Final Thesis

$\Delta \Sigma$-modulation Applied to Switching RF Power Amplifiers

by

Tobias Andersson, Johan Wahlsten

LITH-ISY-EX--07/4010--SE

2007-02-01
ΔΣ-modulation Applied to Switching RF Power Amplifiers

Final Thesis

performed at Electronic Devices
in the Department of Electrical Engineering
at Linköping University

by

Tobias Andersson, Johan Wahlsten

LITH-ISY-EX--07/4010--SE

2007-02-01

Supervisor: Prof. Atila Alvandpour
Department of Electrical Engineering
Electronic Devices
at Linköping University

Examiner: Prof. Atila Alvandpour
Department of Electrical Engineering
Electronic Devices
at Linköping University
ΔΣ-modulation Applied to Switching RF Power Amplifiers

Tobias Andersson, Johan Wahlsten

Background
The task of this thesis is to investigate the possibility of using non-linear high efficiency switching power amplifiers with spectrally efficient varying envelope modulation schemes. And if possible further investigate such a solution on a high level.

This thesis focuses on the theory necessary to understand the technical issues related to power amplifiers and the procedures behind simulating and measuring the characteristics of different power amplifier setups.

Results
Using a ΔΣ-modulated input to a switching amplifier inherently degrades the performance, mainly because of poor coding efficiency and high switching activity. However, by merely using a switching amplifier as a mixer it is shown to be possible of amplifying a non-constant envelope, with available technology.

Conclusion
From this investigation we believe that the widely known technique: pulse width modulation (PWM), together with a tuned switching amplifier and some linearization technique, for example pre-distortion, is a better way to go. Much effort should be put in understanding the fundamental limits and possibilities of an efficient tuned switching power amplifier.

Nykterord
RF, Transmitter, Amplifier, Switching, Power, Delta, Sigma, Modulation, Efficiency
Abstract

Background
The task of this thesis is to investigate the possibility of using non-linear high efficiency switching power amplifiers with spectrally efficient varying envelope modulation schemes and, if possible, further investigate such a solution on a high level.

The thesis focuses on the theory necessary to understand the technical issues related to power amplifiers and the procedures behind simulating and measuring the characteristics of different power amplifier configuration. The thesis also covers basic theory behind ΔΣ-modulators. The theory is needed to draw conclusions about the feasibility of using a ΔΣ-modulator as input to a switching amplifier.

Results
Using a ΔΣ-modulated input to a switching amplifier inherently degrades the performance, mainly because of poor coding efficiency and high switching activity. However, by merely using a switching amplifier as a mixer it is shown to be possible to transmit a non-constant envelope signal, with digital logic. The resulting circuit is, however, not an amplifier and it should not be seen as the final result. As already mentioned: the result lies in the investigation of a using ΔΣ-modulator as input to a switching amplifier.

Conclusion
From this investigation we believe that the widely known technique: pulse width modulation (PWM), together with a tuned switching amplifier and some linearization technique, for example pre-distortion, is a better way to go. Much effort should be put in understanding the fundamental limits and possibilities of an efficient tuned switching power amplifier.
Preface

About the Thesis

This thesis was written as the final project of the “Master of Science in Computer Science and Engineering Programme/Master of Science in Applied Physics and Electronics Programme” at Linköping University in Sweden. The work presented in this thesis was performed at Department of Electrical Engineering - Electronic Devices. The intended audience is people who are studying or have studied basic electronics and communication systems.

Most mathematical formulas and equations are numbered according to their order in each chapter. References in the text are given in square brackets and can be found in the end of the thesis. Moreover the document was typeset in \LaTeX.

Tobias Andersson
Johan Wahlsten
Linköping, Spring 2007
## Contents

1 Introduction 1
  1.1 Purpose ........................................ 1
  1.2 Objectives ...................................... 1
  1.3 Readers Advice .................................. 2

2 Linearity and Units of Measurement 3
  2.1 Linearity ........................................ 3
    2.1.1 Definition ................................... 3
    2.1.2 To model a System ............................. 4
    2.1.3 Three signs of nonlinearities .................. 4
      Harmonics ...................................... 4
      Signal compression ............................... 5
      Intermodulation .................................. 6
    2.1.4 Measuring linearity ............................ 7
      1dB compression point ............................. 7
      Peak to Average Power Ratio ...................... 8
      Third-order intersect point - IP3 ................ 10
      The two tone test ................................ 10
  2.2 Efficiency ...................................... 12
    2.2.1 Drain Efficiency ............................. 12
    2.2.2 Power Added Efficiency - PAE ................ 12
    2.2.3 Coding Efficiency ......................... 12

3 Class A, B, C Amplifiers 14
  3.1 The School Textbook Class A Amplifier ............. 14
  3.2 Deriving a Class A RF Amplifier ................. 15
    3.2.1 A general design ............................ 15
    3.2.2 Power Demands ................................ 17
      Impedance match .................................. 20
  3.3 Class B Amplifier .............................. 20
    3.3.1 Push-Pull .................................... 20
    3.3.2 Class AB ..................................... 21
  3.4 Class C Amplifier ................................ 21
4 Switching PAs
  4.1 A Basic RF Switching Amplifier ................. 23
  4.2 Ideal Switch and Harmonic Short ................ 25
  4.3 Class D Switching Amplifier .................... 27
  4.4 Class E Switching Amplifier .................... 28
    4.4.1 Designability .......................... 29
    4.4.2 Tuning ............................... 29

5 ∆Σ-Modulation
  5.1 Introduction to ∆Σ-Modulators .................. 31
  5.2 Linear Model ................................ 32
    5.2.1 Quantization Noise and Error Model ......... 34
    Rectangular Distribution ....................... 35
    Quantization and Oversampling ................ 38
    Noise Shaping ................................ 38
    Conclusion .................................. 39
  5.2.2 Higher Order ∆Σ-modulators ................ 40
  5.2.3 Optimization of the NTF ..................... 41
  5.2.4 Error Feedback Structure .................... 42
  5.3 Bandpass ∆Σ-modulators ....................... 43

6 Survey
  6.1 Introduction ................................ 45
  6.2 Pulse Width Modulation (PWM) ................... 45
    6.2.1 RF Applications .......................... 45
  6.3 ∆Σ-modulation ................................ 47
    6.3.1 Low pass ∆Σ-modulation .................. 47
    RF PWM by using BP∆Σ-modulation [16] .......... 47
    LP∆Σ-modulation and Mixing [13] ............... 48
  6.3.2 Bandpass ∆Σ-modulation ..................... 49
    1998 [9] ...................................... 49
    2000 [8] ...................................... 51
    2001 [15] ...................................... 52
    2003 [18] ...................................... 54
    2004 [14] ...................................... 55
  6.4 Conclusion .................................. 57

7 Investigation
  7.1 Analogue Amplification ......................... 59
    7.1.1 Matching Network .......................... 61
    7.1.2 Results .................................. 62
    7.1.3 Conclusion ................................ 63
  7.2 High-Level Design of ∆Σ-modulators ............ 64
    7.2.1 Using MatLab to produce the NTF ........... 64
7.2.2 A 4th-order BPΔΣ-modulator using Simulink . . . . . . . 68
7.2.3 A 4th-order BPΔΣ-modulator using VerilogA . . . . . . . 69
    The loop ........................................... 69
    The filter ......................................... 70
7.3 Coding Efficiency ........................................ 72
    7.3.1 Switching Activity ................................ 72
    7.3.2 Coding Efficiency ................................ 73
    7.3.3 Conclusion .................................... 75
7.4 Class E Design Example in CMOS 90nm ....................... 76
    7.4.1 Stationarity and Quasi-Stationarity .................. 79
        Stationarity ....................................... 80
        Quasi-Stationarity ................................ 80
        Application to ∆Σ-modulation ....................... 81
7.5 Digital amplification ..................................... 82
    7.5.1 PWM ............................................ 82
    7.5.2 ΔΣ-modulation .................................. 83
        Suggested Architectures ............................ 84
    7.5.3 Signal Investigation .............................. 88
7.6 Switching PA Stage ....................................... 95
7.7 Comparison ........................................... 97
    7.7.1 Output Power ................................... 98
    7.7.2 One Tone Efficiency .............................. 99
    7.7.3 WCDMA Signal .................................. 100
8 Conclusions .............................................. 103
    8.1 Analogue or Digital Amplification ..................... 103
    8.2 Where to go from here ................................ 103
A 4th order BPΔΣ VerilogA-implementation ..................... 105
B Transistor Analysis; a Small Detour .......................... 108
    B.1 Transistor Modelling ................................ 108
    B.2 Small-Signal Modelling .............................. 110
    B.3 Nonlinear (I-V) i(v, v) ............................. 111
        B.3.1 Conclusion .................................... 116
C C1 and C2 Tuning Procedure ................................ 117
Bibliography .............................................. 119
D Terminology .............................................. 123
Chapter 1

Introduction

The task of this thesis is to investigate the possibility of using non-linear high efficiency switching power amplifiers with spectrally efficient varying envelope modulation schemes and, if possible, further investigate such a solution on a high level.

As modulation schemes become more complex, to allow for higher data rates, a demand for ever increasingly linear power amplifiers arise. Unfortunately increased linearity typically implies lower power efficiency. An example of a non-constant modulation scheme is QPSK, which is the modulation scheme used in 3G. QPSK is a non-constant envelope modulation scheme, which means that the amplifier has to be able to reproduce a modulated amplitude. However, the potentially highly efficient switching amplifiers do not preserve the amplitude. In order to use such a potentially highly efficient amplifier, the non-constant envelope (amplitude) must be represented in another way, better suited for switching power amplifiers.

1.1 Purpose

The purpose of the thesis is to introduce CMOS RF power amplifier research to the department’s agenda. The idea is to increase traditional RF power amplifier efficiency using “digital know-how”. One of the ultimate goals is to simplify RF transmitters by allowing single chip RF transceivers.

1.2 Objectives

Our objectives have been to describe basic RF power amplifier design and associated measurements and investigate the feasibility of using switching amplifiers for RF applications. If possible we should also implement a model of a “promising” architecture in Cadence.
1.3 Readers Advice

The first chapter is introductory and contains the purpose and objectives of the thesis as well as a reader’s advice depicting the contents of the chapters to come.

The second chapter is essential to readers not previously familiar with the term linearity in a signal context. The chapter defines what we mean by linearity and how we measure it, while pointing out some of the related pitfalls and most common mistakes.

The third chapter explains the basic principles of traditional RF amplifier design. A reader familiar with the audio amplifiers will find only a little new here while reader that has just begun to study the topic will find it very helpful.

The fourth chapter introduces switching amplifiers and provides a general explanation of what such an amplifier seeks to achieve. Both Class-D and Class-E amplifiers are treated. A reader that has not heard of switching amplifiers before should find this chapter enlightening.

The fifth chapter introduces a way of creating a switching signal, $\Delta\Sigma$-modulation, from which an analogue waveform can be retrieved. The $\Delta\Sigma$-modulator and some of its properties are explained.

The sixth chapter contains a survey of recently published papers on switching RF amplifiers, comments on the ideas are included.

The seventh chapter presents some of the analogue and digital designs considered during the thesis and simulation results for several different inputs. These results are also discussed and explained. This chapter also compares the results of the simulations and evaluates the techniques used.

The eighth and concluding chapter comments on the results of the previous chapter and relates it to the objectives and purpose of the thesis. Finally it suggests where to make future research efforts in the RF field.
2.1 Linearity

Let us start by looking at the definition of linearity in a mathematical sense, and then see what tools are made available to us for evaluating the mathematical criteria.

2.1.1 Definition

A function, operator, transform or system is linear if it has the following qualities [22]:

1. Homogeneity
2. Additivity

For a homogeneous system with the input \( x(t) \) and the output \( y(t) \) the following must be true for all \( x(t) \):
\[
a \cdot x(t) \rightarrow a \cdot y(t)
\]

For an additive system with the input \( x(t) \) and the output \( y(t) \) the following must be true:

If the input \( x_1(t) \) yields the output \( y_1(t) \) and the input \( x_2(t) \) yields the output \( y_2(t) \), then an input \( x_3(t) = x_1(t) + x_2(t) \) yields the output \( y_3(t) = y_1(t) + y_2(t) \) for all \( x_1(t) \) and \( x_2(t) \).

The two properties above can be formulated as a single criteria:

\[
x(t) = \sum_{i=0}^{\infty} a_i \cdot x_i \iff y(t) = \sum_{i=0}^{\infty} a_i \cdot y_i
\]  \hspace{1cm} (2.1)

It is important to understand that a linear system does not imply a “line-like” \((y(t) = k \cdot x(t) + m)\) relationship between the input and output. For
example $y(t) = \frac{dx(t)}{dt}$ and $y(t) = t \frac{dx(t)}{dt}$ are both linear (homogeneous and additive) but does not have a “line-like” relation between input and output.

Nota Bene: For an electrical system linearity implies that none of the resistances, inductances or capacitances changes as a function of voltage or current.

2.1.2 To model a System

The transfer function of a system can be difficult to derive, therefore it is useful to have a general model we can apply to any time variant system (at least piecewise). The infinite polynomial expression:

$$y(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) + \alpha_4 x^4(t) + ...$$  \hspace{1cm} (2.2)

provides one such a model for us (see calculus textbooks). We do not let the fact that this sum goes on forever bother us, as for most series we can use a simpler model with a finite number of terms and still be reasonably accurate.

