0}}{L_p} (e^{V/V_T} - 1) \quad \text{and} \quad I_{np}(0) = \frac{AeD_n n_{p0}}{L_n} (e^{V/V_T} - 1).$$

The total current flowing across the junction can be written simply as

$$I = I_0 (e^{V/V_T} - 1),$$

where

$$I_0 = \frac{AeD_p p_{n0}}{L_p} + \frac{AeD_n n_{p0}}{L_n}.$$

The derivation explains how the current flow depends exponentially on the voltage V , which is the voltage applied over the pn junction. This dependency between junction voltage and current flow is a fundamental result, which can be applied to describe the basic behaviour of junction currents in semiconductor diodes and transistors. [19, pp. 115–126]

1.3.8 Properties of transistors

Transistors are semiconductor devices, typically made from silicon but earlier also from germanium materials. There exist many different types of transistors. The basic categorisation is usually made between bipolar junction transistors (BJT) and field-effect transistors (FET). The bipolar junction transistors are controlled with input current and the FETs are controlled with input voltage. The basic functionality is still the same in all types of transistors.

Firstly, it is necessary to introduce the symbols used to depict a bipolar junction transistor in a circuit diagram. Figure 1.34 shows the traditional symbols for BJTs of both polarities, npn and pnp . The BJT transistor pins are referenced in the figure as *base* (B), *collector* (C) and *emitter* (E). The currents and voltages in each transistor pin are indicated by a subscript from the pin name. For example,

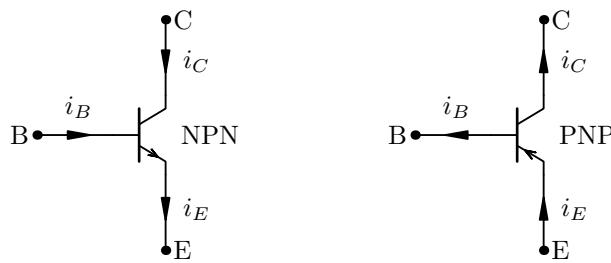


Figure 1.34: Schematic symbols for *npn* and *pnp* bipolar junction transistors

a direct current flowing in the base pin is denoted as I_B and a pure alternating voltage at the collector is written as v_c . Normally the currents in the transistor have both AC and DC components, which is indicated with the notations used in Figure 1.34. The three transistor pins are connected by current equations

$$i_E = i_B + i_C \quad \text{and} \quad i_C = \beta_F i_B,$$

where β_F is the current gain factor of a BJT.

A transistor is used in circuits by *biasing* it in an active operation region by setting suitable DC voltages and currents to all three pins. Voltages and currents in the emitter and collector circuitry are designed to be much higher than on the transistor's base. Weak AC signals applied to the base pin of a transistor get transferred to the higher powered emitter-collector circuitry. When the signal is moving through the base to the emitter-collector circuit, amplification happens as a multiplication process of the original signal with a gain factor. This gain factor is a natural property of transistors and typically it is > 100 by value. Nevertheless, amplification does not come from nowhere, since the DC voltages which enable the voltage and currents levels in the circuit need to be set with external power sources.

A bipolar junction transistor is said to be in *saturation* (fully open) when a maximum amount of current is flowing between the collector and the emitter. When a transistor is saturated, any increase of the base current does not increase the collector-emitter current. Likewise, a BJT is in a *cut-off* state (fully closed) when no current flows between the collector and the emitter. Between these two limiting states the transistor is said to be in an active state.

The gain factor is one of the most important properties of a transistor, so it needs to be explained in detail. The gain factor is divided into the *large-signal* (β), DC (h_{FE} or β_F) and *small-signal* (h_{fe}) values of current gain. The term large-signal refers to currents and voltages near the maximum allowed signal level range that

the transistor can handle. The DC gain parameter is primarily used in biasing analysis, when only direct currents are flowing through the transistor. The small-signal alternative currents as the input to a transistor base pin are typically in the micro or milli prefix range and are truly small as the word suggests.

All of these different types of gain parameters are defined slightly differently, for example, the large-signal gain factor is defined as

$$\beta = \frac{I_C - I_{CB0}}{I_B - (-I_{CB0})},$$

where I_{CB0} is the leakage current between the collector and the base. This type of definition implies that the large-signal current gain is determined for DC signals only.

The DC current gain is officially defined as

$$h_{FE} = \frac{I_C}{I_B} = \beta_F.$$

The notation β_F indicates that this gain parameter refers to a configuration where the transistor is biased in the forward active mode. There also exists the reverse current gain β_R , but it is not commonly used or referred to. The notation β_F is adopted to be used extensively in the following chapters instead of h_{FE} because it seems to be a more commonly used notation in other similar literature. Since the leakage current I_{CB0} is typically very small compared to I_C and I_B , it is normally allowed to approximate that the large-signal gain is the same as the DC gain, i.e. $\beta \approx \beta_F$.

There also exists a relation between the small-signal and large-signal gain factors, namely,

$$h_{fe} = \frac{\partial I_C}{\partial I_B} = \beta + (I_{CB0} + I_B) \frac{\partial \beta}{\partial I_B}.$$

If the approximation is taken that $\beta_F = \beta$, then

$$h_{fe} = \frac{\beta_F}{1 - (I_{CB0} + I_B) \frac{\partial \beta_F}{\partial I_C}}.$$

Generally the DC current gain β_F depends on the biasing setup of the transistor so that changes in I_C will also slightly affect β_F . For a crude approximation it is usually adequate to assume that β_F is constant for all values of I_C . From this assumption it follows that one can take $h_{fe} = \beta_F$. This is because the DC current gain β_F is normally directly given in the transistor datasheets as a single value, and therefore it is easy to also use it for approximating the small-signal

current gain. This approximation is used frequently in this text when equations are derived for pen-and-paper calculations and it should be not taken as the whole truth but only as a crude approximation. [19, pp. 242 – 243].

In the following subsections the focus is on the basics of BJT transistors, since knowledge of the general BJT properties lays a stable foundation for understanding other types of transistors. The design and analysis of transistor circuits consist of two parts. Firstly, the DC levels of the circuit need to be fixed so that the transistor is adequately biased for amplifying AC signals, which are applied to the base pin. As a simple example, let's apply the whole transistor circuit analysis procedure to the basic transistor circuit presented in Figure 1.35.

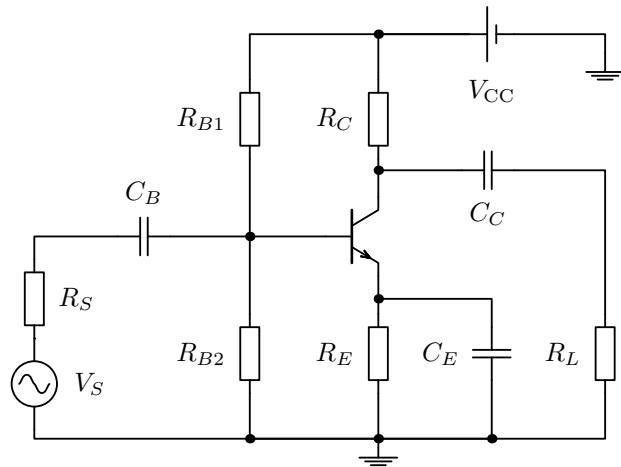


Figure 1.35: A classic amplifier circuit using a single transistor device

1.3.9 Transistor circuit DC analysis

Transistor circuit DC analysis is commonly referred to as *biasing* or calculating the *quiescent values*. DC analysis of transistor circuits is usually considered a much simpler task than AC analysis. This is true to some extent since there are no frequency dependencies in the DC setup, and therefore transistor circuit DC modelling does not have much diversity. On the other hand, it often happens that the systematic circuit analysis methods are not directly applicable when solving DC currents and voltages in transistor circuits.

The general goal of transistor stage DC biasing is to set a stable voltage level to the point where the amplified signal is taken out from the amplifier. This point should be biased in a way that the steady state voltage there is approximately $\frac{V_{CC}}{2}$ if the signal to be amplified is changing symmetrically around the bias point.

The bipolar junction transistor has an equivalent DC circuit presentation, which can be used in place of the actual transistor symbol when doing DC analysis. These so-called *large-signal* models for DC modelling of *npn* and *pnp* transistors are presented in Figure 1.36.

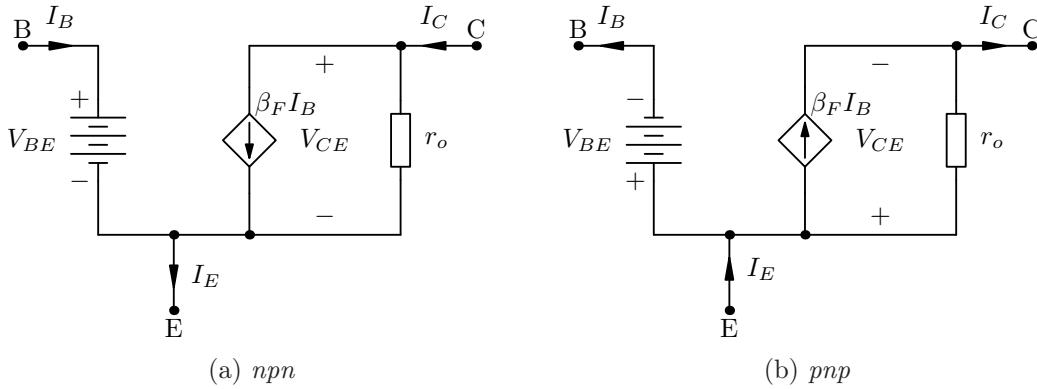


Figure 1.36: DC equivalent circuit models for *npn* and *pnp* BJTs

When DC functionality is concerned, these equivalent models replace the BJT symbols in Figure 1.34. Usually the internal output resistance r_o presented in Figure 1.36 is neglected in pen-and-paper calculations because its large magnitude presents a practical open-terminal situation. Therefore, these models do not contain any internal resistances of transistors, which implies that transistor circuit DC biasing would only depend on external voltages and resistances. In reality this is just an idealised model but it is accurate enough when considering all other inaccuracies in the values of external components. When analysing circuits with computer simulation programs, e.g. SPICE, then all possible internal resistances of the transistor are taken into account in simulations.

However, it is usually not necessary to use these equivalent models in calculations, since basic transistor biasing circuits are not that complex. It is also often better to avoid using systematic circuit analysis methods (e.g. nodal or mesh analysis as they are presented in section 1.3.6) in transistor biasing calculations because source transformations from voltage to current and vice versa are seldom possible. A reasonably systematic method based on the nodal analysis can be however developed and it is presented later in this section

Let's begin the DC analysis on the example circuit presented in Figure 1.35. For tutorial purposes it is better to do the analysis in three different ways, by using the regular transistor symbols (method [a]), using the equivalent circuit models (method [b]) and making use of the principles of the nodal analysis (method [c]).

For DC analysis the circuit in Figure 1.35 is transformed into the circuits shown in Figure 1.37.

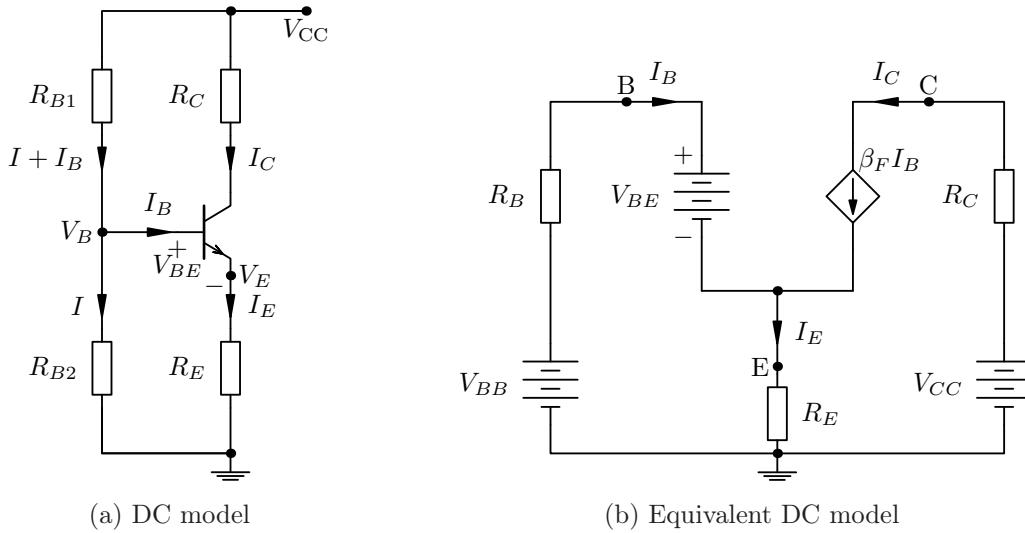


Figure 1.37: Circuit diagrams for DC bias analysis

Method [a]: The example analysis starts from the circuit shown in Figure 1.37a. In this method a suitable set of equations needs to be invented by examining the circuit diagram. At least equations (1.86) – (1.90) based on Kirchhoff's laws can be directly written from schematic 1.37a.

$$I_C = \beta_F I_B \quad (1.86)$$

$$I_E = (\beta_F + 1) I_B \quad (1.87)$$

$$V_{CC} = (I + I_B) R_{B1} + I R_{B2} \quad (1.88)$$

$$V_{CC} = (I + I_B) R_{B1} + V_{BE} + I_E R_E \quad (1.89)$$

$$V_{CC} = I_C R_C + V_{CE} + I_E R_E \quad (1.90)$$

Considering the circuit 1.37a, it might be best to first find an expression for the base current I_B . From there it is possible to calculate voltage V_B at the base pin and proceed from there to solve all the other unknowns. Current I in Figure 1.37a is the current that would normally flow through the resistors R_{B1} and R_{B2} if the transistor would not be connected there. In addition, the base of the transistor needs its current I_B , which is naturally drawn from the power source through the resistor R_{B1} . When current I is solved from equation (1.88),

$$I = \frac{V_{CC} - I_B R_{B1}}{R_{B1} + R_{B2}}.$$

Substituting this expression into (1.89) and using relation (1.87) results in

$$V_{CC} = \frac{V_{CC}R_{B1} - I_B R_{B1}^2}{R_{B1} + R_{B2}} + I_B R_{B1} + V_{BE} + (\beta_F + 1) I_B R_E.$$

A regrouping of terms from one side to the other gives the intermediate form

$$V_{CC} - \frac{V_{CC}R_{B1}}{R_{B1} + R_{B2}} - V_{BE} = I_B \left(R_{B1} + (\beta_F + 1) R_E - \frac{R_{B1}^2}{R_{B1} + R_{B2}} \right).$$

By multiplying the term V_{CC} on left side and the term R_{B1} on right side with $\frac{R_{B1} + R_{B2}}{R_{B1} + R_{B2}}$ (multiplication by 1), the equation is simplified into

$$\frac{V_{CC}R_{B2}}{R_{B1} + R_{B2}} - V_{BE} = I_B \left(\frac{R_{B1}R_{B2}}{R_{B1} + R_{B2}} + (\beta_F + 1) R_E \right),$$

and finally solving for I_B yields

$$I_B = \frac{\frac{V_{CC}}{R_{B1} + R_{B2}} \frac{R_{B2}}{R_{B1} + R_{B2}} - V_{BE}}{\frac{R_{B1}R_{B2}}{R_{B1} + R_{B2}} + (\beta_F + 1) R_E}. \quad (1.91)$$

With the help of I_B one can calculate all the other necessary DC values in the circuit. The goal of the biasing is to set the output DC voltage level to some suitable value. In this case one wants to amplify sinusoidal signals, so the bias voltage on the output should be approximately $V_{CC}/2$ as stated previously. Using I_B one can now determine the voltage V_C at the collector pin of the transistor to be

$$V_C = V_{CC} - I_C R_C = V_{CC} - \beta_F I_B R_C,$$

and the voltage V_{CE} between the collector and the emitter according to equation (1.90) is

$$V_{CE} = V_{CC} - I_C R_C - I_E R_E = V_C - V_E.$$

The collector-emitter voltage V_{CE} has significance in the sense that it can be used to determine whether the transistor is in saturation or not. Values below 0.3 V indicate that the transistor is in saturation, e.g. fully open. Because of this nonzero saturation voltage, the amplified signal cannot fully oscillate between V_{CC} and zero volts.

Method [b]: The second method of transistor DC bias analysis involves the use of the equivalent circuit shown in Figure 1.37b. Here one needs to use a circuit analysis theorem called Thévenin's theorem. According to the theorem, any linear network with respect to a pair of terminals may be replaced by a voltage generator V_{Th} (equal to the open circuit voltage) in series with the resistance R_{Th}

between these terminals. The Thévenin voltage from the base circuitry in Figure 1.35 is calculated as

$$V_{Th} = V_{BB} = \frac{R_{B2}}{R_{B1} + R_{B2}} V_{CC}, \quad (1.92)$$

and the related resistance

$$R_{Th} = R_B = \frac{R_{B1} R_{B2}}{R_{B1} + R_{B2}}, \quad (1.93)$$

with respect to notations used in Figure 1.37b. After determining the Thévenin voltage and resistance, the circuit diagram in Figure 1.37b yields the equation

$$V_{BB} = I_B R_B - V_{BE} + I_E R_E. \quad (1.94)$$

Solving equation (1.94) for I_B gives

$$I_B = \frac{V_{BB} - V_{BE}}{R_B + (\beta_F + 1) R_E}.$$

When substituting the Thévenin equivalent values V_{BB} and R_B with their correspondences from equations (1.92) and (1.93), equality with equation (1.91) has been reached.

Method [c]: In this method one exploits the nodal method to identify the voltage nodes of the circuit and writing the current equations for these nodes. If there are a lot of nodes in the circuit, then the current equations can be written as a matrix equation and the node voltages can be solved systematically using Cramer's rule.

The analysis relies on Figure 1.37a, which indicates the components that affect the static currents and voltages in this BJT amplifier. When replacing the transistor with its equivalent model, one notices that the circuit in Figure 1.37a actually has two voltage nodes, V_B and V_E .

However, since in this case V_B and V_E are related by the *pn*-junction voltage V_{BE} , the circuit has only one distinct voltage node V_B . This means that only one current equation needs to be written for that node and all the bias voltages of the circuit can be solved using that single voltage value. It is important that there are as many independent current equations as there are voltage nodes in the circuit.

According to the Kirchhoff's current rule, the current equation for the voltage node V_B is:

$$I' = I + I_B = I + \frac{I_E}{\beta_F + 1} \quad (1.95)$$

After the current equation is written down, the next step is to express the currents using the supply voltage and the node voltages. The currents appearing in

equation (1.95) can be expressed with respect to the node voltage as:

$$I_E = \frac{V_B - V_{BE}}{R_E} ; \quad I' = \frac{V_{CC} - V_B}{R_{B1}} ; \quad I = \frac{V_B}{R_{B2}}$$

and after substituting these voltage equations to the current equation,

$$\frac{V_{CC} - V_B}{R_{B1}} = \frac{V_B}{R_{B2}} + \frac{V_B - V_{BE}}{R_E(\beta_F + 1)}.$$

From this equation it is possible to solve the node voltage V_B . With the fact that $V_E = V_B - V_{BE}$, one has an equation from where other biasing voltages and currents can be solved.

These examples were given to show that there are several ways to analyse the circuits, and eventually all of them should lead to the same result. It should be also noted that there does not exist any fool-proof step-by-step method for transistor circuit bias analysis. Quite often the equations need to be invented from the schematics and solved in a suitable order to reach the goal of determining the necessary DC voltage levels at each pin of the transistor.

1.3.10 Transistor circuit AC analysis

Just like an equivalent model of the bipolar junction transistor could be presented for DC modelling, there also exists an equivalent transistor model for AC signals. As the DC model is called a large-signal model, the alternating current equivalent is a small-signal model for transistors. This is because the currents to be amplified are typically much smaller than the biasing direct currents of a transistor circuit.

Since the main goal of transistor amplifier circuits is to amplify alternating signals, the AC analysis usually gives the most important information about the properties of the circuit. The result of the AC analysis is an equation that describes frequency dependent behaviour of the circuit. This equation obtained from the small-signal model is the *transfer function* of the circuit, depicting the ratio between the output and input voltages.

Determining a transfer function from AC equivalent models is normally a systematic process but it involves a lot of calculational effort. Usually the nodal method (see section 1.3.6) is chosen as the analysis method, since the transistor circuit equivalent model already contains one current source and the elementary impedances are mostly parallel to each other.

The general transistor small-signal model is depicted in Figure 1.38. The junctions between the positive and negative parts of a semiconductor transistor form

small capacitors (C_μ, C_π) and there are internal resistances (r_b, r_π, r_o) related to all pins.

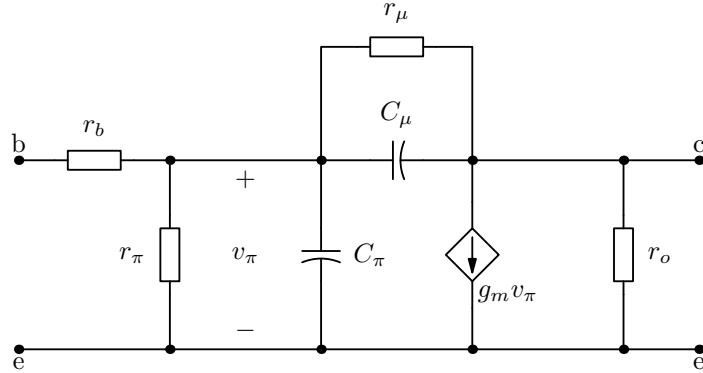


Figure 1.38: The BJT small-signal model for AC analysis

The current-controlled current source (CCCS) in the small-signal model is designated with a parameter g_m , which defines the *transconductance* of the transistor. The resistance r_π together with the base pin resistance r_b form the internal input resistance of a BJT. For the transconductance g_m and the resistance r_π there exists a unique relation,

$$h_{fe} = g_m r_\pi, \quad (1.96)$$

where h_{fe} is the small-signal current gain factor. Equation (1.96) is used very often in pen-and-paper calculations when analysing the small-signal model for some specific transistor circuit. In this book, h_{fe} and β_F are taken as equal, which is not exact, but only a good approximation. To also get numeric results out from the equations, the approximation

$$g_m = \frac{|I_{CQ}|}{V_T} \quad (1.97)$$

is useful. The thermal coefficient $V_T \approx 0.025$ V at room temperature, and I_{CQ} refers to the direct collector current when the transistor is biased to the operating point Q . When the transconductance is evaluated, the current gain factor h_{fe} can be approximated with β_F and equation (1.96) can be used to solve a numerical value for the internal resistance r_π . [15, pp. 414–415]

The general small-signal model in Figure 1.38 needs to be taken into use only if high-frequency amplifiers are designed because the small capacitances only limit the high-frequency response of the model. Audio signals are distinguished as low-frequency signals, and therefore the model in Figure 1.38 can be considerably simplified by neglecting all the capacitances at low frequencies. In the

simplified low-frequency small-signal model, only the internal resistances need to be taken into account when considering audio frequencies. The low-frequency transistor equivalent is drawn in Figure 1.39.

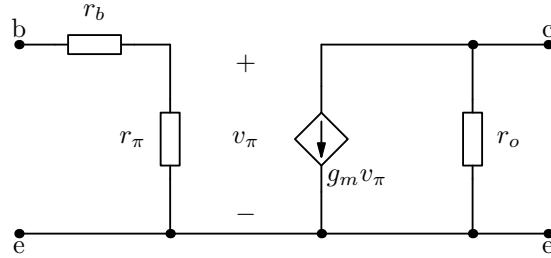


Figure 1.39: A simplified low-frequency small-signal model of a BJT

For pen-and-paper calculations it is suitable to simplify this low-frequency model even further. The only meaningful resistance in the model is r_π , which interacts with the current-controlled current source. The series input resistance r_b is practically very small and the parallel output resistance r_o is relatively large.

The complete small-signal equivalent schematic for the basic transistor amplifier circuit is depicted in Figure 1.40. As the use of the nodal analysis method requires, the input voltage source has been converted to a current source equivalent, where the internal resistance of the source is connected parallel to the source. Moreover, when considering only alternating currents, all DC sources are considered to behave like a ground potential. This is why the resistors R_{B1} and R_C have their other ends connected to the ground level.

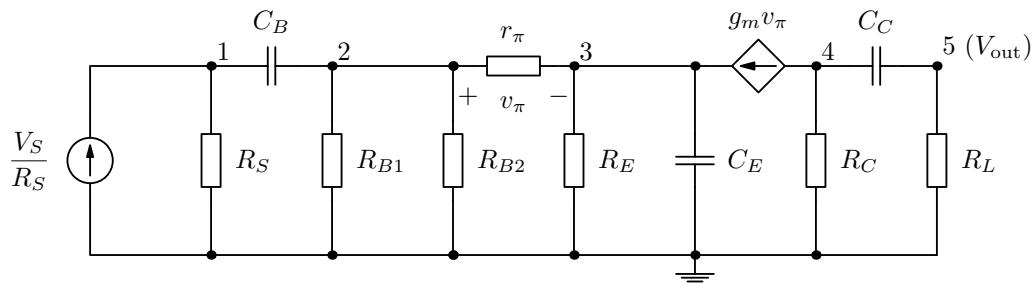


Figure 1.40: A small-signal equivalent of the basic transistor amplifier

The small-signal model in Figure 1.40 is represented by a set of linear equations, where the Laplace transformed impedances are used for capacitors and inductors. By using the systematic method of nodal analysis, the matrix representation (1.98) can be directly written from Figure 1.40.

$$\begin{bmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 \\ 0 & -Y_{32} & Y_{33} & 0 & 0 \\ 0 & 0 & 0 & Y_{44} & -Y_{45} \\ 0 & 0 & 0 & -Y_{54} & Y_{55} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{bmatrix} = \begin{bmatrix} \frac{V_S}{R_S} \\ 0 \\ g_m(V_2 - V_3) \\ -g_m(V_2 - V_3) \\ 0 \end{bmatrix} \quad (1.98)$$

The nonzero elements $Y_{11} \dots Y_{55}$ of the admittance matrix are identified only with their corresponding indices because of limited space. The following listing (1.99) contains the actual terms that should be substituted into the matrix above. Because a sinusoidal input signal is assumed, a direct substitution $s = j\omega$ can be made already at this point of analysis.

$$\begin{aligned} Y_{11} &= \frac{1}{R_S} + j\omega C_B & Y_{12} = Y_{21} &= j\omega C_B \\ Y_{22} &= \frac{1}{r_\pi} + \frac{1}{R_{B1}} + \frac{1}{R_{B2}} + j\omega C_B & Y_{23} = Y_{32} &= \frac{1}{r_\pi} \\ Y_{33} &= \frac{1}{r_\pi} + \frac{1}{R_E} + j\omega C_E & & \\ Y_{44} &= \frac{1}{R_C} + j\omega C_C & Y_{45} = Y_{54} &= j\omega C_C \\ Y_{55} &= \frac{1}{R_L} + j\omega C_C & & \end{aligned} \quad (1.99)$$

Now all the linear equations describing the circuit are in matrix form. There are still terms in the current vector that multiply certain node voltages in the voltage vector. Those terms can (and should) be moved to the admittance matrix before solving the matrix equation. Since matrix equations are generally like normal equations, terms can be added and subtracted from both sides just like in normal equations. The transconductance term g_m needs to be subtracted from the current vector so that those appear at the admittance matrix side.

The positive g_m term in the current vector on row 3 is multiplying voltages V_2 and V_3 . Therefore, this transconductance term should be moved to the admittance matrix on row 3 and to columns 2 and 3 which multiply node voltages V_2 and V_3 respectively in the voltage vector. With some reverse engineering one can verify that the normal matrix equation calculation throws the g_m back to the current vector, and there it is multiplying V_2 and V_3 . From this transformation one has

the modified admittance matrix elements

$$Y'_{32} = -\frac{g_m r_\pi + 1}{r_\pi} \quad Y'_{33} = \frac{g_m r_\pi + 1}{r_\pi} + \frac{1}{R_E} + j\omega C_E.$$

Doing a similar move to the negative g_m on row 4 of the current vector results in two new elements,

$$Y'_{42} = \frac{g_m r_\pi}{r_\pi} \quad Y'_{43} = -\frac{g_m r_\pi}{r_\pi}.$$

In each of the above elements, the product $g_m r_\pi$ is equal to β_F , which is the transistor's current gain factor in the forward-active region. To solve some of the node voltages $V_1 \dots V_5$, one can use the well-known Cramer's rule to solve this system of linear equations systematically. To get a transfer function of the system, the ratio of $V_{\text{out}}/V_{\text{in}}$ needs to be solved. Including the modified admittance matrix elements and using Cramer's rule gives a determinant division,

$$V_5 = V_{\text{out}} = \frac{\begin{vmatrix} Y_{11} & -Y_{12} & 0 & 0 & \frac{V_S}{R_S} \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 \\ 0 & Y'_{32} & Y'_{33} & 0 & 0 \\ 0 & Y'_{42} & Y'_{43} & Y_{44} & 0 \\ 0 & 0 & 0 & -Y_{54} & 0 \end{vmatrix}}{\begin{vmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 \\ 0 & Y'_{32} & Y'_{33} & 0 & 0 \\ 0 & Y'_{42} & Y'_{43} & Y_{44} & -Y_{45} \\ 0 & 0 & 0 & -Y_{54} & Y_{55} \end{vmatrix}}.$$

Notice that to solve the node voltage V_5 , the fifth column of the determinant in the numerator has been replaced with the current vector. This comes from the methodology of the Cramer's rule. Although the determinants have quite many elements, it is still possible to solve this division in symbolic form by writing open the determinants using pen and paper. The benefit of this approach is that from the symbolic transfer function it is possible to analyse what kind of effect a certain component has on the gain properties and the frequency response of the circuit.

To show the complexity of the symbolic expression, the determinant quotient is evaluated and written here explicitly. The numerator expression that comes

directly from the numerator determinant has the form

$$-\beta_F \frac{1}{R_S R_E r_\pi} j\omega C_B j\omega C_C (1 + j\omega C_E R_E),$$

but here all the resistances multiply the denominator, so it is better to also calculate the expression of the denominator and combine the numerator expression with it. Then one has

$$H(j\omega) = \frac{-\beta_F (j\omega)^2 R_L C_B R_C C_C (1 + j\omega C_E R_E)}{(j\omega)^3 a + (j\omega)^2 b + j\omega c + d},$$

where

$$a = r_\pi C_E R_E C_B C_C (R_C + R_L) \left[\left(\frac{1}{R_{B1}} + \frac{1}{R_{B2}} + \frac{1}{r_\pi} \right) R_S + 1 \right]$$

$$\begin{aligned} b = & (\beta_F + 1) R_E \left(\frac{1}{R_{B1}} + \frac{1}{R_{B2}} \right) C_B R_S C_C (R_C + R_L) + \\ & (\beta_F + 1) R_E C_B C_C (R_C + R_L) + r_\pi \left(\frac{1}{R_{B1}} + \frac{1}{R_{B2}} + \frac{1}{r_\pi} \right) \times \\ & [C_E R_E C_C (R_C + R_L) + C_B R_S C_C (R_C + R_L) + C_E R_E C_B R_S] + \\ & r_\pi [C_B C_C (R_C + R_L) + C_B C_E R_E] \end{aligned}$$

$$\begin{aligned} c = & (\beta_F + 1) R_E \left[\left(\frac{1}{R_{B1}} + \frac{1}{R_{B2}} \right) C_B R_S C_C (R_C + R_L) + C_B \right] + \\ & r_\pi \left(\frac{1}{R_{B1}} + \frac{1}{R_{B2}} + \frac{1}{r_\pi} \right) [C_E R_E + C_B R_S + C_C (R_C + R_L)] + r_\pi C_B \end{aligned}$$

$$d = [(\beta_F + 1) R_E + r_\pi] \left(\frac{1}{R_{B1}} + \frac{1}{R_{B2}} \right) + 1.$$

From this symbolic transfer function it can be deduced that it is possible to increase the gain of the example amplifier by increasing the values of resistors R_C and R_L . Time constants as combinations of certain pairs of R and C can also be identified. To draw the actual frequency response curve of the example amplifier, the transfer function can be analysed numerically using Octave or Matlab. The numerical results are presented in the next section, where they are compared to results obtained by using a real circuit simulator program.

1.3.11 Input and output impedances

The input and output impedances are easily determined for a network of purely passive components, since the sum of the impedances seen between the input and output terminals gives the correct value directly. In transistor circuits difficulties arise because the controlled sources are part of the impedance expression.

A systematic way to find the input and output impedances of any network is to use a test voltage source, which is placed between the terminals where the impedance is 'measured'. During the 'measurement', all other voltage sources in the network are short-circuited and current sources open-circuited (any controlled source remains as is). The impedance is evaluated as the ratio of the test voltage and the current that flows through the test source. From the configuration shown in Figure 1.41, the input impedance is obtained as

$$Z_{\text{in}} = \frac{V_{\text{test}}}{I_1},$$

where I_1 is the mesh current obtained by mesh analysis. Similarly from Figure 1.42,

$$Z_{\text{out}} = \frac{V_{\text{test}}}{I_4}.$$

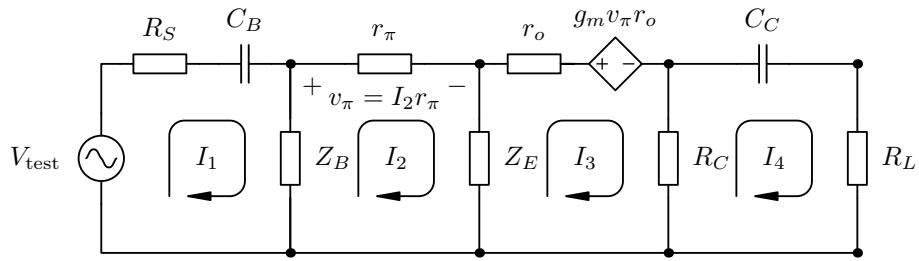


Figure 1.41: A circuit for determining transistor circuit input impedance

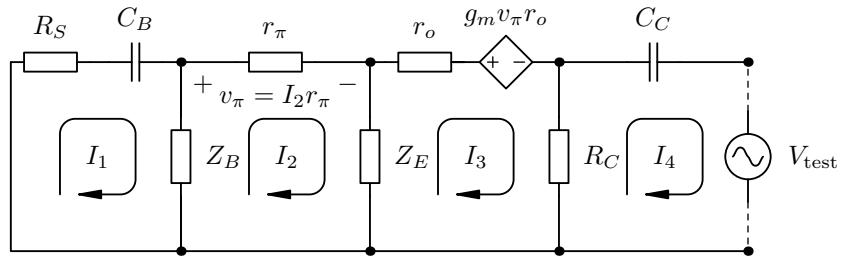


Figure 1.42: A circuit for determining transistor circuit output impedance

Another way to evaluate impedances within circuit networks is to use the Thévenin and Norton theorems, which state that *impedance equals open-circuit voltage*

divided by short-circuit current. To apply this to the transistor circuit output impedance, the open-circuit voltage is typically equal to the output voltage obtained using nodal analysis and the short-circuit current can be evaluated by drawing a short-circuit current loop I_5 after the resistor R_L in Figure 1.41 and then solving the value of the mesh current I_5 .

As a concrete example, the matrix equation for the output impedance circuit in Figure 1.42 can be written from the figure to the form:

$$\begin{bmatrix} R_S + \frac{1}{sC_B} + Z_B & -Z_B & 0 & 0 \\ -Z_B & Z_B + r_\pi + Z_E & -Z_E & 0 \\ 0 & -Z_E & Z_E + r_o + R_C & -R_C \\ 0 & 0 & -R_C & R_C + \frac{1}{sC_C} \end{bmatrix} \times \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -g_m v_\pi r_o \\ -V_{\text{test}} \end{bmatrix}$$

This matrix equation follows directly from the given rules of mesh analysis, which is covered in section 1.3.6. The term $g_m v_\pi r_o$ equals $g_m r_\pi I_2$ as indicated in Figure 1.42. To solve the matrix equation using Cramer's rule, this term should be moved to the impedance matrix on row 3, column 2. To find the output impedance, solve the matrix equation for current I_4 and then solve $\frac{V_{\text{test}}}{I_4}$. Draw the absolute value of this result as a function of frequency to find out the output impedance of the circuit at different frequencies. As a result of numeric evaluation one should notice that in this specific circuit the output impedance at the pass-band approximately equals the value of R_C .

1.3.12 Circuit simulations with SPICE

SPICE comes from the words "Simulation Program with Integrated Circuit Emphasis". Just like the previous chapter described how to form a matrix equation from a circuit, SPICE is designed to do this automatically and solve all the node voltages in an AC, DC and transient simulation mode. This section briefly describes the basics of SPICE using `ngspice` as an example SPICE application. Needless to say, there are several SPICE variations to choose from but the basic idea in all of them is similar. The differences between circuit simulator programs come from the mathematical models used to depict active circuit elements. The used command-set also differs slightly between SPICE versions.

SPICE component models are typically linked with some schematic editor, such as

`gschem` from the gEDA package for Linux. The circuit to be analysed is described to SPICE with a file called *netlist*, which can often be created directly from the schematic editor program. The netlist file is usually a normal text file, but binary files can also be used as a source for simulations. The netlist file contains component names and the node id-tags where the components are connected. SPICE-specific device models and user-defined sub-circuits can also be included to the netlist file to describe the behaviour of more complex components such as transistors and operational amplifiers.

The quiescent currents and voltages of a basic transistor circuit were calculated in section 1.3.9. The equivalent DC simulation mode in SPICE is the operating point (OP) analysis. The SPICE mode for the frequency response calculations presented in section 1.3.10 is naturally AC analysis. SPICE transient analysis is equivalent to solving the set of differential equations which describe the circuit in the time domain. This is the same as taking the Inverse Laplace transform from the *s*-domain model of the circuit.

As an example, a SPICE model for the basic transistor circuit in Figure 1.35 is created. The schematic was first drawn with `gschem` to semi-automate the netlist creation. A valid circuit diagram shown in Figure 1.43 can be extracted directly from the schematic editor. In the `gschem` program, one can define basic

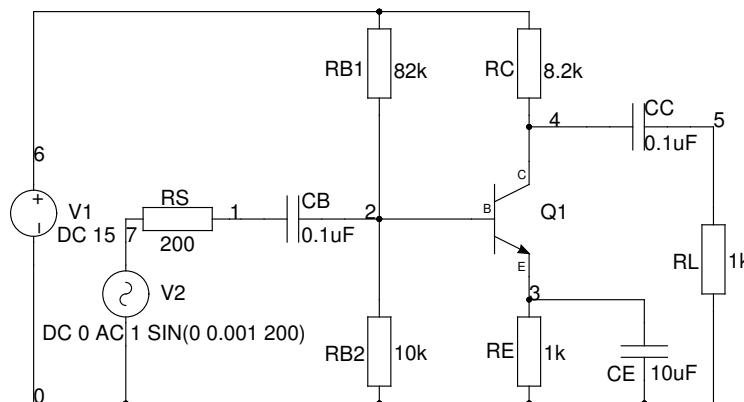


Figure 1.43: A `gschem` model of the classic single-transistor amplifier circuit

components such as resistors, capacitors, inductors and simple voltage sources with a model, where only the 'refdes' and 'value' component parameters need to be defined. An example of the `gschem` parameter definitions for a basic capacitor is shown in Figure 1.44.

Name	Value
device	CAPACITOR
refdes	CB
symversion	0.1
value	0.1uF

Figure 1.44: Defining essential capacitor values in gschem

To be able to obtain the correct transistor models directly to the semi-automatically created netlist file, the SPICE-specific model parameters need to be defined to the gschem component model. This is done by defining a specific 'model-name' and writing the SPICE-specific model parameters into the 'model' field. An example of the parameter definitions for a basic *nPN* transistor is given in Figure 1.45.

Name	Value
device	NPN_TRANSISTOR
refdes	Q1
model-name	NPN1
model	$I_S=1.0e-15$ $B_F=181$

Figure 1.45: Defining essential transistor SPICE parameters in gschem

If the netlist file is created directly from the schematic editor, one must specify names for each individual voltage node in the circuit because SPICE uses those 'netnames' when evaluating the node voltages. When using the gEDA electronics design package in Linux, it is possible to create the netlist file from the gschem schematic. By using the gnetlist application, the SPICE compatible netlist file can be created with the command

```
gnetlist -g spice-sdb -o amp.net amp.sch
```

This command creates a netlist file `amp.net` from the gschem schematic `amp.sch`. The created netlist is as follows:

```
* gnetlist -g spice-sdb -o amp.net amp.sch
*****
* Spice file generated by gnetlist
* spice-sdb version 4.28.2007 by SDB --
* provides advanced spice netlisting capability.
```

```

* Documentation at http://www.brorson.com/gEDA/SPICE/ *
*****
===== Begin SPICE netlist of main design =====
V2 7 0 DC 0 AC 1 SIN(0 0.001 2000)
RL 0 5 1k
V1 6 0 DC 15
Q1 4 2 3 NPN1
.MODEL NPN1 NPN (Is=1.0e-15 Bf=181)
CE 0 3 10uF
CC 4 5 0.1uF
CB 1 2 0.1uF
RS 1 7 200
RB2 0 2 10k
RE 0 3 1k
RC 4 6 8.2k
RB1 2 6 82k
.end

```

This is the basic form of the netlist given by the application. The actual simulations can be specified when running `ngspice` interactively from the command line, but for automation purposes it is better to add the simulation parameters directly to the netlist file. The author prefers to use the `.control` keyword inside the netlist file to specify the used simulation parameters. Assuming that one wants to examine how different values of R_C affect the amplification, the following control sequence can be added to the beginning of the netlist file:

```

*****
.control
ac dec 90 10 100K
alter RC 1k
ac dec 90 10 100K
alter RC 20k
ac dec 90 10 100K
set filetype=ascii
write ampdata.txt db(ac1.v(5)) db(ac2.v(5)) db(ac3.v(5))
gnuplot ampplot db(ac1.v(5)) db(ac2.v(5)) db(ac3.v(5))
alter RC 8.2k
.endc

.OP
===== Begin SPICE netlist of main design =====

```

When the control parameters are ready and saved in the netlist file, the `ngspice` simulation is then run with the command:

```
ngspice -b amp.net
```

This command runs `ngspice` in batch mode, where the result of `.OP` command is printed to the screen and the results of the control sequence are stored in the file `ampdata.txt`. Files for `gnuplot` are also generated, but the `gnuplot` interface in

ngspice version 20 does not seem to be fully functional in terms of generating the plot figure automatically (later versions work fine). Nevertheless, ngspice manages to generate nice skeleton files for plotting with gnuplot.

Concerning the DC analysis, the following print-out is taken directly from the operating point (.OP) analysis results provided by ngspice. It lists the node voltages and all the basic transistor bias values of the static configuration:

```
No. of Data Rows : 1

Node          Voltage
-----
-----
V(1)          0.000000e+00
V(3)          8.764146e-01
V(2)          1.587514e+00
V(4)          7.852887e+00
V(6)          1.500000e+01
V(5)          0.000000e+00
V(7)          0.000000e+00

Source Current
-----
v2#branch     0.000000e+00
v1#branch     -1.03517e-03

BJT: Bipolar Junction Transistor
    device      q1
    model       npn1
    ic          0.000871611
    ib          4.81553e-06
    ie          -0.000876427
    vbe         0.7111
    vbc         -6.26537
    gm          0.0336995
    gpi         0.000186185
    gmu         1e-12
    gx          0
    go          4.52804e-23
    cpi         0
    cmu         0
    cbx         0
    ccs         0
```

The AC simulation results are obtained in the separate gnuplot data file named as ampplot.data as requested in the netlist file. The simulated frequency response curve can be compared with the results of the numerical evaluation of the pen-and-paper calculation shown in the previous section. When comparing the results from the manual calculations and the SPICE simulations, the differences are generally quite small, but the deficiencies of the manual calculation

are revealed when the frequency response is evaluated for several values of the collector resistor R_C . Figure 1.46 shows that a very high value of the collector resistor biases the transistor wrongly and the resulting gain is smaller than expected. The manual evaluation method does not directly reveal this caveat in the frequency response analysis, although the bias calculations need to be done to find out the value of the transistor's internal resistance r_π .

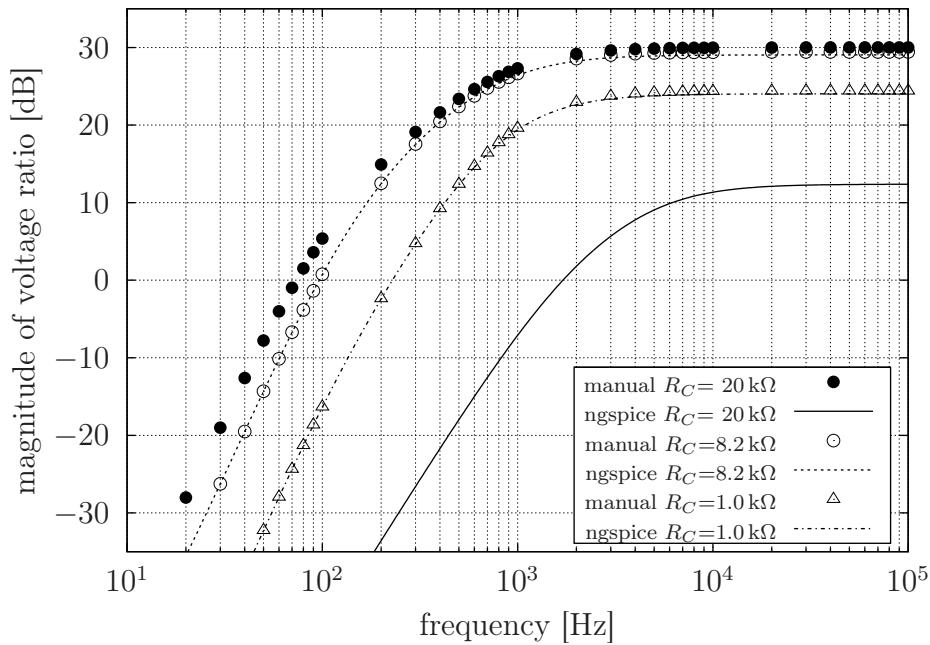


Figure 1.46: Comparison of manual results versus simulation results

What is not that simple by manual analysis methods is that ngspice can create time dependent transient simulation results using 'genuine' test signals, such as a plain sinusoid, square or sawtooth waves. Just to prove that the circuit is actually an amplifier, let's generate a sine wave of 1 mV amplitude and 200 Hz as frequency, then plot the input signal and the output signal to the same plot. For running the transient analysis in ngspice, type the command `tran 1us 30ms` when the netlist is loaded and the ngspice command interpreter is active. This command can also be added to the `.control` section of the netlist file.

From the resulting Figure 1.47, it can be seen that with 200 Hz $V_{\text{out}} \approx 4V_{\text{in}}$. This is in balance with Figure 1.46, where 200 Hz should have 12 dB amplification. As a reminder, a 6 dB increase doubles the voltage amplitude, so that

$$V_{\text{in}} + 6\text{dB} + 6\text{dB} = V_{\text{in}} \cdot 2 \cdot 2 = 4V_{\text{in}}.$$

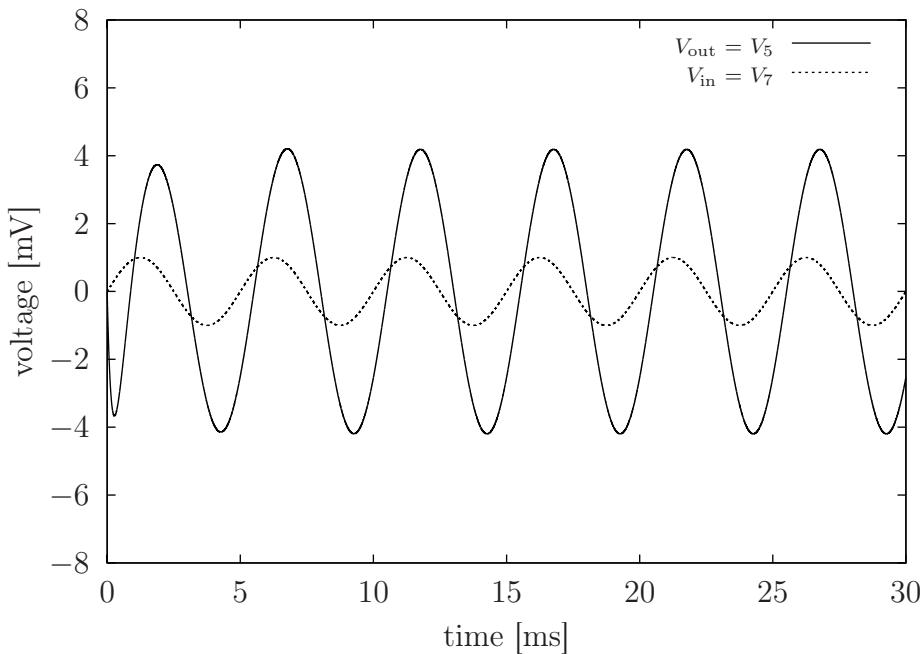


Figure 1.47: Input signal of 1 mV and 200 Hz on the basic amplifier

The output impedance can be simulated by SPICE using the configuration indicated in Figure 1.48.

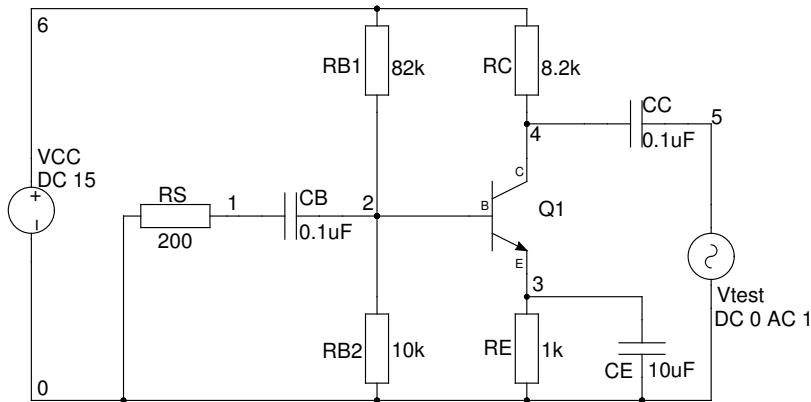


Figure 1.48: Simulating output impedance with gschem

The control section in the simulation file should include the expression to evaluate the output impedance. One example to handle this is:

```
.control
ac dec 90 100 100K
gnuplot outimp abs(ac1.v(5)/ac1.i(Vtest))
.endc
```

CHAPTER 2

VIBES

If there is sound, there is also vibration. A guitar string is one of the most simplest vibrating systems in the world. It might be difficult to imagine, but a piano string would be a more complex system because of the hammer mechanism to set the piano string into motion. One of the most important equations of physics, the one-dimensional wave equation, can be derived from the geometry of a deflected guitar string. The long-haired guitar hero next door, with pants so tight that they stop blood circulation, has surely not realised what his sweaty hands are actually tickling on... Hardcore physics in action!

2.1 A MATHEMATICAL MODEL OF A GUITAR STRING

2.1.1 Setting up the coordinates

Before starting the actual derivation of equations, a decision needs to be made on the dimensions and coordinate system used. It is safest to choose the traditional coordinates, where x refers to the horizontal direction, y denotes vertical values and z brings along the third dimension that expands to orthogonal directions from the plane formed by x and y .

As shown in Figure 2.1, the guitar string is supported from origin ($x = 0$) and from length L . Force \vec{F} is pulling the string as it is normally tuned to have some tension T . The force applied to the string makes the string stretch to length $L + \Delta L$, but the effective length of the string still remains L , since the assumption is made that the supports Δ clamping the string at 0 and L have zero admittance. This means that the string is firmly attached to the supports from both ends, so that no vibrational energy can escape outside the supports and that the vibratory motion of the string is restricted between $0 \leq x \leq L$.



Figure 2.1: The coordinate system used

Equations to model the motion of a guitar string can be mathematically constructed in several ways. A normal guitar string has many nonlinear properties, which are often excluded from the differential equations of motion. These nonlinear properties arise mainly from the stiffness of metal strings, friction against the surrounding air molecules, movement of the end supports and change of tension during vibration. If all of these properties were included in one single differential equation of motion, it would be impossible to find an exact solution to the equation. For this reason, the general approach is to first define a simple model and examine the effect of the nonlinear properties one by one against the basic equation of motion.

As an example, one theoretical model of a guitar string can be presented as a relation between mechanics and electrics. Equivalences in differential equations of mechanical and electrical vibrating systems have evolved into general analogies between mechanical and electrical components. Figure 2.2 makes use of these analogies and depicts an electromechanical model for a string that is set into motion by 'plucking' it with a finger [20, p. 117]. As there are equivalent circuits for electric components, there also exists an equivalent electric circuit of a mechanical string. The quantity β in Figure 2.2 refers to mechanical resistance, m is the mechanical mass and the term $1/k$ is the mechanical compliance written in terms of the spring constant k .

Furthermore in Figure 2.2, the force F that creates the initial displacement for the string is stored as capacitive energy (potential energy) in the string. When the potential energy is released by plucking the string, the switch in the equivalent circuit disconnects the DC source and lets the energy vibrate in the RLC circuitry. This model assumes that the string is built as a sequence of infinitesimal pieces of mechanical vibrators (mass-spring-damper systems), that all have equivalences in terms of inductance, capacitance and resistance. The model also has external

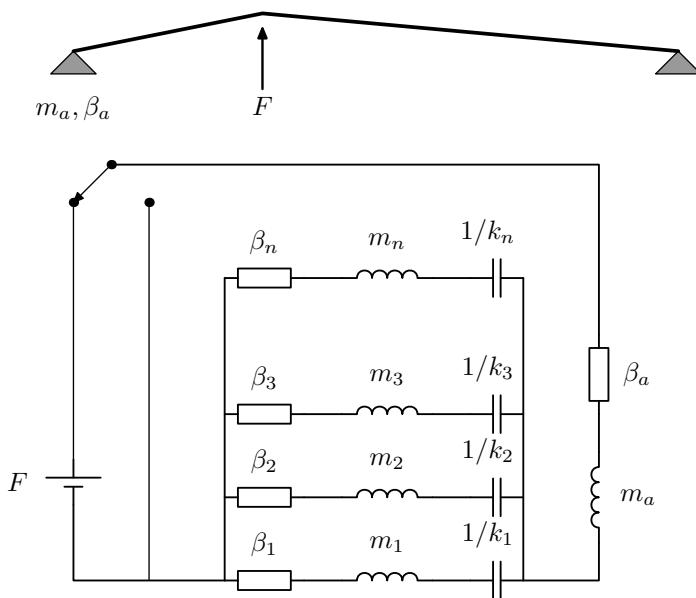


Figure 2.2: An electromechanical equivalence model for a plucked string

energy loss built in terms of air resistance β_a and internal energy loss in terms of β_n , and therefore this is not an adequate approach for *simple* mathematical modelling.

The so-called 'ideal string' model is a linearised standard textbook approach to study the basic physics of the vibrating motion of a string. This model is reduced to be as simple as possible and it is a good starting point for taking the analysis further to investigate the nonlinear behaviour of string motion in guitars and in general.

2.1.2 The ideal string model

The coordinate system in Figure 2.1 and the impedance requirements for the end supports of the string are the first restrictions (or boundary conditions) towards the definition of an idealised mathematical model to describe the motion of a guitar string. The general concept of an ideal string also includes other restrictions that simplify the mathematical modelling of the string as much as possible. The ideal model assumes the following:

1. the string is perfectly flexible (no stiffness) with no internal energy loss
2. the end supports of the string are perfectly rigid and fixed, so that no energy loss happens at the boundaries
3. force F_x (see Fig. 2.1) acting on the string is constant at all times

4. deflection of the string during vibration must be small so that the changes in the vertical force component F_y do not change the tension created by F_x
5. vibratory motion is restricted in the coordinate plane $\{x,y\}$, meaning that no elliptical vibration trajectories are possible
6. the mass of the string is evenly distributed along the length of the string, i.e. the string is homogeneous
7. air friction and effects of gravity are neglected

Considering the number of items on this list, one might think that the ideal model does not represent the real world situation at all. The ultimate question that needs to be answered is concerned with the correlation between a real world string and this idealised model; how accurately does the idealised model simulate a real string, and what are the key factors that make the idealised model fail. Or does it fail after all?

The vibratory motion of a string consists of mechanical waves moving on the string. These waves are set to motion by causing an initial displacement at some point along the string and then releasing the string into motion. An initial displacement at some specific point of the string interacts with the whole length of the string at all times, thereby keeping the string in motion. The frequency of vibration is directly proportional to the velocity of the waves travelling along the string.

Based on the results derived from the ideal model of the string, elementary physics literature [3] presents a mathematical formula to calculate the *harmonic frequencies* (or *upper partials*) of transverse vibrations (deflections in the y -coordinate direction) on a string. The equation for evaluating the harmonic frequencies is

$$f_{Tn} = \frac{nc_T}{2L} = \frac{n}{2L} \sqrt{\frac{F_x}{\rho A}}, \quad (2.1)$$

where the subindex n counts through integer values $[1, 2, 3, \dots, \infty]$ denoting the different harmonic upper partials which are mixed into the motion of the string. The variables that are required to obtain concrete results from equation (2.1) are the length L of the string, the force F_x pulling the string in the x -direction and the physical dimensions of the string; one needs to have the cross-sectional area A and density ρ of the string to define a mass for the string. The wave velocity c_T of the transverse waves moving on the string is then mainly defined by the dimensions of the string.

Although not apparent from the specifications of the ideal model, the string also vibrates in the direction parallel to the string (deflections in the x -coordinate direction). These waves are called longitudinal waves. To depict the harmonic nature of the longitudinal vibration mode, there is a similar equation for longitudinal frequencies,

$$f_{Ln} = \frac{nc_L}{2L} = \frac{n}{2L} \sqrt{\frac{E}{\rho}}, \quad (2.2)$$

corresponding to equation (2.1). Hooke's law is used in equation (2.2) to replace the quotient F/A with Young's modulus E . The longitudinal waves on the string are possible only if the tension of the string is allowed to change during vibration. However, this change of tension is neglected in the ideal model. The basis for the existence of longitudinal waves is explained later when the nonlinear properties of the string are studied. [21, p. 57]

To clarify further the concepts of transverse and longitudinal vibrations, Figure 2.3 depicts the idea and shows the actual vibrational directions and the directions of the wave motion.

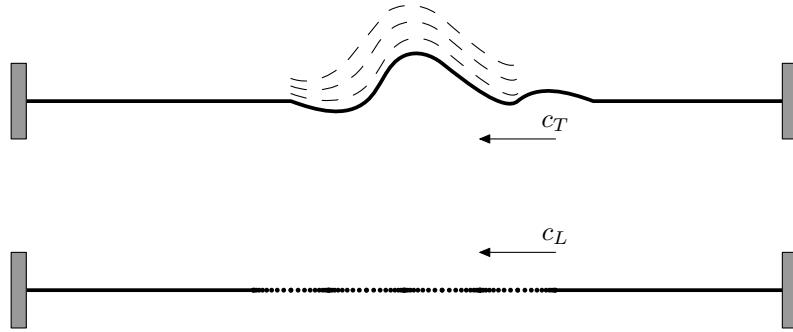


Figure 2.3: Transverse and longitudinal waves travelling on a string

It should be noted from Figure 2.3 that the wave velocities c_T and c_L of equations (2.1) and (2.2) are both parallel to the direction of the string and depend on the physical properties and dimensions of the string. This should be interpreted so that a wave is bouncing back and forth along the string, consisting of either transverse or longitudinal displacement from equilibrium. It is also intuitive to imagine that both modes of vibration can exist at the same wave and that those vibrational modes are connected to each other's motion. Later sections will indicate how the coupling between transverse and longitudinal motion is seen in practise.

When the waves meet the rigidly fixed end supports, they reflect back to the direction where they were coming from. If the supports at both ends of the string were not ideal (mechanical admittance > 0), then energy would escape at the points of reflection and the non-ideal reflection would alter the form of the wave travelling in the string.

When the end supports are ideal (mechanical admittance = 0), the form of the wave stays constant and the reflections of the waves create so-called *standing waves* along the string's length for all harmonic frequencies. Standing waves occur if there is a 180 degree phase shift at the point of reflection. This phenomenon is described in Figure 2.4, where some selected harmonic modes of vibration are drawn.

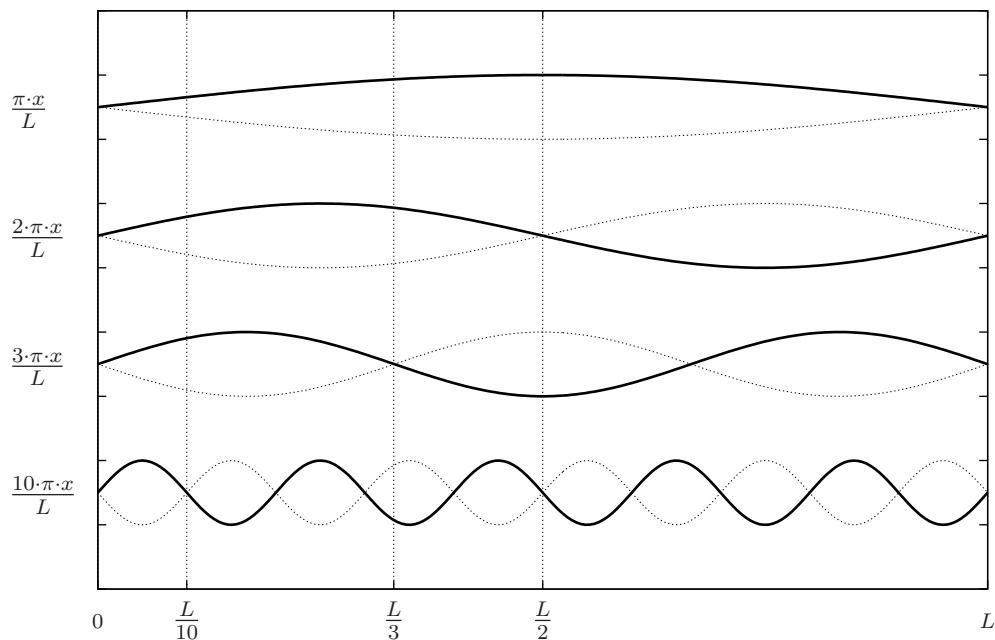


Figure 2.4: Standing wave modes in ideal string vibration

The waves drawn with a dashed line in Figure 2.4 represent the wave that has been reflected from the rigid end support of the string. The points on the string that do not vibrate (the points where the reflection crosses the wave) due to the standing wave phenomenon are called *nodes*. The maximum amplitude points of standing waves are called *antinodes*. These node points divide the string into equally spaced portions, and subsets of these portions are related to the harmonic frequencies of the string, as described by equation (2.1). With respect to the plucking position, these nodes and antinodes have great importance to the magnitudes of the upper partials.

2.1.3 Setting up the equations

Usually some boundaries need to be set before starting to prepare a model for some practical system. If the primary assumptions are well defined, then it is quite straightforward to start the modelling process. The starting point for the following analysis is the model of the ideal string described in section 2.1.2, which is later extended to include some of the nonlinear properties as well.

Mathematical modelling of dynamic mechanical systems usually relies on the basic equation of motion,

$$\vec{F} = m\vec{a}, \quad (2.3)$$

which is also known as Newton's second law of motion. As a special case when the force acts directly in only one coordinate direction, the equation of motion can be written in a non-vectorised form,

$$F = ma, \quad (2.4)$$

where the direction of the force F is usually chosen to be along the x -coordinate direction.

When a mechanical system is at rest, the net force acting on the system is zero, which means that all forces are in equilibrium state. To set the system into motion, force differences causing a non-zero net force need to be applied to the system. As an example, a net force of

$$\vec{F} = \vec{F}_{(x+dx)} - \vec{F}_{(x)} \quad (2.5)$$

between some points x and $x + dx$ of the system describes a situation that is caused usually by some external force, which breaks the balance of the forces in equilibrium. This break from the equilibrium state is exactly what happens when a guitar string is plucked.

When the string is at rest, the system looks like as drawn in Figure 2.1, where the force \vec{F} keeping a steady tension in the string is parallel to the x -coordinate direction, i.e. $\vec{F} = F_x$. To set the string into motion by means of plucking the string, the initial deflection given to the string in the y -coordinate direction results in the triangular form shown in Figure 2.5.

This triangle can be formed mathematically by joining together two straight lines with opposite slopes. The notations a and p in Figure 2.5 refer to the amplitude of the deflection and its position along the string of length L respectively. The mathematical expression $g(x)$ shown in Figure 2.5 is a piecewise continuous function,

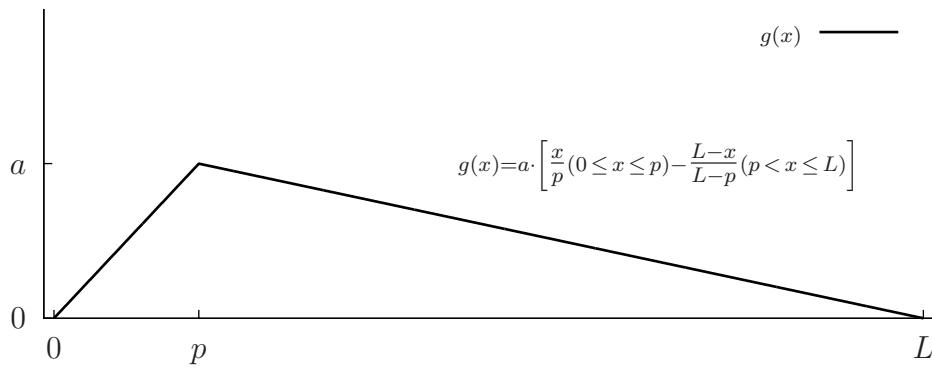


Figure 2.5: The initial triangular form of a plucked string

and with respect to the notations used, it can be used to draw the form of the string as a function of x with the fixed values of a , p and L [22]. This function is needed later to derive an equation of the time independent model of the string.

To set the string in motion, external force is required to create localised force deviations along the string. With the initial deflection shown in Figure 2.5, the total force acting on the string is $\vec{F} = F_x \hat{i} + F_y \hat{j}$, where the F_y component comes from the external force that generated the displacement and F_x is the constant force to keep the string at some tension T . Although force F_y is initially acting only on one point of the string, it still connects to the other points along the string and causes an external force along the whole length of the string.

Figure 2.6 can be thought of as a very small piece of the string sliced off from the string of Figure 2.5. Examining the forces that are at the ends of an infinitesimal piece of the string, it can be noticed that force differences are the result of angle differences within the deflected string.

The assumptions listed for the ideal string require that the deflection amplitude should be small, so that F_y is very small compared to F_x and does not noticeably cause additional tension and stretching to the string. In this case, using the notations of Figure 2.6 in connection with the notations of Figure 2.1, one can set

$$F_{x1} = F_{x2} = F_x$$

because the tension in the x -direction was assumed to be constant at all times. With the external force component F_y added, the magnitude of the total force

$$|\vec{F}| = \sqrt{F_x^2 + F_y^2}, \quad (2.6)$$

and it has a direction towards the tangent line of the string at every value of

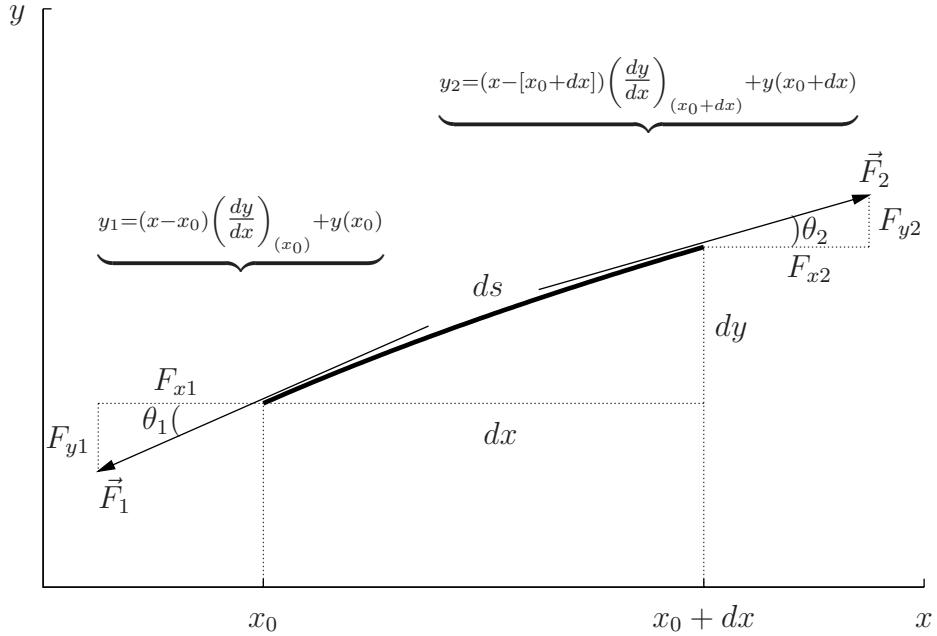


Figure 2.6: Forces acting on a small portion of a deflected string

x . The key idea is that the angle θ of the tangent line against the x -coordinate direction is directly proportional to the deflection of the string and therefore also proportional to the magnitude of the vertical force component F_y . Additionally, F_y is directly proportional to F_x via the angle θ , so that F_y can be completely defined in terms of F_x and θ . As shown in Figure 2.6, if the string is *linearised* at point x_0 with a straight tangent line, the net force F_y can be calculated as the product

$$F_y(x_0) = F_x \tan \theta = F_x \cdot \text{slope at } x_0. \quad (2.7)$$

Based on the reference material presented in the introductory section, it is known that by calculating a derivative at some point of a function, the value of that derivative equals the slope at that specific point. This knowledge can be utilised here and with the notations of Figure 2.6,

$$\begin{aligned} F_{y1} &= F_y(x_0) = F_x \left(\frac{\partial y}{\partial x} \right)_{x_0} \\ F_{y2} &= F_y(x_0 + dx) = F_x \left(\frac{\partial y}{\partial x} \right)_{x_0+dx}. \end{aligned} \quad (2.8)$$

So far, the equations of the orthogonal force components F_x and F_y have been covered, but how about the tangentially directed total force \vec{F} ? With the relation of equation (2.7), the magnitude of the total force from equation (2.6) can be

written in a more informative fashion as

$$|\vec{F}| = F_x \sqrt{1 + \left(\frac{\partial y}{\partial x}\right)^2},$$

which clearly indicates the increase of the total force as a fraction of F_x in connection with the slope caused by the deflection of the string in the y -coordinate direction. The deformation of the string during vibration constantly generates new deflections along the string, thereby keeping the string in motion. Since the ideal model neglected all frictional forces and other losses, the motion goes on repeating itself forever.

Now it is about time to use Newton's second law as given in equation (2.4). Because F_x is assumed constant, the net force acting on the differential piece of string can be defined solely as the difference of the components $F_y(x_0)$ and $F_y(x_0 + dx)$. Applying equations (2.8) and inserting those into equation (2.4),

$$F_x \left[\left(\frac{\partial y}{\partial x} \right)_{x_0+dx} - \left(\frac{\partial y}{\partial x} \right)_{x_0} \right] = ma,$$

and using the properties of differential, the equation of motion becomes

$$F_x \cdot d \left(\frac{\partial y}{\partial x} \right) = F_x \frac{\partial}{\partial x} \left(\frac{\partial y}{\partial x} \right) dx = F_x \frac{\partial^2 y}{\partial x^2} dx = ma.$$

When the string was at rest, i.e. completely parallel to the x -coordinate direction, the mass m of a tiny piece of string with length dx , cross sectional area A and density ρ is $A\rho \cdot dx$. It is also known from basic physics that the acceleration a can be expressed as a second order time derivative. With these reformulations

$$F_x \frac{\partial^2 y}{\partial x^2} dx = \rho A \frac{\partial^2 y}{\partial t^2} dx, \quad (2.9)$$

and the differential multipliers dx cancel each other out from both sides of the equation (2.9). After making a few simplifications to the equation, the final form of the so-called wave equation is written as

$$\frac{\partial^2 y}{\partial t^2} = c_T^2 \frac{\partial^2 y}{\partial x^2}, \quad (2.10)$$

where the transverse wave velocity c_T^2 is equal to $\frac{F_x}{\rho A}$ exactly as in the basic formula (2.1) for the harmonic frequencies of transverse waves. To express the meaning of equation (2.9) in words, it can be said that the acceleration of any small element of the string is directly proportional to the curvature of that element. [23, 24]

Similarly for the longitudinal waves one can write a wave equation that corresponds to equation (2.10). When the string is at rest and only under the influence of F_x , the whole string is parallel to the x -coordinate direction and the length of the infinitesimal piece of the string $ds = dx$. Adding an extra force F'_x stretches ds so that the relative change in ds is $\partial s / \partial x$. Then according to Hooke's law

$$F'_x = EA \frac{\partial s}{\partial x},$$

where E is Young's modulus and A is the cross-sectional area of the string. To get the string vibrating longitudinally, the piece of the string ds needs to experience a force difference at its ends. The difference should appear in the x -coordinate direction, and this is easily calculated as the differential

$$dF'_x = F'_x(x_0 + dx) - F'_x(x_0) = \frac{\partial F'_x}{\partial x} dx = EA \frac{\partial^2 s}{\partial x^2} dx.$$

Again, this expression can be made equal with the mass and acceleration terms, and just like in equation (2.10), the wave equation has the form

$$EA \frac{\partial^2 s}{\partial x^2} dx = \rho A \frac{\partial^2 s}{\partial t^2} dx,$$

which is normally written as

$$\frac{\partial^2 s}{\partial t^2} = c_L^2 \frac{\partial^2 s}{\partial x^2}. \quad (2.11)$$

The longitudinal wave velocity c_L^2 in equation (2.11) is equal to E/ρ , exactly as in the basic formula (2.2) for longitudinal frequencies of vibration. [21, p. 57]

2.1.4 Solving the transverse wave equation

So far, a few partial differential equations have been derived for transverse and longitudinal waves in a string. However, these equations do not give much information in this general form because the same wave equation can describe several different applications of motion. To have a specific solution for a vibrating string, the wave equation needs to be solved with reasonable boundary conditions. The boundary conditions arise from the unique properties of the system, which in this case is the ideally clamped string constrained with the restrictions described in section 2.1.2.

The wave equation (2.10) for transverse waves can be solved using the method of *separating variables*. The starting point is to make the guess that the solution will have the form

$$y(x, t) = F(x)G(t), \quad (2.12)$$

where the solution is a product of separate functions of time and location. To justify the guess, it was simply noticed in the early days of mathematics that many partial differential equations have their solution in this form. Moreover, a general approach would be too complicated for this purpose, albeit an interesting way for solving.

Differentiating equation (2.12) with respect to time t and location x , two new differential equations are obtained as

$$\frac{\partial^2 y}{\partial t^2} = F \ddot{G} \quad (2.13)$$

$$\frac{\partial^2 y}{\partial x^2} = F'' G, \quad (2.14)$$

where the two dots indicate a second derivative of time and the two commas refer to a second derivative with respect to x . These terms, (2.13) and (2.14), can be substituted into the original wave equation (2.10) so that

$$F \ddot{G} = c_T^2 F'' G. \quad (2.15)$$

To separate the variables, equation (2.15) should be divided by term $c_T^2 GF$ to yield

$$\frac{\ddot{G}}{c_T^2 G} = \frac{F''}{F}. \quad (2.16)$$

Now the variables of location and time are separated on different sides of the equation. The most important argument is: what happens if equation (2.16) is evaluated with certain values of x ? For all different values of x , the left side of equation (2.16) should be equal to the right side, which only depends on time. If the left side of the equation is evaluated using different values of x , it will not change the right side, which only depends on time. Therefore, equation (2.16) can only be true if both sides are equal to the same arbitrary constant k . This reasoning leads to two ordinary differential equations,

$$\frac{d^2 F}{dx^2} \frac{1}{F} = k \quad (2.17)$$

$$\frac{d^2 G}{dt^2} \frac{1}{G c_T^2} = k. \quad (2.18)$$

This step is where the boundary conditions come in. The two ordinary differential equations need to satisfy the boundary conditions

$$\text{a) } y(0, t) = 0 = F(0)G(t) \quad \text{and} \quad \text{b) } y(L, t) = 0 = F(L)G(t) \quad (2.19)$$

set by the end supports of the string. In other words, these conditions mean that the string is completely immobilised at the end supports. To find a feasible solution to equation (2.17), it is necessary to go through all the three possibilities for the value of the undetermined constant k .

First it should be noted that in the boundary condition equations (2.19) the product $F(x)G(t)$ is zero in both cases. Here one can use the product rule to deduce that if $G(t)$ is not zero, then $F(0) = 0$ and $F(L) = 0$.

Following the standard procedures of second order differential equations for the solutions of k , it is necessary to examine the auxiliary equation $r^2 + ar + b = 0$, where the roots r_1 and r_2 determine the parameters in the general solution

$$F(x) = Ae^{r_1 x} + Be^{r_2 x} \quad (2.20)$$

of an ordinary homogeneous differential equation of the second order.

If $k = 0$, then $r^2 + 0 = 0$ gives the roots $r_1 = r_2 = 0$. The general solution of (2.17) is then $F(x) = A + B$, but this is not a linearly independent solution, so $F(x) = Ax + B$ is the correct solution. To fix the constants A and B , the general solution is evaluated with the boundary conditions (2.19). In this case the equation pair

$$F(0) = A \cdot 0 + B = 0 \quad F(L) = A \cdot L + B = 0$$

gives $A = 0$ and $B = 0$, which is obviously a useless solution.

If $k = +p^2$, then $r^2 - p^2 = 0$ gives roots $r_1 = r_2 = \pm p$. The general solution of (2.17) is then $F(x) = Ae^{px} + Be^{-px}$. To fix the constants A and B , the solution is evaluated with the boundary conditions (2.19). The resulting pair of equations

$$F(0) = A + B = 0 \quad F(L) = Ae^{pL} + Be^{-pL} = 0$$

gives $A = 0$ and $B = 0$, which is again a useless solution. Now there is only one option left.

If $k = -p^2$, then $r^2 + p^2 = 0$ gives the imaginary roots $r_1 = r_2 = \pm ip$. The general solution of (2.17) is then $F(x) = Ae^{ipx} + Be^{-ipx}$, and using the Euler identity,

$$F(x) = A(\cos px + i \sin px) + B(\cos px - i \sin px).$$

This is a complex-valued solution, but a real-valued solution is needed. According to the theories of linear differential equations, any linear combination of

solutions is also a feasible solution, so let's define the notations

$$y_1 = (C_1 \cos px + iC_1 \sin px) \quad \text{and} \quad y_2 = (C_2 \cos px - iC_2 \sin px),$$

which define the two base functions of the solution with renamed undetermined constants C_1 and C_2 . The new basis for the solution can be arranged by linear combinations (sums and differences) of the form

$$y_1 + y_2 = (C_1 + C_2) \cos px ; \quad (y_1 - y_2) = i(C_1 - C_2) \sin px, \quad (2.21)$$

and after reassigning the constants as $A = C_1 + C_2$ and $B = i(C_1 - C_2)$, the solution in a basis of real-valued functions becomes

$$F(x) = A \cos px + B \sin px, \quad (2.22)$$

where the constants A and B can now generally hold complex quantities. To determine particular values for the constants A and B , the general solution (2.22) is evaluated with the boundary conditions (2.19). In this case,

$$F(0) = A \cos 0 + B \sin 0 = 0 \quad F(L) = A \cos pL + B \sin pL = 0.$$

The case $F(0)$ says that $A = 0$, so then the only solution left is $B \sin pL = 0$, which is true when

$$p = \frac{\pi n}{L} \quad (n = 0, 1, 2, 3, \dots).$$

Let's not care about the value of B just yet because the solution has now advanced to the result $k = -p^2$, which can be substituted forward into equation (2.18) to solve $G(t)$. When substituting k with $-p^2$ in equation (2.18),

$$\frac{d^2G}{dt^2} + \frac{c_T^2 n^2 \pi^2}{L^2} G = 0. \quad (2.23)$$

For a change, this simple second order equation can be solved by using the derivative notation as a differential operator acting on G . Squaring the differential equation (2.23) as

$$\left(\frac{d}{dt} + i \frac{c_T n \pi}{L} \right) \left(\frac{d}{dt} - i \frac{c_T n \pi}{L} \right) G = 0$$

and making use of the product rule, two separate first order differential equations are obtained:

$$\frac{dG}{dt} = -i \frac{c_T n \pi}{L} G \quad (2.24)$$

$$\frac{dG}{dt} = +i \frac{c_T n \pi}{L} G. \quad (2.25)$$

These can be solved easily by the method of separating variables and using direct integration. Noting that the basis of solution has two exponential terms just like the solutions obtained through the auxiliary equation justifies the fact that exponential functions are a quite natural basis for solutions of second order differential equations describing vibratory motion.

Based on the analysis for $F(x)$ that indicated the imaginary roots of the auxiliary equation to be the only sufficient solutions, the use of Euler's formula yields

$$G_n(t) = A_n \cos\left(\frac{c_T n \pi t}{L}\right) + B_n \sin\left(\frac{c_T n \pi t}{L}\right).$$

Now the solutions for $F(x)$ and $G(t)$ have been found. Taking into account that $y(x, t) = F(x)G(t)$, the current complete solution with given boundary conditions can be written as

$$y_n(x, t) = \left[A_n \cos\left(\frac{c_T n \pi t}{L}\right) + B_n \sin\left(\frac{c_T n \pi t}{L}\right) \right] \sin\left(\frac{n \pi x}{L}\right). \quad (2.26)$$

There exists a fundamental theorem for differential equations, which was already used in the derivation of the solution for $F(x)$. The theorem says: *If y_1 and y_2 are solutions of a homogeneous linear partial differential equation, a sum of y_1 and y_2 with constants c_1 and c_2 is also a solution.* This leads to the following conclusion that

$$y(x, t) = \sum_{n=1}^{\infty} \left[A_n \cos\left(\frac{c_T n \pi t}{L}\right) + B_n \sin\left(\frac{c_T n \pi t}{L}\right) \right] \sin\left(\frac{n \pi x}{L}\right). \quad (2.27)$$

The undefined constants A_n and B_n now also multiply the solution of $F(x)$, where the undefined constant B was neglected for a while. This just means that the undefined constant of $F(x)$ is now included in A_n and B_n , which still need to be defined to obtain the complete solution for the equation that describes the motion of a vibrating string.

The boundary conditions for the system have now been included in the current solution of the system, but still the model does not contain any reference to the actual properties of the string. Because the wave equation is a second order differential equation, it is enough to determine the initial ($t = 0$) displacement $y(x, 0) = g(x)$ and velocity $\dot{y}(x, 0) = v(x)$ to fully describe the motion of the string. Setting $t = 0$ in equation (2.27),

$$y(x, 0) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n \pi x}{L}\right), \quad (2.28)$$

and this needs to be equal to the initial shape of the string, modelled by the equation $g(x)$, i.e. one needs to set

$$g(x) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right). \quad (2.29)$$

Unfortunately the constants A_n are still undefined. But the equation on the right side looks awfully familiar. It is the Fourier sine series, where the constants

$$A_n = \frac{2}{L} \int_0^L g(x) \sin\left(\frac{n\pi x}{L}\right) dx. \quad (2.30)$$

For the initial velocity, equation (2.27) needs to be differentiated with respect to time to reach the expression

$$\dot{y}(x, t) = \sum_{n=1}^{\infty} \left[-A_n \frac{c_T n \pi}{L} \sin\left(\frac{c_T n \pi t}{L}\right) + B_n \frac{c_T n \pi}{L} \cos\left(\frac{c_T n \pi t}{L}\right) \right] \sin\left(\frac{n\pi x}{L}\right).$$

The initial velocity is obtained by setting $t = 0$, leading to the equation

$$\dot{y}(x, 0) = \sum_{n=1}^{\infty} B_n \frac{c_T n \pi}{L} \sin\left(\frac{n\pi x}{L}\right) = v(x), \quad (2.31)$$

where the yet undefined constants B_n are determined from equation

$$\begin{aligned} B_n \frac{c_T n \pi}{L} &= \frac{2}{L} \int_0^L v(x) \sin\left(\frac{n\pi x}{L}\right) dx \quad \Big| : \frac{c_T n \pi}{L} \\ B_n &= \frac{2}{c_T n \pi} \int_0^L v(x) \sin\left(\frac{n\pi x}{L}\right) dx. \end{aligned}$$

So what is the practical significance of all this? Considering a string that is plucked with a finger or a pick, the initial velocity is zero, multipliers B_n are all zero and then equation (2.27) becomes

$$y(x, t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{c_T n \pi t}{L}\right), \quad (2.32)$$

where the coefficients A_n will be defined by (2.30), where $g(x)$ depicts the geometrical form of the plucked string just before releasing it into motion. After determining all A_n , it is possible to simulate the vibrations of the string by plotting the function $y(x, t)$ with different values of t . Equation (2.32) can also be written in a more intuitive form,

$$y(x, t) = \frac{1}{2} \sum_{n=1}^{\infty} A_n \sin\left[\frac{n\pi}{L}(x - c_T t)\right] + \frac{1}{2} \sum_{n=1}^{\infty} A_n \sin\left[\frac{n\pi}{L}(x + c_T t)\right]. \quad (2.33)$$

This result is obtained with the help of trigonometric identity formulae, and it clearly shows two identical waves with the same amplitude travelling to opposite directions along the x -coordinate axis. [11, pp. 538–546]

The actual plot of the calculated result reveals interesting properties of the Fourier series and the periodic motion of the ideal string. Figure 2.7 takes the viewpoint from outside the boundaries set by 0 and L to show the idea of the periodicity of the Fourier series. Basically the two triangular waves are periodic functions that continue to eternity, and when the time advances, these waves are travelling in opposite directions. When simulating the waveform of the string in Figure 2.7, only the ten first harmonic components have been taken into the calculation, and that is why the curves look more soft, maybe even more realistic.

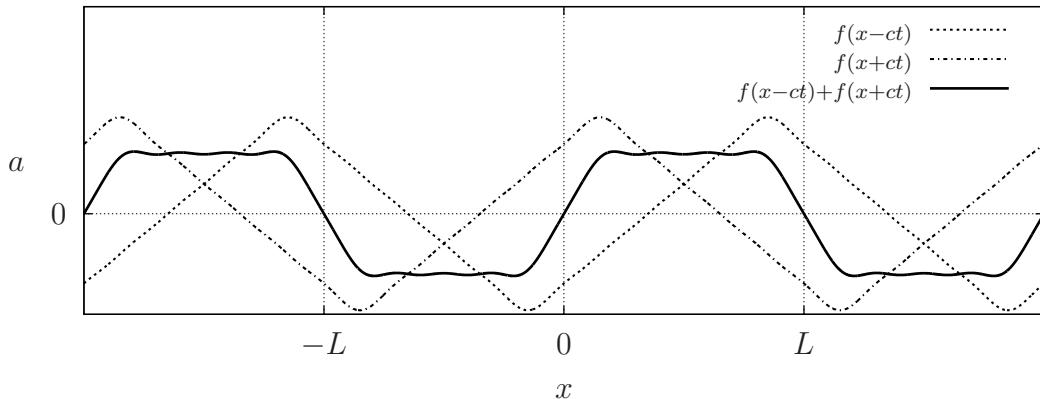


Figure 2.7: Travelling waves outside the boundaries of 0 and L

In Figure 2.7 the string has been plucked initially from the middle of the string and the situation has evolved some units of time, making it possible to notice the trend of the motion. The wave $f(x - c_T t)$ travels to the right and wave $f(x + c_T t)$ to the left. The use of the verb 'to travel' in the context of these wave functions is just an illusion, since immediately from time $t = 0$ the waves exist with all the values of x . A focus on the vertical motion of a single point x_0 at the origin reveals that the movement has the values of a sine wave that can be imagined to be drawn towards the positive or negative direction on the x -axis. The direction is determined by the sign before the time variable t . The wave reflection from the boundary is also an illusion because the sum of the two continuous waves can be imagined to reflect at 0 and L , but actually it is just a new period of a wave that comes from outside the boundaries and sums up with the other one.

If all the restrictions of the ideal string are valid, this motion of travelling waves

would continue to eternity without changing the shape of the waves. If, on the other hand, there was friction and nonideal end supports involved, the waves would change their geometrical shape as time evolves and eventually the friction would cause the waves to fade away completely. This is just to point out that the equations based on the ideal model do not indicate this fact.

Figure 2.8 shows another situation where the string has been initially plucked nearby the other end support. This is to show that when the plucking point is not exactly in the middle of the string, the two waves do not sum up symmetrically, although the shapes of the waves are still symmetric. To show the symmetry of the summed waves, Figure 2.8 should have been extended to cover a more wider view of the x -axis because the full period of the waves $f(x - c_T t)$ and $f(x + c_T t)$ is $2L$. The full period of $f(x - c_T t) + f(x + c_T t)$ is also $2L$, but it is always symmetric about the point $x = L$.

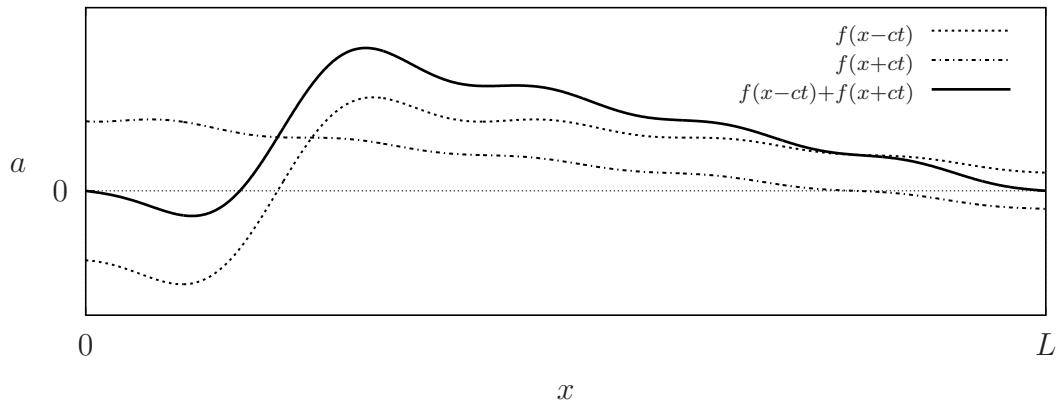


Figure 2.8: Sum of the travelling waves inside the boundaries 0 and L

The Octave script presented in Appendix C can be used to simulate waveforms of the ideal string in the closed interval $[0 : L]$ with different plucking positions and displacement amplitudes.

2.1.5 The significance of A_n : the amplitude spectrum

For the author, the most fascinating result after all this mathematical fireworks is the fact that the geometrical shape of the string determines the sound, tone and timbre of a vibrating string. It is also amazing that the theoretically purely mathematical idea of the Fourier series brings actual physical results from geometry. When considering the event of plucking a guitar string, it is firstly displaced by the amount of a at point p from the equilibrium position, then released to

vibrate freely. The geometrical form of the plucked string was already presented in Figure 2.5. Although that depiction exaggerates the deflection, the triangular form is realistic and valid if a very sharp object is used to deflect the string.

The derivation for the equation of motion for the vibrating string illustrated that the unknown factors A_n in (2.27) are calculated from equation (2.30), where $g(x)$ is the function that describes the geometrical form of the string. Using the notations from Figure 2.5,

$$g(x) = a \cdot \left[\frac{x}{p} (0 \leq x \leq p) - \frac{L-x}{L-p} (p < x \leq L) \right],$$

and this model is quite sufficient to model a variety of initial shapes of plucked strings. If someone wants to go for more accuracy, the obvious discontinuation point between the two slopes at point $x = p$ can be of course modelled to have small roundness, but this complicates the evaluation of the factors A_n .

As stated before, equation (2.30) is equal to the formula for the Fourier sine series. This series assumes that the function $g(x)$ should be periodic and odd [$f(-x) = -f(x)$]. Indeed, $g(x)$ fulfils these requirements since it can be extended to an odd function with a period of $2L$ as indicated in Figure 2.9. Now that

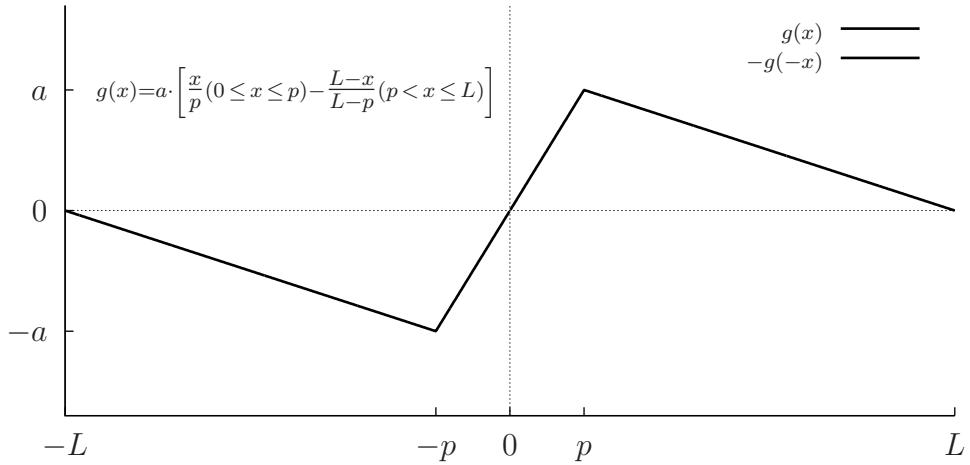


Figure 2.9: Odd periodic extension of the plucked string function $g(x)$

the calculational prerequisites have been justified, one can proceed to solve the factors A_n from equation (2.30). The calculation is relatively straightforward, but quite lengthy. The intermediate steps of the calculation are therefore included in Appendix A, the final result being stated as

$$A_n = \frac{2a}{n^2\pi^2} \left(\frac{L}{p} + \frac{L}{L-p} \right) \sin \left(\frac{n\pi p}{L} \right). \quad (2.34)$$

This result has physical significance because it relates the plucking position p and displacement amplitude a to the tonal properties of the string. It is almost unbelievable that the structure of the harmonic upper partials of the string is solely determined by the plucking position. The result (2.34) can be visualised by plotting an example, where $p = \frac{L}{6}$. According to Figure 2.10, this plucking position creates a spectrum, where every 6th term is zero by amplitude. From another perspective this means that the string has been plucked at a node of the 6th harmonic frequency in the standing wave pattern.

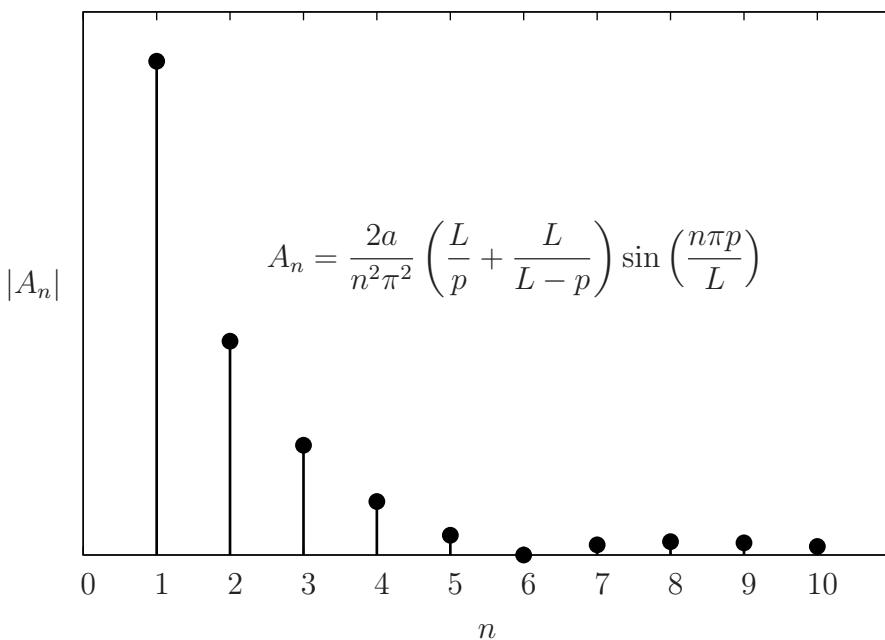


Figure 2.10: A spectrum plot from the theoretical formula with $p = \frac{L}{6}$

Generally the Fourier coefficients are complex-valued and carry also information about the phase of the harmonic frequency components. Because using now only the Fourier sine series, the amplitude coefficients A_n are real numbers. In this case the phase of the frequency components is indicated by + or - signs of A_n , so the phase of the harmonic frequencies can only be 0 or π . This is why the spectrum is plotted using the absolute value $|A_n|$, since with the negative phase the spectrum bins would point downwards.

After substituting the expression of coefficients A_n into the general solution (2.32) for plucked strings, the function describing the motion of the string becomes

$$y(x, t) = \sum_{n=1}^{\infty} \frac{2a}{n^2\pi^2} \left(\frac{L}{p} + \frac{L}{L-p} \right) \sin \left(\frac{n\pi p}{L} \right) \sin \left(\frac{n\pi x}{L} \right) \cos \left(\frac{c_T n \pi t}{L} \right), \quad (2.35)$$

and the equation of motion for a single frequency component is

$$y_n(x, t) = \frac{2a}{n^2\pi^2} \left(\frac{L}{p} + \frac{L}{L-p} \right) \sin\left(\frac{n\pi p}{L}\right) \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{c_T n\pi t}{L}\right). \quad (2.36)$$

These results with all terms written out explicitly in a single equation shows all the variables that can have an effect on the motion of a plucked vibrating string. These results are used later when differentials with respect to x and y of this basic equation of motion are needed.

As already shortly mentioned above, the given equation derived for the harmonic spectrum amplitudes is valid only in that theoretical case where the object used to pluck the string is infinitely sharp and when the string is totally flexible to allow a perfect triangular shape. A more realistic geometric description would round the angle between the two oppositely directed slopes. This would give a better model for plucks made with a finger or a plectrum. It would be mathematically possible to include roundness to the function of string geometry, but then the analytic expressions would not fit on this page. That is the main reason to try and keep things as simple as possible.

The sharp angle in the triangular shape allows infinitely many harmonic components to exist, but e.g. a plectrum with width δ used to pluck a string with length L causes a cut-off for upper partials with indices above L/δ . In other words, the wider the plucking object used, the less harmonics are observed in the resulting vibration and eventually the sound of the string becomes quite dull without the richness of harmonicity. [25]

2.1.6 Alternative solutions of the wave equation

The function $y(x, t)$ given by equation (2.33) is not the only form of solution that satisfies the wave equation (2.10). In physics and engineering it is common to write the sum

$$A \cos \omega t + B \sin \omega t$$

as a complex exponential function with a constant complex factor $C = A + iB$, so that the final result is taken as the real part of

$$\begin{aligned} Ce^{-i\omega t} &= (A + iB)(\cos \omega t - i \sin \omega t) \\ &= A \cos \omega t - Ai \sin \omega t + iB \cos \omega t + B \sin \omega t, \end{aligned}$$

where the real part

$$\Re(Ce^{-i\omega t}) = A \cos \omega t + B \sin \omega t. \quad (2.37)$$

Physicists usually justify this calculation technique by saying that the imaginary part does not have any concrete physical meaning. Moreover, as the imaginary part is added there from nowhere in the first place, surely it can be ignored in the final results.

Considering the ideal string model with wave equation (2.10), the two ordinary second order differential equations are simply solved in the form

$$\begin{aligned} y(x, t) &= F(x)G(t) = (Ae^{ikx} + Be^{-ikx})(Ce^{i\omega t} + De^{-i\omega t}) \\ &= ACE^{i(kx+\omega t)} + BDe^{-i(kx+\omega t)} + ADe^{i(kx-\omega t)} + BCE^{-i(kx-\omega t)}. \end{aligned}$$

By first applying a similar procedure as in equation (2.21) to change the basis to real trigonometric functions of sine and cosine and then transforming it back to a complex exponential notation as in equation (2.37), the general solution of the ideal string wave equation can be written as

$$y(x, t) = \Re(Ee^{-i(kx+\omega t)}) + \Re(Fe^{-i(kx-\omega t)}),$$

where the complex-valued constants $E = e_0 + ie_1$ and $F = f_0 + if_1$.

Alternatively, by performing the multiplication $F(x)G(t)$ as $G(t)F(x)$

$$\begin{aligned} y(x, t) &= G(t)F(x) = (Ce^{i\omega t} + De^{-i\omega t})(Ae^{ikx} + Be^{-ikx}) \\ &= ACE^{i(\omega t+kx)} + BDe^{-i(\omega t+kx)} + BCE^{i(\omega t-kx)} + ADe^{-i(\omega t-kx)}, \end{aligned}$$

and transforming this to the trigonometric basis yields

$$y(x, t) = e_0 \cos(\omega t + kx) + e_1 \sin(\omega t + kx) + f_0 \cos(\omega t - kx) + f_1 \sin(\omega t - kx). \quad (2.38)$$

The complex exponential representation of this is

$$y(x, t) = \Re(Ee^{-i(\omega t+kx)}) + \Re(Fe^{-i(\omega t-kx)}),$$

or when extracting the time dependence as a common factor,

$$y(x, t) = \Re([Ee^{-ikx} + Fe^{+ikx}] e^{-i\omega t}). \quad (2.39)$$

There exist several wave equations similar to the one derived from the geometry of the string. This discussion is added to show different types of functions that are solutions to a wave equation of quite general type. The complex exponential function is often used to present a general solution of a wave equation, and refer to a mathematical model describing sinusoidal motion.

The general solution of wave equation (2.10) can also be derived starting from equation (2.38). To satisfy the boundary condition $y(0, t) = 0$, the constants $e_0 = -f_0$ and $e_1 = -f_1$, so that

$$y(x, t) = e_1 [\sin(\omega t + kx) - \sin(\omega t - kx)] + e_0 [\cos(\omega t + kx) - \cos(\omega t - kx)]. \quad (2.40)$$

Using trigonometric formulae

$$\begin{aligned} \sin(x \pm y) &= \sin x \cos y \pm \cos x \sin y \\ \cos(x \pm y) &= \cos x \cos y \mp \sin x \sin y, \end{aligned}$$

the equation for $y(x, t)$ simplifies to

$$y(x, t) = 2(e_1 \cos \omega t - e_0 \sin \omega t) \sin kx,$$

which is already in the same form as the solution presented in section 2.1.4. The other boundary condition $y(L, t) = 0$ requires from equation (2.40) that $kL = n\pi$, or since $\omega = c_T k$, $\omega = \frac{c_T n\pi}{L}$ and this leads to

$$y_n(x, t) = \left[A_n \cos \left(\frac{c_T n\pi t}{L} \right) + B_n \sin \left(\frac{c_T n\pi t}{L} \right) \right] \sin \left(\frac{n\pi x}{L} \right),$$

where the new constants A_n and B_n have absorbed previous undetermined constants e_0 and e_1 with signs and multipliers of 2. This way exactly the same expression as in equation (2.26) is obtained. [21, p. 39], [24, p. 11]

2.2 NONLINEAR EFFECTS IN STRING VIBRATION

In reality a vibrating string is a source of several interesting phenomena, which mostly relate to nonlinearities in the motion of the string. In that sense the ideal string model is not a realistic simulation model, but some predictions of nonlinear behaviour can be derived by making simple extensions to the linear model. The linear model only approximated that the amplitude of the vibration was small, but other factors that affect the behaviour of the string were completely neglected. Therefore, it is justified to add these neglected factors to the linear model one by one and see what the effect is for each source of nonlinearity. Finally, a general model that correctly simulates also large amplitude string vibrations is presented and evaluated on a simple level.

2.2.1 The effect of amplitude

Starting directly from the linear model, it is possible to analyse the effect of plucking amplitude in the resulting vibrational motion. Large string deflections

mainly affect the vibration frequency by shifting the harmonic frequencies due to increased string tension. Additionally, several coupling phenomena between the longitudinal and transverse modes of vibration are indirect consequences of large displacement amplitudes.

In reality the length of the string changes slightly during vibration, and the larger the displacement amplitude, the more noticeable the change in the length is. The length of the string during vibration can be calculated by using the length element $ds^2 = dx^2 + dy^2$ of a curve and integrating (summing) those elements through the whole length of the string. The actual change of length δL is expressed as a difference to the nominal length L , for which the analytic expression is

$$\delta L = \int_0^L ds - L = \int_0^L \left[\sqrt{1 + \left(\frac{\partial y}{\partial x} \right)^2} - 1 \right] dx. \quad (2.41)$$

To have concrete results from equation (2.41), the function $y(x)$ needs to be inserted and the integral evaluated. The function $y(x)$ is taken from equation (2.32), which makes the integral of (2.41) relatively difficult to solve. Differentiating (2.32) with respect to x ,

$$\frac{\partial y}{\partial x} = \sum_{n=1}^{\infty} \frac{n\pi A_n}{L} \cos \left(\frac{n\pi x}{L} \right) \cos \omega_n t, \quad (2.42)$$

and after inserting the coefficients A_n ,

$$\frac{\partial y}{\partial x} = \frac{2a}{\pi} \left(\frac{1}{p} + \frac{1}{L-p} \right) \sum_{n=1}^{\infty} \frac{1}{n} \sin \left(\frac{n\pi p}{L} \right) \cos \left(\frac{n\pi x}{L} \right) \cos \omega_n t. \quad (2.43)$$

From equation (2.43) one can estimate the behaviour of the summed series. In practise the string does not have any discontinuities in the interval of $[0, L]$, and because of the factor $\frac{1}{n}$, the series converges to the slope (derivative) of the string in $[0, L]$. Furthermore, the slope of the string at any point of the string is always $\ll 1$ or at least < 1 . From here it follows that $(\partial y / \partial x)^2 \ll 1$, and it is justified to use the binomial approximation

$$\sqrt{1 + \left(\frac{\partial y}{\partial x} \right)^2} \approx 1 + \frac{1}{2} \left(\frac{\partial y}{\partial x} \right)^2 - \frac{1}{8} \left(\frac{\partial y}{\partial x} \right)^4 + \dots - \dots + \dots \quad (2.44)$$

As usual, the first two terms of the series are taken as the estimation, and since the series is alternating $(+ \dots - \dots + \dots - \dots)$, the error of the estimation is approximately the first term not taken to the actual estimation, i.e. the fourth order term in this case. So, with a little overestimation (because the fourth order

term has a minus sign), the change in the string's length can be expressed as

$$\delta L \approx \frac{1}{2} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx. \quad (2.45)$$

When the series (2.42) is squared and substituted into (2.45), the equation equals the general form of the inner product of functions. Because the trigonometric basis of the Fourier series is orthogonal, the only terms that are not zero are

$$\frac{n^2\pi^2 A_n^2}{2L^2} \cos^2 \omega_n t \int_0^L \cos^2 \left(\frac{n\pi x}{L} \right) dx.$$

By applying a trigonometric identity formula $\cos^2 x = \frac{1}{2}(1 + \cos 2x)$ to the terms on both sides, the integral reshapes into the form

$$\frac{n^2\pi^2 A_n^2}{4L^2} (1 + \cos 2\omega_n t) \int_0^L \frac{1}{2} \left[1 + \cos \left(\frac{n2\pi x}{L} \right) \right] dx,$$

and evaluation of the integral leads to a value of $\frac{L}{2}$. Therefore, the analytic expression to approximate the change of length of the string during vibration is written as

$$\delta L = \sum_{n=1}^{\infty} \frac{n^2\pi^2 A_n^2}{8L} (1 + \cos 2\omega_n t). \quad (2.46)$$

It is again quite amazing that the use of a trigonometric identity formula brought physical meaning to the expression for δL . Because of the factor of 2 inside the cosine function, the angular frequency for *longitudinal* oscillation due to tension modulation is now $2\omega_n$, i.e. two times the frequency of the transverse waves. This is also an intuitive result since it is clear that the minimum increase of tension due to transverse deflection is when there is no transverse deflection, and the maximum increase comes equally from both sides of transverse deflection.

If the approximation is made that the initial tension force F_x (see Figure 2.1) is proportional to the ratio of stretch ΔL and effective length L , the total tension in the string is

$$F(t) \approx EA \frac{\Delta L}{L} + EA \frac{\delta L}{L} = F_x \left[1 + \sum_{n=1}^{\infty} \frac{n^2\pi^2 A_n^2}{8L\Delta L} (1 + \cos 2\omega_n t) \right]. \quad (2.47)$$

The time related averages of equations (2.46) and (2.47) are

$$L + \overline{\delta L} = L \left[1 + \frac{\overline{\delta L}}{L} \right] = L \left[1 + \sum_{n=1}^{\infty} \frac{n^2\pi^2 A_n^2}{8L^2} \right], \quad (2.48)$$

and

$$F_x + \overline{\delta F} = F_x \left[1 + \frac{\overline{\delta L}}{\Delta L} \right] = F_x \left[1 + \sum_{n=1}^{\infty} \frac{n^2 \pi^2 A_n^2}{8L\Delta L} \right].$$

These averages increase the frequencies of the transverse vibration modes equally, if the assumption holds that the local change of length affects the string equally along the whole distance of the string. This is the same as assuming that the longitudinal waves travel in the string with a much greater velocity than the transverse waves. If this is the case, the change in tension is fast enough to affect the whole string before changing considerably at some local position of the string. [26]

By applying these reformulations to the basic equation (2.1) for the frequencies of the harmonic modes of transverse vibration, with a simplifying approximation

$$\frac{f_{Tn} + \delta f_{Tn}}{f_{Tn}} = \frac{f'_{Tn}}{f_{Tn}} \approx \frac{\sqrt{1 + \frac{\overline{\delta L}}{\Delta L}}}{1 + \frac{\overline{\delta L}}{L}} \approx 1 + \frac{1}{2} \frac{\overline{\delta L}}{\Delta L}. \quad (2.49)$$

This result indicates a relative increase of all harmonic frequencies by a constant factor proportional to the increased tension in the string during vibration.

In related literature this frequency increase (or decrease from the increased state) is also known as *pitch glide* and this effect is commonly observed in all plucked string instruments. Pitch glide is most dominant right after the string is plucked and vanishes rapidly as the vibration amplitude decreases over time. [27]

2.2.2 The effect of friction

If there was no friction or other energy losses involved in the motion of the string, it would never stop vibrating. Friction causes a damping effect that eventually forces the motion of the string to stop and return to a balanced state of rest.

If friction is taken into account in the mathematical model of the ideal string, the general one-dimensional wave equation becomes

$$\rho A \frac{\partial^2 y}{\partial t^2} + D(\omega) \frac{\partial y}{\partial t} = F_x \frac{\partial^2 y}{\partial x^2}, \quad (2.50)$$

where $D(\omega)$ is a general frequency dependent damping coefficient. This type of wave equation is solved using the same method of separating variables, which helped to solve the general wave equation (2.10). The general form of the solution is expected to be the product

$$y(x, t) = F(x)G(t),$$

which is substituted into equation (2.50) to yield

$$\rho A F \ddot{G} + D(\omega) F \dot{G} = F_x G F''. \quad (2.51)$$

Dividing equation (2.51) by the term $F G \rho A$ gives the separated form

$$\frac{\ddot{G}}{G} + \frac{D(\omega)}{\rho A} \frac{\dot{G}}{G} = c_T^2 \frac{F''}{F},$$

and additionally dividing by c_T^2 yields

$$\frac{\ddot{G}}{c_T^2 G} + \frac{D(\omega)}{c_T^2 \rho A} \frac{\dot{G}}{G} = \frac{F''}{F}. \quad (2.52)$$

Again following the already familiar argument that both sides of equation (2.52) have to be equal to a constant, named e.g. k , results in two ordinary differential equations,

$$\frac{d^2 F}{dx^2} \frac{1}{F} = k \quad (2.53)$$

$$\left(\frac{d^2 G}{c_T^2 dt^2} + \frac{D(\omega)}{c_T^2 \rho A} \frac{dG}{dt} \right) \frac{1}{G} = k. \quad (2.54)$$

From here equation (2.53) gives the same solution as in the case of the wave equation (2.10). The modification comes in only from the time dependent equation, which (by following the derivation of equation (2.10)) can be rewritten in the form

$$\frac{d^2 G}{dt^2} + \frac{D(\omega)}{\rho A} \frac{dG}{dt} + \frac{c_T^2 \pi^2 n^2}{L^2} G = 0. \quad (2.55)$$

From here one can proceed to solve the equation as a 'textbook example'. First an auxiliary equation in terms of the order of the differentials is written for variable r and the resulting quadratic equation

$$r^2 + \frac{D(\omega)}{\rho A} r + \frac{c_T^2 \pi^2 n^2}{L^2} = 0$$

is solved to give the roots

$$r_{1,2} = -\frac{D(\omega)}{2\rho A} \pm \sqrt{\left(\frac{D(\omega)}{2\rho A}\right)^2 - \left(\frac{c_T \pi n}{L}\right)^2}.$$

The term involving c_T will be dominant, leading to the complex conjugate roots

$$r_1 = -\frac{D(\omega)}{2\rho A} + i \sqrt{\left(\frac{c_T \pi n}{L}\right)^2 - \left(\frac{D(\omega)}{2\rho A}\right)^2}$$

$$r_2 = -\frac{D(\omega)}{2\rho A} - i \sqrt{\left(\frac{c_T \pi n}{L}\right)^2 - \left(\frac{D(\omega)}{2\rho A}\right)^2},$$

where $i = \sqrt{-1}$ has been taken as a factor in front of the square root term. Using the linear addition of sine and cosine as earlier to get rid of the imaginary part of the solution leads to the general solution

$$G_n(t) = A_n \cos(\omega_n t)e^{-\alpha} + B_n \sin(\omega_n t)e^{-\alpha},$$

where

$$\omega_n = \sqrt{\left(\frac{c_T \pi n}{L}\right)^2 - \left(\frac{D(\omega_n)}{2\rho A}\right)^2} \quad \text{and} \quad \alpha = \frac{D(\omega_n)}{2\rho A}.$$

From here it is already evident that all the factors leading to energy loss in the string are generally lowering the frequencies of the harmonic modes of vibration. The damping term $D(\omega_n)$ is frequency dependent, so that the harmonic frequencies will deharmonise unequally due to different energy loss mechanisms.

In the setup of a vibrating string there are several factors that lead to the damping of the vibration. Considering the string itself, there are two main sources of damping: the effect of air viscosity and internal friction forces. These are associated with individual decay time constants τ_1 and τ_2 respectively. For thin steel strings the more noticeable of these two loss mechanisms is the viscous loss of energy into the air. The internal damping becomes dominant in nylon strings, especially if the nylon strings are over-spun with thin metal wire. [28] [25]

The analytic treatment of the effect of viscous losses is way too complicated to be derived from first principles. The primary work for this has been done by Stokes [29] a very long time ago, and the results are applied to strings of musical instruments by Fletcher [28]. Here the results are just referenced without modifications.

Due to the viscosity of air flow, the frictional retarding force F_R experienced by a cylinder of length L and radius r moving with frequency f and velocity c is given by

$$F_R \approx 2\pi^2 \rho_a f c r^2 L \left(\frac{2\sqrt{2}M + 1}{2M^2} \right),$$

where ρ_a is the density of air and

$$M = \frac{r}{2} \sqrt{\frac{2\pi f}{\mu_a}},$$

where μ_a is the kinematic viscosity of air. In practical strings the value of M is in the order of 1.0 or less. Now, because F_R is proportional to velocity c , the kinetic energy involving F_R will be proportional to the square of the velocity. Evaluation

of the rate of change in energy can be done qualitatively as

$$\frac{dE}{dt} = F_R c = 4\pi \rho_a f \frac{1}{2} \frac{mc^2}{\rho} \left(\frac{2\sqrt{2}M + 1}{2M^2} \right),$$

where the mass m of the string has been forced into the equation using the substitution $\pi r^2 L = \frac{m}{\rho}$ to get the familiar term of kinetic energy to the right side of the equation. Dividing by that kinetic energy term one obtains the energy ratio

$$\frac{dE}{\frac{1}{2} mc^2} = \frac{4\pi \rho_a f}{\rho} \left(\frac{2\sqrt{2}M + 1}{2M^2} \right) dt,$$

where the term on the right side has the units of s^{-1} when the dt term is neglected. The decay time constant due to viscous air flow around the string is therefore the inverse of the right side, namely

$$\tau_1 = \frac{\rho}{4\pi \rho_a f} \left(\frac{2M^2}{2\sqrt{2}M + 1} \right).$$

According to Fletcher and Rossing [21, p. 54], expansion of the term in parentheses leads to a complex dependency between the string radius and frequency of vibration. Due to variations in M , the time constant $\tau_1 \propto \rho r^2$ is independent of frequency at low frequencies. At high frequencies $\tau_1 \propto \rho r^2 / \sqrt{f}$.

Attenuation due to internal friction comes into play when the material of the string is investigated in detail. Advanced level materials physics defines Young's modulus as a complex quantity

$$E = E_1 + jE_2,$$

which is used, for example, in a relaxation formula developed by Debye. According to that theory, the imaginary part of Young's modulus depends on frequency and has a resonance peak at some specific frequency. In a more complex use case, both components E_1 and E_2 are depending on frequency.

If the complex form of Young's modulus is substituted into a regular wave equation that describes the motion of a string, the imaginary term E_2 will result in an imaginary exponential term that automatically is related to energy loss and indicates decay. Fletcher [25] has found that the decay time due to internal damping

$$\tau_2 = \frac{1}{\pi f} \frac{E_1}{E_2},$$

indicating that the decay time is inversely proportional to frequency. It is claimed that the internal friction is not that meaningful in steel strings, but for nylon strings it might contribute as a major source of decay.

In addition to the effects of air and internal damping, a third decay mode giving a time constant τ_3 can be distinguished and it is related to the energy loss through the end supports of the string. This decay mechanism is discussed more thoroughly in section 2.2.4 in the same context along other theories about the effect of end supports.

The three decay mechanisms contribute individually to the total decay time. The additions between the decay times is done invertibly, so that

$$\frac{1}{\tau_{tot}} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}.$$

Figure 2.11 visualises dependency between individual decay times and the frequency of string vibration. The analysis has been made from the viewpoint of the fundamental frequency, hence the increase of frequency is obtained by increasing the tension of the string. Therefore the analysis only gives a qualitative forecast about the decay times of the upper partials that appear along with the fundamental frequency.

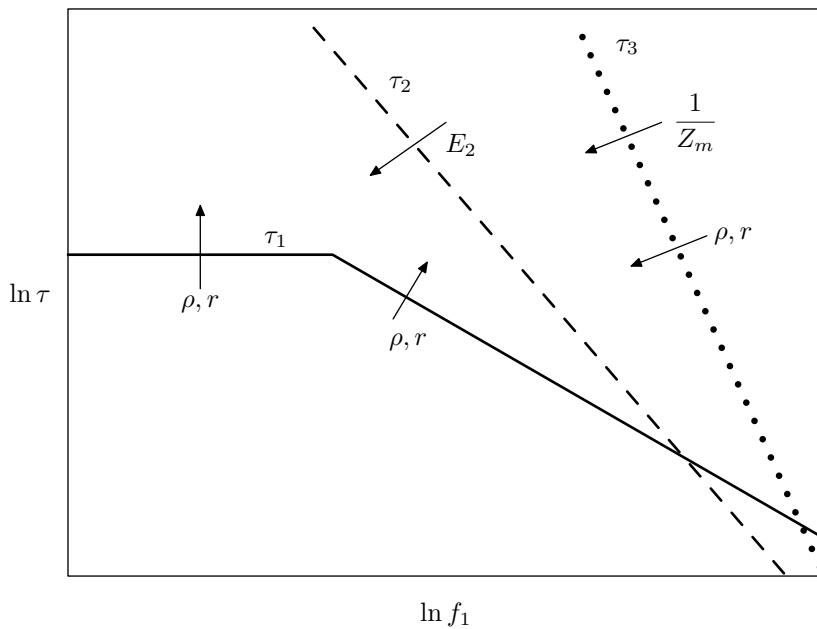


Figure 2.11: Time constants τ_1 , τ_2 and τ_3 with the affected variables depicting frequency dependent string vibration decay. The figure is adaptively redrawn from reference [21].

2.2.3 The effect of stiffness

The linear model that leads to the simple wave equation (2.10) assumes the string to be completely flexible, meaning that the string does not cause any mechanical resistance to bending. In reality it is quite obvious that strings used in musical

instruments have stiffness, which deviates the mathematical model of the string from the ideal approximation.

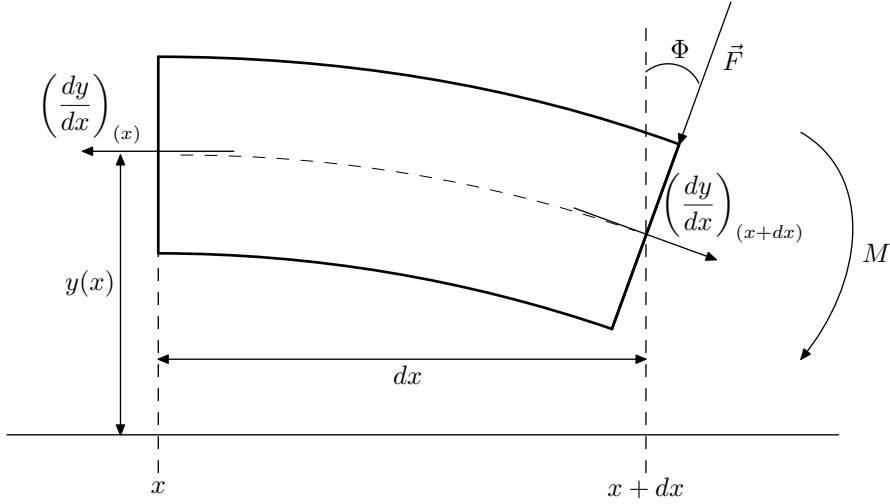


Figure 2.12: Torque M acting on a stiff piece of dx

The situation can be clarified with reference to Figure 2.12, where the infinitesimal piece of string dx is stressed by torque M , which relates to force \vec{F} via length dx . The left side of the string is aligned parallel to the x -axis, and the torque bends the right side to the angle Φ measured against a vertical reference crossing through the centreline of the string. The essential variables that are not marked in Figure 2.12 are the cross-sectional area of the string A , Young's modulus E , and the halved radius ($a/2$) of the string denoted as κ , which refers to the radius of gyration of the circular cross-section of the string.

To bend the length dx of the string by angle Φ , a torque

$$M = \frac{EA\Phi\kappa^2}{dx} \quad (2.56)$$

is needed [23, p. 153]. If the angle Φ is assumed to be small, it can be approximated with the difference of slopes at the ends of the infinitesimal piece of the string. The first term of the Maclaurin series representation of the arcus tangent is used as the approximation because the slope is defined as $\tan \phi = \frac{dy}{dx}$, from where the angle ϕ is the arcus tangent of the slope and $\Phi = \phi_1 - \phi_2$ with a reference to Figure 2.12. The approximation is applied as

$$\Phi \approx - \left[\left(\frac{\partial y}{\partial x} \right)_{(x+dx)} - \left(\frac{\partial y}{\partial x} \right)_{(x)} \right] = -d \left(\frac{\partial y}{\partial x} \right) = - \left(\frac{\partial^2 y}{\partial x^2} \right) dx$$

and after inserting the expression of Φ into equation (2.56),

$$M \approx -EA\kappa^2 \frac{\partial^2 y}{\partial x^2}.$$

It is suitable to question the units of this equation - does the equation really have the units of torque on both sides? Dimensional analysis yields

$$\frac{\text{kg} \cdot \text{m}}{\text{s}^2} \cdot \text{m} = -\frac{\text{kg} \cdot \text{m}}{\text{s}^2 \cdot \text{m}^2} \cdot \text{m}^2 \cdot \text{m}^2 \cdot \frac{\text{m}}{\text{m}^2} = -\frac{\text{kg} \cdot \text{m}}{\text{s}^2} \cdot \text{m}.$$

Both sides are still equal! The dimensions of partial derivatives might be tricky sometimes, but now it was easy since $y(x, t)$ measures the deflection of the string in the y -coordinate direction.

In static equilibrium the difference of torque at the ends of the infinitesimal piece of string dx equals the force acting on length dx as

$$F dx = dM = \frac{\partial M}{\partial x} dx = -EA\kappa^2 \frac{\partial^3 y}{\partial x^3} dx,$$

where the dx cancel each other out from both sides of the equation.

Finally to reach the equation of motion, the difference of forces needs to be evaluated and made equal to the mass and acceleration of the string. This starts as

$$dF = F(x + dx) - F(x) = \frac{\partial F}{\partial x} dx = -EA\kappa^2 \frac{\partial^4 y}{\partial x^4} dx = \rho Adx \frac{\partial^2 y}{\partial t^2},$$

and leads to

$$-\frac{E\kappa^2}{\rho} \frac{\partial^4 y}{\partial x^4} = \frac{\partial^2 y}{\partial t^2}. \quad (2.57)$$

In equation (2.57) the displacement dependent term describes the elasticity of the material when external forces are applied to it. [23, p. 153]

Equation (2.57) can be attached to the basic wave equation (2.10) to get

$$c_T^2 \frac{\partial^2 y}{\partial x^2} - c_L^2 \kappa^2 \frac{\partial^4 y}{\partial x^4} = \frac{\partial^2 y}{\partial t^2}. \quad (2.58)$$

It is irrelevant in this context to try and solve this differential equation, although a solution can still be found with the method of separating the variables and solving the second and fourth order differential equations with respect to some boundary conditions.

Since at the moment there is an interest only in the general behaviour of frequencies for different modes of vibration, dispersion relations can be used to approximate the behaviour of this equation. The essential variables in dispersion relation equations are angular frequency ω and wave number k . These quantities can be used to calculate the wave velocity

$$c_p = \frac{\omega}{k},$$

which is also known as the phase velocity. A differential of the wave velocity leads to a concept of group velocity,

$$c_g = \frac{\partial \omega}{\partial k}.$$

The phase velocity refers to a velocity of a wave with a single frequency, whereas the group velocity is the velocity of an envelope containing a combination of several waves. The propagation of waves that have a nonlinear relation between ω and k is described to be dispersive. [18, p. 4]

In the case of a general solution for wave equation (2.10), the function

$$y(x, t) = A e^{i(kx - \omega t)}$$

depicts an extracted basis of a solution, a single sinusoidal waveform with an arbitrary amplitude constant A , identified with wave number k and angular frequency ω , moving along the string. By substituting this trial solution into the model of (2.58), the behaviour of the trial sinusoidal function is examined in the model of a stiff string. After the substitution, equation (2.58) looks like

$$-k^2 c_T^2 A e^{i(kx - \omega t)} - k^4 c_L^2 \kappa^2 A e^{i(kx - \omega t)} = -\omega^2 A e^{i(kx - \omega t)},$$

from where the minus signs and the trial functions cancel each other out. Now there exists a quadratic equation for k^2 , and using the quadratic formula,

$$k^2 = -\frac{c_T^2}{2c_L^2 \kappa^2} \pm \sqrt{\left(\frac{c_T^2}{2c_L^2 \kappa^2}\right)^2 + \frac{\omega^2}{c_L^2 \kappa^2}}. \quad (2.59)$$

For further analysis the root with the positive sign is chosen because the other root yields negative values. Negative values would make k imaginary, meaning an unphysical solution. With this choice of a positive sign, equation (2.59) simplifies to form

$$k^2 = -\frac{c_T^2}{2c_L^2 \kappa^2} + \frac{c_T^2}{2c_L^2 \kappa^2} \sqrt{1 + \frac{4c_L^2 \kappa^2 \omega^2}{c_T^4}},$$

and binomial approximation is applied to this form on the basis of the quotient term being small compared to unity. The approximation leads to

$$k^2 = -\frac{c_T^2}{2c_L^2 \kappa^2} + \frac{c_T^2}{2c_L^2 \kappa^2} \left(1 + \frac{1}{2} \frac{4c_L^2 \kappa^2 \omega^2}{c_T^4}\right),$$

and this simplifies greatly to

$$\frac{\omega}{k} = c_p \approx c_T,$$

meaning that at low frequencies the wave velocities c_p of the upper partials are directly related to the velocity of the transverse waves, just like in the ideal model.

For high frequencies the term with ω in equation (2.59) dominates, so the approximation is simply

$$k^2 \approx \sqrt{\frac{\omega^2}{c_L^2 \kappa^2}},$$

from where

$$\frac{\omega}{k} = c_p \approx \sqrt{c_L \kappa \omega}.$$

This result indicates that as the frequency grows, the more dispersive the wave motion in the string is.

Approximate solutions for the differential equation (2.58) have been derived by Morse and Ingard [24, p. 190]. The general effect modifying the equation for the harmonic frequencies (2.1) is that the stiffness property of the string alters the harmonic frequencies as

$$f_{Tn} \approx \frac{n}{2L} \sqrt{\frac{F_x}{\rho A}} \left[1 + \frac{2}{L} \sqrt{\frac{EA\kappa^2}{F_x}} + \left(4 + \frac{n^2\pi^2}{2} \right) \frac{EA\kappa^2}{F_x L^2} \right]. \quad (2.60)$$

This result can be visualised by plotting the equation (2.1) for the harmonic frequencies and equation (2.60) side by side for the initial modes of vibration.

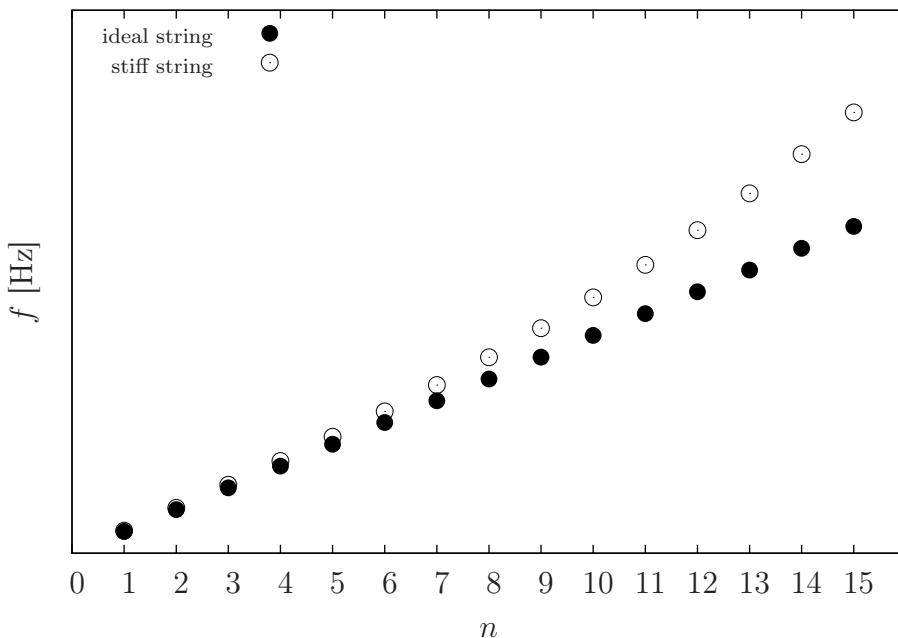


Figure 2.13: Harmonic frequencies versus nonharmonics

2.2.4 The effect of end supports

In the ideal case, the string was assumed to be perfectly clamped from both ends, so that the mechanical admittance of the boundary points was identically zero. This means that no energy can escape through the end supports, and that the supports do not move due to the motion of the string.

In reality a guitar string is only supported by the bridge and nut, where the string is not firmly clamped between massive elements to fasten the string. Due to their low mass and open structure, the end supports of a guitar might slightly follow the motion of the string and let the string move also at the top of the support. Additionally, the bridge and nut supporting the real string dissipate energy as they are mechanically coupled to the body of the guitar.

Figure 2.14 gives an idea of the forces that are affecting the string at the end supports.

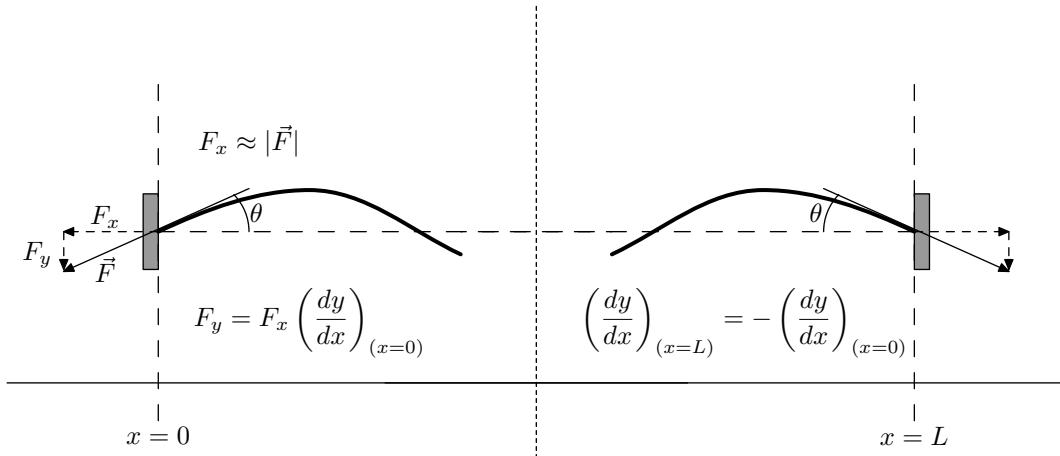


Figure 2.14: Forces at the boundary of the vibrating string

For small amplitude vibration, the angle θ between the string and the alignment of the support is small. Hence, the transverse force F_y can be taken as a linear approximation $F_x \frac{\partial y}{\partial x}$ as indicated in Figure 2.14. The transverse force is therefore defined as a fraction of the force F_x , which is the initial tension force to stretch the string to normal playing tension.

To get an idea of the behaviour of the transverse force of the string impacting the end supports, the equation of motion (2.35) of the plucked string needs to be differentiated with respect to x and evaluated at $x = 0$ or $x = L$ for the bridge and the nut respectively. The result of the differentiation is already presented in

equation (2.43), and evaluation at the bridge, $x = 0$, yields

$$\frac{\partial y}{\partial x} = \frac{2a}{\pi} \left(\frac{1}{p} + \frac{1}{L-p} \right) \sum_{n=1}^{\infty} \frac{1}{n} \sin \left(\frac{n\pi p}{L} \right) \cos \left(\frac{c_T n \pi t}{L} \right). \quad (2.61)$$

From here it follows that the linearised transverse force F_y has a magnitude

$$F_x \frac{\partial y}{\partial x} = F_x \frac{2a}{\pi} \left(\frac{1}{p} + \frac{1}{L-p} \right) \sum_{n=1}^{\infty} \frac{1}{n} \sin \left(\frac{n\pi p}{L} \right) \cos \left(\frac{c_T n \pi t}{L} \right). \quad (2.62)$$

It is interesting to note that when more and more harmonic frequency components are summed together in equation (2.62), the functional shape of the force at the bridge approaches a square wave as shown in Figure 2.15.

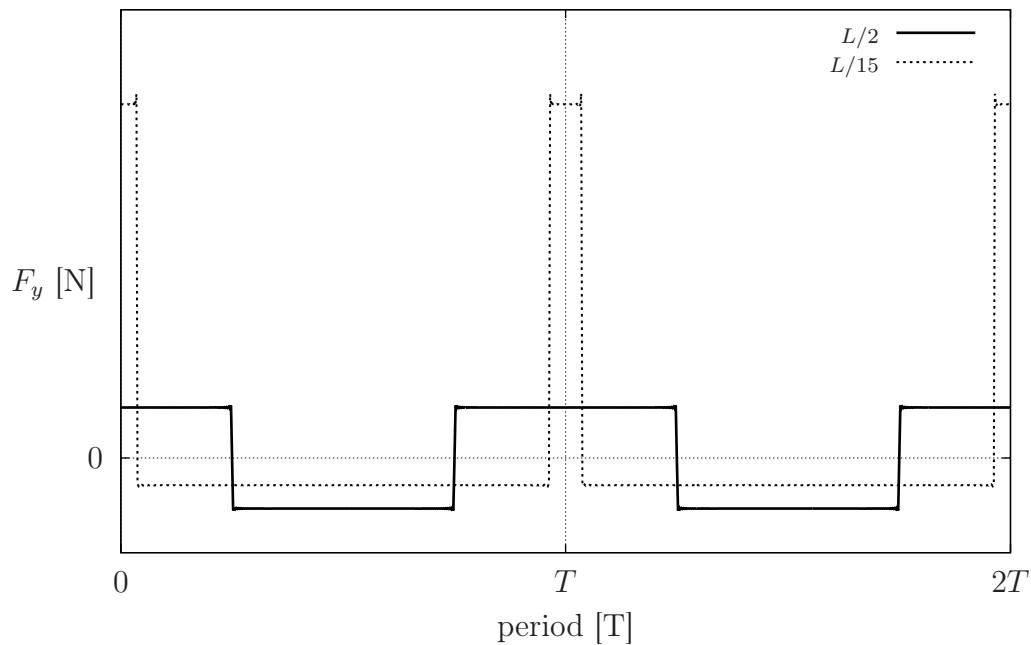


Figure 2.15: Transverse force at the bridge ($x = 0$) for different plucking locations

In Figure 2.15, the symmetric square wave describing the magnitude of the transverse force is a result from plucking the string at the middle. The dashed square wave is calculated from a pluck at distance $\frac{L}{15}$ measured from the bridge. The initial deflection is taken to be in the positive y -axis direction, as indicated in Figure 2.5. If the initial deflection is in the negative direction, then the force in Figure 2.15 also changes its sign. The resulting square wave is not that obvious when looking at equation (2.62), but when all the terms in the Fourier sum are added together, the waveform approaches a square wave with some tiny ripple at the edges. [25]

For the end supports at $x = 0$ and $x = L$ it is possible to define a complex-valued mechanical impedance

$$Z_m = R_m - iX_m = R_m - i \left(m\omega - \frac{k}{\omega} \right), \quad (2.63)$$

where the real part defines a mechanical resistance and the imaginary part defines mechanical reactance. The reactance of the end support can be classified to be related to mass m or stiffness (spring constant) k , which are the mechanical equivalents for inductance and capacitance respectively. The reason for using a minus sign in front of the reactance term follows from the fact that the convention of using $e^{-i\omega t}$ to describe the time dependency of wave motion will be adopted later in this section. Normally in electric engineering one uses $e^{j\omega t}$ as the time dependent factor so that the mechanical impedance would be expressed as $Z_m = R_m + jX_m$. [23, p. 133]

A relation of

$$Z_m = \frac{F}{c} \quad (2.64)$$

exists between mechanical impedance and force F via velocity c . When considering the end supports, the velocity term comes from the time derivative of the function for the string, and the force is the vertical component of the total force affecting the end support. Therefore, the relation of equation (2.64) can be written as

$$Z_m \left(\frac{\partial y}{\partial t} \right) = F_y \approx F_x \left(\frac{\partial y}{\partial x} \right). \quad (2.65)$$

As it was indicated during the derivation of the equation of motion for the ideal string, the resulting waveform can be represented as two similar waves travelling in opposite directions. Based on the discussion in section 2.1.6, the mathematical representation of these two waves in a certain linear combination can be written as

$$y(x, t) = \left(A_+ e^{\frac{i\omega}{c_T} x} + A_- e^{-\frac{i\omega}{c_T} x} \right) e^{-i\omega t}, \quad (2.66)$$

where the continuous waves are travelling to the positive and negative directions of the x -coordinate with complex-valued amplitudes A_+ and A_- respectively. Because of the boundary conditions, the sum of these waves should be 0 at $x = 0$ and $x = L$. Therefore it can be imagined that only one wave is moving between the boundaries and reflecting to the opposite direction from the boundary points. This way of thinking keeps the vibration inside the boundaries, although the mathematical description of the waves is periodically continuing to eternity.

The complex exponential function shown in equation (2.66) is just one linear combination of solutions to the wave equation (2.10). It should be noted that only the real part of the complex exponential representation is physically meaningful. The imaginary part should be carried along with the calculations, but in the end only the real part is extracted as the final result.

The boundary condition at both ends of the string should satisfy equation (2.64), in which the impedance is a complex quantity but not a vector, since the focus is on the impedance and force in the y -coordinate direction only. However, the boundary condition is not the same at locations $x = 0$ and $x = L$. This difference is depicted in Figure 2.14, where identical waves approach the boundaries from opposite directions. The slope of the string at $x = 0$ can be calculated normally in the positive x -direction, but at $x = L$ the slope of the identical wave calculated in the positive x -direction will give an opposite slope. [23, p. 134]

The slope at $x = 0$ is calculated as

$$\frac{f(x_0 + dx) - f(x_0)}{(x_0 + dx) - x_0} = \frac{f(x_0 + dx) - f(x_0)}{dx}$$

and similarly the identical slope is calculated at $x = L$ as

$$\frac{f(x_L - dx) - f(x_L)}{(x_L - dx) - x_L} = \frac{f(x_L - dx) - f(x_L)}{-dx} = -\frac{f(x_L - dx) - f(x_L)}{dx}.$$

The similarity of the identical waves of opposite directions states that $f(x_0) = f(x_L)$ and $f(x_0 + dx) = f(x_L - dx)$. Therefore,

$$\left(\frac{dy}{dx} \right)_{(x=L)} = - \left(\frac{dy}{dx} \right)_{(x=0)}.$$

This leads to the boundary conditions

$$\left(\frac{\partial y(x, t)}{\partial t} \right)_{(x=0)} = \frac{F_x}{Z_{m0}} \left(\frac{\partial y(x, t)}{\partial x} \right)_{(x=0)} \quad (2.67)$$

and

$$\left(\frac{\partial y(x, t)}{\partial t} \right)_{(x=L)} = -\frac{F_x}{Z_{mL}} \left(\frac{\partial y(x, t)}{\partial x} \right)_{(x=L)}. \quad (2.68)$$

Keeping in mind the solution of the wave equation as a sum of two separate waves moving in opposite directions, at the vicinity of the end support the form of the string is defined as the sum of the incident and reflected waves. This is modelled by equation

$$y(x, t) = \left(A_i e^{\frac{i\omega}{c_T} x} + A_r e^{-\frac{i\omega}{c_T} x} \right) e^{-i\omega t}, \quad (2.69)$$

where A_i is the amplitude of the incident wave and A_r is the amplitude of the reflected wave. The wave amplitudes A_+ and A_- from equation (2.66) are renamed in equation (2.69) as A_i and A_r to clarify the reflection process. Equation (2.69) is still the same as equation (2.66), but now it should be thought of as a one single wave reflecting from the end points. Assigning the rightwards moving wave as the incident wave chooses the boundary at $x = L$ as the point of reflection.

Since A_i is now approaching the boundary at $x = L$, equation (2.69) is substituted into the equation for the boundary condition (2.68). After differentiations on the left and right sides, this evaluates to

$$\left(A_i e^{\frac{i\omega}{c_T}x} + A_r e^{-\frac{i\omega}{c_T}x} \right) = \frac{F_x}{c_T Z_{mL}} \left(A_i e^{\frac{i\omega}{c_T}x} - A_r e^{-\frac{i\omega}{c_T}x} \right).$$

A simplification of this result gives

$$A_r \left(\frac{F_x}{c_T Z_{mL}} + 1 \right) e^{-2\frac{i\omega}{c_T}x} = A_i \left(\frac{F_x}{c_T Z_{mL}} - 1 \right),$$

where the exponential term equals 1 at $x = L$ and $x = 0$. A further simplification leads to the quotient

$$\frac{A_r}{A_i} = a + ib = \frac{1 - z}{1 + z}, \quad (2.70)$$

where

$$z = \frac{Z_m}{\rho A c_T}$$

describes the relation of the impedances of the end support Z_m and the nominal mechanical impedance of the string

$$\rho A c_T = \frac{F_x}{c_T} = Z_s.$$

Normally $Z_m > Z_s$, so the derived result simply indicates that the reflection from an unideal end support reduces the amplitude of the wave and causes phase changes to a sinusoidal wave moving along the string. The phase changes appear because the quotient (2.70) has a real and imaginary part, and the phase is calculated in relation to this complex quantity.

If the problem is analysed to the other direction, the amplitudes A_i and A_r change places in equation (2.69) because now the incident wave travels to the negative direction. In this case the boundary condition (2.67) is used for substitution.

The preceding analysis considered only the case where incident and reflected waves consisted only of a single frequency component with wave number k and angular frequency ω . A real plucked string has all of its harmonic frequencies

summed as a single waveform, so the situation is a bit different in that case. To take account of all n frequency components, one just needs to replace the single ω in equation (2.69) with ω_n .

Generally the characteristic impedance of the string and the impedance of the end support vary with frequency. In this case all the wave components in the plucked string would experience a different impedance both at the string and at the end support. Each frequency component would have a different phase and amplitude coefficient. One would have to calculate the reflection properties for all frequency components separately, and at the end sum them all together to form a model of the reflected wave. It would be nice to draw a picture of the shape of a plucked string at the vicinity of an end support that has finite impedance, but it might be a complicated task.

At least one thing is clear; the unideal end supports quickly mess up the triangular form of the two travelling wave components of the plucked string because all the frequency components go out of phase and their amplitude relations change. In addition, unideal end supports also alter the nominal frequencies of the harmonic components of vibration. To prove this, one needs to resolve the wave equation (2.10) using the boundary conditions (2.67) and (2.68). These boundary conditions can be precalculated to a certain degree using the fact that the time derivative of equation (2.69) is $-i\omega_n y(x, t)$, so that the boundary conditions become

$$y(0, t) = -\frac{F_x}{i\omega_n Z_{m0}} \left(\frac{\partial y(x, t)}{\partial x} \right)_{(x=0)} \quad (2.71)$$

and

$$y(L, t) = \frac{F_x}{i\omega_n Z_{mL}} \left(\frac{\partial y(x, t)}{\partial x} \right)_{(x=L)}. \quad (2.72)$$

Now one needs to redefine the factor k in equation (2.17).

Earlier it was noticed that if $k = -p^2$, then the auxiliary equation $r^2 + p^2 = 0$ gives the imaginary roots $r_1 = r_2 = \pm ip$. The general solution of (2.17) is then $F(x) = Ae^{ipx} + Be^{-ipx}$, which can also be written as

$$F(x) = A \cos px + B \sin px.$$

To redetermine the values for the constants A , B and p , the solution is re-evaluated with the boundary conditions of unideal supports. When $x = 0$,

$$F(0) = A \cos 0 + B \sin 0 = -p \frac{F_x}{i\omega_n Z_{m0}} (-A \sin 0 + B \cos 0),$$

so that

$$A = -p \frac{F_x}{i\omega_n Z_{m0}} B.$$

Then at $x = L$,

$$-p \frac{F_x}{i\omega_n Z_{m0}} B \cos pL + B \sin pL = p^2 \frac{F_x}{i\omega_n Z_{mL}} \frac{F_x}{i\omega_n Z_{m0}} B \sin pL + p \frac{F_x}{i\omega_n Z_{mL}} B \cos pL.$$

According to Morse [23, p. 142], one can assume that $\frac{F_x}{i\omega_n Z_m}$ is very small and the square of this term can be neglected in approximation. To satisfy the boundary conditions, the factor p needs to be solved from

$$-p \frac{F_x}{i\omega_n Z_{m0}} \cos pL - p \frac{F_x}{i\omega_n Z_{mL}} \cos pL + \sin pL = 0.$$

Using again the argument that pL will make the value of sine approach zero and the cosine approach to one, a linear approximation of these trigonometric functions can be used. Hence,

$$-p \frac{F_x}{i\omega_n Z_{m0}} - p \frac{F_x}{i\omega_n Z_{mL}} + pL = 0.$$

To make this equation true in the linear approximation of sine and cosine, the impedance terms must disappear to leave only $pL = 0$. To still fulfil this condition that the sine will be zero for all integers n , one needs to define p as

$$pL \left(1 + \frac{F_x}{i\omega_n Z_{m0} L} + \frac{F_x}{i\omega_n Z_{mL} L} \right)$$

and apply this to the equation $pL = n\pi$ as before. Because we now have the result $k = -p^2$, the equation (2.18) can be used to solve $G(t)$.

When substituting k with $-p^2$ into equation (2.18)

$$\frac{d^2G}{dt^2} + \frac{c_T^2 n^2 \pi^2}{L^2} \left(1 + \frac{F_x}{i\omega_n Z_{m0} L} + \frac{F_x}{i\omega_n Z_{mL} L} \right)^2 G = 0,$$

and this differential equation leads to the auxiliary roots

$$r_{1,2} = \pm i \left(\omega_n + \frac{F_x}{iZ_{m0} L} + \frac{F_x}{iZ_{mL} L} \right). \quad (2.73)$$

To make this result plausible, the definition of mechanical admittance is

$$Y_m = \frac{1}{Z_m} = \frac{1}{R_m - iX_m} = \frac{R_m + iX_m}{R_m^2 + X_m^2},$$

following the definition of mechanical impedance given in equation (2.63). Substituting the expression of mechanical admittance into equation (2.73) leads to the roots

$$r_{1,2} = \pm i \left[\omega_n + \frac{F_x}{L} \left(\frac{X_{m0} - iR_{m0}}{R_{m0}^2 + X_{m0}^2} \right) + \frac{F_x}{L} \left(\frac{X_{mL} - iR_{mL}}{R_{mL}^2 + X_{mL}^2} \right) \right].$$

These roots give the set of solutions

$$y_n(x, t) = G(t)F(x) = e^{-\lambda t} (C_n \cos \Omega_n t + D_n \sin \Omega_n t) F(x),$$

where the damping term

$$\lambda = \frac{F_x}{L} \left(\frac{R_{m0}}{R_{m0}^2 + X_{m0}^2} + \frac{R_{mL}}{R_{mL}^2 + X_{mL}^2} \right)$$

and the modified angular frequency term

$$\Omega_n = \omega_n + \frac{F_x}{L} \left(\frac{X_{m0}}{R_{m0}^2 + X_{m0}^2} + \frac{X_{mL}}{R_{mL}^2 + X_{mL}^2} \right).$$

As a general statement, it can be concluded that if the impedance of the end support is equivalent to stiffness (capacitance), $X_m = -\frac{k}{\omega}$, the upper partials are lowered in comparison to the ideal situation. This is because of the minus sign in front of the stiffness related reactance term. The stiffness controlled end support moves in phase with the string's motion. If the end support is equivalent to mass (inductance), $X_m = m\omega$, and the motion of the end support is out of phase with the string and the upper partials are raised. [23, p. 146]

The results also show that mechanical resistance R_m of the support causes a damping term to the equation of motion of a vibrating string. This is also an intuitively correct result and it is nice to notice that it is in alignment with the mathematical evaluation. The factors that lead to specific decay times are presented by Fletcher [25]. The equation (2.64) can be used to give an approximate indication of the quantities that affect the decay times in the end support.

The energy gained when deflecting the string from rest to the initial triangular shape equals the work done over the deflection amplitude a , namely

$$E_0 = a F_x \frac{dy}{dx} \propto a^2 F_x \left(\frac{1}{p} + \frac{1}{L-p} \right), \quad (2.74)$$

where only the dimensional variables have been included in the calculation. With the help of equation (2.1), the force term F_x can be expanded to a group of relevant dimensional variables. Taking along only the fundamental frequency f_1 , the force

$$F_x = f_1^2 4L^2 \pi r^2 \rho, \quad (2.75)$$

and this can be substituted into equation (2.74), which then reads

$$E_0 \propto a^2 f_1^2 L^2 r^2 \rho \left(\frac{1}{p} + \frac{1}{L-p} \right),$$

retaining only the variables with decent dimensional properties, so that the units on both sides of the equation are still equal.

According to article [25], the rate of energy loss at the bridge is

$$\frac{dE}{dt} = -c^2 Z_m,$$

where the velocity term c^2 is intended to be evaluated from the expression (2.32). Fortunately the velocity can also be written using equation (2.62) and equation (2.64) as

$$c^2 = \frac{F_x^2}{Z_m^2} \frac{4a^2}{\pi^2} \left(\frac{1}{p} + \frac{1}{L-p} \right)^2 \frac{1}{n^2} \sin^2 \left(\frac{n\pi p}{L} \right) \cos^2 \left(\frac{c_T n\pi t}{L} \right).$$

In accordance with the previous equations, only the velocity of the fundamental frequency is needed for the analysis. Taking again only the dimensionally relevant terms and using equation (2.75) to expand the force term, the rate of energy loss

$$\begin{aligned} \frac{dE}{dt} &\propto -\frac{1}{Z_m} a^2 f_1^4 L^4 r^4 \rho^2 \left(\frac{1}{p} + \frac{1}{L-p} \right)^2 \\ &\propto -\frac{E_0}{Z_m} f_1^2 r^2 \rho \left(\frac{L^2}{p} + \frac{L^2}{L-p} \right). \end{aligned}$$

From here it is possible to solve the energy relation dE/E_0 , and the remaining set of variables has the unit of s^{-1} . The associated decay time constant is the inverse of this and has the form

$$\tau_3 \propto \frac{Z_m}{f_1^2 r^2 \rho \left(\frac{L^2}{p} + \frac{L^2}{L-p} \right)}.$$

According to this result, the rate of energy loss at the supports depends on the frequency and also the dimensions and material properties of the string. The decay time varies as f^{-2} and this becomes noticeable in situations where the end support is relatively light compared to the mass of the string or when a finger tip is used to stop the string on a fretless bass guitar, for example.

According to another study by Legge and Fletcher [30], unideal end supports are also guilty of coupling energy to harmonic frequency components that are initially at zero amplitude in the spectrum of a vibrating string. The experimental method used a flexible end support and the string was plucked at certain node locations. The observations showed that after the string was released to vibrate, the frequencies originally damped by the effect of the plucking position gained

energy in a timescale of nearly one second and then started to decay. Moreover, energy transfer between all modes of vibration happens more commonly when the end supports are not ideally rigid.

2.2.5 Coupling between directions of vibration

Although the effect of large amplitude string vibrations in the transverse direction was noted to generate longitudinal vibration, it is not the main reason for energy transfer, i.e. coupling, between the transverse and longitudinal modes of vibration.

To focus more comprehensively on the transverse vibrations of a nonlinear string, the results leading to equation (2.49) can be used to construct a more general equation on motion,

$$\frac{\partial^2 \vec{R}}{\partial t^2} + \lambda \frac{\partial \vec{R}}{\partial t} - c_T^2 \left(1 + \frac{1}{2\Delta L} \int_0^L \left| \frac{\partial \vec{R}}{\partial x} \right|^2 dx \right) \frac{\partial^2 \vec{R}}{\partial x^2} = 0. \quad (2.76)$$

The effects of large amplitude vibration are taken into account in equation (2.76) by adding an integral term, which with reference to equation (2.49)

$$\frac{1}{2\Delta L} \int_0^L \left| \frac{\partial \vec{R}}{\partial x} \right|^2 dx = \frac{1}{2} \frac{\delta L}{\Delta L}.$$

This term increases the wave velocity c_T of the transverse waves during the vibration period. The increase is proportional to the stretching of the string by amount δL as a fraction of the stretch ΔL , which measures the initial tension applied to the string to achieve some frequency of vibration. Equation (2.76) is therefore a better approximation than the basic wave equation (2.10) derived from the ideal string model. Equation (2.76) also includes a general damping term λ , which describes the energy losses during the vibration. [31]

The general solution to the differential equation (2.76) would be a Fourier series expansion from the string geometry, where the position vector of the string

$$\vec{R}(x, t) = \sum C_n(t) \sin \left(\frac{n\pi x}{L} \right).$$

However, this is not the exact solution because there does not exist an analytic solution to equation (2.76). A simpler approach is to examine only the solutions for the fundamental frequency component, namely

$$\vec{R}(x, t) = \vec{r}(t) \sin \left(\frac{\pi x}{L} \right).$$

This way all coupling to the upper partials is ignored, and with the help of intermediate results,

$$\frac{\partial^2 \vec{R}(x, t)}{\partial x^2} = \frac{\partial^2}{\partial x^2} \vec{r}(t) \sin\left(\frac{\pi x}{L}\right) = -\vec{r}(t) \sin\left(\frac{\pi x}{L}\right) \frac{\pi^2}{L^2}$$

and

$$\frac{1}{2\Delta L} \int_0^L \sqrt{(\vec{r} \bullet \vec{r}) \cos^2\left(\frac{\pi x}{L}\right) \frac{\pi^2}{L^2}}^2 dx = (\vec{r} \bullet \vec{r}) \frac{\pi^2}{L^2} \frac{1}{4\Delta L} \int_0^L 1 + \cos\left(\frac{2\pi x}{L}\right) dx,$$

the analysed equation (2.76) reduces to the form

$$\frac{\partial^2 \vec{r}}{\partial t^2} + \lambda \frac{\partial \vec{r}}{\partial t} + \omega_0^2 (1 + K \vec{r}^2) \vec{r} = 0, \quad (2.77)$$

where $\omega_0^2 = \frac{c_T^2 \pi^2}{L^2}$ and $K = \frac{\pi^2}{L} \frac{1}{4\Delta L}$.

According to article [31], there are only a few special cases where a stable solution to this differential equation can be found in a stationary frame of reference, which is also known as the standard inertial reference frame. The main idea is to look for solutions in a rotating frame of reference, more generally known as a non-inertial reference frame, which enables to find a general numerical solution to equation (2.77). To understand the concept of a rotating frame of reference, some basic introduction might be needed at this point. The theoretical foundations of a rotating reference frame presented here are adapted quite directly from the textbook of Marion and Thornton [32].

The connection between a stationary reference frame and a rotating reference frame is visualised in Figure 2.16, where the stationary, or fixed, frame is denoted with the coordinate axes x_f, y_f and z_f , and the axes of the moving frame are identified with the normal x, y and z coordinates. With the notations used in Figure 2.16, the point P can be located from the fixed frame with the vector

$$\vec{r}_f = \vec{R} + \vec{r}, \quad (2.78)$$

where vector \vec{R} connects together the origins of the two coordinate systems.

The goal is to derive expressions that allow to locate point P with vectors from both coordinate systems. In this situation the frame with the x, y and z coordinates is rotating and point P has the position vector

$$\vec{r}(x, y, z, t) = a\hat{i} + b\hat{j} + c\hat{k}, \quad (2.79)$$

in the rotating coordinate system. The fundamental question is that if the coefficients a, b, c are all functions of place and time, how to express the motion of

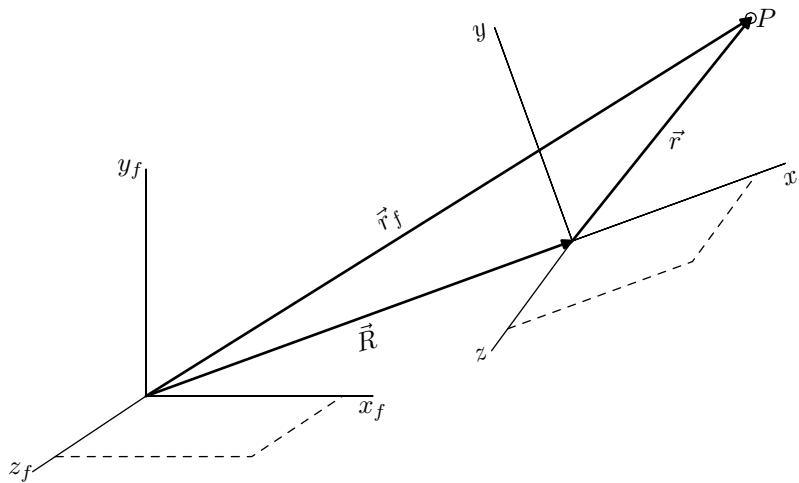


Figure 2.16: Stationary (inertial) and rotating (non-inertial) frames of reference

point P from the view point of the stationary coordinates x_f, y_f, z_f ? Differentiating the position vector (2.79) with respect to time,

$$\begin{aligned} \left(\frac{d\vec{r}}{dt} \right)_{fixed} &= \frac{d}{dt} (a\hat{i} + b\hat{j} + c\hat{k}) = \frac{da}{dt}\hat{i} + \frac{db}{dt}\hat{j} + \frac{dc}{dt}\hat{k} + a\frac{d\hat{i}}{dt} + b\frac{d\hat{j}}{dt} + c\frac{d\hat{k}}{dt} \\ &= \left(\frac{d\vec{r}}{dt} \right)_{rot} + a\frac{d\hat{i}}{dt} + b\frac{d\hat{j}}{dt} + c\frac{d\hat{k}}{dt} \end{aligned}$$

because this gives the rate of change in the stationary frame of reference and the orthogonal unit vectors are also moving with respect to the stationary frame. The subscripts *fixed* and *rot* refer to the coordinate systems where the specific quantity is evaluated. It is essential to keep in mind that vector \vec{r} pointing to P from the origin of the rotating reference frame can be measured in the rotating frame as well as in the fixed frame and the subscripts are used to clarify that. At this point it is still a bit unclear how to evaluate the derivatives of the unit vectors \hat{i}, \hat{j} and \hat{k} .

Figure 2.17 depicts a situation where an infinitesimal rotation vector $\delta\vec{\theta}$ is turned around its axis, so that a position vector \vec{r} attached rigidly to some coordinates is rotated by the amount of angle $\delta\theta$. The goal is to express the small change $\delta\vec{r}$ along curve s in terms of the rotation vector $\delta\vec{\theta}$. A cross-product $\delta\vec{\theta} \times \vec{r}$ is possible, but only when $\delta\vec{r}$ is orthogonal to the directions of $\delta\vec{\theta}$ and \vec{r} . Figure 2.17 is deliberately drawn to show that a large rotation of angle $\delta\theta$ does not provide an orthogonal $\delta\vec{r}$. The orthogonality condition is satisfied only when $\delta\vec{r}$ is tangential to the circumference s of the rotation, that is, when the angle of

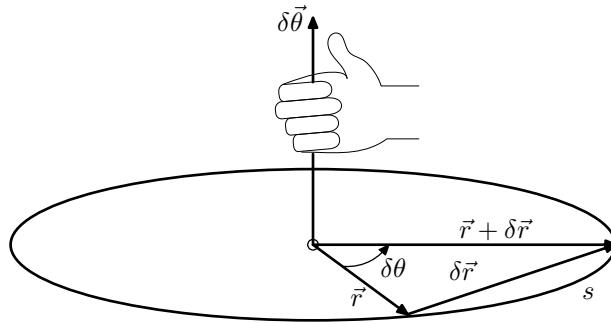


Figure 2.17: Infinitesimal increment $\delta\vec{r}$ by rotation $\delta\vec{\theta}$

rotation is an infinitesimally small differential quantity $\delta\theta$. Then the expression

$$\delta\vec{r} = \delta\vec{\theta} \times \vec{r}$$

is valid to describe the movement of \vec{r} . Based on this jewel of information, the derivative $\frac{d\hat{i}}{dt} = \frac{d\vec{\theta}}{dt} \times \hat{i}$, and similarly to other orthogonal directions. Therefore,

$$\left(\frac{d\vec{r}}{dt} \right)_{fixed} = \left(\frac{d\vec{r}}{dt} \right)_{rot} + \frac{d\vec{\theta}}{dt} \times (a\hat{i} + b\hat{j} + c\hat{k}) = \left(\frac{d\vec{r}}{dt} \right)_{rot} + \vec{\omega} \times \vec{r}, \quad (2.80)$$

where $\vec{\omega}$ is the angular velocity, i.e., the rate of change of the rotation.

By differentiating the equation (2.78) with respect to time,

$$\begin{aligned} \left(\frac{d\vec{r}_f}{dt} \right)_{fixed} &= \left(\frac{d\vec{R}}{dt} \right)_{fixed} + \left(\frac{d\vec{r}}{dt} \right)_{fixed} \\ &= \left(\frac{d\vec{R}}{dt} \right)_{fixed} + \left(\frac{d\vec{r}}{dt} \right)_{rot} + \vec{\omega} \times \vec{r}, \end{aligned}$$

an expression of velocities

$$\vec{c}_f = \vec{C}_f + \vec{c}_r + \vec{\omega} \times \vec{r} \quad (2.81)$$

is obtained. Velocity \vec{c}_f in equation (2.81) is the velocity relative to the stationary (fixed) axes, \vec{C}_f is the linear velocity of the origin of the rotating reference frame, \vec{c}_r is the velocity term relative to the rotating axes, $\vec{\omega}$ is the angular velocity of the rotating axes and $\vec{\omega} \times \vec{r}$ is the velocity due to the rotation of the moving axes. [32, pp. 388-391]

From the expressions of velocity it is possible to differentiate once more with respect to time, giving the equation in terms of acceleration \vec{a} . The expression of acceleration combined with mass m leads to an expression of force

$$\vec{F} = m \left(\frac{d\vec{c}_f}{dt} \right)_{fixed} = m\vec{a}_f.$$

This needs to be done to adapt the familiar Newton's law $\vec{F} = m\vec{a}$ to the non-inertial reference frame. The basic Newton's law is valid only when identifying point P with position vector \vec{r}_f , which is drawn from the origin of the inertial frame. To take the non-inertial frame into use, one needs to express the force using the position vector \vec{r} , which is drawn from the origin of the non-inertial frame. The view to the non-inertial frame can be taken either from the inertial or non-inertial frame as seen later.

The differentiation of velocity equation (2.81) is written explicitly as

$$\left(\frac{d\vec{c}_f}{dt} \right)_{fixed} = \left(\frac{d\vec{C}_f}{dt} \right)_{fixed} + \left(\frac{d\vec{c}_r}{dt} \right)_{fixed} + \frac{d\vec{\omega}}{dt} \times \vec{r} + \vec{\omega} \times \left(\frac{d\vec{r}}{dt} \right)_{fixed},$$

where

$$\begin{aligned} \left(\frac{d\vec{c}_r}{dt} \right)_{fixed} &= \left(\frac{d\vec{c}_r}{dt} \right)_{rot} + \vec{\omega} \times \vec{c}_r \\ \vec{\omega} \times \left(\frac{d\vec{r}}{dt} \right)_{fixed} &= \vec{\omega} \times \left(\frac{d\vec{r}}{dt} \right)_{rot} + \vec{\omega} \times (\vec{\omega} \times \vec{r}). \end{aligned}$$

After notational simplification,

$$\vec{a}_f = \vec{A}_f + \vec{a}_r + \frac{d\vec{\omega}}{dt} \times \vec{r} + 2\vec{\omega} \times \vec{c}_r + \vec{\omega} \times (\vec{\omega} \times \vec{r}). \quad (2.82)$$

The equation (2.82) creates a simple relation to the acceleration terms observed in the stationary frame \vec{a}_f and rotating frame \vec{a}_r together. The essential thing is that both accelerations are now expressed using the position vector \vec{r} of the rotating frame of reference, which enables to change the viewpoint between the reference frames quite easily. With these definitions of acceleration, an observer in the stationary frame experiences a force $m\vec{a}_f$ and an observer in the rotating frame experiences a force $m\vec{a}_r$, where the expression for \vec{a}_r can be solved from equation (2.82). In both cases the actual motion is happening within the rotating frame and it is identified by the position vector \vec{r} of the rotating frame.

The most relevant terms arising from equation (2.82) are the *Coriolis force* term $2m\vec{\omega} \times \vec{c}_r$ and the *centrifugal force* term $m\vec{\omega} \times (\vec{\omega} \times \vec{r})$. The centrifugal force reduces to form $-m\omega^2\vec{r}$ via the vector triple product rule

$$\vec{\omega} \times (\vec{\omega} \times \vec{r}) = (\vec{\omega} \bullet \vec{r})\vec{\omega} - (\vec{\omega} \bullet \vec{\omega})\vec{r} = -\omega^2\vec{r}$$

if $\vec{\omega}$ has a direction that is normal to the radius vector. [32, p. 392]

Now there is enough background information from rotating reference frames to proceed with the equation of motion for the coupled nonlinear string. The

position vector to draw the motion of the string is now moved from the stationary reference frame to a rotating reference frame by placing the acceleration term substitution

$$\vec{a}_f = \vec{a}_r + 2\vec{\omega} \times \vec{c}_r - \omega^2 \vec{r}$$

$$\left(\frac{\partial^2 \vec{r}}{\partial t^2} \right)_{fixed} = \left(\frac{\partial^2 \vec{r}}{\partial t^2} \right)_{rot} + 2\vec{\omega} \times \left(\frac{\partial \vec{r}}{\partial t} \right)_{rot} - \omega^2 \vec{r}$$

into equation (2.77). Due to this substitution, the resulting partial differential equation

$$\frac{\partial^2 \vec{r}}{\partial t^2} + \lambda \frac{\partial \vec{r}}{\partial t} + 2\vec{\Omega} \times \frac{\partial \vec{r}}{\partial t} - \Omega^2 \vec{r} + \omega_0^2 (1 + K\vec{r}^2) \vec{r} = 0 \quad (2.83)$$

has the necessary terms of the Coriolis force and centrifugal force included and the position vector \vec{r} is measured in the rotating reference frame. Since the substituted acceleration term is the one that describes the forces in the stationary frame of reference, the results obtained from this partial differential equation will depict the motion seen by the eyes of a person attached to the fixed inertial frame. Note that the angular velocity regarding the rotating frame is written with $\vec{\Omega}$, because the small ω_0 was already reserved for the angular velocity of the transverse vibrations. Obviously the direction of $\vec{\Omega}$ is perpendicular to the radius vector. To make it more clear,

$$\vec{\Omega} = \Omega_x \hat{i} \quad \text{and} \quad \vec{r} = Y \hat{j} + Z \hat{k}.$$

Regarding both transverse components, the differential equation (2.83) can be separated into two coupled differential equations:

$$\frac{\partial^2 Y}{\partial t^2} + \lambda \frac{\partial Y}{\partial t} - 2\Omega \frac{\partial Z}{\partial t} + \{-\Omega^2 + \omega_0^2[1 + K(Y^2 + Z^2)]\}Y \quad (2.84)$$

$$\frac{\partial^2 Z}{\partial t^2} + \lambda \frac{\partial Z}{\partial t} + 2\Omega \frac{\partial Y}{\partial t} + \{-\Omega^2 + \omega_0^2[1 + K(Y^2 + Z^2)]\}Z. \quad (2.85)$$

This coupled second order differential equation pair can be solved and simulated numerically, using basic methods described by Kreyszig [11, pp. 902–908]. The simplest approach is to apply the Euler method, which is intended for solving first order differential equations numerically. It is also possible to extend the Euler method to second order systems by adding a second pair of variables to denote the derivatives of the first variables. Using the quantities Y and Z from

equations (2.84) and (2.85), the extension yields

$$y_1 = Y, \quad y_2 = \frac{dY}{dt} \quad \Rightarrow \quad \frac{dy_1}{dt} = y_2, \quad (2.86)$$

$$z_1 = Z, \quad z_2 = \frac{dZ}{dt} \quad \Rightarrow \quad \frac{dz_1}{dt} = z_2. \quad (2.87)$$

With similar reasoning, the Euler method can eventually be extended to cover differential equations of any degree. When applying this numerical method to equations (2.84) and (2.85), the following set of equations is obtained:

$$\dot{y}_1 = y_2 \quad (2.88)$$

$$\dot{y}_2 = -\lambda y_2 - 2\Omega z_2 - \{-\Omega^2 + \omega_0^2[1 + K(y_1^2 + z_1^2)]\}y_1 \quad (2.89)$$

$$\dot{z}_1 = z_2 \quad (2.90)$$

$$\dot{z}_2 = -\lambda z_2 + 2\Omega y_2 - \{-\Omega^2 + \omega_0^2[1 + K(y_1^2 + z_1^2)]\}z_1. \quad (2.91)$$

In the preceding equations, the notations \dot{y} and \dot{z} have been used to denote the time derivative of the respective variables.

By choosing suitable values for the parameters K , λ , Ω and ω_0 , different kinds of orbital precession trajectories can be simulated. Figures 2.18 and 2.19 visualise one clear coupling situation, which in practise can happen only with loosely tightened and very flexible strings.

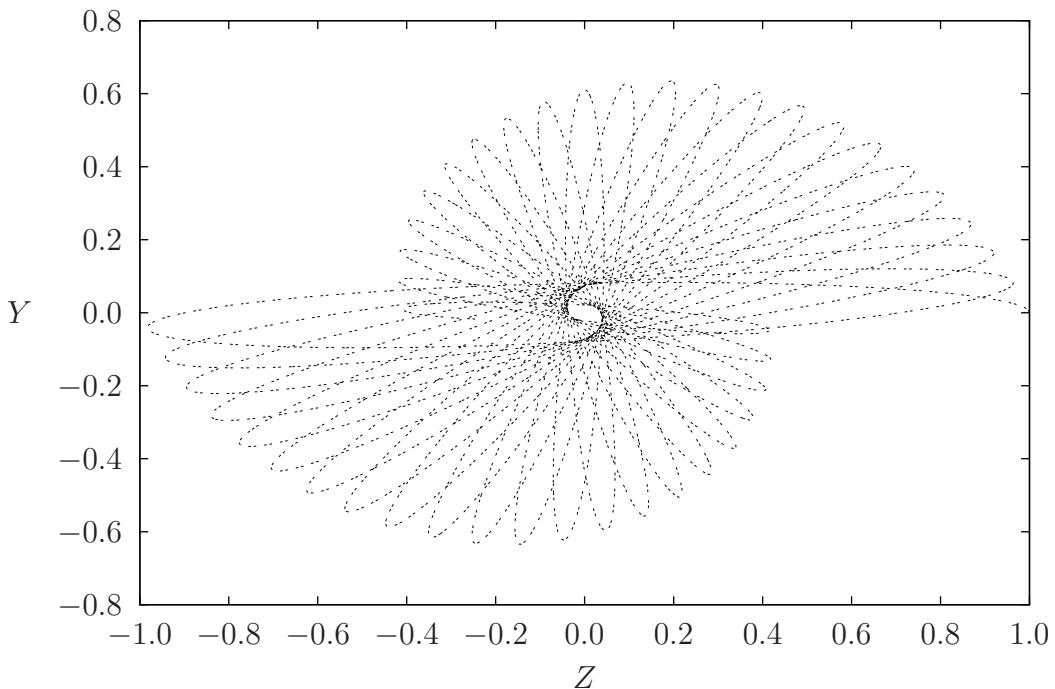


Figure 2.18: A simulated precession trajectory of a nonlinear string vibration

For the sake of clarity, Figure 2.18 shows only the beginning of the vibration

sequence that creates the 'oscilloscope trace' of the waveform shown in Figure 2.19. The precession trajectory is depicted as it would be seen by the eyes of a person attached to the stationary frame, as the substituted acceleration term described the acceleration experienced by a point in the stationary frame. If the motion in the $\{yz\}$ -plane would be observed as attached to the rotating frame, the string would have a constant elliptic trajectory. The preceding motion of the rotating frame at angular frequency Ω is seen by the viewer in the stationary frame and this sums up to be observed as a precessing elliptical motion of the string. Analysis in the inertial frame would only reveal one solution as a special case, but the rotating frame analysis leads to a general solution with numerous different trajectory variations.

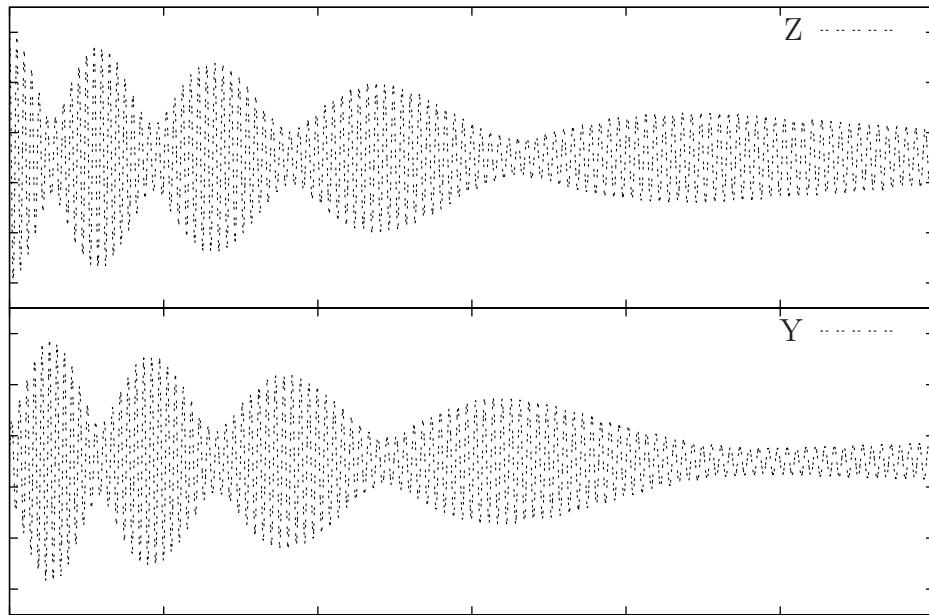


Figure 2.19: Coupling of vibrational modes on yz -plane motion

Figure 2.19 is a projected view from the rotating motion of the string to indicate the vibration amplitudes in both orthogonal transverse directions as a function of time.

Figures 2.18 and 2.19 have been simulated in this context as a trial to reproduce the results obtained by Gough [31]. The units used in the simulation have been scaled to unity measures as in the original experiments. It might be difficult to create a realistic coupling scheme with this simulation model in the first trials. The numerical calculations need to be written with a fast programming language such as C or Fortran because the numerical iterations of differential equations

require many calculational steps.

The coupling between two transverse polarisation modes is obviously not the only coupling mechanism in strings. As already discussed in section 2.2.1, the transverse waves stretch the string longitudinally during vibration so that the tension changes with twice the frequency of the transverse vibration. This longitudinal motion will be periodic in an ideal case, so that the transverse modes can actually create longitudinal vibration with frequencies $2\omega_n$ in situations where the longitudinal motion is originally not present in the string vibration.

Furthermore, as already briefly mentioned in the discussion about the unideal string supports, additional coupling through the support structure of the string can even excite the longitudinal modes of $2\omega_n + \omega_n$, which can lead to very complex multi-coupling combinations between different frequencies. In the ideal case the frequencies remain integer multiples but other nonlinearities can also destroy the harmonic nature of the vibratory motion. [30]

2.3 NONLINEAR EQUATIONS OF MOTION

The wave equation obtained from the model of the ideal string is actually only a reduced and linearised case of a general three-dimensional model. Among other researchers, this model was introduced by Morse and Ingard [24, pp. 856–863] and it is taken as the de facto standard differential equation to model nonlinear string motion. The vectorised differential equation removes the restriction for small deflection amplitudes by allowing the tension of the string to change dynamically. If the vectorised differential equation is split between each coordinate direction, the equation becomes a set of three coupled differential equations. Morse and Ingard also derive an approximate set of differential equations correct to the third order, meaning that the differential terms appearing in the equation are raised to the power of three or lower. This section reveals how to set up the nonlinear differential equations and derive the third order approximation.

The primary goal is to vectorise the equation of force for all three dimensions x , y , and z . This was already partially done when the differential equation for the ideal string was derived in section 2.1.3, where the force was written in the form

$$\vec{F} = F_x \hat{i} + F_y \hat{j} = F_x \hat{i} + F_x \left(\frac{dy}{dx} \right) \hat{j}. \quad (2.92)$$

This implies that the force in the y -coordinate direction is defined as a fraction of the force in the x -coordinate direction, and the magnitude of the transverse

force F_y depends directly on the geometrical slope at each point along the string. The vectorised force describing the total force has a direction to the tangent line of the string for each value of x .

What if the total magnitude of the force and its direction would be separated in two different equations? A position vector to draw the string in the $\{x, y\}$ coordinate plane can be defined as

$$\vec{r}(x, t) = x \hat{i} + y(x, t) \hat{j},$$

and the slope at each point along the x -coordinate direction is then

$$\frac{d\vec{r}}{dx} = 1 \hat{i} + \frac{dy}{dx} \hat{j}. \quad (2.93)$$

The differential of the position vector now points towards the tangent of the string for all values of x . Still what is needed is a unit vector of (2.93), and that is defined as

$$\hat{s} = \frac{d\vec{r}}{\left| \frac{d\vec{r}}{dx} \right|} = \frac{1 \hat{i} + \frac{dy}{dx} \hat{j}}{\sqrt{1 + \left(\frac{dy}{dx} \right)^2}}. \quad (2.94)$$

A unit vector always has a magnitude of unity, so it can be used to multiply other magnitudes without affecting the absolute value.