Common schoolbook practice is to use a third order polynomial, thus modelling nonlinearities dependent on the second and third order terms. That this model is non-linear is easily shown: Let $x(t) = a_1 x_1(t) + a_2 x_2(t)$ in (2.2), as shown below:

$$y(t) = \alpha_1 (a_1 x_1(t) + a_2 x_2(t)) + \alpha_2 (a_1 x_1(t) + a_2 x_2(t))^2 + \alpha_3 (a_1 x_1(t) + a_2 x_2(t))^3$$  \hspace{1cm} (2.3)

By observing the squared term we find that the system is not additive hence not linear:

$$\alpha_2 (a_1 x_1(t) + a_2 x_2(t))^2 \neq \alpha_2 (a_1 x_1(t))^2 + \alpha_2 (a_2 x_2(t))^2$$  \hspace{1cm} (2.4)

However, we also note that should the coefficients $\alpha_2$ and $\alpha_3$ be zero, then the system is linear.

2.1.3 Three signs of nonlinearities

Harmonics

When a signal experiences overtones at multiples of the fundamental frequency it is referred to as harmonic distortion or just harmonics. These tones are connected to even terms in the model given in (2.2). For example modelling the input output behaviour of a system with a transfer function: $y(t) = \alpha_1 x(t) + \alpha_2 x(t)^2$ and inputting $x(t) = A \cos(\omega \cdot t)$ results in the output:
2.1 Linearity

\[ y(t) = \alpha_1 A \cos(\omega \cdot t) + \alpha_2 (A \cos(\omega \cdot t))^2 = \alpha_2 A^2 + \alpha_1 A \cos(\omega \cdot t) + \alpha_2 A^2 \cos(2\omega \cdot t). \] The last term is a cosine with twice the frequency of the input. This term is often referred to as the first overtone or second harmonic (the first one being the original signal).

The physical causes for such overtones can be parasitic impedances, the use components which are nonlinear by design or nature, or filters where resonance has occurred (a design issue).

In general harmonic distortion is distortion at \( n \) times the input frequency \( (n \cdot f) \), where \( n = 2, 3, 4, \ldots \). Harmonic distortion, when applied to the RF field, is usually a lesser problem in terms of linearity as the distortion occurs out of band. In some sense, harmonic distortion can always be filtered out.

**Signal compression**

When the output of a system no longer reaches the expected level in comparison to the level of the input, the system experiences signal compression (also called gain compression). This effect stems from the fact that all physical systems are limited in some way.

Figure (2.1) illustrates how the power out versus power in relationship in a PA differs from a linear behaviour for large input signals. Specifically there exists a point where the actual output power differs by 1dB from the linear output power predicted by the power gain at small input signals.

![Figure 2.1: Illustration of gain compression](image)

Gain compression is connected to the odd terms in the model given in equation (2.2). If the system is modelled as\[ y(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \]
with the input of a single continuous wave, e.g. \( x(t) = A \cos(\omega \cdot t) \) it is possible, by utilizing simple trigonometric rules, to find how the third order term affects the RF spectrum.
A cubic cosine can be expanded like this: \( \cos^3 \omega = \frac{3}{4} \cos(\omega) + \frac{1}{4} \cos(3\omega) \). Hence the output at the fundamental frequency will be: \( y_{\text{fundamental}}(t) = \alpha_1 x(t) + \frac{3}{4} \alpha_3 x(t) \). If \( \alpha_3 \) is negative, then compression of the output signal will occur.

**Intermodulation**

When a system receives a complex signal, which is usually the case, non-linearity causes another phenomenon called intermodulation. This is best demonstrated if we let the input to our system consist of two distinct sinusoids fairly close to each other in frequency. Like before we use the same third order model:

\[
y(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{2.5}
\]

The input is now given by:

\[
x(t) = A(\cos(\omega_1 t) + \cos(\omega_2 t)) \tag{2.6}
\]

Table (2.1) shows the terms yielded by evaluating equation (2.6) in equation (2.5). Inputting two tones to a system, often referred to as two tone test, is a very common and fairly simple way to measure intermodulation distortion. In the section “Measuring linearity” we investigate this test further.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>( \alpha_1 x(t) )</th>
<th>( \alpha_2 x(t)^2 )</th>
<th>( \alpha_3 x(t)^3 )</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( \omega_1 )</td>
<td>1</td>
<td>-</td>
<td>9/4</td>
<td>Gain compression</td>
</tr>
<tr>
<td>( \omega_2 )</td>
<td>1</td>
<td>-</td>
<td>9/4</td>
<td>Gain compression</td>
</tr>
<tr>
<td>2( \omega_1 )</td>
<td>-</td>
<td>1/2</td>
<td>-</td>
<td>Harmonic</td>
</tr>
<tr>
<td>2( \omega_2 )</td>
<td>-</td>
<td>1/2</td>
<td>-</td>
<td>Harmonic</td>
</tr>
<tr>
<td>( \omega_1 \pm \omega_2 )</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2( \omega_1 \pm \omega_2 )</td>
<td>-</td>
<td>3/4</td>
<td>-</td>
<td>Intermodulation</td>
</tr>
<tr>
<td>2( \omega_2 \pm \omega_1 )</td>
<td>-</td>
<td>3/4</td>
<td>-</td>
<td>Intermodulation</td>
</tr>
<tr>
<td>3( \omega_1 )</td>
<td>-</td>
<td>-</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>3( \omega_2 )</td>
<td>-</td>
<td>-</td>
<td>1/4</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.1:** Components in the expanded third degree polynomial

The rows in the table named “Intermodulation” are components appearing in band. Consider \( \omega_1 \) and \( \omega_2 \) to be very close, then 2\( \omega_1 \pm \omega_2 \) and 2\( \omega_2 \pm \omega_1 \) both will be almost equal to the fundamental frequency.

The main thing to remember: **intermodulation distortion is in band distortion; hence no filtering can be done to save the situation.**
2.1 Linearity

2.1.4 Measuring linearity

Once we know what damage nonlinearity can cause, we like to measure the effects of the nonlinearities in system. By doing these measurements we can compare our amplifiers and see how well they perform.

1dB compression point

The most common measurement of compression is the 1dB compression point. By measuring the power in versus power out, i.e. the power gain for very small input signals (before compression can be a factor) we can predict how the output should behave for larger input signals. The prediction is a simple linear extrapolation of the small-signal behaviour.

When the predicted values and the measured values differ by 1dB, the “1dB compression point” is reached. It is advantageous to plot the measured values versus the predicted ones in the same chart to see this clearly.

There are two ways of referring to this point, either by looking at power in or power out. When the 1dB compression point is stated in terms of input power it is called input referred otherwise output referred.

Figure (2.2) illustrates a power in versus power out curve with its 1dB input referred compression point.

![Power in vs. power out and 1dB compression point](image-url)

**Figure 2.2**: Power in vs. power out and 1dB compression point
Peak to Average Power Ratio

The concern this measurement addresses is the increased signal power present in a multiple-frequency-signal compared to a single frequency tone. The increased signal power may cause compression. PAR is helpful to estimate the adequate “back off” required to avoid compression. It is up to the modulation scheme to even the signal power, hence decreasing the PAR. This measurement is, thus, nothing the amplifier designer can adjust. Anyway, as an example, knowing that the amplifier should be used with a signal having large PAR could make the designer to try to increase the 1dB compression point. The definition of PAR is given below:

\[
P_{\text{PAR}} = \frac{\text{PeakEnvelopePower}}{\text{AverageEnvelopePower}} = \frac{\hat{P}}{\bar{P}}
\]  

(2.7)

Once again assume a two tone input signal, repeated here for convenience:

\[
x(t) = A \left( \cos(\omega_1 t) + \cos(\omega_2 t) \right)
\]  

(2.8)

This equation can be rearranged by using trigonometric formulas to get:

\[
x(t) = 2A \left[ \cos(\frac{\omega_1 t - \omega_2 t}{2}) \right] \left[ \cos(\frac{\omega_1 t + \omega_2 t}{2}) \right]
\]  

(2.9)

The peak envelope power (PEP), \(\hat{P}\), present in the signal can now be calculated using equation (2.9) as:

\[
\hat{P} = \left( \frac{\frac{3}{\sqrt{2}}}{R} \right)^2 = \left( \frac{2A}{\sqrt{2}} \right)^2 = \frac{2A^2}{R}
\]  

(2.10)

Further by using equation (2.8), the average envelope power can be calculated by calculating the average power in each term in the sum:

\[
\bar{P} = \left( \frac{A}{\sqrt{2}} \right)^2 + \left( \frac{A}{\sqrt{2}} \right)^2 = 2 \left( \frac{A}{\sqrt{2}} \right)^2 = \frac{A^2}{R}
\]  

(2.11)

In this equal amplitude two tone case, the peak-to-average power ratio is evaluated to a factor of two or equivalently 3dB. This means that the amplitude of a single tone would have to be \(\sqrt{2}\) times the amplitude of the two tones to carry the same power. If we think about it the other way, inputting the same amplitude in a two tone test will imply peaks that are \(\sqrt{2}\) times those of a single tone input. This increase in peak input amplitude will imply greater compression, i.e. the amplifier will operate in a more nonlinear region. In this case it is therefore important to reduce each tone’s input power by 3dB to be able to compare with the single tone case.
2.1 Linearity

Let us try to generalize the input signal a bit. For example an amplitude modulated input signal could be written as:

\[ x(t) = m(t)A_c \cos(\omega_c t) \]  

(2.12)

where \( \omega_c \) represents the carrier frequency and \( m(t) \) the message. If we now let \( m(t) \) be a sum of cosines the input should look like:

\[ x(t) = 2A \cos(\omega_c t)(\cos \omega_m + A \cos 2\omega_m + B \cos 3\omega_m + \cdots) \]  

(2.13)

Let us look at a specific example, again with only two tones, hence \( B, C, D \ldots = 0 \). Figure (2.3) presents four different cases when varying \( A \).

Figure 2.3: Envelope voltage with varying peak-to-average power ratio.

The dotted lines in figure (2.3) are the voltage amplitudes corresponding to the average power. We have made all cases have equal mean power to be able to compare them. The important thing to notice is the difference in the peak-to-average power ratio.

In the upper left and upper right plot, \( A = 1 \) and \( A = 0.25 \) respectively, the time above the average power level is relatively little. However, looking at the lower plots, the signal is constantly above the average power level. If the amplifier starts to compress at the average power level, we have four different cases of when distortion will occur.

Simulations have shown that going deep into compression is very harmful in terms of intermodulation distortion [2]. One simple explanation is
that for high peak-to-average power ratios much of the signal energy will be in the peaks. If much of the signal energy is in the peaks and these exhibits high compression, much of the signal will be distorted.

As can be suspected by the previously presented two-tone examples, an input signal with more than two tones will not make the situation better. At best we can model it statistically and calculate the expected PAR. However, the PAR will vary as a function of the transmitted data and will force us to decrease the input signal. For linear amplifiers reducing the input signal implies lower efficiency. Fortunately moder modulation schemes are designed to have a fairly constant PAR!

To conclude: we have to investigate the peak envelope power, to be able to say anything about the mean power that our amplifier can deliver. It is not sufficient to investigate this with only two tones!

**Third-order intersect point - IP3**

The intermodulation distortion component from table (2.1) is repeated here for convenience:

$$y(t)|_{dB} = \left[\frac{3}{4} \alpha_3 x(t)^3\right]_{dB} = 10 \log\left(\frac{3}{4} \alpha_3 x(t)^3\right) = 3 \times 10 \log\left(\frac{3}{4} \alpha_3 x(t)\right)$$

As can be seen the intermodulation component increases three times as fast as the fundamental tone, if both are expressed in dB. Even though the intermodulation distortion should be much smaller than the fundamental tone in the area of operation, their power levels would eventually intersect if no compression existed. It is also this point of intersection that the term IP3 is an abbreviation of, namely the third-order intersect point.

An example of IP3 measurement is given in figure (2.1.4). As can be seen the intersection point is not a real point on either of the curves (ever). Instead it lays on the extrapolated curves that defines compression. In this case the intermodulation component is extrapolated from -20dBm.

In the same way as 1dB compression point, IP3 can be referred from the output power or the input power. Thus two terms exists, input referred IP3 and output referred IP3.

**The two tone test**

In the same way as we introduced the concept of intermodulation distortion, namely by exciting the model with two tones; the most common way to find the IP3 is to excite the amplifier with two sinusoidal tones with equal amplitude. This kind of test can be done in a physical test bench but of
course also using circuit simulation software. Below follows a few guidelines and considerations to take notice of when attempting to perform such a test using a simulator.

In a circuit simulator, such as the one in Cadence, the IP3 test is most easily performed with help of a periodic steady state (PSS) analysis. Using a PSS analysis is a fast and accurate way to measure the IP3. The requirement is, of course, periodicity. In certain cases, such as ∆Σ-modulation (to be explained later), it is not possible to use the PSS analysis. The alternative is then to rely on the transient response and a preceding DFT analysis.

However, before using either PSS or transient analysis, an AC-analysis should be performed to measure the amplitude characteristics. By doing this it is possible to determine how far away the two input tones can be placed without the AC characteristics affecting the result. The reason for placing the tones at a good interval from each other, and not as close a possible, is to ease the burden of the simulator.

In the PSS analysis, the spacing between the tones decides the step between the harmonics that will be simulated. The smaller the step the more computations have to be done (the more harmonics are needed), needless to say the wish is to keep this number as low as possible to make the simulation run faster.