The magnitude of \vec{F} in equation (2.92) is

$$\left| \vec{F} \right| = \sqrt{\vec{F} \bullet \vec{F}} = F_x \sqrt{1 + \left(\frac{dy}{dx} \right)^2},$$

and the combination

$$\left| \vec{F} \right| \hat{s} = F_x \hat{i} + F_x \left(\frac{dy}{dx} \right) \hat{j}$$

just like in equation (2.92). Neat.

The derivation of the general equation of motion starts by defining a three-dimensional position vector

$$\vec{r}(x, t) = (x + \xi(x, t)) \hat{i} + y(x, t) \hat{j} + z(x, t) \hat{k}, \quad (2.95)$$

which can be used to draw the form of the string in $\{x, y, z\}$ -coordinates as a function of the x -coordinate at any given time t . A visualisation of the meaning of equation (2.95) is presented in Figure 2.20, where the displacement of the string has been magnified to make the position vector stand out more clearly.

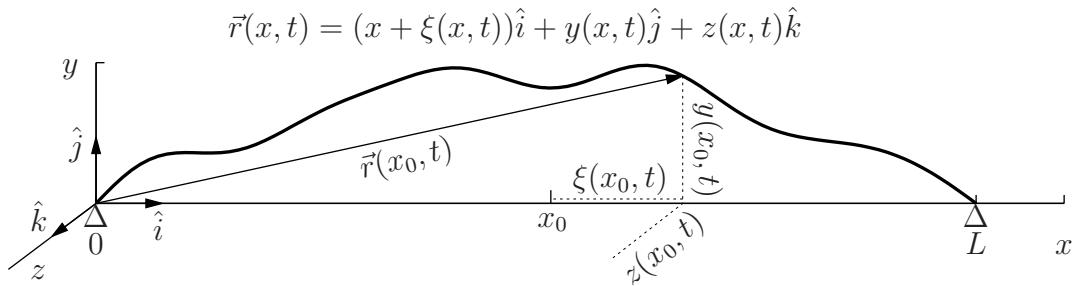


Figure 2.20: The position vector of a vibrating string in three dimensions

To establish a similar analysis of forces that was carried out on the ideal string in section 2.1.2, an expression for the infinitesimal piece of string is needed. In this case the differential element of length can be directly obtained from the definition of the position vector as

$$s = \sqrt{dx^2 + dy^2 + dz^2} = \sqrt{d\vec{r} \bullet d\vec{r}} = \sqrt{\frac{\partial \vec{r}}{\partial x} \bullet \frac{\partial \vec{r}}{\partial x}} dx = \left| \frac{\partial \vec{r}}{\partial x} \right| dx, \quad (2.96)$$

and the magnitude of this element of length is just

$$\left| \frac{\partial \vec{r}}{\partial x} \right| = \sqrt{\left(1 + \frac{\partial \xi}{\partial x} \right)^2 + \left(\frac{\partial y}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial x} \right)^2}. \quad (2.97)$$

A unit vector to define the tangent direction of the string is directly obtained from the element of length (2.96) by dividing that with its magnitude

$$\hat{s} = \frac{\frac{\partial \vec{r}}{\partial x}}{\left| \frac{\partial \vec{r}}{\partial x} \right|} = \frac{\left(1 + \frac{\partial \xi}{\partial x} \right) \hat{i} + \left(\frac{\partial y}{\partial x} \right) \hat{j} + \left(\frac{\partial z}{\partial x} \right) \hat{k}}{\sqrt{\left(1 + \frac{\partial \xi}{\partial x} \right)^2 + \left(\frac{\partial y}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial x} \right)^2}}. \quad (2.98)$$

With this unit vector \hat{s} it is possible to create a vector quantity from the magnitude F pulling the string as the product

$$\vec{F} = F \hat{s}. \quad (2.99)$$

Since \hat{s} includes derivatives to all coordinate directions, equation (2.99) describes the same situation as in the case of the ideal string, which was restricted only to the $\{x, y\}$ -plane. In this general case the total magnitude of the tension force F is defined by equation (2.100), and when combined with \hat{s} , the direction of the total force becomes the tangent direction of the string at all points along the string. The force components F_x , F_y and F_z to orthogonal coordinate directions are again fractions of the total force, proportional to the slope of the string against each coordinate direction.

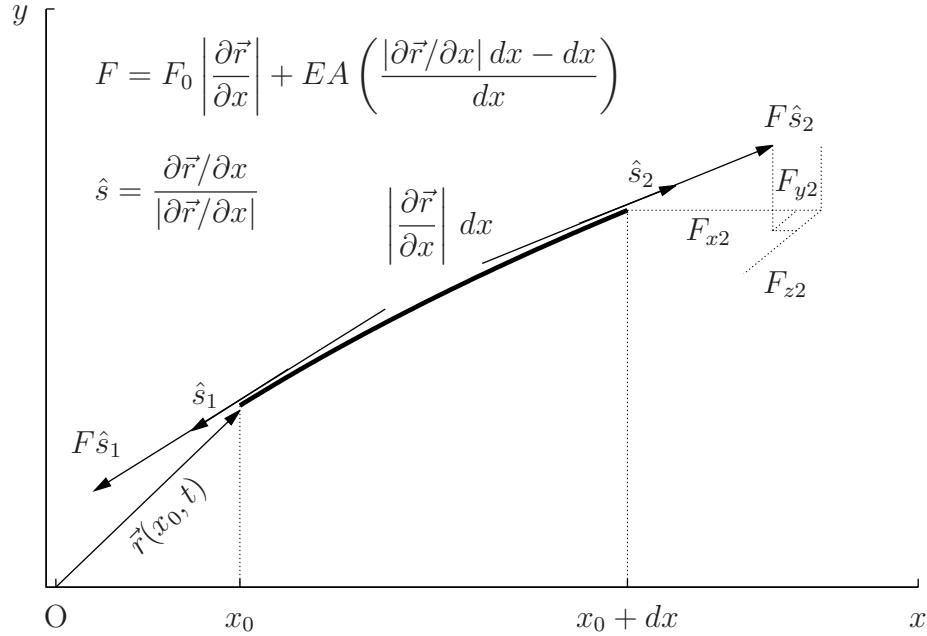


Figure 2.21: Force with directions \hat{s}_1 and \hat{s}_2 on a small piece of string

Now all that is needed is the expression for the magnitude of the tension force. The general three-dimensional model for the vibrating string also takes into account the increase of string tension due to stretching that results from displacing the string from equilibrium. This can be defined directly by Hooke's law, assuming that the dependency is linear. The ideal model required the tension to be constant, i.e F_x , but the enhanced equation for the force pulling the string is

$$F(x, t) = EA \left(\left| \frac{\partial \vec{r}}{\partial x_0} \right| - 1 \right), \quad (2.100)$$

where x_0 defines the relaxed length of the differential element when $F(x, t) = 0$, that is when the string has not yet been stretched to initial tension. The term in parentheses in equation (2.100) describes the relative change of length, where the expression has been simplified a little from the form

$$EA \frac{\Delta L}{L} = EA \frac{\left| \frac{\partial \vec{r}}{\partial x_0} \right| dx_0 - dx_0}{dx_0},$$

which is the familiar form of Hooke's law.

The dummy variable x_0 can be cancelled out from the equation by expanding equation (2.100) as

$$F(x, t) = EA \left(\left| \frac{\partial \vec{r}}{\partial x} \frac{\partial x}{\partial x_0} \right| - 1 \right),$$

and additionally defining that

$$F_0 = EA \left(\frac{\partial x}{\partial x_0} - 1 \right)$$

is the force that stretches the string to initial playing tension. This leads to the final expression for the magnitude of the total force

$$F(x, t) = F_0 \left| \frac{\partial \vec{r}}{\partial x} \right| + EA \left(\left| \frac{\partial \vec{r}}{\partial x} \right| - 1 \right). \quad (2.101)$$

The validity of the term F_0 can be verified by neglecting the EA term and creating the magnitude + unit vector combination with the F_0 term alone. This combination gives a correct expression for the force, just like in equation (2.92). Neat. [33] [34, pp. 9 – 10]

In the case of the ideal string model, the force component F_y was taken as a first order approximation to be a fraction of the constant force F_x , so that the total tangential force could be expressed in terms of F_x only. The generalised model has this feature for all coordinate directions through the term with $F_0 = F_x$ as the multiplier. The transverse motion is mainly defined by this term. The term with EA as the multiplier adds the longitudinal motion of the string, which is based on the changing force due to the extension and compression of the string during vibration. This term is completely missing from the simple ideal model.

The force differentials within the ideal string are due to differences in total force direction at the ends of infinitesimal pieces of the string. This also holds for the general model but in addition the general model contains a force difference term already in the definition of the force of tension [24, p. 858].

For the nonlinear model of the vibrating string, the Newtonian equation of motion can be constructed now as

$$\rho A \frac{\partial^2 \vec{r}}{\partial t^2} = \frac{\partial(F\hat{s})}{\partial x} = \hat{s} \frac{\partial F}{\partial x} + F \frac{\partial \hat{s}}{\partial x}, \quad (2.102)$$

where the left side equals mass × acceleration as usual. After inserting the expression (2.101) of the tension force into equation (2.102),

$$\rho A \frac{\partial^2 \vec{r}}{\partial t^2} = (F_0 + EA) \hat{s} \frac{\partial}{\partial x} \left| \frac{\partial \vec{r}}{\partial x} \right| + (F_0 + EA) \left| \frac{\partial \vec{r}}{\partial x} \right| \frac{\partial \hat{s}}{\partial x} - EA \frac{\partial \hat{s}}{\partial x}. \quad (2.103)$$

To write open the partial derivatives on the right side of equation (2.103), a few intermediate results are needed. The derivative of the magnitude of the position

vector (2.95) looks like

$$\frac{\partial}{\partial x} \left| \frac{\partial \vec{r}}{\partial x} \right| = \frac{\left(1 + \frac{\partial \xi}{\partial x} \right) \frac{\partial^2 \xi}{\partial x^2} + \left(\frac{\partial y}{\partial x} \right) \frac{\partial^2 y}{\partial x^2} + \left(\frac{\partial z}{\partial x} \right) \frac{\partial^2 z}{\partial x^2}}{\sqrt{\left(1 + \frac{\partial \xi}{\partial x} \right)^2 + \left(\frac{\partial y}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial x} \right)^2}}, \quad (2.104)$$

and the partial derivative of the unit vector (2.98) is a true monster,

$$\frac{\partial \hat{s}}{\partial x} = \frac{\frac{\partial^2 \vec{r}}{\partial x^2}}{\left| \frac{\partial \vec{r}}{\partial x} \right|} - \frac{\frac{\partial \vec{r}}{\partial x}}{\left| \frac{\partial \vec{r}}{\partial x} \right|^2} \frac{\left(1 + \frac{\partial \xi}{\partial x} \right) \frac{\partial^2 \xi}{\partial x^2} + \left(\frac{\partial y}{\partial x} \right) \frac{\partial^2 y}{\partial x^2} + \left(\frac{\partial z}{\partial x} \right) \frac{\partial^2 z}{\partial x^2}}{\left| \frac{\partial \vec{r}}{\partial x} \right|}, \quad (2.105)$$

but using equation (2.104) and multiplying both sides of equation (2.105) with the magnitude of the position vector, it simplifies to

$$\left| \frac{\partial \vec{r}}{\partial x} \right| \frac{\partial \hat{s}}{\partial x} = \frac{\partial^2 \vec{r}}{\partial x^2} - \hat{s} \frac{\partial}{\partial x} \left| \frac{\partial \vec{r}}{\partial x} \right|. \quad (2.106)$$

This intermediate result is used to simplify equation (2.103), and after inserting the expression of (2.106) into the second term on the right side of equation (2.103), the final form of the general equation of motion for the vibrating string reads

$$\frac{\partial^2 \vec{r}}{\partial t^2} = (c_L^2 + c_T^2) \frac{\partial^2 \vec{r}}{\partial x^2} - c_L^2 \frac{\partial \hat{s}}{\partial x}, \quad (2.107)$$

where

$$c_T^2 = \frac{F_0}{\rho A} \quad \text{and} \quad c_L^2 = \frac{E}{\rho}$$

just in the same format as those have already appeared earlier ($F_0 = F_x$). Still this model neglects the stiffness of the string and damping caused by friction, but those effects can be added relatively easily since the general form of those terms is already known. The equation of motion (2.107) gives more accurate results only for large amplitude string vibrations.

The final result shown here differs slightly from the equation provided by Morse and Ingard [24, pp. 860–861] but this is due to using a different expression for the force in equation (2.101). The force expression used in [24] assumes that the string is already stretched to initial tension, which leads to a different term for F_0 in equation (2.101). The force term given in this context is compatible with the one-dimensional model of the ideal string in terms of vectorisation, as it is shown in detail. In practise the difference between the two force models is minimal. [34]

In their treatment [24, pp. 860–861], Morse and Ingard present a famous third order approximation, which is separated from equation (2.107) for each vibrational direction. Being third order equations, a lot of approximation is still required when starting from the general equation of motion. There are no intermediate steps given for the approximation calculation in [24], even though there is a long way from equation (2.107) to the given results. As this book tries to give detailed information on all the topics covered, the intermediate steps leading to the third order approximations are presented here.

Each of the orthogonal directions \hat{i}, \hat{j} and \hat{k} is treated separately and partial derivatives with respect to x are written with a subscript x , for example, y_x means a partial derivative of $y(x, t)$ with respect to x . Starting from direction the of \hat{i} :

$$\left(\frac{\partial^2 \xi}{\partial t^2} - [c_L^2 + c_T^2] \frac{\partial^2 \xi}{\partial x^2} \right) \hat{i} = -c_L^2 \frac{\partial}{\partial x} \left[\frac{1 + \xi_x}{\sqrt{(1 + \xi_x)^2 + y_x^2 + z_x^2}} \right] \hat{i}. \quad (2.108)$$

To proceed from here, a binomial approximation is applied to the rightmost term in squared brackets, and as a result,

$$\frac{1 + \xi_x}{(1 + \xi_x) \sqrt{1 + \frac{y_x^2 + z_x^2}{(1 + \xi_x)^2}}} \approx 1 - \frac{1}{2} \frac{y_x^2 + z_x^2}{(1 + \xi_x)^2},$$

where the first two terms have been taken from the binomial series. The approximation still contains a partial derivative in the denominator and clearly the procedure is to get rid of the denominator. That can be accomplished by making an 'elementary school' division:

$$\begin{array}{r} 1 - 2\xi_x + 3\xi_x^2 \dots \\ \hline 1 + 2\xi_x + \xi_x^2 \mid 1 \\ - \quad 1 + 2\xi_x + \xi_x^2 \\ \hline -2\xi_x - \xi_x^2 \\ - \quad -2\xi_x - 4\xi_x^2 - 2\xi_x^3 \\ \hline 3\xi_x^2 + 2\xi_x^3 \\ - \quad 3\xi_x^2 + 6\xi_x^3 + 3\xi_x^3 \\ \hline -4\xi_x^3 - 3\xi_x^3 \\ \dots \end{array}$$

Eventually the division would yield an infinite series, where the resulting terms appear to the topmost row. In this case only the first two terms are needed, so

with the remainder term, the result from the division is

$$\frac{1}{(1 + \xi_x)^2} = 1 - 2\xi_x + \frac{3\xi_x^2 + 2\xi_x^3}{(1 + \xi_x)^2}.$$

A substitution of the result of the division along with the binomial approximation into equation (2.108) gives

$$\left(\frac{\partial^2 \xi}{\partial t^2} - [c_L^2 + c_T^2] \frac{\partial^2 \xi}{\partial x^2} \right) \hat{i} = -c_L^2 \frac{\partial}{\partial x} \left[1 - \frac{1}{2} (y_x^2 + z_x^2) (1 - 2\xi_x) \right] \hat{i}.$$

After differentiating the constant 1 away, the final form of the \hat{i} directional (longitudinal) third order wave equation reads

$$\frac{\partial^2 \xi}{\partial t^2} - [c_L^2 + c_T^2] \frac{\partial^2 \xi}{\partial x^2} = \frac{1}{2} c_L^2 \frac{\partial}{\partial x} [(y_x^2 + z_x^2) (1 - 2\xi_x)]. \quad (2.109)$$

The starting point for the transverse direction of \hat{j} is

$$\left(\frac{\partial^2 y}{\partial t^2} - [c_L^2 + c_T^2] \frac{\partial^2 y}{\partial x^2} \right) \hat{j} = -c_L^2 \frac{\partial}{\partial x} \left[\frac{y_x}{\sqrt{(1 + \xi_x)^2 + y_x^2 + z_x^2}} \right] \hat{j}. \quad (2.110)$$

The binomial approximation from the rightmost term in squared brackets follows as

$$\frac{y_x}{(1 + \xi_x) \sqrt{1 + \frac{y_x^2 + z_x^2}{(1 + \xi_x)^2}}} \approx \frac{y_x}{(1 + \xi_x)} \left[1 - \frac{1}{2} \frac{y_x^2 + z_x^2}{(1 + \xi_x)^2} \right],$$

and then applying the 'elementary school' division to the term $\frac{1}{1 + \xi_x}$:

$$\begin{array}{r} 1 - \xi_x + \xi_x^2 \dots \\ \hline 1 + \xi_x | 1 \\ - \quad 1 + \xi_x \\ \hline -\xi_x \\ - \quad -\xi_x - \xi_x^2 \\ \hline \xi_x^2 \\ - \quad \xi_x^2 + \xi_x^3 \\ \hline -\xi_x^3 \\ \dots \end{array}$$

The result from the division with the remainder term is

$$\frac{1}{1 + \xi_x} = 1 - \xi_x + \xi_x^2 - \frac{\xi_x^3}{1 + \xi_x},$$

from where the first three terms are taken to rewrite the binomial approximation as

$$y_x(1 - \xi_x + \xi_x^2) \left[1 - \frac{1}{2} \frac{y_x^2 + z_x^2}{(1 + \xi_x)^2} \right].$$

The expression can be simplified further to reveal the terms up to the third order. Because

$$(1 - \xi_x + \xi_x^2) = (1 + \xi_x)^2 - 3\xi_x,$$

the binomial expression simplifies to

$$\begin{aligned} y_x \left[(1 - \xi_x + \xi_x^2) - \frac{1}{2} \frac{(y_x^2 + z_x^2)[(1 + \xi_x)^2 - 3\xi_x]}{(1 + \xi_x)^2} \right] = \\ -\frac{1}{2} \left[-2y_x + 2y_x\xi_x(1 - \xi_x) + y_x(y_x^2 + z_x^2) - \frac{3\xi_x y_x (y_x^2 + z_x^2)}{(1 + \xi_x)^2} \right]. \end{aligned}$$

The last term of this expression is already of the fourth order, so finally the wave equation for the transverse direction of \hat{j} can be written as

$$\frac{\partial^2 y}{\partial t^2} - [c_L^2 + c_T^2] \frac{\partial^2 y}{\partial x^2} = -c_L^2 \frac{\partial^2 y}{\partial x^2} + \frac{1}{2} c_L^2 \frac{\partial}{\partial x} [y_x(y_x^2 + z_x^2) + 2y_x\xi_x(1 - \xi_x)],$$

and after small simplifications,

$$\frac{\partial^2 y}{\partial t^2} - c_T^2 \frac{\partial^2 y}{\partial x^2} = \frac{1}{2} c_L^2 \frac{\partial}{\partial x} [y_x(y_x^2 + z_x^2) + 2y_x\xi_x(1 - \xi_x)]. \quad (2.111)$$

The derivation for the equation for the second direction \hat{k} of transverse vibration is exactly the same as for \hat{j} so that

$$\frac{\partial^2 z}{\partial t^2} - c_T^2 \frac{\partial^2 z}{\partial x^2} = \frac{1}{2} c_L^2 \frac{\partial}{\partial x} [z_x(y_x^2 + z_x^2) + 2z_x\xi_x(1 - \xi_x)]. \quad (2.112)$$

This is the method for deriving the three equations of motion (2.109), (2.111) and (2.112), which are in agreement with the equations presented in [24, p. 860], except that because of the different definition of force, the wave propagation velocity terms c_L^2 and c_T^2 are differently placed. The results given in [24] can be obtained using the substitution $c_L^2 = c_L^2 - c_T^2$. Unfortunately Morse and Ingard have left a small typo in their resulting equations, namely concerning the term c_T^2 on the left side of equation (2.111), but maybe this is clear from the context.

If one looks more closely at the left hand sides of all three equations, one notices that the two transverse equations of y and z have the linear differential equation with the transverse velocity right in place. This tempts to assume that the longitudinal equation would have the linear equation on the left side with the

longitudinal velocity alone. But there is $c_L^2 + c_T^2$ instead of c_L^2 . This is the result from defining the force in the way of equation (2.101), and the conclusion is that the longitudinal wave velocity is in fact written as

$$c_L^2 = \frac{EA + F_0}{\rho A}.$$

This is not meaningful in practise because EA is always much greater than F_0 , but an interesting outcome in any case.

The three nonlinear differential equations are still in differential form, and although these are rough approximations of the general equation of motion, these equations do not have an analytic solution. On the other hand, these equations can be used to prove the fact that in the general case each of the directions of vibration are coupled to each other, and that the longitudinal-transverse mode coupling is associated with each equation.

As a special case, it is possible to derive an expression for the $\{xy\}$ -plane motion by combining the nonlinear equations for transverse and longitudinal vibration and neglecting equation (2.112) for the $\{xz\}$ -plane motion. By dividing equation (2.109) with c_L^2 and assuming that $c_L = \infty$ and approximating that $1 - 2\xi_x \approx 1$ and that $z = 0$, the equation

$$\frac{\partial^2 \xi}{\partial x^2} = -\frac{1}{2} \frac{\partial}{\partial x} \left(\frac{\partial y}{\partial x} \right)^2$$

is obtained [33]. Obviously this is a very far fetched approximation, but at least this can be solved analytically to some extent.

To solve this differential equation, it can be directly antiderivatived twice. The first direct antiderivation is of the form

$$d \left[\partial \xi + \frac{1}{2} \left(\frac{\partial y}{\partial x} \right)^2 dx \right] = d(C_1)$$

and the second one continues as

$$d \left[\xi + \frac{1}{2} \int_0^x \left(\frac{\partial y}{\partial z} \right)^2 dz \right] - xC_1 = d(C_2),$$

which leads to an equation for the longitudinal waves, where

$$\xi(x, t) = -\frac{1}{2} \int_0^x \left(\frac{\partial y}{\partial z} \right)^2 dz + xC_1 + C_2.$$

The constants C_1 and C_2 are evaluated using the boundary conditions $\xi(0, t) = 0$:

$$\xi(0, t) = 0 = -\frac{1}{2} \int_0^0 \left(\frac{\partial y}{\partial x} \right)^2 dx + 0 + C_2 \quad \Rightarrow \quad C_2 = 0,$$

and $\xi(L, t) = 0$:

$$\xi(L, t) = 0 = -\frac{1}{2} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx + LC_1 + C_2 \quad \Rightarrow \quad C_1 = \frac{1}{2L} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx.$$

The detailed solution satisfying the boundary conditions is therefore

$$\xi(x, t) = -\frac{1}{2} \int_0^x \left(\frac{\partial y}{\partial z} \right)^2 dz + \frac{x}{2L} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx.$$

The derivative with respect to x is

$$\frac{\partial \xi(x, t)}{\partial x} = -\frac{1}{2} \left(\frac{\partial y}{\partial x} \right)^2 + \frac{1}{2L} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx + 0. \quad (2.113)$$

Neglecting ζ_x and assuming that $(1 - \xi_x) \approx 1$ in the equation (2.111) for the transverse motion yields

$$\frac{\partial^2 y}{\partial t^2} - c_T^2 \frac{\partial^2 y}{\partial x^2} = \frac{1}{2} c_L^2 \frac{\partial}{\partial x} \left\{ \frac{\partial y}{\partial x} \left[\left(\frac{\partial y}{\partial x} \right)^2 + 2\xi_x \right] \right\}. \quad (2.114)$$

Substitution of the derivative $\frac{\partial \xi(x, t)}{\partial x}$ from equation (2.113) into equation (2.114) gives

$$\frac{\partial^2 y}{\partial t^2} - c_T^2 \frac{\partial^2 y}{\partial x^2} = \frac{1}{2} c_L^2 \frac{\partial}{\partial x} \left\{ \frac{\partial y}{\partial x} \left[\frac{1}{L} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx \right] \right\}, \quad (2.115)$$

and after simplification this results in

$$\frac{\partial^2 y}{\partial t^2} = c_T^2 \frac{\partial^2 y}{\partial x^2} + c_L^2 \left[\frac{1}{2L} \int_0^L \left(\frac{\partial y}{\partial x} \right)^2 dx \right] \frac{\partial^2 y}{\partial x^2}. \quad (2.116)$$

If the result derived by Morse and Ingard [24] is resumed with a substitution $c_L^2 = c_L^2 - c_T^2$, this result reveals the coupling of the transverse and longitudinal modes via the multiplier $(c_L^2 - c_T^2)$ and the length modulation during vibration is included in the integral term. This is quite similar to equation (2.76) used for deriving the equations to simulate the orbital motion of the nonlinear string in section 2.2.5 and it is commonly referred to as the Narasimha's equation, which was first introduced by Carrier [35].

2.4 MEASURING THE STRING VIBRATIONS

The purpose of these measurements is to visualise the waveforms and spectrums of a plucked guitar string and to compare the outcome of the measurements to the results obtained from theoretical models. The string is treated in these measurements as a self-contained vibrating unit, without the nonlinearities that would arise if the string was connected to a guitar and/or measured through a nonlinear magnetic pickup. Analysis of the measurement results offers insight on the properties of the string as a source of sound and tone colour. To be able to understand how the whole signal chain of electric guitar accessories changes the sound of the string, it is essential to know the harmonic content of the upper partials present in the pure tone structure originating from a string.

A typical construction of an electric guitar string is depicted in Figure 2.22, which shows a sliced model of an *overwound* string. Light gauge strings do not have the winding around them, so those can be represented as 'core-only' strings. The core wire of a typical steel string is just an uniform piece of metal without any discontinuities. However, steel strings become impractically stiff if the thickness of the core is increased over a certain limit. To increase the mass of the string delicately without considerably increasing the stiffness of the string, heavier gauges are constructed by wrapping an extra layer of metal wire around the core. Typical materials used as the outer winding are nickel plated steel, pure nickel and stainless steel; sometimes even silver or gold.

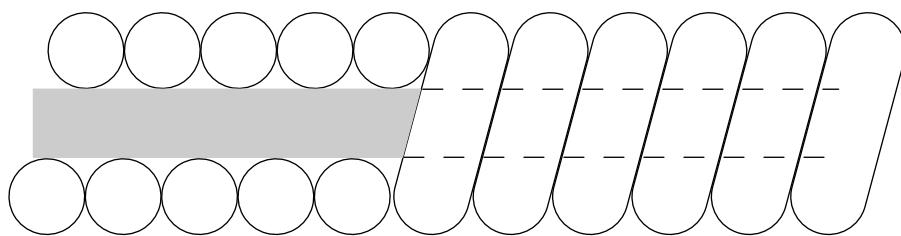


Figure 2.22: The inner core and outer winding of a roundwound guitar string

The winding of an electric guitar string can be shaped as *flatwound*, *halfwound* or *roundwound*. The geometry of a round winding is shown in Figure 2.22; obviously roundwound strings have a round cross-sectional area. The flat winding has more like a rectangular cross-sectional shape, and the halfwound strings have a round inner layer and a flat outer layer. The wound strings are claimed to have a tendency to enhance the second and third upper partials compared to an unwound string. However, this detail will not be verified in the measurements

since the theoretical analysis of wound strings is not covered in the previous sections. It is essential that the measurement results are somehow comparable to the theoretical foundations already established in the previous sections.

2.4.1 Properties of the measured string

The measurements were made using a regular unwound steel core string designed for an electric guitar. A nylon string was also used as a reference to study how the effect of stiffness is seen in the measurement results. Usually string manufacturers give details only on the string diameter, but some also specify the tension needed to tighten the string to its nominal frequency within some predefined scale length. The total tension caused by tuned strings is useful information for all guitarists, since the neck construction of the guitar will bend more as the tension in the strings increases. The tension specifications also relate directly to the fundamental frequency of a string.

For example, the steel string chosen for the following experiments was specified with a diameter of 0.017 inches, which equals 0.43 mm. The manufacturer has defined this to be used as the *B* string, and with a nominal scale length of 648 mm, a tension equal to a weight of 11.93 kg is needed to stretch the string to the note *B*₃ with a frequency of 247 Hz. This can be verified with equation (2.1), using the density of iron, $\rho_{\text{iron}} \approx 7870 \frac{\text{kg}}{\text{m}^3}$. A substitution of the known values leads to the expression

$$f_1 = \frac{1}{2 \cdot 0.648 \text{ m}} \sqrt{\frac{9.81 \frac{\text{m}}{\text{s}^2} \cdot 11.93 \text{ kg}}{7870 \frac{\text{kg}}{\text{m}^3} \cdot \pi \cdot 0.000215^2 \text{ m}^2}}$$

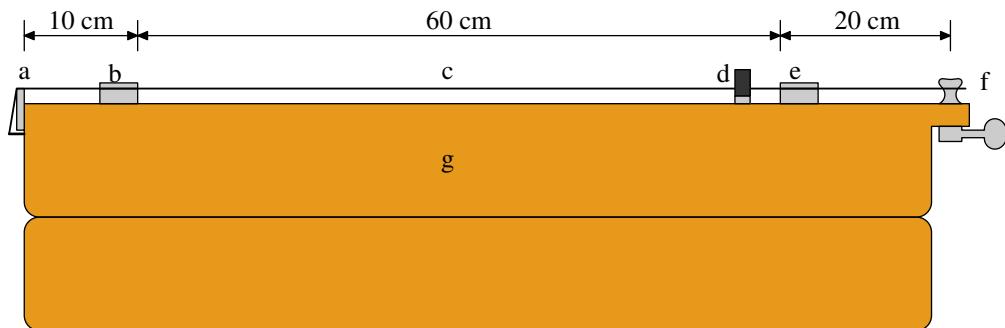
which evaluates to $f_1 = 246.92$ Hz. Nice.

A steel string stretches typically only a few millimetres when it is tightened to a normal playing tension. Nylon strings are clearly more flexible, and they can stretch even for a few centimetres under the standard tension.

2.4.2 The measurement setup

In the experiments, the string was mounted on a self-built test bench, which tries to provide ideal conditions for the end supports of the string by means of clamping the string between rigid metal plates. The construction of the test bench was primarily inspired by a measurement setup described by R. M. French [36, p. 108], where a sturdy piece of wooden log was used as the platform

for mounting the string. Only the massive wood structure as the measurement platform has been adopted for this purpose, all the other things are a result of improvisation to put all the necessary pieces together. The wooden platform was handcrafted by attaching two 4 by 4 inch stocks on top of each other by glue and sawing a small ledge for the tuning peg attachment. A barebone sketch of the measurement platform used is given in Figure 2.23.



- | | |
|--------------------------------|----------------------------|
| a = metal plate as end support | e = rightmost string clamp |
| b = leftmost string clamp | f = tuning peg |
| c = string | g = sturdy piece of wood |
| d = transducer | |

Figure 2.23: An illustrative sketch of the measurement platform

Primarily only the tuning peg and the vertical plate at the end of the platform were installed in place and the string was prestretched to some initial tension. After the string had stabilised to the initial stretch so that no noticeable drift in the frequency was observed, a pack of metal plates was installed underneath the string as the rightmost string support. The metal plates were attached together by glue and afterwards the complete stack of plates was glued to the wooden platform. The string was then clamped to this support by a top metal plate, which was tightened by four screws to strongly clamp the string against the set of metal plates.

A discrete component of an H21B1 photodarlington optical interrupter switch was used as the actual measuring instrument to transduce the string vibration to an electric signal. The optoswitch was firstly mounted about three centimetres away from the end support. This location was chosen after many trials to maximise the signal to noise ratio at modest plucking amplitudes of a few millimetres. Measurement systems are never completely noise free, not even optical systems, but as the intensity of the light is increased, the signal will have less noise. Unfortunately a high-intensity laser system was not available for these

low budget, do-it-yourself measurements.

The electrical and physical properties of the optoswitch are covered in detail within the context of optical pickups (see section 3.3), since it also offers a practical and simple example of optical pickups used in guitars. As a measurement device it offers the cheapest and simplest way to measure the movement of the string with good accuracy. The fundamental ideology in the optical measurement is to place the string in the middle of a beam of light so that when the string moves, it modulates the light beam that is feeding photoelectric current to a photoactive component, such as a phototransistor. This ideology is visualised in Figure 3.38 in section 3.3.1.

In comparison to a basic magnetic pickup, the optoswitch has the advantage of wide range linearity (see Figure 3.44) and a very small sensor area of 1 mm. Restricting the measurement location to such a small area gives the possibility to measure even very high-frequency components accurately. In addition, the optoswitch directly measures the displacement of the string, as opposed to the magnetic pickup, which measures the velocity of the string. The trace of the string displacement obtained with the optical transducer is easily transformed to show the exact waveform and spectrum depicting the motion and harmonic content of a guitar string.

In the practical measurement setup the optoswitch needed an external voltage supply of 5 V, which was regulated from a basic 9 V battery as shown in Figure 2.24. The resistors R_D and R_E were chosen to keep the current in the LED circuit

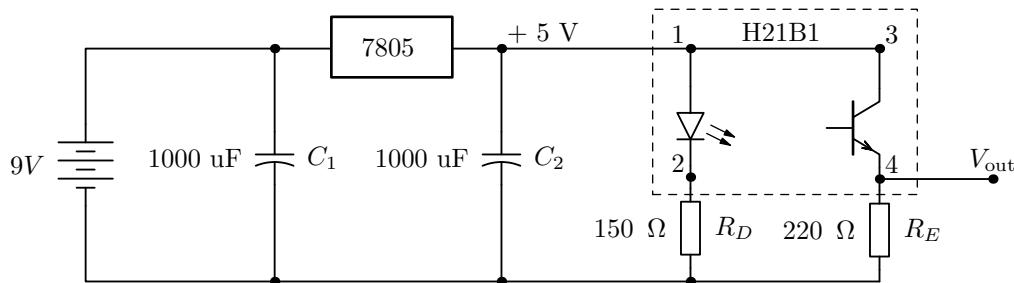


Figure 2.24: A wiring schematic for the optoswitch

sufficiently large, and allow almost a maximum voltage swing (0 V – 4.2 V) at the output [37]. The phototransistor depicted in Figure 2.24 is actually a darlington pair, but for simplicity it is drawn here as a single transistor.

The output from this circuit was simultaneously measured by an oscilloscope

and a basic built-in computer sound card. The oscilloscope provided a sampling frequency of 100 kHz and the PC sound card was able to record with a 96 kHz sampling rate. For a vibration frequency of about 200 Hz, these sampling rates will give approximately 450 discrete measurement points per period for further analysis. Surprisingly the sound card offered a better voltage measurement resolution with its 16-bit DA conversion. The oscilloscope had a far more weaker resolution, which is seen as unnecessary fluctuation in the measurement results.

A first set of measurements was executed with this setup, where only the right side of the string was clamped and the optoswitch was at a three centimetre distance from the clamped support. In this setup the effective length of the string was 700 mm. A sharp metal needle was used as a plectrum to pluck the string, but as a reference, plastic picks of different widths were used as well. After finishing the first set of measurements, the left string clamp was installed, reducing the effective length to 600 mm, and then a second set of measurements was done. Later the optoswitch was also moved closer to the support to get better measurements from the higher upper partials.

2.4.3 Error limit calculations

The purpose of error analysis is to reflect the uncertainty of the measured values compared to the expected values provided by theoretical equations. Instead of comparing the results obtained by measurements to the exact theoretical values, it is better to determine a range of theoretical values that are possible to obtain with the limited accuracy of the measured values. The error in the measured values propagates to results that are calculated using the measured values as variables. Error analysis estimates the total error in the final result due to different inaccuracies in the associated variables. [38]

In the measurements of the vibrating string, the primary interest is to find out the amplitudes of the harmonic frequency components and to show their relative magnitudes with respect to the magnitude of the fundamental frequency. Theoretically this harmonic spectrum is calculated from equation (2.34), and with the weight of the transducer added, the theoretical values for relative spectral magnitudes are calculated from equation

$$A_n(p, P, L) = \frac{\sin\left(\frac{n\pi p}{L}\right) \sin\left(\frac{n\pi P}{L}\right)}{n^2 \sin\left(\frac{\pi p}{L}\right)}, \quad (2.117)$$

from where the relative amplitudes are obtained by dividing the amplitude of

all the frequency components with the amplitude of the fundamental frequency. The weighting created by the location of the transducer is explained in detail in section 3.1.

The variables in equation (2.117) that will suffer from the limited accuracy of measurement are the plucking location p , the pickup location P and the length of the string L . If the position measurement is done using a device that has tick marks at every millimetre (a tape measure or a ruler, for example), the error in the measurement can be taken as ± 1 mm. In this specific case this level of uncertainty is used for all measured values of p , P and L .

When defining the relative magnitudes of the spectrum coefficients with equation (2.117) using these measured values, the uncertainty from all measured values will propagate to the final result. Assuming that the relative error in the measured values is small, the normal procedure is to calculate the total differential of $A_n(p, P, L)$ with respect to all measured variables. That will lead to the following equations:

$$\frac{\partial A_n}{\partial p} = |A_n| \frac{n\pi}{L} \frac{\cos\left(\frac{n\pi p}{L}\right)}{\sin\left(\frac{n\pi p}{L}\right)}$$

$$\frac{\partial A_n}{\partial P} = |A_n| \frac{n\pi}{L} \frac{\cos\left(\frac{n\pi P}{L}\right)}{\sin\left(\frac{n\pi P}{L}\right)}$$

$$\frac{\partial A_n}{\partial L} = |A_n| \frac{n\pi}{L^2} \left[\frac{p \cos\left(\frac{n\pi p}{L}\right)}{\sin\left(\frac{n\pi p}{L}\right)} + \frac{P \cos\left(\frac{n\pi P}{L}\right)}{\sin\left(\frac{n\pi P}{L}\right)} \right].$$

By writing the equations in this form suggests that the problem with this approach is the division by zero at the node points of certain harmonic frequencies. This is not the actual problem because the sines will disappear from the denominator if the term $|A_n|$ is written open. However, there is still a problem with the node locations. The cosine term will be zero at the node points, so the error would then be zero. Obviously these are not suitable formulae for error analysis.

One approach to get around the node problem is to rewrite the derivative of sine

in the form

$$\frac{d \sin(x_0)}{dx} = \frac{\sin(x_0 + dx) - \sin(x_0)}{dx},$$

so that even if the sine or the cosine at x_0 would be zero, the difference quotient will have a value other than zero. Defining the errors this way would lead to expressions such as

$$\partial A_n = \frac{\left[\sin\left(\frac{n\pi(p + \delta p)}{L}\right) - \sin\left(\frac{n\pi p}{L}\right) \right] \sin\left(\frac{n\pi P}{L}\right)}{n^2 \sin\left(\frac{\pi p}{L}\right)} dp,$$

as an example of the error term with respect to p . After the error terms have been calculated for each variable p , P and L , the total error can be obtained by summing the three terms quadratically as the standard procedure of error propagation for independent errors requires.

The most crude way to determine the error limits is to just use the error magnitudes of the measured quantities to find a maximum and minimum of the expression (2.117) for A_n . The main difference compared to the approach explained above is that the error terms are not summed quadratically, so that this procedure will produce slightly larger error limits. It was seen as suitable to use the crude method for the following analysis, because the error limits would become adequately small in that way as well.

2.4.4 Measurement results - Amplitude spectrum

The first measurements were done using only the right string clamp and the other end was resting on top of the end support plate. The optoswitch was attached with screws on top of the string so that the direction of vibration was forced to vertical motion. This arrangement was easier to accomplish with this measurement platform. The Earth's gravity clearly affects the vibration unequally in this way, but just keep this in mind, since within this experiment no comparison will be done to the horizontal vibrations.

Additionally, this setup with only one end clamped forced the effective length of the string to 700 ± 1 mm, and for this length the string was initially tuned as close as possible to 200 Hz. This is because the Discrete Fourier analysis requires the frequency ratio of the sampling frequency and the fundamental to be an exact integer. However, in practical measurements this requirement is almost impossible to fulfil.

The first object of interest was to see how well the measured waveforms correlate

with theoretical expectations. If the waveforms are not in agreement with the theoretical expectations, then it is of no use to proceed to calculate the spectrum representations. Figure 2.25 shows an oscilloscope trace of the first two periods from a vertical pluck with a displacement of about 1 mm applied at the middle of the string. The dashed waveform shows the ideal theoretical waveform for direct comparison purposes.

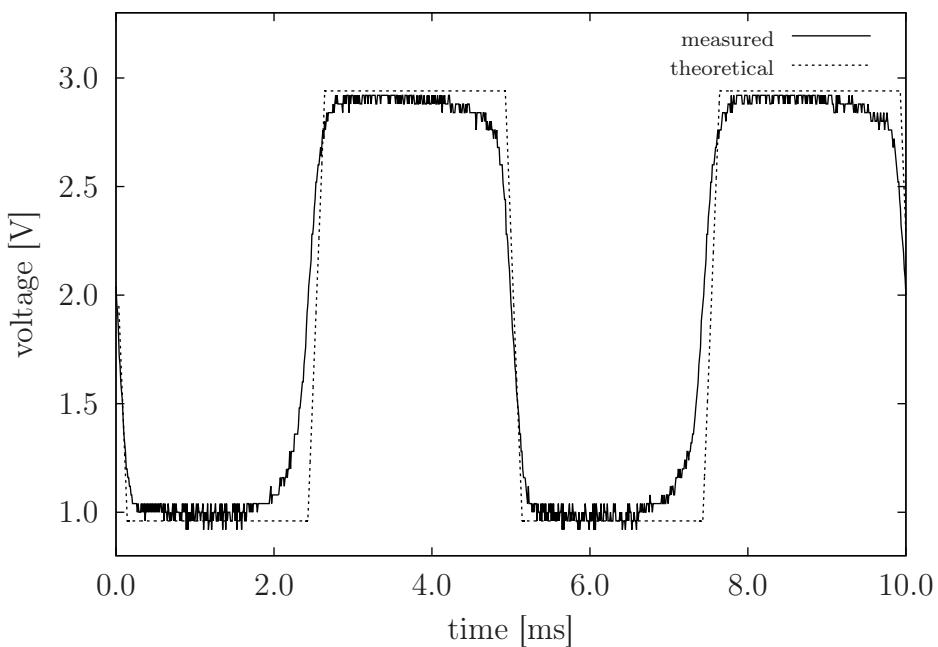


Figure 2.25: The first two periods of string vibration plucked at the centre

The waveforms are quite similar, but what was seen in all measured waveforms was the peculiar roundness at uneven corners of the almost squared wave. Later on, this rounding effect was investigated more to find out the reason that is causing it. At first thought it was assumed to relate to the stiffness of the steel string, but measurements with a nylon string indicated that the stiffness was not the primary explanation for this effect, although clearly the stiffness aspect plays a small role here as well.

The main reason for the roundness was found to be the nonideality of the plucking object. The sharp metal needle chosen as the plectrum to pluck the strings is as close as one can get to an ideally narrow plucking device, but apparently it was not a sharp enough needle. Another important thing to consider here is the force used for plucking, which is related to the stiffness of the string. For plucks with small displacement amplitudes of approximately 1 – 2 mm, the force on the string was found to be too small to make the initial shape of the string

to resemble the theoretical triangular shape. Eventually the best experimental effort was to still use the sharp metal needle for plucking, but increase the initial displacement to somewhere around 5 ± 1 mm or more, and this resulted in the waveform shown in Figure 2.26. This figure compares the best experimental result to a pluck with a plastic plectrum and the theoretical expectation curve.

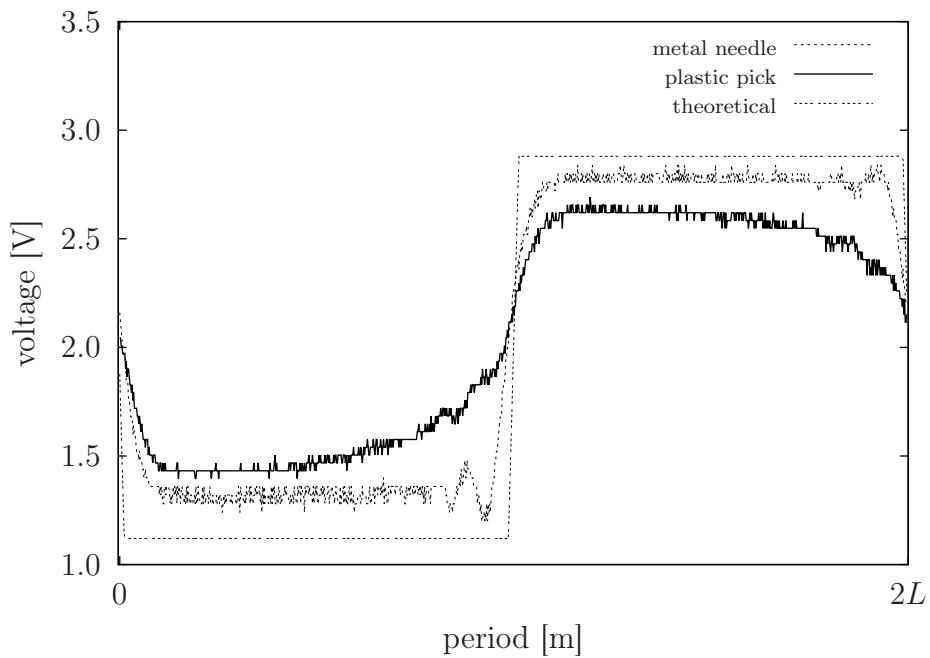


Figure 2.26: Investigations on the roundness effect

The softer and wider plastic plectrum clearly caused the roundness to increase. This is also clear from the theoretical point of view, since at the moment of plucking a wider plectrum covers more node points of high-frequency components than a sharp metal needle. Clearly there is still the effect of stiffness present because the waveform is not totally balanced, but it has some ripple at the edges. This would suggest that the frequency components are not all exact integer multiples of the fundamental frequency.

To get an overall view of how the waveform depends on a certain plucking position, the waveforms measured from plucks in different positions can be drawn as a function of plucking distance starting from the leftmost end of the string. Figure 2.27 presents the measured results along with the theoretical expectations on the shapes of the waveforms. Clearly the duty cycle between the high and low states of the square wave changes as the plucking position advances but the square-like shape remains the same. Near the left end support the wave is mostly in the high state and takes a short dip to the low state. As the plucks proceed

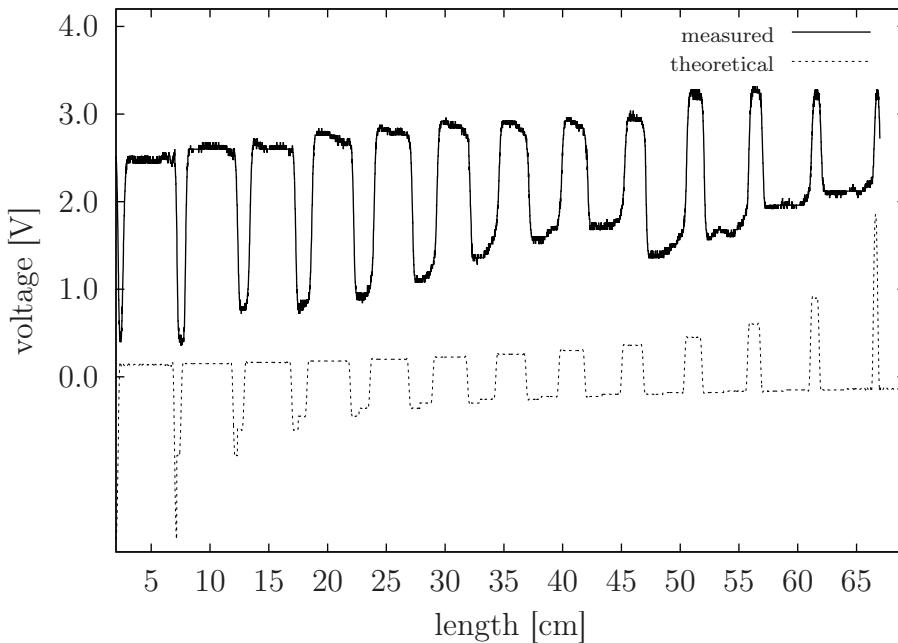


Figure 2.27: Waveforms as a function of the plucking position

towards the right end support, the low state becomes more and more dominant. The symmetry with respect to the mid point is easy to notice.

From the data of several measured plucks the best ones were chosen for further analysis to determine the spectrum of the waveforms. The most successful plucks were measured when both ends of the string were tightly clamped between the metal supports. In this setup the effective length of the string was 600 ± 1 mm, and with this length the fundamental frequency was about 230 ± 1 Hz. Additionally, to get a large number of harmonics measured, the optoswitch was mounted at 8 ± 1 mm from the leftmost end support of the string.

The spectrums were calculated by applying the Discrete Fourier Transform to the very first period of vibration because the external unideal factors such as the air resistance will affect the propagation of the wave in the string extremely fast.

Before going through the results of the spectral analysis, a good question is to ask how it is possible that the DFT calculated from the square-like signal is expected to give exactly the same amplitude components as the analytical formula, which is derived from a triangular shaped string. The equivalency between the triangular and squared forms can be proven semi-analytically by creating the static triangular string model from the analytical Fourier series and using the time dependent equation (2.33) to numerically create the trajectory of the string during

one full period of vibration. A sample can be provided to do this in Octave:

```
% number of components from Fourier's series
n = 1:1000;
% length of the string (between supports) in centimeters
L = 60;
% plucking position measured as a fraction of the string length
p = L/6;
% plucking amplitude in centimeters
a = 1;
% pickup position with respect to the length
x = 2;
% time goes through one full period
t = L/2:0.01:2*L + L/2;

% create the triangle from Fourier's series
% and calculate for duration of t observing from constant x
FACTOR = 2*a*(L/p + L/(L-p))/pi^2;
Z1 = sin(n'*pi*(x-t)/L)';
Z2 = sin(n'*pi*(x+t)/L)';
Z = (sin(n*pi*p/L)./n.^2)';
K = 0.5*FACTOR*(Z1+Z2)*Z;

% take FFT from the square-like wave, normalise amplitude
Y = 2*abs( fft(K,(length(K)-1)) )/(length(K)-1);
% reference amplitudes from theoretical triangle formula
X = FACTOR*Z;

% divider term to cancel the measurement location weighting
DIV = (sin((1:6)*pi*x/L))';
% two columns side by side for easy comparison for similarity
BUF = [Y(2:7)./DIV X(1:6)]
```

The result in variable K is now the trajectory of the string within one full period, and the amount of samples is several thousands. From these samples one can calculate the DFT and the actual magnitude of a spectrum bin is obtained when the weight function of the pickup position, as indicated in section 3.1, is divided away. This example will result in numerical values given in Table 2.1.

The results are practically identical. The only differences come from the numerical accuracy used in the DFT calculation, where rounding errors are likely to occur, but even this is so small that the differences are beyond the 16th decimal. This result proves that in an ideal situation the triangular and square-like waveforms produce amplitude components of the same magnitude. The only difference is that the triangular form generates some components with a negative amplitude, which justifies the use of the absolute values in the preceding equations. This proof gives confidence to assume that measurements on an ideal string should give results that can be directly compared to the theoretical spectrum.

Table 2.1: Amplitude components from two different shapes

n	square	triangle
f_1	0.729512522224832	0.729512522224832
f_2	0.315888188312782	0.315888188312782
f_3	0.162113893827740	0.162113893827740
f_4	0.078972047078195	0.078972047078195
f_5	0.029180500888993	0.029180500888993
f_6	1.9353×10^{-17}	2.2961×10^{-17}

To get a good view of how accurately the measurements agree with a large set of harmonics, it is best to pluck the string close to the end support. This plucking position will excite all harmonic frequencies as equally as possible. As the theoretical calculations have already indicated, a pluck close to the middle will result in a quite nondiverse spectrum with all the upper partials very close to zero amplitude.

Figure 2.28 shows the spectrum that resulted from a pluck at 20 ± 1 mm distance from the rightmost end support. In this measurement the optoswitch transducer was mounted at 8 ± 1 mm from the leftmost end support of the string. This location of the transducer is obligatory to get all the harmonics measured without hitting any node points. When drawn on a linear axis, the results seem to be a good match to the theoretically calculated values, where the error limits have been calculated with the uncertainties indicated above along with the plucking position and the transducer location. Some of the first spectral bins fit nicely inside the error limits, although the general behaviour is that the amplitudes of the upper partials follow the lower error limit curve. It is to be assumed that there have been some issues resulting from the measurement setup that affect all the measurements this way.

The next step is to see how well the measurements identify that the pluck has been executed on a node of some upper partial. To compare to the previous theoretical spectrums, the string was plucked at the node of the 6th upper partial, which should result in a zero amplitude spectral bin at $n = 6$ and all its integer multiples.

The plot in Figure 2.29 has been prepared using a logarithmic axis in the y -

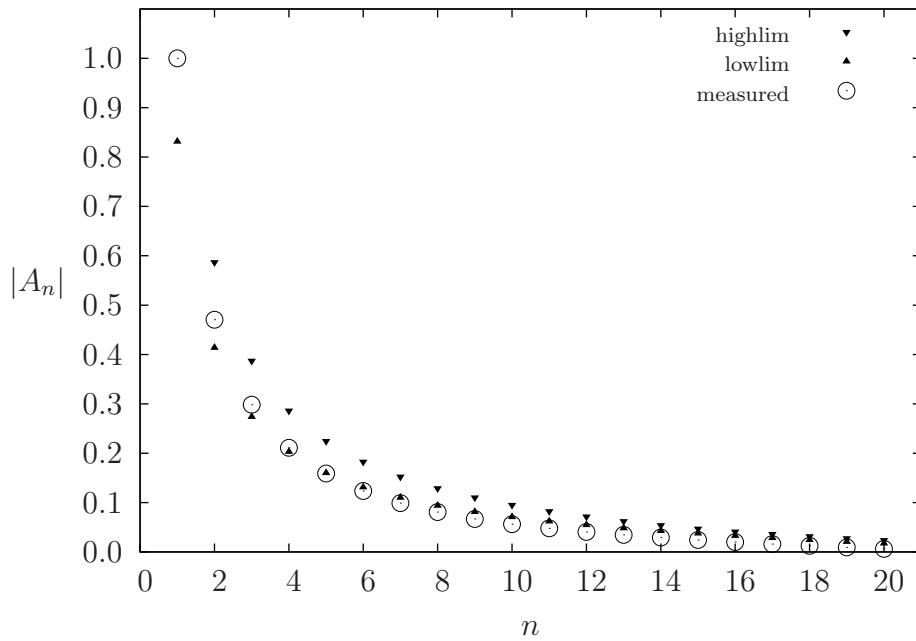


Figure 2.28: A steel string plucked from $\frac{L}{30}$

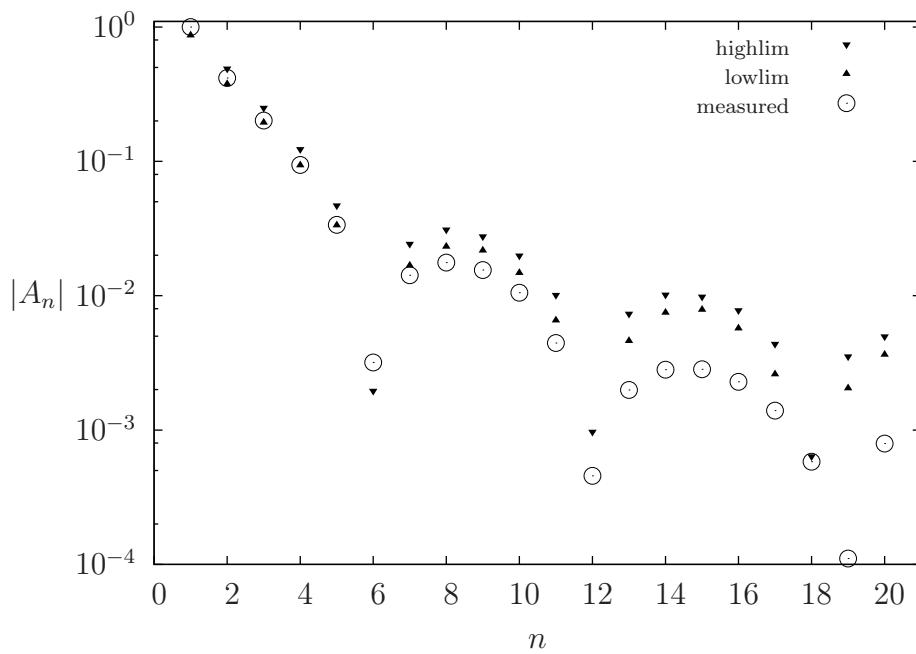


Figure 2.29: A steel string plucked from $\frac{L}{6}$

coordinate direction because this will show more accurately the small differences in the very low-amplitude spectral bins. The first and second nodes hit the 6th and 12th upper partials quite accurately, but the third node affects the 19th upper partial more instead of the expected 18th. No evidence was found on the root

cause for this frequency 'glide' to happen, as it also is in conflict with the stiffness-related theory of the raised frequencies of the higher upper partials. Most likely the plucking location was measured with a small bias towards the 19th node, which would still generate the 6th and 12th nodes reasonably accurately but would show up as inconsistency at the higher upper partials where the distances separating the nodes are small.

It turned out to be a difficult task to get a successful pluck at the middle of the string. Firstly there is more uncertainty in measuring the exact mid-point of the string and secondly it seemed that the actual pluck execution had to be done purely. This means that the initial displacement needed to be relatively large so that the shape of the string was close as possible to the ideal triangle. The stiffness of the string caused it to resist being molded towards the ideal triangle shape on the plucks with small amplitude displacement. For this special case the difference between horizontal and vertical plucks was examined. It seemed that horizontally directed plucks provided slightly better results than the vertically directed plucks but the difference was quite minimal. This observed difference is most likely due to the uneven effect of the gravitational force in the vertically directed motion.

Eventually, after several attempts, Figures 2.30 and 2.31 show the best effort from the measurement sessions.

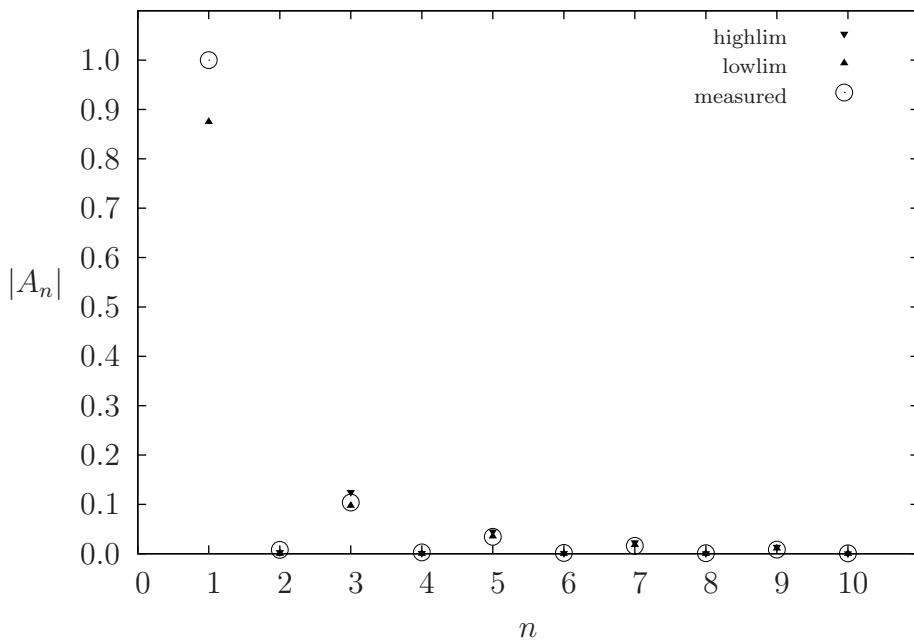


Figure 2.30: A steel string plucked from $\frac{L}{2}$

The linear y-axis plot in Figure 2.30 does not reveal the small deviations from the theoretical values. These are only noticed on the logarithmic plot. According to the logarithmic plot, the first ten or so upper partials are quite nicely in place, but then it seems that the upper partials are shifted in frequency so that they are no longer integer multiples of the fundamental frequency and therefore the harmonic Fourier analysis results in smaller amplitudes for the higher frequencies.

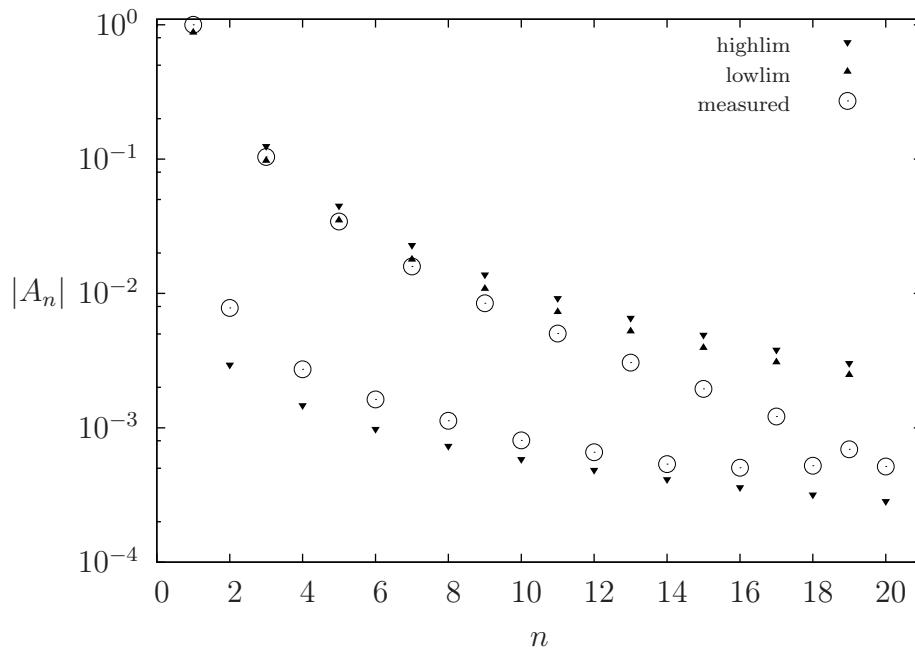


Figure 2.31: A steel string plucked from $\frac{L}{2}$, a logarithmic scale plot

A nylon string provided a great counterpart to the steel string to investigate the true effects of the stiffness of the string. If the same frequency shifts are noticed in the nylon string, then it is not possible to completely blame the stiffness for that observation. Figure 2.32 summarises the results nicely. A similar frequency shift is noticed as with the steel string and the resulting spectrum is similar compared to the steel string. Maybe the only subtle difference is that the node points are more distinctly closer to actually being zero, and the very high upper partials still create the expected form of the logarithmic spectrum.

According to these results, it seems that the nylon string would provide spectrums that are closer to the ideal string than the steel string. Unfortunately this was not the case when analysing the nylon string spectrum for the middle-pluck waveforms. In that case the nylon string did not provide any better results compared to the steel string.

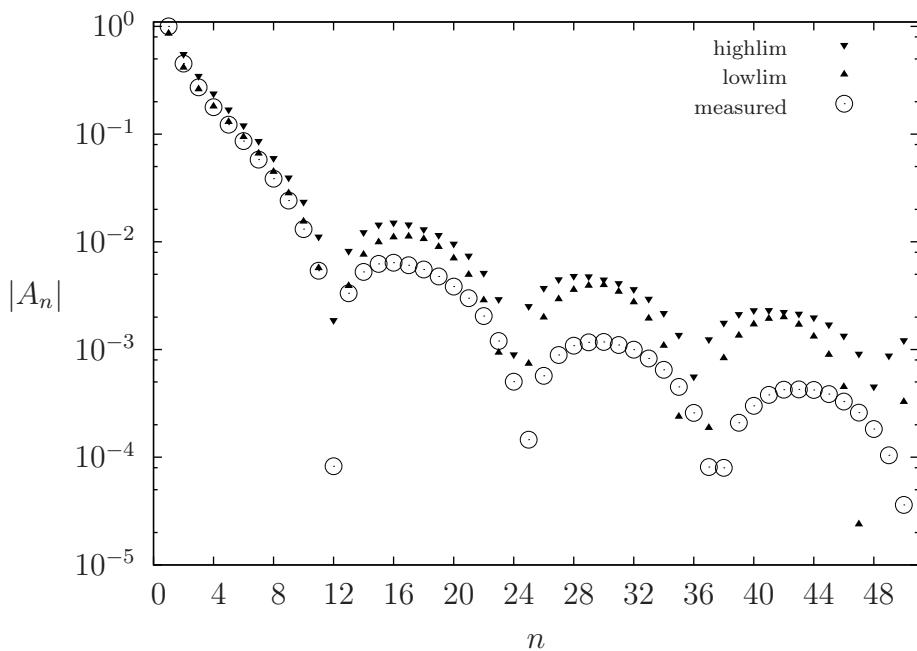


Figure 2.32: A nylon string plucked from $\frac{L}{12}$

2.4.5 Measurement results - Decay of upper partials

The decay times of different frequency components were analysed from the measurement data obtained from plucking the string near the end support. In this situation the high-frequency components are initially the strongest compared to the fundamental frequency and it is possible to follow the relative decay times of the high-frequency components compared to the decay time of the fundamental frequency.

The data used for analysis was recorded from the same pluck-fest sessions as was the data for the spectrum analysis part, but instead of using the oscilloscope, longer data storage was possible only with the computer sound card. Since the focus is on the relative decay of the frequency components, the absolute accuracy of the measurement device is not that significant.

The analysis method was to write a computer program that reads .wav files and locates all such zero-crossings that are constraining one full period of vibration within. This was done by analysing an average amount of data points within one period and allowing the amount of data points to vary approximately by 20 data points around the preanalysed amount of data points within one full period. Another constraint was that the periodicity should continue to the three following zero-crossings, so that three sequences of data points having a period

approximately equal to the preanalysed value were located. This is not a fool-proof method to discard false periods, but clearly in the majority of cases the method works fine.

When one period of data was isolated from the .wav file, the program simultaneously calculated the Discrete Fourier Transform to find out the amplitudes of the first 30 upper partials. The amplitude values and the time value of the located period were written to a text file and the program proceeded to locate the next period in the .wav file. This way a set of time values related to a certain set of amplitude values was obtained.

The decay times are expected to follow the exponential law, so the easiest method to linearise the decay curves is to take a natural logarithm of the amplitude values before plotting. Then the decay time τ can be obtained by fitting a line to the linearised data and calculating the slope k of the fitted line. This follows from the basic equation for the line

$$y_1 - y_0 = k(t_1 - t_0),$$

where a natural logarithm is taken from the y -values as

$$\ln y_1 - \ln y_0 = k(t_1 - t_0).$$

After solving for y_1 this leads to

$$y_1 = y_0 e^{k(t_1 - t_0)},$$

so that when following the general laws of decay, $k = \frac{1}{\tau}$.

Firstly, it is better to visualise the decay curves to show a bigger picture of how the different frequency components behave in steel and nylon strings. Figure 2.33 shows the decay curves for the steel string having a fundamental frequency of about 230 Hz. The string was plucked at distance $\frac{L}{30}$ and the optical pickup was located 8 mm from the end support of the string. All the 29 frequency components obtained from analysis are included in the figure.

Although Figure 2.33 may have more artistic than scientific value, it indicates that the 10 first upper partials are relatively stable and the decay proceeds at a constant rate. The higher frequency components (< 2000 Hz) behave more and more chaotically and they do not form any straight line, but looking closely at the figure, the decay trend is still clearly visible for all frequency components.

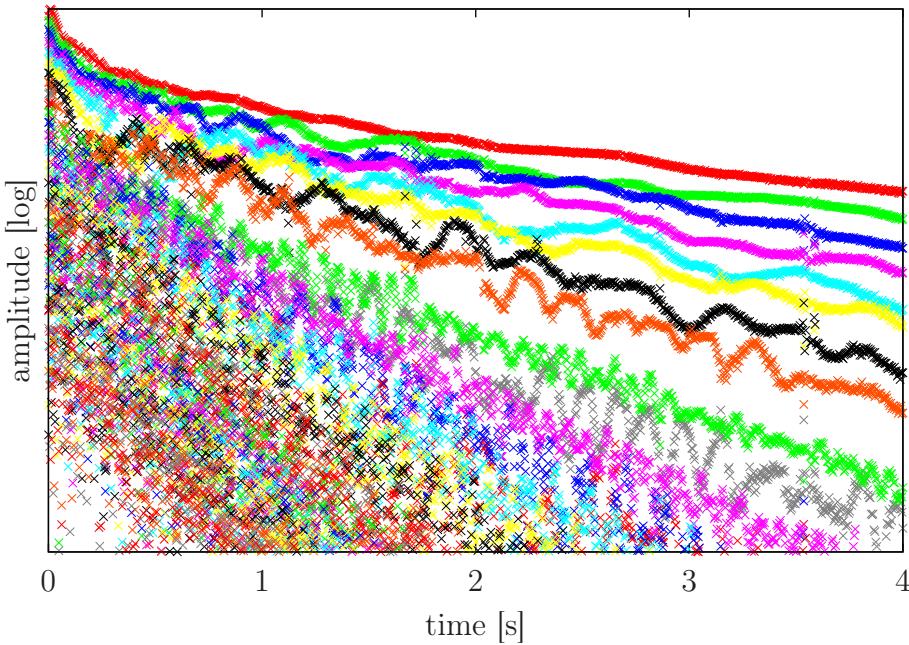


Figure 2.33: Decay of amplitude components in a steel string

As one might easily guess, the nylon string behaves a bit differently from the steel string. In Figure 2.34 the eye focuses on the periodical alternations in the amplitude decay curves. To acquire this set of measurement data, the nylon string with a fundamental frequency of 163 Hz was plucked at $\frac{L}{12}$. Therefore, it is not possible to directly compare this figure to the one representing the decays of the steel string.

The fundamental wave seems to decay at an extremely steady pace in the nylon string, but all the upper partials have a tendency to occasionally decay almost to a minimum, and then gain the amplitude back after a short while. This might be a sign of energy transfer between the frequency components, or then it is just that the higher harmonics are travelling faster in the string (dispersion), and therefore they are not moving within the full period determined by the fundamental wave. Compared to the steel string, the amplitudes of the higher harmonics are more stable and the changes observed in the decay curves are predictable from the earlier data points.

One explanation for the bumpiness of the nylon string decay curves might come from the measurement setup itself. The nylon string was not properly attached from its other free end because there was no similar metal ball at the end as there was on the steel string. Therefore, the nylon string was fixed from the other end only by the clamping mechanism. This affected the string so that it could not be

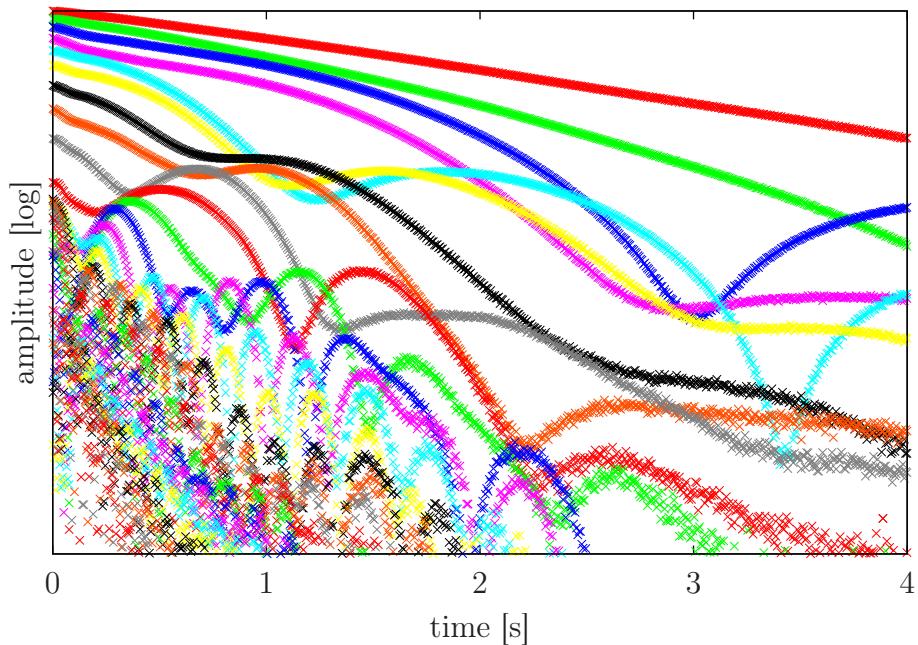


Figure 2.34: Decay of amplitude components in a nylon string

stretched to a normal playing tension, and eventually 160 Hz was seen to be a good compromise at the time of measurement. Both of these reasons could lead to a loose fixation of the string, which could allow the string to move more in the vertical direction, enabling the coupling of the horizontal and vertical modes of vibration. But this is just a guess, because there are no other data available to confirm these suspicions.

The actual goal of these measurements was to prove the predictions of the theory that the higher frequencies decay faster than the lower frequencies. This was already visually noted from the two figures above but more proofs can be provided. After the decay times have been determined by doing a simple line fit to the data and processing k as indicated earlier, the obtained time constants τ can be plotted along a frequency axis. This plot has been prepared in Figure 2.35.

The plot has been drawn to a log-log plot because there the differences can be seen more clearly. One can also imagine connecting the lower frequency data points with a line of some slope k_l and the higher frequency data points with another line having some deeper slope k_h . These lines, however, are not fitted to the figure because there is no clear indication of some specific corner frequency that would divide the lower and higher frequencies apart.

If some line fitting were to be done to the two data series depicted in Figure 2.35,

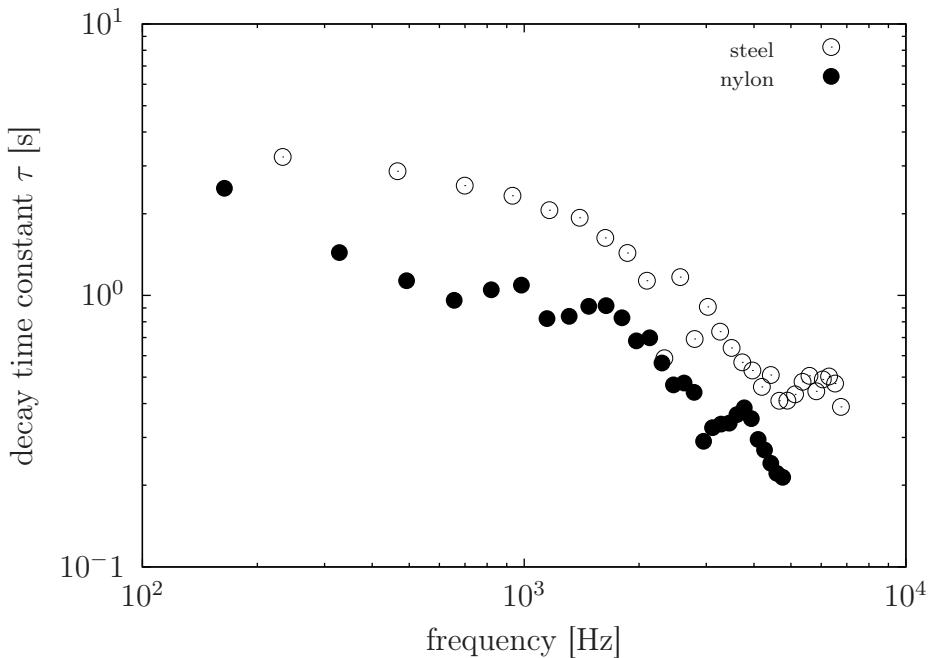


Figure 2.35: Decay time constant versus frequency

this would require taking logarithms from both y' -values and x' -values, namely

$$\ln \tau_1 - \ln \tau_0 = k(\ln f_1 - \ln f_0).$$

After solving for τ_1 , this leads to

$$\tau_1 = \tau_0 \left(\frac{f_1}{f_0} \right)^k$$

so that the slope constant k now defines the exponential behaviour of the ratio of the two consecutive frequencies.

The reason for the two humps in the data series of the nylon string can be easily explained with the fact that when plucking at $\frac{L}{12}$, the nodes are the upper partials 12 and 24. Around these node points the amplitudes are generally very low and they are disappearing into the noise. For this reason no clear continuation of the data series is obtained in the vicinity of the node points.

2.4.6 Measurement results - Pitch glide

Some data from a set of earlier measurements has been reused to show how the vibration frequency changes as the initial plucking displacement grows larger. These measurements are intended to verify the validity of the theoretical formulae derived in section 2.2.1. The main focus is on the results obtained from

equation (2.49), which gives the relative frequency increase as a function of the sum of the amplitude components A_n .

The ΔL in equation (2.49) refers to the length that the string is initially stretched to under the normal playing tension due to F_x , as shown in Figure 2.1. To be able to compare the experimental and theoretical results, ΔL had to be measured when the end point of the string was stretched from its effective resting length L to the total length of $L + \Delta L$. In this experiment, the measured value for the effective length was (845 ± 1) mm and ΔL was (2 ± 1) mm for a steel string and (46 ± 2) mm for a nylon gut string. The nylon string had a silverplated overwound that increased the mass and also the stiffness of the string a tiny amount.

The displacement coefficient a in equation (2.34) refers directly to the displacement of the peak of the initial triangular form of the string before releasing it into motion. This coefficient is connected to the spectral amplitude coefficients A_n in equation (2.46), which describes the length increase of the string due to the initial plucking displacement.

It is possible to derive some reasonable error limits for the measurement results on the displacement dependent frequency from the theoretical equations using the standard method of error propagation [38, p. 75]. For the uncertainty in the measured value of ΔL , one has the error formula

$$\left| \frac{\delta f'_n}{\delta \Delta L} \right| = f'_n \frac{\delta L}{2\Delta L^2 \left(1 + \frac{\delta L}{\Delta L} \right)}.$$

The derivation of the error formula for the displacement variable a requires the use of the chain rule, since the variable is indirectly included in the equation for δL , so that

$$\left| \frac{\delta f'_n}{\delta a} \right| = \left| \frac{\delta f'_n}{\delta \Delta L} \frac{d\Delta L}{da} \right| = f'_n \frac{1}{2\Delta L \left(1 + \frac{\delta L}{\Delta L} \right)} \frac{d\Delta L}{da},$$

where

$$\frac{d\Delta L}{da} = \frac{aL}{\pi^2} \left(\frac{1}{p} + \frac{1}{L-p} \right)^2 \sum_n \frac{1}{n^2} \sin^2 \left(\frac{n\pi p}{L} \right).$$

Another factor of uncertainty comes from the way how the vibration frequency is defined from the number of data samples within one fundamental period. If the vibration frequency is directly calculated as

$$f(N) = \frac{f_s}{N},$$

where N is the number of samples inside one fundamental period bounded by zero-crossings at each side, the error in the calculated frequency is then

$$|df(N)| = |f(N + dN) - f(N)| \approx \left| \frac{df(N)}{dN} \right| dN = \frac{f_s}{N^2}.$$

This uncertainty is mostly overcome by using linear interpolation at the zero-crossing boundaries. The linear interpolation is used in the following data analysis, and therefore the error term arising from the discrete nature of the measurement process can be neglected.

To perform the measurements, a regular computer sound card was used to record the output signal obtained from the optoswitch, which was used to measure the motion of the string. The resulting waveforms were saved as .wav audio files for later analysis. The frequency corresponding to each vibration amplitude was determined as an average value over the first ten periods of vibration. The initial vibration periods are assumed to have approximately the same amplitude as was measured for the plucking displacement when releasing the string to vibrate. The measurement results correlate with the theoretical predictions as indicated by Figures 2.36 – 2.38.

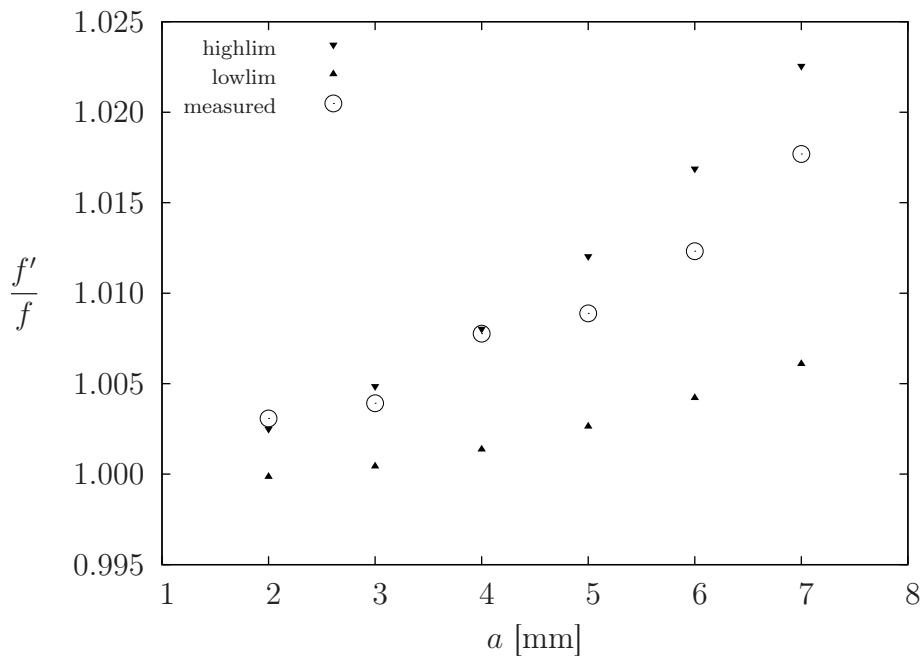


Figure 2.36: Amplitude dependent frequency changes in the measured steel string

The frequency increases measured on the steel string fit quite nicely inside the error limits. The error limits do not have a constant deviation between them because both components of the error propagation formula contribute relatively

equally to the limits. As the amplitude grows, the more it affects the error limits as well. The measured error in the amplitude term was chosen quite large in this case, but despite that, the trend from the theoretical values follows the measured results.

As for the nylon string, at first it seems that the measured values are far from the theoretical calculations. There are two possible outcomes because of this observation. Either the theoretical formula for the frequency change does not hold in practical situations, or there might have been a mistake when determining the initial stretch of the nylon string. The error term arising from measuring the initial stretch is clearly a more dominant term that defines the error limits in this case.

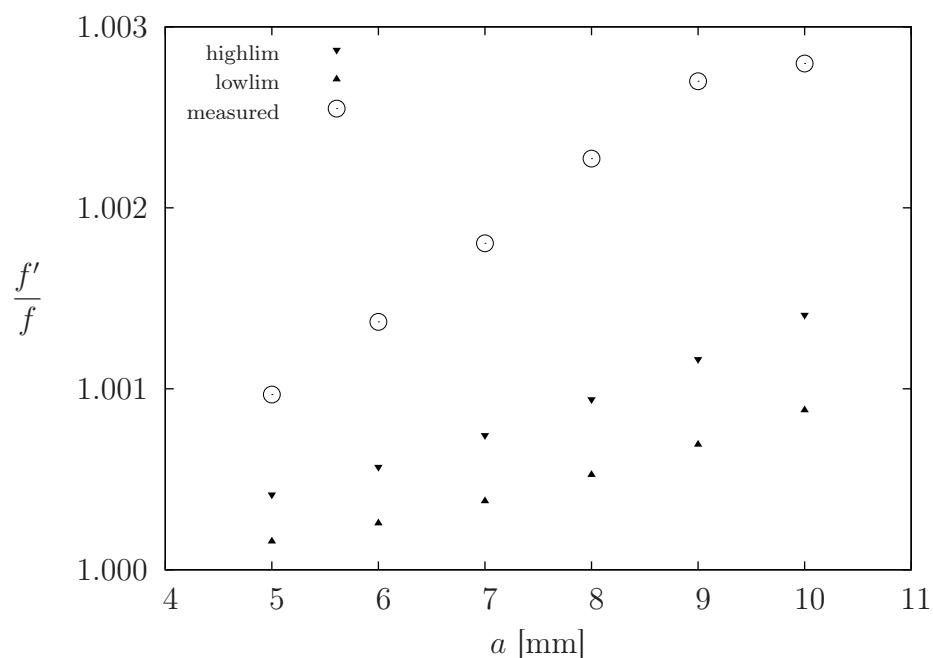


Figure 2.37: Amplitude dependent frequency changes in the measured nylon string

To verify this, the value for the initial stretch used in the calculations is modified to see if there is a value that theoretically gives a behaviour similar to the measurements. The value for the corresponding initial stretch was found to be 16 mm, which produces the plot in Figure 2.38. This stretch is considerably less than the value actually measured, but since the nylon string is so flexible, it is difficult to find the difference between the exact resting length and the length under initial tension. On the basis of this reasoning, it can be assumed that the equations derived in section 2.2.1 are valid and they describe quite accurately the real change in the vibration frequency with respect to a certain plucking am-

plitude.

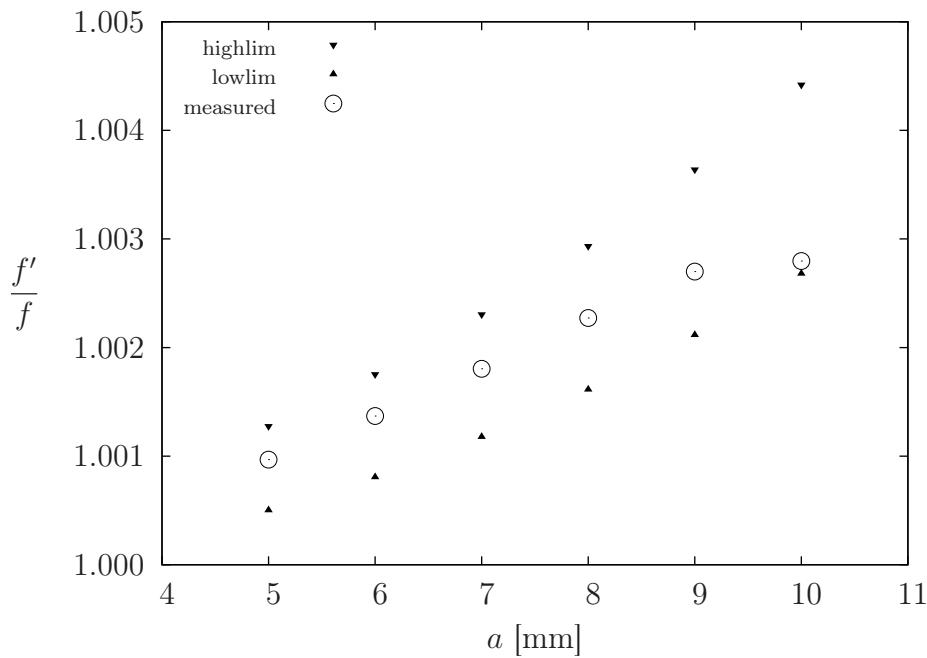


Figure 2.38: Replotted frequency changes in the measured nylon string

The practical significance of these measurements is to indicate the actual frequency shifts occurring with different vibration amplitudes. From the measurements it is easy to draw the conclusion that the nylon strings do not notably change their frequency with modest plucking amplitudes. The steel strings, however, are quite sensitive to the plucking amplitude and in some cases the change in the vibration frequency can be noticed even by ear. This unideality of the steel strings relates directly to the difficulties in tuning the strings. Basically the correct way to tune the strings would be to use very small plucking amplitudes, but then again aggressive playing would result in intonation problems. However, the exponential decay reduces the initial vibration amplitude very fast so that after a few tenths of a millisecond the amplitude has attenuated considerably. From this point of view, despite any given initial displacement amplitude, the string vibrates most of the time in the small amplitude range where the relative frequency changes are negligible.

CHAPTER 3

GUITAR PICKUPS

Guitar pickups sense the motion of the strings and transduce that mechanical motion into an electrical signal. There are different types of pickups available; the most common in the electric guitar being the electromagnetic transducer. Optical transducers are also sometimes applied especially in bass guitars, and to make acoustic guitars electric, contact microphones such as piezoelectric pickups are used. The transfer characteristics vary significantly between displacement sensitive and velocity sensitive pickups. But what is actually meant by displacement and velocity type of transducers anyway? The following sections will clarify this and many other interesting details of guitar pickups.

Before proceeding any further to identify properties of different pickup types, one common factor for all pickups needs to be covered.

3.1 THE EFFECT OF PICKUP POSITIONING

Different pickup types have naturally different transfer characteristics as far as the frequency response in the electrical domain is concerned. However, regardless of the pickup type, the pickup placement relative to the guitar string has a significant effect on the reproduction of the string's tone colour. The influence of the placement is the same for all pickup types, therefore it can be treated as a common topic.

If a pickup is placed at a location P along the length of the string, the pickup senses the vibration of the string only in that specific position. Remembering the different modes of vibration and the standing wave phenomenon of the ideal string, the harmonic modes of vibration have different peak-to-peak amplitudes depending on the point of observation.

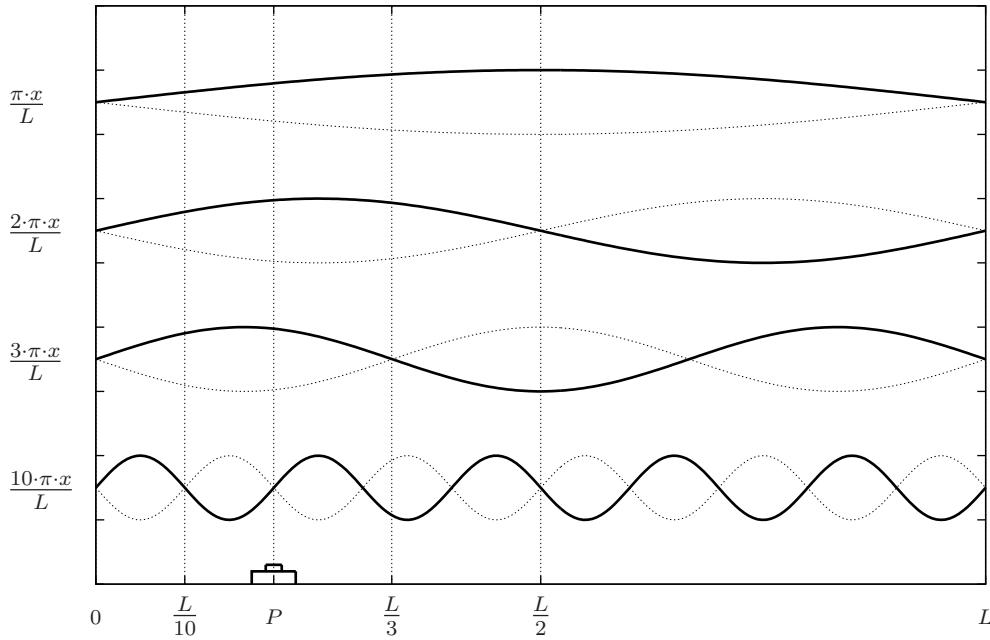


Figure 3.1: Pickup position at $P = \frac{L}{5}$

In other words, the pickup acts as a weight function for the amplitude spectrum components of string vibration. As an example, a pickup placed close to the bridge of the guitar will sense a peak-to-peak displacement of only a few percent from the maximum deflection of the fundamental frequency measured at the middle of the string. Figure 3.1 visualises the significance of placing a pickup to a certain location relative to the length of the string.

As an example of the weight function properties of a pickup, equation (2.34) giving the amplitudes of the upper partials of a plucked string with a pickup positioned at point P along the string would read

$$A_n(P) = A_n \sin\left(\frac{n\pi P}{L}\right) = \frac{2a}{n^2\pi^2} \left(\frac{L}{p} + \frac{L}{L-p}\right) \sin\left(\frac{n\pi p}{L}\right) \sin\left(\frac{n\pi P}{L}\right), \quad (3.1)$$

where $A_n(P)$ is considered to be a function of the pickup position P . The right-most term can have values in the range from 0 to 1, giving the relative amplitude of a sinusoid with a harmonic index n at location P relative to the length of the string. This term will scale the amplitudes of all upper partials with a fraction of their maximum amplitude at $\frac{P}{L}$.

Figure 3.2 gives a visual example of how much the pickup position and type actually modify the natural harmonic content of an ideal string. The theoretical upper partial structures in Figure 3.2 have been evaluated using equation (3.1). Comparing to the spectrum of an ideal string, the theoretical amplitude spectrum

is shown for an optical pickup and a magnetic pickup, which are both placed at $\frac{L}{15}$ and the pluck is executed at the same location $\frac{L}{15}$.

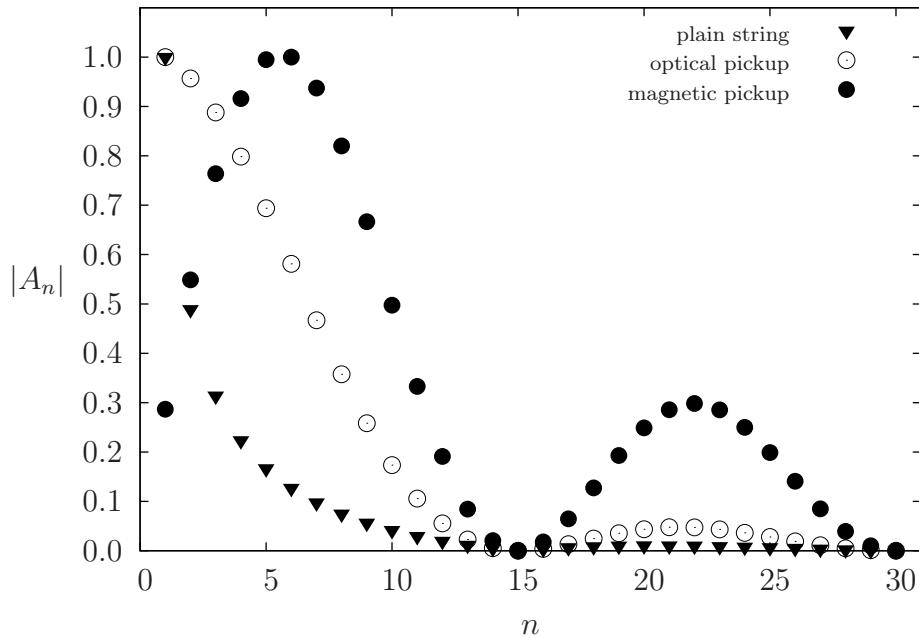


Figure 3.2: Pickup position at $P = \frac{L}{15}$

The analysis above is based on the assumption that the pickup is an infinitesimal point at some distinct position P . Similarly as the finite width δ of a plectrum or a finger introduces a cut-off for upper partials above $\frac{L}{\delta}$, the width of the pickup creates the same effect on its behalf. Furthermore, if the pickup senses the motion of the string from relatively wide area, it is not exactly comparable to the situation where the pickup could be assigned a single-valued coordinate of position as if it was some analytic function. Although these facts are not covered in the included analytic treatment, one should bear these unidealities in mind when observing the results of theoretical derivation and practical measurements.

3.2 MAGNETIC PICKUPS

The key element to make the guitar electric was the invention of the magnetic pickup. The first models were patented already in the 1930's, but the golden era of magnetic pickups started in the late 1950's. Nowadays there are dozens of companies that design and manufacture magnetic pickups for electric guitars. Some of the most well-known manufacturers have been able to earn a reputation of magnetic巫医 doctors because of unrevealed secrets of magnetic pickups. Are these truly secrets, or is it just the basic laws of nature that give that magic touch?

3.2.1 The basic construction and function of a magnetic pickup

Figure 3.3 shows the elementary components included in the construction of a magnetic pickup. The central components are the magnetic polepieces and the coil winding around the magnets. The magnetic polepieces can be separately positioned for each string as depicted in the figure, or there can be just one large bar magnet to cover the area beneath all strings. To keep the whole ensemble together, the magnets and the winding are packed between the top and bottom plates. Usually the coil wire is wound around a plastic bobbin, which has placeholders for the magnets inside.

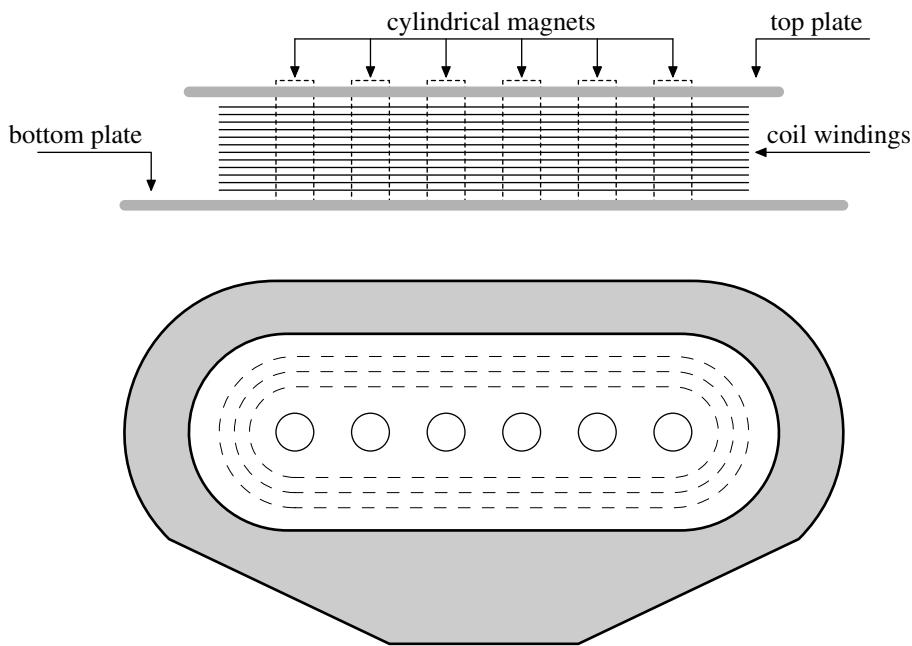


Figure 3.3: The standard construction of a magnetic pickup

The magnetic pickup is a transducer that generates an electric signal from the mechanical motion of the string. As the strings vibrate within the magnetic field of the pickup magnets, the flux of the magnetic field through the coil changes. The changing flux induces an electromotive force in the coil, which causes a current to flow in the coil. This current is an electric representation of the movement of the strings.

In the following sections the parts of the pickup are given more in-depth analysis, starting from the static magnetic field generated by the permanent cylindrical magnets, and ending up to the induced current in the coil. This analysis presents the properties of the magnetic pickup as a signal source. Afterwards the magnetic pickup is modelled as a normal inductor as a part of an electric circuit.

3.2.2 The magnetic field of a pickup magnet

Guitar pickups are normally built around permanent magnets, such as alnico (aluminum, nickel, cobalt) or ceramic magnets. Neodymium and samarium cobalt are also sometimes used as basis materials for guitar pickup magnets. The development trend of pickup magnets started from alnico magnets and has eventually evolved towards ceramic materials, although alnico magnets are still most commonly used. Magnets assembled from these materials create a very strong magnetic field, which extends relatively far away from the surface of the magnet. The stronger the magnetic field, the more sensitive the pickup is to sense the movement of a vibrating string. [39, p. 30]

Over the years, several manufacturers have varied the performance of magnetic pickups by modifying the size, shape and material of the magnetic core of the pickup. All these modifications alter the direction and strength of the magnetic field around the pickup, which has a noticeable effect on the 'sound' of the pickup. The range of used magnet shapes have reached from bulky horseshoe magnets to compact plate magnets and elegant cylindrical bar magnets. A combination of these magnet types is sometimes used to increase the strength of the magnetic field around the pickup. The simplest structure of a magnetic pickup from the point of view of analysis is the one shown in Figure 3.3, which in basic principles fortunately represents a major part of commercial magnetic pickups.

The magnetic field of a regular bar magnet oriented according to the north (N) and south (S) poles is visualised in Figure 3.4. The magnetic field spreads around the magnet and magnetic field lines are used to describe the area where the influence of the magnetic field is reaching. For cylindrical-shaped magnets, the side view of the field is the same; the extension to the third dimension would only reveal the same pattern going around the vertical axis of the magnet.

The magnetic field \vec{B} (also referred to as magnetic flux density) is a vector quantity and it is measured in the unit of tesla [T], named after the Serbian physicist Nicolai Tesla. A typical numerical value for \vec{B} of an average pickup bar magnet can be assumed to be around $0.1 \text{ T} = 1000 \text{ gauss}$, measured within 1 cm's distance from the top surface of the magnet [40]. The magnitude $|\vec{B}|$ depends somewhat exponentially on the distance measured from the surface of the magnet.

A relatively simple mathematical model for the magnetic field of cylindrical bar

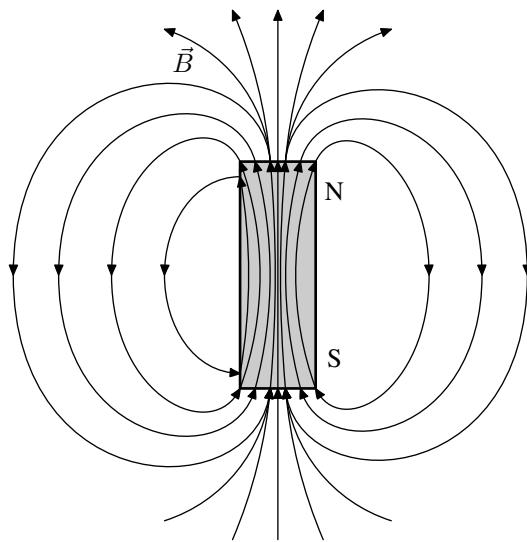


Figure 3.4: Magnetic field lines of a (cylindrical) bar magnet

magnets has been suggested and proven to be relatively accurate in practical experiments of theoretical research [40]. This model states that the main contribution to the changing magnetic field through the coil windings of the pickup comes from the changes of the magnetic field component B_z , which is oriented towards the z -coordinate direction. This statement is based on the idea of magnetic flux through a closed surface, shown in Figure 3.5.

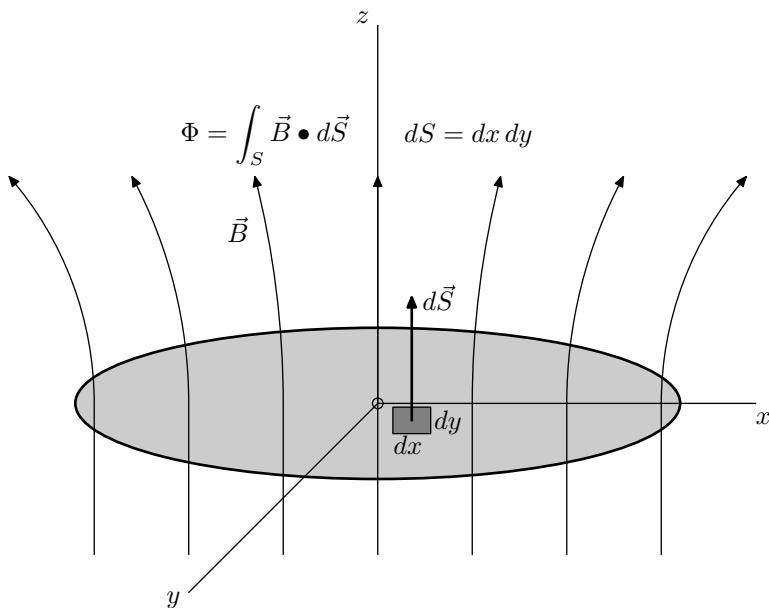


Figure 3.5: Magnetic flux Φ through a closed surface with area S

The total magnetic flux through any given surface is mathematically defined as

$$\Phi = \int_S \vec{B} \bullet d\vec{S}.$$

The differential surface element vector $d\vec{S}$ has a different form depending on the coordinates used. In the basic Cartesian coordinates the surface element vector is defined to be

$$d\vec{S} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ dx & 0 & 0 \\ 0 & dy & 0 \end{vmatrix},$$

which equals $dx dy \hat{k}$. This means that the magnitude of the surface element equals the area of a rectangle with sides dx and dy and the direction of $d\vec{S}$ is perpendicular to the plane formed by linearisations dx and dy . For other coordinate systems, this basic element can be transformed via the Jacobian matrix.

The derivation of the model for the magnetic field of a cylindrical magnet begins from the equation

$$\vec{B} = B_0 \frac{1}{r^2} \hat{r} \quad (3.2)$$

for a magnetic induction field due to a magnetic point charge. The idea of a point charge has been adopted from the definition of electric charge. The fundamental theory of electromagnetism does not directly validate the existence of a magnetic monopole, which this model refers to with the idea of a point charge. In equation (3.2),

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}, \quad (3.3)$$

and \hat{r} is a unit vector of r to give a direction to the magnitude B_0 of the magnetic field \vec{B} . The unit vector is written as

$$\begin{aligned} \hat{r} &= \frac{\vec{r}}{|\vec{r}|} = \frac{(x - x_0)\hat{i} + (y - y_0)\hat{j} + (z - z_0)\hat{k}}{\sqrt{\vec{r} \bullet \vec{r}}} \\ &= \frac{(x - x_0)\hat{i} + (y - y_0)\hat{j} + (z - z_0)\hat{k}}{\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}}, \end{aligned}$$

where the denominator simplifies to r according to (3.3). The vector component to the z -coordinate direction is therefore

$$B_z = B_0 \frac{z - z_0}{r^3}.$$

To expand this ideology from a point charge to a charged surface which would model the top plane of the cylindrical magnet, the infinitesimal areas of the

surface constructing of a group of point charges need to be summed up via an integral. The distribution of the point charges is handled with a function of magnetic charge density σ , which can be considered to be constant for the sake of simplicity. Assuming σ to be constant equals a uniform charge distribution throughout the whole surface, which is a quite realistic assumption.

To form a general model of the situation, the top surface plane of the theoretical cylindrical magnet needs to be attached to a decent coordinate system. The system shown in Figure 3.6 gives an idea of the chosen coordination setup. The

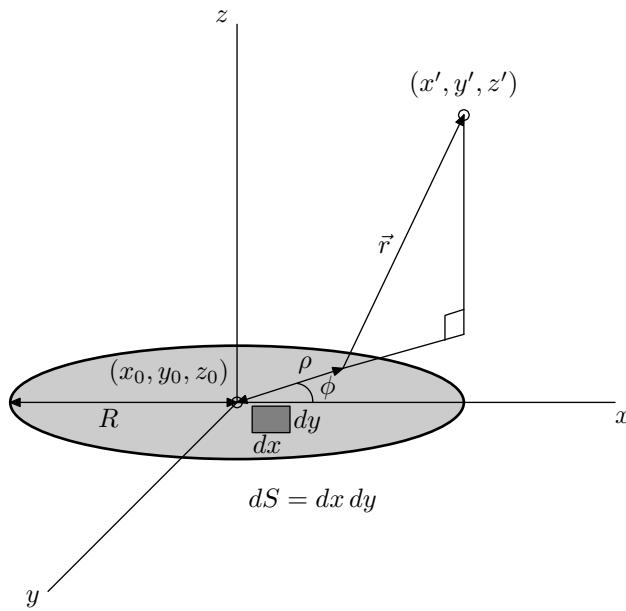


Figure 3.6: Magnetic field vector from a magnetised surface

centre of the surface plane in Figure 3.6 is identified with coordinates (x_0, y_0, z_0) and an arbitrary point (x', y', z') in space experiences the field arising from the magnetised surface.

In Cartesian coordinates the integral to calculate the field component $B_z(x', y', z')$ is

$$B_z = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{\sigma(z' - z_0)}{[(x' - [x_0 - x])^2 + (y' - [y_0 - y])^2 + (z' - z_0)^2]^{\frac{3}{2}}} dy dx, \quad (3.4)$$

where σ represents the constant value of the magnetic charge density. If one is interested only in relative changes of B_z , the magnetic charge density can be assigned to the value of unity. The integral in equation (3.4) is evaluated first with respect to variable y and then with respect to variable x .

Apparently there is no analytical result to this integral, but it is common procedure in engineering sciences to solve these kinds of difficult equations numerically. When inserting values to variables x' , y' and z' in equation (3.4), and solving the integral numerically, it will give the z -directional field strength as the answer. To make the equation simpler, the constants x_0 , y_0 and z_0 can be set to point to the origin of the coordinate system so that they are all zero. Then the point (x', y', z') where the magnetic field is evaluated is measured as a distance from the origin.

To make equation (3.4) fit to the notations used in Figure 3.6, a transformation to a polar coordinate system is in order. To achieve this, substitutions $x = \rho \cos(\phi)$ and $y = \rho \sin(\phi)$ are needed, and the infinitesimal surface area element changes from $dS = dy dx$ to $dS = \rho d\rho d\phi$. Similarly the limits of integration change from the Cartesian domain to the polar domain. With these changes, the integral (3.4) becomes

$$B_z = \int_0^{2\pi} \int_0^R \frac{\sigma(z' - z_0)\rho}{[(x' - [x_0 - \rho \cos(\phi)])^2 + (y' - [y_0 - \rho \sin(\phi)])^2 + (z' - z_0)^2]^{\frac{3}{2}}} d\rho d\phi.$$

This should give the same numerical results as (3.4), but maybe the notation is more intuitive in this form.

Some experimental results of a numerical simulation based on the derived model are presented in Figure 3.7. In this simulation the magnetic field strength B_z of a cylindrical magnet is calculated as a function of distance to the z - and y -coordinate directions from the origin (x_0, y_0, z_0) , which is fixed to the centre of the top surface of the cylindrical magnet. The diameter of the magnet in the numerical analysis was set as 13 mm and the magnetic charge density σ had a constant value of 1.

The purpose of the analysis is to show the relative changes in the magnetic field as a function of the distance from the centre of the magnet, and not to obtain exact values for the magnitude of the magnetic field. The results obtained from the theoretical model have been verified to be in excellent agreement with measurements on actual cylindrical magnets [40]. The measurements in the experiments were made using a Hall probe DC magnetometer to validate the theoretical results.

So far the treatment has concerned only a single cylindrical magnet. In the case of several magnets, the contribution from each magnet is approximately a direct sum of individual field strengths.

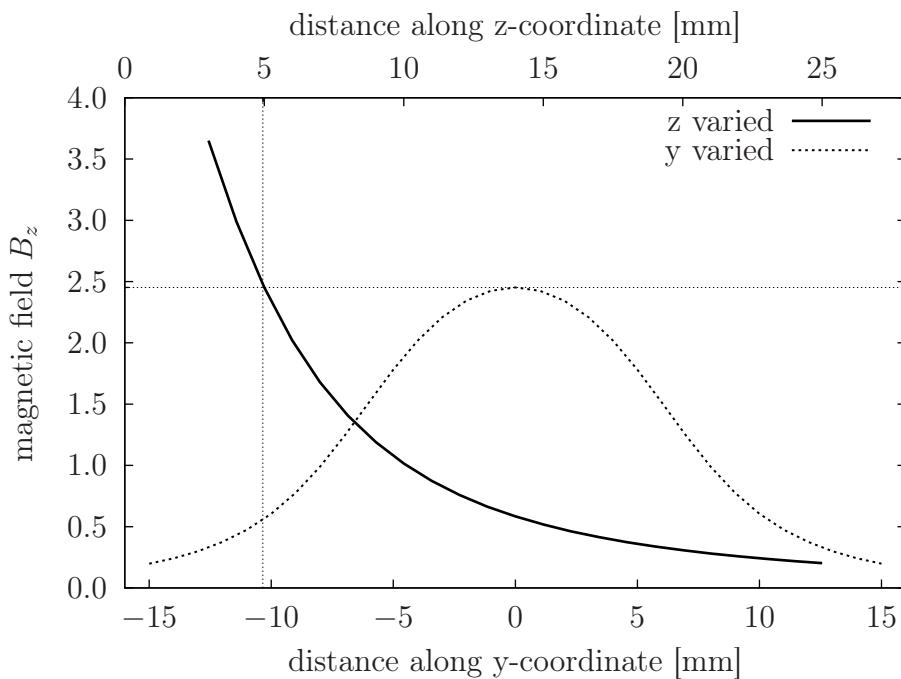


Figure 3.7: Simulated magnetic field of a cylindrical pickup magnet

3.2.3 Magnetic interaction between a pickup and strings

The physical actuator that generates the signal from the magnetic pickup is the interaction of the strings with the magnetic field of the pickup's magnets. This interaction is extremely material dependent, and only certain metals are sufficient for producing decent signal levels. Materials are classified as diamagnetic, paramagnetic or ferromagnetic, based on the material's internal behaviour under the influence of an external magnetic field. Strings for electric guitars are made from ferromagnetic metals such as steel (iron) and nickel because ferromagnetic materials interact most powerfully with external magnetic fields.

The magnetic properties of materials are determined by atomic-level magnetic dipole moments $\vec{\mu}$, which are vectors because of their directional dependency on external fields. Initially in a demagnetised piece of material, the magnetic dipole moments are randomly oriented all around the material and their magnitudes basically cancel each other out. When the demagnetised piece of material is brought close to a permanent magnet that generates a constant magnetic field, the magnetic dipoles align themselves to the direction of the external field, and their individual magnitudes contribute to the same direction. This is the thing that makes the initially demagnetised piece of material magnetic in the atomic level.

Well, that's a fine piece of theoretical jargon, but how does all this apply to guitar pickups? When the guitar strings are not in contact with the magnetic field generated by the magnetic polepieces of the guitar pickup, the strings are demagnetised with the directions of the dipole moments $\vec{\mu}$ randomly scattered.

After the initially demagnetised string is placed on top of the cylindrical magnet of a guitar pickup, the situation is somewhat similar as is drawn in Figure 3.8.

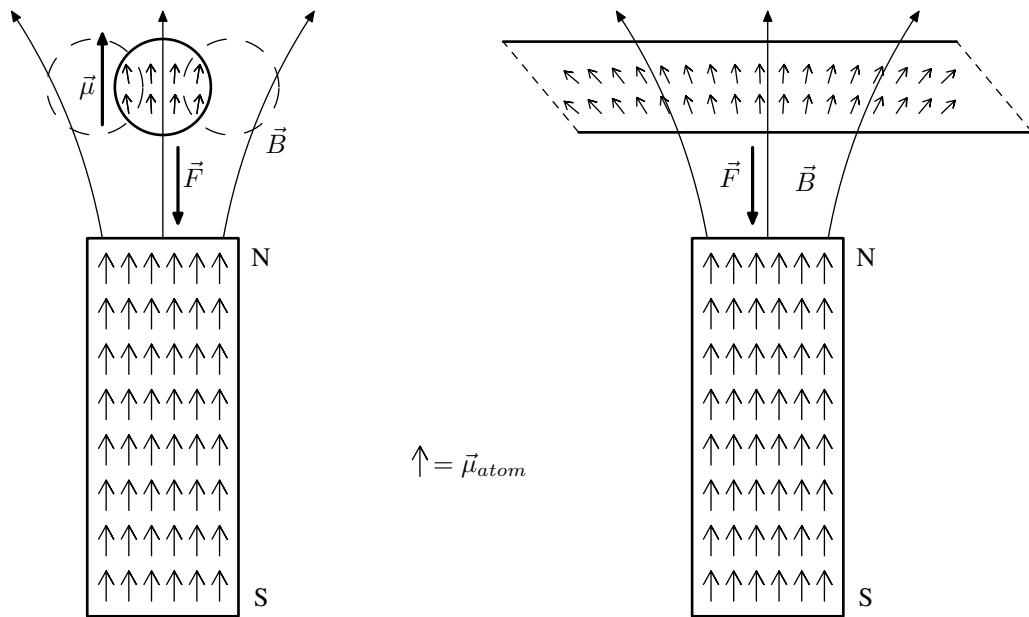


Figure 3.8: String magnetisation within a magnetic field of a pickup magnet

Figure 3.8 tries to visualise the alignment of the magnetic dipole moments $\vec{\mu}$ inside the guitar string after the string is placed into the constant magnetic field \vec{B} of the pickup magnet. The part of the string that is influenced by the magnetic field also becomes magnetised with the same orientation of north (N) and south (S) poles as the pickup magnet has. The similar polarity of magnetisation causes a magnetic attraction force \vec{F} to affect between the pickup magnet and the string. This force pulls the string slightly towards the permanent magnet.

After the string has been magnetised by the pickup magnet, the string and the pickup form a common magnetic unit. The field lines of this magnetic system become connected to each other's south and north poles as shown in Figure 3.9. Motion of the string on top of the pickup changes the total magnetic field of the system and the changes in the magnetic field induce an electromotive force ε to the coil of the pickup.

Returning to the simple model described by Horton [40], the same article extends

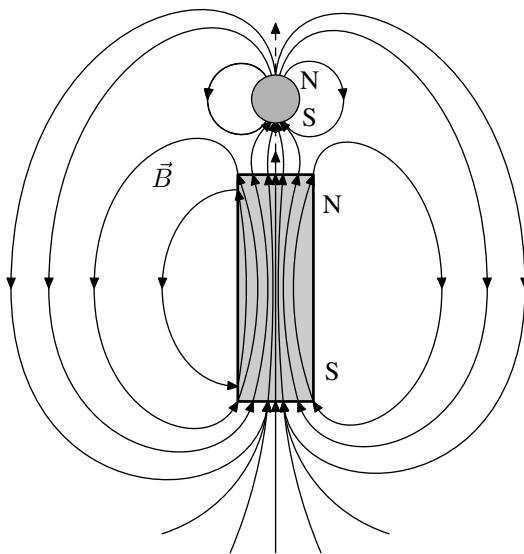


Figure 3.9: The effect of guitar string magnetisation on the field \vec{B} of the pickup

the magnetic field calculations to include a ferromagnetic guitar string above the cylindrical magnet. The ideology presented by the article is quite clever, although it is an oversimplified model of the situation.

The main idea is to eventually be able to simulate the e.m.f. induced in the coil of the pickup. For this objective it is necessary to observe the relative changes of the magnetic field in the direction parallel to the centre axis of the cylindrical magnet. The major assumptions are that the magnetic field of the cylindrical magnet is much stronger than the *local* field of the magnetised string and the magnetisation of a small piece of the string at coordinates (x', y', z') is proportional to the magnetic field of the cylindrical magnet evaluated at (x', y', z') . The total magnitude of that *local* magnetic field at a *specific point of the string* is

$$|\vec{B}_S| = \int_0^{2\pi} \int_0^R \frac{\sigma \rho}{[(x' - [x_0 - \rho \cos(\phi)])^2 + (y' - [y_0 - \rho \sin(\phi)])^2 + (z' - z_0)^2]} d\rho d\phi,$$

where the notation \vec{B}_S is used to indicate the local field of the string at coordinates (x', y', z') , and σ still describes the magnetic charge density, which is taken as a constant in this case due to the assumption of uniform density. This mathematical model follows directly from the geometry of the system. At the location of the string it is not relevant to concentrate only on the z -coordinate direction, but it is absolutely necessary to take the total field to describe the field at the string.

The relative changes of the magnetic field along the z -coordinate direction are

meaningful at point (x_0, y_0, z_p) because according to the analysis of Horton, the total magnetic flux through the pickup coil is *proportional* to the magnetic field at the centre axis of the cylindrical magnet. In reality the total flux would be a sum of all B_z values at locations (x_p, y_p, z_p) within the inner diameter of the pickup's coil, but for the sake of simplicity it is assumed here that in a symmetrical situation the relative changes in the total flux can be estimated by taking the flux as $\Phi_B = B_z(x_0, y_0, z_p)$.

The magnetic field of the small piece of string at location (x', y', z') has influence to the point (x_0, y_0, z_p) according to the geometry of the system. The distance is measured between the points and the unit vector \hat{r} is giving the direction, so that

$$B_z(x_0, y_0, z_p) = \gamma \left| \vec{B}_S \right| \frac{(z' - z_p)}{[(x' - x_0)^2 + (y' - y_0)^2 + (z' - z_p)^2]^{\frac{3}{2}}}. \quad (3.5)$$

The variable γ in equation (3.5) is just a scaling factor, describing the magnetic susceptibility of the string. Equation (3.5) means that changes in the local magnetic field and the relative displacement of the string are reflected to the field component B_z observed at the centre axis of the pickup magnet.

The whole assembly, including the top surface of the cylindrical magnet and the string above the magnet along the x -axis with the vectors connecting the relevant points, is visualised in Figure 3.10. Numerical simulations can be carried out

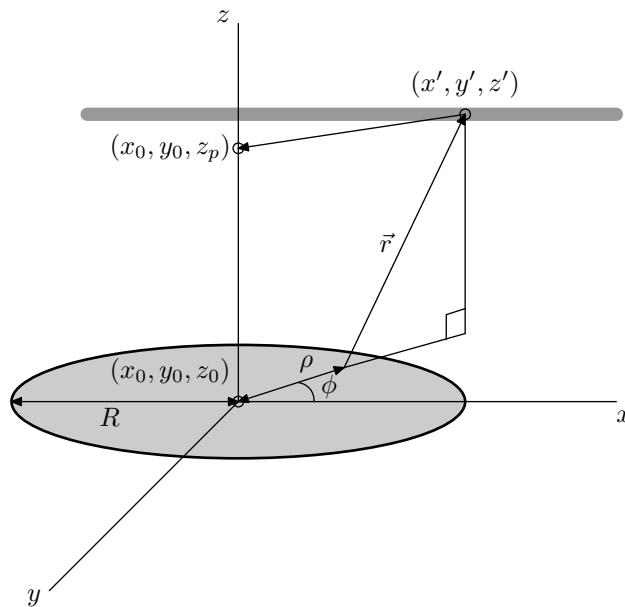


Figure 3.10: The string's magnetic field at (x', y', z') reflects to point (x_0, y_0, z_p)

using equation (3.5) to approximate the (relative) changes in the magnetic flux

in the direction perpendicular to the top surface of the cylindrical magnet and the pickup coil.

However, it is extremely important to note that when using equation (3.5), the string is now localised to a single point (x', y', z') . This is accurate enough if the interest is to examine only relative changes in the magnetic field, since the motion of an ideal string is linear when observing it from one single point (x', y', z') . From the point of view of numerical simulations, the string can be extended in length by adding more localised points along the x -axis. This extension requires that equation (3.5) has to be calculated several times with different values of (x', y', z') and summing the results together. Equation (3.5) can be made even more general by assigning $x_0 = x_p$ and $y_0 = y_p$, where x_p and y_p refer to arbitrary coordinates in x and y directions respectively, and adding together the field changes in several points on the surface of the magnet.

The relative changes of the magnetic field caused by the string at (x_0, y_0, z_p) can be plotted as a function of the horizontal displacement distance from (x_0, y_0, z_s) , where z_s refers to the height of the string above the surface of the magnet. Figure 3.11 draws the calculated relative magnetic field alongside the calculated field of the magnet without a string for comparison purposes. Clearly the changes caused

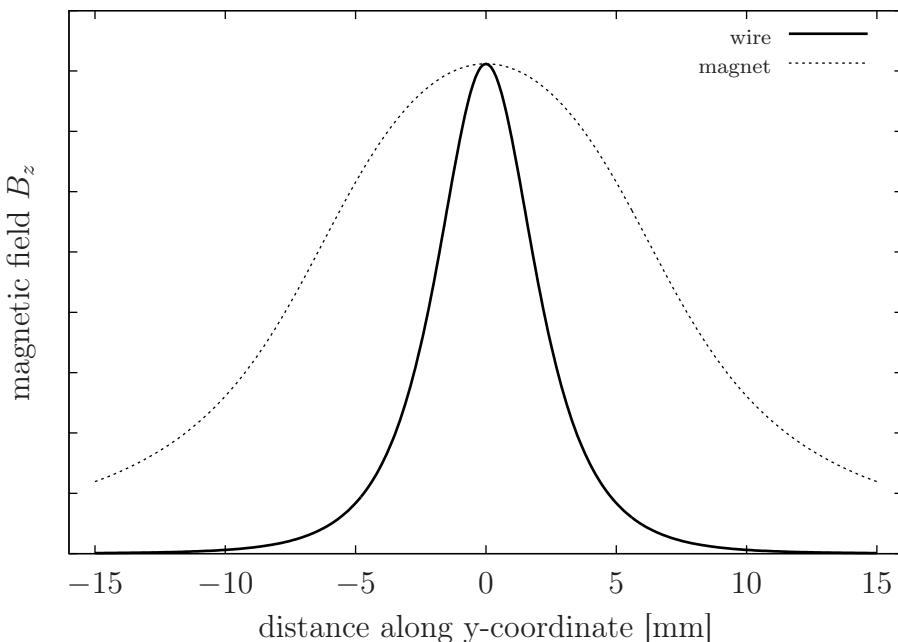


Figure 3.11: Relative changes in the magnetic field at x_0, y_0, z_p due to string displacement

by the string's local field at point (x_0, y_0, z_p) decay more abruptly to zero with

increasing distance compared to the plain field of the magnet. For visualisation purposes the two curves have been scaled to an equal magnitude at 0 mm, but in reality the magnitudes will differ significantly.

Without any simulations, a major conclusion can already be made from the exponential format of the simulation model itself. The magnetic pickup is not a linear transducer for the motion of the string. The displacement of the string is transformed by a 'square root' mapping to the changes of the magnetic flux through the coil of the pickup. This type of distortion is depicted in Figure 3.12, where the solid line represents the purely horizontal vibration with a 4 mm amplitude, and the dashed line indicates the vertical direction of vibration with a 0.75 mm amplitude. The numerical calculations were carried out using a 13 mm magnet diameter and the string was placed 5 mm above the centre of the magnet's top surface. The changes in the magnetic field were observed at 2 mm above the magnet.

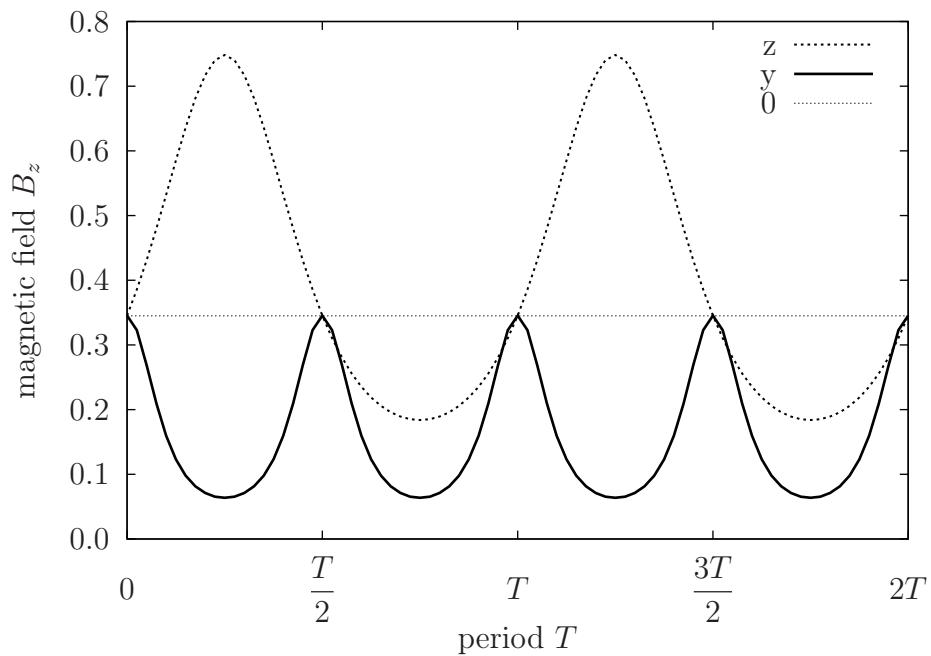


Figure 3.12: Simulated magnetic field changes due to string vibration

Clearly the waveforms are not purely sinusoidal, although in this case the amplitudes have been chosen to be unrealistically large to enhance the effect of nonlinearity. Similar observations have also been made with other mathematical pickup models, as indicated in article [41].

An interesting artifact in the results shown in Figure 3.12 is the relatively large

difference in sensitivity between the horizontal and vertical directions. A small deflection to the vertical direction changes the field considerably more than the corresponding deflection to the horizontal direction. Furthermore, the period of the vibration of the horizontal deflection seems to be half from the vertical period. The periodicity of the horizontal vibration will not yet become clear from this context because this model does not include the effect of the plucking position. The doubling of frequency is true only for symmetric vibration about the centre of the magnet, which is theoretically obtained by placing the pickup and plucking position to $\frac{L}{2}$. Later it is observed that asymmetry caused by the plucking position will reveal the true period of the horizontal vibration. Anyhow it is quite shocking that in certain theoretical cases the fundamental frequency is completely absent from the transduced signal.

A small numerical simulation session was carried out to find out more about the nonlinearity of the guitar pickup. Numerical analysis was done with the model by setting the centre of the magnet's top surface at the origin of the coordinates, and the string to 3 mm above the magnet with a diameter of 13 mm. The motion of the string went through an analytical sine wave period, which in case of a linear transducer should be measured as a pure single-frequency sine wave in the magnetic field oscillation. The changes in the magnetic field component B_z were calculated at (x_0, y_0, z_p) , where z_p was 1 mm above the midpoint of the magnet. The vibration amplitudes used were 0.1 mm, 1.1 mm, 2.1 mm and 3.1 mm. The simulation was executed for one full vibration period and the Discrete Fourier Transformation was used to calculate the harmonic spectrum of the curve describing the change of B_z . The numerical results of the DFT analysis are presented in Table 3.1. The amplitude of the primary component has been normalised to unity so that the amplitudes of the upper partials are percentual values from the primary component.

Because of the symmetry in the simulation setup, the fundamental frequency is missing from the results and the second harmonic frequency is now taken as the fundamental. That is why every second amplitude is zero in the set of tabulated values. The results show that only vibration amplitudes below 0.1 mm will be mapped to a relatively clean sine wave by the pickup magnet. In practise this means that the notes played on the electric guitar will be slightly distorted for a few tenths of a second after which the amplitude of the string vibration will decay to a displacement level where the pickup gives an almost linear response to the played note.

Table 3.1: The frequency response of B_z for a sine wave input

n	0.1 mm	1.1 mm	2.1 mm	3.1 mm
f_1	0.00000000	0.00000000	0.00000000	0.00000000
f_2	1.00000000	1.00000000	1.00000000	1.00000000
f_3	0.00000000	0.00000000	0.00000000	0.00000000
f_4	0.00078408	0.08284069	0.23003543	0.36968213
f_5	0.00000000	0.00000000	0.00000000	0.00000000
f_6	0.00000054	0.00638874	0.04932321	0.12774577
f_7	0.00000000	0.00000000	0.00000000	0.00000000
f_8	0.00000012	0.00047469	0.01019443	0.04260305
f_9	0.00000000	0.00000000	0.00000000	0.00000000
f_{10}	0.00000004	0.00003447	0.00205996	0.00138987

3.2.4 Electromagnetic induction in a magnetic pickup

Now that it is approximately clear how the magnetic field of the magnetic pickup behaves, it is time to examine how the changes in the magnetic field are transformed into an electric signal. The fundamental interaction between the magnetic field and the induced e.m.f. is specified by Faraday's law of induction. When using a magnetic pickup as a transducer, the frequency of magnetic field oscillation is directly related to the frequency of the string vibration, whereas the amplitude of the induced electromotive force ε is related to the *velocity* of the string vibration. This follows directly from Faraday's law,

$$\boxed{\varepsilon = -\frac{d\Phi_B}{dt}}, \quad (3.6)$$

where the rate of change of the magnetic flux Φ_B is directly proportional to the induced electromotive force ε . If the coil consists of several turns of wire, as definitely is the case for a magnetic pickup, the induced electromotive force is scaled by the number of turns N , namely

$$\varepsilon = -N \frac{d\Phi_B}{dt}.$$

The link between magnetic flux Φ_B and the magnetic field \vec{B} is defined as

$$\Phi_B = \int \vec{B} \bullet d\vec{S}, \quad (3.7)$$

in the sense that the total magnetic field through some surface S sums up as the magnetic flux Φ_B . The dependency between the flux and the magnetic field is illustrated for reference in Figure 3.13. The idea of a flux was already used when examining the model of the pickup magnet.

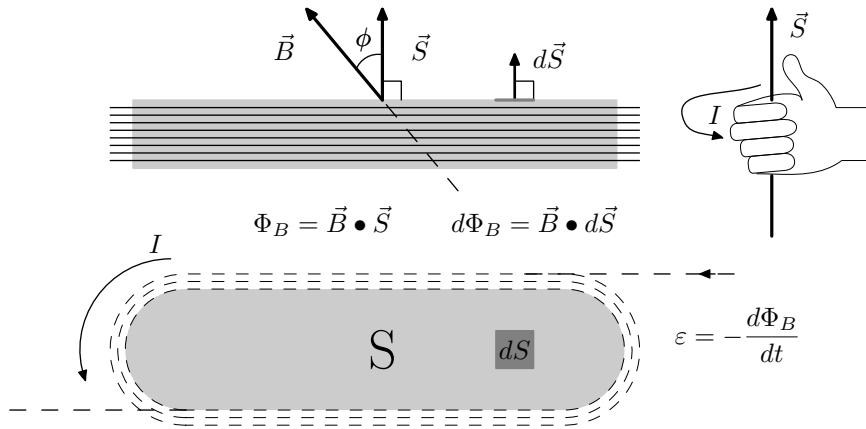


Figure 3.13: Magnetic flux through pickup windings

When the amplitude of the induced e.m.f. depends on the string's velocity instead of displacement, significant differences occur in terms of the spectrum of the string vibration. Equation (2.32) in Chapter 2 describes the motion of the string and it is based on displacement in the $\{xy\}$ -plane. To represent the velocity of the string, a time derivative of equation (2.32) is needed. Differentiation of equation (2.32) with respect to time yields

$$\frac{\partial y(x, t)}{\partial t} = - \sum_{n=1}^{\infty} \frac{c_T n \pi A_n}{L} \sin\left(\frac{n \pi x}{L}\right) \sin\left(\frac{c_T n \pi t}{L}\right), \quad (3.8)$$

and inserting the multipliers A_n ,

$$\frac{\partial y(x, t)}{\partial t} = - \frac{2 a c_T}{\pi} \left(\frac{L}{p} + \frac{L}{L-p} \right) \sum_{n=1}^{\infty} \frac{1}{n} \sin\left(\frac{n \pi p}{L}\right) \sin\left(\frac{n \pi x}{L}\right) \sin\left(\frac{c_T n \pi t}{L}\right).$$

Now the amplitudes of the harmonic components fade away with the leading term $\frac{1}{n}$ instead of $\frac{1}{n^2}$ of equation (2.34). This means that the higher spectral harmonics of the string vibration will be more dominant when using the magnetic pickup as the electrical transducer device. As mentioned in section 3.1, the positioning of the pickup makes an additional contribution to the spectrum. Figure 3.14 gives an idea of the situation by plotting the amplitude components of string vibration when the string is plucked at $\frac{L}{6}$ and the magnetic pickup is located at $\frac{L}{9}$ along the string. The difference to the plain spectrum of the plucked string shown in Figure 2.10 is enormous.

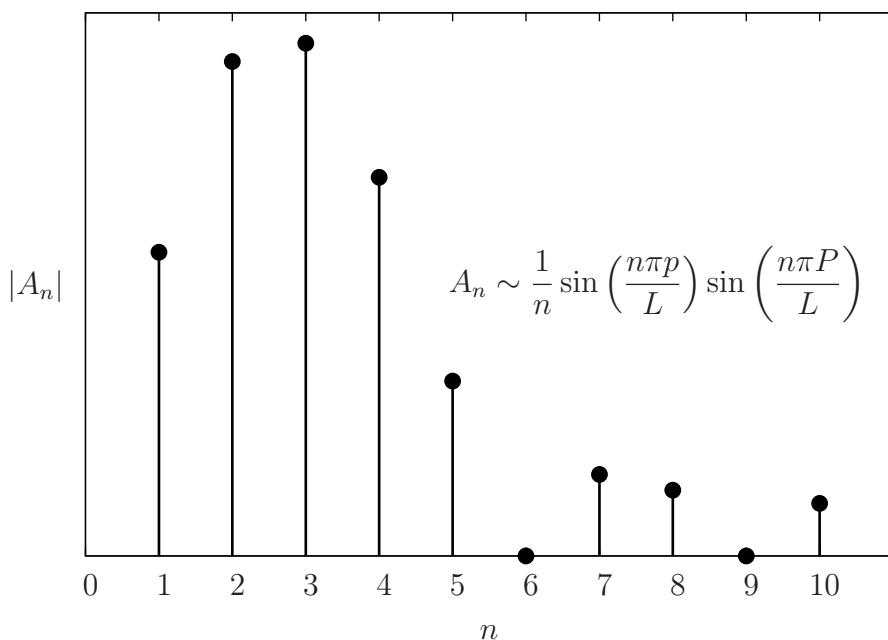


Figure 3.14: Amplitude components of a string; measured from a magnetic pickup

In relation to the earlier treatment of changes in the magnetic flux due to string vibration, the model of equation (3.5) can be used with a real string displacement model to simulate the waveform of the induced electromotive force to the coil of the pickup. The induced e.m.f. is obtained from equation (3.5) by differentiating it with respect to time - just like Faraday's law says. Figure 3.15 depicts a situation where the pickup is placed at position $\frac{L}{4}$ and the string is released to vibrate in a mixed horizontal/vertical direction by plucking at position $\frac{L}{6}$ along the length of the string.

The vibration sequence is modelled by the ideal string model, where the two identical waves are moving to opposite directions and reflecting without losses from the boundaries. Only ten upper partials were used to generate the form of the string to have some 'realistic' roundness to the waveforms. In a pure theoretical case when an infinite amount of upper partials are included in the calculations, the waveforms would have sharp edges and long straight line segments. The initial conditions for the numerical analysis are similar as in the earlier simulations. The string was located at 5 mm above the magnet and changes in the magnetic field were observed at 2 mm above the magnet. The magnetic field component B_z at the centreline of the magnet was chosen to represent the relative flux of the whole magnet, so no additional sums of B_z at different locations need to be calculated.

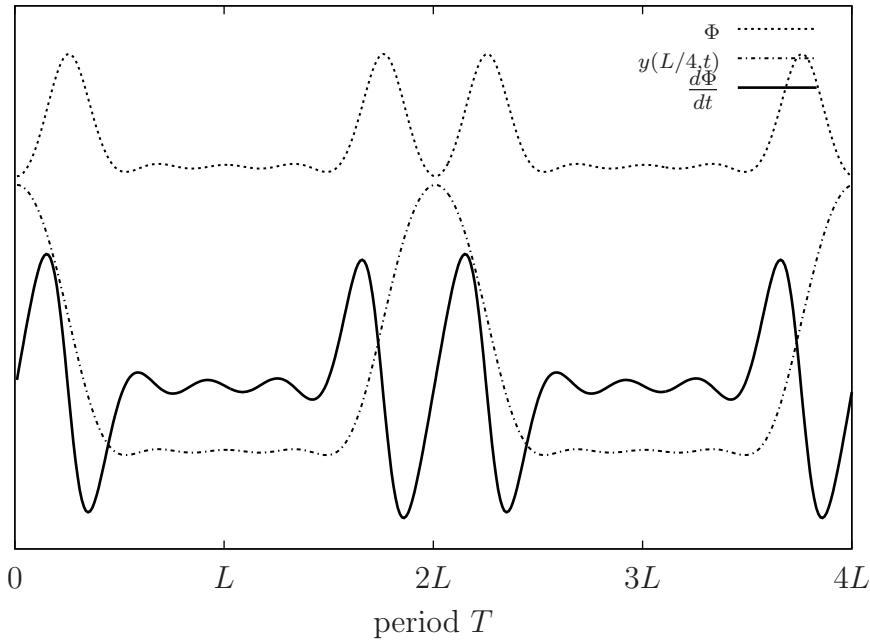


Figure 3.15: String displacement versus induced e.m.f.

The period T of the vibration sequence can be deduced from equation (2.1) by the relation

$$f = \frac{1}{T} \quad \Rightarrow \quad T = \frac{1}{f} = 2Lc_T,$$

which indicates the time that the wave moves a distance $2L$ with a velocity c_T of the transverse waves propagating in the string. This clears out the obscurities in the period of the horizontal vibration, which were indicated in Figure 3.12.

The situation identical to Figure 3.14 can be reproduced from the perspective of the model provided by Horton. By initialising the numerical analysis to calculate the pickup response for a string plucked at $\frac{L}{6}$ with a 2 mm amplitude and the pickup placed at $\frac{L}{9}$, a harmonic spectrum can be plotted from the graph of the current induced to the coil of the pickup. The numerical calculations are exactly the same as when simulating the curves in 3.15, but now 1000 spectrum components were taken along to the analysis. For comparison purposes, the analytic differential of string displacement describing the velocity of the string is shown along with the simulation results in Figure 3.16.

Otherwise the waveforms describing the induced e.m.f. appear to be similar, but whereas the idealistic analytical model builds up as a square wave, the simulated model results in more like a shark's fin wave.

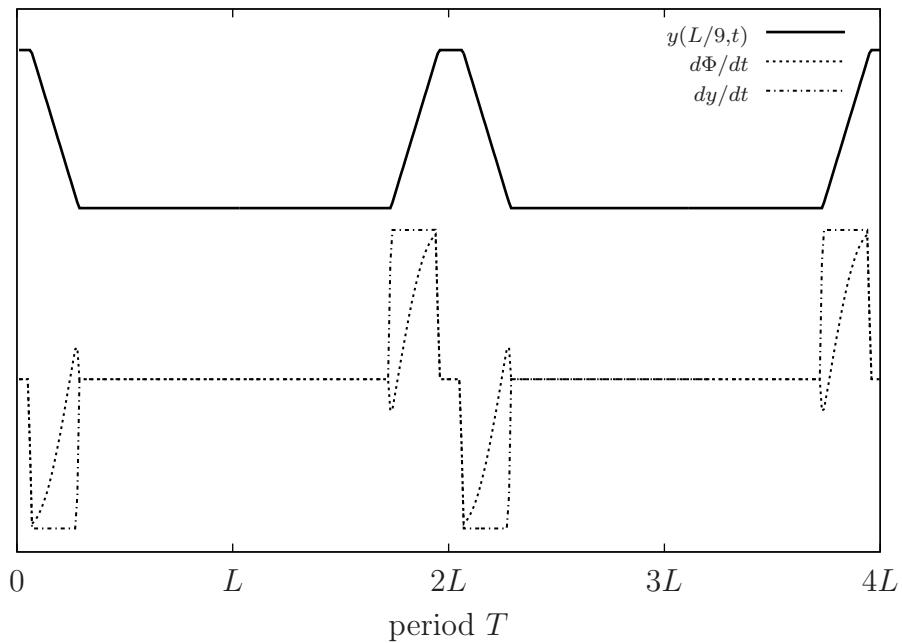


Figure 3.16: String displacement versus simulated e.m.f. and theoretical result

Now it is of course interesting to see how well the amplitude spectrums of these models match each other. Figure 3.17 indicates that the analytical differential model predicts correctly the nodes caused by the plucking position and pickup location. The simulated model does not locate either of these nodes and it does not follow the theoretical expectations very well either.

To defend the validity of the simulation model, it has to be noted that the simulation was run with a relatively large string displacement which creates additional nonlinearity to the results. In addition, the initial position of the string was defined to be in the middle of the magnetic polepiece. This way the simulated horizontal pluck vibrates on both sides of the centreline, also creating additional nonlinearity, which can be noticed as the small step before the shark's fin in Figure 3.16. If the vibration amplitude is kept small and the resting position of the string slightly off from the centreline, the simulation model would in fact get very close to the ideal model.

It is necessary to investigate the linearity issue also from the viewpoint of induced e.m.f. A similar analysis was conducted as in the case of magnet response; the string was modelled as if it had been plucked from the midpoint, and only the fundamental frequency component was retained to give a pure sine wave motion. If the response of the pickup is linear, this simulation setup should produce an ideal sine wave as the induced e.m.f. of the pickup. In the simulation setup, the

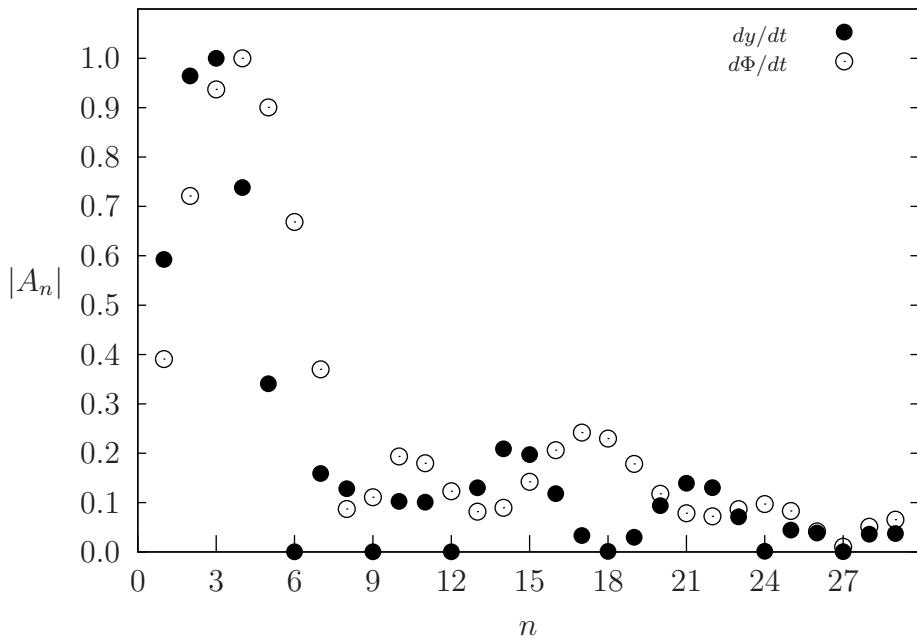


Figure 3.17: The amplitude spectrum of the waveform induced to the pickup coil

string was vibrating horizontally 3 mm above the magnet and the changes in the magnetic field were recorded at 1 mm above the magnet. The amplitudes of the horizontal string displacements were 0.1 mm, 1.1 mm and 3.1 mm. Figure 3.18 draws a nice conclusion of the numerical analysis results.

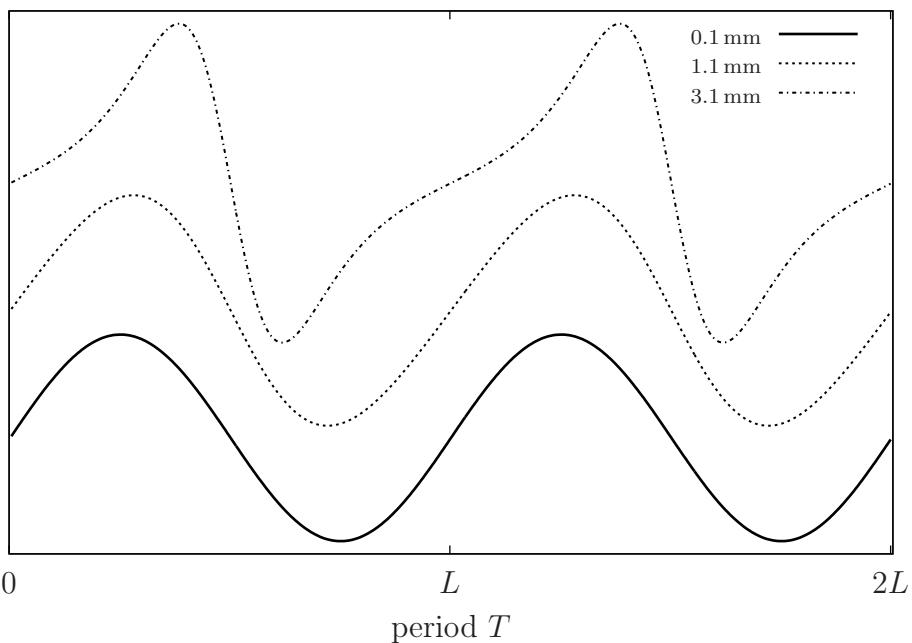


Figure 3.18: Induced waveforms for different amplitude sine wave input

It seems that the simulation model would still give a reasonably linear e.m.f. response even for 1 mm displacement amplitudes. As the amplitude grows, the sine wave starts to get distorted by kind of falling towards the right.

Although these results seem to be very rational, it is not possible to say for sure that the absolute limit for an undistorted response is exactly the 1 mm displacement amplitude. These simulations are extremely good for indicating relative differences between different amplitudes, but undoubtedly they cannot predict the absolute values of vibration amplitudes that will become distorted by the magnetic pickup. The simulations used unity values for the magnetic susceptibility γ of the string and for the magnetic charge density σ . In reality, these and many other small details that are now neglected will have an effect on the results.

3.2.5 The magnetic field generated by the coil of a pickup

Based on the analogy between electric and magnetic fields, the current induced to the coil of the pickup should by assumption generate a similar magnetic field that initially caused the current to flow in the coil. This assumption can be analysed by deriving an expression for the magnetic field generated by a current carrying solenoid.

When a current is flowing in a solenoid or in any piece of wire in general, a magnetic field is generated. According to basic textbook physics, in case of a solenoid the magnetic field due to current flow is quite similar to the field generated by a permanent cylindrical magnet. For basic analysis, it is enough to examine the field generated by one current carrying loop of a solenoid, depicted in Figure 3.19, to see whether the expressions have similarity to equations presented in section 3.2.2. To obtain an expression for the magnetic field of current carrying wire shown in Figure 3.19, one can use the Biot-Savart law. This law states that

$$\vec{B}(\vec{r}) = \frac{\mu_0 I}{4\pi} \oint_S \frac{d\vec{l}}{|\vec{r} - \vec{r}_0|^2} \times \frac{\vec{r} - \vec{r}_0}{|\vec{r} - \vec{r}_0|}, \quad (3.9)$$

meaning that the total field at some point of space is obtained by integrating around a closed path S in space. In equation (3.9) it is assumed that the point (x_0, y_0, z_0) is not fixed to the origin of the coordinate axes. For simplicity, in this case one can take point (x_0, y_0, z_0) as the origin so that $\vec{r}_0 = 0$. Then the Biot-Savart law is simply

$$\vec{B}(\vec{r}) = \frac{\mu_0 I}{4\pi} \oint_S \frac{d\vec{l}}{|\vec{r}|^2} \times \frac{\vec{r}}{|\vec{r}|} = \frac{\mu_0 I}{4\pi} \oint_S \frac{d\vec{l}}{|\vec{r}|^2} \times \hat{r}, \quad (3.10)$$

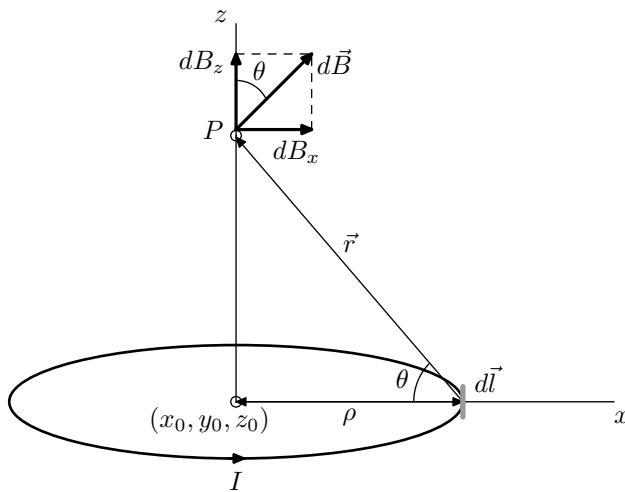


Figure 3.19: The magnetic field at a distance P from the centre of a current loop

where \hat{r} is a unit vector in the direction of \vec{r} .

In the situation depicted in Figure 3.19, the small $d\vec{l}$ piece of conducting wire is always perpendicular to the position vector \vec{r} . Therefore, the differential element of length can be handled as a non-vector quantity, and then an infinitesimal slice of \vec{B} has the magnitude

$$|d\vec{B}| = \frac{\mu_0 I}{4\pi} \frac{dl}{|\vec{r}|^2} \frac{|\vec{r}|}{|\vec{r}|} = \frac{\mu_0 I}{4\pi} \frac{dl}{|\vec{r}|^2},$$

which is obtained simply by taking the absolute value of \vec{r} and noting that the differentiation cancels out the integral from the right side.

From the geometrical setup of Figure 3.19 it also follows that the vector product $\vec{r} \bullet \vec{r} = \rho^2 + z^2$, so that

$$|d\vec{B}| = \frac{\mu_0 I}{4\pi} \frac{dl}{(\rho^2 + z^2)}.$$

The differential magnetic field vector $d\vec{B}$ can be separated into the axial components of dB_x and dB_z , from which the component

$$dB_z = \frac{\mu_0 I}{4\pi} \frac{dl}{(\rho^2 + z^2)} \cos \theta = \frac{\mu_0 I}{4\pi} \frac{dl}{(\rho^2 + z^2)} \frac{\rho}{(\rho^2 + z^2)^{\frac{1}{2}}}.$$

From here the total field B_z can be integrated to reveal the equation

$$B_z = \frac{\mu_0 I}{4\pi} \frac{\rho}{(\rho^2 + z^2)^{\frac{3}{2}}} \int dl = \frac{\mu_0 I \rho}{2(\rho^2 + z^2)^{\frac{3}{2}}}, \quad (3.11)$$

which is similar to the field-equation of a cylindrical magnet. The most important term to note here is the $(\rho^2 + z^2)^{\frac{3}{2}}$ term in the denominator. This term arises from the geometry of the system and is therefore similar to the expression of

a cylindrical magnet field. The factorial multipliers in front of equation (3.11) are not directly transferable to depict magnets but similar scaling terms can be identified from magnet analysis.

To apply the solenoid model to cylindrical magnet modelling, the effect of a continuous set of loops covering the whole distance of ρ as depicted in Figure 3.19 needs to be included in the model. For this purpose, another integral needs to be added to integrate the top plane of the magnet along ρ . The model of equation (3.11) only gives the magnitude of B_z at a distance along the axis of the solenoid, so the extension to an arbitrary position in space would need to be added.

3.2.6 A single-coil pickup vs. a humbucker

The primary reason for the development of the so-called 'humbucker' pickup is related to electromagnetic noise problems of the single-coil pickups. When talking about 'hum' in the context of magnetic guitar pickups, one usually means the electromagnetic radiation arising from the 50/60 Hz power line, which is present in every building and of course nearby all power transmission line networks. The fact that electric guitars need electricity to work as intended makes the problem of hum inavoidable, but with some clever tricks it is possible to reduce the hum considerably.

The basic physics behind the theory of electromagnetic waves is rather complicated and it is based on a set of wave equations derived from Maxwell's theory of electromagnetism. These wave equations have a similar form as the wave equation of the ideal string, the format being the same for both \vec{E} and \vec{B} . The wave equation for the electric field is

$$\nabla^2 \vec{E} = \epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

and for the magnetic field,

$$\nabla^2 \vec{B} = \epsilon_0 \mu_0 \frac{\partial^2 \vec{B}}{\partial t^2},$$

where the differential operator *nabla* is a shorthand notation of the form

$$\nabla = \frac{\partial}{\partial x} \hat{i} + \frac{\partial}{\partial y} \hat{j} + \frac{\partial}{\partial z} \hat{k}.$$

The solutions of the differential equations for \vec{E} and \vec{B} are sinusoidal waves, just like in the case of an ideal string. One of Maxwell's four equations says that

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}.$$

Without further derivation this links the electric and magnetic fields together, the factor of proportionality being $\frac{1}{c}$,

$$\vec{B} = \frac{1}{c}(\hat{k} \times \vec{E}).$$

In another words, where there exists an electric field, there is also a magnetic field present. Figure 3.20 is drawn to depict the relation between fields \vec{E} and \vec{B} .

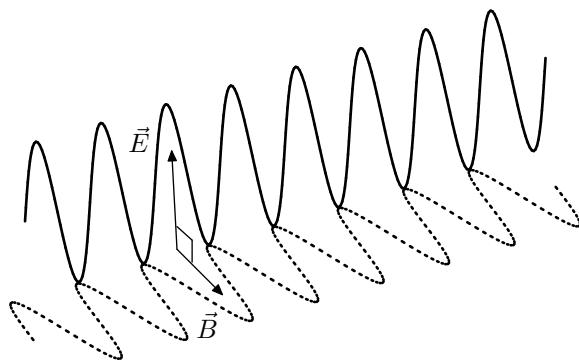


Figure 3.20: The polarized electromagnetic wave of fields \vec{E} and \vec{B}

The other fact of the matter is that these electromagnetic waves are propagating everywhere. The velocity of propagation depends on the medium where the waves travel but generally electromagnetic waves move with the speed of light. Although the transverse vibrations of the electromagnetic waves would have a low frequency, the waves still propagate in air and in other substances at extremely high velocity. The velocity of propagation and the frequency of transverse vibrations of \vec{E} and \vec{B} are not connected in any way, in contrast to the case of the vibrating string. One simply cannot hide the pickup from these waves, and because the magnetic field component of electromagnetic wave changes with time, it induces a current to every inductor on its path of propagation, including guitar pickup coils.

One solution to protect the pickup from the low-frequency hum would be to enclose the pickup inside a metal box because sufficiently thick pieces of highly conducting material reflect most of the electromagnetic radiation. The portion of the wave that is not reflected away penetrates into the material and slowly attenuates as it penetrates deeper. A quantity δ called the skin depth gives a measure of the distance of penetration where the wave has attenuated by a factor

of e^{-1} . In mathematical terms, the skin depth is defined as

$$\delta = \sqrt{\frac{1}{\pi f \mu \sigma}}, \quad (3.12)$$

where μ is the magnetic permeability and σ describes the conductivity of the material. To fully attenuate the penetrating signal of electromagnetic interference, the thickness of the shielding material should be ten times the value given by equation (3.12). For 50 Hz radiation it would require a metal plate of over 50 mm to fully attenuate the electromagnetic interference. Surely this is not a feasible solution to block the hum. [9, p. 431]

A delicate solution for the hum problem is depicted in Figure 3.21, which presents the idea of a humbucker pickup. In principle, the humbucking pickup consists of two single-coil pickups connected together in series.

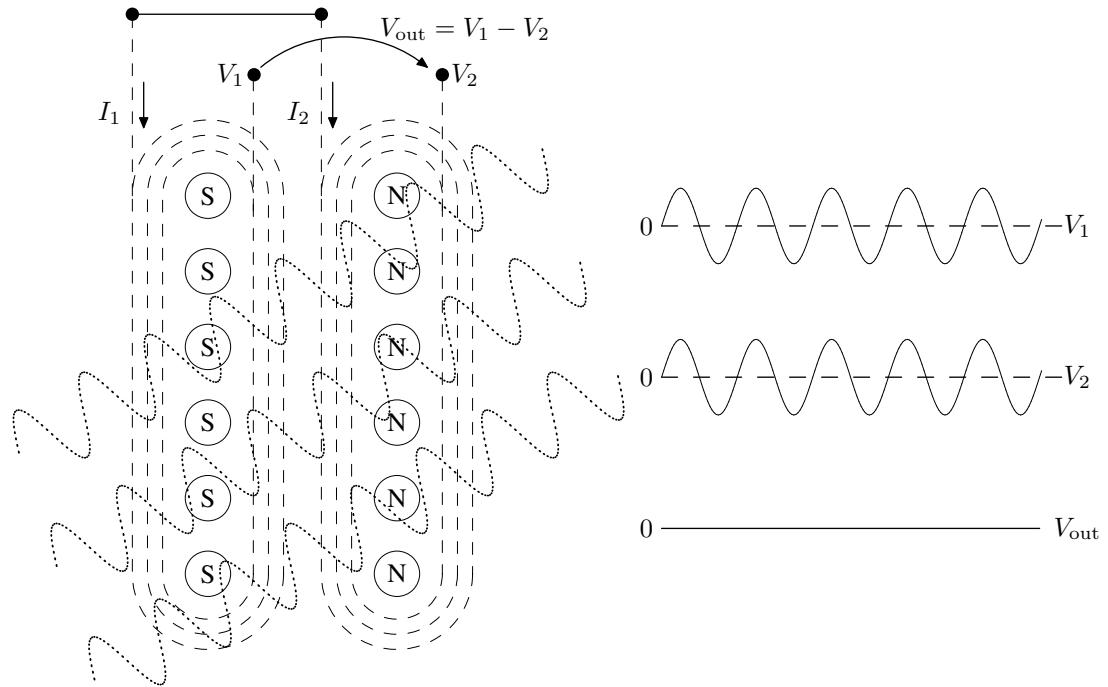


Figure 3.21: Induced voltage due to electromagnetic interference

When the electromagnetic radiation hits the coils of a humbucker pickup, a similar (but not exactly the same) interference signal is induced to both coils. Then it is only a matter of choice how to connect the coils so that the voltage measured at the output would result in noise cancellation. Figure 3.21 has the idea that the output voltage is taken from the same terminal of both coils so that the potential difference would always fluctuate in phase. This would mean that the ground of the circuitry would also fluctuate, which makes this a bit obscure solution. But in

any case, this is the ideology behind cancelling the hum by connecting the coils in series in a manner as described in Figure 3.21.

What about the string vibration then? Would it not also be cancelled this way? No, because the direction of the magnet polarities are different in both pickups. This way the induced electromotive force is pushing the current to the opposite directions in the coils, and the series-connected coils make sure that these oppositely induced electromotive forces add up to increase the output level of the string vibration. This is visualised in Figure 3.22.

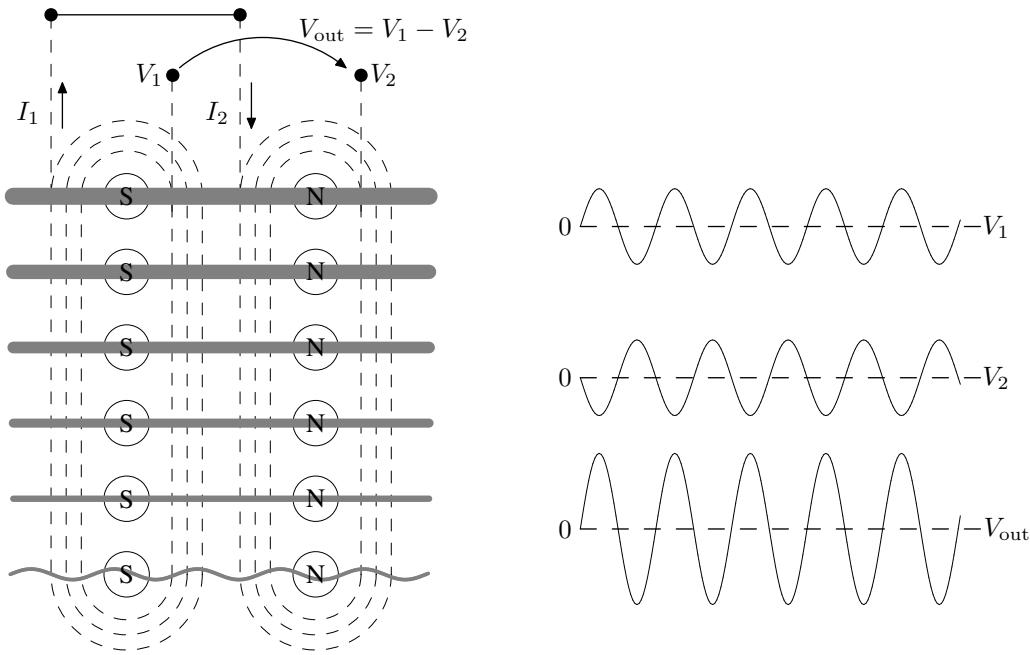


Figure 3.22: Induced e.m.f. in the humbucker pickup due to string vibration

Several rewiring modifications are often suggested for modifying the tone of the humbucking pickups [42, pp. 79–82]. Unfortunately most of them, possibly even all of them, break the humbucking property of the pickup. This loss of humbuck is due to the fact that the coils are reconnected in a way that the common-mode signals do not cancel themselves out as a voltage difference at the output terminals of the pickup.

3.2.7 A magnetic pickup as part of an electric circuit

As an electronic component, the magnetic pickup acts as a normal magnetic core inductor. The main parameters that define the properties of the pickup are directly related to the equivalent circuit [43] [44] of an inductor shown in Figure 3.24a. It is clear that the resistance of the coil wire is in series with the

inductance, but it is not that evident that there also appears a capacitor in parallel with the inductance and the resistance.

The coil winding can be thought of as a parallel wire transmission line where the nearby wires interact capacitively with each other. The capacitance builds up all the way along the parallel turns of the winding and it is therefore a sum of capacitances of individual loops of the inductor. There also exists winding-to-core capacitance for conductive core inductors and layer-to-layer capacitance for multilayer inductors [43]. Figure 3.23 indicates the turn-to-turn and turn-to-core capacitances between individual turns of the coil windings and to the conductive core of the inductor. These types of capacitances are commonly referred to as *stray capacitances*.

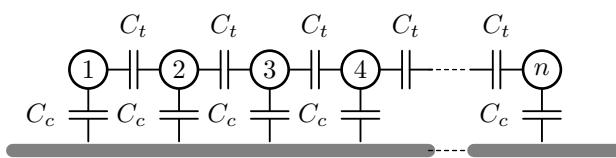


Figure 3.23: Stray capacitances between inductor windings

It is quite another question how the stray capacitance in the inductor windings is modelled mathematically [43] [44]. Needless to say that the theoretical formulae to calculate the total capacitance of a certain type of inductors from a geometrical layout is out of the scope of this book. It is still possible to create a simple model for a pickup capacitance without the need to evaluate the different types of capacitances separately.

The relevant piece of information is simply that the total capacitance of the inductor is actually connected in parallel with the coil and it can be 'measured' between the opposite ends of the inductor. Therefore, the total impedance of the pickup is a combination of R_L , L and C_L of the pickup, as depicted in Figure 3.24a. The subscripts in this particular case are there to remind about the fact that the resistance and the capacitance are internal quantities of the inductor.

The RLC equivalent circuit of a magnetic pickup behaves like a second order filter. A simple transfer function for a general magnetic pickup can be derived from Figure 3.24b, where an AC source has been added to model the electromotive force induced to the inductor as a result of electromagnetic induction due to changing magnetic field. Equating the current loop in Figure 3.24b leads to the

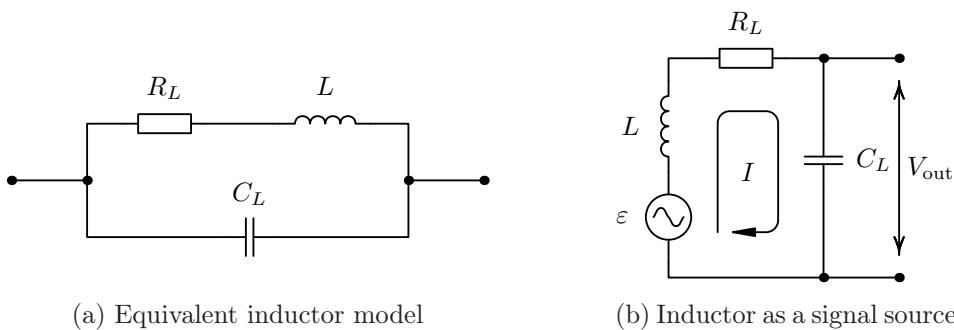


Figure 3.24: Practical inductor models

expression

$$\left[sL + R_L + \frac{1}{sC_L} \right] I(s) = \varepsilon(s).$$

The output voltage V_{out} of the pickup equals the voltage over the capacitor $V_c(s)$ so that the transfer function of the magnetic pickup can be written as

$$\frac{V_c(s)}{\varepsilon(s)} = \frac{V_{\text{out}}}{\varepsilon(s)} = I(s) \frac{1}{sC_L} = \frac{1}{LC_L \left(s^2 + s \frac{R_L}{L} + \frac{1}{LC_L} \right)}. \quad (3.13)$$

This transfer function can be used to evaluate the frequency response of a typical magnetic core inductor as a guitar pickup is.

A median value for the resistance of a typical magnetic pickup is somewhere between $7 \text{ k}\Omega$ - $10 \text{ k}\Omega$, and typical inductance values are in the range of a few henries [45, pp. 91–94]. This means that the inductance of a magnetic pickup is quite large. According to Lemme [46], the capacitance of the coil of a guitar pickup commonly varies between 80 pF to 240 pF, but in some cases it can also exceed this range a little, depending on the geometry of the whole pickup construction. For averaged values of $L = 5 \text{ H}$, $R_L = 10 \text{ k}\Omega$ and varying C_L of 80 pF to 320 pF, the transfer characteristics of a magnetic pickup are as shown in Figure 3.25.

Magnetic pickups are advertised to have a clear resonant peak in their transfer characteristics. This is true according to Figure 3.25, but a resonance peak can be obtained with basically any normal inductor. Still, the resonant peak of the guitar pickup is claimed to have something magical in it. It does not. It is just like any other inductor. Somehow the ear is quite sensitive to the changes in the sound caused by this resonance peak of the magnetic pickup and it seems to appeal to the hearing sensation in a positive manner.

A typical transfer function of the magnetic pickup yields a resonance frequency

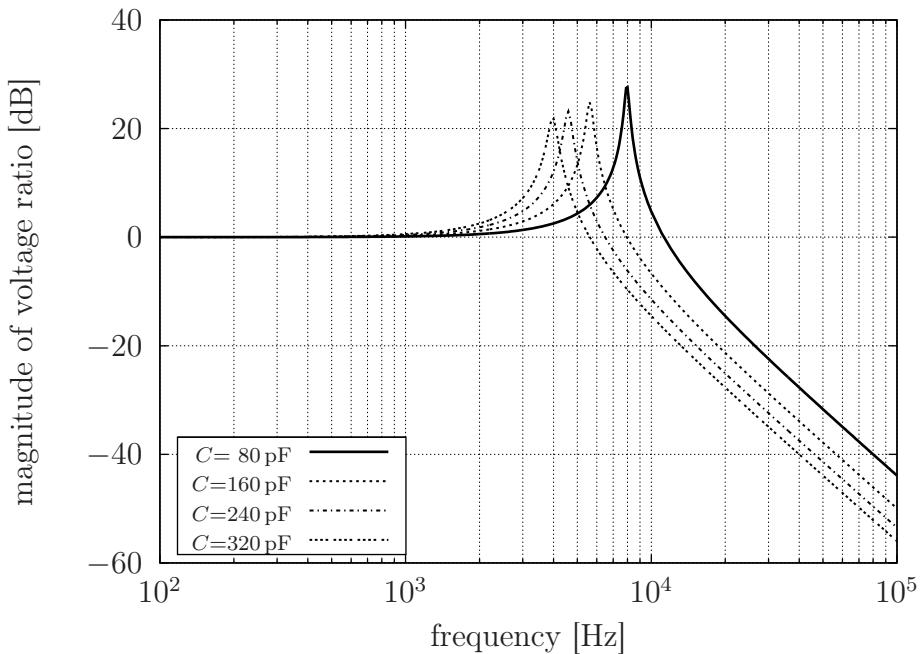


Figure 3.25: Frequency response curves of a magnetic pickup with varying C_L

of a few kilohertz. The resonance frequency f_r for a regular RLC circuit can be calculated with the analytical formula

$$f_r = \frac{1}{2\pi} \frac{1}{\sqrt{LC_L}},$$

meaning that the resistance does not affect the resonance frequency at all. The most surprising thing in the simulation results was the gain of about 20 dB at the resonance frequency. Passive circuits are not famous for signal amplification.

The location and shape of the resonance peak is determined from the equivalent model of the inductor as the resonance frequency of the basic RLC circuit. The effect on the resonance peak from changes in each variable, namely R_L , L and C_L , can be summarised with the following bullets:

- the peak moves towards higher frequency as inductance L increases
- the peak moves towards lower frequency as capacitance C_L increases
- the peak becomes smaller and wider as resistance R_L increases

Obviously the decrease in each value has a reverse effect on the resonance peak. A delicate tone modification can be implemented by connecting an additional capacitor in parallel with the pickup. In principle the capacitance of this additional capacitor will directly sum up with the internal capacitance of the pickup,

thereby lowering the resonance frequency by some amount. The resistance and the inductance go hand in hand, since the resistance is effectively the resistance of the inductor wire. Increasing the number of turns increases both inductance and resistance. The only way to alter the inductance without touching the resistance is to change the material or size of the permanent core magnet.

If both the resistance and inductance of the pickup are known, the transfer function of a magnetic pickup can be easily simulated with a certain degree of accuracy using the transfer function (3.13). The most critical effect comes from the capacitance, and unfortunately the exact value of that capacitance is quite difficult to determine. Usually manufacturers document only the resistance value quite accurately, but the value of the inductance is unknown and the capacitance is never even mentioned anywhere.

The output impedance of a magnetic pickup can be evaluated from the equivalent circuit in Figure 3.24b. The effective impedance of a pickup is measured as seen from the output terminals. The impedance is evaluated as a parallel connection of the capacitor with the inductor and resistor so that

$$Z = \frac{\frac{1}{sC_L}(R_L + sL)}{\frac{1}{sC_L} + R_L + sL},$$

which simplifies to

$$Z = \frac{R_L + sL}{LC_L \left(s^2 + s\frac{R_L}{L} + \frac{1}{LC_L} \right)} = \frac{V_{\text{out}}}{\varepsilon(s)}(R_L + sL). \quad (3.14)$$

After identifying the resonance frequency term $\frac{1}{LC_L} = \omega_0^2$ and making the substitution $s = j\omega$, as we are naturally dealing with sinusoidal signals, the expression for the impedance of the pickup is

$$Z = \frac{\omega_0^2(R_L + j\omega L)}{(\omega_0^2 - \omega^2) + j\omega \frac{R_L}{L}}. \quad (3.15)$$

Since the impedance Z defined via equation (3.15) is a complex number, it is not possible to directly analyse the form of the impedance as a function of frequency. The absolute value (or magnitude) of a complex number, on the other hand, is a real-valued quantity. In the case of equation (3.14), one can apply the general rule of calculating the absolute value of a ratio of complex numbers z_1 and z_2 as

$$\left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|},$$

meaning that the absolute values of the numerator and denominator can be evaluated separately. A similar simplification rule exists for the phase (or argument) of a ratio of complex numbers

$$\arg \frac{z_1}{z_2} = \arg z_1 - \arg z_2,$$

so that the phases of the numerator and denominator can be evaluated separately and combined together with a minus sign.

The magnitude of the pickup's impedance and the corresponding phase shift are drawn as a function of frequency in Figure 3.26. Realistic averaged values obtained from the pickup specifications were used to generate the magnitude and phase plots so that the obtained impedance values are realistic. The impedance

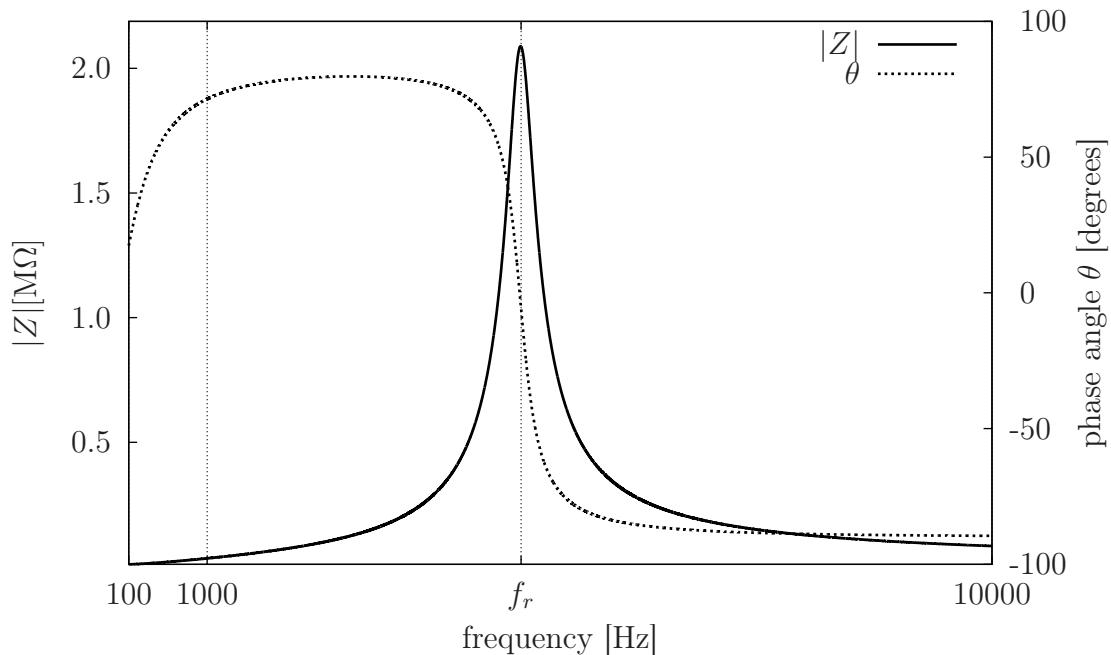


Figure 3.26: Impedance curves of a magnetic pickup

of the pickup will reach a maximum value at the resonance frequency, where the reactances of the inductor and the capacitor will have equal values. This is also seen from the curve describing the phase shift, which hits zero at the resonance frequency. The impedance calculation reveals that different frequencies will experience different impedance at the pickup, and also if the pickup is connected to an external circuit (normally the tone and volume control circuit), it cannot be modelled as a constant output resistance for all frequencies.

The two megohm impedance at the resonance frequency appears so huge... How does it come about? At low frequencies below 500 Hz, the impedance approx-

imately equals the value of the coil resistance because the impedance of the inductor is low compared to the resistance, and the impedance of the capacitor is large. So in the low frequencies it is just the capacitor impedance parallel to the relatively small resistance, which in total equals the resistance due to the coil winding.

At the resonance frequency, the impedance of the inductor and the capacitor are large and also equal, so that the effective impedance comes from the parallel connection of the inductor and the capacitor. This has a relatively huge value, as is seen from Figure 3.26.

For modelling the humbucker pickups, it is sufficient to use an equivalent circuit of Figure 3.27, where two single-coil circuits are connected in series. The output

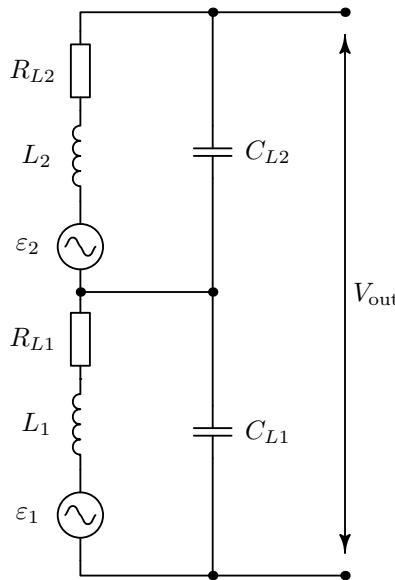


Figure 3.27: An equivalent circuit of a humbucker pickup

impedance calculated from this equivalent circuit becomes $Z_1 + Z_2$, where the expression for both impedances is obtained from equation (3.15).

The changes in the transfer function are not so evident since the two pickup sections are interacting with each other. However, evaluation of the mesh matrix

$$\begin{bmatrix} sL_1 + R_{L1} + \frac{1}{sC_{L1}} & 0 \\ 0 & sL_2 + R_{L2} + \frac{1}{sC_{L2}} \end{bmatrix} \times \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

proves that the transfer function of a humbucker pickup is also a sum of two

individual transfer functions of single-coil pickups shown in equation (3.13).

3.2.8 Measurements on a magnetic pickup

Unfortunately, without special and expensive equipment it is not possible to measure any magnetic properties such as the strength of the magnetic field of pickup magnets. The measurements are therefore restricted to observing only the output waveforms of a magnetic pickup as the guitar string is vibrating on top of it. The measurements were done using a basic digital storage oscilloscope, which is directly connected via a $10\text{ M}\Omega$ probe to the coil wire of the pickup. The single-shot functionality of the oscilloscope was used to isolate the first few periods of vibration, from where the initial spectrum can be evaluated.

In these measurements the focus is not to emphasise the actual spectrum of the string but only to concentrate on the spectrum obtained from a magnetic pickup located at a certain position along the string. Therefore, the weighting property of the pickup location is not divided away when drawing the amplitude spectrum. If error limits need to be evaluated, the theoretical expectation values are calculated with the certain weight factor to come in terms with the correct expectation values. This is done to give an idea of the real signal components obtained from the guitar pickup.

The pickup used in the measurements was a cheap replacement pickup, which was bought as a spare part from a nearby electronics store. The pickup seemed to be a single-coil TeleCaster clone intended to be used as the bridge pickup. Figure 3.3 gives a good idea of how the pickup looks like on the outside. The specifications for the pickup claim a resistance of $7.1\text{ k}\Omega$ and an inductance of 5.9 H . The resonance frequency of the pickup was informed to be 3 kHz , but this value has been obtained using a 680 pF load, which is intended to simulate the effect of a guitar cable.

In the measurement session, the pickup was mounted at about 50 mm from the left clamping support of the same measurement platform (Fig. 2.23), which was used in the vibrating string measurements. The guitar string was adjusted to vibrate about $1 - 2$ millimetres above the centreline of a single polepiece of the pickup shown in Figure 3.3. The effective length of the string was 600 mm so that the pickup created a node at $\frac{L}{12}$.

The first set of measurements was conducted to investigate the induced initial waveforms from horizontal and vertical plucks and to compare the measured

results to the theoretical expectations. A sharp plastic pick was used to pluck the string at different locations and the amplitude was kept relatively modest, but not especially restricting the amplitude to certain limits. Figure 3.28 depicts the measurement results of the first few periods obtained from a pluck at $\frac{L}{6}$.

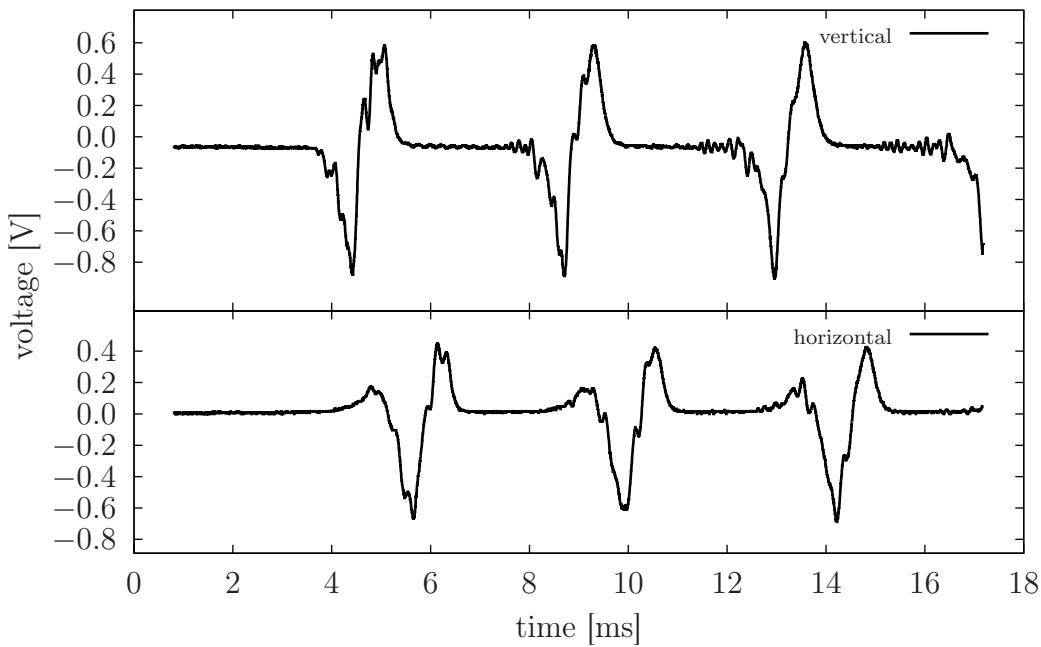


Figure 3.28: Measured waveforms from vertical and horizontal plucks at $\frac{L}{6}$

The ideal waveform evaluated as the time derivative of equation (2.32) would have one negative and one positive square pulse at the locations where the measured waveforms have the rounded spikes. The waveform from the vertical pluck seems more closer to the theoretical predictions because the horizontal pluck gave an extra 'hill' just before the negative spike.

Eventually the 'hill' was identified as a consequence of placing the string close to the centreline of the pickup magnet pole. As the string crosses the centreline, it starts to reverse the waveform that was obtained when approaching the centreline. Obviously the string was not exactly in the middle of the pickup pole, so this is why the 'hill' is smaller than the preceding rounded spike.

The measured waveforms shown in Figure 3.28 can be reproduced quite nicely using the simulation model of Horton [40]. Only 20 upper partials from the ideal spectrum of a string were used in the simulations to get that nice roundness to the waveform. The simulation results surprisingly validate the usefulness of the theoretical model based on the idea of magnetic point charges. The results

also verify the mechanism leading to the extra 'hill' observed in the horizontal vibration waveform.

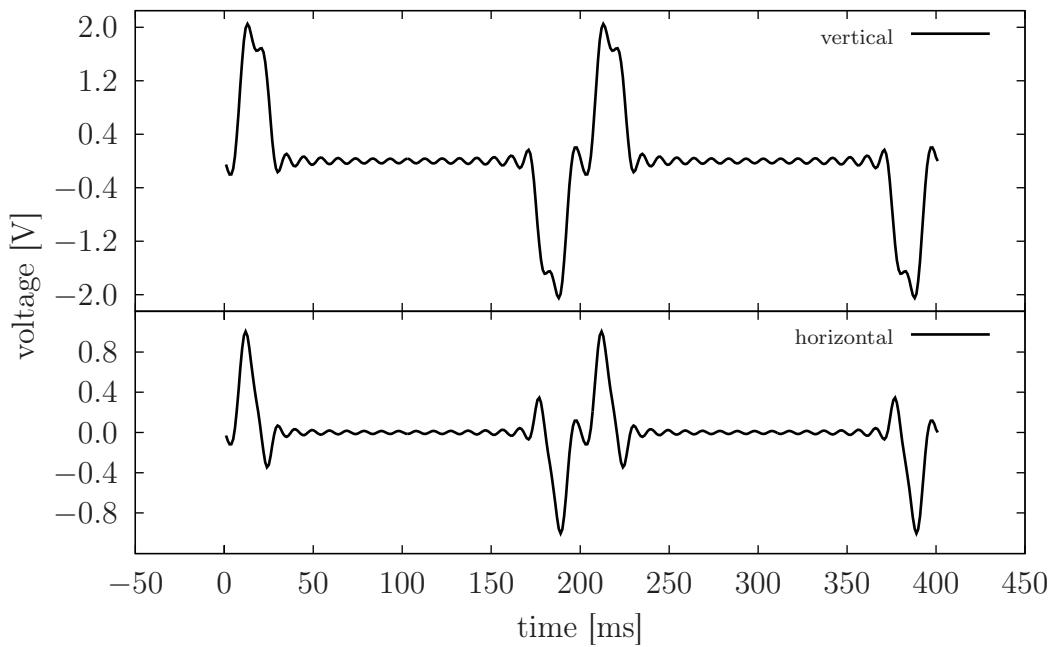


Figure 3.29: Simulated waveforms from vertical and horizontal plucks at $\frac{L}{6}$

The simulations also indicate the clearly stronger signal obtained from the vertical motion of the string. This effect was not seen in the measurement results of the real magnetic pickup. Anyhow, the plucking displacements were not controlled during the measurements, so there is at least one source of uncertainty which prevents official verification of the stronger signal observed in vertical motion.

The upper partial structures calculated from the measured waveforms of Figure 3.28 are shown in Figure 3.30. The spectrum analysis indicates that the nonlinearities of the magnetic pickup can noticeably affect the observed magnitudes of the upper partials. Although the waveform of the vertical pluck seems relatively theoretical, the spectrum reveals that the nonlinearities of the pickup cause the node at the plucking position (6th upper partial) to gain energy and not indicate the node properly.

However, despite the extra 'hill' in the horizontal waveform, the node caused by the plucking position is reproduced correctly in the spectrum. The node at the 12th upper partial due to the placement of the pickup and the plucking position is not that clearly visible in either spectrum. These spectrums are not explicitly

compared to the theoretical values at this point because the focus is to compare the differences between the vertical and horizontal plucks.

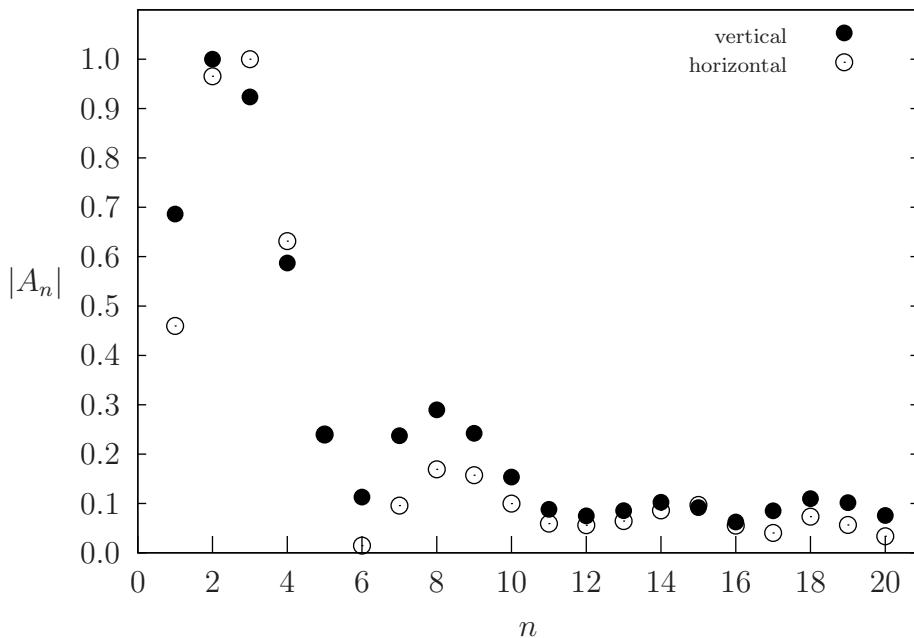


Figure 3.30: The calculated spectrums for vertical and horizontal plucks at $\frac{L}{6}$

The waveforms for the vertical and horizontal vibration directions were also measured when the string was plucked at the middle of the string. Figure 3.31 shows that the vertically directed pluck results in a relatively clean and symmetrical waveform. The horizontal waveform, on the other hand, suffers even more from the centreline problem. In this case, the second upper partial almost becomes the fundamental frequency so that the sound is theoretically one octave higher because of the centreline crossing effect.

The calculated spectrums from the midpoint plucks reveal the expected fact that the nonlinearities observed in Figure 3.31 lead to a spectrum where every second upper partial does not have minimum amplitudes as the theory predicts. Due to the frequency doubling effect in the horizontal waveform, the upper partials below the 10th harmonic are clearly emphasised compared to the expected spectrum.

To produce a decent spectrum plot which would be comparable to the theoretical values, the initial deflection of the string prior to plucking has to be directed to the horizontal direction, and the crossing of the polepiece centreline during vibration should be avoided. The plucking amplitude also needs to be smaller

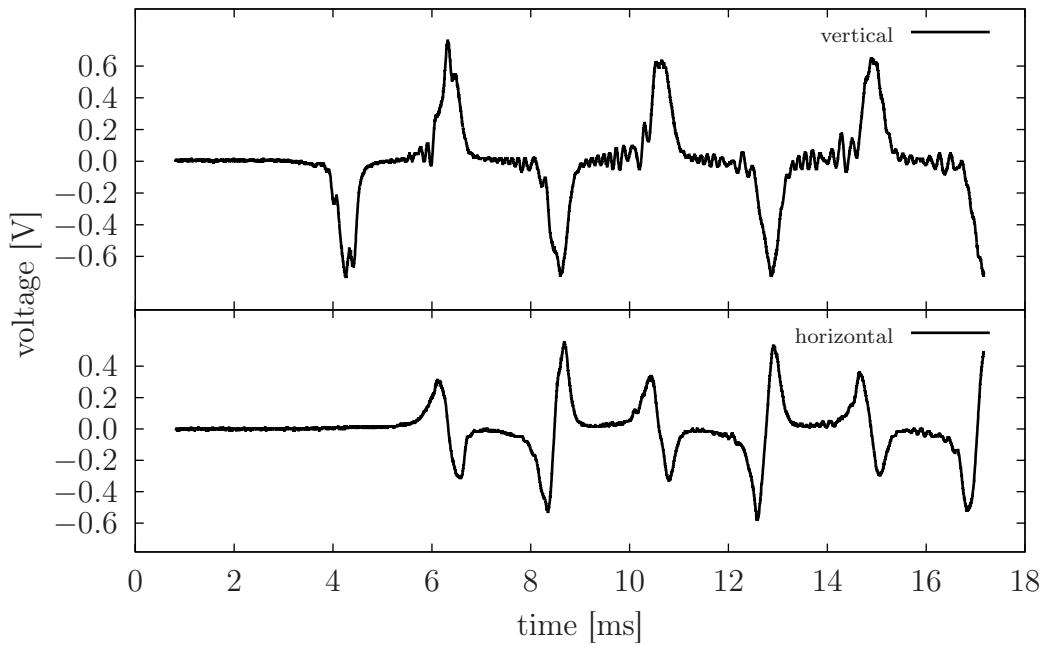


Figure 3.31: The measured waveforms from vertical and horizontal plucks at $\frac{L}{2}$

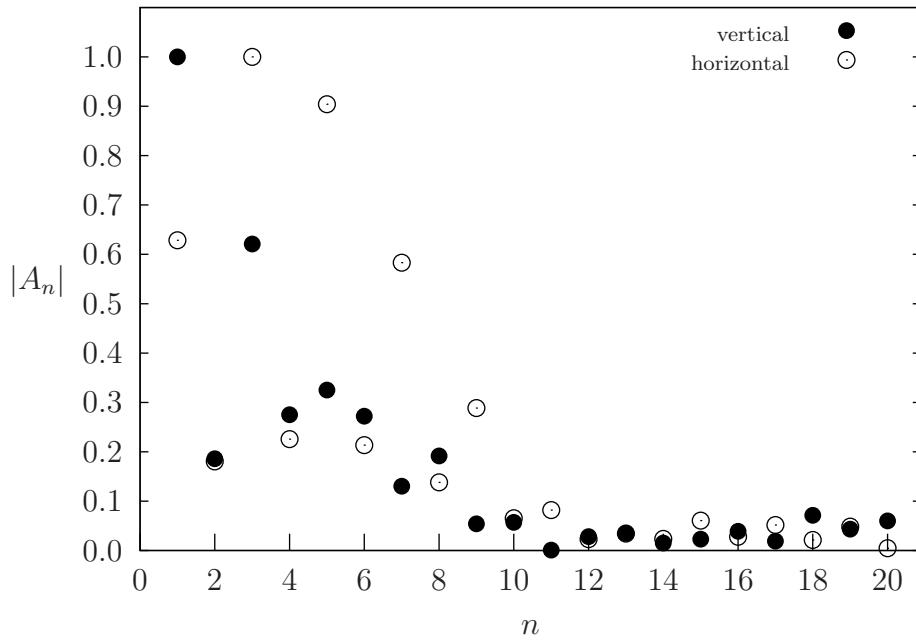


Figure 3.32: The calculated spectrums for vertical and horizontal plucks at $\frac{L}{2}$

than in the previous experiments to reduce unnecessary distortion due to nonlinearities of the pickup.

After several trials, one successful waveform was recorded. Figure 3.33 shows

the measured waveform of a pluck executed at a distance $\frac{L}{6}$ from the right clamping support. The extra hill is almost completely missing and the upward and downward spikes seem to be relatively symmetrical in shape.

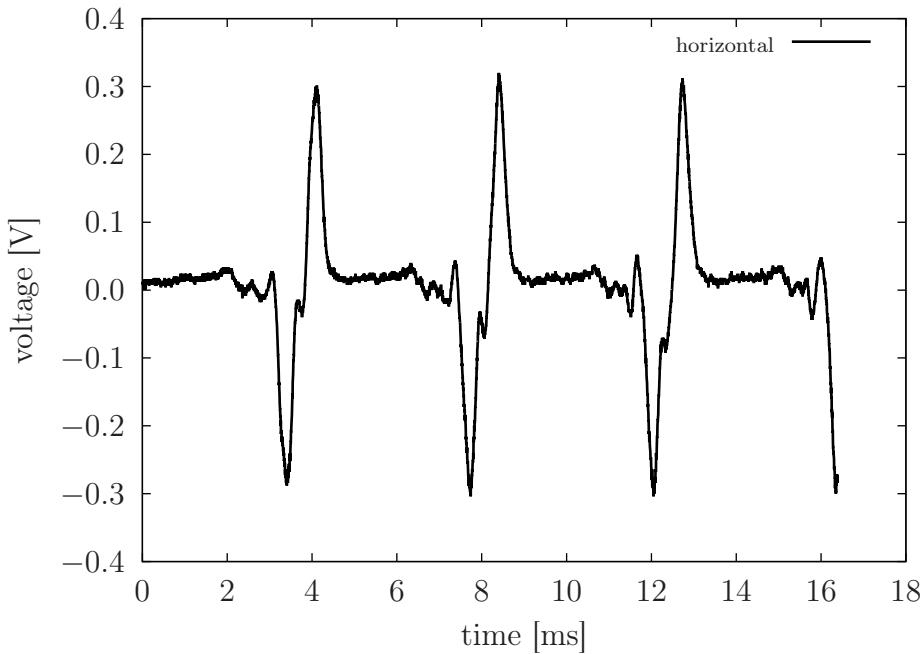


Figure 3.33: The waveform resulting from a 'successful' pluck at $\frac{L}{6}$

The corresponding spectrum with theoretical error limits looks similar to the one drawn in Figure 3.34. The error limits have been defined in the same way as for the measurements for the amplitude spectrums of string vibration. The example script can be seen from Appendix C. The part where the effect of the pickup location had been divided away has now been removed for this error analysis. The upper and lower limits have been iterated numerically as the maximum values that can be obtained with the given uncertainties in the measurements.

From the resulting spectrum, it is observed that the theoretical spectrum coefficients obtained as a time differential of equation (2.32) gave a good forecast to the actual measured spectrum. Therefore, the graphical analysis derived from the initial triangular shape of the string extends as a relatively good model for the magnetic pickup as well, as long as the induced e.m.f. in the pickup stays within a linear region. The measurement results also indicate that the model obtained from the magnetic point charge might exaggerate the nonlinearities in the magnetic transducer. However, if the plucking amplitude is kept very small and the centreline crossing avoided in the simulations, the magnetic point charge model is eventually in agreement with the triangular shape derivation.

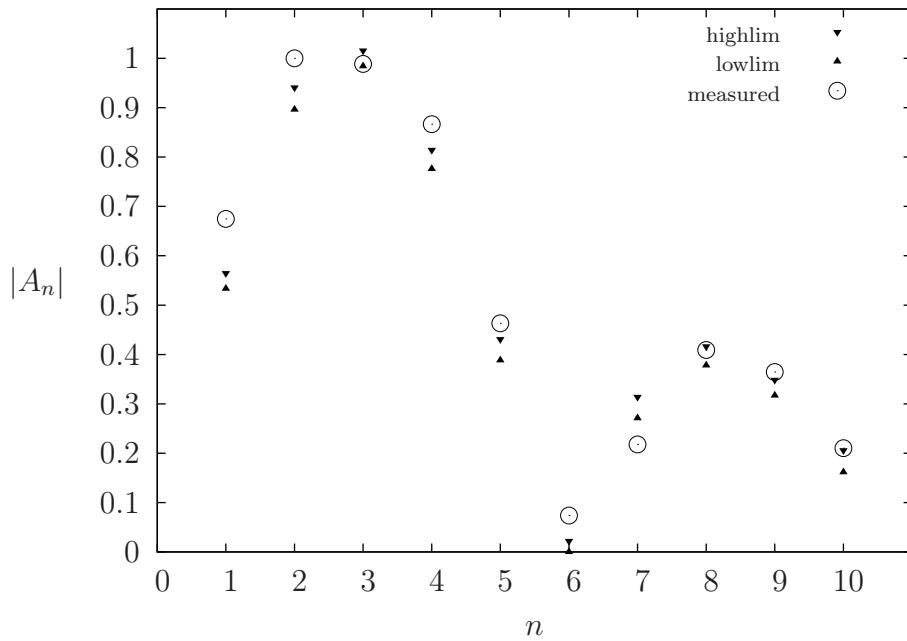


Figure 3.34: The harmonic spectrum resulting from a pluck at $\frac{L}{6}$

The resistance, inductance and capacitance which define the essential electrical characteristics of a magnetic pickup can be determined relatively easily. The resistance of a pickup can be measured with a regular ohmmeter. The inductance and capacitance can be measured indirectly with the help of the measurement setup shown in Figure 3.35, where the pickup is connected in series with a voltage source and a shunt resistor R_S . The voltage source v_g should be a function generator or equivalent and the device to measure v_{RS} and other potential differences in the circuit can be a regular multimeter or an oscilloscope.

When determining the inductance, the alternating voltage source v_g should produce relatively low-frequency (100 – 500 Hz) oscillations. At such low frequencies the capacitive reactance of the pickup is so large that it can be approximated to be disconnected from the equivalent model of the pickup. When the capacitance is neglected, the pickup can be modelled as an inductor and its internal resistance connected in series. If a multimeter is used for the measurements, the voltage equation for the circuit in Figure 3.35 is written using RMS values,

$$V_g = V_{\text{pickup}} + V_{RS}, \quad (3.16)$$

where all values are assumed to be measured values from the circuit. The voltage components in equation (3.16) in terms with current I_S are

$$V_{RS} = I_S R_S \quad \text{and} \quad V_{\text{pickup}} = I_S \sqrt{R_L^2 + \omega L^2}.$$

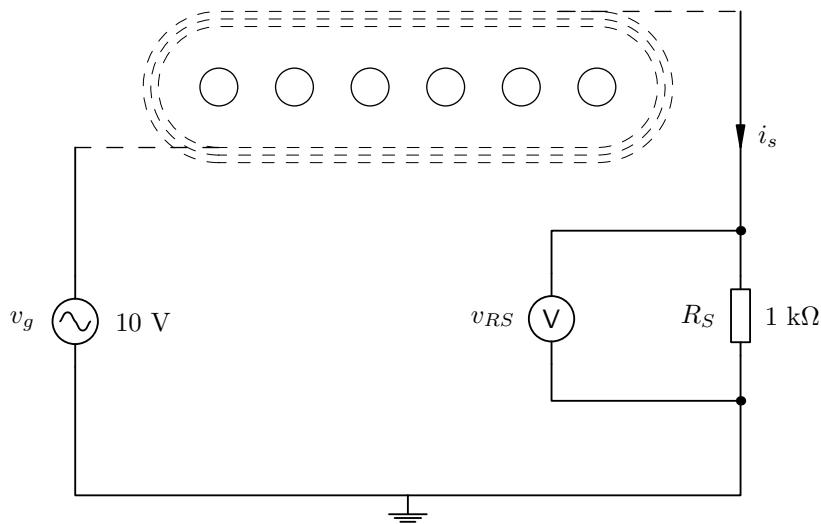


Figure 3.35: A rudimentary measurement arrangement to extract values for the inductance and resonance frequency of a magnetic pickup

When the common current I_S is eliminated by dividing the other voltage expression with the other, the inductance L can be solved from the resulting equation as

$$L = \frac{1}{2\pi f} \sqrt{\left(\frac{V_{\text{pickup}}}{V_{RS}} R_S \right)^2 - R_L^2}.$$

This method requires two voltages to be measured from the circuit in 3.35: the voltage over the shunt resistor V_{RS} and the voltage over the pickup V_{pickup} .

The pickup capacitance can be calculated when the inductance and the resonance frequency of the pickup are known. To measure the resonance frequency, one simply needs to adjust the frequency of the voltage source v_g and simultaneously measure the voltage v_{RS} with a multimeter or an oscilloscope. The frequency where the measured voltage has a minimum value is the resonance frequency of the pickup. After the inductance L and resonance frequency f_r have been determined, the capacitance can be calculated from the equation

$$C = \frac{1}{4\pi^2 f_r^2 L}$$

following from the definition of LC circuit resonance. [47]

3.3 OPTICAL PICKUPS

Optical pickups make use of a coupled system consisting of an infrared LED and a photodetector between which the guitar string is placed. The first Light Emitting Diode, a.k.a. LED, was developed in 1962. This electro-optical device was

based on GaAsP (Gallium-Arsenide-Phosphide) layers, which emitted red light. GaAsP is a semiconductor material, an alloy of gallium arsenide and gallium phosphide. Nowadays, commercial LEDs are available for 247-1550 nm wavelengths, a range which covers approximately all the possible colours visible to the human eye.

A *photodetector* is a device that senses emitted light and transduces it to electric current. When a LED is coupled with a photodetector, a compact measuring device or an optomechanical switch is obtained. The most common photodetector device is a *photodiode*, although light dependent resistors (LDR) and other type of LEDs are also used as detectors of emitted light. Photodiodes are popular because of their rapid response time and wide linear range. A *phototransistor* is created by combining a photodiode with an amplifying element. [48] [49]

The first optical guitar pickups were introduced already back in 1969, and several more advanced models have been developed ever since. Optical pickups have the advantage of having a compact, low power and low cost structure, which employs commonly used techniques to build optical switches as discrete components. Optical guitar pickups were primarily developed to tackle hum problems and eliminate the string-attractive force of magnetic pickups. Optical pickups are also more linear transducers compared to magnetic pickups. These properties make the optical pickup quite superior over the traditional magnetic pickup ... but still the optical pickups are very seldom used as guitar pickups. The reason could not be the sound of the pickup, or could it?

3.3.1 An optical transducer as a guitar pickup

Reusing existing commercially available components for applications that they are not initially intended for is efficient and environmentally friendly. This is the case with optical guitar pickups, which in the most simplest cases can be implemented using readily available photointerrupter or photoreflector modules. The internal structure of these standard components is depicted in Figure 3.36.

The photointerrupter consists of an infrared light emitting diode (IR LED) and a phototransistor. The light rays generated by the LED are focused directly to the detecting surface of the phototransistor, making the transistor fully saturated with the current generated by the incident light. The switching operation is achieved by inserting an opaque object to block the line of sight between the LED and the phototransistor. The current flow in the base of the phototransistor

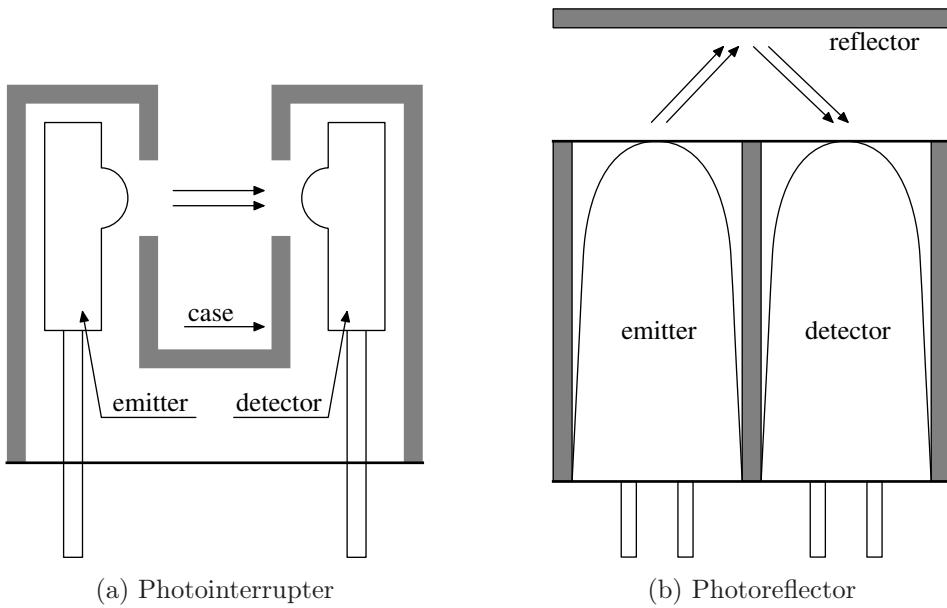


Figure 3.36: Commercially available optoelectric components

can be controlled by casting a partial shadow over the phototransistor's detector area. The dependency between the shadow ratio and the amount of base current is almost linear as is shown later.

A practical application of using a photointerrupter component as a guitar pickup is shown in Figure 3.37. In a static configuration, the guitar string is placed be-



Figure 3.37: An optical interrupter switch used as a guitar pickup

tween the LED and the phototransistor so that approximately half of the emitted light is blocked from reaching the detector area. When the string is vibrating on top of the photodetector, the shadow ratio is modulated by the motion of the string. The geometrical details of this implementation are shown in Figure 3.38.

If the photodetector is very small, the string displacement will reach very close to the edges of the detector area. In these cases the performance of the interrupter

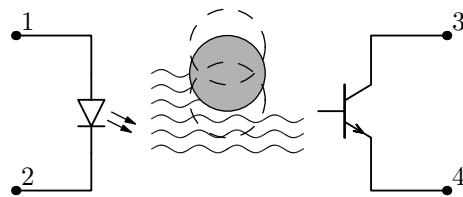


Figure 3.38: The ideology of the optical interrupter switch pickup

technique can be enhanced by using two LEDs in parallel to even out the intensity of light observed at the edges of the detector area [50].

One reason for interrupter-style pickups not being favoured by guitarists is that when placing a light/sensor pair on opposite sides of the string, it will disturb important playing techniques such as palm-damping near the bridge of the guitar.

The use of a photoreflector module as a sensor of string vibration has also been studied to some extent [51]. In the experiment described in article [51], the reflections were taken directly from the surface of the moving string, not employing the use of the shadow casting technique as in the case of the photointerrupter module.

Efficient implementation using the photoreflector as a guitar pickup requires quite a lot of advanced theoretical studies because the light is reflected in different angles from the cylindrical surface of the string even when the string is not moving. When the motion of the string is added, the reflection pattern becomes quite chaotic, and it is not that clear whether the amount of reflected light actually correlates linearly to the motion of the string. One visualisation of the actual reflective pickup setup is given in Figure 3.39, and it is trying to emphasise the scattering of the reflected light rays from the surface of the string.

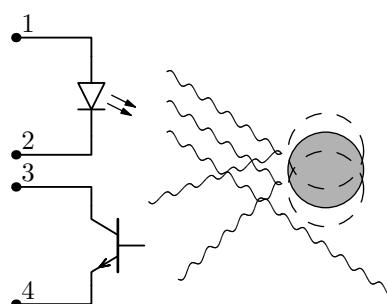


Figure 3.39: The ideology of the optical reflector pickup

The reflecting method can also make use of the shadowing technique. When used

in this way, a mirror or an equivalent highly reflective surface is used to create a perfect reflection and the reflected light is then obstructed by the moving string. Many projects have tried to make use of the reflective technique but eventually failed to get a decent signal out from the system. A simpler approach would be to use solutions based on the interrupter technique.

3.3.2 An optical pickup as a signal source

A thing to consider in the emitted light rays of the LED is the direction dependency of the light's intensity. Depending on the geometry of the LED window surface, the light does not spread evenly in all directions. The geometry of the windowed LED surface at the air interface boundary determines the shape of the radiation pattern of the emitted light. If the surface of the LED window is of a planar type, the LED radiates light in a pattern called the *lambertian emission pattern*. Other common LED surface types are the hemispherical and the parabolic surface. Figure 3.40 presents the surface geometries of the planar and hemispherical LED windows.

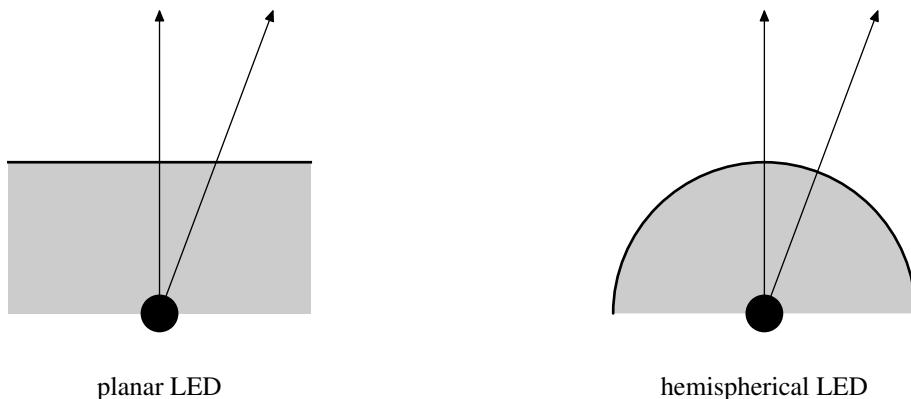


Figure 3.40: The surface geometries of planar and hemispherical LEDs

The lambertian pattern obtained from the planar LED is analysed here as an example because it is presumably the most common emission pattern. Analysis of the emission pattern is needed to verify that the LED emits light sufficiently evenly in the required angular directions so that in all use cases of interrupting and reflecting the light, the LED will not be a source of nonlinearity from the viewpoint of the photodetector device.

A formula that describes the lambertian pattern can be derived by using Figures 3.41 and 3.42 as references.

As shown in Figure 3.41, the light emerging from the light source is refracted at

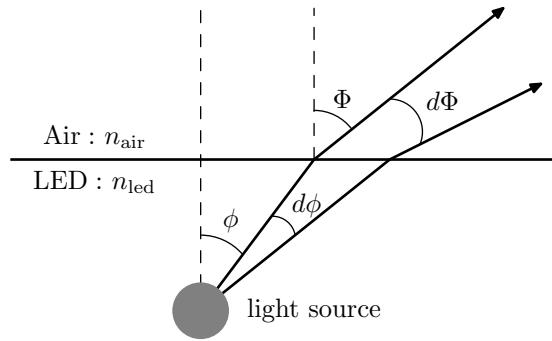


Figure 3.41: Refraction at the LED-air interface of a planar LED

the LED-air boundary. Assuming that Snell's law of refraction is obeyed,

$$n_{\text{led}} \phi = n_{\text{air}} \sin \Phi, \quad (3.17)$$

and taking a differential with respect to the angles yields

$$n_{\text{led}} d\phi = n_{\text{air}} \cos \Phi d\Phi \quad \Rightarrow \quad d\Phi = \frac{n_{\text{led}}}{n_{\text{air}}} \frac{1}{\cos \Phi} d\phi. \quad (3.18)$$

An application of the energy conservation law requires that for the light intensities in air and in the LED, the relation

$$I_{\text{led}} dA_{\text{led}} = I_{\text{air}} dA_{\text{air}}$$

is valid. This states that the intensity of light passing through an infinitesimal area element at the surface of the LED is directly proportional to the intensity through some area element of air. Both of the area elements are assumed to be part of a spherical surface that depicts the surroundings viewed from a point-like source. Figure 3.42 illustrates the area element related to the surface which the emitted light passes through.

The area element of a general spherical surface is calculated using the parametric representation

$$\vec{r}(\phi, \theta) = R \sin \phi \cos \theta \hat{i} + R \sin \phi \sin \theta \hat{j} + R \cos \phi \hat{k}$$

of a sphere with radius R . The parameter θ spans over the horizontal angles and ϕ goes through the angles in the vertical direction. The vector differentials to angular directions ϕ and θ are

$$\frac{\partial \vec{r}(\phi, \theta)}{\partial \phi} = R \cos \phi \cos \theta \hat{i} + R \cos \phi \sin \theta \hat{j} - R \sin \phi \hat{k}$$

$$\frac{\partial \vec{r}(\phi, \theta)}{\partial \theta} = -R \sin \phi \sin \theta \hat{i} + R \sin \phi \cos \theta \hat{j}.$$

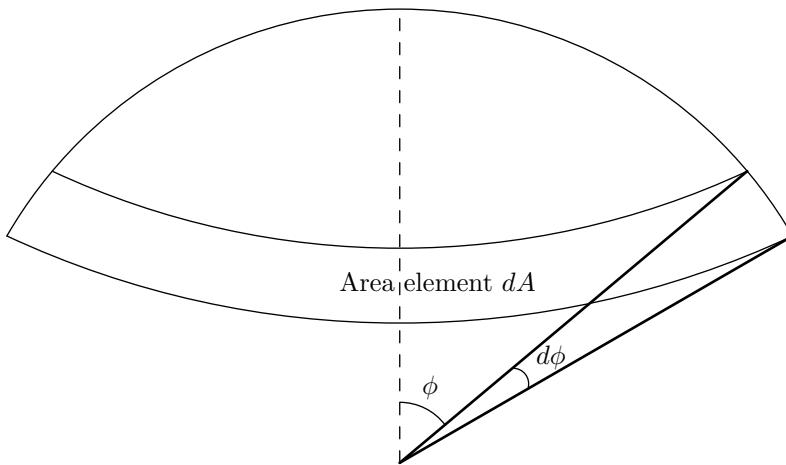


Figure 3.42: An area element of a spherical surface

These are needed to calculate the surface normal vector $\vec{N}(\phi, \theta)$, which is defined as a vector product

$$\begin{aligned}\vec{N}(\phi, \theta) &= \frac{\partial \vec{r}(\phi, \theta)}{\partial \phi} \times \frac{\partial \vec{r}(\phi, \theta)}{\partial \theta} \\ &= R^2 \sin^2 \phi \cos \theta \hat{i} + R^2 \sin^2 \phi \sin \theta \hat{j} + R^2 \cos \phi \sin \phi \hat{k},\end{aligned}$$

from where the surface area element is evaluated as

$$dA = \left| \frac{\partial \vec{r}(\phi, \theta)}{\partial \phi} \times \frac{\partial \vec{r}(\phi, \theta)}{\partial \theta} \right| d\phi d\theta.$$

After writing open the vector product and taking the magnitude, the expression

$$dA = R^2 \sin \phi d\phi d\theta$$

is obtained for the spherical surface area element. Since the whole circumference of the sphere is involved, the area element can be evaluated further by integrating over the horizontal angle to reach a differential element,

$$dA = \int_{\theta=0}^{2\pi} d\theta R^2 \sin \phi d\phi = 2\pi R^2 \sin \phi d\phi,$$

depicted in Figure 3.42. From here the differential surface area element for air becomes

$$dA_{\text{air}} = 2\pi R^2 \sin \Phi d\Phi \quad (3.19)$$

and after substituting equations (3.17) and (3.18) into equation (3.19), the expression

$$dA_{\text{air}} = 2\pi R^2 \frac{n_{\text{led}}^2}{n_{\text{air}}^2 \cos \Phi} \phi d\phi$$

is obtained.

Similarly for the LED, the surface area element is given by

$$dA_{\text{led}} = 2\pi R^2 \sin \phi d\phi \approx 2\pi R^2 \phi d\phi \quad (3.20)$$

under the assumption that the angle ϕ is very small.

The intensity of light from the light source is given by the total source power, which spreads like a spherical wave and attenuates over distance. Therefore, at distance R in the LED, the intensity

$$I_{\text{led}} = \frac{P_{\text{source}}}{4\pi r^2}.$$

Collecting all the previous equations together, it is possible to express the intensity of the emitted light in the surrounding air as

$$I_{\text{air}} = \frac{P_{\text{source}} n_{\text{air}}^2}{4\pi r^2 n_{\text{led}}^2} \cos \Phi. \quad (3.21)$$

[49, pp. 93–95]

Equation (3.21) can be visualised as a radar plot, where the intensity is drawn as a function of radiation angle. Figure 3.43 indicates that the lambertian source

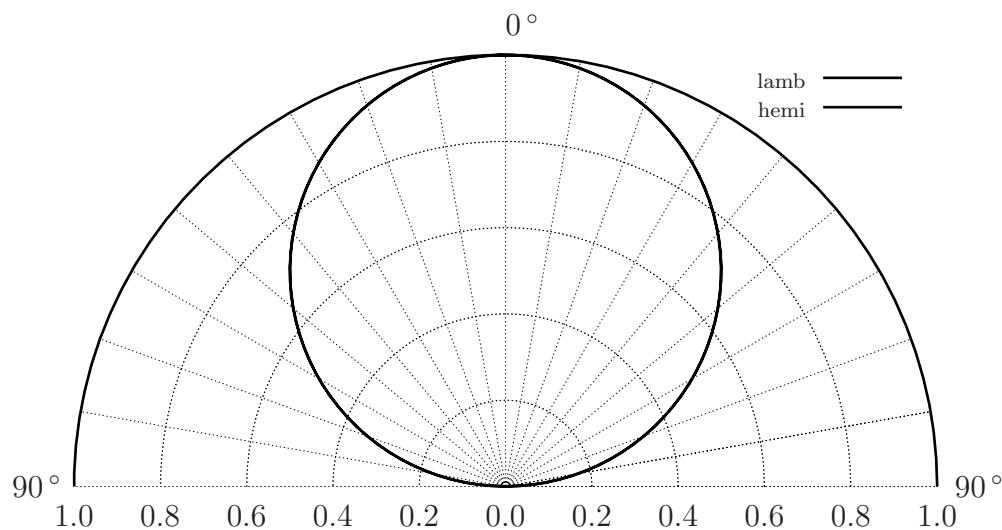


Figure 3.43: The far-field patterns of planar and hemispherical LEDs

gives a reasonably equal intensity in a ± 10 degree angle around the zero angle.

If the LED window is of the hemispherical type, then the intensity spreads evenly to all directions. This should be enough to be able to trust the LED as a linear source of light at least in the photointerrupter device, where the light rays are not required to spread far from the direct line of sight.

The linearity aspect of the photoreceptor side and at the same time of the whole pickup can be approached by practical methods of measurement. Figure 3.44 shows results from an experiment where a small light blocking object was moved gradually over the photoreceptor area of a photointerrupter switch component H21B1. The measurement results indicate a clear linear region, which could be employed as the area to project the motion of string vibration on top of the photodetector window. The linear region shown in Figure 3.44 extends only for a

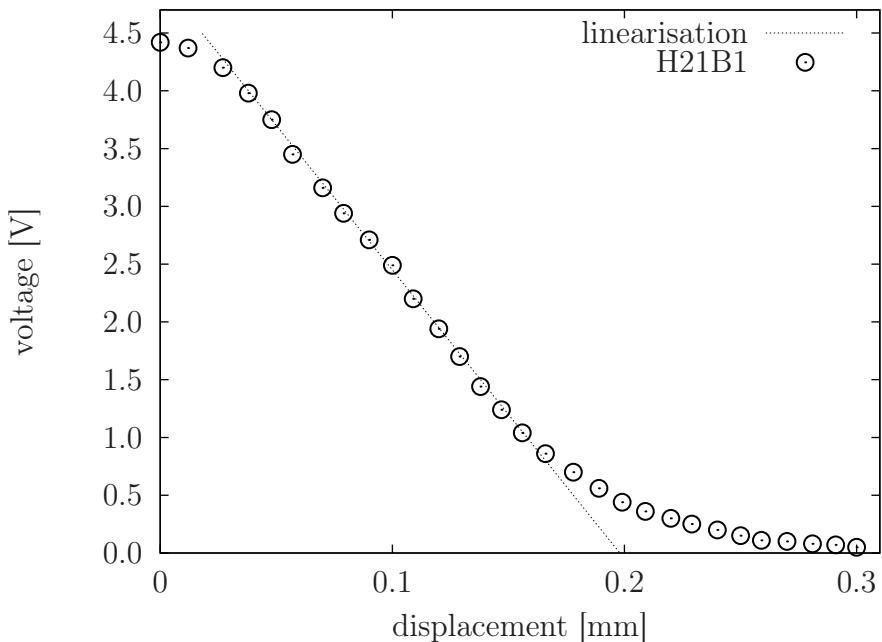


Figure 3.44: The linear region of operation of an optical switch. The graph is adaptively redrawn from results provided in article [37]

distance of a few tenths of a millimetre. Since the diameter of the photodetector window of H21B1 is 1 mm, the linear region is found approximately within 0.1 mm's distance from the middle of the receiver window. In practise, this requires an extremely high degree of accuracy in the placement of a string between the LED and the photodetector. Light gauge strings often require reduction of the photodetector sensor area because strings with a small diameter do not cover exactly half of the receiver window. In this situation, excess light leaks from the photoreceptor edges. The sensor area can be reduced using opaque tape, for

example.

As magnetic pickups are sensing the velocity of the string vibration, optical pickups are an example of displacement sensitive transducers. This property makes a big difference to the spectrum content of the signal produced by the optical pickup. Since the analytic formula (2.34) for the amplitude spectrum of the ideal string is based on the displacement of the string, the same formula combined with the pickup weight function describes the spectrum obtained from the optical pickup. In comparison to the magnetic pickup, the higher harmonic content of the optical pickup is quite poor. This is compensated in practise by placing the optical transducers in the vicinity of the bridge of the guitar. As explained in section 3.1, placing the pickup near the string's end support increases the amplitudes of the high-frequency upper partials of the string vibration.

Figure 3.45 gives an idea of the spectrum of an optical pickup by plotting the amplitude components of string vibration when the string is plucked at $\frac{L}{6}$ and the optical pickup is located at $\frac{L}{9}$ along the string. There is a clear difference compared to the spectrum of the magnetic pickup, shown in Figure 3.14.

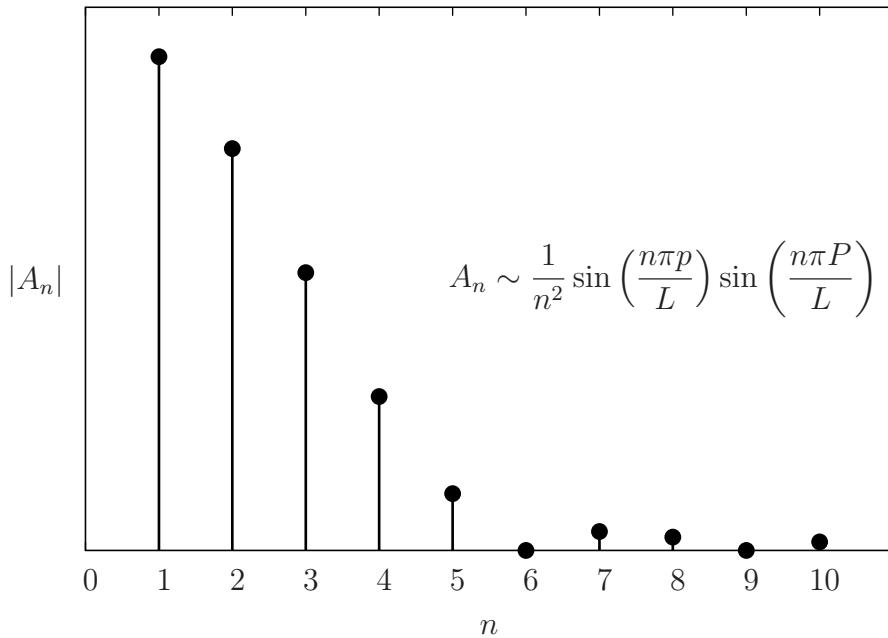


Figure 3.45: The upper partial structure of an optical pickup

The nodes in the analytic spectrum of the optical pickup are of course identified at the 6th and 9th upper partials as they should be. The key feature of a displacement sensing pickup is that in every situation the fundamental frequency

has the highest amplitude and the following harmonics are scaled by the multiplier $\frac{1}{n^2}$. Since the optical pickup forms a linear projection on the motion of the string, elliptic trajectories and perpendicular motion in general are not correctly transduced by the optical pickup. This is a clear defect compared to the magnetic pickup, where the motion of the string is detected in all three dimensions.

3.3.3 An optical pickup as part of an electric circuit

The part of the optical pickup which couples with the following stages of tone and volume control is the photodetector circuit, which is typically implemented using a single phototransistor component. A phototransistor is generally a normal bipolar junction transistor where a photodiode has been connected between the base and the collector. This idealisation model is shown in Figure 3.46.

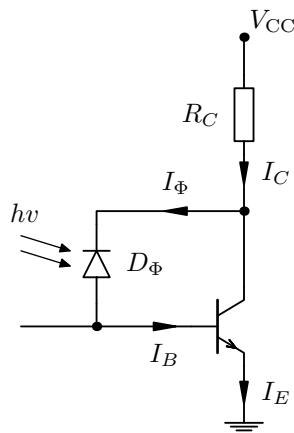


Figure 3.46: A phototransistor as a photodetector device

A small-signal model based on the idealised phototransistor is depicted in Figure 3.47. Basically the small-signal model is equivalent with a basic BJT small-signal model but the photodiode appears as a linear current source between the collector and the base. When the transition capacitances C_π and C_μ are included in the model, the frequency response curve will have a high-frequency limit where the gain of the phototransistor will start to decay. The transition capacitances will not affect the low-frequency response in any way. [52, pp. 127 – 132]

To examine the high-frequency response of a typical bipolar phototransistor, the transfer function can be derived using the nodal matrix representation of the small-signal model. Since the base resistance r_b is usually very small, it is rational to neglect it altogether to simplify the setup to two voltage nodes V_1 and V_2 , which are labelled as nodes 1 and 2 in Figure 3.47. With this modification, the

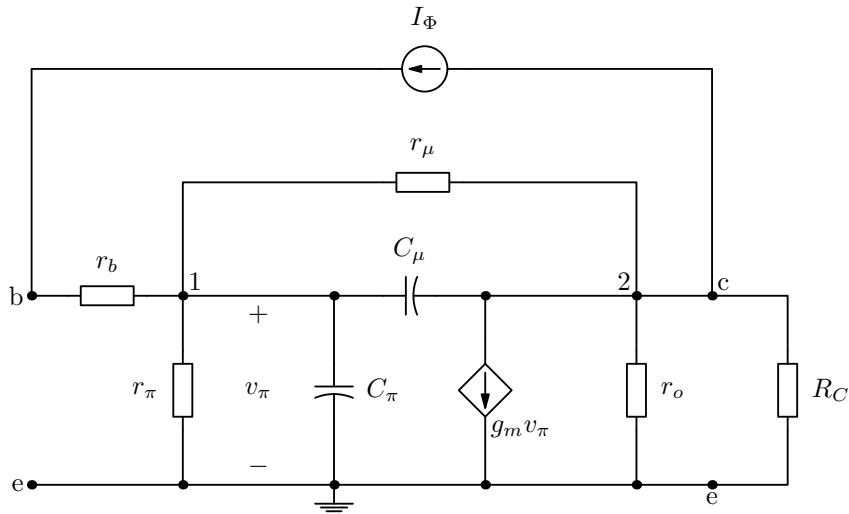


Figure 3.47: The phototransistor small-signal model

matrix equation becomes

$$\begin{bmatrix} \frac{1}{r_\pi} + \frac{1}{r_\mu} + j\omega C_\pi + j\omega C_\mu & -j\omega C_\mu - \frac{1}{r_\mu} \\ -j\omega C_\mu - \frac{1}{r_\mu} & \frac{1}{r_o} + \frac{1}{r_\mu} + \frac{1}{R_C} + j\omega C_\mu \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} I_\Phi \\ -I_\Phi - g_m V_1 \end{bmatrix}. \quad (3.22)$$

To solve the matrix equation, the transconductance term from the current vector is moved to the admittance matrix and then Cramer's rule is used to solve an expression for the voltage at node 2. Since the signal source is a pure current source, it is not possible to directly solve a transfer function as the ratio of the output and input voltages. However, since $\frac{V_2}{R_C} = I_{\text{out}}$, it is possible to solve the ratio of the output and input currents.

A direct application of Cramer's rule to matrix equation (3.22) leads to the expression for the node voltage V_2 :

$$\frac{-I_\Phi R_C (1 + j\omega C_\pi r_\pi + \beta_F)}{(1 + j\omega C_\pi r_\pi) \left(1 + \frac{R_C}{r_o} + \frac{R_C}{r_\mu} + j\omega C_\mu R_C \right) + (1 + j\omega C_\mu r_\mu) \left(\frac{r_\pi}{r_\mu} \left[1 + \frac{R_C}{r_o} \right] + \beta_F \frac{R_C}{r_\mu} \right)}.$$

After dividing by $I_\Phi R_C$, the transfer function

$$\frac{I_o}{I_\Phi} = \frac{-(1 + j\omega C_\pi r_\pi + \beta_F)}{(1 + j\omega C_\pi r_\pi) \left(1 + \frac{R_C}{r_o} + \frac{R_C}{r_\mu} + j\omega C_\mu R_C \right) + (1 + j\omega C_\mu r_\mu) \left(\frac{r_\pi}{r_\mu} \left[1 + \frac{R_C}{r_o} \right] + \beta_F \frac{R_C}{r_\mu} \right)}$$

is obtained. The denominator has been written in a compressed form because of lack of space. The essential thing to learn from here is that the time constants from all the mixed RC combinations have an effect on the slope and the corner frequency where the high-frequency attenuation begins.

Figure 3.48 visualises the effect of the load resistance R_C to the high-frequency cut-off. In this specific configuration, high load resistance values decrease the bandwidth of the phototransistor, but still the response includes the audio range even with very high resistance loads. The response curves have been normalised to a 0 dB amplitude in the passband to focus on the differences in the cut-off area. The absolute values of the gain are not important in this case, as only the relative differences between the curve shapes are meaningful.

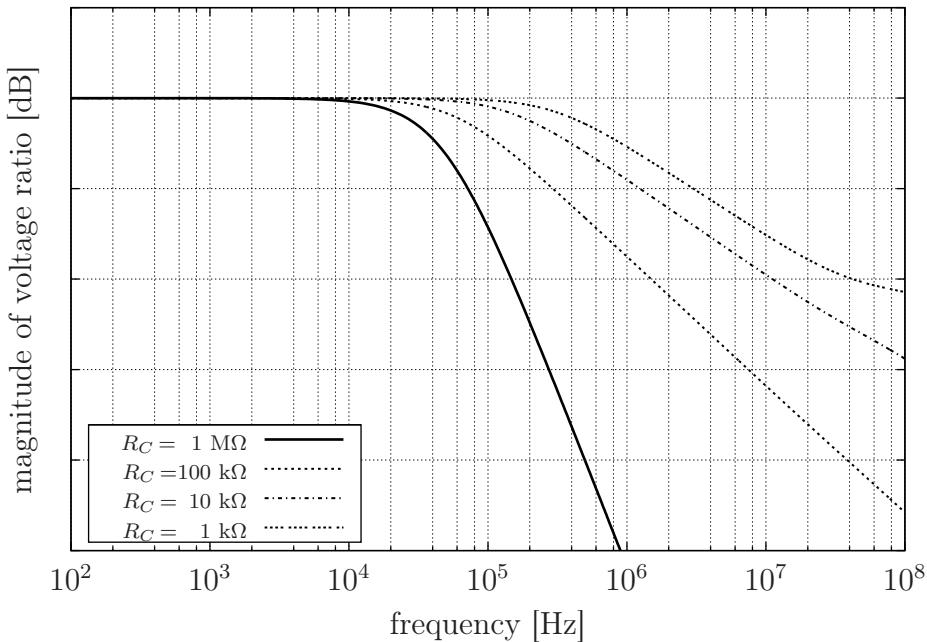


Figure 3.48: Frequency response curves with varying R_C

3.3.4 Measurements on an optical pickup

The measurement data originally used to analyse the waveforms and amplitude spectrums of plucked strings in Chapter 2 is reused here to present the true spectrum obtained from a typical optical pickup. In the plain string measurements in section 2.4, the assumed weight function of the pickup was divided away to reveal the realistic amplitude spectrum of a typical plucked string. When analysing these results to investigate the response of an optical pickup, the measurement data is kept in its original form and the analytical error limits are multiplied with the weight function of the pickup at the specific location.

The optical device used as a pickup was a photointerrupter switch H21B1; a component that can be purchased directly from a well equipped electronics store. The detector area of this specific interrupter switch has a diameter of only 1 mm,

so the width will not cause any unnecessary cut-off for any meaningful upper partial. The interrupter switch was hooked up to an external circuit, where it was connected as shown in Figure 2.24. The output signal was taken from the emitter of the phototransistor and measured with a digital storage oscilloscope.

Measurements were made by plucking the string at different locations and also from near both end supports. The measurement platform shown in Figure 2.23 was used to support the string, and the optical interrupter switch was placed relatively close to the end support. The interrupter switch was mounted with screws to the measurement platform and it was supported underneath to the suitable height with respect to the string. This mounting orientation forced the plucking direction to be vertical, although horizontal plucks could be imitated to some degree as vertical plucks by turning the platform 90 degrees sideways.

The theoretical results related to the signal obtained from an optical pickup are derived in equation (2.35) with the additional term describing the pickup weighting taken from section 3.1. Figure 3.49 compares the calculated theoretical waveform to the waveform recorded from the first few periods of the string's motion. The string was plucked at $\frac{L}{6}$ and the pickup was placed at 3 cm's distance from the right end support shown in Figure 2.23. The reason for concentrating on the first periods after the pluck is to avoid the uneven decay of upper partials, which causes the measured waveforms to deviate from the theoretical ones.

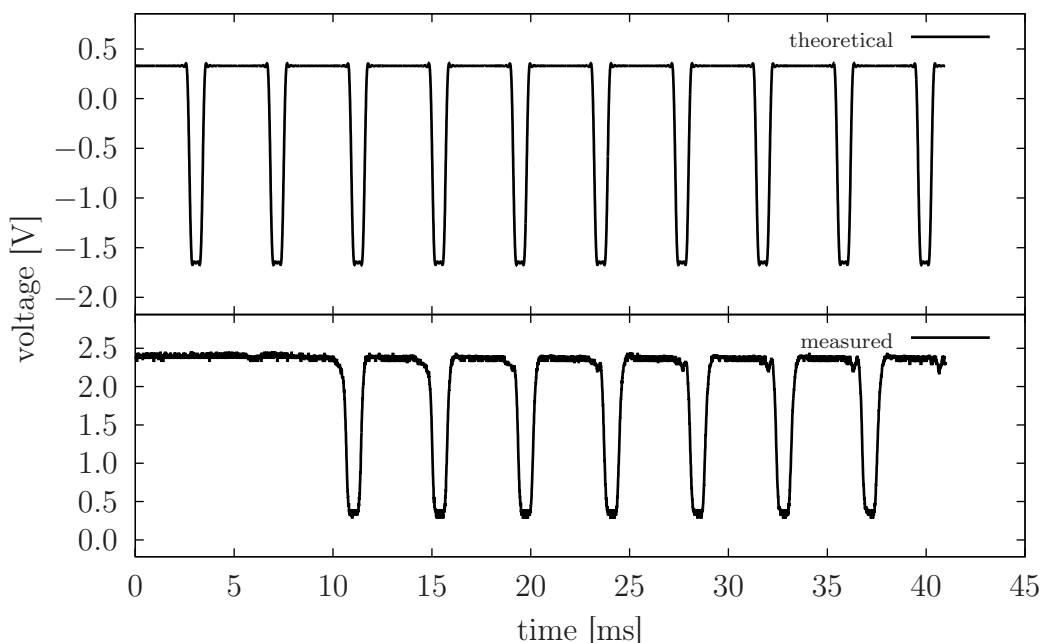


Figure 3.49: The measured and theoretical waveforms from a pluck at $\frac{L}{6}$

The waveforms in Figure 3.49 are quite similar, which proves reasonably well that the analytical model is sufficient enough to forecast the output signal of an optical pickup. The plucking position is sensed adequately as can be seen by the matching duty cycles between the theoretical expectation and the measured signal.

Figure 3.50 compares the waveforms from the experimentally most obvious pluck at the middle of the string. In this case the duty cycle of the measured waveform is expected to be fifty-fifty, which it actually is. The measurement results from the

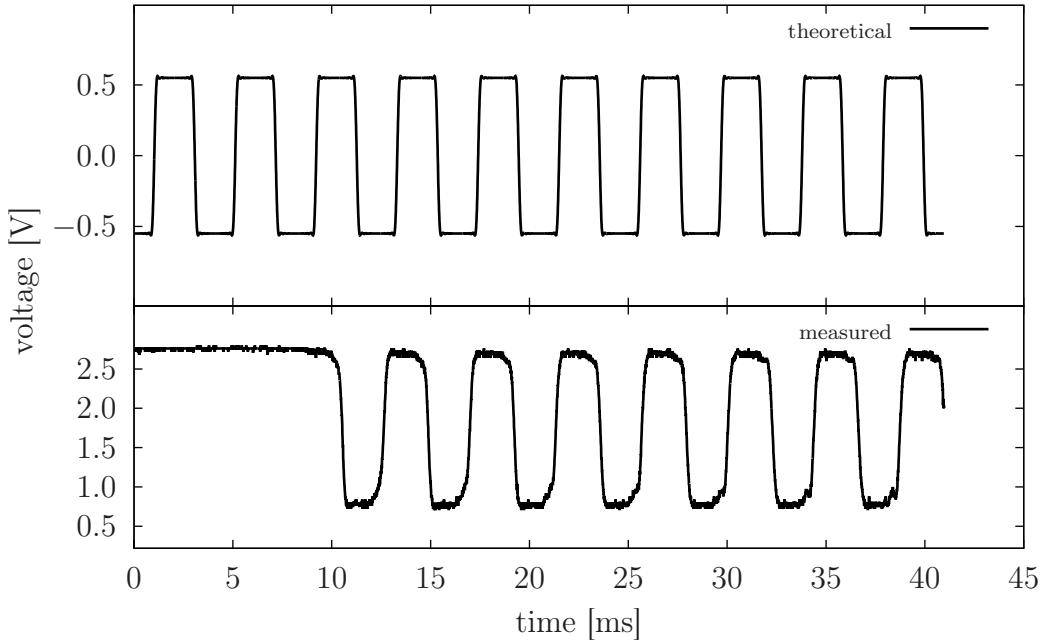


Figure 3.50: The measured and theoretical waveforms from a pluck at $\frac{L}{2}$

middle pluck show a good correspondence to the theoretical waveform, although the edges of the square pulses in the measured waveform are a little bit curved. The curviness indicates that some of the higher upper partials are either missing due to insufficient dynamics of the measurement setup or that the stiffness effects have covered some of the nodes of the higher upper partials, not allowing them to vibrate freely.

Figure 3.51 shows an example of the amplitude spectrum obtained using an optical pickup. The plucking position $p = \frac{L}{6}$ has been chosen so that the spectrum would be comparable to some extent to the spectrum measured from the magnetic pickup. On the other hand, in this measurement the optical pickup was placed at 8 mm's distance from the left boundary that clamped the string firmly.

Therefore, as the magnetic pickup was placed at 5 cm from the left end support, there is a big difference in the placement of the pickup.

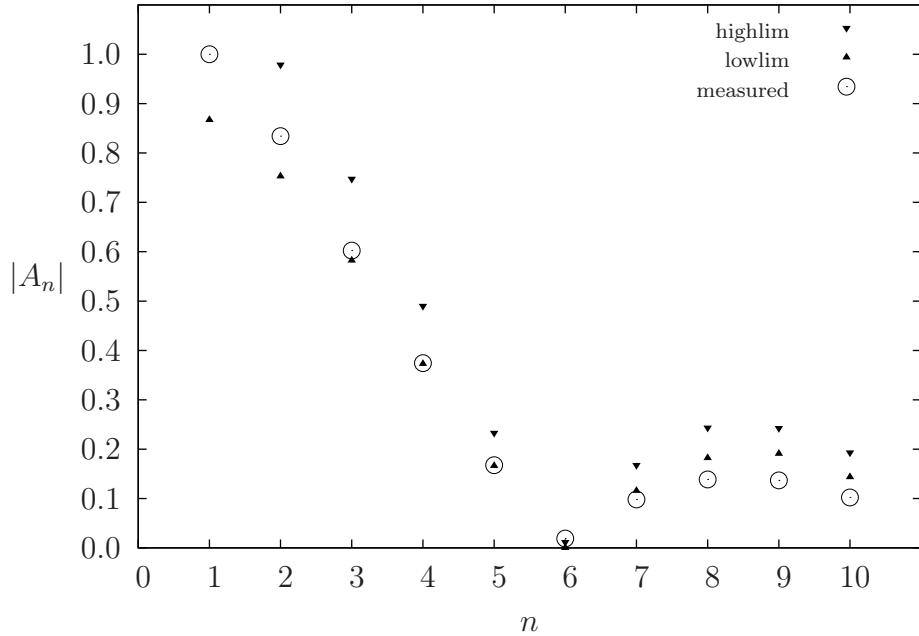


Figure 3.51: The amplitude spectrum resulting from a pluck at $p = \frac{L}{6}$

The amplitudes of the upper partials are in agreement with the theoretical error limits, although the higher harmonics have a smaller amplitude than expected. The amplitude spectrum indicates clearly that even when the optical pickup is so close to the boundary, the harmonic content is still clearly more richer in the magnetic pickup. The n^2 term in the denominator in equation (2.35) is orderly attenuating the upper partial response of the displacement sensing optical pickup and eventually cutting off the higher harmonics from the spectrum. The magnetic pickup can generate waveforms where the second or the third component have the highest amplitude in the initial spectrum, but for optical pickups the fundamental is *always* the most dominant.

The extreme case, where the plucking position is $p = \frac{L}{30}$, is plotted in Figure 3.52 to show a realistic maximum upper partial content obtained with an optical pickup. Unfortunately the measurement data turned out to be out of range with respect to the error limits, but the decreasing trend of the upper partials is still valid to give realistic information on the spectral content.

The measurement data in Figure 3.52 is the same where the spectrum for the string vibration was calculated. From this viewpoint, where the effect of the

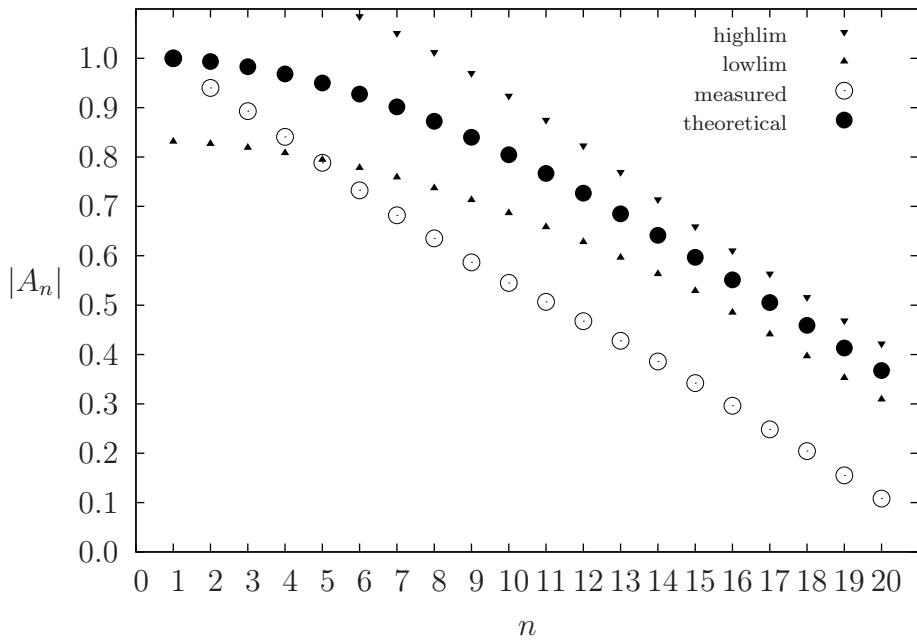


Figure 3.52: Plucking position at $p = \frac{L}{30}$

pickup location has not been divided away, it is clearly seen that the pluck has not been a successful one. Basically none of the upper partial amplitudes are within the error limits and the slight curvature of the sequential amplitudes is not visible in the plot. If the upper partials are evaluated after the 20th partial, it is noticed that the node appears already at the 23rd component. This indicates that the stiffness of the string this close to the end of the string is so strong that the exact plucking position is somewhat obscured. It is not very likely that there would have been such a large error in the plucking position, since the 23rd node is at 26 mm from the 'origin' and the 30th node is at 20 mm's distance.

3.4 PIEZOELECTRIC PICKUPS

Piezoelectric pickups are typically used to electrify acoustic guitars. There are many different designs of piezoelectric pickups available, the most common being an under-saddle mounted pickup, which can also be used in electric guitars. To sense the force of the string at the bridge of a guitar is the most reasonable solution to utilise the piezoelectric transducer because the theoretical modelling and analysis can be carried out relatively easily, as seen in section 2.2.4. Figure 3.53 shows a schematical view of how the piezoelectric pickup is mounted under the saddle, which supports the strings in the bridge of the guitar.

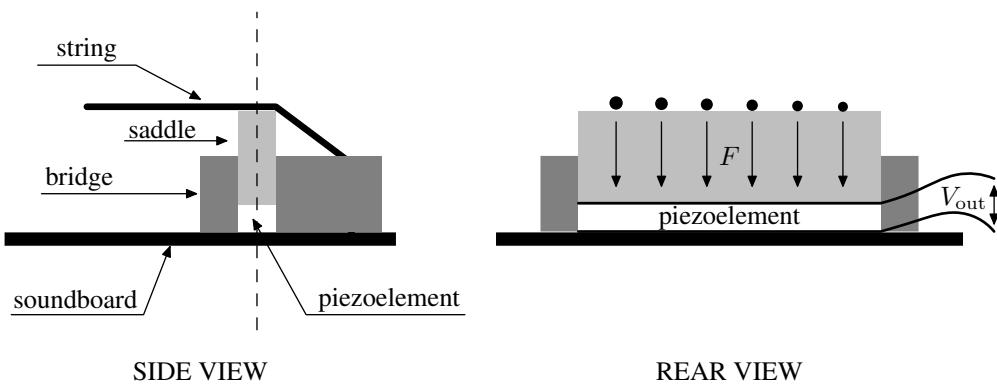


Figure 3.53: Under-saddle installation of a piezoelectric transducer

Some commercially available piezoelectric pickups are intended to be mounted in direct contact on the top or bottom side of the soundboard surface to be used more like a regular piezoelectric accelerometer. Since the soundboard of an acoustic guitar has resonance peaks at several locations for different frequencies, it might be difficult to find a suitable mounting location if linear operation is desired. The analysis of the properties of these *seismic pickups* would concern only acoustic guitars, which is out the scope of this book. However, the general principles of the under-saddle piezoelectric pickup are also applicable for these surface mounting solutions.

3.4.1 The piezoelectric effect

Piezoelectric pickups rely on the special properties of certain materials that charge up electrically due to deformation of the material under external stress. The *piezoelectric effect* was discovered by Jacques and Pierre Curie already back in 1880. To realise the piezoelectric effect, a mechanical force is applied to a piece of material having a certain crystal atomic structure. Under the influence of an external force, the material becomes electrically polarised and the degree of polarisation is proportional to the applied force. This process is also reversible. In the *inverse piezoelectric effect*, a material exposed to an external electric field changes its shape according to the applied voltage.

Figures 3.54 and 3.55 illustrate the fundamental properties of the piezoelectric effect. The reversible dualism of force and voltage is evident. From the figures, one can imagine either a force to be applied and the voltage measured or vice versa. A steady force applied in an equilibrium state is a special case where the output voltage of the piezo element is zero, as shown in Figure 3.54. [53, p. 2–4] [54, ch. 2]

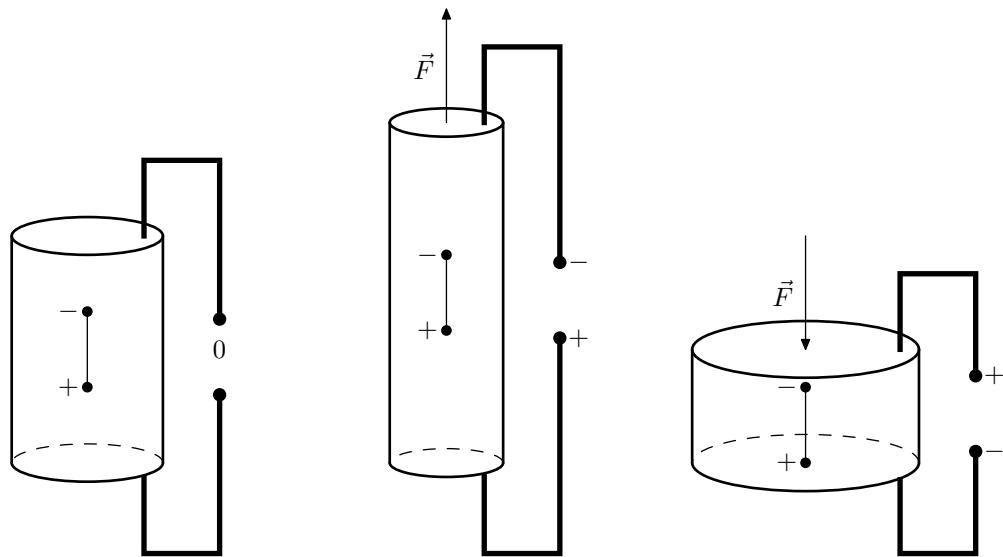


Figure 3.54: Voltage parallel to piezoelectric polarisation

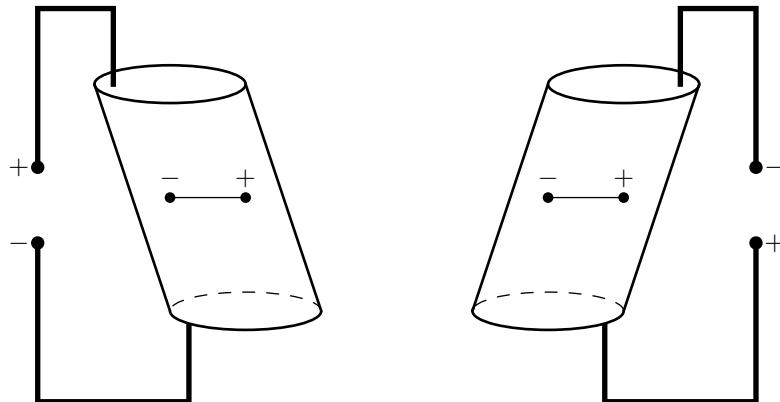


Figure 3.55: Voltage perpendicular to piezoelectric polarisation

In the guitar pickup application, the piezoelectric transducer is used as a force sensor that transforms the applied mechanical forces into an electrical signal.

3.4.2 A piezoelectric transducer as a signal source

In the under-saddle implementation, the piezoelectric pickup is used as a *fixed-reference transducer* where one terminal of the transducer is attached to a point that is fixed in space (bridge) and the other terminal is attached mechanically to the point whose force or acceleration is to be measured (saddle). When the piezoelectric pickup is placed under the saddle, it mainly experiences a dynamic compressional force as the string vibrates, but some shear forces are also evidently present in the horizontally polarised motion of the string. Figure 3.56 illustrates the basic situation where a piezoelectric transducer is under a direct

force of compression. This simple model is used as the foundation for further analysis.

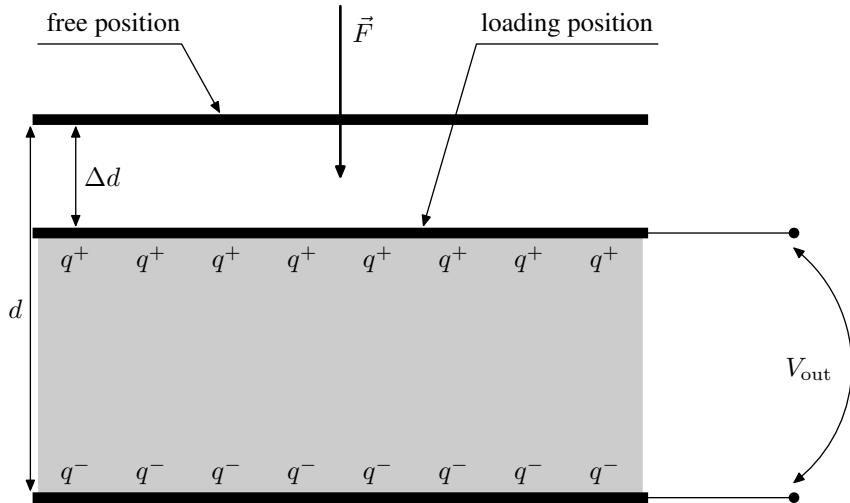


Figure 3.56: Deformation of a piezo element between conducting plates

When external forces affect the piezoelectric crystal, the charges inside the crystal momentarily lose balance and a net surface charge appears on top of the crystal. The parallel metal plates are there to collect the surface charge of the crystal using a capacitive charging process.

For the piezoelectric pickup to be trusted as a linear transducer, the issues of linearity need to be analysed with respect to variations in force magnitude and frequency, since both of these have to have linear dependency on the generated electrical signal. In this context, the focus of the analysis is to examine the piezoelectric transducer as a force sensor where the signal is taken in at the mechanical terminals of the sensor and the output is taken from the electrical terminals.

On a circuit level the piezoelectric transducer is modelled as a mixture of mechanical and electrical components. The equivalent electromechanical model is drawn in Figure 3.57, where the mechanical and the electrical parts of the piezoelectric transducer are coupled together via an electromechanical transformer. The transformer has a transform ratio of N , which carries the units of volts per newton. The transformer is intended to be understood as an ideal transformer, which only does a mathematical transformation of physical quantities as

$$V_1 = \frac{N}{1} F_2 \quad \text{and} \quad i_1 = \frac{1}{N} c_2,$$

where the subscripts refer to the primary and secondary side of the transformer. The practical idea is that voltage comes out from the primary side because of

force input, and velocity comes out from the secondary side because of current input.

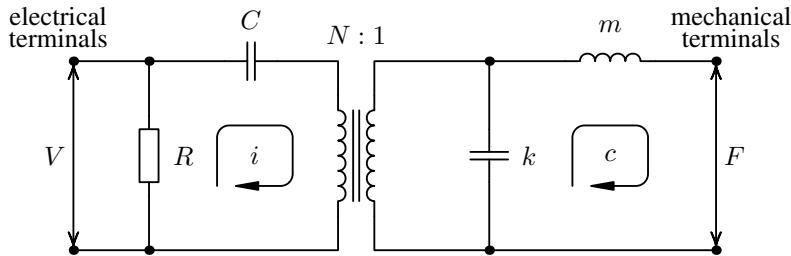


Figure 3.57: An electromechanical model of a piezoelectric transducer

The part that couples with external forces at the mechanical terminals is composed of electromechanical analogues as mass m and mechanical stiffness k , which is measured in the units of newtons per meter. The stiffness in this context is the same as the familiar spring constant k . The impedances of the mechanical elements are $j\omega m$ for the 'mass-inductor' and $\frac{k}{j\omega}$ for the 'spring-capacitor'. The quantity c that circulates around the mechanical circuit is velocity, which is measured as meters per second. Clearly the mechanical part is formed from elements that will lead to resonance effects in a certain frequency range.

To clarify the reasoning behind the electromechanical analogues, a simple explanation can be given based on the equivalent units of power in equation

$$Vi = Fc = \left[\frac{\text{kg} \cdot \text{m}^2}{\text{s}^3} \right], \quad (3.23)$$

where V is voltage, i current, F force and c velocity. From this equation it is possible to make two different choices for the electromechanical quantity pairs; voltage and force are quite obvious pairs and this is the preferred choice, but it is also possible, and sometimes even necessary, to choose velocity to pair with voltage and current with force.

When piezoelectric transducers are concerned, it is better to pair voltage with force, since both represent some sort of a potential and relate to energy in a similar fashion. After this decision from equation (3.23), the analogue pair of current and velocity is automatically obtained to represent the flow-type quantities. Moreover, since

$$i = \frac{dq}{dt} \quad \text{and} \quad c = \frac{dy}{dt}, \quad (3.24)$$

a relation between charge q and displacement y can be identified. Using the definitions of inductance and capacitance, the mechanical analogues can be eval-

ated from equation (3.23), for example,

$$Vi = L \frac{d^2q}{dt^2} i = Fc = m \frac{d^2y}{dt^2} c \quad (3.25)$$

connects mass m as the mechanical analogue of inductance L . Table 3.2 collects the most common electromechanical equivalences resulting from the choice of pairing voltage with force.

Table 3.2: Electromechanical equivalences when voltage is paired with force

electrical	symbol	mechanical	symbol
charge	q	displacement	y
current	i	velocity	c
inductance	L	mass	m
capacitance	C	compliance	$1/k$
resistance	R	damping coeff.	β
voltage	V	force	F

On the electrical side the piezoelectric transducer is basically a capacitor followed by a very large resistance R , which is seen between the electrical terminals of the transducer. The capacitance of the capacitor C is, to a good approximation, determined from the basic formula of capacitance between two parallel conducting plates [55, p. 241]. Although the electrical side does not have a pair of components to create resonances at certain frequencies, it still has the capacitor as a frequency dependent component.

To address the frequency dependent linearity of the mechanical part, it is possible to derive a transfer function for the right side 'circuit' of Figure 3.57. Similarity with the equivalent circuit of the magnetic pickup is evident, so the transfer function is expected to be in the same form. Going around the velocity loop in Figure 3.57 leads to the force equation

$$m \frac{dc}{dt} + k \int c dt = F(t), \quad (3.26)$$

which in the Laplace domain is written as

$$\left(ms + \frac{k}{s} \right) c = F(s),$$

from where

$$c(s) = \frac{F(s)}{ms + \frac{k}{s}}.$$

The output force at the terminals of the electromechanical transformer is the force over the stiffness-related capacitor. This force is written as

$$F_{\text{out}} = c(s) \frac{k}{s} = \frac{F(s) \frac{k}{m}}{s^2 + \frac{k}{m}},$$

which leads to the transfer function

$$\frac{F_{\text{out}}}{F_{\text{in}}} = \frac{\omega_0^2}{s^2 + \omega_0^2}, \quad (3.27)$$

where $\omega_0^2 = \frac{k}{m}$, indicating the resonance frequency. The result has the same form as in the case of the magnetic pickup, only the resistive term is now missing. Since the format of the transfer function is the same, then for a sinusoidal frequency sweep the response curves obtained from equation (3.27) are exactly the same as shown in Figure 3.25. However, the real world piezoelectric transducers are not that ideal, and the actual approximate frequency response curve looks something like shown in Figure 3.58.

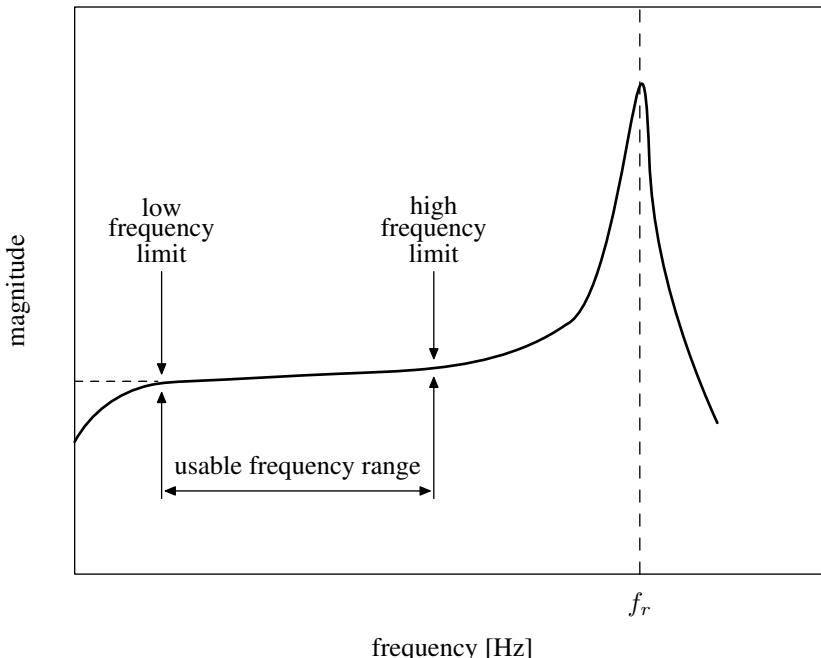


Figure 3.58: The linear region of a piezoelectric transducing element. Figure is redrawn with adaptations from reference [56].

Compared to the theoretical curves, the actual linear part is reduced by a limited low-frequency response and some excessive nonlinearities near the resonance frequency. Luckily there is still a relatively wide linear region in the response

curve, which indicates that the piezo element can be approximately used as a linear transducer as far as the frequency response is concerned. The resonance frequencies of piezoelectric transducers are typically in the ultrasonic frequency range so that normal audio frequencies are not affected by the nonlinearities of the resonance region.

The linearity between the magnitude of the input force and the resulting electric signal is not that evidently proved. However, since the whole network itself is a linear circuit model, and considering the fact that the stiffness related term is the dominating mechanical impedance at audio frequencies, it follows from equation (3.26) that

$$kx = F \quad \Rightarrow \quad x = \frac{1}{k}F. \quad (3.28)$$

Multiplication of equation (3.28) with the inverse of the electromechanical transformation ratio N gives

$$\frac{1}{N}x = \frac{1}{Nk}F,$$

and simplification leads to relation

$$q = \frac{1}{Nk}F = CF,$$

where q is the electric charge and C reflects capacitance with corresponding units, but this should not be confused with the capacitor noted with C in the electromechanical circuit of Figure 3.57. Since the quotient $\frac{q}{C}$ equals voltage, there is now a direct relation between the input force and the electromotive force generated to the primary coils of the electromechanical transformer.

Obviously, because expression for electric current i is obtained by differentiating the charge q with respect to time, the current is directly related to the changes in the applied force

$$\frac{dq}{dt} = i = \frac{1}{Nk} \frac{dF}{dt}. \quad (3.29)$$

Equation (3.29) clearly indicates that the piezoelectric transducer cannot be used to measure static forces since the current depends on the changes in the applied force. This is also why the piezoelectric force sensor is preferred to be modelled as a current source in the electrical domain.

3.4.3 A piezoelectric pickup as part of an electric circuit

Figure 3.59 shows the two possible equivalent circuits for a piezoelectric sensor as seen *in the electrical domain*. The signal source in Figure 3.59a having the magnitude $\left| \frac{Nkc}{j\omega} \right|$ indicates the voltage obtained from the electromechanical

transformer in Figure 3.57. The current source model shown in Figure 3.59b emphasises the piezoelectric transducer as a source of charge, which under a constantly varying force becomes a current source with a magnitude that equals the rate of change of the charge.

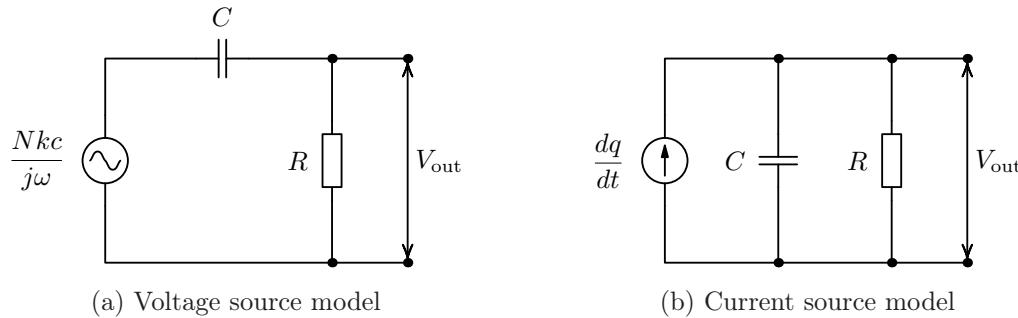


Figure 3.59: Force sensor equivalent circuits using voltage source and current source

As a good approximation, the electrical capacitance C of a force sensor having the structure as shown in Figure 3.56 is evaluated from the geometry of the transducer, where an insulation is separating two conductive plates. The capacitance is evaluated with the piezoelectric crystal in an unloaded state (air pressure as the only load), using the well-known formula

$$C = \frac{\epsilon\epsilon_0 A}{d}$$

for a plate capacitor. Variable A defines the surface area of each plate and d refers to the distance between the plates as shown in Figure 3.56. The constant ϵ refers to the material dependent value of relative permittivity and ϵ_0 is the permittivity of air.

Manufacturers of piezoelectric pickups do not bother to document the capacitance values of their products into any marketing material or datasheet, so the information needs to be iterated from unreliable sources in the Internet. A good guess appears to be around 500 – 1000 pF for under-saddle piezoelectric pickups. This capacitance value range can be used to simulate the piezoelectric pickup.

The internal resistance R of the piezo crystal is so huge that in many cases it can be neglected altogether. If the resistance is considered to be infinite, the output voltage V_{out} seen in the equivalent circuits is basically the voltage of the capacitor with some charge q :

$$V_{\text{out}} = \frac{q}{C} = \frac{1}{Nk} \frac{F}{C}.$$

Output voltage V_{out} refers to the actual voltage measured from the electrical output terminal of the force sensor, and it seems to be directly proportional to the applied force F . This proportionality can be analysed more generally from the equivalent circuits to see the effect of the resistance R . Direct transformation to the Laplace s domain gives the voltage equation

$$V_{\text{out}} = \frac{R \frac{1}{sC}}{R + \frac{1}{sC}} i = \frac{R}{1 + sCR} i,$$

and when using equation (3.29) a very interesting transfer function

$$\frac{V_{\text{out}}}{F} = \frac{sR \frac{1}{Nk}}{1 + sCR} = \frac{sCR \frac{1}{kNC}}{1 + sCR} \quad (3.30)$$

is obtained to link the output voltage to the input force. A time constant RC can be identified and the quotient $\frac{1}{kNC}$ defines a static sensitivity factor of the force sensor. As the factor $\frac{1}{kN}$ was noted to have the units of capacitance, the static sensitivity factor is a unitless quantity, which it needs to be to keep the derived equations sensible. [57, p. 585]

By taking the value of R very large in equation (3.30), a direct proportionality between the input force and output voltage is retained. However, if the resistance is less than $10^{10} \Omega$, then it starts to have a considerable effect on the frequency response: instead of a linear response, the relation between the input force and the output voltage becomes frequency dependent.

If the piezoelectric force sensor is used as a guitar pickup, it will be connected in parallel with a tone control resistance or an amplifier with finite input impedance Z_L , which is seen as a load from the perspective of the piezo element. Since the input impedance Z_L of the following circuit stage is parallel to R , the total impedance will equal Z_L because in practice Z_L will always be much smaller than the internal resistance R of the piezo crystal.

Figure 3.60 visualises the magnitude ratio between the input force and the output voltage assuming that the input force is kept at constant magnitude at all frequencies. Clearly only a $10 \text{ M}\Omega$ resistance seems to be sufficient to give a steady frequency response in the audio range. A capacitor value of 500 pF was used when calculating the gain curves. The gain curves in Figure 3.60 are normalised to 0 dB for clarity and they do not give realistic information about the absolute signal levels obtained from a typical piezoelectric pickup.

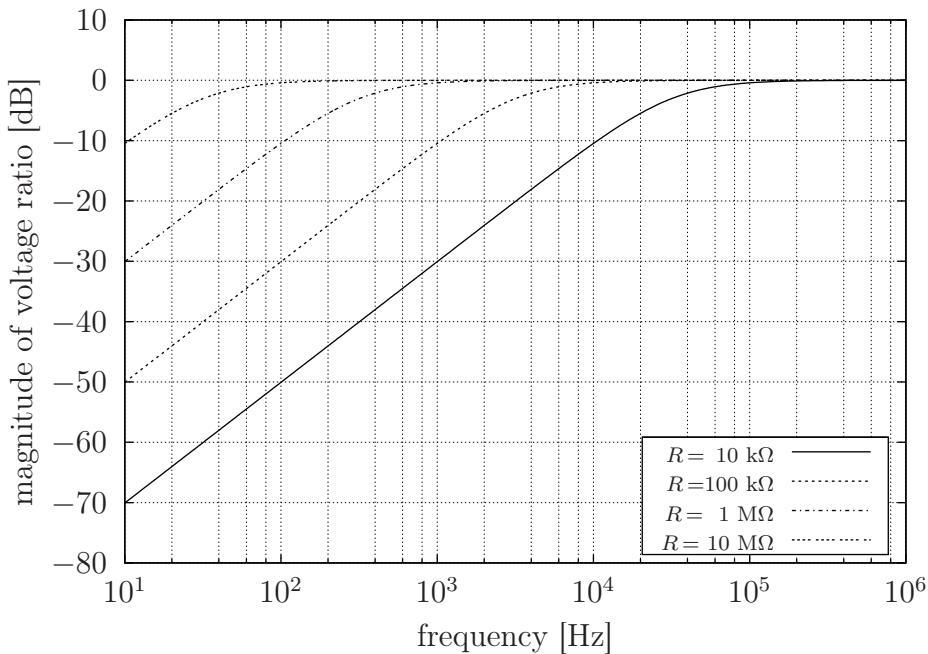


Figure 3.60: The frequency response of a piezoelectric pickup with varying load resistance

Based on the impedance analysis above, the piezoelectric pickup is not ready to be directly connected to a normal guitar amplifier or tone control circuit because the relatively low input impedance will distort the frequency response of the piezoelectric pickup. To isolate the pickup from the following circuit sections and to make a decent impedance matching, a preamplifier buffer circuit is needed directly after the piezoelectric pickup. There are several simple preamplifier implementations available in the Internet and Figure 3.61 shows one of them. The presented preamplifier is called a 'quick and dirty piezo preamp' and it is designed by Francis Deck. The preamplifier circuit is intended to be operated by a 9 V battery. In a static configuration, resistor R_2 self-biases the FET to a suitable range of operation and resistor R_1 determines the input impedance. Capacitor C_1 affects the low-frequency response of the circuit; larger capacitance values enhance the low-frequency response of the preamplifier stage.

Another good choice in addition to JFET buffers is to use dedicated *charge amplifiers*, basic op-amp circuits, which are presented in general literature on measurement and instrumentation. The reasoning for using a charge amplifier is obvious, since charge q is used as the current source in the equivalent circuit of a general piezoelectric transducer.

The small-signal model from a situation where the pickup is connected to the

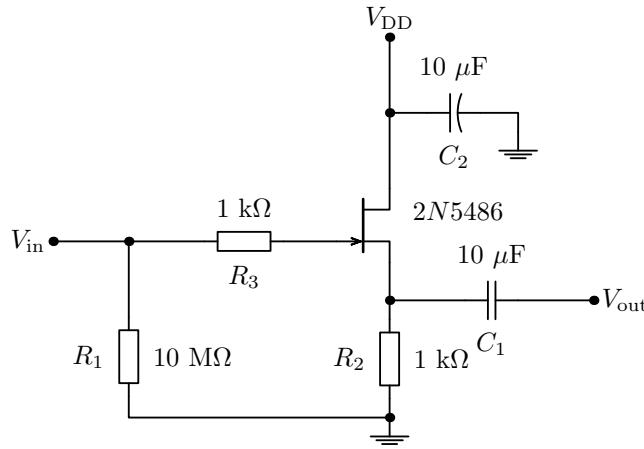


Figure 3.61: A preamplifier buffer for a piezoelectric guitar pickup

'quick and dirty' preamp is shown in Figure 3.62. The current source equivalent is chosen to represent the piezoelectric pickup and the value of the source is identified deliberately with the time derivative of charge to link directly to the time derivative of the input force. The output voltage is taken from node 5, where one can imagine a load resistance to be connected to represent the input impedance of a tone control circuit or a guitar amplifier. The resistor R_{0+1} refers to the parallel combination of the internal piezo crystal resistance and the resistor R_1 shown in Figure 3.61. Since the internal resistance of the piezo crystal should be much larger than R_1 , the effective value of R_{0+1} is close to $10 \text{ M}\Omega$.

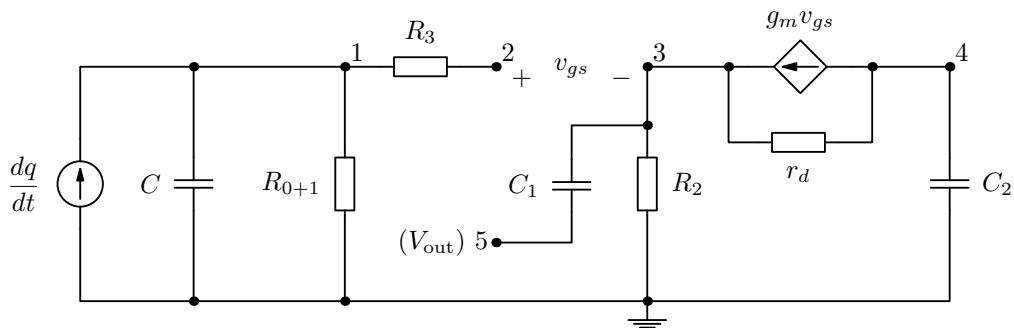


Figure 3.62: A small-signal model for a piezoelectric guitar pickup + preamp

The matrix equation (3.31) indicates the structure of the linear equations that need to be solved in order to calculate the frequency response of the combined

piezoelectric pickup and preamplifier buffer.

$$\begin{bmatrix} Y_{11} & Y_{12} & 0 & 0 & 0 \\ Y_{21} & Y_{22} & 0 & 0 & 0 \\ 0 & Y_{32} & Y_{33} & Y_{34} & Y_{35} \\ 0 & Y_{42} & Y_{43} & Y_{44} & 0 \\ 0 & 0 & Y_{53} & 0 & Y_{55} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{bmatrix} = \begin{bmatrix} j\omega q \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (3.31)$$

In the matrix representation the constant current source terms $g_m(V_2 - V_3)$ and $-g_m(V_2 - V_3)$ from rows 3 and 4 of the current vector have already been transferred to the admittance matrix. The expressions of the Y_{ij} elements in the admittance matrix are indicated in listing (3.32).

$$\begin{aligned} Y_{11} &= \frac{1}{R_{0+1}} + \frac{1}{R_3} + j\omega C & Y_{12} = Y_{21} &= -\frac{1}{R_3} \\ Y_{22} &= \frac{1}{R_3} & Y_{32} &= -\frac{\mu}{r_d} \\ Y_{33} &= \frac{1}{R_2} + \frac{\mu + 1}{r_d} + j\omega C_1 & Y_{34} &= -\frac{1}{r_d} \\ Y_{42} &= \frac{\mu}{r_d} & Y_{43} &= -\frac{\mu + 1}{r_d} \\ Y_{44} &= \frac{1}{r_d} + j\omega C_2 & Y_{35} = Y_{53} &= -j\omega C_1 \\ Y_{55} &= j\omega C_1 \end{aligned} \quad (3.32)$$

For JFETs, there exists the relation

$$\mu = g_m r_d,$$

which is equivalent to the equation $\beta_F = g_m r_\pi$ given for bipolar junction transistors. The value for internal drain resistance r_d and for the transconductance g_m should always be calculated from the biasing conditions, but a good guess is to have r_d somewhere between a few tens of kilohms and a few hundreds of kilohms [15, p. 426]. In this case the interest is towards the shape of the frequency response curve. Hence, the actual value of r_d is not meaningful, since it mainly affects the gain of the circuit.

In the matrix equation the charge based current source $\frac{dq}{dt}$ has been transformed into the Laplace domain as sq , from where in case of sinusoidal test signals the direct substitution of $s = j\omega$ is valid and shown in the current vector as $j\omega q$. This

way it is possible to obtain the transfer function in the form of output voltage divided by input force.

To compare the input and output voltages in the preamplifier circuit, the voltages at nodes 1 and 5 need to be solved. Numerical analysis with Octave results in the graph shown in Figure 3.63. According to the curves, the preamplifier buffer preserves the shape of the response curve for any external impedance to be connected to node 5. No additional gain is obtained since the preamplifier is in a source follower configuration. On the contrary, the signal attenuates just like predicted by the theoretical analysis of a source follower circuit.

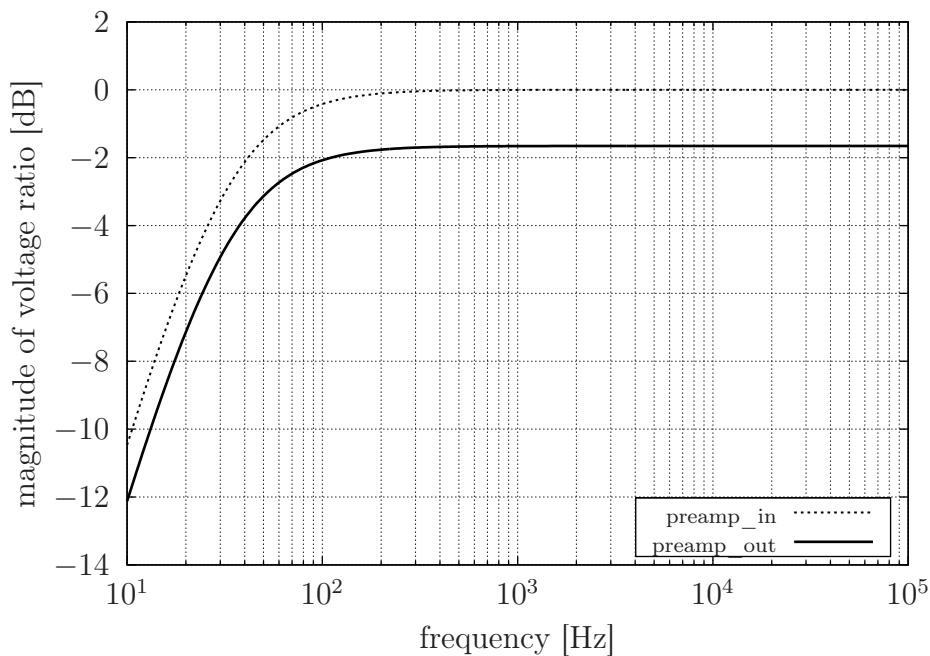


Figure 3.63: The frequency response of a piezoelectric pickup with a preamp

It should be questioned whether it is reasonable to include the resistor R_1 to fix the input impedance of the amplifier buffer. If R_1 is left out, the impedance is determined by the parallel connection of the internal JFET input impedance and the output resistance of the piezoelectric transducer. Both of these resistances should be an order or two higher than the $10 \text{ M}\Omega$ resistance.

Resistor R_1 has possibly been included to have a known value as the input impedance, since it is not evident what the actual resistances of the piezo element and the JFET input are. Without R_1 the low-frequency response of the preamplifier circuit should approach a flat line, which would be good news for those who want to have the bass tones maximised. On the other hand, the use of R_1 will stabilise the resistance because the linearity of the JFET and piezo crystal

resistances is not known in detail.

3.4.4 Measurements on a piezoelectric pickup

For measurement purposes, the cheapest possible under-saddle type piezoelectric pickup was purchased from the local electronics dealer. The same measurement platform as shown in Figure 2.23 was again utilised, and in this case the piezoelectric transducer was glued on top of the rightmost end support. Instead of imitating a realistic under-saddle installation, the piezoelectric pickup was used as the saddle itself so that the string was resting on top of the piezosensitive area. In this setup the piezoelectric pickup is used as a true force sensor for measuring the force generated by the vibrating string at the end support.

The string was set into motion by plucking it in the vertical direction. With this choice of direction, a vertical force component is generated at the 'saddle' to push and pull the string on top of the piezo element. By doing the measurements this way, the theoretically easiest modelling scheme can be directly compared to the measurement results. The theoretical equations to calculate the force at the end support were given in section 2.2.4 and those equations were used to create reference waveforms to be compared to the measurement results.

A regular steel string was used as the signal source and a digital storage oscilloscope was used in the single-shot configuration to record the first few periods of the string's vibration sequence. The $10 \text{ M}\Omega$ input resistance provided by the oscilloscope measurement probes is adequately large to not cause any frequency related distortion to the measurement results, since the nominal vibration frequency of the string was approximately 230 Hz.

Figure 3.64 depicts a measured waveform from a vertical pluck at $\frac{L}{6}$. The piezoelectric pickup was located at L because it was mounted on top of the rightmost end support. The measured signal is drawn alongside a waveform indicating the expected outcome of the measurement. The measured waveform matches very nicely to the theoretical expectation waveform, which was drawn using the first 20 upper partials from the harmonic spectrum of the string. The theoretical analysis predicts square pulses with an oscillating top section. The measured waveform has visible oscillation, which is more dominant than the theory predicts. Otherwise the duty cycles of the square pulses are in agreement. The theoretical waveform has not been frequency synced with the measured string, so therefore there is a small frequency difference between the theoretical and

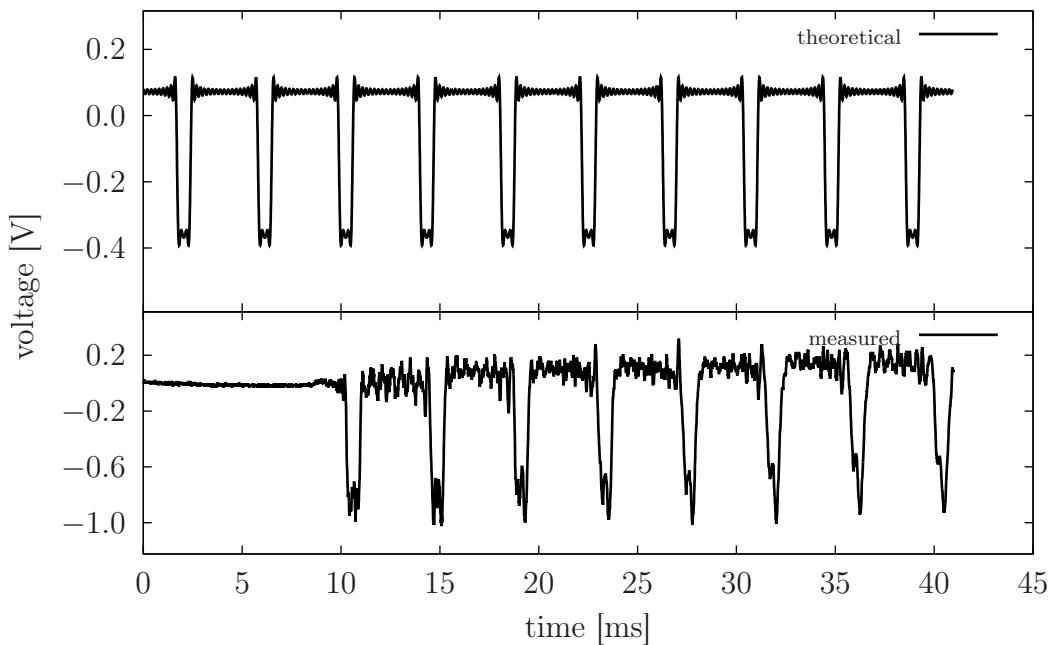


Figure 3.64: The measured and theoretical waveforms from a pluck at $\frac{L}{6}$

measured waveforms.

Figure 3.65 presents the theoretically most obvious pluck at the middle of the string. In this case, the duty cycle of the waveform is expected to be fifty-fifty, which it approximately is. This is a relatively successful measurement with a good correspondence to the theoretical waveform.

The initial deflection and release process of the string creates an additional transient force which gradually fades away. This effect was most noticeable when plucking from the middle of the string and it seems as though the average potential of the pulse train is gradually lifting upwards.

The shape of the waveform resulting from the force measurement at the end support is almost the same as the waveforms obtained by measuring with an optical pickup nearby the end support. The only visual difference is the amount of ripple at the top of the square pulses. Because of this, it is reasonable to expect similar spectrums from the piezo pickup measurements.

Theoretical analysis reveals that the expected spectrums resulting from measuring the force on top of the end support and measuring the shadow of the string in the vicinity of the end support should give the same magnitude spectrum for the most significant upper partials. There are differences only at the high end of

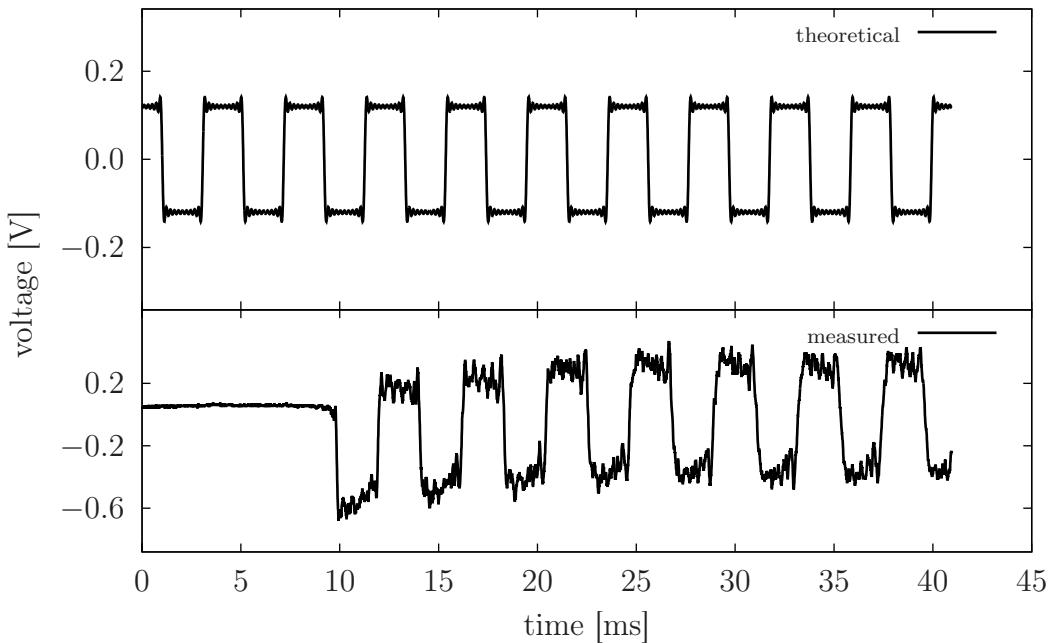


Figure 3.65: The measured and theoretical waveforms from a pluck at $\frac{L}{2}$

the upper partials, which is almost impossible to verify by measurements.

Figure 3.66 shows a spectrum that is calculated from the waveform shown in Figure 3.64. Considering the large amount of ripple in the measured waveforms,

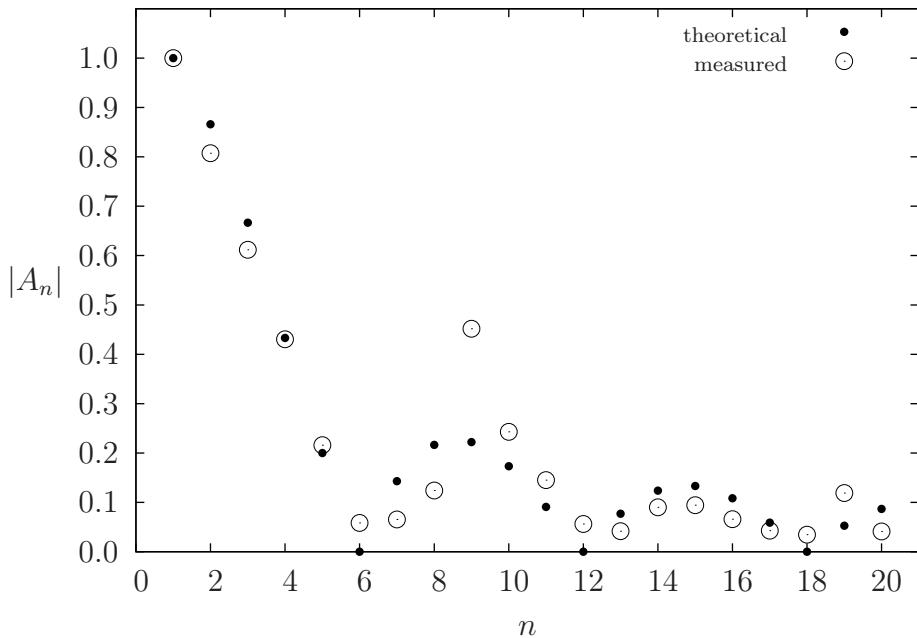


Figure 3.66: The measured and theoretical spectrums from a pluck at $\frac{L}{6}$

the amplitudes of the upper partials follow nicely the theoretically calculated amplitudes. Apparently the excessive ripple was caused by the hyperactive 9th upper partial, which is clearly stronger than the theory would expect. The reason for this behaviour was not investigated any further, but possibly the measurement setup had some resonances at that frequency range.

Figure 3.67 illustrates the difference in the waveforms between the vertical and horizontal plucking directions. When the force on top of the piezo element is due to parallel motion instead of perpendicular motion, the waveform is a not that distinct square wave but it is still faintly identifiable as a centre pluck from the 50 percent duty cycle.

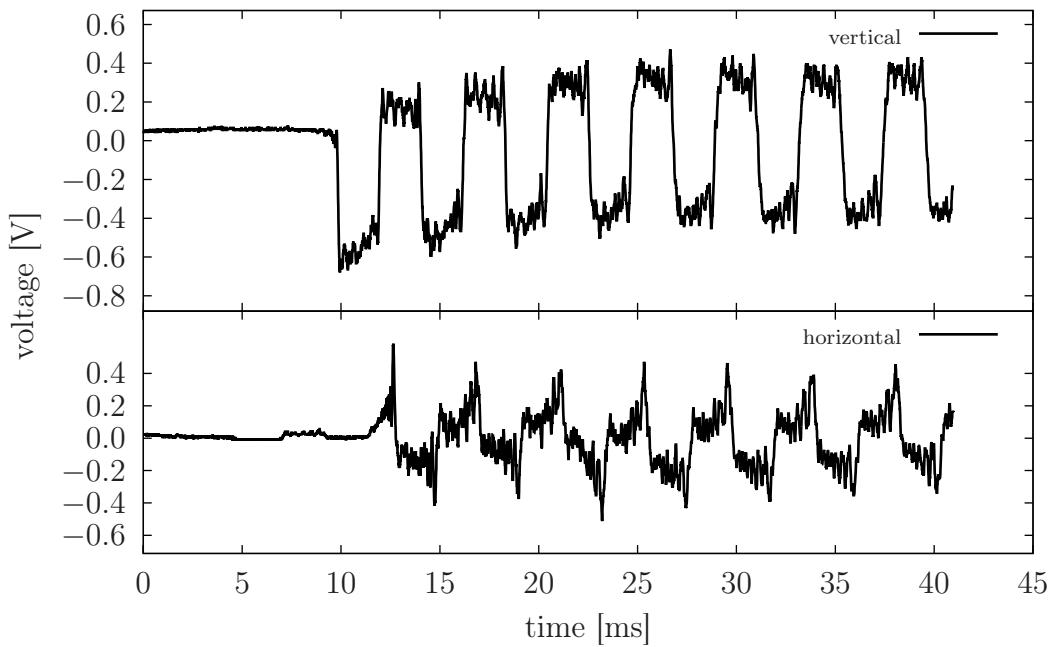


Figure 3.67: Comparing vertical and horizontal plucks at $\frac{L}{2}$

CHAPTER 4

GUITAR TONE CONTROL CIRCUITS

On the journey through the signal path of the electric guitar, after the transducer stage there comes a tone control stage. The tone control circuit is a potentiometer controlled filter which can be used to cut out certain frequency bands from the output signal of the pickup. Usually the tone control circuit is a very basic first order low-pass filter that just attenuates high frequencies as a function of an adjustable potentiometer resistance.

Nothing prevents from using complex filter structures as the tone control section of the guitar, but this is seldom necessary since the amplifier normally has its own tone control circuitry as a built-in feature. Another fact of the matter is that the more components are added to the passive network following the pickup, the more the guitar's output signal will be attenuated. In every practical design it is always better to keep things as simple as possible and minimise the amount of components to save unnecessary costs.

4.1 A BASIC TONE CONTROL CIRCUIT

According to Anderton [42, p. 75], the most basic tone control circuit used in electric guitars is similar to the circuit in Figure 4.1. This kind of tone control solution is normally installed by default to all affordable standard type guitars usually bought by beginner musicians. In Figure 4.1 the tone control stage has been 'soldered' to the equivalent circuit model of the pickup. The pickup characteristics are defined by resistance R_P , inductance L_P and capacitance C_P . The tone control stage is parallel to the pickup and it is represented by a series connection of resistance R_T and capacitance C_T . For the sake of completeness, the volume control potentiometer R_V has also been included in this model.

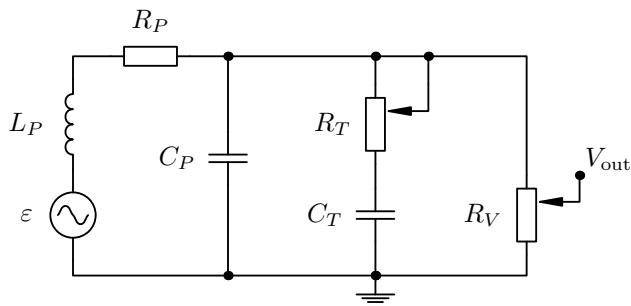


Figure 4.1: A magnetic pickup with a standard tone and volume control circuit

In this tone control circuit, potentiometer R_T controls the amount of signal current that is directed to ground through capacitor C_T . If the resistance R_T is large, most of the current will travel towards the volume control stage. When the resistance R_T is decreased, more current can escape to ground via the tone capacitor. The capacitor, on the other hand, acts as a smaller resistance for high frequencies than for low frequencies, and therefore the high frequencies are attenuated more than the low frequencies.

Typical resistance values for the tone and volume potentiometers are $250\text{ k}\Omega - 500\text{ k}\Omega$, and the value of the tone capacitor is usually between $0.01\text{ }\mu\text{F} - 0.1\text{ }\mu\text{F}$. In audio applications, logarithmic potentiometers are almost always used as the volume control potentiometers, whereas the type of the tone control potentiometer is chosen according to design. Anderton [42, p. 75] suggests a $250\text{ k}\Omega$ potentiometer for R_T and a $0.02\text{ }\mu\text{F}$ capacitor for C_T . These values are something that can be found from the standard line of guitars with single-coil pickups. Guitars equipped with more powerful humbucker pickups normally have these values doubled to

$$R_T = 500\text{ k}\Omega \quad C_T = 0.047 \times 10^{-6}\text{ F} \quad R_V = 500\text{ k}\Omega.$$

These values are used in this context for further analysis. The value of R_T will be varied throughout its possible range to see how it affects the signal.

The values used to define the pickup characteristics are

$$R_P = 10\text{ k}\Omega \quad C_P = 240 \times 10^{-12}\text{ F} \quad L_P = 5\text{ H},$$

thereby presenting a reasonably high impedance pickup with a resonance peak close to 5 kHz. The value for C_P has been tweaked deliberately for these simulations to reach a 5 kHz resonance because then the effect of the tone control circuit becomes more clear as it is possible to notice the gradual disappearance of the pickup's natural resonance peak when R_T is adjusted.

To be able to use nodal analysis to derive a transfer function for the tone control circuit, an equivalent small-signal model is redrawn in Figure 4.2, where the source of the electromotive force of the pickup has been transformed into a current source and the volume potentiometer has been divided in two parts, R_{V1} and R_{V2} . In the simulations, the volume has been set to the middle position, i.e. $R_{V1} = R_{V2} = \frac{R_V}{2}$, and the effect of varying R_T is examined.

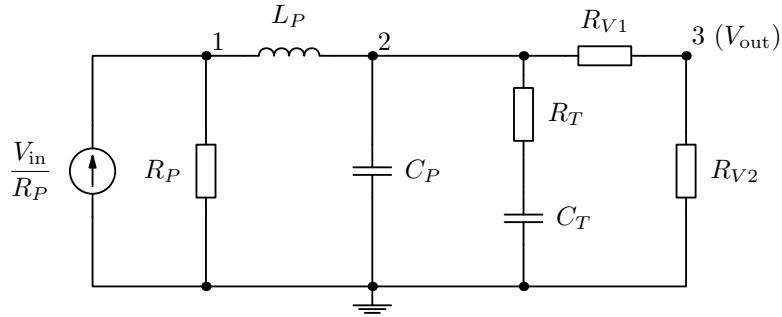


Figure 4.2: The tone control network redrawn for nodal analysis

From the redrawn small-signal model of the tone control circuit, it is easy to directly write down the matrix equation describing the circuit. Because the circuit has three nodes plus one common ground, the matrix equation will have three rows representing the three linearly independent equations derived from Figure 4.2.

$$\begin{bmatrix} \frac{1}{R_P} + \frac{1}{sL_P} & -\frac{1}{sL_P} & 0 \\ -\frac{1}{sL_P} & \frac{1}{sL_P} + sC_P + \frac{sC_T}{1+sC_T R_T} + \frac{1}{R_{V1}} & -\frac{1}{R_{V1}} \\ 0 & -\frac{1}{R_{V1}} & \frac{1}{R_{V1}} + \frac{1}{R_{V2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_P} \\ 0 \\ 0 \end{bmatrix}$$

The transfer function of the circuit can be evaluated from the matrix equation using Cramer's rule. Since the output signal is taken from node 3 of the small-signal model, column number 3 in the numerator determinant is replaced by the current vector to solve the node voltage V_3 . After some lengthy algebraic festivities, the transfer function has evolved into the form

$$H(s) = \frac{V_3}{V_{in}} = \frac{1 + sC_T R_T}{As^3 + Bs^2 + Cs + D},$$

where the coefficients

$$A = \left(1 + \frac{R_{V1}}{R_{V2}}\right) C_P L_P C_T R_T$$

$$B = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_P L_P + C_T L_P + C_P R_P C_T R_T) + C_T L_P \frac{R_T}{R_{V2}}$$

$$C = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_T R_P + C_P R_P + C_T R_T) + C_T R_T \frac{R_P}{R_{V2}} + \frac{L_P}{R_{V2}}$$

$$D = \frac{R_{V1}}{R_{V2}} + \frac{R_P}{R_{V2}} + 1.$$

The simulation results of the circuit presented in Figure 4.3 reveal that the tone control circuitry with the chosen component values clearly attenuates the resonance peak in the pickup response curve already at the maximum value of R_T . The high-frequency roll-off after the -3 dB point is comparable to a second order low-pass filter because the first order comes from the pickup stage and the second order is due to the tone control stage.

If the resonance peak of the pickup had been closer to 10 kHz, it would have faded away almost completely due to the tone control section. It is therefore obvious that the impedance characteristics of the pickup must be considered carefully when designing the tone control circuit. Sensible design would allow the natural resonance peak of the pickup to be audible at least with maximum bandwidth of the low-pass filter. If the resonance peak of the pickup is located higher than 5 kHz in this example, the tone control circuit should be modified to raise the cut-off frequency of the low-pass filter to preserve the resonance peak. Unfortunately this setup of R_T and C_T does not form a decent low-pass filter to be able to efficiently modify the cut-off frequency at the maximum value of R_T . Changing R_T to a 1 MΩ potentiometer helps a little, since then the capacitor C_T does not have much effect when R_T is at maximum value.

Trial simulations were also conducted to examine the effect of different volume potentiometer settings (half, max, min) but this did not cause any noticeable change in the frequency response. This is good because obviously the volume control should modify only the volume level and the tone control only the tone characteristics.

As noted already in the pickup analysis, the frequency of the pickup resonance peak can be controlled by changing the value of pickup capacitance C_P . Additional capacitance parallel to the pickup lowers the frequency of the resonance peak, and in connection with the tone control circuit, lowering the resonance frequency of the pickup enhances the effect of the pickup resonance at certain frequencies.

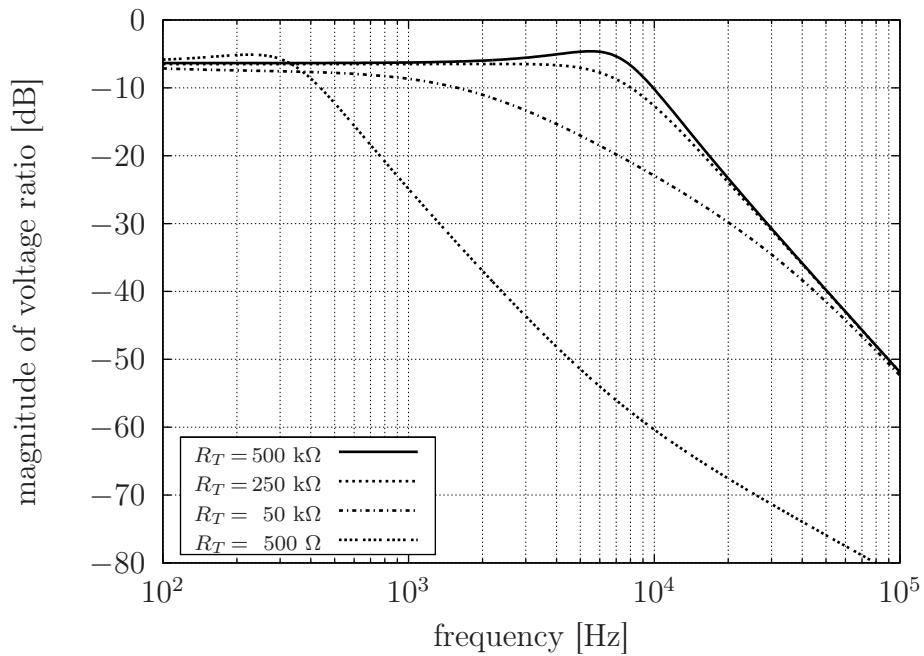


Figure 4.3: The calculated frequency response at the tone section output

What comes to the output impedance of the circuit 4.1, changes in the tone potentiometer R_T affect the output impedance of the guitar as shown in Figure 4.4. The equation from where the impedance curves in Figure 4.4 are drawn is

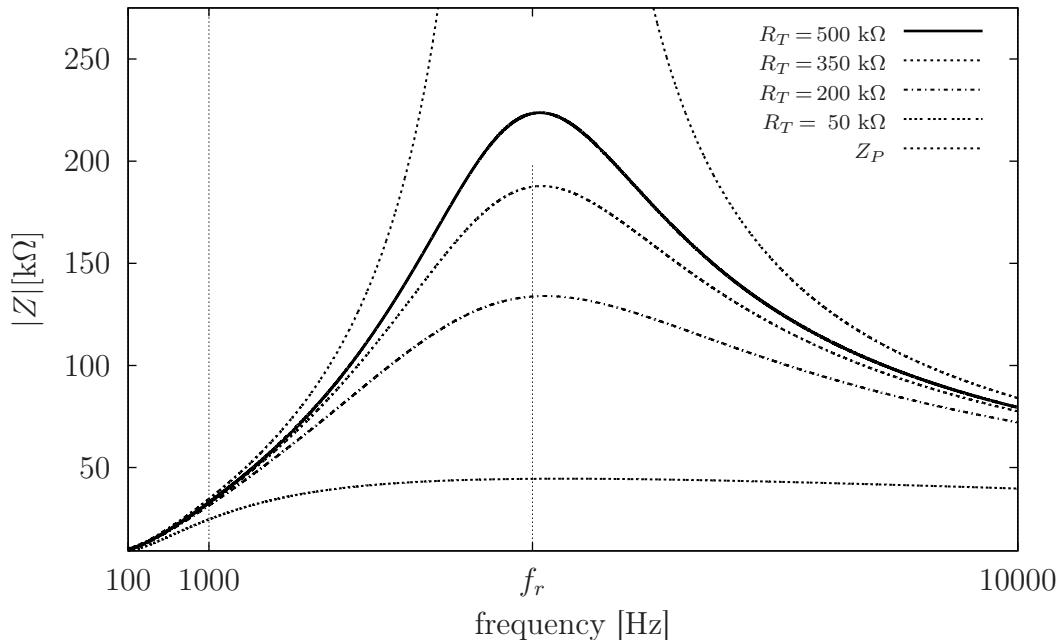


Figure 4.4: Impedance curves with a changing value of R_T

$$Z_{PTV} = \frac{\left[\frac{Z_P Z_T}{Z_P + Z_T} + R_{V1} \right] R_{V2}}{\frac{Z_P Z_T}{Z_P + Z_T} + R_{V1} + R_{V2}} \quad (4.1)$$

where the notation Z_{PTV} refers to the common impedance of the pickup, tone and volume control circuitry. In equation (4.1),

$$Z_P = \frac{\frac{1}{sC_P}(R_P + sL_P)}{\frac{1}{sC_P} + R_P + sL_P}$$

refers to the impedance of the pickup section and

$$Z_T = R_T + \frac{1}{sC_T}$$

is the impedance of the tone section.

The impedance peak at resonance frequency is preserved for values of R_T above 100 kΩ but the height of the peak is considerably smaller compared to the impedance of the separated pickup stage. At the edges it still follows the impedance curve of the unconnected pickup, which is drawn as a reference Z_P to Figure 4.4.

Similarly, changes in the volume potentiometer R_V affect the output impedance of the guitar as shown in Figure 4.5. When turning down the volume, the total

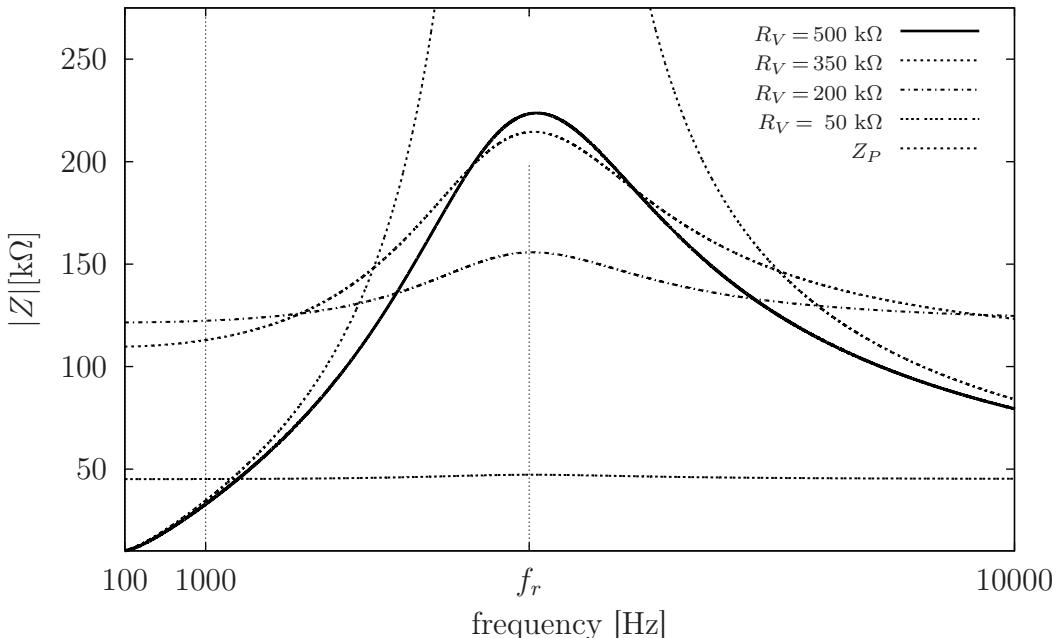


Figure 4.5: Impedance curves with changing a value of R_V

impedance will equal the resistance of R_{V2} , which refers to the part of the volume control potentiometer that is parallel to the output jack of the guitar. When the volume is gradually turned louder, the total impedance starts to resemble the normal impedance curve of the pickup. At least now we know that if someone wants to play with constant output impedance for all frequencies, just turn down the volume ... But as the evaluation of the frequency response revealed, the volume potentiometer will not affect the tone at all. Strange.

As a conclusion, in a normal playing situation where the volume control is normally turned to full volume (maximum resistance), the output impedance measured from the output jack of the guitar is a few hundred kilohms in the kilohertz frequency range. For low fundamental frequencies the impedance is much lower. The huge impedance peak of the separate pickup section is evened out considerably by the tone and volume control section.

4.2 MODIFIED TONE CONTROL CIRCUITS

The basic tone control circuit in Figure 4.1 can be extended to offer more variable control with relatively minor modifications. The controllability increases by one dimension by adding one capacitor to the circuit. The solution presented in Figure 4.6 places the additional capacitor along with the tone control potentiometer, and this way it activates the unused second branch of potentiometer R_T . Suggested values for the two capacitors are $0.01 \mu\text{F}$ for one and $0.02 \mu\text{F} - 0.05 \mu\text{F}$

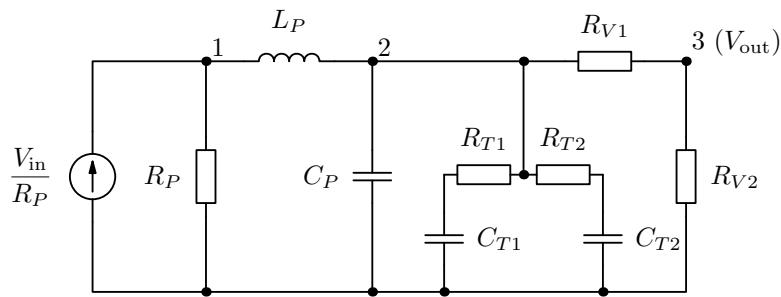


Figure 4.6: The modified tone control network for nodal analysis

for the other one, as given by Anderton [42, p. 75]. The tone control potentiometer is recommended as a linear $500 \text{ k}\Omega$ pot, but $250 \text{ k}\Omega$ is sufficient enough as well. For this analysis, the capacitor values were chosen as $C_{T1} = 0.047 \mu\text{F}$ and $C_{T2} = 0.01 \mu\text{F}$. Other component values are the same as in the analysis of the basic tone control circuit, which was presented in Figure 4.1.

The matrix equation derived from the small-signal model of the modified tone control circuit of Figure 4.6 is

$$\begin{bmatrix} \frac{1}{R_P} + \frac{1}{sL_P} & -\frac{1}{sL_P} & 0 \\ -\frac{1}{sL_P} & \frac{1}{sL_P} + sC_P + \frac{sC_{T1}}{1+sC_{T1}R_{T1}} + \frac{sC_{T2}}{1+sC_{T2}R_{T2}} + \frac{1}{R_{V1}} & -\frac{1}{R_{V1}} \\ 0 & -\frac{1}{R_{V1}} & \frac{1}{R_{V1}} + \frac{1}{R_{V2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_P} \\ 0 \\ 0 \end{bmatrix}.$$

An evaluation of the Cramer's determinants to solve the voltage at node 3 leads to the transfer function

$$H(s) = \frac{V_3}{V_{in}} = \frac{(1+sC_{T1}R_{T1})(1+sC_{T2}R_{T2})}{As^4 + Bs^3 + Cs^2 + Ds + E},$$

where the coefficients

$$A = \left(1 + \frac{R_{V1}}{R_{V2}}\right) C_P L_P C_{T1} R_{T1} C_{T2} R_{T2}$$

$$B = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_P L_P C_{T1} R_{T1} + C_P L_P C_{T2} R_{T2} + C_{T2} L_P C_{T1} R_{T1} + C_{T1} L_P C_{T2} R_{T2} + C_P R_P C_{T1} R_{T1} C_{T2} R_{T2}) + C_{T1} R_{T1} C_{T2} R_{T2} \frac{L_P}{R_{V2}}$$

$$C = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_P L_P + C_{T1} L_P + C_P R_P C_{T1} R_{T1} + C_P R_P C_{T2} R_{T2} + C_{T2} R_P C_{T1} R_{T1} + C_{T1} R_P C_{T2} R_{T2} + C_{T1} R_{T1} C_{T2} R_{T2}) + C_{T1} L_P \frac{R_{T1}}{R_{V2}} + C_{T2} L_P \frac{R_{T2}}{R_{V2}} + C_{T1} R_{T1} C_{T2} R_{T2} \frac{R_P}{R_{V2}}$$

$$D = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_{T1} R_P + C_P R_P + C_{T1} R_{T1} + C_{T2} R_P + C_{T2} R_{T2}) + C_{T1} R_{T1} \frac{R_P}{R_{V2}} + C_{T2} R_{T2} \frac{R_P}{R_{V2}} + \frac{L_P}{R_{V2}}$$

$$E = \frac{R_{V1}}{R_{V2}} + \frac{R_P}{R_{V2}} + 1.$$

The addition of one component of complex impedance increased the degree of the denominator of the transfer function by one to the fourth degree. The simulation results for varied values of potentiometer R_T are depicted in Figure 4.7.

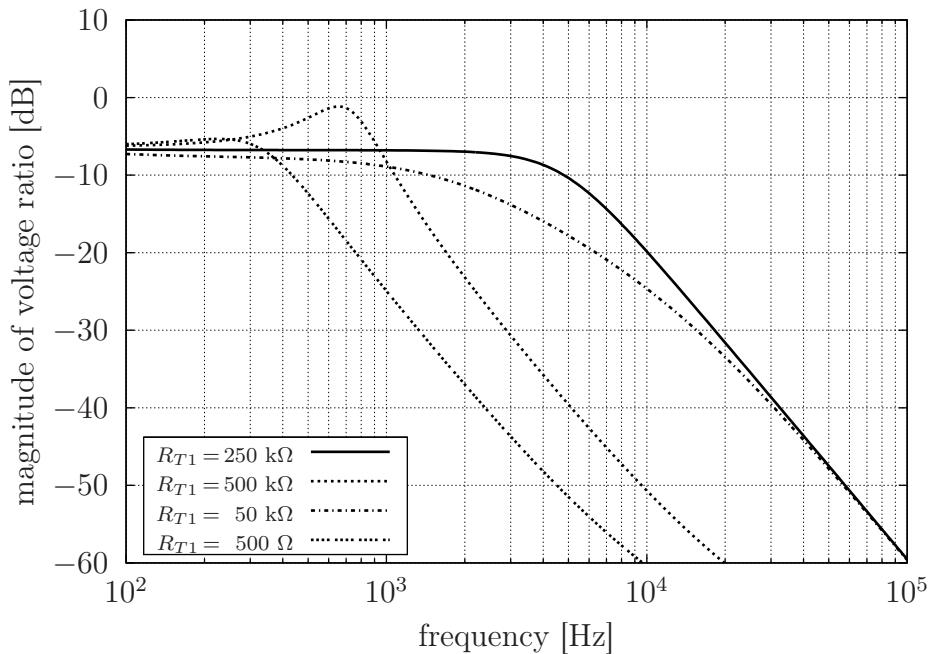


Figure 4.7: Frequency response curves of the modified tone control section

According to the simulation results, this solution does not offer anything drastically new compared to the basic tone control circuit. On the contrary, this modification effectively eats away the rest of the pickup's resonance peak, but at least some small changes of tone control can be achieved with this approach.

The idea of the circuit is that the neutral setting of the tone potentiometer is now at the middle position. Turning the potentiometer towards both extremes will increase the attenuation of the high frequencies, but now there are basically two different tone potentiometers combined into the one single tone control knob. The usefulness of this implementation should be questioned because it only gives two approximately similar low-pass filter functions to use. However, this is only how it looks like as a diagram, the ear might be more sensitive to notice the slight difference.

As another slight modification to the basic tone control circuit of section 4.1, Figure 4.8 adds again one capacitor to the circuit, this time in direct contact with the volume control potentiometer. The circuit is drawn directly as a small-signal model where the volume potentiometer has been divided into two separate resistors. The pickup is modelled as a current source so that direct use of nodal analysis would be possible. The idea of this circuit is to enhance the high-frequency portion of the output with respect to the low-frequency range. As the volume is turned down, the low frequencies are attenuated more than the high

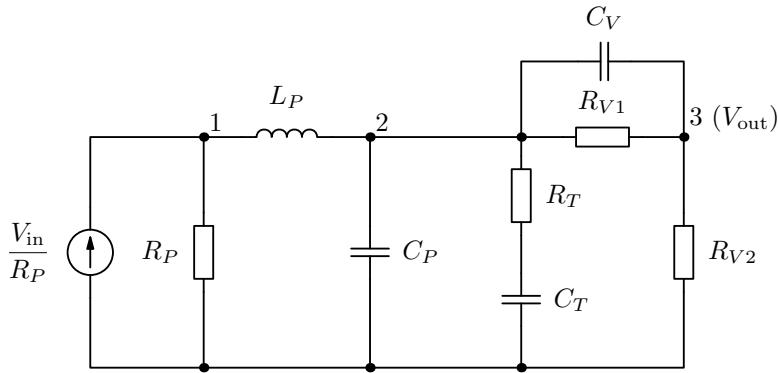


Figure 4.8: A combined control section for tone and volume

frequencies, which are bypassing the volume control potentiometer through the additional capacitor.

An application of the nodal analysis method to the small-signal model of Figure 4.8 results in a matrix equation

$$\begin{bmatrix} \frac{1}{R_P} + \frac{1}{sL_P} & -\frac{1}{sL_P} & 0 \\ -\frac{1}{sL_P} & \frac{1}{sL_P} + sC_P + \frac{sC_T}{1+sC_T R_T} + \frac{1}{R_{V1}} + sC_V & -\frac{1}{R_{V1}} - sC_V \\ 0 & -\frac{1}{R_{V1}} - sC_V & \frac{1}{R_{V1}} + \frac{1}{R_{V2}} + sC_V \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_P} \\ 0 \\ 0 \end{bmatrix}.$$

The output voltage at node 3 is obtained in a similar fashion as in the previous sections and the transfer function can be solved by dividing the node voltage by the input voltage V_{in} . In this case the ratio of output and input voltages as a function of the Laplace variable s reads

$$H(s) = \frac{V_3}{V_{in}} = \frac{(1+sC_T R_T)(1+sC_V R_{V1})}{As^4 + Bs^3 + Cs^2 + Ds + E},$$

where the coefficients

$$A = C_P C_V R_{V1} L_P C_T R_T$$

$$B = \left(1 + \frac{R_{V1}}{R_{V2}}\right) C_P L_P C_T R_T + C_V L_P C_T R_T \frac{R_{V1}}{R_{V2}} + C_T L_P C_V R_{V1} + C_P L_P C_V R_{V1} + C_P R_P C_V R_{V1} C_T R_T$$

$$C = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_P L_P + C_T L_P + C_P R_P C_T R_T) + C_T L_P \frac{R_T}{R_{V2}} + C_V L_P \frac{R_{V1}}{R_{V2}} + C_T R_T C_V R_{V1} + C_T R_T C_V R_P \frac{R_{V1}}{R_{V2}} + C_P R_P C_V R_{V1} + C_T R_P C_V R_{V1}$$

$$D = \left(1 + \frac{R_{V1}}{R_{V2}}\right) (C_T R_P + C_P R_P + C_T R_T) + C_T R_T \frac{R_P}{R_{V2}} + \frac{L_P}{R_{V2}} + C_V R_P \frac{R_{V1}}{R_{V2}} + C_V R_{V1}$$

$$E = \frac{R_{V1}}{R_{V2}} + \frac{R_P}{R_{V2}} + 1.$$

To calculate the frequency response for sinusoidal frequencies, the Laplace variable s in the transfer function is substituted with the complex angular frequency term $j\omega$ and the transfer function is evaluated for a range of frequencies. The calculated frequency response of the basic tone control circuit with the capacitively bypassed volume control potentiometer is presented in Figure 4.9.

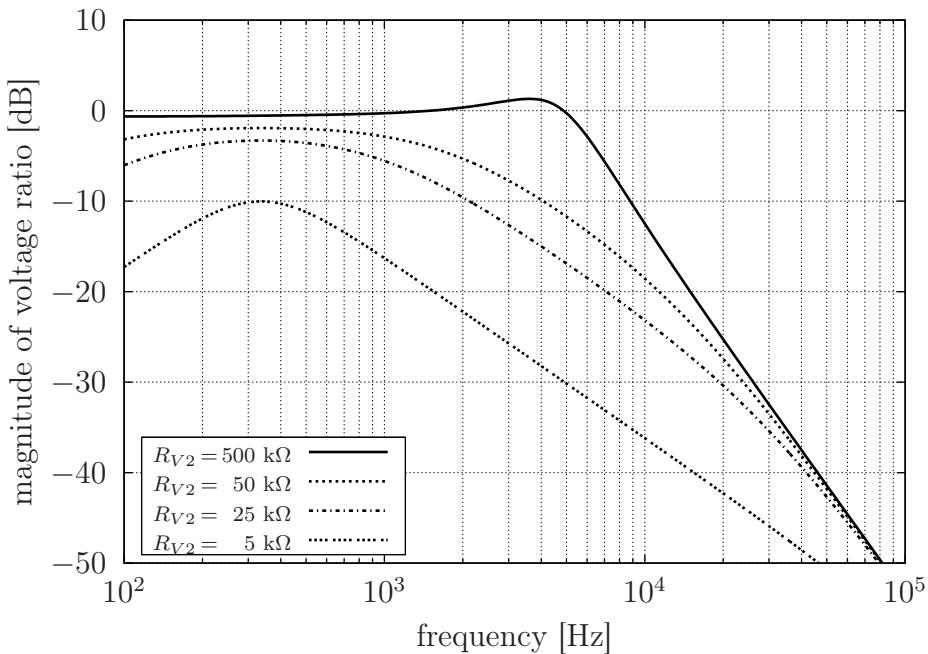


Figure 4.9: Calculated frequency response of the modified volume control section

The circuit works as expected since according to the response curves in Figure 4.9, low resistance values of the volume potentiometer R_V are attenuating the signal and at the same time cutting out the low frequencies. However, it is difficult to imagine that this kind of modification would be very useful because the attenuation at the low frequencies is not that noticeable compared to the total attenuation of the signal caused by the volume potentiometer. The relative attenuation of the low frequencies becomes clear only when the volume potentiometer is near to its minimum value.

An obvious extension to the basic tone control circuit is to add a high-pass filter

along with the existing low-pass filter. The solution of Figure 4.10 is adding the high-pass filter as a separate block, which needs one additional potentiometer to be added to the guitar's control knobs.

Clearly this implementation method is not very rational because the need to add another knob to the guitar would require some drilling of one's valuable instrument. A more elegant solution for implementing simultaneous low/high-pass filtering is presented in the next section.

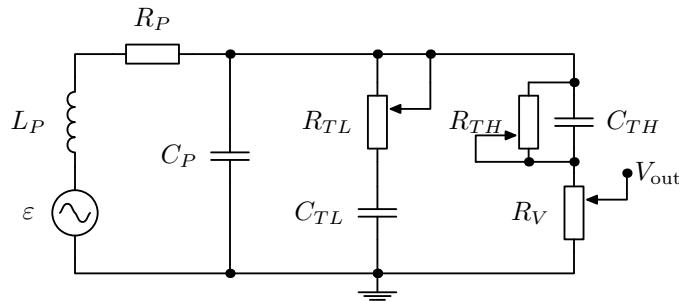


Figure 4.10: Tone control with an added high-pass filter section

4.3 THE 'BIG MUFF π' TONE CONTROL CIRCUIT

As already noticed, the tone control system embedded into the guitar is usually the simplest possible filter circuit. This is understandable because the amplifier normally has a so-called 'tone stack', which offers plenty of possibilities for tone control.

A model of a bit more versatile filter circuit which could be used for guitar tone control can be borrowed from a popular distortion and sustain effect pedal, 'Big Muff π'. This pedal was manufactured under the brand name of Electro-Harmonix during the years 1970 – 1984. The tone control section covers only a small part of the 'Big Muff' schematic [58] and can be easily extracted as an independent tone circuit. The tone section isolated from the 'Big Muff' schematic is drawn in Figure 4.11.

Clearly the circuit has two branches where one branch is a simple high-pass filter and the other one is a simple low-pass filter. The subscript *B* in the circuit diagram 4.11 refers to the components on the 'bass-cut' side and the subscript *T* refers to the components making the 'treble-cut'. These two filters are connected together with a potentiometer, which is used to adjust the filtering somewhere in between the two extremes of the high-pass and low-pass modes.

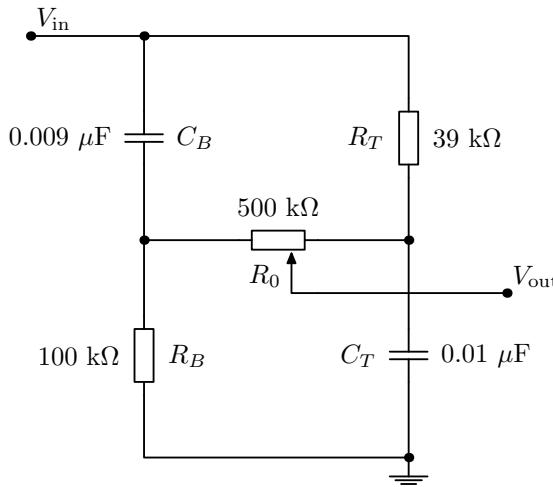


Figure 4.11: The circuit diagram of the 'Big Muff' tone section

To analyse the frequency response of the 'Big Muff' tone control section, the circuit is redrawn in Figure 4.12 to prepare for writing the circuit in the matrix form required by the nodal analysis method. The potentiometer has been divided into two separate resistors and the input voltage is split into two current sources to feed both input branches of the circuit.

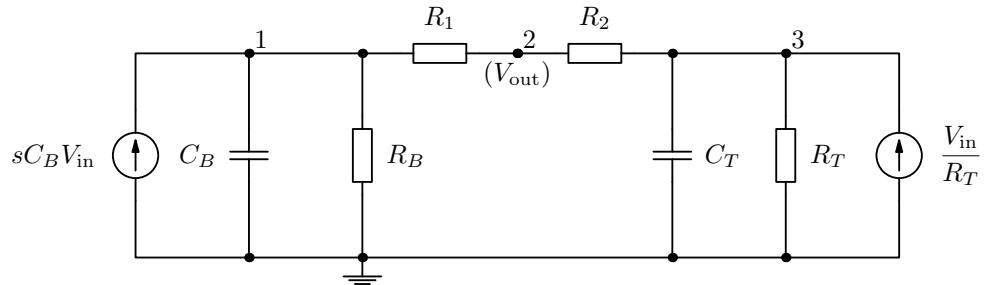


Figure 4.12: The 'Big Muff' tone control network redrawn for nodal analysis

The circuit diagram of Figure 4.12 is represented by a matrix equation (4.2), from where all the node voltages and branch currents of the circuit can be solved either symbolically by using Cramer's rule or numerically by using Octave or Matlab.

$$\begin{bmatrix} \frac{1}{R_B} + \frac{1}{R_1} + sC_B & -\frac{1}{R_1} & 0 \\ -\frac{1}{R_1} & \frac{1}{R_1} + \frac{1}{R_2} & -\frac{1}{R_2} \\ 0 & -\frac{1}{R_2} & \frac{1}{R_2} + \frac{1}{R_T} + sC_T \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} sC_B V_{in} \\ 0 \\ \frac{V_{in}}{R_T} \end{bmatrix} \quad (4.2)$$

Since the matrix equation for this circuit is not tremendously huge, the transfer function can be processed as a pen-and-paper calculation using Cramer's rule to solve the expression for the output voltage. After solving the voltage at node 2, the explicit expression for the 'Big Muff' transfer function $H(s) = \frac{V_2}{V_{\text{in}}}$ as a function of the Laplace variable s becomes

$$\frac{s^2 C_T C_B R_2 R_T + s C_B (R_1 + R_2 + R_T) + \frac{R_1}{R_B} + 1}{s^2 C_T R_T C_B (R_2 + R_1) + s \left[C_T R_T \left(1 + \frac{R_1 + R_2}{R_B} \right) + C_B (R_1 + R_2 + R_T) \right] + \frac{R_1 + R_2 + R_T}{R_B} + 1}$$

To use the transfer function for simulating the frequency response with a sine wave input signal, one can make the substitution $s = j\omega$ as explained in section 1.2.6. Suitable component values for practical guitar output signal filtering are indicated in the 'Big Muff' schematic 4.11. Simulations with these component values result in a filter function that changes from a low-pass to a high-pass filter as a function of the voltage divider resistors R_1 and R_2 . These resistors together form the potentiometer R_0 in Figure 4.11.

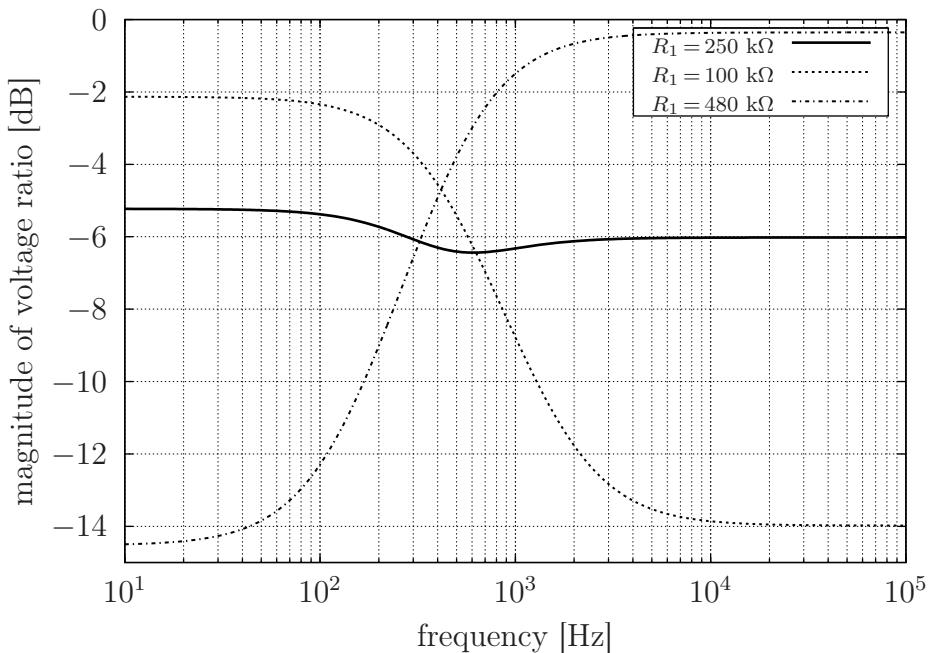


Figure 4.13: Frequency response curves of the 'Big Muff' tone circuit

The frequency response of the 'Big Muff' tone control section indicates that the circuit offers quite symmetric low-pass and high-pass filtering depending on the resistance of the potentiometer R_0 . However, when measured in decibels, the attenuation at the stop-band is not that deep compared to the centre value of

–6 dB. When the potentiometer is at the middle position, the circuit passes all frequencies with equal attenuation.

Until now, the 'Big Muff' tone control section has been analysed as a separate circuit element without embedding the circuit as a part of a bigger design. In the original 'Big Muff' schematic, the tone section is isolated between two amplifier sections that do not cause a significant impedance loading to the tone control section. The situation is different if the 'Big Muff' tone control stage is used in a guitar. Because the magnetic pickup of an electric guitar is a passive component, it does not provide ideal input and output impedances to connect to the following circuit stages. It is therefore expected that the filtering properties of the 'Big Muff' tone control section will be changed when connected directly after a magnetic pickup.

Figure 4.14 shows a circuit diagram where the 'Big Muff' tone control is connected after the equivalent circuit of a magnetic pickup. A voltage divider to simulate the volume control is also included in the circuit diagram after the 'Big Muff' tone control section. The equivalent circuit for nodal analysis purposes is not drawn separately, since it is easily modified from the schematic of Figure 4.14.

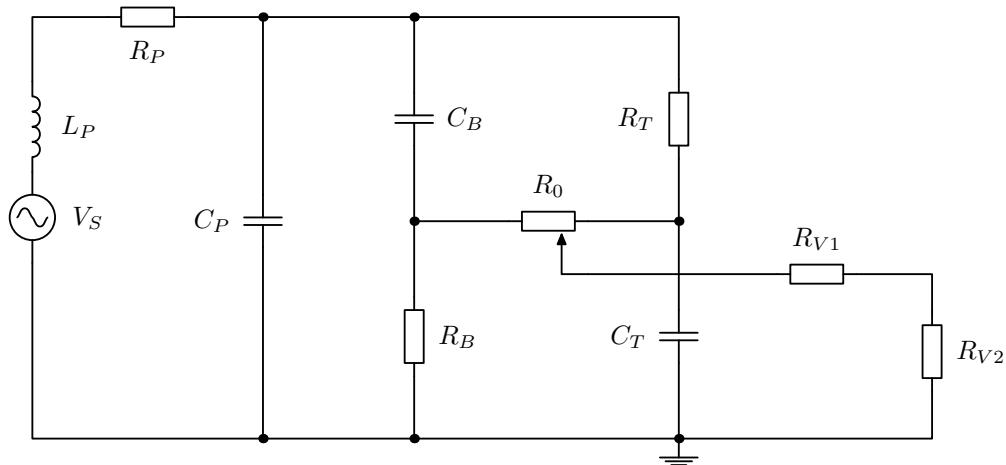


Figure 4.14: The 'Big Muff' tone control circuit connected to pickup

Due to the direct connection to the magnetic pickup, the filtering characteristics of the circuit change quite a lot as indicated in Figure 4.15. The magnetic pickup seems to add in its low-pass characteristics so that the high-frequency cut-off of the whole circuit combination is visibly increased. Otherwise at the low frequencies the circuit behaves as the separated 'Big Muff' circuit. To enhance the

functionality of the tone control, it should be isolated between two amplifier sections as in the original 'Big Muff' design. This would require battery operated active electronics to be mounted onto the guitar, which is not a tempting idea.

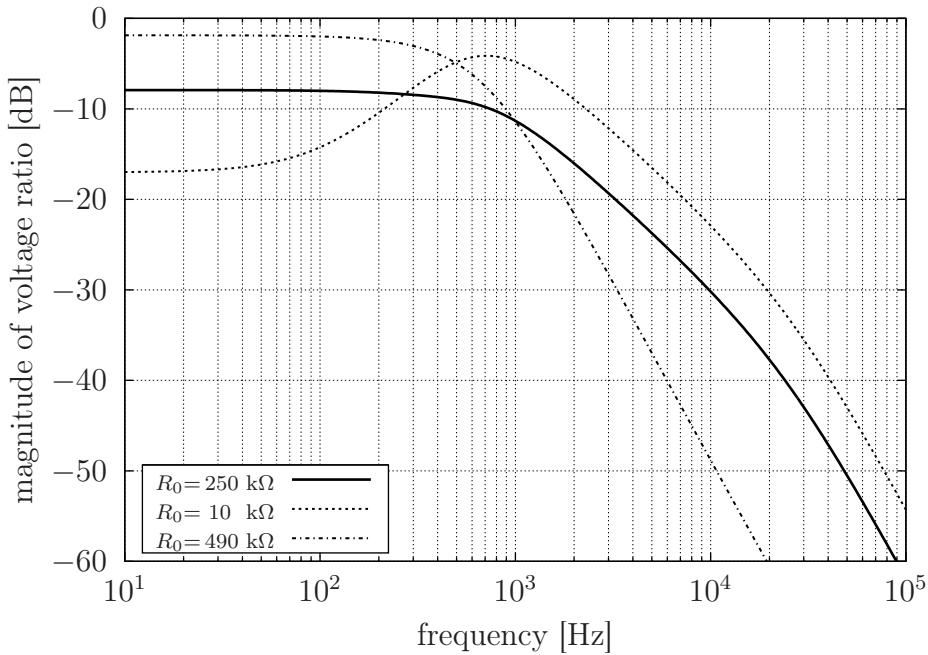


Figure 4.15: Frequency response curves of the 'Big Muff' tone control circuit when connected to the pickup

4.4 AMPLIFIER TONE CONTROLS FOR GUITAR MOUNTING

A random assortment of different guitar amplifier tone control circuits is presented here in the light that the separated tone control sections could be used as the guitar's tone control system. In the following analysis, the selected tone control circuits are taken through a frequency response simulation in SPICE to visualise the control properties of the circuits. This section therefore offers an overview of different tone control solutions used in commercial designs. The input voltage to the circuits is taken directly from an ideal voltage source, which means that if these circuits are built after the magnetic pickup on the guitar side, their filtering characteristics are bound to change. This was the case with the 'Big Muff' tone control presented in the previous section.

Some of the old classic amplifiers have been so simple that tone control circuits consisting of only one potentiometer have been designed. One of these 'one-tone-pot' guitar amplifiers is a small 'Moonlight' tube amplifier designed by Simcha

Delft. The schematic of this amplifier have been published as a do-it-yourself project and it is not to be taken as a commercial amplifier. The tone control section used in 'Moonlight' has been extracted to Figure 4.16. In this circuit diagram, potentiometer R_T is used to control the tone and potentiometer R_V is only used to control the volume. The circuit has many similarities with the basic guitar tone control circuit. From this perspective, it could be directly tested as a guitar tone control system.

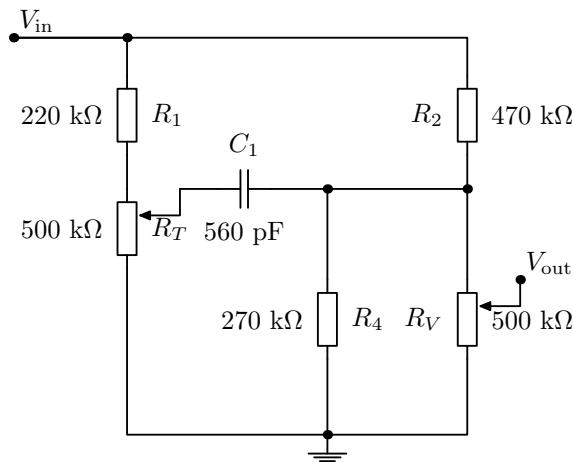


Figure 4.16: The tone control circuit of the 'Moonlight' amplifier

The frequency response curves as a function of the tone potentiometer R_T of the 'Moonlight' tone control section are drawn in Figure 4.17. The circuit seems to be a basic low-pass filter, but at the other end of the scale the circuit has the ability to change itself to a state of slight high-frequency boost. Clearly this tone control solution has a more variable frequency response than the most basic tone control circuit in an electric guitar. However, the circuit seems to cause significant attenuation to the signal, which is not good when considering the low signal levels obtained from a magnetic pickup.

Figure 4.18 shows a so-called 'Baxandall' tone control circuit, which is a general design commonly used in several audio amplifier implementations. The circuit has been published in a magazine article [59] in 1949 by E. J. James, but a feasible implementation of this circuit was given by P. J. Baxandall [60] a few years later. This implementation uses two tone control potentiometers, the one for bass control (R_B) and the other for treble control (R_T). The output voltage is taken between the two potentiometers. The limiting cases of bass-cut and treble-cut in the frequency response of the 'Baxandall' tone control section are depicted in Figure 4.19. The tone control limits are simulated having the other

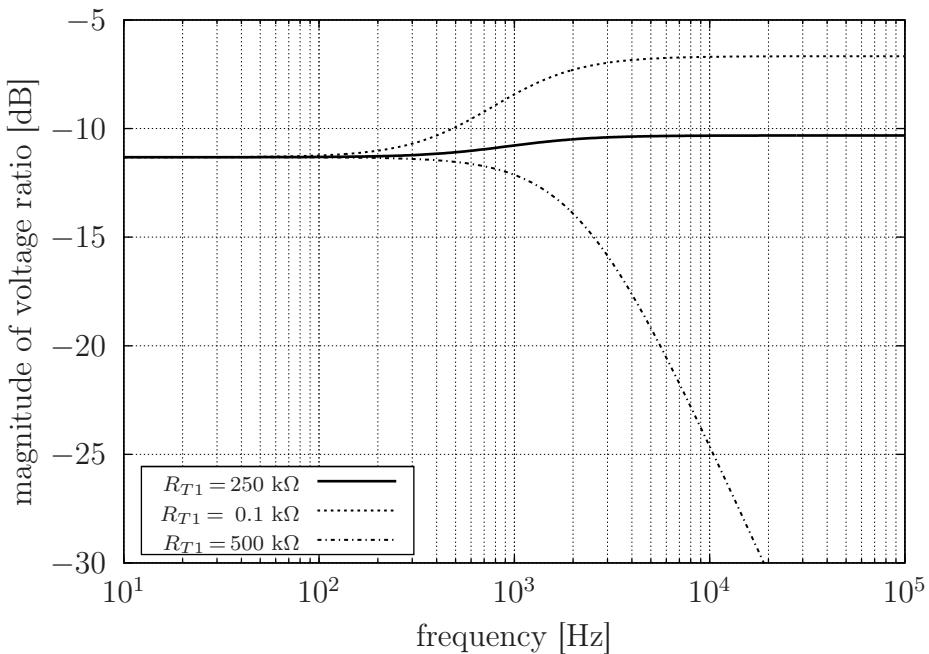


Figure 4.17: Moonlight frequency response curves with a changing value of R_T

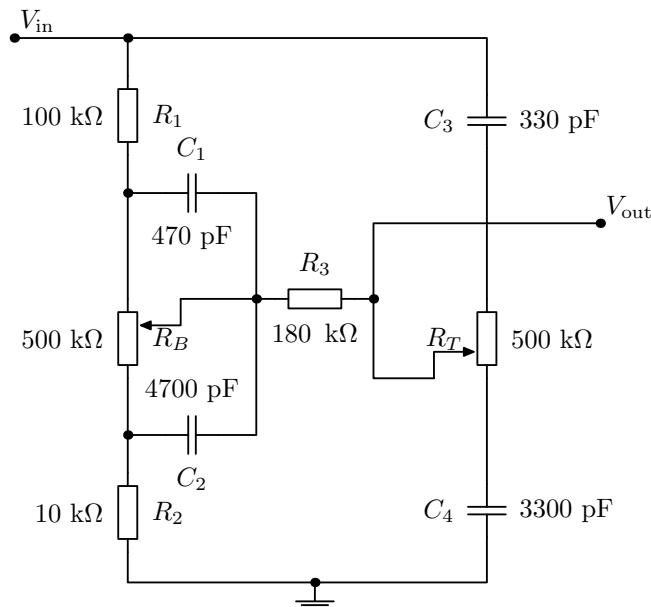


Figure 4.18: The Baxandall tone control circuit

potentiometer at a limiting position while the other is at the middle position. The attenuation provided by the 'Baxandall' circuit in the bass-cut and treble-cut modes is much more effective compared to the single-potentiometer solutions of 'Moonlight' and 'Big Muff'. The response curves are also reasonably symmetric around the 'notch' frequency when examining the frequency response in the logarithmic scale.

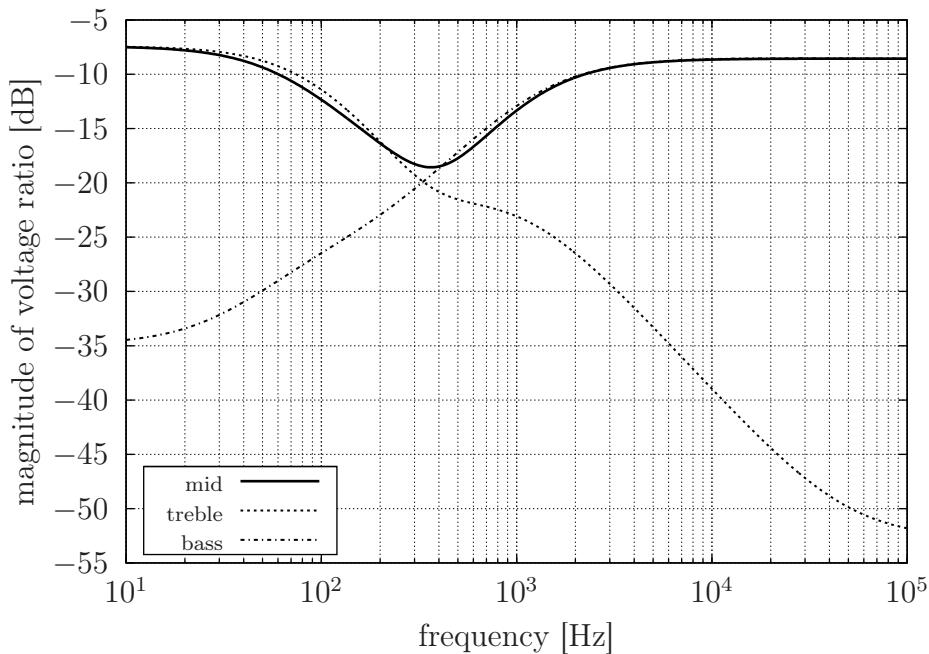


Figure 4.19: Baxandall frequency response curves with potentiometers R_B and R_T at the middle, bass-cut and treble-cut positions

The schematic in Figure 4.20 shows the tone control section of an old Fender Pro 6G5 amplifier. This implementation is relatively similar to the 'Baxandall' tone control section but it is implemented using less components. The two potentiometers are separated for controlling bass R_B and treble R_T .

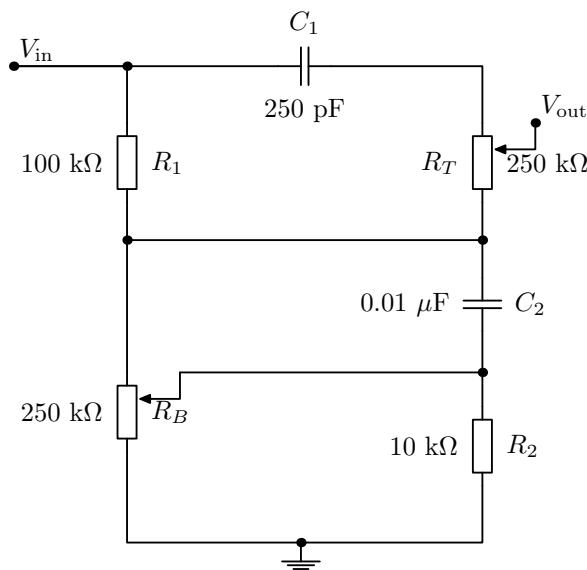


Figure 4.20: The Fender Pro 6G5 tone control circuit

The limiting cases of bass-cut and treble-cut in the frequency response of the

Fender Pro 6G5 tone control section are depicted in Figure 4.21. The tone control limits are simulated having one potentiometer at a limiting position while the other one is at the middle position. The treble-cut and the bass-cut response curves cross very symmetrically at around 1 kHz and the middle position setting of both potentiometers gives a small notch filter kind of a response. If the two potentiometers are simultaneously in extreme positions, a nice band-pass filter response is obtained. Similar response curves were also achieved with one potentiometer in the 'Big-Muff' implementation.

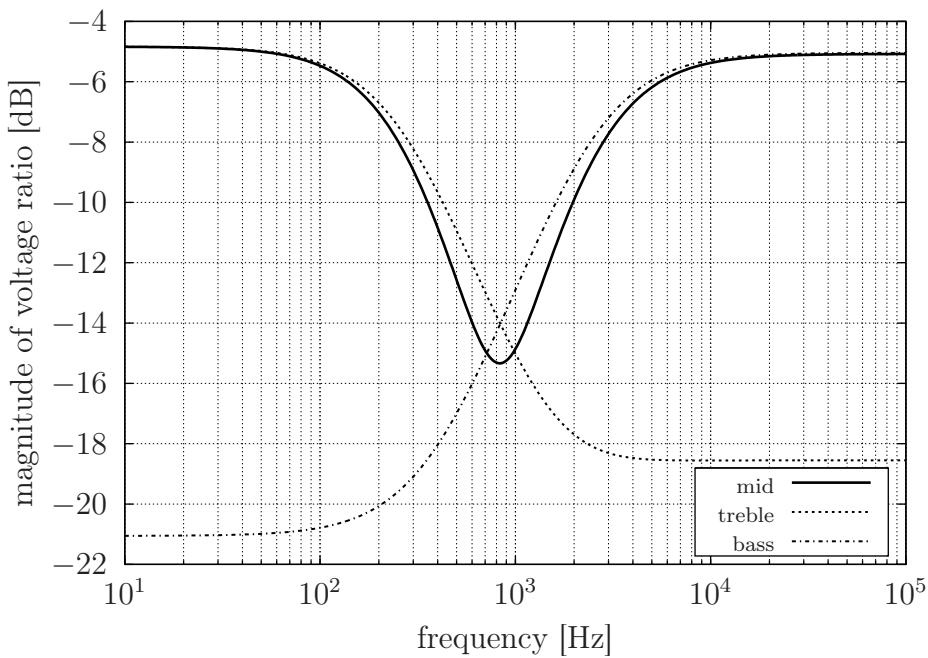


Figure 4.21: Fender Pro 6G5 frequency response curves with potentiometers R_B and R_T at the middle, bass-cut and treble-cut positions

CHAPTER 5

GUITAR CABLES

The majority of technical solutions in every sector of industry use cables to carry signals between electrical devices and to connect different electrical interfaces. In general, cables are electric components which can be described by common mathematical models and equivalent circuits. The term *lumped element model* in the context of cables and transmission lines in general is used to indicate that the model is constructed from a group of small identical elements.

As guitar accessories, cables are mostly annoying. They are always laying around on the floor at the players' feet trying to get them to stumble and injured. After a gig, the maintenance crew will get tennis elbows from the effort of sorting out a truckload of cables to neat and tidy rolls. Ever so often, the cables decide to break a wire and leave for early retirement. But does anyone really know how much cables really affect the tone and what are the properties of the cables that need to be looked out for to get that killer tone? Read on and find out what is inside a basic instrument cable and how to model it mathematically.

5.1 THE GENERAL TRANSMISSION LINE THEORY

Figure 5.1 describes a general equivalent model for a transmission line, commonly known as a cable. A mathematical simulation model divides the cable into small pieces of length δx , so that a sum of these small pieces eventually builds up the whole cable. Each individual piece consists of inductance L' , resistance R' , capacitance C' and conductance G' . All of these units are expressed as per unit length, meaning that $R'\delta x = R$, where R is the resistance of the cable with length δx . [61, pp. 10–15]

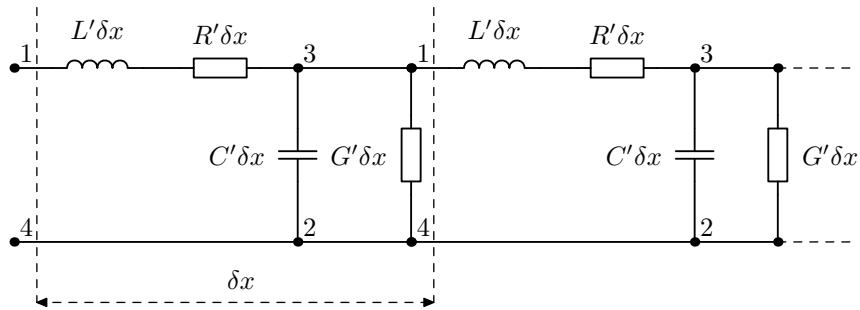


Figure 5.1: An equivalent circuit for a general transmission line

From Figure 5.1, relations for voltage and current can be written as

$$V(x + \delta x, t) - V(x, t) = \frac{\partial V}{\partial x} \delta x = -R' \delta x I - L' \delta x \frac{\partial I}{\partial t} \quad (5.1)$$

$$I(x + \delta x, t) - I(x, t) = \frac{\partial I}{\partial x} \delta x = -G' \delta x V - C' \delta x \frac{\partial V}{\partial t}, \quad (5.2)$$

from where the infinitesimal lengths δx cancel out from both sides. By differentiating the differential equation for voltage with respect to x and the equation for current with respect to time,

$$\frac{\partial}{\partial x} \frac{\partial V}{\partial x} = \frac{\partial^2 V}{\partial x^2} = -R' \frac{\partial I}{\partial x} - L' \frac{\partial^2 I}{\partial x \partial t} \quad (5.3)$$

$$\frac{\partial}{\partial t} \frac{\partial I}{\partial x} = \frac{\partial^2 I}{\partial x \partial t} = -G' \frac{\partial V}{\partial t} - C' \frac{\partial^2 V}{\partial^2 t}. \quad (5.4)$$

When substituting equations (5.4) and (5.2) into equation (5.3), a general partial differential equation to describe an infinitesimal piece of a transmission line is obtained in the form

$$\frac{\partial^2 V}{\partial x^2} = R' G' V + (R' C' + G' L') \frac{\partial V}{\partial t} + L' C' \frac{\partial^2 V}{\partial^2 t}. \quad (5.5)$$

It is not relevant to start solving this equation since there are so many different possibilities to choose for special initial and boundary conditions. However, it is possible to analyse equation (5.5) in the frequency domain by applying the Laplace transform to equation (5.5) with respect to time [11, pp 594–595]. Taking the transform first from the left side of the equation gives

$$\begin{aligned} \mathcal{L} \left\{ \frac{\partial^2 V(x, t)}{\partial x^2} \right\} &= \int_0^\infty e^{-st} \frac{\partial^2 V(x, t)}{\partial x^2} dt = \frac{\partial^2}{\partial x^2} \int_0^\infty e^{-st} V(x, t) dt = \\ &\quad \frac{\partial^2}{\partial x^2} \mathcal{L}\{V(x, t)\} = \frac{\partial^2}{\partial x^2} V(x, s), \end{aligned}$$

assuming that it is allowed to change the order of integration and differentiation. For the right side of equation (5.5) the Laplace transform yields

$$\begin{aligned} R'G'\mathcal{L}\{V(x, t)\} + (R'C' + G'L')[s\mathcal{L}\{V(x, t)\} - V(x, 0)] + \\ L'C'[s^2\mathcal{L}\{V(x, t)\} - sV(x, 0) - V_t(x, 0)], \end{aligned}$$

and if the initial conditions $V(x, 0)$ and $V_t(x, 0)$ are taken to be zero, the final form of the transformed equation reads

$$\frac{\partial^2 V(x, s)}{\partial x^2} = [R'G' + (R'C' + G'L')s + L'C's^2]V(x, s). \quad (5.6)$$

The standard auxiliary equation defined for second order differential equations gives the roots

$$r_{1,2} = \pm \sqrt{(R' + sL')(G' + sC')},$$

and this yields a general solution of the form

$$V(x, s) = Ae^{-\gamma x} + Be^{\gamma x}, \quad (5.7)$$

where the propagation constant $\gamma = \sqrt{(R' + sL')(G' + sC')}$. If the propagational properties of a sinusoidal signal in the transmission line are investigated, one can make the substitutions for $s = j\omega$ and $\gamma = \sqrt{(R' + j\omega L')(G' + j\omega C')}$.

Impedance is generally defined as the ratio of voltage and current V/I . To reach the expression of general transmission line impedance, one can also apply the Laplace transform to equation (5.1), which leads to

$$\frac{\partial V(x, s)}{\partial x} = -(R' + sL')I(x, s). \quad (5.8)$$

By substituting the result (5.7) into (5.8), performing the differentiation of $V(x, s)$ and embedding the minus sign into the arbitrary (meaning not yet fixed) constants A and B , one has

$$\frac{\partial V(x, s)}{\partial x} = \gamma V(x, s) = \sqrt{(R' + sL')(G' + sC')}V(x, s) = (R' + sL')I(x, s).$$

The expression for the Laplace domain voltage $V(x, s)$ is now $A'e^{-\gamma x} + B'e^{\gamma x}$, where $A' = A$ and $B' = -B$ because of the embedding of the minus sign. This does not have any significance at this point since the constants have not been defined against reasonable boundary conditions. In any case, the result for the transmission line impedance Z_t reduces to

$$Z_t = \frac{V(x, s)}{I(x, s)} = \sqrt{\frac{R' + sL'}{G' + sC'}},$$

which in the case of sinusoidal signals can be written as

$$Z_t = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}}. \quad (5.9)$$

The problem can also be approached from the viewpoint of Maxwell's equations

$$\nabla \bullet \vec{D} = \rho_f \quad (5.10)$$

$$\nabla \bullet \vec{B} = 0 \quad (5.11)$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (5.12)$$

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}, \quad (5.13)$$

which can be used to describe almost all phenomena in the field of electromagnetism. All the field vectors \vec{D} , \vec{B} , \vec{E} and \vec{H} are functions of position \vec{r} and time t , although this is not explicitly indicated in the following equations.

The equation of motion for an electromagnetic wave can be derived from Maxwell's equations. Especially in metal substances the relation $\vec{J} = \sigma \vec{E}$ holds. By also using the equalities $\vec{D} = \epsilon \epsilon_0 \vec{E}$ and $\vec{H} = \vec{B}/\mu \mu_0$, equation (5.13) becomes

$$\frac{1}{\mu \mu_0} \nabla \times \vec{B} = \sigma \vec{E} + \epsilon \epsilon_0 \frac{\partial \vec{E}}{\partial t}. \quad (5.14)$$

In a similar fashion as in the derivation of transmission line equations, one uses equations (5.12) and (5.13) as an equation pair to derive the wave impedance of electromagnetic waves. The wave impedance is defined as the ratio \vec{E}/\vec{H} by a common agreement. For starters, the equation (5.12) is differentiated with respect to location using the relation

$$\nabla \times \nabla \times \vec{E} = -\nabla^2 \vec{E} + \nabla(\nabla \bullet \vec{E}),$$

where the latter term is zero because the analysis focuses on conducting materials. Therefore, equation (5.12) becomes

$$\nabla \times \nabla \times \vec{E} = -\nabla^2 \vec{E} = -\frac{\partial}{\partial t}(\nabla \times \vec{B}),$$

and with equation (5.14),

$$\nabla^2 \vec{E} = \epsilon \epsilon_0 \mu \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2} + \mu \mu_0 \sigma \frac{\partial \vec{E}}{\partial t}. \quad (5.15)$$

Similarly as in the case of transmission line derivation, the Laplace transform of equation (5.15) yields

$$\nabla^2 \vec{E}(s) = (\epsilon \epsilon_0 \mu \mu_0 s^2 + \mu \mu_0 \sigma s) \vec{E}(s), \quad (5.16)$$

where the notation $\vec{E}(s)$ has been used to distinguish the Laplace transformed vector field $\mathcal{L}\{\vec{E}\} = \vec{E}(s)$. The auxiliary equation gives the roots

$$r_{1,2} = \pm \sqrt{s\mu\mu_0(s\epsilon\epsilon_0 + \sigma)},$$

so that the general solution to equation (5.16) is

$$\vec{E}(s) = Ce^{-\lambda\vec{r}} + De^{\lambda\vec{r}}, \quad (5.17)$$

where $\lambda = \sqrt{s\mu\mu_0(s\epsilon\epsilon_0 + \sigma)}$.