To be able to detect the intermodulation distortion when performing a DFT analysis the simulation time must be enough. For example, if the spacing between the tones is 1MHz then at least 1µs have to be analysed.
Further, if the frequencies of interest are around 2GHz, as in the case of 3G, then around 2000 RF cycles need to be analysed. Hence, by separating the tones we can reduce the simulation time needed.

The conclusion is that the tones should be chosen so that the intermodulation terms appear inside a frequency region, which is unaffected by AC characteristics. If the frequency gap between the tones is chosen too wide, the IP3 figures could appear better as a consequence of the out of band filtering that occurs in the system. Real intermodulation distortion appears in band, no filtering can remove them, so our simulation has to reflect that situation to be meaningful.

### 2.2 Efficiency

To measure the efficiency of a RF power amplifier the two most common measurements are: drain efficiency and power added efficiency.

#### 2.2.1 Drain Efficiency

The simplest to use and also most often used measurement of efficiency is **drain efficiency**. Drain efficiency is simply the power delivered to the load divided by the DC power in the power amplifier. Ideally we want all of the DC power to appear as RF power in the load, in that case the drain efficiency becomes 100%. We define the drain efficiency as:

\[
\eta = \frac{P_{\text{load}}}{P_{\text{DC}}} \tag{2.15}
\]

#### 2.2.2 Power Added Efficiency - PAE

A more accurate way to measure the efficiency is to take all parts of the system in consideration, not just the power amplifier. With power added efficiency we also measure the power dissipated in the stages preceding the power amplifier. For example, we may need to input 0.1W to get 1W at the load. This implies an effective output power of 0.9W, which should be compared to the DC power in the transistor. We define power added efficiency as:

\[
\eta = \frac{P_{\text{load}} - P_{\text{input}}}{P_{\text{DC}}} \tag{2.16}
\]

#### 2.2.3 Coding Efficiency

After investigating linear amplification and before starting to experiment with switching power amplifiers we studied the subject of $\Delta\Sigma$-modulation.
(chapter 5) and constructed a way to simulate them (chapter 7.2). The purpose was to have a signal source to apply to the switching amplifier and we thought of \( \Delta \Sigma \)-modulation as being one way of achieving that. Unfortunately, after the theory and design of the test benches we understood the great pitfall of having to filter the \( \Delta \Sigma \)-modulated output. Afterwards, applying extensive filtering at the output of a power amplifier should have made us a bit suspicious. Heavy filtering and power amplification even sounds quite contradictory. Unfortunately, it turned out, that most of the energy in the \( \Delta \Sigma \)-modulated signal is filtered away. In [12] the term \textit{coding efficiency} is defined as:

\[
\eta = \frac{P_{\text{before filter}}}{P_{\text{after filter}}} \tag{2.17}
\]

The results of our investigation of coding efficiency as a function of \( \Delta \Sigma \)-modulator and input signal is found in chapter 7.3.
In this chapter we will take a closer look at the characteristics and structure of the Class A, B and C amplifiers.

3.1 The School Textbook Class A Amplifier

Your average textbook would present you with a Class A amplifier stage looking somewhat different from what you would desire for RF purposes. Below is a basic schematic of what is called a “Common emitter stage”. This is the arch type and absolute favourite of textbooks writers, but not an optimal solution to RF amplification.

![Diagram of a Common Emitter Stage]

**Figure 3.1:** The most common and useful PA, if you go by the occurrences in textbooks
3.2 Deriving a Class A RF Amplifier

The needs of a RF amplifier are not those of the average audio spectrum amplifier. The method of transmitting data using a modulated carrier wave at a high frequency requires special consideration from a design point view. The following chapter will give an intuitive motivation to the various components of a class A RF amplifier.

The goal is still the same as for the lower-frequency power amplifier, namely to deliver a higher power to a load than the original source was able to do, using a DC source for increased power, but preserving the frequency spectrum of the original source.

3.2.1 A general design

We begin with the core element, the transistor. By changing the potential at the gate we can steer the current flowing from drain to source. The idea is to use this possibility to guide a current through another component or device. Initially several ways of placing such a load in relation to the transistor comes to mind. We will take a closer look at only one of them.

This option is popularly called a common source step and requires “something” that allows the transistor to affect the potential of the point $V_D$, otherwise it would be locked to the potential $V_{DD}$ imposed by the DC-source.

By adding an inductor (popularly named RF-choke, RFC) we can isolate $V_D$ from $V_{DD}$ for high frequencies since the impedance of the RFC will become very large for such frequencies. This will allow the transistor, when increasing its conduction, to lower the potential of $V_D$, if the frequency is
high enough. If your transistor is large and your load small, a large current change through the load occurs as a response to the lowered potential of $V_D$, while a large load causes a large potential loss of $V_D$ instead.

![Figure 3.3: $V_D$ isolated from $V_{DD}$ at high frequencies by a RFC](image)

Adding the inductance however, does not solve the fact that for all intents and purposes we have a DC path to the load constantly leaking a DC current. This is easily discouraged by ways of adding a (large) capacitance in series with the load. The impedance will be effectively zero for high frequencies while still blocking DC.

![Figure 3.4: Decoupling capacitance added](image)

To ensure class A operation the signal we apply to the gate needs to always stay above the threshold voltage of the transistor, $V_T$. We achieve
this by biasing the gate of our transistor to a DC voltage suitably high to allow our signal through without lowering the “overall” gate potential below $V_T$, using two resistors in the kilo ohm range. It is important not to use too small resistors, as that would mean an unnecessary DC power loss from the DC current through them. $V_{bias}$ should be at least as high as $V_T + V_{in}$, or some clipping will occur. At the same time we isolate the input source from our new DC level with a large capacitance.

![Biasing to DC-level of $2/3$ of $V_{DD}$, $R_2 = 2R_1$](image)

For simulation purposes, and later on for some practical uses, it is important that the signal source has the same output impedance as the input impedance of the amplifier (to avoid signal reflections). We want to construct an AC-path perceiving the same impedance as the input source’s while maintaining the DC biasing.

By adding a resistance of the same size as the output resistance of the input source, in series with one of the resistors used to bias the transistor gate, then decoupling the new resistance with an inductance and the other one with a capacitor, we create different paths for DC and AC, thus creating the conditions we stated above.

Choosing the correct input impedance and output impedance is very much part of the balancing the amplifier for linearity and correct power output.

### 3.2.2 Power Demands

The first thing we should be aware of is that it is far from certain that a single CMOS transistor amplifier can accommodate our power needs. It is increasingly hard to drive the gate of large transistors at high frequencies. The impedance of the gate decreases with size just as it decreases with
higher frequency, at some point we might find ourselves trying to drive a one-ohm-sized impedance to a potential of several volts (a 10mm transistor for example). This is just short of impossible not to mention a waste of time since that will require more energy than our amplifier will be able to deliver to our load, thus creating a PA that delivers less power than it requires from the input to drive it.

If our power expectations are realistic there is still the matter of linearity. A CMOS FET transistor is far from the ideal model of a transistor. To achieve the desired linearity even at low power levels, we might have to sacrifice the efficiency as far as below 1%.

Finding a suitable transistor size is somewhat of an iterative process but it helps considerably to have a general idea of how the $I_D/V_{GS}$ and $I_D$/transistor size charts looks.

Using no linearizing feedback or adaptive filters we can never expect a better linearity than what the $I_D/V_{GS}$ chart allows, meaning we have to find a region where the following holds:

1. if an increase $V_{GS}$ yields $I_D$ then an increase $K \cdot V_G$ has to yield $K \cdot I_D$ regardless of $V_{DS}$.

This roughly means we have to take care not to use too large input signals or allow too large variations of $V_{DS}$. The chart illustrates.

By looking at such a chart or by experimenting a little we can find an approximate region in which we can allow ourselves to operate. Note that doing this just gives us a reasonably good place to start our simulations, nothing more. Using the desired swing of $V_D$ we can calculate the expected power dissipation of our load as:
3.2 Deriving a Class A RF Amplifier

Figure 3.7: $I_D/V_{GS}$ Chart

\[ V_D = \hat{U} \]  \hspace{1cm} (3.1)
\[ \frac{U^2}{R \cdot L} = 2 \cdot P_{AVG} \]  \hspace{1cm} (3.2)
\[ (3.3) \]

If this is more than or exactly the power you wanted you are more or less done now, or you can decrease your input signal (and biasing) to lower the power delivered to your load even further. Choose a transistor size from your $I_D$/transistor size chart and you are ready to simulate the design to verify that your linearity was good enough. However, if this power is smaller than what you need there is still a lot of work to be done.

As shows there is a way to increase the power delivered to your load by decreasing the impedance (resistance) of your load. Usually this is a fixed value (a surprising number of things does not come with a potentiometer built in) but by using impedance transformation it is possible to reduce
the impedance of your load seen from D ($V_D$). Using equation (3.2) backwards we can arrive at a new load that would give us the desired power while still not increasing $V_D$ beyond the nominally linear region identified in the charts.

$$\frac{U^2}{2} \cdot P_{AVG} = R_{L_{new}}$$  \hspace{1cm} (3.4)

Assuming we can construct a net to transform the impedance to the desired value we move on to the transistor and its input signal. $V_G$ and $V_D$ were determined by visually inspecting the $I_D/V_{GS}$ chart.

The transistor has to be big enough to support both the current $I_D$ and the leaking component $I_{DC}$ caused by the biasing. This leakage is what ultimately ruins the efficiency of your amplifier. Choose the transistor size such that $V_G \cdot I_D = \hat{U}/R_L$.

Having done this, then all we need to do is create the impedance transformation and simulate the circuit and see if we can tune things a bit.

**Impedance match**

We would like our resistive load to look like an impedance of e.g. lesser value, preferably still resistive. By putting a capacitance in parallel with the load resistance it is possible to lower the impedance if we also add an inductance in series it is possible to remove the reactance and thus have a resistive load once more.

### 3.3 Class B Amplifier

Though structurally similar to the Class A amplifiers (filtering may be added), the Class B amplifiers only amplify half of the input wave cycle. Without filtering or tricks (3.3.1) they cause a large amount of signal distortion, but their efficiency improves dramatically. The efficiency improvement stems from the amplifying element being switched off altogether half of the time, and so do not dissipate power.

#### 3.3.1 Push-Pull

A widely used circuit using Class B elements is the “push-pull” arrangement. Complementary devices, preferably identical in all aspects, amplify the opposite halves of the input signal, which of course consumes more power than the example above. The two halves are then recombined to create the output. This arrangement still gives excellent efficiency, while reducing the distortions of the single stage. However recombining the two waveforms seamlessly difficult and the design is prone to small glitches at the “joins”
between the two halves of the signal. This phenomenon is unique for the push-pull set up and is called crossover distortion.

Figure 3.8: Class B Amplifier demonstrating the signal properties

3.3.2 Class AB

A solution to the aforementioned distortion issue is to bias the two devices a little past the brink of conduction, rather than firmly off, when they are not amplifying. An amplifier where this has been done is called a Class AB Amplifier or said to use Class AB operation.

This causes each device to be operated in a non-linear region which however is linear over half the waveform. Contrary to former example the stages still conducts a small amount on the “off-half”. While behaving as a class A amplifier in the region where both devices are in the linear region, the circuit is not a class A since the signal passes outside the linear region, hence the name AB. The transients causing the typical crossover distortion of class B operation will occur for each of the halves but when the two halves are combined, the total crossover distortion is greatly minimized or eliminated completely.

3.4 Class C Amplifier

Also Class C amplifiers share the global topology of the Class A stage (filtering may be added). Conduction is once more the primary difference between the stages, a class C amplifier conducts less than half of the input
signal, and thus the distortion of the output is extreme. However the upside is that efficiencies of up to 90\% can be reached.

Rather than trying to find ways of compensating for the distortion in the general case, as was AB operation to class B, class C is used together with strongly tuned filters. This limits the bandwidth capabilities but for RF signals with small bandwidth and tuned loads or low grade audio applications, such as sirens, this may be enough.

Figure 3.9: A Class C Amplifier and signal example
Chapter 4

Switching PAs

This chapter intends to briefly introduce the subject of switching power amplifiers. The major reason for the current large interest in switching amplifiers is the will to increase the power efficiency, for example a base station transmitter. The problem is, however, not to build the switching amplifiers, but rather to make them do something useful. The most efficient switching amplifiers are so called tuned switching amplifiers. These are tuned to a certain frequency and are therefore very happy with just that frequency. So again, the problem is to force it to modulate the amplitude and/or the phase to make it do something useful. One way is to modify the switching input signal, commonly known as pulse width modulation, were the pulse widths change as a function of the amplitude. We hoped that another method would be to use a ∆Σ-modulated signal as input. We have, however, learned that a ∆Σ-modulated input signal unfortunately excludes the use of standard tuned switching amplifiers. To get feeling for what a high-efficiency amplifier means we have studied the theory and finally built and simulated a class E amplifier.

4.1 A Basic RF Switching Amplifier

In figure (4.1) a simple schematic of a switching mode amplifier is presented together with the switching waveforms. Some elements are familiar from our previous (linear) derivations, that is the RFC and a DC-blocked RF load resistor.

As can be seen in figure (4.1b) one result of the ideal switch is complementary current- and voltage waveforms. This means that no power is dissipated in the switch, the current and voltage are never simultaneously high. This thought of having an ideal switch, which dissipates no power, is easily interpreted as the optimum solution for a switching amplifier. However, this is not completely true. The viewpoint we have chosen to understand this misconception is to look at the fourier expansion of a
Figure 4.1: Basic RF Switching amplifier (a) and its waveforms (b) [2]

square wave, which is the output from the amplifier in this case. The fourier expansion is given in equation (4.1).

\[
\frac{4}{\pi} \sum_{n=1,3,5,...}^{\infty} \frac{1}{n} \sin \left( \frac{n\pi x}{T/2} \right)
\]  

(4.1)

The RMS for a square wave with amplitude \(A\) and period \(T\) is given by:

\[
RMS_{\text{square\,wave}} = \sqrt{\frac{1}{T} \int_{\tau}^{\tau+T} V^2(t) \, dt}
\]  

(4.2)

\[
= \sqrt{\frac{1}{T} \int_{0}^{T/2} A^2 \, dt + \frac{1}{T} \int_{T/2}^{T} (-A)^2 \, dt}
\]  

(4.3)

\[
= \sqrt{\frac{2A^2}{T} \int_{0}^{T/2} 1 \, dt}
\]  

(4.4)

\[
= \sqrt{\frac{2A^2 T}{T}} = A
\]  

(4.5)

From equation (4.1) it can be seen that the fundamental component has an amplitude of \(\frac{4}{\pi}\). To evaluate the RMS of this fundamental sinusoidal we only have to divide the amplitude with the factor \(\sqrt{2}\). By doing this we
can compare the power in the fundamental tone relatively the total power in the square wave.

\[
\frac{P_{\text{fundamental}}}{P_{\text{squarewave}}} = \left(\frac{\frac{4}{\sqrt{2}\pi}}{1^2/R}\right) = \frac{16}{2\pi^2} = \frac{8}{\pi^2} \approx 81\% \tag{4.7}
\]

This comparison tells us that no matter how efficient the switch is, the is efficiency is always limited to 81%. To be honest, we have cheated a bit and only cared about the symmetrical case, with 50% duty cycle. This is, however, the most beneficial case, which is shown in [2]. All other duty cycles will given even worse efficiency.

### 4.2 Ideal Switch and Harmonic Short

If the harmonics are eliminated a new situation occurs. The easiest way to eliminate the first even order harmonic is to place a so called tank circuit in parallell with the load resistor. The task of the tank circuit is to create an harmonic short. In figure (4.2) the ideal operation of an inverse class F amplifier is given, where the current waveform remains square and the voltage waveform takes sinusoidal form. In the ideal case, the load impedance represents a short for all even order harmonics and high impedance for all odd order harmonics.

![Figure 4.2: Ideal switch and harmonic short waveforms (ideal inverse class F operation)](image)

The DC current component is clearly \(I_0\). The DC voltage component can be calculated by evaluating the expectation integral:
\[ V_{DC} = V_s \int_0^{0.5} t \sin(2\pi t) \, dt \quad (4.8) \]

\[ = \frac{V_s}{2\pi} \left[ -t \cos(2\pi t) \right]_0^{0.5} + \frac{V_s}{2\pi} \int_0^{0.5} \cos(2\pi t) \, dt \quad (4.9) \]

\[ = \frac{V_s}{2\pi} \left( -0.5 \cos(\pi) \right) + \frac{V_s}{4\pi^2} \left[ \sin(2\pi t) \right]_0^{0.5} \quad (4.10) \]

\[ = \frac{V_s}{\pi} + \frac{V_s}{4\pi^2} \sin(\pi) \quad (4.11) \]

\[ = \frac{V_s}{\pi} \quad (4.12) \]

The DC power hence becomes:

\[ P_0 = \frac{I_0 V_s}{\pi} \quad (4.13) \]

By looking at the fourier expansion of a half rectified sine wave, given in equation (4.14) the fundamental component can be found to be \( \frac{1}{2} \sin(\omega t) \) (also the DC component is visible as \( \frac{1}{\pi} \) as was derived):

\[ \frac{1}{\pi} + \frac{1}{2} \sin(\omega t) - \frac{2}{\pi} \sum_{n=2,4,...} \frac{1}{n^2 - 1} \cos(\omega nt) \quad (4.14) \]

The fundamental voltage component hence becomes \( \frac{V_s}{2} \). For the fundamental current component we can look at the fourier expansion of the square wave, which gives us an amplitude of \( \frac{4I_0}{\pi} \). The fundamental power (RMS) is thus given by:

\[ P_1 = \frac{V_s}{2} \frac{4I_0}{\sqrt{2} \sqrt{2}} = \frac{V_s I_0}{\pi} = P_0 \quad (4.15) \]

The theoretical drain efficiency becomes:

\[ \eta = \frac{P_1}{P_0} = 100\% \quad (4.16) \]

This result unfortunately poses practically impossible harmonic impedance conditions. It is impossible, in real hardware, to design harmonic conditions so that an infinite sum of even and odd order harmonics gives a half rectified sinusodial and a square wave respectively. The best and often done is to provide harmonic short for at least two even order terms and harmonic peaking for three odd order terms, which gives an efficiency of approximately 85% [5].

The conclusion is that it is now, at least in theory, possible to achieve 100% efficiency by having an ideal switch.
4.3 Class D Switching Amplifier

Figure 4.3: Class D RF amplifier (a) and its waveforms (b) [2]

Figure (4.3) presents a circuit diagram and the switching waveforms of a class D amplifier. For this configuration two things are assumed:

i. The repetition cycle matches the resonant frequency of the LCR circuit.

ii. The LCR circuit has a high Q-factor (strong inertia), which forces the waveforms to remain sinusoidal.

Actually, we could end the analysis here because the fundamental assumptions for the class D amplifier goes against our requirements. We want to amplify a pulse train with varying repetition cycle, such as a $\Delta\Sigma$-modulated signal. Hence, the repetition cycle does not always match the resonant frequency of the LCR circuit. For completeness, we show the theoretical efficiency. The fundamental voltage component at the load is $\frac{4}{\pi} \cdot \frac{V_{dc}^2}{2}$ due to the switching voltage being square ($\frac{V_{dc}}{2}$ is the amplitude of the square wave, if it would be centered around zero, for which the fourier series expansion given in equation (4.1) holds). The fundamental voltage thus equals:

$$V_1 = \frac{2V_{dc}}{\pi} \quad (4.17)$$

Further the fundamental component of the load current can be found by realizing that the output current is a sum of two half rectified sinusoids. The switches named “A” and “B” seen in figure (4.3 a) conduct a positive half sinewave and a negative half sinewave respectively. As already derived in the previous section the fundamental component in a half rectified sinewave is $\frac{1}{2}I_{pk}$, hence the fundamental current component equals $I_{pk}$. The fundamental power (RMS) becomes:
Switching PAs

\[ P_1 = \frac{V_1 I_1}{2} = \frac{2V_{dc} I_{pk}}{2\pi} = \frac{V_{dc} I_{pk}}{\pi} \]  

(4.18)

and the dc supply power is

\[ P_{dc} = V_{dc} I_{dc} = \frac{V_{dc} I_{pk}}{\pi} \]  

(4.19)

Both equations are equal and therefore the drain efficiency is 100%. Also, there is an increase of \( \frac{4}{\pi} \) or roughly 1dB in output power. Finally it should be mentioned that the realizability of this type of amplifier is often reduced by the introduction of two switches. Particularly it is the high-side (ungrounded PMOS) device, that poses difficulties at higher frequencies.

There exists two architectures of the class D amplifier: voltage mode and current mode class D. To achieve high efficiency, for high frequency operation, the current mode class D is beneficial. We have studied and simulated one such amplifier as given in [11]. However, the simplicity and designability of a class E amplifier made us decide not to further investigate the class D mode of operation.

4.4 Class E Switching Amplifier

One of the major benefits with class E is that it can be thought of as a switching type amplifier but there is not so strong demands on the switch. Class E can be implemented with a switching device that, for example, includes a linear region. The transient response is forced to maximize power efficiency, with carefully chosen design criterias; even with a not so good switch. This is also expressed in the first article introducing the class E: “...the load network shapes the voltage and current waveforms to
prevent simultaneously high voltage and high current in the transistor; . . . , especially during the switching transitions [23].

4.4.1 Designability

One attractive feature of class E amplifiers is that they are designable, which means that there exist design equations and a tuning method. The first original set of equations rely on an infinite loaded $Q$, which follows from assuming the current in $C_2$ and $L_2$ to be sinusoidal. The equations can be made more accurate by adjusting them to a loaded $Q$ and using the $Q$ factor as a design parameter. Presented below are the improved design equations found in [23], where the original equations have been modified by means of a polynomial fit:

\begin{align*}
V_{DD} &= \frac{V_{breakdown}}{3.56} \times SF \\
\frac{P}{R} &= \frac{(V_{cc} - V_0)^2}{P} - 0.576801(1.001245 - \frac{0.451759}{Q_L} - \frac{0.402444}{Q_L^2}) \\
R &= \frac{(V_{cc} - V_0)^2}{P} - 0.576801(1.001245 - \frac{0.451759}{Q_L} - \frac{0.402444}{Q_L^2}) \\
C_1 &= \frac{1}{34.2219 f R} (0.99866 + \frac{0.91424}{Q_L} - \frac{1.03175}{Q_L^2}) + \frac{0.6}{(2\pi f)^2 L_1} \\
C_2 &= \frac{1}{2\pi f R} \frac{1}{Q_L - 0.104823(1.00121 + \frac{1.01468}{Q_L - 1.7879}) - \frac{0.2}{(2\pi f R)^2 L_1}} \\
L_2 &= \frac{Q_L R}{2\pi f}
\end{align*}

Some comments on the equations, could be done. L1 should, as usual, present a high reactance at the tuned frequency. Equation (4.20) is the supply voltage we should use. It is calculated by knowing that the maximum drain voltage for a class E amplifier is roughly 3.56 times the chosen supply voltage (for 50% duty cycle). Furthermore, SF is a safety factor. For example if the breakdown voltage for the chosen transistor is 1.2 volts and we want to have 10% safety margin then $V_{DD}$ should equal approximately 0.3 volt. This is a small voltage and it shows one of the main disadvantages of the class E amplifier. The transistor has to withstand relatively high voltages, which forces the designer to lower the supply voltage and thus the available output power. The maximum voltage is also a function of the conduction angle (duty cycle) [2]. Larger conduction angles mean more power in the tuning network and will cause larger voltages. This is a problem with for example pulse width modulation.

4.4.2 Tuning

Once the design parameters are evaluated using equations (4.20) to (4.25), the class E amplifier could be tuned for higher efficiency (correct operation).
In figure (4.5a) a typical mistuned class E $V_{DS}$ waveform is showed. Our experience have showed that there is quite little work of tuning the amplifier to reasonable drain efficiency, in the region of 80 to 90 percent. Figure (4.5b) summarizes the effect of adjusting parameters in the load-network. For example, increasing both $C_2$ and $L_2$ moves the crest down and to the right. A complete tuning guide is found in appendix (C).

**Figure 4.5:** Typical mistuned class E (a) and effects of adjusting load-network components (b) [23]
This chapter focuses on the method we mainly have investigated, as a way to generate the switching signal to be used in conjunction with a switching amplifier. The method is called $\Delta \Sigma$-modulation and is a way to convert a digital signal into analog or an analog signal into digital. In our case we think of the $\Delta \Sigma$-modulation as a digital to analog conversion; it is the analog version we finally want to reach the load. We look at the $\Delta \Sigma$-modulation as a translation from an amplitude varying signal representation to a two symbol representation. The two symbol representation might be suitable for a switching amplifier.

The first sections of this chapter tries to explain the functioning of the $\Delta \Sigma$-modulator by means of a common linearized model. This linear model gives some important insight and with help of this model the basic properties of a $\Delta \Sigma$-modulator is derived. The second last section in this chapter is used to exemplify the design of a $\Delta \Sigma$-modulator using MatLab and Simulink. The last section describes the procedure we have used to translate the Simulink model into a Verilog-model.

5.1 Introduction to $\Delta \Sigma$-Modulators

Understanding the behaviour of a $\Delta \Sigma$-modulator is quite tedious and for us it required many simulations to eventually get a feeling for what is happening. We see the $\Delta \Sigma$-modulator as a non-linear control system. It is non-linear because it makes use of both an ADC and a DAC. Anyhow, it is a control system because it uses feedback. Further the $\Delta \Sigma$-modulator is an oversampling converter. This means that it uses higher frequency than the (required) Nyquist frequency.

The standard approach for explaining the $\Delta \Sigma$-modulator is to replace the non-linear elements with a linear noise source, representing quantization noise. With this simplification we end up with a linear control system and therefore linear analysis can be performed. Especially, the transfer function
from the input to the output can be derived. From a system point of view, this transfer function characterizes the system.

Figure (5.1) shows a simple \( \Delta\Sigma \)-modulator and its linear z-domain model. This \( \Delta\Sigma \)-modulator works as an analog to digital converter. The input is analog and the integrator \((1/s)\) is commonly implemented with switched capacitor techniques [10]. In a digital to analog converter, the modulator loop is commonly implemented using a digital signal processor.

As we will see; reaching high resolution in an data-converter can be done in mainly three ways. Increasing the number of levels in the quantizer is the most logic approach. A fine grained quantizer increases the resolution. There may, however, be difficulties building a fine grained quantizer. The other two ways are oversampling and noise shaping. The \( \Delta\Sigma \)-modulator uses a coarse quantizer with oversampling and noise shaping. Why it works and the benefits from: quantizer resolution, oversampling and noise shaping will be derived. First, let us start with what we have: the \( \Delta\Sigma \)-modulator; and analyze it in the linear domain.