Applying the Laplace transform also to equation (5.12) gives

$$\nabla \times \vec{E}(s) = -s\vec{B}(s) = -s\mu\mu_0\vec{H}(s),$$

and after substituting (5.17) and performing the differentiation in a similar fashion as in the case of transmission line analysis,

$$\nabla \times \vec{E}(s) = \lambda\vec{E}(s) = s\mu\mu_0\vec{H}(s).$$

This gives an expression for the wave impedance as

$$Z_w = \frac{\vec{E}}{\vec{H}} = \sqrt{\frac{s\mu\mu_0}{\sigma + s\epsilon\epsilon_0}},$$

and in the case of sinusoidal signals one can set $s = j\omega$, so that

$$Z_w = \sqrt{\frac{j\omega\mu\mu_0}{\sigma + j\omega\epsilon\epsilon_0}}.$$

Comparing the expressions of transmission line impedance Z_t and electromagnetic wave impedance Z_w implies that $C' = \epsilon\epsilon_0$ and $L' = \mu\mu_0$. This follows from the assumption of a lossless cable, for which the impedance equations reduce to the forms

$$Z_t = \sqrt{\frac{L'}{C'}}, \quad Z_w = \sqrt{\frac{\mu\mu_0}{\epsilon\epsilon_0}},$$

based on the assumption that one can set $Z_t = Z_w$. This is a quite significant result because now if only a general expression Z_0 with the parameters of ϵ and μ is known for some specific type of cable, it is then possible to find out the inductance per unit length and capacitance per unit length via relations

$$C' = \frac{\sqrt{\mu\mu_0\epsilon\epsilon_0}}{Z_0} \quad \text{and} \quad L' = Z_0\sqrt{\mu\mu_0\epsilon\epsilon_0}.$$

This obviously leads to the connection

$$L' = \frac{\mu\mu_0\epsilon\epsilon_0}{C'} \quad (5.18)$$

between capacitance and inductance. This equivalence will have practical significance later.

5.2 PRACTICAL GUITAR CABLES

The main transmission line types used to build guitar cables are the coaxial type and the parallel wire type. Guitar cables that imitate the construction of coaxial cables are commonly referred to as shielded instrument cables. These cables are not manufactured according to such official specification standards that the regular coaxial cables follow but they can still be categorised as high quality cables. The parallel wire cable cannot even be considered as a serious instrument cable, but still the poor musician might fall for the temptation of a cheaper priced cable at the cost of quality. At least the poor author has fallen for it. It is a good enough cable if only the sound comes through it, isn't it? Isn't it really?

5.2.1 A shielded instrument cable

Coaxial cables are often used as measurement cables in laboratories because they have a proper shielding and a clearly defined characteristic impedance. The general construction of a coaxial cable is imitated in musical instrument cables to produce a rigidly shielded cable that is more flexible and adapts to the terminating impedances of an assortment of musical accessories. Electric guitar cables which are referred to as shielded instrument cables have a construction similar to the coaxial cable. A typical construction of a shielded guitar cable is depicted in Figure 5.2. This model is adapted from the datasheets provided by the instrument cable manufacturer DiMarzio [62].

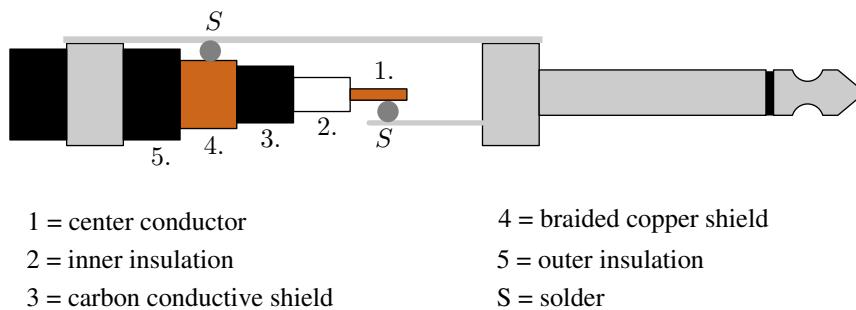


Figure 5.2: Construction of a high class shielded guitar cable

It depends largely on the manufacturer how thoroughly the cables are documented. Quite often no data is given about the electrical characteristics of the cables, although in the case of shielded cables, it is possible to manufacture the cables according to certain specifications that determine the essential characteristics. For high quality cables there is usually information available on the resistance and the capacitance of the cable. This information is enough for de-

termining the characteristic impedance since the theoretical impedance can be deduced from the capacitance value with the help of equation (5.18).

At least one patent application [63] has been submitted for a shielded instrument cable, where the construction of the cable has been described in great detail and the effects of each layer in the cable are discussed thoroughly. The patent application document also lists measured capacitances and other electrical characteristics of several commercially available guitar cables.

The most crucial electrical property in a guitar cable seems to be the capacitance of the cable because it modifies the resonance frequency of the pickup. To analytically determine the characteristic impedance of a coaxial cable, the expression for either the capacitance or the inductance needs to be determined. The capacitance and the inductance of a coaxial cable can be calculated directly from the dimensions of the cross-sectional layout of the cable. The main parameters to specify the essential dimensions of the cable are shown in Figure 5.3.

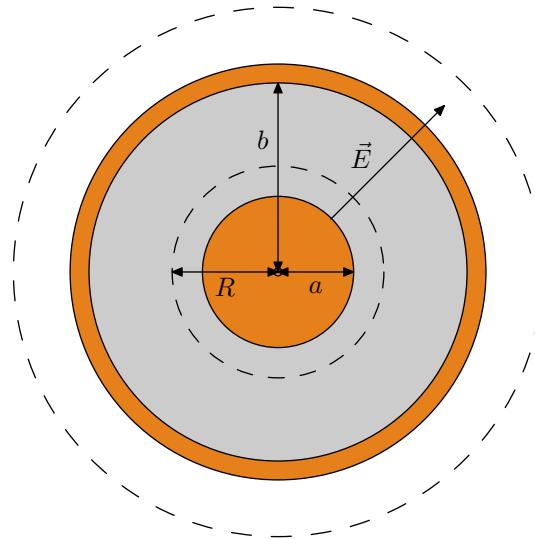


Figure 5.3: A cross-sectional view of a basic coaxial cable

The derivation of the capacitance equation starts from the divergence theorem of Gauss. An application of this theorem states that

$$\int_S \vec{D} \bullet d\vec{S} = \int_V \nabla \bullet \vec{D} dV,$$

saying that the flux of the electric displacement vector \vec{D} through a surface equals the divergence of \vec{D} within volume V . The divergence is related to free charge density ρ_f via the relation

$$\nabla \bullet \vec{D} = \rho_f,$$

so that

$$\int_S \vec{D} \bullet d\vec{S} = \int_V \rho_f dV, \quad (5.19)$$

where the volume integral now equals the total (net) free charge enclosed within surface S . If there is no net charge inside the volume V , as is the case for an electrically neutral piece of matter, the result of the volume integral equals zero. The volume integral of net free charge is also zero when evaluated within a conductor under the influence of an external electric field. The idea would now be to examine the charges related to the inner conductor of the coaxial cable with a radius a .

The total electric field inside a volume V enclosed by some arbitrary boundaries is defined not only by the volume charge density ρ , since the surface charge density σ is also contributing to the total charge. The volume charge density ρ refers to the charges inside a volume and the surface charges are contributions from induced charges on the surface boundary, which cannot be taken as a part of the volume V . Therefore, the total charge is the sum

$$\int_V \rho_f dV + \int_S \sigma_f dS,$$

so that equation (5.19) becomes

$$\int_S \vec{D} \bullet d\vec{S} = \int_V \rho_f dV + \int_S \sigma_f dS.$$

When the coaxial cable is used to carry signals, an external voltage is applied between the terminals of the cable. The applied voltage induces surface charges to the surface of the inner conductor (with radius a), while the electric field \vec{E} inside the inner conductor is zero. From this it follows that the total free volume charge ρ_f inside the conductor is zero, and the only contribution to the net charge comes from the surface charge density, i.e.

$$\int_S \vec{D} \bullet d\vec{S} = \int_S \sigma_f dS. \quad (5.20)$$

The cross-sectional geometry of a general coaxial type cable is a cylinder with some arbitrary constant radius, noted as R in this context. This geometrical shape has a parametric representation

$$\vec{r}(R, \theta, z) = R \cos \theta \hat{i} + R \sin \theta \hat{j} + z \hat{k}, \quad (5.21)$$

so that the volume element from the Jacobian matrix yields

$$dV = \begin{bmatrix} \cos \theta dR & \sin \theta dR & 0 \\ -R \sin \theta d\theta & R \cos \theta d\theta & 0 \\ 0 & 0 & dz \end{bmatrix} = R dR d\theta dz. \quad (5.22)$$

The surface element vector for a cylinder with a constant radius R is

$$d\vec{S} = \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ -R \sin \theta d\theta & R \cos \theta d\theta & 0 \\ 0 & 0 & dz \end{bmatrix} = R \cos \theta d\theta dz \hat{i} + R \sin \theta d\theta dz \hat{j}, \quad (5.23)$$

with a direction to the radial component of the cylinder and normal to the surface. The magnitude of this surface element is

$$|d\vec{S}| = \sqrt{d\vec{S} \bullet d\vec{S}} = R d\theta dz. \quad (5.24)$$

The electric displacement vector \vec{D} is also normal to the surface, so that for a unity length ($z \in [0, 1]$) piece of cable

$$\int_S \vec{D} \bullet d\vec{S} = \int_S D dS = \int_0^1 \int_0^{2\pi} D R d\theta dz,$$

and since \vec{D} depends only on R , the integral evaluates to

$$DR \int_0^1 \int_0^{2\pi} d\theta dz = D2\pi R. \quad (5.25)$$

Referring again to the dimensions given in Figure 5.3, the surface charge density is located at a distance of a constant radius a from the centre of the cable and the charge is assumed to be evenly distributed over the whole area. With these definitions applied to a unity length ($z \in [0, 1]$) piece of cable, the integral on the right side of equation (5.20) yields an expression for the total charge

$$Q = \sigma_f a \int_0^1 \int_0^{2\pi} d\theta dz = \sigma_f 2\pi a, \quad (5.26)$$

so that from equation (5.20) the electric displacement vector

$$D(R) = \frac{a\sigma_f}{R} \quad \Rightarrow \quad E(R) = \frac{a\sigma_f}{\epsilon\epsilon_0 R}.$$

To calculate the capacitance of the coaxial cable directly from the definition of the capacitance $C = Q/V$, the potential difference between the two conductors in the cable needs to be evaluated. This can be calculated as

$$V = \phi(a) - \phi(b) = - \int_b^a \vec{E} \bullet d\vec{l} = - \int_b^a \frac{a\sigma_f}{\epsilon\epsilon_0 R} dR = \frac{a\sigma_f}{\epsilon\epsilon_0} \ln \left(\frac{b}{a} \right), \quad (5.27)$$

since the electric field vector \vec{E} is pointing towards the same direction as $d\vec{l}$, which is equal to dR by equation (5.21).

Using equations (5.26) and (5.27), the capacitance per unit length of a coaxial cable

$$C = \frac{Q}{V} = \frac{\sigma_f 2\pi a}{\frac{a\sigma_f}{\epsilon\epsilon_0} \ln\left(\frac{b}{a}\right)} = \epsilon\epsilon_0 \frac{2\pi}{\ln\left(\frac{b}{a}\right)}. \quad (5.28)$$

Although the units do not indicate the length dimension, the result can be taken to be per unit length because of the unity integration made in equation (5.26). A better way to approach this would have been to leave the integration undone at that phase. A shortcut via equation (5.18) gives the inductance per unit length of the coaxial cable

$$L = \mu\mu_0 \frac{\ln\left(\frac{b}{a}\right)}{2\pi}.$$

When the capacitance, inductance and resistance are known, one can apply equation (5.9) to calculate the characteristic impedance of the cable. [64, pp. 74–75 & 331–332]

The equation (5.28) for coaxial cable capacitance holds exactly only for a cable where the centre conductor is built from one solid wire and the outer shield is a uniform piece of metal foil or equivalent. To get more flexibility to the instrument cables, the centre conductor is often made of several wires wound together. The effective radius a is defined in this situation via approximate multipliers as

$$\begin{aligned} a &= 2.84r \quad (n = 7) & a &= 3.99r \quad (n = 12) \\ a &= 4.90r \quad (n = 19) & a &= 6.86r \quad (n = 37), \end{aligned}$$

where r is the geometrical radius of a single centre conductor wire and n is the number of individual wires wound together. Similarly there are approximate multipliers for the outer conductor if the conductor is not a uniform piece of metal. In guitar cables the outer shield is often braided or served to further increase the flexibility of the cable. In these special cases the effective radius b becomes

$$b = b + 1.5r \quad (\text{braided}) \quad ; \quad b = b + 0.8r \quad (\text{served}),$$

where r is the geometrical radius of a conductor wire that the shield is wound from. [63]

Additionally the patent application [63] strives to minimise the capacitance per unit length for the benefit of increasing the bandwidth. All of the proposed methods relate to altering some variable in equation (5.28), but every aspect modified in that equation has its advantages and disadvantages. A reduction

of the dielectric constant leads to more expensive materials and manufacturing processes. An increase of the effective radius of the outer shield also adds more material costs and can only be done within certain limits due to the available connector sizes. A reduction of the effective radius of the centre conductor leads to more breakable cables. The resulting cable measures as a compromise of all details mentioned above.

To enhance the durability of the cable, the patent application [63] defines a measure of braid angle, which is composed of several variables related to the composition of the braided shield. By changing some parameters in the shield construction, it is possible to enhance the overall strength and stiffness of the cable. Breakability is indeed a big problem in guitar cables and the enhanced braid shielding is one way to improve the durability of the cable.

5.2.2 A parallel wire cable

Parallel wire cables are used for transmitting signals for relatively short lengths. The wires are often twisted around each other to minimise external electric and magnetic fields around them. Figure 5.4 shows a basic junction of a parallel wire cable and the 6.3 mm guitar plug revealing the inner construction of the cable. If the parallel wire transmission lines are being used as instrument cables, they

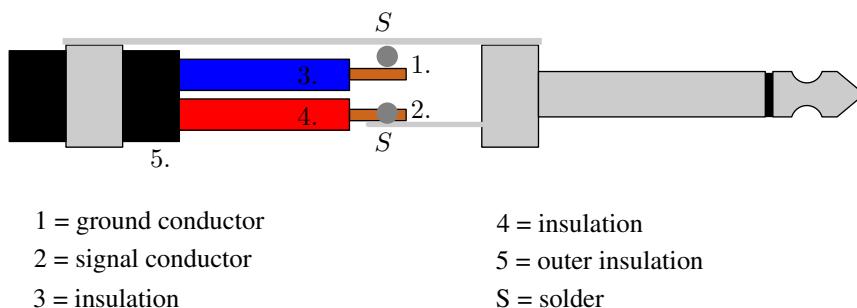


Figure 5.4: Construction of a parallel wire guitar cord

are not considered as high quality as the coaxial type transmission lines. This is because the coaxial construction offers better shielding from external disturbance signals and has more clearly defined electrical characteristics.

If the characteristic impedance of the parallel wire cable is calculated, firstly one needs to determine the expression for either the capacitance or the inductance of the geometrical construction. To calculate the capacitance per unit length of a parallel wire cable, a special method of 'image charges' is used. This method works only if the equipotential lines of the linearised charges have an identical

shape with the conductor. This requirement is fulfilled nicely for the cylindrical shaped conductors of a parallel wire cable.

Figure 5.5 depicts the geometrical variables needed to derive a mathematical model for the cable. The starting point for the analysis is to reduce the sur-

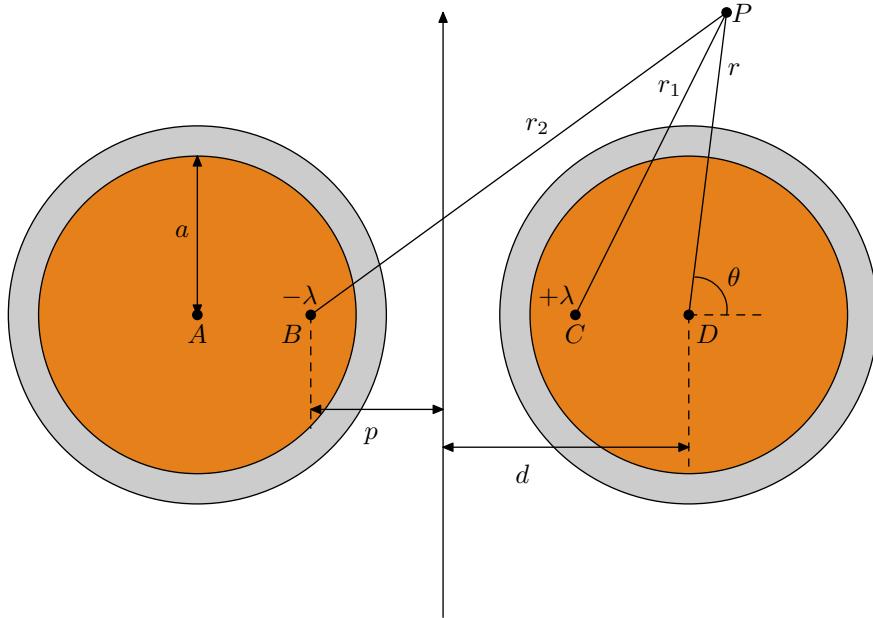


Figure 5.5: Geometry of a parallel wire transmission line

face charges on the conductor to line charges $-\lambda$ and $+\lambda$ inside the conductor. These line charges are shown as points B and C in Figure 5.5. The fictitious line charge is called the 'image' of the real surface charge of the conductor. Gauss's divergence theorem in this context is

$$\int_S \vec{D} \bullet d\vec{S} = \oint \lambda dl.$$

where the left side evaluates to $D2\pi R$ as in equation (5.25). The right side is the total charge per unit length, which is now calculated from a constant line charge density

$$Q = \lambda \int_0^1 dz = \lambda.$$

From here it is evident that

$$D(R) = \frac{\lambda}{2\pi R} \quad \text{and} \quad E(R) = \frac{\lambda}{2\pi\epsilon_0 R}, \quad (5.29)$$

and the potential ϕ is related to the electric field \vec{E} by

$$\vec{E} = -\nabla\phi, \quad (5.30)$$

so that integration of the left side of equation (5.29) gives

$$\int E(R) dR = \int -\nabla \phi(R) dR = -\phi(R).$$

The integration on the right side of equation (5.29) yields

$$\int \frac{\lambda}{2\pi\epsilon\epsilon_0 R} dR = \frac{\lambda}{2\pi\epsilon\epsilon_0} \ln(R) + \ln(C),$$

where $\ln(C)$ is the constant of integration. Hence, the potential

$$\phi(R) = -\frac{\lambda}{2\pi\epsilon\epsilon_0} \ln(R) + \ln(C) = \frac{\lambda}{2\pi\epsilon\epsilon_0} \ln\left(\frac{C}{R}\right). \quad (5.31)$$

In this expression $C > a$, where a is the radius of the conductor.

The line charges create a potential to point P in Figure 5.5, which according to equation (5.31) is

$$\begin{aligned} \phi_P &= \frac{\lambda}{2\pi\epsilon\epsilon_0} \ln\left(\frac{1}{r_1}\right) - \frac{\lambda}{2\pi\epsilon\epsilon_0} \ln\left(\frac{1}{r_2}\right) \\ &= \frac{\lambda}{4\pi\epsilon\epsilon_0} \ln\left(\frac{r_2^2}{r_1^2}\right). \end{aligned}$$

With the help of the law of cosines, from triangle BDP

$$r_2^2 = (d+p)^2 + r^2 + 2(d+p)\cos\theta = 2(d+p)(d+r\cos\theta) + r^2 + p^2 - d^2.$$

In the above expression for the law of cosines, the even function properties of $\cos(\pi-\theta) = -\cos\theta$ are used to simplify the expression. Similarly, for the triangle CDP

$$r_1^2 = (d-p)^2 + r^2 + 2(d-p)\cos\theta = 2(d-p)(d+r\cos\theta) + r^2 + p^2 - d^2.$$

The expressions for r_1^2 and r_2^2 can be substituted into the potential equation:

$$\phi_P = \frac{\lambda}{4\pi\epsilon\epsilon_0} \ln\left(\frac{2(d+p)(d+r\cos\theta) + r^2 + p^2 - d^2}{2(d-p)(d+r\cos\theta) + r^2 + p^2 - d^2}\right).$$

When the location of the image charges are chosen so that $p^2 = (d^2 - a^2)$ and the potential of point P is evaluated at $r = a$,

$$\phi_P(a) = \frac{\lambda}{4\pi\epsilon\epsilon_0} \ln\left(\frac{d+p}{d-p}\right),$$

and the potential difference between the two conductors is

$$\begin{aligned} V &= \phi_{P+}(a) - \phi_{P-}(a) = \frac{\lambda}{4\pi\epsilon\epsilon_0} \ln\left(\frac{d+p}{d-p}\right) - \frac{-\lambda}{4\pi\epsilon\epsilon_0} \ln\left(\frac{d+p}{d-p}\right) \\ &= \frac{\lambda}{2\pi\epsilon\epsilon_0} \ln\left(\frac{d+p}{d-p}\right). \end{aligned}$$

From here the capacitance can be calculated using the basic relation

$$C = \frac{Q}{V} = \frac{2\pi\epsilon\epsilon_0}{\ln\left(\frac{d+p}{d-p}\right)} = \frac{2\pi\epsilon\epsilon_0}{\ln\left[\frac{(d+p)(d+p)}{(d-p)(d+p)}\right]} = \frac{2\pi\epsilon\epsilon_0}{\ln\left[\frac{(d+p)^2}{d^2-p^2}\right]}.$$

Using the earlier definition of $p^2 = (d^2 - a^2)$, the expression for the capacitance per unit length becomes

$$C = \frac{\pi\epsilon\epsilon_0}{\ln\left[\frac{d + \sqrt{d^2 - a^2}}{a}\right]} = \frac{\pi\epsilon\epsilon_0}{\ln\left(\frac{d}{a} + \sqrt{\frac{d^2}{a^2} - 1}\right)}.$$

When the capacitance per unit length is known, one can use equation (5.18) to calculate the inductance per unit length and equation (5.9) to finally find out the characteristic impedance of a cable with a parallel wire construction. These formulae will only give rough estimates about the electrical properties of the cable since the parallel wires are often twisted around each other. The analytical expressions will be more accurate in the case of the coaxial cable, where the structure of the cable is more rigid and controllable.

Usually the characteristics of more difficult cable geometries are determined by measurements. The measurements are based on the theoretical fact that

$$Z_0 = \sqrt{(Z_{\text{open}})(Z_{\text{short}})},$$

which means that the characteristic impedance of the cable can be obtained by determining the input impedance of the cable when it is connected to a voltage source with the other end open (Z_{open}) and in short circuit (Z_{short}). Unfortunately this measurement is not easily accomplished by basic equipment, so the details for this procedure are not covered in this context.

5.3 IMPEDANCE MATCHING

The obvious practical application for cables is to use them to carry signals from one device to another. Every electrical device has its own input impedance to which the cable is connected to. This input impedance is seen as a load from the viewpoint of the output terminals of the other device and the cable.

The combination of the characteristic impedance Z_0 of the cable and the loading impedance Z_L generated by input impedance of some device affect the propagation of the signal in the cable. General theory states that the signal travelling in

the cable is completely transmitted to the device only when $Z_0 = Z_L$. If Z_0 has a different value from Z_L , part of the signal is reflected back from the junction between the cable and the input terminals of the device.

The depiction of the junction of the cable and the load impedance is given in Figure 5.6, where the common agreement seems to be that the junction is located at $x = 0$ and the beginning of the cable is at $x = -l$. The waves are generated and delivered from the beginning of the cable to the load, where they are either completely absorbed by the load impedance or partially reflected back towards the negative x -coordinate direction.

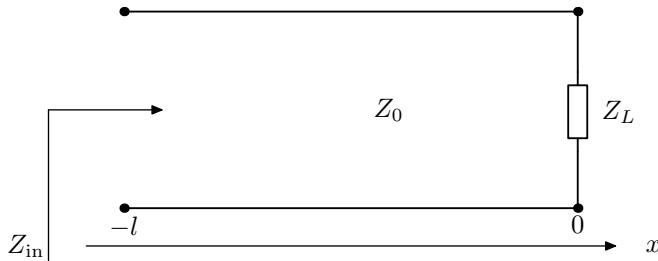


Figure 5.6: Junction of cable impedance Z_0 with load Z_L

The mathematical analysis of the situation is identical to the vibrating string versus the end support problem discussed in section 2.2.4. Consider a cable which has a characteristic impedance Z_0 and is terminated with a device input impedance $Z_L = Z_0$. A typical wave travelling in the positive x -direction in this cable has a mathematical representation for the voltage

$$V = V_0 e^{j(\omega t - kx)},$$

and for the current

$$I = \frac{V_0}{Z_0} e^{j(\omega t - kx)}.$$

NOTE! Here is a distinct difference to the waves in the vibrating string, namely regarding the use of j instead of i as the imaginary symbol and the use of the positive time dependent exponential function $e^{j\omega t}$ instead of the negative $e^{-i\omega t}$. According to Morse [24, p. 12], this is only a matter of sign convention, but when one of these conventions is chosen, then it should be used all the way through. Here it is better to adopt the use of $e^{j\omega t}$, because otherwise the following results will differ somewhat from the results obtained in the standard literature for electric engineering. For example, if the convention of the negative exponential

$e^{-i\omega t}$ was adapted in this context, the impedance of resistance and inductance connected in series would be $R - i\omega L$ instead of the common $R + j\omega L$.

When solving the partial differential equation (5.5) for the general transmission line, the solution was calculated using the Laplace transform. Equation (5.5) is anyhow separable into a product of time t and location x . Therefore, equation (5.7) can be connected to time by writing

$$V(x, t) = [A_+ e^{-\gamma x} + A_- e^{\gamma x}] e^{j\omega t}, \quad (5.32)$$

where A_+ is a complex-valued amplitude of a wave propagating to the positive direction on the x -axis, and A_- is the amplitude of the wave travelling in the negative direction. The propagation constant $\gamma = \sqrt{(R' + sL')(G' + sC')}$. Notice also the effect of the convention to use the positive exponential for the time dependence, since now the signs in the exponents of the two wave components do not match the direction of propagation.

Just like in the case of the unideal string supports, unevenly matched impedances in cable junctions cause reflection of the propagating signal. In the vicinity of the junction of the cable and the device input terminals, the voltage waveform can be modelled as

$$V(x, t) = Ae^{(j\omega t - \gamma x)} + K A e^{(j\omega t + \gamma x)},$$

where A is the complex-valued amplitude of the wave and K is a voltage reflection coefficient that determines the amplitude of the reflected wave as a fraction of the original amplitude A . The reflection equation for the current is

$$I(x, t) = \frac{1}{Z_0} [Ae^{(j\omega t - \gamma x)} - K A e^{(j\omega t + \gamma x)}].$$

The minus sign appears, because the direction of the reflecting current is opposite to the current of the transmitted signal.

At the boundary $x = 0$, where the cable connects with the load impedance, the expression for the voltage is

$$V(0, t) = Ae^{j\omega t} (1 + K), \quad (5.33)$$

and the current

$$I(0, t) = \frac{A}{Z_0} e^{j\omega t} (1 - K). \quad (5.34)$$

The boundary condition at the junction of the cable and the load impedance requires that $V = IZ_L$. Hence,

$$\frac{V(0, t)}{I(0, t)} = Z_0 \frac{1 + K}{1 - K} = Z_L,$$

and from here one can solve the expression for the reflection coefficient as

$$K = \frac{Z_L - Z_0}{Z_L + Z_0}. \quad (5.35)$$

The impedances Z_L and Z_0 are of course complex numbers and carry the information about the magnitude and phase.

When the reflected wave sums up with the transmitted wave, it generates a standing wave formation. For transmission lines, there exists a definition called *voltage standing wave ratio*,

$$\text{VSWR} = \frac{1 + K}{1 - K}.$$

This quantity expresses the amplitude ratio of the maximum and minimum amplitudes of the standing wave in a transmission line. Figure 5.7 depicts the standing wave and the locations of maximum and minimum amplitude.

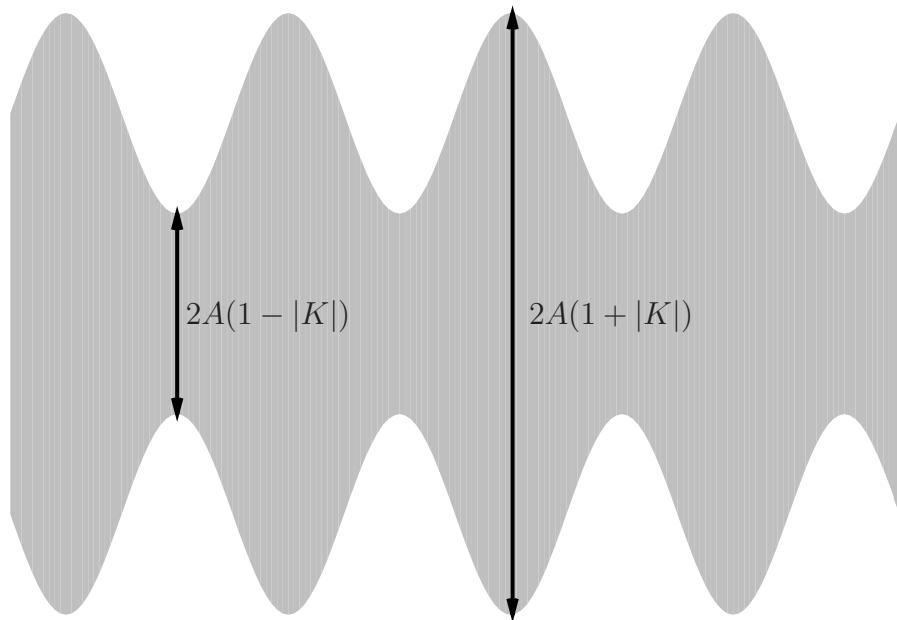


Figure 5.7: Standing waves in a transmission line

The input impedance of a load terminated transmission line can be determined using equations (5.33) and (5.34), but evaluated at $x = -l$, which is the location of the input according to Figure 5.6. Therefore,

$$V(-l, t) = A e^{j\omega t} [e^{\gamma l} + K e^{-\gamma l}], \quad (5.36)$$

and the current

$$I(-l, t) = \frac{A}{Z_0} e^{j\omega t} [e^{\gamma l} - K e^{-\gamma l}]. \quad (5.37)$$

From here the input impedance is calculated as

$$Z_{\text{in}} = \frac{V(-l, t)}{I(-l, t)} = Z_0 \frac{e^{\gamma l} + K e^{-\gamma l}}{e^{\gamma l} - K e^{-\gamma l}},$$

and a substitution of K from equation (5.35) gives

$$Z_{\text{in}} = Z_0 \frac{Z_L \cosh(\gamma l) + Z_0 \sinh(\gamma l)}{Z_0 \cosh(\gamma l) + Z_L \sinh(\gamma l)},$$

where the equivalence formulae

$$e^x - e^{-x} = 2 \sinh x \quad \text{and} \quad e^x + e^{-x} = 2 \cosh x$$

are used to transform the exponential functions into hyperbolic functions of sinh and cosh. The input impedance in the case of a lossless transmission line is

$$Z_{\text{in}} = Z_0 \frac{Z_L \cos(\beta l) + j Z_0 \sin(\beta l)}{Z_0 \cos(\beta l) + j Z_L \sin(\beta l)},$$

where $\beta = \frac{2\pi}{\lambda}$ and refers to the phase constant from the expression of the propagation constant γ .

As a special case of the derived transmission line impedances, the so-called *quarter-wave* transmission line has properties, which are used extensively in radio frequency applications. The name 'quarter-wave' refers to the length of the line in terms of the wavelength λ of the signal travelling along the transmission line. A *short-circuited* transmission line with a length close to $\frac{\lambda}{4}$ is seen as

an inductor when $l < \frac{\lambda}{4}$ and $\frac{2\lambda}{4} < l < \frac{3\lambda}{4}$

and as a

a capacitor when $\frac{\lambda}{4} < l < \frac{2\lambda}{4}$ and $\frac{3\lambda}{4} < l < \frac{4\lambda}{4}$.

At the limit, where $l = \frac{n\lambda}{4}$, the transmission line is a series *LC* resonator circuit.

Similarly an *open-circuited* transmission line with a length $\frac{\lambda}{4}$ is seen as

a capacitor when $l < \frac{\lambda}{4}$ and $\frac{2\lambda}{4} < l < \frac{3\lambda}{4}$

and as

an inductor when $\frac{\lambda}{4} < l < \frac{2\lambda}{4}$ and $\frac{3\lambda}{4} < l < \frac{4\lambda}{4}$.

At the limit, where $l = \frac{n\lambda}{4}$, the transmission line is a parallel *LC* resonator circuit. This behaviour partially explains why short wires in electric circuits are sensitive to picking up radio frequency noise. [61, pp. 32 – 37]

5.4 A CABLE AS PART OF A GUITAR TONE CONTROL CIRCUIT

So far the signal from a vibrating string has been transformed to an electrical signal in the pickup and filtered through the tone control circuit. To continue building the model for the electrical characteristics of the electric guitar, a cable is connected to the guitar's output jack, which is soldered to the volume control potentiometer of the guitar. This construction is modelled with the equivalent circuit in Figure 5.8, which is directly drawn in the suitable format for nodal analysis. The characteristics of the cable are modelled with resistance R_C , capac-

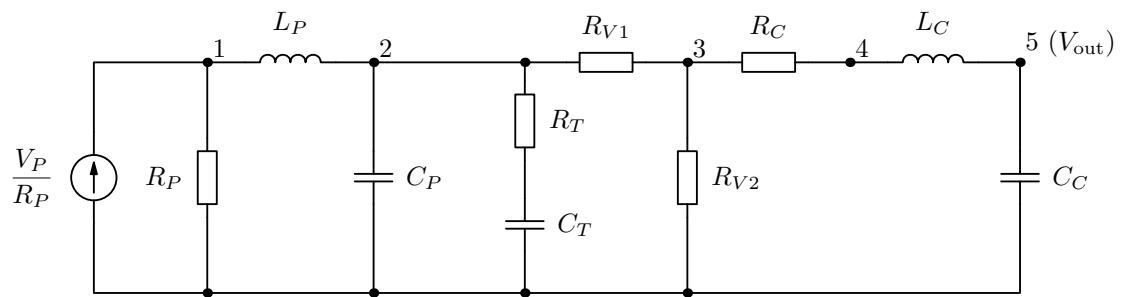


Figure 5.8: A circuit model of a cable connected to the guitar

itance C_C and inductance L_C . The volume control potentiometer R_V is modelled as two regular resistors and the components with the subscript T refer to the tone control section. The guitar pickup equivalent circuit is connected before the tone controls. The complete cable-loaded guitar output section shown in Figure 5.8 can be modelled by the matrix equation

$$\begin{bmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 \\ 0 & -Y_{32} & Y_{33} & -Y_{34} & 0 \\ 0 & 0 & -Y_{43} & Y_{44} & -Y_{45} \\ 0 & 0 & 0 & -Y_{54} & Y_{55} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{bmatrix} = \begin{bmatrix} \frac{V_P}{R_P} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

All the nonzero elements in the admittance matrix are indicated by the admittance symbol Y with subscripts referring to the element in the matrix as $Y_{\text{row},\text{column}}$. The actual matrix elements derived from the equivalent circuit are

given in listing (5.38).

$$\begin{aligned}
 Y_{11} &= \frac{1}{R_P} + \frac{1}{sL_P} & Y_{12} = Y_{21} &= \frac{1}{sL_P} \\
 Y_{22} &= \frac{1}{sL_P} + sC_P + \frac{sC_T}{1+sC_T R_T} + \frac{1}{R_{V1}} & Y_{23} = Y_{32} &= \frac{1}{R_{V1}} \\
 Y_{33} &= \frac{1}{R_{V1}} + \frac{1}{R_{V2}} + \frac{1}{R_C} & Y_{34} = Y_{43} &= \frac{1}{R_C} \\
 Y_{44} &= \frac{1}{R_C} + \frac{1}{sL_C} & Y_{45} = Y_{54} &= \frac{1}{sL_C} \\
 Y_{55} &= \frac{1}{sL_C} + sC_C
 \end{aligned} \tag{5.38}$$

Typical resistance and capacitance values for shielded guitar cables can be found from manufacturers' datasheets, sales brochures [62] and sometimes even from patent applications [63]. According to Lemme [46], the total capacitances of typical guitar cables vary between 300 pF – 1000 pF, and according to Anderton [42, p. 79], the springy coil type cables usually have the largest capacitance value per unit length. The values for cable resistance R_C and inductance L_C taken to the calculations are

$$R_C = 0.02 \frac{\Omega}{\text{m}} \quad \text{and} \quad L_C = 0.2 \times 10^{-6} \frac{\text{H}}{\text{m}}.$$

The cable is taken to be 5 meters long and the total capacitance C_C of the cable is varied from 100 pF to 900 pF to see the effect on the overall frequency response with the pickup and the tone control circuit included.

The simulation results with these values are presented in Figure 5.9. The component values have been kept identical compared to the response curves calculated from the output of the tone circuit in section 4.1. According to the results, the cable has a quite significant contribution to the overall frequency response in the higher frequency range. As the capacitance increases, the cut-off frequency of the filter assembly regarding the high frequencies moves towards lower frequencies and at the same time attenuates the resonance peak of the pickup.

The impedance curves in section 4.1 were presented to underline the effect of the volume and tone control potentiometers on the total impedance seen from the output jack of the guitar. The same impedance curves can be reproduced to indicate the effect of the cable connected at the end of the tone and volume control sections. From Figure 5.8, the impedance seen at the end of the cable is

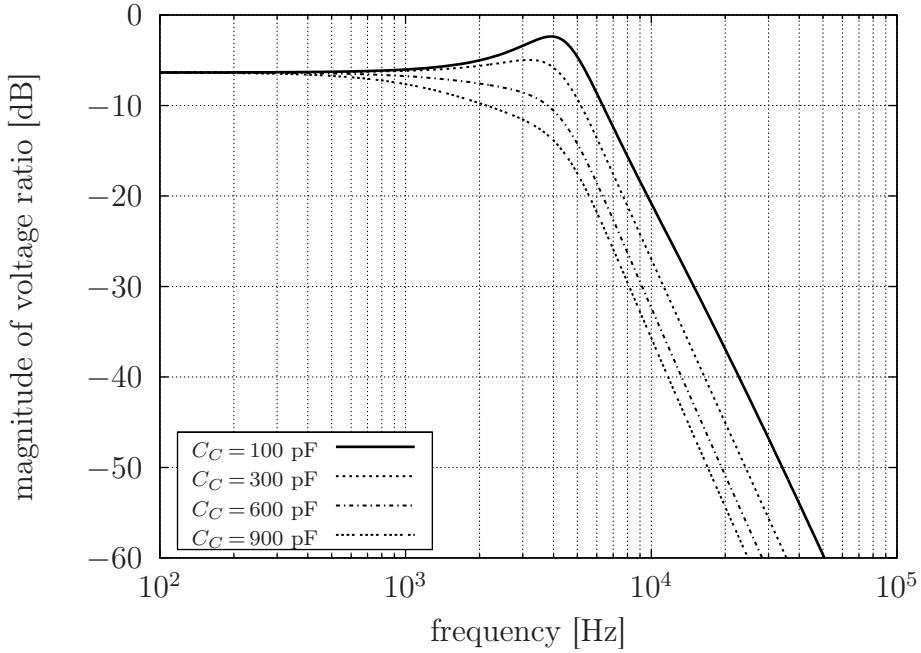


Figure 5.9: Frequency response at cable output

calculated as

$$Z_{PTVC} = \frac{(Z_{PTV} + R_C + sL_C) \frac{1}{sC_C}}{Z_{PTV} + R_C + sL_C + \frac{1}{sC_C}}, \quad (5.39)$$

where Z_{PTV} is the total impedance of the tone and volume control circuit including the pickup. The formula to evaluate Z_{PTV} was given in equation (4.1). Based on equation (5.39), the total impedance with the cable included reveals that the capacitance of the cable lowers the resonance frequency compared to the circuit without the cable.

Figure 5.10 depicts the impedance curves when the resistance of the volume control potentiometer R_V is changed. The total capacitance C_C of the cable was fixed to 500 pF and all the other values remained unchanged so that direct comparison to Figure 4.5 is possible. The original impedance curve of the unconnected pickup is drawn as a reference to indicate the shift in the resonance frequency.

Similarly, Figure 5.11 is directly comparable with Figure 4.4, which shows the effect of changing the resistance of the tone control potentiometer R_T . The only difference seems to be the change in the resonance frequency due to the small capacitance of the cable. The value of the cable capacitance C_C was fixed to 500 pF as in the case of the volume control impedance calculations shown above.

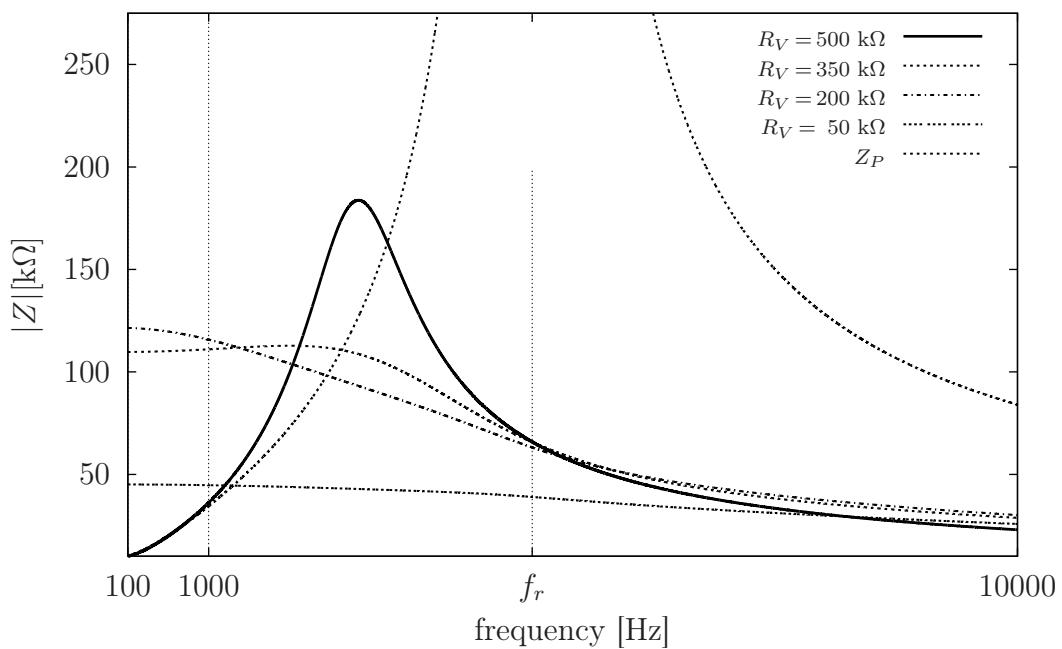


Figure 5.10: Impedance seen at the cable output when controlling the volume

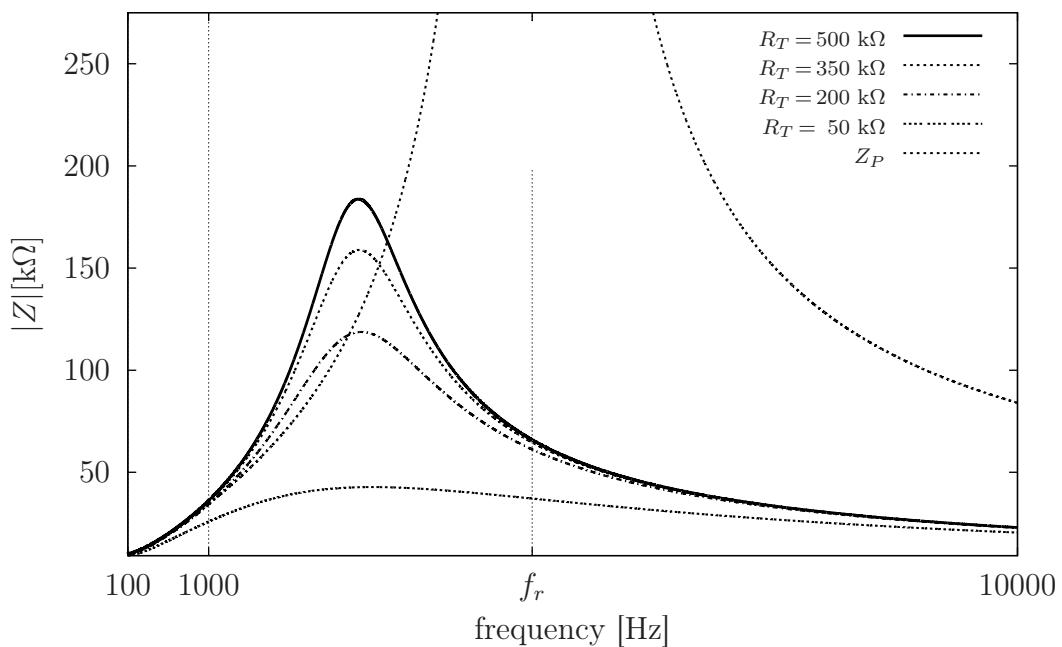


Figure 5.11: Impedance seen at the cable output when controlling the tone

CHAPTER 6

EFFECT DEVICES

As noted in the previous chapters, the strings of the electric guitar act almost as a constant source of tone. In practise this means that variations in the physical properties of the strings do not noticeably alter the spectrum of sound, although the human ear might be sensitive enough to differentiate the sound of certain types of strings. Using electricity to transfer the sound of guitar strings to an amplifier gives a million possibilities to alter and modify the sound before it reaches the loudspeaker. For that purpose, effect pedals, a.k.a. stomboxes, were invented. Soon after their invention, these pedals became wet dreams of all guitar players alike.

All physical measures and quantities involved in signal processing theory can be used as a basis for guitar effects. This scientific point of view has been chosen to categorise effects into subcategories of amplitude A , frequency f , phase ϕ and time t . All of these quantities can be used as parameters in a single function,

$$A \sin(2\pi ft + \phi).$$

This chapter presents only the most simplest effect circuits from each category. The chosen effect circuits are analysed from a scientific point of view as much as possible. It just happens that these circuits easily get too complex to handle them in manual calculations, and therefore the decision to analyse only the simple circuits is justified. Many of the following circuits have features that cover the field of basic electronics quite effectively. To also present the related theoretical background, subsections of 'side effects' have been included within some sections to explain a certain part of the circuit in more detail.

The treatment does not include digital nor software based effects. This is because the book is about science, not fiction ... The most interesting part in analogue

guitar effect devices is the way how properties of basic materials of nature affect and modify electric signals to produce elegant sounds. This is true science in direct connection with the laws of nature. Surely modern theories of signal processing allow to implement much more complex effects only by means of programming, but those things should be treated in other books.

6.1 AMPLITUDE EFFECTS

The most common effect type related directly to electric guitars is the so-called distortion effect. In this effect type the sound coming from the guitar is modified so that the *amplitude* waveform of the signal changes from its original shape. The most common reshaping method is to cut off the rounded corners of the waveform so that the signal starts to resemble a square wave. The square wave is considerably rich in harmonics compared to a sinusoidal signal, thereby giving more possibilities to control the sound by different playing techniques. Sine-to-square conversion can be considered a special case of harmonic distortion, since for periodic signals the period stays the same and only the upper partial texture of the signal is changed. Common methods to create squared amplitude distortion effects include:

- using comparators or Schmitt triggers to shoot the signal to maximum amplitude until a reference voltage is crossed by the input voltage
- using gain stages with overly large amplification so that the signal is clipped either symmetrically or asymmetrically
- using diodes to clip the amplitude peaks of the waveform (logarithmic amplification)

Distortion effects based on these waveform squaring techniques typically cause a rough sound that is widely used in all areas of music. [65] [66]

In the stompbox world, distortions come in two flavours: basic distortion and overdrive. The difference between these two is subtle but still noticeable. Overdrive effects are amplitude dependent so that a low-amplitude input signal gets distorted less than high-amplitude input. This allows the guitarist to control the effect with the style of playing. Basic distortion devices do not have this sensitivity control, as they just distort everything in an equal manner. As an example of amplitude distortion devices, the classic fuzz effect is analysed in section 6.2.

A tremolo effect can also be categorised as an amplitude effect. The underlying mechanism to produce typical tremolo tones is to apply amplitude modulation techniques to the signal obtained from the guitar. The amplitude modulation on a circuit level is implemented using a voltage-controlled amplifier block. The gain of the amplifier is modulated by a voltage signal from a low-frequency oscillator. The modulating waveforms can include sine, triangle, ramp, square or any other periodic signals that can be generated by a self-oscillating circuit section. The voltage-gain coupling is typically implemented using opto-isolators or field-effect transistors as voltage-controlled resistors. A basic analogue tremolo circuit which makes use of the resistive properties of field-effect transistors is taken under more detailed analysis in section 6.3.

The third major effect type under the amplitude category consists of devices that produce compression, expansion and noise gating. These effect types are most often used in a recording studio to even out the signal level within the dynamic range of the recording devices. All of these dynamic effects are based on voltage-controlled amplification combined with constant probing of the input signal voltage level. The voltage-gain coupling can be implemented using opto-isolators or field-effect transistors as voltage-controlled resistors, just like in the case of the tremolo effect. A more detailed analysis on an old compressor effect pedal, the 'Orange Squeezer', is presented in section 6.4.

6.2 FUZZ BOX VOODOO

One of the first commercial distortion pedals was the 'Fuzz-Tone' pedal, which was manufactured under the brand name 'Maestro' in the early years of 1960's. The bipolar junction transistor was invented in 1947, so it did not take long for it to find its way into a variety of entertainment devices - including guitar effect pedals. The first textbooks of semiconductor engineering were written based on the properties of germanium transistors, which were still more popular than silicon transistors at that time. Due to this, the first distortion pedals were designed to use germanium transistors.

Surely the fuzz effect is the most popular do-it-yourself (DIY) project for guitarists with interests towards electronics. The 'original' fuzz effect schematic is shown in Figure 6.1. As a first guess, the circuit diagram of the fuzz effect looks like a standard preamplifier section, which could be used in several different electronics projects. Actually, almost exactly the same circuit is given as a text-

book example by Cowles [67, p. 101] to demonstrate a basic 'direct-coupled pair' transistor amplifier with DC feedback. The same circuit implemented using AC feedback is also discussed in the same context.

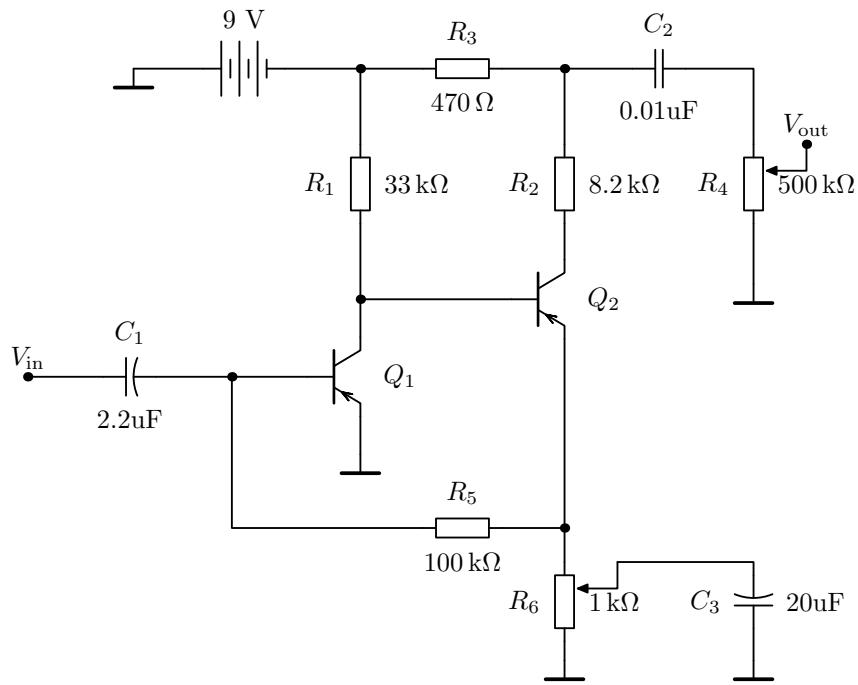


Figure 6.1: The fuzz effect circuit schematic

The thing that makes this circuit differ from a regular preamplifier or a standard textbook example is that the design has been deliberately made erroneously, including the most obvious failure in biasing the two transistors to a balanced operating point. This undergraduate student kind of biasing is analysed in detail in section 6.2.1.

From a user perspective, the device is as simple as it can get, but this also means that the variational properties of this device are only restricted to the distortion effect. The basic user of this effect device can adjust the level of distortion (over-amplification or clipping) with potentiometer R_6 and the output level of the signal with potentiometer R_4 . The DC feedback loop in the circuit is created with resistor R_5 , which connects the emitter of transistor Q_2 to the base of transistor Q_1 . The feedback loop is there to stabilise the gain properties of the circuit. Other resistors are included to complete the transistor biasing, and capacitors C_1 and C_2 are isolating the possible DC voltages from preceding and following circuits.

Basically there are so many variations of this classic circuit that it is impossible to

say which is the original version. This design is adapted directly from the website [68] maintained by the authoritative guitar effect specialist R.G. Keen. A similar schematic has also been published in the literature by Hunter [69, p. 29]. The first fuzz circuits used ancient germanium *pnp* transistors, which bring a fresh breeze from the first decades of electronics. There are also other versions of the fuzz effect that use *npn* transistors and some have even used silicon transistors. The genuine fuzz that Jimi Hendrix used was built with germanium transistors, so that is the way to go in this case.

6.2.1 The DC bias analysis of the fuzz effect

This section presents two slightly different approaches to analyse the biasing configuration of the fuzz effect. When using the first method, one has freedom to invent the equations directly by examining the different branches of the circuit. The second method relies on the nodes of the circuit to construct the biasing equations. The second method is more systematic and often the preferred approach to solve similar problems.

To make the DC analysis easier, it is a good habit to draw the DC part of the circuit separately. At first look at Figure 6.2, it does not seem to be easy to calculate the quiescent values for this circuit because of the feedback loop between the transistors Q_1 and Q_2 .

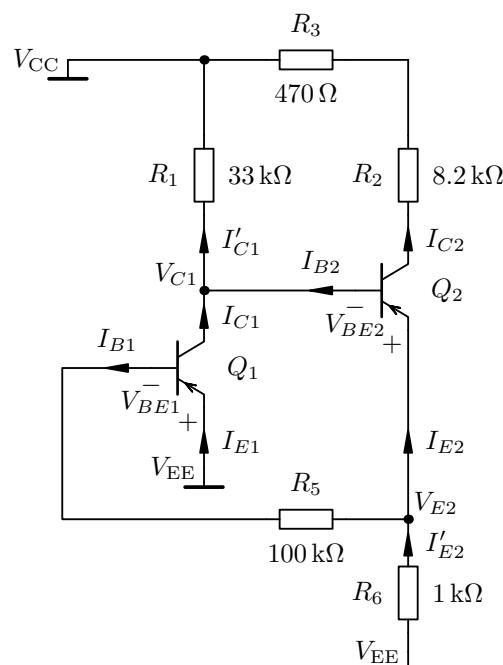


Figure 6.2: The fuzz circuit redrawn for DC bias analysis

Because *pnp* transistors are used, V_{EE} in Figure 6.2 is connected to +9 V and V_{CC} is serving as the ground potential. The current directions have been chosen to flow from higher potential towards lower potential, according to the common universal agreement of current flow.

To demonstrate the first analysis method, one needs to invent feasible current and voltage equations from the circuit. From Figure 6.2 it is possible to write at least the following transistor circuit current equations:

$$I_{C1} = \beta_{F1} I_{B1} \quad (6.1)$$

$$I_{E1} = (\beta_{F1} + 1) I_{B1} \quad (6.2)$$

$$I_{C2} = \beta_{F2} I_{B2} \quad (6.3)$$

$$I_{E2} = (\beta_{F2} + 1) I_{B2} \quad (6.4)$$

$$I'_{E2} = I_{E2} - I_{B1} \quad (6.5)$$

$$I'_{C1} = I_{C1} + I_{B2}. \quad (6.6)$$

Voltage equations are obtained by summing all voltage drops occurring in a certain loop. The loops can be taken from V_{EE} to ground, or back to V_{EE} , as long as all the branches have been included in the set of voltage equations. The equations are not unique by any means, and the following are only one set of possible equations:

$$V_{EE} = V_{EE} - V_{BE1} - I_{B1}R_5 + I'_{E2}R_6 \quad (6.7)$$

$$0 = V_{EE} - V_{BE1} - I_{B1}R_5 - V_{BE2} - I'_{C1}R_1 \quad (6.8)$$

$$V_{C2} = I_{C2}(R_2 + R_3). \quad (6.9)$$

A good starting approach is to solve first the base current of transistor Q_1 . A substitution of equation (6.5) into (6.7) leads to

$$I_{B1} = \frac{I_{E2}R_6 - V_{BE1}}{R_5 + R_6}. \quad (6.10)$$

This is a relatively simple expression for I_{B1} and it can be substituted further into equation (6.8). Including also (6.6) in (6.8) gives

$$0 = V_{EE} - V_{BE1} - I_{E2} \frac{R_5 R_6}{R_5 + R_6} - V_{BE1} \frac{R_5}{R_5 + R_6} - V_{BE2} - I_{C1}R_1 - I_{B2}R_1.$$

The next step is to find an expression that has only the current variable I_{B2} in it. Thus, one needs to use equations (6.1) and (6.4) and once more substitute (6.10)

into the remaining reference to I_{B1} . Now the equation solved for I_{B2} reads

$$I_{B2} = \frac{V_{EE} + V_{BE1} \left(\frac{R_5 + \beta_{F1} R_1}{R_5 + R_6} - 1 \right) - V_{BE2}}{R_1 + (\beta_{F2} + 1) \left(\frac{R_5 R_6 + \beta_{F1} R_1 R_6}{R_5 + R_6} \right)}. \quad (6.11)$$

The expressions for these two currents I_{B1} and I_{B2} can be used to solve all other unknown currents and voltages.

For example, by substituting the result of (6.11) into equation (6.3), the value for the quiescent collector current I_{CQ2} is found. This value is needed in the AC analysis phase to determine the value for the transistor's internal resistance r_π by using equations

$$g_m = \frac{|I_{CQ}|}{V_T} \quad \text{and} \quad r_\pi = \frac{\beta_F}{g_m}, \quad (6.12)$$

where the thermal coefficient $V_T \approx 25$ millivolts.

When inserting the actual component values into the derived quiescent equations, one can have some idea about the features of this circuit. Regardless of the β_F current gain values of the transistors, the first transistor is overbiased close to V_{EE} and the second stage is underbiased close to the ground level. At normal amplitudes of the input signal, this uneven biasing causes clipping of the input signal, which is heard as distortion at the output of the fuzz effect. It is the peculiar DC biasing of the transistors which creates the fuzzing sound.

The numerical results obtained from the analytical quiescent value calculations are collected to a common table in section 6.2.5, where those can be easily compared to results from SPICE simulations and prototype measurements.

The second bias analysis method builds up the current and voltage equations in a slightly different, yet more organised, manner. The direct current model in Figure 6.2 has two distinct voltage nodes, labelled as V_{C1} and V_{E2} . The output node is identified in this case as V_{C2} , but it is not needed in the bias analysis. It is also beneficial to notice that $V_{E2} = V_{C1} + V_{BE2}$. This reduces the number of distinct voltage nodes to one.

Based on Kirchhoff's current law, the sum of currents entering a node equals the sum of currents leaving a node. Therefore, the current equations for the voltage

nodes V_{E2} and V_{C1} are:

$$\begin{aligned} I'_{E2} &= I_{E2} - I_{B1} = (\beta_{F2} + 1)I_{B2} - I_{B1} \\ I'_{C1} &= I_{C1} + I_{B2} = \beta_{F1}I_{B1} + I_{B2}. \end{aligned}$$

These equations can be combined to eliminate the base current I_{B2} and then the current equation reads

$$I'_{C1} = \beta_{F1}I_{B1} + \frac{I'_{E2}}{\beta_{F2} + 1} + \frac{I_{B1}}{\beta_{F2} + 1}.$$

These currents can be expressed with the node voltages as:

$$I'_{E2} = \frac{V_{E2}}{R_6} ; \quad I'_{C1} = \frac{V_{C1} - V_{CC}}{R_1} ; \quad I_{B1} = \frac{V_{EE} - V_{BE1} - V_{E2}}{R_5}$$

and after substituting the voltage equations,

$$\frac{V_{C1} - V_{CC}}{R_1} = \beta_{F1} \frac{V_{EE} - V_{BE1} - V_{E2}}{R_5} + \frac{V_{E2}}{R_6(\beta_{F2} + 1)} + \frac{V_{EE} - V_{BE1} - V_{E2}}{R_5(\beta_{F2} + 1)}.$$

With the fact that $V_{E2} = V_{C1} + V_{BE2}$, one has an equation from where the voltage V_{C1} can be solved. Based on this analysis, all the other biasing voltages and currents can also be solved.

6.2.2 The AC analysis of the fuzz effect

The small-signal model for the fuzz effect is drawn by replacing the transistors by their small-signal equivalents as explained in section 1.3.10. To prepare for the use of nodal analysis, all external alternating voltage sources are converted into current sources and all DC voltage sources are considered to be at ground potential from the viewpoint of alternating signals.

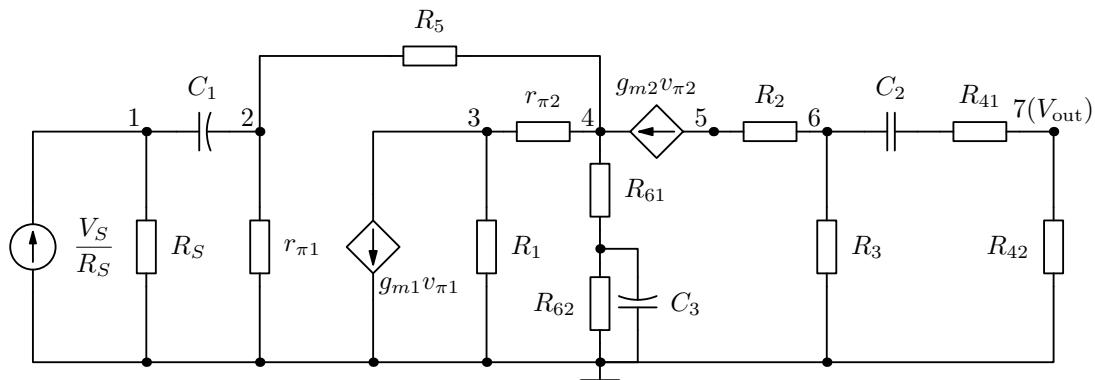


Figure 6.3: A small-signal model of the fuzz effect circuit

Figure 6.3 shows the circuit diagram of the small-signal model. The current source on the left and its internal resistance R_S should represent the electric

guitar as a signal source. The resistors $r_{\pi 1}$ and $r_{\pi 2}$ represent the internal input resistances of transistors Q_1 and Q_2 . Similarly, g_{m1} and g_{m2} are the transconductances of the transistors and the small-signal voltages $v_{\pi 1}$ and $v_{\pi 2}$ represent the potential differences acting over the input resistances $r_{\pi 1}$ and $r_{\pi 2}$. The potentiometers R_6 and R_4 have been divided into two separate resistors. The voltage nodes are indexed with numbers from 1 to 7, where the node 7 is the output node of the circuit. The circuit of Figure 6.3 has a matrix representation,

$$\begin{bmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & 0 & -Y_{24} & 0 & 0 & 0 \\ 0 & 0 & Y_{33} & -Y_{34} & 0 & 0 & 0 \\ 0 & -Y_{42} & -Y_{43} & Y_{44} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & Y_{55} & -Y_{56} & 0 \\ 0 & 0 & 0 & 0 & -Y_{65} & Y_{66} & -Y_{67} \\ 0 & 0 & 0 & 0 & 0 & -Y_{76} & Y_{77} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \\ V_7 \end{bmatrix} = \begin{bmatrix} \frac{V_S}{R_S} \\ 0 \\ -g_{m1}V_2 \\ g_{m2}(V_3 - V_4) \\ -g_{m2}(V_3 - V_4) \\ 0 \\ 0 \end{bmatrix}.$$

The nonzero elements of the admittance matrix are identified only with their corresponding indices because of limited space. The following listing (6.13) contains the actual terms that should be substituted into the matrix above.

$$\begin{aligned} Y_{11} &= \frac{1}{R_S} + j\omega C_1 & Y_{12} = Y_{21} &= j\omega C_1 \\ Y_{22} &= \frac{1}{r_{\pi 1}} + \frac{1}{R_5} + j\omega C_1 & Y_{24} = Y_{42} &= \frac{1}{R_5} \\ Y_{33} &= \frac{1}{r_{\pi 2}} + \frac{1}{R_1} & Y_{34} = Y_{43} &= \frac{1}{r_{\pi 2}} \\ Y_{44} &= \frac{1}{r_{\pi 2}} + \frac{1}{R_5} + \frac{j\omega C_3 + \frac{1}{R_{62}}}{1 + j\omega C_3 R_{61} + \frac{R_{61}}{R_{62}}} & & (6.13) \\ Y_{55} &= \frac{1}{R_2} & Y_{56} = Y_{65} &= \frac{1}{R_2} \\ Y_{66} &= \frac{1}{R_2} + \frac{1}{R_3} + \frac{j\omega C_2}{1 + j\omega C_2 R_{41}} & Y_{67} = Y_{76} &= \frac{j\omega C_2}{1 + j\omega C_2 R_{41}} \\ Y_{77} &= \frac{j\omega C_2}{1 + j\omega C_2 R_{41}} + \frac{1}{R_{42}} \end{aligned}$$

At this phase, the active transconductance parameters g_{m1} and g_{m2} have not yet been included in the admittance matrix, since this form of the equation is just to show the primary matrix that is directly written down from the small-signal model. To reach a sufficient form for the admittance matrix to solve the node

voltages, the transconductances need to be transferred to the admittance matrix. In the current vector, transconductance g_{m1} on row 3 is multiplying voltage V_2 . Therefore, g_{m1} is moved to the admittance matrix on row 3 to column 2, which multiplies the node voltage V_2 in the voltage vector. The move of g_{m1} leads to a modified admittance matrix element

$$Y'_{32} = -\frac{g_{m1}r_{\pi1}}{r_{\pi1}}.$$

Doing a similar move to g_{m2} on rows 4 and 5 of the current vector results in four elements:

$$\begin{aligned} Y'_{43} &= -\frac{g_{m2}r_{\pi2} + 1}{r_{\pi2}} & Y'_{44} &= \frac{g_{m2}r_{\pi2} + 1}{r_{\pi2}} + \frac{1}{R_5} + \frac{j\omega C_3 + \frac{1}{R_{62}}}{1 + j\omega C_3 R_{61} + \frac{R_{61}}{R_{62}}} \\ Y'_{53} &= \frac{g_{m2}r_{\pi2}}{r_{\pi2}} & Y'_{54} &= -\frac{g_{m2}r_{\pi2}}{r_{\pi2}}. \end{aligned}$$

In each of the above equations $g_{m1}r_{\pi1} = \beta_{F1}$ and $g_{m2}r_{\pi2} = \beta_{F2}$, and normally the β_F current gain parameter is used in the calculations. Basically there are slight differences between the DC current gain β_F and the AC current gain h_{fe} as explained in section 1.3.8, but as a rough estimate the β_F is sufficient to be used in place of h_{fe} . With these additions, the use of Cramer's rule leads to the determinant division

$$V_7 = V_{\text{out}} = \frac{\left| \begin{array}{ccccccc} Y_{11} & -Y_{12} & 0 & 0 & 0 & 0 & \frac{V_S}{R_S} \\ -Y_{21} & Y_{22} & 0 & -Y_{24} & 0 & 0 & 0 \\ 0 & Y'_{32} & Y_{33} & -Y_{34} & 0 & 0 & 0 \\ 0 & -Y_{42} & Y'_{43} & Y'_{44} & 0 & 0 & 0 \\ 0 & 0 & Y'_{53} & Y'_{54} & Y_{55} & -Y_{56} & 0 \\ 0 & 0 & 0 & 0 & -Y_{65} & Y_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & -Y_{76} & 0 \end{array} \right|}{\left| \begin{array}{ccccccc} Y_{11} & -Y_{12} & 0 & 0 & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & 0 & -Y_{24} & 0 & 0 & 0 \\ 0 & Y'_{32} & Y_{33} & -Y_{34} & 0 & 0 & 0 \\ 0 & -Y_{42} & Y'_{43} & Y'_{44} & 0 & 0 & 0 \\ 0 & 0 & Y'_{53} & Y'_{54} & Y_{55} & -Y_{56} & 0 \\ 0 & 0 & 0 & 0 & -Y_{65} & Y_{66} & -Y_{67} \\ 0 & 0 & 0 & 0 & 0 & -Y_{76} & Y_{77} \end{array} \right|}.$$

Basically it is possible to write open the determinants and obtain the transfer

function of the fuzz circuit in symbolic form but it takes about two weeks and many sheets of paper (experience has shown ...). The brute force method is obviously not suitable in this case. However, the determinant division can be solved effortlessly using Octave or Matlab, where the determinants can be evaluated numerically. The results of the numerical analysis to obtain a theoretical frequency response of the fuzz circuit are shown in section 6.2.5.

6.2.3 Side effect: feedback in transistor circuits

Since the fuzz effect circuit is a textbook example of a typical feedback amplifier, it is convenient to explore the concept of feedback more thoroughly in this context. Feedback mainly affects the DC and AC gains and the input and output impedances of the circuit. If the feedback loop is excluded from the fuzz circuit, the remaining circuit would be a simple direct-coupled pair shown in Figure 6.4. Here the circuit has been drawn using *npn* transistors. In the upcoming analysis the circuit is treated as a current amplifier by adding a current-shunt feedback loop and as a voltage amplifier by adding a voltage-series feedback loop.

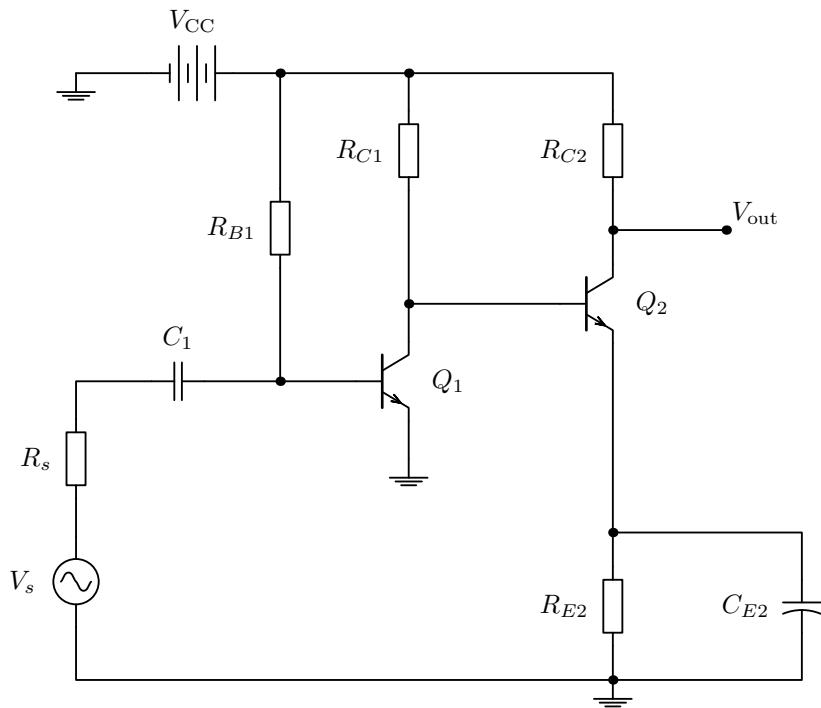


Figure 6.4: A bipolar junction transistor direct-coupled pair amplifier

Feedback is a fundamental concept in all fields of engineering. In electronics negative feedback is used for stabilising gain, modifying impedance, extending bandwidth and reducing noise. References to the terms of DC and AC feedback

imply that the feedback stabilises and enhances either DC or AC behaviour of the circuit. Typically the feedback is a mixture of AC and DC feedback. The most important parameters to describe feedback are the return ratio T and the feedback factor β . The following discussion and analysis aims to explain how these parameters are evaluated and what do they actually mean in practice.

Before getting down to business, it is worth mentioning a good tool for analysing feedback in electronics and in general: the signal flow graph. Figure 6.5 introduces the most basic elements used for visualising a complete signal chain as a flow graph. The key point is to understand how the multipliers G_x between two

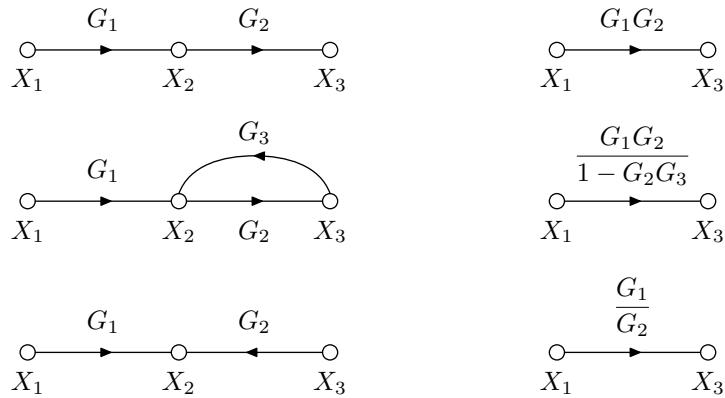


Figure 6.5: Basic forms of signal flow functions and their 'simplifications'

junctions are related to each other when the flow graph is simplified by reducing the number of branches. The mathematical 'proof' of these multiplier combinations is simple; for example the bottom signal flow is evaluated as

$$\begin{aligned} X_2 &= G_1 X_1 \\ X_2 &= G_2 X_3 \quad \rightarrow \quad X_3 = \frac{G_1}{G_2} X_1. \end{aligned}$$

The multiplier is applied to the junction from where the arrow leaves and the result of this multiplication is the junction where the arrow points to.

A general feedback cycle is easily demonstrated with the signal flow graph of Figure 6.6. The form of presentation is adapted from the textbook of Millman [15]. The idea is to introduce a new set of 'two-port' parameters $t_{11}, t_{12}, t_{21}, t_{22}$, which act as multipliers between the branches of the feedback circuit. The method of evaluating the multipliers is to utilise the idea of superposition. When evaluating each multiplier one by one, some of the junctions (voltage or current sources) are given a zero value to isolate a specific section of the circuit. The dashed line between X_i and \hat{X}_i should be interpreted as if the connection between these

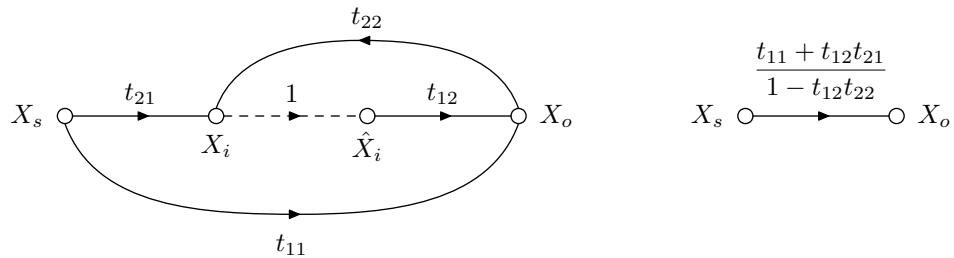


Figure 6.6: Signal flow graph to explain feedback

two points had been broken. Since \hat{X}_i is just a parameter related to the controlled source, the connection between X_i and \hat{X}_i in the small signal model can be directly considered as 'broken' and the circuit configuration does not need any changes for calculations. Detailed examples of this method are to follow later on.

The simplified equation for the feedback signal flow is derived by writing the equations

$$X_i = t_{21}X_s + t_{22}X_o \quad (6.14)$$

$$\hat{X}_i = X_i \quad (6.15)$$

$$X_o = t_{12}\hat{X}_i + t_{11}X_s \quad (6.16)$$

and then simply solving the relation between the input X_s and the output X_o . When comparing to the "general equation" of feedback,

$$X_o = \frac{t_{11} + t_{12}t_{21}}{1 - t_{12}t_{22}} X_s = \frac{A_{OL}}{1 + T} X_s, \quad (6.17)$$

each of the parameters are given their special meaning. The mathematical definitions for the parameters are stated as:

$$\begin{aligned} t_{11} &= \left. \frac{X_o}{X_s} \right|_{\hat{X}_i=0} & t_{12} &= \left. \frac{X_o}{\hat{X}_i} \right|_{X_s=0} \\ t_{21} &= \left. \frac{X_i}{X_s} \right|_{X_o=0} & t_{22} &= \left. \frac{X_i}{X_o} \right|_{X_s=0}. \end{aligned}$$

On the right side of each definition is an indication which source should be set to zero when evaluating the parameter. The source can be current or voltage depending on the feedback configuration. Often the trickiest part is to handle the output source correctly when evaluating t_{21} . It helps to think that when setting the output source to zero, the output signal should be isolated so that no signal goes to the feedback loop. When considering parameters t_{12} and t_{21} , the feedback part of the circuit is typically seen as a load and therefore is a natural part of these parameters.

With respect to these parameters, the return ratio has a relation of

$$-T = \left. \frac{X_i}{\hat{X}_i} \right|_{X_s=0} = t_{12}t_{22} = t_{12}t_{21} \frac{t_{22}}{t_{21}} \approx A_{OL} \beta.$$

The last approximation is valid only if the feedforward path is neglected in the feedback network. The return ratio T is a very useful parameter, which is especially needed when evaluating the impedance changes due to the feedback configuration.

Typically circuits with feedback loops are analysed by extracting the amplifier gain path and the feedback path as separate parameters. By the rules indicated in figure 6.5, the gain path forward transmission transfer function is

$$\frac{X_o}{X_s} = t_{12}t_{21}. \quad (6.18)$$

If the feedforward path t_{11} is neglected (as is typically safe to do), then the so called open loop gain $A_{OL} = t_{12}t_{21}$. With the same reasoning the feedback path reverse transmission transfer function is

$$\frac{X_s}{X_o} = \frac{t_{22}}{t_{21}} = \beta. \quad (6.19)$$

So the general idea is to evaluate a transfer function between input and output to both directions of signal flow and indicate them as A_{OL} and β . The most interesting feedback parameter is the *feedback factor* β , which describes the reverse transmission of the feedback network. Quite often approximate feedback analysis methods give a false meaning to β , as if it would be a factor related to t_{22} path only. Generally β describes the reverse transfer function where the output node X_o is taken as the signal source and the source node X_s is treated as the end point of the feedback signal.

The input and output impedances of a circuit with feedback can be evaluated using the Blackman's impedance formula

$$Z_F = Z_D \frac{1 + T_{SC}}{1 + T_{OC}}, \quad (6.20)$$

where Z_D is the input or output impedance evaluated from a 'dead system' (gain is set to zero) and T_{SC}, T_{OC} are the return ratio T evaluated in short-circuited (SC) and open-circuited (OC) conditions respectively. Typically it is enough to evaluate T once according to its definition and then in the obtained equation set values 0 (short) or ∞ (open) to the input and terminating impedances (R_s and R_L). The 'dead system' is obtained by suppressing the controlled source ($\hat{X}_i = 0$)

just like it is done when evaluating the t_{11} feedforward parameter. The 'dead system' input and output impedances Z_{ID} and Z_{OD} are then evaluated from the zero-gain system. These impedances are scaled with the return ratio depending on the feedback configuration used. Concrete examples will follow later on.

The current-shunt feedback is typically applied in current amplifiers, where the output current is sampled and fed back to the input in a parallel (shunt) connection. This feedback configuration is also known as the 'shunt-series' type because the output current is probed by a series connection (parallel probing of current is not possible). In the feedback analysis the amplifier element is modelled as a general 'two-port' system (see Figure 6.7), from where the series and shunt connections are easily distinguished.

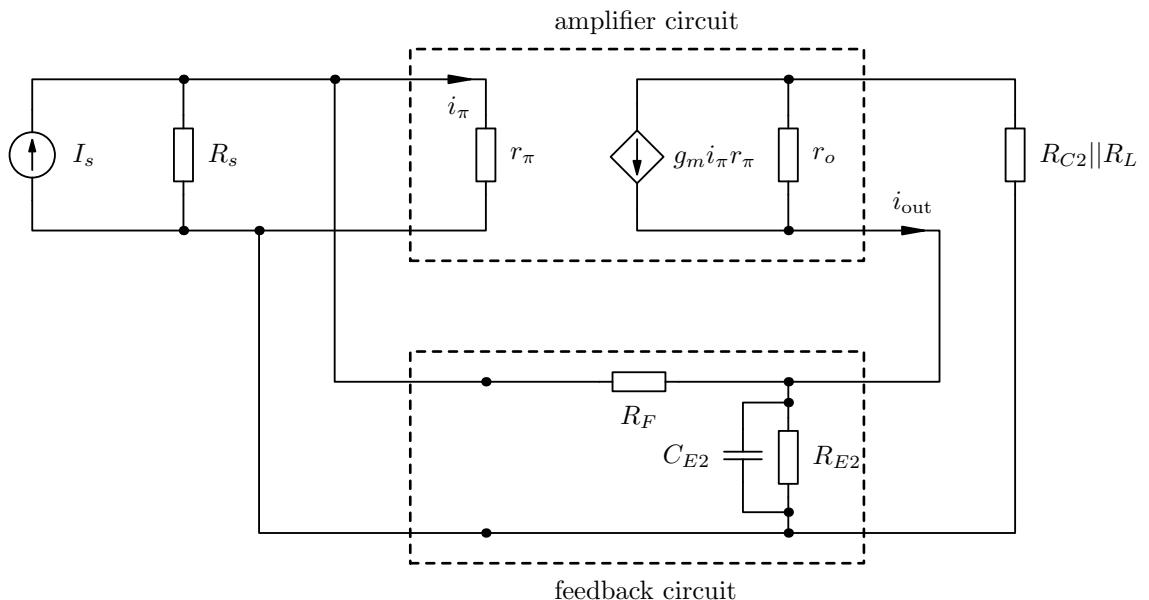


Figure 6.7: A current amplifier with shunt feedback

The feedback network in Figure 6.7 is implemented as a simple passive current divider, where the output current is divided between resistors R_F and R_{E2} . The capacitor C_{E2} acts as a short circuit for AC currents so that the feedback loop in this case is closed for DC current only. The feedback configuration is drawn specifically this accurately to relate to the circuit under analysis.

An example amplifier structure using the current-shunt configuration is shown in Figure 6.8. This configuration is chosen because it is used as a basis in the early 'fuzz-face' guitar effect devices. The feedback loop is implemented by connecting the emitter of the second amplifier stage to the base of the first amplifier stage via resistor R_F . As already mentioned, the feedback loop is closed only for direct

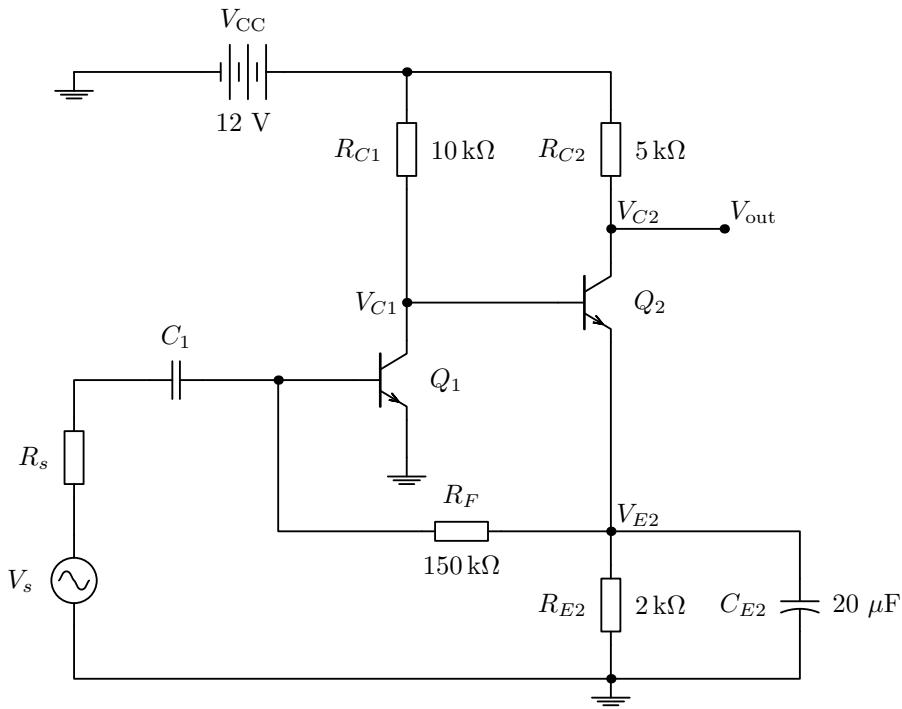


Figure 6.8: A direct-coupled pair amplifier with current-shunt feedback

current. Therefore, the feedback is used here mainly to enhance the biasing stability of the amplifier.

To see the effects of the feedback loop, the amplifier should be given an in-depth analysis from the viewpoint of DC and AC operation. The DC model is redrawn for analysis purposes in Figure 6.9. Clearly the feedback loop is used for giving the needed base current for transistor Q_1 .

A reasonably systematic method for the DC bias analysis is to examine the independent voltage nodes of the circuit. The direct current model in Figure 6.9 has two distinct voltage nodes, labelled as V_{C1} and V_{E2} . It is also beneficial to notice that $V_{E2} = V_{C1} - V_{BE2}$. This reduces the number of distinct voltage nodes to one.

Based on Kirchhoff's current law, the sum of currents entering a node equals the sum of currents leaving a node. Therefore, the current equations for the voltage nodes V_{E2} and V_{C1} are:

$$\begin{aligned} I'_{E2} &= I_{E2} - I_{B1} = (\beta_{F2} + 1)I_{B2} - I_{B1} \\ I'_{C1} &= I_{C1} + I_{B2} = \beta_{F1}I_{B1} + I_{B2}. \end{aligned}$$

These equations can be combined to eliminate the base current I_{B2} and then the

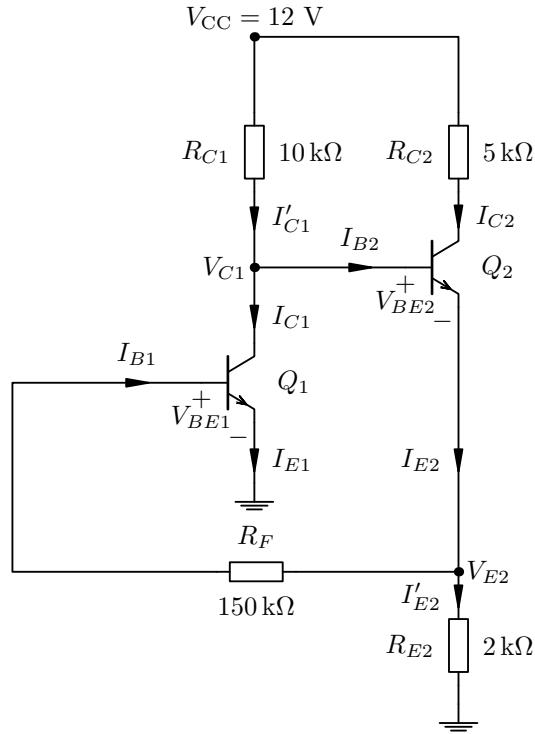


Figure 6.9: A DC model of the amplifier with current-shunt feedback

current equation reads

$$I'_{C1} = \beta_{F1} I_{B1} + \frac{I'_E2}{\beta_{F2} + 1} + \frac{I_{B1}}{\beta_{F2} + 1}.$$

These currents can be expressed with the node voltages as:

$$I'_{E2} = \frac{V_{E2}}{R_{E2}} ; \quad I'_{C1} = \frac{V_{CC} - V_{C1}}{R_{C1}} ; \quad I_{B1} = \frac{V_{E2} - V_{BE1}}{R_F}$$

and after substituting the voltage equations,

$$\frac{V_{CC} - V_{C1}}{R_{C1}} = \beta_{F1} \frac{V_{E2} - V_{BE1}}{R_F} + \frac{V_{E2}}{R_{E2}(\beta_{F2} + 1)} + \frac{V_{E2} - V_{BE1}}{R_F(\beta_{F2} + 1)}.$$

With the fact that $V_{E2} = V_{C1} - V_{BE2}$, one has an equation from where the voltage V_{C1} can be solved. Based on this analysis, all the other biasing voltages and currents can also be solved.

Next the focus moves to the AC feedback analysis, which for the given circuit is simple: there is no AC feedback because of capacitor C_{E2} . But since it would be nice to introduce the tools for feedback analysis, let's remove all the capacitors from the circuit and analyse the feedback properties of the circuit at mid-frequency range.

Considering the frequency response (AC) analysis, it is necessary to redraw the circuit using the small-signal model of the BJT. A low-frequency BJT model is

sufficient in this case because the focus is on the audio frequency analysis only. The small-signal model in Figure 6.10 still has the capacitors drawn in place, but the following analysis assumes those are removed from the circuit.

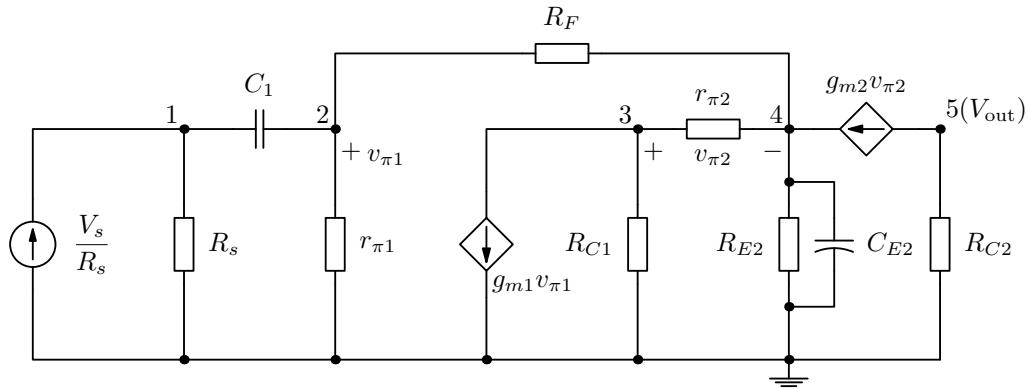


Figure 6.10: A small-signal model of the direct-coupled pair amplifier with current-shunt feedback

Let's apply the general feedback analysis to this circuit and find equations for t_{12} , t_{21} , T and β . Since the feedback signal is current, it is convenient that the controlled sources are transformed into 'current mode' using the expansion

$$g_m v_\pi = g_m r_\pi \frac{v_\pi}{r_\pi} = \beta_F i_\pi. \quad (6.21)$$

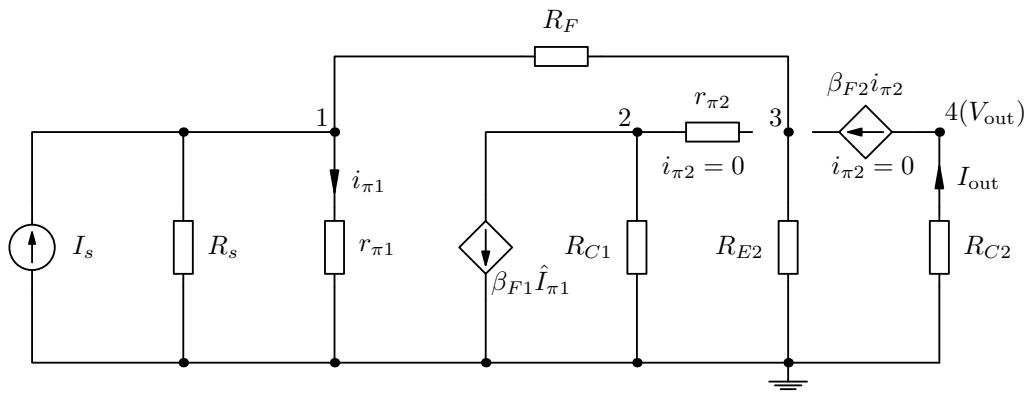
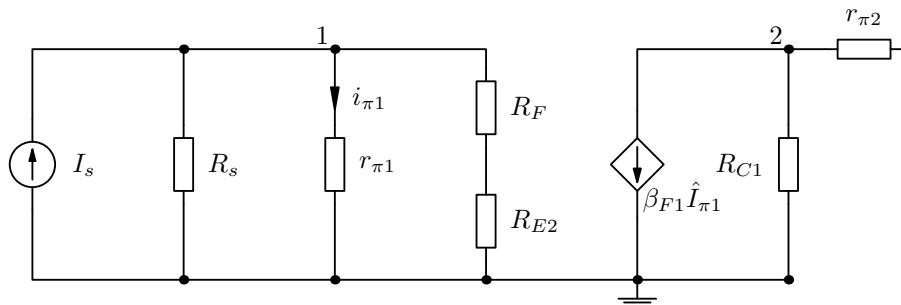
When following this feedback analysis it is necessary that one controlled source be chosen as an independent source. This independent source is identified later on by using the notation

$$\beta_F \hat{I}_\pi. \quad (6.22)$$

The return ratio T and t_{12} are obtained from the same circuit model where $X_s = I_s$ is set to zero. To evaluate t_{21} , a circuit with the output source $X_o = I_o = 0$ needs to be drawn. This circuit is shown in Figure 6.11 and simplified further in Figure 6.12.

The matrix equation describing the circuit in Figure 6.12 is given by equation (6.23).

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi1}} + \frac{1}{R_F + R_{E2}} & 0 \\ 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} I_s \\ -\beta_F \hat{I}_{\pi1} \end{bmatrix} \quad (6.23)$$

Figure 6.11: Circuit model for evaluating the t_{21} feedback parameterFigure 6.12: Simplified circuit model for evaluating the t_{21} feedback parameter

This matrix equation can be written directly by inspecting the circuit model and using the rules of the nodal analysis. The nodal analysis is used, because the second controlled current source cannot be transformed as a voltage source due to a missing parallel resistance. According to the rules of the nodal analysis, the admittance terms are collected to the main matrix and the current source terms are placed into the current vector. The voltage nodes in the voltage vector are the variables to be solved from this equation.

For evaluation purposes it is convenient that all controlled source terms from the current vector are moved to the admittance matrix. Because the matrix equation is just a set of linear equations, the controlled source terms are subtracted or added to both sides of the equation. It is also necessary to identify to which voltage node variables the controlled source term relates to. Based on this discussion, the matrix equation (6.24) is readily modified in the desired form.

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} + \frac{1}{R_F + R_{E2}} & 0 \\ \frac{\beta_{F1}}{r_{\pi 1}} & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} I_s \\ 0 \end{bmatrix} \quad (6.24)$$

The feedback parameter t_{21} was defined as the fraction $\frac{X_i}{X_s}$, where the source terms in this current amplifier case are defined as currents. Therefore, the resulting equation for t_{21} is

$$t_{21} = \frac{i_{\pi 1}}{I_s} = \frac{V_1}{r_{\pi 1} I_s} = \frac{R_s(R_{E2} + R_F)}{(r_{\pi 1} + R_s)(R_{E2} + R_F) + R_s r_{\pi 1}}. \quad (6.25)$$

As a single parameter this does not yet describe much of the circuit properties, it is just a multiplier between the input source and the input of the amplifier.

The circuit model needs to be redrawn for analysing the return ratio T and the feedback parameter t_{12} . Figure 6.13 shows the correct model, where the input source is removed due to the rule $X_s = I_s = 0$. When the value of a current source is set to 0, it describes an open-circuit condition and the branch containing the current source is simply removed from the circuit as seen from Figure 6.13.

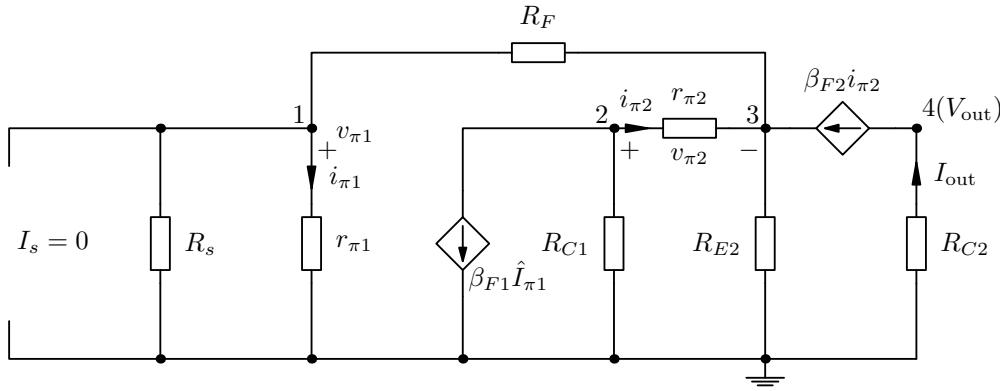


Figure 6.13: Circuit model for evaluating feedback parameters T and t_{12}

The matrix equation describing the circuit model of Figure 6.13 is written as

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} + \frac{1}{R_F} & 0 & -\frac{1}{R_F} & 0 \\ 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} & -\frac{1}{r_{\pi 2}} & 0 \\ -\frac{1}{R_F} & -\frac{1}{r_{\pi 2}} & \frac{1}{R_F} + \frac{1}{R_{E2}} + \frac{1}{r_{\pi 2}} & 0 \\ 0 & 0 & 0 & \frac{1}{R_{C2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\beta_{F1} \hat{I}_{\pi 1} \\ \beta_{F2} i_{\pi 2} \\ -\beta_{F2} i_{\pi 2} \end{bmatrix}.$$

This equation needs to be modified so that it is easy to solve the fractions $-\frac{i_{\pi 1}}{\hat{I}_{\pi 1}}$ and $\frac{I_{\text{out}}}{\hat{I}_{\pi 1}}$ as required by the definitions of the feedback parameters. Therefore

the independent controlled source term $\hat{I}_{\pi 1}$ is left to the current vector, while all other current source terms are moved to the admittance matrix. The resulting matrix equation is of the form shown below and the unknown voltages can be relatively easily solved from this matrix equation using the Cramer's rule.

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} + \frac{1}{R_F} & 0 & -\frac{1}{R_F} & 0 \\ 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} & -\frac{1}{r_{\pi 2}} & 0 \\ -\frac{1}{R_F} & -\frac{\beta_{F2} + 1}{r_{\pi 2}} & \frac{1}{R_F} + \frac{1}{R_{E2}} + \frac{\beta_{F2} + 1}{r_{\pi 2}} & 0 \\ 0 & \frac{\beta_{F2}}{r_{\pi 2}} & -\frac{\beta_{F2}}{r_{\pi 2}} & \frac{1}{R_{C2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\beta_{F1}\hat{I}_{\pi 1} \\ 0 \\ 0 \end{bmatrix}.$$

After solving the fraction $-\frac{i_{\pi 1}}{\hat{I}_{\pi 1}}$ one has the equation for the return ratio

$$T = -\frac{i_{\pi 1}}{\hat{I}_{\pi 1}} = \frac{R_P}{r_{\pi 1}} \frac{\beta_{F1}R_{C1}(\beta_{F2} + 1)R_{E2}}{(R_{E2} + R_F + R_P)(R_{C1} + r_{\pi 2}) + (\beta_{F2} + 1)R_{E2}(R_F + R_P)},$$

where a parallel resistance term $R_P = R_s || r_{\pi 1}$ has been taken into use to simplify the equations. In a similar fashion the feedback parameter t_{12} is solved using the same model, and the result is

$$t_{12} = \frac{V_4}{R_{C2}\hat{I}_{\pi 1}} = -\frac{\beta_{F1}R_{C1}\beta_{F2}(R_{E2} + R_F + R_P)}{(R_{E2} + R_F + R_P)(R_{C1} + r_{\pi 2}) + (\beta_{F2} + 1)R_{E2}(R_F + R_P)}.$$

If the approximation is made that $\beta_{F2} + 1 \approx \beta_{F2}$, then the multiplier t_{22} will simplify considerably to the form

$$t_{22} = -\frac{T}{t_{12}} = \frac{R_S R_{E2}}{(r_{\pi 1} + R_S)(R_F + R_{E2}) + R_S r_{\pi 1}},$$

and then the equation for the feedback factor becomes

$$\beta = \frac{t_{22}}{t_{21}} = \frac{R_{E2}}{R_{E2} + R_F}. \quad (6.26)$$

For evaluating the input impedance one must draw a circuit model for the so called 'dead-system'. The dead-system impedance was needed for using the Blackman's impedance formula (6.20). Figure 6.14 shows the dead-system model with the independent controlled source suppressed. Figure 6.14 includes also an ideal test current source I_T , which does not have any internal source impedance. Hence, the input impedance is determined by sourcing current and measuring the resulting voltage at voltage node V_1 .

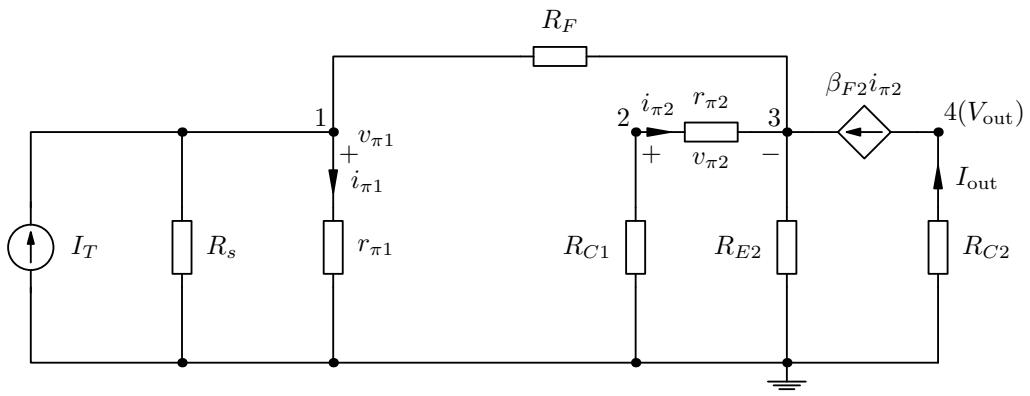


Figure 6.14: Circuit model for evaluating the dead-system input impedance

The impedance is evaluated by solving V_1 from the matrix equation below and dividing by the test current I_t .

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} + \frac{1}{R_F} & 0 & -\frac{1}{R_F} & 0 \\ 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} & -\frac{1}{r_{\pi 2}} & 0 \\ -\frac{1}{R_F} & -\frac{\beta_{F2} + 1}{r_{\pi 2}} & \frac{1}{R_F} + \frac{1}{R_{E2}} + \frac{\beta_{F2} + 1}{r_{\pi 2}} & 0 \\ 0 & \frac{\beta_{F2}}{r_{\pi 2}} & -\frac{\beta_{F2}}{r_{\pi 2}} & \frac{1}{R_{C2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} I_t \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then one will get the dead-system input impedance with R_s included. The true input impedance of the amplifier itself does not contain R_s . The input impedance without R_s is obtained from the result by making the resistance of R_s infinitely large (in numerical analysis 100 megohms is enough). For getting a decent expression for the dead-system input impedance it is not necessary to solve the matrix equation, because a direct inspection of Figure 6.14 shows the input impedance to be approximately the parallel connection of the branches $r_{\pi 1}$ and $R_F + R_{E2}$.

The Blackman's formula requires also the return ratio to be evaluated in short-circuit and open-circuit conditions. Basically this is simple, since the methodology states that R_s and R_L are not part of the amplifier feedback cycle when considering the input and output connections. Therefore, one can get the short-circuit and open-circuit conditions by inspection from the already evaluated expression for T . Setting $R_s = 0$ in the return ratio of the current-shunt feedback system creates the short-circuit condition. For this condition $T(R_s \rightarrow 0) = 0$, so

the only meaningful value for the return ratio must come from setting $R_s = \infty$. Hence,

$$T(R_s \rightarrow \infty) = \frac{\beta_{F1} R_{C1} (\beta_{F2} + 1) R_{E2}}{(R_{E2} + R_F + r_{\pi1})(R_{C1} + r_{\pi2}) + (\beta_{F2} + 1) R_{E2} (R_F + r_{\pi1})},$$

and the use of the Blackman's impedance formula for the input impedance with feedback gives

$$Z_{IF} = \frac{Z_{ID}}{1 + T(R_s \rightarrow \infty)} \approx \frac{r_{\pi1} || (R_F + R_{E2})}{1 + T(R_s \rightarrow \infty)}.$$

So the current-shunt feedback reduces the input impedance by the amount determined by the open-circuited return ratio.

As its name suggests, **the voltage-series feedback** is typically applied in voltage amplifiers. In this feedback type the output voltage is sampled and fed back to the input in series with the signal source. This feedback configuration is also known as the 'shunt-series' type because the output voltage is probed by a shunt connection (probing of voltage with a series connection is not possible). In the feedback analysis the amplifier element and the feedback network are modelled as a general 'two-port' systems (see Figure 6.15), from where the series and shunt connections are easily distinguished.

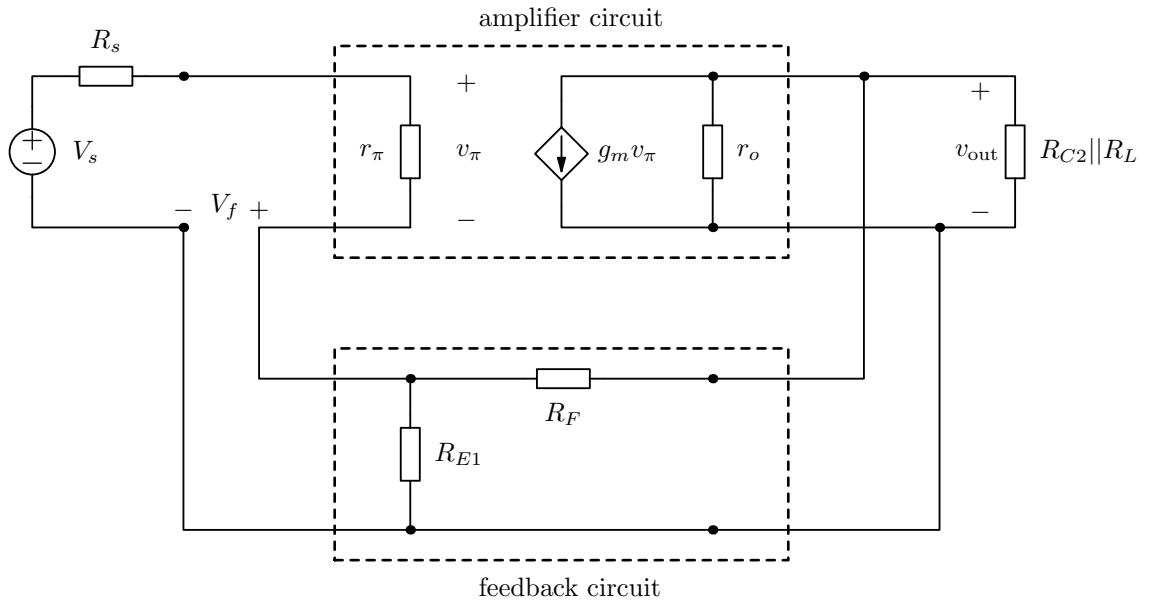


Figure 6.15: A voltage amplifier with series feedback

An example amplifier structure using the voltage-series configuration is shown in Figure 6.16. The feedback loop is implemented by connecting the collector of

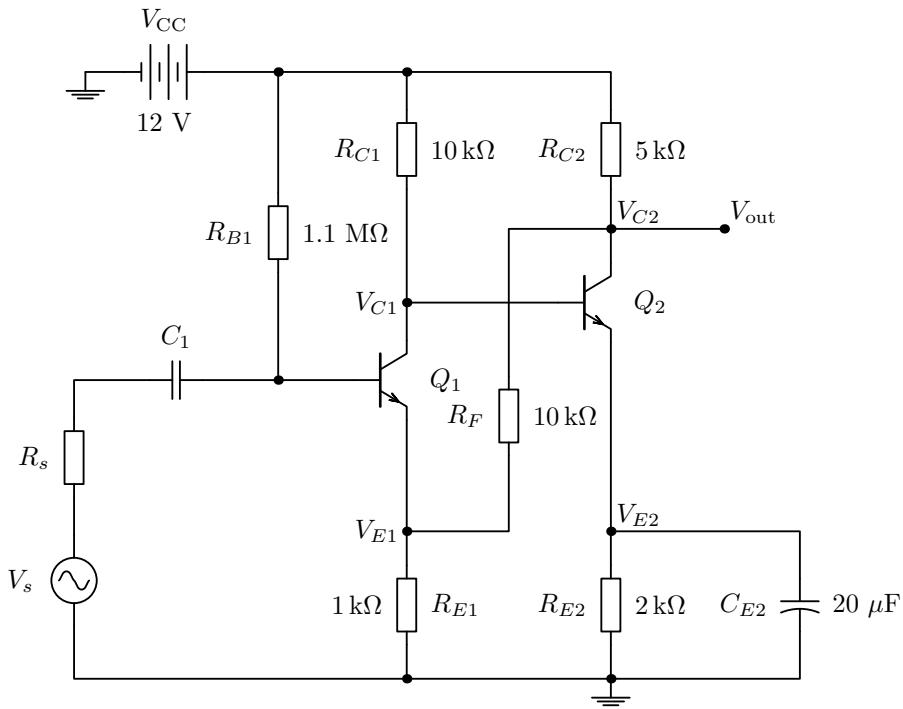


Figure 6.16: A direct-coupled pair amplifier with voltage-series feedback

the second amplifier stage to the emitter of the first amplifier stage via resistor R_F .

To see the effects of the feedback loop, the amplifier should be analysed from the viewpoint of DC and AC operation. The DC model is redrawn for analysis purposes in Figure 6.17. The analysis assumes that both of the transistors are biased in the forward-active region. The biasing is however sensitive to the current levels and either of the transistors is easily saturated in this connection. In this case, the feedback loop affects very little to the DC biasing, because the feedback only affects the potential V_{E1} and does not interfere with any significant transistor bias current.

The circuit in Figure 6.9 has three distinct voltage nodes, labelled as V_{C1} , V_{E1} and V_{C2} . Based on Kirchhoff's current law, the current equations for the nodes are:

$$I'_{E1} = I_{E1} + I_F = (\beta_{F1} + 1)I_{B1} + I_F$$

$$I'_{C1} = I_{C1} + I_{B2} = \beta_{F1}I_{B1} + I_{B2}$$

$$I'_{C2} = I_{C2} + I_F = \beta_{F2}I_{B2} + I_F$$

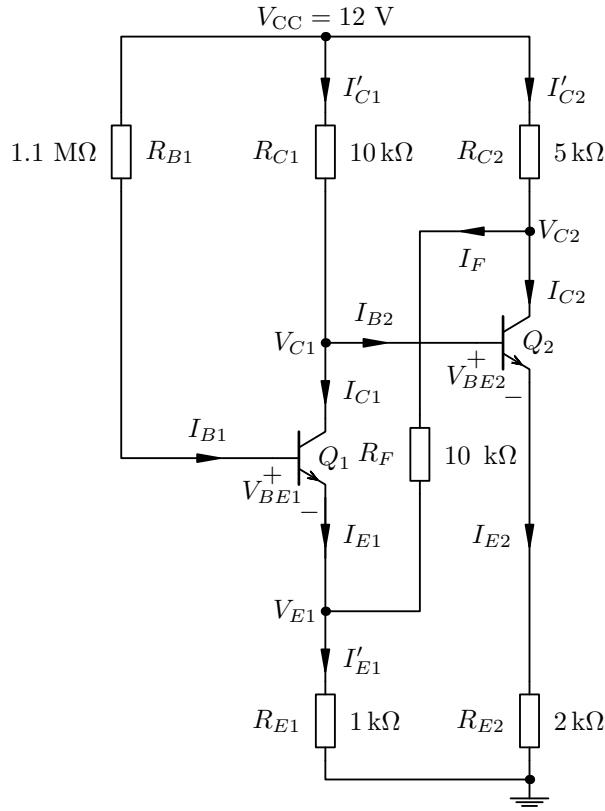


Figure 6.17: A DC model of the amplifier with voltage-series feedback

The essential currents can be expressed with the node voltages as:

$$\begin{aligned} I'_{E1} &= \frac{V_{E1}}{R_{E1}} & I'_{C1} &= \frac{V_{CC} - V_{C1}}{R_{C1}} & I_{B1} &= \frac{V_{CC} - V_{BE1} - V_{E1}}{R_{B1}} \\ I_F &= \frac{V_{C2} - V_{E1}}{R_F} & I'_{C2} &= \frac{V_{CC} - V_{C2}}{R_{C2}} & I_{B2} &= \frac{V_{C1} - V_{BE2}}{R_{E2}(\beta_{F2} + 1)}. \end{aligned}$$

Substitutions of the voltage equations into the current equations leads to the following three equations

$$\frac{V_{E1}}{R_{E1}} = (\beta_{F1} + 1) \frac{V_{CC} - V_{BE1} - V_{E1}}{R_{B1}} + \frac{V_{C2} - V_{E1}}{R_F}$$

$$\frac{V_{CC} - V_{C1}}{R_{C1}} = \beta_{F1} \frac{V_{CC} - V_{BE1} - V_{E1}}{R_{B1}} + \frac{V_{C1} - V_{BE2}}{R_{E2}(\beta_{F2} + 1)}$$

$$\frac{V_{CC} - V_{C2}}{R_{C2}} = \frac{\beta_{F2}}{\beta_{F2} + 1} \frac{V_{C1} - V_{BE2}}{R_{E2}(\beta_{F2} + 1)} + \frac{V_{C2} - V_{E1}}{R_F}.$$

These equations can be organised into a matrix equation, from where each of the node voltages can be solved using the Cramer's rule. The matrix equation is:

$$\begin{bmatrix} \frac{1}{R_{E1}} + \frac{\beta_{F1} + 1}{R_{B1}} + \frac{1}{R_F} & 0 & -\frac{1}{R_F} \\ -\frac{\beta_{F1}}{R_{B1}} & \frac{1}{R_{C1}} + \frac{1}{R_{E2}(\beta_{F2} + 1)} & 0 \\ -\frac{1}{R_F} & \frac{\beta_{F2}}{\beta_{F2} + 1} \frac{1}{R_{E2}} & \frac{1}{R_{C2}} + \frac{1}{R_F} \end{bmatrix} \times \begin{bmatrix} V_{E1} \\ V_{C1} \\ V_{C2} \end{bmatrix} = \begin{bmatrix} (\beta_{F1} + 1) \frac{V_{CC} - V_{BE1}}{R_{B1}} \\ \frac{V_{CC}}{R_{C1}} + \beta_{F1} \frac{V_{BE1} - V_{CC}}{R_{B1}} + \frac{V_{BE2}}{R_{E2}(\beta_{F2} + 1)} \\ \frac{V_{CC}}{R_{C1}} + \frac{\beta_{F2}}{\beta_{F2} + 1} \frac{V_{BE2}}{R_{E2}} \end{bmatrix}.$$

Based on this analysis, all the other biasing voltages and currents can be solved using the magnitudes of the node voltages.

Then one proceeds towards the AC analysis. Figure 6.18 shows the complete small-signal model of the system, but to ease out the calculations, the simplified model of Figure 6.19 is adequate. The simplifications concern the removal of DC-blocking capacitors and the high-valued biasing resistor R_{B1} .

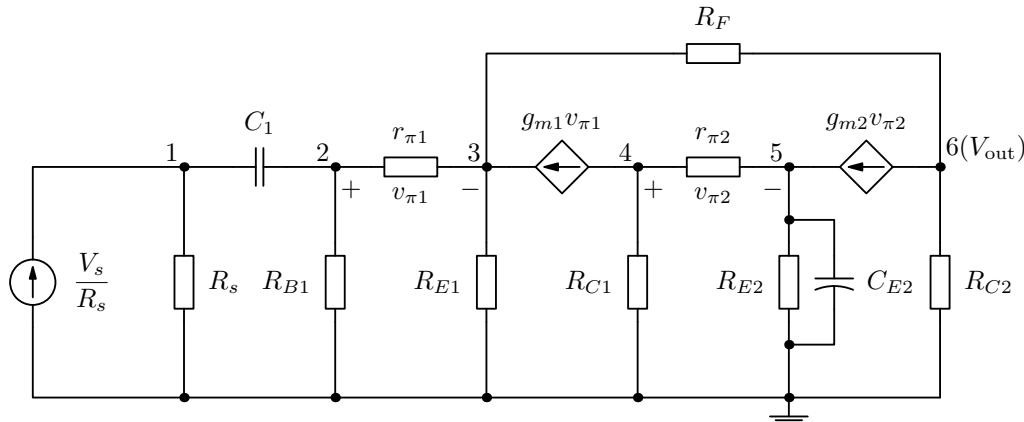


Figure 6.18: A small-signal model of the direct-coupled pair amplifier with voltage-series feedback

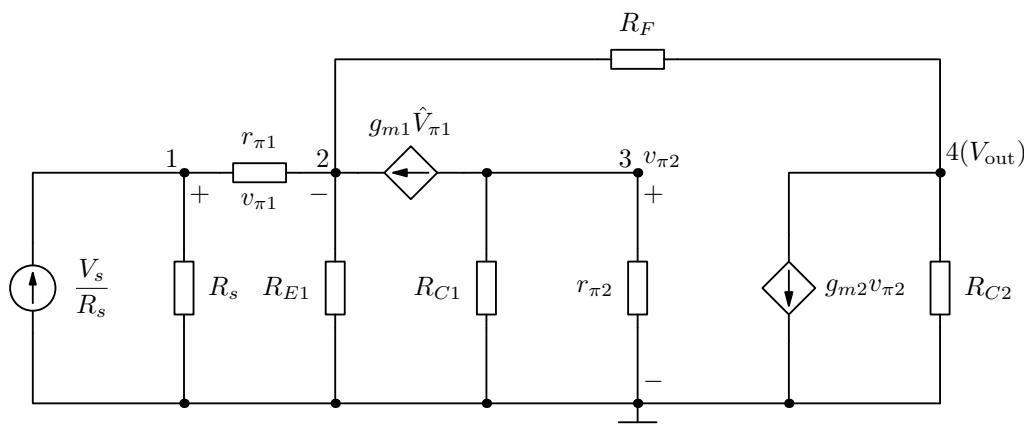


Figure 6.19: A simplified model for the mid-band frequencies

The analysis approach for the voltage-series case is the same as for the current-shunt feedback configuration. First, evaluate the feedback parameter t_{21} from its own small-signal model, where the output voltage source is set to zero ($X_o = V_o = 0$). This small-signal model is drawn in Figure 6.20, where the feedback from the output voltage has been eliminated by setting the output voltage to zero. The elimination of the output voltage enables to analyse the situation where only the input source V_s is active and hence utilising the principle of superposition.

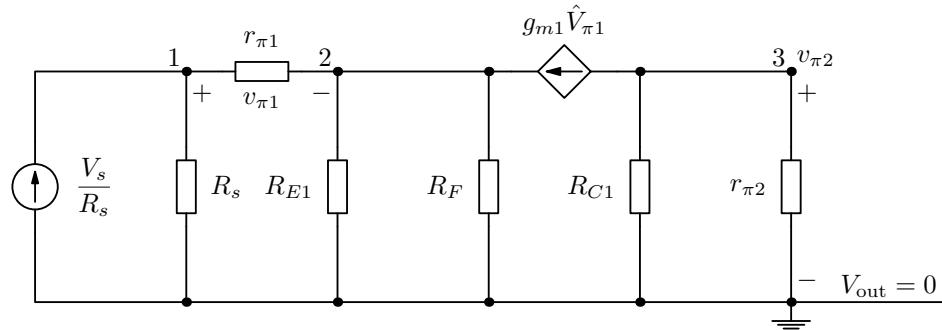


Figure 6.20: Small-signal model for evaluating the feedback parameter t_{21}

The dependent controlled source is chosen to be the first current-controlled current source and is identified in the figure and in the equations by the capitalised notation, $g_{m1}\hat{V}_{\pi1}$. The small-signal model of Figure 6.20 is mapped to the following matrix representation, where the admittance terms are collected to the main matrix and the current source terms are placed into the current vector. The voltage nodes in the voltage vector are the variables to be solved from this equation.

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi1}} & -\frac{1}{r_{\pi1}} & 0 \\ -\frac{1}{r_{\pi1}} & \frac{1}{r_{\pi1}} + \frac{1}{R_{E1}} + \frac{1}{R_F} & 0 \\ 0 & 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_s}{R_s} \\ g_{m1}\hat{V}_{\pi1} \\ -g_{m1}\hat{V}_{\pi1} \end{bmatrix} \quad (6.27)$$

For evaluation purposes it is convenient that all controlled source terms from the current vector are moved to the admittance matrix. Because the matrix equation is just a set of linear equations, the controlled source terms are subtracted or added to both sides of the equation. It is also necessary to identify to which

voltage node variables the controlled source term relates to. Based on this discussion, the matrix equation (6.28) is readily modified in the desired form.

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} & -\frac{1}{r_{\pi 1}} & 0 \\ -\frac{\beta_{F1} + 1}{r_{\pi 1}} & \frac{\beta_{F1} + 1}{r_{\pi 1}} + \frac{1}{R_{E1}} + \frac{1}{R_F} & 0 \\ \frac{\beta_{F1}}{r_{\pi 1}} & -\frac{\beta_{F1}}{r_{\pi 1}} & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_s}{R_s} \\ 0 \\ 0 \end{bmatrix} \quad (6.28)$$

Since $v_{\pi 1}$ is not directly any node variable, it has to be calculated as the difference $V_1 - V_2$. The feedback parameter t_{21} is then found to be

$$t_{21} = \frac{v_{\pi 1}}{V_s} = \frac{V_1 - V_2}{V_s} = \frac{r_{\pi 1}(R_{E1} + R_F)}{(r_{\pi 1} + R_s)(R_F + R_{E1}) + (\beta_{F1} + 1)R_{E1}R_F}. \quad (6.29)$$

As already noted, as a single parameter the t_{21} does not tell much about the properties of the circuit. Because the aim is to find the expressions for β and T , the battle with the equations continues. For evaluating T and t_{12} the input source needs to be suppressed ($X_s = V_s$) and let the output source be active in turn. The small-signal model of Figure 6.21 provides a suitable model for the job.

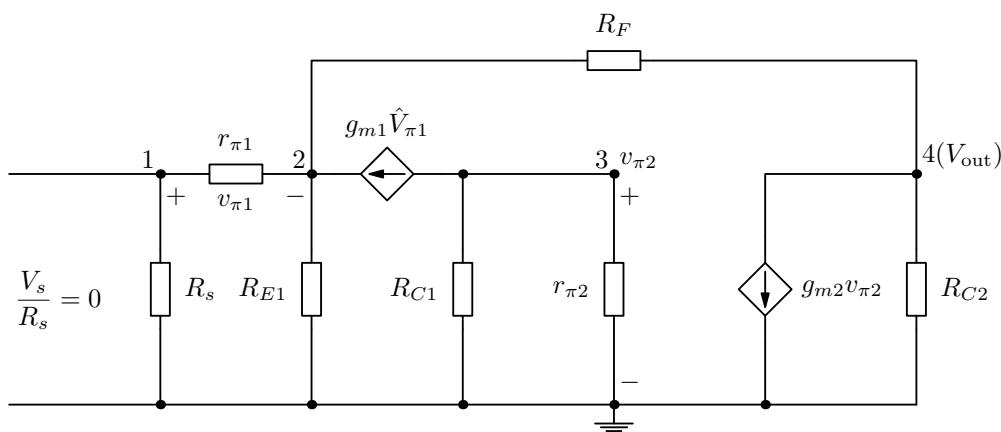


Figure 6.21: Small-signal model for feedback parameters T and t_{12}

From this model, the following matrix equation is written:

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} & -\frac{1}{r_{\pi 1}} & 0 & 0 \\ -\frac{1}{r_{\pi 1}} & \frac{1}{r_{\pi 1}} + \frac{1}{R_{E1}} + \frac{1}{R_F} & 0 & -\frac{1}{R_F} \\ 0 & 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} & 0 \\ 0 & -\frac{1}{R_F} & 0 & \frac{1}{R_{C2}} + \frac{1}{R_F} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 0 \\ g_{m1}\hat{V}_{\pi 1} \\ -g_{m1}\hat{V}_{\pi 1} \\ -g_{m2}\hat{V}_{\pi 2} \end{bmatrix}.$$

To ease out the evaluation process, all but one of the dependent controlled source terms are moved from the current vector to the admittance matrix. After moving the terms, the node variables are ready to be solved from matrix equation using the Cramer's rule.

$$\begin{bmatrix} \frac{1}{R_s} + \frac{1}{r_{\pi 1}} & -\frac{1}{r_{\pi 1}} & 0 & 0 \\ -\frac{\beta_{F1} + 1}{r_{\pi 1}} & \frac{\beta_{F1} + 1}{r_{\pi 1}} + \frac{1}{R_{E1}} + \frac{1}{R_F} & 0 & -\frac{1}{R_F} \\ 0 & 0 & \frac{1}{R_{C1}} + \frac{1}{r_{\pi 2}} & 0 \\ 0 & -\frac{1}{R_F} & g_{m2} & \frac{1}{R_{C2}} + \frac{1}{R_F} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -g_{m1}\hat{V}_{\pi 1} \\ 0 \end{bmatrix}.$$

To find the return ratio T , the fraction $-\frac{v_{\pi 1}}{\hat{V}_{\pi 1}}$ needs to be solved. The result of the lengthy calculation is

$$T = \frac{\beta_{F1}R_{C1}\beta_{F2}R_{C2}}{R_{C1} + r_{\pi 2}} \frac{R_{E1}}{(R_{C2} + R_{E1} + R_F)(R_s + r_{\pi 1}) + (\beta_{F1} + 1)R_{E1}(R_{C2} + R_F)}.$$

The same circuit model can be reused to evaluate the feedback parameter t_{12} , which is defined as $\frac{V_4}{\hat{V}_{\pi 1}}$. The equation for t_{12} is

$$t_{12} = \frac{\beta_{F1}R_{C1}\beta_{F2}R_{C2}}{R_{C1} + r_{\pi 2}} \frac{\frac{1}{r_{\pi 1}}[(r_{\pi 1} + R_s)(R_F + R_{E1}) + (\beta_{F1} + 1)R_{E1}R_F]}{(R_{C2} + R_{E1} + R_F)(R_s + r_{\pi 1}) + (\beta_{F1} + 1)R_{E1}(R_{C2} + R_F)}.$$

The t_{22} parameter is obtained by dividing T by t_{12} . Hence,

$$t_{22} = -\frac{T}{t_{12}} = -\frac{r_{\pi 1}R_{E1}}{(r_{\pi 1} + R_s)(R_F + R_{E1}) + (\beta_{F1} + 1)R_{E1}R_F}.$$

Then finally the feedback factor β is obtained by dividing t_{22} by t_{21} and the result of this calculation is

$$\beta = \frac{t_{22}}{t_{21}} = -\frac{R_{E1}}{R_{E1} + R_F}.$$

So here we got the expected result, which could have been guessed by directly looking at the circuit diagram. Here one did not even need to make any simplifications during the evaluation process, therefore the resistive feedback network does not interact with any of the surrounding passive components. One interesting fact is that the feedback factor is negative in this case. This actually a good thing because the output voltage in this circuit is connected as positive feedback to the emitter of the first transistor. From there the signal is inverted to the input side and from there comes the minus sign. In the current-shunt example the output signal was already inverted with respect to the input, that is why the feedback factor there was positive.

Finally the input impedance of the voltage-series feedback circuit is analysed just for the sake of completeness. Following the standard procedure, the dead-system impedance is evaluated from the model where the dependent controlled source is suppressed. This model is drawn into Figure 6.22.

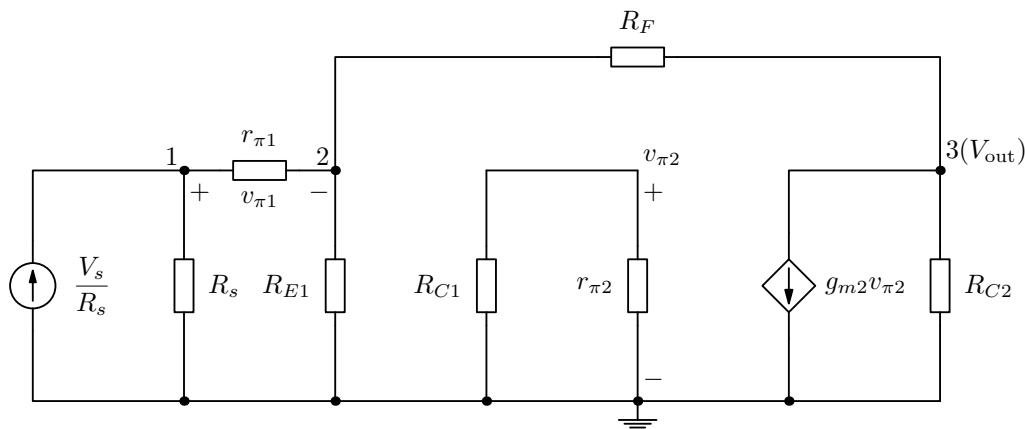


Figure 6.22: Small-signal model for evaluating the input impedance

By direct inspection one can easily see that the dead-system input impedance, when looking into the amplifier input terminals, is $r_{\pi1} + R_{E1}||(R_F + R_{C2})$. The middle part of the circuit is completely isolated and connected to ground, so the potential between R_{C1} and $r_{\pi2}$ is at the ground level. Note that if the dependent source had been chosen to be the second one, the resulting equation for the input impedance would have been different. This is a bit difficult to accept, since the main feedback parameters, T , A_{OL} and β are the same even when choosing the

second controlled source as the dependent one.

Anyway, next one needs to analyse the situations for $R_s \rightarrow \infty$ and $R_s \rightarrow 0$ in the expression of the return ratio. Setting $R_s = \infty$ results to $T = 0$ and $R_s = 0$ results to a reasonable value. Hence, the input impedance with feedback is defined as $Z_{IF} = Z_{ID}(1 + T(R_s \rightarrow 0))$, where

$$T(R_s \rightarrow 0) = \frac{\beta_{F1} R_{C1} \beta_{F2} R_{C2}}{R_{C1} + r_{\pi 2}} \frac{R_{E1}}{(R_{C2} + R_{E1} + R_F)r_{\pi 1} + (\beta_{F1} + 1)R_{E1}(R_{C2} + R_F)}.$$

A direct conclusion is that voltage-series type of feedback increases the input impedance by the factor of $(1 + T)$. However, according to this analysis, the effect of the emitter follower, which also increases the input impedance, is lost when the global feedback loop is applied. This conclusion might be wrong, if a mistake was made when analysing the dead-system input impedance.

From another point of view, Figure 6.23 indicates a situation where the Miller theorem is used to break the feedback loop into two separate resistors.

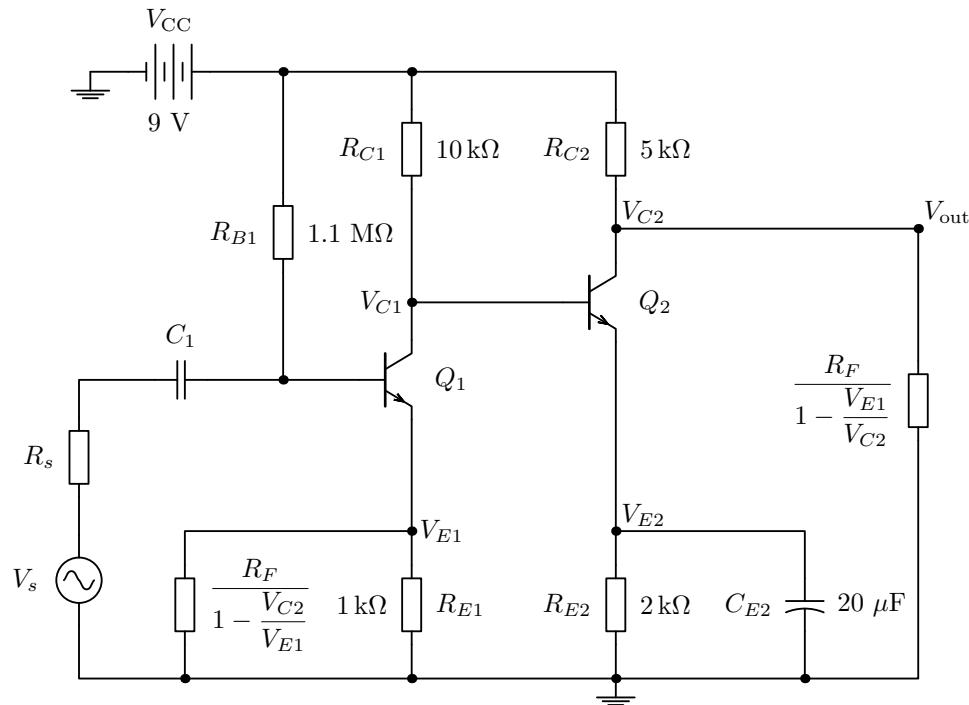


Figure 6.23: The feedback impedance split according to the Miller theorem

The rescaled feedback resistor R_F appears in parallel with the load resistor (look at R_{C2} in the small-signal model in this case) and also in parallel with the emitter resistor R_{E1} . The positive feedback reflects R_F as *negative impedance* in parallel with R_{E1} . This negative impedance increases the total emitter resistance. The

final step towards negative feedback to the input terminals is arranged as a *local* feedback loop from the emitter of Q_1 .

The concepts *global* and *local* feedback basically relate to the amount of components which the feedback loop encloses within. If the feedback loop is taken from the output to the input over one single transistor, the feedback is localised and therefore the feedback is commonly referred to as local feedback. If the feedback loop extends over a complete circuit stage with several amplifier elements, then the feedback is categorised as global feedback.

Both of the examples already given in this context are more or less examining the feedback in the global scope, although the voltage-series example was a combination of global and local feedback networks. The most common type of local feedback is found from a simple single-stage BJT amplifier with an emitter resistor, which is drawn in Figure 6.24. In practical transistor amplifiers the emitter resistor is typically bypassed for AC-signals with a capacitor, but when the capacitor is not there, the effective input signal is facing a feedback loop from the emitter to the input.

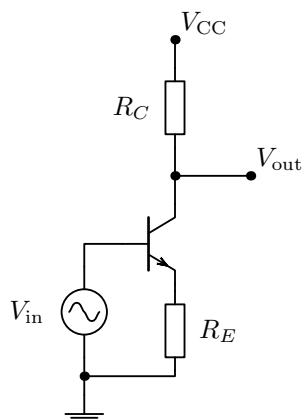


Figure 6.24: A BJT amplifier with an unbypassed emitter resistor

Figure 6.25 shows a small-signal model of the common-emitter amplifier and indicates that the local feedback structure of an unbypassed CE-stage is of the series-series type, where the current from the output is sampled and converted into voltage in the feedback loop. The voltage due to the output current over the emitter resistor is therefore seen as inverted voltage at the input side of the transistor.

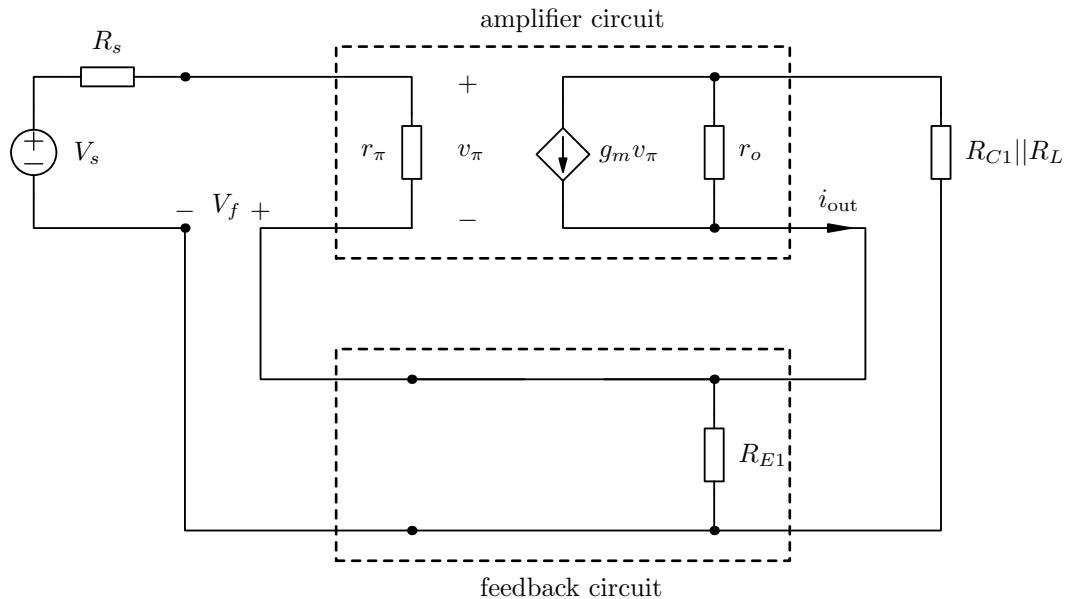


Figure 6.25: A series-series feedback network of the BJT with an emitter resistor

6.2.4 A SPICE model for the fuzz circuit

The fuzz circuit schematic shown in Figure 6.26 was redrawn with the gschem schematic editor, from where it is easy to prepare a netlist for SPICE simulations. The schematic includes all the components assigned with their actual values. The potentiometers R_4 and R_6 are split into two separate resistors so that the voltage divider functionality can be simulated by changing simultaneously the values of the two resistors.

In addition to the real resistor and capacitor values, all other necessary component parameters, such as transistor gains and saturation currents, need to be defined in gschem so that it would be straightforward to generate a complete netlist for ngspice simulations. The way to assign these additional model parameters is explained in section 1.3.12. The signal source parameters should also be defined to suit both direct current and alternative current simulations.

When using the gEDA design tools collection in Linux, the netlist for SPICE is created from the schematic of Figure 6.26 by using the command

```
gnetlist -g spice-sdb -o fuzz_net.net fuzz.sch
```

where the output file named as `fuzz_net.net` is obtained from the schematic file `fuzz.sch`. Because the fuzz circuit has only a small amount of components, the barebone netlist looks relatively simple, but ...

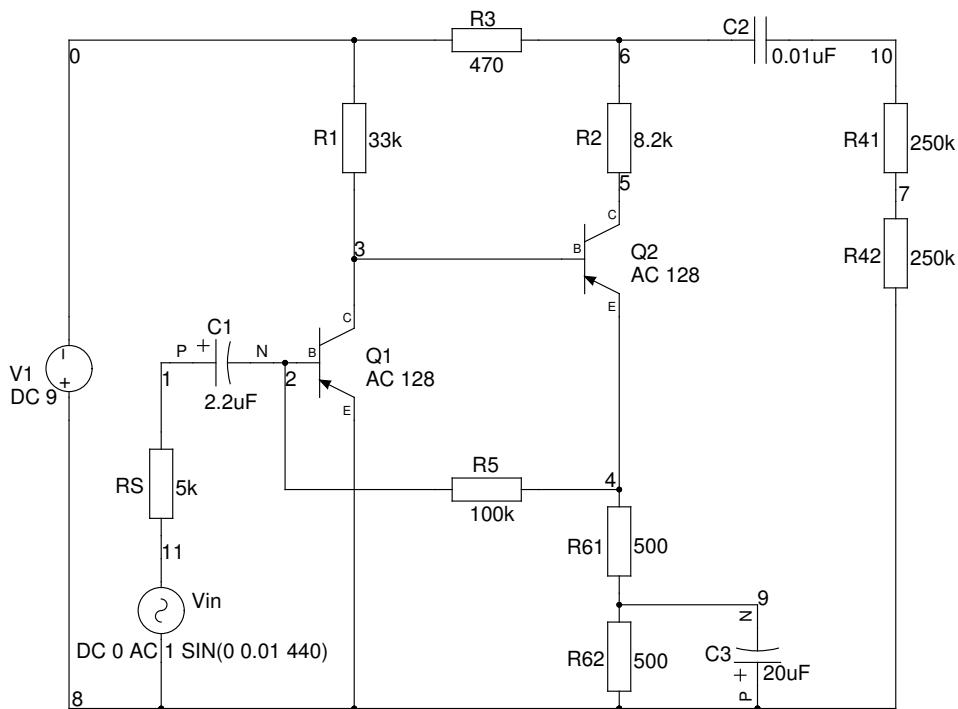


Figure 6.26: The fuzz schematic drawn with gschem

```

=====
 Begin SPICE netlist of main design =====
 Vin 8 11 DC 0 AC 1 SIN(0 0.01 440)
 RS 11 1 5k
 R42 8 7 250k
 R41 7 10 250k
 R62 8 9 500
 R61 9 4 500
 V1 8 0 DC 9
 C2 6 10 0.01uF
 C3 8 9 20uF
 Q2 5 3 4 AC128
 R2 6 5 8.2k
 R3 0 6 470
 R1 3 0 33k
 C1 1 2 2.2uF
 Q1 3 2 8 AC128
 .MODEL AC128 PNP (????)
 R5 2 4 100k
 .end

```

... there is still one problem to solve. The component model for the AC 128 BJT is deliberately replaced with question marks because there is no official model available. The SPICE model for this component needs to be defined by other means.

The corner stone to simulate the fuzz effect is to have a realistic component model for the old and leaky germanium transistors. For a novice junior scientist, it is quite difficult to start creating a SPICE model to depict the behaviour of germanium transistors. More difficulties arise from the fact that the default transistor parameters in SPICE are aimed to model an average silicon transistor, since all discrete transistors are nowadays made of silicon.

Because the aim is to run simulations at audio frequencies, one can neglect the differences of germanium and silicon transistors in the high-frequency range. These high-frequency differences are mainly due to the junction capacitances, which in germanium transistors limit the high-frequency response quite significantly.

After neglecting the high-frequency properties, the only things to be determined are the parameters that control the DC behaviour of the transistor. These parameters in SPICE3 standard used by ngspice are the saturation current parameter 'IS', forward current gain control parameters 'BF', 'NF', 'ISE', 'IKF', and 'NE', the corresponding reverse current gain control parameters 'BR', 'NR', 'ISC', 'IKR' and 'NC', and the Early voltages 'VAF' and 'VAR' for the forward- and reverse-active modes respectively.

These parameters control the curves of collector bias current I_C and base bias current I_B , which change as a function of the base-to-emitter voltage V_{BE} . In simplified manual calculations, these curves are expected to behave in a similar fashion since the forward current gain factor β_F is assumed to be constant so that the relation $I_C = \beta_F I_B$ holds at all times. To replicate this simple 'constant β_F ' model in SPICE, the only parameters that need to be defined are IS, BF and VAF for the forward-active mode. The parameters that are left undefined in the SPICE model will take the default values, which should support the 'constant β_F ' approach.

In reality, β_F changes for different values of I_C , and this dependency between β_F and I_C is clearly indicated in every BJT datasheet one can find. For very low and very high values of I_C , the current gain factor β_F is lower than it is in the middle range of the I_C values. In SPICE3 based simulation programs it is possible to simulate this situation by defining the extra leakage parameters ISE, ISC along with NE and NC and the high current roll-off parameters IKF and IKR. Figure 6.27 explains the meaning of these extra parameters.

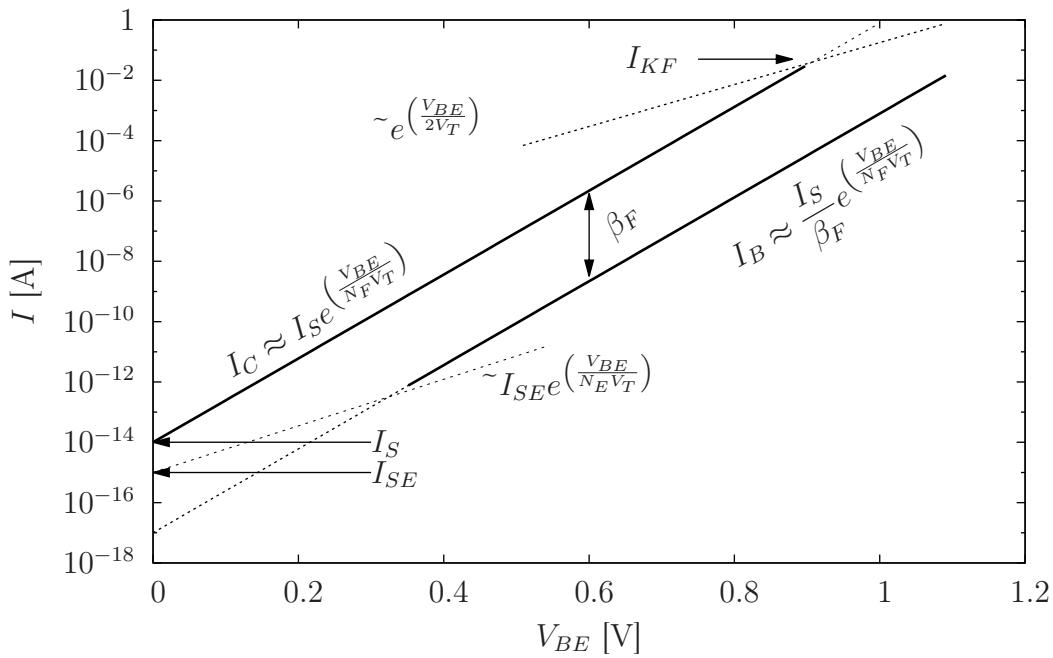


Figure 6.27: SPICE parameters controlling β_F

With reference to Figure 6.27, IS is the y -axis intercept point of the I_C curve at $V_{BE} = 0$, and NF defines the slope of the line. In SPICE, NF is commonly assumed to have a value of 1, although some BJT SPICE models set NF as slightly larger than one. The leakage currents at low values of collector current introduce an additional slope to the plot of I_B . This effect can be controlled by the parameters ISE and NE. IKF controls the point where β_F starts to drop at high collector currents due to the internal resistances of the transistor *pn*-junctions.

The primary step for modelling a certain transistor is to find a datasheet for it. The AC 128 is still in production at the time of writing and one relatively informative datasheet by Valvo Radioröhrenfabrik GmbH was found from their Internet databases. The datasheet does not contain any predefined SPICE model, so one needs to construct the model based on the curves presented in the datasheet. An example to extract SPICE parameters from BJT datasheets is given by N. Malik [70], and that procedure is adapted partially also in this case.

The best way to get started is to set a few basic parameters for DC operation and see how the fuzz circuit biases with those parameters. The most important parameters to start with are the saturation current I_S and the forward active mode current gain factor β_F (sometimes also referred to as h_{FE}). The average peak current gain of AC 128 is mentioned in the datasheet as $\beta_F = 90$, and it stays at this level in the range from $I_C = 50$ mA to $I_C = 450$ mA. Outside of this

given range of collector current, the value of β_F is slightly lower. The saturation current is not stated explicitly but it can be determined graphically from the graph of I_C versus V_{BE} given in the datasheet.

The simplest way is to pick one point close to the operating point value of I_C and make use of the formula defined in the SPICE2 standard, where

$$I_C \approx I_S e^{\left(\frac{V_{BE}}{N_F V_T}\right)} \left(1 + \frac{V_{CE}}{V_{AF}}\right).$$

This equation is given by Sedra [71, p. 146] as a first order approximation of the Gummel-Poon collector current SPICE equation for BJTs. More detailed equations for SPICE2 are given by Nagel [72], and for SPICE3 by Steer [73], for example. These documents are available in .pdf format for reading on the Internet. For further approximations, the Early voltage V_{AF} can be set to infinity so that the expression for the collector current is then almost identical to the so-called Shockley equation for diodes. The following analysis uses this more radical approximation of

$$I_C \approx I_E \approx I_S e^{\left(\frac{V_{BE}}{N_F V_T}\right)}$$

from where

$$I_S \approx I_E e^{\left(\frac{-V_{BE}}{N_F V_T}\right)}.$$

The data of an average I_E - V_{BE} curve was determined from the graph of the AC 128 datasheet by tabulating the values of I_C on constant increments of V_{BE} . The collected data is plotted in Figure 6.28 to show what kind of a device we are dealing with.

The exponential curve in Figure 6.28 is drawn through the points determined from the datasheet and the same points are also drawn after taking a logarithm from the values of I_E . A linear fit is applied to the logarithmic points and from the y -axis intersect one can determine the saturation current I_S . According to the graph, a typical saturation current for AC 128 seems to be about $20 \mu\text{A}$. The datasheet also provides deviation graphs, which indicate that the saturation current can be inside a range of $9 - 25 \mu\text{A}$. These are very high values in comparison to the values of silicon BJTs, which typically have the saturation current as 10^{-14} amperes in the order of magnitude.

Additionally, the Early voltage V_A can also be determined from the graphs given in the AC 128 datasheet. Figure 6.29 depicts the idea of graphically determining the Early voltage from a given set of V_{CE} - I_C - I_B curves. The Early voltage

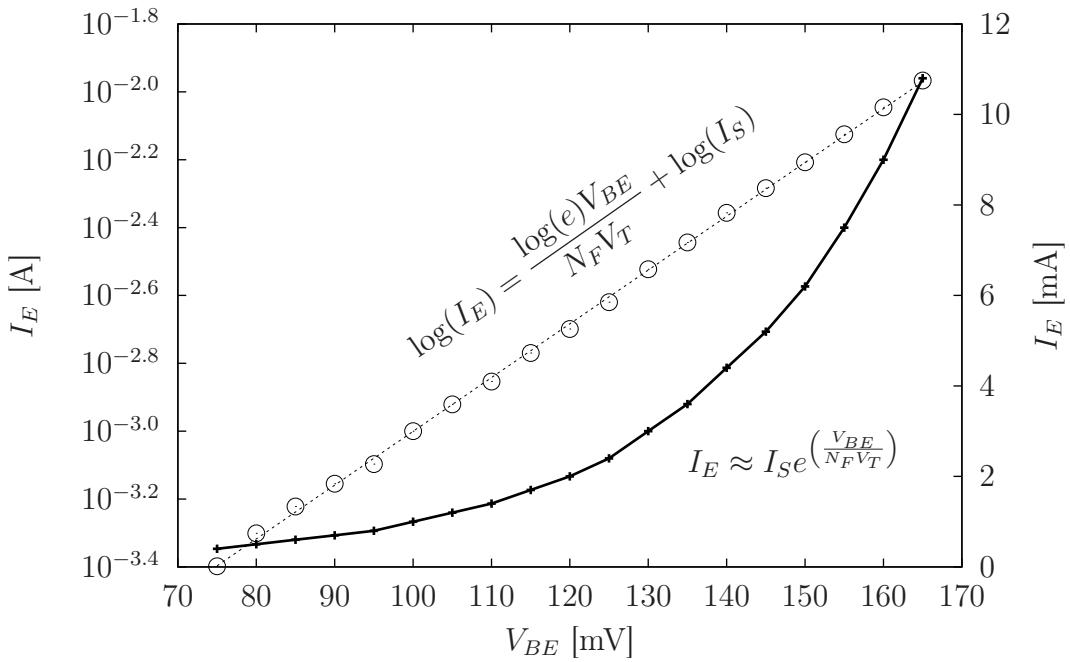


Figure 6.28: Determining the saturation current from a datasheet

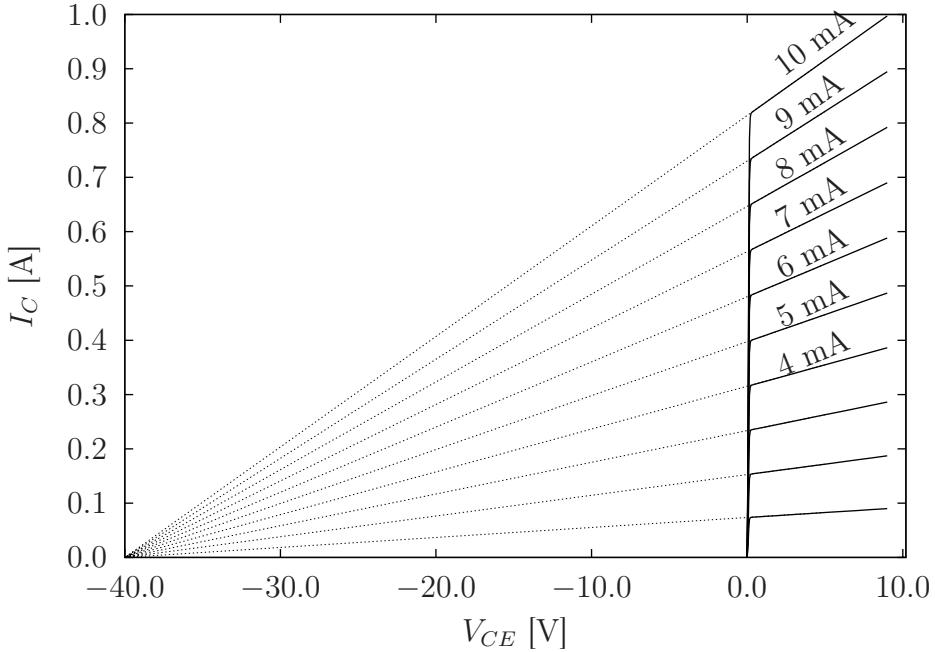


Figure 6.29: The Early voltage V_A versus current-voltage characteristics

is the x -axis intercept point of the continued I_B curves. In this case, V_{AF} was approximated to be -40 V for the AC 128 germanium transistor. This value is set as a positive value to the SPICE parameter VAF.

Now all the basic parameters are determined and the AC 128 BJT model can be

tested in SPICE with the values $IS = 20e-6$, $BF = 90$, $VAF = 40$ and $NF = 1$. The simplest way to test the rationality of the model is to create a circuit diagram that depicts the actual situation to define the $I_E - V_{BE}$ and $I_C - V_{CE}$ graphs. Figures 6.30a and 6.30b show simple circuits for this purpose. The measurement setup to determine the $I_C - V_{CE}$ characteristics is often referred to as a *curve tracer* arrangement [71, p. 149].

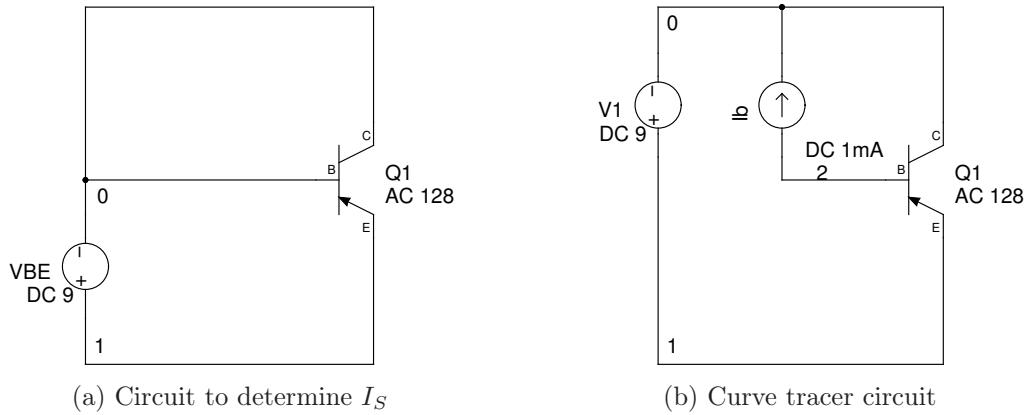


Figure 6.30: Circuits to simulate BJT characteristics

The SPICE input files are created from these circuits using a `gnetlist` command, and a command sequence for running a specific set of simulations is placed to both netlist files. The primary interest is to see whether the defined values lead to a similar $I_E - V_{BE}$ dependency in the simulation as in the datasheet graph. The complete simulation file for the circuit in Figure 6.30a is

```

.control
dc VBE 0.075V 0.180V 0.005V
gnuplot saturat -dc1.I(VBE)
.endc
***** Begin SPICE netlist of main design ****=
VBE 1 0 DC 9
Q1 0 0 1 PNP1
.MODEL PNP1 PNP (IS=20u BF=90 VAF=40 NF=1)
.end

```

where a DC sweep with increments of 0.005 V is made for the DC source acting as the potential difference V_{BE} . The current values measured from that source indicate the current I_E , which is eventually plotted as a function of V_{BE} . The simulation result was in agreement with the datasheet curve, so after this test the model is still acceptable.

Next comes the curve tracer test. In the SPICE simulation, the voltage source is given a DC sweep from 0 V to 9 V and the current source is swept from 1 mA to 10 mA.

```
.control
dc V1 0V 9V 1mV Ib 1m 10m 1m
.endc
***** Begin SPICE netlist of main design *****
Ib 2 0 DC 1mA
V1 1 0 DC 9
Q1 0 2 1 PNP1
.MODEL PNP1 PNP (IS=20u NF=1 BF=90 VAF=40)
.end
```

The resulting set of curves are in accordance with the datasheet. Based on these two tests, the model is a good representation of the AC 128 bipolar junction transistor. The next step is to use this model in the actual fuzz circuit description file, which was presented in the same context with the gschem schematic drawing 6.26. To test the DC bias voltages of the circuit, it is enough to run the operating point (OP) analysis for the fuzz circuit netlist. The simplest way to run this analysis is to add a new line at the end of the SPICE circuit description file and write '.OP' there. Then run the analysis using the command

```
ngspice -b fuzz_net.net
```

and the operating point information is printed to the screen.

From the operating point information it is easy to see that the model is not working correctly in the actual fuzz circuit because in the stable biasing state the V_{BE} of Q_2 is only about 0.02 V. The model needs to be extended to cover the leakage currents I_{CB0} and I_{EB0} , which refer to the collector and emitter currents when the transistor is in a cut-off state. According to the datasheet, these currents can be as high as 20 μ A but typically about 4 μ A at room temperature. This value is almost of the same order of magnitude as the saturation current I_S , so it plays a significant role in the SPICE model. Along with these cut-off currents, one needs to define a reverse current gain factor β_R , which is much smaller than the forward current gain, typically in the range from 1 to 5.

The test circuits for determining the leakage currents I_{CB0} and I_{EB0} are presented in Figure 6.31. The tweaking of the leakage currents started a long sequence of trial and error. Basically it has to be admitted that without knowing the true V_{BE} that was measured from the prototype circuit, these empirically determined

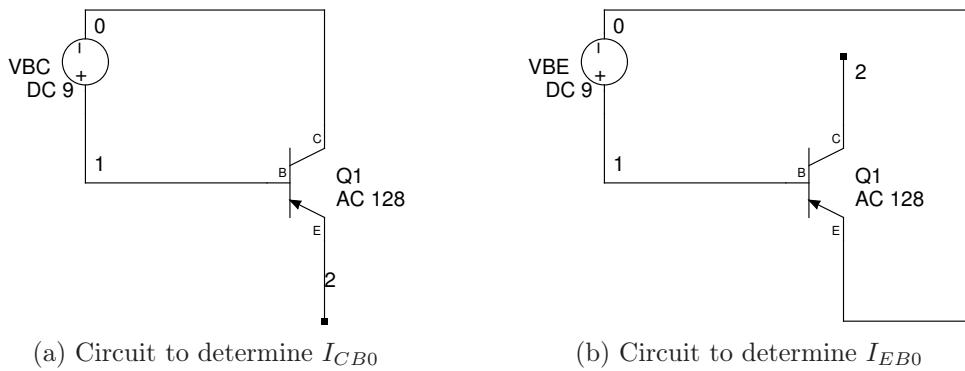


Figure 6.31: Circuits for measuring leakage currents I_{CB0} and I_{EB0}

SPICE parameters would never have given a realistic behaviour of the fuzz circuit. This is because both of the transistors in the fuzz circuit are biased so close to the saturation and cut-off states where the leakage effects and other anomalies have a relatively large effect on the bias voltages of the circuit. Eventually, after several hundred rounds of tweaking the leakage values and testing the model with the saturation, curve tracer and leakage measurement circuits, a model of

```
.MODEL AC128 PNP (IS=10u NF=1 NR=1 ISE=0.5u ISC=1u NC=1.5 NE=1.5 BF=90
+ BR=5 VAF=40 VAR=40 EG=0.67 VJE=0.25 VJC=0.25 VJS=0.25)
```

was accepted. The parameter values of VAR, EG, VJE, VJC and VJS have been included as theoretical values for the sake of 'completeness', but they do not have any effect on the biasing. This model gives a V_{BE} of about 0.09 V in the fuzz circuit, the I_{CB0} is around $3.5 \mu\text{A}$ and the set of curves given by the curve tracer are in a relatively good agreement with the graph presented in the datasheet. With the accepted model, the saturation curve is still a little bit out of range from the nominal curves given in the datasheet. Figure 6.32 shows the extreme curves along with the medium curve obtained from the datasheet, and the accepted SPICE model ends up a bit outside the right extreme curve. After all the effort put into this, this will have to do.

Another way to reach $V_{BE} \approx 0.1 \text{ V}$ is simply to adjust NF to 0.5, which is half of the default value, and then take IS somewhere around 10^{-8} . This kind of model is basically justified since in [15, p. 49] it is written that in germanium diodes I_S is a couple of orders of magnitude larger than in silicone diodes and that NF is about half the value of silicone diodes. In the same reference it says that $V_D \approx 0.2$, whereas silicon diodes have $V_D \approx 0.6$. It has to be emphasised that the results obtained from this approach are not in agreement with the datasheet curves,

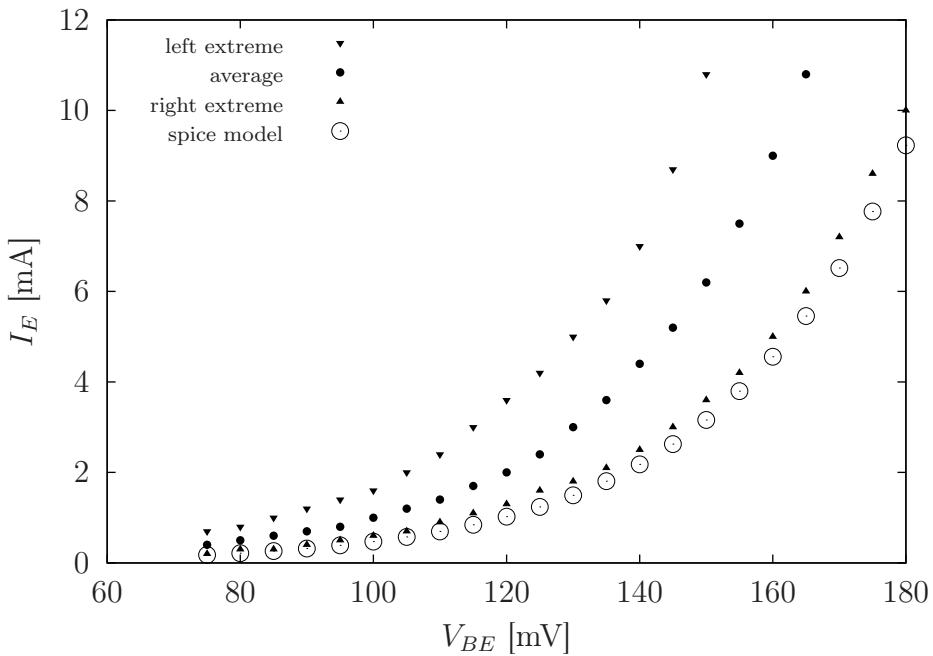


Figure 6.32: The accepted model against the datasheet expectations

but the alternative method works as a simple approximation. More theoretical information about the differences of germanium and silicon transistors is given in old transistor circuit analysis books, such as [74], [75] and [67].

The actual simulations with the fuzz circuit concern the operating point analysis for the transistors and a frequency response (AC) analysis with different values of potentiometer R_6 . Also a transient analysis needs to be run with different values of the input voltage level to show the actual waveform obtained from the simulated circuit. To run all of these analyses automatically with a single netlist file, the following control parameters are added to the beginning of the netlist file.

```
.control
ac dec 90 10 100K
alter R61 100
alter R62 900
ac dec 90 10 100K
alter R61 900
alter R62 100
ac dec 90 10 100K
set filetype=ascii
write fuzzdata.txt db(ac1.v(7)) db(ac2.v(7)) db(ac3.v(7))
gnuplot fuzz_ac db(ac1.v(7)) db(ac2.v(7)) db(ac3.v(7))
```

```

alter R61 200
alter R62 800
tran 1us 30ms
alter Vin sin [ 0 0.1 440 ]
tran 1us 30ms
gnuplot fuzz_tran tran3.v(7) tran4.v(7)
.endc

.OP

```

Once the control parameters are in place, the actual simulation run is started from the command line with the command

```
ngspice -b fuzz_net.net
```

This command starts to read the command sequence from the netlist file, which runs the analyses automatically and stores the results into the files specified in the netlist commands. The results for the operating point DC voltages are printed to the command line terminal at the end of the analysis. It has to be noted that the gnuplot command does not work properly with version 20 of ngspice but it creates nice data files for further processing. The command sequence also includes a command for writing the results into an ASCII encoded text file, but for plotting purposes the data file created by the gnuplot command is better.

6.2.5 Results of simulations, calculations and measurements

Regarding the transistor bias analysis of section 6.2.1, the component values from schematic 6.2 were inserted to the equations. The numerical results were obtained for all transistor pin voltages and for the voltage node enumerated as 6 in Figure 6.3. The internal potential difference V_{BE} of the AC 128 transistors in the fuzz circuit was measured with a multimeter to be approximately 0.1 V, and this value was also used in the theoretical calculations. The measured base-emitter voltage matches relatively well to the approximation made from the graphs of the datasheet when using a 1 mA collector bias current.

For the SPICE simulation part, the netlist from the circuit in Figure 6.26 was loaded into ngspice and analysed with the operating point (.OP) command, which was written inside the netlist file. The simulation results for DC quiescent voltages were read directly from the output listing provided by ngspice.

For measurement purposes, the fuzz circuit using AC 128 transistors and 10 % precision carbon composition resistors and electrolytic capacitors was built onto

a small solderless breadboard, a.k.a. a plugboard, where the components can be replaced easily just by plugging them to the board. A regular multimeter was used to measure the transistors' quiescent voltages.

The data of the simulated, calculated and measured quiescent voltages in the static configuration are tabulated in Table 6.1 for comparison purposes. The

Table 6.1: Fuzz circuit quiescent voltages

value	simulated	calculated	measured
V_{B1}	8.91	8.90	8.92
V_{C1}	8.49	8.51	8.49
V_{E1}	9.00	9.00	9.02
V_{B2}	8.49	8.51	8.49
V_{C2}	3.60	3.30	3.80
V_{E2}	8.58	8.61	8.62
V_6	0.20	0.18	0.21

results seem to be relatively equal, except for the collector voltage V_{C2} of transistor Q_2 . Most likely this difference appears because of the leakage current of the AC 128 transistors since small changes in ISE and ISC parameters had a significant effect on this value in the SPICE simulation. The actual AC 128 transistors were bought 'off the shelf' from a nearby electronics store and they were used as is. At least one out of four transistors did not eventually work in this circuit at all, so manual selection is often needed when using these old germanium transistors. Based on the amount of tweaking needed for the SPICE model to work correctly, the manual calculation in this case is the more effortless way to obtain reasonably accurate results.

Electronics designers did not have SPICE in the 60's when the fuzz effect circuit was designed. It is therefore clear that the designer calculated the biasing conditions manually just by taking the β_F as 90 from the datasheet and approximating the order of magnitude of the collector currents in order to get the value of $V_{BE} \approx 0.1$ volts from the datasheet. The AC 128 is primarily intended to be used in power amplifier applications. When used in a battery operated circuit, the low collector currents result in a low value of V_{BE} . This is why in the case of the fuzz effect one cannot assume typical $V_{BE} \approx 0.2$ volts. Was this intentional from the designer to use a power amplifier with low currents, who knows, but

the main thing is that the effect works like a charm.

The AC analysis of the fuzz circuit is not very informative because the DC clipping of the signal also distorts the frequency response of the circuit. The frequency content of the clipped waveform is much more rich in the harmonic upper partials, and this is not seen from the results of the analytical AC analysis. Only if the high-frequency cut-off starts early in the frequency response curves given by the AC analysis, then one knows that part of the high-frequency harmonic content created by clipping the signal is damped out at the output of the fuzz. This would indicate that the full potential of the circuit is not unleashed.

It is anyhow interesting to simulate how the adjustment of the 'fuzz' level via potentiometer R_6 affects the amplification and the low-frequency response of the circuit. Figure 6.33 depicts the effect of three different adjustments of R_6 to

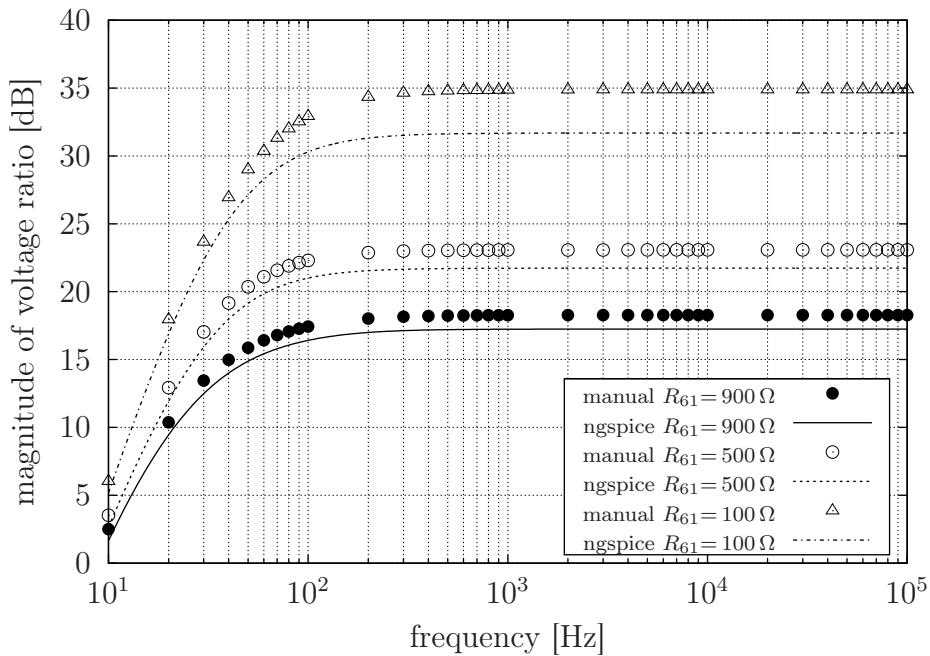


Figure 6.33: Comparing the fuzz data from the simulations and manual calculations

the general frequency response. It is easy to notice that the manual calculations have the same shape with the ngspice simulated results, but the gain values are different. This is because of the added leakage parameters, which reduce the full amplification potential of the circuit.

The frequency response measurements on the prototype circuit were not done because it was not seen as feasible due to the heavy clipping of the signal. The simulation results indicate that the 'fuzz' potentiometer should be a logarithmic

pot, since the gain does not change linearly with the resistance value of R_6 . One evident conclusion is also that a lower value of R_6 causes a higher level of gain.

A SPICE transient simulation was also applied to the fuzz circuit to find out the actual response waveform of the circuit to simple sine wave input signals. These simulations can be compared to the actual input-output prototype measurements, which were done using a digital storage oscilloscope. The benefit of using such a scope is that it can store the data in a comma separated list (.csv) format for later plotting.

Figure 6.34 reveals the simulated response of the fuzz circuit for sinusoidal test signals of different levels. In the simulation setup, the volume potentiometer was set close to the maximum, and R_6 was set so that $R_{61} = 200 \Omega$ and $R_{62} = 800 \Omega$. This setting brings more gain to the system as already noticed from the frequency response curves, and therefore it also indicates the fuzz feature more clearly in the low signal levels. The test signal had a frequency of 440 Hz and the signal voltage levels were 10 mV and 100 mV so as to imitate the true output signal levels from a typical electric guitar with single-coil pickups. As seen in Figure 6.34, the ground level of the circuit is 9.0 volts because of the need to use inverted operation voltages for *pnp* transistors.

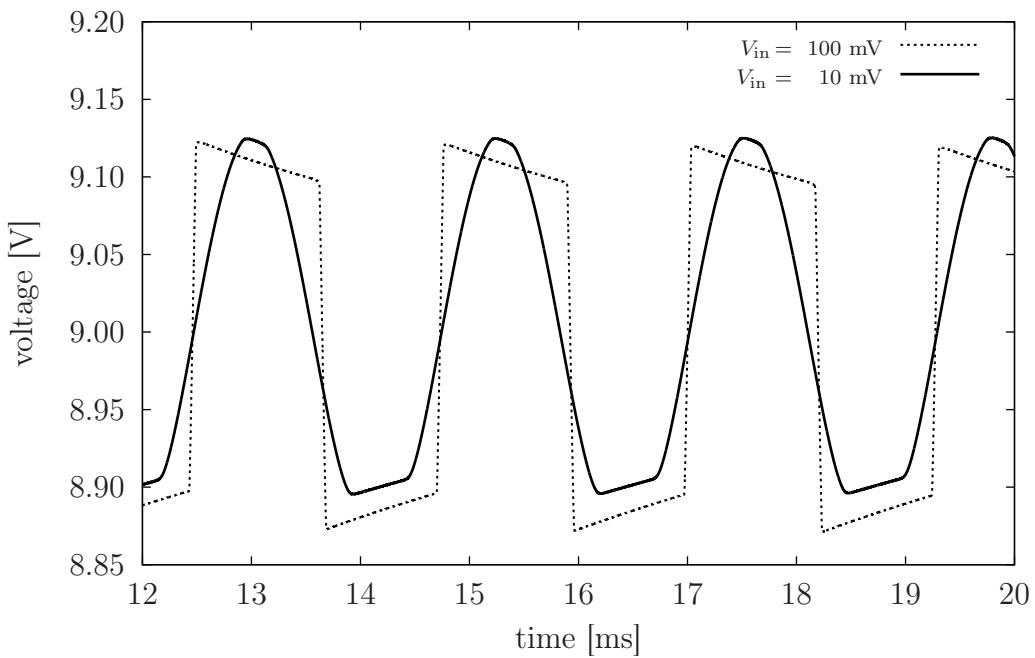


Figure 6.34: A transient fuzz simulation using a sinusoidal input signal

The distinctive difference between the low and high level responses was found

to be asymmetric clipping of the top and bottom parts of the low level signals, whereas higher level signals were clipped more symmetrically from the top and bottom. According to the simulations, the range between the points where the signal starts to be chopped off from the top and bottom or not clipped at all is quite narrow, in the order of a few millivolts. Because of the asymmetric clipping, a skilled guitar player can control the effect by the style of playing, since softly played notes are distorted differently from hard hammered notes. This gives a third dimension for controlling the fuzz effect device.

The general scope of the measurements on the do-it-yourself prototype device was to roughly verify the correlation between the simulations based on mathematical models and the real world prototype response to a sinusoidal test signal. The signal source used for the measurements was chosen to be a real electric guitar, Ibanez GRG 170 DXL equipped with two humbucker pickups and one single-coil pickup. The signal from the guitar was generated by playing a fingered note on the *B*-string close to the 12th fret of the fretboard and using the humbucker pickups closest to the fretboard. This combination created a relatively clean sine wave with a frequency quite close to 440 Hz. This frequency was chosen to match closely to the simulations, where the pure sine wave is the easiest test signal to generate.

The first thing that was noticed in the prototype measurements was that when an electric guitar was used as the signal source, the fuzz circuit itself had some kind of a loading effect on the waveform of the signal. This is basically expected, since the current-shunt feedback configuration in the fuzz circuit decreases the input impedance and therefore draws more current from the pickup circuitry of the guitar. Figure 6.35 shows the difference by comparing a direct measurement at the end of the guitar cable before and after connecting the cable to the fuzz input jack. It might also be a case where the second transistor clips the signal by saturation and causes some distortion also at the input. But this theory is purely speculation.

The actual measurements simply tried to verify the similarity to the simulated waveforms presented in Figure 6.34. The potentiometers of the fuzz prototype were set to a level where the output potentiometer was giving full output, and the potentiometer controlling the level of the fuzz was set so that the level of fuzz is close to maximum. This setup was trying to imitate the settings used in the transient simulations but the accurate values used in the measurements for the fuzz level control potentiometer were not measured. This measurement setup

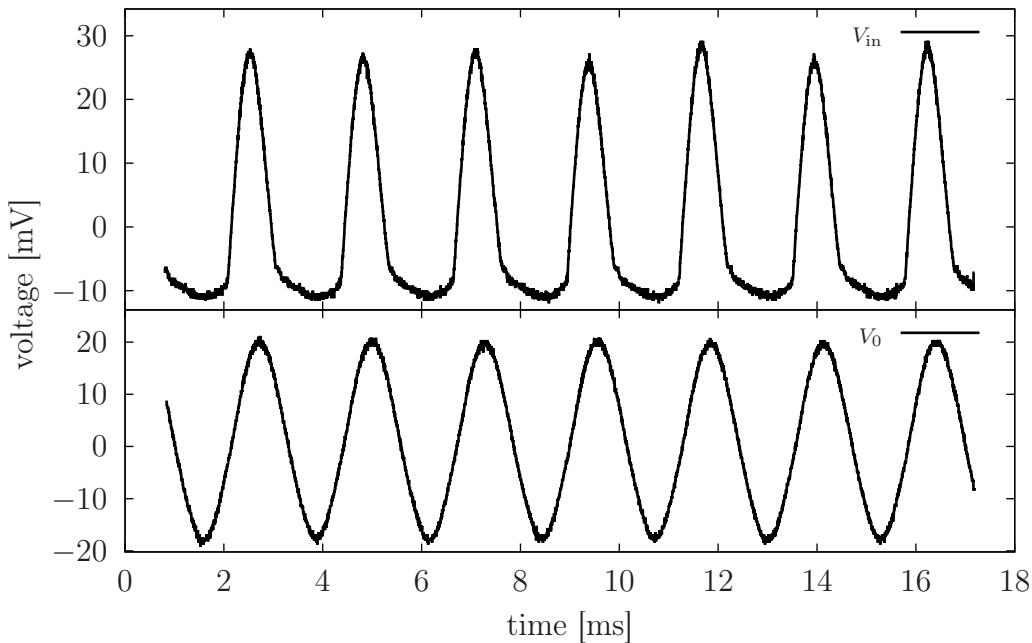


Figure 6.35: The change in the input voltage due to fuzz circuit loading

might therefore have been slightly different compared to the simulation setup.

For hard hammered notes, where the input signal level is high, the measured waveform from the prototype came out according to Figure 6.36. It was quite amazing to see that the measurements on a prototype circuit gave almost exactly the same output waveform as in the simulation results. The output level of the signal was also in a relatively good agreement with the simulated results. The output signal in the prototype measurements stayed longer in the low state than in the simulations. This was due to the signal distortion observed already at the input.