**Figure 5.1:** A delta sigma modulator used as an ADC and a linear z-domain model.

### 5.2 Linear Model

As stated earlier, figure (5.1) includes a linear model of the \( \Delta\Sigma \)-modulator. However the model can be made more intuitively understandable by replacing the integration made up by \( \frac{1}{z-1} \) with the configuration shown in figure (5.2). In this case the modulator is entirely made up by simple delay elements and summing nodes. Having this simple structure it is easy to use
z-domain analysis. First, by simply looking in figure (5.2), we can derive equation (5.1):

$$Y(z) = z^{-1}Y(z) + U(z) - z^{-1}V(z) \quad (5.1)$$

and equation (5.2):

$$V(z) = Y(z) + E(z) \quad (5.2)$$

Combining equations (5.1) and (5.2) gives us equation (5.3), as shown below.

$$V(z) = Y(z) + E(z)$$
$$= z^{-1}Y(z) + U(z) - z^{-1}V(z) + E(z)$$
$$= U(z) + E(z) - z^{-1}(V(z) - Y(z))$$
$$= U(z) + E(z) - z^{-1}E(z)$$
$$= U(z) + (1 - z^{-1})E(z) \quad (5.3)$$

As a hint to the high resolution possible for a \(\Delta\Sigma\)-modulator, let us take a look at DC-input. Remembering that the \(z\)-transform can be obtained from the discrete Fourier transform by the \(z = e^{j\omega}\) substitution it can be seen that the DC value is obtained for \(\omega = 0 \leftrightarrow z = 1\). If the error, \(e\), is finite then equality \(U(1) = Y(1)\) follows directly from equation (5.3). This means that the output is exactly equal to the input, hence very high resolution is possible for DC-input. Now, we want to simplify equation (5.3) by introducing two common functions as shown below:

$$V(z) = STF(z)U(z) + NTF(z)E(z). \quad (5.4)$$

The functions introduced are the signal transfer function (STF) and the noise transfer function (NTF). It is interesting to analyze both functions in
the frequency domain, by performing the $z = e^{j\omega}$ substitution. The signal transfer function equals unity and no filtering occurs. The noise transfer function becomes:

$$NTF(e^{j\omega}) = 1 - e^{-j\omega}$$ (5.5)

To see the magnitude response, we look at the squared magnitude:

$$|NTF(e^{j\omega})|^2 = |1 - e^{-j\omega}|^2$$
$$= |1 - \cos \omega - j \sin \omega|^2$$
$$= \sqrt{(1 - \cos \omega)^2 + (\sin \omega)^2}^2$$
$$= 1 - 2 \cos(\omega) + \cos^2(\omega) + \sin^2(\omega)$$ (5.6)
$$= 2 - 2 \cos(\omega) = 2(1 - \cos(\omega))$$
$$= 4 \sin^2\left(\frac{\omega}{2}\right)$$
$$= 4 \sin^2(\pi f) = \left(2 \sin^2(\pi f)\right)^2.$$

For $f = 0$ equation (5.6) equals zero and for frequencies close to zero it is approximately equal $(2\pi f)^2$. The noise transfer function clearly exhibits a high pass behaviour. This means that the quantization noise is filtered away from the low pass region, where the signal is located. This is a highly desirable property and it is often referred to as noise shaping. Figure (5.3) illustrates the noise filtering function of a first order $\Delta\Sigma$-modulator.

By analyzing the $\Delta\Sigma$-modulator in the linear $z$-domain, we have now found that there is a lot of filtering going on. In fact the whole construction of a $\Delta\Sigma$-modulator can be seen as a filter problem: we want to construct the desirable signal- and noise transfer functions. When we adopt to this viewpoint, because of understandability, it must be remembered that the $\Delta\Sigma$-modulator is not linear and there is much more complexity than revealed when performing linear analysis. In fact, linearize then analyze, is a common engineering behaviour, but in this case careful simulation and verification is needed to make sure that the $\Delta\Sigma$-modulator works.

### 5.2.1 Quantization Noise and Error Model

We have already introduced an error source to represent the error done when quantizing. If we let the error $e(n)$ equal $y(n) - v(n)$, i.e. the difference between the input and output, then the model is not an approximation. It is when we start making assumptions about the error source, that the linear model becomes approximate. One common assumption is that the error signal is a stochastic signal, for example independent white-noise [10].
5.2 Linear Model

![Noise transfer function for a first order ΔΣ-modulator](image)

**Figure 5.3:** Noise shaping for a first order modulator.

**Rectangular Distribution**

Let us see what happens if we assume that the quantization error is rectangular distributed. This means that the probability of the quantization error is equal for all values, i.e. there is no quantization error that is more or less probable. This assumption can be used for a very active input signal [10]. The assumption is denoted in equation (5.7) and the probability density function is depicted in figure (5.4).

\[ e(n) \in \text{rect}\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \quad (5.7) \]

It is easily shown that the probability density function is constant and equal to \( \frac{1}{\Delta} \), as shown in equation (5.8).
Mathematically the quadrature mean value is defined as: $\int x^2 f_X(x) \, dx$ [7]. Physically the quadrature mean value is a measurement of mean power. The quantization noise power is therefore given by equation (5.9).

$$\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} f_X(x) \, dx = 1 \leftrightarrow \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} A_x \, dx = 1 \leftrightarrow \left[ A_x \right]_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} = 1 \leftrightarrow 2A_x \frac{\Delta}{2} = 1 \leftrightarrow A_x = \frac{1}{\Delta}$$

Knowing that the noise power is $\Delta^2/12$ we can calculate the brick wall two-sided spectral density as in equation (5.10). $S_e(f)^2$ denotes the power spectral density of the noise signal. (When we integrate the power spectral density it should sum up to the noise power.)

$$\int_{-\frac{L_f}{2}}^{\frac{L_f}{2}} S_e^2(f) \, df = \int_{-\frac{L_f}{2}}^{\frac{L_f}{2}} A^2 f_s = \frac{\Delta^2}{12}$$

**Figure 5.4:** Probability density function for rectangular distributed noise.
This gives us

\[ A_e = \frac{\Delta}{\sqrt{12} \sqrt{f_s}} \quad (5.11) \]

It is interesting to stop for a while and reflect over our derived results. First the total quantisation noise power is a function of \( \Delta \) which is the difference between two adjacent quantization levels. This implies that it is a function of the number of bits in the quantizer. Increasing the numbers bits will decrease \( \Delta \) and hence decrease the total quantization noise power. Let us see what the theoretically maximum signal to noise ratio (SNR) is for a quantizer with \( N \) bits. To achieve the maximum SNR the largest input signal should be used. First we need two relationships regarding the quantizer. The largest output is:

\[ Y_{\text{max}} = \frac{L \cdot \Delta}{2} \quad (5.12) \]

and the number of levels, \( L \), for a \( N \)-bit quantizer is:

\[ L = 2^N \quad (5.13) \]

The signal power for a sinusoidal is given by:

\[ P_{\text{signal}} = V_{\text{rms}}^2 = \frac{Y_{\text{max}}^2}{\sqrt{2}} = \left( \frac{L \cdot \Delta}{2} \right)^2 \frac{1}{2} = \frac{L^2 \Delta^2}{8} \quad (5.14) \]

We now know both the signal power and the noise power, so the signal to noise ratio (SNR) can be calculated:

\[
SNR = 10 \log \left( \frac{P_{\text{signal}}}{P_{\text{error}}} \right) \\
= 10 \log \left( \frac{L^2 \Delta^2}{8} \right) \\
= 10 \log \left( \frac{12L^2 \Delta^2}{8 \Delta^2} \right) \\
= 10 \log \left( \frac{3}{2} L^2 \right) \\
= 20 \log \left( \sqrt{\frac{3}{2} L} \right) \\
= 20 \log \left( \sqrt{\frac{3}{2} 2^N} \right) \\
\approx 6.02N + 1.76 \quad (5.15)
\]

As can be seen by equation (5.15) the SNR is improved by approximately 6dB by each new bit added.
Quantization and Oversampling

It is evident from the previous section that we can improve the SNR of the quantizer by adding more levels to it. However, in ∆Σ-modulators the number of bits in the quantizer is relatively few. In our case we use a one bit quantizer. Fortunately, in ∆Σ-modulators the two most important properties are oversampling and noise shaping. Let us investigate what oversampling can do for us. The assumption of rectangular distributed quantization noise, with white power spectral density is still valid. The oversampling ratio is defined as:

\[ OSR = \frac{f_s}{2f_0} \]  

where \( f_s \) is the sampling frequency and \( f_0 \) is the signal frequency. Two times the signal frequency is often referred to as the Nyquist frequency. The key to understand the benefit of oversampling is to note that the quantization noise is spread over a wider frequency range. The quantization noise is still located between \(-f_s/2\) and \(f_s/2\), but the sampling frequency has been increased. Therefore, in the signal bandwidth the quantization noise power equals:

\[
\int_{-f_s/2}^{f_s/2} S_e^2(f)|H(f)|^2 \, df = \int_{-f_0/2}^{f_0/2} A_e^2 \, df = \frac{2f_0}{f_s} \frac{\Delta^2}{12} = \frac{1}{OSR} \frac{\Delta^2}{12} \]  

In equation (5.17) \( H(f) \) is a brick wall filter, used to represent the signal band with. Note that by using double the required sampling frequency the in band quantization noise power is halved, corresponding to 3dB gain in SNR. Once again assuming the largest input signal is used the, maximum achievable SNR for a sinusoidal input becomes:

\[
SNR = 6.02N + 1.76 + 10 \log(\text{OSR})
\]  

Noise Shaping

The final property left investigating is to see what happens when noise shaping is introduced. We now know that the maximum achievable SNR is dependent upon both the number of quantization levels and the oversampling ratio. For each new bit added to the quantizer approximately 6dB of SNR is gained and for each doubling of the oversampling rate the SNR improves by 3dB. Remember that the benefit from using oversampling could be derived from the fact that the quantization noise was spread over a larger frequency. To arrive at the noise power, we integrated the power density function only over the signal bandwidth. This was accomplished by adding a brick wall filter, with a bandwidth equal that of the signal.
Now, we simply change this brick wall filter into our noise shaping filter and perform the calculation of the quantization noise once more.

According to equation (5.6) the NTF of a first order ΔΣ-modulator equals: \(2 \sin(\frac{\pi f_0}{f_s})\). If we assume the OSR to be quite high, then \(\frac{f_0}{f_s} << 1\) and we may use the approximation \(2 \sin(\frac{\pi f_0}{f_s}) \approx 2\left(\frac{\pi f_0}{f_s}\right)^2\), which results from Taylor expanding the sin-function. The noise power calculation is shown in equation (5.19).

\[
\int_{-f_0}^{f_0} S_c^2(f)|H(f)|^2 \, df = \\
\int_{-f_0}^{f_0} S_c^2(f)|NTF(f)|^2 \, df \approx \\
\int_{-f_0}^{f_0} \frac{\Delta^2}{12} \frac{1}{f_s} \left(\frac{2\pi f_0}{f_s}\right) \, df = \\
\frac{\Delta^2}{12} \frac{1}{f_s} \frac{4\pi^2}{f_s^2} \int_{-f_0}^{f_0} f_0^2 \, df = \\
\frac{\Delta^2}{12} \frac{1}{f_s} \frac{4\pi^2}{f_s^2} \int_{-f_0}^{f_0} f_0^2 \, df = \\
\frac{\Delta^2}{12} \frac{1}{f_s^3} \frac{4\pi^2}{3} f_0^3 = \\
\frac{\Delta^2}{12} \frac{\pi^2}{3} \left(\frac{2f_0}{f_s}\right)^3 = \\
\frac{\Delta^2}{36} \pi^2 \left(\frac{1}{OSR}\right)^3
\]

Finally, with this estimation of the quantization noise power the maximal achievable SNR can be estimated. Once again we assume that the signal power is that of a sinusoidal. The result is shown below:

\[
SNR = 6.02N + 1.76 - 5.17 + 30 \log(OSR) \quad (5.20)
\]

As can be seen in equation (5.20) doubling the oversampling rate now increases the SNR by 9dB! This result should be compared to oversampling without noise shaping, which only increased the SNR with 3dB.

**Conclusion**

It seems like the ΔΣ-modulator is a good approach to achieve high signal to noise ratio. It spreads the quantization noise by oversampling. Further, it uses quantization noise shaping to gain even more from the oversampling. Finally, we do not need to construct a fine grained quantizer. We could, for example, use an inherently linear one bit quantizer. Of course, for
application, the linearity of the one bit quantizer is not the main benefit.
It is rather the two-level representation, which we are aiming for.

5.2.2 Higher Order $\Delta \Sigma$-modulators

Here we will briefly go through the maximum possible SNR for a second order $\Delta \Sigma$-modulator. The assumption is, as before, that the quantization error is rectangular distributed as given in equation (5.7). We also assume that the oversampling ratio is high enough to be able to approximate the sinusoidal NTF with the first term in a Taylor-expansion.

As can be seen in equation (5.19) the increase of SNR as a function of the oversampling ratio, is dependent upon the NTF. A linear model for a second order $\Delta \Sigma$-modulator is given in figure (5.5).

![Linear model for a second order $\Delta \Sigma$-modulator.]

We can find the NTF by finding the transfer function from the input to the output, as was done in equation (5.3). In the frequency domain the squared magnitude of the NTF, for a second order $\Delta \Sigma$-modulator is found to be [21]

$$|NTF(e^{j2\pi f})|^2 = (2 \sin \pi f)^4 \approx (2\pi f)^4, f << 1 \quad (5.21)$$

If we derive the quantization noise power with this NTF, we find the signal to quantization noise ratio with a sinusoidal input to be [10]

$$SNR = 6.02N + 1.76 - 12.9 + 50 \log(OSR) \quad (5.22)$$

With a second order modulator the gain in SNR is 15dB for each doubling of the oversampling ratio. This could be compared with the 9dB gain of a first order modulator. In figure (5.6) the estimated maximum possible SNR with a sinusoidal input is given. In other words, equation (5.22) and equation (5.20) are plotted.

As can be seen relatively high oversampling rates are required for a first order $\Delta \Sigma$-modulator, to get good SNR. In general it can be shown [10] that an Lth-order noise-shaping modulator improves the SNR by $6L +$
5.2 Linear Model

**Definition 5.1.** A binary $\Delta\Sigma$-modulator with an $NTF = H(z)$ is likely to be stable if $\max |H(e^{j\omega})| < 1.5 \forall \omega$.

This is also the criteria we have used when designing the NTF, of course together with simulations.

Finally there exists many structures for constructing higher order modulators. As we are not to build the $\Delta\Sigma$-modulator we leave that topic. Instead we focus on just one structure called error-feedback structure, for its theoretical high SNR. A good book for a more complete discussion is for example: [21].

### 5.2.3 Optimization of the NTF

To further optimize the noise shaping behaviour optimization of the NTF’s zeros and poles can be applied [21]. For a simple low pass $\Delta\Sigma$-modulator the NTF has all its zeros at DC, hence at DC quantization noise will be

---

**Figure 5.6:** Maximum theoretically possible SNR for a first and second order $\Delta\Sigma$-modulator.

3dB/octave. The major problem with higher-order $\Delta\Sigma$-modulators is however stability. For a single bit $\Delta\Sigma$-modulator there exists [21] some conservative and some liberal approximate criteria for designing a stable NTF. One widely used and simple criterion is the Lee Criterion [21].
filtered away. By spreading the zeros over the signal range improved signal to quantization noise ratio can be achieved [21]. The same applies for the poles, which should surround the signal band. Spreading the zeros reduces the total in band quantization noise power and moving the poles toward the zeros reduces the out-of-band gain, which improves stability.

5.2.4 Error Feedback Structure

There exists many different structures for ∆Σ-modulators. Specifically one is called error feedback structure [21]. This structure has one drawback and one advantage. The drawback is that it is sensitive to variations in the filter coefficients. This property implies that the error feedback structure is not suited for analog implementations. It could however be useful for digital implementation, where filter coefficients do not vary. The advantage of using this structure is that it is theoretically possible to exactly represent the input. In other words the signal transfer function equals unity. The error feedback structure is given in figure (5.7). Note that the loop filter is $1 - NTF(z)$, which means that we still only have to construct the noise transfer function to be able to arrive at the loop filter. Construction of the NTF is considered in section 7.2.1.

![Error feedback structure](image)

**Figure 5.7:** Error feedback structure.

Performing linear analysis we can derive the input to output relation:

$$Y(z) = X(z) + E(z)$$  \hspace{1cm} (5.23)

Also the node variable $X(z)$ can be expressed as a function of the input and output:

$$X(z) = U(z) + H(z)(X(z) - Y(z)) \Leftrightarrow$$

$$X(z)(1 - H(z)) = U(z) - H(z)Y(z) \Leftrightarrow$$

$$X(z) = \frac{U(z) - H(z)Y(z)}{1 - H(z)}$$  \hspace{1cm} (5.24)
Combining equation (5.23) and (5.24) gives:

\[ Y(z) = U(z) + (1 - H(z))E(z) \]  

(5.25)

As can be seen by equation (5.25) the signal transfer function equals unity and the noise transfer function equals \(1 - H(z)\), hence \(NTF(z) = 1 - H(z)\) or equivalently the loop filter becomes \(H(z) = 1 - NTF(z)\).

## 5.3 Bandpass \(\Delta \Sigma\)-modulators

Now that we know the basic functioning and properties of a \(\Delta \Sigma\)-modulator it is time to try to make it more useful for our application. First of all building a radio system implies the use of radio frequencies rather than base band frequencies around dc. It would be good if we could transform the theory and properties into higher frequencies. Once again we rely on linear analysis. Remember that the most important property of the \(\Delta \Sigma\)-modulator is noise shaping, i.e. the ability to filter away the quantization noise at the signal of interest. By linear analysis this property can be derived from the noise transfer function. The noise transfer function, in turn, can be described as a filter. This means that we have an opportunity to apply a low pass-to-bandpass transformation on this filter. A standard and simple transformation is to let \(z = -z^2\).

For a low pass \(\Delta \Sigma\)-modulator the NTF has zeros close to or equal to \(z = 1\), i.e. for frequencies close to dc because \(z = 1 = e^{j\omega} \rightarrow \omega \approx 0\). The idea is to move the zeros away from dc toward higher frequencies. Let us see where the zeros, of the NTF, are placed if we apply the \(z = -z^2\) transformation.

\[-z^2 = -(e^{j\omega})^2 = -e^{2j\omega} = -\cos(2\omega) + j\sin(2\omega) = 1 \rightarrow \]  

(5.26)

\[
\begin{align*}
\text{sin}(2\omega) = \sin(4\pi f) = 0 & \rightarrow f = 0, \frac{1}{4} \\
-\cos(2\omega) = -\cos(4\pi f) = 1 & \rightarrow f = \frac{1}{4}
\end{align*}
\]

As can be seen by equation (5.26) the normalized frequency of \(f = \frac{1}{4}\) now corresponds to the zeros of the NTF. Hence, with this type of transformation, the zeros are placed at \(f = \frac{1}{4} f_s\). As this is a simple solution to transform a low pass \(\Delta \Sigma\)-modulator into a bandpass \(\Delta \Sigma\)-modulator it has a drawback. It requires a sampling frequency of four times the signal frequency. For example, for a signal frequency at 2GHz the required sampling frequency becomes 8GHz, which is very high. One solution to this problem is summarized in section (6.3.2).

Finally it should be noted that by transforming a low pass \(\Delta \Sigma\)-modulator with \(z = -z^2\) the same properties of the low pass modulator are inherited into the bandpass modulator, even though the order of the filter is doubled.
[21]. This means, for example, that when we are talking about a fourth order bandpass $\Delta\Sigma$-modulator it has the same properties as a second order low pass $\Delta\Sigma$-modulator.
Chapter 6

Survey

6.1 Introduction

This chapter intends to introduce the methods and ideas, for improved power amplifier efficiency, that we have come across during our research for this thesis. Basically this chapter can be seen as a kind of summary of the papers we have found in the IEEE-database.

6.2 Pulse Width Modulation (PWM)

The first applications with a switching power amplifier were based on pulse width modulation (PWM). The idea of PWM is to let the widths of the pulses represent the amplitude of the signal. This is illustrated in figure (6.1), where a sawtooth signal is used to compare the amplitude of the input signal and create pulses with varying widths.

As can be seen this method also requires a kind of oversampling, namely the sawtooth signal frequency must be many times higher than the input signal. However, for audio applications even a couple of MHz may be enough, which could be realizable. For example if we assume the audio signal bandwidth to be 44kHz then a 2.2MHz sawtooth signal frequency would be a factor of 50 times higher than the input signal.

6.2.1 RF Applications

When trying to use the same ideas in the RF-field we directly encounter the problem of generating the PWM pulses. RF pulsewidth modulation requires a switching frequency larger than or equalling the carrier frequency (assuming subharmonics not to be used) [17]. A good thing is that spurious products are concentrated around the harmonics of the carrier frequency.

To get a linear relationship between the pulsewidth and the corresponding amplitude it is necessary to vary the pulsewidth according to an inverse
sine of the modulating signal \([17]\), as the fundamental component of a square wave is proportional to a sine. If phase modulation is desirable, then also the phase component needs to be predistored \([19]\) and should vary the period of the square wave. Specifically, the amplitude should be pre-distorted according to:

\[
a^*(t) = \frac{\arcsin a(t)}{n}
\]  

(6.1)

and the phase needs to be pre-distorted according to:

\[
\phi^*(t) = \frac{\phi(t)}{n}
\]  

(6.2)

By using this type of pre-distortion the components in the Fourier series for the pulsewidth modulated signal becomes:

\[
h^*_n(t) = (-1)^n \frac{4A}{n\pi} a(t) \cos(n\omega_s t + \phi(t))
\]  

(6.3)

Let us make two conclusions. First, by using the pre-distortion in equations (6.1) and (6.2) we can map the baseband signal onto a higher frequency \((\omega_a)\) component. Second, the frequency of which we map the baseband can be the carrier frequency or a subharmonic of the carrier frequency, by having \(\omega_c = n\omega_s\). However, the signal amplitude reduces in proportion to the chosen subharmonic, as can be seen by the \(\frac{4A}{n\pi}\) term in equation (6.3).
There exists two types of pulsewidth modulated signals: uniform PWM and bipolar PWM. Uniform PWM is a two-level signal, for example Gnd and \( V_{dd} \), which implies a DC-component with this type of modulation. One way to eliminate this DC-component is to use bipolar PWM or tri-state PWM [19].

6.3 \( \Delta \Sigma \)-modulation

The main focus in our project has been to investigate the possibilities of using a two level \( \Delta \Sigma \)-modulated signal as input to a switching amplifier. During our research we have mostly found results concerning BP\( \Delta \Sigma \)-modulation and its intended applications, however we start with two suggestions of how to use a low pass \( \Delta \Sigma \)-modulator. Using a low pass \( \Delta \Sigma \)-modulator has the great advantage of requiring a much lower sampling frequency.

6.3.1 Low pass \( \Delta \Sigma \)-modulation

RF PWM by using BP\( \Delta \Sigma \)-modulation [16]

The article ”Generation of RF pulsewidth modulated microwave signals using delta-sigma modulation“ by Keyzer, Uang, Sugiyama, Iwamoto, Galton and Asbeck presents an idea of how to generate a RF-PWM signal (as in [19]).

The \( \Delta \Sigma \)-modulators are intended to be used in conjunction with high speed logic. However, the \( \Delta \Sigma \)-modulators are working on baseband data.
and could therefore easily be built. They are also very accurate. The high speed logic is inaccurate. It is, however, stated that the phase noise produced by the inaccurate processing will be shaped out of band because of the noise shaping in the ΔΣ-modulator.

The idea presented is interesting. Using a PWM-signal as input to a tuned switching amplifier is more feasible than using a ΔΣ-modulated signal. However, this feasibility is reduced when both phase and amplitude are to be modulated, because of lost periodicity. Also the circuitry needed for generating the different pulse widths and pulse delays must be made in at least \( 8f_c \). For a carrier frequency of 2.14GHz, this translates into 17.2GHz. This fact together with poor understanding of tuned switching amplifiers, made us to continue our research with ΔΣ-modulation.

LPΔΣ-modulation and Mixing [13]

![Figure 6.3: Generation of RF digital switching signal, by low pass ΔΣ-modulation [13].](image)

In the book "RF and Baseband Techniques for Software Defined Radio“ by Kenington another idea is given on how to use a low pass ΔΣ-modulator. Again, the reason is to avoid the need for very high speed complex digital circuits. The block diagram is presented in figure (6.3). The author states that the I- and Q-signals could be masked, for example by using \( \{1, 0, 0, 1\} \) and \( \{0, 1, 1, 0\} \), to be able to add them.

We have seen three problems with this architecture. First digital mixing will cause massive unwanted mixing. The problem is worsened by the fact that the out of band noise is amplified by the noise shaping in the ΔΣ modulators. The second problem comes from the addition of the I- and Q-signals. Finally, the third problem regards the DC-level of the ΔΣ-modulators. If Gnd and \( V_{dd} \) are used as the output signals from the ΔΣ-modulators, then a massive DC-component will exist. This DC-component
will be up-converted to the carrier frequency (see further chapter (7.5.3)).

We show a slightly modified architecture in chapter (7.5.2) to get rid of this problem.

### 6.3.2 Bandpass ∆Σ-modulation

The articles below are summarized in a chronological order. Our research traced back to an article from 1998. This article seems to be the first to propose the use of a BPΔΣ-modulator.

**1998 [9]**

In the article “Linear High-Efficiency Microwave Power Amplifiers Using Bandpass Delta-Sigma Modulators” authors Jayaraman, Chen, Hanington Larson and Asbeck, describe what they call “a novel amplifier configuration” using a BP ∆Σ-Modulator and a switching amplifier. The goal stated by the authors of the article is to achieve high efficiency and high linearity. Their “schematic structure” (block diagram) can be found in figure 6.4.

The authors have simulated the amplifier in a double-hetrojunction bipolar GaAs transistor process using SPICE. Images are supplied in the article of an 800MHz sine wave at several points in the signal path, both in time and frequency domains. The final output signal has a noise margin of 40dB as far as it is possible to tell from the picture.

For a closer description of the ΣΔ-modulator the authors refers to a previous publication “BPΔΣ-Modulator With 800 MHz Centre Frequency” by the same authors from 1997. The class-S amplifier used can be seen in figure 6.5.

The authors present a five item list contributing to lowering the Power Added Efficiency:

i. Power losses in the transistors during switching

![System architecture when using a BPΔΣ-modulation for digital RF generation.](image-url)
Figure 6.5: Suggestion of PA to amplify the $\text{BP} \Delta \Sigma$-modulated signal.

ii. Charging and discharging parasitic capacitances

iii. The power needed to operate the BPDSM

iv. Power lost in diodes and transistors associated with their on voltage

v. Power lost in passive components such as the output filter

Items 1 to 3 is commented further upon as highly dependant on the signal clock frequency and a after further analysis an efficiency of 85% for full scale signals is derived when disregarding the BPDSM power demands.