For softly played notes, there were some differences noticed when comparing to the simulated results. The asymmetrical clipping was more distinct in the actual prototype and the range between full clipping and no clipping was clearly wider than in the simulations. In any case, the resemblance of the waveforms in the asymmetrical clipping state is still quite amazing. In Figure 6.37 one can see that with an approximately 1 mV input signal the signal is clipped only from the bottom part and the top part is reproduced without clipping. This indicates that the prototype circuit had probably more gain than the simulated circuit, or that the leakage currents behaved differently in the prototype.

The previous measurements were made directly from the output terminals of the

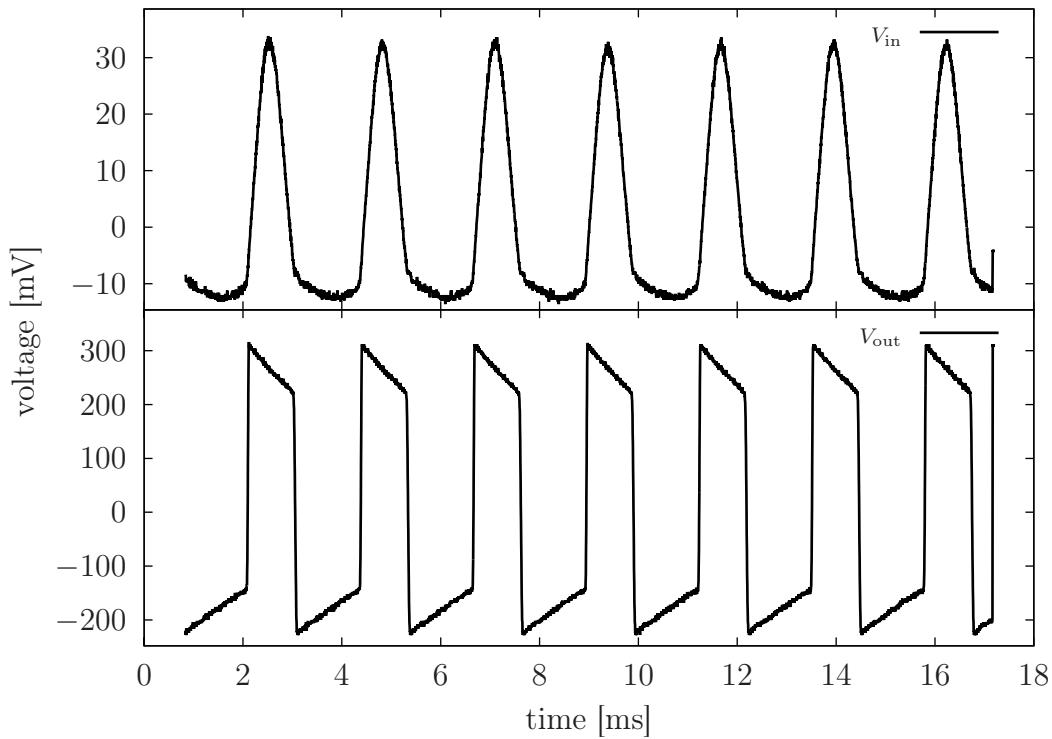


Figure 6.36: Heavy clipping with a large input voltage level

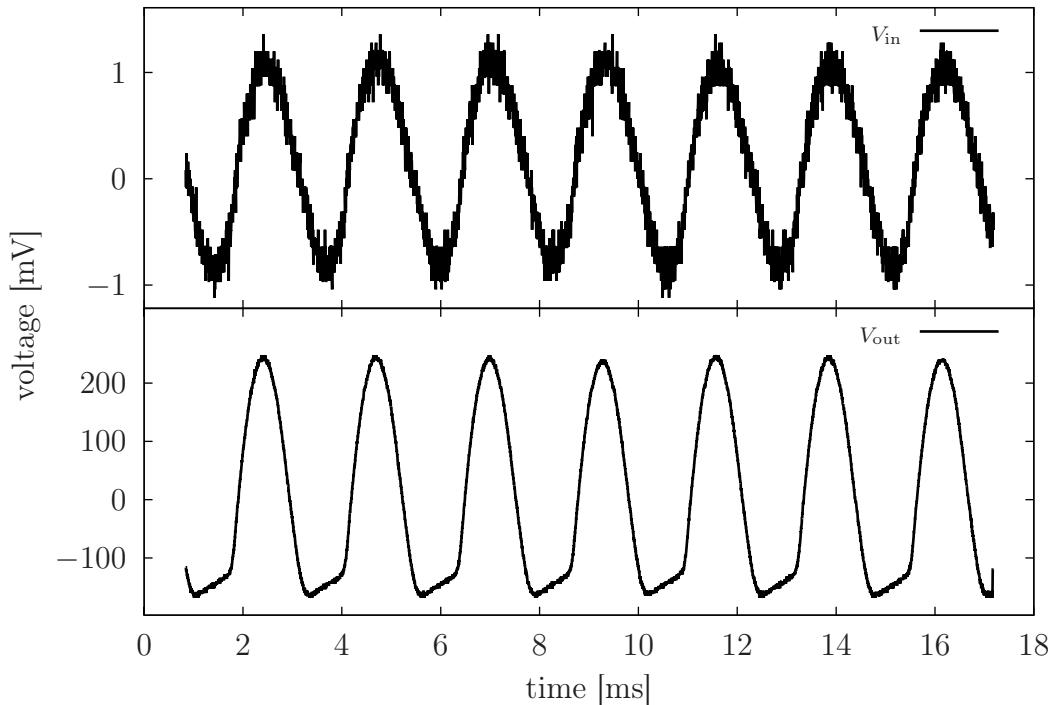


Figure 6.37: Soft clipping with a small input level

fuzz device when there was no cable attached to the output jack. Previously it was already noticed that the fuzz circuit itself created a loading effect on the

signal from the guitar. The situation is similar at the output terminals of the fuzz circuit. A cable connection from the fuzz circuit to an amplifier causes a distinct difference in the output signal of the fuzz effect. The measured waveform in the output terminals of the fuzz circuit changes as indicated in Figure 6.38. This means that the measurement results have a dependency on the external connections on the fuzz circuit.

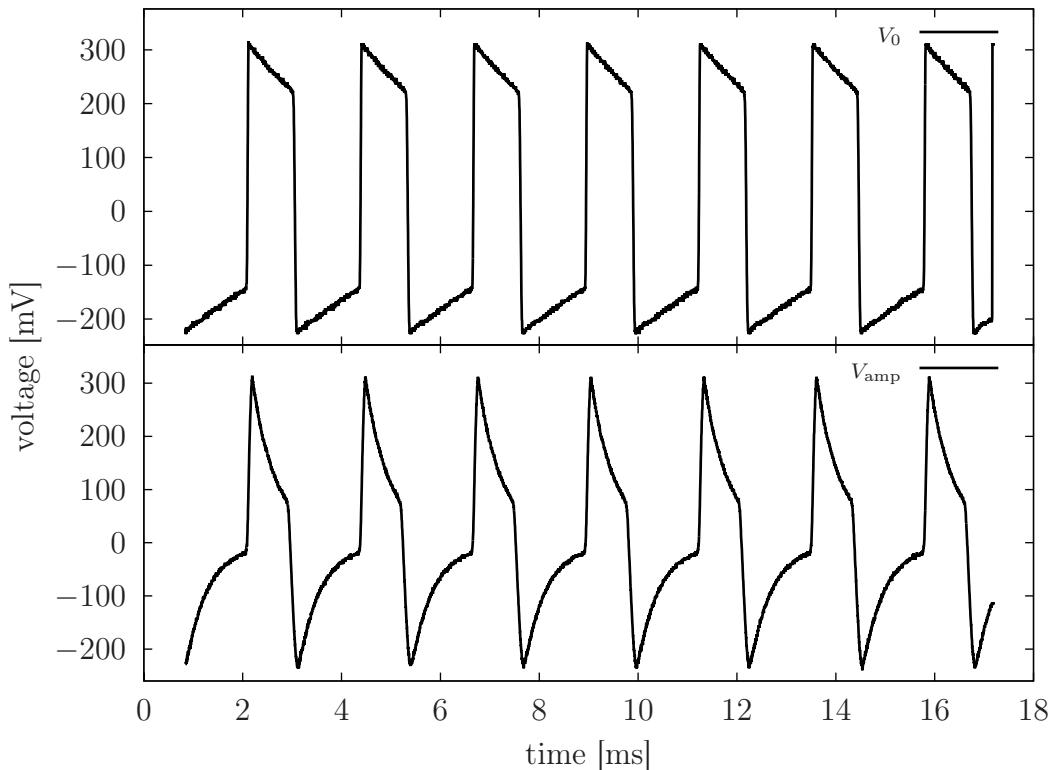


Figure 6.38: The effect of the amplifier connection

One important fact to point out is that a computer sound card is not a reliable measurement device. In the case of the fuzz effect, the output signal is clipped to look almost like a square wave. Square waves of low audio frequencies cannot be measured reliably with a sound card because the input terminal of the sound card includes a capacitor that alters the DC level signal at the top and bottom of the square wave waveform.

6.3 TREMOLO = AMPLITUDE MODULATION

The tremolo effect creates a sensation of oscillating volume control, which in technical terms means amplitude modulation of the input signal. The terms tremolo and vibrato are often mixed together to mean the same thing, but whereas

tremolo refers to the oscillating volume, vibrato is specified as the oscillating frequency. A common method in electrical engineering to generate an amplitude modulated signal is to modify the amplifier supply voltage to alter the gain. This produces an 'envelope' where the higher frequency is closed inside the frames of the lower frequency [14, p. 574].

According to the theory of amplitude modulation, the original signal $\sin \omega_c t$ and the modulating signal $\sin \omega_m t$ are mixed together to form a modulated signal,

$$M(t) = (1 + \alpha \sin \omega_m t) \sin \omega_c t, \quad (6.30)$$

where α is a numeric value between 0 and 1 and it is known as the modulation index. Generally it is assumed that $\sin \omega_m t < \sin \omega_c t$, i.e. the modulating signal is of much lower frequency than the original signal. Amplitude modulation is most commonly used in radio frequency applications where the signal of the higher frequency is named as a carrier signal. This naming convention is also used consistently in this context with subscript c , although it is quite misleading for an audio effect application.

From equation (6.30) it is evident that the mixing in amplitude modulation is done by multiplication, not as a regular sum. However, multiplication is not the only way to create modulated signals. According to the trigonometric identity between products and sums,

$$\sin \omega_c \sin \omega_m = \frac{1}{2} [\cos(\omega_c - \omega_m) - \cos(\omega_c + \omega_m)], \quad (6.31)$$

the multiplication can be replaced by a sum where a difference and a sum of frequencies, i.e. $\omega_c - \omega_m$ and $\omega_c + \omega_m$, are added together.

Figure 6.39 depicts different ways to create a modulated signal using a mixture of multiplication and sums. The signals in the figure have functional descriptions

$$f_m(\omega_m t) = \sin(2\pi t) \quad \text{and} \quad f_c(\omega_c t) = \sin(2\pi 20t),$$

with a reference to the modulating signal and the carrier signal respectively, the carrier having a distinctly higher frequency. The notations f_{c-m} and f_{c+m} refer to the sums arising from the product $f_c \cdot f_m$. If the modulation is taken as a direct product, the modulated signal $M(t)$ is at its smallest peak-to-peak amplitude when the modulating signal $f_m(t)$ is close to zero. This is seen in the left side graphs of Figure 6.39. The right side graphs present the modulation as defined by equation (6.30). The most attenuated amplitude of the modulated signal $M(t)$

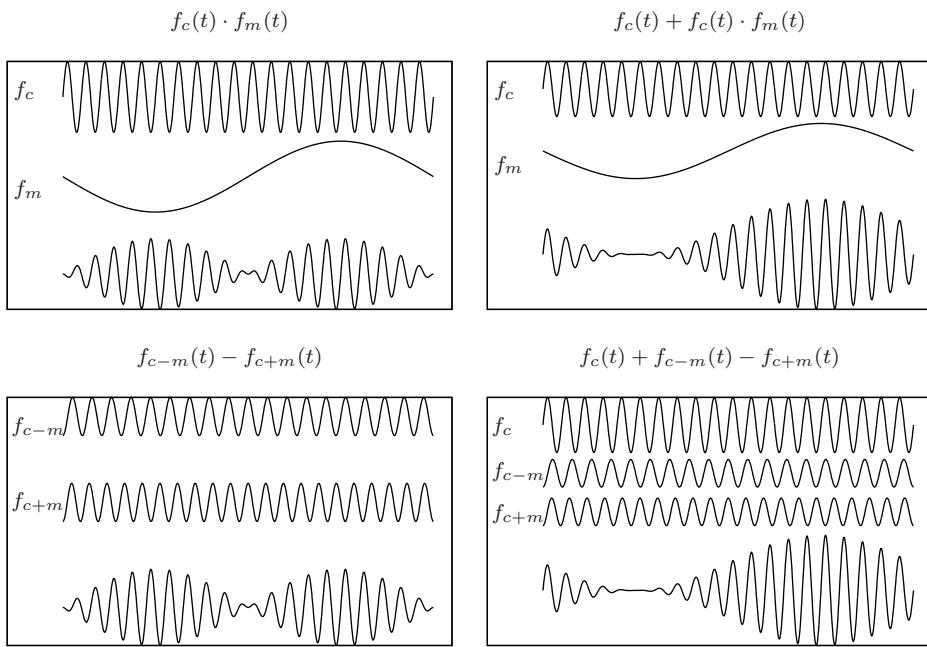


Figure 6.39: Four ways to set up amplitude modulation

occurs when the modulating signal $f_m(t)$ has a negative maximum, and the maximum peak-to-peak amplitude of the modulated signal $M(t)$ is obtained with a positive maximum value of the modulating signal $f_m(t)$.

Amplitude modulation is fundamental stuff in the theory of signal processing and it will surely create a cool sound effect, but how to realise multiplication of signals with resistors, capacitors and transistors? The schematic 6.40 of a simple tremolo effect will answer this question, at least after it is thoroughly analysed.

The schematic has been adopted from the *Tonepad* website [76], where at the time of writing it is readily available for everyone as a do-it-yourself vintage effect project. Apparently this circuit has originally appeared as a project article in an electronics magazine, 'Electronics Australia', which has not been published anymore in recent years. This simple schematic contains many interesting details of basic electronics in action, so that is why a detailed analysis of the circuit is in order. A general description of the circuit is given now and extended later in the following sections.

So how does the tremolo circuit actually work? The input signal travels a very short path to the output through one buffering transistor Q_1 , which has a quite minimalistic gain. The output signal is taken from the potentiometer R_7 where the voltage division only affects the amplitude of the output voltage, and not in

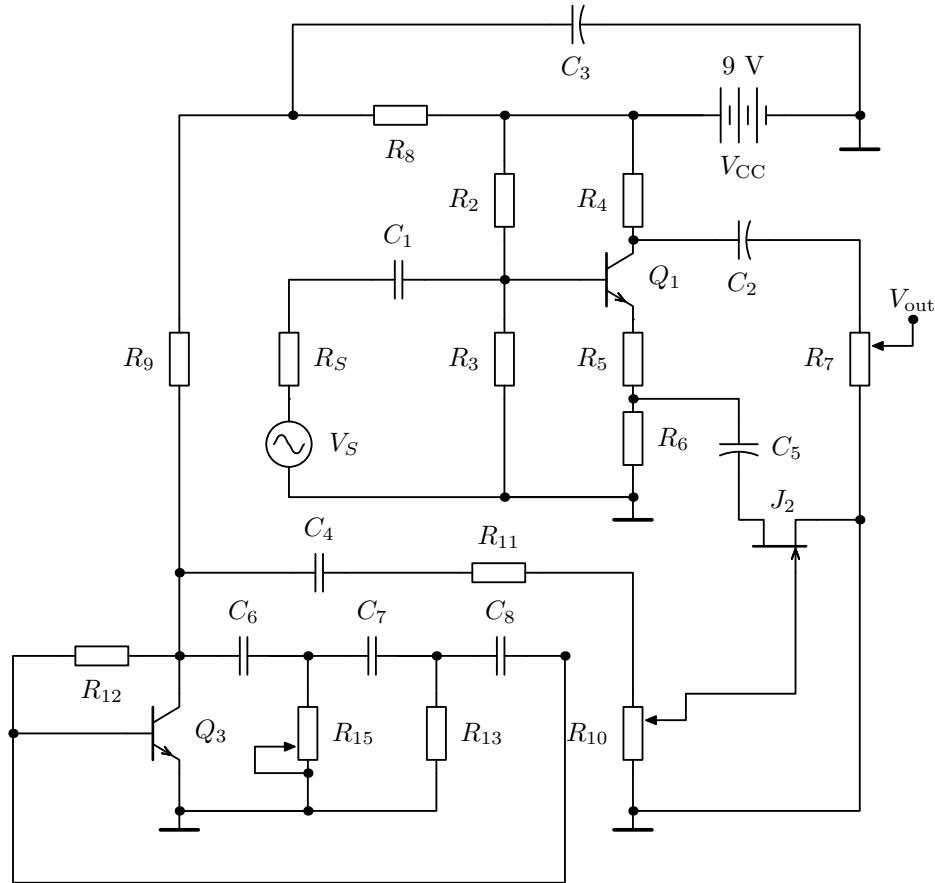


Figure 6.40: A circuit diagram of the tremolo effect

any frequency dependent property of the circuit. The gain of the transistor Q_1 is designed to keep the output signal almost on the same level as the input signal.

The actual modulating signal is created with the bipolar junction transistor Q_3 , which is hooked up in an oscillator configuration. The frequency of the modulation signal can be adjusted from 2 Hz to 8 Hz with potentiometer R_{15} so that there are only a few full vibrational periods within a second interval. The in-depth analysis of this so-called RC phase-shift oscillator is treated separately as a side effect since it seemed to be quite a fascinating realisation of a simple audio frequency oscillator.

The modulation signal is connected to the input signal via a junction field-effect transistor (JFET) J_2 , which is biased in the *ohmic region* by setting the drain-to-source DC potential difference V_{DS} to zero with capacitor C_5 . In the ohmic region, the JFET can be used as a voltage-controlled resistor. The modulation signal controls the gate voltage v_{GS} of JFET J_2 , which thereby changes its resistance as a function of the modulating voltage. The details of using a JFET as a

resistor are treated in a separate side effect section.

Eventually the change in the emitter resistance of transistor Q_1 due to JFET J_2 modifies the AC amplification factor of Q_1 at the rate of the oscillation produced with Q_3 . This generates the modulation effect according to equation (6.30). The modulation depth can be controlled with potentiometer R_{10} , which affects the self-biasing configuration of J_2 .

6.3.1 The DC bias analysis of the tremolo effect

The analytical biasing calculations for transistor Q_1 can be handled with equation (1.91), which was presented already in the introductory section 1.3.9. In the static DC configuration, the JFET J_2 is hidden behind the capacitor C_5 so that it will have no effect on the quiescent voltages of Q_1 . The biasing of this circuit is set like a textbook example of a basic common emitter amplifier, and therefore no distortion is expected out of this effect.

To find out the biasing currents and voltages of the BJT Q_3 , some new equations are still needed in addition to what is presented in section 1.3.9. This time the JFET J_2 and the potentiometer R_{10} controlling the modulation depth are hidden behind the capacitor C_4 and are therefore not affecting the DC biasing of Q_3 . For the voltage V_C at the collector of Q_3 one can write two equations,

$$V_C = V_{CC} - I_C(R_8 + R_9) \quad (6.32)$$

$$V_C = V_{BE} + I_B R_{12} = V_{BE} + \frac{I_C}{\beta_F} R_{12}, \quad (6.33)$$

from where one can solve an expression for the collector current,

$$I_C = \frac{V_{CC} - V_{BE}}{R_8 + R_9 + \frac{R_{12}}{\beta_F}}.$$

This current can be used to solve the base current I_B and all voltages at the terminals of Q_3 .

Here are all the equations that are needed to calculate the essential biasing voltages by means of manual analysis. The actual component values are substituted into these equations and the resulting voltage levels are compared to the measurements and the simulation results in section 6.3.6.

6.3.2 The AC analysis of the tremolo effect

To simplify the AC analysis of the tremolo effect, the oscillator is ruled out and the JFET is replaced by a normal resistor R_J . By varying the value of R_J within the analytical limits of JFET resistance, the frequency response and small-signal amplification of the circuit can be analysed easily and without loss of accuracy.

With these simplifications, the small-signal model for the tremolo effect is as shown in Figure 6.41. This small-signal model is almost the same as the model of

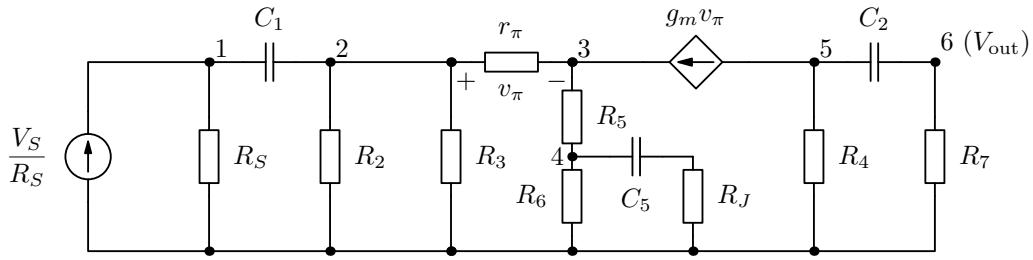


Figure 6.41: A partial small-signal model of the tremolo circuit

the basic transistor amplifier 1.40. The only difference is the relatively complex emitter impedance of the tremolo circuit. Although the emitter impedance could have been included in the model as one branch without adding the extra node 4, for manual calculations it was seen as a simplification to add the node 4.

A set of linear equations can be directly written from the small-signal model into a matrix form. Since the nodal analysis method is used, the resistors and capacitors are represented as admittances, which multiply the node voltages $V_1 - V_6$. The admittance matrix is of a size 6x6 and the whole equation is written as

$$\begin{bmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 & 0 \\ 0 & -Y_{32} & Y_{33} & -Y_{34} & 0 & 0 \\ 0 & 0 & -Y_{43} & Y_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & Y_{55} & -Y_{56} \\ 0 & 0 & 0 & 0 & -Y_{65} & Y_{66} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \end{bmatrix} = \begin{bmatrix} \frac{V_S}{R_S} \\ 0 \\ g_m (V_2 - V_3) \\ 0 \\ -g_m (V_2 - V_3) \\ 0 \end{bmatrix}.$$

The admittances indicated by general subscripted notations $Y_{11} \dots Y_{66}$ in the preceding matrix equation are given explicitly in listing (6.34).

$$\begin{aligned}
 Y_{11} &= \frac{1}{R_S} + j\omega C_1 & Y_{12} = Y_{21} &= j\omega C_1 \\
 Y_{22} &= \frac{1}{r_\pi} + \frac{1}{R_2} + \frac{1}{R_3} + j\omega C_1 & Y_{23} = Y_{32} &= \frac{1}{r_\pi} \\
 Y_{33} &= \frac{1}{r_\pi} + \frac{1}{R_5} & Y_{34} = Y_{43} &= \frac{1}{R_5} \\
 Y_{44} &= \frac{1}{R_5} + \frac{1}{R_6} + \frac{j\omega C_5}{1 + j\omega C_5 R_J} & \\
 Y_{55} &= \frac{1}{R_4} + j\omega C_2 & Y_{56} = Y_{65} &= j\omega C_2 \\
 Y_{66} &= \frac{1}{R_7} + j\omega C_2
 \end{aligned} \tag{6.34}$$

To calculate the final form of the transfer function $H(j\omega)$, the transconductance terms need to be moved to the admittance matrix and then Cramer's rule is used to solve the output voltage at node 6. This procedure has already been shown during the AC analysis of the basic transistor amplifier in section 1.3.10, so it is not repeated here. The resulting transfer function is similar as in the case of the basic transistor amplifier. Because of the tremendous amount of work needed to write open the determinants in Cramer's rule, the manually obtained approximate model of the tremolo circuit is analysed only numerically using Octave. In Octave or Matlab it is possible to directly solve the two determinants numerically without the need to solve a symbolical expression for the transfer function resulting from the division of the two determinants.

To get at least some idea of how the tremolo circuit works in the frequency domain as far as the gain is concerned, the frequency response of the circuit is calculated for four different values of R_J , which theoretically represent the JFET resistance at some specific gate-to-source voltage v_{GS} . An approximate value for the transistor's internal resistance r_π is evaluated using equations (6.12), where the quiescent collector current I_C is obtained from the DC analysis.

According to Figure 6.42, the AC gain of the circuit changes as the resistance R_J is varied. Small values of R_J give a higher gain than large R_J . Now we just need to understand better how the resistance of the JFET changes with the modulation signal, and what the actual values to be used for R_J are to comprehensively simulate the dynamic resistance of a JFET. Because the resistor R_6 is connected parallel to the resistance depicted by R_J , the value of R_6 limits the maximum emitter resistance close to its nominal resistance when R_J is much larger than R_6 . Alternatively, when resistance R_J is smaller or about equal to R_6 , the total resistance of the parallel connection is defined mostly by the magnitude of R_J .

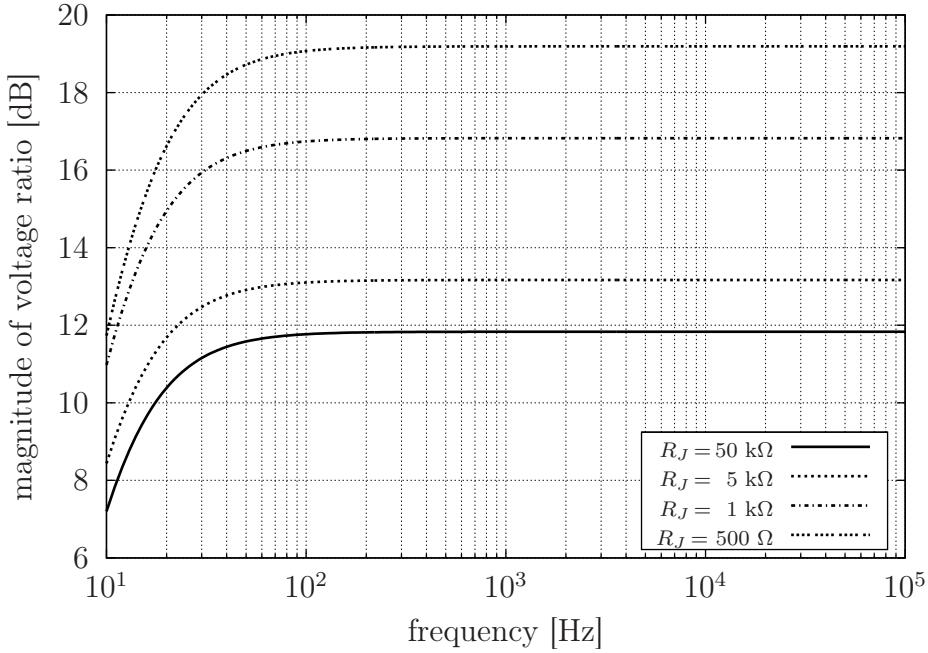


Figure 6.42: Initial analysis of the tremolo circuit frequency response

alone.

6.3.3 Side effect: the JFET as a resistor

In the tremolo effect, the multiplication operation of the main signal with the modulating signal needed for realising equation (6.30) is done with the JFET J_2 . The implementation makes use of the unique property of JFETs to act as voltage-controlled resistors. To understand how to estimate the resistance of a JFET and use it effectively in applications of discrete component electronics, a closer look at the basic theory of JFETs is required.

The graph of curves representing the $i_D - v_{DS}$ characteristics of a JFET with different values of v_{GS} can be roughly divided into two distinctly different areas called the *ohmic* and *saturation* regions. Therefore, the drain current i_D in a JFET has two functional descriptions

$$i_D = \frac{I_{DSS}}{V_P^2} [2(v_{GS} - V_P)v_{DS} - v_{DS}^2] (1 + \lambda v_{DS}) \quad (\text{ohmic}) \quad (6.35)$$

$$i_D = \frac{I_{DSS}}{V_P^2} (v_{GS} - V_P)^2 (1 + \lambda v_{DS}) \quad (\text{saturation}) \quad (6.36)$$

where the ohmic region is restricted by the limits ($v_{GS} > V_P; v_{DS} < v_{GS} - V_P$) and the saturation region is constrained with ($v_{GS} > V_P; v_{DS} > v_{GS} - V_P$) [71, p. 214]. The parameter λ is a channel length modulation parameter, and usually this term is so close to zero that it can be neglected. The parameter V_P is the pinch-off

voltage of a JFET referring to a value of V_{GS} which fully closes the conductive channel of the JFET, while I_{DSS} is the saturation current when the conductive channel is fully open at $v_{GS} = 0$. These equations assume that the drain current i_D , gate-to-source voltage v_{GS} and the drain-to-source voltage v_{DS} are carrying AC and DC components, as should be clear from the choice of notation.

The following analysis concentrates on the more commonly used *n*-type JFET, which is always biased to operation by having the gate-to-source voltage v_{GS} reverse-biased. This reverse biasing means that if the source is connected to ground, the JFET is fully open at $v_{GS} = 0$ and fully closed at $v_{GS} = V_P$. In the *n*-type JFET, V_P is a negative voltage, e.g. -3 V, therefore making the gate-source junction reverse-biased. The *p*-type JFET works analogously to the *n*-type JFET but with changed voltage polarities.

The static analysis of JFET current-voltage characteristics using only DC components is shown in Figure 6.43, where the drain current I_D of a specific JFET with $V_P = -1.8$ V is drawn as a function of the drain-to-source voltage V_{DS} for selected values of gate-to-source voltage V_{GS} . The figure also highlights the ohmic region of the JFET with grey shading.

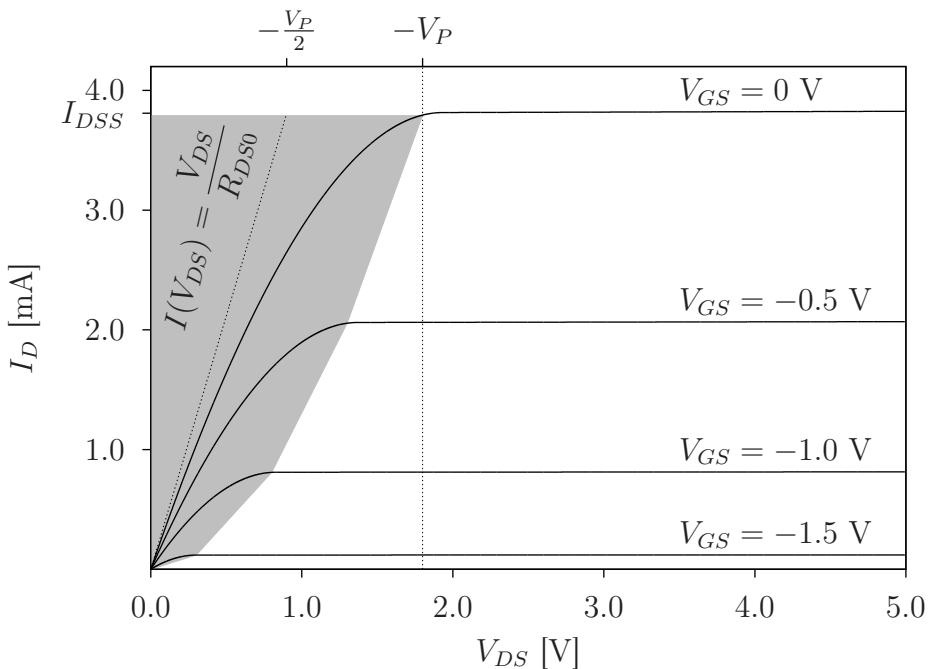


Figure 6.43: Characteristic curves of a JFET

From equation (6.35) the reciprocal value of the drain-source resistance R_{DS} in the ohmic region is obtained by dividing the drain current i_D by the potential

difference v_{DS} over the drain and source. Taking the reciprocal of this quotient and neglecting the term with λ leads to

$$R_{DS} = \frac{v_{DS}}{i_D} = \frac{V_P^2}{I_{DSS} [2(v_{GS} - V_P) - v_{DS}]} \quad (6.37)$$

Especially if v_{DS} is very close to zero, a sufficient approximation is

$$R_{DS} \approx \frac{V_P^2}{2I_{DSS} (v_{GS} - V_P)} \quad (6.38)$$

Furthermore, as a special case when $v_{GS} = 0$,

$$R_{DS} = R_{DS0} \approx \frac{-V_P}{2I_{DSS}} \quad (6.39)$$

The reciprocal value of resistance R_{DS0} is seen in Figure 6.43 as the slope of the tangential line drawn beside the $V_{GS} = 0$ curve at small values of V_{DS} . The tangential line extends to the point ($V_{DS} = -V_P/2, I_D = I_{DSS}$), which can be used to define the value of R_{DS0} .

The equation for R_{DS} can also be obtained analytically by deriving an expression for transconductance g_m , a familiar small-signal transistor parameter, which in the context of JFETs is defined as

$$g_m = \left. \frac{\partial i_D}{\partial v_{GS}} \right|_{v_{DS}=V_{DSQ}}$$

In the saturation region, the drain current of a JFET has been experimentally found to depend on v_{GS} and V_P as

$$i_D = I_{DSS} \left(1 - \frac{v_{GS}}{V_P} \right)^2 \quad (6.40)$$

The exponent of 2 is rounded from an experimental value to a nice integer, thereby being an experimental approximation and not an exact theoretical law [67, p. 260]. After differentiating the expression of i_D with respect to v_{GS} , the expression for transconductance is

$$g_m = \frac{-2I_{DSS}}{V_P} \left(1 - \frac{V_{GSQ}}{V_P} \right), \quad (6.41)$$

where V_{GSQ} is the DC gate-to-source voltage at the chosen static bias point of the JFET.

Another experimental fact is that the transconductance g_m at zero gate voltage in the saturation region is approximately equal to the channel conductance g_{ds}

between drain and source in the ohmic region when $V_{GSQ} = 0$ and V_{DSQ} is close to zero volts [67, p. 260]. Mathematically this fact states that

$$g_m(V_{GSQ} = 0) = g_{m0} = g_{ds0} = \frac{1}{R_{DS0}} = \frac{-2I_{DSS}}{V_P}.$$

Hence,

$$R_{DS0} = \frac{-V_P}{2I_{DSS}}.$$

This is the common way to analyse the JFET drain resistance, but not the only way since the equation for the drain current i_D changes for different doping density distributions of the JFET substrate. The model of equation (6.40) represents one limit where the doping density is approximated with an exponential function [77, p. 21]. The other limit is obtained when the doping density is assumed to be uniform, and this case has been theoretically analysed by Shockley [78]. The derivations [77, pp. 13–17] following from the uniform doping distribution lead to an equation

$$i_D = I_{DSS} \left[1 - 3 \left(\frac{V_{GS}}{V_P} \right) + 2 \left(\frac{V_{GS}}{V_P} \right)^{\frac{3}{2}} \right] \quad (6.42)$$

for the drain current in the saturation region. Applying the definition of the transconductance to equation (6.42),

$$g_m = \frac{-3I_{DSS}}{V_P} \left[1 - \left(\frac{V_{GSQ}}{V_P} \right)^{\frac{1}{2}} \right]. \quad (6.43)$$

To reach a similar equation as (6.38), which would be slightly wrong reasoning in this case, the equation for the drain-to-source resistance at very small v_{DS} would read

$$R_{DS} \approx \frac{-V_P}{3I_{DSS} \left[1 - \left(\frac{V_{GSQ}}{V_P} \right)^{\frac{1}{2}} \right]}. \quad (6.44)$$

A similar result has been derived by Millman [19, p. 390 and p. 414], also using Shockley's analysis as a basis.

When the JFET is biased with the drain and source at approximately the same potential (i.e. $V_{DS} \approx 0$), the current versus voltage characteristics of the JFET at some specific V_{GS} behave linearly. Therefore, the JFET looks just like a normal resistor from the signal's point of view. In practise this means that small sinusoidal oscillation of the drain current i_D about the origin creates a similar sinusoidal voltage v_{DS} across the JFET. If the oscillations are not small, or if the DC component of v_{DS} is significant, the sinusoidal form of the signal would be

distorted because the resistance of the JFET would not behave linearly anymore for these largely offset signals.

The value of the JFET resistance can be controlled by changing the voltage on the gate of the JFET. If a reference resistance R_{DSR} in the drain-source channel is known for some specific gate-to-source voltage V_{GSR} , all other values of R_{DS} with some V_{GS} can be approximated with

$$R_{DS} = \frac{V_{GSR} - V_P}{V_{GS} - V_P} R_{DSR}. \quad (6.45)$$

This is derived simply by dividing equation (6.38) with itself.

When talking about a real JFET component bought from an electronics store, the simplest setup to find out R_{DS0} is the case when V_{GS0} is set to zero and $V_{DS} \approx 0$. Then one can directly use equation (6.39), where R_{DS0} depends only on pinch-off voltage V_P and the zero-gate voltage drain current I_{DSS} . Unfortunately one cannot benefit much from the values given in manufacturers' datasheets for I_{DSS} and V_P , because there is too much variation in the specifications. If the JFET is to be used as a resistor with a specified range of resistance, the values for I_{DSS} and V_P should be determined by measurements. Even approximately measured values lead to a much more accurate value of R_{DS0} than directly using the values given in the datasheet.

Luckily, the most basic measurement procedures for I_{DSS} and V_P are relatively simple, so it is worthwhile trying it out – especially when building hand-made vintage effects. The measurement setup to obtain an approximate value of I_{DSS} is described in Figure 6.44a and the setup for V_P measurement is depicted in Figure 6.44b.

Both measurements are done by biasing the JFET into the saturation region by choosing the battery to be a few volts over the expected absolute value of V_P and then measuring the voltage drop between points a and b . In the I_{DSS} measurement one should use a very low-valued resistor and then convert the voltage drop over a to b into current using Ohm's law. Another way is to simply connect an ammeter between the drain pin and the battery. To determine the pinch-off voltage V_P , a resistance of about $1\text{ M}\Omega$ is placed between the source pin and the ground. This will self-bias the JFET, and because the current flowing through the JFET channel will be very small due to the high resistance of the chosen R_S , the potential difference between points a and b will be close to the actual pinch-off voltage. In true pinch-off the current in the JFET channel is diminishingly small.

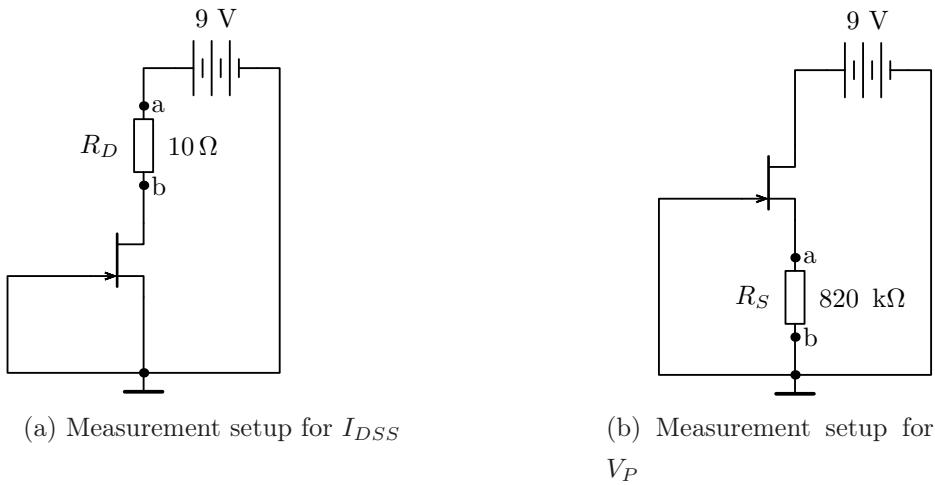


Figure 6.44: Measurement setups to determine the necessary JFET parameters

Therefore, the higher resistance used as R_S in the measurement of V_P , the closer one will get to the actual value.

The measurement probes of the measurement device should also have as high a resistance as possible. At least oscilloscopes have a $10\text{ M}\Omega$ probe impedance, which is quite sufficient already. The difference to using a $1\text{ M}\Omega$ probe is clearly noticeable and in some cases a $1\text{ M}\Omega$ probe can give misleading results when measuring JFETs. In reality, the measured value for the pinch-off voltage will always be slightly smaller (typically a few hundred millivolts smaller) than the absolute value of true V_P .

In some applications it is required to use JFETs with identical values of I_{DSS} and V_P . The methods described above are suitable for finding similar JFETs, but to get more accurately matched JFETs, it is better to measure the I_{DSS} with a few different operating voltages (V_{GS}) and pick the best matches from a group of measurements. To match V_P it would also be good to measure V_{GS} with several different resistors to get a set of values, and compare these several values between different JFETs.

Although the JFET channel current i_{DS} behaves linearly as a function of v_{DS} , creating a constant resistance R_{DS} when v_{DS} deviates only a little from zero value, the channel resistance as a function of V_{GS} is completely nonlinear as can be seen from Figure 6.45. Both equations (6.38) and (6.44) have been plotted into Figure 6.45 to see the actual difference of the two limiting cases. The actual resistance curve of some practical JFET device is most likely somewhere in between these two curves. The only difference between the two equations is one square root

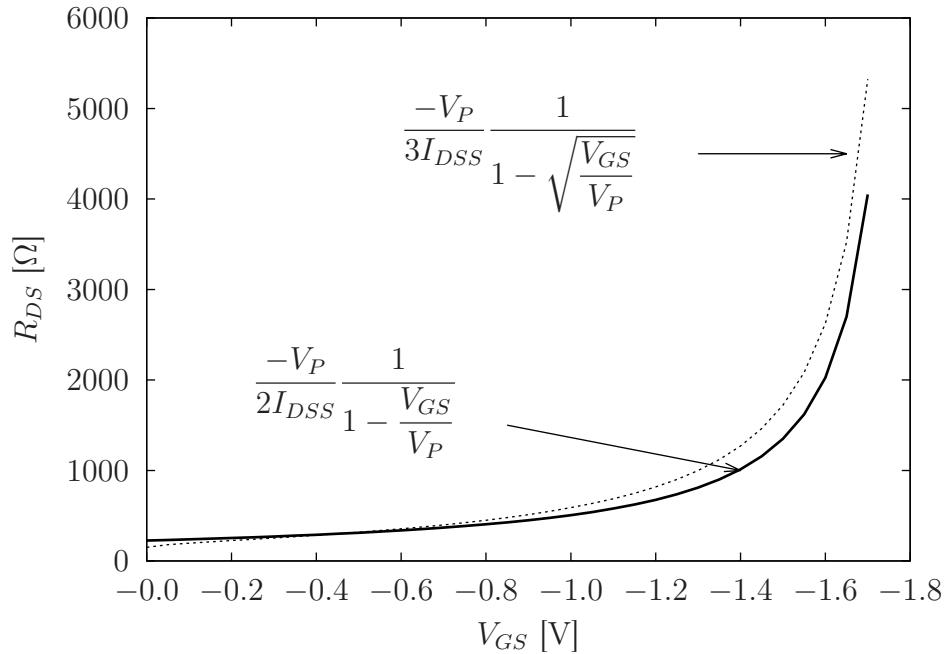


Figure 6.45: JFET resistance as a function of V_{GS} when $V_{DS} \approx 0$

term and a constant multiplier of 2 in one and 3 in the other. Generally these differences are not that distinct because the behaviour of the curves is roughly the same. The function with the square root term has been derived from the JFET channel geometry, and therefore represents a more theoretical way of obtaining the equation for the resistance curve.

As a general conclusion from Figure 6.45, for a wide range of V_{GS} values the channel resistance R_{DS} is quite small and changes very slowly. When approaching the pinch-off voltage ($V_P = -1.8$ V in this case), the resistance suddenly starts changing more and more rapidly and finally ends at the theoretically infinite resistance at $V_{GS} = V_P$. Therefore, it is a bit difficult to use the JFET resistance effectively to implement an accurately controlled change of resistance for practical applications. A common way to use the JFET as a voltage-controlled resistance is to connect the JFET parallel to some reference resistance, as has been done here in the tremolo circuit.

6.3.4 Side effect: the RC phase-shift oscillator

If an amplifier circuit has positive feedback, it is quite certain that it will start oscillating. Nevertheless, it is not that easy to generate steady and controlled oscillations with simple circuits. For low-frequency oscillations, it is irrational to use LC (inductor - capacitor) based circuits because large-valued inductors are

hard to find, large in size and also expensive. That is why most of the audio applications use mainly RC (resistor - capacitor) based oscillators or replace the inductors with operational amplifiers.

One of the simplest low-frequency oscillator circuits is the so-called "RC phase-shift oscillator" [15, p. 661]. The basic construction of a phase-shift oscillator uses a network of three or more resistor-capacitor pairs and an amplifier element, such as a BJT, a FET or an operational amplifier [79] [80]. The version of the oscillator that is used in this tremolo effect circuit is drawn independently in Figure 6.46 to support a more in-depth analysis.

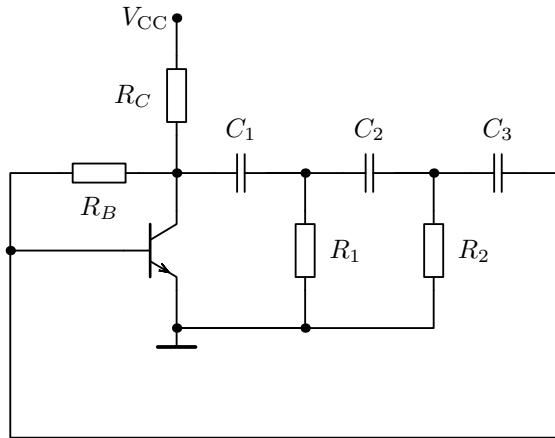


Figure 6.46: A BJT RC phase-shift oscillator circuit

The idea of the circuit is to get a phase shift of 180 degrees from the transistor output, and feed the signal back to the input of the transistor with an additional 180 degree phase shift from the RC network – a full 360 degree shift in total. The feedback to the amplifier is definitely positive since the signal always comes back to the input having the same phase as it started. The oscillations are started with any sudden impulse, such as switching on the power to the circuit. Another important condition for oscillations is that the amplifier must have a gain at least equal to the attenuation of the RC network.

In common literature, there are examples of the analysis of the RC network without the amplifier, and if the amplifier is included, it is mostly assumed to be ideal with large input impedance and small output impedance. In this specific oscillator used in the tremolo circuit, the transistor's internal resistance r_π plays a considerable role because it is part of the RC network. What complicates the analysis even more is the fact that the amplifier cannot be assumed to be ideal in this case because the RC network is affecting the gain of the transistor configura-

tion. It will be beneficial to analyse the ideal case first and then add in the things breaking the ideal model to see the differences.

This version of the BJT phase-shift oscillator uses an elegant biasing arrangement, where the base bias current is taken from the collector of the transistor. The bias network is drawn in Figure 6.47. The feedback loop from collector to

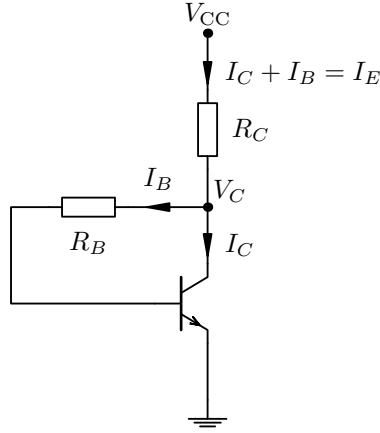


Figure 6.47: DC bias network of the BJT phase-shift oscillator

base helps to stabilise the bias point and therefore is an alternative choice to using an emitter resistor for bias stabilisation. The reason for using collector-to-base biasing is to try to maximise the gain and avoid having a bypass capacitor, which would be needed in the emitter resistor biasing scheme. When designing low-frequency oscillators it is beneficial to try to avoid the use of any bypass capacitors, because unless they have very large capacitance values, they have a significant effect to the oscillator behaviour.

It is possible to develop a relatively simple design model for biasing V_C to half way point of the operating voltage V_{CC} . From Figure 6.47 one has

$$V_{CC} = 0 + V_{BE} + I_B R_B + I_E R_C = I_E \left(R_C + \frac{R_B}{\beta_F + 1} \right) + V_{BE} \quad (6.46)$$

and from here the quiescent emitter current is

$$I_{EQ} = I_E = \frac{V_{CC} - V_{BE}}{R_C + \frac{R_B}{\beta_F + 1}}. \quad (6.47)$$

Let's assume that one wants to use a transistor with a certain β_F and collector resistor R_C . Then, from (6.46) and from (6.47) with $V_{CC}/2$ one has

$$R_B = R_C(\beta_F + 1) \left[\frac{2(V_{CC} - V_{BE})}{V_{CC}} - 1 \right]. \quad (6.48)$$

Substitution of this base resistance value R_B into equation (6.47) gives the quiescent emitter current and then

$$V_C = V_{CC} - I_{EQ}R_C = \frac{V_{CC}}{2}. \quad (6.49)$$

Theoretically this procedure yields the value of R_B , which biases the collector voltage to the optimal operating point when R_C and β_F are known.

As for the AC analysis, the small-signal equivalent circuit for the phase-shift oscillator of Figure 6.46 is depicted in Figure 6.48. The small-signal model can be

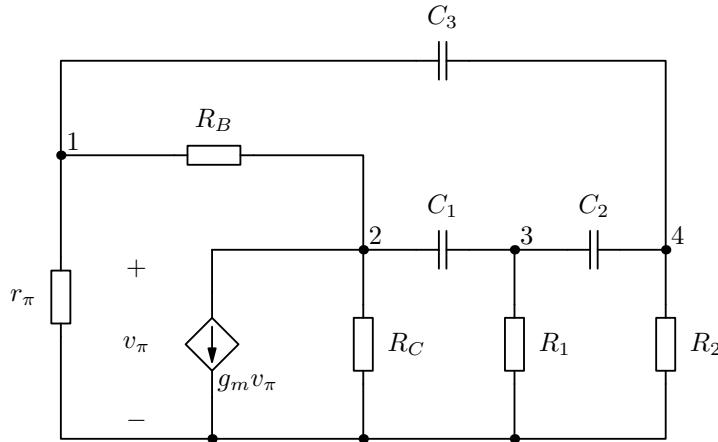


Figure 6.48: The small-signal equivalent of the RC phase-shift oscillator

described as a matrix representation of the form $Y \times V = I$, where Y refers to the admittance matrix, V to the node voltage vector and I to the current source vector. In explicit terms, the matrix equation is

$$\begin{bmatrix} \frac{1}{r_\pi} + \frac{1}{R_B} + sC_3 & -\frac{1}{R_B} & 0 & -sC_3 \\ -\frac{1}{R_B} & \frac{1}{R_C} + \frac{1}{R_B} + sC_1 & -sC_1 & 0 \\ 0 & -sC_1 & \frac{1}{R_1} + sC_1 + sC_2 & -sC_2 \\ -sC_3 & 0 & -sC_2 & \frac{1}{R_2} + sC_2 + sC_3 \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -g_m V_1 \\ 0 \\ 0 \end{bmatrix},$$

where by inspection from Figure 6.48, $v_\pi = V_1$ in the current source vector.

Starting with the analysis of the small-signal model of Figure 6.48, it is reasonable to assume that the voltage at node 2 has a 180 degree phase shift in relation to the input. This assumption is valid in the case of an ideal amplifier, where the

RC network is not seen as a load that affects the amplification. To determine the additional phase shift of the RC network, one needs to solve the voltage for nodes 2 and 1. Using the general nodal analysis method to start solving the voltage at node 1, one can write

$$\frac{V_1 \cdot \det(Y)}{g_m V_1} = \det(V_1) = \begin{vmatrix} -\frac{1}{R_B} & 0 & -sC_3 \\ -sC_1 & \frac{1}{R_1} + sC_1 + sC_2 & -sC_2 \\ 0 & -sC_2 & \frac{1}{R_2} + sC_2 + sC_3 \end{vmatrix} \quad (6.50)$$

and similarly for the voltage at node 2,

$$\frac{V_2 \cdot \det(Y)}{-g_m V_1} = \det(V_2) = \begin{vmatrix} \frac{1}{r_\pi} + \frac{1}{R_B} + sC_3 & 0 & -sC_3 \\ 0 & \frac{1}{R_1} + sC_1 + sC_2 & -sC_2 \\ -sC_3 & -sC_2 & \frac{1}{R_2} + sC_2 + sC_3 \end{vmatrix}. \quad (6.51)$$

To reach an explicit form of a transfer function, the determinants on the right sides of equations (6.50) and (6.51) should be written open and then form the expression for $\frac{V_1}{V_2}$. The ratio of these two voltages gives the attenuation factor γ for the RC network. When the imaginary part of this expression is zero, an equality of $-V_1 = \gamma V_2$ should hold.

After working out the determinant calculations from equations (6.50) and (6.51) and making the substitution $s = j\omega$ that holds for sinusoidal oscillations, one has the relation

$$\frac{V_1}{V_2} = \frac{a + jb}{c + jd}, \quad (6.52)$$

where

$$a = 1 - R_1 R_2 (C_1 C_2 + C_1 C_3 + C_2 C_3) \omega^2$$

$$b = R_1 (C_1 + C_2) \omega + R_2 (C_2 + C_3) \omega - R_B R_1 R_2 C_1 C_2 C_3 \omega^3$$

$$c = \left(1 + \frac{R_B}{r_\pi}\right) [1 - R_1 R_2 (C_1 C_2 + C_1 C_3 + C_2 C_3) \omega^2] - R_B R_1 (C_1 C_3 + C_2 C_3) \omega^2 - R_B R_2 C_2 C_3 \omega^2$$

$$d = \left(1 + \frac{R_B}{r_\pi}\right) [R_1(C_1 + C_2)\omega + R_2(C_2 + C_3)\omega] + \\ R_B C_3 \omega - R_B R_1 R_2 C_1 C_2 C_3 \omega^3$$

The next step is to set the imaginary part of equation (6.52) to zero. This needs further elaborate calculations because the equation (6.52) has to be multiplied with its complex conjugate in order to get the imaginary coefficient separated from the real part. The complex conjugate expansion yields

$$\frac{V_1}{V_2} = \frac{(a + jb)}{(c + jd)} \frac{(c - jd)}{(c - jd)} = \frac{(a + jb)(c - jd)}{c^2 - d^2}, \quad (6.53)$$

and now the imaginary part can be separated from the equation. To meet the requirement of a 180 degree phase shift, the imaginary part has to equal zero, i.e.

$$-j(ad) + j(bc) = 0. \quad (6.54)$$

After performing the multiplications, one ends up with a biquadratic equation

$$A\omega^4 - B\omega^2 - 1 = 0 \quad (6.55)$$

where

$$A = R_1 R_2 C_1 C_2 \left[\frac{R_B}{r_\pi} R_1 R_2 (C_1 C_2 + C_1 C_3 + C_2 C_3) + R_B R_1 (C_1 C_3 + C_2 C_3) + R_B R_2 C_2 C_3 \right]$$

$$B = R_1 R_1 (C_1 C_1 + 2C_1 C_2 + C_2 C_2) + 2R_1 R_2 C_2 C_2 + R_2 R_2 (C_2 C_2 + C_2 C_3) + \left(1 + \frac{R_B}{r_\pi}\right) R_1 R_2 C_1 C_2$$

It is possible to derive a solution formula for equation (6.55) by making the substitution $x = \omega^2$. Proceeding from here,

$$\begin{aligned} Ax^2 - Bx - 1 &= 0 & | \cdot 4A \\ 4A^2 x^2 - 4ABx &= 4A & | + B^2 \\ (2Ax)^2 - 4ABx + B^2 &= B^2 + 4A \\ (2Ax - B)^2 &= B^2 + 4A \\ x &= \frac{B + \sqrt{B^2 + 4A}}{2A}. \end{aligned}$$

This equation looks familiar. It is almost the same as the quadratic formula. It was better to write it all here because the result needs clarification. Basically the

answer should have the term $B \pm \sqrt{B^2 + 4A}$ but because no resistor or capacitor value is negative, the square root term is *always* larger by value than B . If the minus sign is taken along, then one would allow negative frequencies as the answer, and that is absurd.

Having clarified this point, it can be concluded that from the four possible solutions to equation (6.55), only one of them is always physically meaningful for this specific case. Returning the substitution between x and ω^2 , the solution has the form

$$\omega = \sqrt{\frac{B + \sqrt{B^2 + 4A}}{2A}}$$

and since $\omega = 2\pi f$, a true frequency value is obtained from equation

$$f_{\text{osc}} = \sqrt{\frac{B + \sqrt{B^2 + 4A}}{8\pi^2 A}}. \quad (6.56)$$

Now that the expression for the oscillating frequency of the RC phase-shift oscillator has been found, the value of ω should be substituted back to equation (6.52) to find out the attenuation caused by the impedances in the RC network in the specific oscillating frequency. This back substitution also verifies that there have not been any mistakes in the previous calculations. The result from equation (6.52) should be negative and real-valued (although the possible imaginary part in numerical calculations can be diminishingly small).

So how accurately does equation (6.56) predict the actual frequency of oscillation for the circuit in Figure 6.46? Unfortunately not very accurately. Firstly, in the derivation of the equation the amplifier element was assumed to be ideal, and secondly, the transistor's obscure internal resistance r_π makes the results even more undeterministic. The circuit that fits exactly to the equations derived so far for oscillation frequency is depicted in Figure 6.49. Here the transistor's controlled current source $g_m v_\pi$ is replaced by a constant current source I_C .

In a general case, a theoretical condition for sinusoidal oscillations requires that the real part of a feedback amplifier return ratio $T(j\omega)$ should equal to -1 and the imaginary part of $T(j\omega)$ should equal zero. This condition is called the *Barkhausen Criterion* [15, p. 660]. Practical oscillators can, and should have, a loop gain that is greater than one since the oscillation amplitude is limited by nonlinear effects. When using a gain which makes the return ratio very close to unity in the phase-shift oscillator, the oscillating signal is an extremely clean sine wave. If the gain is increased to make return ratio grow above unity, the

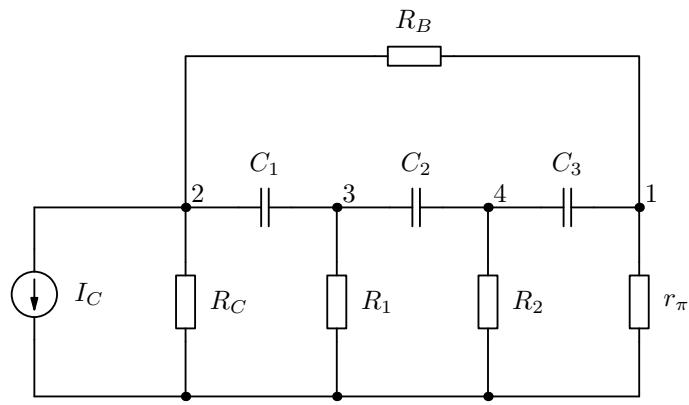


Figure 6.49: The RC phase-shift network without a transistor

sine wave becomes more and more distorted, and it is eventually clipped at the operating voltage limits. However, even if clipping occurs, the RC network filters the signal so that it can still be recognised as a sine wave, but not a very clean one.

Concerning the ongoing phase-shift oscillator analysis, the return ratio of the circuit is, by definition, evaluated as

$$T = -\frac{V_1}{\hat{V}_1} = -\frac{V_1}{V_1}, \quad (6.57)$$

where the notation \hat{V}_1 is used for differentiating the input and output signals at voltage node V_1 . The return ratio can be taken as a regular transfer function, where \hat{V}_1 is the input signal and V_1 is the output signal. Therefore, the correct way to evaluate the oscillation frequency of the circuit is to solve the output voltage V_1 from the 4×4 matrix equation using Cramer's rule and divide it by the term $g_m V_1$, which holds the input voltage \hat{V}_1 coming to node V_1 :

$$V_1 = g_m V_1 \frac{\det(V_1)}{\det(Y)}. \quad (6.58)$$

The calculation procedure is similar to the one already shown. As a result the transfer function is again of the form:

$$\frac{V_1}{g_m V_1} = -\frac{R_{xD}^2 a + jb}{R_B c + jd}, \quad (6.59)$$

where

$$a = 1 - R_1 R_2 (C_1 C_2 + C_1 C_3 + C_2 C_3) \omega^2$$

$$b = R_1 (C_1 + C_2) \omega + R_2 (C_2 + C_3) \omega - R_B R_1 R_2 C_1 C_2 C_3 \omega^3$$

$$c = 1 - C_1 C_2 [R_1 R_2 + (R_1 + R_2) R_{xA}] \omega^2 - C_1 C_3 (R_1 R_2 + R_1 R_{xB} + R_2 R_{xA} + R_{xD}^2) \omega^2 - C_2 C_3 [R_1 R_2 + (R_1 + R_2) R_{xB}] \omega^2$$

$$d = [C_1 (R_1 + R_{xA}) + C_2 (R_1 + R_2) + C_3 (R_2 + R_{xB})] \omega - C_1 C_2 C_3 [(R_1 + R_2) R_{xD}^2 + R_1 R_2 R_{xC}] \omega^3.$$

To shorten the equations, the following resistance combinations have been used:

$$R_{xA} = \frac{R_C(r_\pi + R_B)}{R_C + r_\pi + R_B}; R_{xB} = \frac{r_\pi(R_C + R_B)}{R_C + r_\pi + R_B}; R_{xC} = \frac{R_B(r_\pi + R_C)}{R_C + r_\pi + R_B}; R_{xD}^2 = \frac{R_C R_B r_\pi}{R_C + r_\pi + R_B}.$$

Without further elaborate derivations, the transfer function (6.59) can be used to find out the frequency of oscillation. The exact theoretical oscillation frequency can be evaluated with the help of the general quadratic formula, and the resulting equation is

$$f_{\text{osc}} = \sqrt{\frac{-B + \sqrt{B^2 - 4AC}}{8\pi^2 A}}, \quad (6.60)$$

where

$$A = R_B R_1 R_2 C_1 C_2 C_3 \left[(C_1 C_2 + C_1 C_3 + C_2 C_3) \left\{ R_1 R_2 \left(1 - \frac{R_{xC}}{R_B} \right) - (R_1 + R_2) \frac{R_{xD}^2}{R_B} \right\} + C_2 (C_1 R_{xA} + C_3 R_{xB}) (R_1 + R_2) + C_1 C_3 (R_1 R_{xB} + R_2 R_{xA} + R_{xD}^2) \right]$$

$$B = [(R_{xC} - R_B) R_1 R_2 - (R_1 + R_2)(R_1 R_{xB} + R_2 R_{xA})] C_1 C_2 C_3 - C_2 (C_1 R_{xA} + C_3 R_{xB}) [C_2 (R_1 + R_2)(R_1 + R_2) + (R_1 C_1 R_1 + R_2 C_3 R_2)] - R_2 (R_2 R_{xA} + R_{xD}^2) C_1 C_3 C_3 - R_1 (R_1 R_{xB} + R_{xD}^2) C_1 C_1 C_3$$

$$C = -(R_{xA} C_1 + R_{xB} C_3).$$

It would also be nice to be able to directly calculate the value for β_F that could be taken as a lower limit for producing a sufficient voltage gain for oscillation. An analytic expression for this purpose can be derived from equation (6.50), from where the unity gain relation

$$\frac{V_1}{V_1} = \frac{\beta_F}{r_\pi} \frac{\det(V_1)}{\det(Y)}$$

can be solved. Setting an equality $\frac{V_1}{V_1} = 1$ requires that the voltage arriving to node 1 equals the voltage that left that node. This leads to the useful equation

$$\beta_F \geq \frac{\det(Y)}{\det(V_1)} r_\pi \quad (6.61)$$

for the current gain factor. Equation (6.61) gives a theoretical minimum value for β_F to get the oscillator running for a certain component configuration. In another words, this value of β_F obtained from equation (6.61) gives theoretically a voltage gain of 1 when going around the circle in Figure 6.48 from node 1 back to node 1. In terms of the transfer function (6.59) the same minimum current gain factor is evaluated as

$$\beta_F \geq -\frac{R_B r_\pi}{R_{xD}^2} \frac{c}{a}. \quad (6.62)$$

Notice how only the terms a and c are included to the equation from the transfer function. This is because one can assume that at oscillation frequency the imaginary part of the transfer function vanishes completely.

In order to calculate the voltage gain of the transistor between nodes 1 and 2 of Figure 6.48, equation (6.51) gives the voltage relation

$$\frac{V_2}{V_1} = \frac{-\beta_F}{r_\pi} \frac{\det(V_2)}{\det(Y)}, \quad (6.63)$$

which can be used to estimate the voltage gain of the transistor with arbitrary component values in this specific configuration. Because of the impedance of the RC network, the phase shift in the amplifier element from node 1 to node 2 is not exactly 180 degrees at any frequency of oscillation. It is as if the amplifier would cause a phase 'drift' that lowers the frequency. The general effect of an unideal amplifier is that the gain is smaller and the oscillation frequency is typically lower in comparison to the ideal case already derived to give equation (6.56).

Typically design equations for tuning the oscillator to a certain frequency are given with the assumption that all capacitors and all resistors have the same value. This choice leads to a very simple equation to determine the oscillation frequency from. It is undoubtedly beneficial to select the capacitors to have the same value, because in addition to simplicity, the needed current gain is optimised by choosing equal capacitor values. To make the oscillator tunable by a potentiometer requires the design equations as derived above, because with these equations all the necessary parameters can be evaluated mathematically. An example Octave script for design aid is given in Appendix C.

The oscillation frequency in the tremolo circuit can be controlled by taking R_1 as a potentiometer. This is a nice idea, but there is one disadvantage when using the phase-shift oscillator as a potentiometer-controlled source of vibration. The voltage gain needed to keep the oscillator running varies with the values of R and C , which are used for tuning the phase-shift network to a certain frequency.

The voltage gain should compensate the attenuation caused by the RC network at the frequency of oscillation. Figure 6.50 shows the need for changing the voltage gain with different values of bias resistor R_B when the resistor R_1 is varied from $1 \text{ k}\Omega$ to $100 \text{ k}\Omega$.

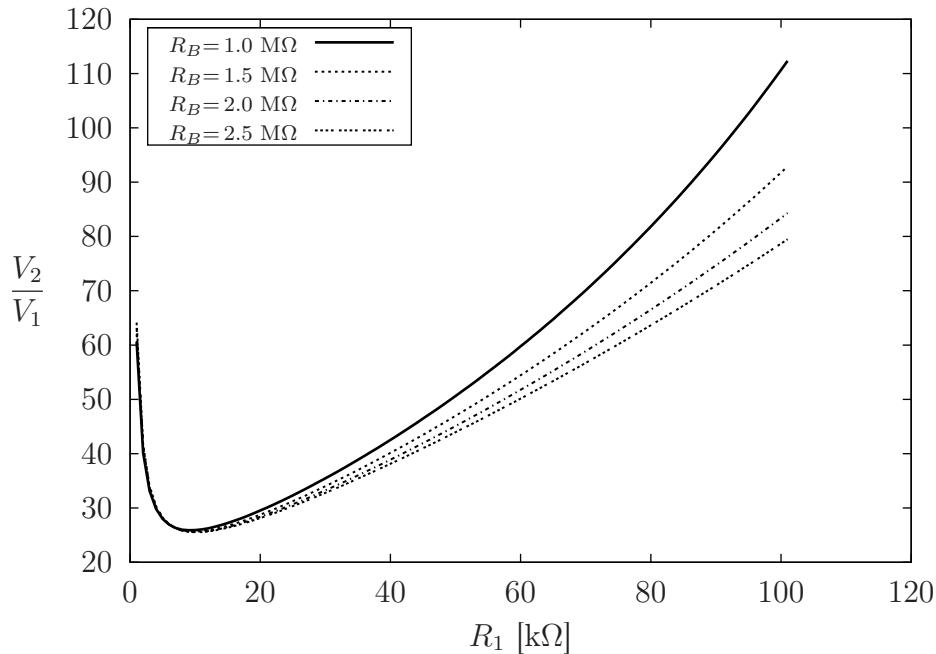


Figure 6.50: Voltage gain needed for oscillations

As R_1 is a potentiometer, the changes in its value reflect to the frequency of oscillation nonlinearly, as shown by the solid line in Figure 6.51. When the resistance range of R_1 is drawn using a logarithmic axis, the frequency dependency becomes linear. This linearised dependency is represented by a dashed line in Figure 6.51. If a linear potentiometer was to be used as R_1 , it would be difficult to set the higher frequencies accurately because a very small change in the resistance of R_1 would cause a relatively large change in frequency. It is therefore clear that a logarithmic potentiometer should be chosen as R_1 .

The transistor for the phase-shift oscillator needs to be chosen so that it has a large enough β_F to produce the required voltage gain. If the transistor is set up to give the maximum gain needed at the upper and lower limits of resistance R_1 , the middle values of R_1 will lead to a distorted sine wave output. This happens because the phase-shift oscillator gives a nice and clean sine wave if and only if the voltage gain provided by the transistor is only slightly larger than the attenuation of the RC network. Using a potentiometer to vary the frequency of oscillation forces to allow a distorted sine wave output at some range of frequen-

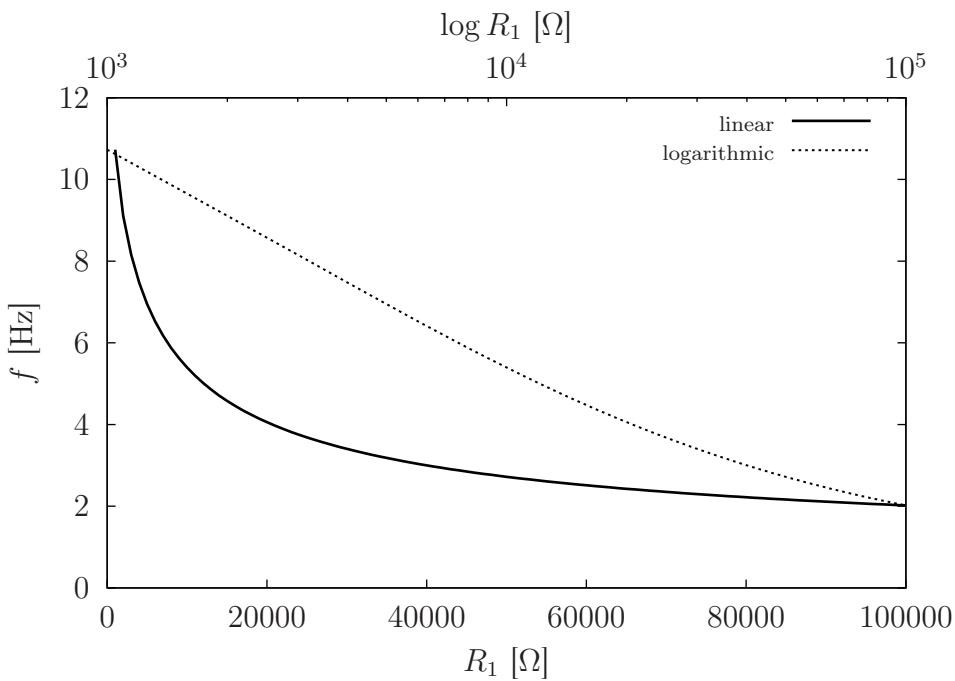


Figure 6.51: The oscillation frequency as a function of R_1

cies. At the limits of R_1 , the peak-to-peak amplitude of the generated sine wave might be lower than in the middle section due to the different gain requirements.

The different gain factors (voltage gain, required current gain and attenuation of the RC network that indicates the minimum required voltage gain) are visualised in Figure 6.52. The gain curves have been calculated using the actual tremolo circuit component values and a fixed value of 160 for the current gain factor β_F . Figure 6.52 indicates that the oscillations will occur through the whole range of R_1 values. In the location where the gap between the actual gain and the required gain is at its largest, the output signal of the oscillator will be a slightly distorted sine wave. If the curve of the required gain goes higher than the actual gain at some value of R_1 , then oscillations would fade out nearby that area.

It needs to be emphasised that the previous analysis was made from a purely theoretical point of view and it used a few simplifications in the transistor model. The situation is a little bit different when building the oscillator circuit in practise, and it might turn out that the theoretical value for the unity gain β_F is not enough to make the circuit oscillate. This configuration is especially difficult because the oscillation frequency and also the voltage gain depend on the transistor's internal resistance r_π , the value of which depends on the DC biasing conditions and is difficult to determine accurately by theoretical means.

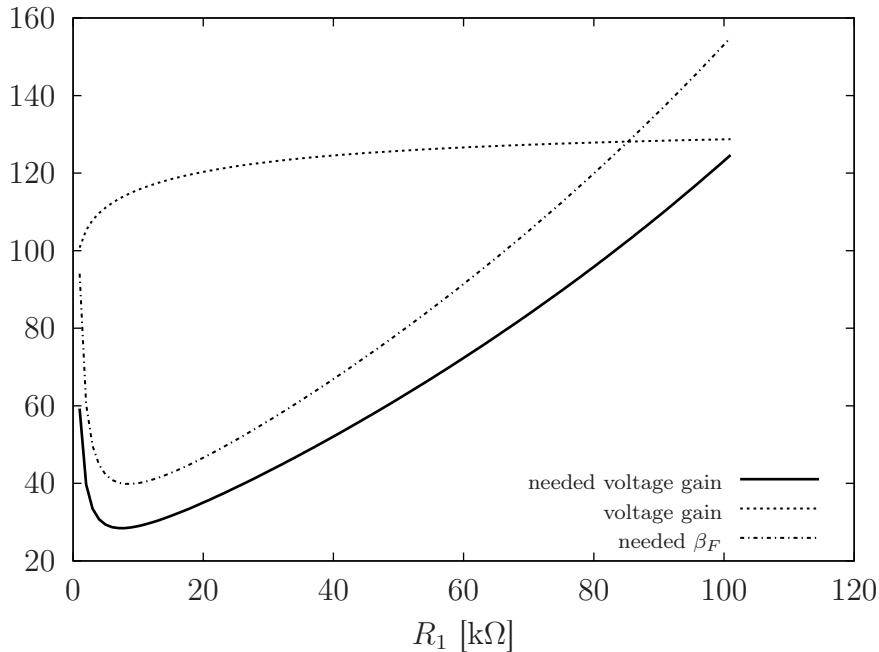


Figure 6.52: Gain curves for RC phase-shift oscillator design

6.3.5 A SPICE model for the tremolo circuit

The tremolo schematic was redrawn using gschem to be able to directly generate the netlist file for ngspice simulations. The schematic includes all the components assigned with their actual values. The potentiometers in the circuit are split into two separate resistors to be able to simulate the voltage divider functionality by changing simultaneously the values of the two resistors. As a slight difference to the schematic 6.40, there is one additional $2.2\text{ M}\Omega$ resistor R_1 at the input of the circuit, and R_{14} has been added in series with potentiometer R_{15} to add a base resistance to compensate the situation where the potentiometer R_{15} is short-circuited. The transistors used in this model are 2N5088 for both bipolar transistors and 2N5457 for the JFET.

When using the gEDA design tools collection in Linux, the netlist for the SPICE simulations is created from the schematic of Figure 6.53 using the command

```
gnetlist -g spice-sdb -o tremolo_net.net tremolo.sch
```

where the output file named as `tremolo_net.net` is created from the schematic file `tremolo.sch`. The created netlist is printed here for convenience to show the model parameters used for the 2N5088 (`.model NPN1`) bipolar junction transistor and the 2N5457 (`.model NFET1`), although these SPICE models are readily available in the datasheets provided by the manufacturers, which can easily be

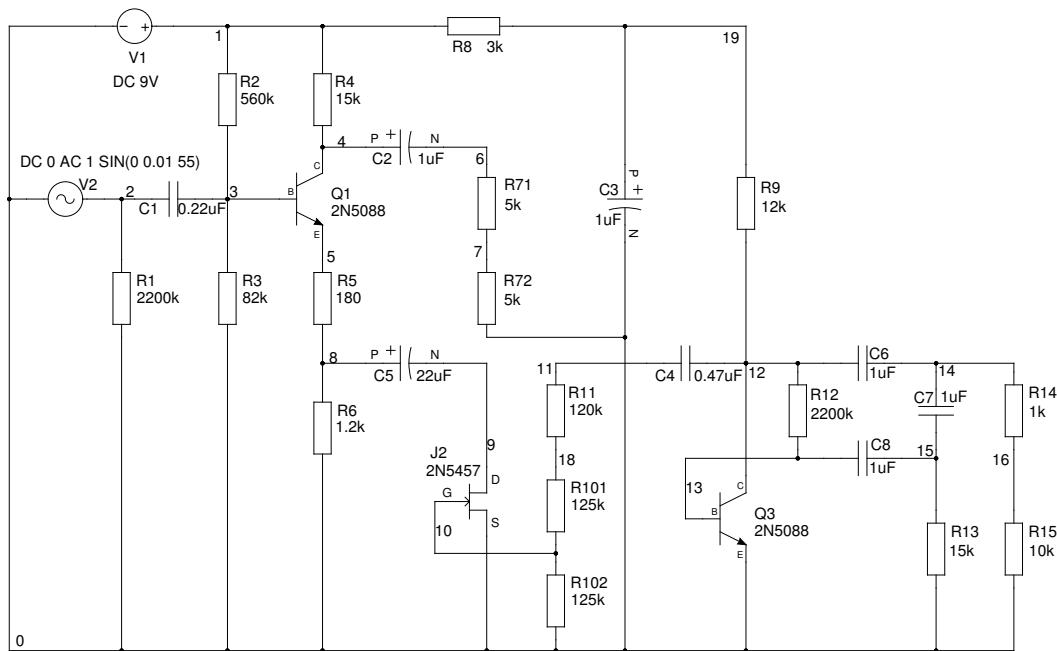


Figure 6.53: The tremolo schematic drawn with gschem

found from the Internet.

```
=====
 Begin SPICE netlist of main design =====
J2 9 10 0 NFET1
.MODEL NFET1 NJF (VTO=-1.8 BETA=0.00135 LAMBDA=0.001 RD=35 RS=31.5
+ CGS=2.25E-12 CGD=6E-12 KF=6.5E-17 AF=0.5)
V2 0 2 DC 0 AC 1 SIN(0 0.01 55)
V1 1 0 DC 9V
R102 0 10 125k
R9 12 19 12k
C3 19 0 1uF
R71 7 6 5k
R15 0 16 10k
C7 15 14 1uF
R12 13 12 2200k
R101 10 18 125k
Q3 12 13 0 NPN1
Q1 4 3 5 NPN1
.MODEL NPN1 NPN (Is=5.911f Xti=3 Eg=1.11 Vaf=62.37 Bf=1.122K Ne=1.394 Ise=5.911f Ikf=14.92m
+ Xtb=1.5 Br=1.271 Nc=2 Isc=0 Ikr=0 Rc=1.61 Cjc=4.017p Mjc=.3174 Vjc=.75
+ Fc=.5 Cje=4.973p Mje=.4146 Vje=.75 Tr=4.673n Tf=821.7p Itf=.35 Vtf=4 Xtf=7 Rb=10)
C5 8 9 22uF
C2 4 6 1uF
C8 15 13 1uF
C4 11 12 0.47uF
C6 12 14 1uF
C1 2 3 0.22uF
R14 16 14 1k
R13 0 15 15k
R11 18 11 120k
R72 0 7 5k
R8 1 19 3k
R1 0 2 2200k
R6 0 8 1.2k
R5 8 5 180
R4 4 1 15k
R3 0 3 82k
R2 3 1 560k
.end
```

With this tremolo circuit configuration, one is mainly interested in the transient simulation since the frequency response is already handled by the manual approximate form of analysis in section 6.3.2.

To perform the transient simulations for the tremolo circuit, adequate control commands are added to the beginning of the netlist file. In this specific case it was noticed that the RC phase-shift oscillator needed about 20 seconds of simulation time to properly get up and running, so therefore the simulation is done in the interval from 19 to 20 seconds. The simulation is run twice with different settings of the potentiometer R_{10} that controls the modulation depth. The operating point analysis is also added to see how the transistors are biased in the static DC configuration. The simulation control commands are now set as

```
.control
tran 100us 20s 19s
alter R101 1k
alter R102 250k
tran 100us 20s 19s
gnuplot tre_tran tran1.v(7) tran1.v(10) tran2.v(7) tran2.v(10)
.endc

.OP
```

and the results are requested to be saved to a gnuplot formatted data file. Eventually, the simulation is run using the command

```
ngspice -b tremolo_net.net
```

which prints the operating point information to the computer screen and creates the data file `tre_tran.data` to be plotted later with gnuplot.

6.3.6 Results of simulations, calculations and measurements

The actual prototype of the simulated tremolo circuit 6.53 was built on a medium size solderless breadboard. It was not easy to find the 2N5088 and 2N5457 components from the local dealers, so replacements had to be found for these transistors. A bipolar junction transistor BC549C was used instead of 2N5088, and 2N3904 was also experimented in the phase-shift oscillator circuit. The BF245B and BF245A field-effect transistors were examined as replacements for the 2N5457. The BF245B does not seem to be a direct replacement for 2N5457 since the measured values for I_{DSS} and V_P of BF245B were 11 mA and -3.6 V respectively. The simulation model of 2N5457 had $V_P = -1.8$ V, and it matches much better with the BF245A, which was measured to have a pinch-off voltage of -1.7 V. The parameter measurements of BF245B and BF245A were made as

described in Figures 6.44a and 6.44b.

Despite the replacement parts used in the prototype, the simulations were run using the SPICE models for 2N5088 and 2N5457 because these models were easily obtained from the manufacturer's datasheet. One can also argue on the validity of these models since many of the JFET's parameters cannot be exactly defined. For example, even if the model gives $V_P = -1.8$ V for the 2N5457, the real component might have $V_P = -0.8$ V due to large variations in the value of the pinch-off voltage. So again, it is better to just throw in some approximate values and see what happens. Luckily the tremolo circuit is not that sensitive to a certain JFET parameter. The functionality of the effect depends only on the pinch-off voltage V_P .

The results of the DC biasing calculations of the tremolo circuit are gathered into Table 6.2 along with the simulated and measured values. The quiescent values of transistor Q_1 are referenced by using a subscript 1, and the quiescent values of transistor Q_3 used in the oscillator stage are referenced with a subscript O (as Oscillator). The voltage node 8 refers to the indexed wiring junctions marked in the gschem schematic 6.53.

Table 6.2: Tremolo circuit quiescent voltages

value	simulated	calculated	measured
V_{B1}	1.11	1.10	1.03
V_{C1}	3.93	3.55	3.32
V_{E1}	0.47	0.50	0.51
V_8	0.40	0.43	0.44
V_{BO}	0.64	0.60	0.58
V_{CO}	2.34	2.36	2.56

A V_{BE} of 0.6 volts was assumed in the numerical calculations. The current gain factor β_F was 550 as indicated as an average value for the BC549C transistor, which was chosen to be used in the prototype build. The collector voltage V_{CO} of the oscillator transistor depends largely on the current gain factor β_F , so if different transistors are used, the value of V_{CO} will change. Generally the results for the bias voltages are in agreement, although the operating point analysis in the simulations was done with the SPICE model of the 2N5088 transistor, which has a β_F of over a 1000.

The transfer characteristics of the tremolo circuit have already been demonstrated in Figure 6.42. The frequency response passes all relevant audio frequency components nicely and the change in the gain does not alter the frequency characteristics. The value of the input capacitor C_1 can be increased if very low frequencies are involved.

A transient analysis from the simulated tremolo circuit produces the waveform shown in Figure 6.54. When simulating the transient behaviour, the input signal was set to a sine wave with a 55 Hz frequency (the open A-string of a bass guitar) and the amplitude to 0.01 volts. This frequency was chosen because of the visualisational benefits of showing the signal clearly under the modulation envelope. The bottom part of Figure 6.54 shows that the 55 Hz sine is inside the modulating envelope, which is generated by 'multiplying' the original input signal with the low-frequency oscillator signal.

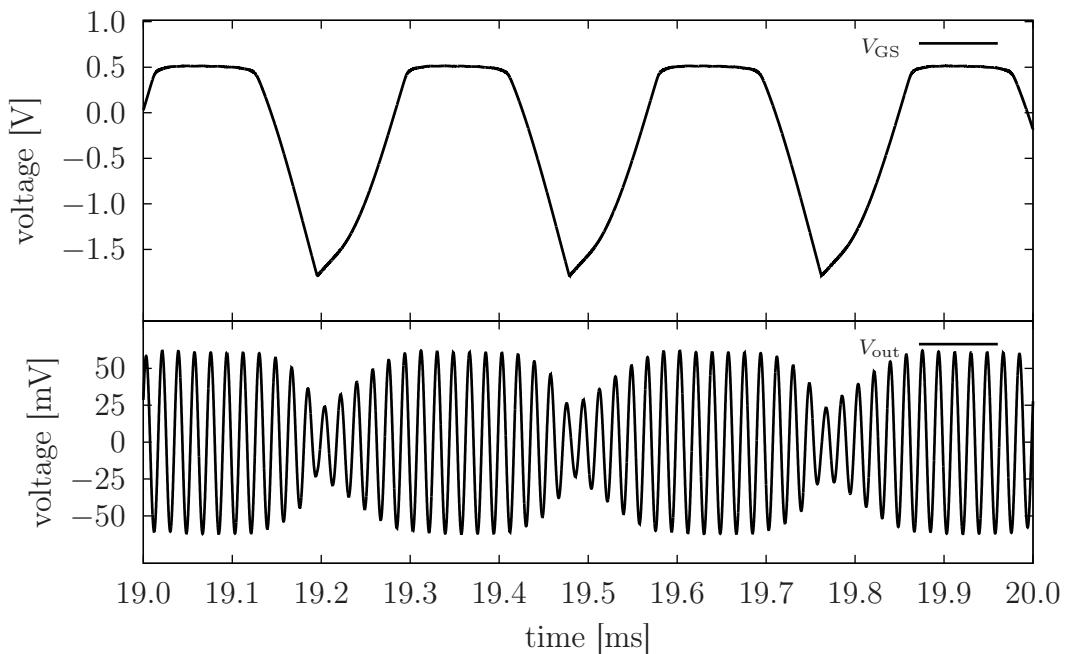


Figure 6.54: A transient simulation of the tremolo circuit

The top part of Figure 6.54 also presents the voltage at the gate pin of J_2 . The voltage from the phase-shift oscillator has a peak-to-peak voltage of approximately 8 volts, and after capacitor C_4 this voltage swings ± 4 volts around zero volts. The potentiometer R_{10} is used to scale the oscillator output voltage to the gate of J_2 . Because the waveform of the gate voltage is so far from a clean sine wave, in this case the gain of the oscillator transistor is too high with respect to the Barkhausen criterion.

When the oscillator voltage is negative, it reverse biases JFET J_2 with a negative v_{GS} . When the voltage of the JFET gate goes down close to the pinch-off voltage, the parallel resistance of J_2 and R_6 grows to the maximum, which equals the value of R_6 alone. The maximum resistance reduces the small-signal gain of Q_1 on the input signal. The oscillating resistance on the emitter of Q_1 produce the roller-coaster form of output from this tremolo circuit. It is interesting to see that the envelope has the sinusoidal form, although the resistance of the JFET changes so nonlinearly with the changing v_{GS} . Apparently the parallel resistor R_6 has been chosen very carefully to linearise the JFET resistance changes.

The depth of the modulation can be controlled by potentiometer R_{10} . The maximum modulation depth is obtained by adjusting the potentiometer to give a maximum resistance of $250\text{ k}\Omega$ between the JFET gate and ground. At the maximum modulation depth, the peak-to-peak voltage of the phase-shift oscillator at the gate of J_2 is not attenuated at all by R_{10} . This change in the modulation depth is visualised in Figure 6.55. It is clearly seen that when V_{GS} hits the pinch-

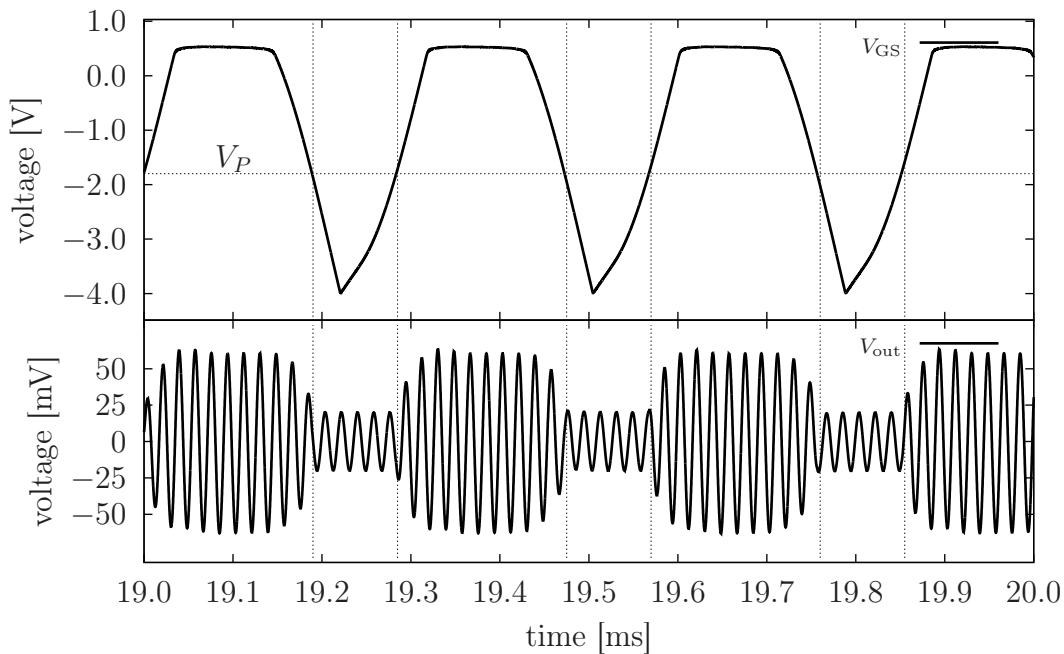


Figure 6.55: Maximum modulation depth

off voltage and beyond, the emitter resistance is effectively equal to the constant value of R_6 because the resistance of the JFET is so much larger than R_6 . The modulated output signal envelope in this case looks more like a result of a square wave modulation than the actual sinusoidal modulation.

One can also observe from the simulations that the gain changes with a factor of

2 as the JFET changes state from closed to open. The factor of 2 in a voltage gain means a 6 dB change in decibels. This correlates relatively nicely with the gain curves presented in Figure 6.42. This describes the relevant parts in the transient behaviour of the tremolo effect in a nutshell.

The prototype build of the tremolo effect circuit was first tested with a complete setup of a guitar and an amplifier to verify that the circuit actually sounds as intended. Luckily it did, after correcting some minor wiring errors on the plug-board. As usual, nothing ever works straight away on the first trial. The best sounding effect was achieved when the oscillator was giving a modulation signal of about 3 - 5 Hz. This aurally tuned potentiometer setting was chosen to be used also for the official measurements.

To be able to compare with the simulation results, a 55 Hz sine wave was played through a computer sound card as an official test signal for the prototype measurements. Just as in the case of the simulations, the frequency of 55 Hz was chosen for the test signal because the signal inside the modulation envelope was adequately visible with this frequency. The output of the sound card was directly connected to the input terminals of the prototype tremolo circuit with a regular guitar cable. The output terminals of the tremolo circuit were not connected to an amplifier during the measurements, but instead a digital storage oscilloscope was used to trace down the waveforms in the tremolo circuit.

Figure 6.56 shows the measured input and output waveforms from the prototype circuit using the BF245B JFET. The potentiometer R_{10} was set to the full 250 k Ω value to make the amplitude modulation work sufficiently. The measured output waveform is similar with the simulated amplitude modulated waveform of Figure 6.54.

As a general comment with reference to Figures 6.54 and 6.56, the amplitude gain of the prototype circuit is about 10, which correlates with the simulations. The measurement of the output signal was done over a fixed 10 k Ω resistor, whereas the simulation output signal was recorded with the output potentiometer in the halfway range, i.e. 5 k Ω . This difference attenuates the simulation results by a factor of 2, which makes the results between the simulations and measurements similar. The voltage gain of the circuit was higher than expected, so most likely a transistor with a smaller β_F would be better to replace the BC550C as Q_1 . The signal fed from the computer sound card was set to an unnecessary high level, therefore it was not perfectly simulating the signal coming from the

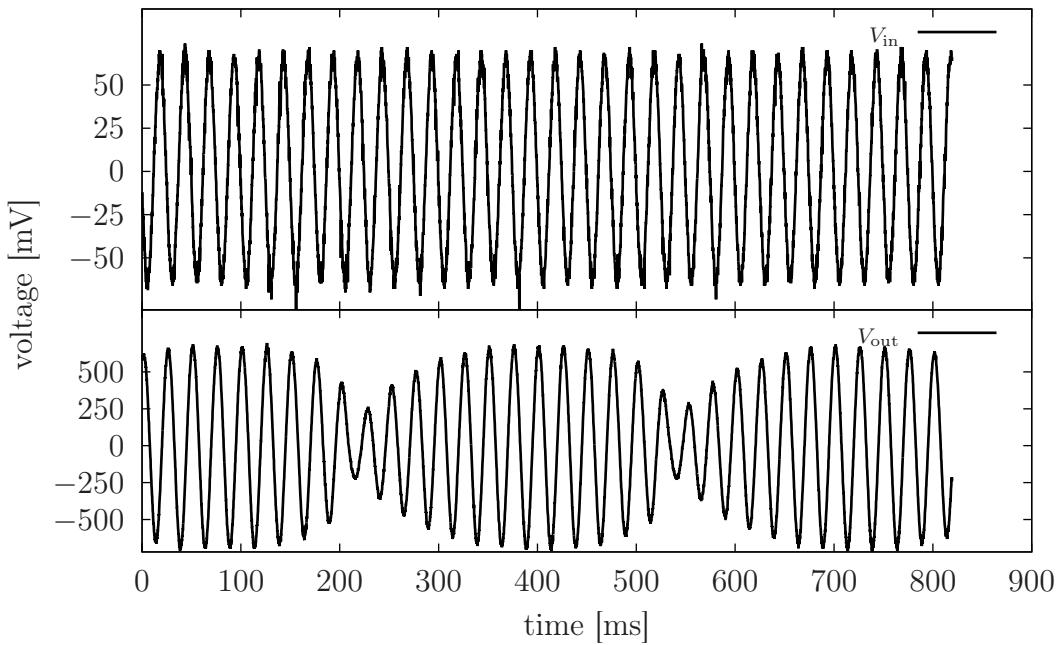


Figure 6.56: The measured input and output voltages of the tremolo effect

electric guitar output jack.

Figure 6.57 shows how the sine wave generated by the phase-shift oscillator relates to the output signal of the tremolo circuit. The amplification does indeed change at the rate set by the oscillator. Moreover, what is essential to note in Fig-

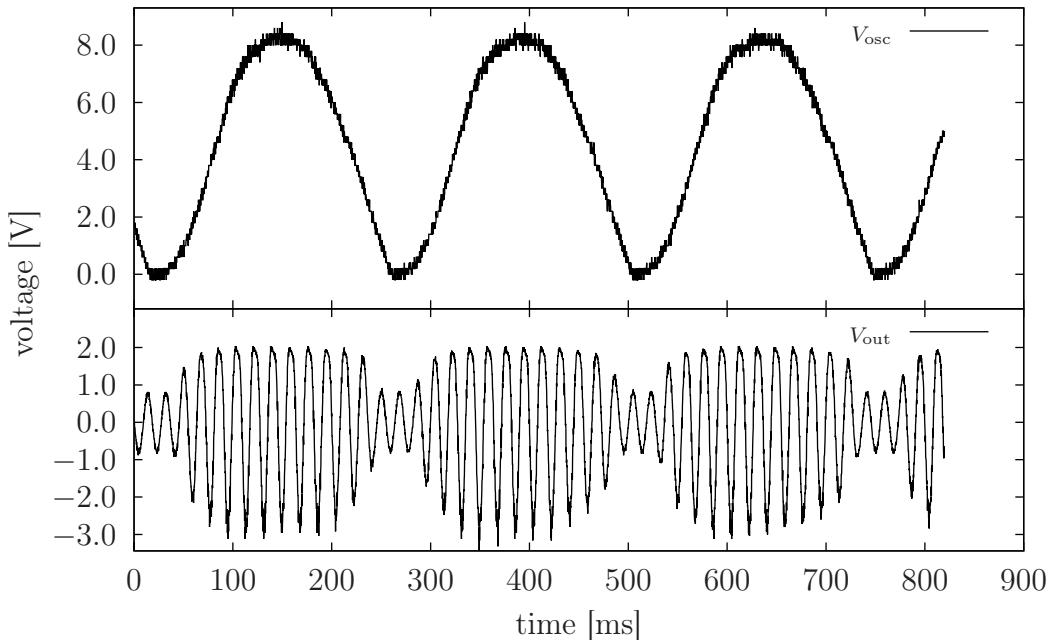


Figure 6.57: Oscillator output versus tremolo output

ure 6.57 is the voltage range of the oscillator's sine wave. When using transistors with a high β_F , the peak-to-peak voltage of the oscillator is about the same as V_{CC} . If the gain of the transistor is not very high but still sufficient to keep the oscillator running, V_{p-p} might be a few volts under V_{CC} . Therefore, with different transistors the changes in the oscillator's peak-to-peak voltage might lead to a nonuniform modulation depth within the adjustable frequency range provided by potentiometer R_{15} .

To compare the effect of specific pinch-off voltages of the used JFETs, the gate voltage and the tremolo output waveforms were recorded for both BF245B and BF245A. According to Figures 6.58 and 6.59, a JFET with a pinch-off voltage of approximately -2 volts works best for this circuit, since the depth can then be controlled in a more versatile way. Clearly the V_P of -3.6 V of BF245B is too high and does not provide a good adjustability for the modulation depth. The BF245A

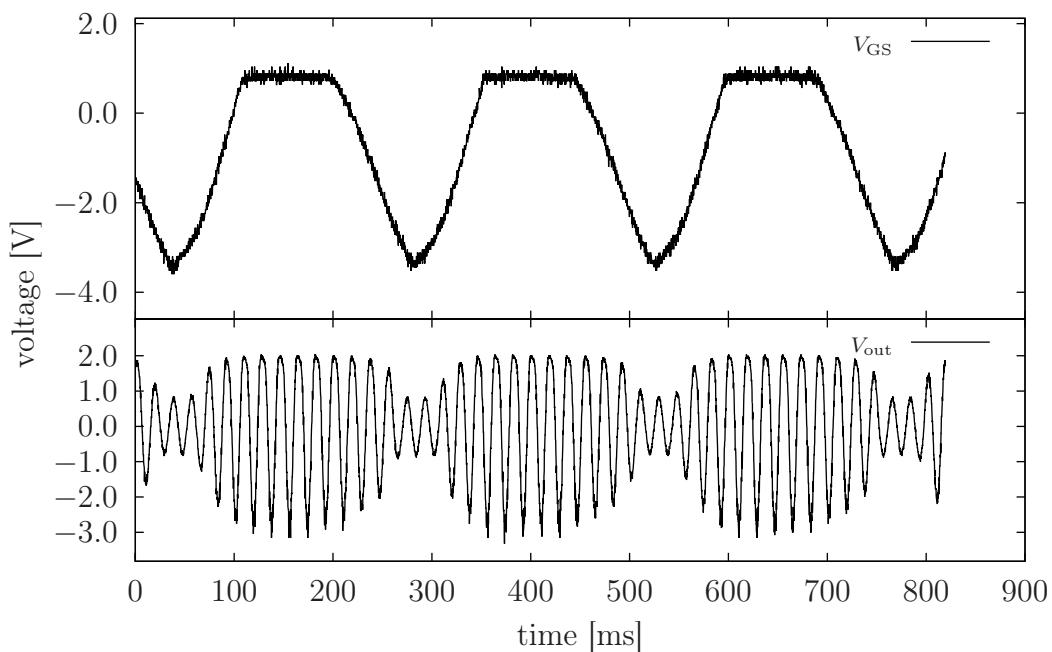


Figure 6.58: JFET BF245B gate voltage versus tremolo output

behaves almost exactly like the 2N5457 in the simulations. Clearly the pinch-off voltage is lower in BF245A than in BF245B, and therefore leads to better control on the modulation depth.

The phase-shift oscillator was also measured as a separate circuit, because of the wide theoretical treatment given in section 6.3.4. Transistors 2N2369, 2N3904 and BC549C having a different β_F were tested to find out if the theoretical analysis could predict the frequency of oscillation and the gain properties correctly.

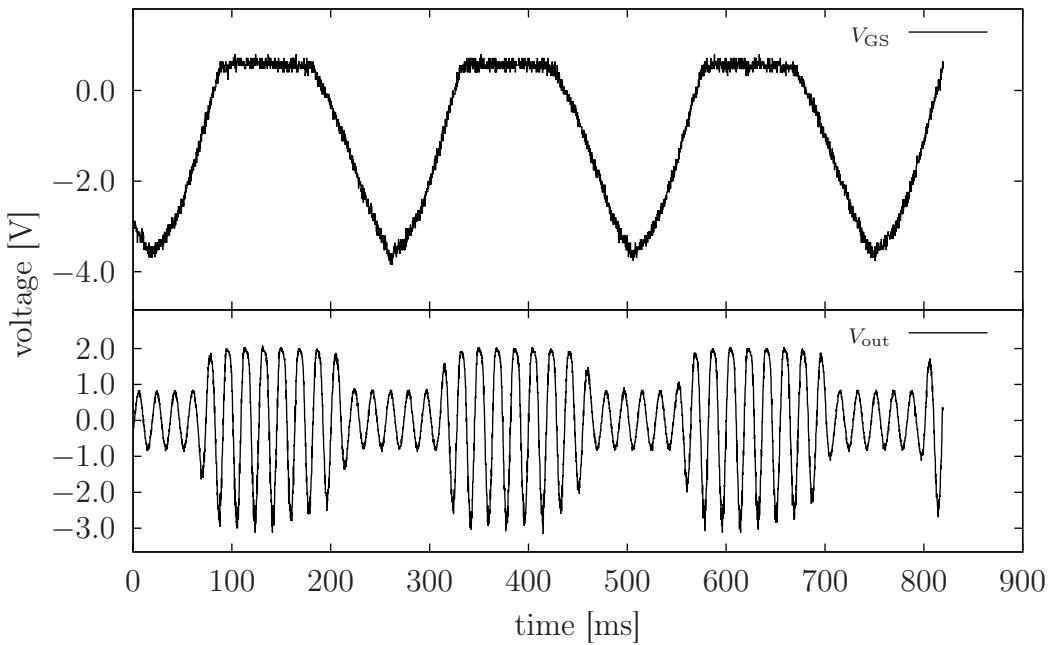


Figure 6.59: JFET BF245A gate voltage versus tremolo output

For comparing the measured oscillation frequencies to the analytical values, equation (6.60) was used for determining the theoretical upper limit for oscillation frequencies within the range of potentiometer R_{15} set at maximum and minimum value. The measurements were made for a few selected values of R_{15} so that normal resistors with fixed resistance values were used instead of a potentiometer. It was already noted that the nonideal amplifier causes the actual oscillation frequency to be lower than the theoretical frequency of the isolated RC network. The frequency value provided by equation (6.60) was used for calculating the theoretical values of minimum β_F and gain with respect to the varying resistance provided by potentiometer R_{15} .

The specific properties of the transistors used were obtained primarily by measurements. The current gain β_F was measured using a basic multimeter with h_{FE} measurement capability. The value of r_π for the analytical calculations was determined by measuring the voltage at the collector of Q_3 and determining the bias current I_{CQ} with regular Ohm's law calculations. Then, from here one can proceed by combining the approximate BJT formulae $g_m V_T \approx I_{CQ}$ and $g_m r_\pi = \beta_F$ as

$$r_\pi \approx \frac{\beta_F V_T}{I_C}, \quad (6.64)$$

where V_T is approximately 25 mV at room temperature. The value of I_{CQ} was determined using the results of the manual bias calculations for the oscillator

transistor Q_3 presented in section 6.3.1. This theoretical result was compared to the value of I_{CQ} obtained by the measurements. Table 6.3 summarises the results of the measurements.

Table 6.3: Phase-shift oscillator measurements

	2N2369		2N3904		BC549C	
value	meas.	calc.	meas.	calc.	meas.	calc.
β_F	77	-	232	-	553	-
r_π	12.2 kΩ	10.7 kΩ	23.3 kΩ	18.1 kΩ	34.5 kΩ	33.3 kΩ
A_V	65	70	115	114	168	184
f_3 kΩ	—	—	5.20 Hz	5.61 Hz	5.45 Hz	5.31 Hz
f_{10} kΩ	3.91 Hz	4.51 Hz	3.44 Hz	4.07 Hz	3.56 Hz	3.82 Hz
f_{15} kΩ	3.43 Hz	3.96 Hz	3.00 Hz	3.60 Hz	3.09 Hz	3.37 Hz
f_{22} kΩ	3.06 Hz	3.47 Hz	2.66 Hz	3.16 Hz	2.72 Hz	2.97 Hz
f_{39} kΩ	2.64 Hz	2.80 Hz	2.26 Hz	2.58 Hz	2.25 Hz	2.43 Hz
f_{47} kΩ	—	—	2.16 Hz	2.40 Hz	2.10 Hz	2.27 Hz
f_{68} kΩ	—	—	1.98 Hz	2.10 Hz	1.86 Hz	2.01 Hz
f_{82} kΩ	—	—	1.88 Hz	1.96 Hz	1.78 Hz	1.87 Hz

In Table 6.3, the gain A_V has been measured as the ratio of the peak-to-peak voltages at the base and collector of the transistor. The measurement has been made at the highest value of R_{15} because then the needed gain has been close to the actual gain, and the faulty values due to clipping have been hopefully avoided. The gain results seem reasonable and are also in good agreement with the calculated values from equation (6.63).

The calculated values for the frequency of oscillation are still a bit overshooting but the consistency is good, meaning that there is not a value that would be completely out of the scope. Therefore, the frequency values predicted by the calculations are relatively accurate and this means that the method used to derive the equations can be used for designing similar phase-shift oscillators for arbitrary frequencies. The gain of transistor 2N2369 was too low for some of the values of R_{15} and this was also nicely seen in the calculations, since the minimum β_F given by equation (6.61) went over the actual β_F of transistor 2N2369 for all the missing values.

6.4 A COMPRESSOR AND AN EXPANDER

Compression and expansion techniques can be categorised mainly as recording studio effects, but some compression devices do also exist as effect pedals for electric guitars. The idea of a compressor is simple: if some notes are played so strongly that their amplitude exceeds a preset limit, the compressor device becomes active and attenuates the overly loud note. A compressor can also give more gain to very low level signals to even out the average amplitude level of the signal in the long run. The compressor effect simply reshapes the amplitude envelope of sharp and quickly decaying sounds into a more constant amplitude envelope. As an example, the amplitude envelope of a plucked guitar string can be varied in many ways by using the compressor with different settings. Figure 6.60 shows a waveform of a sharp pluck of a single string and the compression parameters relating to that waveform.

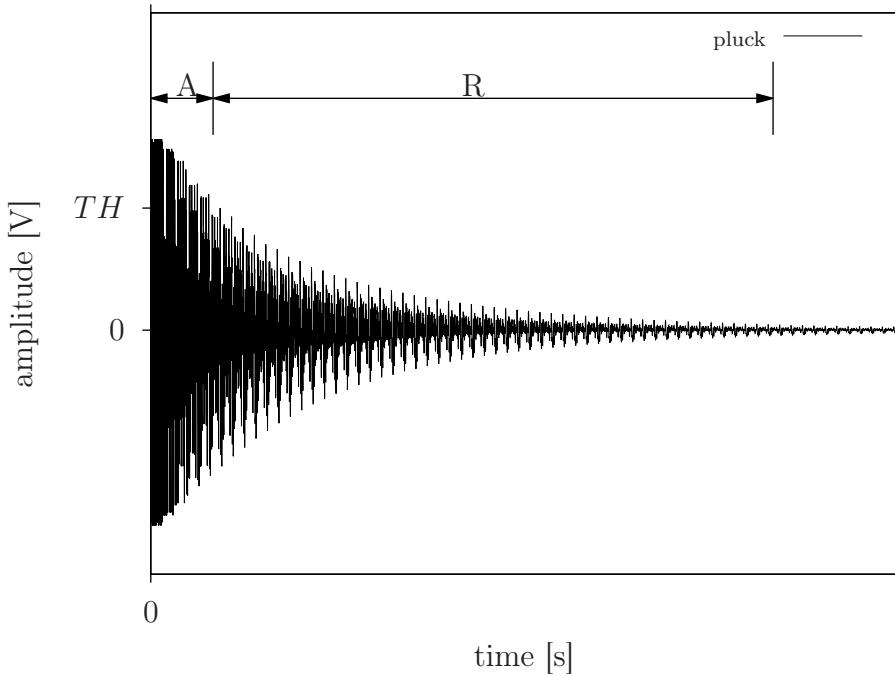


Figure 6.60: The attack (A) and release (R) times of a sharp pluck

There are four basic parameters that can be controlled on all compressors: *threshold level*, *attack time*, *release time* and *compression ratio*. Additionally, the compressor device might also include a gain section. When related to compression, the gain is referenced as make-up gain. Three of the basic compressor parameters are shown in Figure 6.60. The compression process starts when the amplitude of the signal goes over the threshold level (TH).

The attack time (A) describes how fast the compressor reacts after the signal

has exceeded the threshold level and is able to attenuate the amplitude back towards and below the threshold level. Short attack times make the compressor react to fast crossings of the threshold level, while longer attack times allow short impulses to go over the threshold without compression; only longer transcending signals are acknowledged by the compressor.

The release time (R) refers to the period where the compressor starts to deactivate itself by decreasing the attenuation and finally ending up to an uncompressed state, where the total gain of the system is back to normal. The release time is typically clearly longer than the attack time.

At the time when the output signal is over the threshold level, the compression ratio determines the amount of attenuation applied to the input signal. As an example, a 4:1 compression ratio is reached in a situation where the input signal is four times higher above the threshold than the output above the threshold. If the input is 4 dB above threshold, then the output is 1 dB above the threshold with a 4:1 compression ratio. A bit strange definition, but that is the way it has been defined.

The make-up gain is a factor that is used to give some extra gain to keep the average amplitude close to the threshold level when the compressor is not attenuating the amplitudes cutting across the threshold. [66, pp. 131 – 134]

One of the first analogue guitar effect pedals that created amplitude compression is called an 'Orange Squeezer', originally designed by the analogue effects guru Dan Armstrong. The naming of the effect pedal follows the same line of his colourful product names such as the 'Green Ringer' and 'Red Ranger'. The implementation of the Orange Squeezer is simple, but nonetheless it is an interesting and very powerful device. That is why it is good to go through the functionality of the circuit and learn something about compression. The schematic of the original Squeezer pedal is presented in Figure 6.61 and it is redrawn from the schematic found from the Tonepad website [81].

The Orange Squeezer circuit works in a feedback mode so that it monitors the output level and attenuates the input signal when the output voltage exceeds a certain preset level. Another option to implement a compressor would be to use a feedforward mode where the level of the input signal is monitored, but the feedback mode is usually a more robust solution. The gain block of a general compressor device is often referred to as a voltage-controlled amplifier (VCA).

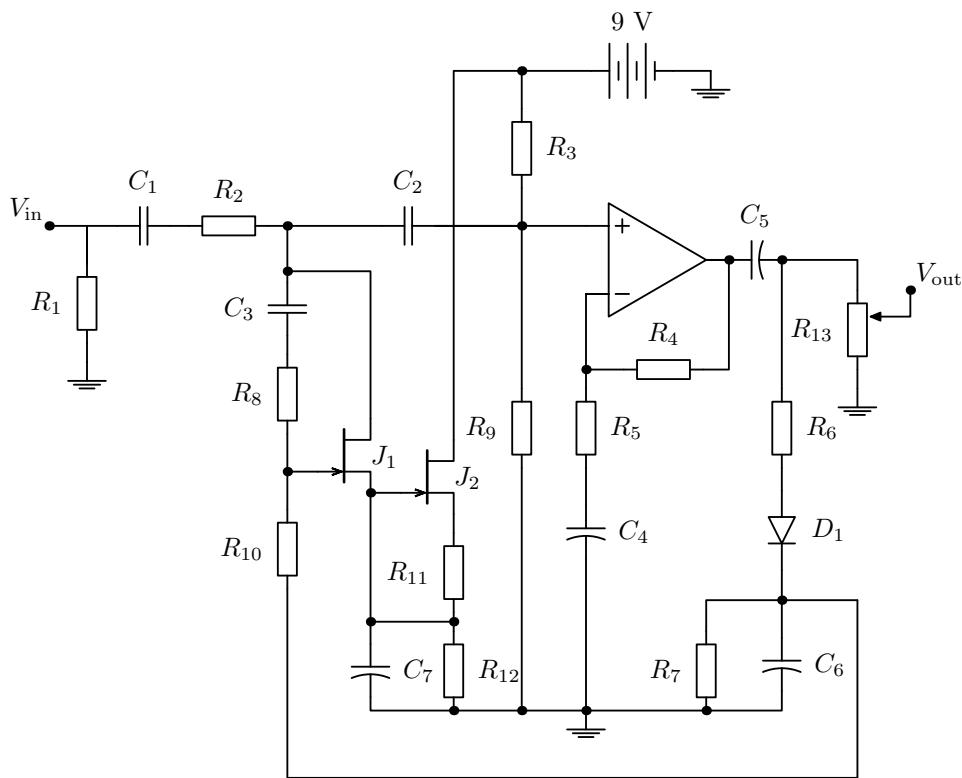


Figure 6.61: A circuit diagram of the Orange Squeezer effect pedal (for component values, see appendix D)

The capacitor C_6 acts as the level monitor at the output side and it is charged through resistor R_6 . The job of diode D_1 is to pass only the positive voltage pulses to charge C_6 and force the discharging process of the capacitor through R_7 instead of the combination of R_6 and R_7 . This is not that elegant a feature because the diode misses the negative pulses, and additionally creates an unnecessary voltage drop across itself. Therefore, it is necessary to use a germanium diode as D_1 , because the voltage drop is then only about 0.2 volts instead of the 0.6 volts provided by silicon diodes.

After C_6 has charged close to the peak voltage, the voltage level of C_6 is seen at the gate of the JFET J_1 . Here again the JFET is used as a voltage-controlled resistor, which lowers its resistance as the voltage at C_6 goes higher. The resistance of J_1 along with R_2 make up a voltage divider, so when the resistance of J_1 is low, more current is directed to ground instead of the amplifier and the input signal is attenuated this way.

When the input signal attenuates, it also attenuates the output voltage of the op-amp and the capacitor C_6 starts to discharge through resistor R_7 . To say it in other words using official compression terms, the attack time (which is approxi-

mately the time constant of R_6 and C_6) of this compressor can be controlled by changing the value of R_6 and the release time is determined by the time constant of C_6 and R_7 . The operational amplifier is hooked up in a typical non-inverting configuration, which offers a constant gain to the input signal. Basically the op-amp should provide the make-up gain, but in this setup the gain control comes from the combination of the attack and release times.

Figure 6.62 illustrates the output voltage peak detector configuration used in the Orange Squeezer circuit. The waveforms are simulated from an idealised test circuit where a square wave generator is feeding the capacitor through a $1\text{ k}\Omega$ resistor and a diode with a 0.2 V potential drop. The capacitor is connected in parallel with a $1\text{ k}\Omega$ resistor just like in schematic 6.61.

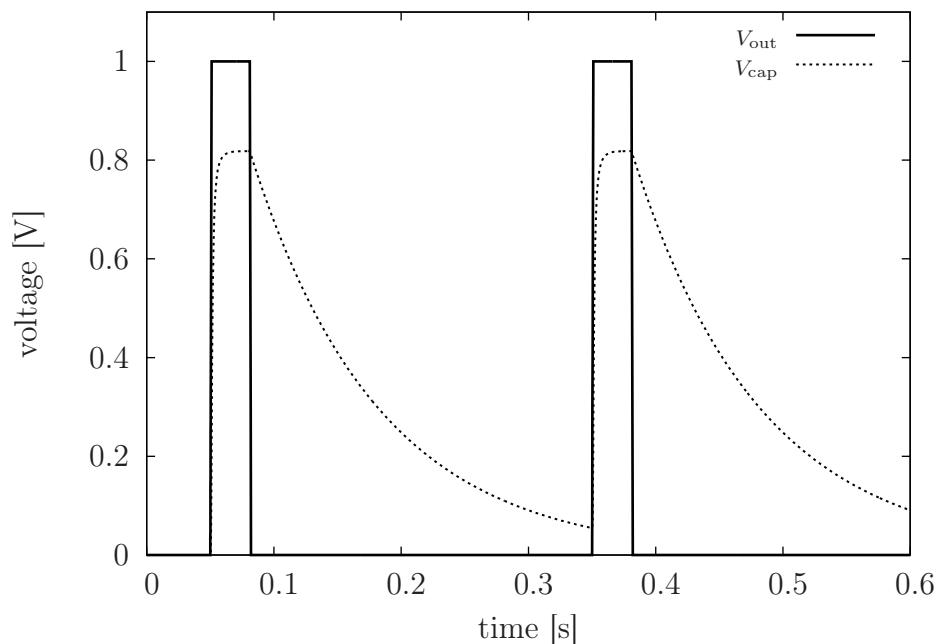


Figure 6.62: Charging and discharging process of a peak detector capacitor

A short description of Figure 6.62 is that when the short pulse of the output voltage V_{out} rises to 1 volt, the voltage V_{cap} across the capacitor builds up very rapidly close to V_{out} (delimited by the 0.2 volts of the diode). After the capacitor has reached full charge, the voltage across the capacitor stays at a constant level because V_{out} stays at a constant level in this imaginary example. When V_{out} drops lower, the capacitor starts discharging slowly. The discharged current from the capacitor cannot flow past the reverse-biased diode, so the resistor R_7 consumes all the current. Since the value of R_7 is large, it takes a very small amount of current from the capacitor and the discharge process is relatively slow. The

capacitor is charged again when the diode is forward-biased by large V_{out} . This happens when V_{out} is about 0.2 volts higher than the voltage across the capacitor.

The threshold level in the Orange Squeezer is set with the resistor R_{12} , which defines the voltage at the source of JFET J_1 . The voltage over R_{12} is generated with the constant current provided by JFET J_2 . In the original device, resistor R_{12} is a trimmer potentiometer. It is only used once to trim a single suitable level to correlate with the pinch-off voltage of J_1 , and it is not intended to be adjusted after the primary setting. Basically R_{12} could be built as a user-controlled potentiometer to allow small adjustments on the compression threshold level.

A transient simulation from the Orange Squeezer circuit produces quite different results depending on the frequency of the input signal and the value of resistor R_4 . It is relatively unrealistic to use a sine wave as a test signal in this case, but it gives a clean example of the functionality of the circuit. Figure 6.63 shows a situation where the input signal is a 1 V sine wave, which at the input terminal of the op-amp is as shown in the top part of Figure 6.63. At the beginning of the waveform, the input amplitude at the op-amp is so large that the output is amplified to rails and the signal is clipped. At this point the capacitor C_6 is charged up and the result of the potential rise at the gate of JFET J_1 is quickly seen as the attenuated sine wave at the op-amp input. As the input signal attenuates, this also stabilises the output, which after a while settles to a constant amplitude sine wave.

The voltage levels at the gate and source pins of J_1 are visualised in Figure 6.64. The reference voltage at the source stays constant as it is controlled by the value of R_4 . As the potential at the gate rises higher so that the pinch-off voltage of J_1 is passed, the resistance of the JFET J_1 drops and the effect of this on the input signal is shown in Figure 6.63.

As a conclusion, the user of the Orange Squeezer pedal basically cannot control any of the basic compressor parameters without making some own modifications. The attack and release times can be made adjustable by replacing a few of the resistors with potentiometers. On the other hand, it has been a good decision from the designer of the Orange Squeezer to leave out the possibility to adjust the compression parameters. Compression as an effect is really difficult to master by a novice and it needs several years of experience to know the best settings. It might also be that the pedal has only one specific sweet spot that works best for the waveforms generated by the guitar.

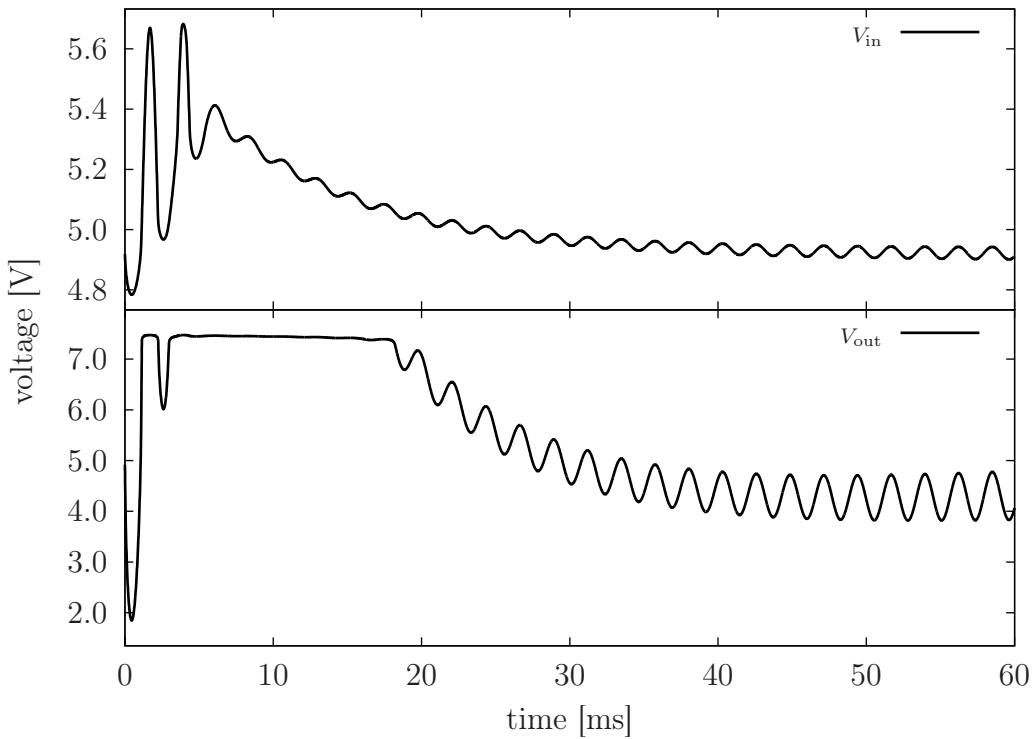


Figure 6.63: A transient response of the actual compressor circuit

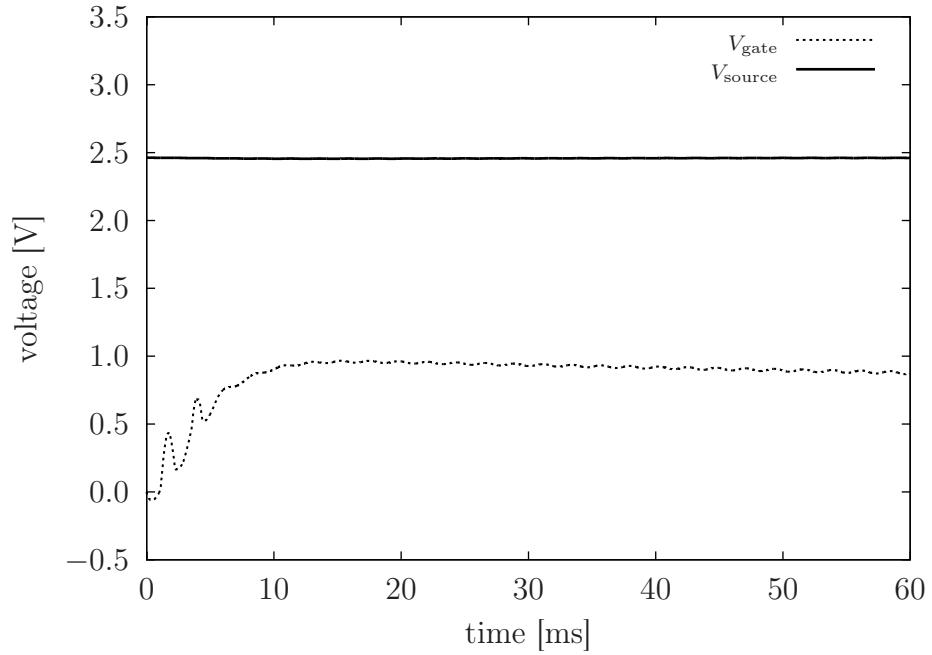


Figure 6.64: Gate voltage versus the reference voltage level at the source

The Orange Squeezer circuit could be improved to also sense the negative amplitude peaks by placing a unity gain inverting operational amplifier directly to the output of the first operational amplifier. This will turn the negative peaks

into positive peaks and the output from the other op-amp can be taken through another diode to charge the capacitor C_6 . In that case, D_2 should be the same model as D_1 , and the resistor R_6 should be moved after the two diodes because otherwise the other diode will bypass the resistor and the charging current from that branch will be different.

The basic technique of using a JFET as a key component in an amplitude limiter circuit is commonly used and it has also been described in a scientific paper [82]. Therefore, the implementation of the Orange Squeezer is a good example of how a limiter circuit is constructed from basic components.

Expansion is the counterpart for compression in the sense that compression decreases the dynamic range of the amplitude waveform, whereas expansion increases the dynamic range. Obviously the threshold for the expander is set close to the noise level, so it is on the opposite side as compared to the compressor.

Expander effects can be identified as upward or downward expanders. Upward expanders start to process the audio when the signal level is over the threshold. The goal is to amplify the low-amplitude sound and increase the signal-to-noise ratio. Downward expanders seek for the parts where the audio drops below the threshold and therefore they try to adjust the (assumed) noise signal to a minimum level. This kind of functionality is also known as *noise gating*. Both approaches actually increase the dynamic range of an audio signal.

In general, expansion is very rarely used as a guitar effect so there is no real need to dig further into this topic. Compressors do not dramatically change the sound of the guitar, so in that sense they cannot be counted as true guitar effect devices. In a broader sense, compressors and expanders belong to the group of general signal processing devices.

6.5 FREQUENCY EFFECTS

Frequency based effects mainly use the properties of filters to alter the frequency content and the phase of an audio signal. A basic graphic equaliser is already a good example of a frequency 'effect'. To make life more exciting, quite often frequency effects are implemented as sweep type of effects where some filter function (= equaliser pattern) is swept through a range of frequencies, boosting and attenuating a set of different frequencies over time. This creates an unique sound that is easily identified as an effect in the frequency domain.

The wah effect pedal is a good example of a guitar effect that boosts a certain frequency range acting as a foot-switch movable equaliser. A guitar player using the wah effect can choose to enhance the higher or lower upper partials of the guitar signal by selecting a constant setting of the wah pedal. A more common way of using the wah pedal is to rhythmically sweep up and down the frequency range to produce a distinct 'wah-wah' sound. The frequency boost in the wah pedal is based on the resonance properties of a basic RLC circuit. By adding an envelope follower to automatically control the frequency sweeping of a regular wah pedal, the result is commonly known as 'auto-wah'. An in-depth analysis of a simple wah effect circuit is given in section 6.6.

Other distinguishable effect types in the frequency category are pitch shifters, which change the frequency of the played notes. While the more advanced pitch shifters allow the user to choose the amount of applied shift, the simpler ones just shift notes up or down in octave multiples. The 'Green Ringer' effect pedal presented in section 6.7 is a good example of an 'octave-up' type of effect.[69]

6.6 THE ORIGINS OF THE WAH-WAH SOUND

Just like the fuzz pedal is considered the king of amplitude effects, the wah effect is one of the oldest and most widely used effect in the series of frequency categorised effects. The first wah pedals were built in the middle of the 1960's carrying the brand name 'Vox Wah-Wah' among the line of other pedals manufactured under the Vox brand. The Vox Wah-Wah circuit appears to be a very simple design with a small amount of components, so it is a very suitable target for further analysis.

Figure 6.65 shows one version of a circuit diagram imitating the original Vox Wah-Wah circuit [83]. Naturally many variations of this circuit exist, but this specimen is considered to be the original one. At least it is the simplest wah circuit, so it is the ideal circuit to learn about the basic electronics that makes it all happen. From the user's perspective the wah effect is relatively simple to use because it has only one potentiometer that the user can adjust to filter out a certain band of frequencies. The practical stompbox (pedal) itself uses an up-down moving lever, which the user needs to foot-kick down and release back up to control the frequency sweep. Therefore, the wah effect is a true stompbox in the very literal meaning of the word.

As a quick introduction to the signal path of the circuit, the underlying transistor

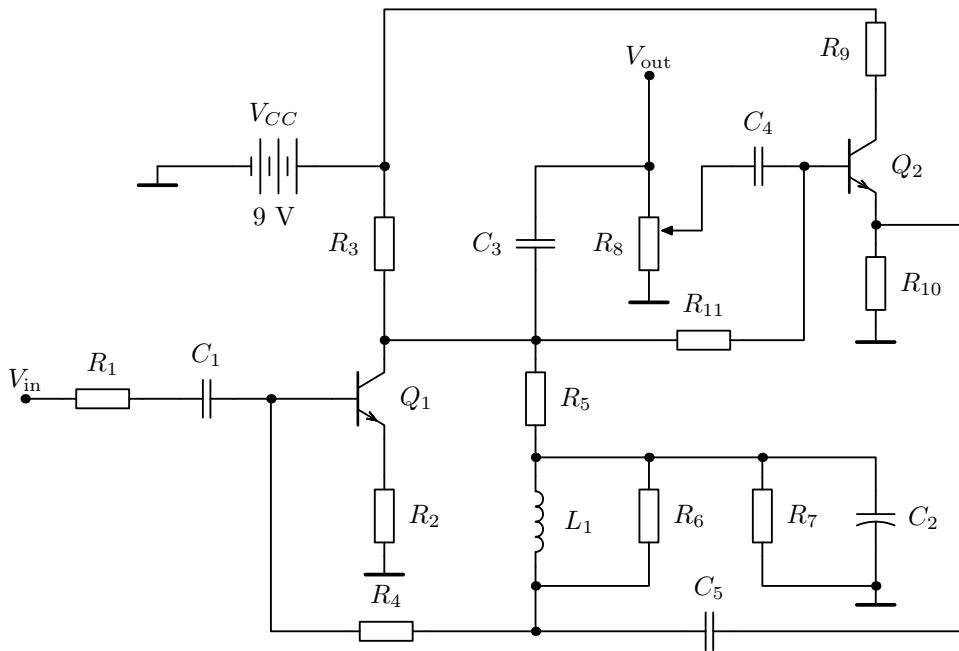


Figure 6.65: A circuit diagram of one of the Vox Wah-Wah models

configuration is very similar to the one used in the fuzz effect. The input signal goes through the first gain stage consisting of a bipolar junction transistor Q_1 and biasing resistors R_1 , R_2 and R_3 . From the signal's point of view, the resistors R_5 and R_{11} are very large and the impedance of C_3 is very small. Therefore, the alternating signal continues from the collector of Q_1 through the path C_3 , R_8 and C_4 to the base of the second transistor Q_2 , from where the signal is fed back to the base of transistor Q_1 . Between the feedback loop connecting transistors Q_1 and Q_2 , there is a basic RLC circuit consisting mainly of the components R_6 , L_1 and C_5 . This RLC resonance circuit is the heart of the wah effect, which is based on moving the resonance frequency of the RLC section by controlling the potentiometer R_8 .

6.6.1 The DC analysis of the wah effect

Since the basic circuit structure is similar to the fuzz effect, the same kind of approach can be used to calculate quiescent values for the transistors of the wah circuit. The most problematic part in this case is to handle the current through resistor R_7 correctly. To get a reasonable relation to the current I_{B1} , one 'extra' voltage equation is needed.

After removing those parts of the circuit that are hidden behind capacitors, the wah effect DC equivalent circuit simplifies a great deal in comparison to the full

circuit 6.65. Figure 6.66 shows the extracted DC schematic of the wah circuit. From the viewpoint of DC signals, the parallel connection of inductor L_1 and

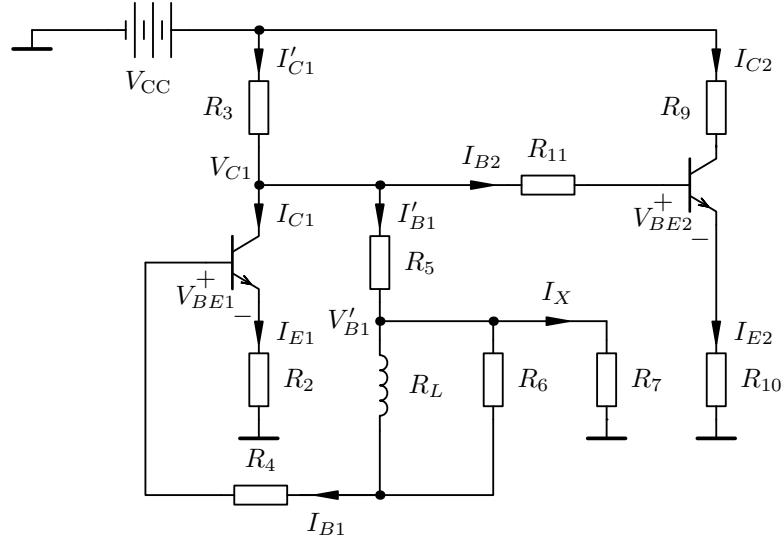


Figure 6.66: The wah effect DC biasing circuit model

resistor R_6 has resistance R_L , which refers to the internal resistance of the inductor. This is because the resistance of the inductor comes from the resistance of the coiled inductor wire, which is typically a few ohms or even less. It would be perfectly acceptable to completely neglect resistances R_L and R_6 in the following derivations.

From the equivalent DC model of the wah circuit, one can write down the ordinary transistor circuit current equations. The analysis results in six current equations,

$$I_{C1} = \beta_{F1} I_{B1} \quad (6.65)$$

$$I_{E1} = (\beta_{F1} + 1) I_{B1} \quad (6.66)$$

$$I_{C2} = \beta_{F2} I_{B2} \quad (6.67)$$

$$I_{E2} = (\beta_{F2} + 1) I_{B2} \quad (6.68)$$

$$I'_{B1} = I_{B1} + I_X \quad (6.69)$$

$$I'_{C1} = I_{C1} + I'_{B1} + I_{B2}, \quad (6.70)$$

and the three following voltage equations can also be written as

$$0 = V_{CC} - I'_{C1} R_3 - I'_{B1} R_5 - I_{B1}(R_L + R_4) - V_{BE1} - I_{E1} R_2 \quad (6.71)$$

$$0 = 0 + I_X R_7 + I'_{B1} R_5 - I_{B2} R_{11} - V_{BE2} - I_{E2} R_{10} \quad (6.72)$$

$$0 = 0 + I_X R_7 - I_{B1}(R_L + R_4) - V_{BE1} - I_{E1} R_2. \quad (6.73)$$

The obvious solution method is to find an expression for base current I_{B1} . After determining the value of I_{B1} , all the other quiescent voltages of the circuit can then be calculated using the value of I_{B1} . Substituting current relations (6.69) and (6.70) into voltage equation (6.71),

$$V_{CC} = I_{C1}R_3 + I_{B1}R_3 + I_XR_3 + I_{B2}R_3 + I_{B1}R_5 + I_XR_5 + I_{B1}(R_L + R_4) + V_{BE1} + I_{E1}R_2,$$

and gathering all terms under similar current variables yields

$$\begin{aligned} V_{CC} = I_{B1} [R_5 + R_L + R_4 + (\beta_{F1} + 1)(R_3 + R_2)] &+ I_X(R_3 + R_5) \\ &+ I_{B2}R_3 + V_{BE1}. \end{aligned} \quad (6.74)$$

Solving I_{B2} from equation (6.72) says that

$$-I_{B2} = \frac{V_{BE2} - I_X(R_7 + R_5) - I_{B1}R_5}{R_{11} + (\beta_{F2} + 1)R_{10}},$$

and substituting this expression into (6.74) brings an equation

$$\begin{aligned} V_{CC} = I_{B1} \left[R_5 + R_L + R_4 + (\beta_{F1} + 1)(R_3 + R_2) + \frac{R_5R_3}{R_{11} + (\beta_{F2} + 1)R_{10}} \right] \\ + I_X(R_3 + R_5) - \frac{V_{BE2} - I_X(R_7 + R_5)}{R_{11} + (\beta_{F2} + 1)R_{10}} R_3 + V_{BE1}. \end{aligned} \quad (6.75)$$

There is still that unknown current I_X giving us trouble. Using equation (6.73),

$$I_X = \frac{I_{B1}(\beta_{F1} + 1)R_2 + I_{B1}(R_L + R_4) + V_{BE1}}{R_7},$$

and this needs to be inserted into equation (6.75). Doing this insertion and solving for I_{B1} gives the final expression in the form

$$I_{B1} = \frac{\text{numerator}}{\text{denominator}}, \quad (6.76)$$

where the numerator is

$$V_{CC} + \frac{V_{BE2}R_3}{R_{11} + (\beta_{F2} + 1)R_{10}} - V_{BE1} - V_{BE1} \frac{R_3 + R_5}{R_7} - \frac{V_{BE1}(R_7 + R_5)R_3}{R_7(R_{11} + (\beta_{F2} + 1)R_{10})},$$

and the denominator is written as

$$\begin{aligned} R_5 + R_L + R_4 + (\beta_{F1} + 1)(R_3 + R_2) + \frac{R_3R_5}{R_{11} + (\beta_{F2} + 1)R_{10}} \\ + \frac{(R_L + R_4 + (\beta_{F1} + 1)R_2)}{R_7} \left[\frac{(R_7 + R_5)R_3}{(R_{11} + (\beta_{F2} + 1)R_{10})} + (R_3 + R_5) \right]. \end{aligned}$$

Another more systematic way to analyse the circuit is to use the circuit nodes instead of the branches as anchors for the analysis. The obvious voltage nodes

of the circuit are V_{C1} and V'_{B1} but those will not give enough information about all the branches of the circuit. In this case it is better to utilise the DC equivalent model of the BJT and add both emitter voltage nodes V_{E1} and V_{E2} to the set.

Based on Kirchhoff's current law, the sum of currents entering a node equals the sum of currents leaving a node. Therefore, the current equations for the voltage nodes V_{C1} , V'_{B1} , V_{E1} and V_{E2} are:

$$I'_{C1} = I_{C1} + I_{B2} + I'_{B1}$$

$$I'_{B1} = I_{B1} + I_X$$

$$I_{E1} = (\beta_{F1} + 1)I_{B1}$$

$$I_{E2} = (\beta_{F2} + 1)I_{B2}$$

The essential currents in the above equations can be expressed with the node voltages as:

$$\begin{aligned} I'_{B1} &= \frac{V_{C1} - V_{B1}}{R_5} & ; & \quad I'_{C1} = \frac{V_{CC} - V_{C1}}{R_3} & ; & \quad I_{B1} = \frac{V'_{B1} - V_{BE1} - V_{E1}}{R_4 + R_L} \\ I_X &= \frac{V'_{B1}}{R_7} & ; & \quad I_{B2} = \frac{V_{C1} - V_{BE2} - V_{E2}}{R_{11}} & ; & \quad I_{E1} = \frac{V_{E1}}{R_2} \end{aligned}$$

and I_{E2} is similar to I_{E1} . Substitutions of the voltage equations into the current equations leads to the following four equations

$$\frac{V_{CC} - V_{C1}}{R_3} = \frac{\beta_{F1}}{\beta_{F1} + 1} \frac{V_{E1}}{R_2} + \frac{V_{C1} - V_{BE2} - V_{E2}}{R_{11}} + \frac{V_{C1} - V'_{B1}}{R_5}$$

$$\frac{V_{C1} - V'_{B1}}{R_5} = \frac{V'_{B1} - V_{BE1} - V_{E1}}{R_4 + R_L} + \frac{V'_{B1}}{R_7}$$

$$\frac{V_{E1}}{R_2} = (\beta_{F1} + 1) \frac{V'_{B1} - V_{BE1} - V_{E1}}{R_4 + R_L}$$

$$\frac{V_{E2}}{R_{10}} = (\beta_{F2} + 1) \frac{V_{C1} - V_{BE2} - V_{E2}}{R_{11}}.$$

These equations can be organised into a matrix equation, from where each of the node voltages can be solved using Cramer's rule. The matrix equation is of the form:

$$\begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & 0 \\ 0 & Y_{32} & Y_{33} & 0 \\ Y_{41} & 0 & 0 & Y_{44} \end{bmatrix} \times \begin{bmatrix} V_{C1} \\ V'_{B1} \\ V_{E1} \\ V_{E2} \end{bmatrix} = \begin{bmatrix} \frac{V_{CC}}{R_3} + \frac{V_{BE2}}{R_{11}} \\ \frac{V_{BE1}}{R_4 + R_L} \\ \frac{V_{BE1}}{R_4 + R_L} \\ \frac{V_{BE2}}{R_{11}} \end{bmatrix}$$

where the non-zero terms in the admittance matrix are

$$\begin{aligned} Y_{11} &= \frac{1}{R_3} + \frac{1}{R_{11}} + \frac{1}{R_5} & Y_{12} = Y_{21} &= -\frac{1}{R_5} \\ Y_{13} &= \frac{\beta_{F1}}{\beta_{F1} + 1} \frac{1}{R_2} & Y_{14} &= -\frac{1}{R_{11}} \\ Y_{22} &= \frac{1}{R_5} + \frac{1}{R_4 + R_L} + \frac{1}{R_7} & Y_{23} &= -\frac{1}{R_4 + R_L} \\ Y_{32} &= \frac{1}{R_4 + R_L} & Y_{33} &= -\frac{1}{R_4 + R_L} - \frac{1}{R_2(\beta_{F1} + 1)} \\ Y_{41} &= \frac{1}{R_{11}} & Y_{44} &= -\frac{1}{R_{11}} - \frac{1}{R_{10}(\beta_{F2} + 1)} \end{aligned}$$

Based on this analysis, all the other biasing voltages and currents can be solved using the magnitudes of the node voltages.

Based on the experience so far, the analytical equations easily become too long to handle in pen-and-paper style calculations. If one wants to continue to also solve the frequency response equations numerically, then it is necessary to solve the quiescent conditions to find out the value of r_π at the DC operating point. The symbolic equations can be written to a mathematical software spreadsheet and solved there numerically, but still the SPICE simulation software offers a much faster way to obtain the biasing conditions and the frequency response. On the other hand, extensive use of SPICE might not give that much insight into the functionality of the analysed circuit.

Anyway, now that the equations for the DC analysis are derived, it is time to move on to conduct the AC analysis to find out the frequency response of the system. The equations presented in this section will be used for calculating the actual biasing voltages in section 6.6.6, where the results are compared to values obtained by simulations and prototype measurements.

6.6.2 The AC analysis of the wah effect

To be able to derive the transfer function of the wah circuit, a small-signal model needs to be constructed. The small-signal model is derived directly from the original schematic 6.65 of the wah effect and it is drawn in Figure 6.67. The transistors in the small-signal model are depicted by equivalent circuits as explained in section 1.3.10. All DC voltage sources are grounded because only alternating signals are analysed. External alternating voltage sources have been converted into current sources so that the nodal method can be used to analyse the circuit.

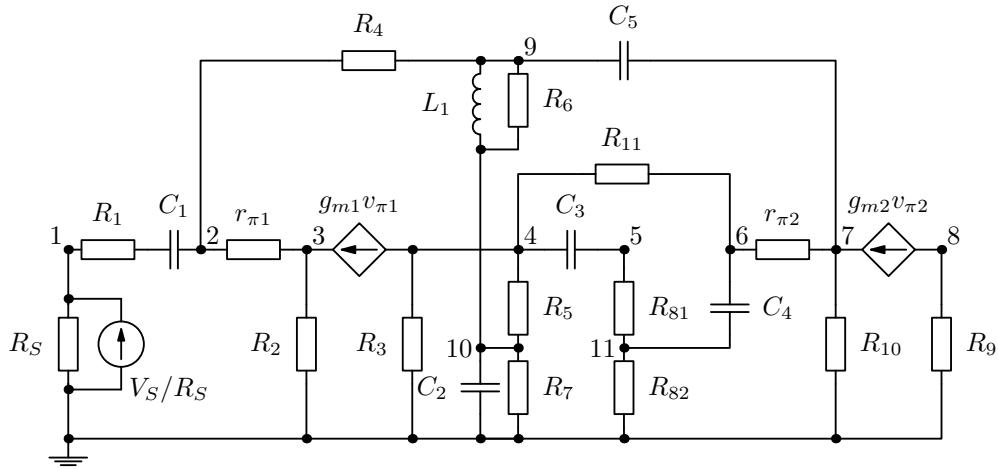


Figure 6.67: The AC equivalent circuit model of the wah pedal

The leftmost current source represents the signal source that drives the signal to the effect device. The resistance R_S depicts the internal resistance of the signal source. The possibility for the signal source to have capacitive or inductive reactance is not considered here, but if test signals of a single frequency are used to simulate the circuit, the frequency dependent impedance of the source is not that meaningful. The value of the transistor's internal resistance r_π is evaluated analytically by using equations

$$g_m = \frac{|I_{CQ}|}{V_T} \quad \text{and} \quad r_\pi = \frac{\beta_F}{g_m},$$

where the thermal coefficient $V_T \approx 25$ millivolts and the quiescent collector current I_{CQ} is obtained from the biasing calculations.

As explained in section 1.2.6 covering the Laplace transform, for sinusoidal test signals one does not need to care about the Laplace transform of the sinusoidal source itself if one is only interested in the frequency response.

The mathematical equations based on the small-signal model of the wah circuit can be written directly in matrix format as $Y \times V = I$. Because there are 11 individual voltage nodes in the small-signal model, the admittance matrix will become a 11x11 square matrix,

$$Y = \begin{bmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 & 0 & 0 & 0 & -Y_{29} & 0 & 0 \\ 0 & -Y_{32} & Y_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Y_{44} & -Y_{45} & -Y_{46} & 0 & 0 & 0 & -Y_{410} & 0 \\ 0 & 0 & 0 & -Y_{54} & Y_{55} & 0 & 0 & 0 & 0 & 0 & -Y_{511} \\ 0 & 0 & 0 & -Y_{64} & 0 & Y_{66} & -Y_{67} & 0 & 0 & 0 & -Y_{611} \\ 0 & 0 & 0 & 0 & 0 & -Y_{76} & Y_{77} & 0 & -Y_{79} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & Y_{88} & 0 & 0 & 0 \\ 0 & -Y_{92} & 0 & 0 & 0 & 0 & -Y_{97} & 0 & Y_{99} & -Y_{910} & 0 \\ 0 & 0 & 0 & -Y_{104} & 0 & 0 & 0 & 0 & -Y_{109} & Y_{1010} & 0 \\ 0 & 0 & 0 & 0 & -Y_{115} & -Y_{116} & 0 & 0 & 0 & 0 & Y_{1111} \end{bmatrix}$$

and the related voltage and current vectors are

$$V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \\ V_7 \\ V_8 \\ V_9 \\ V_{10} \\ V_{11} \end{bmatrix}; \quad I = \begin{bmatrix} \frac{V_S}{R_S} \\ 0 \\ g_{m1}(V_2 - V_3) \\ -g_{m1}(V_2 - V_3) \\ 0 \\ 0 \\ g_{m2}(V_6 - V_7) \\ -g_{m2}(V_6 - V_7) \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The matrix equation is so large that the nonzero elements are only indicated by the admittance symbol Y with subscripts referring to the element in the matrix as $Y_{\text{row},\text{column}}$. All the matrix elements derived from the small-signal model are listed below and they are intended to be substituted into the 11x11 admittance matrix shown above.

$$\begin{aligned}
Y_{11} &= \frac{1}{R_S} + \frac{j\omega C_1}{1 + j\omega C_1 R_1} & Y_{12} = Y_{21} &= \frac{j\omega C_1}{1 + j\omega C_1 R_1} \\
Y_{22} &= \frac{1}{R_4} + \frac{1}{r_{\pi 1}} + \frac{j\omega C_1}{1 + j\omega C_1 R_1} & Y_{23} = Y_{32} &= \frac{1}{r_{\pi 1}} \\
Y_{29} &= Y_{92} = \frac{1}{R_4} & Y_{33} &= \frac{1}{r_{\pi 1}} + \frac{1}{R_2} \\
Y_{44} &= \frac{1}{R_3} + \frac{1}{R_5} + \frac{1}{R_{11}} + j\omega C_3 & Y_{45} = Y_{54} &= j\omega C_3 \\
Y_{46} &= Y_{64} = \frac{1}{R_{11}} & Y_{410} = Y_{104} &= \frac{1}{R_5} \\
Y_{55} &= \frac{1}{R_{81}} + j\omega C_3 & Y_{511} = Y_{115} &= \frac{1}{R_{81}} \\
Y_{66} &= \frac{1}{R_{11}} + \frac{1}{r_{\pi 2}} + j\omega C_4 & Y_{67} = Y_{76} &= \frac{1}{r_{\pi 2}} \\
Y_{611} &= Y_{116} = j\omega C_4 & Y_{77} &= \frac{1}{r_{\pi 2}} + \frac{1}{R_{10}} + j\omega C_5 \\
Y_{79} &= Y_{97} = j\omega C_5 & Y_{88} &= \frac{1}{R_9} \\
Y_{99} &= \frac{1}{R_4} + \frac{1}{R_6} + \frac{1}{j\omega L_1} + j\omega C_5 & Y_{910} = Y_{109} &= \frac{1}{R_6} + \frac{1}{j\omega L_1} \\
Y_{1010} &= \frac{1}{R_5} + \frac{1}{R_6} + \frac{1}{R_7} + \frac{1}{j\omega L_1} + j\omega C_2 & Y_{1111} &= \frac{1}{R_{82}} + \frac{1}{R_{81}} + j\omega C_4
\end{aligned}$$

To be able to solve the node voltages, the transconductance terms from the right side of this matrix equation need to be moved to the admittance matrix on the left side. The transconductance term g_{m1} in the current vector on rows 3 and 4 is multiplying voltages V_2 and V_3 . Therefore, g_{m1} can be moved to the admittance matrix on rows 3 and 4 to columns 2 and 3 which multiply voltages V_2 and V_3 in the voltage vector. The move of g_{m1} leads to the modified admittance matrix elements

$$\begin{aligned}
Y'_{32} &= -\frac{g_{m1}r_{\pi 1} + 1}{r_{\pi 1}} & Y'_{33} &= \frac{g_{m1}r_{\pi 1} + 1}{r_{\pi 1}} + \frac{1}{R_2} \\
Y'_{42} &= \frac{g_{m1}r_{\pi 1}}{r_{\pi 1}} & Y'_{43} &= -\frac{g_{m1}r_{\pi 1}}{r_{\pi 1}}.
\end{aligned}$$

A similar move for g_{m2} on rows 7 and 8 of the current vector results in four elements

$$\begin{aligned}
Y'_{76} &= -\frac{g_{m2}r_{\pi 2} + 1}{r_{\pi 2}} & Y'_{77} &= \frac{g_{m2}r_{\pi 2} + 1}{r_{\pi 2}} + \frac{1}{R_{10}} + j\omega C_5 \\
Y'_{86} &= \frac{g_{m2}r_{\pi 2}}{r_{\pi 2}} & Y'_{87} &= -\frac{g_{m2}r_{\pi 2}}{r_{\pi 2}}.
\end{aligned}$$

In each of the above equations, $g_{m1}r_{\pi 1} = \beta_{F1}$ and $g_{m2}r_{\pi 2} = \beta_{F2}$. The output voltage is taken from the voltage node 5 as defined in Figure 6.67. The use of Cramer's rule is already explained in previous sections, and due to the huge size of the admittance matrix, the determinant division is not written here. The

explicit expression of the transfer function would also be so lengthy that there is absolutely no point to solve it by hand. To solve the output node voltage V_5 , the determinant division following from Cramer's rule can be evaluated numerically using mathematical software. An example of using Octave to evaluate transfer functions is given in Appendix C.

6.6.3 Side effect: RLC circuit resonance

The heart of the wah effect is a basic RLC circuit that creates a resonance phenomenon to boost a certain range of frequencies. As the location of the resonance in the frequency axis is swept from one frequency range to the other and back again, it creates the sound typical to a wah pedal. From another point of view, the wah effect can also be considered a moving band-pass filter having a centre frequency that is changed with a potentiometer. The filtering effect is achieved by the resonance properties of a RLC circuit.

To back up the introduction to the basic properties of RLC circuits, Figure 6.68 shows the series- and parallel-connected RLC circuits. Using the notations pro-

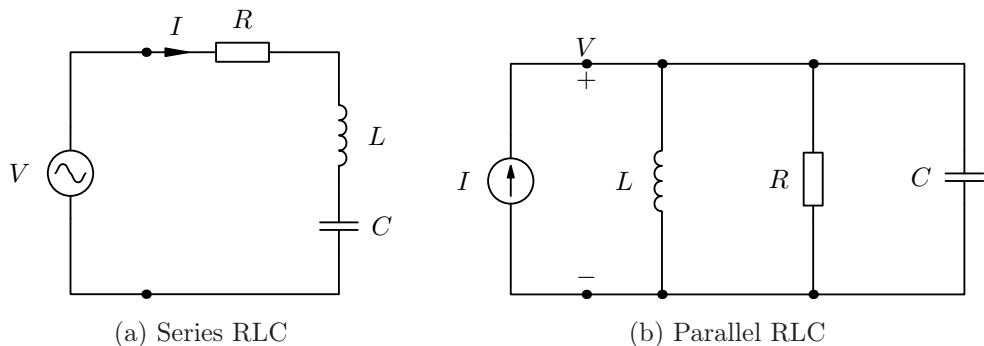


Figure 6.68: Series and parallel RLC networks

vided in Figure 6.68a, the series-connected RLC circuit is represented by the differential equation

$$L \frac{dI}{dt} + RI + \frac{1}{C} \int I dt = V.$$

This differential equation is actually just a voltage equation that follows from the basic Kirchhoff's voltage law. The current loop mesh analysis method states that the sum of all voltage drops within a loop are equal to the source voltage connected to the loop. After applying the Laplace transform, the differential equation is written

$$\left(sL + R + \frac{1}{sC} \right) I = V, \quad (6.77)$$

where the three terms inside the parentheses represent impedances from each individual component.

If the parallel-connected RLC circuit is analysed, it is obvious to choose the nodal analysis method relating to Kirchhoff's current law. The nodal analysis method uses admittances instead of impedances because the differential equation describing the circuit is written as

$$C \frac{dV}{dt} + \frac{1}{R} V + \frac{1}{L} \int V \, dt = I.$$

The Laplace transformed equation of the parallel circuit reads

$$\left(sC + \frac{1}{R} + \frac{1}{sL} \right) V = I, \quad (6.78)$$

where the three terms inside the parentheses represent admittances from each individual component.

To identify the resonance frequency, the impedance and admittance properties of RLC circuits need to be analysed. From the equations using the Laplace variable s , the impedance of the series circuit is $Z_s = R + sL + \frac{1}{sC}$. In the special case of sinusoidal input signals, $s = j\omega$ so that

$$Z_s = R + j \left(\omega L - \frac{1}{\omega C} \right).$$

Clearly the impedance needs to be minimised in order to get a maximum current flowing in the circuit. The maximum state of current flow is reached when the imaginary part of the impedance expression vanishes. Therefore, the resonance frequency can be determined from equation

$$\omega L = \frac{1}{\omega C} \quad \Rightarrow \quad \omega = \frac{1}{\sqrt{LC}}.$$

Because the angular frequency ω equals $2\pi f$, the actual resonance frequency of the series circuit is

$$f_s = \frac{1}{2\pi\sqrt{LC}}.$$

The frequency response of resonance circuits is commonly analysed by plotting the power magnitude of the circuit on a decibel scale. The resonance properties are often evaluated by measuring the width of the resonance area between -3 dB frequencies around the resonance frequency ω_0 . The -3 dB point in a frequency response graph corresponds to a frequency where the signal *power* is half of the maximum value at the resonance peak. Around the resonance peak, the

bandwidth $\delta\omega$ is measured as the difference $\omega_H - \omega_L$, where ω_H is the higher and ω_L the lower -3 dB frequency. The quotient

$$\frac{\delta\omega}{\omega_0} = \frac{1}{Q},$$

defines a quantity Q named as the *quality factor*. Without further derivation, the quality factor Q that defines the sharpness of the resonance peak in the series resonance circuit is defined as

$$Q_s = \frac{1}{R} \sqrt{\frac{L}{C}}.$$

In practical applications, resistance is the major contributor to the sharpness. The smaller the resistance is, the higher and better Q_s is.

The reasoning for finding the resonance frequency is similar in the case of the parallel circuit, except that instead of minimising impedance, one needs to minimise admittance from the equation

$$Y_p = \frac{1}{R} + j \left(\omega C - \frac{1}{\omega L} \right). \quad (6.79)$$

However, despite the slightly different approach, in case of ideal components this leads to the same resonance frequency term

$$\omega C = \frac{1}{\omega L} \quad \Rightarrow \quad \omega = \frac{1}{\sqrt{LC}},$$

$$f_p = \frac{1}{2\pi\sqrt{LC}}.$$

The quality factor for the parallel circuit is the inverse of the series circuit, namely

$$Q_p = R \sqrt{\frac{C}{L}}.$$

A lot of matters are still unclear. Is the RLC configuration in the wah circuit series or parallel? And how is it possible to change the resonance frequency of an RLC circuit with one potentiometer since the resonance frequency does not depend on resistance at all? Let Mr. Miller answer all of these questions.

6.6.4 Side effect: the Miller theorem

The *Miller effect*, a.k.a. the *Miller theorem*, states that any impedance connected between voltage nodes V_i and V_j in a circuit can be reduced to two separate impedances between nodes $V_i - V_N$ and $V_j - V_N$, where V_N represents the ground reference of the circuit. The Miller-transformed impedances are scaled with a

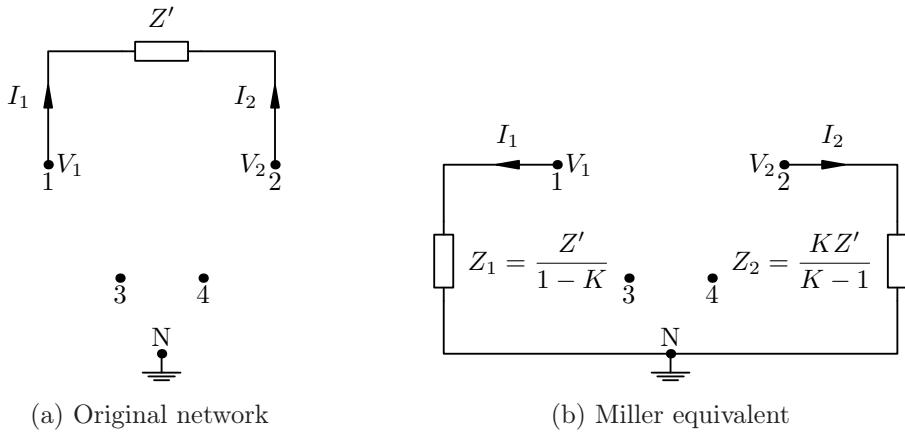


Figure 6.69: The Miller theorem explained

fractional potential difference between nodes V_i and V_j . To make it more clear, Figure 6.69 is drawn to back up the explanation of the Miller theorem.

The circuit of Figure 6.69 has N nodes with node voltages $V_1, V_2, V_3, \dots, V_N$, where the ground node $V_N = 0$. Nodes 1 and 2 are interconnected with impedance Z' . The expression for current I_1 can be tweaked as

$$I_1 = \frac{V_1 - V_2}{Z'} = \frac{V_1(1 - K)}{Z'} = \frac{V_1}{Z'} = \frac{V_1}{Z_1}, \quad (6.80)$$

where $K = \frac{V_2}{V_1}$. For current I_2 one can apply similar reasoning:

$$I_2 = \frac{V_2 - V_1}{Z'} = \frac{V_2 \left(1 - \frac{1}{K}\right)}{Z'} = \frac{V_2}{KZ'} = \frac{V_2}{Z_2}. \quad (6.81)$$

The current equations shown above also define the new impedance factors that scale the original impedance Z' with the factor K , which is the voltage ratio between the two nodes.

Since the nodal equations derived from circuits 6.69a and 6.69b are identical, the two networks are equivalent. The essential result of the Miller theorem is that the equivalent impedances Z_1 and Z_2 separated between the two nodes are scaled by the voltage ratio of the two nodes. When any impedance is used for connecting the input and output nodes of an amplifier circuit, the impedance changes dynamically with the gain of the amplifier element.

This is what happens in the wah circuit. The value of capacitance C_5 is changed by controlling the base pin voltage of Q_2 with potentiometer R_8 . Therefore, the

voltage ratio $\frac{V_7}{V_9}$ changes and directly affects the impedance of the capacitance C_5 , which in the Miller equivalent circuit is connected parallel to the inductor L_1 and the resistor R_6 . This makes the RLC configuration of the wah circuit to be in parallel, at least approximately, since there are other passive components in direct contact with the RLC section. Figure 6.70 indicates the Miller equivalent small-signal model of the wah circuit.

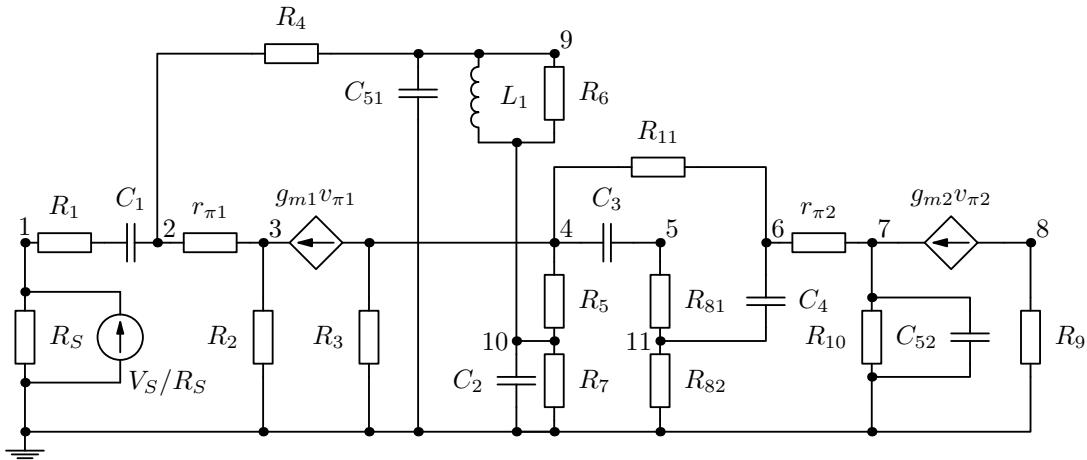


Figure 6.70: The wah small-signal model using the Miller equivalent capacitor

Changes in the resistance of the potentiometer R_8 reflect signal level changes to node 6. Node 6, on the other hand, defines the voltage $v_{\pi 2}$, which controls the current gain of the transistor. As the current at node 7 changes, the voltage over resistor R_{10} changes accordingly. This directly affects the voltage ratio $K = \frac{V_7}{V_9}$, which controls the impedance of capacitors C_{51} and C_{52} in the Miller equivalent circuit. Based on equations (6.80) and (6.81), the formulae to evaluate the values for the Miller capacitances are

$$C_{51} = C_5(1 - K) \quad \text{and} \quad C_{52} = C_5 \frac{K - 1}{K},$$

where in practise the gain factor K is evaluated from the original small-signal model. Therefore, the Miller capacitances will be complex-valued. Sometimes a good approximation of real capacitance values are obtained by taking the absolute value of K , but this should not be taken as granted.

Based on the analysis of basic RLC circuits and the Miller theorem, the resonance effects in the wah circuit could be approximated by evaluating the impedance curves of the parallel RLC circuit. The impedance of the parallel circuit is evaluated simply as the inverse of the admittance expression given in equation (6.79).

Figure 6.71 visualises a few of the impedance curves evaluated using the component values $L = 0.5 \text{ H}$ and $R = 33 \text{ k}\Omega$, which are taken directly from the wah circuit. The capacitance is varied around the nominal value of C_5 , which is also indicated in the wah circuit.

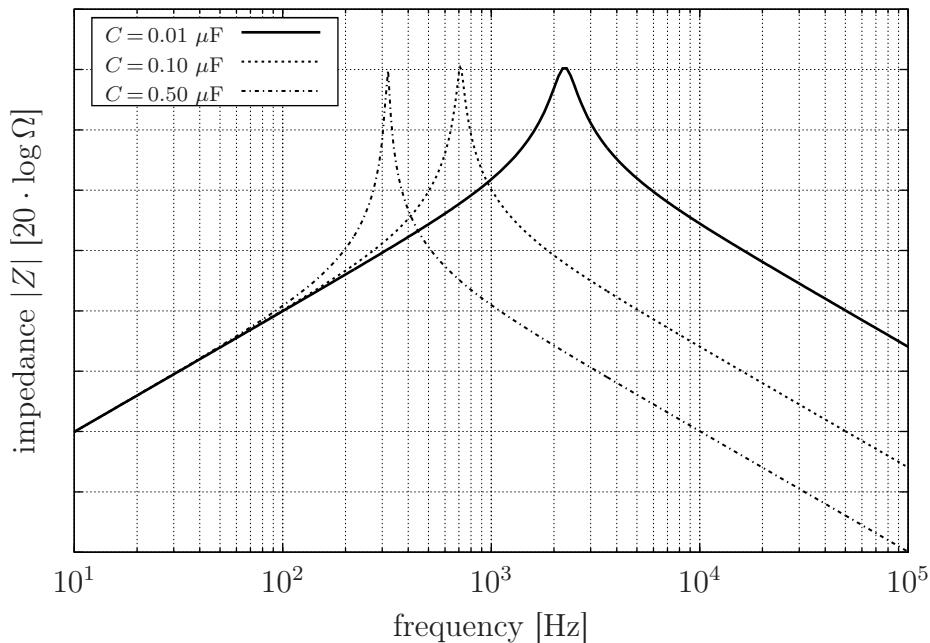


Figure 6.71: Impedance curves of a parallel RLC circuit

Because the gain of the amplifier stage Q_2 changes with frequency, the Miller capacitance values also change with frequency. This means that in this kind of resonance setup it is at least theoretically possible to get even two resonance peaks with a fixed resistance value of R_8 .

6.6.5 A SPICE model for the wah circuit

To prepare a netlist for SPICE simulations, it is better to use a dedicated schematic editor to draw the circuit and to generate the netlist automatically. For this reason, the schematic shown in Figure 6.72 was drawn using the `gschem` schematic editor. The schematic includes all the components assigned with their actual values. The potentiometer R_8 is split into two separate resistors R_{81} and R_{82} , so that the voltage divider functionality can be simulated by changing simultaneously the values of the two resistors.

In addition to real resistor and capacitor values, the transistor gains and saturation currents need to be defined as device models in `gschem` to depict the true

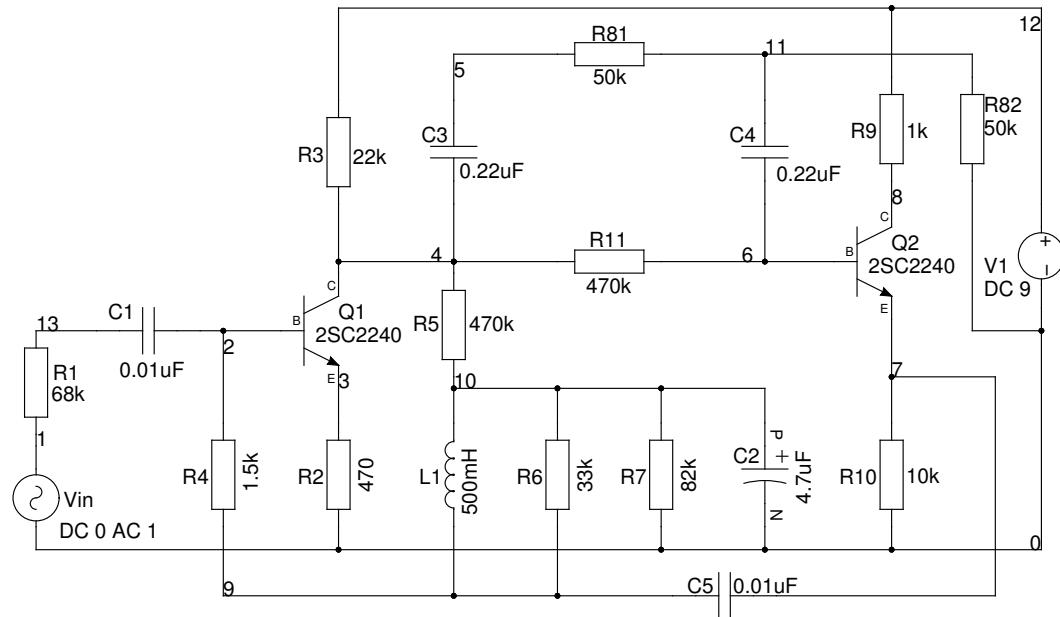


Figure 6.72: The wah effect schematic drawn with gschem

transistor behaviour. When the transistor models are included in the schematic editor, the automatically created netlist file will have the correct model included. This time no fancy transistor parameter evaluation is needed, since in the wah circuitry one can use basic *nPN* transistors with SPICE models provided by manufacturers.

For the wah effect, it is better to choose transistors with a current gain factor of about 200 or over because otherwise the variations in the gain controlled capacitance might be too small. This is why the 2SC2240 transistors were chosen for this construction. This transistor has a β_F of approximately 220 and it is especially designed for low-noise audio applications. The equivalent transistor model corresponding to 2SC2240 is BC550, which can also be used in this circuit. Usually the BC550 are sold a bit cheaper.

Unfortunately an exact SPICE model for 2SC2240 could not be found from the datasheet of the manufacturer. Luckily, the transistor parameters are not critical in this application, so it is enough to just define the saturation current and the average β_F in the SPICE transistor model. The other parameters can be left to default SPICE values.

When using the gEDA design tools collection in Linux, the netlist can be created from the schematic of Figure 6.72 by using the command

```
gnetlist -g spice-sdb -o wah_net.net wah.sch
```

where the output file named as `wah_net.net` is created from the schematic file `wah.sch`. The automatically created netlist file is shown here for convenience.

```
***** Begin SPICE netlist of main design *****
R82 11 0 50k
Vin 0 1 DC 0 AC 1 SIN(0 0.01 440)
R81 11 5 50k
C4 6 11 0.22uF
C3 4 5 0.22uF
R1 1 13 68k
C2 10 0 4.7uF
C5 9 7 0.01uF
C1 13 2 0.01uF
R6 9 10 33k
L1 9 10 500mH
V1 12 0 DC 9
R7 0 10 82k
Q2 8 6 7 NPN2
.MODEL NPN2 NPN (Is=1.0e-14 Bf=220)
Q1 4 2 3 NPN1
.MODEL NPN1 NPN (Is=1.0e-14 Bf=220)
R10 0 7 10k
R9 8 12 1k
R11 4 6 470k
R5 10 4 470k
R3 4 12 22k
R2 0 3 470
R4 9 2 1.5k
.end
```

Obviously one is interested in the frequency response of the circuit with different values of potentiometer R_8 . The transient response is also interesting in the sense that one sees how cleanly does the circuit reproduce a sinusoidal test signal.

To perform the needed analysis with the help of a SPICE engine, the adequate control commands are added to the beginning of the netlist file. The frequency response simulation is run three times for different settings of the potentiometer R_8 which adjusts the centre frequency of the wah-filter. The operating point analysis is also added to see how the transistors are biased in the static DC configuration. The simulation control commands are now set as

```
.control
ac dec 90 10 100K
alter R81 1k
alter R82 99k
ac dec 90 10 100K
alter R81 99k
alter R82 1k
ac dec 90 10 100K
set filetype=ascii
write wahdata.txt db(ac1.v(5)) db(ac2.v(5)) db(ac3.v(5))
gnuplot wah_ac db(ac1.v(5)) db(ac2.v(5)) db(ac3.v(5))
alter R81 50k
alter R82 50k
tran 1us 30ms
gnuplot wah_tran tran3.v(5) tran3.v(1)
```

```
.endc
.OP
```

and the results are requested to be saved to a gnuplot formatted data file. Eventually, the simulation is run using the command

```
ngspice -b wah_net.net
```

which prints the operating point information to the computer screen and creates a data file `wah_tran.data` to be plotted later with gnuplot.

6.6.6 Results of simulations, calculations and measurements

Regarding the DC analysis of section 6.6.1, the component values indicated in the schematic 6.72 were inserted into the equations. The numerical results were obtained for all transistor pin voltages. The internal potential difference V_{BE} of the 2SC2240 transistors in the wah circuit was estimated to be approximately 0.6 V, and this value was also used in the calculations.

For the DC simulation part using SPICE, the netlist from the circuit of Figure 6.72 was loaded into ngspice and analysed with the operating point (.OP) command, which was written inside the netlist file. The simulation results for DC bias voltages were read from the output listing provided by ngspice.

To be able to compare the theoretical results to prototype measurements, the wah circuit using 2SC2240 transistors and 10 % precision resistors and capacitors was built to a small solderless breadboard, a.k.a. a plugboard, where the components can be replaced easily just by plugging them onto the board. A regular multimeter was used for measuring the transistors' quiescent voltages. The 500 mH inductor is relatively large-valued, so understandably it was difficult to find one directly from the shelf. The solution for the inductor problem was to buy three 150 mH inductors and connect them in series (still lacking that 50 mH). The only thing that needs to be taken care of in this kind of solution is that the physical distance between the inductors on the plugboard has to be as large as possible to avoid unnecessary mutual inductance between the inductors.

The simulated, calculated and measured quiescent values of the transistors Q_1 and Q_2 of the wah circuit are gathered into Table 6.4 for comparison purposes. As it almost always seems to be the case, the results obtained analytically and experimentally are generally in good agreement, but there is always that one black sheep in the family of results. This time it is the base voltage of Q_2 , which differs

quite significantly from the analytical results. Actually, this measured voltage seems to fight against the laws of nature, since the base voltage is lower than the emitter voltage. This definitely should not be the case when *npn* transistors are used. Anyhow, this result is accepted and documented, because no real bugs were found from the connections in the prototype circuit. The explanation for this phenomenon is not known.

Table 6.4: Wah circuit transistor quiescent values

value	simulated	calculated	measured
V_{B1}	0.69	0.60	0.68
V_{C1}	5.01	5.15	5.00
V_{E1}	0.08	0.08	0.08
V_{B2}	4.24	4.39	3.05
V_{C2}	8.64	8.64	8.66
V_{E2}	3.62	3.79	3.61

Next it was the time to examine the frequency response of the circuit. Firstly, a comparison between the manual calculations and the SPICE simulations was made by plotting the response curves into the same graph. Figure 6.73 can be drawn from the results of the simulations and the calculations.

The simulations and the calculations of the frequency response are this time in perfect agreement. This is not surprising because a very basic SPICE model was used for the transistors and that leads to the expectation of similar results. According to the simulations, the resistance value of the potentiometer R_8 shifts the resonance frequency of the circuit between 400 Hz and 2000 Hz. The resonance peaks at low frequencies are sharper than at the higher frequencies. The potentiometer should be chosen as a logarithmic audio taper, since then the frequency shift would occur linearly with the movement of the potentiometer. It was a huge surprise to notice how similar these frequency response curves are to the impedance curves of the parallel RLC circuit in Figure 6.71. It seems that the wah circuit is actually working quite close to an ideal parallel RLC circuit. This is certainly not evident when looking at the wah schematic.

As a bonus from the SPICE simulations, one can get the time dependent signal waveform for a simple test signal as the basic sine wave. The results of this transient simulation are shown in Figure 6.74.

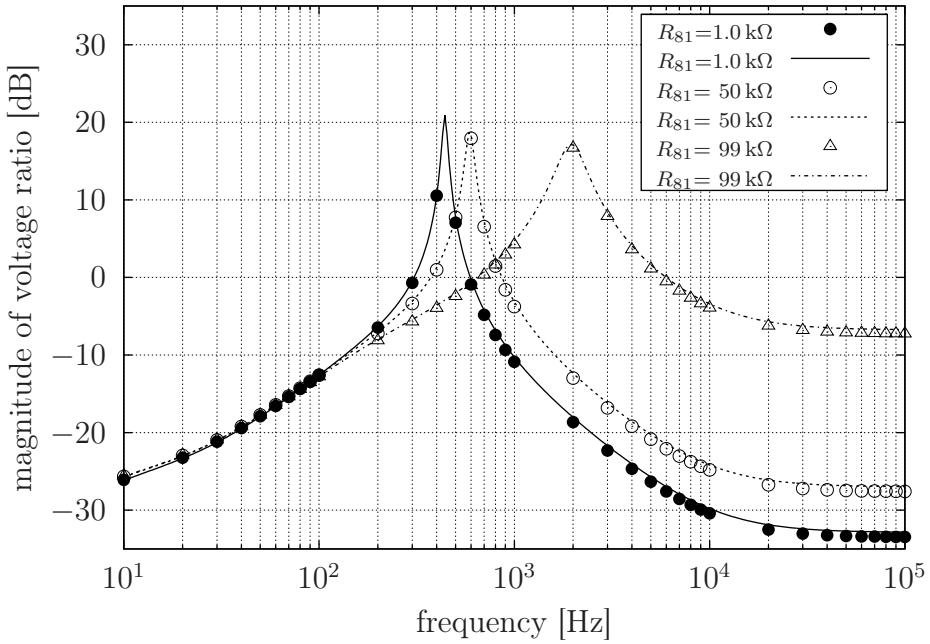


Figure 6.73: Comparing the results of the simulations and manual calculations

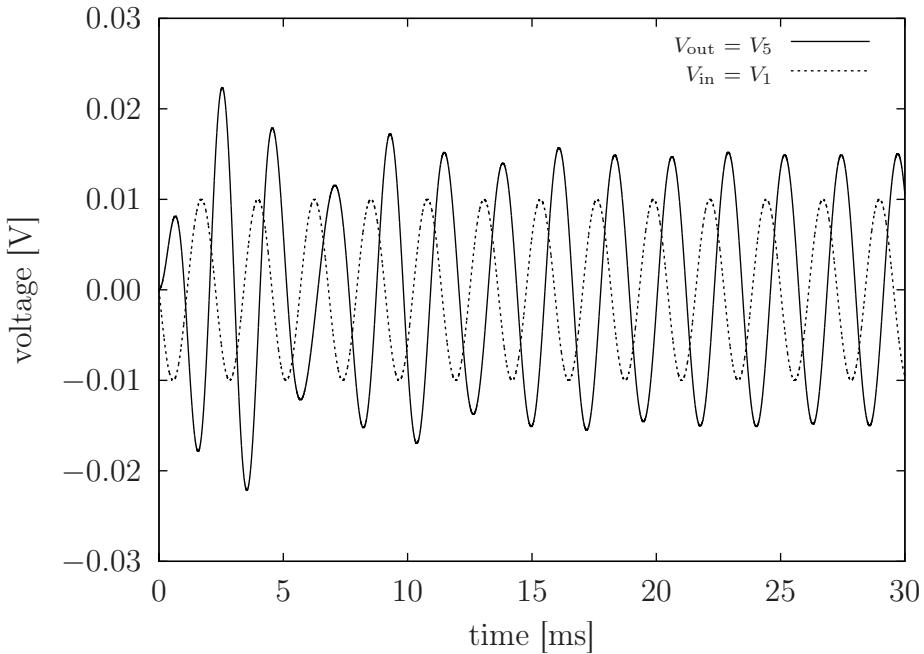


Figure 6.74: A transient simulation of a sine wave input to the wah circuit

The transient simulation is run using some random frequency and random potentiometer setting because the initial idea was just to see whether the wah circuit distorts the input signal in any way. These simulated waveforms can be compared directly to the measured waveforms, but unfortunately the functionality of the effect cannot be sufficiently verified in the time domain.