The article concludes that the proposed amplifier technology is very promising.

Although novel at the time there is a fundamental flaw present in this article. The authors have not sufficiently evaluated their suggestion using a more complex signal, utilizing the full 25MHz bandwidth proposed. Such an evaluation would have found that the BPDSM will seldom work with a “full scale signal” and thus producing a less efficiently encoded signal. This in turn leads to an increased loss during output filtering, reducing efficiency dramatically. Furthermore it is doubtful if the 40dB noise margin documented is sufficient for modern applications. The suggestion is not as promising as the authors hope.
In the article “Bandpass ∆Σ Class-S Amplifier” from 2000 the authors, Iwamoto, Jayaraman, Hanington, Chen, Bellora, Thornton, Larson and Asbeck, demonstrates a PA design utilizing a Class-S amplifier and a BP∆Σ-modulator. The goal the authors give is to eliminate the high speed circuitry present in similar PAs based on PWM architectures.

The authors’ suggestion of Class-S amplifier can be seen in figure (6.6), a CMOS inverter with free-wheeling diodes. The function of the free-wheeling diodes is stated to be to provide a current flow when the corresponding transistor is off. The transistors used are discrete, commercially available, off-the-shelf transistors.

For the BP∆Σ modulator design the authors refers to the article “Linear High-Efficiency Microwave Power Amplifiers Using Bandpass ∆Σ Modulators” for the plausibility of such a device. The implementation of the Class-S amplifier is tested at 10MHz using a preprogrammable pattern generator. The authors comment that the modest speed is due to limitations of the pattern generator available.

Finally the authors conclude that the last Class-S amplifier demonstrated an efficiency of 33% and a -40.38dBc intermodulation level at a 26dBm output, 23dBm each for two tones 22kHz apart. This is, according to the authors demonstrated the possibility of attaining both high linearity and efficiency in narrowband RF power amplifier applications.

The authors support the following list of causes lowering efficiency:

i. $R_{DS}$, on resistance of the FETs (p-FET=10Ω)

ii. Rise and fall times (5ns)
iii. Output capacitance

iv. Filter insertion loss (0.8dB)

Finally hints are given on how to increase the efficiency, as summarized below

i. Use devices with faster switching times and lower on resistance

ii. Use a n-channel FET only design, by using baluns

iii. Investigate different filter topologies

The article should be read with the word “narrowband” in mind. The tested bandwidth is 22kHz at 10MHz which today has to be considered as very narrow indeed. It is also a very low transmission frequency compared to modern radio communications. Should the speed be increased beyond the 10MHz the discrete transistors would become less and less viable due to parasitics and should a more complex signal be used, or wider, the efficiency will plummet due to the increasingly heavy filtering (see coding efficiency chapter 7.3). The article is not as promising as the authors believe.

2001 [15]

![Suggested architecture](image)

**Figure 6.7:** Suggested architecture in [15]

In the article “Transmitter Architecture Using Digital Generation of RF Signals” authors Rode, Hinrichs and Asbeck focuses on the same approach as in the first article from 1998. However, the input signal has been changed into IS-95 CDMA, to be more realistic. The architecture is shown in figure (6.7). The architecture is made up by one digital part and one analog part. The digital part is mainly two digital signal processors and a BPΔΣ-modulator. Also there is a digital up converter responsible for up converting the inphase and the quadrature components of the IS-95 CDMA signal. The
analog part is the switching power amplifier and the post filter. Ideally the up conversion is performed according to [15]:

\[ y(t) = x_i(t) \sin(2\pi f_0 t) + x_q(t) \cos(2\pi f_0 t) \]  

(6.4)

In the digital case the up conversion operation is provided by [15]:

\[ y(n) = x_i(n)s_1(n) + x_q(n)s_2(n) \]  

(6.5)

\[ s_1(n) = \{1, 1, -1, -1, \ldots\} \]  

(6.6)

\[ s_2(n) = \{-1, 1, 1, -1, \ldots\} \]  

(6.7)

The two signals \( s_1(n) \) and \( s_2(n) \) is local oscillator signals. The reason they become so simple is because the sampling frequency is exactly four times the desired carrier frequency.

The group has experimentally verified the basic functionality by using the available equipment, which meant that the carrier frequency was limited to 800MHz. The group used a pattern generator capable of 200Mb/s and 16 channels. Further the 16 channels were serialized to achieve an output bit data rate at 3.2Gb/s. As the sampling frequency of the chosen BP\( \Delta \Sigma \)-modulator is \( f_s = 4f_0 \), this implies that \( f_0 = 3.2/4 = 800MHz \).

Finally the group filtered the bit-stream through a Toko four pole dielectric filter. A plot indicates noise at approximately -45dBc. It is, however, stated that the noise floor of the spectrum analyzer masks the suppressed quantization noise in the receive band.

For curiosity we examined the basis functions: by letting \( \phi_1(t) \) and \( \phi_2(t) \) be the time continuous equivalences. Then it can be shown that these two functions are orthogonal according to the orthogonal criterion given in equation (6.8) [3].

\[ \int_0^T \phi_1(t)\phi_2(t) \, dt = 0 \]  

(6.8)

In this case we have

\[ -\int_0^{T/4} 1 \, dt + \int_{T/4}^{T/2} 1 \, dt - \int_{T/2}^{3T/4} 1 \, dt + \int_{3T/4}^T 1 \, dt = 0 \]  

(6.9)

This paper makes no suggestion of how to amplify the resulting \( \Delta \Sigma \)-modulated signal, which is a major design problems. Also we feel unsure about the filtering problem. During our simulations we have had the same problem: it seems as the noise floor masks the suppressed quantization noise, even though we have used ideal filters and components.
In the article "Transmitter Architecture Using Digital Generation of RF Signals" again by Rode, Hinrichs and Asbeck almost the same architecture as in the previous article [15] is presented. The difference is the added coarse gain control via a DSP controlled DC-to-DC converter. It is stated that gain control also can be added via scaling of the digital input to the transmitter, but this method will reduce the SNR. This time the focus is on the implementation of a BPΔΣ-modulator and four topics are covered. These are presented below together with a brief summary.

i. Modulator order
   A modulator order greater than two improves the performance, but is not necessary to meet the specifications. Further higher order modulators increase power dissipation.

ii. Modulator arrangement
    By having $f_s = 4f_0$ it is possible to let all coefficients in the modulator be powers of two. Further, regarding delay elements, they can be incorporated into logical operations by means of pipelining.

iii. Modulator numerical representation
    By simulations it is shown that using anything less than 10-bit representation corrupts the signal. With 12-bit representation satisfactory performance is given. Further, the group used signed (two’s-complement) integers because of simplicity and speed of operations carried out with integers.

iv. Dithering
    The group has used dithering but states that the necessity is not that high because of the pseudorandom CDMA input. The dithering operation is implemented by a linear feedback shift register.
v. Overflow
The modulator should be made insensitive to overflow. Further the whole bit-depth should be utilized as much as possible. It is given some suggestions on how to solve these problems.

vi. Integrator implementation The function implemented is \(1/(1 + z^{-2})\). This integration is a result of applying \(z = -z^2\) to the \(z\)-transform integration \(1/(1 - z)\). The key in the chosen architecture is parallelism. All the operations are carried out independent of each other. For low power dissipation and speed logic the domino logic was selected. Dual-rail was selected because of its completeness. The group has estimated the transistor count for the dual-rail domino logic integrator to 1300.

Again we are left with only the digital part of the system. Even though this article aims at a “complete” transmitter, the amplification part is left out.

2004 [14]
The article "Transmitter utilising bandpass delta-sigma modulator and switching mode power amplifier" by Ketola, Sommarek and Halonen presents an idea to generate a digital RF wave, by BPΔΣ-modulation but focuses on reducing the needed sample rate. The target is a WCDMA base station transmitter. Also, some non-idealities in the modulator are treated.

The need to reduce the sampling frequency is illustrated with the carrier frequency of WCDMA (2.14GHz), which implies a four times higher sampling frequency of 8.56GHz. The idea presented is to reduce the sampling frequency by making use of the images produced by the BPΔΣ-modulator. For a base station the downlink band is 60MHz. The OSR then becomes \(8.56GHz/(2 \times 60MHz) = 71.33\ldots\), which is relatively high. By using an image and a lower sampling frequency the OSR is decreased but may be enough. It is stated that for a BPΔΣ-modulator with sampling frequency \(4f_0\) the images are located at

\[
f_{image} = \frac{f_s}{4} \pm N \cdot \frac{f_s}{2} = (1 \pm 2N) \cdot \frac{f_s}{4}, \quad N = 1, 2, 3 \ldots
\]  

(6.10)

The main problems with using an image, is the reduction of OSR and the attenuation of the images. Theoretically the attenuation is not a problem, because the quantization noise is reduced in a similar manner. In reality it is problem because of other nonidealities, which do not reduce in the same manner. The non-idealities discussed are summarized below:
• Jitter
The timing jitter is one of the most signal quality degrading non-idealities in ∆Σ-modulators. To be able to meet 50dB ACLR, a 8.56GHz sampling clock should suffer at most 0.88ps rms jitter. The conclusion is that not only very high sampling frequency is required but also a very pure clock.

• Periodic jitter
By having digital parts operating in multiple sampling frequencies divided from the output sampling clock, can cause periodic jitter with frequency $f_s/2$. This means that there is more jitter at every other sample period. This may cause phase modulation components.

• Asymmetric rise and fall times
These effects are stated to be similar to the noise caused by jitter. A solution for avoiding asymmetric errors is to use a differential topology.

One specific example of using an image is presented. The chosen image is the first, hence $N=1$ in equation (6.10), which evaluates into a carrier frequency of $\frac{42.14GHz}{3} = 2.8533 \ldots$. The required sampling frequency is hereby reduced by a factor of three. The output spectrum was finally filtered by a modeled 6th order Chebyshev LC-filter.

We feel that this idea has very limited practical use. Image attenuation together with poor coding efficiency implies poor efficiency and low output power. Unfortunately, no comments are given on the chosen filter.
6.4 Conclusion

The main obstacles for using a ∆Σ-modulator:

i. High sampling frequency
   On the orders of 8GHz for WCDMA.

ii. Hard to amplify
    Can not be used in conjunction with a standard tuned switching amplifier.

iii. Poor coding efficiency
    We have found the coding efficiency to be below acceptable levels. Even with a 100% efficient power amplifier the coding efficiency degrades the performance. See further chapter 7.3.

iv. High switching activity
    Using a modulator with a sampling frequency of $4 \times f_c$ will introduce a very high switching activity. See further chapter 7.3.
Chapter 7

Investigation

We were given the task of reading up on recently suggested digital solutions to RF transmission and, if possible, design and simulate such a contraption using the institution’s CMOS design tools. The thesis document was requested to be an educational document with focus on presenting the necessary background information and relevant studies to enable an outsider to rapidly gain insight in the issues regarding linear amplification with switching power amplifiers.

We were given the suggestion to start by creating and investigating a simple Class A amplifier to learn the design and measuring tools. We were also given a loose aim toward a base station intended for 3G. Thus our approach became:

i Class A - Design for Comparison

ii High-Level Design of $\Delta\Sigma$-modulator

iii Coding Efficiency

iv Class E - Amplifier Design

v Digital Amplification - Our Suggestions

vi Result
7.1 Analogue Amplification

In this section we will construct a reference Class-A Amplifier which we can compare the digital amplifier we have innovated.

We start by claiming that a Class-A amplifier, with such high linearity requirements as -45dBc ACLR, without any feedback or other linearization techniques is hard to construct. It requires that the transistor is well backed off from its compression point, which in turn implies that the transistor is operating in a small signal regime rather than large signal (contrary to what would be expected of a power amplifier). Standard design equations are therefore not applicable. For instance, it is not possible to state that the swing at drain is $2 \times V_{dd}$ anymore. If the swing were allowed to reach such magnitude, the resulting linearity would be very poor. As we do not know what signal amplitudes we can tolerate, to achieve -45dBc ACLR, we cannot decide the size of the transistor. However, a larger transistor has larger gain, and gain, loosely speaking, can be traded for linearity, because the input voltage can be made smaller, which decreases the area used in the $V_{ds} - I_{ds}$ characteristics plot. However, increasing the transistor size will decrease the efficiency as the bias current increases.

To be able to compare the results of the $\Delta\Sigma$-modulator amplifier and this Class-A amplifier, we need some common measures. For example, we aim at -45dBc ACLR in a bandwidth of 5MHz relative the carrier. For a larger bandwidth -50dBc should be fulfilled. This is in accordance with 3GPP specifications given in [1]. Second we can estimate the maximum output power of a single sinusoidal for a $\Delta\Sigma$-modulator amplifier. The RMS voltage of a square wave with amplitude $A$ is simply $A$ according to

![Figure 7.1: Final schematic of Class A amplifier](image-url)
equation (4.6). This implies a power of \( \frac{A^2}{R_{load}} \), of which 50% will be filtered away according to the coding efficiency chapter (7.3). As the \( V_{dd} \) should be around 1.2 volt, for this 90nm process, the maximum power could be approximately \( \frac{(0.6)^2}{50} \approx 3.6mW \). Assuming that we do not reach full swing, but only ±0.5 volt, gives us \( \frac{(0.5)^2}{50} \approx 2.5mW \). (50Ω load impedance and 50% coding efficiency are assumed). We try to design the Class-A amplifier for:

i. \( P = 3mW \)

ii. \( \text{ACLR} = -45\text{dBc} \ (5\text{MHz}), -50\text{dBc} \ (>10\text{MHz}) \)

iii. \( R_{load} = 50\Omega \)

Ignoring impedance matching and therefore choosing the load resistor to 50Ω and sweeping the size of the transistor gives us figure (7.2). It can be seen that no larger transistor than approximately 200\( \mu \)m could be chosen for ACLR -45dBc.