Figures 6.75 and 6.76 present some selected oscilloscope traces measured from the prototype circuit. A computer sound card was used as the function generator to feed in a relatively large amplitude sine wave to see how easily the circuit starts to distort. The locations of the resonance peaks were approximately determined by playing a logarithmic sine sweep from 50 Hz to 4000 Hz and checking from the oscilloscope screen at what frequency the maximum amplitude output occurs.

Figure 6.75 shows the input and output waveforms when the wah potentiometer is at minimum position. In this case the highest amplitude in the output was measured at 375 Hz, and the waveform in the figure is a snapshot taken at this frequency.

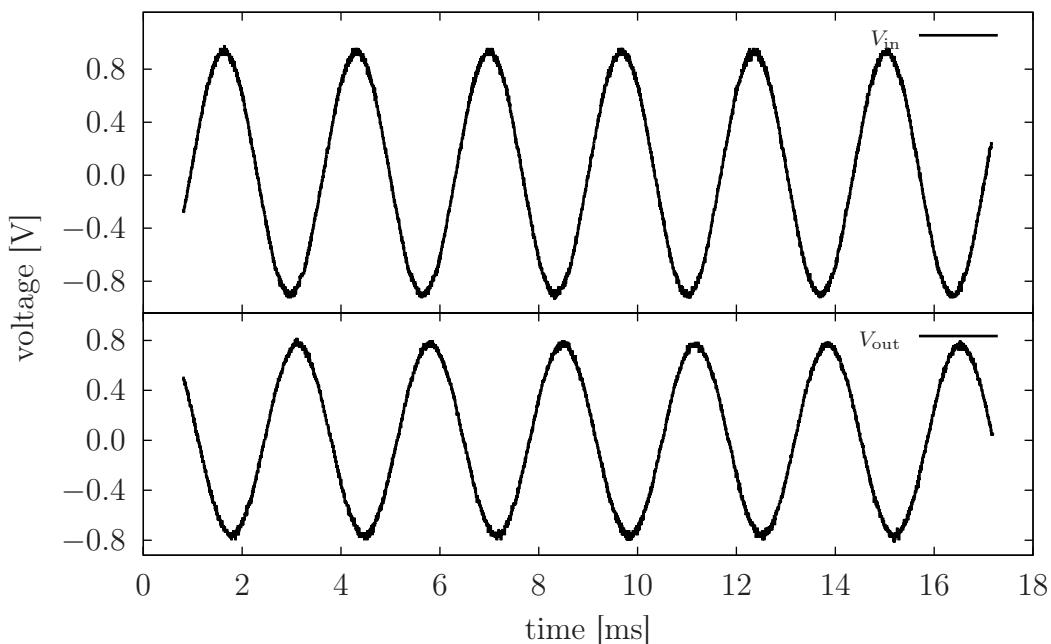


Figure 6.75: The wah effect potentiometer at minimum

From Figure 6.75 it is easy to see that the gain of the circuit is almost exactly unity, although the simulated gain at this setup should have been about 20 dB. This indicates that the prototype build was not very successful, but still has the lowest resonance peak at the correct frequency. It seems as if the peak of the prototype filter is not nearly as sharp as the simulations indicated. The only nice thing is that the minimum resonance frequency is almost exactly the same as the simulations predicted.

As a reference, Figure 6.76 shows the input and output waveforms when the

wah potentiometer is at maximum position. In this case the highest amplitude of the output signal was measured at 2475 Hz, and the waveform in the figure is a snapshot taken at this frequency.

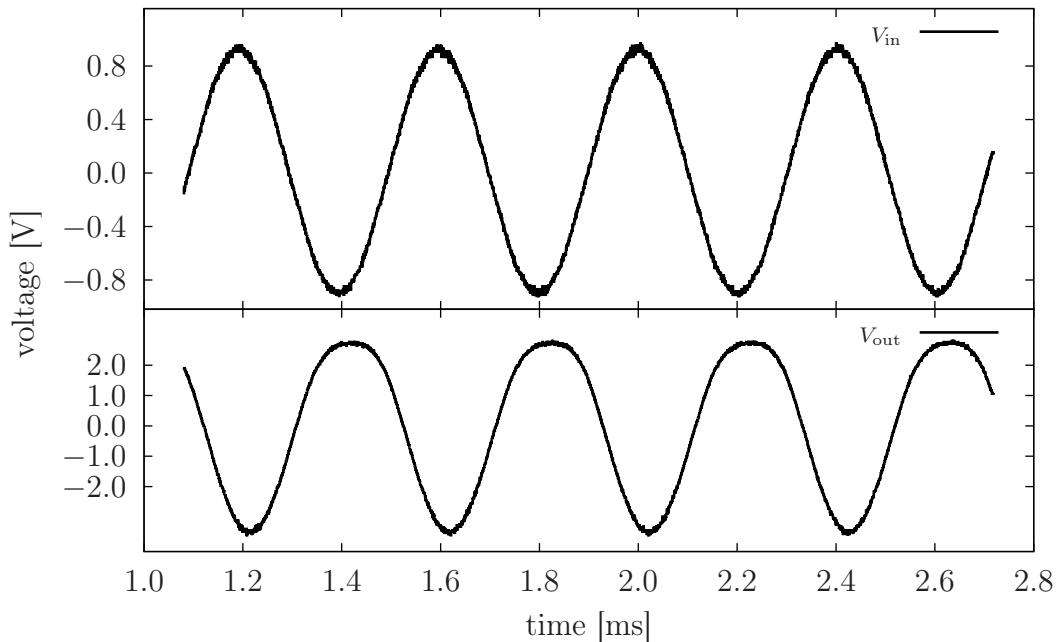


Figure 6.76: The wah effect potentiometer at maximum

The resonance peak is found at the correct frequency and now the circuit gives a larger gain, but still the gain is smaller compared to what was expected after the simulations. Also the gain at the higher frequencies should have been a little bit smaller than at the lower frequencies. These observations indicate that the prototype build did not succeed as expected.

Due to a lack of decent measuring equipment, a basic computer sound card was used for recording the full range frequency response of the wah prototype circuit. A sound card was temporarily used as the measurement instrument because the available oscilloscope did not have enough memory to record the full frequency range. The sound card was also used as a signal generator, from where the sine sweep was played as an audio file. The sine sweep was a constant amplitude logarithmic frequency sweep from 50 Hz to 4000 Hz, just to cover the expected tuning range of the wah prototype.

Figure 6.77 shows the results of the sweep 'measurement'. The upper waveform is recorded with the wah potentiometer at minimum, and the lower waveform is with the wah potentiometer at maximum. Unfortunately, the sound card had

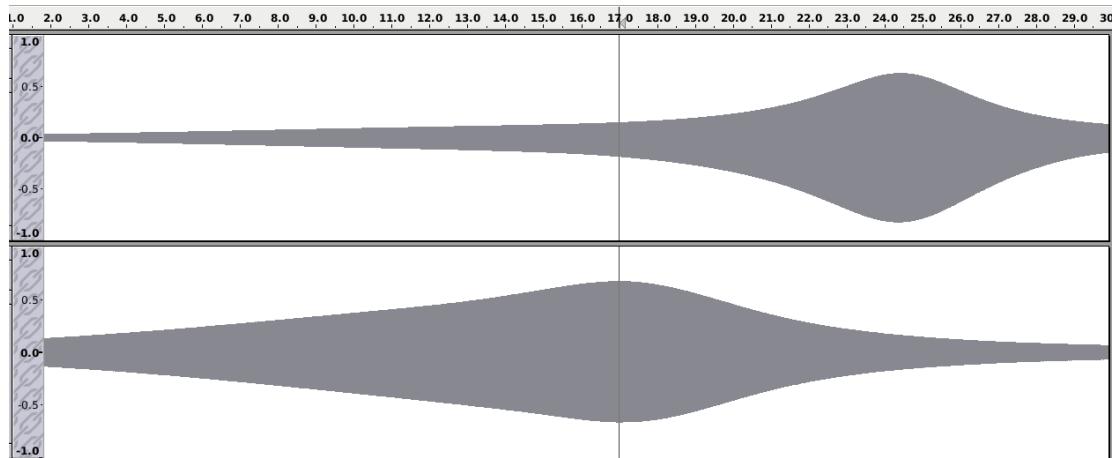


Figure 6.77: Wah effect frequency sweep waveforms

a loading effect on the wah prototype circuit, and the results for the lower resonance frequency of the circuit were not the same as measured with the oscilloscope. The lower resonance frequency according to the sound card recording was about 750 Hz, and the higher resonance frequency was the same 2400 Hz as it was with the oscilloscope, but the sine wave was distorted because of the loading effect of the sound card. Somehow the low-frequency resonance did actually miss the sharp peak and the resonance area was much more widespread than the simulations indicated.

Furthermore, from Figure 6.77 it is only possible to compare the sweep waveforms between the low and high resonance peaks. While doing the recording with the sound card, the attenuation factor was different for both of the waveforms, and this cancels the possibility to directly compare the actual amplitude of the low- and high-frequency sweep envelopes. Nevertheless, Figure 6.77 shows reasonably well the feature of the wah effect; basically it can be considered as a parallel RLC circuit, where the resonance frequency can be controlled with a simple potentiometer. This is the complete functional description of the wah effect in a nutshell. Nothing fancy from theoretical point of view (except for the clever use of the Miller effect), but still the wah is an awesome sounding effect.

6.7 AN OCTAVE DOUBLER EFFECT

There exist many different pitch changing effects, but only a few of them are simple analogue based effect devices. One good example of a simple analogue design is the so-called 'Green Ringer' pedal designed by Dan Armstrong at the beginning of 1970's. The schematic of the 'Green Ringer' circuit was obtained

from the Tonepad website [84] and is drawn in Figure 6.78. The Green Ringer pedal is categorised as an 'octave-up' effect, which is expected to double the frequency of the input signal, at least if the signal is a simple sinusoid. Further analysis indicates that for more complex signals it is not that evident that all frequency components get doubled.

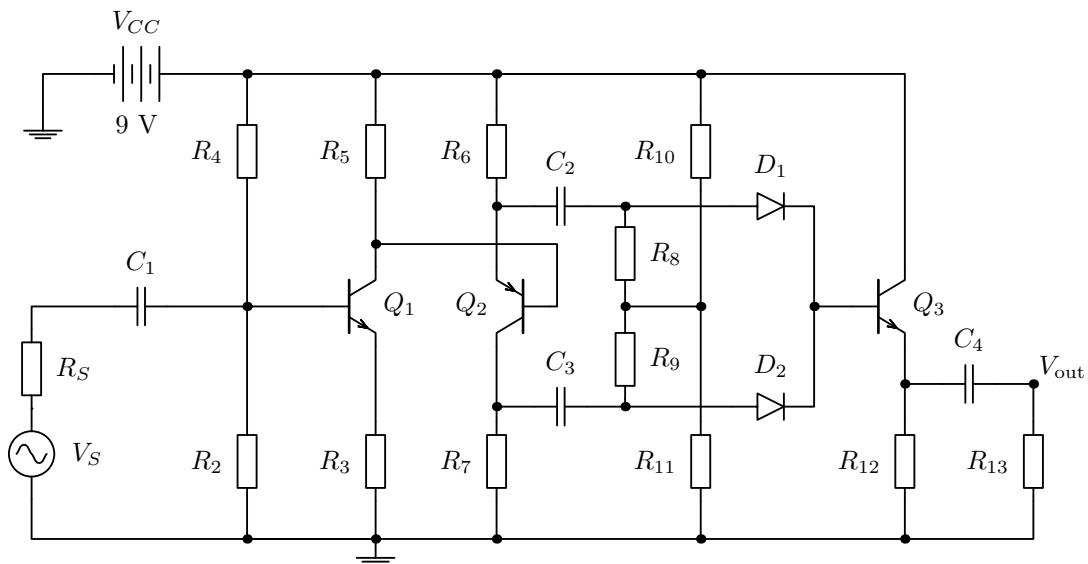


Figure 6.78: A circuit diagram of the Green Ringer effect

A short description of the signal path is as follows: the input signal from the guitar is delivered to the base of transistor Q_1 through a DC coupling capacitor C_1 . Transistor Q_1 is connected in a basic amplifier configuration with biasing resistors R_2 , R_3 , R_4 and R_5 . Transistor Q_2 is used for splitting the signal into two separate paths that have a phase difference of 180 degrees with equal gain. If the input signal is a simple sine wave, this phase difference mirrors the signal to positive and negative halves. The capacitors C_2 and C_3 isolate the DC voltage levels on each side, since the voltage divider system using resistors R_8 , R_9 , R_{10} and R_{11} sets the bases of the diodes D_1 and D_2 at equal potential. The circuit uses the two equally biased diodes to conduct the positive and the negative half of the input signal as positive halves, so eventually the frequency of the signal at the base of Q_3 is doubled to the next octave. The Green Ringer stompbox is powered by a 9 volt battery, which is typical for the majority of effect pedals.

6.7.1 The DC analysis of the Green Ringer effect

The analytical biasing calculations for transistor Q_1 can be handled with equation (1.91) presented already in the introductory section. The biasing scheme of

this circuit follows the standards of a good amplifier design, and therefore no distortion is expected out of this effect.

From the schematic 6.78 it is seen that the base voltage at Q_2 equals the quiescent collector voltage V_{CQ1} of transistor Q_1 . From here it is easy to identify that

$$V_{EQ2} = V_{CQ1} + V_{BEQ2},$$

and the emitter current

$$I_{EQ2} = \frac{V_{CC} - V_{EQ2}}{R_6},$$

which can be used for solving the collector voltage

$$V_{CQ2} \approx R_7 I_{EQ2},$$

because the assumption is that $I_{EQ2} \approx I_{CQ2}$.

The voltages around Q_3 are more difficult to solve. There might be a more exact way to determine the biasing of Q_3 , but lack of experience sometimes forces to crude approximate methods. In the active mode, the emitter current of Q_3 has a maximum value

$$I_{EQ3(max)} \approx \frac{V_{CC}}{R_{12}}.$$

With the choices of $V_{CC} = 9$ V and $R_{12} = 10$ k Ω , this maximum emitter current will be about 1 mA. If the DC current gain factor β_F of Q_3 is approximately 100, then the base current I_{BQ3} (moving towards the emitter) is about 10 μ A. This is the maximum current that the base sucks from the node point between resistors R_8 and R_9 . If this point is indexed as node 8 and the corresponding voltage at that node is labelled as V_8 , the voltage

$$V_8 = V_{CC} - (I + I_{BQ3})R_{10} = IR_{11}, \quad (6.82)$$

and the unknown current must be determined. From equation (6.82)

$$I = \frac{V_{CC} - I_{BQ3}R_{10}}{R_{10} + R_{11}}, \quad (6.83)$$

and therefore

$$V_8 = \frac{V_{CC} - I_{BQ3}R_{10}}{R_{10} + R_{11}} R_{11}. \quad (6.84)$$

From here onwards just calculate the voltage drop over R_8 and R_9 since the base current I_{BQ3} is divided equally between these two resistors. Then use the voltage drop of the ideal diode for D_1 , D_2 and for the base-emitter junction of Q_3 . This approximate procedure will reveal the emitter voltage V_{EQ3} .

6.7.2 The AC analysis of the Green Ringer effect

At first thought, the frequency response of the Green Ringer circuit is difficult to analyse by means of transfer functions because of the diodes in the circuit. Diodes are nonlinear components, so that if the analytical diode model

$$I_d = I_S (e^{V/V_T} - 1)$$

is used in the nodal analysis that solves a set of linear equations, the solution cannot be found in symbolic form. For this reason, some approximate model would be needed.

For an approximate small-signal model, it is possible to linearise the diode current equation by taking the first two terms from the power series expansion of the exponential function. From there it follows that for very small alternating currents the diode acts approximately as a normal resistor parallel with a normal semiconductor junction capacitance. The junction capacitance affects only in the high-frequency range, and in audio applications the capacitance can be neglected. Therefore, in the small-signal model one could replace the diodes with resistors.

Even when it is possible to create a decent small-signal model of the whole Green Ringer circuit, the troublesome pen-and-paper AC analysis is skipped altogether, because the Green Ringer circuit does not offer any drastically new tricks in the frequency domain. By making the AC analysis, one could only check whether the low-frequency response of the circuit is designed adequately. This does not tell much about the functionality of the circuit. The main idea of the circuit relies on the transient properties of the diodes when using them on the verge of cut-off to conduct half of the waveform and cut the other half. This would not be revealed in the traditional frequency response analysis.

As a related topic to transient time domain diode modelling, research articles [85] and [86] describe numerical approximation methods to achieve realistic diode modelling equations in distortion effect circuits, where the small-signal approximation is not valid anymore due to larger signal magnitudes.

6.7.3 A SPICE model for the Green Ringer circuit

To prepare for SPICE simulations, the Green Ringer circuit presented in Figure 6.78 is redrawn with the gschem schematic editor. Figure 6.79 shows the Green Ringer schematic including the component values used in the simulations.

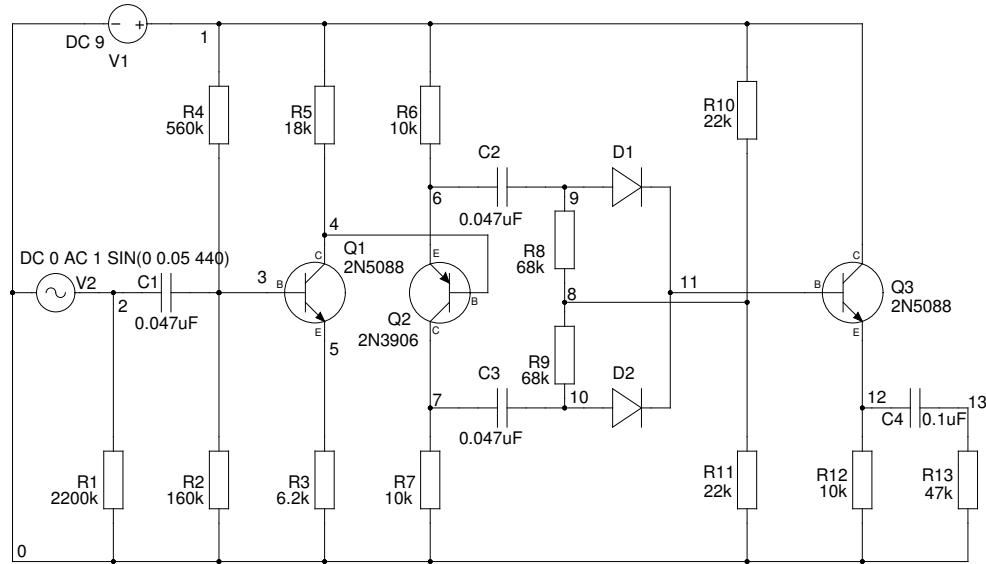


Figure 6.79: The Green Ringer effect schematic drawn with gschem

In addition to real resistor and capacitor values, the transistor gains and saturation currents need to be defined as model parameters to the gschem transistor properties sheet. Both 2N3906 and 2N5088 are very basic silicon transistors, and the SPICE parameters for them are easily found from component-specific datasheets. The diodes used here can be almost any diodes as long as they have quite accurately defined voltage drop across them. The SPICE model used for the diodes just defines the saturation current, so basically the default SPICE diode model is used in the simulations.

When using the gEDA design tools collection in Linux, the netlist for SPICE simulations is created from the gschem schematic of Figure 6.79 with the command

```
gnetlist -g spice-sdb -o greenrng_net.net greenrng.sch
```

where the output file named as `greenrng_net.net` is created from the schematic file `greenrng.sch`. The automatically created netlist file is as follows:

```
=====
 Begin SPICE netlist of main design =====
 Q2 7 4 6 PNP1
 .MODEL PNP1 PNP (Is=1.41f Xti=3 Eg=1.11 Vaf=18.7 Bf=180.7 Ne=1.5 Ise=0 Ikf=80m Xtb=1.5
 + Br=4.977 Nc=2 Isc=0 Ikr=0 Rc=2.5 Cjc=9.728p Mjc=.5776 Vjc=.75 Fc=.5
 + Cje=8.063p Mje=.3677 Vje=.75 Tr=33.42n Tf=179.3p Itf=.4 Vtf=4 Xtf=6 Rb=10)
 V2 0 2 DC 0 AC 1 SIN(0 0.05 440)
 V1 1 0 DC 9
 D2 10 11 Dido
 .MODEL Dido D (Is=1.0e-15)
 D1 9 11 Dido
 R8 8 9 68k
```

```

R9 10 8 68k
R1 0 2 2200k
R4 3 1 560k
R2 0 3 160k
Q3 1 11 12 NPN1
Q1 4 3 5 NPN1
.MODEL NPN1 NPN (Is=5.911f Xti=3 Eg=1.11 Vaf=62.37 Bf=1.122K Ne=1.394 Ise=5.911f Ikf=14.92m
+ Xtb=1.5 Br=1.271 Nc=2 Isc=0 Ikr=0 Rc=1.61 Cjc=4.017p Mjc=.3174 Vjc=-.75 Fc=.5
+ Cje=4.973p Mje=.4146 Vje=-.75 Tr=4.673n Tf=821.7p Itf=.35 Vtf=4 Xtf=7 Rb=10)
R13 0 13 47k
R12 0 12 10k
C4 12 13 0.1uF
C3 7 10 0.047uF
C2 6 9 0.047uF
C1 2 3 0.047uF
R10 8 1 22k
R6 6 1 10k
R11 0 8 22k
R7 0 7 10k
R3 0 5 6.2k
R5 4 1 18k
.end

```

The frequency response would not reveal anything special in this case, so the only meaningful analysis of the circuit would be the transient simulation. For this purpose, the simulation control commands are set for the transient simulation by adding the control statements

```

.control
tran 1us 20ms
gnuplot grn_tran tran.v(13) tran.v(2)
.endc

.OP

```

where the results of the simulation are requested to be saved to a gnuplot formatted data file. Eventually, the simulation is executed with ngspice using the command

```
ngspice -b greenrng_net.net
```

which prints the operating point information to the computer screen and creates a data file grn_tran.data to be plotted later with gnuplot.

6.7.4 Results of simulations, calculations and measurements

Regarding the DC analysis of section 6.7.1, the component values indicated in the schematic 6.79 were inserted into the equations and evaluated numerically using Octave. Numerical results were obtained for all transistor and diode pin voltages. The internal potential difference V_{BE} of the 2N3906 and 2N5088 transistors in the Green Ringer circuit was estimated to be approximately 0.6 V, and this value was also used in the calculations.

Regarding the DC simulation in SPICE, the netlist from the circuit of Figure 6.79 was loaded into ngspice and analysed with the operating point (.OP) command, which was written inside the netlist file as indicated in the previous section. The simulation results for DC bias voltages were read from the output listing provided by ngspice.

For real world measurements, the Green Ringer circuit shown in Figure 6.79 was built on a small solderless breadboard, a.k.a. a plugboard, where the components can be replaced easily just by plugging them onto the board. While keeping the circuit in a static DC configuration, a regular multimeter was used for measuring the transistors' quiescent voltages. The simulated, calculated and measured quiescent values are gathered in Table 6.5 for comparison purposes.

Table 6.5: DC operating point voltages of the Green Ringer circuit

value	simulated	calculated	measured
V_{B1}	1.95	1.95	1.71
V_{C1}	5.20	5.09	5.37
V_{E1}	1.32	1.35	1.33
V_{B2}	5.20	5.09	5.37
V_{C2}	3.11	3.31	2.82
V_{E2}	5.87	5.69	6.08
V_{B3}	3.97	3.45	3.94
V_{E3}	3.33	2.85	3.63
V_{D1}	4.47	4.05	4.13
V_{D2}	4.47	4.05	4.13
V_8	4.49	4.39	4.41

Mostly the analytically obtained voltages are similar to the measured ones, but of course a few deviations exist. Especially the voltage at the diode pins is something that differed significantly for all methods of analysis. The approximate derivation of the biasing voltages of transistor Q_3 was also not very successful, as the manually calculated value of V_{E3} clearly differs from the simulated and measured values.

Transient simulations in SPICE were carried out using a 440 Hz sine wave as the test signal. The results of the simulation are shown in Figure 6.80, which verifies that the output signal has a frequency of two times the input signal. At

the beginning of the waveform there seems to be an initial transient effect that alters the behaviour of the output signal for a short while.

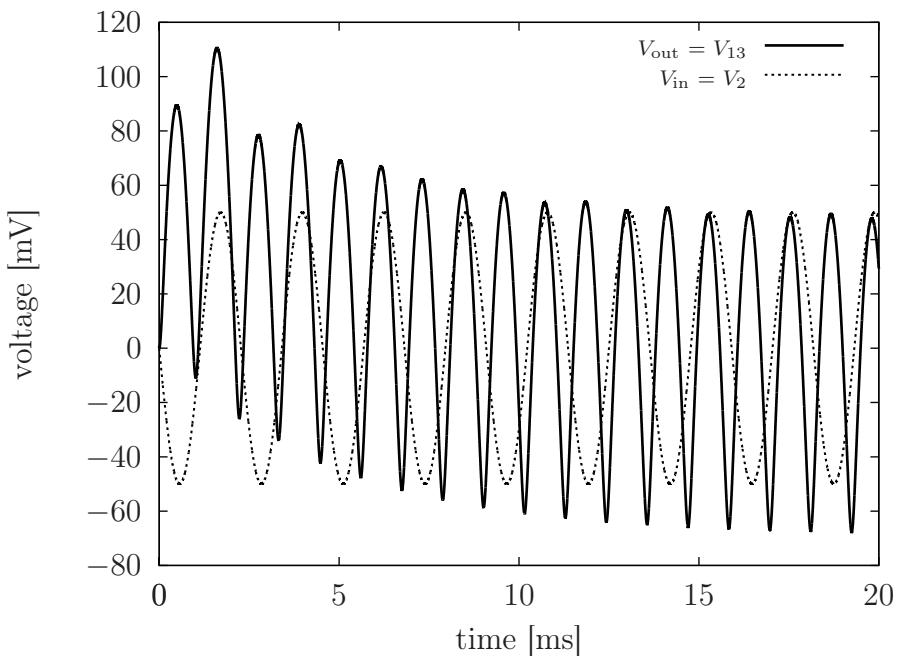


Figure 6.80: The simulated input and output signals of the Green Ringer circuit

In the prototype measurements, a computer sound card was used as a signal source, from where a 440 Hz sine wave was played out to the input terminals of the Green Ringer circuit. The input and output waveforms were measured with a digital storage oscilloscope and saved in a .cvs file format for later plotting with gnuplot. Figure 6.81 indicates that the circuit works as the simulations predicted, although the DC biasing of the prototype circuit differed from the simulation values. A controlled test for verifying the Green Ringer behaviour with more complex input signals was not done, but the ear indicates that the octave doubling effect is not achieved in this case.

The 'ad-lib' testing of the prototype circuit revealed that the ringer-effect can also be applied without clipping with relatively high input voltages close to 1 V. The Green Ringer raises the pitch of the tone by one octave, but clearly it only works as intended for nearly theoretical signals that resemble more or less pure a sine wave. Therefore, if the effect is used when playing chords, the effect sounds quite horrible. Unfortunately this effect device does not offer much material for theoretical analysis from the scientific perspective.

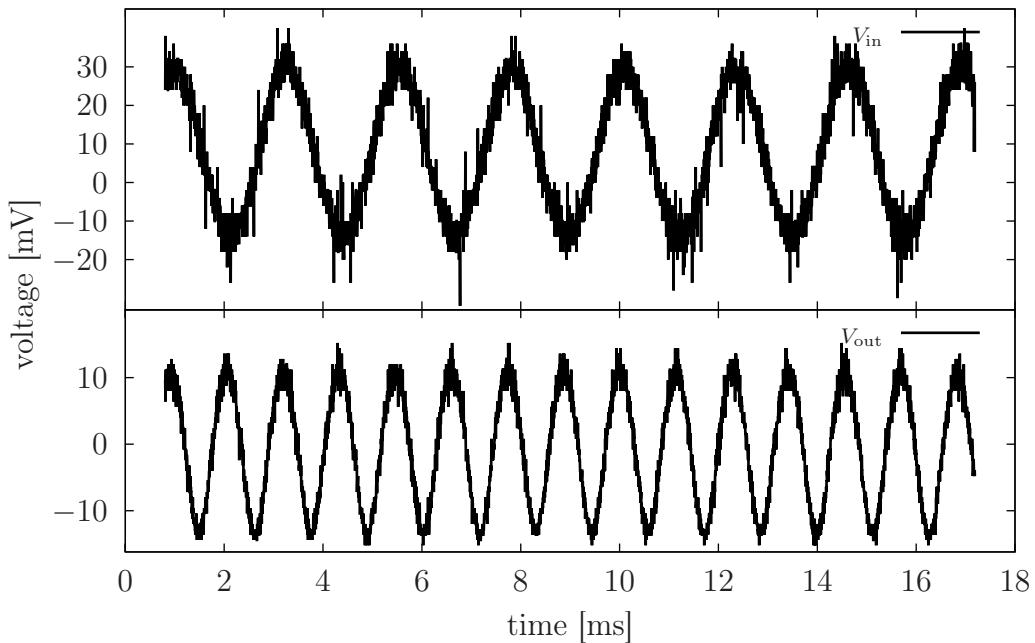


Figure 6.81: Measured input and output waveforms of the Green Ringer effect

6.8 PHASE EFFECTS

When stepping towards more complicated circuitry, up comes the almighty phaser and the even mightier flanger. These two effects are the ones to cover under the topic of phase effects, although the flanger is more of a time-based delay effect than an actual phase-shifting circuit. But since the basic idea of notch filters is used both in phasers and flangers, it is better to analyse both effects under a common section.

Although phase shifting had been used as an effect in studio recordings for some time already, the first phaser and flanger guitar pedals came to the markets in early 1970's. As a separate effect unit, these pedals clearly offer more possibilities to tweak the sound because the distortion effect and basic tone control are usually built in as separate sections into the amplifiers.

6.9 A PHASE SHIFTER

Due to the complexity of implementation by purely analogue components, only a few simple phase shifters have been manufactured. Good examples for further analysis are the commercial effect pedals carrying brand names 'MXR Phase 45' and 'MXR Phase 90', the big brother of MXR Phase 45. These popular effect pedals were manufactured by a company called MXR Innovations during the

years 1974 – 1980 along with other guitar effect pedals in the categories of distortions, delays and flangers [69, p. 82].

Just like all the effects already covered, a phase shifter is based on basic techniques of signal manipulation. In a general phase-shifting effect, the sound is modified using a stack of voltage-controlled all-pass filters, which do not filter out any frequencies but only alter the phase of the frequency components of the signal. This phase-modified signal is summed to the original input signal, which is taken to the output through a separated signal path. The idea can be made more understandable using a signal flow graph, which shows the route of the signal and the basic operations done to the signal as it progresses through the system. A simplified black box flow graph representing the phase shifter effect is depicted in Figure 6.82.

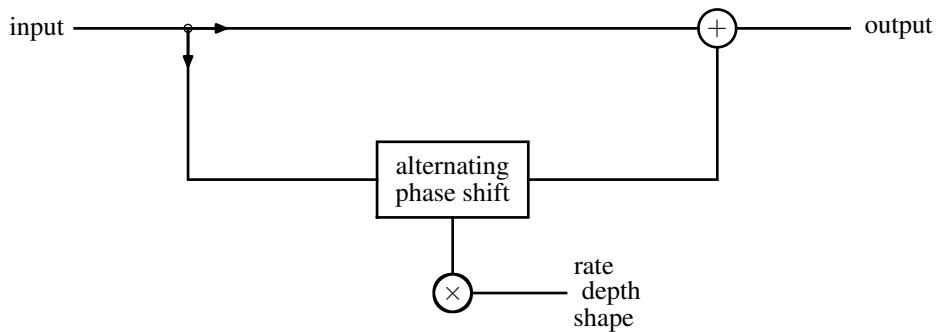


Figure 6.82: A signal flow graph of a typical phase-shifting effect

In a typical phase-shifting effect the magnitude of the phase shift is different and periodically changing for all frequencies. This functionality gives a frequency sweep type of an effect, although in this context it can be called a phase sweep. Figure 6.82 also shows the most common modulation alternatives for the phase-shifted signal. The user of a phase shifter effect can typically control the rate of the periodic phase sweep, the magnitude ratio of the original and phase-shifted signals and the waveform (sine, triangle, square, etc.) used to periodically modulate the phase sweep.

The MXR Phase 45 guitar effect schematic is presented in Figure 6.83. Most likely this schematic has been originally leaked to the public domain by someone who has directly traced out the circuit board of the original device. One could easily assume that the number related to each phaser model manufactured under the brand of MXR refers to the amount of phase shift the device produces. The following analysis will reveal whether this assumption is correct or false.

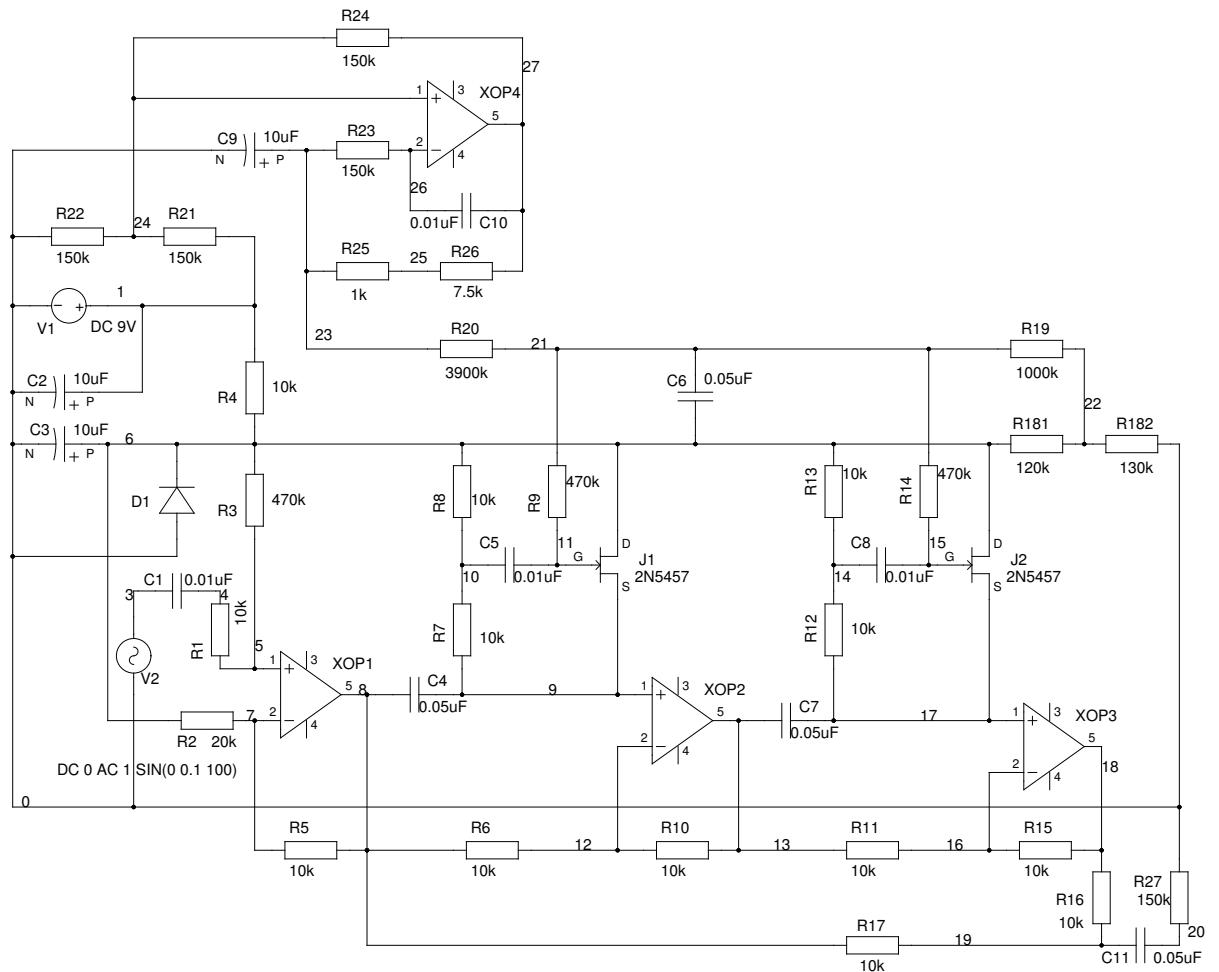


Figure 6.83: Phase 45 effect pedal schematic drawn with gschem

As a brief overview of the circuit, the input signal goes first through a buffering amplifier XOP1 and is then split into two different paths at node 8. One path goes to the output stage through a filter section formed by two op-amps (XOP2, XOP3) and the other path leads directly to the output stage. The separated signals are added back together via resistors R_{16} and R_{17} , which in this case mix the signals with even balance (10/10). The output signal is taken from node 20, right after the DC blocking capacitor C_{11} .

The signal going through the phase-shifting stage is affected by a constantly changing amount of phase shift for each frequency component. This alternating phase shift is accomplished by varying the filter section resistance by dynamically modulating the gate-to-source voltage of the JFETs J_1 and J_2 near the pinch-off voltage – just like in the tremolo effect. The modulating signal comes from a basic oscillator circuit, which can be identified as a textbook example of a relaxation oscillator. The properties of the relaxation oscillator will be covered as a

side effect in section 6.9.1.

The phase shift network is connected to a virtual ground reference voltage, which is mainly generated by the zener diode D_1 . This zener diode is used for setting the magnitude of the virtual ground to about half the value of the operating voltage, which is the typical 9 volt battery. The DC voltage level of the modulating signal at the gate of the JFETs is biased by a trimmer potentiometer R_{18} . This biasing should be so that if the virtual ground is at 4.5 V and the pinch-off voltage V_P of the JFETs is -1.8 V, potentiometer R_{18} should be trimmed to give a DC voltage of approximately $4.5\text{ V} - 1.8\text{ V} = 2.7\text{ V}$ to the gate of the JFETs. The signal from the oscillator stage will oscillate around the pinch-off voltage, thereby modifying the drain-to-source resistance of the JFETs.

The design requires that the two op-amp filters behave identically to give exactly the same phase shift at all times. Therefore, the components used to construct the filters should have as similar values as possible. The most critical components are the JFETs because of the large variation in their pinch-off voltages. The JFETs used in the circuit need to be manually measured to have similar values of pinch-off voltage V_P and saturation current I_{DSS} to unleash all the power that this Phase 45 effect pedal has to offer.

6.9.1 Side effect: the relaxation oscillator

The schematic of the Phase 45 effect contains two interesting examples of using the operational amplifier. Firstly the op-amp is used for creating low-frequency oscillations as an relaxation oscillator, and secondly it can be used for realising a huge assortment of different kinds of filter transfer functions.

To concentrate first on the oscillator implementation, the oscillator solution used in the Phase 45 effect is isolated from the main schematics. Figure 6.84 presents a theoretical textbook example and the Phase 45 modified version of the relaxation oscillator circuit side by side. These kind of oscillators are relatively difficult to analyse because their functionality is based on the transient behaviour of the capacitor charging and discharging process. Therefore, it is not possible to directly calculate a transfer function of the circuit that would describe all the oscillatory properties.

The operational amplifier in the relaxation oscillator configuration is wired as a Schmitt trigger. In this setup a threshold voltage level is set to the positive input, and as the negative input crosses this threshold voltage, the output saturates

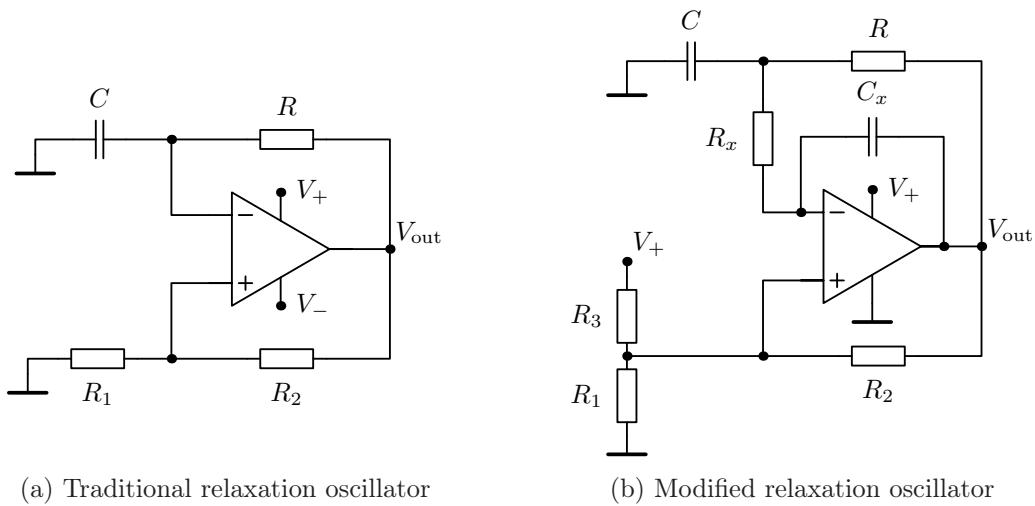


Figure 6.84: Relaxation oscillators

immediately to full operating voltage. Rising edge cross of negative input voltage over the threshold causes the output to swing to negative saturation and vice versa. This happens because of the feedback to the positive input together with the huge open-loop gain of the operational amplifier.

The simplest way to analyse this basic circuit is to assume that the output voltage of the circuit is a clean square wave signal. This square wave feeds the capacitor C through R , and depending on the state (V_+ or V_-) of the output voltage V_{out} , the capacitor is either charging or discharging with some time constant RC . The transient function of capacitor charging needs to be calculated to find out how long it takes to reach the switching levels that change the output from positive saturation to negative saturation and back to positive again. This time constant controlled process produces the square wave output oscillation of a certain frequency f_{rlx} .

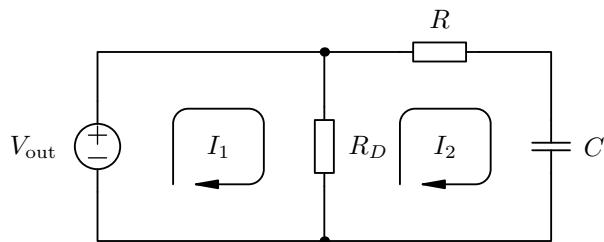


Figure 6.85: A simple relaxation oscillator RC network without an op-amp

Figure 6.85 shows the initial setup for the analysis. The precondition for the analysis is that the voltage source V_{out} gives a steady DC voltage of V_+ and that the capacitor C is fully discharged. Resistor R_D models the voltage divider section

presented in Figure 6.84a. Previously the nodal analysis method was used to produce the matrix equations of circuits, but in this case it is not very suitable because there is no series resistance to convert the voltage source to a current source. It is about time to introduce the power of mesh analysis. For mesh currents I_1 and I_2 of Figure 6.85, one can write

$$\begin{bmatrix} R_D & -R_D \\ -R_D & R_D + R + \frac{1}{sC} \end{bmatrix} \times \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} \frac{V_{\text{out}}}{s} \\ 0 \end{bmatrix}.$$

The voltage v_c over the capacitor is $I_2 Z_C$, where Z_C is the impedance of the capacitor. After solving the previous matrix equation

$$I_2(s) = \frac{V_{\text{out}}}{R} \frac{1}{s + \frac{1}{RC}},$$

and voltage over the capacitor as a function of the Laplace variable s is

$$v_c(s) = I_2(s) \frac{1}{sC} = \frac{V_{\text{out}}}{RC} \frac{1}{s \left(s + \frac{1}{RC} \right)}. \quad (6.85)$$

Next the method of partial fractions is used to modify the expression to a suitable form for making the inverse Laplace transform. According to the directives of partial fractions,

$$\frac{V_{\text{out}}}{RC} \frac{1}{s \left(s + \frac{1}{RC} \right)} = \frac{A}{s} + \frac{B}{s + \frac{1}{RC}},$$

and multiplication by $s \left(s + \frac{1}{RC} \right)$ yields

$$\frac{V_{\text{out}}}{RC} = A \left(s + \frac{1}{RC} \right) + Bs. \quad (6.86)$$

From (6.86) a pair of equations can be formed to solve values for the multipliers A and B . Equating similar powers of s gives

$$As + Bs = 0s \quad \Rightarrow \quad A = -B \quad (6.87)$$

$$\frac{A}{RC} + 0B = \frac{V_{\text{out}}}{RC} \quad \Rightarrow \quad A = V_{\text{out}}. \quad (6.88)$$

Now the parameters A and B are defined and equation (6.85) becomes

$$v_c(s) = \frac{V_{\text{out}}}{s} - \frac{V_{\text{out}}}{s + \frac{1}{RC}}. \quad (6.89)$$

The inverse Laplace transform for equation (6.89) is easily obtained from tabulated standard formulae given in Appendix E. The voltage v_c in the time domain can now be expressed as

$$v_c(t) = V_{\text{out}} \left(1 - e^{-\frac{t}{RC}} \right). \quad (6.90)$$

This model describes only the situation where the initially fully discharged capacitor is charged from a constant voltage V_{out} to full charge via the RC feedback network. It is to be noted that this charging process does not depend on the resistance R_D of the voltage divider network used for creating the threshold voltage levels for output voltage switching.

Taking into account the charging thresholds provided by the voltage divider R_1 and R_2 of Figure 6.84a, the voltage v_i between the positive and negative inputs of the op-amp is

$$v_i = v_c - \frac{R_1}{R_1 + R_2} V_{\text{out}}.$$

In this case the initial charge in the capacitor has a voltage $\frac{R_1}{R_1 + R_2} V_{\text{out}}$ and

$$v_c(t) = V_{\text{out}} \left[1 - \left(1 + \frac{R_1}{R_1 + R_2} \right) e^{-\frac{t}{RC}} \right]. \quad (6.91)$$

From here it is possible to solve the oscillation period T , using the fact that when $t = T/2$, the voltage measured over the capacitor reaches the higher threshold voltage. Then from (6.91),

$$\frac{R_1}{R_1 + R_2} V_{\text{out}} = V_{\text{out}} \left[1 - \left(1 + \frac{R_1}{R_1 + R_2} \right) e^{-\frac{T}{2RC}} \right]. \quad (6.92)$$

Solving this equation for T leads to the time for the oscillating period

$$T = 2RC \ln \left(1 + \frac{2R_1}{R_2} \right). \quad (6.93)$$

This applies only to the case where the supply voltages of the op-amp are $\pm V_S$. More generally, equation (6.92) can be written as

$$V_{\text{limH}} = V_{\text{out}} - (V_{\text{out}} - V_{\text{limL}}) e^{-\frac{T}{2RC}}, \quad (6.94)$$

which gives for the period of oscillation

$$T = 2RC \ln \left(\frac{\frac{1 - \frac{V_{\text{limL}}}{V_{\text{out}}}}{1 - \frac{V_{\text{limH}}}{V_{\text{out}}}}}{\frac{1 - \frac{V_{\text{limL}}}{V_{\text{out}}}}{1 - \frac{V_{\text{limH}}}{V_{\text{out}}}}} \right). \quad (6.95)$$

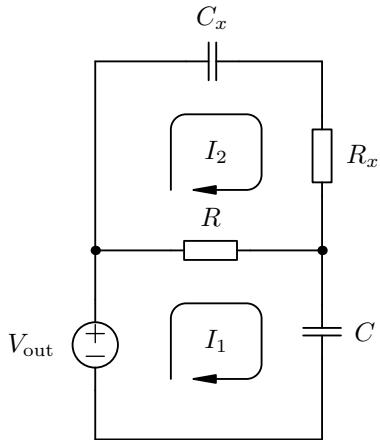


Figure 6.86: The modified relaxation oscillator RC network without an op-amp

Moving on to analyse the modified version of the relaxation oscillator. The equivalent network for mesh analysis is given in Figure 6.86. The voltage divider resistances have been left out from this model because in the previous analysis they were noted to have no effect on the final result. Writing out the equations for the mesh currents gives the matrix equation

$$\begin{bmatrix} R + \frac{1}{sC} & -R \\ -R & R + R_x + \frac{1}{sC_x} \end{bmatrix} \times \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} \frac{V_{\text{out}}}{s} \\ 0 \end{bmatrix}.$$

In the Phase 45 circuit the output is taken on top of C but the input for the op-amp comes on top of R_x . The preference is to analyse the voltage at C . One way to get an expression for the voltage over C is to calculate the current flowing through R , and then $v_c = V_{\text{out}} - (I_1 - I_2)R$. From the matrix equation

$$I_1(s) = \frac{\frac{V_{\text{out}}}{s} \left(R + R_x + \frac{1}{sC_x} \right)}{R \left(R_x + \frac{1}{sC_x} \right) + \frac{1}{sC} \left(R + R_x + \frac{1}{sC_x} \right)} \quad (6.96)$$

$$I_2(s) = \frac{\frac{V_{\text{out}}}{s} R}{R \left(R_x + \frac{1}{sC_x} \right) + \frac{1}{sC} \left(R + R_x + \frac{1}{sC_x} \right)}. \quad (6.97)$$

After taking the difference $I_1 - I_2$ and simplifying, the current through R can be

written as

$$I_R(s) = I_1(s) - I_2(s) = \frac{\frac{V_{\text{out}}}{R} \left(s + \frac{1}{R_x C_x} \right)}{s^2 + \left(\frac{1}{RC} + \frac{1}{R_x C_x} + \frac{1}{R_x C} \right) s + \frac{1}{RC R_x C_x}}. \quad (6.98)$$

Usually the procedure of the inverse Laplace transform is to try to square the denominator by adding and subtracting the needed terms, but in this case it is not a proper method. The rough method is to use the quadratic formula to find the roots of the denominator and to use those roots to write the denominator as a product of two terms. In this case, this method gives the roots

$$s_{1,2} = \left(\frac{1}{2RC} + \frac{1}{2R_x C_x} + \frac{1}{2R_x C} \right) \pm \sqrt{\left(\frac{1}{2RC} + \frac{1}{2R_x C_x} + \frac{1}{2R_x C} \right)^2 - \frac{1}{RC R_x C_x}}.$$

Clearly a substitution is needed to avoid excessively long equations. In the following formulations the terms of roots s_1 and s_2 are addressed as

$$b = \frac{1}{2RC} + \frac{1}{2R_x C_x} + \frac{1}{2R_x C} \quad \text{and} \quad c = \frac{1}{RC R_x C_x}.$$

With these substitutions, equation (6.98) can be written as

$$I_R(s) = \frac{\frac{V_{\text{out}}}{R} \left(s + \frac{1}{R_x C_x} \right)}{(s + b + \sqrt{b^2 - c})(s + b - \sqrt{b^2 - c})}. \quad (6.99)$$

A similar approach as used previously is adopted here to try and modify the equation into a form where one can apply the inverse Laplace transform directly from tables. A partial fraction expansion of equation (6.99) starts with

$$\frac{\frac{V_{\text{out}}}{R} \left(s + \frac{1}{R_x C_x} \right)}{(s + b + \sqrt{b^2 - c})(s + b - \sqrt{b^2 - c})} = \frac{A}{s + b + \sqrt{b^2 - c}} + \frac{B}{s + b - \sqrt{b^2 - c}},$$

and this gives a pair of equations,

$$A + B = \frac{V_{\text{out}}}{R} \quad (6.100)$$

$$A(b - \sqrt{b^2 - c}) + B(b + \sqrt{b^2 - c}) = \frac{V_{\text{out}}}{RR_x C_x}, \quad (6.101)$$

which give the factors A and B the values

$$B = \frac{V_{\text{out}}}{R} \frac{1}{2\sqrt{b^2 - c}} \left[\frac{1}{R_x C_x} - (b - \sqrt{b^2 - c}) \right] \quad (6.102)$$

$$A = \frac{V_{\text{out}}}{R} \left(1 - \frac{1}{2\sqrt{b^2 - c}} \left[\frac{1}{R_x C_x} - (b - \sqrt{b^2 - c}) \right] \right). \quad (6.103)$$

Finally, the solution for the time dependent voltage at the capacitor is

$$v_c(t) = V_{\text{out}} - (V_{\text{out}} - V_{\text{limL}}) (Ae^{-\tau_A t} + Be^{-\tau_B t}), \quad (6.104)$$

where τ_A and τ_B are those very complicated expressions involving a mixture of time constants due to R, C, R_x and C_x .

Things got rather complicated again after adding two components to the original circuit. So why have R_x and C_x been added to the circuit in the first place? The author's guess is that those are there to prevent hysteresis, because basic Schmitt triggers are known to suffer from hysteresis at the transition region. It is possible to get a hysteresis problem especially with low-frequency changes around the trigger point. The trick with the extra capacitor C_x is to use it with a considerably faster time constant so that the extra charge in capacitor C_x triggers the change and empties out fast to prevent the hysteresis effect. The original large-valued capacitor never reaches over the triggering limits thanks to this extra trigger capacitor. Additionally, capacitor C_x might slightly increase the rise times of the output square wave.

The next step is to show that the oscillation times for the modified circuit can still be calculated using the formula of the traditional relaxation oscillator. Because the product $R_x C_x$ is so much smaller than the product RC , the term previously assigned as b can be approximated with

$$b = \frac{1}{2RC} + \frac{1}{2R_x C_x} + \frac{1}{2R_x C} \approx \frac{1}{2R_x C_x},$$

and with this approximation the roots s_1 and s_2 can be written as

$$s_{1,2} = -\frac{1}{2R_x C_x} \left(1 \pm \sqrt{1 - \frac{4R_x C_x}{RC}} \right).$$

Here the square root can be expanded as a binomial approximation containing the first two terms of the binomial series

$$\sqrt{1 - \frac{4R_x C_x}{RC}} \approx 1 - \frac{2R_x C_x}{RC},$$

so that the roots s_1 and s_2 become

$$s_1 = -\frac{1}{R_x C_x} + \frac{1}{RC} \approx -\frac{1}{R_x C_x} \quad (6.105)$$

$$s_2 = -\frac{1}{RC}. \quad (6.106)$$

With these roots the partial fraction expansion finally leads to expression

$$I_R(s) = \frac{A}{s + \frac{1}{RC}} + \frac{B}{s + \frac{1}{R_x C_x}},$$

where $B = 0$ and $A = \frac{V_{\text{out}}}{R}$. Now the voltage over the capacitor C can be written

$$v_c(t) = V_{\text{out}} - RI_R(t) = V_{\text{out}} \left(1 - e^{-\frac{t}{RC}} \right). \quad (6.107)$$

The final formula for the voltage of the capacitor C is eventually approximated to be the same as for the textbook version of the relaxation oscillator. Hence, when choosing the component values for R and C in this modified oscillator, the design can be made using the familiar formulae already derived for the traditional relaxation oscillator.

6.9.2 Side effect: an op-amp all-pass filter

Operational amplifiers can be used to construct a variety of filter functions, where the so-called biquad filters are the most commonly used. Biquad filters can be used for realising second order filter functions. There exist simple operational amplifier realisations also for simpler first order filter functions. The filter used in the Phase 45 effect is an all-pass filter, which only modifies the phase of the signal. There are a few ways to analyse the transfer function of this all-pass filter.

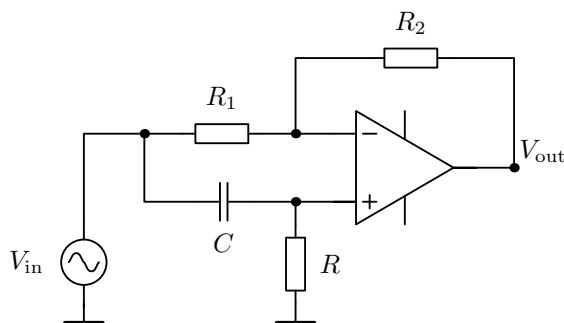


Figure 6.87: A simple all-pass filter realisation using an operational amplifier

One way relies on the very basic Kirchhoff's current law. The second method is again the more general nodal analysis method, but with slight modifications due to the fact that the input terminals of the ideal operational amplifier are assumed to have the same voltage at all times. In this simple case, let's analyse the all-pass filter circuit using Kirchhoff's current law. The introduction to the systematic and more general matrix-based nodal analysis of operational amplifier circuits is presented later.

Based on the statement that the sum of currents leaving a node is zero, the following equations can be written from the schematic of Figure 6.87:

$$\frac{V_- - V_{\text{in}}}{R_1} + \frac{V_- - V_{\text{out}}}{R_2} = 0 \quad (6.108)$$

$$(V_+ - V_{\text{in}}) sC + \frac{V_+}{R} = 0, \quad (6.109)$$

where the voltages V_- and V_+ are the voltages at the input terminals – and + of the operational amplifier. Because $V_+ = V_-$, it is reasonable to solve V_- from equation (6.108) and V_+ from equation (6.109)

$$V_- = \frac{V_{\text{in}} R_2 + V_{\text{out}} R_1}{R_1 + R_2}$$

$$V_+ = \frac{V_{\text{in}} sRC}{1 + sRC},$$

and assign these expressions as equal. After simplifying the expressions,

$$\frac{V_{\text{in}} R_2 + V_{\text{out}} R_1}{R_1 + R_2} = \frac{V_{\text{in}} sRC}{1 + sRC}. \quad (6.110)$$

To reach the transfer function, the ratio of $\frac{V_{\text{out}}}{V_{\text{in}}}$ must be solved. Equation (6.110) reveals that the ratio

$$\frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{\frac{R_2}{R_1} - sRC}{1 + sRC}. \quad (6.111)$$

This was reasonably easy to evaluate, but when the circuits have more components, it eventually becomes impossible to use this basic method to evaluate the transfer functions in symbolic form. For that purpose, it is necessary to investigate how the ideal op-amp circuits are analysed using the systematic method of nodal matrix equations.

6.9.3 Side effect: the op-amp in the nodal matrix method

Unfortunately the extremely useful nodal matrix analysis does not work directly with circuits using ideal op-amps. This is due to equal voltages at the input terminals of the operational amplifier. However, with a few additional steps it is possible to analyse op-amp circuits efficiently using the nodal matrix method of analysis.

As an example, the op-amp filter of section 6.9.2 is reanalysed. Firstly it is redrawn to Figure 6.88, so that the voltage source is transformed as a current source as is required for the nodal analysis. From this figure the following matrix

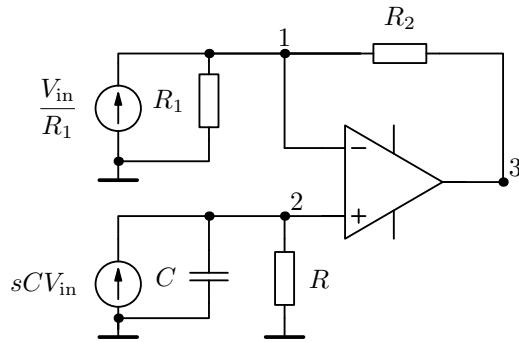


Figure 6.88: The op-amp filter circuit ready for nodal analysis

equation can be deduced:

$$\begin{bmatrix} \frac{1}{R_1} + \frac{1}{R_2} & 0 & -\frac{1}{R_2} \\ 0 & \frac{1}{R} + sC & 0 \\ -\frac{1}{R_2} & 0 & \frac{1}{R_2} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ sCV_{in} \\ I_0 \end{bmatrix}.$$

This matrix needs to be reduced a little to yield direct results. Because in the ideal op-amp $V_1 = V_2$, all the terms in column 1 can be added to column 2 and after this the whole column 1 can be thrown away from the matrix. This column deletion leaves one row too many to the matrix. Row 3 is connected to the output where we have put just some undeterministic output current I_0 . This current can be evaluated later when the voltages V_1 and V_2 have been solved. So out goes row 3. The reduced matrix is now

$$\begin{bmatrix} \frac{1}{R_1} + \frac{1}{R_2} & -\frac{1}{R_2} \\ \frac{1}{R} + sC & 0 \end{bmatrix} \times \begin{bmatrix} V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ sCV_{in} \end{bmatrix}.$$

The output voltage is marked as node 3 in Figure 6.88, so that is what is going to be solved:

$$V_3 = \frac{\frac{V_{in}sC}{R_1} + \frac{V_{in}sC}{R_2} - \frac{V_{in}}{R_1}\frac{1}{R} - \frac{V_{in}sC}{R_1}}{\left(\frac{1}{R} + sC\right)\frac{1}{R_2}}.$$

After dividing V_{in} to the left side of the equation and some amount of simplifica-

tions, the result is obtained as

$$\frac{V_3}{V_{\text{in}}} = \frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{\frac{R_2}{R_1} - sRC}{1 + sRC}, \quad (6.112)$$

which is the same result as in equation (6.111). Section 9.1 gives more examples on the use of the operational amplifier nodal analysis in matrix form.

With reference to Figure 6.89, the general rules for applying the nodal method of analysis in circuits containing ideal operational amplifiers are as follows:

1. Write the complete admittance matrix equation of the analysed circuit
2. Remove equation (row) related to node C
3. If node A(B) is connected to ground, remove column B(A). Otherwise add column B(A) to column A(B) and remove column B(A)

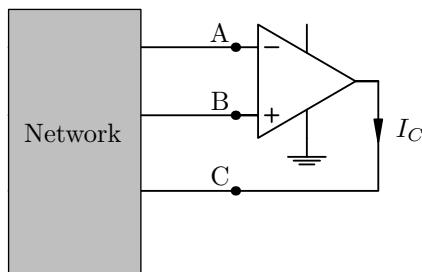


Figure 6.89: The ideal operational amplifier in the nodal method of analysis

6.9.4 The DC analysis of the Phase 45 effect

The DC analysis can be carried out in this case without any detailed calculations, since the analysis can mostly rely on the low current properties of the JFETs and the operational amplifiers. The first thing to distinguish is the virtual ground set by a 4.7 V zener diode. From there it is possible to figure out that the JFET drain-to-source voltage $V_{DS} = 0$ because no current is flowing to the op-amp inputs. Since there is practically no current going through the two $10\text{ k}\Omega$ resistors, many of the locations inside the filter constructions have the same DC voltage in the static configuration.

With reference to the circuit diagram of Figure 6.83, the capacitors C_5 and C_8 are open circuits at constant voltages, and therefore the DC voltage to the JFET gates is taken through a voltage divider system formed by R_{18} , R_{19} and R_{20} .

The relaxation oscillator is like a separate unit in this circuit, and because it will be constantly oscillating, it is not analysed for DC voltages.

6.9.5 The AC analysis of the Phase 45 effect

The Phase 45 circuit of Figure 6.83 has two all-pass filters connected in series. How does that change the transfer function from equation (6.111)? The all-pass filters are the actuators for the phase shift, so the AC analysis will focus on investigating the behaviour of these filters. The circuit to be analysed in this case is depicted in Figure 6.90.

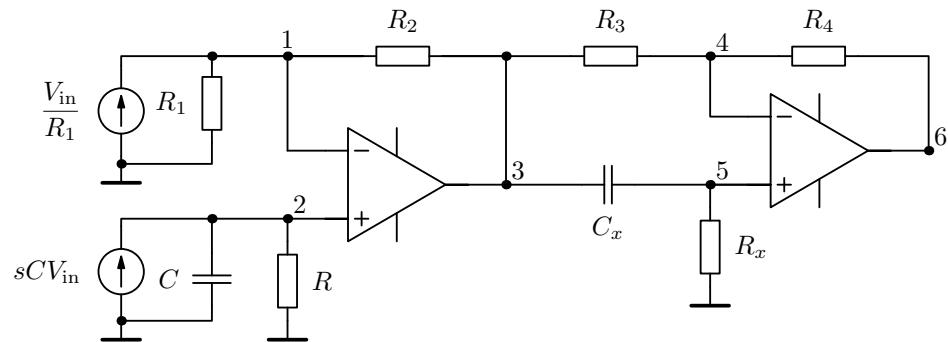


Figure 6.90: The all-pass filter section of the Phase 45 circuit

The nodal analysis of this circuit proceeds similarly as shown in section 6.9.3. To save one intermediate step, the matrix representation (6.113) of this circuit is directly written in the reduced form with two rows and columns removed.

$$\begin{bmatrix} \frac{1}{R_1} + \frac{1}{R_2} & -\frac{1}{R_2} & 0 & 0 \\ \frac{1}{R} + sC & 0 & 0 & 0 \\ 0 & -\frac{1}{R_3} & \frac{1}{R_3} + \frac{1}{R_4} & -\frac{1}{R_4} \\ 0 & -sC_x & \frac{1}{R_x} + sC_x & 0 \end{bmatrix} \times \begin{bmatrix} V_2 \\ V_3 \\ V_5 \\ V_6 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ sCV_{in} \\ 0 \\ 0 \end{bmatrix}. \quad (6.113)$$

After solving V_6 from the previous matrix equation using Cramer's rule, the transfer function of the circuit presented in Figure 6.90 is

$$\frac{V_6}{V_{in}} = \frac{\left(\frac{R_2}{R_1} - sRC\right)\left(\frac{R_4}{R_3} - sR_xC_x\right)}{(1 + sRC)(1 + sR_xC_x)}. \quad (6.114)$$

If one chooses to have $R_1 = R_3$, $R_2 = R_4$, $R = R_x$ and $C = C_x$, then

$$\frac{V_6}{V_{\text{in}}} = \frac{\left(\frac{R_2}{R_1} - sRC\right)^2}{(1 + sRC)^2}. \quad (6.115)$$

For ideal op-amps this transfer function gives a unity gain for all frequencies, but the phase shift depends on the frequency of the input signal. The phase shift variation of this two-stage all-pass filter is from 0 to 180 degrees because both capacitors will give a maximum of a 90 degree shift. The total phase shift reaches the maximum 180 degrees at frequency

$$f_p = \frac{1}{2\pi RC}. \quad (6.116)$$

The input signal of the Phase 45 circuit is combined to the phase-shifted signal via 10 kΩ resistors, which generate a direct sum with equal weights. The frequency response curve of this summed signal will then have one 'notch' at frequency f_p . Since the filters are designed to be adjustable via equal resistors R , the frequency f_p of the notch will change as a function of R . The location of the notch in the frequency domain is depicted in Figure 6.91 for a few selected resistance values of R , while the other components are fixed to have the same values as in the phaser circuit 6.83.

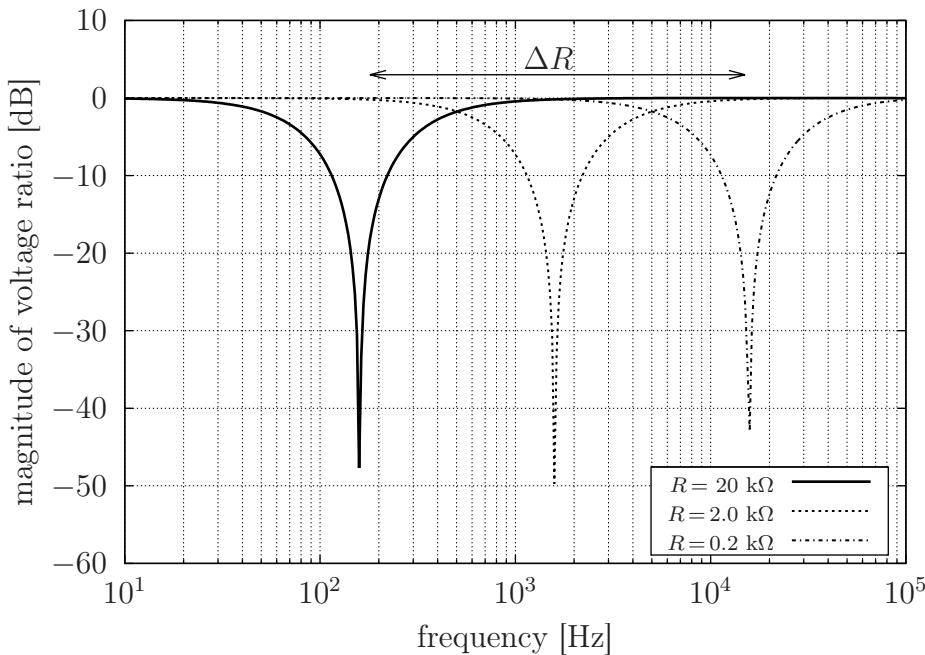


Figure 6.91: The frequency response after summing the input and filtered signals

In this context it should be mentioned that the buffer amplifier XOP1 in front

of the filter stage in the schematic 6.83 is essential to produce evenly balanced notches for all frequencies in the range of operation frequencies. Without the buffer amplifier, the internal impedance of the signal source (guitar pickup + tone control in this case) would have a considerable effect on the frequency response of the sum of signals. The buffering amplifier can be taken away only in the case where the internal impedance of the signal source is 1 for all frequencies, but this is not very realistic for any practical signal source.

Figure 6.92 shows the resistance part of the all-pass filter of the Phase 45 circuit. A proof should be given that the effective resistance acting as R in the all-pass filter is equal to the parallel resistance of JFET internal drain resistance R_D and the series connection of R_1 and R_2 . The capacitor C_x added in the middle of this resistance assembly affects extremely little or not at all the total impedance of the setup.

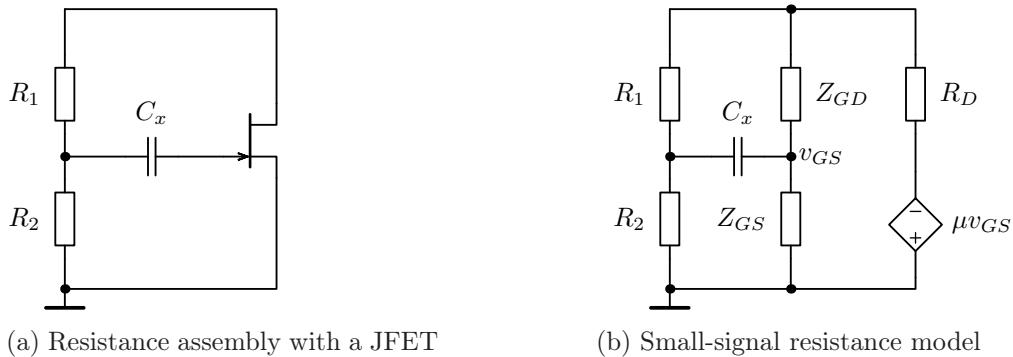


Figure 6.92: The variable resistance circuit extracted from the all-pass filter

From the viewpoint of R_1 and R_2 , C_x is in series with the internal impedances (resistance + capacitance) Z_{GS} and Z_{DS} of the JFET. Impedances Z_{GS} and Z_{DS} are tremendously huge at audio frequencies, which means that the parallel connections of R_1 or R_2 with $C_x + Z_{in}$ have the effective values of R_1 and R_2 . Since the JFET practically does not draw any current, the current through C_x is minimal at all frequencies. From there it follows that the voltage drop across C_x is also minimal at all frequencies. Therefore, C_x does not filter out any input signal frequencies and it does not affect the total series impedance of $R_1 + R_2$.

Since the op-amp in the all-pass filter assembly does not draw any current either, the current through the JFET's drain-source channel is extremely small. This limited current keeps v_{DS} close to zero at all times and enables the possibility to use the JFET as a voltage-controlled resistance.

But what exactly is the purpose of having the capacitor C_x there in the middle of everything? Apparently it is used for transporting part of the alternating voltage to modulate the JFET at the rate that the signal source voltage is varying. This arrangement tries to stabilise the resistance value of the JFET during the low-frequency modulation. One might think it is a clever solution, but it seems to be more like a hack. If the low-frequency modulation was allowed to have a larger peak-to-peak voltage at the gate of the JFET, the resistance would surely change between the lower and upper limits and the phase sweep would extend to a wider range of frequencies. As the resistance of the JFET goes low, the notch in the frequency response will increase to several kilohertz.

As a conclusion to the resistance discussion, the total resistance of the impedance assembly in Figure 6.92 is effectively having the JFET resistance in parallel with the series-connected R_1 and R_2 . The resistance values used for calculating the locations of the notches in Figure 6.91 have been chosen to imitate the limits of the total resistance R . When v_{GS} at the JFET gate is at the pinch-off voltage, the resistance R_D is very high and the total resistance $R = R_1 + R_2$, which is 20 k Ω according to Figure 6.83. When v_{GS} changes a few hundred millivolts towards the source voltage, the resistance of the JFET drops dramatically to about 1 k Ω , as shown by the theoretical analysis in Figure 6.45. The limit of having 200 Ω as the total resistance is just a theoretical minimum and it only illustrates the maximum range where the notch can move. When considering the voltage levels obtained from the relaxation oscillator through the 3.9 M Ω resistor to the gate of the JFET, the total resistance R of the all-pass filters change between 20 k Ω and 2 k Ω . This means that the notch in the frequency response will sweep through an approximate range from 200 Hz to 2000 Hz.

If the name of the Phase 45 effect is not completely misleading, one would assume the number 45 to mean the relative phase shift obtained using the oscillator-controlled variable resistance. Figure 6.93 shows the effect of changing the resistance value within the range of 1 – 10 k Ω . Obviously a constant 45 degree phase shift is not achievable only by changing the value of R . The phase shift is clearly smaller at very low and very high frequencies than it is at the centre frequencies.

It would be relatively simple to extend the Phase 45 to a Phase 90 circuit by adding a second double all-pass filter section at the end of the filter section of

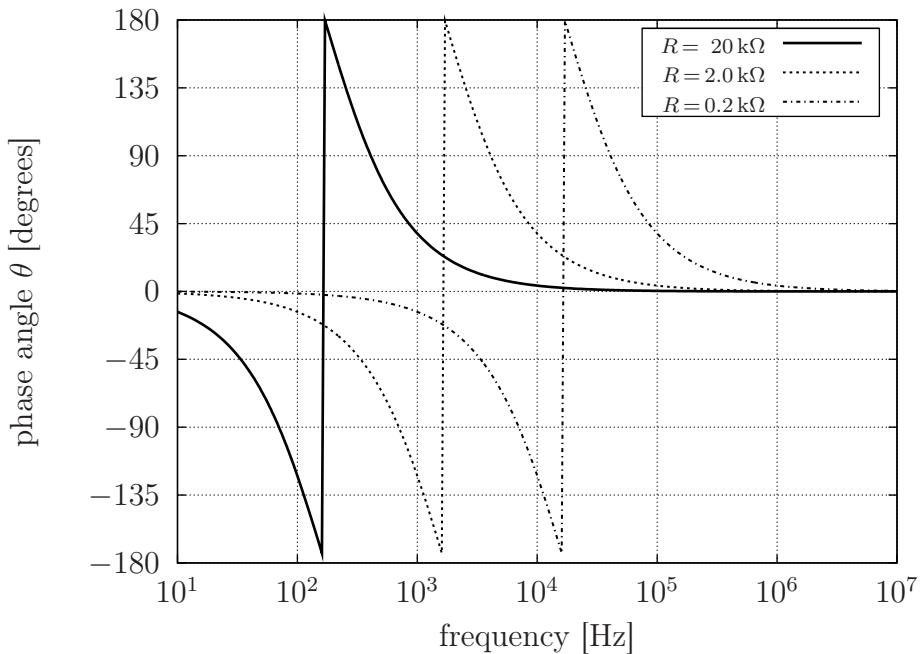


Figure 6.93: Phase shift curves of the Phase 45 effect with different values of R

the Phase 45 circuit. In that case, the transfer function would be

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{\left(\frac{R_2}{R_1} - sRC\right)^4}{(1 + sRC)^4}, \quad (6.117)$$

and the maximum phase shift from the capacitors would reach 360 degrees. In this shifting range the critical 180 degree shift is encountered twice so that there will be two notch points in the frequency response curve at locations

$$f_{p1} = f_p - \frac{f_p}{2} \quad f_{p2} = f_p + \frac{f_p}{2},$$

where f_p is given by equation (6.116). The phase diagram is presented for convenience in Figure 6.94.

6.9.6 A SPICE model for the Phase 45 circuit

The circuit diagram of the Phase 45 effect drawn with the gschem schematic editor was already presented in Figure 6.83. This is the first time in this chapter when operational amplifiers are encountered in the circuit diagram. Unfortunately op-amps create some small problems for SPICE simulations. Although all schematic editors typically contain an op-amp circuit element, the op-amp does not have any built-in model description in SPICE. Therefore, the op-amps are labelled in the schematic as XOP, where the prefix X indicates to SPICE that the component model is given as a sub-circuit block in the netlist file.

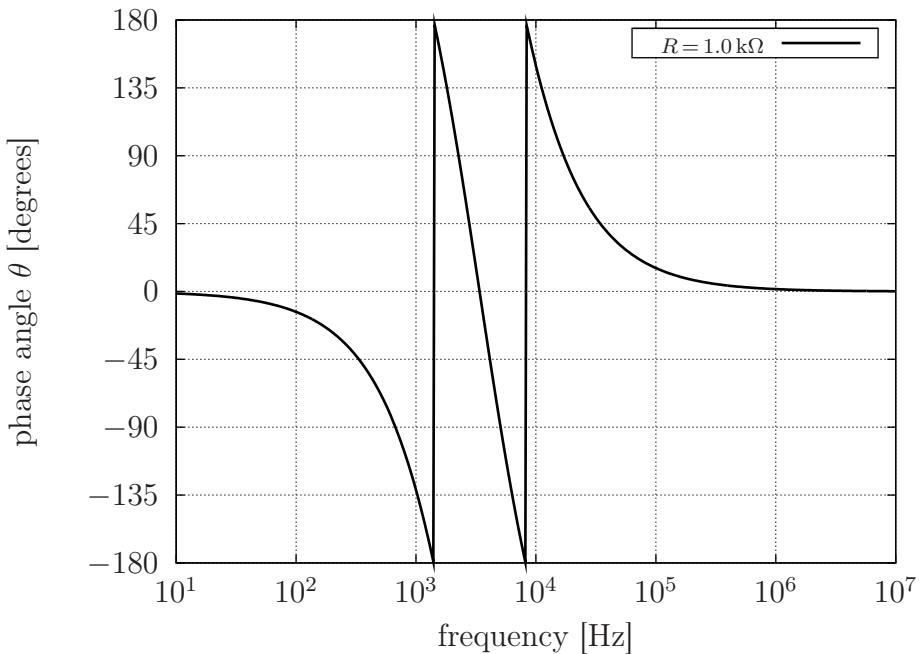


Figure 6.94: A phase shift curve of the Phase 90 effect

The simplest sub-circuit model for the op-amp is the ideal model, where the inputs do not draw any current and the voltage gain is controlled by the voltage between the input pins. This model is depicted in Figure 6.95. The SPICE sub-

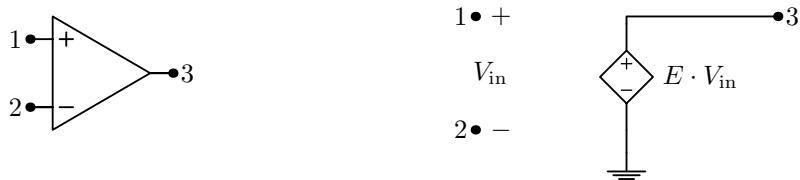


Figure 6.95: An ideal operational amplifier equivalent circuit

circuit macro model for this ideal op-amp circuit is

```
* IDEAL OPAMP MACRO MODEL
* connections:    non-inverting input
*                  |    inverting input
*                  |    |    output
*                  |    |    |
.SUBCKT OPAMP1    1    2    3
*
Iopen1   1 0 0A
Iopen2   2 0 0A
EGAIN   3 0 1 2 1000K
.ENDS
```

If a component in the circuit diagram is labelled as XOP1, the call to the sub-circuit

in the netlist file is done with the line XOP1 0 2 3 OPAMP1, where the tag OPAMP1 defines the specific sub-circuit model to use. The netlist lines starting with letter 'E' (EGAIN in this case) define a voltage-controlled voltage source component in SPICE. The description of this controlled source requires to define four indices: the first two are to define the output nodes and the following two indices define the input nodes. The output voltage depends on the voltage difference at the input nodes, which is amplified by the factor given after the pin connections. In this example, the gain factor is 1000000, which is an approximation of the huge open-loop gain of a typical operational amplifier.

Another similar model that takes into account the theoretical input and output resistances of the op-amp is presented in Figure 6.96. For this op-amp equivalent

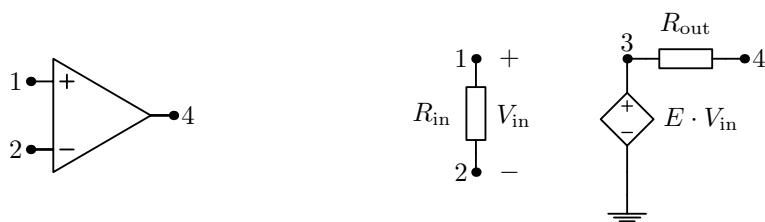


Figure 6.96: An op-amp equivalent circuit with input and output resistances

circuit, the corresponding SPICE sub-circuit model is written as

```
* OPAMP MACRO MODEL
* connections:      non-inverting input
*                  |      inverting input
*                  |      |      output
*                  |      |      |
.SUBCKT OPAMP1      1      2      4
*
RIN    1 2      10000k
EGAIN  3 0 1 2  100K
ROUT   3 4      100
.ENDS
```

These two models are sufficient enough for simulating audio frequency oscillations with limited gains. Problems occur, for example, if the op-amp is used as an oscillator where the output voltage swings between the operating voltages. The models presented in Figures 6.96 and 6.95 have nothing to limit the output voltage to the operating voltages. Instead, the output voltage of this model can swing to infinity as the ideal model of an op-amp suggests.

Luckily component manufacturers have provided tailor-made SPICE sub-circuit models for several types of components, including op-amps. Therefore, it is bet-

ter to just use what has been provided and not start re-inventing the wheel, or the op-amp in this case.

Usually the sub-circuit models are stored as library (.lib) files which have a huge amount of different sub-circuit definitions. It is relatively simple to include these libraries directly to the SPICE netlist file and use the models as they would be written in the actual netlist file. The inclusion of a library file into a SPICE netlist is done with the .include command. For example, the command

```
.include /usr/share/gEDA/models/spice/tex_inst.lib
```

would include all the sub-circuits from a Texas Instruments library file into the netlist. The .include command is normally written to the beginning of a netlist file before the simulation control commands.

The models provided by the manufacturers allow realistic simulations and limit the output voltage so that it will not exceed the operating voltages. The op-amp model provided by Texas Instruments is used in the simulations of the Phase 45 circuit. It might not be easy to find sub-circuit models defined by the component manufacturers because normally the SPICE simulation software packages are not available for free, unlike gEDA and ngspice. Only a few essential libraries can be found from the Internet as 'open source' deliveries.

With the gEDA design tools collection in Linux, the SPICE specific netlist file can be created from the schematic of Figure 6.83 by using the command

```
gnetlist -g spice-sdb -o phaser45_net.net phaser45.sch
```

where the output file named as phaser45_net.net is created from the schematic file phaser45.sch. The somewhat lengthy but yet complete netlist file of the Phase 45 circuit ready for a simulation run is included in Appendix B.

When simulating circuits containing operational amplifiers, the transient analysis of the op-amp output voltage is the most interesting one, so at least that should be included in the control section of the netlist file. The obvious simulation stores the waveforms of the input and output signals to see how the signal changes inside the Phase 45 circuit.

```
.control
tran 10us 990ms 810ms
set filetype=ascii
write phasdata.txt tran1.v(3) tran1.v(20)
gnuplot phas_tran tran1.v(3) tran1.v(20)
.endc
```

Additional analysis could include the AC analysis, but there it should be noted that the analysis is executed using initial values where the JFET is fully closed. This will give the notch close to 200 Hz. A suitable initial voltage should be set to V_{GS} to obtain the frequency response when the JFET is partially open.

6.9.7 Results of simulations, calculations and measurements

The prototype of the Phase 45 circuit shown in Figure 6.83 was built on a solderless breadboard using LM224J to replace the two TL071 op-amp components of the original schematic and BF245A JFETs to replace the 2N5457 JFETs. To save extra costs, some resistor values were approximated to use some resistors that were already used in previous projects. These approximations and their effects are explained later in this section.

The netlist prepared for SPICE simulations from circuit 6.83 was loaded into ngspice and analysed with the operating point (.OP) command to obtain the DC values to be compared with the measured values. The calculations based on theoretically derived equations were not done for this circuit, because the static configuration of the circuit is explained in section 6.9.4. The data of the simulated and measured DC values is gathered to Table 6.6 where it can be easily compared for differences.

The measured and simulated DC values differ significantly in absolute value because the virtual ground was measured to be unexpectedly low in the prototype. The values are a bit more equal if the results are compared relative to the virtual ground level. The value 4.67 V is repeated many times in the simulated results, and it can be taken as the virtual ground voltage for the circuit, about half of the full operating voltage. The corresponding virtual ground for the prototype circuit seems to be 3.86 V. Possibly the zener diode had some deviation from the nominal value, which might have caused the observed difference in the virtual ground voltage between the real world prototype and the ideal simulated model. The high input impedance of the JFETs and the op-amps might also cause the deviation.

In the transient simulations, the first thing to verify is that the phase shift obtained from the all-pass filter at some specified frequency of the input signal agrees with the calculated shift curves of Figure 6.93. The input signal was chosen to be a 100 Hz sine wave (close to the open G-string in a bass guitar), because with this frequency the visualisation of the effect is relatively simple. Figure 6.97

Table 6.6: Phase 45 circuit DC values

value	simulated	calculated	measured
V_{D0}	4.67	-	3.86
V_{G1}	2.87	-	2.15
V_{D1}	4.67	-	3.86
V_{S1}	4.67	-	3.80
V_{G2}	2.87	-	2.82
V_{D2}	4.67	-	3.86
V_{S2}	4.67	-	3.80
V_{10}	4.67	-	3.82
V_{21}	2.87	-	1.14
V_5	4.67	-	2.60
V_7	4.67	-	3.86
V_8	4.67	-	3.86
V_{12}	4.67	-	3.86

shows the phase shift of the Phase 45 circuit in the situation where the resistance of the JFETs change from maximum resistance to lower resistance. In this case the total resistance R in the filter sections is about $20\text{ k}\Omega$ at the beginning and it changes roughly to $1\text{ k}\Omega$ towards the end.

The theoretical phase shifts predicted by Figure 6.93 are approximately -120 degrees for the $20\text{ k}\Omega$ resistance and it approaches zero degrees with lower values of the resistance. These phase shift predictions are in good agreement with the simulated results in Figure 6.97, where the dashed curve presents the voltage measured from node 18, which is identified in Figure 6.83. In this perspective the theoretical transfer functions derived in section 6.9.5 are valid.

The results of the time domain transient simulation in Figure 6.98 show the phase shift of the output signal at node 20 against the input signal. This situation is a little bit different compared to Figure 6.97, since the output signal is a sum of the input signal with the phase-shifted signal measured from node 18.

The simulation results indicate that the sum of the input signal with the phase-shifted signal will in general attenuate the actual output signal by some amount depending on the phase shift obtained in each frequency. The relative phase shift

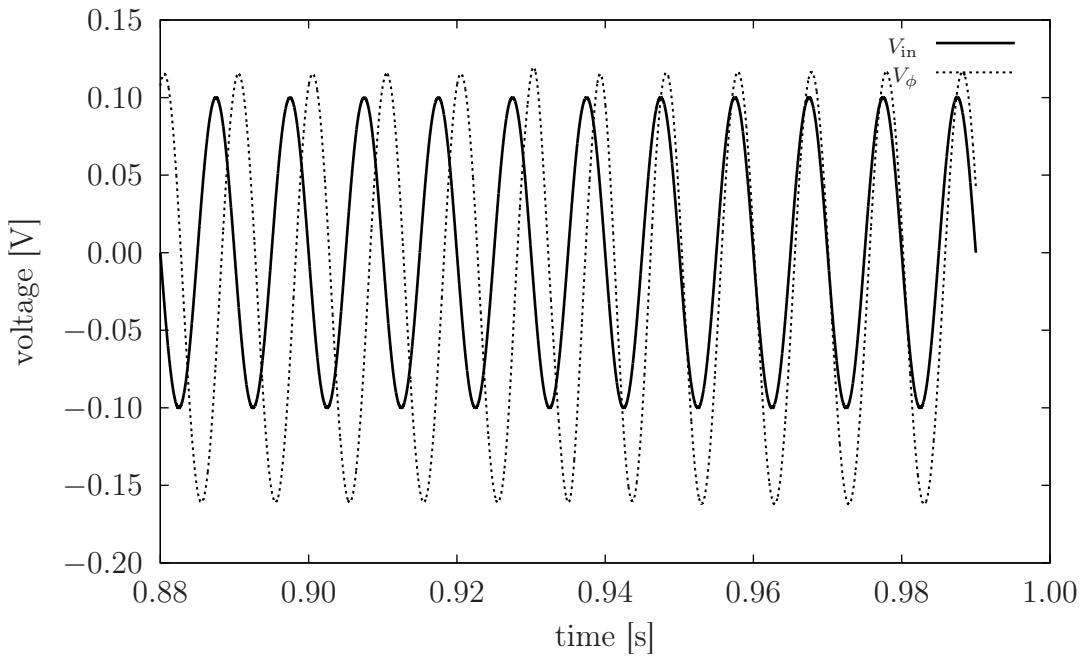


Figure 6.97: The input vs. the phase-shifted signal of the Phase 45 circuit

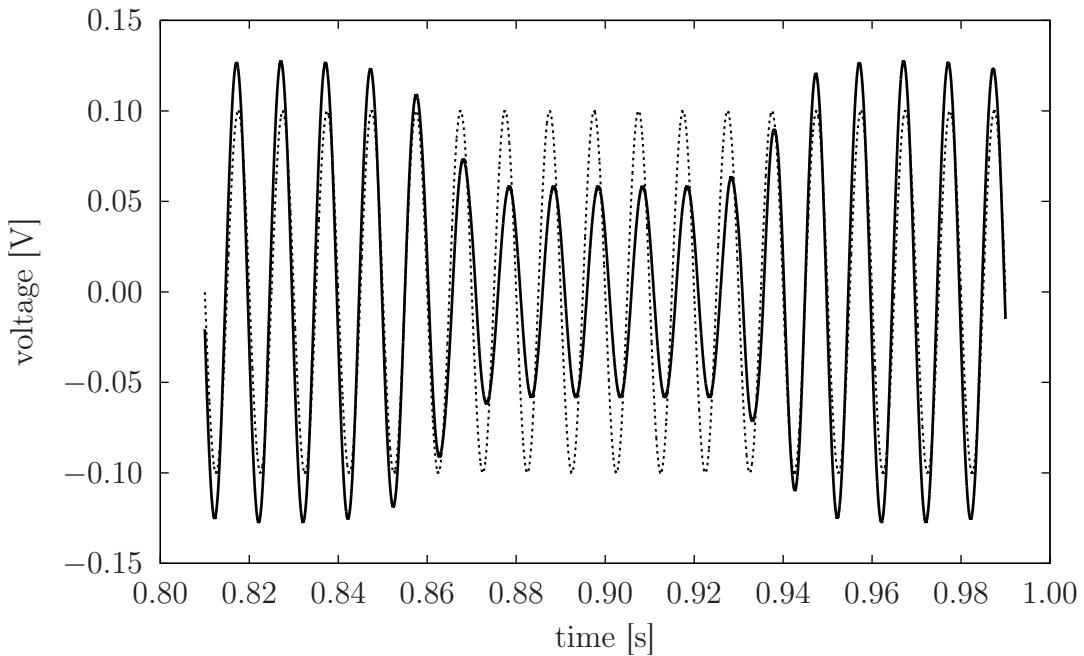


Figure 6.98: The input signal vs. the output signal of the Phase 45 circuit

of the output signal as the JFET resistance moves from maximum resistance to lower resistance should be equal to the relative phase shift indicated by Figure 6.93. In this case it should be about -110 degrees. This verifies at least the fact that the name 'Phase 45' does not refer to a 45 degree phase shift, since the phase shift obtained from the circuit varies for each frequency and in general is more

than 45 degrees.

In the prototype measurements, a computer sound card was used as a signal source, from where a 100 Hz sine wave was played out. This frequency was chosen because it is the same test signal as in the simulations. The input and output signals were measured with a digital oscilloscope. Two separate measurements were made, where in the first measurement the output waveform had a maximum negative shift to the left side of the input signal, and the second output waveform had a maximum positive shift to the right. The results are plotted in Figure 6.99. From Figure 6.99 it is evident that the circuit actually works as

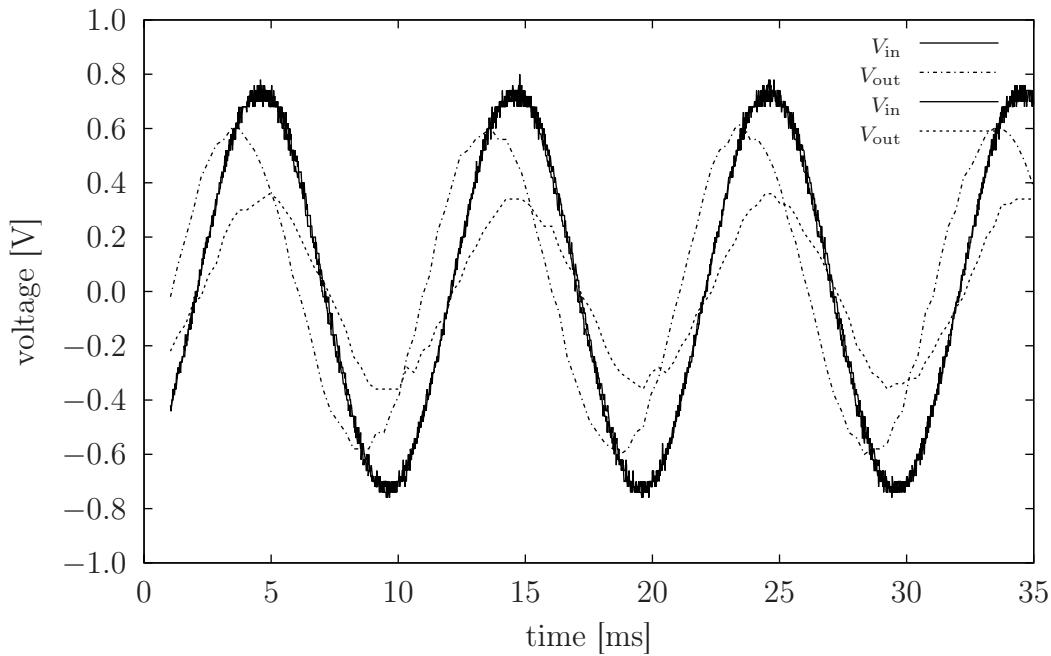


Figure 6.99: Measured input and output voltage waveforms of the Phase 45 effect

expected and the functionality is also similar to the simulations. Although the visualisation is in a little bit different form, this plot from the measurements can be compared to Figure 6.98 to see that the output signal moves with the same shift as in the simulations. Roughly it does work the same way. The main thing, however, is that the effect sounded good; all the fancy curves are secondary in contrast to the actual tones obtained from the effect.

Considering the prototype circuit as a moving notch filter, the low-frequency limit was determined by scanning a frequency where the maximum negative phase shift to the left completely attenuated the signal to zero amplitude. This frequency was measured to be about 200 Hz. The maximum positive phase shift to the right completely attenuated the signal at about 2000 Hz, indicating the

high-frequency limit of the notch filter. The theory predicts limits of 160 Hz and 2200 Hz, so the prototype measured close to the theoretical values and was therefore working as expected.

To draw a better picture from the measurements while still retaining comparability to Figure 6.98, the overview on the input and output curves in Figure 6.100 shows that although the phase shift seems to be similar to the simulated reference, the output signal does not change its amplitude as much as in the simulations. This indicates that the phase shift is not reaching the full theoretical range in the prototype circuit.

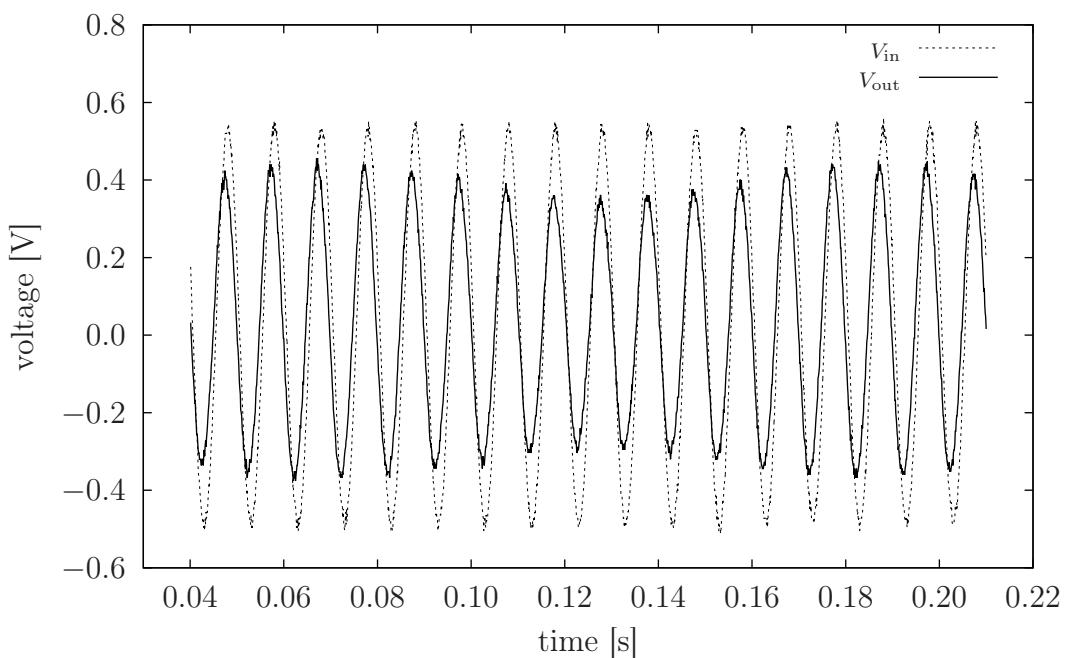


Figure 6.100: A measurement on the general output signal waveform

This lag of efficiency leads to a discussion about the JFETs. For the Phase 45 circuit to give maximum shifting properties, the JFETs should be handpicked to be a matched pair. This means that the drain current I_{DSS} at zero gate voltage and the pinch-off voltage V_P should be as close to equal as possible in the filter section of the Phase 45 circuit. To truly get matched FETs, it is necessary to measure I_{DSS} with a few different operating voltages (V_{DD}) and pick the best matches from a group of measurements. To determine V_P it would be good to measure V_{GS} with several different resistors to get a set of points. To this set of measured V_{GS} values, one should fit the parabola in the form of equation (6.36), because the measurements are made in the saturation region. The essential thing is that the matching is done with respect to several values. Matching to one

measured value, as suggested by Keen [87], usually works fine but does not yet guarantee a fool proof match.

It is also essential to cover the signal connection between the voltage at the relaxation oscillator and the gate-to-source voltage of the JFETs. For starters, Figure 6.100 depicts the measured voltage waveforms from the input and output of the oscillator. At the oscillator output, the voltage waveform is a square wave with a

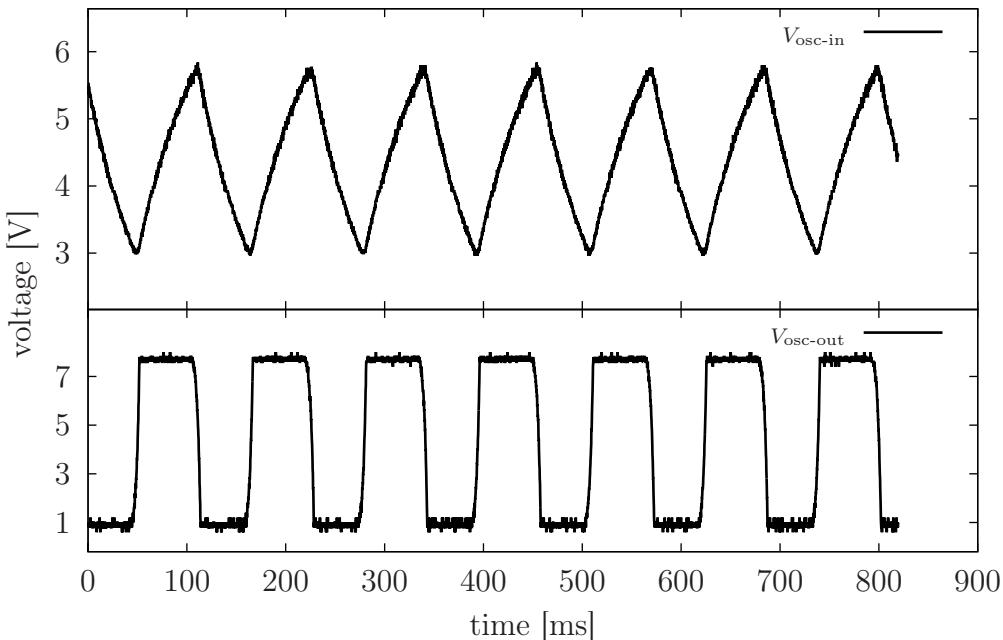


Figure 6.101: Measured oscillator input and output waveforms

peak-to-peak value close to the operating voltage of the op-amp. It depends on the op-amp type as to how close to the 'rails' the output voltage can swing.

The input voltage of the oscillator shows how the capacitor C_9 sequentially charges up and then discharges. This triangular voltage at the node 23 is fed through the voltage divider formed by the $3.9\text{ M}\Omega$ and $1\text{ M}\Omega$ resistors. The capacitor C_6 at the voltage divider filters out the highest frequencies, so the voltage eventually going through the $470\text{ k}\Omega$ resistors to the gate of the JFETs looks almost like a sine wave with a relatively small peak-to-peak voltage of about 200 mV in this specific measurement.

The use of the trimmer potentiometer R_{18} seems an awkward way to adjust the reference voltage to the hot spot where the effect starts working. This potentiometer sets the DC voltage on top of which the 200 mV signal from the oscillator is swinging. As the pinch-off voltage of the BF245A was measured to be

approximately -1.7 V, the voltage at the gate of the JFETs should swing from -1.7 V to -1.5 V, assuming the 200 mV alternating control voltage from the oscillator. With these specifications as a reference to Figure 6.45, the resistance of the JFETs would change from ∞ to about $2\text{ k}\Omega$ (theoretically). Figure 6.102 depicts the measured voltage at the JFET gate as the potentiometer R_{18} is adjusted to its minimum and maximum values. Somewhere between these limiting values a certain DC voltage is found that makes the circuit rock.

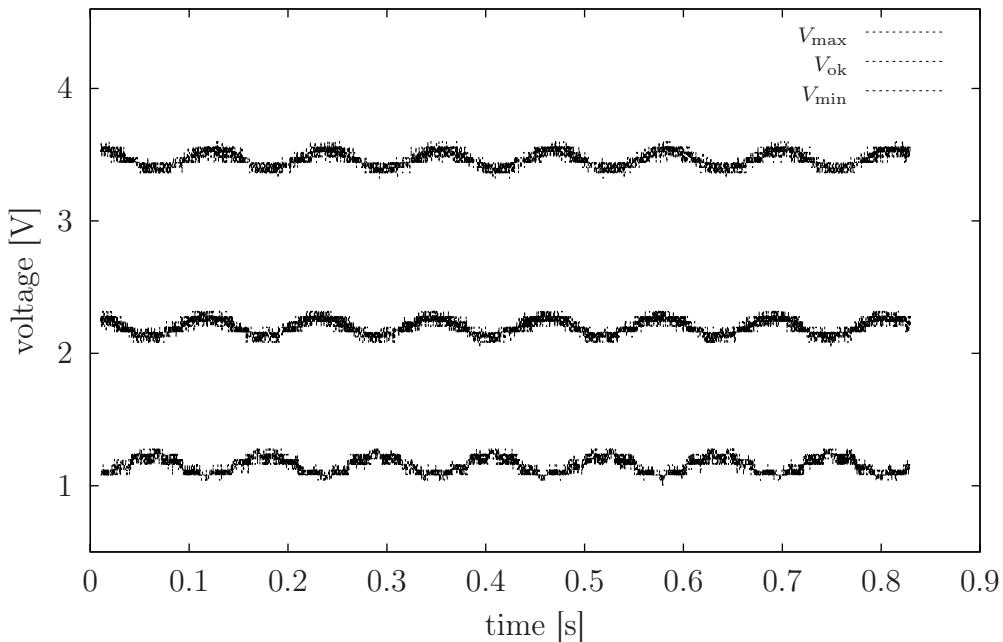


Figure 6.102: Potentiometer-adjusted voltage levels at the JFET gate

The voltage difference between the maximum DC voltage in Figure 6.102 (which should equal the virtual ground level), and the voltage with which the effect works, should approximately equal the pinch-off voltage of the used JFETs.

Because the maximum range of DC voltage values in the scope of potentiometer R_{18} is only $3.8\text{ V} - 1.0\text{ V} = 2.8\text{ V}$, it is clear that the BF245B JFET with the measured V_P of -3.6 V would have not worked in this circuit. To be able to use JFETs with higher pinch-off voltages in this Phase 45 circuit, the zener diode should have a bigger value. For example, a 5.6 V zener could be used to give a wider range of DC voltage adjustability via R_{18} .

The relaxation oscillator was studied as a separate circuit because a lot of effort was wasted in deriving all those theoretical formulae in section 6.9.1. It is interesting to see how well equation (6.95) predicts T , the period of oscillation. Only

a few measurements were made on the oscillation period, mainly using the maximum and minimum setting of R_{25} , which in the prototype circuit was a $470\text{ k}\Omega$ potentiometer. The measurement results are documented in Table 6.7 beside the calculated values from equation (6.95) with $R = R_{25} + R_{26}$.

Table 6.7: Measured relaxation frequencies against analytical estimates

$R_{25} + R_{26}$ [kΩ]	$T_{\text{calculated}}$ [s]	T_{measured} [s]
$R_{\max} = 470 + 6.8$	5.72	5.84
$R_{\min} = 0.0 + 6.8$	0.094	0.116*

The results need some explanations. The peak-to-peak output voltage with the largest resistance value was measured to be 8.1 V, V_{limL} was measured as 3.0 V, and V_{limH} was measured as 5.3 V. These were the values used for calculating the theoretical frequency of oscillation in the case of maximum resistance R_{\max} . The minimum resistance changed the limiting voltages a bit, the output voltage being approximately 7.5 V, $V_{\text{limL}} \approx 3.0$ V and $V_{\text{limH}} \approx 5.3$ V. These values were used for calculating the case of minimum resistance R_{\min} .

The relative error in the minimum resistance case is clearly too high. The reason is depicted in Figure 6.103. The capacitor C_{10} was noticed to be more harmful

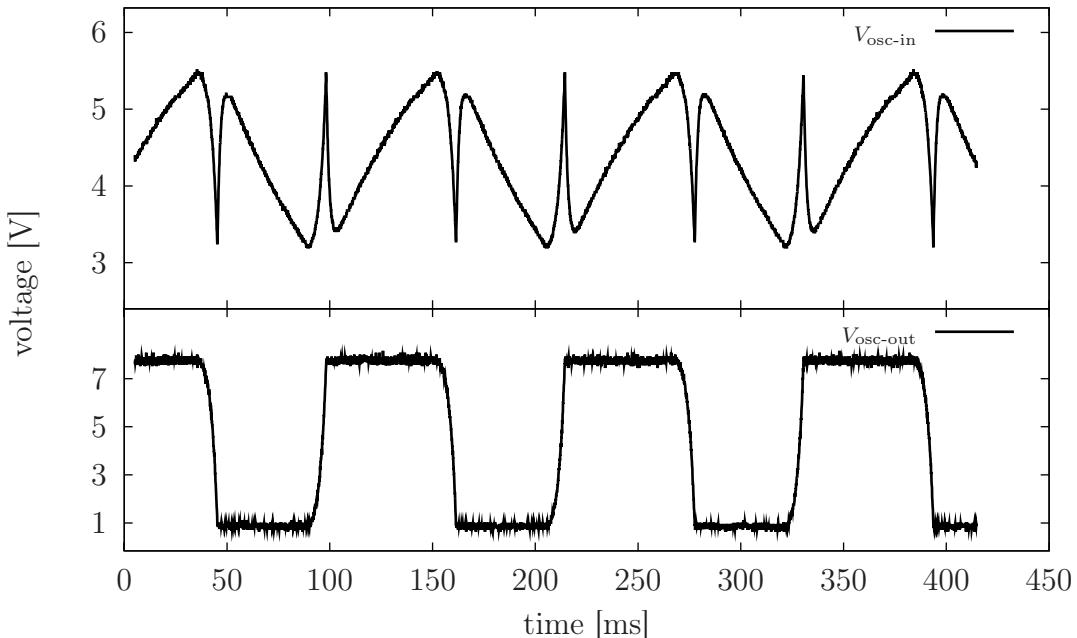


Figure 6.103: Transient spikes because of capacitor C_{10}

than useful. The capacitive coupling with the output signal at higher frequencies messed up the behaviour of the oscillator and caused error to the period of oscillation. When C_{10} was removed, the obscure spikes disappeared and the oscillation period was measured to be 0.092 seconds. This result is in excellent agreement with the theoretical result.

Because of the unexpected behaviour of capacitor C_{10} , it was not possible to study the effect of the added components on the oscillation periods of the basic relaxation oscillator of Figure 6.84a. But according to the calculations, the difference would have been very small and unnoticeable in practical situations.

According to this testing and analysis, it would be better to remove resistor R_{23} and capacitor C_{10} from the oscillator solution of the Phase 45 circuit. In addition, if someone wants to save money, the circuit would also work without capacitor C_6 , which is only used to make the signal from the oscillator more round by filtering out the higher harmonics. This does not affect much the overall functionality of the circuit. The capacitors C_5 and C_8 could also be removed if the signal coming from the oscillator has a peak-to-peak voltage close to 0.5 V. The signal level of the modulating oscillator voltage can be increased by replacing the $3.9\text{ M}\Omega$ resistor by a smaller one, but then one needs to make sure that the $1\text{ M}\Omega$ resistor is still in the correct range.

As a conclusion, a phase shifter effect consists of many interesting scientific details, which strongly relate to many other technical applications. It is extremely useful to go through the details of the circuit and eventually make own modifications. Phase shifters are the favourite effects of many guitarists, so to build the Phase 45 circuit as a do-it-yourself project is highly recommended.

6.10 A FLANGER

To be able to build a flanger effect pedal requires much more than basic discrete analogue components can offer. Special integrated circuits are needed to produce enough delay to create a true flanging effect, although in principle it could be possible to imitate the functionality of the integrated circuit by several hundred discrete analogue components. Due to the lack of 'all-analogue' properties, the flanger is not covered from a scientific point of view. Only the principles of operation are described by words and a little mathematics.

Figure 6.104 describes the signal flow inside a typical flanger effect unit. Firstly,

the input signal is split into two separate propagation paths just like in the phase shifter effect unit. The first signal path leads directly to the output, and the second path goes through a delay line, where the delay can be controlled with external modulation. The delayed signal is summed to the clean signal just before the output.

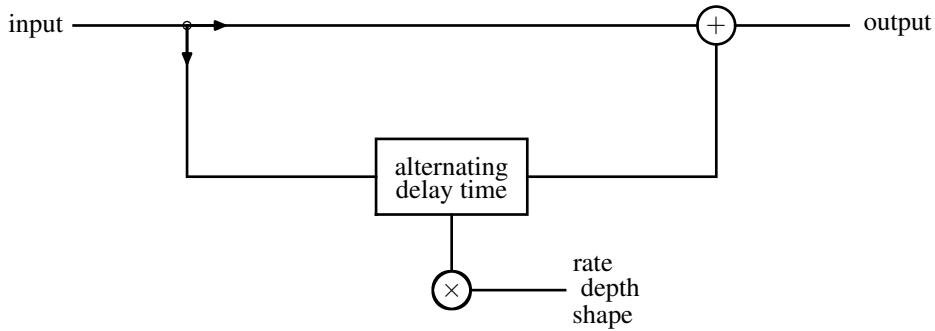


Figure 6.104: A signal flow graph at the core of a typical flanger effect

The flanger would not be a flanger without the possibility to modulate the rate, depth and shape of the delay acting on the signal. In the flanger effect, the delay time is constantly changing and the rate controls how quickly the delay time is changed between its limiting values. The depth parameter controls the range of the delay time limits, and the shape indicates the modulation (sine, square, triangle, etc.) how the delay time oscillates within its limits. [66, p. 211]

The essential signal shaping mechanism of the flanger effect is also good to understand in a theoretical level through a simplified example where the input signal to the flanger is a set of harmonic frequencies with equal or arbitrary amplitudes. This theoretical test signal can be, for example, the signal obtained from a vibrating string. The functional description of the signal is simplified here to the form

$$y(t) = \sum_{n=1}^N \frac{1}{n^2} \sin(n2\pi f_0 t), \quad (6.118)$$

where f_0 defines the fundamental frequency and n is the ordinal number of the harmonic overtone. In this trigonometric series, the fundamental frequency has the largest amplitude and others are lower by the square of the harmonic number, just like in the case of the vibrating string.

It is absolutely necessary to highlight the difference between a flanger and a phase shifter. With a simple two-stage phase shifter one can adjust the maximum 180 degree phase shift to only one frequency at a time. As a concrete example,

one could have a signal as defined by equation (6.118) with a fundamental frequency $f_0 = 50$ Hz and one would design the single notch in a phase shifter to cancel out the 20th harmonic frequency, which is 1000 Hz in this specific case. According to the transfer function of the phase shifter, the frequency response would have only one downward directed spike of a 'comb filter' at 1000 Hz.

The same case is more versatile when implemented with a flanger. Now one approaches the problem from the viewpoint of a time delay. The question is that how much delay is needed to shift the 1000 Hz frequency component 180 degrees out of phase? The answer is 0.5 ms. The phase shifter will give this same delay at 1000 Hz. The essential difference to the phaser is that the time delay of the flanger equally affects all frequency components in the signal. Therefore, in the theoretical case where the test signal consists of only harmonic frequencies, all the frequencies that are harmonic to the first cancelled frequency (1000 Hz in this example) are also cancelled. This way the comb filter transfer function created by the flanger is much more complex compared to the phase shifter. Figure 6.105 depicts the example situation where the set of harmonic frequencies with $f_0 = 1000$ Hz is delayed by 0.5 ms and then summed to the original signal.

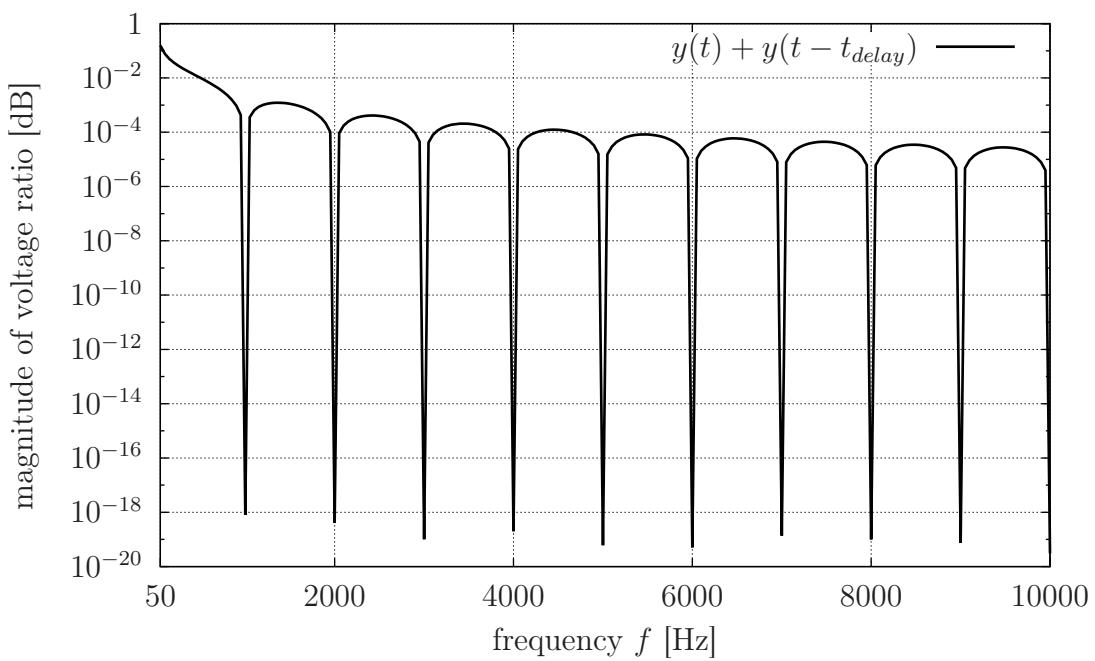


Figure 6.105: A constant-delay flanger effect applied to a theoretical test signal

The comb structure in this case is equally spaced. In the general case where the input signal can be a mixture of any frequencies, the comb structure will become extremely complex and diverse.

The actual flanger effect does not keep the time delay constant. Only frequent changes in the delay time create the hearing sensation that is commonly recognised as the flanger. In the theoretical case where the input signal is defined by equation (6.118), the time delay needed to cancel a certain harmonic frequency component n and all its integer multiple frequencies can be calculated via the formula

$$t_{\text{delay}} = \frac{1}{2nf_0}, \quad (6.119)$$

where f_0 is the fundamental frequency.

6.11 TIME EFFECTS

The effects in the time category rely on the power of mixing delayed signals with the original signal to generate different effect types such as echo, reverberation, chorus and flanger effects. To distinguish different effect types from each other, the length of the delay is used as a parameter to categorise the effect into different subgroups. A rough division into short and long delays assigns short delays to be shorter than 15 ms, and if the delay is longer than 55 ms, the delay is considered to be long. Table 6.8 summarises all the necessary information about different delay types and related effects.

Table 6.8: A categorisation of delay effects with respect to delay time

delay type	delay length	effect type
short delay	< 15 ms	flanger, pitch shift
medium delay	15 ms – 55 ms	chorus, pitch shift
long delay	> 55 ms	echo, reverberation

The different delay effect types definitely need careful explanation because the differences in some cases are diminishingly small. Let's start from the group of **long delays**, which consists of echo and reverberation effects. Echo can be distinguished as a discrete repeating pattern of a delayed sound. The sounds that work best when echoed are short impulsive type sounds repeated at relatively long intervals. Reverberation is a more general echo of several continuous sounds in a certain type of controlled environment, such as a room or concert hall. In these two effects the delayed signal is added to the original signal so that it is still recognisable as a copy of the original sound pattern. By changing the delay time, reverberation in different kind of environments can be simulated. The

theoretical means to create the reverberation effect is presented in section 6.13 and an actual analogue reverberation effect device is analysed in section 6.15.

In the group of **medium length** delay effects, chorus refers to the slightly asynchronous sound of two simultaneously playing instruments. This effect can also be referred to as double tracking in a studio recording session. By summing a single guitar tone to itself with e.g. a 30 ms delay, it sounds like two guitarists are playing the same tune simultaneously. Furthermore, adding the same tone with several different medium length delays creates the sensation that several guitars are playing simultaneously.

Flangers are more elegant **short length** effect devices as far as mathematics is concerned. Flangers take advantage of the properties of the constructive and destructive interference of signals, where the interference comes from summing two or more signals. Since any type of continuous signal can be constructed as a sum of sine waves, every sinusoidal component interacts with the delayed components, thereby creating a comb filter to boost certain frequencies and cancel some of the others. The flanger is therefore a moving comb filter that continuously alters the delay in the interval of a short delay time, cancelling and boosting different frequencies as it sweeps through the range. This effect creates a hearing sensation that is best described as a whooshing or a ringing sound. A more specific analysis on the flanger effect is given in section 6.10. [66, pp. 231 – 234]

Pitch shifting happens when the delay time is modulated to continuously change its length. Since flangers and chorus effects are typically based on modulated delay, those effects have the pitch shifting property as a side effect. The pitch shift and the modulated delay have connections to a mathematical method called *correlation*, more accurately autocorrelation and cross-correlation. The theoretical basis of the pitch shift effect is opened up more extensively in section 6.12.

6.12 THEORETICAL PITCH SHIFT

Figure 6.106 illustrates how a constantly increasing time delay causes a downward shift in the pitch of a note. Having a delay ratio of 2:1, the delayed sine wave on the lower part of the figure has a period twice as long as the reference sine wave. The nodes identified with letters *a*, *b*, *c*, *d* and *e* are used for indicating a few strategical locations within one period of a clean sine wave in both waveforms in Figure 6.106. Then all that is needed is to pay attention to the amount of delay between the marked points in the two sine waves. This is what

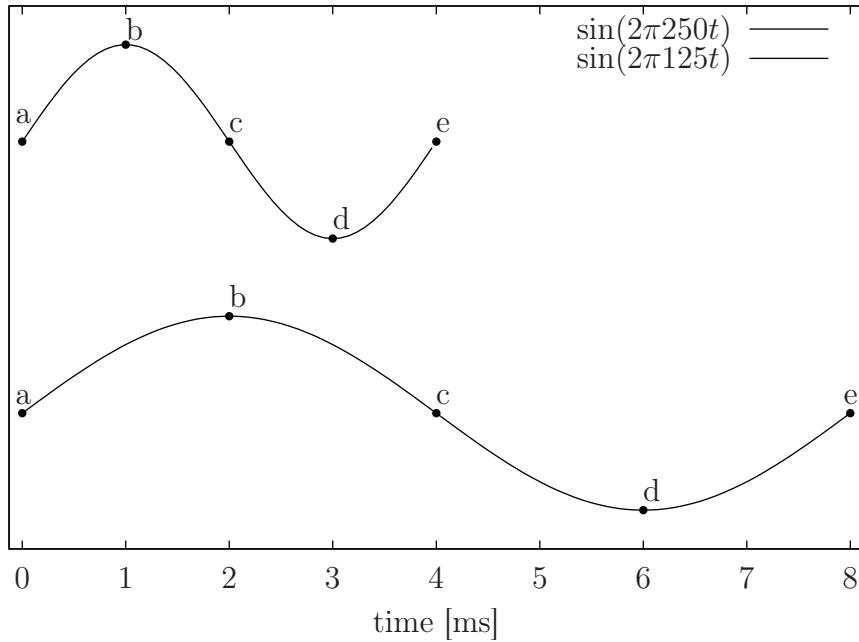


Figure 6.106: Pitch shift

is noticed on the delay for each point:

$$a \rightarrow a = 0 \text{ ms}$$

$$b \rightarrow b = 1 \text{ ms}$$

$$c \rightarrow c = 2 \text{ ms}$$

$$d \rightarrow d = 3 \text{ ms}$$

$$e \rightarrow e = 4 \text{ ms}$$

The delay is not constant, but the rate of change of the delay between two neighbouring points stays constant. In general mathematical terms this could be expressed as

$$t_{\text{total}} = t_{\text{normal}} + t_{\text{delay}}, \quad (6.120)$$

which means that the total playback time t_{total} consists of the playback time t_{normal} of normal speed playback and the delay time t_{delay} . Equation (6.120) can also be expressed as ratios to some reference unit of time. Then one is dealing with playback rates defined as

$$\frac{\Delta t_{\text{total}}}{\Delta t} = \frac{\Delta t_{\text{normal}}}{\Delta t} + \frac{\Delta t_{\text{delay}}}{\Delta t}, \quad (6.121)$$

which is a bit more understandable from the practical point of view. As an example, a reference unit Δt of 1 ms is adopted and the delay time is chosen as $\frac{1 \text{ ms}}{1 \text{ ms}}$.

Then from equation (6.121) the total playback time would be

$$\frac{1 \text{ ms}}{1 \text{ ms}} + \frac{1 \text{ ms}}{1 \text{ ms}} = \frac{2 \text{ ms}}{1 \text{ ms}},$$

meaning that with the delay added, it takes 2 milliseconds to advance the 'tape' by 1 millisecond. According to this evaluation, the total playback time will be doubled if the normal playback is played with unity ratio, and the constantly increasing delay will be equal to the reference unit. This also means that all the tones in the playback will have their pitch lowered by one octave. So how much delay is needed to lower the pitch by two octaves? In that case the total playback ratio will be 4 to 1 and from equation (6.121)

$$\frac{\Delta t_{\text{delay}}}{\Delta t} = \frac{\Delta t_{\text{total}}}{\Delta t} - \frac{\Delta t_{\text{normal}}}{\Delta t} = \frac{4 \text{ ms}}{1 \text{ ms}} - \frac{1 \text{ ms}}{1 \text{ ms}} = \frac{3 \text{ ms}}{1 \text{ ms}},$$

meaning that the delay ratio should be 3 to 1, if the normal playback is done with a 1 to 1 scaling.

As a general result, to decrease the pitch of all audio by controlling the playback speed the delay should have a constant increase by a fraction of some reference unit of time. In other words, the total delay with respect to some fixed reference point will grow continuously. To increase the pitch, the delay should have a constant decrease by a fraction of some reference unit of time, meaning that the rate of delay in equation (6.121) has a negative value. What all this confusing definition hides behind it is that if an audio tape is played with slower speed, the pitch will drop, and vice versa for higher speed playback. The rate of change of the delay determines a factor, which scales the pitch of each frequency in equal manner. [66, pp. 250 – 253]

6.13 A THEORETICAL REVERBERATION EFFECT

Three essential parameters that define the reverberation texture are *reverberation time*, *bass ratio* and *predelay*. The reverberation time indicates the duration of a decaying sound of reverberation. The bass ratio compares the decay times of low-frequency components to middle frequency components. The predelay is the time difference between directly arrived sound and the first arrival of reflected sound.

Convolution is often introduced in scientific literature in the context of several physical phenomena, but yet there does not seem to be a distinct application that directly yields results based on the convolution calculation. Luckily reverberation

happens to give a clear meaning to convolution, and by using a set of impulse functions it is possible to theoretically create a simple reverberation effect.

The discrete convolution of functions $f(n)$ and $g(n)$ is defined as

$$h(n) = f * g = \sum_{m=0}^n f(m)g(n-m). \quad (6.122)$$

The calculation process of the convolution function can be made more understandable by creating a table with the corresponding indices of f and g that are added together as $h(n)$.

Table 6.9: A grid of convolution components

	$f_0 g_0$	$f_0 g_1$	$f_0 g_2$	$f_0 g_3$
+	...	$f_1 g_0$	$f_1 g_1$	$f_1 g_2$	$f_1 g_3$
+	$f_2 g_0$	$f_2 g_1$	$f_2 g_2$	$f_2 g_3$...
+	$f_3 g_0$	$f_3 g_1$	$f_3 g_2$	$f_3 g_3$
=	$h(0)$	$h(1)$	$h(2)$	$h(3)$	$h(4)$	$h(5)$	$h(6)$

Table 6.9 is based on data sets $f = [f_0, f_1, f_2, f_3]$ and $g = [g_0, g_1, g_2, g_3]$. When the convolution is used for creating reverberation, the other data set will contain one or more (unit) impulse points. By setting $f = [1, 0, 0, 0]$, the result of the convolution against g will be $h = [g_0, g_1, g_2, g_3, 0, 0]$, which equals g . If $f = [0, 1, 0, 0]$, then $h = [0, g_0, g_1, g_2, g_3, 0]$, which means that g has been delayed by one data step. Reverberation is created by setting $f = [1, 0, 1, 0, 1]$, because then delayed copies of the original signal are summed to the clean signal just like indicated in Table 6.10.

Table 6.10: Reverberation effect via convolution

	g_0	g_1	g_2	g_3
+	...	0	0	0	0
+	g_0	g_1	g_2	g_3
+	0	0	0	0	...
+	g_0	g_1	g_2	g_3
=	$h(0)$	$h(1)$	$h(2)$	$h(3)$	$h(4)$	$h(5)$	$h(6)$	$h(7)$

The delay between reverberation components can be controlled by adding more zeros between the (unit) impulses. The unit impulses need not to equal one, since more realistic reverberation is achieved by slightly attenuating the reverberation components as time passes.

Creating a reverberation theoretically using the convolution also suggests that reverberation can be considered a set of echoes that are generated recursively [8, p. 167]. However, this kind of definition is in conflict with some definitions that characterise the difference between reverberation and echo.

6.14 ANALOGUE DELAY ACTUATORS

Delay and reverberation types of effect devices can be built using analogue tapes, mechanical springs, mechanical plates and so-called Bucket Brigade devices (BBD) to generate the desired type of delay to the signal. [65, p. 144][66, Ch. 9]

A delay line built by using analogue tape is useful because it can be easily modified from some specific home stereo tape recorders as described by Anderton [42, pp. 95–97]. All you need is a three-head cassette deck (to play and record simultaneously) and a few simple mixer circuits – and the tape delay unit is ready for use. Tape delays are most suitable for implementing the echo effect.

The Bucket Brigade device is fading the limits of analogue and digital electronics. It creates a delay line from several MOSFET - capacitor pairs (see Figure 6.107), where the signal moves from capacitor to capacitor at a speed determined by an external clock circuit feeding the Bucket Brigade. The capacitors are used as sample-and-hold devices, which at each clock pulse quickly charge up (or down) to the voltage level of the preceding capacitor. When the MOSFETs are closed, their resistance is so large that the capacitors will not discharge between the clock pulses.

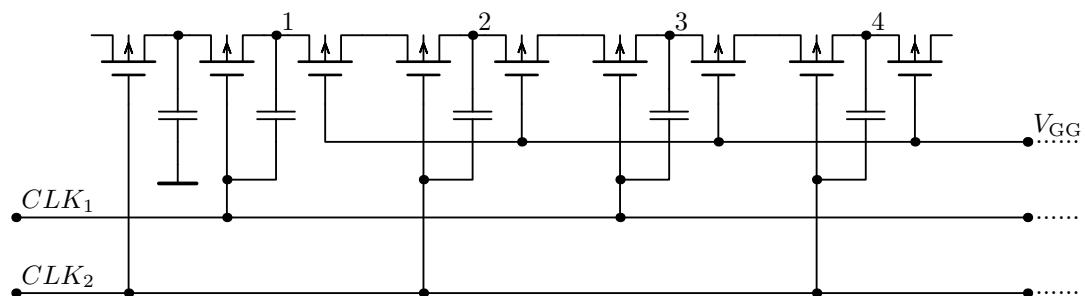


Figure 6.107: The start sequence of a Bucket Brigade Device

Bucket Brigade devices are most suitable for generating chorus, flanger and reverberation effects. The practical BB-devices are designed to operate within a range of clock frequencies limited by maximum f_{\max} and minimum f_{\min} frequency values. The maximum time delay given by a BBD is determined by the number of stages N and the minimum clock frequency with equation

$$T_{\max} = \frac{N}{2f_{\min}}, \quad (6.123)$$

and likewise the minimum time delay

$$T_{\min} = \frac{N}{2f_{\max}}. \quad (6.124)$$

Springs and plates offer a mechanical propagation path, where the propagation time is related to the physical dimensions of a spring or a plate. Mechanical wave propagation in solid objects is always slower than electrical wave propagation in wires. By using a two-way electromechanical transducer, the electric signal from a guitar can be bypassed to a mechanical propagation path and the same signal can be transformed later back to the electrical domain. Between these two transforms, the signal has been delayed in comparison to the directly travelled electrical signal, and a sum of these two signals creates the desired effect. Because the image of a signal at some point in time will 'stay alive' in the spring and plate for quite long before decaying, these delay mechanisms are only suitable for reverberation effects.

Even though the easiest way to build an analogue delay effect would be to use a BBD or tape delay, it is more convenient to choose the spring reverberation device to be taken under detailed investigation in the following section as an example of a time-based effect. The idea from a technical point of view is fascinating, because the implementation mixes electronics and mechanics in a very ingenious way.

6.15 A REVERBERATION EFFECT USING MECHANICAL SPRINGS

The ultimate analogue version of a reverberation device has been introduced a long time ago as a built-in effect block inside old tube amplifiers of the 1960's. The implementation of the effect is again quite simple, but also extremely ingenious. The electrical signal originating from the electric guitar is directed to a system of coiled springs that act as a mechanical transmission line. The propagation mechanism provided by the springs introduces a set of repeatedly delayed

signals as the waves bounce back and forth along the springs. The reflecting signal is restored to the electrical domain in the receiving end and finally summed with the original 'bypassed' signal to create the reverberation effect.

6.15.1 Wave propagation in a helical spring

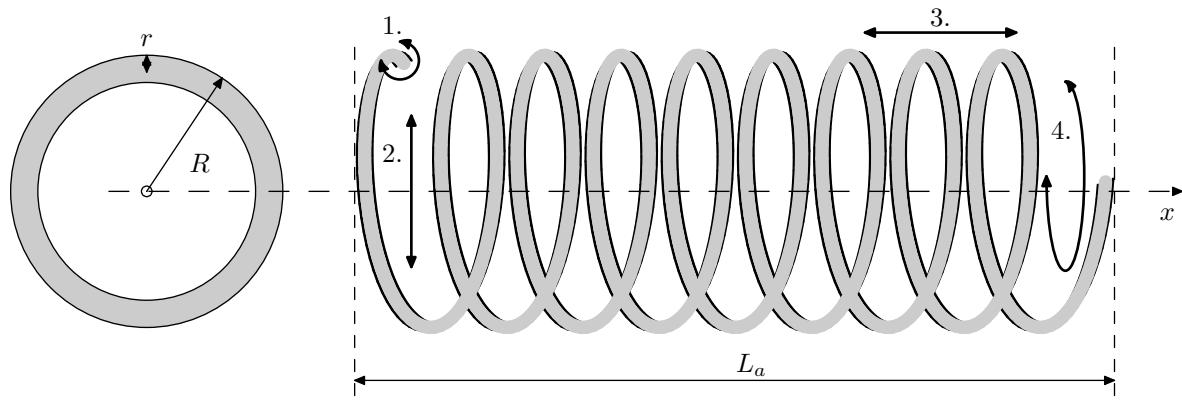
From a scientific point of view it is extremely interesting to understand how mechanical vibrations are propagating in the spring in general. To be more exact, the springs in this context should be referred to as helical springs, since that is by far the most common type of a spring used to create the reverberation effect. The dilemma of wave motion in helical springs turns out to be more complex than one would assume. Mechanical waves can propagate in a spring in several ways through basic vibration mechanisms that are originally derived to model wave motion within elastic rods.

In elastic rods there are three different types of waves possible: longitudinal waves, transverse waves and torsional waves. The most obvious of these propagation mechanisms are the transverse and longitudinal wave motion, which are already familiar from the analysis of the theoretical guitar string model. The torsional wave is set about by twisting the elastic rod around its axis of revolution so that the angle differences of twist propagate along the rod.

In springs the typical elastic rod propagation mechanisms appear in doubles. Primarily, for the sake of mathematical modelling, the helical spring is considered in this context as a hollow elastic rod with the same radius as the helix of the spring. Additionally the wire itself, which is curled to form the helical structure of the spring, can be considered an elastic rod. Therefore, there exists the longitudinal, transverse and torsional modes of wave propagation for the complete helix structure and separately the same three modes for the wire alone [88]. Figure 6.108 visualises four of the six possible propagation mechanisms in a helical spring with helix radius R , wire diameter r and axial helix length L_a . The axis of revolution of the helix is chosen to be aligned with the x -coordinate axis.

At first thought, one might assume that in the spring reverberation unit the signal travels along the spring as a longitudinal wave. The most significant question to be answered is that how long delays does a helical spring with certain physical properties create? In the case of longitudinal waves, the velocity of propagation can be calculated from a formula:

$$c_L = L_a \sqrt{\frac{k}{m}}, \quad (6.125)$$



- 1 = torsional motion of the wire
2 = transverse motion of the helix
3 = longitudinal motion of the helix
4 = torsional motion of the helix

Figure 6.108: The basic geometrical dimensions of a helical spring

where L_a is the length, m is the mass and k is the force constant of the spring.

The result of equation (6.125) is analogous to the equation of longitudinal waves in a string, which is already derived in section 2.1.3. In this case, the starting equation for the force is

$$F = -kx, \quad (6.126)$$

which is familiar from any book of basic physics. To keep the derivation robust for springs of any axial length L_a , the force in equation (6.126) should arise from stretching (or compressing) the spring the from the resting length L_a to length $L_a \frac{\partial s}{\partial x}$, i.e.

$$F = -k \left(L_a - L_a \frac{\partial s}{\partial x} \right).$$

Here ∂x is the infinitesimal unit of length and ∂s is the stretched length of the infinitesimal piece of spring. From this initial setup, the spring is set to motion by arranging a force difference between the ends of the infinitesimal piece of the spring, namely

$$dF = F(x + dx) - F(x) = \frac{\partial F}{\partial x} dx = k L_a \frac{\partial^2 s}{\partial x^2} dx.$$

The mass is again defined for the length dx , so that the Newtonian equation of motion is

$$k L_a \frac{\partial^2 s}{\partial x^2} dx = \rho A \frac{\partial^2 s}{\partial t^2} dx,$$

where the lengths dx cancel from both sides and one is left with $\rho A = \frac{m}{L_a}$, which is the mass of the spring if the length is divided away. This type of definition for

mass is commonly referred to as mass per unit length. Hence, the final simplified form of the equation of motion reads

$$\frac{k}{m} L_a^2 \frac{\partial^2 s}{\partial x^2} = \frac{\partial^2 s}{\partial t^2},$$

and the velocity of the longitudinal waves travelling along the helical structure of the spring can be assigned as $c_L^2 = L_a^2 \frac{k}{m}$.

Figure 6.109 is drawn to explain the torsional motion. The figure shows a typical model of an elastic rod, which is twisted around its axis of revolution by the moment of force M .

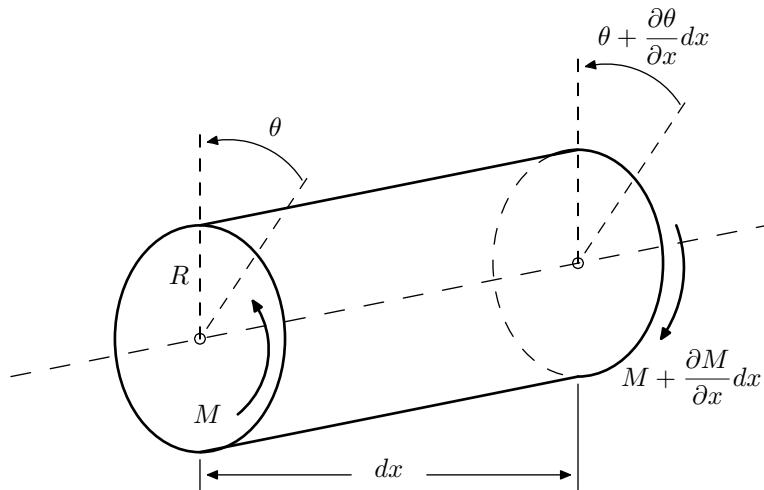


Figure 6.109: A torsion model of an elastic rod approximating a helical spring

Yet again, for a simple analysis, the linear models derived for elastic rods are applicable also for helical springs. For the torsional motion of the helix, the starting equation comes from the definition of the moment of force,

$$M = Fx = J \frac{\theta}{x} = I\alpha = \rho I_p \alpha, \quad (6.127)$$

where F is the force and moment M is measured at a point where the force is influencing through distance x . The alternative equality, where J is the torsional rigidity and θ refers to the angle of twist along the axis, follows directly from the theory of elasticity [89] and it can be derived from geometry. The mass acceleration term follows from the mass moment of inertia I and the angular acceleration α . The other measure for inertia introduced in equation (6.127) is the polar moment of inertia I_p , which is more often used along with torsional forces.

The angular acceleration is commonly defined as the second time derivative of the rotation angle, i.e.

$$\alpha = \frac{\partial^2\theta}{\partial t^2}.$$

Combined with the mass moment of inertia I , angular acceleration creates a rotational analogue to the classic Newtonian force equation $F = ma$. The inertia moments for several geometrical shapes are tabulated in common textbooks of physics, and if the helical spring is considered a hollow cylindrical rod, then $I_{helix} = mR^2$, where R is the radius of the helix.

Considering now a situation where the spring is fixed at the other end and the other end is rotated around the axis of revolution of the spring by some angle θ , the moment is created according to equation (6.127). The differences in the angular twist between the ends of a segment under analysis can be linearly reduced within the infinitesimal distance dx of the spring. In mathematical terms, this yields an equation

$$M = J \frac{\partial\theta}{\partial x} = k_r L_a \frac{\partial\theta}{\partial x},$$

where the latter part also suggests that the length of the spring changes as the rotation angle changes between the ends of the spring segment. The rotational force constant (or stiffness constant) k_r now affects the relative distance that the spring compresses or stretches during an infinitesimal rotation. [90]

Changes in the moment of force create motion to the spring, and therefore one needs the differential

$$dM = M(x + dx) - M(x) = \frac{\partial M}{\partial x} dx = k_r L_a \frac{\partial^2\theta}{\partial x^2} dx.$$

Since the moment of inertia I of an infinitesimal piece of spring is $\rho A R^2 dx$, the Newtonian equation of motion can be written as

$$k_r L_a \frac{\partial^2\theta}{\partial x^2} dx = \rho A R^2 \frac{\partial^2\theta}{\partial t^2} dx,$$

where the lengths dx cancel from both sides, and leaves $\rho A = \frac{m}{L_a}$, which equals the mass per unit length of the helical spring. Therefore,

$$\frac{k_r L_a^2}{m R^2} \frac{\partial^2\theta}{\partial x^2} = \frac{\partial^2\theta}{\partial t^2},$$

and the velocity of the torsional waves in the helix structure can be assigned as $c_R^2 = \frac{k_r L_a^2}{m R^2}$. The naming convention with the subscript R might be misleading in this context, but this comes from the need to separate the torsional motions of the helix and the wire itself. Hence, the torsional waves are imagined to create

rotational motion of the spring, and the naming of the velocity term c_R follows from this rotation.

The differential equation for the torsional waves moving *along the wire* of the spring is derived exactly like the equation for the torsional waves in the helix, the only difference being that the torsional rigidity and the moment of inertia are evaluated for the wire that forms the spring. This means that the radius R in Figure 6.109 refers now to the radius r depicted in Figure 6.108. To get confused even more, different parameters from equation (6.127) are preserved to yield the final result. By choosing the torsional rigidity J instead of $k_r L_a$ and ρI_p instead of the mass moment of inertia I , the differential equation is of the form

$$\frac{J}{\rho I_p} \frac{\partial^2 \theta}{\partial x^2} = \frac{\partial^2 \theta}{\partial t^2},$$

where

$$c_{To}^2 = \frac{J}{\rho I_p} = \frac{J}{\rho A} \frac{1}{\kappa^2} = \frac{GI_p}{\rho A} \frac{1}{\kappa^2} = \frac{G}{\rho}. \quad (6.128)$$

The constant κ refers to the radius of gyration and G is the shear modulus.

Because in some cases the values for the force constants k and k_r are unknown, it is better to derive approximate formulae to calculate the velocity of wave propagation in the spring. According to Wittrick [90], the force constants k and k_r can be written as

$$k = \frac{J}{L_{\text{tot}} R^2} \quad \text{and} \quad k_r = \frac{EA\kappa_1^2}{L_{\text{tot}}},$$

where J is the torsional rigidity, L_{tot} is the total length of the wire in the spring and κ_1 is the radius of gyration along the axis where the bending moment is affecting. These expressions are valid if there is no coupling between the longitudinal and torsional wave motion. Otherwise these can be taken as a good approximations from the actual values of the force constants.

The idea to proceed from here is to concentrate on the actual propagation times

$$t_L = \frac{L_a}{c_L} = \sqrt{\frac{m}{k}} \quad \text{and} \quad t_R = \frac{L_a}{c_R} = \sqrt{\frac{mR^2}{k_r}}$$

for the longitudinal and torsional waves in the helix respectively. It is important to note that these propagation times are independent of the axial length of the helix. These propagation times can be reduced to the case where the wave travels along the curled wire of the spring where the total length of propagation is L_{tot} , instead of along the axial length L_a of the helix. This reduction is justified

because the wave must primarily propagate along the wire, and from there it causes the axial or rotational motion of the helix structure.

With the alternative forms for the force constants, the expressions for these reduced propagation times are

$$t_L = L_{\text{tot}} R \sqrt{\frac{\rho A}{J}} \quad \text{and} \quad t_R = L_{\text{tot}} \sqrt{\frac{\rho A R^2}{E A \kappa_1^2}},$$

where the substitution $m = \rho A L_{\text{tot}}$ has been made. This substitution is very valid because it actually results in the total mass of the spring wire.

In these reduced cases the wave velocities become

$$c_L = \frac{L_{\text{tot}}}{t_L} = \frac{1}{R} \sqrt{\frac{J}{\rho A}} \quad \text{and} \quad c_R = \frac{L_{\text{tot}}}{t_R} = \sqrt{\frac{E A \kappa_1^2}{\rho A R^2}}$$

and for the torsional waves in the helix along the circular wire with $\kappa_1 = \frac{1}{2}r$, the result

$$c_R = \frac{r}{2R} \sqrt{\frac{E}{\rho}}$$

is directly obtained by simplification.

For the longitudinal case, it is beneficial to note that from equation (6.128),

$$c_{To}^2 = \frac{1}{\kappa^2} \frac{J}{\rho A},$$

so that

$$c_L = \frac{\kappa}{R} c_{To},$$

because of the common term $\frac{J}{\rho A}$.

For a wire with a circular cross section, $\kappa = \frac{r}{\sqrt{2}}$, and this leads to a relation between the velocities of longitudinal waves along the helix and torsional waves along the wire as

$$c_L = \frac{r}{\sqrt{2}R} c_{To},$$

where r is the radius of the wire and R is the radius of the helix. This is important because now all the parameters in the expressions for the velocities are easily measured from the actual spring.

In addition to the linkage with the longitudinal motion of the helix and the torsional motion of the wire, there also exists the relation

$$c_R = \sqrt{\frac{k_r}{k R^2}} c_L = \sqrt{\frac{EI_d}{GI_p}} c_L$$

between the axial and torsional propagation velocities in the helix. Here I_d refers to the diametrical moment of inertia, E is the Young's modulus, G is the shear modulus and I_p is the polar moment of inertia. This relation can be used for approximating the propagation velocity of torsional waves in the helix structure if the longitudinal propagation velocity is known. [90] [91] [92]

A practical example of the validity of the equations presented above must be shown. Let us assume a spring where the dimensions are measured as indicated in Table 6.11.

Table 6.11: The measured dimensions of a typical helical spring

<i>helix length L_a</i>	<i>helix radius R</i>	<i>wire radius r</i>	<i>turn count N</i>
0.0163 m	0.0022 m	0.000175 m	303

In addition to these measured values, from tables of physical constants one can find the values for Young's modulus, shear modulus and density of steel as $E = 200 \cdot 10^9$ Pa, $G = 80 \cdot 10^9$ Pa and $\rho = 7800 \frac{\text{kg}}{\text{m}^3}$ respectively. With these numerical values, the propagation velocity of torsional waves along the helix of the example spring becomes

$$c_R = \frac{r}{2R} \sqrt{\frac{E}{\rho}} = \frac{0.000175 \text{ m}}{2 \cdot 0.0022 \text{ m}} \sqrt{\frac{200 \cdot 10^9 \frac{\text{kg} \cdot \text{m}}{\text{m}^2 \cdot \text{s}^2}}{7800 \frac{\text{kg}}{\text{m}^3}}} = 201.4 \frac{\text{m}}{\text{s}}.$$

The velocity of torsional waves travelling in the wire of the spring is

$$c_{To} = \sqrt{\frac{G}{\rho}} = \sqrt{\frac{80 \cdot 10^9 \frac{\text{kg} \cdot \text{m}}{\text{m}^2 \cdot \text{s}^2}}{7800 \frac{\text{kg}}{\text{m}^3}}} = 3202 \frac{\text{m}}{\text{s}},$$

and with the help of this value it is possible to calculate the velocity of the longitudinal waves along the helical structure of the spring, where the velocity is

$$c_L = \frac{r}{\sqrt{2}R} c_{To} = \frac{0.000175 \text{ m}}{\sqrt{2} \cdot 0.0022 \text{ m}} \cdot 3202 \frac{\text{m}}{\text{s}} = 180 \frac{\text{m}}{\text{s}}.$$

These results indicate that the velocity of the torsional waves along the helix is slightly larger compared to the velocity of the longitudinal waves along the helix.

To determine the propagation times, it is necessary to evaluate the total length L_{tot} of the wire in the spring. Here one can use the standard methods of arc length calculations for vector functions. A helix with a radius R can be parametrised with the vector equation

$$\vec{r}(t) = R \cos t \hat{i} + R \sin t \hat{j} + bt \hat{k}, \quad (6.129)$$

where the multiplier b defines how much the helix curve proceeds upwards (or forward) for each turn of the wire. Obviously each full turn requires that t increases by 2π . The definition of the arc length element ds is

$$ds = \left| \frac{d\vec{r}(t)}{dt} \right| dt,$$

which can be integrated over the range defined for the parametrisation variable t to give the total length of s .

The arc length element for the helix is obtained from equation (6.129) by first calculating the derivative

$$\frac{d\vec{r}(t)}{dt} = -R \sin t \hat{i} + R \cos t \hat{j} + b \hat{k}$$

and then forming the expression of the arc length element

$$\left| \frac{d\vec{r}(t)}{dt} \right| dt = \sqrt{R^2 \sin^2 t + R^2 \cos^2 t + b^2} dt = \sqrt{R^2 + b^2} dt. \quad (6.130)$$

Before integrating the total length, the parameter b should be determined to reflect the actual rise of the helix curve at each turn of the wire. So, if the helix has N turns, t increases by 2π with each turn, and the axial length of the helix is L_a , then

$$b2\pi N = L_a \quad \Rightarrow \quad b = \frac{L_a}{2\pi N}.$$

After substitution into equation (6.130), this results in an arc length element of the form

$$\frac{1}{2\pi N} \sqrt{(2\pi RN)^2 + L_a^2} dt$$

and integration over the whole range of t yields

$$L_{\text{tot}} = \int_{t=0}^{2\pi N} \frac{1}{2\pi N} \sqrt{(2\pi RN)^2 + L_a^2} dt = \sqrt{(2\pi RN)^2 + L_a^2}.$$

The use of this formula in this example gives a total length

$$L_{\text{tot}} = \sqrt{(2\pi RN)^2 + L_a^2} = \sqrt{(2 \cdot \pi \cdot 0.0022 \text{ m} \cdot 303)^2 + (0.0163 \text{ m})^2} \approx 4.19 \text{ m},$$

where the variables are as indicated in Table 6.11.

Using the calculated value of the total length L_{tot} , the propagation time of the torsional waves along the helix sums up as

$$t_R = \frac{L_{\text{tot}}}{c_R} = \frac{4.19 \text{ m}}{201.4 \frac{\text{m}}{\text{s}}} \approx 0.0208 \text{ s}$$

and similarly the time for the longitudinal wave to travel once through the helical structure is

$$t_L = \frac{L_{\text{tot}}}{c_L} = \frac{4.19 \text{ m}}{180.0 \frac{\text{m}}{\text{s}}} \approx 0.0233 \text{ s.}$$

These are the times for the waves to travel from the input side of the spring to the output side. This means that these are the primary delay times measured on the first reflection at the output. The other reflections travel twice the distance, so the actual delay time obtained with the example spring is around 40 – 50 ms, which is exactly the amount of delay needed for the reverberation effect (see Table 6.8). Therefore, the calculation procedures presented above for the wave velocities are to be trusted to give good approximations on the wave propagation times in the spring.

Unfortunately the differential equations from where the delay times are derived are again naively simple and they do not represent a realistic model of the spring. These simple models give the impression that waves of all frequencies travel along the spring with the same velocity. In reality the helical spring is very dispersive, meaning that higher frequencies travel slower than low frequencies. Additionally, a practical spring has a cut-off frequency of about 4000 Hz. This means that frequencies above this limit are not propagating in the spring at all. This follows directly from a dispersion relation equation, where the local maximum of the dispersion relation corresponds with the point of minimum propagation velocity. [93]

Figure 6.110 visualises the dispersion curve for a situation where a set of waves with a large number of 'harmonic' velocities/frequencies is travelling along a spring. In an ideal case of linear dependency between the wave number and the frequency, Figure 6.110 would show a straight line, but now it is quite far from being linear. The first upper partials with low wave numbers behave somewhat like in a string, where the dispersion relation gives almost a linear dependency for the wave number and the frequency. After reaching the cut-off frequency f_c at the wave number n_c , the higher upper partials start to decrease in frequency, which is

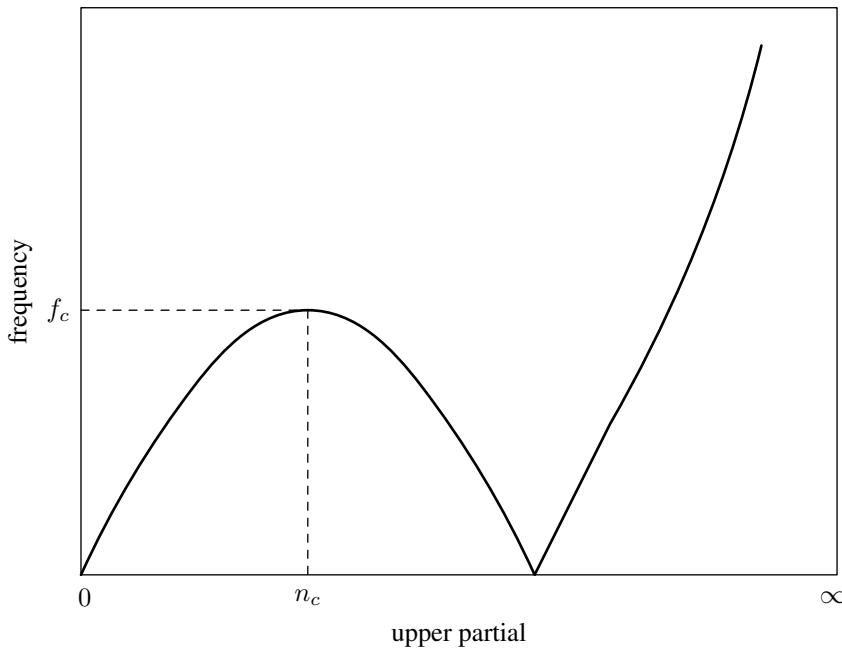


Figure 6.110: Dispersion of wave propagation in a spring

kind of difficult to understand at first. Basically this just means that the velocities of waves with a certain frequency are decreasing. Eventually the dispersion curve reaches a second minimum, whereafter the behaviour is approximately linear for very high frequencies. The general shape of the dispersion curve varies with geometrical properties, such as the helix radius and the increment angle of the helical structure. [94]

Generally speaking, each individual spring used in a commercial reverberation unit offers a slightly different propagation time. For units with two springs, the other spring has a relatively short delay time of approximately 30 milliseconds and the other has a longer delay time of approximately 45 milliseconds. If there is a third spring in the unit, then that normally has a medium delay time of about 37 milliseconds. These values are obtained from specifications provided by Accutronics. Differences to this delay arrangement might exist in reverberation units manufactured by different companies.

6.15.2 Interfacing electrical and mechanical vibrations

There exist companies that sell spring reverberation units as a ready-made ensemble so that the do-it-yourself enthusiast does not have to build all the mechanics from scratch. The only thing that is needed to integrate the reverberation effect to the propagation path of the guitar output signal is to design the electronics around the reverberation unit.

The most common type of a spring reverberation unit [95] ever used was originally designed by the Hammond Organ company in the beginning of 1960's. The basic principles of this mechanism are depicted in Figure 6.111. Hammond had patented a few other spring mechanisms already before this solution, but those were not as nearly as robust and compact as this design is.

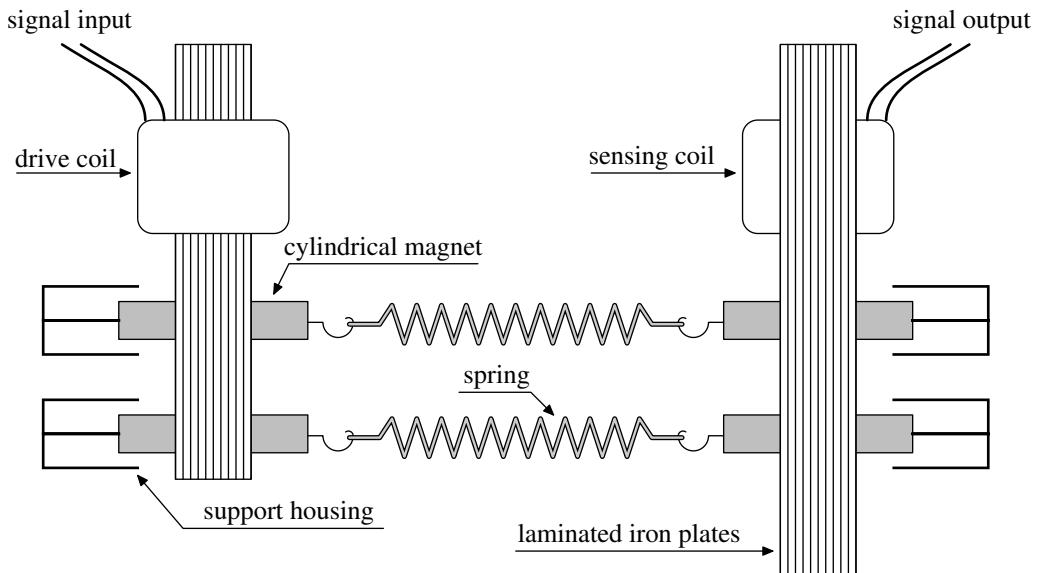


Figure 6.111: Top view of the coil - magnet - spring coupling mechanism

In the reverberation unit depicted in Figure 6.111, the signal from the guitar is fed as a current to the drive coil of the spring unit. The current in the coil creates a magnetic field to the iron core inside the coil and alternating currents of the guitar output signal create a varying magnetic field. The springs are attached to small ferromagnets, which are magnetically coupled with the magnetised iron core. The changing magnetic field in the iron core causes *rotational* movement to the ferromagnets that couple the mechanical vibrations to the springs. The vibrational wave front travels along the springs at the velocity defined approximately by equation (6.125), and when the signal reaches the output terminal of the spring unit, part of the signal reflects back towards the input and back to the output again. This way a decaying series of echoes is achieved.

The output side of the spring unit has the same mechanical construction as the input. The springs are attached to small ferromagnets, and the motion of those magnets creates a changing magnetic field to the iron core that goes through the output coil. This changing magnetic field generates a current to the output coil. The waveform of the output current is a replica of the input current. To finally create the actual reverberation effect, the delayed signal recovered from

the springs is combined with the original signal.

Figure 6.111 gives the impression that only one complete spring is used between the input and output. The system works fine with one spring, but using two springs coupled in series as shown in Figure 6.112 is mentioned to have some benefits over the one spring approach.

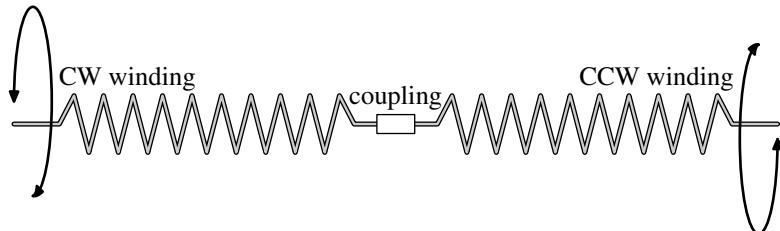


Figure 6.112: Coupling of two springs wound in alternative directions

As shown in Figure 6.112, the other spring is mounted so that its windings appear to rotate clockwise (CW) and the other in the opposite direction of the counter-clockwise (CCW) rotation of windings. The torsional waves travelling in the first spring will also excite longitudinal wave motion to the spring. The problem is now that the torsional and longitudinal waves propagate at slightly different velocities and their internal coupling can cause unwanted resonance effects in some cases. In this two-spring setup, the longitudinal waves are damped, since as the other spring compresses longitudinally due to the torsional rotation, the other spring expands and somewhat cancels the longitudinal motion of the coupled spring system [92]. A clever workaround, but sometimes the coupling of longitudinal and torsional waves can create interesting and wanted side effects along with the pure reverberation. So eventually it is just a matter of taste to use one spring or two longitudinally coupled springs as the gateway between the input and output terminals of the reverberation unit.

Figure 6.113 shows a 90 degree projection on the drive transducer and the pickup transducer. The figure shows more clearly how the small cylindrical magnets are aligned with the yoke, which is made of laminated iron plates. The drive and pickup transducers have differently shaped iron yokes, because the requirements for shielding are different in the input and output sides. The signal is very weak at the output pickup, so therefore it needs better magnetic shielding to protect against external disturbances. The almost closed outer construction of the yoke provides the extra shielding in this case.

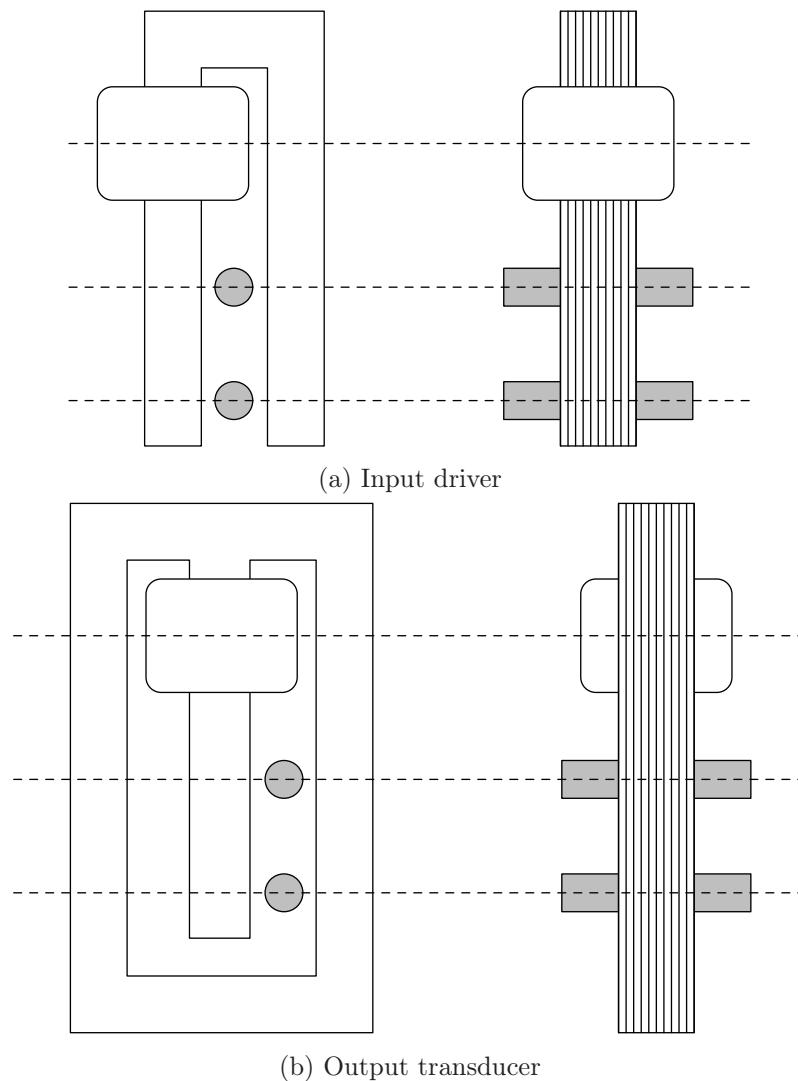


Figure 6.113: Projections of the drive (input) and pickup (output) transducers

Figure 6.114 indicates the magnetisation of the iron core due to the current flowing in the coil. The direction of the magnetic field is defined by the polarity of the top and bottom bars of the U-shaped iron core. The direction of the magnetic field changes with the direction of the alternating current in the coil. This system works as an electromagnet where the iron core is magnetised only when current is flowing in the coil.

The small cylindrical magnets that are connected to the springs are permanent magnets and they are initially magnetised in the transverse direction so that the long sides of the magnet have the magnetic polarity of N or S. As the magnetic polarity of the iron core interchanges between N and S, the cylindrical magnets align themselves according to the rules of magnetic attraction. Therefore, the cylindrical magnets transfer the alternating current signal as a rotational move-

ment to the springs.

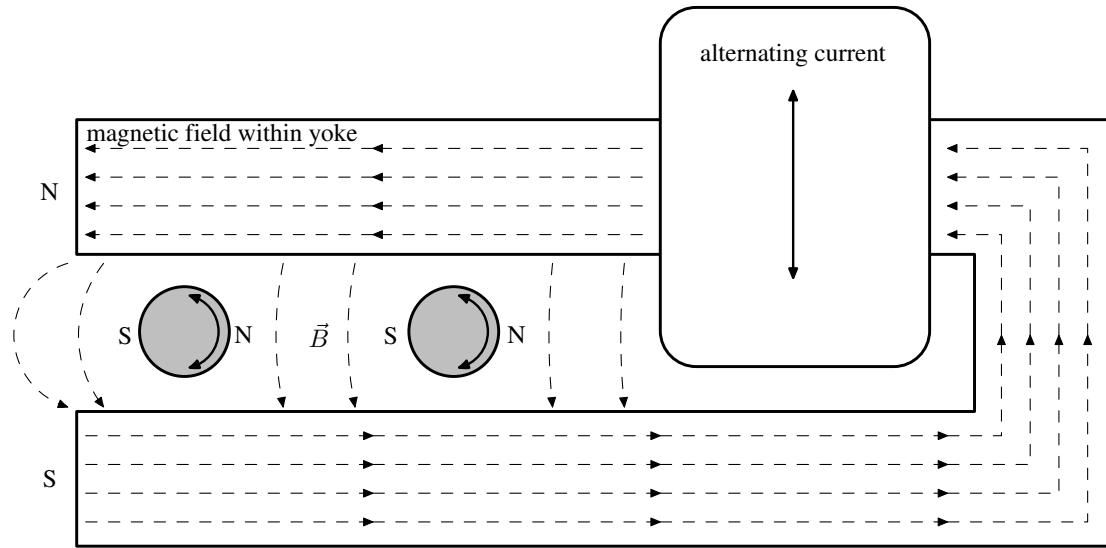


Figure 6.114: Magnetic interactions at the input transducer

A rotation at the end of the spring compresses the spring, and hence a force difference is created between the nearby curls of the spring. This force difference creates longitudinal motion in the spring that propagates towards the other end of the spring. Although the motion is not purely longitudinal, but more like rotational, the equation for the longitudinal speed of propagation can still be used to predict the time that the signal is propagating to the other end. The rotational motion unwraps from the spring when the force pulse has reached the end of the spring.

6.15.3 A drive and recovery circuit for a reverberation unit

Figure 6.115 illustrates an example of the interface to feed the drive coil of the springs with an amplified signal from the guitar and to recover the signal after it has travelled through the springs to the sensing coil. This design is copied directly from an article by Craig Anderton, published in the September edition of the Guitar Player magazine back in the year 1976 [96]. This design uses a basic quad op-amp IC as a central component to drive a reverberation unit with the guitar signal and to recover and sum the delayed echoes to the original signal. The actual component values for this circuit are listed in Appendix D.

The magnitude of the magnetic field in the iron yoke depends on how much current is flowing in the drive coil of the reverberation unit. With this in mind, it is worth noting that basic op-amps are not able to create reasonably large output

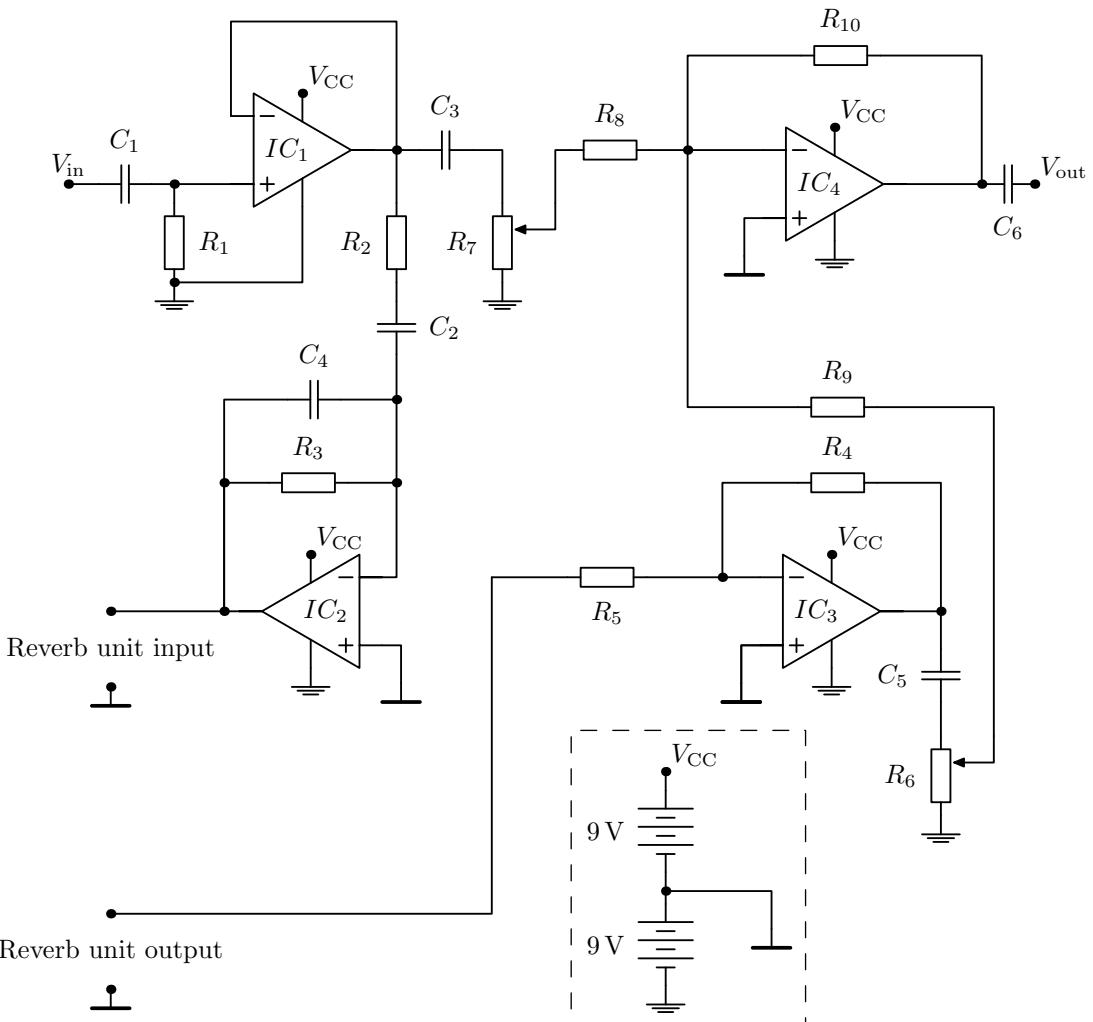


Figure 6.115: A circuit diagram of the 'Stage Center Reverb' effect

currents. That is why this particular design cannot drive reverberation units which have a low input impedance. Luckily there are a variety of reverberation units available with plenty of input impedance alternatives. A suitable nominal input impedance of the reverberation unit used with the circuit of Figure 6.115 would be from 150Ω to 400Ω . The input impedance of the reverberation unit is almost completely inductive, and the value of the given nominal input impedance is measured at 1 kHz input signal frequency.

To be able to use reverberation units with e.g. 8Ω input impedance, the interface with the drive coil should be enhanced with a power amplifier block. The IC LM386 should be quite suitable for getting a larger drive current to the reverberation unit, but the additional drive section can also be built after a basic op-amp section using two discrete transistors in a push-pull configuration.

In the schematic 6.115, the first operational amplifier IC_1 is used only as a buffer

interface between the guitar and the reverberation drive section without any amplification. The second operational amplifier IC_2 that feeds the current to the reverberation unit is the most significant component in this circuit. Because the impedance of the drive coil is growing with frequency about 20 dB per decade, the gain given by the drive circuit must somehow provide similar frequency dependent amplification. This is achieved with a biquad filter construction depicted in Figure 6.116.

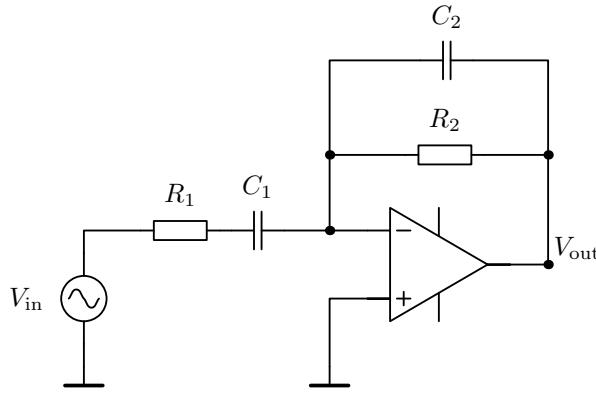


Figure 6.116: A simple band-pass biquad filter

The transfer function of the band-pass filter of Figure 6.116 is

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{\frac{1}{-sC_2R_1}}{s^2 + s\left(\frac{1}{C_1R_1} + \frac{1}{C_2R_2}\right) + \frac{1}{C_1C_2R_1R_2}}. \quad (6.131)$$

This transfer function can be derived directly from the amplification formula of the inverting op-amp configuration, which states that

$$V_{\text{out}} = -\frac{Z_2}{Z_1}V_{\text{in}},$$

where in this specific case substitutions

$$Z_2 = \frac{R_2 \frac{1}{sC_2}}{R_2 + \frac{1}{sC_2}} \quad ; \quad Z_1 = R_1 + \frac{1}{sC_1}$$

are made.

These substitutions lead to the equation

$$V_{\text{out}} = \frac{R_2}{(sC_2R_2 + 1)\left(R_1 + \frac{1}{sC_1}\right)}V_{\text{in}},$$

from where the final result (6.131) can be simplified.

In addition to the purpose of providing the frequency dependent amplification, the gain of the circuit 6.115 is intended to be adapted to reverberation units with a different input impedance by adjusting the value of R_2 . Figure 6.117 shows what this actually means in practise and what kind of transfer characteristics the filter has in general.

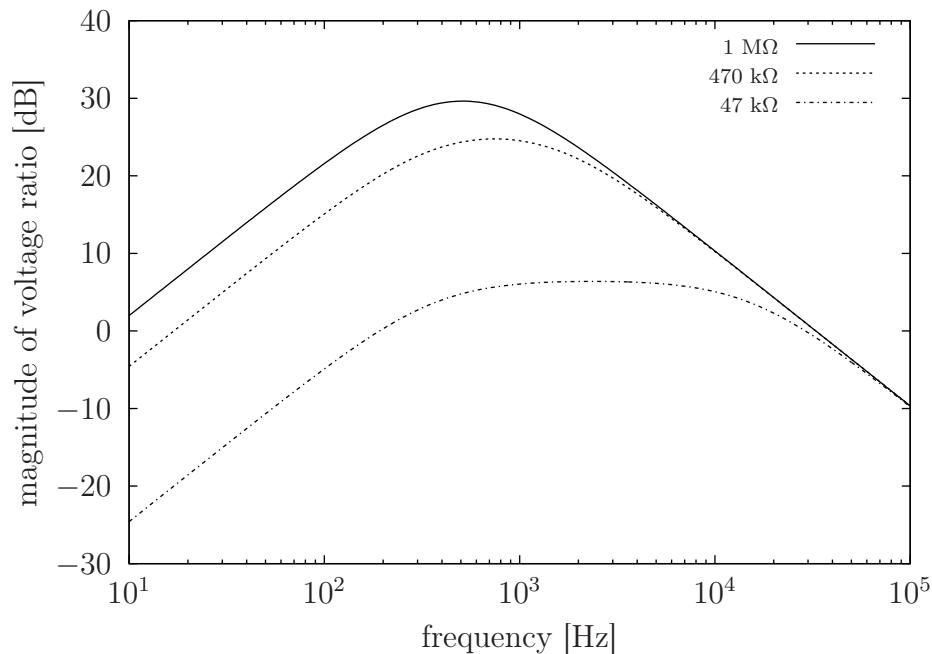


Figure 6.117: Drive stage transfer characteristics with different values of R_2

According to the frequency response curves of Figure 6.117, the filter seems to provide the rising 20 dB per decade gain as needed, but considering the ~ 4000 Hz cut-off frequency of the springs, it seems that the centre frequency of the filter has been fixed too low. It is difficult to say whether this has been intentional from the designer, but more tonal properties could be obtained from the reverberation unit if the centre frequency of the band-pass filter would be risen closer to 4000 Hz. This can be done by decreasing the values of R_1 , C_1 and C_2 and adapting the gain by increasing the value of R_2 ($= R_3$ in schematic 6.115). However, there are problems in keeping the bandwidth of the filter in a reasonable range when changing the values of any component. This can be seen directly from the transfer function, where the same components are defining both the centre frequency and the quality factor of the filter. Otherwise this combination of gain and filtering is an adequate solution for this purpose; not perfect, but adequate.

At the recovery side of the reverberation unit, the operational amplifier IC_3 is just a standard gain stage, which must have a very large gain to actually amplify the weak recovered signal from the mechanical propagation path. From the recovery gain section the signal proceeds to the operational amplifier IC_4 , which is just a summing amplifier to mix the 'dry' (original) and 'wet' (reverberated) signals together. Potentiometers R_6 and R_7 determine the user controllable mixing ratio of the dry and wet signals.

6.15.4 Measurements on the reverberation device

To test how the spring reverberation mechanism works in practise, two reverberation units were purchased and connected to the circuit 6.115. The reverberation units were cheap copies from the original Accutronics models 8AB2A1B and 1BB2C1B, which have input impedances of $10\ \Omega$ and $150\ \Omega$ respectively. The input impedances are reported by the manufacturer and the impedances are applicable for a 1 kHz test signal. Naturally the impedance changes with frequency, since the drive coil is almost purely a reactive component.

The interface circuit shown in Figure 6.115 was built onto a solderless breadboard to ease out the measurements and possible modifications. A LM224J quad op-amp chip was chosen to be used as the central component, only for the reason that it had already been bought for an earlier project and it was currently the only one readily available without having to buy any new ones. The nominal output current from LM224J is around 40 mA, and that is the only significant op-amp parameter for this circuit.

The very first finding from the measurements was that the reverberation unit 8AB2A1B with a $10\ \Omega$ input impedance cannot be used with the circuit of Figure 6.115 because the current output obtained from the driving operational amplifier is not large enough. Therefore, the prototype measurements were done using the reverberation unit 1BB2C1B only.

The primary point of interest was to experimentally find out the propagation time of the signal in the springs. By creating a very short impulse to the input of the reverberation unit, it is then possible to measure the propagation of the signal in the springs and to see as the impulse is reflected back and forth in the springs. In this home-made experiment, one end of a guitar cable was plugged into the input terminals of the circuit, and the short impulse was created by quickly hitting the open end of the cable with a thumb. There are surely better ways to create short

impulses, but the thumb was the quick and dirty way of tackling the problem this time.