![Figure 7.2: Swept width, \( R_l = 50\Omega \)](image)

As can be seen in figure (7.2) it is not possible to reach the wanted output power. Decreasing the output impedance will decrease the gain, which will increase the linearity as explained earlier. It will also decrease the output power, because of the decreased gain. As we wanted more output power, we have to increase the transistor size. The formula seems
7.1 Analogue Amplification

to be: decrease load resistance and increase transistor size. Sweeping the width for $R_L=12\Omega$ gives the result shown in figure (7.3).

![Graph showing swept width, $R_l = 12\Omega$](image)

Figure 7.3: Swept width, $R_l = 12\Omega$

According to figure (7.3) approximately a $600\mu m$ wide transistor could be chosen. With this choice the IM3 suppression is also more than the needed 45 dB. This is desired and possibly even necessary, as the, not yet designed, matching network will degrade the IM3 suppression substantially, as it increases the voltage swing at the transistor drain.

7.1.1 Matching Network

A low-to-high impedance matching network is needed. Structurally simple and giving the freedom to chose a Q-value, is the π-network. The π-network is seen, preceding the load, in figure (7.1). The design equations are given below [5]:

\[
\begin{align*}
\omega C_s &= Q_1 / R_s \quad (7.1) \\
\omega C_l &= Q_2 / R_l \quad (7.2) \\
\omega L &= R_s (Q_1 + Q_2) / (1 + Q_1^2) \quad (7.3) \\
Q_2 &= \sqrt{\frac{R_l}{R_s}} (1 + Q_1^2) - 1 \quad (7.4) \\
Q_1^2 &> \frac{R_s}{R_l} - 1 \quad (7.5)
\end{align*}
\]

Solving the equations for $Q=3$, $R_s=12\Omega$, $R_l=50\Omega$ and $\omega=2$GHz gives $L=1.51\text{nH}$, $C_s=7.13\text{pF}$ and $C_l=4.34\text{pF}$. 
7.1.2 Results

As a final verification we perform a periodic steady state analysis. The result is seen in figure (7.4). As we are using a two-tone input we must look at the output power $3\text{mW}/2=1.5\text{mW}=1.76\text{dBm}$. The lines which gives the IM3 suppression shows almost 50dB.

![Figure 7.4: PSS analysis](image)

Finally a two-tone transient analysis gives the following result (IIP3 is taken from the PSS analysis):

i. $P_{RMS} = 3\text{mW}$

ii. $\text{ACLR} = -44.55\text{dB} (5\text{MHz})$

iii. $R_{load} = 50\Omega$

iv. $\eta_{\text{drain}} = 1.04\%$

v. $IIP_3 = 9.96\text{dBm}$

The last simulation is to verify the WCDMA performance. In this case it is difficult to decide an accurate measure of the ACLR. The performance
seems to be better with WCDMA than with just two tones. For this amplifier IM3 suppression of 45dB is reached with $V_g \approx 0.8 \pm 0.1$. The result of applying a similar restriction to the WCDMA input signal is shown in figure (7.5).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{class_a_wcdma_data}
\caption{WCDMA data applied to the Class A amplifier}
\end{figure}

The results from applying WCDMA data to the class A amplifier becomes:

i. $P_{RMS} = 5.7\text{mW}$

ii. $\eta_{\text{drain}} = 1.98\%$

\subsection{7.1.3 Conclusion}

No more than a couple of percent drain efficiency can be expected from a non-linearized class A amplifier, which fulfills the requirements for 3G.
7.2 High-Level Design of $\Delta \Sigma$-modulators

To be able to simulate the suggested architectures involving switching amplifiers and $\Delta \Sigma$-modulators we needed high-level models. First, we made a model using MatLab and Simulink. Second, we translated this (graphical) model into hardware descriptive language (VerilogA-code).

Constructing a $\Delta \Sigma$-modulator involves a number of steps and choices. Common design variables are [21]:

i. Selection of the NTF

ii. Modulator order

iii. Number of quantization levels

iv. Choice of low pass, bandpass or quadrature modulation

In our case the third choice is fixed. We want two quantization levels, to get a “digital” signal. By requirements such as SNR, complexity and power consumption the modulator order can be chosen. By looking at a plot like figure (5.6), it is possible to choose the suitable order given the SNR requirement. Because of complexity and coding efficiency the order should be kept as low as possible. A second order modulator is often a good choice.

Once the design variables are chosen the next step is to produce a suitable NTF and start simulate the modulator to verify that its functionality.

7.2.1 Using MatLab to produce the NTF

There exists a toolbox\textsuperscript{1} for MatLab, which may be used to produce the NTF. This toolbox is used in this section to exemplify a design. Using the error feedback structure, the major task is to design the loop filter. Remember that when the NTF is constructed, it is easily modified into the loop filter by the equation: $H(z) = 1 - NTF(z)$. The code to produce the NTF is shown below.

```matlab
addpath '/toolbox'
order = 2; %Order of the NTF
OSR = 64; %Oversampling Ratio
opt = 1; %Optimize zero positions
f0 = 0; %DC
Hinf = 1.5; %Max gain at large frequencies
NTF = synthesizeNTF(order,OSR,1,Hinf,f0);
plotPZ(NTF); %Plot the pole zero configuration
```

The resulting pole zero configuration is shown in figure (7.6 a). Note that the zeros are not placed exactly at DC, which means that we have an optimized NTF as explained in section (5.2.3).

\textsuperscript{1}The toolbox is available for download and can be found at [20]
7.2 High-Level Design of $\Delta\Sigma$-modulators

As soon as the NTF is constructed we are ready to transform it into the error feedback loop filter. The function `tfdata()` is used to retrieve the denominator and the numerator to be used in conjunction with the function `freqz()`, which plots the frequency response. The only strictly required code left is: `H = 1-NTF`. The frequency response is found in figure (7.6 b).

```matlab
H = 1-NTF; %Construct the loop filter
[numNTF,denNTF,Ts] = tfdata(tf(NTF));

%Plot the frequency response of the NTF:
figure('name','Lowpass NTF');
freqz(numNTF{1},denNTF{1});

[num,den,Ts] = tfdata(tf(H));
%Plot the frequency response of the loopfilter:
figure('name','Lowpass H = 1-NTF,...
loopfilter for Error feedback structure.');
freqz(num{1},den{1});
```

We now have a Mat-Lab function that represents the loop filter. A very time efficient way to retrieve a realization of the filter is to use `dfilt`, which is Matlab’s digital filter implementation routines. By typing: `Hd = dfilt.df1(num{1},den{1});` a filter object `Hd` is retrieved. The filter is realized using one of the available filter structures. In the example above “direct form I” is used, according to `df1`. Further, the command `realizemdl(Hd);` automatically realizes the filter object into a Simulink object. The realization is shown in figure (7.7).

![Figure 7.6: Pole zero configuration (a) and frequency response (b) for a second order LP NTF.](image-url)
\% Using Matlab's digital filter implementation (dfilt).
\% Construct a direct form I (df1) filter.
\% Type "help dfilt/structures" to get a complete list of structures.
Hd = dfilt.df1(num{1},den{1});

if (isstable(Hd))
    disp('The lowpass filter is stable.');
else
    disp('The lowpass filter is not stable.');
end;

realizemdl(Hd); \% Realize the model into a Simulink object.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.7.png}
\caption{Direct form I realization of the low pass loop filter.}
\end{figure}

As we are also interested in bandpass $\Delta\Sigma$-modulators. The next step is therefore to apply the $z = -z^2$ transformation on the NTF to give it bandpass characteristics instead of low-pass characteristics. We have not found any really good way to do this. The process involves manual replacement of all occurrences of $z$ with $-z^2$. The pole zero configuration of the bandpass NTF is shown in figure (7.8 a). The frequency response for the bandpass NTF is shown in figure (7.8 b). Figure (7.8 b) should be compared with figure (7.6 b) to verify that the properties of a second order modulator are inherited by the fourth order bandpass modulator, after this transformation.
% The code below is written to convert a second order lowpass NTF into a fourth order bandpass, by applying the transformation \( z = -z^2 \).
\[
z = \text{zpk('z')}; \quad \% \text{Make } z \text{ a symbolic variable}
\]
\[
% \text{NTF} \% \text{Prints the NTF so that we can transform it.}
********\text{REMEMBER TO CHANGE ACCORDING TO THE NTF} ********
\]
\[
\text{NTF}_\text{BP} = \frac{(-z^2)^2 - 1.999 \cdot (-z^2) + 1}{(-z^2)^2 - 1.225 \cdot (-z^2) + 0.4413};
\]
**********\text{REMEMBER TO CHANGE ACCORDING TO THE NTF} **********
\[
H_\text{BP} = 1 - \text{NTF}_\text{BP};
\]
\[
[\text{numNTF}_\text{BP},\text{denNTF}_\text{BP},Ts] = \text{tfdata}(\text{tf(NTF}_\text{BP}));
\text{figure(}‘name’,’Bandpass NTF’\text{);}\quad \%	ext{Frequency response}
\]
\[
[\text{num}_\text{BP},\text{den}_\text{BP},Ts] = \text{tfdata}(\text{tf(H}_\text{BP}));
\text{figure(}5\text{);}\quad \%	ext{Frequency response}
\]

Figure 7.8: Pole zero configuration (a) and frequency response (b) for the bandpass NTF.

Again, the final step is to realize the filter with a given structure, in order to simulate it in Simulink. Once again we have chosen the “direct form I structure”, to be able to compare the realizations.

% Using MatLab's digital filter implementation (dfilt).
% Construct a direct form I (df1) filter.
% Type “help dfilt/structures” to get a complete list of structures.
% if(isstable(Hd_BP))
% disp(’The 4th order bandpass filter is stable.’);
% else
% disp(’The 4th order bandpass filter is not stable.’);
% end;
realizemdl(Hd_BP); \%	ext{Realize the model into a Simulink object.}
7.2.2 A 4th-order BPΔΣ-modulator using Simulink

As the required loop filter is automatically generated, it is now a simple task to build a model in Simulink to verify the functionality of a ΔΣ-modulator. One example of a test-bench in Simulink is found in figure (7.9).

![Diagram of a 4th-order BPΔΣ-modulator in Simulink](image)

**Figure 7.9:** Testbench of fourth order BPΔΣ-modulator in Simulink

The transient analysis result is shown in figure (7.10). The figure consists of five subplots. The plot showing the filter input and the plot showing the difference node output is illustrated just to make a point about the complexity of what is going on inside the loop. The input and the filtered output looks pretty much the same, except for a phase shift and some reduction in amplitude. The good thing is that both these errors seem to be constant, which we can tolerate. Finally looking at the digital RF wave, it is a pulse-width modulated square wave as was desired.

Figure (7.11a) shows an FFT made on the modulator output, to show the spectral characteristics and especially the noise shaping. The FFT uses $2^{15} = 32768$ points and a sampling frequency of $4 \times 2^{32} \approx 17\text{GHz}$.

As can be seen by the FFT, figure (7.11a), noise shaping is happening as the quantization noise is filtered at $f_s/4 = 2^{31}/4 \approx 1.07\text{GHz}$. The FFT analysis also displays many large tones over the entire spectrum. We are not entirely sure about this phenomena, but we suspect that the tones are so called idle tones. Idle tones are found in a low-pass modulator when dc is used as input [21]. In the bandpass version we suspect that idle tones may be generated when an input with constant frequency is used, such as in this example. The solution to this problem is called dithering [21]. Dithering typically means the act of introducing some random signal into the modulator [10]. The most suitable place for this addition is prior the quantizer, so that the dithering becomes noise shaped as well. Note that dithering is not used to mask out the noise, instead it is used to lower the probability for idle tones to occur, by breaking up the tones [10]. We have introduced noise as can be seen in figure (7.9). With the use of dithering the
7.2 High-Level Design of $\Delta\Sigma$-modulators

7.2.3 A 4th-order BP $\Delta\Sigma$-modulator using VerilogA

To be able to simulate our $\Delta\Sigma$-modulator in Cadence, a high-level model was implemented using VerilogA. The goal is to make it possible to easily build a $\Delta\Sigma$-modulator in VerilogA.

**The loop**

As can be seen in figure (7.12), illustrating the error feedback structure, the operations inside the loop are simple if we ignore the loop filter. Seen as program instructions we can make a list of necessary instructions:

i. Sample the input

ii. Add the input $u(n)$ with the filter output to get $x(n)$

iii. Quantize $x(n)$ to get the output $y(n)$

iv. Subtract $y(n)$ from $x(n)$, to build the filter input

These steps can be seen in the VerilogA code snippet below. The sampling takes place on the rising edge of the clock.

---

**Figure 7.10:** Representative transient analysis
Figure 7.11: FFT analysis of digital output, (a) no dithering, (b) with dithering.

Figure 7.12: Error feedback structure.

#define (cross(V(clk)-Vdd/2, +1)) begin
  in_real = V(signal_in); //i. Sample input.
  sum1 = in_real + sum_output_nodes; //ii. Add input and filter output.
  //iii. 2-level DAC:
  if (sum1>=0) begin
    DAC_out = 0.5;
    mod_out_r = 1; //