Figure 6.118 shows the first and second reflections of the impulse signal, and from the figure it is possible to determine the delay time of the two springs and notice the difference between them. The reflected signals from each of the

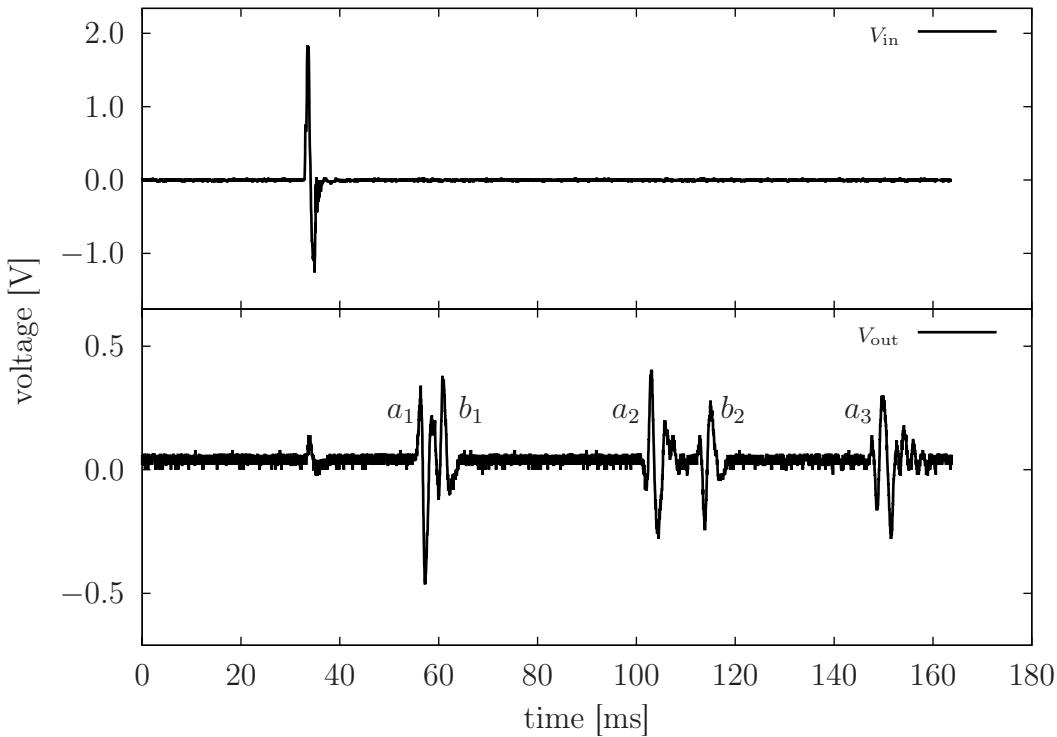


Figure 6.118: The first reflections from the two springs, a and b

two springs are identified in the figure with letters a and b , and their subscripts indicate the ordinal number of reflection, i.e. a_1 refers to the first reflection of the signal in spring a . From the first reflection it can be seen that the propagation time for the signal to travel from input to output is approximately 23 ms for spring a and 29 ms for spring b . These are the primary delay times of the springs.

The following reflections defining the secondary delay times take double the times mentioned above, since the signal travels twice the length of the spring before reaching the output again. Therefore, the actual delays obtained from the springs are approximately 46 ms and 58 ms, and these times are in a relatively good agreement with the definition of the reverberation type of delay indicated in Table 6.8.

Already from the reflection identified as a_3 in Figure 6.118 it is possible to notice

some smearing of the reflected pulse. This is most likely due to the fact that the spring is a very dispersive propagation medium, so the high frequencies travel slower in the spring compared to the low frequencies. Since the input pulse was not a pure sine pulse, it is constructed of several different frequencies that propagate at different velocities in the springs.

Figure 6.119 offers a view from a longer time period where the effect of dispersion is more clearly seen. In addition to the smearing of the wave in a single spring, at some instances of time the wave of the other spring reflects at the same time as the other. This created even more chaotic smearing of the original simple input pulse.

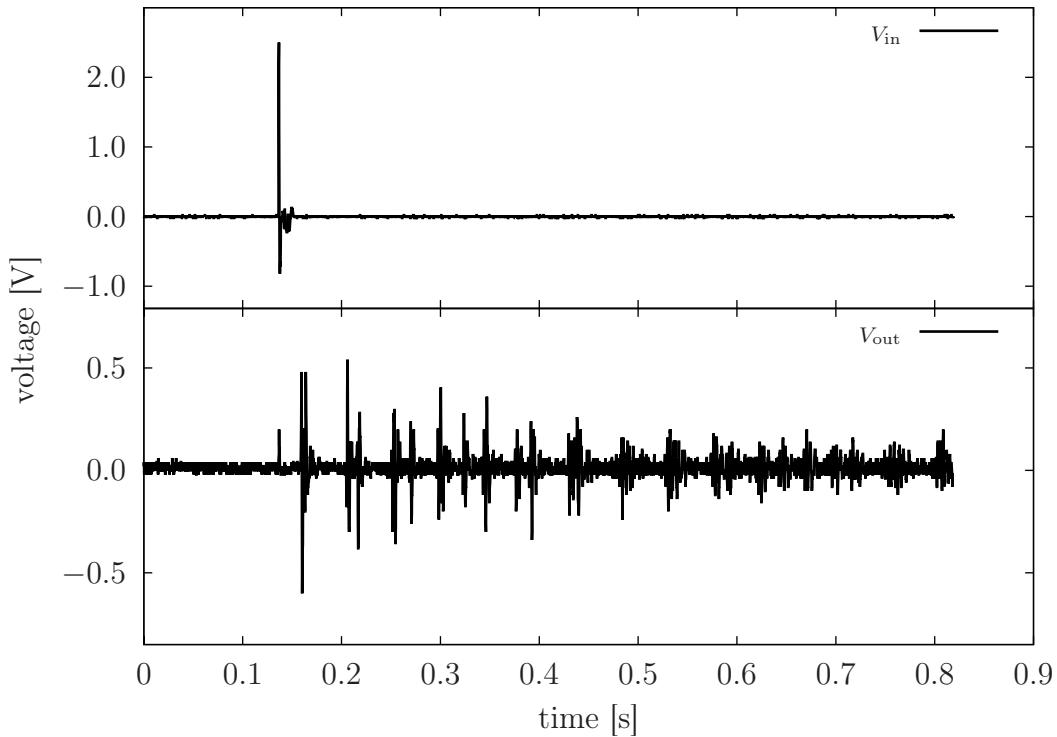


Figure 6.119: A broader view on the reflective reverberation process

In a more general scope, the interest is also towards the decay time of the reverberation. This was tested by giving a slightly longer test pulse to the circuit to create a continuous reverberation waveform with several reflections summed up together. The actual shape of the input signal and the resulting decay waveform are depicted in Figure 6.120. From the decaying envelope of the reverberation output it is possible to approximate the decay time and also verify that the reverberation unit is actually working as it should, since the manufacturers usually give the typical decay times in the datasheets.

For this specific reverberation unit, the decay time was reported as a range from 1.75 seconds to 3 seconds in the datasheet, but no specific information is given to what level the signal decays within the given time frame. The measured decay in Figure 6.120 appears to be somewhere inside these limits, so the reverberation unit seems to be working properly with the setup used.

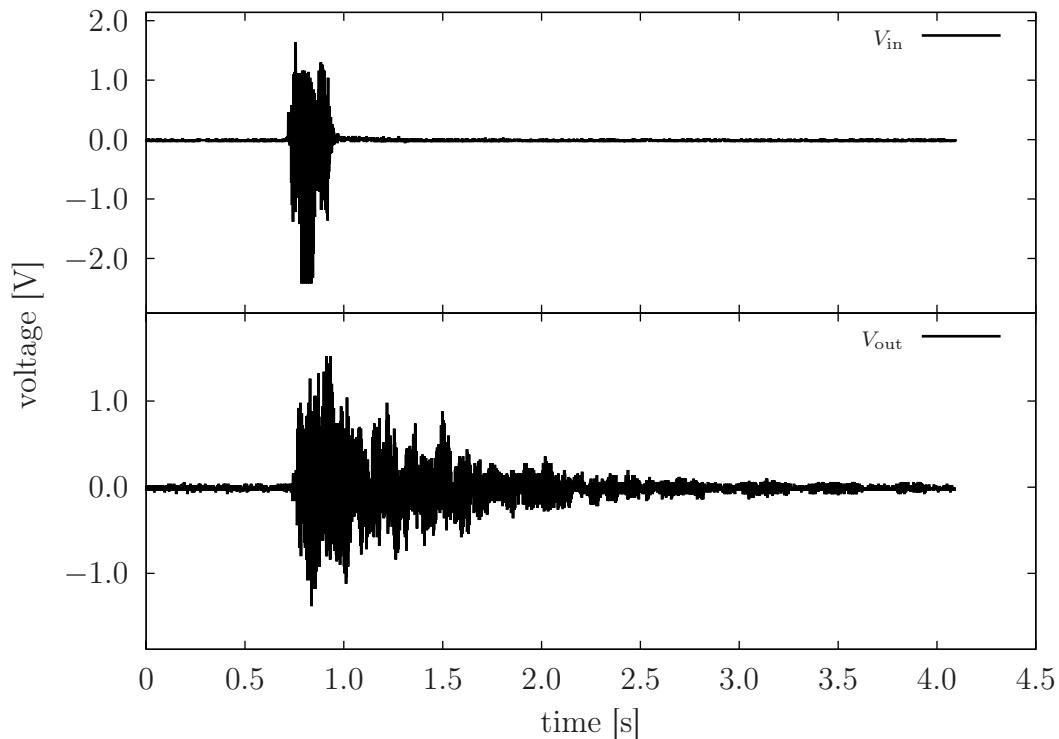


Figure 6.120: The approximate decay time for a burst of sound

For some input frequencies the reverberation waveform starts to resemble a standing wave with equally spaced crests and notches. Most likely this is due to the mixing of the similar signals in both springs. The standing wave effect is visualised in Figure 6.121. This happens more clearly with certain frequencies, while some other frequencies might not show any tendency to the standing wave phenomenon. There might also be some resonance effects behind this type of behaviour.

A note on the linearity of the reverberation unit: although a helical spring is a very dispersive platform for wave motion, single frequencies can propagate in the spring without any distortion what so ever. Figure 6.122 depicts a situation where a continuous sine wave is transmitted to the reverberation circuit. Surprisingly the output is also a clean sine wave and there is no evidence of reverberation noticed.

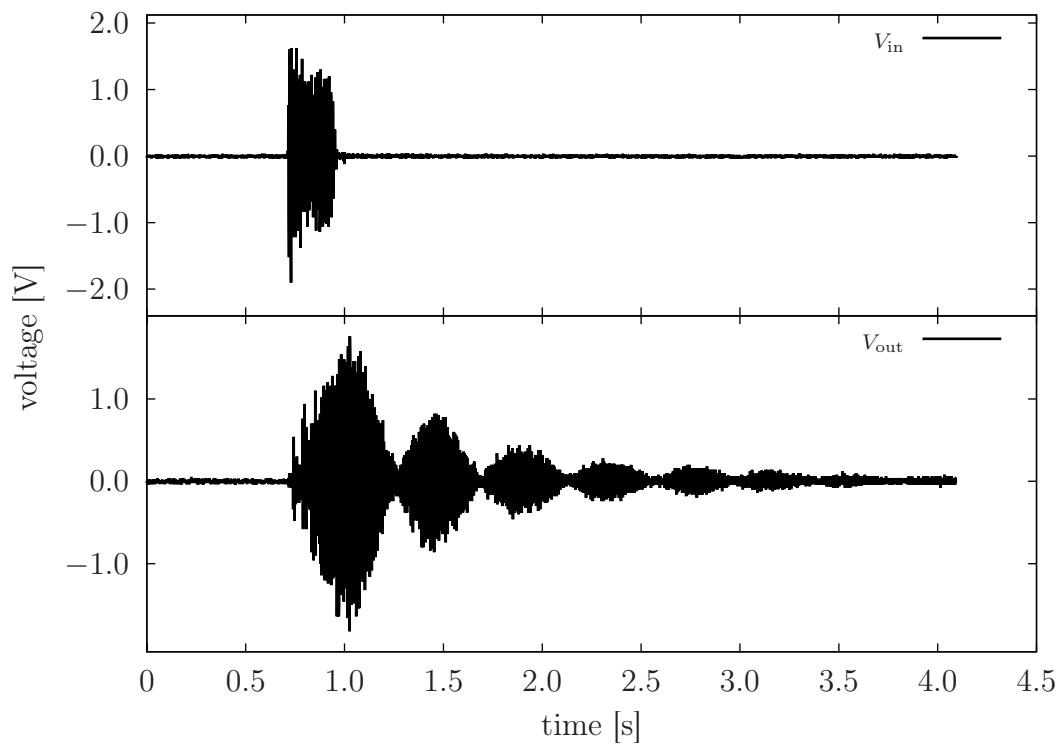


Figure 6.121: Beats in the reverberation waveform

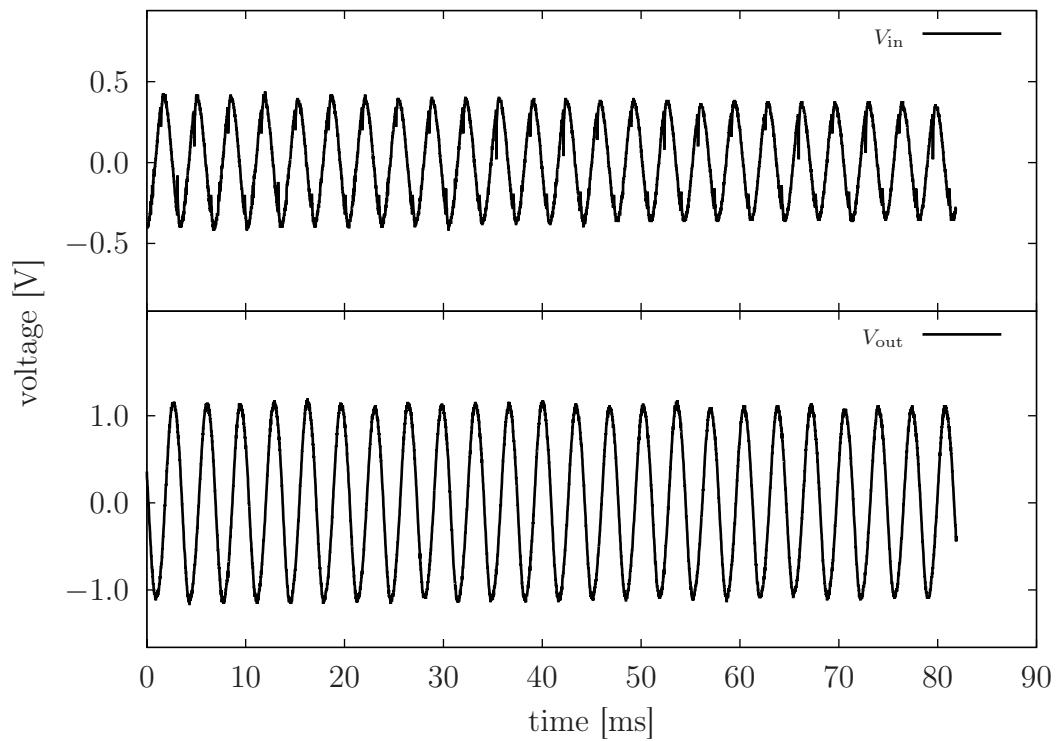


Figure 6.122: A spring as a linear medium of propagation

In this kind of situation, the reflected sine waves sum up with a different phase, but the period is still the same. With any phase difference this situation results in a wave with the same period, but additionally if the phase difference is in a suitable range, the wave keeps the shape of an ideal sine. However, it has to be mentioned that this kind of clean mixing of the waves does not happen for all phase differences, and most likely in the majority of cases the sum of reflections will cause additional bumps to the waveform. Nevertheless, in all cases of pure sine wave reflections, the period of the output signal is the same as for the input signal.

CHAPTER

7

GUITAR AMPLIFIERS

Based on topics discussed in section 1.3.8, it is already known that transistors are used for signal amplification. A quick conclusion might be to assume that guitar amplifiers contain a lot of transistors. This assumption is partially correct, but there also exists an alternative amplification device. Before transistors were invented, early radio receivers used some special components to amplify the weak signal of radio transmission. These radios of the old days are often called tube radios and the amplification devices used on them are called *electronic tubes* or *vacuum tubes*. These electronic tubes are still manufactured these days and the most common commercial products to use these tubes are guitar amplifiers. So again it is not always about using the most modern technology to achieve the best results, it is the quality of sound that matters and for that purpose tubes are good building blocks for guitar amplifiers.

Because of the two different choices over the amplification device, guitar amplifiers can be roughly divided into classes of tube amplifiers and solid-state amplifiers, where the term solid-state covers all transistor types. Although the tube based guitar amplifiers are claimed to offer the ultimate tone, transistors as general amplifier elements have many advantages over tubes. Some main advantages are that transistors do not need to be heated, they are smaller and lighter, operate with low voltages and are mechanically more rugged compared to the vacuum tubes. The disadvantages of transistors are temperature sensitivity, relatively low maximum values for output power and voltage swing and a wide spread of characteristics of a given type of transistor; in these categories the tube offers better performance over the transistor. [19, pp. 321 – 322]

As far as guitar amplifiers are concerned, the transistor based implementation seems to be a safer choice for a small low-power amplifier used at home. For a

large outdoor concert session, a stack of high-power tube amplifiers might be the better option. In any case, both approaches provide fully functional amplifiers for electric guitars.

7.1 COMMON GUITAR AMPLIFIER BUILDING BLOCKS

The most essential parts of a typical guitar amplifier are the preamplifier, tone control and power amplifier sections. In between these stages there might be a few additional voltage amplifier sections to increase the total voltage gain of the amplifier. This kind of amplifier flow model is aligned with the general model of an audio amplifier system, where the input and intermediate stages are aimed to amplify a small input excitation to a large enough level to drive the output stage. The output stage is a current delivering buffer that typically feeds an inductive transducer, which in the case of guitar amplifiers is a low-impedance loudspeaker or even an array consisting of several loudspeakers.

Figure 7.1 shows a block diagram view to one practical assortment of sequential amplifier stages inside a guitar amplifier. The main objective at the input side is to get as much voltage gain as possible. At the amplifier output, the loudspeakers need a lot of current instead of voltage, so therefore a power amplifier stage is needed there to provide a large current flow and a suitable impedance matching.

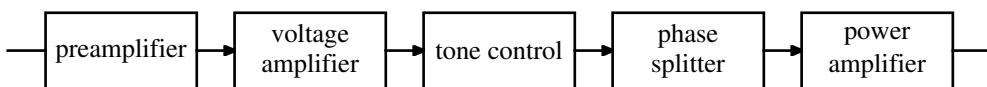


Figure 7.1: A block diagram of typical guitar amplifier stages

There is of course plenty of room inside the amplifier to extend the design to include different kinds of effects and switches. Typically many commercial amplifiers contain a built-in effects unit offering different kinds of distortion effects and reverberation. Since the effect devices have been covered already in the previous section, here the focus is on the most basic amplifier assembly. The following sections reveal more closely on a component level what actually is inside each amplifier block by analysing a few simple circuit elements using tubes and solid-state devices.

7.2 OPERATION MODE CLASSES OF AUDIO AMPLIFIERS

In general terms, different operation modes of the power amplifier output buffer section have been divided into different alphabetically named *classes*. These classes of operation are strongly related to different ways to bias the tubes or transistors at the output stage in a so-called 'push-pull' configuration. Details of this push-pull mode will follow in later sections. The most commonly used operation classes in audio amplifiers are the *A*, *B* and *AB* classes.

In the **class A** operation mode, the amplifying elements are biased so that the input signal can swing peak-to-peak without saturating or cutting off the amplifier element (tube or transistor). The operating point and the input signal level are designed so that the current in the plate (or collector) circuit of each amplifying element flows all the time.

Class *A* is the most commonly used class in basic voltage amplifier stages, and it can be considered the traditional method of amplification. Class *A* amplifiers are mainly intended to be used in the linear region of their characteristic properties to avoid excessive distortion arising from the nonlinear region. For large-signal amplification, nonlinear distortion from a class *A* configuration is unavoidable. Amplifiers in this class do not necessarily need to be in a push-pull configuration, so that a single amplifier element can be used to implement a class *A* output buffer.

The second important amplifier class is **class B**, which requires that the output buffer has the push-pull configuration implemented with matched amplifier elements. In the DC quiescent state, both of the amplifying elements have been biased exactly on the verge of cut-off so that ideally no idle currents are flowing without the input signal. This is a big benefit when considering the total dissipative power consumption of the amplifier.

When an input signal is applied to the class *B* output buffer, the positive parts of the input voltage activate the other element from the cut-off state and the negative parts activate the other element. The uneven biasing arrangement in the class *B* amplifiers increases the effective range of voltage levels which the push-pull configuration can take as input. If the input voltage is sinusoidal, amplification takes place only for a half cycle, as shown in Figure 7.2.

In practice it is almost impossible to arrange the amplifier elements to have identical cut-off biasing. For this reason the class *B* amplifiers suffer from quite severe

distortion (called *crossover distortion*) at the transition region, where the other element cuts off and the other goes to the active region. Because of this distortion, high fidelity audio amplifiers very rarely use a pure class *B* construction in the output buffer. On the other hand, class *B* amplifiers benefit from the push-pull configuration to reduce harmonic distortion. The reasons leading to this distortion reduction are explained later in section 7.3.

The **class AB** is a delicate mixture or a harsh compromise of the *A* and *B* classes and it is clearly the most popular class used in audio amplifiers in general. In this class the amplifier elements have still been biased close to cut-off like in class *B*, but they allow both positive and negative input voltages to be amplified partially in both amplifier elements. Depending on the amount of mixing, the other half of the input signal will not be fully amplified because the amplifier element goes to the cut-off state at some point along the way. This arrangement can be used to avoid the crossover distortion of class *B* amplifiers but still take advantage of the increased dynamic range over the pure class *A* operation and retain the distortion cancelling benefits of class *B*. [97, p. 60] and [19, p. 451]

Figure 7.2 visualises the response of each of the three classes to a sinusoidal input signal. The black and silver colours refer to push and pull elements separately to clarify the differences between the operation classes. The other signal (silver colour) has been deliberately shifted very slightly to the right so that both signals can be identified during the whole operation cycle.

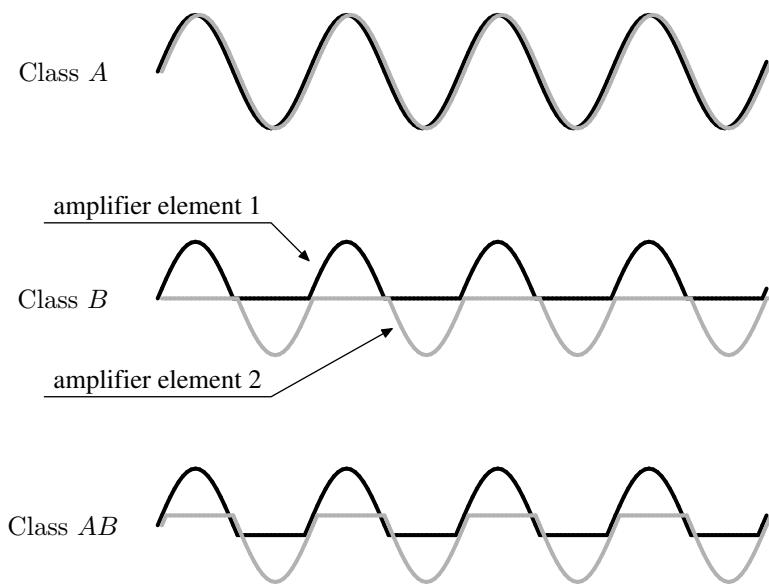


Figure 7.2: Power amplifier operation classes compared

7.3 DISTORTION IN AUDIO POWER AMPLIFIERS

All amplifiers exhibit various forms of distortion. Especially since the power amplifier section operates on large signals instead of small signals (the linear equivalent model), the output stage of a power amplifier is prone to distorting the signal due to the nonlinear transfer characteristics of transistors and tubes. The most common distortion mechanisms for amplifiers in general are *nonlinear distortion*, *frequency distortion* and *delay distortion*. The nonlinear distortion in the forms of harmonic and intermodulation distortion is the most noticeable source of distortion in power amplifiers. [19, p. 452]

The nonlinear distortion within the amplifying element is the result of rejecting the inaccurate small-signal assumption that the output plate/collector current i_o would depend linearly on the input signal x , i.e. $i_o = Gx$, where the parameter G represents a constant factor. Instead, for large amplitude signals it is more realistic to assume that the input and output signals are related through the parabolic equation

$$i_o = G_1 x + G_2 x^2, \quad (7.1)$$

where G_1 and G_2 are general constants that specify the actual shape of the parabola.

Now, if the input signal is a sinusoid of the form

$$x = X_m \cos \omega t, \quad (7.2)$$

a substitution into (7.1) leads to

$$i_o = G_1 X_m \cos \omega t + G_2 X_m^2 \cos^2 \omega t. \quad (7.3)$$

Application of the trigonometric formula $\cos^2 \omega t = \frac{1}{2} + \frac{1}{2} \cos 2\omega t$ to the quadratic sinusoid reduces the equation to the linear form:

$$i_o = B_1 \cos \omega t + B_2 \cos 2\omega t,$$

where the constants B_1 and B_2 include all possible constant factors and which can be evaluated in terms of the G parameters.

The point of all this is to show how extra harmonic frequencies appear in large-signal amplification applications and that those harmonics can be considered as noise on top of the original signal. In a general case the large-signal transfer characteristics of a tube or a transistor should be modelled with a power series

$$i_o = G_1 x + G_2 x^2 + G_3 x^3 + G_4 x^4 + \cdots + G_N x^N,$$

which after substituting the cosine input signal from equation (7.2) becomes

$$i_o = B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + B_4 \cos 4\omega t + \dots \quad (7.4)$$

Hmm, it appears that the Fourier series is everywhere. [19, pp. 543 – 549]

The Fourier components B_n in equation (7.4) represent the amplitudes of the output current components. Numerical values for these amplitude components can be obtained with a method described by Espley [98]. In this method, equidistant points from the transfer curves are graphically chosen and used to create a set of equations from where the terms B_n can be solved. However, the use of this method here is irrelevant since the aim is only to understand the fundamental sources of harmonic distortion.

From the amplitude terms B_n , the distortion components are evaluated as

$$D_2 = \frac{|B_2|}{|B_1|}, \quad D_3 = \frac{|B_3|}{|B_1|}, \quad D_4 = \frac{|B_4|}{|B_1|} \quad \dots,$$

where the D_s ($s = 2, 3, 4, \dots$) represents the distortion of the s th harmonic. The total root-mean-square (RMS) power obtained from the distorted signal is

$$P = (B_1^2 + B_2^2 + B_3^2 + \dots) \frac{R_L}{2} = (1 + D_2^2 + D_3^2 + \dots) P_1,$$

where P_1 is the power of the undistorted signal and R_L represents a load resistance where the power is measured. The total harmonic distortion can be evaluated as a squared sum of the distortion components

$$D = \sqrt{D_2^2 + D_3^2 + D_4^2 + D_5^2 + \dots}$$

so that

$$P = (1 + D^2) P_1.$$

As an example, a total distortion of 10 percent ($D = 0.1$) leads to a total power of

$$P = (1 + [0.1]^2) P_1 = 1.01 P_1,$$

which indicates a relatively low impact of harmonic distortion on the total output power.

Relating to the distortion issues, the reason to favour the push-pull configuration as the power amplifier output stage construction is the fact that it eliminates a great deal of harmonic distortion. In the push-pull configuration consisting of amplification devices Q_1 and Q_2 , an input signal

$$x_1 = X_m \cos \omega t$$

delivered to the amplification device identified as Q_1 creates an output current

$$i_1 = I_0 + B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \dots,$$

where the harmonic distortion components are present. The push-pull ideology demands that the input signal or the output signal of Q_2 is phase inverted with the signal of Q_1 , therefore

$$i_2 = I_0 + B_0 - B_1 \cos(\omega t + \pi) + B_2 \cos 2(\omega t + \pi) + \dots,$$

which is equal to

$$i_2 = I_0 + B_0 - B_1 \cos \omega t + B_2 \cos 2\omega t - B_3 \cos 3\omega t + \dots.$$

Now that the total output signal is formed as the difference between the output signals of Q_1 and Q_2 ,

$$i_o = k(i_1 - i_2) = 2k(B_1 \cos \omega t + B_3 \cos 3\omega t + B_5 \cos 5\omega t + \dots)$$

with a constant of proportionality k .

The conclusion is that with Q_1 and Q_2 being identical, a push-pull configuration will delete all even numbered harmonic distortion components from the output signal. This is the main reason why almost all audio amplifier output stages are implemented with the push-pull configuration, usually referring to the use of class B . But since class B introduces the problem of crossover distortion, class AB is the most popular output configuration in all audio power amplifiers that feed signal to loudspeakers.

The harmonic distortion occurred when the input signal was a single sine wave. Taking the input signal from a guitar when playing chords generates a set of different frequencies that are not integer multiples of each other. When this kind of signal is substituted into equation (7.1) and the same evaluation is done as in the case of a single input frequency, the result is an additional set of frequencies, which are sums and differences of the original input frequencies. These additional non-harmonic frequencies with respect to amplifiers are referred to as *intermodulation distortion*. When considering musical signals in general, intermodulation distortion is naturally present at all times, but how the amplifier enhances these distortion components is completely another question.

7.4 TUBE AMPLIFIERS

As already noted, tubes are the semiconductors of the old days. Tubes come in different flavours and they are not to be taken only as plain amplification devices.

The simplest type of a tube is called a diode, and basically its functionality is the same as for the semiconductor diode. For the tube to act as a semiconductor transistor replacement, there exist triode and pentode tubes. In the basic operation mode tubes also need to be heated, so therefore there are two additional heating wires coming out from all different tube types.

Figure 7.3 illustrates some common tubes and indicates the heater (H) wires in each type of tube. Normally when drawing schematics for tube circuits the heater wires are not included in the drawing.

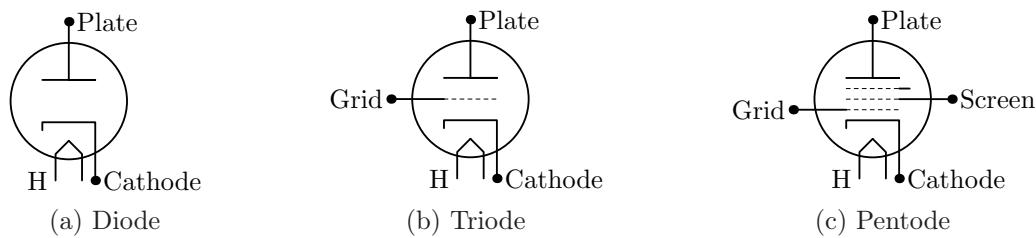


Figure 7.3: Common vacuum tube types

7.4.1 The anatomy of different tube types

The tube diode shown in Figure 7.3a consists of two main elements: the cathode and the plate. If the tube diode is forward-biased, the cathode emits electrons to the plate, which is kind of a collector of electrons. On the other hand, if the diode is biased on the negative direction, it prevents the current from flowing. Therefore, the most common application for the tube diode is to act as a rectifier on the power supply circuitry, and similarity to the semiconductor diode is evident.

The tube triode depicted in Figure 7.3b has the transistor as its semiconductor relative. The basic construction is the same as in the tube diode, but the third wire is added to form a so-called *control grid*. This control grid is located between the plate and the cathode, making it equivalent to the base of the transistor. Electrons that flow from the cathode to the plate pass the control grid, which can therefore be used to control the flow of electrons; this is just what happens in the semiconductor transistor, but with the exception that the tube triode is a voltage-controlled device, whereas the basic bipolar junction transistor is a current-controlled device. In this sense the field-effect transistors as voltage-controlled current sources are triode equivalents on the semiconductor side.

The tube pentode adds still two more elements to the tube triode. These elements are called the *suppressor grid* and the *screen grid*. Both of these grids insert themselves between the control grid and the plate. The suppressor grid is usually connected directly to the cathode, which is the reason for not drawing the pin extension for it in Figure 7.3c. The suppressor grid and the screen grid stabilise and increase the gain of the pentode tube and add more high-frequency functionality by eliminating some excessive capacitance between the tube junctions. There also exists a tetrode tube, which is like the pentode but without the suppressor grid. Pentodes have displaced the tetrodes as more advanced devices, but the basic idea of screening is the same in both types.

Pentodes are typically used in radio-frequency and video amplifiers because the pentode has a relatively good high-frequency response. Pentodes are also used as audio power amplifiers because they provide higher gain compared to the triode. [97, pp. 22–28] [19, pp. 160–171]

For analysis and design purposes, the three most important vacuum tube parameters are the *plate resistance* r_p , the *mutual conductance* g_m and the *amplification factor* μ . These are dependent on each other via equation

$$\mu = r_p g_m, \quad (7.5)$$

which is similar to the BJT parameter relation $\beta_F = r_\pi g_m$. [19, p. 164]

The small-signal model of a vacuum tube is typically drawn using a voltage-controlled voltage source (VCSV), but to be able to use the often more elegant nodal analysis method, the small-signal model can also be transformed into a voltage-controlled current source (VCCS). The VCCS representation unifies the analysis of tube amplifiers and BJT transistor amplifiers in a simple fashion, and this approach is favored in the following sections. Figure 7.4 shows the voltage-controlled voltage source and the voltage-controlled current source models applied for a tube triode element. [19, pp. 188 –196]

The more general small-signal model of a tube also includes capacitances C_{gp} between grid and plate, C_{gk} between grid and cathode and C_{pk} between plate and cathode. These capacitances reduce the active bandwidth of the tube amplifier in the high-frequency range. In audio applications these capacitances are normally not a problem, so those can be left out from the analysis.

For practical calculations the tube datasheets often provide values for μ , r_p and g_m . The data might be given as averaged constant values or in the form of a graph

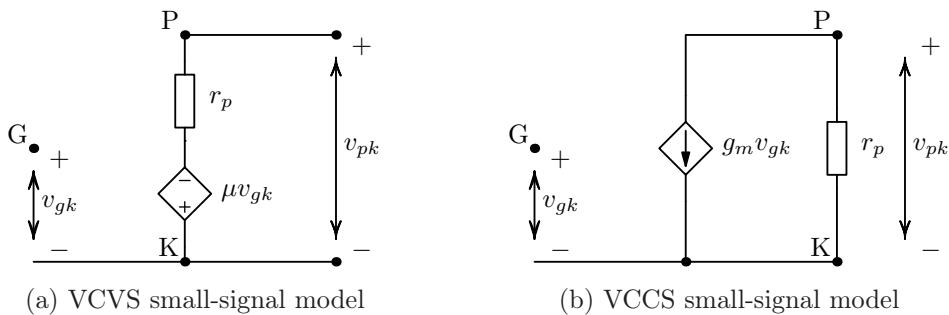


Figure 7.4: Low-frequency small-signal models for triode tubes

where the dependencies of the parameter values over the plate current are more accurately indicated. The value of r_p can also be determined analytically from the biasing currents of the tube circuit under consideration. This is analogous to the analytic determination of r_π with respect to a certain value of collector current I_C in a BJT amplifier circuit.

7.4.2 A tube preamplifier

Typically the preamplifier stage of a tube guitar amplifier consists of only a single triode, which is connected in a common-cathode configuration. At this stage the aim is to maximise the voltage gain, which is achieved by considering the maximum plate current with the applied operating voltage, and choosing the plate resistor R_L according to that restriction.

A typical preamplifier circuit of a tube amplifier is drawn in Figure 7.5. In this implementation, a tube of the type 12AX7A is often used because it has a relatively large amplification factor ($\mu = 100$) compared to other similar triodes. In the amplifier user interface the preamplifier gain is made adjustable by adding a potentiometer after the output capacitor C_L . This potentiometer is labelled as the 'volume' control knob in the front panel of the amplifier control board. The master volume control, on the other hand, is typically linked to the second voltage gain stage in the amplifier. A second voltage gain section is needed because the passive tone control section following the preamplifier usually attenuates the signal a decent amount.

To get some idea of the gain that a preamplifier stage provides to the signal coming from the guitar, it is necessary to analyse the transfer function based on a small-signal model. The preamplifier circuit of Figure 7.5 has a small-signal equivalent as shown in Figure 7.6. The tube is presented by a voltage-controlled current source, where the plate resistance r_p is used to create the current source

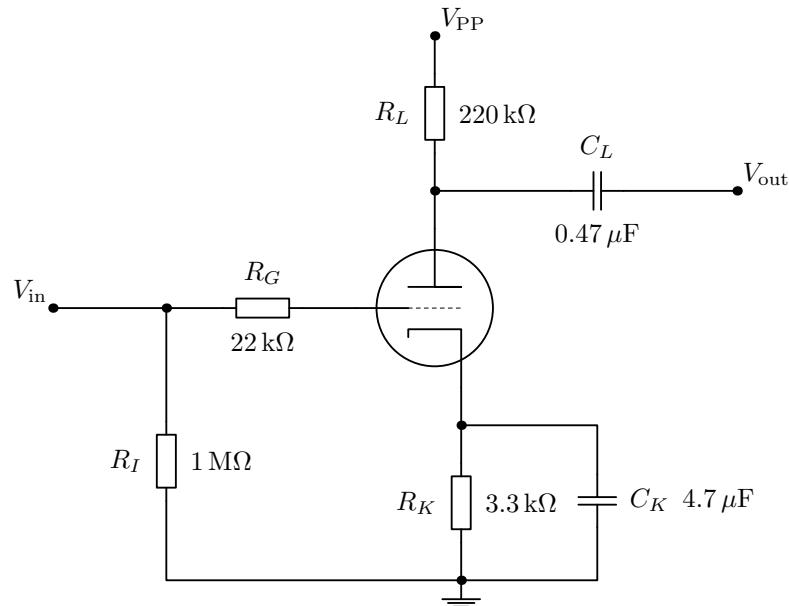


Figure 7.5: A simple preamplifier stage

equivalent of the internal model of the tube. This way the small-signal model is directly compatible with the requirements of the nodal method of analysis and is in alignment with the transistor amplifier analysis.

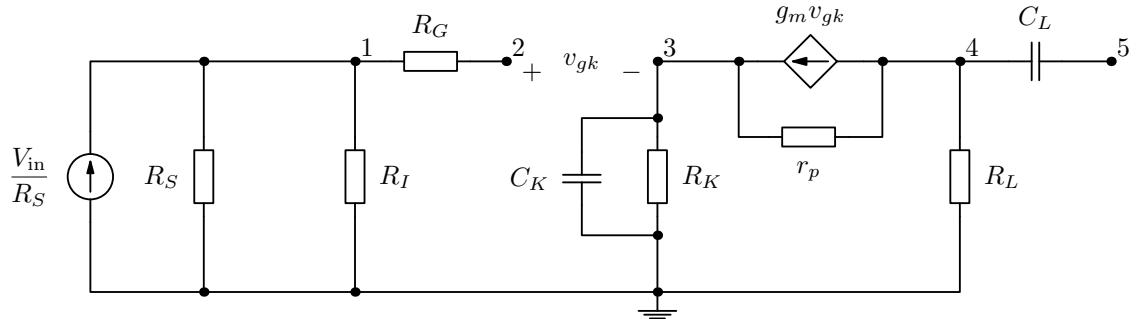


Figure 7.6: The preamplifier small-signal model

The nodal voltage equations can be written down directly as a matrix equation where on the left side an admittance matrix multiplies a voltage vector equating to a current vector on the right side. Controlled current sources appear of course to the current vector, but since the controlled sources normally depend on the node voltages, terms can (and should) be moved to the admittance matrix before solving any of the node voltages.

The nodal equations obtained from Figure 7.6 lead to a matrix representation of equation (7.6), where the controlled current source terms are already transferred from the current vector to the admittance matrix. This move saves one interme-

diate step, but for the sake of completeness it breaks the fluent continuity of the calculation process. To get an idea of all the intermediate steps involved, it is better to study the previous effects chapter, where all the steps are included in a similar transistor circuit analysis.

$$\begin{bmatrix} Y_{11} & Y_{12} & 0 & 0 & 0 \\ Y_{21} & Y_{22} & 0 & 0 & 0 \\ 0 & Y_{32} & Y_{33} & Y_{34} & 0 \\ 0 & Y_{42} & Y_{43} & Y_{44} & Y_{45} \\ 0 & 0 & 0 & Y_{54} & Y_{55} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_S} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (7.6)$$

Because of limited space, the non-zero matrix elements in equation (7.6) are indicated with the admittance symbol Y with a subscript that refers to the row and column index of the specific admittance value. The corresponding admittance values are referenced in the aligned listing (7.7).

$$\begin{aligned} Y_{11} &= \frac{1}{R_S} + \frac{1}{R_I} + \frac{1}{R_G} & Y_{12} = Y_{21} &= -\frac{1}{R_G} \\ Y_{22} &= \frac{1}{R_G} & Y_{32} &= -\frac{\mu}{r_p} \\ Y_{33} &= \frac{1}{R_K} + \frac{\mu+1}{r_p} + j\omega C_K & Y_{34} &= -\frac{1}{r_p} \\ Y_{42} &= \frac{\mu}{r_p} & Y_{43} &= -\frac{\mu+1}{r_p} \\ Y_{44} &= \frac{1}{r_p} + \frac{1}{R_L} + j\omega C_L & Y_{45} = Y_{54} &= -j\omega C_L \\ Y_{55} &= j\omega C_L \end{aligned} \quad (7.7)$$

The matrix equation (7.6) can be solved numerically using Cramer's rule and taking e.g. Octave as the number cruncher. For numerical calculations the component values are chosen as indicated in Figure 7.5 and the tube parameters are taken as $\mu = 100$ and $r_p = 62.5 \text{ k}\Omega$, with signal source resistance $R_S = 100 \Omega$. In reality the internal plate resistance r_p of the tube depends on the quiescent plate current, and in this case the true value of the internal resistance has not been calculated. Normally the resistance r_p can be determined graphically from the characteristic curves of a specific tube given in its datasheet.

With the values mentioned, the gain curve of the preamplifier stage will be

shaped as shown in Figure 7.7. The -3 dB point of this amplifier, indicating the low-frequency cut-off, is found to be below 20 Hz, which is a relatively low frequency in general. From the perspective of the frequency response, the analysed preamplifier is almost perfect for an audio amplifier. The gain of ≈ 38 dB

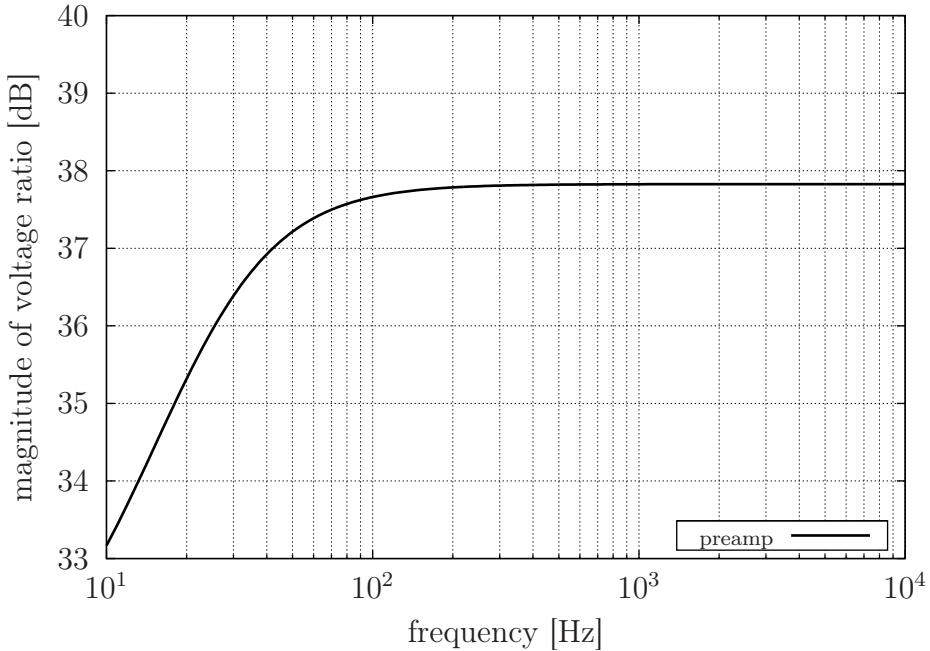


Figure 7.7: Gain curve of the analysed preamplifier

is transformed to decibels from a real multiplier of -77 , meaning that after the preamplifier stage the amplitude of the input signal voltage has been multiplied by 77 and inverted 180 degrees in phase.

The analysis of the preamplifier did not take notice that there might be several different input paths to the preamplifier. However, adding several input paths does not change the setup from the preamp's point of view. If the amplifier has two input jacks, they typically have slightly different input impedances to allow a choice between higher impedance pickups versus lower impedance pickups. The input impedance arrangement in a tube amplifier might look something like depicted in Figure 7.8, which defines different resistance values for R_I and R_G of the preamplifier depending on which of the two inputs is used.

The arrowheads in Figure 7.8 represent switches, which are normally closed but open up when a guitar cable plug is attached to the input. If the plug is inserted into input 1, the $1\text{ M}\Omega$ resistor is connected to ground via the switch of input 2. This setup leads to preamplifier circuit resistance values of $R_I = 68\text{ k}\Omega$ and $R_G = 68\text{ k}\Omega$. If the plug is connected to input 2, the $68\text{ k}\Omega$ resistors are connected

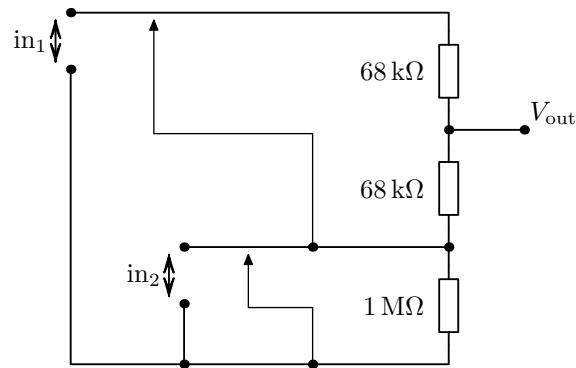


Figure 7.8: A typical input impedance arrangement in a guitar amplifier

parallel via the switch of input 2. Therefore, $R_I = 1\text{ M}\Omega$ and $R_G = 34\text{ k}\Omega$ in the preamplifier circuit of Figure 7.5.

7.4.3 A tube amplifier tone control circuit

From the preamplifier stage, the signal is usually directed to a tone control circuit, which besides tone shaping causes some attenuation of the signal. The tone control section adds two or three adjustable knobs more to the user interface of the amplifier. The most needed tone shapers are the bass and treble controls, but it is also common to have a control knob to the mid-range frequencies.

A typical tone control circuit in a guitar amplifier might look something like the one shown in Figure 7.9. The actual circuit that gets its input from the preamplifier is depicted in Figure 7.9a and Figure 7.9b shows an equivalent circuit ready for mesh analysis.

At first look, the circuit seems to have three high-pass filters stacked on top of each other and each capacitor would determine the cut-off frequency for each filter. The input signal is separated into three different paths, which are then summed together again with the two potentiometers. However, the idea of three separate high-pass filters is not the whole truth in this case because all the stacked sections are interconnected and they cannot be treated independently. A similar tone control implementation is presented in connection of the solid-state amplifier analysis in section 7.5, but there the separation of signal paths is more clearly indicated.

As opposed to the more often used nodal analysis, this tone control circuit calls for the current mesh analysis because the network has much more voltage nodes than closed current loops. By choosing the mesh analysis, the size of the impedance

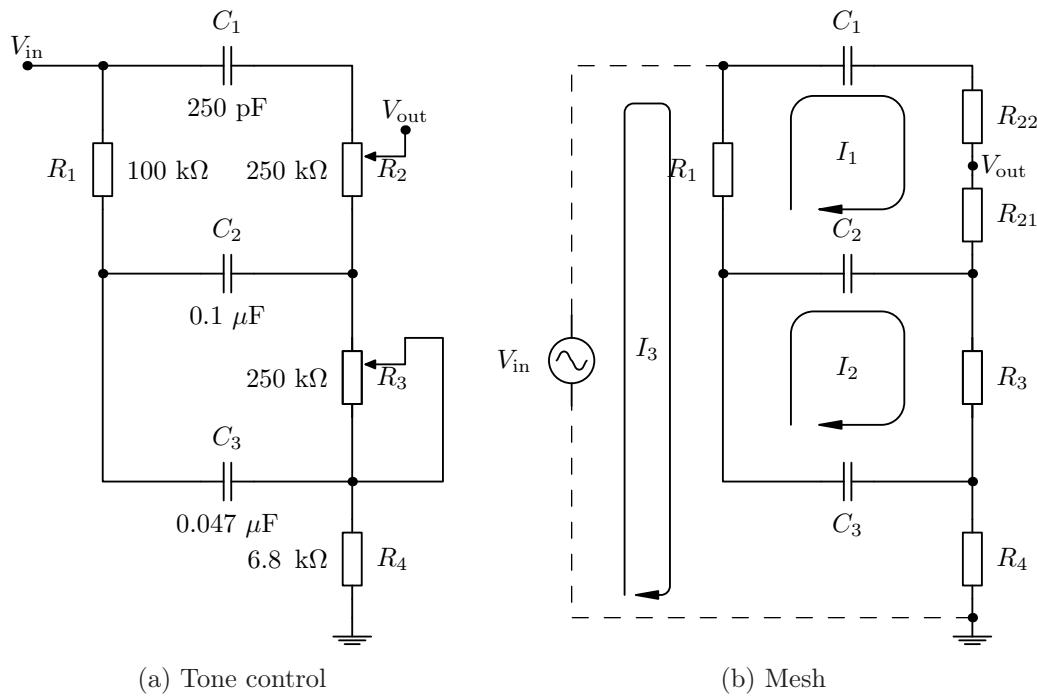


Figure 7.9: A tone control circuit of a guitar amplifier

matrix is reduced from a 6×6 matrix to a 3×3 matrix. The complete matrix equation derived from circuit 7.9b is

$$\begin{bmatrix} R_1 + R_{21} + R_{22} + \frac{1}{sC_1} + \frac{1}{sC_2} & -\frac{1}{sC_2} & -R_1 \\ -\frac{1}{sC_2} & R_3 + \frac{1}{sC_2} + \frac{1}{sC_3} & -\frac{1}{sC_3} \\ -R_1 & -\frac{1}{sC_3} & R_1 + R_4 + \frac{1}{sC_3} \end{bmatrix} \times \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ V_{in} \end{bmatrix},$$

where the Laplace domain variable s can be directly replaced by $j\omega$ for simulating the circuit with a sinusoidal input signal.

From Figure 7.9b it is clear that the output voltage expressed with respect to the mesh currents is

$$V_{\text{out}} = R_{21}I_1 + R_3I_2 + R_4I_3, \quad (7.8)$$

which means that all the mesh currents need to be solved from the matrix equation to get an expression for the output voltage V_{out} . Pen-and-paper calculation requires a big effort in this case, but it is still manageable. The expressions for the three mesh currents are

$$\frac{R_{21}I_1}{V_{\text{in}}} = \frac{(j\omega)^3 R_1 R_3 R_{21} C_1 C_2 C_3 + (j\omega)^2 R_1 R_{21} C_1 (C_2 + C_3) + j\omega R_{21} C_1}{(j\omega)^3 A + (j\omega)^2 B + j\omega D + 1}$$

for mesh loop 1,

$$\frac{R_3 I_2}{V_{\text{in}}} = \frac{(j\omega)^2 (R_3 C_1 C_2 (R_1 + R_{21} + R_{22}) + R_1 R_3 C_1 C_3) + j\omega R_3 (C_1 + C_2)}{(j\omega)^3 A + (j\omega)^2 B + j\omega D + 1}$$

for mesh loop 2 and

$$\frac{R_4 I_3}{V_{\text{in}}} = \frac{T}{(j\omega)^3 A + (j\omega)^2 B + j\omega D + 1}$$

for mesh loop 3, where the dummy variable T holds the expression

$$\begin{aligned} T = & (j\omega)^3 R_3 R_4 C_1 C_2 C_3 (R_1 + R_{21} + R_{22}) + \\ & (j\omega)^2 [(C_1 C_2 + C_1 C_3) R_4 (R_1 + R_{21} + R_{22}) + R_3 R_4 (C_2 C_3 + C_1 C_3)] + \\ & j\omega R_4 (C_1 + C_2 + C_3). \end{aligned}$$

The denominator is the same for all of the three mesh equations. The terms A , B and D in the denominator expression are

$$A = R_3 C_1 C_2 C_3 [R_1 (R_{21} + R_{22}) + R_4 (R_1 + R_{21} + R_{22})]$$

$$\begin{aligned} B = & (C_1 C_2 + C_1 C_3) [R_1 (R_{21} + R_{22}) + R_4 (R_1 + R_{21} + R_{22}) + R_1 R_3] + \\ & R_1 R_3 C_2 C_3 + C_1 C_2 R_3 (R_{21} + R_{22}) + R_3 R_4 (C_2 C_3 + C_1 C_3) \end{aligned}$$

$$D = C_1 (R_{21} + R_{22} + R_3 + R_4) + C_2 (R_1 + R_3 + R_4) + C_3 (R_1 + R_4).$$

The output/input voltage ratio can now be calculated as a sum of the mesh current expressions as indicated in equation (7.8). The voltage ratio can then be used to simulate the frequency response of the tone control section and to see how different values in the potentiometers affect the output. The tone control circuit is analysed here without any connection to the preceding gain stage nor the stages after the tone control section. This treatment therefore assumes that the output impedance of the preamplifier stage is very small, and the input impedance of the following stage is very large, which is usually true at least in the case of the input impedance of the following stage.

Figure 7.10 depicts some results of the calculations by showing the limiting cases where the potentiometers are in a maximum or minimum position. The notation 'bass' in Figure 7.10 indicates that R_3 is at maximum and R_{21} at minimum. The notation 'treble' is the opposite of 'bass', and in the 'middle' configuration both of the potentiometers are at maximum.

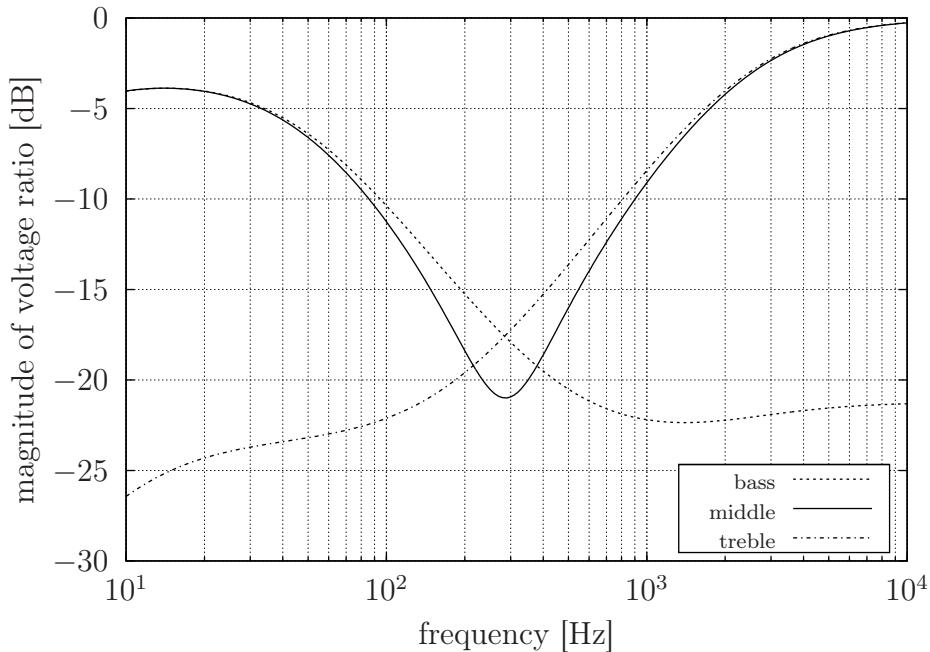


Figure 7.10: Limiting cases in the tone control settings

When either the bass control or the treble control are separately at maximum, the frequency response looks pretty nice. When both potentiometers are at maximum resistance, the middle position gets a relatively deep notch around 300 Hz. It is kind of weird to cancel these frequencies out, but most likely that is the price that has to be paid to keep the amount of components at minimum in this implementation.

The output impedance of the tone control section varies as the potentiometers are turned. Because of the alternating output impedance, it is not advisable to locate the tone control stage right before the power amplifier stage. The normal procedure is to include a second gain stage after the tone control to compensate the slight attenuation due to the passive components in the tone section and to stabilise the impedance seen by the power amplifier stage. [97, p. 48]

7.4.4 Tube phase splitters

Since there is no *pnp/npn* variation in tubes, implementation of a push-pull configuration requires that the signal path is split into two phase inverted routes before the power amplifier stage. With the separated signals it is possible to use two tubes at the output buffer instead of a single tube and get more signal power to the loudspeaker. If a guitar amplifier is designed to use only a single tube as a current buffer before the loudspeaker, this phase splitting stage is not needed.

There are two common and simple solutions available to implement a phase splitter circuit with tubes, although the same methods apply equally to transistor-based designs as well. The simplest phase splitter variant is depicted in Figure 7.11, and it is commonly referred to as the *cathodyne* (or *concertina*) phase splitter [97, p.49].

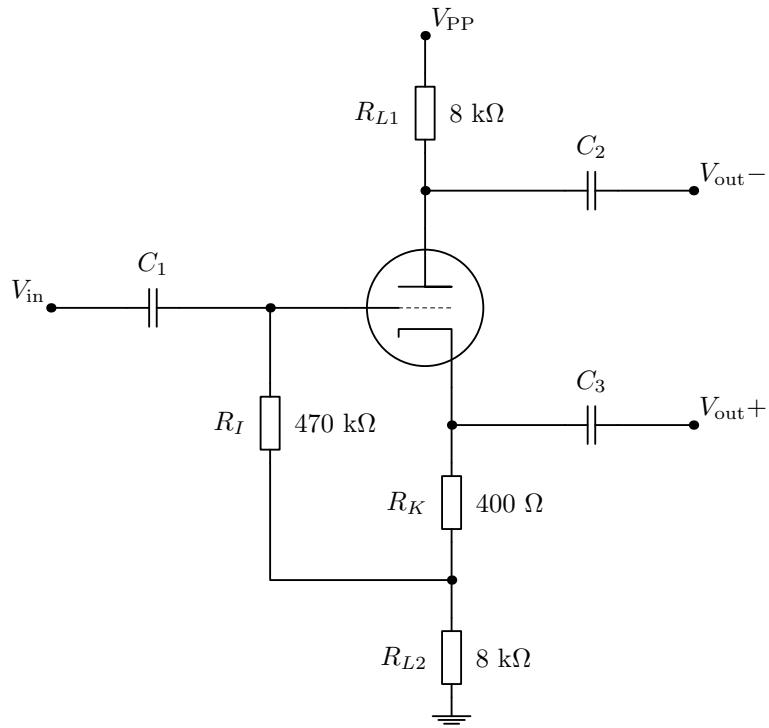


Figure 7.11: The cathodyne phase splitter

The small-signal model of the cathodyne phase splitter is depicted in Figure 7.12. The equivalent circuit is drawn without the DC coupling capacitors C_1 , C_2 and C_3 because the aim of the analysis is to find out the expression for the gain and identify the actual phase split obtained from this circuit. The resistor R_S in the small-signal model represents the output impedance of the preceding stage.

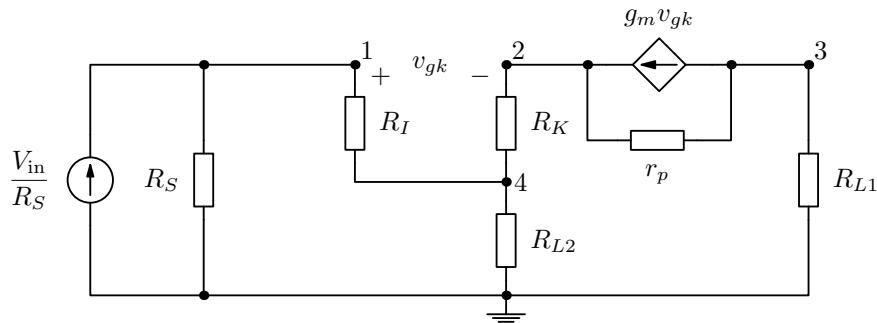


Figure 7.12: The small-signal model of the cathodyne phase splitter

As was the case with the preamplifier small-signal analysis, the matrix equations from the small-signal model of Figure 7.12 are one step ahead regarding the intermediate steps. Therefore, the controlled current source terms $g_m(V_1 - V_2)$ and $-g_m(V_1 - V_2)$ from rows 2 and 3 are already transferred from the current vector to the admittance matrix. After this step, the matrix equation of the cathodyne phase splitter circuit is

$$\begin{bmatrix} Y_{11} & 0 & 0 & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \\ Y_{31} & Y_{32} & Y_{33} & 0 \\ Y_{41} & Y_{42} & 0 & Y_{44} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} \frac{V_{\text{in}}}{R_S} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The nonzero admittance elements are expanded in the listing (7.9) and these elements are intended to be substituted into the cathodyne phase splitter admittance matrix presented above.

$$\begin{aligned} Y_{11} &= \frac{1}{R_S} + \frac{1}{R_I} & Y_{14} = Y_{41} &= -\frac{1}{R_I} \\ Y_{21} &= -\frac{\mu}{r_p} & Y_{22} &= \frac{1}{R_K} + \frac{\mu+1}{r_p} \\ Y_{23} &= -\frac{1}{r_p} & Y_{24} = Y_{42} &= -\frac{1}{R_K} \\ Y_{31} &= \frac{\mu}{r_p} & Y_{32} &= -\frac{\mu+1}{r_p} \\ Y_{33} &= \frac{1}{r_p} + \frac{1}{R_{L1}} & Y_{44} &= \frac{1}{R_K} + \frac{1}{R_I} + \frac{1}{R_{L2}} \end{aligned} \quad (7.9)$$

Although there is not much theoretical interest towards this simple circuit, the analytical expression of the transfer function is evaluated here in symbolic form. The transfer function expressions for the node voltages V_2 , V_3 and V_4 indicated in Figure 7.12 are

$$\frac{V_2}{V_{\text{in}}} = \frac{R_{L2}(\mu R_I + R_{L1} + r_p + \mu R_K) + \mu R_I R_K}{Z}$$

for node 2,

$$\frac{V_3}{V_{\text{in}}} = \frac{-R_{L1}(\mu R_I - R_{L2})}{Z}$$

for node 3 and

$$\frac{V_4}{V_{\text{in}}} = \frac{R_{L2}(\mu R_I + R_{L1} + r_p + (\mu+1)R_K)}{Z}$$

for node 4. All the transfer functions above share the same impedance expression

$$Z = R_{L1}(R_I + R_{L2}) + r_p(R_I + R_{L2}) + R_S(R_{L1} + r_p + R_{L2}) + (\mu + 1)(R_I R_{L2} + R_K[R_I + R_{L2} + R_S])$$

in the denominator.

Since the denominator is the same for each transfer function, the only difference in the gain between nodes 2, 3 and 4 comes from the numerator. Clearly, but surprisingly, the high input impedance R_I together with the amplification factor μ is the dominant term for all nodes. To get equal gain from the plate and the cathode output nodes, it must hold that $R_{L1} \approx R_{L2}$. The approximate equality just indicates that there is a very small difference, but in practise it is best to choose identical values for R_{L1} and R_{L2} .

The numerical results for the gain of the cathodyne phase splitter with the values indicated in Figure 7.11 are

$$\frac{V_2}{V_{\text{in}}} = 0.914 \quad ; \quad \frac{V_3}{V_{\text{in}}} = -0.869 \quad ; \quad \frac{V_4}{V_{\text{in}}} = 0.870,$$

where the averaged values of $\mu = 100$, $R_S = 100 \Omega$ and $r_p = 62.5 \text{ k}\Omega$ have been used in the numerical evaluation. It should be questioned whether it would be better to take the non-inverting output from node 4 instead from node 2 because there the output gain is closer to the gain of node 3. But anyway, the cathodyne phase splitter does its job and splits the phase with practically equal amplitude. A minor problem of this solution is that instead of giving some gain to the signal, the circuit actually attenuates the signal. Because of the attenuation, an extra gain stage would be needed to compensate the lack of gain. The output impedances are also different when comparing the plate output and the cathode output. For these reasons, some better solution should be found to implement a working phase splitter with decent gain.

The more popular method to construct the phase split functionality is to use a circuit called the *long-tail pair*, which besides splitting the phase also provides a relatively large gain. Figure 7.13 lays down the schematic details of the circuit, which is basically a textbook example of a basic differential amplifier, but now it is used by keeping the other input steadily grounded and feeding the input signal solely from the other input.

This version also includes a feedback input (V_{FB}), which is connected to the output buffer stage of the amplifier to stabilise the gain performance of the long-tail pair and the whole power amplifier in general. The feedback resistances R_{FB1}

and R_{FB2} have a very small effect on the basic functionality of the long-tail pair so that they can be neglected when analysing the long-tail pair alone without feedback. Eventually, when the feedback loop is closed with the output buffer of the power amplifier, the functionality of the long-tail pair does not change at all, but the feedback resistors define the gain of the whole power amplifier stage on which the long-tail pair is acting as the signal input path. This might seem confusing at first but a detailed explanation is given in the context of solid-state amplifiers in section 7.5.2.

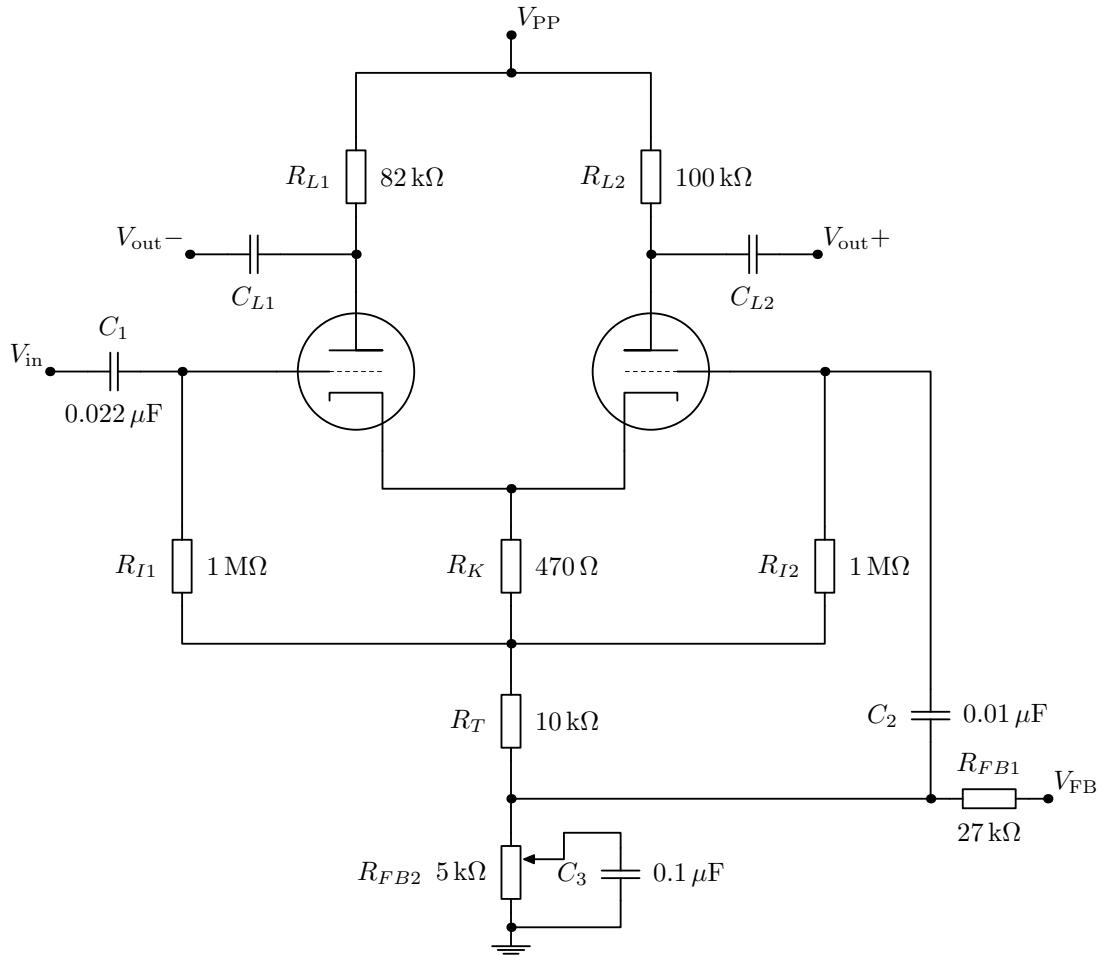


Figure 7.13: A long-tail pair phase splitter

From Figure 7.13 it is impossible to understand why the circuit works as a phase splitter. At first thought it seems that there are just two common-cathode amplifiers side by side, which would result in inverted output from both sides.

The secret of the long-tail pair is revealed when the circuit is drawn in another way. Figure 7.14 indicates that the circuit is actually a combination of common-cathode and common-grid amplifiers. Most of the surrounding components have been removed from this redrawn version but the essential core functionality is

still retained with the remaining set of components. The common-cathode section produces the phase-inverted output and the common-grid section keeps the signal in phase.

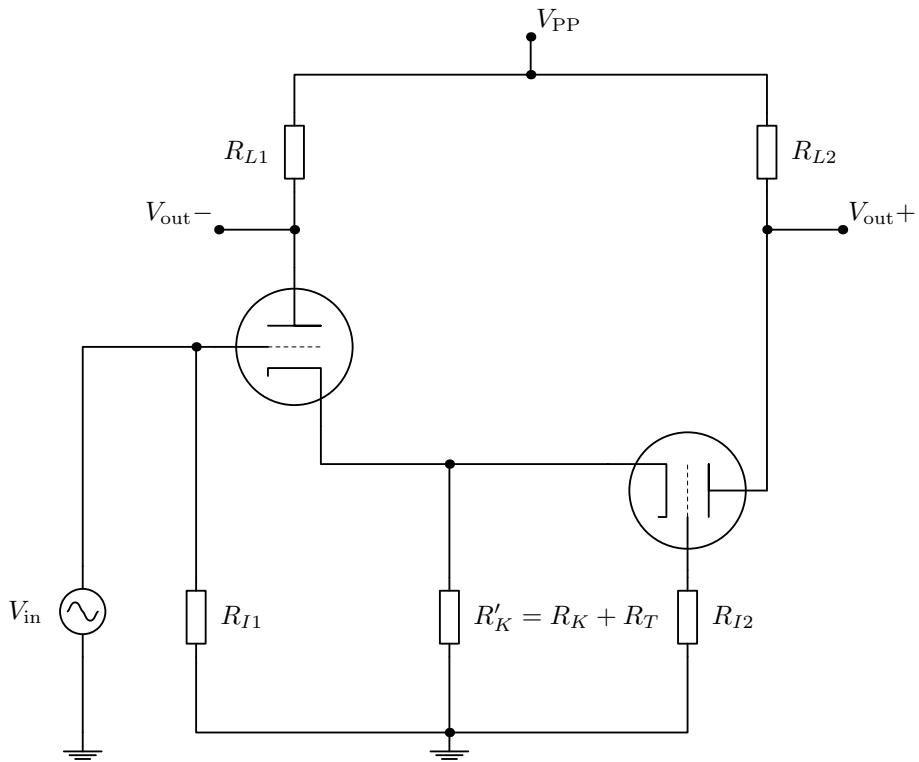


Figure 7.14: The long-tail circuit redrawn as a common-cathode common-grid pair

From Figure 7.14 it is also more convenient to proceed to analysing the small-signal model of the circuit. The reduction of the component count is done to keep things relatively simple, to focus on the main idea of the circuit, which is uneven amplification and phase splitting between the two outputs of the circuit. It should be noted that the cathode resistance R'_K is now taken as the sum of R_K and R_T indicated in Figure 7.13. This is done so that the calculated numerical results for the gain would be comparable to the gain obtained by Kuehnel [99, p. 96].

Figure 7.15 shows the small-signal model of the reduced version of the long-tail pair. The grid of the common-cathode configuration is taken as node 1, and the grid of the common-grid tube is at node 3. The outputs of the common-grid and common-cathode circuits are nodes 4 and 5 respectively. Although the circuit was clarified in Figure 7.14 by drawing it as a combination of common-cathode and common-grid amplifiers, the small-signal model still looks like as if there would be two common-cathode amplifiers joined together. Maybe it is that way in reality, because then one could assume that the gain difference would be

significant between the two outputs.

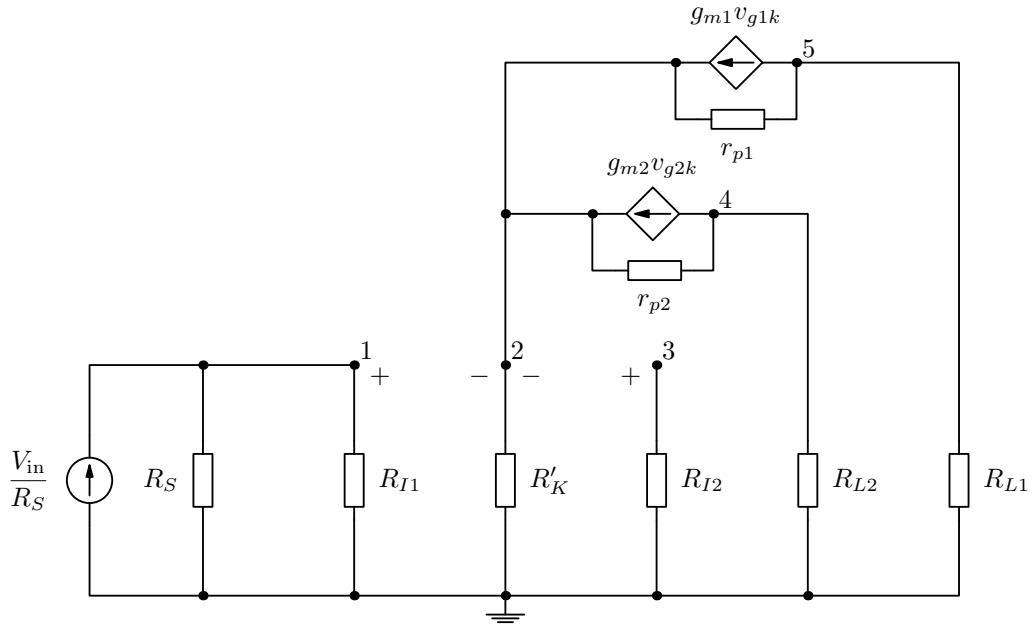


Figure 7.15: The small-signal equivalent circuit of the long-tail pair

The matrix equation (7.10) from the long-tail pair small-signal model is written directly with the transconductance terms already moved to the admittance matrix and equation (7.5) is used for changing between the transconductance g_m and amplification factor μ . This saves one intermediate step of calculation but might complicate the understanding of the analysis process.

The nonzero elements of the admittance matrix are marked with their corresponding indices. The aligned listing (7.11) contains the actual terms that should be substituted into the admittance matrix. Because there is only node linkage between nodes 2 and 4 plus 2 and 5, there is not that much symmetry in the matrix elements. After inserting the actual admittance terms from the aligned listing (7.11), the node voltages V_1, \dots, V_5 of the matrix equation (7.10) given below are ready to be solved using Cramer's rule.

$$\begin{bmatrix} Y_{11} & 0 & 0 & 0 & 0 \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} & Y_{25} \\ 0 & 0 & Y_{33} & 0 & 0 \\ 0 & Y_{42} & Y_{43} & Y_{44} & 0 \\ Y_{51} & Y_{52} & 0 & 0 & Y_{55} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{bmatrix} = \begin{bmatrix} \frac{V_{\text{in}}}{R_S} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (7.10)$$

$$\begin{aligned}
Y_{11} &= \frac{1}{R_S} + \frac{1}{R_{I1}} & Y_{21} &= -\frac{\mu_1}{r_{p1}} \\
Y_{22} &= \frac{1}{R'_K} + \frac{\mu_1 + 1}{r_{p1}} + \frac{\mu_2 + 1}{r_{p2}} & Y_{23} &= -\frac{\mu_2}{r_{p2}} \\
Y_{24} &= -\frac{1}{r_{p2}} & Y_{25} &= -\frac{1}{r_{p1}} \\
Y_{33} &= \frac{1}{R_{I2}} & Y_{42} &= -\frac{\mu_2 + 1}{r_{p2}} \\
Y_{43} &= \frac{\mu_2}{r_{p2}} & Y_{44} &= \frac{1}{r_{p2}} + \frac{1}{R_{L2}} \\
Y_{51} &= \frac{\mu_1}{r_{p1}} & Y_{52} &= -\frac{\mu_1 + 1}{r_{p1}} \\
Y_{55} &= \frac{1}{r_{p1}} + \frac{1}{R_{L1}}
\end{aligned} \tag{7.11}$$

After using Cramer's rule to solve node voltages V_4 and V_5 , which are the non-inverting and inverting output nodes respectively, the expressions for the gain factors become

$$\frac{V_4}{V_{\text{in}}} = \frac{\frac{1}{R_S} R'_K R_{L2} \mu_1 (\mu_2 + 1)}{Z} \tag{7.12}$$

and

$$\frac{V_5}{V_{\text{in}}} = \frac{-\frac{1}{R_S} R'_K R_{L1} \mu_1 (\mu_2 + 1) \left[1 + \frac{R_{L2} + r_{p2}}{(\mu_2 + 1) R'_K} \right]}{Z}, \tag{7.13}$$

where

$$\begin{aligned}
Z = & \left(\frac{1}{R_S} + \frac{1}{R_{I1}} \right) [R_{L1} R_{L2} + r_{p2} R_{L1} + r_{p1} R_{L2} + \\
& r_{p1} r_{p2} + R'_K (\mu_1 + 1) (R_{L2} + r_{p2}) + R'_K (\mu_2 + 1) (R_{L1} + r_{p1})].
\end{aligned}$$

If the transfer functions (7.12) and (7.13) are compared, they have much in common if symmetry is used for choosing the component values for the redrawn long-tail pair. To emphasise the symmetry by choosing $R_{L1} = R_{L2}$, the gain of the common-cathode section will be larger by a factor of

$$1 + \frac{R_{L2} + r_{p2}}{(\mu_2 + 1) R'_K}.$$

This gain difference is compensated in practise by choosing the value of R_{L1} to be slightly smaller than R_{L2} . It is also evident that as the tail resistance is increased, the difference in the gain becomes smaller. An infinitely large R'_K would give the same gain from both outputs with identical plate resistances.

Because R_S is clearly much smaller than R_{I1} , the internal source resistance R_S representing the output resistance of the previous stage will cancel itself out from the equations and it will not affect the results obtained from the calculations.

If the component values in the reduced model of the long-tail pair are chosen as $R'_K = R_K + R_T = 10.470 \text{ k}\Omega$ and

$$R_{L1} = R_{L2} = 100 \text{ k}\Omega \quad ; \quad r_{p1} = r_{p2} = 57.7 \text{ k}\Omega \quad ; \quad \mu_1 = \mu_2 = 100,$$

the actual numerical values for the gains at nodes 4 and 5 become

$$\frac{V_4}{V_S} = 29.5 \quad \text{and} \quad \frac{V_5}{V_S} = -33.9$$

and the factor making the common-cathode gain larger has a value 1.15. Kuehnel [99] had another approach to the gain analysis and he got gain values of 29.9 and -34.3 for the same reduced model, so the results obtained here are nicely in the same ballpark.

Now the aim would be to equalise the gains from both outputs by reducing the value of R_{L1} by 15 percent, which comes directly from the value of the extra gain factor. The reduced plate resistance $R_{L1} = 87 \text{ k}\Omega$, and the numerical values of the gains are now

$$\frac{V_4}{V_S} = 30.866 \quad \text{and} \quad \frac{V_5}{V_S} = -30.858.$$

This suggests that the value of the factorial can be used to reduce the gain of the common-cathode stage to be equal with the common-grid stage. Since $87 \text{ k}\Omega$ is not a standard resistance value, $82 \text{ k}\Omega$ is typically chosen as the closest standard resistance value in the original long-tail pair circuit 7.13.

7.4.5 A tube power amplifier

The power amplifier stage of a tube amplifier can be constructed with a single-ended one tube setup or a push-pull setup using two matched tubes. Pentode tubes are preferred over triodes in the power amplifier stage because the pentode device can provide more gain and it can be used with higher currents than the triode.

The single-ended construction is used only if the cost of the amplification system needs to be minimised. In the single-ended setup the tube is used close to its maximum ratings regarding the quiescent currents and plate power dissipation capabilities. Figure 7.16 shows an illustration of a typical single-ended

power amplifier stage using a pentode tube. In this setup, the output signal is taken from the plate and it is coupled to the load through an impedance matching transformer. The load resistance R_L represents the nominal resistance of a loudspeaker.

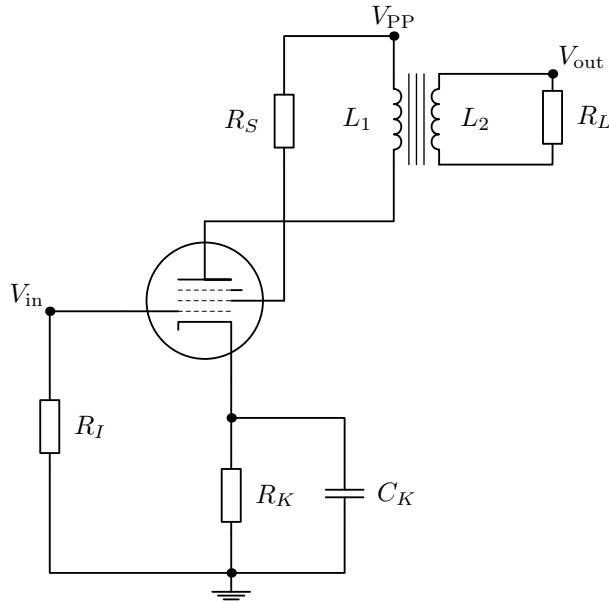


Figure 7.16: A single-ended tube power amplifier stage

The single-ended output stage works fine if the power demand at the load is not that high. The problems relating to single-ended output stages are nonlinearities that lead to harmonic distortion, and obviously one tube can deliver less power to the loudspeaker compared to two tubes. If a single-ended solution is used, there is no need to use the phase splitter either because phase split signals are only required in the double tube push-pull setup. Because the majority of guitar amplifiers use the push-pull configuration with a pair of output buffer active elements, the single-ended configuration is not explored any further.

Figure 7.17 shows a typical push-pull power amplifier section of a tube guitar amplifier with approximate component values. The input signal for the push-pull power amplifier output buffer comes directly from the long-tail pair phase splitter (Figure 7.13) through the DC coupling capacitors C_{L1} and C_{L2} . In some cases the phase splitting is done with a centre-tapped transformer, but the use of active element phase splitters such as the long tail-pair is a more elegant and cheaper solution.

By adjusting the voltage V_{bias} , the push-pull stage can be biased either in class *B* or class *AB* operation mode. The advantages of using a push-pull stage to

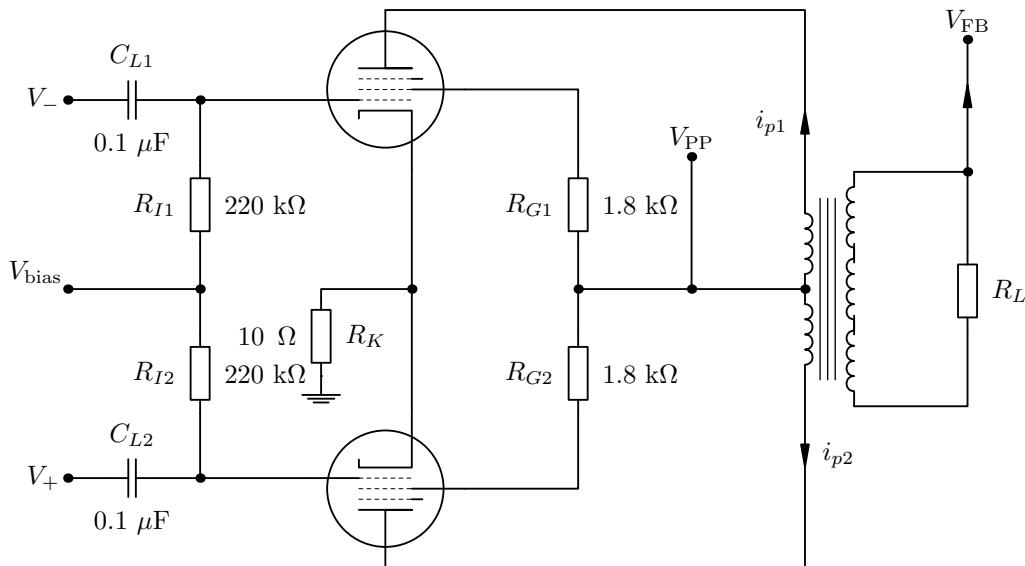


Figure 7.17: A push-pull power amplifier stage

couple through a transformer is that the DC biasing currents cancel themselves out in the primary coil and therefore they do not cause any excess magnetisation of the iron core of the transformer. Another advantage is that the even harmonic distortion coefficients cancel out in the push-pull configuration as explained in section 7.3. The common-mode hum from the power supply is also reduced by the push-pull stage, but only when the hum is added directly from the operating voltage of the push-pull tubes. If the earlier voltage amplifier stages have added power supply hum to the effective signal, then it is not cancelled by the push-pull setup. [19, p. 560]

The power amplifier section of a guitar amplifier deals with large-amplitude signals, so therefore small-signal models are not suitable for depicting the behaviour of the tubes in this amplifier stage. Quite often the power amplifier design and analysis is done graphically using experimentally determined device output characteristics. The graphical methods are applicable for both DC biasing and AC gain design.

As explained in section 7.5.2 covering the solid-state version of the power amplifier, the gain of the power amplifier stage is with good approximation determined by feedback resistors R_{FB1} and R_{FB2} , which are shown in Figure 7.13. The feedback voltage V_{FB} is taken from the loudspeaker input signal as shown in Figure 7.17. The combination of the long-tail pair and the power amplifier stage can be considered a discrete component version of an operational amplifier where the closed loop gain is determined solely by external impedances. Because the

signal after the output transformer is in the same phase as the signal coming to the long-tail pair, the gain provided by the power amplifier stage is calculated using the non-inverting op-amp gain formula

$$G = 1 + \frac{R_{FB1}}{R_{FB2}}.$$

The total gain of the whole guitar amplifier is a sum of the gain from the preamplifier plus the gain from the power amplifier. The tone control section causes some attenuation, which needs to be considered when defining the value for the total gain.

The tube power amplifier is typically coupled to the loudspeaker via a transformer as shown in Figure 7.17. This is done to prevent excess waste of power, which happens if the DC quiescent plate current flows directly through a low-ohmic load resistance, as the loudspeaker would be to the plate circuitry. Another reason is simply that it is not wise to let the DC component through to the loudspeaker or any other output device in general. To get the best power transfer efficiency between the power amplifier and the loudspeaker, the transformer is also used to achieve the required impedance matching.

An efficient power transfer between circuit stages requires that the output impedance of the previous stage and the input impedance of the following stage are equal. With a slight taste of contradiction, audio amplifier designs usually aim to make the output impedance of the power amplifier as small as possible [100, p. 25]. The reason for this is that with output impedance close to zero, the amplifier output is unaffected by the loading of the loudspeaker impedance, which is very nonlinear due to speaker cone resonances and voice coil inductance.

The voltage and current transfer characteristics of an ideal transformer shown in Figure 7.18 are written as

$$V_1 = \frac{N_1}{N_2} V_2 \quad \text{and} \quad I_1 = \frac{N_2}{N_1} I_2, \quad (7.14)$$

where the subscript 1 refers to the primary coil of the transformer and the subscript 2 to the secondary coil of the transformer. The voltage and current are transformed according to the ratio of the turn counts N_1 and N_2 . The ratio of voltage and current equations (7.14) yield

$$\frac{V_1}{I_1} = \frac{N_1^2}{N_2^2} \frac{V_2}{I_2},$$

which can also be written as

$$Z'_L = \frac{N_1^2}{N_2^2} Z_L,$$

where Z'_L represents the effective input impedance and Z_L is the output impedance. By using a transformer with a different amount of turns on each side, it is possible to create the required impedance matching.

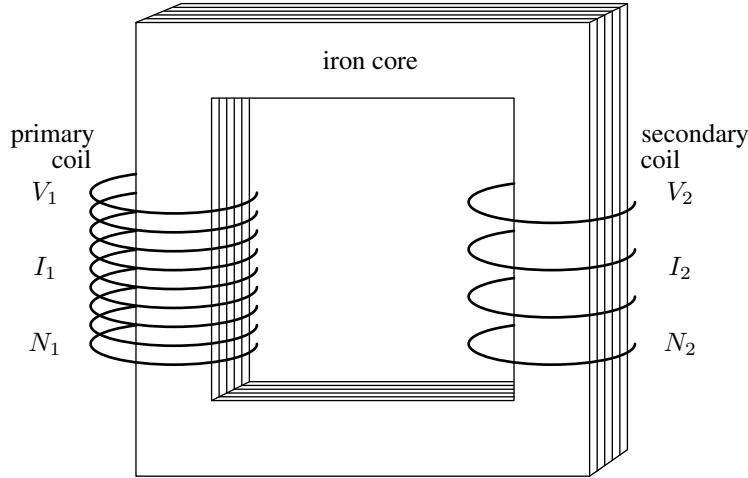


Figure 7.18: A textbook model of an ideal transformer

The actual push-pull power amplifier circuit resembles the long-tail pair circuit, but now in the power amplifier the differential amplifier construction is actually used as a real differential amplifier. A simple example of a differential amplifier with two active inputs and outputs is depicted in Figure 7.19. The amplifier amplifies the difference $V_{\text{in}1} - V_{\text{in}2}$ in such a way that $V_{\text{out}1}$ gives the inverted output and $V_{\text{out}2}$ gives the non-inverted output.

In mathematical terms, without going too deep into the details, the gain properties of the differential amplifier can be expressed as

$$V_{\text{out}1} = -A(V_{\text{in}1} - V_{\text{in}2}) \quad \text{and} \quad V_{\text{out}2} = +A(V_{\text{in}1} - V_{\text{in}2}),$$

where A is referring to some general gain factor.

Because the phase splitter is feeding the signal to the differential stage of the power amplifier, the input voltage $V_{\text{in}2} = -V_{\text{in}1}$, and the difference of these input voltages is

$$V_{\text{in}1} - V_{\text{in}2} = V_{\text{in}1} - (-V_{\text{in}1}) = 2V_{\text{in}1}.$$

Therefore, the power amplifier combines the phase split signal from the long-tail pair at the input so that the signal is again a replica of the original signal received from the guitar. But even after the signal recombination at the input of the power amplifier, the two output paths of the differential power amplifier again cause a new 180 degree phase shift to the signals that eventually propagate towards the loudspeaker.

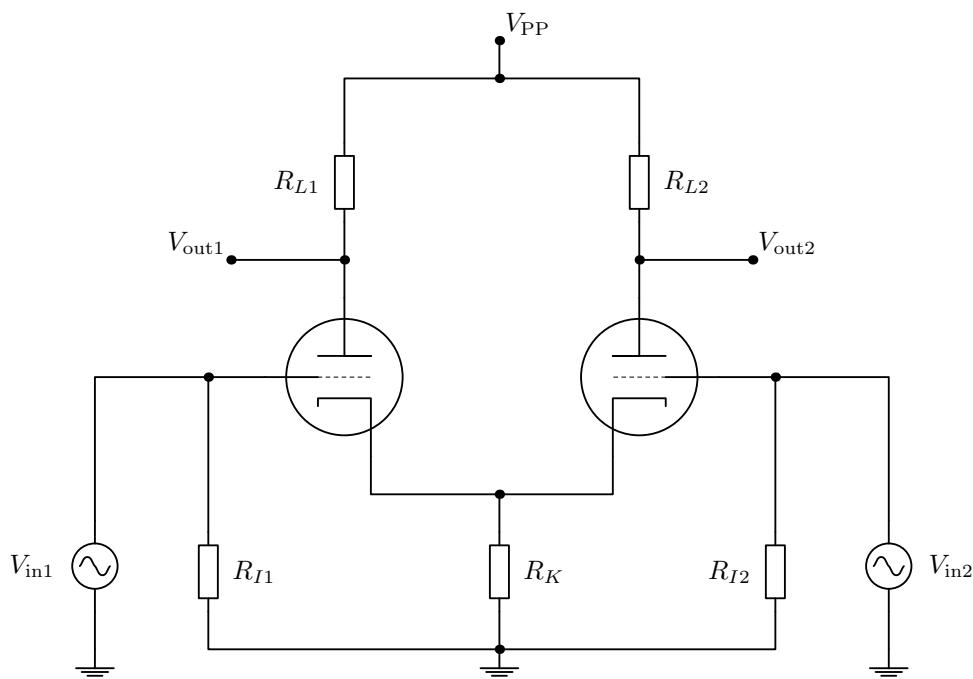


Figure 7.19: A textbook model of a differential amplifier with two inputs

Luckily there is that transformer coupling between the loudspeaker and the power amplifier. Because the transformer coil is centre-tapped at the power amplifier side, the recombination of the differential output signal happens at the transformer almost automatically. For common-mode signals, the currents in the two different plate circuits are flowing in opposite directions, and this creates a magnetic flux to the transformer that in an ideal situation completely cancels itself out. This is because the centre-tapped primary transformer coil is wound in the same direction all the way, but the currents are flowing in opposite directions around the centre-tapped coil in the case of common-mode signals. But since in this case the effective signal is in a differential mode after the differential amplifier, the current flows in the same direction on both sides of the centre tap, and this recreates a complete replica of the input signal to the secondary coil of the transformer. [101, pp. 435–441]

7.5 TRANSISTOR AMPLIFIERS

Transistor amplifiers are more generally categorised as solid-state amplifiers, because integrated circuits such as operational amplifiers are often used in modern guitar amplifier implementations. For the do-it-yourself fanatics, the solid-state guitar amplifiers are the best first project candidates because they run with relatively low voltages and are therefore safer to build than tube amplifiers. Another

benefit of solid-state amplifiers is that the full functionality of the amplifier can be easily simulated by SPICE to ensure the validity of the design. The best way to learn the basics of solid-state guitar amplifiers is to trace out a full working solution and analyse it further. For this purpose, a cheap but working Aria AB-30 50 watt bass amplifier circuit (made in Korea) was traced directly from the circuit board and drawn into a readable schematic.

After some searches in the wonderful world of the Internet, there seems to be a Japanese company Aria that mainly makes guitars, but they also have an amplifier called Aria AB-30 in the markets. The Aria AB-30 amp taken to pieces in this session had been bought from a Finnish music store in January 1993, and it came with a cheap Fender Precision bass clone manufactured by Washburn. From this background it is reasonable to suspect that this amplifier is just a cheap clone of some unknown amplifier. But since it is cheap, it is also simple and therefore a good candidate for further analysis.

Figure 7.20 shows an illustration of the front panel of the Aria AB-30 amplifier. This amplifier offers only the most essential control knobs to adjust the volume and the tone. It does not get simpler than this.

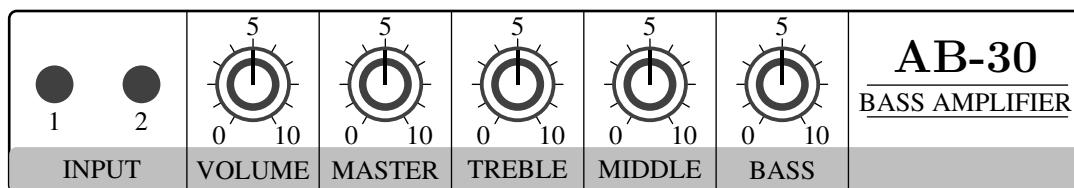


Figure 7.20: The front panel of the Aria AB-30 bass amplifier

In general terms, the upper level block diagram is the same for solid-state guitar amplifiers as it is for the tube amplifiers. There is a preamp followed by a tone control section and a second voltage amplifier section. Then the signal goes to a differential amplifier and finally it reaches the power amplifier section. Most of the differences between solid-state and tube amplifiers are in the preamplifier section and in the power amplifier section.

7.5.1 A solid-state preamplifier

As the tube preamplifier uses discrete tubes as the amplifying elements, it is not convenient to settle for a similar solution in solid-state using discrete transistors in the preamplifier. The gain is most easily obtained using operational amplifiers because they are almost ideal amplifier components, and they can offer a large

gain with relatively small operating voltages.

Typically a dual op-amp IC is used as a connection to the volume and master volume potentiometers at the user interface side. The first op-amp circuitry takes care of the prevolume control, and this stage is commonly followed by a tone control section. After the tone control, the second op-amp circuit handles the master volume settings and delivers the amplified signal to the differential amplifier preceding the power amplifier circuitry.

Figure 7.21 shows the schematic of the whole preamp + tone control section of the solid-state guitar amplifier under investigation. The component values for the schematic are listed in Appendix D. The preamplifier stage is quite similar to the one presented and analysed by Rod Elliott [102].

The connection configuration to obtain different input impedances to the two input jacks is similar to the input configuration in a tube amplifier. The switches that open up when a guitar cord is plugged in make sure that the two input resistors connect in series or in parallel, depending on the input jack used.

Because the op-amps are so ideal devices, there is not that much to analyse in the preamplifier from the scientific point of view. Transfer functions for both op-amp sections can be easily derived using the general formula

$$G = 1 + \frac{Z_F}{Z_G}$$

for the gain of the non-inverting operational amplifier. The term Z_F refers to the feedback impedance that connects the output to the negative input and Z_G to the grounding impedance that connects between the negative input and ground.

For the first op-amp section, the impedance expressions are identified as

$$Z_F = \frac{R_{103}}{1 + sC_{103}R_{103}} \quad \text{and} \quad Z_G = \frac{1}{sC_{105}} + R_{106} + R_v,$$

where R_v is the resistance of the volume control potentiometer. These impedances lead to the transfer function

$$H(s) = 1 + \frac{sC_{105}R_{103}}{s^2C_{105}C_{103}R_{103}(R_{106} + R_v) + s(C_{103}R_{103} + C_{105}[R_{106} + R_v]) + 1}.$$

If a sinusoidal signal is used as a test signal, the Laplacian variable s can be substituted with the complex angular frequency term $j\omega$. With this substitution and evaluation of the transfer function with a range of input frequencies, the frequency response of the first op-amp section with the volume control becomes as

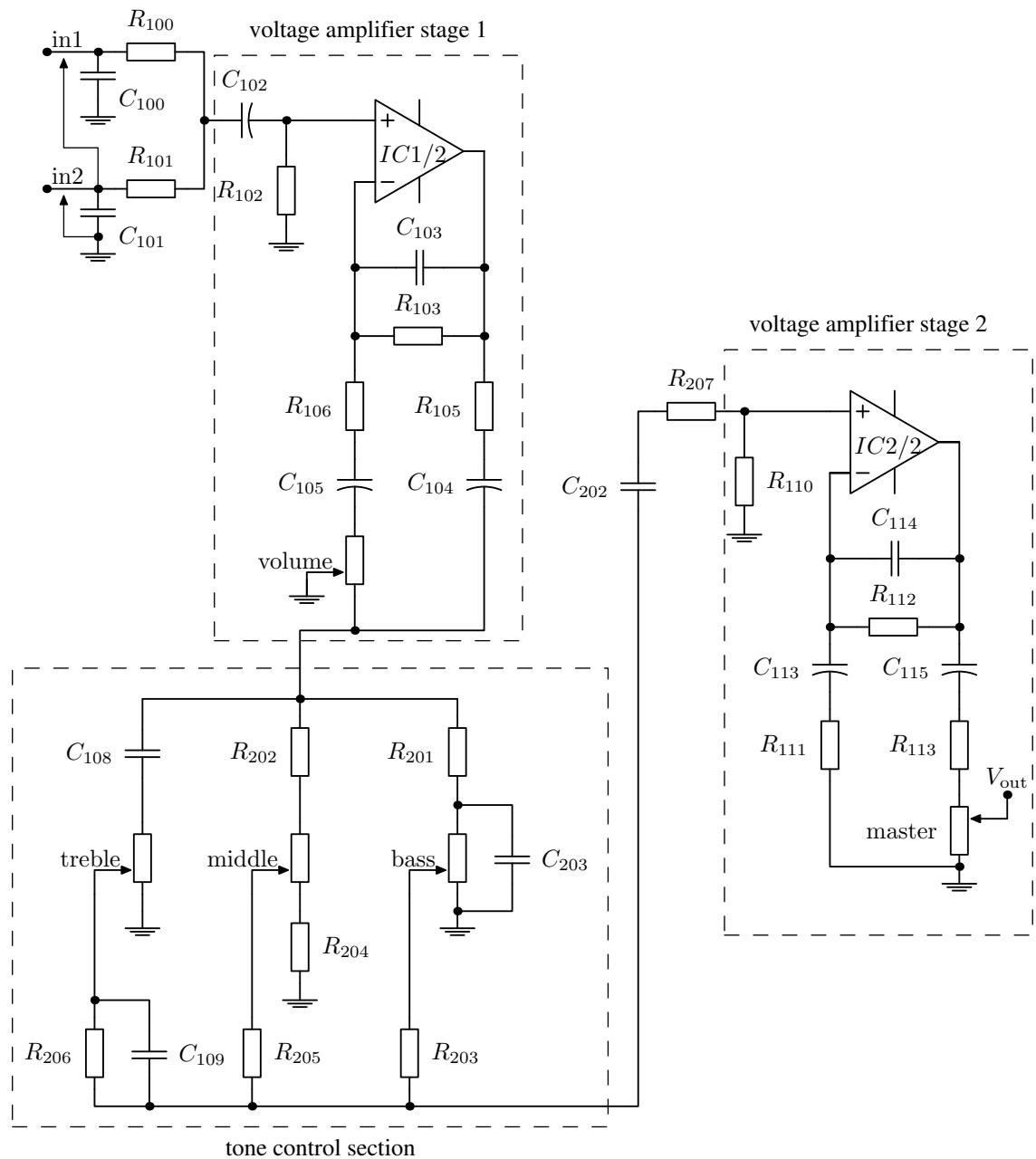


Figure 7.21: The preamp and tone control section of Aria AB-30 bass amplifier

shown in Figure 7.22. The response curves have been drawn for three different values of the volume control potentiometer to show how the gain depends on the volume knob settings. Clearly the volume control potentiometer should be a logarithmic one, since the gain does not change linearly with the potentiometer resistance. Also, the bandwidth of the first preamp stage changes when the volume control potentiometer is adjusted. Considering the fact that this is a bass amplifier, the low-frequency cut-off could be a bit lower with full volume, but then again there might be issues when the amplifier is connected to the loudspeaker system, which might have some unwanted resonances at those very low

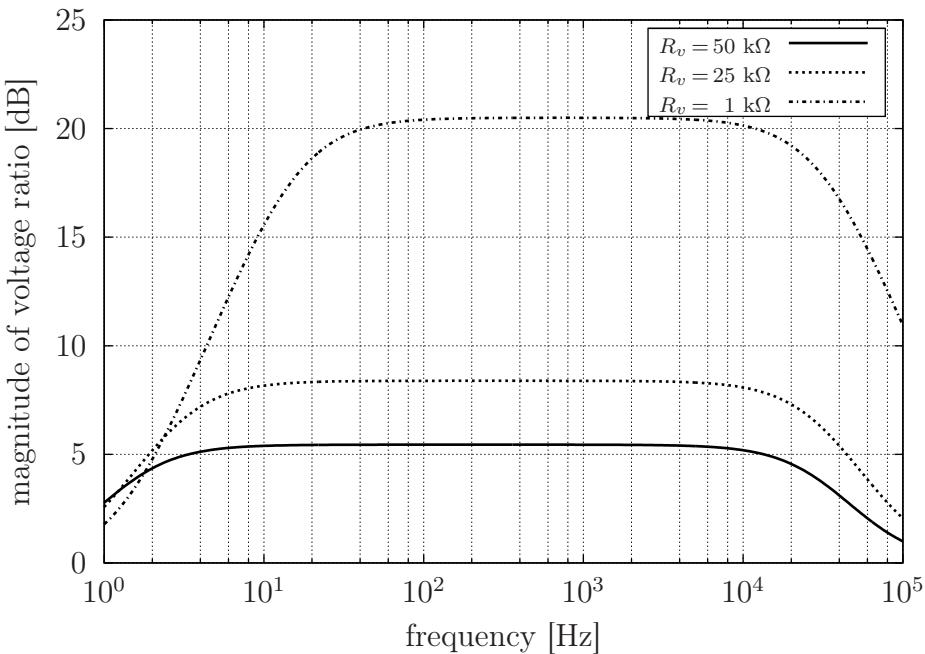


Figure 7.22: The gain of the first op-amp stage with varying volume control

frequencies.

After the first gain stage comes yet another variation of a very simple tone control system. This simple tone control setup can be analysed intuitively without too much details. At the beginning of the tone control stage, the signal is divided to three different propagation paths, one for high frequencies (treble), one for low frequencies (bass), and one for all frequencies (middle).

The treble control is a simple voltage divider (or a high-pass filter), where the high frequencies get more emphasis since the capacitor C_{108} passes the high frequencies with less attenuation. The bass control works with the same principle, but in an opposite fashion: the capacitor C_{203} bypasses the high frequencies to ground and the low frequencies are passed onwards from the potentiometer. The middle frequencies are controlled with a simple voltage divider setup and not filtered at all. At the end of the tone control system, the three signal paths are summed together via the weighting resistors R_{203} , R_{205} and R_{206} .

The second op-amp section has the same construction as the first one and can therefore be directly analysed by replacing the component values into the transfer function of the first op-amp section. The replacements lead to the equation

$$H(s) = 1 + \frac{sC_{113}R_{112}}{s^2C_{113}C_{114}R_{112}R_{111} + s(C_{114}R_{112} + C_{113}R_{111}) + 1},$$

and the visualisation of this equation is presented in Figure 7.23.

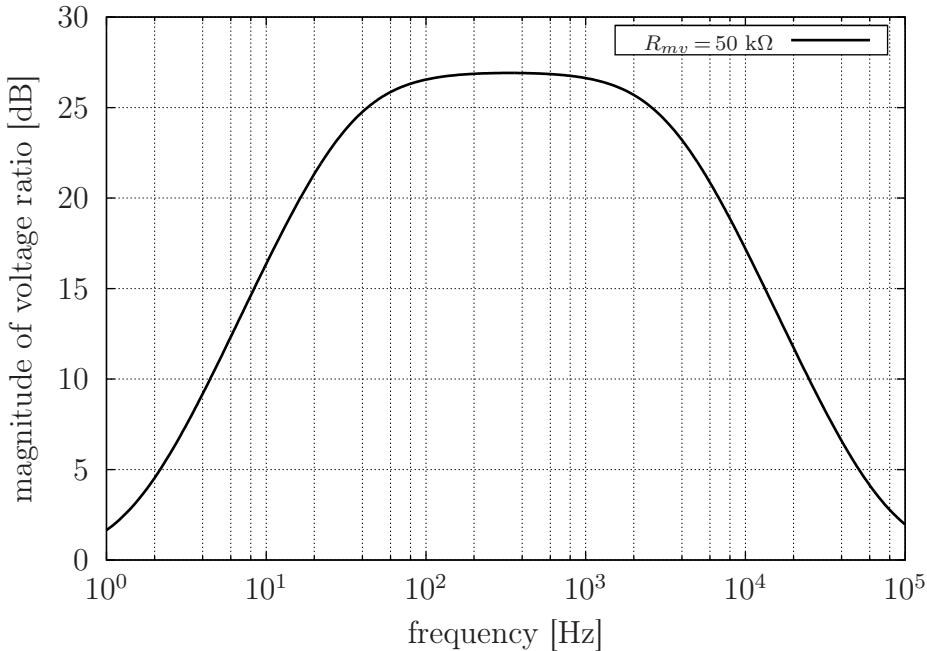


Figure 7.23: The maximum gain obtained from the second op-amp stage (master volume)

According to Figure 7.23, the bandwidth of this second gain stage is narrowed down from the first gain stage by increasing the value of C_{114} . Now the problem is that the high-frequency cut-off comes too early. The tone control section should not affect the two gain sections because the op-amps isolate the tone control from the gain control sections.

7.5.2 A solid-state power amplifier

The most common solid-state power amplifier structure follows the so-called three-stage architecture model [100, p. 32]. In this model the primary circuit element is a basic differential amplifier, which is followed by a voltage gain stage and an output buffer stage. For gain and stability purposes, there almost always exists a negative feedback path from the output buffer to the second input of the differential amplifier.

The analysis of the power amplifier section can proceed the hard way or through a shortcut. This time it is better to choose the shortcut, because from there it is almost too easy to understand the essentials of audio power amplifiers. The key idea is to think that the whole power amplifier stage is replaced by one operational amplifier. This analogy is justified because the input of the op-amp is

a differential amplifier and the output is a current buffer just like in the discrete component power amplifier. The reason for building a power amplifier from discrete components is the need to deliver large currents and maximise heat dissipation. This would not be possible with basic operational amplifier components.

Before digging deeper into the power amplifier schematic, some key ideas need to be explained using the op-amp as an example. As shown in Figure 7.24, the op-amp can be used in an open-loop mode and in a feedback mode.

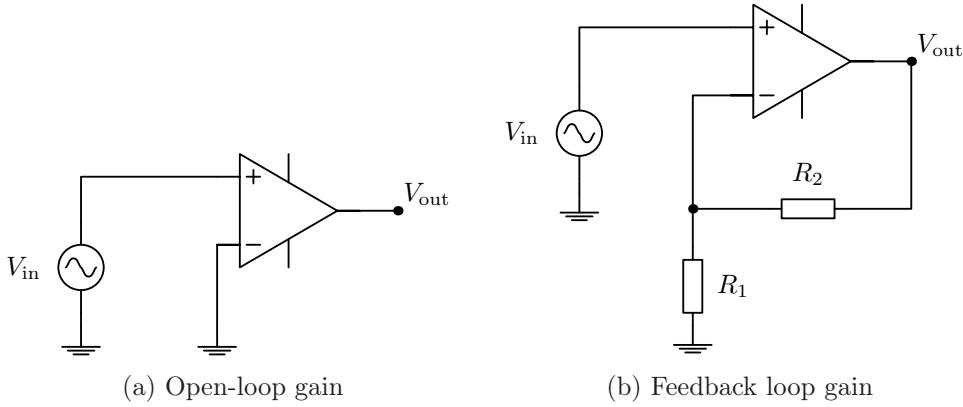


Figure 7.24: Two different use modes of the operational amplifier

The gain of the open-loop mode is huge and depends only on the internal connections inside the op-amp IC. The actual value for the open-loop gain might be very difficult to determine analytically. The huge gain makes the open-loop mode very unstable and prone to malfunction.

The gain of the feedback mode is much more stable and it can be determined solely by the components that make up the (external) feedback loop. There is a requirement relating to this, and in every case it needs to be checked if the requirement is fulfilled. To be able to determine the gain accurately using only the feedback loop, the open-loop gain must be much larger than the feedback gain. This is almost always true in operational amplifiers, but in power amplifiers this might not be the case.

To derive the dependency between the feedback gain and the open-loop gain, one can use the basic inverting op-amp circuit of Figure 7.25 as an example. The open-loop gain parameter A is identified as the internal gain parameter associated with the voltage-controlled voltage source of the op-amp.

After the input voltage source and the voltage-controlled voltage source have been converted to current sources with resistor R_1 and internal output resistance

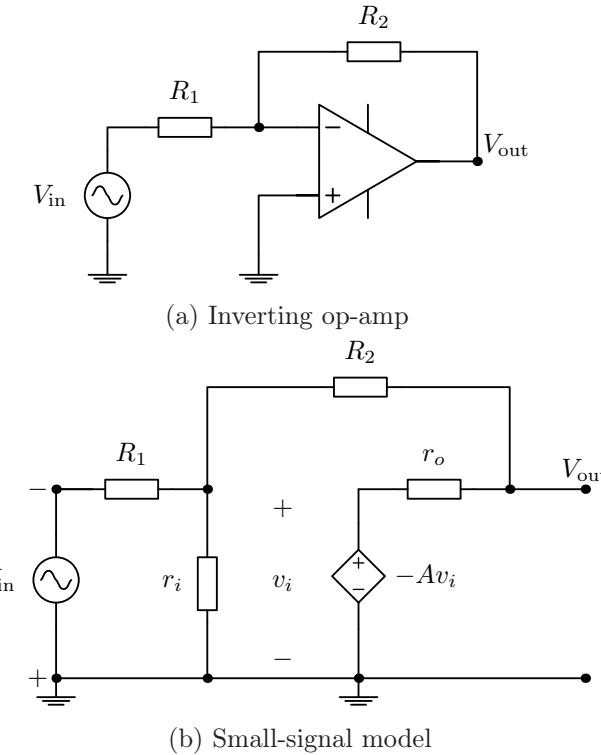


Figure 7.25: The inverting op-amp and its small-signal model

r_o respectively, the small-signal model is represented by the matrix

$$\begin{bmatrix} \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{r_i} & -\frac{1}{R_2} \\ -\frac{1}{R_2} & \frac{1}{R_2} + \frac{1}{r_o} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ -\frac{Av_i}{r_o} \end{bmatrix}.$$

The gain term $-\frac{Av_i}{r_o}$ still needs to be transferred from the current vector to the admittance matrix. Because $V_1 = v_i$, the final form of the matrix equation is

$$\begin{bmatrix} \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{r_i} & -\frac{1}{R_2} \\ -\frac{1}{R_2} + \frac{A}{r_o} & \frac{1}{R_2} + \frac{1}{r_o} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ 0 \end{bmatrix}.$$

Solving for the node voltage V_2 , which is the output voltage V_{out} , yields a fractional gain term

$$\frac{V_2}{V_{in}} = \frac{-R_2 + \frac{r_o}{A}}{R_1 + \frac{1}{A} \left(r_o + R_1 + R_2 + \frac{r_o R_1}{r_i} + \frac{R_1 R_2}{r_i} \right)}.$$

Clearly, if the open-loop gain A is very large, the solution reduces to the well-known formula for the gain of the inverting operational amplifier, where the gain is solely determined by the feedback resistances. But ... if the discrete components inside the amplifier block fail to make A large, then the approximation relying on the feedback gain is not that accurate anymore. This is the dependency between the open-loop gain and the feedback gain. The general goal in designing robust amplifiers is to get A very high in a certain amplifier block so that a feedback loop from the output buffer to the input defines the gain according to the feedback gain approximation.

Now it is the right time to attack the schematics of the Aria AB-30 power amplifier section. This implementation follows the standard three-stage architecture. Figure 7.26 illustrates the differential amplifier consisting of two identical bipolar junction transistors and also the voltage amplifier block, which consists of only a single transistor Q_{104} with a capacitive shunt feedback implemented with capacitor C_{122} .

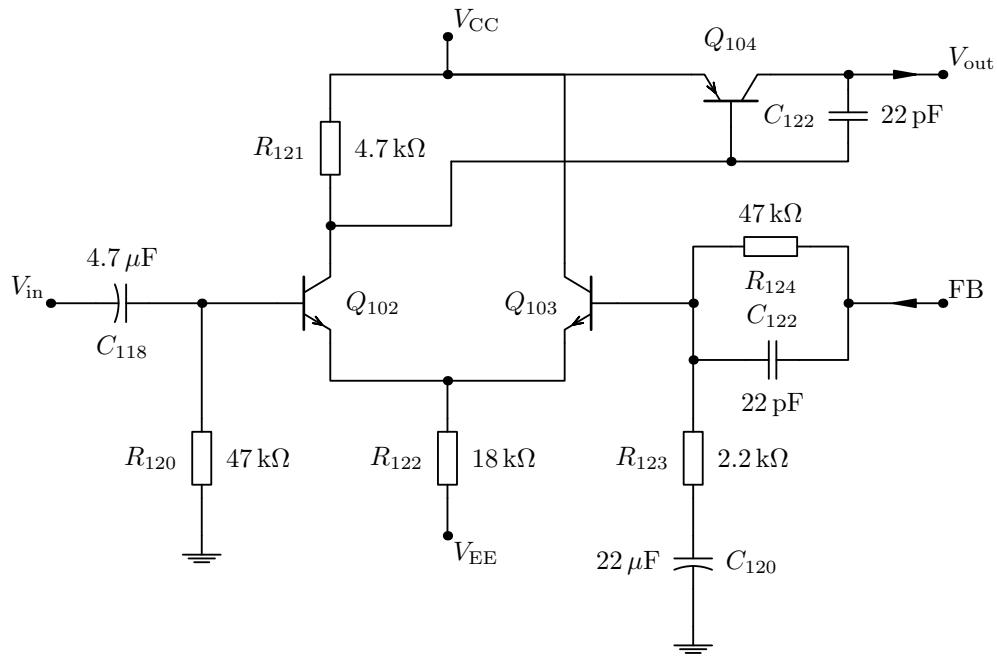


Figure 7.26: The differential amplifier with feedback and voltage amplifier sections

The idea with the feedback capacitor is to make use of the Miller theorem (see section 6.6.4) to stabilise and linearise the frequency response of the amplifier at high frequencies. The voltage gain stage is also called the transimpedance stage, which creates a dominant pole in the frequency characteristics of the amplifier. Since this adds a *local* negative feedback loop to the circuit, it also affects the

global negative feedback loop, but mainly outside the audible frequency range. [100]

The global feedback loop is closed by taking a small part of the output signal to the differential input stage. If the assumption is made that the total gain provided by the differential amplifier, voltage gain transistor and the output buffer stage is high enough to fulfil the requirement of high open-loop gain, the total voltage gain of the power amplifier at audio frequencies is

$$G = 1 + \frac{R_{124}}{R_{123}} = 1 + \frac{47}{2.2} = 22.4 = 27 \text{ dB}. \quad (7.15)$$

This result can be easily verified by a SPICE simulation. Using a general transistor model with $\beta_F = 100$ for each transistor, a 0.1 V input signal to the differential amplifier gives an output voltage swing of 2.2 V. It is very surprising that the gain of the relatively complicated power amplifier can be defined so accurately with only two resistors. However, this approximation can sometimes fail at high frequencies, where the local Miller capacitance feedback makes the global feedback network somewhat frequency dependent [100]. The output buffer stage that connects to Figure 7.26 is shown in Figure 7.27.

The gain formula of the non-inverting op-amp is used in equation (7.15), because the feedback signal taken from the output buffer (Figure 7.27) is in the same phase as the input signal. The phase differences between the input and output signals can be determined by following the signal path from the differential amplifier to the output buffer and by adding up the phase modifications, which mainly come from the amplifier sections. In a general case for different power amplifier stages, it is not directly clear which input of the differential amplifier is the inverting input because the whole three-sectional architecture has an effect on this.

The total gain for the whole amplifier can be calculated as a sum of all individual gains in decibels:

$$G_{\text{tot}} = G_V + G_{MV} + G_P = 20 \text{ dB} + 27 \text{ dB} + 27 \text{ dB} = 74 \text{ dB}.$$

This corresponds to a multiplier of 5000, which means that an input signal of 0.1 volts would be 500 volts at the output. Is this possible? Of course not. Firstly the operating voltages limit the output voltage to a certain level, which is normally less than 30 volts in a typical solid-state guitar amplifier, and signals stronger than this get clipped. Secondly the input configuration and the tone control section attenuate the signal significantly so that the gain multiplier drops somewhere

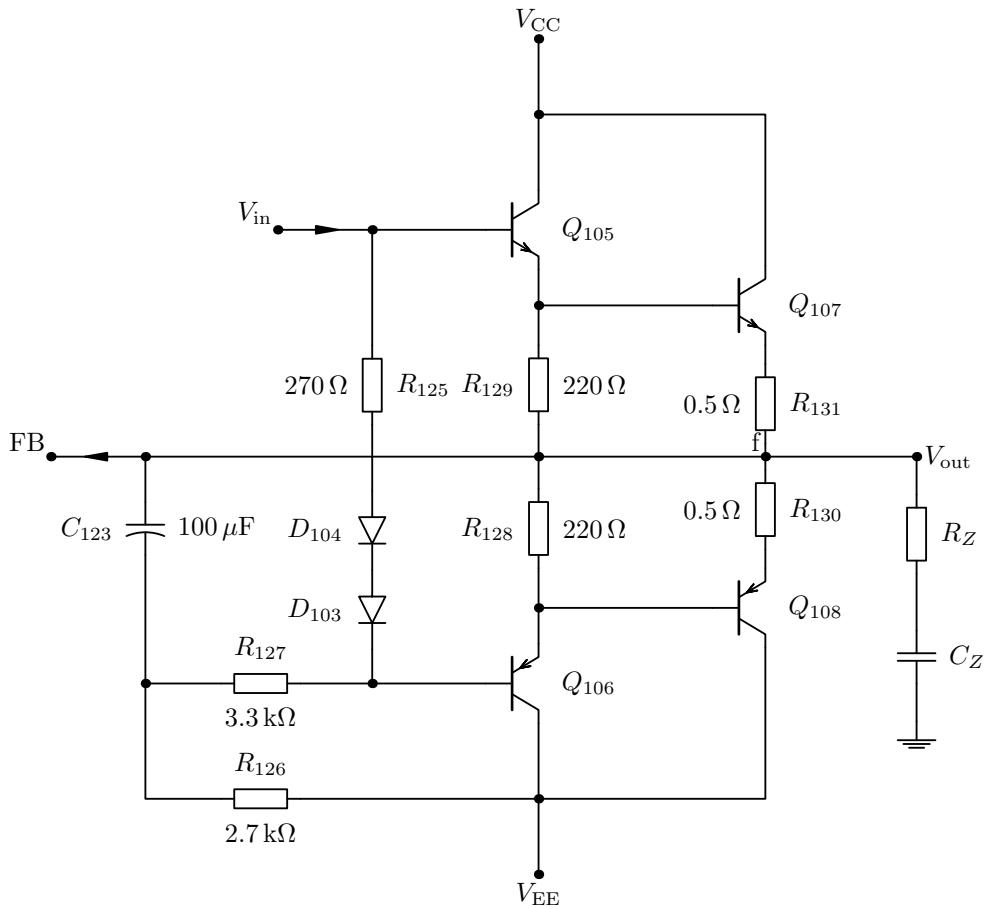


Figure 7.27: The output buffer stage of the solid-state bass amplifier

around 400. This would still mean that a 0.1 volt input signal with full gain might drive the output buffer into saturation and distort the sound.

The output buffer shown in Figure 7.27 is implemented as a quasi-complementary type, which refers to the symmetrical layout of the four output transistors. This kind of setup defines the power amplifier as a class *AB* amplifier where the upper *npn* transistors conduct when the signal is over the V_{BE} voltage drop that forward-biases those transistors. Similarly, the lower *pnp* transistors start to conduct when the signal is below the V_{BE} voltage drop.

To fill the empty gap between $+V_{BE}$ and $-V_{BE}$, the two diodes have been added to bias the transistors properly. The diodes lift the potential of the *npn* transistors so that when the signal at the *pnp* transistors' base pin crosses the $-V_{BE}$ barrier, the signal at the *npn* transistors' base pin crosses the $+V_{BE}$ barrier. This biasing arrangement avoids the crossover distortion that is typical to class *B* amplifiers, although in this specific case the feedback loop is the major solution that also corrects the problem of crossover distortion.

In this Aria AB-30 amplifier, there is an extra feedback loop to the *pnp* transistors that breaks the symmetry of the system. This feedback loop is not familiar from the common literature, so the true meaning of C_{123} , R_{126} and R_{127} remains a mystery. At least they do not break the functionality of the amplifier, but there is a doubt as to whether that extra feedback loop is really necessary.

7.6 INTERFACING AMPLIFIERS WITH LOUDSPEAKERS

As common practise, the output impedance of any solid-state amplifier connecting to moving-coil loudspeakers (see Chapter 8) is designed as low as possible, ideally zero. This makes the amplifier seem almost like an ideal voltage source to the loudspeaker so that the amplifier output is unaffected by loading. This approach tries to linearise the effects of speaker cone resonance and voice coil inductance by making the loudspeaker impedance variation appear to be small compared to the internal frequency response variations of the amplifier. The ratio of output and input impedance can be considered to have a damping factor

$$DF = \frac{Z_{\text{out}}}{Z_{\text{in}}}, \quad (7.16)$$

where large damping factors are beneficial to attenuate the loudspeaker imperfections in the frequency domain. [100, p. 26]

Due to the very low output impedance of a typical amplifier, it is better to avoid connecting long cables with significant resistance between an amplifier and a loudspeaker. From the loudspeaker's point of view, the cable resistance increases the output impedance of the amplifier and all the benefits of the damping methodology are easily lost with careless cabling solutions. [103, pp. 166–174]

The implementation method of low output impedance is not ideal when considering the signal flow (reflections, etc.). The signal flow can be improved by increasing the amplifier output impedance with a suitable series resistor so that the amplifier would act more like a current source. This implementation is not directly compatible with a typical loudspeaker, which is designed to be coupled with a voltage source. The current-drive method also does not reduce the speaker cone resonances, so additional filters should be added at the amplifier side to compensate for the loudspeaker nonlinearities. [104]

In addition to the output impedance compensation method, components R_Z and C_Z in schematic 7.27 form a so-called *Zobel network* [100, p. 198] [105, p. 124], which is intended to compensate for the rising high-frequency impedance of the

loudspeaker voice coil. The Zobel network is not necessarily required to make the amplifier work properly, and therefore the existence and meaning of R_Z and C_Z is only briefly mentioned in this context.

The situation is different when considering tube amplifiers because it is impossible to get a low output impedance from any tube amplifier configuration. The output impedance of tube amplifiers is scaled lower by connecting a transformer between the output tubes of the amplifier and the loudspeaker. However, the common approach has been to scale the output impedance close to true impedance matching, which means that the tube amplifiers are not using the damping factor to compensate for the nonlinearities of the loudspeaker.

7.7 GUITAR AMPLIFIER SIMULATIONS WITH SPICE

At least the transistor based amplifier presented in the previous section can be fully simulated using SPICE. Only the operational amplifiers might need a specific model to indicate possible clipping of the signal. For the sake of simplicity, all the transistors can be modelled as dummy transistor models (only define $\beta_F = 100$ to the component model) if they are used in the power amplifier, difference amplifier and the voltage amplifier stages. The use of dummy models is possible mainly for two reasons: firstly the gain of the power amplifier stage is defined by the feedback loop and not by the gain of individual transistors, and secondly because the SPICE parameters do not define the maximum allowed collector current of the transistors. Therefore, the simulation model does not 'burn' even if several amperes are flowing in the output buffer transistors.

As a concrete example, the Aria AB-30 bass amplifier can be drawn to a single schematic file with `gschem`, translated into a netlist file with `gnetlist`, and simulated with `ngspice`. The only things that need to be added manually to the netlist file are the sub-circuit model for the op-amps and obviously the simulation parameters for a frequency sweep and transient analysis.

The modelling of tube amplifiers in SPICE is a bit more difficult task for a few reasons. The primary reason is that the tube needs to be modelled as a general voltage-controlled current element in SPICE, so a new component model possibly needs to be created. Secondly, the tube amplifiers use transformers in the power amplifier stage, which again causes more challenges to the SPICE model. All of this is of course doable, but it will take some more effort than the modelling of solid-state amplifiers.

An approximate SPICE model for vacuum tubes can be created from the empirical 'three-halves-power' law that imitates the dependency between the plate current and grid voltage as

$$i_P = g_m \left(v_G + \frac{v_P}{\mu} \right)^n, \quad (7.17)$$

where n usually is very close to $\frac{3}{2}$. In a small-signal mode of operation, the exponent n would have a value 1, which linearises the plate current equation. In equation (7.17), g_m refers to the transconductance and μ is the amplification factor so that the quotient $\frac{g_m}{\mu}$ represents the inverse of the dynamically changing internal plate resistance r_p .

A relatively detailed discussion on vacuum tube SPICE models is given by M. Leach [106], but unfortunately the research paper does not explain the origins of the model very thoroughly. In the research paper the power law (7.17) is written in the form

$$i_P = K (\mu v_G + v_P)^n, \quad (7.18)$$

where K is used for denoting the quotient $\frac{g_m}{\mu}$. The parameters K and μ are used as basis for the given tube SPICE models. As an example from the research paper, a triode equivalent SPICE model is shown in Figure 7.28.

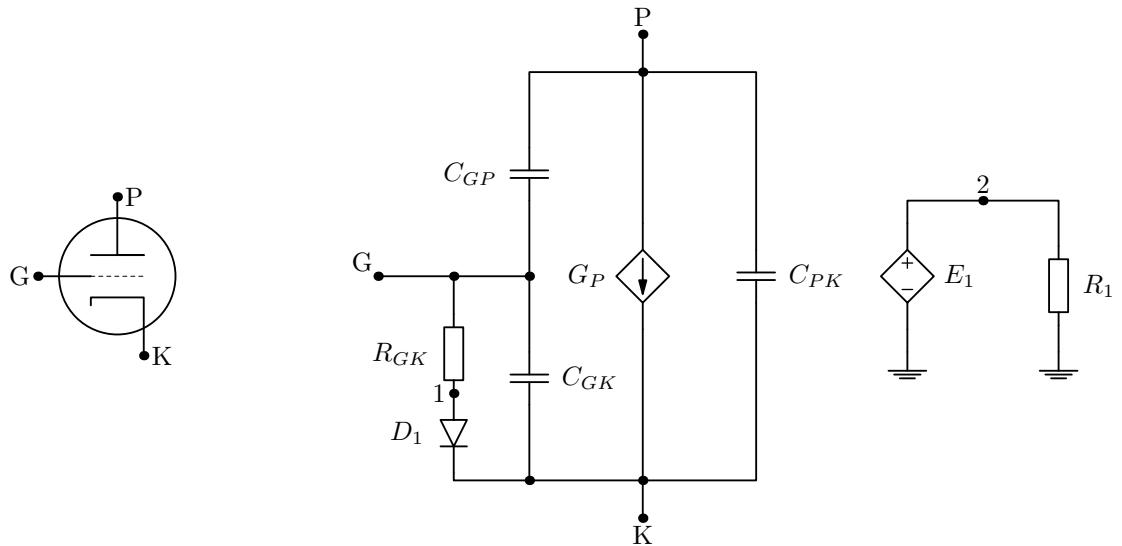


Figure 7.28: A triode on the left and its SPICE equivalent model on the right

The general form of the SPICE netlist corresponding to Figure 7.28 is

```
* Connections: Plate
* | Grid
```

```

*           | | Cathode
*
*.SUBCKT TRIODE P G K
E1 2 0 VALUE={V(P,K) + "u"*V(G,K)}
R1 2 0 1.0K
Gp P K VALUE={"K)*(PWR(V(2),1.5) + PWRS(V(2),1.5))/2}
RGK G 1 "Rgk"
D1 1 K DM
.MODEL DM D
CGK G K "Cgk"
CGP G P "Cgp"
CPK P K "Cpk"
.ENDS TRIODE

```

where G_p and E_1 are controlled current and voltage sources as defined by the SPICE specifications.

The model described by Leach is quite close to the regular small-signal model of a triode, but with the exception that the effects of the internal variable resistance r_p are 'hidden' to the separated controlled voltage source circuit. This is because the value of r_p depends on the plate current in each case, and therefore this arrangement makes it possible to dynamically determine the value of r_p from the plate current.

The website <http://tdsl.duncanamps.com/dcigna/tubes/spice/> offers some SPICE models for the most common tubes, all of which are more or less based on the model provided in the research paper [106]. As an example, the model for a 12AX7A vacuum tube without the diode branch is

```

* Connections: Plate
*           | Grid
*           | | Cathode
*           | | |
*.SUBCKT 12AX7A P G K
E1 2 0 VALUE={45+V(P,K)+95.43*V(G,K)}
R1 2 0 1.0K
Gp P K VALUE={1.147E-6*(PWR(V(2),1.5)+PWRS(V(2),1.5))/2}
Cgk G K 1.6P
Cgp G P 1.7P
Cpk P K 0.46P
.ENDS 12AX7A

```

The model seems to be very simple to use, but troubles may arise with different

SPICE versions because of the use of the nonlinear controlled sources Gp and $E1$. This model is intended to be used with PSpice, but for example ngspice uses a slightly different syntax for the nonlinear dependent sources, so it is advised to read the SPICE manual first before simulations. The model presented above does not make use of the diode D_1 and resistor R_1 presented in Figure 7.28, and therefore this model is restricted to situations where $v_{gk} < 0$.

The given model was tested against an example provided in the research paper [106]. The test circuit is shown in Figure 7.29.

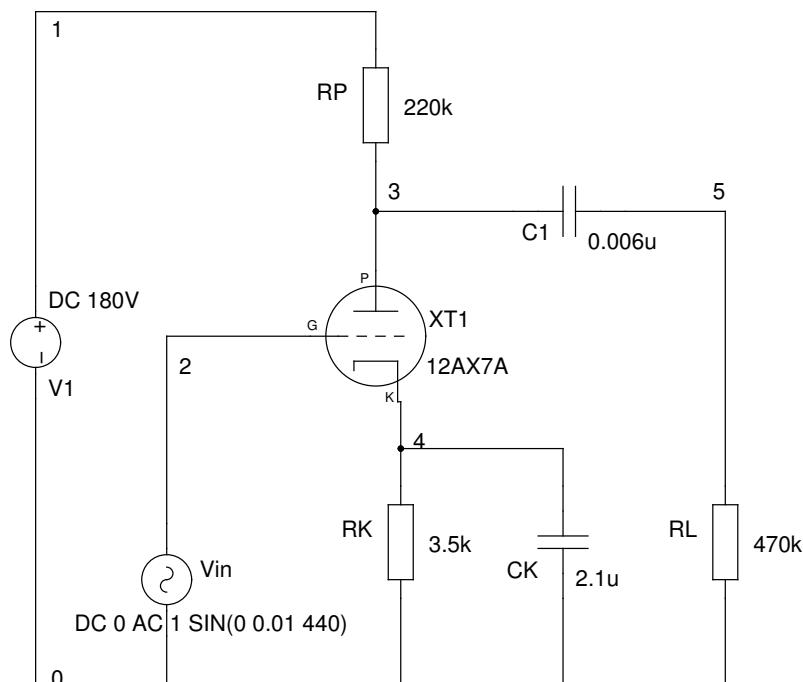


Figure 7.29: A simple tube amplifier using the 12AX7A vacuum tube

The netlist file created from the schematic 7.29 is given below as a complete listing including also the simulation control commands. This example introduces how the nonlinear controlled source expressions in the 12AX7A sub-circuit are modified to suite the ngspice command structure.

```
***** Begin SPICE simulation control section *****
.control
.ac dec 90 10 10000K
```

```

.endc
=====
 Begin SPICE subcircuit definitions =====
.SUBCKT 12AX7A P G K
E1 2 0 vol='45+V(P,K)+95.43*V(G,K)'
R1 2 0 1.0K
Gp P K cur='1.147E-6*( V(2)**1.5 + (V(2)**1.5)*V(2)/abs(V(2)) )/2'
Cgk G K 1.6P
Cgp G P 1.7P
Cpk P K 0.46P
.ENDS
=====
 Begin SPICE netlist of main design =====
RK 0 4 3.5k
C1 5 3 0.006u
CK 0 4 2.1u
RL 0 5 470k
XT1 3 2 4 12AX7A
Vin 0 2 DC 0 AC 1 SIN(0 0.01 440)
V1 1 0 DC 180V
RP 3 1 220k
.end

```

The test circuit shown in Figure 7.29 is a very simple amplifier circuit, from where it is also possible to derive a small-signal model for numerical matrix calculations. The goal is to validate the correctness of the given SPICE model of 12AX7A by comparing the frequency response to results from the numerical calculation based on the general tube small-signal model. Figure 7.30 illustrates the small difference between the numerical results and the SPICE model simulation.

What is surprising in Figure 7.30 is the very large bandwidth of the frequency response curve. Many would possibly have made other assumptions on the bandwidth of tube amplifiers.

The pentode tube SPICE model is more complex compared to the triode model. The three-halves-power equation for pentodes is

$$i_P = aK (\mu_c v_G + \mu_s v_S + v_P)^n, \quad (7.19)$$

for the plate current and

$$i_S = (1 - a)K (\mu_c v_G + \mu_s v_S + v_P)^n, \quad (7.20)$$

for the screen grid current. The parameter a in the previous equations describes a fraction of the current in the plate. Typically a depends on the voltage ratio

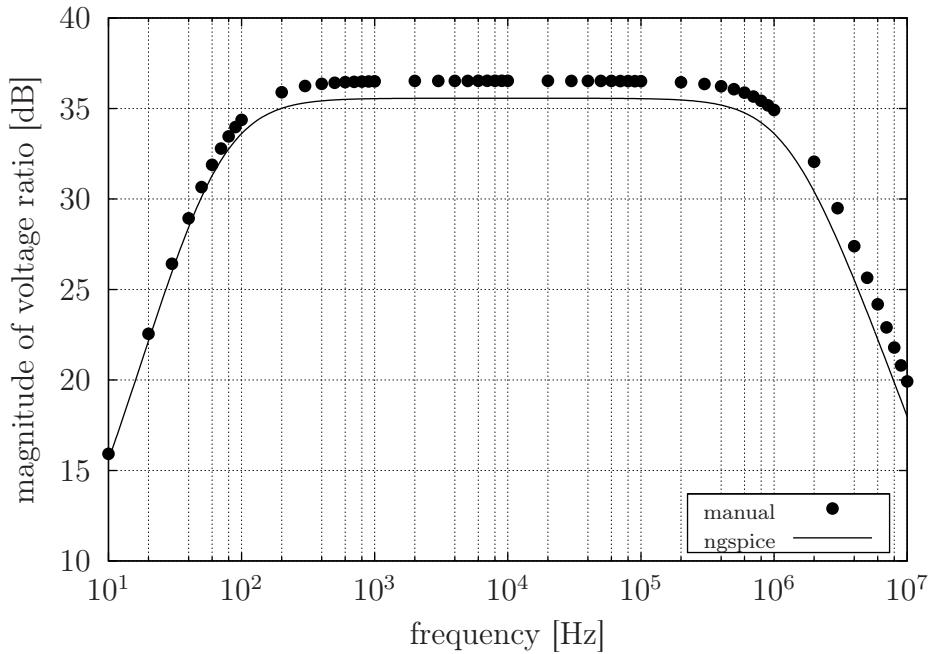


Figure 7.30: A SPICE simulation versus a manual numerical calculation of 12AX7A

$\frac{v_P}{v_S}$, but for simplicity a can be taken as a constant value for each tube. The two separate amplification factors μ_c and μ_s relate to the control grid and the screen grid respectively. Typically μ_s is a lot smaller than μ_c . [106]

Another example given by Leach is the pentode equivalent SPICE model, which is shown in Figure 7.31.

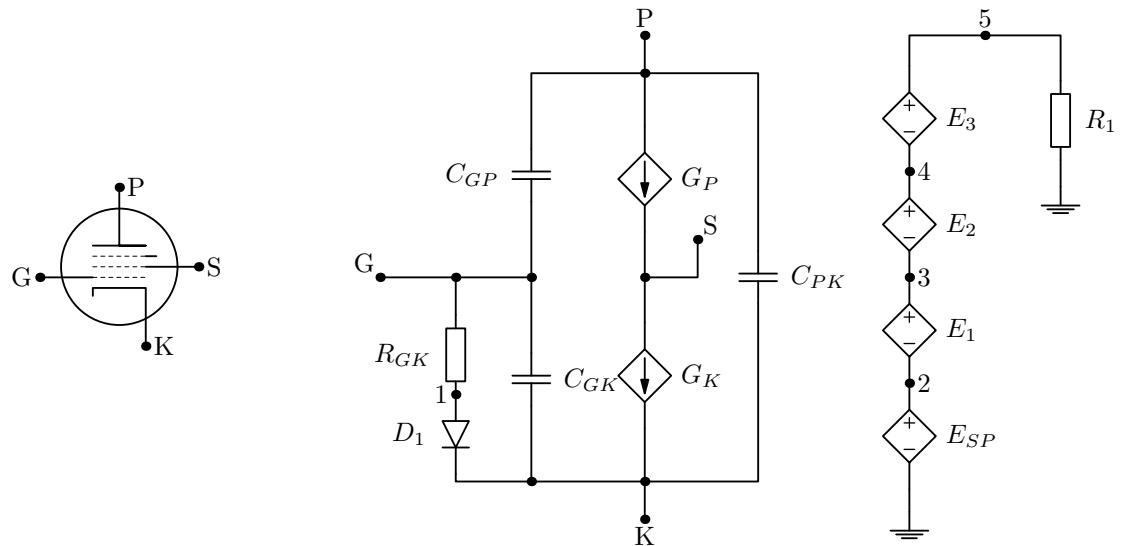


Figure 7.31: A pentode on the left and its SPICE equivalent model on the right

The general form of the SPICE netlist corresponding to Figure 7.28 is

```
* Connections: Plate
*           | Screen
*           | | Grid
*           | | | Cathode
.SUBCKT PENTODE P S G K
RGK G 1 "Rgk"
D1 1 K DM
.MODEL DM D
ESP 2 0 VALUE={V(P,K) + "us"*V(S,K) + "uc"*V(G,K)}
E1 3 2 VALUE={"K)*(PWR(V(2),1.5) + PWRS(V(2),1.5))/2}
E2 3 4 VALUE={"K)*(PWR("us"*V(S,K),1.5)*V(P,K)/"Va"}
E3 5 4 VALUE={(1 - V(4,2)/ABS(V(4,2) + .001))/2}
R1 2 0 1.0K
Gk S K VALUE={V(3,2)}
Gp P S VALUE={"a)*(V(3,4)*(1-V(5,4)) + V(3,2)*V(5,4)}
CGK G K "Cgk"
CGP G P "Cgp"
CPK P K "Cpk"
.ENDS PENTODE
```

where V_a determines the plate-cathode voltage at the point where the slope of the curve takes a sharp turn. In this model, the characteristics of a certain tube are defined by setting values to a , u_c , u_s , K and V_a . From the tube datasheets it might be troublesome to get the correct values for a and V_a , so the use of this model is not that straightforward as the triode model. The pentode model is not used in this context because the author does not fully understand the pentode model.

Some predefined pentode models using this approach are given on the website <http://tdsl.duncanamps.com/dcigna/tubes/spice/> and they can be used directly by assigning the correct pin ordering to the predefined sub-circuit. However, the presented tube models are not super-accurate since they are based on purely empirical mathematical models. By using the presented models for the triode and the pentode, it is possible to execute approximative SPICE simulations for tube based guitar amplifier circuits.

CHAPTER 8

LOUDSPEAKERS

The last obstacle in the guitarist's signal chain is the loudspeaker. Loudspeakers are responsible for converting the electric signal back to the acoustic domain so that the notes played on the guitar are actually heard by the audience. Therefore, loudspeakers are essential devices in the signal chain and they definitely cannot be bypassed if problems occur. The loudspeaker consists of electrical and mechanical interfaces, and additionally an acoustical interface. One can only hope that the loudspeaker acts as a linear transducer ...

The loudspeaker element used to accompany guitar amplifiers is called a *moving-coil* loudspeaker describing the internal construction of the loudspeaker element. It might be the most common type of a loudspeaker element, because they are also used with all home entertainment system loudspeakers. The general word loudspeaker refers to the mechanical loudspeaker element and the acoustical box surrounding the element. In this context, the focus is on the loudspeaker element only, neglecting the acoustical effects of the box. In professional terms the loudspeaker element is called a *driver*, the flat board where the driver is mounted is called a *baffle* and the actual loudspeaker cabinet is referred to as an *enclosure*. A *loudspeaker system* is a combination of driver + baffle or driver + enclosure [107]. Theoretically the term *infinite baffle* means a very large baffle, but in some cases it can also refer to the enclosure cabinet [103, p. 65].

8.1 THE CONSTRUCTION OF A MOVING-COIL LOUDSPEAKER ELEMENT

A sliced model of a complete moving-coil loudspeaker element is visualised in Figure 8.1, where most of the essential parts of the driver are identified. The voice coil is swirled between a radial magnet structure and it is also in direct contact with the speaker cone horn. The speaker cone is connected to a sup-

porting metal structure with elastic joints, which are modelled as springs in this illustrative sketch. The centre part of the speaker cone is typically protected by a metallic dust cap.

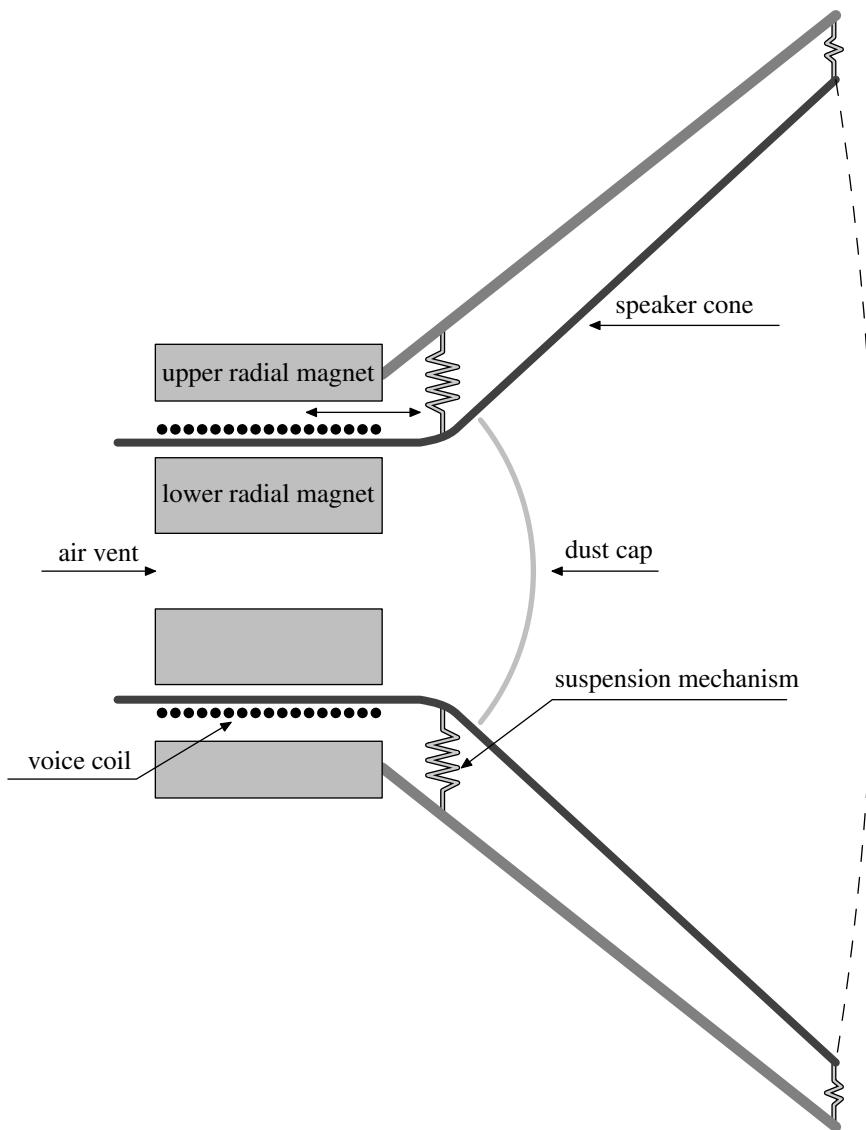


Figure 8.1: A sketch of a basic moving-coil driver

The permanent magnets create a radially spreading magnetic field. This magnetic field is oriented perpendicularly to the voice coil, which is placed between the magnets. When current is flowing in the voice coil, a mechanical force is affecting the coil and since the voice coil is loosely attached, it will move along with the force. Furthermore, the coil is connected mechanically to the speaker cone, which will replicate the motion of the voice coil. Eventually the moving speaker cone interacts with air to create acoustical wave motion.

A good starting point for modelling the loudspeaker element is the interaction

between the voice coil and the radial magnets. Figure 8.2 shows an idealised sketch of the coil-magnet system.

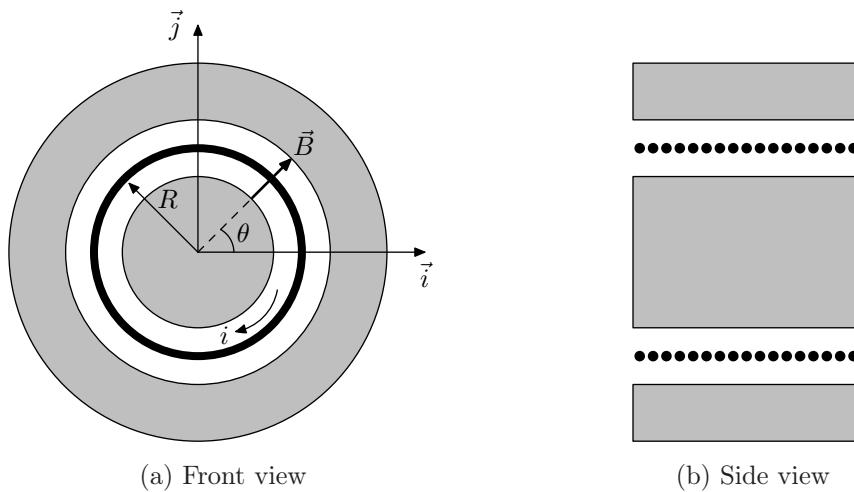


Figure 8.2: A voice coil in a nonuniform magnetic field

The force affecting the voice coil can be calculated analytically using the vectorised equation

$$\vec{F} = i \vec{l} \times \vec{B},$$

which in a differential form is written as

$$d\vec{F} = i d\vec{l} \times \vec{B}. \quad (8.1)$$

To calculate the length of the coil under the influence of the magnetic field, the geometry of the coil can be vectorised with the equation

$$\vec{l} = R \cos \theta \hat{i} - R \sin \theta \hat{j},$$

using the notations given in Figure 8.2. The direction of rotation is defined by the direction of the current i , which is now chosen to circulate clockwise in the coil. The differential of the equation of vectorised length evaluates as

$$d\vec{l} = -R \sin \theta d\theta \hat{i} - R \cos \theta d\theta \hat{j},$$

and the magnetic field can also be expressed using the same angle θ as

$$\vec{B} = B_0 \cos \theta \hat{i} - B_0 \sin \theta \hat{j}.$$

The reason for choosing the signs this way is to get a positive result as the answer, but basically it is not necessary to start twiddling the signs.

Now everything is in place to use equation (8.1) to calculate the magnitude and the direction of the force affecting the voice coil. Solving the cross product

$$d\vec{l} \times \vec{B} = \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ -R \sin \theta \, d\theta & -R \cos \theta \, d\theta & 0 \\ B_0 \cos \theta & -B_0 \sin \theta & 0 \end{bmatrix} = B_0 R \, d\theta \, \hat{k},$$

which means that the force will be directed only towards the direction of the unit vector \hat{k} . The magnitude is obtained by integrating both sides of equation (8.1) so that the limits in the right integral lead to the real length of the coil:

$$F = i B_0 R \int_0^{N2\pi} d\theta.$$

Eventually the most basic form

$$F = i B_0 l, \quad (8.2)$$

is recovered to model the force that affects the voice coil. Using the electromechanical transformation factor N to multiply both sides of equation (8.2), the relation

$$v = c B_0 l \quad (8.3)$$

between the voltage v and velocity c is obtained. The factor $B_0 l$ cross-couples the electromechanical variables in a very evil way, causing lots of harm to maintaining a robust derivation of the electromechanical model of the driver.

It is better to build the driver model piece by piece and try to explain the details understandably. The easiest part is to create the electrical model for the voice coil of the loudspeaker element. Since the voice coil is already an electrical component, it is easily modelled as an ideal inductor with a resistor representing the internal resistance of the coil wire.

8.2 MODELLING THE VOICE COIL

But wait, things are not that easy, because the voice coil is definitely not an ideal inductor. Even if the resistance of the coil wire is modelled as a separate ideal resistor R_0 , which represents the nominal impedance of the driver, it is not enough to designate the remaining inductance as an ideal inductor [108]. The presence of strong magnets used in loudspeaker elements makes the inductance

of the voice coil behave nonlinearly as far as the frequency response is concerned. Analysis of measurement data obtained from real world drivers has provided an empirical model for the voice coil equivalent circuit. A suitable equivalent circuit to simulate the impedance of the voice coil is shown in Figure 8.3.

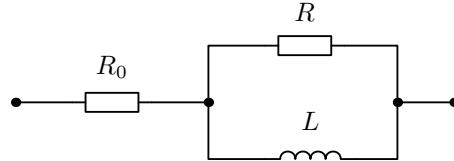


Figure 8.3: An empirical voice coil circuit

The total impedance of the equivalent circuit shown in Figure 8.3 is

$$Z = R_0 + \frac{sL}{1 + \frac{sL}{R}},$$

which yields a step-like impedance increase defined by the parallel system consisting of a resistor and an inductor.

In reality the impedance of the loudspeaker voice coil increases more steadily with frequency than one individual step at some specific frequency. Therefore, a better approximation would be to use several parallel combinations of inductors and resistors connected in series [104, p. 135]. Figure 8.4 illustrates the ideology to combine the impedance sections together. The transfer function in this case would be

$$Z = R_0 + \frac{sL_1}{1 + \frac{sL_1}{R_1}} + \frac{sL_2}{1 + \frac{sL_2}{R_2}} + \frac{sL_3}{1 + \frac{sL_3}{R_3}}.$$

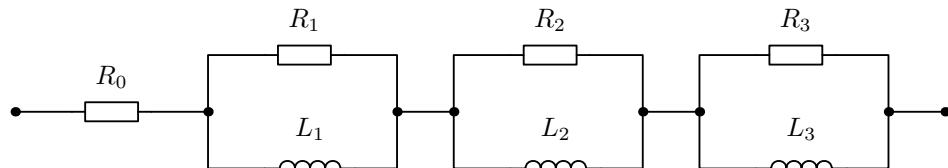


Figure 8.4: A set of voice coil models connected in series

Each individual parallel impedance block changes the total impedance at a certain frequency range. If the resistance of the parallel connection is altered, the step-like impedance variation can be moved in the frequency axis to the desired

location. As shown in Figure 8.5, increase in the resistor value increases the frequency where the step increase in the total impedance is obtained. Alternatively, increase in the inductor value gives an increase to the total impedance.

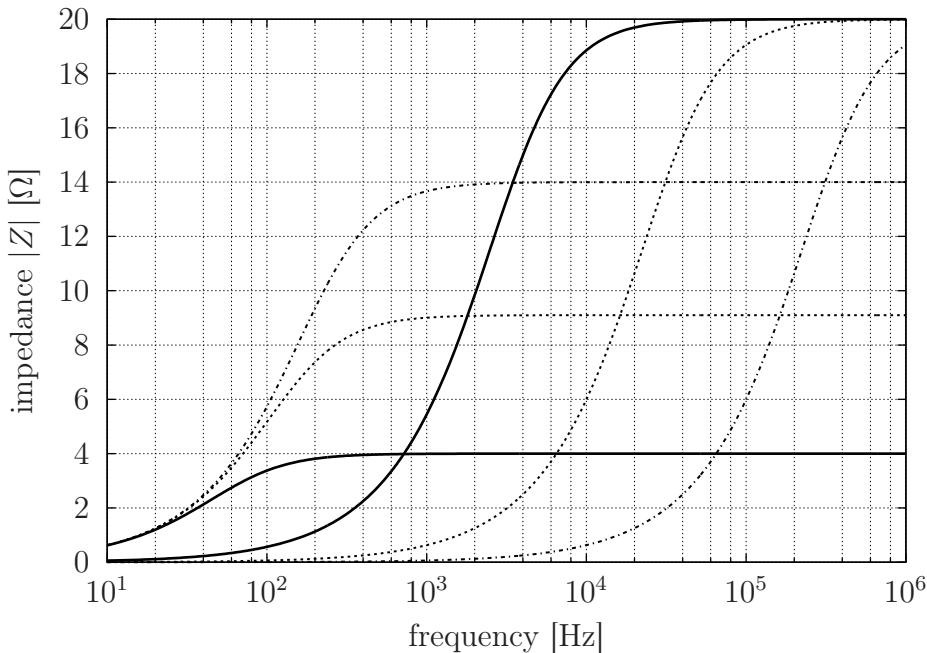


Figure 8.5: Altering resistance R and inductance L in the parallel configuration

A suitable combination of step increases at certain frequencies builds up an exponentially increasing total impedance as shown in Figure 8.6, where the influence of three different parallel circuits have been plotted as cumulative curves. The impedances in Figure 8.6 are

$$Z_1 = \frac{j\omega 180 \cdot 10^{-6}}{1 + \frac{j\omega 180 \cdot 10^{-6}}{2}} \quad ; \quad Z_2 = \frac{j\omega 100 \cdot 10^{-6}}{1 + \frac{j\omega 100 \cdot 10^{-6}}{7}} \quad ; \quad Z_3 = \frac{j\omega 60 \cdot 10^{-6}}{1 + \frac{j\omega 60 \cdot 10^{-6}}{20}}.$$

This possibility to empirically imitate the impedance of any loudspeaker voice coil is what is obtained by using a chain of parallel connected systems of a resistor and inductor. To create a specific analytical model for a certain loudspeaker, some data from impedance measurements are needed. If no measurement data is available, then using a voice coil model of one inductor with approximate inductance L is better than nothing.

8.3 ELECTROMECHANICAL ANALOGUES REVISITED

The loudspeaker consists of electrical, mechanical and acoustical interfaces, but through electromechanical and electroacoustical analogues the whole loudspeaker

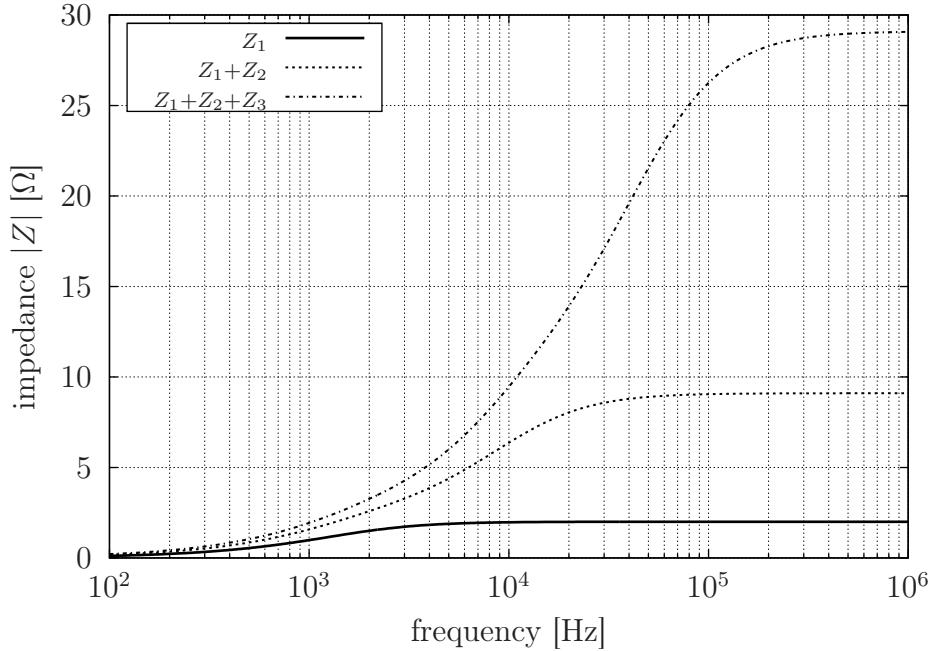


Figure 8.6: Total impedance due to cumulatively added parallel RL blocks

can be modelled by a single electric circuit, which consists of resistors, capacitors, inductors and a signal source. Unfortunately the whole concept of electromechanical analogues is unclearly explained in related literature; at least the author has been very confused about connecting a mass-damper-spring system to a resistor-capacitor-inductor system. This section presents a methodology that should be easy to understand, but it might be in conflict with standard textbooks.

The main idea in identifying an electrical equivalence of a mechanical quantity is to find the correct impedance pairs between electrical and mechanical models and identify the need to use admittance instead of impedance. These equivalences can be approached directly through the electrical circuit theories, which can be extended to cover mechanical models as well.

Figure 8.7 introduces series- and parallel-connected RLC circuits. Using the notations provided in the figure, the series-connected RLC circuit is represented by the differential equation

$$L_s \frac{di_s}{dt} + R_s i_s + \frac{1}{C_s} \int i_s dt = v_s.$$

After applying the Laplace transform, the differential equation is written as

$$\left(sL_s + R_s + \frac{1}{sC_s} \right) i_s = v_s, \quad (8.4)$$

where the three terms inside the parentheses represent impedances from each

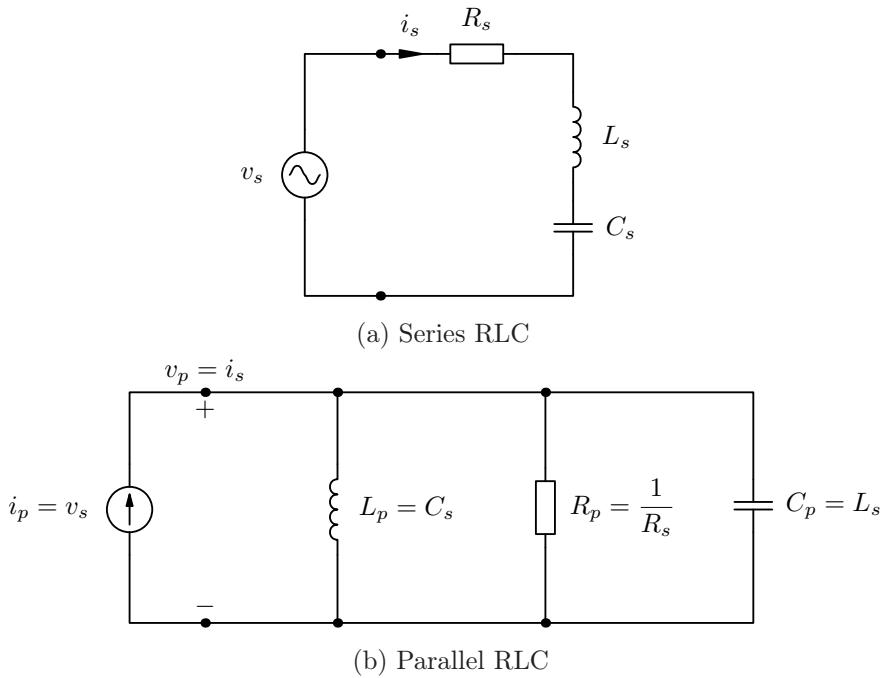


Figure 8.7: Electrical networks representing the duality of impedance and admittance models

individual component. If the units are written open, every impedance term has the units of ohm. This equation relates the current i_s with the voltage v_s via the impedance terms. Essentially in this case, the current has been *chosen* as the *flowing* quantity in the circuit, and the current is also the quantity that has an equal magnitude for all components in the circuit. Circuits of this type are commonly solved using the mesh analysis method.

If the parallel-connected RLC circuit is analysed, it is obvious to choose the nodal analysis method, which uses admittances instead of impedances. This follows from the differential equation

$$C_p \frac{dv_p}{dt} + \frac{1}{R_p} v_p + \frac{1}{L_p} \int v_p dt = i_p,$$

where voltage has now taken the place of the flow-type variable. Voltage is also seen as an equal magnitude by all components in the circuit. The Laplace transformed differential equation reads

$$\left(sC_p + \frac{1}{R_p} + \frac{1}{sL_p} \right) v_p = i_p, \quad (8.5)$$

where the three terms inside the parentheses represent admittances from each individual component. The components in the parallel circuit of Figure 8.7 have been assigned with values which relate them to the series circuit. By substituting

these values into equation (8.5), an equation identical to (8.4) is obtained. This explains the concept duality, which at first thought is not that obvious.

The duality of the series and parallel circuits can be examined in a more general fashion by rewriting the impedance and admittance terms for the series and parallel circuits. For the series circuit, the impedance term can be written as

$$Zi = L \left(s + \frac{R}{L} + \frac{1}{sLC} \right) i = v$$

and for the parallel circuit the admittance is rewritten as

$$Yv = C \left(s + \frac{1}{RC} + \frac{1}{sLC} \right) v = i.$$

When interchanging quantity pairs $Z, Y ; C, L ; R, \frac{1}{R}$ and v, i in the equation for the parallel circuit, the equation of the series circuit is obtained [109, p. 22]. It is slightly confusing that the sources only change from one quantity to another, but the magnitude of the changed source remains the same. When a voltage source is changed to its current source dual, the magnitude is preserved as $v = 10 \rightarrow i = 10$. This means that the voltage measured in the dual circuit is not the same voltage after transformation, only the quantity representation of the source is changed from voltage to current.

For mechanical systems the corresponding Laplace transformed equations are

$$\left(sm + \beta + \frac{1}{s}k \right) c = F, \quad (8.6)$$

for impedance representation and

$$\left(s\frac{1}{k} + \frac{1}{\beta} + \frac{1}{sm} \right) F = c, \quad (8.7)$$

for admittance representation.

Figure 8.8 depicts a mechanical model that represents the response of the speaker cone to external forces. Mass m represents the mass of the speaker cone and the spring constant k is related to the compliance of the cone suspension mechanism. The damping coefficient β refers to all factors that resist the motion of the speaker cone.

The ultimate question is, should this system be modelled as a parallel network or a series network? In the model described by Figure 8.8, the mechanical components of mass, spring and damper seem to be parallel, but in related literature

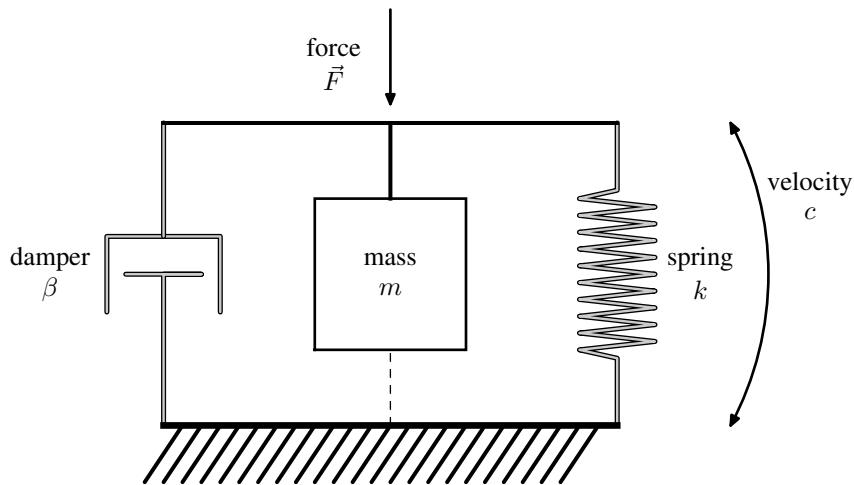


Figure 8.8: Parallel-connected mass, spring and damper

the system is usually taken to be in series, because velocity is equal for all components.

Figure 8.9 shows the dualism in mechanical circuits using electrical components with the logic that the term with s (derivator) connects to the inductor and $\frac{1}{s}$ (integrator) connects to the capacitor. The series system is modelled with equation (8.6) and the parallel circuit is described with equation (8.7).

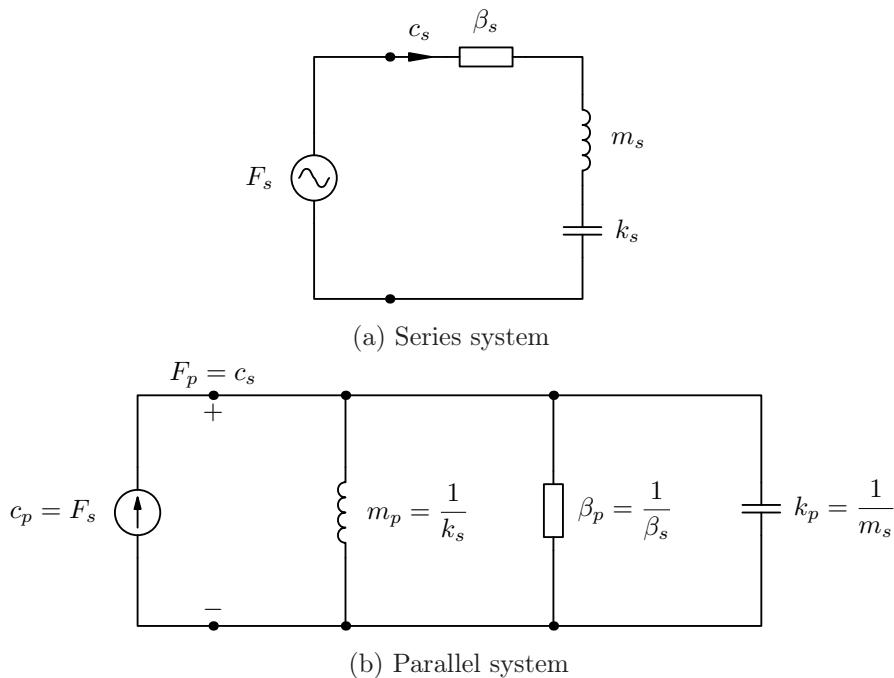


Figure 8.9: Mechanical networks representing the duality of the impedance and admittance models

The aim is now to model the mechanical circuit as an electrical equivalent. For this purpose it is enough to choose either equation (8.6) or (8.7) to represent the speaker cone and relate that equation to equation (8.4) or (8.5). To acknowledge common procedures, equation (8.6) is chosen to represent the mechanical circuit. To build a bridge between force and current as suggested by equation (8.2), it is natural to choose the parallel electronic model to connect with the mechanical model. In both equations the source term is on the right side, which are now related together, implying equality between current and force.

Transformation factors between mechanical and electrical units can be derived by dividing appropriate impedance and admittance equations together so that the force-current relationship is recovered. For example, a division of equation (8.7) by equation (8.4) gives the direct relation

$$\frac{F}{i} \left(sC_p + \frac{1}{R_p} + \frac{1}{sL_p} \right) = \frac{c}{v} \left(sm + \beta + \frac{1}{s}k \right)$$

$$B_0^2 l^2 \left(sC_p + \frac{1}{R_p} + \frac{1}{sL_p} \right) = \left(sm + \beta + \frac{1}{s}k \right)$$

between the mechanical and electrical quantities. Equations (8.2) and (8.3) have been used for digging out the impedance conversion parameter $B_0^2 l^2$. To use it, just choose the multipliers related to similar powers of s and use the factor $B_0^2 l^2$ to convert between electrical admittance and mechanical impedance. If the current term is multiplied to the other side of the equation, it is clear that the current is directly proportional to the force.

Another useful conversion formula is obtained by multiplying together equations (8.2) and (8.3), leading to

$$i = \frac{F}{c} \frac{v}{B_0^2 l^2}, \quad (8.8)$$

which connects the current, force, velocity and voltage together via $B_0^2 l^2$. These formulae should be enough to describe the mechanical circuit of Figure 8.8 as an electrical circuit with resistances, inductances and capacitances. The electrical equivalent circuit is shown in Figure 8.10, where electrical admittances are equivalent to mechanical impedances.

Because the current is a direct replica of the force, it is possible to designate the physical roles of currents i_{p1} and i_{p3} in Figure 8.10. Since force

$$F_1 \propto k \frac{1}{s} c = k \int c \, dt = kx,$$

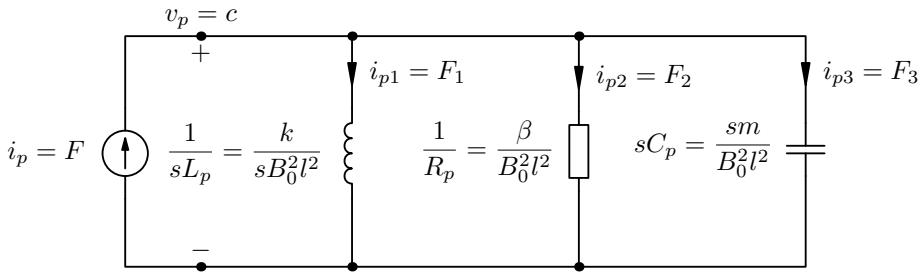


Figure 8.10: An electrical model for simulating the mechanics of the speaker cone

current i_{p1} flowing through the inductor is directly related to the displacement of the speaker cone. Likewise,

$$F_3 \propto msc = k \frac{d}{dt} c = ma,$$

which makes current i_{p3} represent the acceleration of the speaker cone. The acceleration of the cone generates the force which is eventually related to the acoustic pressure created by the cone movement.

The matter of confusion in this model is related to the mechanical impedances, because now there is a conflict compared to the parallel presentation of duality in Figure 8.9b. The impedances and admittances have changed places because force and velocity are interchanged as the signal source. This confusion is forced into the model by the transformation factor $B_0 l$ in a natural way, so it is better just to accept it and continue the analysis.

8.4 AN ELECTRICAL MODEL OF THE DRIVER

Using the concepts of ideal electromechanical and mechanoacoustical transformers, it is possible to create a model of the complete loudspeaker element in the form of an electric circuit. Figure 8.11 reveals the whole electromechanoacoustical model of the moving-coil loudspeaker. All components representing any reactance have been labelled using the Laplace impedance factor s to reflect the true impedance value of the component. The voltage source V represents the output of a guitar amplifier while R_L and L are a simplified model for the voice coil. The transformer with the factor $B_0 l$ is an electromechanical transformer, which couples the electrical circuit to the mechanical part. The current from the electrical side is directly coupled to force F on the mechanical side due to the familiar transformer rule for current

$$i_1 = \frac{1}{B_0 l} i_2 \quad \Rightarrow \quad i_2 = F = i_1 B_0 l.$$

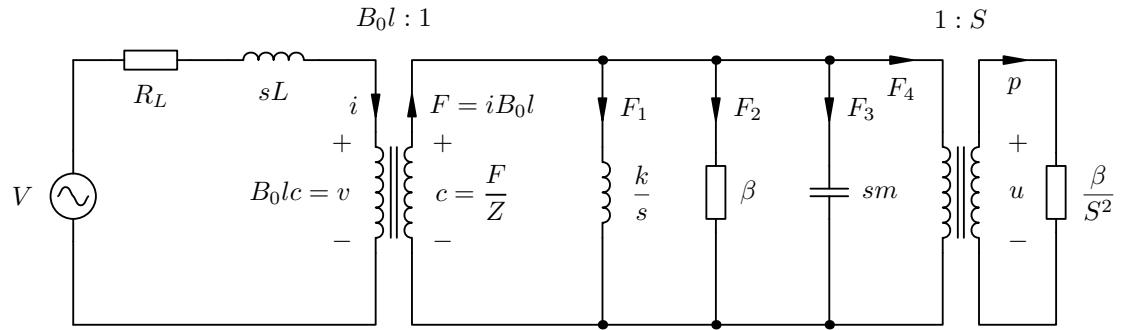


Figure 8.11: An electromechanical equivalent circuit of a moving-coil loudspeaker

As the force 'flows' through the impedances, it builds up as the velocity potential c . The total impedance Z of the mechanical part is formed by the parallel connected impedance of mass, mechanical resistance and spring constant. The explicit expression for the parallel impedance is

$$Z = \beta \left(\frac{\frac{s}{\beta}}{s^2 + \frac{k}{\beta}s + \frac{k}{m}} \right),$$

from where the resonance frequency ω_0 and the quality factor Q are identified as $\frac{k}{m} = \omega_0^2$ and $\frac{k}{\beta} = \frac{\omega_0}{Q}$. The velocity on the secondary side of the transformer is coupled to the primary side due to the transformer rule for voltage,

$$v_1 = \frac{B_0 l}{1} v_2 \quad \Rightarrow \quad v_1 = B_0 l c.$$

The voltage on the primary side is directly proportional to velocity, which justifies the use of the voltage transformer rule.

The electromotive force created to the voice coil due to the movement of the speaker cone has a significant impact on the characteristics of the driver. Since the e.m.f. is created in the opposite direction with the voltage received from the amplifier, the additional e.m.f. is actually reducing and distorting the voltage signal obtained from the amplifier.

The second transformer in Figure 8.11 with the factor $1 : S$ is an mechanoacoustical transformer that couples force F to pressure p according to equation

$$F_1 = \frac{S}{1} F_2 \quad \Rightarrow \quad F_2 = p = \frac{F_1}{S}.$$

The factor S is the area of the loudspeaker cone. The use of current transformer formula is justified, because force is flowing as a current on the primary side of

the transformer. Since force is related to current, pressure on the secondary side should also be considered as a current related quantity.

The acoustical impedance relation is similar to the mechanical impedance and the elementary quantities involved are pressure p , acoustical impedance Z_a and volume velocity, which is defined as

$$u = cS = \frac{FS}{Z_a} = \frac{pS^2}{Z_a}.$$

As a first approximation it is enough to simulate the radiation impedance experienced by the speaker cone as a pure resistive component.

Figure 8.12 indicates the equivalent circuit of the driver when the electrical model of the speaker cone and the electrical representation of acoustical impedance are connected to the voice coil model. The voltage excitation V_{in} comes directly to the circuit from the output terminals of a guitar amplifier and the current i flows through the whole circuit from input to 'ground'. The component val-

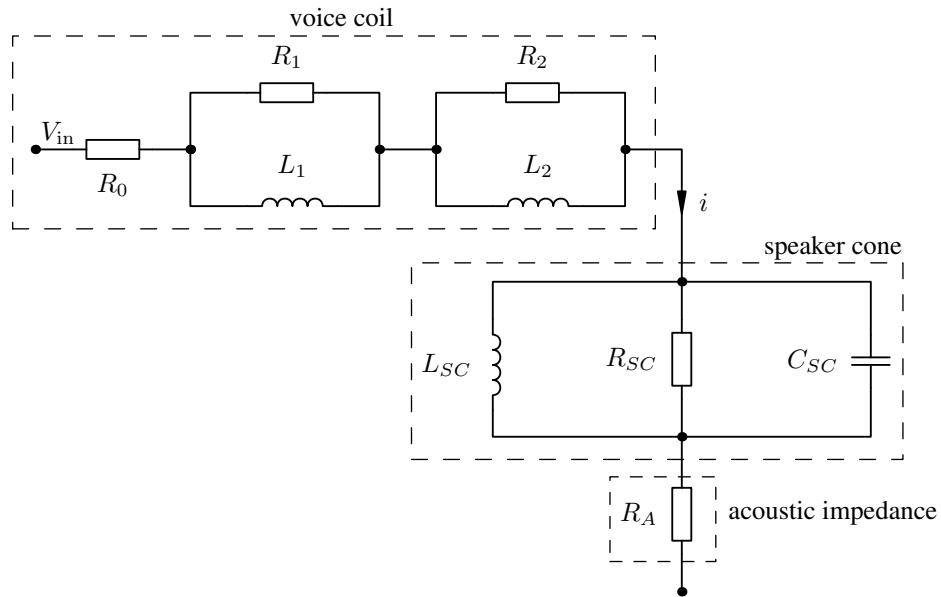


Figure 8.12: An electrical model combining the voice coil, speaker cone and acoustic impedance

ues for the speaker cone model are calculated through the electromechanical impedance equations shown in Figure 8.10.

To indicate the relatively large variance in the driver impedance at different frequencies, the impedance curve in Figure 8.13 is drawn from the circuit of Figure

8.12 using the realistic values

$$R_0 = 8 \Omega$$

$$R_1 = 2 \Omega$$

$$R_2 = 7.1 \Omega$$

$$L_1 = 0.18 \text{ mH}$$

$$L_2 = 0.10 \text{ mH}$$

$$L_{SC} = 25 \text{ mH}$$

$$C_{SC} = 240 \mu\text{F}$$

$$R_{SC} = 18 \Omega.$$

The air resistance was neglected to indicate how the base level of the impedance

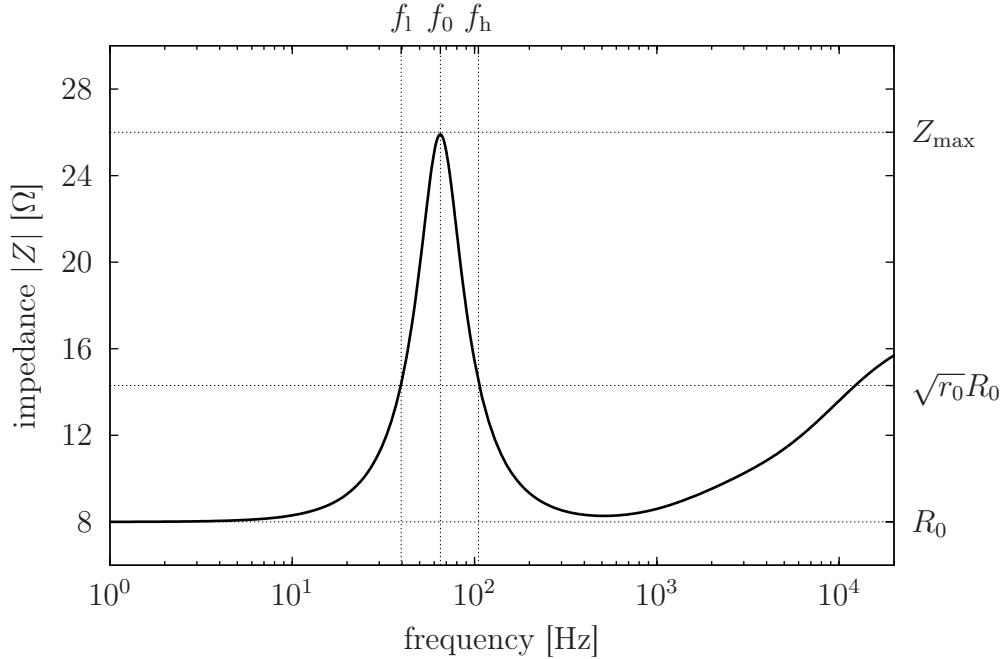


Figure 8.13: A typical impedance curve for a moving-coil loudspeaker

is created by the nominal voice coil resistance R_0 . The complete equation to calculate the total impedance is

$$Z = R_0 + \frac{j\omega L_1}{1 + \frac{j\omega L_1}{R_1}} + \frac{j\omega L_2}{1 + \frac{j\omega L_2}{R_1}} + R_{SC} \left(\frac{\frac{1}{j\omega R_{SC} C_{SC}}}{(j\omega)^2 + j\omega \frac{1}{R_{SC} C_{SC}} + \frac{1}{L_{SC} C_{SC}}} \right),$$

where the last term describes the impedance of the speaker cone. Using the standard substitutions of the angular resonance frequency ω_0 and the quality factor Q , the speaker cone impedance

$$Z_{SC} = R_{SC} \left(\frac{\frac{j\omega \frac{\omega_0}{Q}}{(j\omega)^2 + j\omega \frac{\omega_0}{Q} + \omega_0^2}}{(j\omega)^2 + j\omega \frac{\omega_0}{Q} + \omega_0^2} \right). \quad (8.9)$$

The quality factor Q and angular resonance frequency ω_0 in the impedance expression have a huge practical significance when measuring the electrical and mechanical driver properties [107]. A simple measurement procedure to determine ω_0 and Q for any moving-coil driver is described in more detail in section

8.5. The mechanical resonance peak also gives an indication of the driver's transfer function properties at low frequencies.

Typically the frequency dependent impedance curve of the driver starts from the nominal voice coil resistance R_0 as shown in Figure 8.13. The resonance peak in the lower frequencies comes from the speaker cone parameters, and the gradually ascending tail impedance is due to the varying voice coil impedance. Parameters Z_{\max} , $\sqrt{r_0}R_0$, f_0 , f_1 and f_h frame the bandwidth inside the resonance area. These parameters are important when experimentally determining the essential (Thiele-Small) driver parameters, which are referenced in section 8.5.

Since the impedance of the driver is frequency dependent, it is clear that all frequency components are not reproduced identically by the speaker. The nonlinear frequency response of the driver can be partially compensated at the amplifier side using simple filter circuits like the Zobel network. The use of the Zobel network is illustrated in Figure 7.27 and it is used for compensating the nonlinearities of the voice coil inductance by attenuating high frequencies. The resonance peak of the speaker cone is compensated by keeping the output impedance of the amplifier as low as possible, since then the relative impedance variations between the amplifier output and driver input are reduced to some extent. The ratio of the input impedance of the driver and the output impedance of the amplifier defines a parameter called the *damping factor*.

The idea of the damping factor can be examined with an oversimplified example. The output stage of a typical audio amplifier is connected in a doubled emitter follower configuration where the driver is connected in parallel with the output impedance of the amplifier. Figure 8.14 presents a test circuit where the driver is connected in parallel with the emitter resistor R_E of a simple emitter follower circuit. The low-frequency response of the amplifier section is determined by the input capacitor C_1 . The frequency response of the amplifier with the driver input connected to the node V_{out} of the amplifier is evaluated with different values of resistance R_E .

The results of the frequency response simulation are shown in Figure 8.15. When R_E is large, the driver impedance in parallel with R_E is dominant, because it is seen as such a small impedance. The effects of the nonlinear impedance of the driver are clearly visible in the frequency response curve. With a very small value of R_E , the constant emitter resistance is the dominant impedance and the effects of the nonlinear driver impedance are not seen in the frequency response

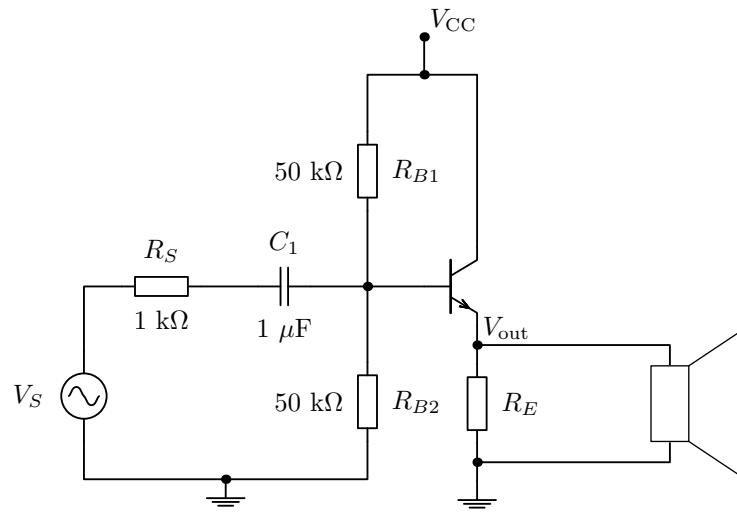


Figure 8.14: A test circuit to examine the effect of the damping factor

curve. The side effect is that the magnitude of the output voltage V_{out} measured at the emitter is attenuated due to the damping factor. By making the amplifier

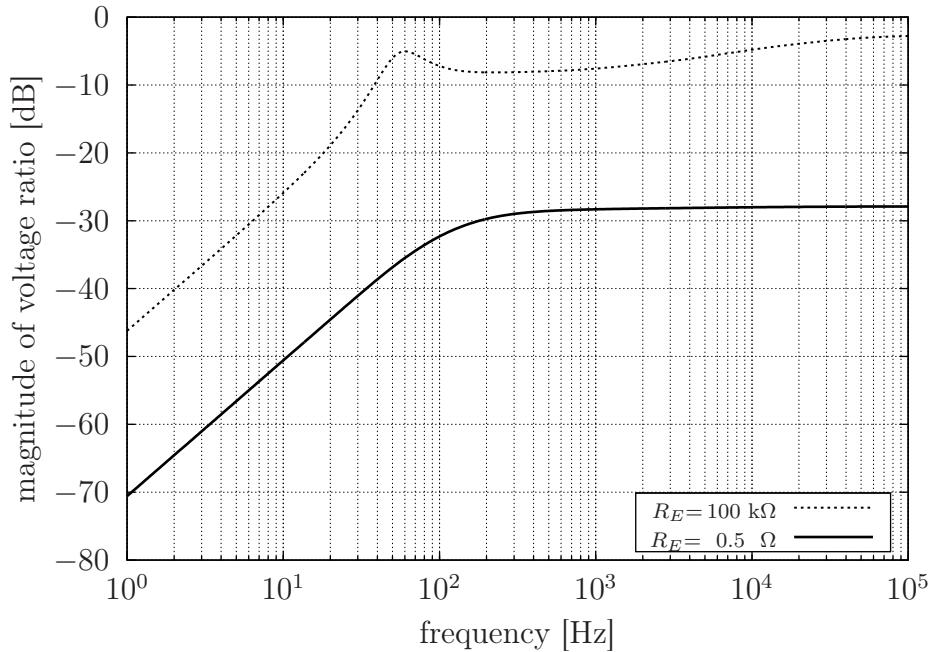


Figure 8.15: An example of using the damping factor to linearise the frequency response of an amplifier with a loudspeaker load

impedance as low as possible, the amplifier is adequate to drive almost all kinds of loudspeakers with a relatively linear frequency response.

The loudspeaker modelling in this section has only taken account of a loudspeaker element that is not enclosed in a speaker cabinet. When the loudspeaker

is placed inside a box, the acoustical properties of the combined system are changed. This usually affects the impedance curve by moving and reshaping the area of the resonance peak. Other resonance effects might be added to the combined frequency response due to the enclosure of the driver. If the driver is mounted onto a large flat plate (an infinite baffle), the acoustic properties can be approximately described by a second order high-pass filter function as derived by Small [107]. This mounting mechanism can be used to reduce the resonance peak in the transfer characteristics of a loudspeaker element.

8.5 MEASURING THE ESSENTIAL DRIVER PARAMETERS

Based on the impedance expression (8.9) of a moving-coil driver, most of the mechanical speaker cone properties can be determined if the resonance frequency ω_0 and the quality factor Q are known. The presented measurement procedure, as described by Small in reference [107], aims to determine these quantities so that the obtained results could be used directly in equation (8.9). In addition to measuring the properties of the speaker cone, the measurements try to estimate the voice coil inductance.

A simple approach to measuring the frequency response of a driver is to use a signal generator to provide sine waves of known frequencies and measure the voltage at the terminals of the voice coil. The driver is intended to be measured in a free-standing state, so that it should be removed from the speaker cabinet and the baffle board. The complete measurement setup is shown in Figure 8.16, where the voltage v_g from the signal generator is fed through a $1\text{ k}\Omega$ resistor R_S to the driver input terminals. An oscilloscope or a digital volt meter is used for measuring the voltage v_{R0} over the voice coil. In this setup the voltage at the voice coil terminals is directly proportional to the absolute value of the driver impedance, because the current is kept almost constant with the series resistor R_S . The current i_s flowing through each component in the circuit can be determined by measuring voltage v_{RS} over resistor R_S and then using Ohm's law to calculate the current.

When measuring average size drivers, the resonance frequency is found when the voltage v_{RS} reaches a minimum value somewhere in the frequency range of 10 – 100 Hz. With this in mind, a detailed measurement guide is presented:

1. measure accurate values for resistor R_S and voice coil resistance R_0 using a multimeter

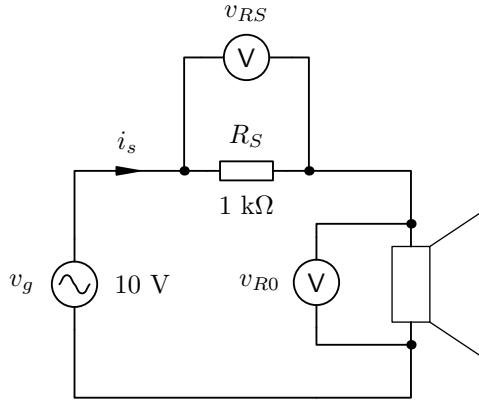


Figure 8.16: A measurement setup to determine the driver frequency response

2. by adjusting the frequency of the input voltage v_g , find the frequency where voltage v_{RS} has a minimum value. Write down the resonance frequency as f_r and the voltages v_{RS} and v_{R0}
3. using the values obtained at step 2, calculate current $i_s = \frac{v_{RS}}{R_S}$ at the resonance frequency
4. using the value of i_s obtained from step 3, calculate the loudspeaker impedance at the resonance frequency as $|Z_{\max}| = \frac{v_{R0}}{i_s}$
5. calculate $r_0 = \frac{|Z_{\max}|}{R_0}$
6. calculate $\sqrt{r_0}R_0$ to obtain the -3 dB impedance value
7. by adjusting the frequency of the input voltage, identify the -3 dB frequencies f_l and f_h on both sides of the resonance frequency. Note that for best accuracy, the measured impedance needs to be evaluated each time as in steps 3 and 4. Therefore, finding the exact frequencies where the measured impedance equals $\sqrt{r_0}R_0$ might take some time.

After the resonance frequency and the -3 dB frequencies are found, it is possible to calculate the quality factor Q from the measured frequencies. The evaluation based on the formulae given by Small [107] is surprisingly easy. First evaluate the mechanical quality factor Q_{MS} using equation

$$Q_{MS} = \frac{f_s \sqrt{r_0}}{f_h - f_l}. \quad (8.10)$$

Then obtain the electrical quality factor from equation

$$Q_{ES} = \frac{Q_{MS}}{r_0 - 1}. \quad (8.11)$$

With the help of these two quality factor values it is possible to determine the electrical inductance L_{CES} , capacitance C_{MES} and resistance R_{ES} in the speaker cone equivalent circuit, which are shown as L_{SC} , C_{SC} and R_{SC} in Figure 8.12.

The mechanical and electrical quality factors Q_{MS} and Q_{ES} are also related to the electrical equivalence components as

$$Q_{\text{MS}} = \omega_0 C_{\text{MES}} R_{\text{ES}} \quad \text{and} \quad Q_{\text{ES}} = \omega_0 C_{\text{MES}} R_0.$$

From here, one can solve

$$R_{\text{ES}} = \frac{Q_{\text{MS}}}{Q_{\text{ES}}} R_0, \quad (8.12)$$

and using this resistance value, one can use the relation $\frac{\omega_0}{Q} = \frac{1}{RC}$ from the general parallel resonance circuit to calculate the capacitance

$$C_{\text{MES}} = \frac{Q_{\text{MS}}}{\omega_0 R_{\text{ES}}}. \quad (8.13)$$

Note that the mechanical quality factor is used here, because reference [107] has defined the speaker cone impedance equation (8.9) using Q_{MS} as the general quality factor Q . Finally, the value of the inductance L_{CES} can be solved from the resonance frequency equation as

$$L_{\text{CES}} = \frac{1}{\omega_0^2 C_{\text{MES}}}. \quad (8.14)$$

With these equations, all values of the electrical equivalence components in the speaker cone resonance circuit can be determined.

To show that the equations really do work accurately, let's pretend that the impedance curve in Figure 8.13 is obtained by measurements from a real loudspeaker element. The impedance values measured from the driver are

$$|Z_{\text{max}}| = 25.9 \Omega \quad \text{and} \quad R_0 = 8.0 \Omega,$$

so that based on these values, $r_0 = 3.2375$ and $\sqrt{r_0} R_0 = 14.4 \Omega$. The latter value determines the -3 dB frequencies, which are

$$f_1 = 39.6 \text{ Hz} \quad \text{and} \quad f_h = 105 \text{ Hz},$$

as shown in Figure 8.13. The resonance frequency at the maximum impedance is $f_r = 65.0 \text{ Hz}$.

To find the mechanical quality factor, a substitution of numerical values to equation (8.10) yields

$$Q_{\text{MS}} = \frac{65 \text{ Hz} \cdot \sqrt{3.2375}}{105 \text{ Hz} - 39.6 \text{ Hz}} = 1.7883,$$

and similarly the electrical quality factor

$$Q_{\text{ES}} = \frac{1.7883}{3.2375 - 1} = 0.79924.$$

Now the resistance R_{ES} can be calculated as

$$R_{\text{ES}} = \frac{1.7883}{0.79924} \cdot 8 \Omega = 17.9 \Omega,$$

and finally the values for the capacitance and inductance become

$$C_{\text{MES}} = \frac{1.7883}{2 \cdot \pi \cdot 65 \text{ Hz} \cdot 17.9 \Omega} = 244.6 \mu\text{F},$$

and

$$L_{\text{CES}} = \frac{1}{(2 \cdot \pi \cdot 65 \text{ Hz})^2 \cdot 244.6 \mu\text{F}} = 24.5 \text{ mH}.$$

The obtained values are very close to the actual values from where the curve in Figure 8.13 was drawn.

The mass m , spring constant k and mechanical resistance β of the speaker cone can be calculated from the electrical values if the force factor $B_0 l$ is known. Unfortunately the value for the force factor is usually not known, so more measurements are needed. Small [107] introduces a set of parameters:

- C_{MS} = mechanical compliance of the loudspeaker element
- M_{MS} = mechanical mass of the loudspeaker element
- R_{MS} = mechanical resistance of the loudspeaker element
- C_{AS} = acoustical compliance of the loudspeaker element ($= C_{\text{MS}} S^2$)
- M_{AS} = acoustical mass of the loudspeaker element ($= \frac{M_{\text{MS}}}{S^2}$)
- R_{AS} = acoustical resistance of the loudspeaker element ($= \frac{R_{\text{MS}}}{S^2}$)
- V_{AS} = equivalent volume of air of the loudspeaker element

A link between the electrical and acoustical parameters of a driver is created by expressions of V_{AS} , namely by equations

$$V_{\text{AS}} = \rho_0 c^2 C_{\text{AS}} = \rho_0 c^2 C_{\text{MS}} S^2 \quad (8.15)$$

and

$$V_{\text{AS}} = V_T \left[\frac{f_{\text{CT}} Q_{\text{ECT}}}{f_r Q_{\text{ES}}} - 1 \right]. \quad (8.16)$$

In equation (8.15) ρ_0 is the density of air and c is the speed of sound in air. From these two equations the latter one (8.16) is intended to be used along with the measurement results to determine an experimental value for V_{AS} . The parameter V_T refers to the net internal volume of a speaker cabinet, meaning that another identical set of measurements for the loudspeaker element needs to be done where the driver is mounted in a box which volume is known. When the driver is placed inside a speaker cabinet, the acoustical frequency response of the cabinet changes the resonance properties of the stand-alone driver. The parameters f_{CT} and Q_{ECT} in equation (8.16) refer to the resonance frequency and the electrical quality factor measured from the combined system of the driver and the cabinet.

When the value of V_{AS} has been determined experimentally, equation (8.15) can be used to calculate C_{MS} . Then equation

$$\omega_0^2 = \frac{1}{C_{MS}M_{MS}}, \quad (8.17)$$

where ω_0 is the resonance angular frequency of the free-standing driver, leads to the value of M_{MS} , and finally the equation

$$Q_{ES} = \omega_0 C_{MES} R_E = \omega_0 R_E \frac{M_{MS}}{B_0^2 l^2} \quad (8.18)$$

reveals the value of the force factor $B_0 l$.

Another way to evaluate the mass and compliance of the speaker cone is to increase the mass of the cone in a controlled manner by placing additional mass (10 – 20 grams of blue tack or equivalent) on the speaker cone. Then repeat the round of measurements to locate the resonance frequency and –3 dB frequencies. The mechanical properties of the cone can be evaluated from the changes observed in the resonance frequency and the corresponding quality factors due to the added mass.

The inductance of the speaker coil can be estimated directly from the measurement results. Using the same procedure as shown in Figure 8.16, the impedance is evaluated in the high-frequency range (> 500 Hz) all the way up to 10 kHz or even more. Then the impedance curve can be estimated by a chain of parallel connected resistor and inductor combinations. Alternatively a single value for the inductance can be roughly estimated by fitting a straight line on top of the high-frequency impedance curve and using the basic formula $X_L = \omega L$ of reactance to find L .

CHAPTER 9

TUNING DEVICES

Before going any further into this topic, it must be emphasised that the best way to tune any guitar is by ear. To achieve the best results, simply pluck an open string on the guitar against a reference sound and minimise the beats in the heard sound. Good sound references are the classic tuning fork, or even the 50/60 Hz power line hum heard with some cheap amp/loudspeaker cabinets. When one string is successfully tuned to some external reference, all the other strings can be tuned against the string that is in tune.

The absolute pitch of instruments is never that meaningful, as the essential thing is that all the instruments are in tune with each other when playing in a group. That is another reason why the quickest, easiest and most elegant way to tune is by ear. The ear is also the final judge of the sensation of in-tunedness, which is eventually only a subjective matter.

The other approach is to tune a guitar like an engineer, a technology addict who communicates more fluently with machines than other people. To design a device that measures the pitch of a played note on the guitar needs all the information that is covered in the previous chapters of this book. To be consistent with all the other chapters, the goal is to avoid any digital signal processing. The only allowed integrated circuit is the operational amplifier.

This chapter introduces a simplified redesign from an analogue guitar tuner circuit which was patented in the middle of the 1970's. The aim is mostly to learn useful things from the field of analogue signal processing and present a simple approach, which does not guarantee super accurate tuning results in every condition. If one is going to build and use this tuner in practise, it is suggested to either improve the design or simply buy a cheap (digital) tuner from a music

store. At least remember that you have been warned.

To fully understand the inner beauty of this basic analogue guitar tuner device, as a prerequisite it is a must to know how to analyse a bit more complicated op-amp based filter circuits. In analogue guitar tuners it is usually a necessity to first filter out most of the upper partials of the string vibration signal to ease out the frequency detection of the played note.

After a general analysis method for the filters is covered, a useful multivibrator circuit block is analysed, because it is also later used in the actual tuning device.

9.1 STATE VARIABLE BIQUAD FILTERS

The term *biquad* comes from a transfer function of the general type

$$H(s) = \frac{a_2 s^2 + a_1 s + a_0}{s^2 + b_1 s + b_0}, \quad (9.1)$$

which is called a *biquadratic function* since the numerator and the denominator are both quadratic functions of the Laplacian variable s . This transfer function can be used for modelling second order low-pass, high-pass, band-pass and notch filters. The functional forms of these filters are given in Table 9.1, where K is a constant gain parameter and ω_0 defines the centre frequency of the band-pass filter or the cut-off frequency of the low/high-pass filters. The coefficient Q is a quality factor, which indicates how sharply the filter defines the transition between the passband and the stopband. In ideal filters the value of Q is infinitely large, so the bigger the better is the way to go with this parameter.

Table 9.1: Basic biquad transfer functions

<i>low-pass</i>	<i>high-pass</i>	<i>band-pass</i>	<i>notch</i>
$\frac{K}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2}$	$\frac{Ks^2}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2}$	$\frac{Ks}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2}$	$\frac{K(s^2 + \omega_r^2)}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2}$

Generally among filter circuits the width of the passband is measured as a difference of the frequencies at the -3 dB points. In band-pass or notch filters the term $\frac{\omega_0}{Q}$ is directly related to the width of the passband. Therefore, the basic properties of biquad filters are quite easily determined by examining the coefficients in the transfer function (9.1).

One basic building block of the tuning circuit to be presented is a filter section consisting of three operational amplifiers. This circuit can be used to realise the transfer functions of low-pass, band-pass and high-pass biquad filters. It would be possible to build a biquad filter with a much smaller amount of components, but then the tunability of the filter would not be as good as in the more general state variable filters. If the tuner is designed to tune all six strings of a guitar, then it needs to be easily tunable for all corresponding frequencies.

The amplifier stages of filters which consist of several operational amplifiers are separately connected as simple integrators and inverters. The fundamental idea can be derived directly from equation (9.1). For example, the low-pass transfer function

$$H(s) = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{-a_0}{s^2 + b_1 s + b_0}$$

can be rearranged as

$$\begin{aligned} (s^2 + b_1 s + b_0) V_{\text{out}} &= -a_0 V_{\text{in}} \\ \left(s + b_1 + \frac{b_0}{s} \right) V_{\text{out}} &= -\frac{a_0}{s} V_{\text{in}} \\ \left(1 + \frac{b_0}{s(s + b_1)} \right) V_{\text{out}} &= -\frac{a_0}{s(s + b_1)} V_{\text{in}} \\ V_{\text{out}} &= -\frac{b_0}{s(s + b_1)} V_{\text{out}} - \frac{a_0}{s(s + b_1)} V_{\text{in}}. \end{aligned}$$

Furthermore, this can be forced to be written as

$$V_{\text{out}} = (-1) \left(-\frac{K_1}{s} \right) \left[-\frac{K_2}{s + b_1} V_{\text{out}} - \frac{a_0}{K_1} \frac{1}{s + b_1} V_{\text{in}} \right], \quad (9.2)$$

where a substitution of $b_0 = K_1 K_2$ has been made. [110, pp. 243 – 248]

The basic op-amp blocks can be identified by inspection from equation (9.2). There is the basic inverter with the factor -1 , the block $-\frac{K_1}{s}$ is a basic integrator, and the remaining two terms can also be realised as integrators. The block diagram of this rearrangement is shown in Figure 9.1.

A circuit that realises the signal flow presented in block diagram 9.1 can now be constructed. Figure 9.2 presents a typical op-amp filter that is built from three basic op-amp blocks, which each separately have very basic functionalities as integrators or inverting gain stages, but together they form a robust filter with an easily adjustable centre frequency ω_0 and quality factor Q . This filter implementation is also known as the Tow-Thomas biquad filter topology.

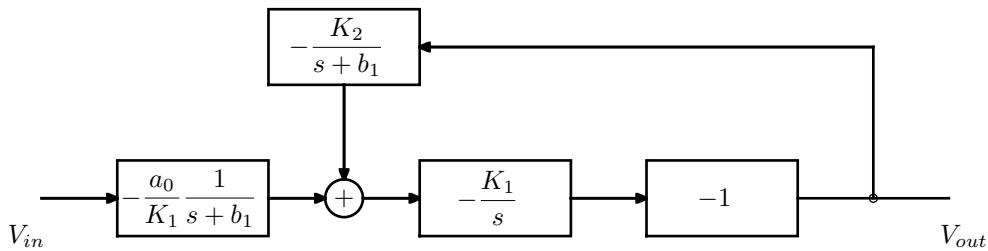


Figure 9.1: A block diagram of the rearrangement of a biquad function

With reference to Figure 9.1 and equation (9.2), the first op-amp realises the integrator terms of type $-\frac{K}{s + b_1}$, the second op-amp realises the integrator term of type $-\frac{K}{s}$ and the third op-amp is just a basic inverter giving the term -1 . The required summation is done at the input of the first operational amplifier using resistors R_4 and R_3 .

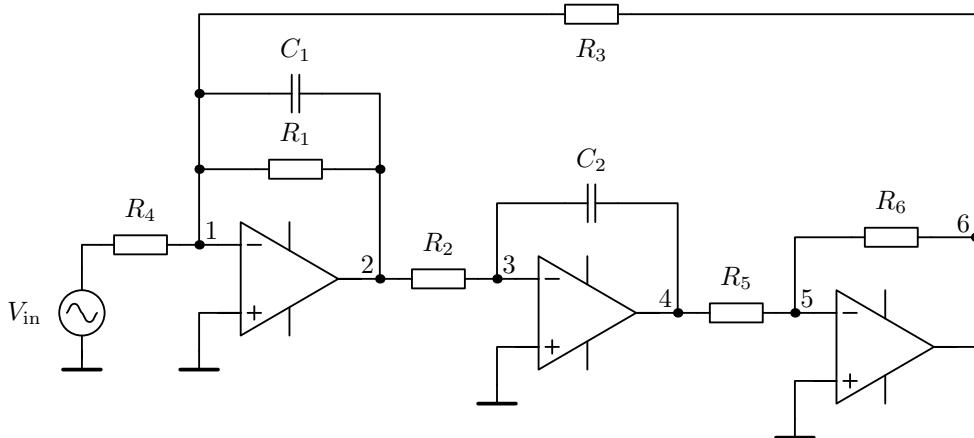


Figure 9.2: The Tow-Thomas biquad filter topology

If an efficient band-pass filter is going to be used at the input of a guitar tuning device, the centre frequency of the filter should be easily adjustable to the specific string to be tuned. In biquad filters that are using the three op-amp construction, it is usually possible to adjust the centre frequency by changing the value of one resistor without affecting much the quality factor of the filter.

In addition, the transfer functions of multi op-amp filters are often surprisingly easy to evaluate directly from the schematics. For example, the circuit of Figure 9.2 can be modelled with the matrix equation

$$\begin{bmatrix} Y_{11} & -Y_{12} & 0 & 0 & 0 & -Y_{16} \\ -Y_{21} & Y_{22} & -Y_{23} & 0 & 0 & 0 \\ 0 & -Y_{32} & Y_{33} & -Y_{34} & 0 & 0 \\ 0 & 0 & -Y_{43} & Y_{44} & -Y_{45} & 0 \\ 0 & 0 & 0 & -Y_{54} & Y_{55} & -Y_{56} \\ -Y_{61} & 0 & 0 & 0 & -Y_{65} & Y_{66} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \end{bmatrix} = \begin{bmatrix} \frac{V_{\text{in}}}{R_4} \\ I_1 \\ 0 \\ I_2 \\ 0 \\ I_3 \end{bmatrix},$$

where the current terms I_1 , I_2 and I_3 refer to the output currents of the operational amplifiers. The nonzero admittance elements of the matrix are marked with subscripts, which refer to their location in the admittance matrix in terms of rows and columns. The listing (9.3) contains the actual admittance terms that should be substituted into the matrix above.

$$\begin{aligned} Y_{11} &= \frac{1}{R_4} + \frac{1}{R_1} + \frac{1}{R_3} + sC_1 & Y_{12} = Y_{21} &= \frac{1}{R_1} + sC_1 \\ Y_{22} &= \frac{1}{R_1} + \frac{1}{R_2} + sC_1 & Y_{23} = Y_{32} &= \frac{1}{R_2} \\ Y_{33} &= \frac{1}{R_2} + sC_2 & Y_{34} = Y_{43} &= sC_2 \\ Y_{44} &= \frac{1}{R_5} + sC_2 & Y_{45} = Y_{54} &= \frac{1}{R_5} \\ Y_{55} &= \frac{1}{R_5} + \frac{1}{R_6} & Y_{56} = Y_{65} &= \frac{1}{R_6} \\ Y_{66} &= \frac{1}{R_6} + \frac{1}{R_3} & Y_{61} = Y_{16} &= \frac{1}{R_3} \end{aligned} \tag{9.3}$$

The reduction rules of the special op-amp nodal analysis can (and must) be applied to the matrix equation to simplify the calculations and to get correct results. The basic rules of the op-amp nodal analysis method are given in section 6.9.3. In the case of ideal operational amplifiers, the node voltages V_1 and V_3 and V_5 are at ground potential, and therefore the respective columns 1, 3 and 5 can be removed from the admittance matrix. To reshape the matrix to a square, all the rows that are aligned with the currents I_1 , I_2 and I_3 are removed. With these reductions, the 6x6 admittance matrix is now a 3x3 matrix:

$$\begin{bmatrix} -\frac{1}{R_1} - sC_1 & 0 & -\frac{1}{R_3} \\ -\frac{1}{R_2} & -sC_2 & 0 \\ 0 & -\frac{1}{R_5} & -\frac{1}{R_6} \end{bmatrix} \times \begin{bmatrix} V_2 \\ V_4 \\ V_6 \end{bmatrix} = \begin{bmatrix} \frac{V_{\text{in}}}{R_4} \\ 0 \\ 0 \end{bmatrix}.$$

From this matrix it is possible to solve the output voltages of all three op-amps. Depending on which output of V_2 , V_4 or V_6 is chosen, the resulting transfer function will be in either a low-pass or band-pass filter format. The tuning circuit takes the output from node 2, which will give the response of the band-pass filter. In this case, the transfer function

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{V_2}{V_{\text{in}}} = \frac{-\frac{1}{R_4 C_1} s}{s^2 + \frac{1}{R_1 C_1} s + \frac{1}{R_2 R_3 C_1 C_2} \frac{R_6}{R_5}}. \quad (9.4)$$

The other transfer functions for this circuit are two low-pass forms,

$$\frac{V_4}{V_{\text{in}}} = \frac{\frac{1}{R_2 R_4 C_1 C_2} \frac{R_6}{R_5}}{s^2 + \frac{1}{R_1 C_1} s + \frac{1}{R_2 R_3 C_1 C_2} \frac{R_6}{R_5}}$$

and

$$\frac{V_6}{V_{\text{in}}} = \frac{-\frac{1}{R_2 R_4 C_1 C_2} \frac{R_6}{R_5}}{s^2 + \frac{1}{R_1 C_1} s + \frac{1}{R_2 R_3 C_1 C_2} \frac{R_6}{R_5}}.$$

According to the general form of biquad functions shown in Table 9.1, the nominal angular frequency ω_0 for all three filters can be identified as

$$\omega_0 = \sqrt{\frac{1}{R_2 R_3 C_1 C_2} \frac{R_6}{R_5}},$$

and the quality factor can be derived from the multiplier of s in the denominator to yield

$$Q = \sqrt{\frac{C_1}{R_2 R_3 C_2} \frac{R_6}{R_5}} R_1,$$

so that the ratio $\frac{\omega_0}{Q}$ gives the multiplier $\frac{1}{R_1 C_1}$ for s .

As a band-pass filter, this circuit provides sharp band-pass filtering with a stable quality factor for a wide range of values of resistor R_2 . The procedure for determining the suitable component values to tune the filter to a certain frequency is

given in reference [110, p. 250]. By assigning the biquad parameters of equation (9.4) as

$$K = \frac{1}{R_4 C_1} \quad ; \quad \frac{\omega_0}{Q} = \frac{1}{R_1 C_1} \quad ; \quad \omega_0^2 = \frac{1}{R_2 R_3 C_1 C_2},$$

the unscaled component values should be chosen according to

$$C_1 = C_2 = 1 \quad \text{and} \quad R_2 = R_3 = R$$

$$R_1 = \frac{Q}{\omega_0} \quad ; \quad R = \frac{1}{\omega_0} \quad ; \quad R_4 = \frac{1}{K}.$$

After the value relations have been determined as indicated above, the actual impedance scaling to realistic impedance values should be done. The resistors R_5 and R_6 should be chosen to be equal, but their values do not depend on the impedance scaling of the other components.

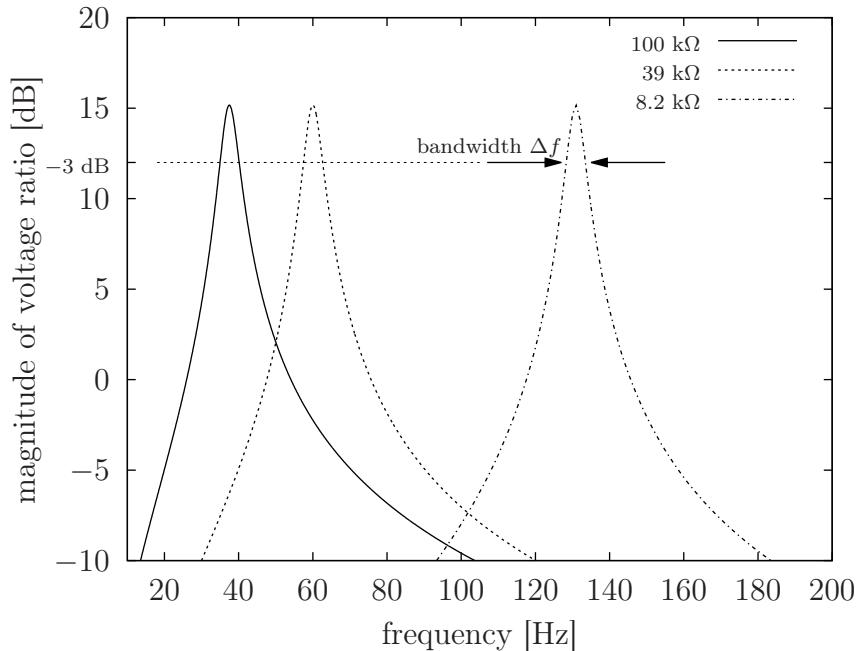


Figure 9.3: The frequency response of the resistance adjustable band-pass filter

An example of the frequency response of the filter with a few selected values of R_2 is shown in Figure 9.3. The circuit was initially tuned to have a centre frequency of 55 Hz, which is the frequency of the open A-string of a bass guitar. The figure shows that the width of the passband does not change when altering the value of R_2 . This means that the centre frequency can be changed while the other parameters of the filter remain fixed.

Figure 9.4 presents a more versatile three op-amp biquad filter. This filter can provide low-pass, band-pass and high-pass transfer functions from nodes 3, 5

and 7. This kind of filter can be referred to as a state variable filter because it

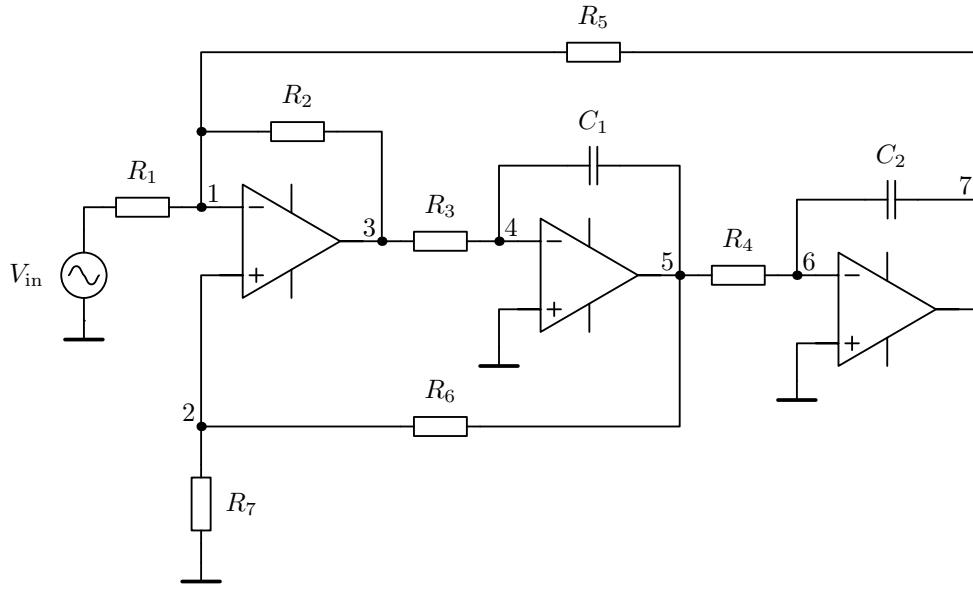


Figure 9.4: An enhanced state variable filter

can be used to construct all the common filter functions without changing the component configuration of the circuit. The filter of Figure 9.2 could not create the high-pass transfer function, and therefore it cannot be taken as a true state variable filter. The circuit in Figure 9.4 can be modelled with the matrix equation

$$\begin{bmatrix} Y_{11} & 0 & -Y_{13} & 0 & 0 & 0 & -Y_{17} \\ 0 & Y_{22} & 0 & 0 & -Y_{25} & 0 & 0 \\ -Y_{13} & 0 & Y_{33} & -Y_{34} & 0 & 0 & 0 \\ 0 & 0 & -Y_{43} & Y_{44} & -Y_{45} & 0 & 0 \\ 0 & -Y_{52} & 0 & -Y_{54} & Y_{55} & -Y_{56} & 0 \\ 0 & 0 & 0 & 0 & -Y_{65} & Y_{66} & -Y_{67} \\ -Y_{71} & 0 & 0 & 0 & 0 & -Y_{76} & Y_{77} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \\ V_7 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ 0 \\ I_1 \\ 0 \\ I_2 \\ 0 \\ I_3 \end{bmatrix}.$$

The nonzero elements $Y_{i,j}$ of the admittance matrix are marked with their corresponding indices indicating the row and column. The listing (9.5) contains the actual terms that should be substituted to the admittance matrix above.

$$\begin{aligned}
Y_{11} &= \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_5} & Y_{13} = Y_{31} &= \frac{1}{R_2} \\
Y_{22} &= \frac{1}{R_6} + \frac{1}{R_7} & Y_{25} = Y_{52} &= \frac{1}{R_6} \\
Y_{33} &= \frac{1}{R_2} + \frac{1}{R_3} & Y_{34} = Y_{43} &= \frac{1}{R_3} \\
Y_{44} &= \frac{1}{R_3} + sC_1 & Y_{45} = Y_{54} &= sC_1 \\
Y_{55} &= \frac{1}{R_4} + sC_1 & Y_{56} = Y_{65} &= \frac{1}{R_4} \\
Y_{66} &= \frac{1}{R_4} + sC_2 & Y_{67} = Y_{76} &= sC_2 \\
Y_{77} &= \frac{1}{R_5} + sC_2 & Y_{71} = Y_{17} &= \frac{1}{R_5}
\end{aligned} \tag{9.5}$$

In the case of ideal operational amplifiers, the node voltages V_1 and V_2 are at the same potential, and therefore column 2 can be added to column 1 and column 2 is removed. Also, because V_4 and V_6 are grounded, the columns multiplying these voltages can be removed. To reshape the matrix into a square, all the rows that are aligned with the currents I_1 , I_2 and I_3 are removed. With these reductions, the 7x7 matrix is now a 4x4 matrix:

$$\left[\begin{array}{cccc} \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_5} & -\frac{1}{R_2} & 0 & -\frac{1}{R_5} \\ \frac{1}{R_6} + \frac{1}{R_7} & 0 & -\frac{1}{R_6} & 0 \\ 0 & -\frac{1}{R_3} & -sC_1 & 0 \\ 0 & 0 & -\frac{1}{R_4} & -sC_2 \end{array} \right] \times \begin{bmatrix} V_1 \\ V_3 \\ V_5 \\ V_7 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

From this matrix equation, one can solve the transfer functions for every op-amp output node using the Cramer's rule. The rule leads to an equation having determinants in the numerator and in the denominator.

After the admittance determinant from the denominator has been written open, the expression in the denominator for all of the transfer functions becomes

$$s^2 \frac{C_1 C_2}{R_2} \left(\frac{1}{R_6} + \frac{1}{R_7} \right) + s \frac{C_2}{R_3 R_6} \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_5} \right) + \frac{1}{R_3 R_4 R_5} \left(\frac{1}{R_6} + \frac{1}{R_7} \right).$$

After solving the numerator determinants and simplifying with the denominator,

the expressions for the transfer functions are

$$\frac{V_7}{V_{\text{in}}} = \frac{-\frac{1}{C_1 C_2 R_3 R_4} \frac{R_2}{R_1}}{s^2 + s \frac{R_7(R_2 R_5 + R_1 R_5 + R_1 R_2)}{C_1 R_1 R_3 R_5 (R_6 + R_7)} + \frac{1}{C_1 C_2 R_3 R_4} \frac{R_2}{R_5}}$$

for low-pass,

$$\frac{V_5}{V_{\text{in}}} = \frac{s \frac{1}{C_1 R_3} \frac{R_2}{R_1}}{s^2 + s \frac{R_7(R_2 R_5 + R_1 R_5 + R_1 R_2)}{C_1 R_1 R_3 R_5 (R_6 + R_7)} + \frac{1}{C_1 C_2 R_3 R_4} \frac{R_2}{R_5}}$$

for band-pass, and

$$\frac{V_3}{V_{\text{in}}} = \frac{-s^2 \frac{R_2}{R_1}}{s^2 + s \frac{R_7(R_2 R_5 + R_1 R_5 + R_1 R_2)}{C_1 R_1 R_3 R_5 (R_6 + R_7)} + \frac{1}{C_1 C_2 R_3 R_4} \frac{R_2}{R_5}}$$

for a high-pass filter.

From the filter functions, the nominal angular frequency ω_0 can be identified as

$$\omega_0 = \sqrt{\frac{1}{C_1 C_2 R_3 R_4} \frac{R_2}{R_5}},$$

and the quality factor can be derived from the multiplier of s in the denominator to yield

$$Q = \sqrt{\frac{C_1 R_3 R_2 R_5}{C_2 R_4}} \frac{R_1 (R_6 + R_7)}{R_7 (R_2 R_5 + R_1 R_5 + R_1 R_2)}.$$

As the transfer function for this state variable filter is more complicated than for the Tow-Thomas biquad, it is not easy to give a straightforward procedure for choosing the component values for certain requirements. The expressions of nominal angular frequency and the quality factor can be used in numerical analysis to iterate the component values suitable for one's needs. One good initial condition is to choose the capacitances to have value 1 and the resistances of the integrator to have the same value, which could also be initially set as 1.

9.2 A MONOSTABLE MULTIVIBRATOR

A multivibrator is a circuit that can be used to generate pulses of different lengths. If the multivibrator generates a continuous sequence of pulses, it is called an astable multivibrator. If the multivibrator has two stable states which can be controlled on and off with external pulses, it is called a bistable multivibrator. A

monostable multivibrator is categorised somewhere in between the astable and bistable multivibrators.

The monostable multivibrator has a stable state and an unstable state. When no external control pulses are provided to the circuit, the circuit rests in its stable state. Normally this stable state means that the output terminal of the multivibrator is at zero volts. When the external trigger pulse is encountered in the input terminals of the circuit, the output of the multivibrator switches to the unstable state, which in terms of digital electronics is usually +5 volts. This unstable state remains for some predetermined time and then the output returns to the stable state. This so-called *single-shot* or *one-shot* sequence can be repeated after the stable state has been reached.

The time that the multivibrator spends in the unstable state is commonly controlled by a capacitor charging process, where the time constant of a single *RC* stage is used to determine the timings. Figure 9.5 shows the simplest implementation of a monostable multivibrator with the timing components resistor R and capacitor C .

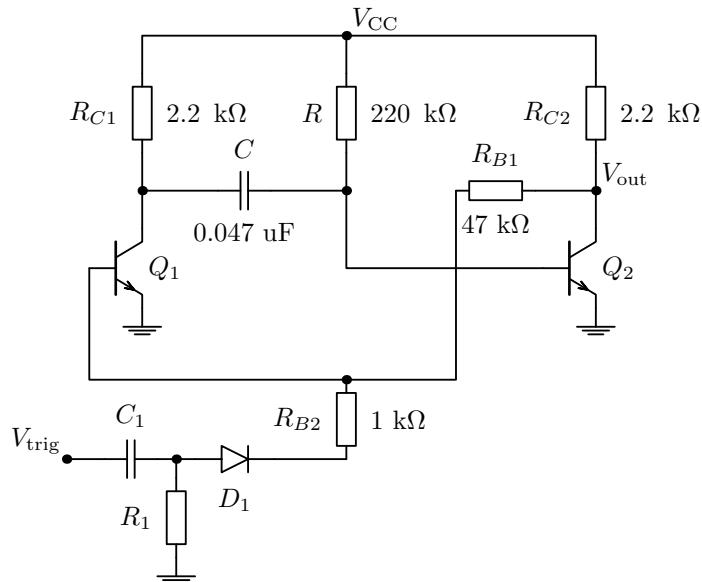


Figure 9.5: A simple monostable multivibrator circuit

When a battery is connected between the V_{CC} and ground terminals of circuit 9.5, transistor Q_2 is conducting with its base-emitter connection forward-biased. The transistor Q_1 is not conducting because of the voltage divider system of resistors R_{C2} , R_{B1} and R_{B2} provides a base voltage from the collector of Q_2 , which is at the ground potential. Obviously the ground voltage is lower than the 0.6 volts

needed to bias Q_1 into the conducting state. Since Q_1 is not conducting, the voltage at its collector approximately equals V_{CC} , and the voltage at the base of Q_2 is equal to the normal base-emitter voltage drop of V_{BE2} . The capacitor between these terminals is therefore charged through resistor R_{C1} to a voltage $V_{CC} - V_{BE2}$ with a time constant $R_{C1}C$. This is the stable state of the circuit.

To trigger the unstable state, a short voltage pulse V_{trig} is fed to the input terminal so that V_{BE1} becomes forward-biased for a short time. This pulls the collector of Q_1 very close to ground and sets the left plate of capacitor C approximately to zero volts. Since the capacitor in this configuration is initially charged so that its left side has the positive charge and the right side has the negative charge, the potential at the base of Q_2 drops to a voltage $-(V_{CC} - V_{BE2})$ and cuts off Q_2 . Now the voltage at the collector of Q_2 is high enough to keep Q_1 forward-biased through the voltage divider resistors R_{B1} and R_{B2} . Even if the trigger pulse fades away rapidly, Q_1 is conducting until the voltage at the terminals of capacitor C has risen just over the base-emitter voltage drop of V_{BE2} to make Q_2 conduct again. In the unstable state, capacitor C is charged from V_{CC} through resistor R , and this makes up the time constant RC that defines the unstable operation time of the circuit. The charging process of the capacitor is a reverse charge, where the positive charge is transferred from the left capacitor plate to the right side. During the unstable period, the output signal at the collector of Q_2 builds up as a square wave.

It is important to note that the process cannot be re-triggered instantly after the unstable period has ended, because the capacitor needs to be fully reverse charged again to match the stable state. This recovering period is determined by the time constant $R_{C1}C$.

The idea of using a capacitor and a diode at the input of the trigger signal is to get an extremely sharp trigger pulse independent of the actual trigger signal used. Actually capacitor C_1 and resistor R_1 put together a high-pass filter which alters long trigger pulses into sharp spikes. The diode passes only the positive part of these spikes to the base of transistor Q_1 .

The semi-analytic derivation for the duration T of the unstable state is relatively similar to the derivation for the period of the relaxation oscillator in section 6.9.1. When the start of the unstable state is taken at time $t = 0$, then the voltage at the base of transistor Q_2 is described with the equation

$$v_{BE2}(t) = V_{CC} - (2V_{CC} - V_{BE2})e^{-t/RC}. \quad (9.6)$$

This equation is valid right after the moment transistor Q_1 starts to conduct. Equation (9.6) means that the capacitor is charged from V_{CC} , which would be the maximum voltage of the fully charged capacitor and that the charging begins at $t = 0$ from the voltage

$$v_{BE2}(0) = V_{CC} - (2V_{CC} - V_{BE2}) = -V_{CC} + V_{BE2},$$

which is intuitively true, as it is in agreement with the previous description of the circuit.

Now to find out the expression for T , all that is needed is to solve it from equation (9.6) at $t = T$. At that time Q_2 starts to conduct, which requires that $v_{BE2}(T) \geq V_{BE2}$. Therefore,

$$V_{BE2} = V_{CC} - (2V_{CC} - V_{BE2})e^{-T/RC},$$

and solving for T gives

$$T = RC \ln \left(\frac{2V_{CC} - V_{BE2}}{V_{CC} - V_{BE2}} \right). \quad (9.7)$$

For a crude approximation this can be simplified to

$$T = RC \ln 2,$$

which gives very slightly shorter times than the more accurate formula (9.7).

A simulation of the simple multivibrator results in the waveforms depicted in Figure 9.6. The capacitor voltage has been measured from the base pin of Q_2 , so therefore the maximum voltage is limited to V_{BE2} and the minimum voltage is limited to $-(V_{CC} - V_{BE2})$. The output voltage is taken from the collector of Q_2 .

The previous example presented a more do-it-yourself way of building a working multivibrator circuit. If there are no restrictions to using integrated circuits in the design, then it is better to get a 555-timer IC and add an external capacitor and resistor to the timer circuit to get better functionality than the discrete component version offers. Because the scientific view is appreciated in this context, the discrete design offers a better platform to understand the inner beauty of multivibrators.

9.3 AN ANALOGUE GUITAR TUNING DEVICE

The first electronic guitar tuning devices have been designed already in the late 1950's. At least one implementation has been patented [111] at that time but

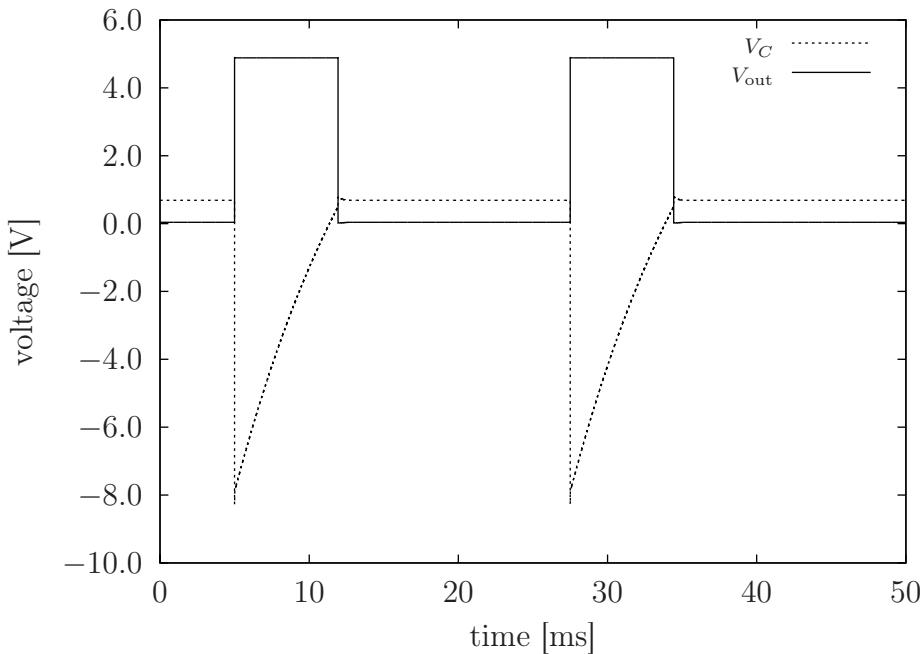


Figure 9.6: The capacitor voltage V_C and V_{out} of the multivibrator circuit

this circuit was intended as a more general frequency measurement device than a guitar tuner. The idea of modifying a sinusoidal input signal to a square wave is introduced in this patent and the same method was later used in several commercial guitar tuning devices. One patent application [112] published in later years is based on the ideas given by [111]. The following discussion gives a short introduction on the implementation presented in patent [112].

The guitar tuner circuit designed by F. G. Allen [112] can be built as a battery operated device, which consists solely of discrete analogue components. In this device the input signal is first band-pass filtered to force the signal to resemble a pure single-frequency sine wave. After the filtering stage the sine wave is over-amplified so that it is clipped at the operating voltage limits and therefore shaped as a square wave. A square wave has much more clearly defined transitions from a positive pulse to a negative pulse and vice versa, so therefore it is suitable for determining the exact frequency of the signal.

The rising edge of the square wave signal is used to trigger a monostable multivibrator circuit which is tuned to have its unstable 'on'-period approximately equal to the half-period of the frequency of the note that the selected guitar string is desired to be tuned to. As the time that is spent at the 'on'-state (T_{ON}) stays constant, the length of the 'off'-state (T_{OFF}) depends on the period of the string's fundamental frequency (T_F) and it is determined as the time difference $T_F - T_{ON}$.

The combination of the rising edge trigger and the fixed time monostable multivibrator creates a continuous square wave, whose *duty cycle* varies with the trigger frequency. The concept of a duty cycle is explained in Figure 9.7.

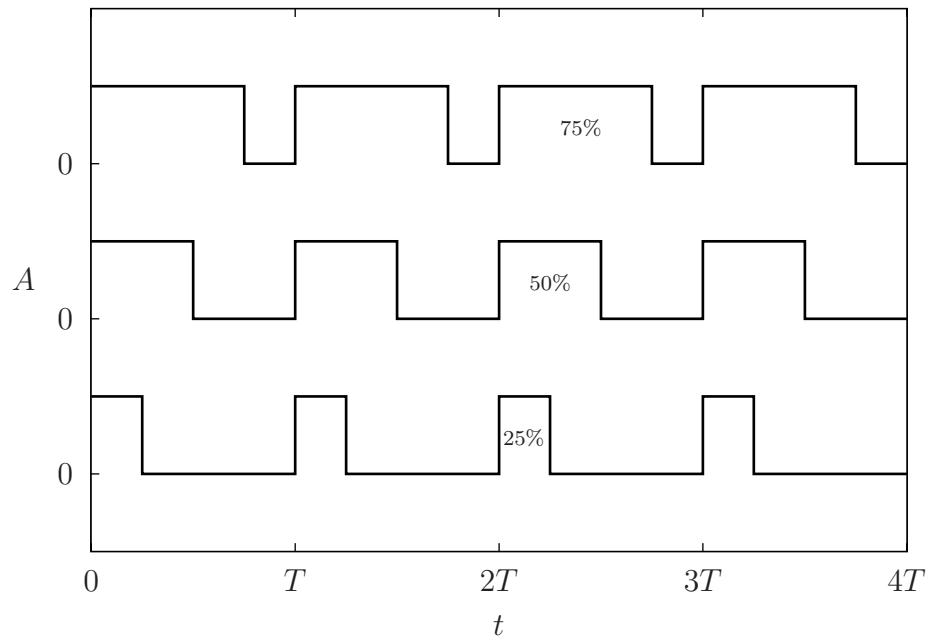


Figure 9.7: A visualisation of square wave duty cycles

The duty cycle of a square wave is defined as the time that the wave spends in active state during one full period of vibration. The duty cycle is reported as a fraction of the period T . Figure 9.7 shows three different periodic square wave signals with duty cycles of 25% 50% and 75%.

As noted, the over-amplification and a triggered monostable multivibrator together create a more or less unevenly duty cycled square wave from the sinusoidal input signal. To convert the actual frequency of the input signal to a unique voltage value, the duty cycle modified square wave is used to charge a simple averaging filter, which consists of a combination of a resistor and a capacitor. The average voltage level at the capacitor depends linearly on the duty cycle of the square wave.

Finally, the average voltage level of the capacitor is directed to a differential amplifier, and the output of this amplifier is used to drive a null meter that shows whether the played note is in tune or not. The indication is given for both high and low notes, but due to the filter, the device has clear limits for how much out-of-tune notes can be accurately detected and tuned to the correct frequency.

This was the functionality description of the tuner presented in [112]. To learn more about this tuning method, the original implementation is redesigned in this context with slight modifications but still using the same basic idea of determining the fundamental frequency of a guitar string from the duty cycle of a sequence of square pulses.

Figure 9.8 shows the schematic of the redesigned circuit. The implementation has been designed to tune only the *A*-string of an electric guitar. The extension to all six strings can be achieved by adding five resistors to the filter and the multivibrator and a switch that can switch between these resistors simultaneously. However, the rationale of this extension should be challenged, since the implementation requires very accurate resistor values to give good tuning results.

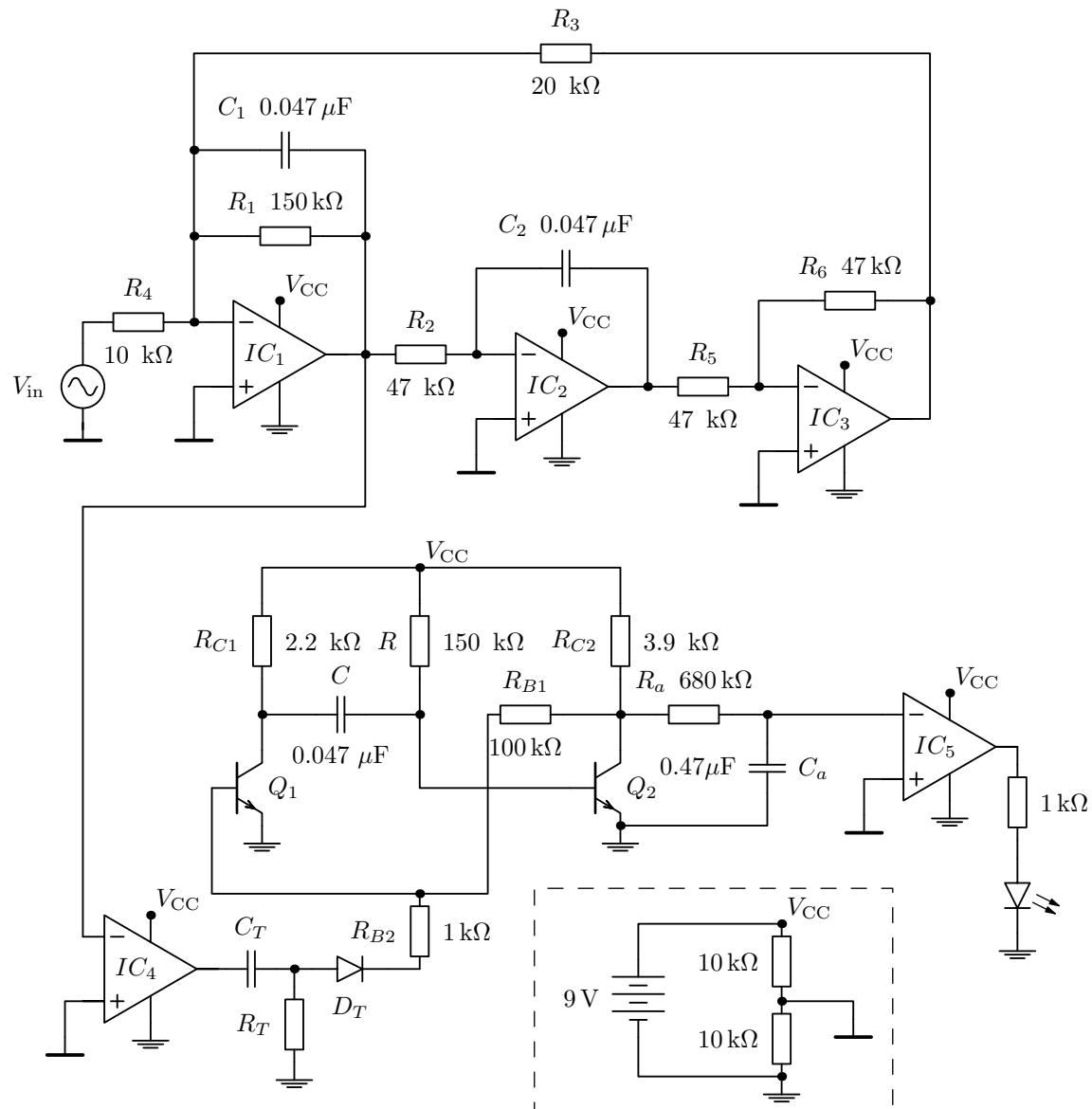


Figure 9.8: A complete circuit diagram of an analogue guitar tuner

The biggest change compared to the original design is to use a LED to indicate if the note is too high or too low. In the stable state when there is no signal coming from the guitar, the LED is lighted to indicate that the power is on. When a note is played, the LED stays on if the note is too low and turns off if the note is too high. The tuning is intended to be done by finding the transition frequency where the LED changes from on to off and the final touch to the tuning is done so that the tension of the slightly low tuned string is tightened slowly until the LED is completely off. This is not that user friendly implementation, but it saves a few components from the design.

After the high level functional description, the next step is to analyse this redesigned circuit in detail by simulations and actual measurements. At first it might seem that the voltage averaging method does not appear to be accurate enough for acceptable tuning. Therefore, the redesigned circuit is just an exploratory journey to learn about the different building blocks of this circuit and to see how accurate tuning can actually be accomplished with this approach. Another goal of the redesign is to minimise the amount of components to simplify the original patented circuit, but not necessarily to make it better.

The filter does not require accurate component values, and it can also be used to amplify the input signal significantly so that even very small amplitude string vibration will create a signal of a few volts. If the gain of the filter is increased significantly, some components are already saved in the beginning compared to the original design.

Figure 9.9 shows the calculated transfer characteristics of the filter with the values given in schematic 9.8. Theoretically the gain of the filter section reaches almost 30 dB, which gives a voltage multiplier of 32 to boost the input signal. In this case the centre frequency is set close to the expected frequency of 110 Hz, but preferably the centre frequency should stay a little under the expected frequency, because then the filter works better when the string is clearly low from ideal tuning. A concrete example is that if the centre frequency of the filter was to be set exactly at 110 Hz (A), then an input signal of 82 Hz (E) would already let the second harmonic slightly through from the filter and this would affect the accuracy to detect a significantly low note.

From the filter, the signal is taken directly to yet another op-amp, which is used as a simple comparator without external resistors, as shown in Figure 9.10. As the rising or lowering edge of the input signal crosses the reference voltage, the

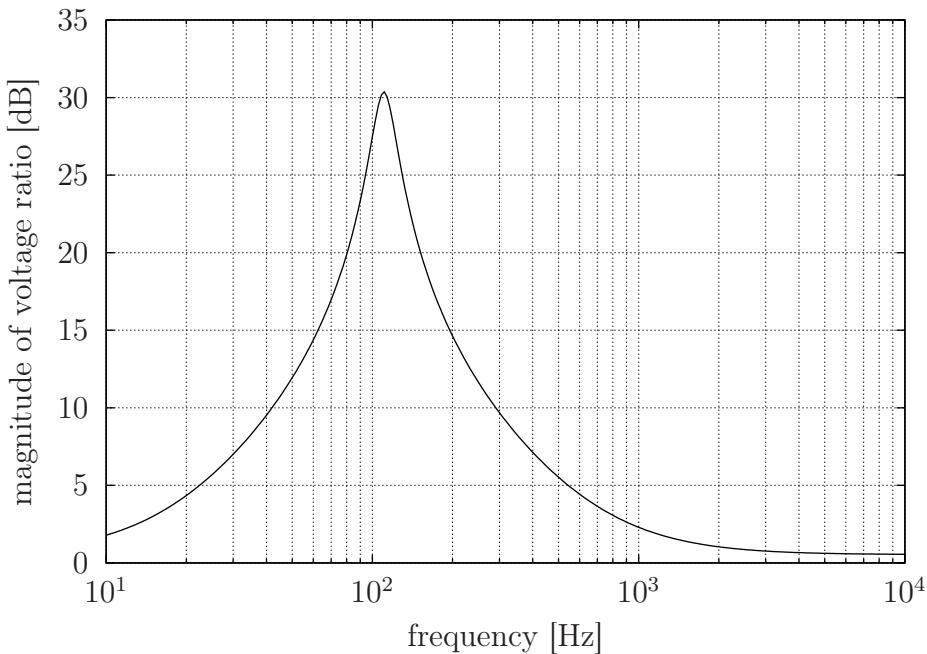


Figure 9.9: Gain obtained from the filter section

output of the comparator shoots close to the operating voltage limits. This op-amp comparator is the key element for creating the sine-to-square conversion.

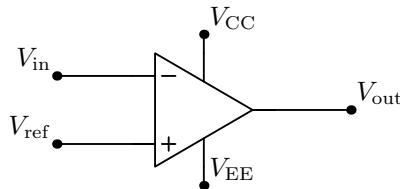


Figure 9.10: Using the operational amplifier as a comparator

A simulation run using the comparator circuit with a sinusoidal input signal gives almost perfect square pulses as shown in Figure 9.11. Because of the simulation setup, the input signal starts out with a small transient effect, which attenuates the amplitude of the first periods of the sine wave. However, the square wave might not be this ideal in practise, since the square wave output is directed to the capacitor at the input of the multivibrator, which acts as a load to the comparator circuit. This capacitive load makes the square wave a bit rounder from the edges.

The reason for using a capacitor at the input of the multivibrator is that it gives a fast spike trigger instead of the square wave coming from the comparator circuit. The advantage of this approach is that now the trigger signal never stays in the high state longer than the pulse provided by the multivibrator.

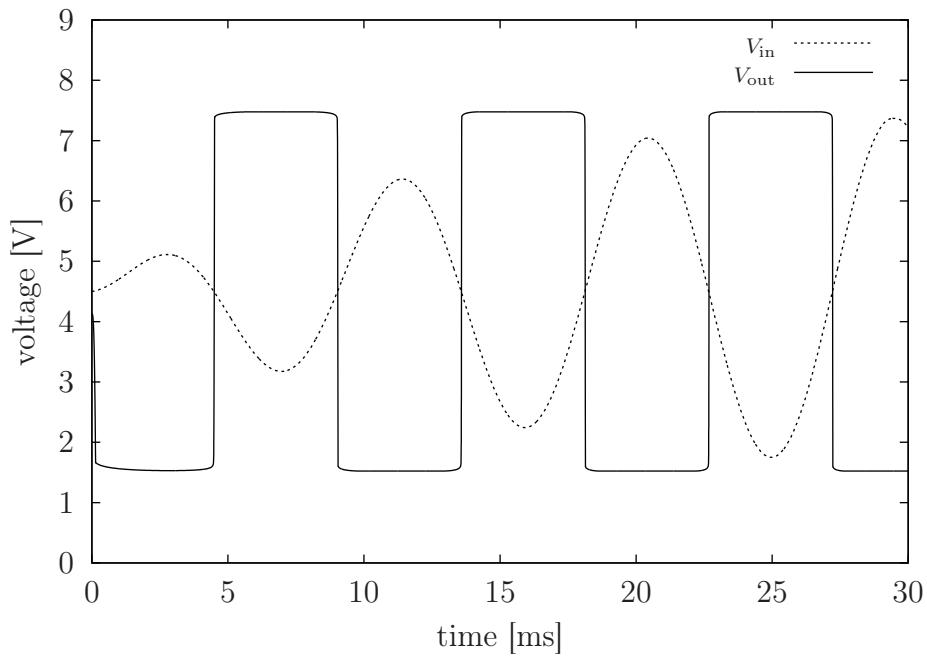


Figure 9.11: The simulated input and output waveforms of the comparator

Figure 9.12 illustrates the trigger signal that propagates to the multivibrator. The simulated measurement has been taken before and after the capacitor C_T . The trigger signal has an upward spike when the comparator shoots high and a

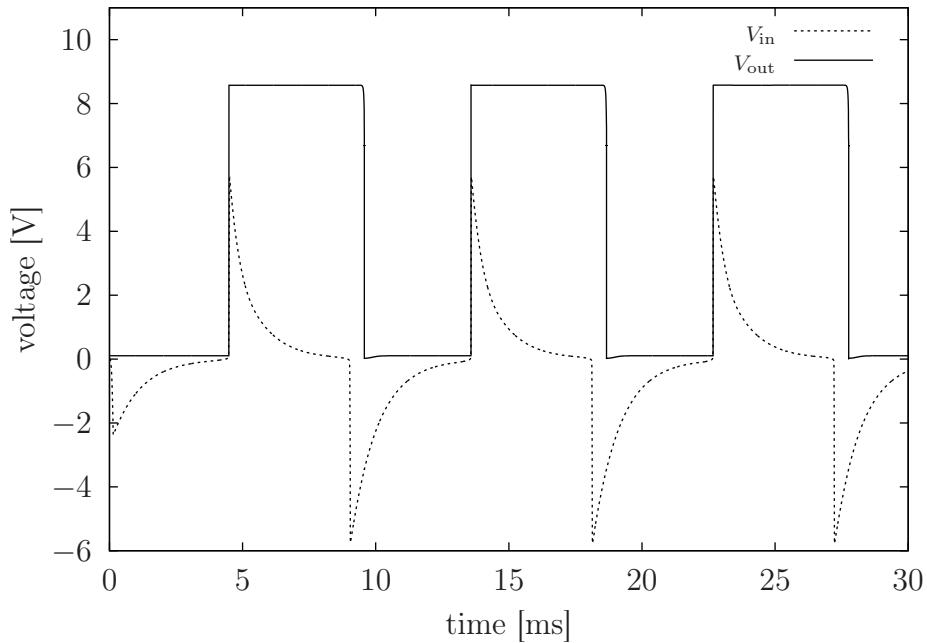


Figure 9.12: The simulated input and output waveforms of the multivibrator

downward spike when the comparator shoots low. The diode D_T is used to filter

out the negative spikes. Figure 9.12 also shows the output signal obtained from the collector of Q_2 of the multivibrator. An important aspect to notice here is that the output does not reach the operating voltage but is a few hundred millivolts short from V_{CC} and ground. This makes things a little more complicated, since now the circuit cannot be directly configured to be in tune with the average voltage of $V_{CC}/2$.

The duty cycle modified output of the multivibrator is taken to a basic integrator circuit, which calculates the average value of the signal. Figure 9.13 presents an integrator circuit made purely from passive components. This realisation effectively does the same job as the alternative version of the integrator using the op-amp in between the resistor and the capacitor.

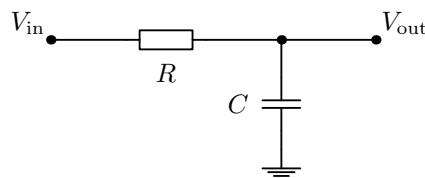


Figure 9.13: A simple RC integrator circuit

The functionality of this critical averaging filter needs to be analysed more carefully, since it affects significantly to the accuracy of the frequency detection. The simulation results of the average voltage measured over the capacitor are illustrated in Figures 9.14 and 9.15. In Figure 9.14, square waves with different duty cycles are driven through the averaging filter and the voltage over the capacitor is measured. It can be seen that the rise time is relatively long, but the average value is reached quite accurately in each case.

The amount of ripple and the rise time of the averaged voltage are connected to each other through the time constant of the RC circuit. Figure 9.15 illustrates the difference when the time constant of the averaging filter is changed. If the time constant RC is short, then the rise time is faster but the ripple is larger. For time constants of longer duration the ripple gets smaller but the rise time leads to a slower response.

It should be mentioned that the component value of R in the multivibrator needs to be fine-tuned experimentally, since the period of the unstable state of the multivibrator does not necessarily directly relate to the target frequency of tuning. This obviously depends on the chosen reference voltage used in the final comparator after the voltage averaging. Furthermore, the fact that the output square

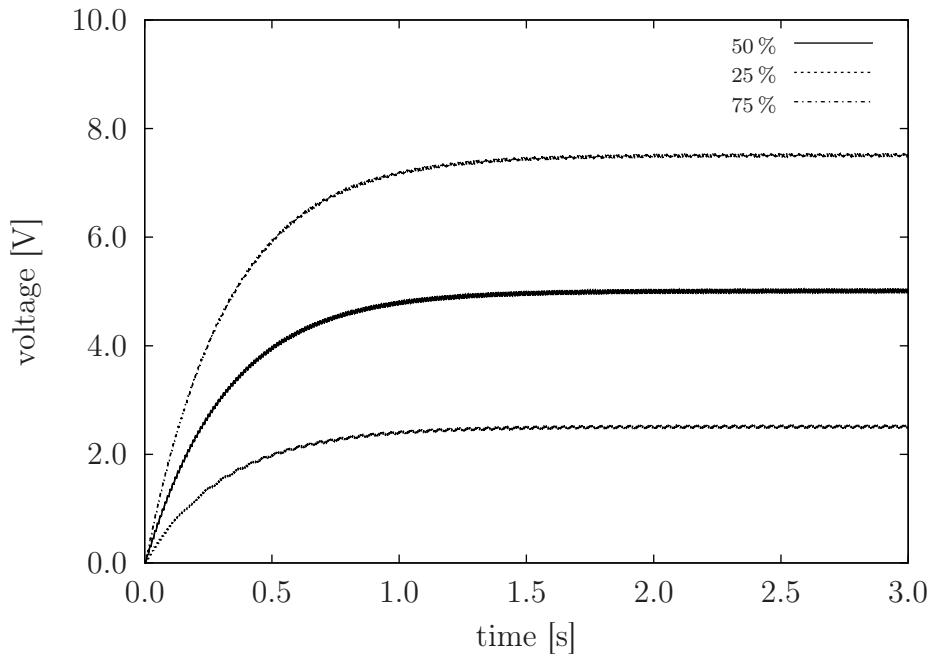


Figure 9.14: Average voltages from different duty cycles

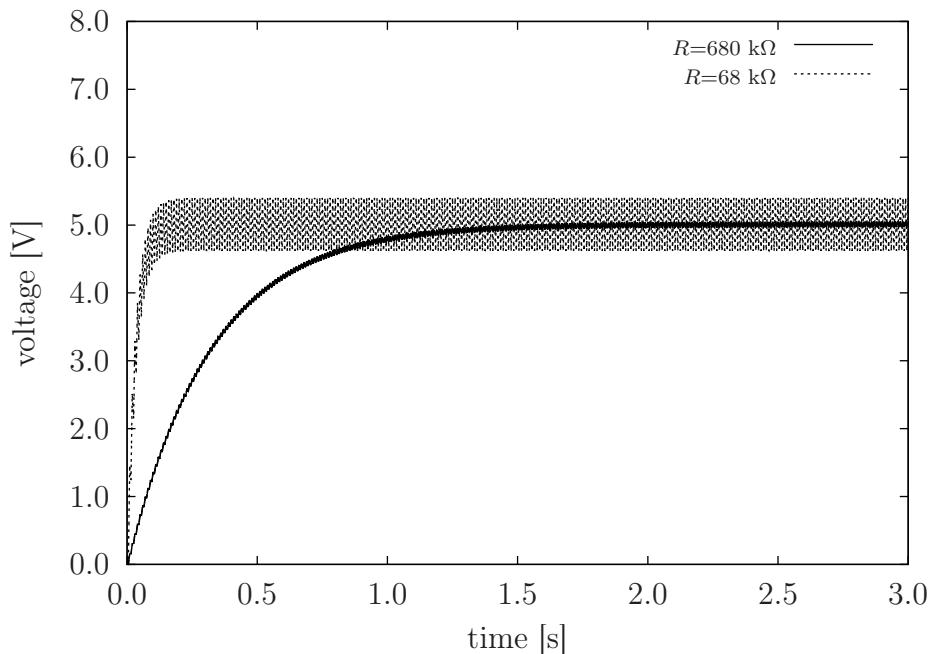


Figure 9.15: Average fluctuation with different resistance values

wave from the multivibrator does not quite reach the limits set by the operating voltages affects the need to adjust R . The need of initial manual tuning of the built tuner circuit itself definitely needs to be considered when building this device.

9.4 MEASUREMENTS ON THE TUNING CIRCUIT

The prototype of the tuning circuit was built onto a solderless breadboard using the uA741 op-amp as a quad LM348 and a single ICs. The possibility to use uA741 indicates that all available op-amp types can be used to realise this circuit. The transistors used in the multivibrator circuit are of the basic 2N3904 type. The capacitors were normal ceramic capacitors and the resistors were typical carbon composition resistors with a 5 % – 10 % tolerance. For better accuracy it is advised to use metal film resistors and capacitors in analogue filters and other timing related circuits.

One of the few things that could not be easily simulated with SPICE was the functionality of the filter with signals that have a large amount of upper partials. Figure 9.16 shows that in the vicinity of the centre frequency the filter performs a perfect cut-off for the overtones.

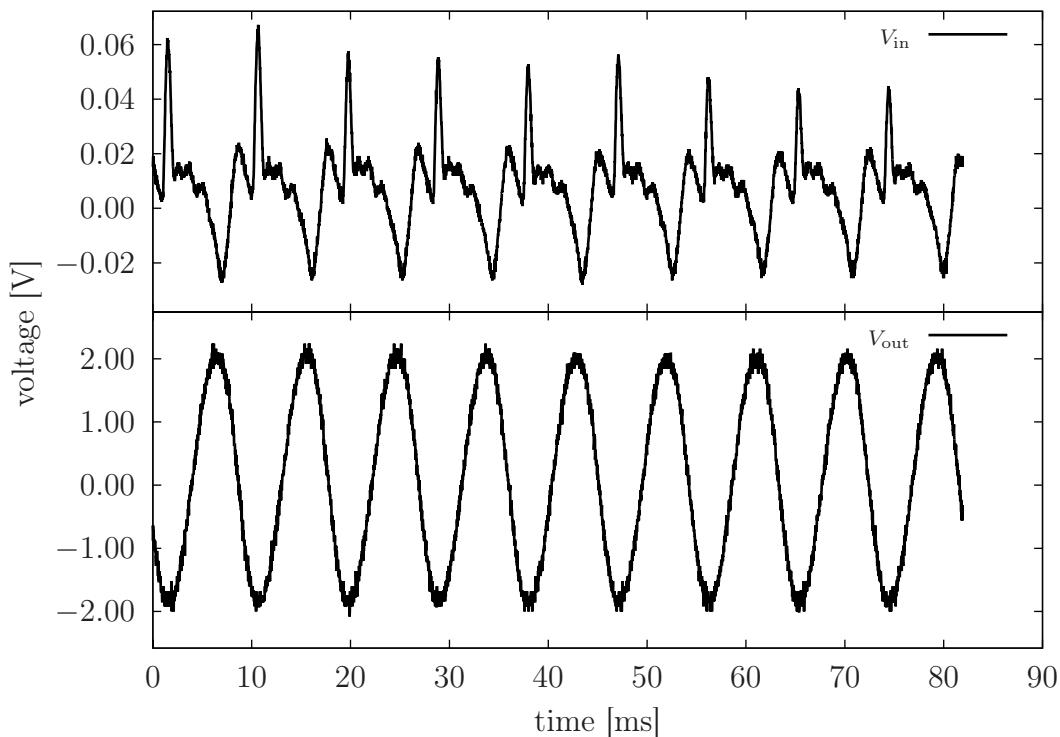


Figure 9.16: Performance of the real-world filter

Although the performance of the filter seems perfect in this figure, the reality of the tuning situation is that the string to be tuned might have a very low frequency compared to the target frequency of tuning. In these cases it is clear that the filter lets a few of the first upper partials through and this usually gives false triggering to the multivibrator in the middle of the fundamental period. To avoid this case,

it would be better to design the centre frequency of the filter a little bit lower than the target frequency.

The gain of the filter can be estimated from the figure to be about 50 ($= 4/0.08$), which is about 34 decibels. This is a bit more than the simulations predicted in Figure 9.9, but this difference is mainly due to the fact that the resistance R_4 used in the measured circuit was lower than the simulated 10 k Ω resistor to get some extra gain out of the filter circuit.

The next step was to verify the square shaping properties of the comparator stage IC_4 . Figure 9.17 indicates that the measured square wave is a bit more rounded from the corners than the simulations predicted. Still the shape of the square wave is not important in this stage of the circuit, but only the sharpness of the rising edge is meaningful to give an exactly timed trigger for the multivibrator stage. It is not critical if the peak-to-peak voltage does not reach the limits of operation voltages at this stage. Therefore, it is possible to use any readily available op-amp model as IC_4 .

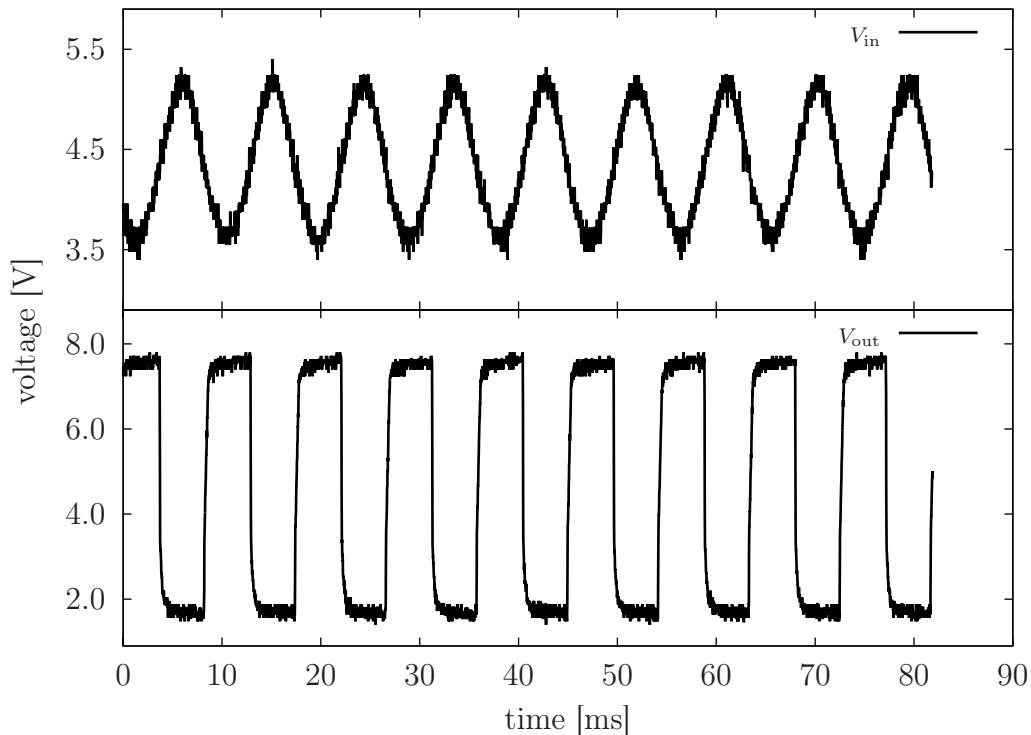


Figure 9.17: The input and output waveforms of the comparator stage

Figure 9.18 shows how the voltage at the input and output of the trigger capacitor looks like. The positive-going spike is aligned with the rising edge of the comparator output signal and the negative-going spike of the trigger capacitor is

aligned with the falling edge of the comparator output. The positive trigger pulse rises only just below two volts because of the loading effect of the multivibrator stage. Still, the two volts are clearly enough to trigger the multivibrator, since it is only necessary to exceed the V_{BE} potential drop of the transistor to be able to successfully trigger the monostable state.

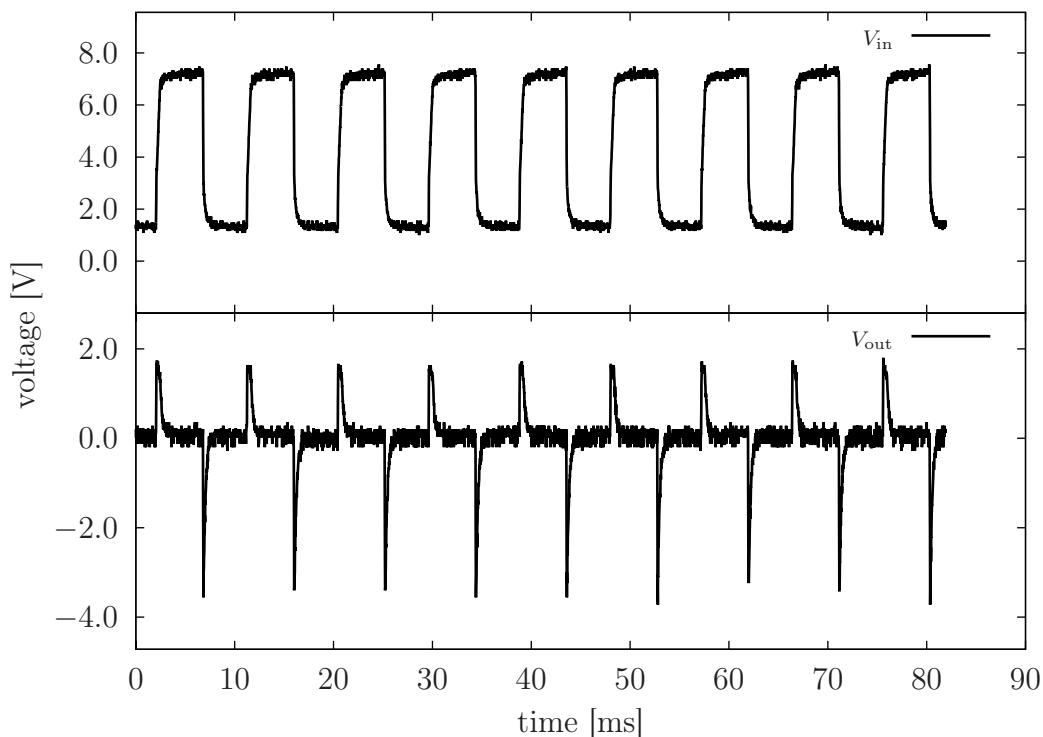


Figure 9.18: The voltage waveform over the multivibrator input capacitor

When the multivibrator is triggered to the unstable state, the voltage over the timing capacitor C drops close to $-V_{CC}$ and starts to charge towards V_{CC} . The relation of the timing capacitor voltage V_C and the output voltage V_{out} of the multivibrator is visualised in Figure 9.19.

It is already clear that the multivibrator output voltage cannot reach the operating voltages, but it can be adjusted closer by making R_{C2} smaller within certain limits. If the value of R_{C2} is taken too small, then it starts to affect the loading time of the timing capacitor C and this is certainly not good.

If there is a real need to get the voltage before the averaging filter closer to the operating voltages, then it is better to add one switching transistor after the output of the multivibrator. Even this addition will not get the voltage swing exactly to the rails because of the excess 0.2 V voltage drop of a BJT transistor in the saturation mode. Although the switching transistor could increase the

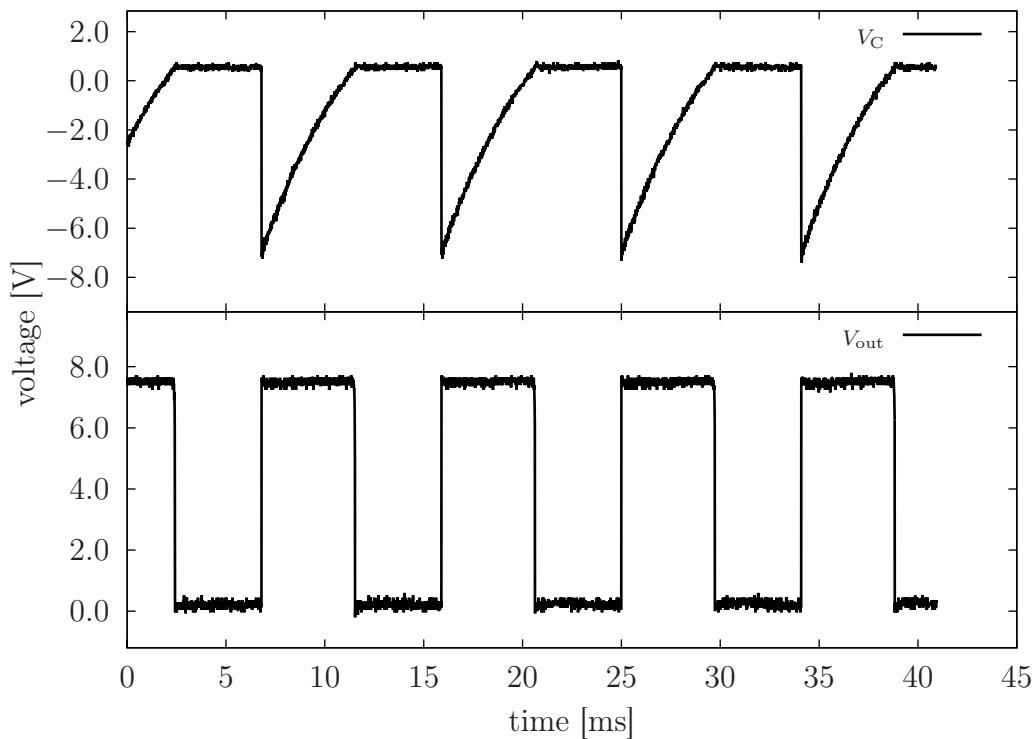


Figure 9.19: The capacitor voltage and the output voltage of the multivibrator

performance of the tuner, it is not added to this solution because of the need to minimise the amount of components.

The final stage of the measurement session was to investigate the voltage waveform between the averaging filter output and the comparator input. The averaged voltage was measured along with the comparator reference voltage from the inputs of the final comparator stage. Figure 9.20 shows two different time captures from both of the inputs measured together.

The upper half of Figure 9.20 shows a still-capture right after the string has been plugged. As expected, the averaged voltage has a steady ripple due to the capacitor charging process. The unexpected thing is that the amplitude of the guitar signal at the filter is so strong that it seems to leak from the common ground point to the reference voltage. The lower half visualises the capture a few seconds after the pluck. At that time the disturbances at the reference voltage have attenuated, but the noise is still larger than the expected symmetric ripple in the averaged voltage.

The measurement results indicating the fluctuating reference level reveal yet another vulnerability in the circuit design. To overcome this issue and to get the

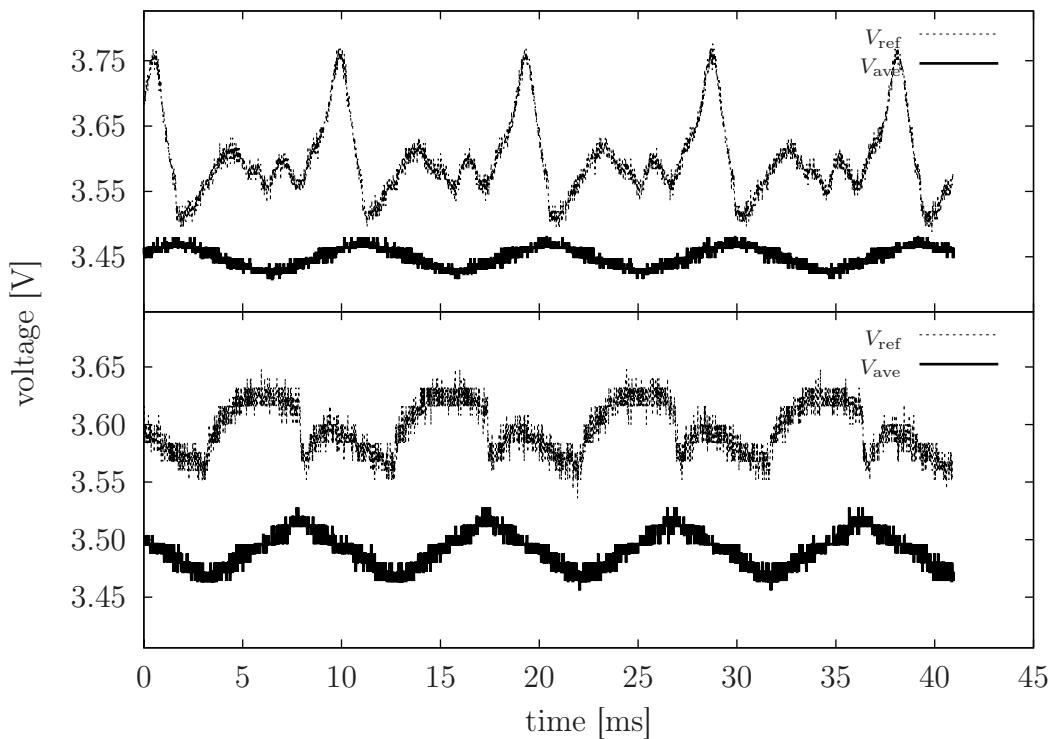


Figure 9.20: The averaged voltage and reference voltage during the tuning process

reference voltage more stable, a more robust way of creating the reference voltage to the circuit should be implemented. Clearly the simple voltage divider trick with two resistors is not adequate in this situation.

In the original patent [112], it is claimed that the tuning result is independent of the operating voltage V_{CC} . This is basically true if a switching transistor is used to ensure that the squared waveform to the averaging circuit has a peak-to-peak voltage range that equals V_{CC} . Then the threshold level to indicate correct tuning and the averaged train of square pulses are both defined via the same voltage. This ensures that changes in the operating voltage alter the threshold and the peak-to-peak voltage with the same fractional amount.

The operating voltage also has a significant effect on the timing periods of the monostable multivibrator. Commercial multivibrator ICs, such as the 555-timer IC, are guaranteed to create constant timing intervals independent of the operating voltage. However, in this particular redesign, the timing period of the naively simple multivibrator solution depends on the operating voltage, as clearly indicated by equation (9.7). This dependency on the operating voltage leads to the data presented in Table 9.2.

Table 9.2: Tuning results with different operating voltages

$V_{CC} = 9.5 \text{ V}$	$V_{CC} = 8.0 \text{ V}$	$V_{CC} = 7.0 \text{ V}$
110.5 Hz	109.9 Hz	109.2 Hz

The tuning reliability was investigated using three different 9 volt batteries, one very old and used battery (7.0 V), a sparingly used battery (8.0 V) and a completely new battery (9.5 V). The variation in the tuning result is basically not that severe, but the differences in the frequencies can be heard by ear. This is definitely not acceptable.

Prior to identifying the problems relating to the operating voltage level, the inconsistent tuning results were suspected to arise from the fact that the multivibrator does not provide an output signal that would reach from zero volts to V_{CC} . The output voltage swing is left short from the rails mainly due to the $3.9 \text{ k}\Omega$ resistor R_{C2} . This problem was tried to be corrected by a simple transistor switch, which is drawn in Figure 9.21. The base resistance R_B of the switch needs to be determined so that the base gets enough current to fully saturate the transistor when the input signal is high. When the input signal is low, the transistor goes to the cut-off state and the resistance of R_C should be relatively small to prevent any unnecessary voltage drop due to possible excess current flow at the collector. As a general note, the switch inverts the output signal of the multivibrator. If the transistor switch is used, the input to the comparator IC_5 should be changed to the non-inverting pin (+), and the inverting pin (-) should be grounded.

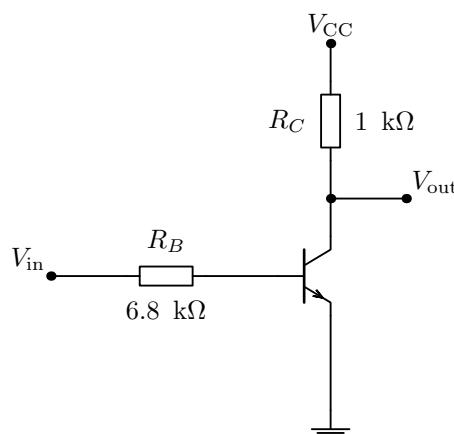


Figure 9.21: Using a bipolar junction transistor as a simple switch

By using the transistor switch it is possible to get a square wave ranging very close to the operating voltages, but still a BJT transistor normally has about a 0.2 volt drop across it in saturation. Making R_C smaller than 1 k Ω seemed to be a bad choice in the actual circuit, because then the circuit was taking too much current from the battery. When using the weakest battery, the high current consumption lead to an additional drop of V_{CC} when the switching transistor was in saturation. For similar reasons it was not possible to make R_{C2} in the multivibrator output smaller than 3.9 k Ω , because when tried, it lead to obscure transient effects in the capacitor charging process and caused a noticeable timing error.

After adding the switching transistor between the multivibrator and the averaging filter in the redesigned circuit, the problem of V_{CC} dependent tuning results did not vanish. Therefore, the next guilty component under suspicion was the multivibrator, and indeed when inserting actual values into equation (9.7), the timing period grows slightly as the operating voltage is decreased. This increase in the timing period leads to a small increase in the duty cycle, and this slightly increases the averaged voltage. This way the lowered operating voltage causes the slightly raised averaged voltage, which switches the state of the comparator sooner than expected. Most likely the use of a commercial timer IC as the multivibrator would solve this issue and the tuner would be more stable, accurate and most of all more usable.

If a sequential tuning is performed over a short time period and the tuning is always performed by tightening the string until the LED goes off, then the practical accuracy of this tuner can reach something like ± 0.5 Hz for a 110 Hz target frequency. To be on the safe side, a maximum error of ± 1 Hz is more reasonable in normal tuning. The absolute value of the error grows if the target frequency is raised, but the relative error stays approximately constant.

In the enhanced version of this analogue tuning device, the signal leakage to ground should be prevented and a more accurate comparator solution to the last section of the tuner should be developed. With these corrections the tuner could compete in accuracy with the commercially available digital guitar tuners.

9.5 THE STROBOTUNER

The main idea of the Strobotuner is to visually indicate the correct tuning by controlling flashing light bulbs with the signal of the device to be tuned and syn-

chronously rotating a motor at user-chosen speed. The speed of motor rotation is related to the reference frequency of the tuned note. The motor is spinning a disk that lets the light through it at the chosen frequency, and the flashing lights behind the disk go on and off at the rate determined by the fundamental frequency of the instrument. The sound of the instrument is sensed by recording the sound via a microphone.

The phenomenon behind the idea of the Strobotuner is familiar, for example, from old cowboy movies, where the wheels of a moving wagon suddenly stop rotating when the speed of rotation and the frame rate of the film are 'in tune'. When the instrument is in tune, the stroboscopic effect in the tuner shows a semicircle of motionless lines around the disc. If the tone is sharp, the lines will appear to revolve to the right or clockwise. If the tone is flat, the lines in the disc will appear to rotate to the left or counter-clockwise.

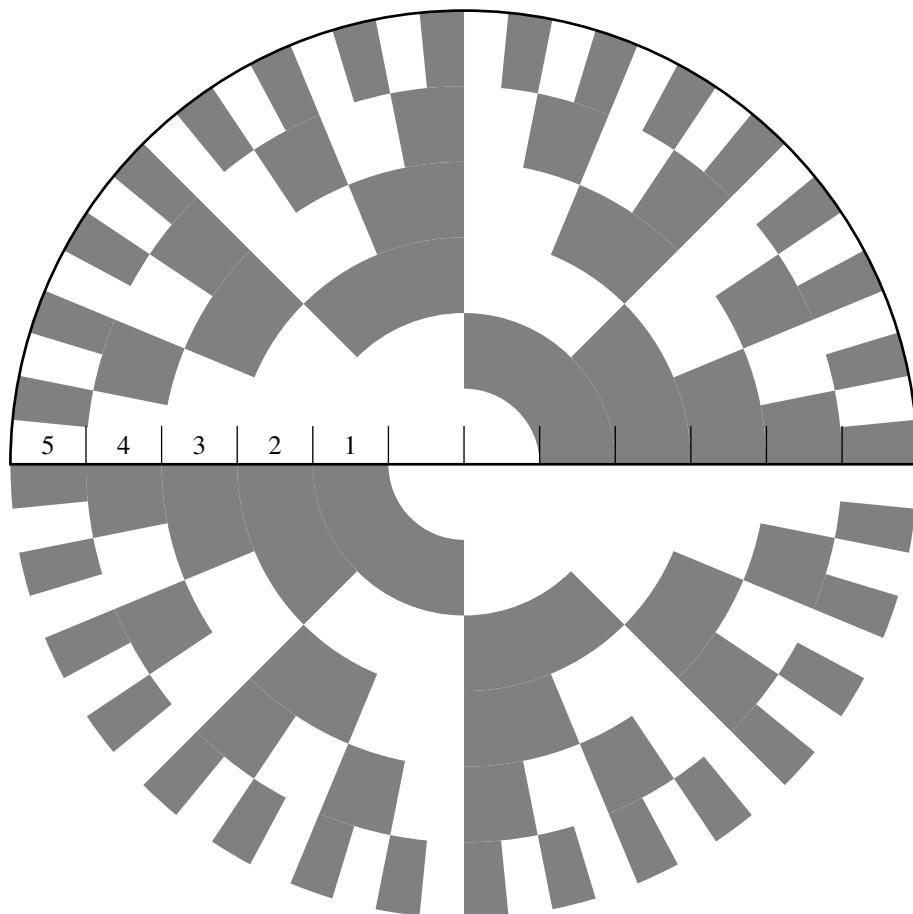


Figure 9.22: The rotating disk of the Strobotuner

The geometrically interesting disk of the Strobotuner is depicted in Figure 9.22 and it is the part of the device that makes it all happen. As the disk rotates, the

lights flicker at a certain frequency, and when the rotation speed of the disk and the flickering speed of the lights are in tune, the instrument is in tune as well. The figure shows only five octave bands on the disk because of drawing limitations, but the original Strobotuner is designed for eight octave bands, which can be extended even up to ten octaves with a special switch.

Figure 9.23 reveals the user interface of the Strobotuner. The sound from the instrument to be tuned is recorded by a microphone attached to 'mic in'. The sensitivity of the microphone can be controlled by the 'gain' potentiometer. The gain should be adjusted high enough to clearly see the illumination of the lights behind the strobodisk when a note is played from the instrument.

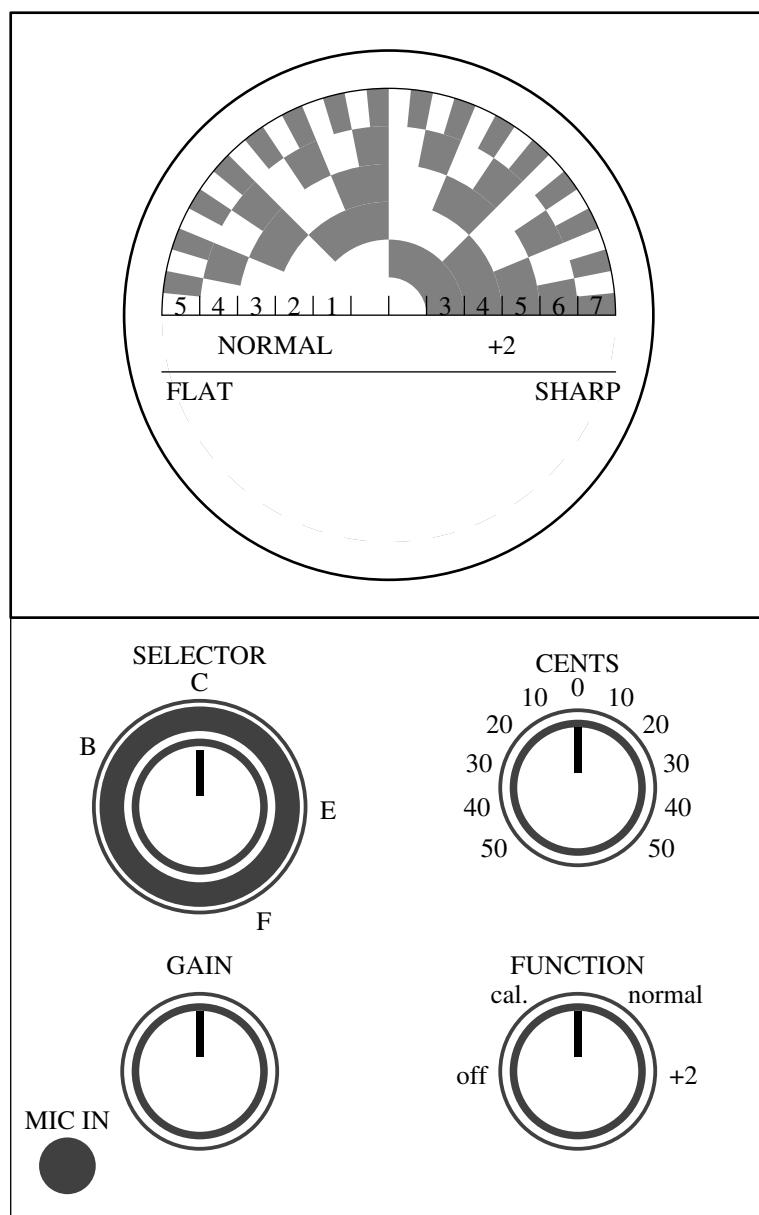


Figure 9.23: The front panel of the Strobotuner device

The main mode of operation is chosen from the 'function' switch. The calibration mode is used prior to the actual tuning to ensure that the device is working as intended in a specific tuning environment. The need to change the calibration mainly arises from ambient temperature changes, because the ambient temperature affects both the tuning device and the instrument to be tuned. The 'cents' switch is used in the calibration phase to make fine adjustments so that the calibration signal appears to be in tune in the specific tuning environment. After the calibration, the 'cents' switch is not adjusted during the actual tuning process.

The actual tuning is done in the 'normal' mode and extended range tuning can be performed in the '+2' mode. The extended range offers the possibility to expand the octave range of the Strobotuner to two octaves higher, so ultimately the eight octave range can be expanded as a ten octave range.

In the 'normal' mode, the user can choose the note to be tuned from the 'selector' switch. The selector switch is shown in full detail in Figure 9.24.

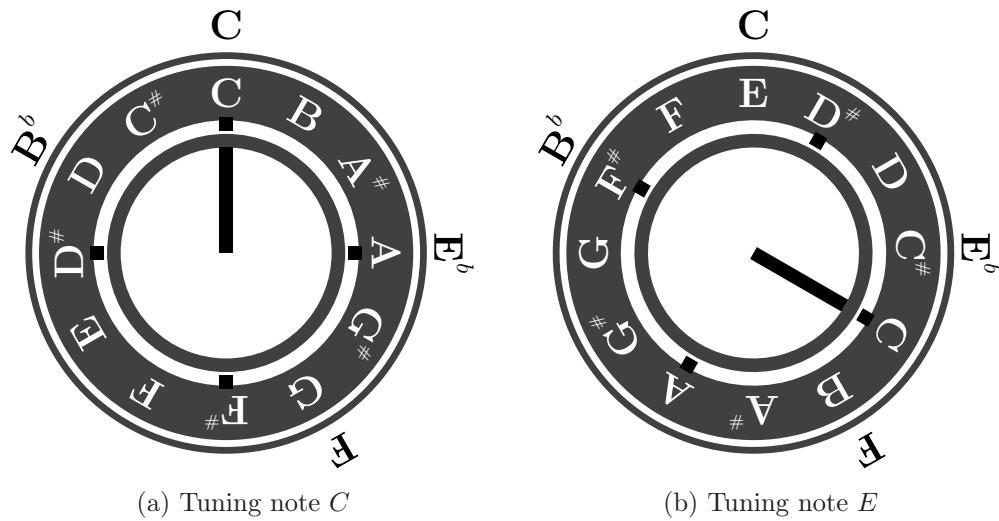


Figure 9.24: Selecting the tuned note from the selector switch

For tuning guitars or basses to normal tuning, the selector switch should be positioned so that the note to be tuned (marked at the rotating selector switch) is aligned with the letter *C* that is stationarily marked on the front panel. Figure 9.24a shows the selector switch when tuning to note *C* in the normal scale and Figure 9.24b indicates the positioning of the selector switch when tuning to note *E* in the normal scale.

Typically the majority of instruments are tuned using the *C* alignment, which can be considered the normal tuning. An instrument that uses a *C* tuning plays out

a frequency that is the same as the standard definition of the note *C* frequency. A guitar is definitely a *C* instrument, also referred to as a *non-transposing* instrument.

Some instruments require special tunings and typically these are called *transposing* instruments. If a special tuning is needed, then the selector switch can be aligned with the three other possibilities *B^b*, *E^b* or *F*. This way the selector switch can be used to transpose different instruments to the required special tuning by aligning the tuned note to *F*, for example, instead of the typical *C*.

The instruments that normally require transpositions include the wind instruments, which do not produce the same pitch that is read from the staff. A regular clarinet is a transposing instrument of *B^b*, which means that if the *C* note is played with a clarinet, it would sound like *B^b*. This is why clarinet players cannot read the same notes as guitarists if they want to play in the same tune together. [113]

The Strobotuner is another example of an analogue tuning device that cleverly makes use of controlling the rotation speed of an electric motor at a certain angular frequency to generate a reference for the tuned note. A clear visual indication using flashing lights is given to the user to verify the status of tuning. If someone has interests towards the detailed implementation of the Strobotuner, the circuit diagram of the device and other fun facts are given in the Strobotuner user manual [113].

9.6 AUTOMATIC TUNING SYSTEMS FOR GUITARS

This section includes a few hand waving notes and remarks about some of the commercially available self-tuning guitars. The technology behind these gadgets is described without detailed mathematical analysis because the systems rely solely on digital signal processing - the only feasible way to create a system that actually works fast and accurately.

At the time of writing, there are two elegant systems in the markets that automatically tune all strings at once with one strum of the strings. The more obvious implementation is the 'Robot Guitar', where small motors are used for turning the tuning pegs while a separate microprocessor is analysing the fundamental frequencies of each string and controlling the motors accordingly. The other solution extends the 'Autotune' technology from vocal pitch corrections to guitar

string pitch correction without mechanically tuning the strings.

To tune all strings at the same time reliably requires a separate sensor to measure the fundamental frequency for each string. The 'Robot Guitar' uses six individual piezoelectric transducers for each string. The piezosensors are used as the saddle, just like described in section 3.4.4 concerning the measurements of a piezoelectric pickup. This implementation allows the guitar to have normal magnetic pickups, and it does not interfere with the actual electromechanical behaviour of the guitar in any way.

The 'Autotune' guitar uses a hex-pickup in place of normal magnetic pickups. The hex-pickups are also designed to sense each string individually, and they have been earlier used in some MIDI-guitar applications. The downside of this solution is that the hex-pickup cannot be replaced by standard single-coil or humbucker pickups, which clearly restricts the freedom of making do-it-yourself mods to the guitar later on.

A crude block diagram for an automatic guitar tuning system would contain only two blocks, one for determining the frequency of a string and the other one for controlling the specific actuator that corrects the tuning.

In both implementations, the 'Robot Guitar' and the 'Autotune', the fundamental frequencies of the strings are determined using fragments of code that handle the audio signal post processing in the digital domain. If the signal has even a slight tendency for periodicity, an autocorrelation calculation can be used to determine the fundamental frequency of the signal. The autocorrelation (or cross-correlation) function is normally used to determine a delay between similar signals. In the case of periodic signals, autocorrelation can be used to find out the delay between one period of a periodic signal.

After the fundamental frequencies of the strings have been determined, the next step is to make the required corrections towards target frequencies. The 'Autotune' system uses a pitch shifting algorithm to modify the frequency content of the signal in the digital domain and then transfer this frequency corrected digital signal back to an analogue signal. It is quite confusing to think that the strings in the guitar can be way off from the correct tuning, but the electric signal obtained from the guitar is in perfect tune.

The 'Robot Guitar' uses small motors to do the required tension adjustments to the string as it would normally be manually tuned. The fundamental frequencies

of all strings are constantly monitored by the built-in microcontroller and the strings are tightened or loosened according to the calculated delta to the target frequencies. The motors can be placed at the tuning pegs or at the bridge of the guitar, from where it is also possible to tighten the strings by using an alternative tuning mechanism. Using motors to tune the guitar is considerably slower than using the pitch shift algorithm to dynamically alter the frequencies of the strings.

CHAPTER

10

OUTRO

The previous chapters have covered the whole signal transmission chain of the notes played on an electric guitar. This signal path served as a natural basis for this book and created a rationally organised structure where the material of previous chapters helps to understand the topics of the following chapters. In addition, the electric guitar and related accessories contain so much fundamental applications of basic scientific research that this approach can be used as a platform to learn a major part of university level physics. The amount of scientific theories and applications related to electric guitars and especially music in general is so enormous that this book only scratches the surface of the world of musical science and electronics.

Simplified mathematical models were used in each section to show that the designs of electric guitar related technologies can be efficiently analysed and simulated at home using only a PC and some freeware programs. Detailed analysis of simple effect devices should lay a solid foundation for advancing towards more complicated implementations. Hopefully the contents of this book encourages all readers to develop their skills and knowledge beyond the treatment presented in this book. Ultimately, ideas and competence for completely new designs have been adopted. The following paragraphs summarise each chapter, the key topics and ideas of the book in a short and leisurely fashion.

The analysis of the vibrating string focused on presenting the string as a source of tone with specific harmonic content. The variables that affect the structure of the upper partial magnitudes were all identified to relate to the geometrical shape of the string. In the initial state, the geometry of a plucked string is mainly determined by the plucking position, displacement amplitude, width of the plectrum and stiffness properties of the string. The higher harmonics in the string are

more noticeable when the string is plucked near the end support of the string, but even then the fundamental frequency is the most dominant in the spectrum of upper partials.

Many factors, such as length modulation during vibration, stiffness, unideal end supports and different types of friction, cause the string to become a nonlinear medium of propagation for transverse and longitudinal waves. The nonlinearities also affect the tonal properties of the string by slightly modifying the upper partial content in the harmonic spectrum presentation. Despite all the nonlinearities, the behaviour of the string to a certain accuracy can be modelled mathematically using the linear model of an ideal string.

In an electric guitar, the mechanical vibratory motion of the string is transformed as a signal in the electrical domain using transducers, which are commonly referred to as 'guitar pickups'. Three different types of pickups (magnetic, piezoelectric and optical) used in modern electric guitars were treated as general transducers and analysed from the viewpoint of linearity and electrical properties. Despite being the most nonlinear pickup, the familiar magnetic pickup enhances the strength of the higher harmonics of the string and makes the sensation of tone livelier and richer. The magnetic pickup is sensitive to string vibrations in all directions, whereas the optical and piezoelectric pickups fail to produce evenly changing signals when the string vibrates in elliptical trajectories in normal playing situations.

The optical and piezoelectric pickups reproduce the most accurate replica of the string's harmonic spectrum, because both pickups can be categorised as displacement sensitive pickups. The magnetic pickup, on the other hand, is a velocity sensitive pickup, whose signal is not directly related to the actual displacement trajectories of the string's motion. Therefore, the magnetic pickup creates its own spices to the sound of the string, making it the best choice among musical pickups for electric guitars.

The optical pickup can be considered an active pickup because it requires a separate voltage source to operate. The magnetic and piezoelectric pickups are self-contained passive pickups that create the electric signal directly from the physical interaction with the motion of the string.

The user interface in the electric guitar normally contains knobs for controlling the tone and volume. The passive tone control circuits encountered in electric

guitars are typically very simple first order filter circuits, which can only be used to cut out certain frequencies from the signal received from the pickups. The most common filter type in electric guitars is the first order low-pass filter, where the high frequencies are cut off at a certain corner frequency. To control the tone, a guitar player can adjust the corner frequency of the low-pass filter with a potentiometer. The low-pass filter implementation indicates that the higher upper partials are the ones that most dramatically change the sound of the strings. The default tone control configuration in the electric guitar can be modified to a certain degree until the complexity progresses towards active filters, which are more commonly encountered in guitar amplifiers.

The guitar's volume control interface is always implemented using a single potentiometer. The output impedance of the guitar, as seen by the amplifier, is approximately equal to the impedance of the volume pot. However, when magnetic pickups are used, the impedance of the pickup is connected parallel to the volume control potentiometer. The volume potentiometer normally has such a high resistance that the output impedance of the guitar is dominated by the low impedance of the pickup at low frequencies. The impedance of a magnetic pickup changes radically near the resonance frequency of a few kilohertz. At low frequencies the impedance of the pickup is dominated by the DC resistance of the pickup, which is about $10\text{ k}\Omega$. Near the resonance frequency, the impedance is dominated by the inductance and capacitance of the magnetic pickup, which can be as high as a few megohms.

It is easy to think that the cables used to connect the guitar to the amplifier or to the effect devices would have no effect on the sound of the electric guitar. By theoretical calculations it is possible to prove that the capacitance of a guitar cable can attenuate the frequency response of the higher upper partials in the frequency range of a few kilohertz. Additionally, the cable capacitance is responsible for shifting the resonance frequency of the magnetic pickup towards lower frequencies. The construction of high quality guitar cables imitates the internal structure of coaxial cables, but shielded instrument cables are more flexible and have different nominal impedances. It is not advised to use a paired cable as an instrument cable because the quality of tone will be reduced considerably due to lousy shielding, crosstalk and other related issues.

The field of analogue guitar effects is extremely widespread. Almost all possible studio effects have been implemented as stand-alone guitar effects pedals. Many of the popular effects built in the early days of analogue electronics are extremely

simple but yet effective devices. They are also fruitful for studying and learning all the basic elements of analogue electronics. A variety of effect devices have been designed over the years for manipulating all elementary signal properties of amplitude, frequency, phase, and even time.

Typically the field of professional signal processing strives to retain the quality of the signal at the best possible level, but the guitar effects usually do just the opposite. Most of the effect devices create distortion, frequency sweeps, amplitude modulation, reverberation and phase shifts; all the possible tricks to mess up the original signal of the guitar. A guitar effects designer could not work in the same building with a Hi-Fi audio amplifier designer. When the effects designer is content with any Lo-Fi audio heard from the guitar amplifier, both of them are pulling the rope in opposite directions.

Still, quite many agree that effect devices create tones that make the guitar sound more interesting as an instrument. At least the songwriter has more freedom to play with different sounds to come up with new ideas for the next hit single. As an example, an old classic song can be coloured up and revitalised by playing it through a set of carefully chosen sound effects.

An electric guitar is nothing without an amplifier, so the musician buying a guitar is also forced to buy a heavy-duty amplifier. Transistors and tubes are the essential electrical components to enable amplification of electric signals. In guitar amplifiers, these components are used in their natural habitat as discrete amplification elements. Again, old and simple guitar amplifiers lay down the best practical guidelines for learning about the basics of audio amplification technology.

The golden rule of thumb is: never talk to a Hi-Fi fanatic about loudspeakers, because the discussion will eventually develop into a monologue of unbelievable jargon which will ruin your whole day. Actually, never discuss anything related to music with a Hi-Fi guy. But anyway, loudspeakers are needed as the final electromechanoacoustical transducer stage to transmit the notes played on the guitar to be received by the ears of the audience.

Guitar tuners are your best friends if you want to play in tune. It is also possible to survive without a tuner device only by trusting your own ears. All that is needed is a sound source of a known reference note. All the other tunings can be made by ear relative to the one fixed reference. If you are an engineer, collect

all the possible tuning devices in the world and take them apart to see how they are actually made. When using the knowledge of the upper partial structure of a plucked guitar string, one can develop a tuning technique based on the plucking position and the pickup location that will directly generate a relatively clean sine wave. A pure sine wave is an optimal signal when trying to detect the correct frequency for fine-tuning the instrument. Applications of modern technology have finally produced feasible solutions for building self-tuning guitars. Automatic tuning systems are interesting from the scientific point of view, but obviously not necessarily required to play the guitar in tune.

In addition to all the guitar related information, many examples of practical circuit analysis methods have been given. Especially the systematic matrix methods to solve transfer functions from small-signal models in symbolic form are very useful in many ways. The use of the matrix method does not require the skills to 'invent' adequate current and voltage equations from a circuit diagram. Instead, it offers a very straightforward method of solving expressions for currents and voltages in all sections of the circuit under analysis. From the symbolic expressions it is often possible to identify the key components that affect the gain and the cut-off frequency of the analysed circuit. If the matrix equation is too massive to be solved in symbolic form, mathematical software such as Octave or Matlab can be used to solve numerical values for the necessary determinants.

Furthermore, many practical examples have been provided from SPICE simulations of analogue circuits. Especially in analogue electronics, SPICE is a very useful tool for simulating the essential functionality of almost any circuit. With one simulation run one can relatively easily solve the quiescent voltages, frequency response and transient behaviour of a circuit. When looking under the hood, SPICE is more or less based on the same matrix methods that are used in nodal analysis, which was used extensively in almost all chapters of this book.

This concludes the life journey of a ♪ which successfully made its way through the complete signal chain of an electric guitar.

REFERENCES

- [1] Rahn, J. (1983) *A Theory for All Music*. University of Toronto Press, ISBN 0-8020-5538-9.
- [2] Blatter, A. (2007) *Revisiting music theory : a guide to the practice*. Routledge, ISBN 0-415-97440-2.
- [3] Young, H. D. and Freedman, R. A. (2008) *University Physics with Modern Physics*. Pearson Addison-Wesley, 12th edn., ISBN 0-321-50130-6.
- [4] Beament, J. (2001) *How we hear music: the relationship between music and the hearing mechanism*. Boydell Press, ISBN 0-85115-940-0.
- [5] Helmholtz, H. (1954) *On the Sensations of Tone*. Dover Publications, The 2nd English edition with respect to the 4th german edition of 1877.
- [6] Loy, G. (2006) *Musimathics: the mathematical foundations of music*, vol. 1. The MIT Press, ISBN 0-262-12282-0.
- [7] Fourier, J. B. J. (1955) *The Analytical Theory of Heat*. Dover Publications, The English translation of the original publication.
- [8] Loy, G. (2007) *Musimathics: the mathematical foundations of music*, vol. 2. The MIT Press, ISBN 0-262-12285-5.
- [9] Davis, D. and Patronis, E. (2006) *Sound System Engineering*. Focal Press / Elsevier, 3rd edn., ISBN 978-0-240-80830-7.
- [10] Hassani, S. (2006) *Mathematical Physics - a modern introduction to it's foundations*. Springer, ISBN 0-387-98579-4.
- [11] Kreyszig, E. (2006) *Advanced Engineering Mathematics*. John Wiley and Sons, 9th edn., ISBN 0-471-72897-7.
- [12] Ogata, K. (1997) *Modern Control Engineering*. Prentice Hall, 3rd edn., ISBN 0-13-261389-1.
- [13] Adams, R. A. (2006) *Calculus : A Complete Course*. Pearson Education, 6th edn., ISBN 0-321-27000-2.
- [14] Horowitz, P. and Hill, W. (1980) *The art of electronics*. Cambridge University Press.
- [15] Millman, J. and Grabel, A. (1987) *Microelectronics*. McGraw-Hill, 2nd edn., ISBN 0-07-100596-X.
- [16] Sze, S. M. (1981) *Physics of Semiconductor Devices*. John Wiley and Sons, 2nd edn., ISBN 0-471-09837-X.
- [17] Ashcroft, N. W. and Mermin, N. D. (1976) *Solid State Physics*. CBS Publishing Asia, ISBN 0-03-049346-3.
- [18] Lonngren, K. E. (1988) *Introduction to Physical Electronics*. Allyn and Bacon, ISBN 0-205-11409-1.
- [19] Millman, J. and Halkias, C. C. (1967) *Electronic Devices and Circuits*. McGraw-Hill, Library of Congress Catalog Card Number 67-16934.
- [20] Olson, H. (1967) *Music, Physics and Engineering*. Dover Publications, 2nd edn.
- [21] Fletcher, N. and Rossing, T. (1999) *The Physics of Musical Instruments*. Springer-Verlag New York, 2nd edn.
- [22] Traube, C. and Smith, J. O. (2000) Estimating the plucking point on a guitar string. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pp. DAFX1–DAFX6.
- [23] Morse, P. M. (1981) *Vibration and Sound*. Acoustical Society of America, 2nd edn., ISBN 0-88318-876-7.
- [24] Morse, P. M. and Ingard, U. K. (1986) *Theoretical Acoustics*. Princeton University Press, Originally published by McGraw-Hill, 1968.
- [25] Fletcher, N. H. (1976) Plucked strings - a review. *Catgut Acoustical Society Newsletter*, vol. 26, pp. 13–17.
- [26] Gough, C. E. (2000) The mass-loaded and nonlinear vibrating string problem revisited. *European Journal of Physics*, vol. 21, pp. L11–L14.
- [27] Lee, N., Smith, J. O., Abel, J., and Berners, D. (2009) Pitch glide analysis and synthesis from recorded tones. *Proceedings of the 12th International Conference on Digital Audio Effects*, pp. DAFX1–DAFX8.

- [28] Fletcher, N. H. (1977) Analysis of the design and performance of harpsichords. *Acustica*, **vol. 37**, pp. 139–147.
- [29] Stokes, G. G. (1851) On the effect of internal friction of fluids on the motion of pendulums. *Trans. Cambridge Phil. Soc.*, **vol. 9**.
- [30] Legge, K. A. and Fletcher, N. (1984) Nonlinear generation of missing modes on a vibrating string. *Journal of the Acoustical Society of America*, **vol. 76**, pp. 5–12.
- [31] Gough, C. E. (1984) The nonlinear free vibration of a damped elastic string. *Journal of the Acoustical Society of America*, **vol. 75**, pp. 1770–1776.
- [32] Thornton, S. T. and Marion, J. B. (2004) *Classical Dynamics of Particles and Systems*. Brooks/Cole - Thomson Learning, 5th edn., ISBN 0-534-40896-6.
- [33] Rowland, D. R. (2004) Parametric resonance and nonlinear string vibrations. *American Journal of Physics*, **vol. 72**, pp. 758–766.
- [34] Bank, B. (2006) *Physics-based Sound Synthesis of String Instruments Including Geometric Nonlinearities*. Ph.D. thesis, Budapest University of Technology and Economics, H-1117 Budapest.
- [35] Carrier, C. F. (1945) On the non-linear vibration problem of the elastic string. *Q. Appl. Math.*, **vol. 3**, pp. 157–165.
- [36] French, R. M. (2009) *Engineering the Guitar - Theory and Practice*. Springer, ISBN 978-0-387-74368-4.
- [37] Hanson, R. J. (1987) Optoelectronic detection of string vibration. *The Physics Teacher*, **vol. 25**, pp. 165–166.
- [38] Taylor, J. R. (1997) *An Introduction to Error Analysis*. University Science Books, 2nd edn., ISBN 0-935702-75-X.
- [39] Hunter, D. (2008) *The Guitar Pickups Handbook the start of your sound*. Backbeat Books, ISBN 978-0-87930-931-2.
- [40] Horton, N. G. and Moore, T. R. (2009) Modeling the magnetic pickup of an electric guitar. *American Journal of Physics*, **vol. 77**, pp. 144–150.
- [41] Lemarquand, G. and Lemarquand, V. (2007) Calculation method of permanent-magnet pickups for electric guitars. *IEEE Transactions on Magnetics*, **vol. 43**, pp. 3573–3578.
- [42] Anderton, C. (1995) *Do-It-Yourself Projects For Guitarists*. Miller Freeman Books, ISBN 0-87930-359-X.
- [43] Massarini, A. and Kazimierczuk, M. K. (1997) Self-capacitance of inductors. *IEEE Transactions on Power Electronics*, **vol. 12**, pp. 671–676.
- [44] Yu, Q. and Holmes, T. W. (2001) Stray capacitance modeling of inductors by using the finite element method. *IEEE Transactions on Electromagnetic Compatibility*, **vol. 43**, pp. 88–93.
- [45] Milan, M. (2007) *Pickups - Windings and Magnets... and the Guitar Became Electric*. Centerstream, ISBN 978-1-57424-209-6.
- [46] Lemme, H. E. W. (1986) The secrets of electric guitar pickups. *Electronic Musician*, **12**, pp. 66–72.
- [47] Jungmann, T. (1994) *Studies on the Behaviour of Electric Guitar Pickups*. M. eng. thesis, Helsinki University of Technology, Helsinki, Finland.
- [48] O'Toole, M. and Diamond, D. (2008) Absorbance based light emitting diode optical sensors and sensing devices. *Sensors*, **vol. 8**, pp. 2453–2479.
- [49] Schubert, E. F. (2006) *Light-Emitting Diodes*. Cambridge University Press, 2nd edn., ISBN 978-0-521-86538-8.
- [50] Cho, S.-K. and Park, Y.-W. (2011) Development of an optical transducer for an electro-acoustic guitar. *Proceedings of Meetings on Acoustics - 162nd Meeting*, **vol. 14**, pp. 1–6.
- [51] Leroy, N., Bevilacqua, F., and Fléty, E. (2006) Reflective optical pickup for violin. *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06)*, pp. 204–207.
- [52] Decoster, D. and Harari, J. (2009) *Optoelectronic Sensors*. ISTE ltd / Wiley, ISBN 978-1-84821-078-3.
- [53] Arnau, A. (2010) *Piezoelectric Transducers and Applications*. Springer, revised 2nd edn., ISBN 978-3-642-09624-2.
- [54] Morgan Technical Ceramics (2011) *Piezoelectric Ceramics – Properties and Applications*.
- [55] Beranek, L. L. (1988) *Acoustical Measurements*. Acoustical Society of America, revised 1st edn., ISBN 0-88318-583-0.
- [56] Harris, C. M. and Piersol, A. (2002) *Harris' Shock and Vibration Handbook*. McGraw-Hill, 5th edn., ISBN 978-0-07-137081-3.

- [57] Dyer, S. A. (2001) *Wiley Survey of Instrumentation and Measurement*. John Wiley and Sons, ISBN 978-0-471-39484-6.
- [58] Peña, F. (2005) Big muff circuit schematics. [online], p. <http://www.tonepad.com/>, Referenced on 19.06.2011.
- [59] James, E. J. (1949) Simple tone control circuit: Bass and treble, cut and lift. *Wireless World*, february, pp. 48–50.
- [60] Baxandall, P. J. (1952) Negative feedback tone control, independent variation of bass and treble without switches. *Wireless World*, october, pp. 402–405.
- [61] Wadell, B. C. (1991) *Transmission Line Design Handbook*. Artech House, ISBN 0-89006-436-9.
- [62] DiMarzio (2011) Instrument and studio cable sales page. [online], p. <http://www.dimarzio.com/>, Referenced on 31.05.2011.
- [63] Allen, J. C., Neitzel, E., and Theorin, C. (2009) *Cable for Stringed Musical Instruments*. Gore Enterprise Holdings, Inc, US Patent application no. 12/392,494.
- [64] Grant, I. S. and Phillips, W. R. (1990) *Electromagnetism*. John Wiley and Sons, 2nd edn., ISBN 0-471-92712-0.
- [65] Boscorelli, N. (1999) *The Stomp Box Cookbook*. Guitar Project Books, 2nd edn., ISBN 0-9663824-1-2.
- [66] Case, A. (2007) *Sound FX*. Focal Press / Elsevier, ISBN 978-0-240-52032-2.
- [67] Cowles, L. G. (1966) *Analysis and Design of Transistor Circuits*. D. Van Nostrand Company.
- [68] Keen, R. G. (1998) The technology of the fuzz face. [online], p. <http://www.geofex.com/>, Referenced on 31.05.2011.
- [69] Hunter, D. (2004) *Guitar Effects Pedals the practical handbook*. Backbeat Books, ISBN 978-0-87930-806-3.
- [70] Malik, N. R. (1990) Determining spice parameter values for bjt's. *IEEE Transactions on Education*, vol. 33, pp. 366–368.
- [71] Roberts, G. W. and Sedra, A. S. (1992) *Spice for Microelectronic Circuits*. Oxford University Press, 3rd edn., ISBN 0-19-510584-2.
- [72] Nagel, L. W. (1975) *SPICE2: A Computer Program to Simulate Semiconductor Circuits*. Ph.D. thesis, EECS Department, University of California, Berkeley.
- [73] Steer, M. B. (2007) *SPICE: User's Guide and Reference*. fREEDA, 1.3 edn.
- [74] Cirovic, M. M. (1971) *Semiconductors - Physics, Devices and Circuits*. Prentice Hall, ISBN 13-806307-9.
- [75] Joyce, M. V. and Clarke, K. K. (1961) *Transistor Circuit Analysis*. Addison-Wesley.
- [76] Peña, F. (2009) Ea electronics tremolo circuit schematics. [online], p. <http://www.tonepad.com/>, Referenced on 13.08.2011.
- [77] Sevin, L. J. (1965) *Field-Effect Transistors*. McGraw-Hill, Library of Congress Catalog Card Number 64-8624.
- [78] Shockley, W. (1952) A unipolar field-effect transistor. *Proceedings of the I.R.E.*, vol. 40, pp. 1365–1376.
- [79] Sherr, S. (1954) Generalized equations for rc phase-shift oscillators. *Proceedings of the IRE*, vol. 42, pp. 1169 – 1172.
- [80] Locke, W. M. (1963) A new analysis of the transistor phase-shift oscillator. *Radio and Electronic Engineer*, vol. 25, pp. 145 – 153.
- [81] Peña, F. (2004) Orange squeezer circuit schematics. [online], p. <http://www.tonepad.com/>, Referenced on 26.12.2011.
- [82] Leach, W. M. (2008) An fet audio peak limiter. [online], p. <http://users.ece.gatech.edu/mleach/papers/limiter.pdf>, Referenced 04.01.2012.
- [83] Plunkett, B. J. and Kushner, L. L. (1970) *foot-controlled continuously variable preference circuit for musical instruments*. Warwick Electronics, Inc, US Patent no. 3,530,224.
- [84] Peña, F. (2005) Green ringer circuit schematics. [online], p. <http://www.tonepad.com/>, Referenced on 19.06.2011.
- [85] Yeh, D. T., Abel, J. S., and Smith, J. O. (2007) Simplified, physically-informed models of distortion and overdrive guitar effects pedals. *Proceedings of the 10th International Conference on Digital Audio Effects*, pp. DAFX1–DAFX8.

- [86] Yeh, D. T., Abel, J. S., and Smith, J. O. (2007) Simulation of the diode limiter guitar distortion circuits by numerical solution of ordinary differential equations. *Proceedings of the 10th International Conference on Digital Audio Effects*, pp. DAFX1–DAFX7.
- [87] Keen, R. G. (2002) Jfet matching for effects. [online], p. http://www.geofex.com/article_folders/fetmatch/fetmatch.htm, Referenced on 31.08.2011.
- [88] Fletcher, N. H., Tarnopolskaya, T., and de Hoog, F. R. (2001) Wave propagation on helices and hyperhelices: a fractal regression. *Proceedings of the Royal Society*, **vol. 457**, pp. 33–43.
- [89] Roark, R. J. (1954) *Formulas for Stress and Strain*. McGraw-Hill, 3rd edn.
- [90] Wittrick, W. H. (1966) On elastic wave propagation in helical springs. *International Journal of Mechanical Sciences*, **vol. 8**, pp. 25–47.
- [91] Pietra, L. D. and Valle, S. D. (1982) On the dynamic behaviour of axially excited helical springs. *Meccanica*, **vol. 17**, pp. 31–43.
- [92] Abel, J. S., Berners, D. P., Costello, S., and Smith, J. O. (2006) Spring reverb emulation using dispersive allpass filters in a waveguide structure. *Audio Engineering Society Convention Paper*, pp. 1–7.
- [93] Parker, J. and Bilbao, S. (2009) Spring reverberation: A physical perspective. *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*, pp. DAFX 1–6.
- [94] Bilbao, S. and Parker, J. (2010) Perceptual and numerical aspects of spring reverberation modeling. *Proc. of the 20th Int. Symposium on Music Acoustics (ISMA-2010)*, pp. 1–7.
- [95] Young, A. C. (1963) *Artificial Reverberation Unit*. Hammond Organ Company, US Patent no. 3,106,610.
- [96] Anderton, G. (1976) Stage center reverb unit. *Guitar Player Magazine*, September issue.
- [97] Megantz, R. (2009) *Design and Construction of Tube Guitar Amplifiers*. TacTec Press, ISBN 978-0-651-29180-6.
- [98] Espley, D. C. (1933) The calculation of harmonic production in thermionic valves with resistive loads. *Proceedings of the I.R.E.*, **vol. 21**, pp. 1439–1446.
- [99] Kuehn, R. (2009) *Circuit Analysis of a Legendary Tube Amplifier The Fender Bassman 5F6-A*. Pentode Press, 3rd edn., ISBN 978-0-9769-8225-8.
- [100] Self, D. (2006) *Audio Power Amplifier Design Handbook*. Elsevier, 4th edn., ISBN 978-0-7506-8072-1.
- [101] Gray, T. S. (ed.) (1954) *A First Course in Electronics, Electron Tubes and Associated circuitry*. John Wiley and Sons, 2nd edn., Massachusetts Institute of Technology Dept. of Electrical Engineering.
- [102] Elliott, R. (2008) 100w guitar amplifier mk ii. [online], p. <http://sound.westhost.com/project27.htm>, Referenced on 01.04.2012.
- [103] Newell, P. and Holland, K. (2007) *Loudspeakers For Music Recording and Reproduction*. Focal Press / Elsevier, ISBN 0-2405-2014-9.
- [104] Meriläinen, E. (2010) *Current-Driving of Loudspeakers*. Print-On-Demand, ISBN 978-1-450-54400-9.
- [105] Dailey, D. J. (2011) *Electronics for Guitarists*. Springer, ISBN 978-1-4419-9535-3.
- [106] Leach, W. M. (1995) Spice models for vacuum tube amplifiers. *Journal of the Audio Engineering Society*, **vol. 43**, pp. 117–126.
- [107] Small, R. H. (1971) Direct-radiator loudspeaker system analysis. *IEEE Transactions on Audio and Electroacoustics*, **vol. AU-19**, pp. 269–281.
- [108] Leach, W. M. (2002) Loudspeaker voice-coil inductance losses: Circuit models, parameter estimation, and effect on frequency response. *Journal of the Audio Engineering Society*, **vol. 50**, pp. 442–449.
- [109] Bleuler, E. and Haxby, R. (eds.) (1964) *Methods in Experimental Physics - Volume 2: Electronic Methods*. Academic Press, ISBN 978-0-12-475902-2.
- [110] Alasaarela, E. (1980) *Elektroniikan suodattimet*. KS-teletietokirjat, A Finnish translation of Principles Of Active Network Design by G. Daryani.
- [111] Motz, J. W. (1959) *Frequency Measurement Instrument*. US Patent no. 2,901,699.
- [112] Allen, F. G. (1975) *Electronic Guitar Tuner*. F. G. Allen Associates, Inc, US Patent no. 3,881,389.
- [113] Conn, C. G. (1982) *Owner's Manual and Facts About Strobotuner Model ST-11*.

NOMENCLATURE

<i>Notation</i>	<i>Quantity</i>	<i>Unit name</i>	<i>Special unit</i>	<i>Basic form</i>
a	acceleration	-	-	$\frac{1}{\text{s}^2}$
α	angular acceleration	-	-	$\frac{1}{\text{s}^2}$
A, S	area	-	-	m^2
A_n	amplitude component	-	-	1
β	mechanical resistance	-	-	$\frac{\text{kg}}{\text{s}}$
β_F, h_{FE}	DC current gain	-	-	1
\vec{B}	magnetic field	tesla	T	$\frac{\text{kg}}{\text{A} \cdot \text{s}^2}$
C	capacitance	farad	F	$\frac{\text{A}^2 \cdot \text{s}^4}{\text{kg} \cdot \text{m}^2}$
c	velocity	-	-	$\frac{\text{m}}{\text{s}}$
\vec{D}	electric displacement	-	-	$\frac{\text{A} \cdot \text{s}}{\text{m}^2}$

Continued on the next page...

Notation	Quantity	Unit name	Special unit	Basic form
E	Young's modulus	pascal	Pa	$\frac{\text{kg}}{\text{m} \cdot \text{s}^2}$
\vec{E}	electric field	-	-	$\frac{\text{kg} \cdot \text{m}}{\text{A} \cdot \text{s}^3}$
E	total energy	joule	J	$\frac{\text{kg} \cdot \text{m}^2}{\text{s}^2}$
ε	electromotive force	volt	V	$\frac{\text{kg} \cdot \text{m}^2}{\text{A} \cdot \text{s}^3}$
ϵ	permittivity	-	-	$\frac{\text{A}^2 \cdot \text{s}^4}{\text{kg} \cdot \text{m}^3}$
\vec{F}	force	newton	N	$\frac{\text{kg} \cdot \text{m}}{\text{s}^2}$
f	frequency	hertz	Hz	$\frac{1}{\text{s}}$
G	conductivity	siemens	Ω^{-1}	$\frac{\text{A}^2 \cdot \text{s}^3}{\text{kg} \cdot \text{m}^2}$
g_m	transconductance	siemens	Ω^{-1}	$\frac{\text{A}^2 \cdot \text{s}^3}{\text{kg} \cdot \text{m}^2}$
G	shear modulus	pascal	Pa	$\frac{\text{kg}}{\text{m} \cdot \text{s}^2}$
h_{fe}	AC current gain	-	-	1
I, i	current	ampere	-	A
I	intensity	watt	W	$\frac{\text{kg} \cdot \text{m}^2}{\text{s}^3}$
I	moment of inertia	-	-	$\text{kg} \cdot \text{m}^2$

Continued on the next page...

<i>Notation</i>	<i>Quantity</i>	<i>Unit name</i>	<i>Special unit</i>	<i>Basic form</i>
I_p	polar moment of inertia	-	-	m^5
I_d	diametral moment of inertia	-	-	m^5
\vec{J}	current density	-	-	$\frac{\text{A}}{\text{m}^2}$
J	torsional rigidity	-	-	$\frac{\text{kg} \cdot \text{m}^3}{\text{s}^2}$
k	spring constant	-	-	$\frac{\text{kg}}{\text{s}^2}$
k	wave number	-	-	$\frac{1}{\text{m}}$
κ	radius of gyration	metre	-	m
L, l	length	metre	-	m
L	inductance	henry	H	$\frac{\text{kg} \cdot \text{m}^2}{\text{A}^2 \cdot \text{s}^2}$
M	torque	-	-	$\frac{\text{kg} \cdot \text{m}^2}{\text{s}^2}$
m	mass	kilogram	-	kg
N	sample length	-	-	1
N	electromechanical factor	-	-	$\frac{\text{m}}{\text{A} \cdot \text{s}}$
n_x	index of refraction	-	-	1

Continued on the next page...

<i>Notation</i>	<i>Quantity</i>	<i>Unit name</i>	<i>Special unit</i>	<i>Basic form</i>
p	pressure	pascal	Pa	$\frac{\text{kg}}{\text{m} \cdot \text{s}^2}$
P	power	watt	W	$\frac{\text{kg} \cdot \text{m}^2}{\text{s}^3}$
Φ	magnetic flux	weber	Wb	$\frac{\text{kg} \cdot \text{m}^2}{\text{A} \cdot \text{s}^2}$
Q	quality factor	-	-	1
Q, q	electric charge	coulomb	C	$\text{A} \cdot \text{s}$
ρ	resistivity	-	Ωm	$\frac{\text{kg} \cdot \text{m}^3}{\text{A}^2 \cdot \text{s}^3}$
ρ	density	-	-	$\frac{\text{kg}}{\text{m}^3}$
r	radius	metre	-	m
R, r	resistance	ohm	Ω	$\frac{\text{kg} \cdot \text{m}^2}{\text{A}^2 \cdot \text{s}^3}$
T	tension	newton	N	$\frac{\text{kg} \cdot \text{m}}{\text{s}^2}$
T	temperature	kelvin	-	K
T	period	second	-	s
t	time	second	-	s
τ	time constant	second	-	s

Continued on the next page...

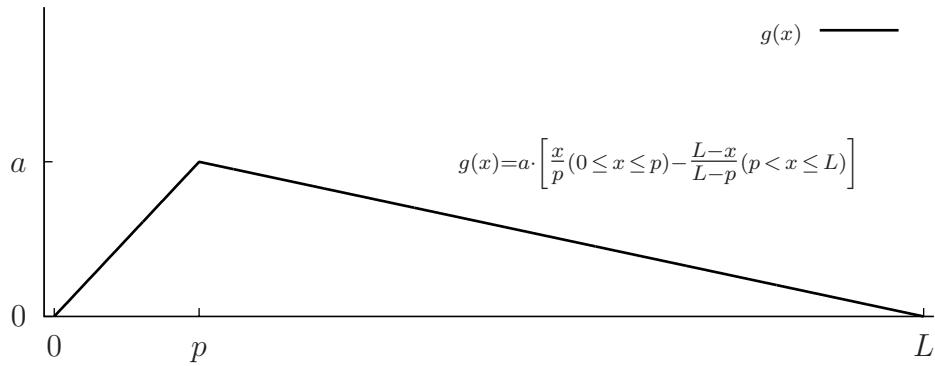
<i>Notation</i>	<i>Quantity</i>	<i>Unit name</i>	<i>Special unit</i>	<i>Basic form</i>
u	volume velocity	-	-	$\frac{\text{m}^3}{\text{s}}$
μ	amplification factor	-	-	1
μ	permeability	-	-	$\frac{\text{kg} \cdot \text{m}}{\text{A}^2 \cdot \text{s}^2}$
U	potential energy	joule	J	$\frac{\text{kg} \cdot \text{m}^2}{\text{s}^2}$
V, v	voltage	volt	V	$\frac{\text{kg} \cdot \text{m}^2}{\text{A} \cdot \text{s}^3}$
ω	angular velocity	-	-	$\frac{\text{m}}{\text{s}}$
X_n	Fourier component	-	-	1
x_n	data value	-	-	1
X	reactance	ohm	Ω	$\frac{\text{kg} \cdot \text{m}^2}{\text{A}^2 \cdot \text{s}^3}$
Y	admittance	siemens	Ω^{-1}	$\frac{\text{A}^2 \cdot \text{s}^3}{\text{kg} \cdot \text{m}^2}$
Z	impedance	ohm	Ω	$\frac{\text{kg} \cdot \text{m}^2}{\text{A}^2 \cdot \text{s}^3}$

ABBREVIATIONS

e.m.f.	= electromotive force
KVL	= Kirchhoff's Voltage Law
KCL	= Kirchhoff's Current Law
DFT	= Discrete Fourier Transform
IDFT	= Inverse Fourier Transform
BJT	= Bipolar Junction Transistor
JFET	= Junction Field-Effect Transistor
MOSFET	= Metal Oxide Semiconductor Field-Effect Transistor
AC	= Alternating Current
DC	= Direct Current
LED	= Light Emitting Diode
LDR	= Light Dependent Resistor
VSWR	= Voltage Standing Wave Ratio
DIY	= Do It Yourself
SPICE	= Simulation Program with Integrated Circuit Emphasis
op-amp	= Operational Amplifier
IC	= Integrated Circuit

APPENDIX A – MATHEMATICAL DERIVATIONS

A step-by-step derivation of the magical expression for the theoretical amplitude spectrum components of a vibrating string is presented here. The meaning of the variables used in the derivation are indicated in the figure showing the deflected string having a triangle form.



According to the theories of the Fourier series, the amplitude spectrum bin magnitudes can be evaluated as

$$A_n = \frac{2}{L} \int_0^L g(x) \sin\left(\frac{n\pi x}{L}\right) dx,$$

where the function $g(x)$ is formed as a piecewise function according to the geometric form of the deflected string

$$g(x) = a \left[\frac{x}{p} (0 \leq x \leq p) + \frac{L-x}{L-p} (p < x \leq L) \right].$$

Based on this theoretical jewel, the expression for A_n can be solved in symbolic form by using a common integration method of integration by parts. All the intermediate steps leading to the final result are written down in the following listing:

$$\begin{aligned}
A_n &= \frac{2}{L} \int_0^p \frac{ax}{p} \sin\left(\frac{n\pi x}{L}\right) dx + \frac{2}{L} \int_p^L \frac{a(L-x)}{L-p} \sin\left(\frac{n\pi x}{L}\right) dx \\
&= \frac{2}{L} \int_0^p \frac{ax}{p} \frac{L}{n\pi} \sin\left(\frac{n\pi x}{L}\right) \frac{n\pi}{L} dx + \frac{2}{L} \int_p^L \frac{a(L-x)}{L-p} \frac{L}{n\pi} \sin\left(\frac{n\pi x}{L}\right) \frac{n\pi}{L} dx \\
&= \frac{2}{L} \int_0^p -\frac{ax}{p} \frac{L}{n\pi} \cos\left(\frac{n\pi x}{L}\right) - \frac{2}{L} \int_0^p -\frac{a}{p} \frac{L}{n\pi} \cos\left(\frac{n\pi x}{L}\right) dx + \\
&\quad \frac{2}{L} \int_p^L -\frac{aL-ax}{L-p} \frac{L}{n\pi} \cos\left(\frac{n\pi x}{L}\right) - \frac{2}{L} \int_p^L \frac{a}{L-p} \frac{L}{n\pi} \cos\left(\frac{n\pi x}{L}\right) dx \\
&= -\frac{2}{L} \frac{aL}{n\pi} \cos\left(\frac{n\pi p}{L}\right) + 0 + \frac{2}{L} \int_0^p \frac{a}{p} \frac{L^2}{n^2\pi^2} \cos\left(\frac{n\pi x}{L}\right) \frac{n\pi}{L} dx - 0 + \\
&\quad \frac{2}{L} \frac{a(L-p)}{L-p} \frac{L}{n\pi} \cos\left(\frac{n\pi p}{L}\right) - \frac{2}{L} \int_p^L \frac{a}{L-p} \frac{L^2}{n^2\pi^2} \cos\left(\frac{n\pi x}{L}\right) \frac{n\pi}{L} dx \\
&= -\frac{2}{L} \frac{aL}{n\pi} \cos\left(\frac{n\pi p}{L}\right) + \frac{2}{L} \int_0^p \frac{a}{p} \frac{L^2}{n^2\pi^2} \sin\left(\frac{n\pi x}{L}\right) + \\
&\quad \frac{2}{L} \frac{aL}{n\pi} \cos\left(\frac{n\pi p}{L}\right) - \frac{2}{L} \int_p^L \frac{a}{L-p} \frac{L^2}{n^2\pi^2} \sin\left(\frac{n\pi x}{L}\right) \\
&= \frac{2}{L} \frac{a}{p} \frac{L^2}{n^2\pi^2} \sin\left(\frac{n\pi p}{L}\right) - 0 - 0 + \frac{2}{L} \frac{a}{L-p} \frac{L^2}{n^2\pi^2} \sin\left(\frac{n\pi p}{L}\right) \\
&= \boxed{\frac{2a}{n^2\pi^2} \left(\frac{L}{p} + \frac{L}{L-p} \right) \sin\left(\frac{n\pi p}{L}\right)}
\end{aligned}$$

APPENDIX B – SPICE NETLISTS

A COMPLETE SPICE NETLIST FOR THE PHASE 45 CIRCUIT

The circuit diagram of the Phase 45 effect pedal is shown in Figure 6.83

```
* gnetlist -g spice-sdb -o phaser_net.net phaser45.sch
*****
* Spice file generated by gnetlist *
* spice-sdb version 4.28.2007 by SDB --
* provides advanced spice netlisting capability.
* Documentation at http://www.brorson.com/gEDA/SPICE/
*****
.include /usr/share/gEDA/models/spice/tex_inst.lib

.control
tran 10us 990ms 810ms
set filetype=ascii
write phasdata.txt tran1.v(3) tran1.v(20)
gnuplot phas_tran tran1.v(3) tran1.v(20)
.endc

.OP

===== Begin SPICE netlist of main design =====
R20 21 23 3900k
XOP1 5 7 1 0 8 TL071/301/TI
XOP2 9 12 1 0 13 TL071/301/TI
XOP3 17 16 1 0 18 TL071/301/TI
XOP4 24 26 1 0 27 TL071/301/TI
V2 0 3 DC 0 AC 1 SIN(0 0.1 100)
V1 1 0 DC 9V
R27 0 20 150k
C11 19 20 0.05uF
R25 23 25 1k
R24 24 27 150k
D1 0 6 DI01
.MODEL DI01 D (Is = 1.0f Bv = 4.7)
R17 19 8 10k
R23 26 23 150k
R26 27 25 7.5k
R22 0 24 150k
R21 24 1 150k
R3 5 6 470k
R4 6 1 10k
R182 0 22 130k
R19 22 21 1000k
R181 22 6 120k
R7 9 10 10k
```

```

R8 10 6 10k
R9 11 21 470k
R14 15 21 470k
R13 14 6 10k
R12 17 14 10k
R16 19 18 10k
R15 16 18 10k
R11 13 16 10k
R10 12 13 10k
R6 8 12 10k
R5 7 8 10k
R2 6 7 20k
R1 5 4 10k
J2 6 15 17 2N5457
J1 6 11 9 2N5457
.MODEL 2N5457 NJF (VTO=-1.8 BETA=0.00135 LAMBDA=0.001 RD=35 RS=31.5
+ CGS=2.25E-12 CGD=6E-12 KF=6.5E-17 AF=0.5)
C3 6 0 10uF
C2 1 0 10uF
C9 23 0 10uF
C6 6 21 0.05uF
C10 26 27 0.01uF
C8 14 15 0.01uF
C5 10 11 0.01uF
C7 13 17 0.05uF
C4 8 9 0.05uF
C1 3 4 0.01uF
.end

```

A SPICE NETLIST FOR THE BAXANDALL TONE CONTROL CIRCUIT

The circuit diagram of the Baxandall tone control system is presented in Figure 4.18

```

* gnetlist -g spice-sdb -o baxandal.net baxandal.sch
*****
* Spice file generated by gnetlist *
* spice-sdb version 4.28.2007 by SDB -- *
* provides advanced spice netlisting capability. *
* Documentation at http://www.brorson.com/gEDA/SPICE/ *
*****
.control
ac dec 90 10 100K
alter RT1 100
alter RT2 500k
ac dec 90 10 100K
alter RT1 500k
alter RT2 1k
ac dec 90 10 100K
alter RT1 250k
alter RT2 250k
alter RB1 100
alter RB2 500k
ac dec 90 10 100K

```

```

alter RB1 500k
alter RB2 1k
ac dec 90 10 100K
gnuplot baxall db(ac1.v(4)) db(ac2.v(4)) db(ac3.v(4)) db(ac4.v(4)) db(ac5.v(4))
.endc

***** Begin SPICE netlist of main design *****
C4 0 7 3300pF
C3 6 1 330pF
R2 0 5 10k
C1 3 2 470pF
C2 3 5 4700pF
RB2 5 3 250k
RB1 3 2 250k
RT2 7 4 250k
RT1 4 6 250k
R3 4 3 180k
VP 0 1 DC 0 AC 1
R1 2 1 100k
.end

```

A SPICE NETLIST FOR THE MOONLIGHT TONE CONTROL CIRCUIT

The circuit schematic of the Moonlight single-pot tone control system is presented in Figure 4.16.

```

* gnetlist -g spice-sdb -o moonlight.net moonlight.sch
*****
* Spice file generated by gnetlist *
* spice-sdb version 4.28.2007 by SDB -- *
* provides advanced spice netlisting capability. *
* Documentation at http://www.brorson.com/gEDA/SPICE/ *
*****

.control
ac dec 90 10 100K
alter RT1 100
alter RT2 500k
ac dec 90 10 100K
alter RT1 500k
alter RT2 1k
ac dec 90 10 100K
gnuplot moon db(ac1.v(4)) db(ac2.v(4)) db(ac3.v(4))
.endc

***** Begin SPICE netlist of main design *****
R2 4 1 470k
RT2 0 3 250k
RT1 3 2 250k
RV2 0 5 250k
RV1 5 4 250k
R3 0 4 270k
VP 0 1 DC 0 AC 1

```

```
R1 2 1 220k
C1 4 3 560pF
.end
```

A SPICE NETLIST FOR THE FENDER PRO 6G5 TONE CONTROL CIRCUIT

The circuit schematic of the Fender-style tone control system is presented in Figure 4.20.

```
* gnetlist -g spice-sdb -o fenton.net fenton.sch
*****
* Spice file generated by gnetlist *
* spice-sdb version 4.28.2007 by SDB --
* provides advanced spice netlisting capability.
* Documentation at http://www.brorson.com/gEDA/SPICE/
*****
.control
ac dec 90 10 100K
alter RT1 100
alter RT2 250k
ac dec 90 10 100K
alter RT1 250k
alter RT2 1k
ac dec 90 10 100K
alter RT1 125k
alter RT2 125k
alter RB1 100
alter RB2 250k
ac dec 90 10 100K
alter RB1 250k
alter RB2 1k
ac dec 90 10 100K
gnuplot fenton db(ac1.v(4)) db(ac2.v(4)) db(ac3.v(4)) db(ac4.v(4)) db(ac5.v(4))
.endc

=====
 Begin SPICE netlist of main design =====
R2 0 3 10k
C2 3 2 0.01uF
C1 5 1 250pF
RB2 0 3 125k
RB1 3 2 125k
RT2 2 4 125k
RT1 4 5 125k
VP 0 1 DC 0 AC 1
R1 2 1 100k
.end
```

APPENDIX C – OCTAVE SCRIPTS

THE MOTION OF AN IDEAL STRING

An Octave script to visualise the motion of the ideal string as a sum of two identical triangles moving in opposite directions.

```
clear all
clc

% initialise variables
n = 1:100;
% length of the string chosen as 100 cm
L = 100;
% plucking position defined as a fraction of the string's length
p = L/9;
% the string is drawn from 0 to L in the x direction
x = 0:0.1:L;
% plucking displacement is chosen as 6mm
a = 0.6;
% drawing the initial configuration at t = 0
t = 0*L;

% calculate the initial shape of the string as
% two identical triangles K1 and K2 and their sum K3
FACTOR = 2*a*(L/p + L/(L-p))/pi^2;
Z1 = sin(n'*pi*(x-t)/L)';
Z2 = sin(n'*pi*(x+t)/L)';
Z = (sin(n*pi*p/L)./n.^2)';
K1 = 0.5*FACTOR*Z1*Z;
K2 = 0.5*FACTOR*Z2*Z;
K3 = K1 + K2;

% setup a figure to the screen, the plot in it will be updated dynamically
figur = plot(x,K3,'LineWidth',5, x,K2,'LineWidth',5, x,K1,'LineWidth',5);
set(gca,'ylim',[-a*1.1 a*1.1]);
set(gca,'ytick',[-a 0 a]);
set(gca,'yticklabel',{'-a','0','a'});
set(gca,'xlim',[0,L]);
set(gca,'xtick',[0 p L]);
set(gca,'xticklabel',{'0','p','L'});
set(gca,'fontsize',14);
legend('a*[f(x-ct) + f(x+ct)]','f(x+ct)','f(x-ct)');

% loop for few full periods of vibration, advance time
for s = 1:400
    t = (s/100)*L;

    % calculate the shape of the string again as the time has changed
```

```

FACTOR = 2*a*(L/p + L/(L-p))/pi^2;
Z1 = sin(n'*pi*(x-t)/L)';
Z2 = sin(n'*pi*(x+t)/L)';
Z = (sin(n*pi*p/L)./n.^2)';
K1 = 0.5*FACTOR*Z1*Z;
K2 = 0.5*FACTOR*Z2*Z;
K3 = K1 + K2;

% update data in the figure window
set(figur(1), 'ydata', K3);
set(figur(2), 'ydata', K2);
set(figur(3), 'ydata', K1);

% draw the updated data
drawnow;
end;

```

THEORETICAL ERROR LIMITS

An Octave script to calculate theoretical error limits for the amplitude spectrum of a plucked string to be compared with measurement results.

```

clear all
clc

% the length of the string is actually 600 mm
% but the scale is expanded here
L = 6000;
% plucking location as a fraction of the length
P = L/12;
% uncertainty in pluck location measurement (+- 1mm)
dP = 10;
% transducer location from the bridge
X = 80;
% uncertainty in transducer location measurement (+- 1mm)
dX = 10;

% arrays to hold maximum, minimum and exact (R) values
MAX = [];
MIN = [];
R = [];

% calculate for 50 upper partials
for n = 1:50

PV = P-dP:P+dP;
XV = X-dX:X+dX;

% the size of this matrix is PV x XV
Y = sin(n*pi.*PV./L).*sin(n*pi.*XV/L)/( (n^2)*sin(pi*p/L)*sin(n*pi*X/L) );

% locate the max and min values from Y (DIM 2)
MAX = [MAX max( max(abs(Y)) )];
MIN = [MIN min( min(abs(Y)) )];
R = [R sin(n*pi*p/L)*sin(n*pi*X/L)/( (n^2)*sin(pi*p/L)*sin(n*pi*X/L) ) ];


```

```

end;

% print out the normalised values to the screen
MAX'/max(abs(R))
MIN'/max(abs(R))
abs(R')/max(abs(R))

```

SIMULATING THE INDUCED E.M.F. IN A MAGNETIC PICKUP

An Octave script to visualise the magnetic field and the induced e.m.f. at the pickup with respect to a specific displacement pattern of an ideal string. Theoretical e.m.f. from the form of the string is also calculated and visualised.

```

clear all
clc
% 100 upper partials are used for constructing the form of the string
n=1:100;
% the lenght of the string... the units are meters :)
L=100;
% plucking position measured as a fraction of the string length
p=L/2;
% pickup position with respect to the length
x=L/4;
% amplitude of the initial pluck
a = 4;
% time defined relative to the vibration periods (full period = 2L)
t = 0:1:4*L;

% building the string from the components of the Fourier series.
% K will hold the values of the form of the string at time t.
FACTOR = 2*a*(L/p + L/(L-p))/pi^2;
Z1 = sin(n'*pi*(x-t)/L)';
Z2 = sin(n'*pi*(x+t)/L)';
Z = (sin(n*pi*p/L)./n.^2)';
K = 0.5*FACTOR*(Z1+Z2)*Z;

% differential of the string displacement with respect to time.
% in theory this describes the velocity a string sensed by a magnetic pickup
% and reflects waveform of the induced e.m.f.
dKdt = diff(K) ./ diff(t');

% calculating the magnetic field changes in the pickup
% with respect to the location of the string given by K
% string is vibrating 5 mm above the pickup magnet and
% the changes in the magnetic field are measured 2 mm above the string
% x0,y0,z0 define the origin
% x1,y1,z1 define the location of a string in rest above the magnet
% xp,yp,zp define the point of measuring the changes in
% the magnetic field component Bz
sigma = 1;
y1 = 0;
x1 = 0;
x0 = 0;
y0 = 0;
z0 = 0;

```

```

z1 = 5.0;
xp = 0;
yp = 0;
zp = 2.0;

B = [] ; % empty storage array for measured magnetic field values
T = [] ; % empty storage array for time values
S = [] ; % empty storage array for string deflection values

% a for loop progresses through all the string displacement values which
% initially calculated as a function of time so each displacement value
%represents the dislocation of the string as time advances
for q = 1:length(K)

% x-coordinate (and z-coordinate) values updated as
% a function of the string displacement
x1 = K(q);
%z1 = 5.0 - ( K(q)/4 );

% integrating the magnetic field at the string, the size of the
% pickup magnet is defined by the integration limit as 2*6.5 mm
Bs = dblquad(@(r,p) (sigma.*r)./( (x1 - (x0 - r.*cos(p))).^2 + ...
(y1 - (y0 - r.*sin(p))).^2 + (z1 - z0)^2 ) ,0,6.5,0,2*pi);

% the magnetic field observed at 2 mm above the pickup magnet
Bz = Bs * (z1-zp)/( ( (x1-xp)^2 + (y1-yp)^2 + (z1-zp)^2 )^(3/2) );

B = [B Bz]; % update a new value to array on each round
T = [T q];
S = [S K(q)];
end; % end for loop

% calculate the derivative of the magnetic field values
% to obtain the induced e.m.f.
dB = diff(B') ./ diff(T');
% store the simulated e.m.f. values to an array
EMF = [0 dB'];

% constructing four separate figures to visualise all the evaluated data.
% separate figures used to avoid scaling of different magnitudes to the same plot
figure (1)
plot(T,B)
set(gcf,'name','Magnetic Field B by simulation model','numbertitle','off')
set(gca,'xlim',[0,4*L]);
set(gca,'xtick',[0 4*L]);
set(gca,'xticklabel',{'0','4L'});
set(gca,'fontsize',14);
xlabel('time')

figure (2)
plot(T,EMF)
set(gcf,'name','Induced e.m.f. by simulation model','numbertitle','off')
set(gca,'xlim',[0,4*L]);
set(gca,'xtick',[0 4*L]);
set(gca,'xticklabel',{'0','4L'});
set(gca,'fontsize',14);
xlabel('time')

figure (3)

```

```

plot(T,[0 dKdt'])
set(gcf,'name','Induced e.m.f. by analytical expression','numbertitle','off')
set(gca,'xlim',[0,4*L]);
set(gca,'xtick',[0 4*L]);
set(gca,'xticklabel',{'0','4L'});
set(gca,'fontsize',14);
xlabel('time')

figure (4)
plot(T,S)
set(gcf,'name','String displacement as a function of time','numbertitle','off')
set(gca,'xlim',[0,4*L]);
set(gca,'xtick',[0 4*L]);
set(gca,'xticklabel',{'0','4L'});
set(gca,'fontsize',14);
xlabel('time')

```

SOLVING CIRCUITS NUMERICALLY WITH OCTAVE

An example Octave script to evaluate an admittance matrix equation numerically using Cramer's rule. Run command: preamp(90,1,10000,1)

```

function z = preamp(deviation, startfreq, stopfreq, mode)

% define the step size for frequency vector
% adder is the step for arithmetic series (mode = 1)
% multiplr is the step for geometric series (mode = 2)
adder = (10-1)/deviation;
multiplr = 10^(1/deviation);

% determine the exponent k according to input variable stopfreq
k = 0;
while 10^k < stopfreq
k = k + 1;
end;

% buffer to store the frequency and the magnitude data for plotting
PLOTBUF = [];

% define the component values used in the matrix equation
RI    = 1000000;
RG    = 22000;
RK    = 3300;
RL    = 220000;
rp    = 62500;
RS    = 100;
CK = 4.7e-6;
CP = 0.47e-6;
u     = 100;

% buffers to hold magnitude (H) and frequency (X) data
H=[];
X=[];

%initial value for the start frequency

```

```

f = 1;

% loop over the given frequency range
for j = 0:k
while f <= (10^(j+1) - adder)

if (f >= (startfreq - 0.00001) && f <= stopfreq)

w = f*2*pi;

% define the matrix elements
Y11 = 1/RS + 1/RI + 1/RG;
Y12 = Y21 = -1/RG;
Y22 = 1/RG;
Y32 = -u/rp;
Y33 = 1/RK + (u+1)/rp + i*w*CK;
Y34 = -1/rp;
Y42 = u/rp;
Y43 = -(u+1)/rp;
Y44 = 1/rp + 1/RL + i*w*CP;
Y45 = Y54 = -i*w*CP;
Y55 = i*w*CP;

NIMx = [Y11 Y12 0 0 0; ...
         Y21 Y22 0 0 0; ...
         0 Y32 Y33 Y34 0; ...
         0 Y42 Y43 Y44 Y45; ...
         0 0 0 Y54 Y55];;

OSx = [Y11 Y12 0 0 1/RS; ...
         Y21 Y22 0 0 0; ...
         0 Y32 Y33 Y34 0; ...
         0 Y42 Y43 Y44 0; ...
         0 0 0 Y54 0];;

% calculate the magnitude in decibels
H=[H 20*log10(abs(det(OSx)/det(NIMx)))];;
X = [X f];
endif;

% determine the next frequency according to the chosen mode
if (mode == 1)
f = f + adder*10^j;
else
f = f*multiplr;
endif;

end; % while
end; % for

% store data to be saved and/or plotted
PLOTBUF = [PLOTBUF X' H'];

% save the data as a text file for plotting later with Gnuplot
% ngspice generates similar data file,
% so comparison and plotting against that is simple.
save -ascii preamp_octa.data PLOTBUF

```

DESIGNING A BJT PHASE-SHIFT OSCILLATOR WITH OCTAVE

An example Octave script to help in Phase-Shift oscillator design.

```
% numerical (iteration) and analytical calculation of
% oscillating frequency of the phase shift oscillator
clear all; clc;
% CALCULATE RB IF RC, VCC AND BETA ARE KNOWN
VT = 0.025;
VCC = 9
VBE = 0.65
B1 = 200
RC = 15000
RB = RC*(B1+1)*(2*(VCC - VBE)/VCC - 1)
% SET COLLECTOR BIAS VOLTAGE TO VCC/2
IEQ = (VCC - VBE)/(RC + RB/(B1+1))
VC = VCC - IEQ*RC
% ONCE BIAS IS SET, DETERMINE rpi
ICQ = IEQ*B1/(B1+1);
gm = ICQ/VT;
rpi = B1/gm
rpi = B1*VT/ICQ
% SET FREQUENCY WITH COMPONENT VALUES
R1 = 15000
R2 = 33000
C1 = 2.2e-06
C2 = 5.5e-08
C3 = 7.7e-07
% DEFINE HELPER RESISTANCES
RxA = RC*(rpi + RB)/(rpi + RB + RC);
RxB = rpi*(RC + RB)/(rpi + RB + RC);
RxC = RB*(rpi + RC)/(rpi + RB + RC);
RRxD = rpi*RB*RC/(rpi + RB + RC);
% CALCULATE FREQUENCY f USING QUADRATIC FORMULA WITH PARAMS A,B and C
C = -(RxA*C1+RxB*C3);

B = ((RxC-RB)*R1*R2 - (R1+R2)*(R1*RxB+R2*RxA))*C1*C2*C3-...
      (C1*RxA+C3*RxB)*((R1+R2)*(R1+R2)*C2 + (R1*C1*R1+R2*C3*R2))*C2-...
      R2*(R2*RxA+RRxD)*C1*C3*C3 - R1*(R1*RxB+RRxD)*C1*C1*C3;

A = RB*R1*R2*C1*C2*C3*...
      ( (C1*C2+C1*C3+C2*C3)*(R1*R2*(1-RxC/RB) - (R1+R2)*RRxD/RB)+ ...
      C2*(C1*RxA+C3*RxB)*(R1+R2) + C1*C3*(R1*RxB+R2*RxA+RRxD) ) ;

f = sqrt( (-B+sqrt(B*B-4*A*C))/(8*pi*pi*A) )
w = 2*pi*f;
% HAND-CALCULATED DETERMINANTS TO BE COMPARED TO NUMERIC DETERMINANTS LATER
CRAMER_V1 = 1 - w*w*(C1*C2+C1*C3+C2*C3)*R1*R2 + ...
i*(R1*(C1+C2)*w + R2*(C2+C3)*w - RB*R1*R2*C1*C2*C3*w*w*w);

CRAMER_V2 = 1 + RB/rpi - ...
w*w*(R1*R2*(1+RB/rpi)*(C1*C2+C1*C3+C2*C3) + C1*C3*RB*R1 + C2*C3*RB*(R1+R2))+...
i*w*(R1*(1+RB/rpi)*(C1+C2) + R2*(1+RB/rpi)*(C2+C3) + C3*RB)-...
i*w*w*w*R1*R2*RB*C1*C2*C3;

DET_Y      = 1 - ...
w*w*(C1*C2*(R1*R2+(R1+R2)*RxA))-...
w*w*(C1*C3*(R1*R2+R1*RxB+R2*RxA+RRxD) + C2*C3*(R1*R2+(R1+R2)*RxB))+...
```

```

i*w*(C1*(R1+Rx A) + C2*(R1+R2) + C3*(R2+Rx B)) - ...
i*w*w*C1*C2*C3*((R1+R2)*RRxD + R1*R2*Rx C);

% PRINT RESULTS TO CONSOLE
B_NEEDED = -(RB*rpi/RRxD)*...
( 1 - w*w*(C1*C2*(R1*R2+(R1+R2)*Rx A) + C1*C3*(R1*R2+R1*Rx B+R2*Rx A+RRxD)+...
C2*C3*(R1*R2+(R1+R2)*Rx B)) ) / (1 - R1*R2*(C1*C2+C1*C3+C2*C3)*w*w)

GAIN      = B1*RRxD*CRAMER_V2/(rpi*RB*DET_Y)
GAIN_ABS = abs(GAIN)
NODE_V1   = -(RRxD/RB)*CRAMER_V1/DET_Y
NODE_V2   = -(RRxD/RB)*CRAMER_V2/DET_Y

% INCREASE FREQUENCY SLIGHTLY AND LOOP DOWN UNTIL IMAGINARY PART CHANGES SIGN
% EVALUATE DETERMINANTS ON EACH ITERATION
% THIS IS FOR VERIFYING CORRECTNESS OF MANUAL CALCULATIONS
f = f + 0.05; angled = -1;
while angled < 0
f = f - 0.00001;
w = 2*pi*f;

Y11 = 1/rpi + 1/RB + i*w*C3;
Y12 = Y21 = 1/RB;
Y14 = Y41 = i*w*C3;
Y22 = 1/RC + 1/RB + i*w*C1;
Y23 = Y32 = i*w*C1;
Y33 = 1/R1 + i*w*C1 + i*w*C2;
Y34 = Y43 = i*w*C2;
Y44 = 1/R2 + i*w*C2 + i*w*C3;
% V1/V1
CRAMER_V1=[ -Y12      0      -Y14 ; ...
            -Y32      Y33      -Y34 ; ...
            0      -Y43      Y44 ];
% V2/V1
CRAMER_V2=[  Y11      0      -Y14 ; ...
            0      Y33      -Y34 ; ...
            -Y41      -Y43      Y44 ];
% Common denominator
DET_Y=[  Y11      -Y12      0      -Y14 ; ...
            -Y21      Y22      -Y23      0      ; ...
            0      -Y32      Y33      -Y34 ; ...
            -Y41      0      -Y43      Y44 ];

angled = arg(det(CRAMER_V1)/det(DET_Y));
end; % while
% PRINT RESULTS TO CONSOLE
f
B_NEEDED = (det(DET_Y)/det(CRAMER_V1))*rpi
GAIN      = (B1/rpi)*det(CRAMER_V2)/det(DET_Y)
GAIN_ABS = abs(GAIN)
NODE_V1   = det(CRAMER_V1)/det(DET_Y)
NODE_V2   = -det(CRAMER_V2)/det(DET_Y)

```

APPENDIX D – PARTS LISTS FOR SELECTED CIRCUITS

ORANGE SQUEEZER COMPRESSOR PARTS LIST

Resistors:

$$\begin{array}{lll} R_1 = 4.7 \text{ M}\Omega & R_2 = 82 \text{ k}\Omega & R_3 = 390 \text{ k}\Omega \\ R_4 = 220 \text{ k}\Omega & R_5 = 10 \text{ k}\Omega & R_6 = 1.5 \text{ k}\Omega \\ R_7 = 100 \text{ k}\Omega & R_8 = 470 \text{ k}\Omega & R_9 = 470 \text{ k}\Omega \\ R_{10} = 470 \text{ k}\Omega & R_{11} = 2.4 \text{ k}\Omega & R_{12} = 10 \text{ k}\Omega \text{ (trim. pot.)} \end{array}$$

Capacitors:

$$\begin{array}{lll} C_1 = 0.047 \mu\text{F} & C_2 = 0.047 \mu\text{F} & C_3 = 0.0022 \mu\text{F} \\ C_4 = 4.7 \mu\text{F} & C_5 = 4.7 \mu\text{F} & C_6 = 4.7 \mu\text{F} \\ C_7 = 4.7 \mu\text{F} & & \end{array}$$

Diode: 1N100 or 1N34A

Operational amplifier: 4558

JFETs: 2N5457 or BF245A.

STAGE CENTER REVERB PARTS LIST

Resistors:

$$\begin{array}{lll} R_1 = 1 \text{ M}\Omega & R_2 = 22 \text{ k}\Omega & R_3 = 470 \text{ k}\Omega \\ R_4 = 2.2 \text{ M}\Omega & R_5 = 47 \text{ k}\Omega & R_6 = 50 \text{ k}\Omega \\ R_7 = 50 \text{ k}\Omega & R_8 = 10 \text{ k}\Omega & R_9 = 10 \text{ k}\Omega \\ R_{10} = 33 \text{ k}\Omega & & \end{array}$$

Capacitors:

$$\begin{array}{lll} C_1 = 0.02 \mu\text{F} & C_2 = 0.02 \mu\text{F} & C_3 = 0.22 \mu\text{F} \\ C_4 = 220 \text{ pF} & C_5 = 0.22 \mu\text{F} & C_6 = 0.22 \mu\text{F} \end{array}$$

Operational amplifier: no special requirements.

Reverberation unit: nominal input impedance of 150 Ω or higher.

ARIA AB-30 BASS AMPLIFIER PARTS LIST**Resistors**

$R_{100} = 12 \text{ k}\Omega$	$R_{101} = 12 \text{ k}\Omega$	$R_{102} = 220 \text{ k}\Omega$	$R_{103} = 47 \text{ k}\Omega$
$R_{105} = 1 \text{ k}\Omega$	$R_{106} = 3.9 \text{ k}\Omega$	$R_{110} = 470 \text{ k}\Omega$	$R_{111} = 2.2 \text{ k}\Omega$
$R_{112} = 47 \text{ k}\Omega$	$R_{113} = 4.7 \text{ k}\Omega$	$R_{120} = 47 \text{ k}\Omega$	$R_{121} = 4.7 \text{ k}\Omega$
$R_{122} = 18 \text{ k}\Omega$	$R_{123} = 2.2 \text{ k}\Omega$	$R_{124} = 47 \text{ k}\Omega$	$R_{125} = 270 \Omega$
$R_{126} = 2.7 \text{ k}\Omega$	$R_{127} = 3.3 \text{ k}\Omega$	$R_{128} = 220 \Omega$	$R_{129} = 220 \Omega$
$R_{130} = 0.5 \Omega$	$R_{131} = 0.5 \Omega$	$R_{201} = 33 \text{ k}\Omega$	$R_{202} = 47 \text{ k}\Omega$
$R_{203} = 100 \text{ k}\Omega$	$R_{204} = 47 \text{ k}\Omega$	$R_{205} = 1 \text{ M}\Omega$	$R_{206} = 220 \text{ k}\Omega$
$R_{207} = 12 \text{ k}\Omega$			

Capacitors

$C_{100} = 560 \text{ pF}$	$C_{101} = 560 \text{ pF}$	$C_{102} = 0.1 \mu\text{F}$	$C_{103} = 100 \text{ pF}$
$C_{104} = 10 \mu\text{F}$	$C_{105} = 2.2 \mu\text{F}$	$C_{108} = 0.001 \mu\text{F}$	$C_{109} = 0.001 \mu\text{F}$
$C_{202} = 0.1 \mu\text{F}$	$C_{203} = 0.22 \mu\text{F}$	$C_{113} = 2.2 \mu\text{F}$	$C_{114} = 0.001 \mu\text{F}$
$C_{115} = 2.2 \mu\text{F}$	$C_{118} = 4.7 \mu\text{F}$	$C_{120} = 22 \mu\text{F}$	$C_{121} = 150 \text{ pF}$
$C_{122} = 22 \text{ pF}$	$C_{123} = 100 \mu\text{F}$		

Potentiometers

volume: = 50 kΩ (log)	master: = 50 kΩ (lin)
treble: = 100 kΩ (log)	middle: = 100 kΩ (log)
bass: = 100 kΩ (log)	

Transistors

$Q_{102} = KTC3198$	$Q_{103} = KTC3198$	$Q_{104} = A1266$
$Q_{105} = C1627A$	$Q_{106} = A1274$	$Q_{107} = D1352$
$Q_{108} = B989$		

Operational amplifier: JRC NJM 4558

APPENDIX E – TABLES OF LAPLACE TRANSFORMS

$F(s) = \mathcal{L}[f(t)]$	$f(t)$
$\frac{1}{s}$	1
$\frac{1}{s^2}$	t
$\frac{1}{s^n}$	$\frac{t^{n-1}}{(n-1)!}$
$\frac{1}{s-a}$	e^{at}
$\frac{1}{(s-a)^2}$	te^{at}
$\frac{1}{(s-a)^n}$	$\frac{1}{(n-1)!} t^{n-1} e^{at}$
$\frac{\omega}{s^2 + \omega^2}$	$\sin \omega t$
$\frac{s}{s^2 + \omega^2}$	$\cos \omega t$
$\frac{1}{(s-a)^2 + \omega^2}$	$\frac{1}{\omega} e^{at} \sin \omega t$
$\frac{s-a}{(s-a)^2 + \omega^2}$	$e^{at} \cos \omega t$

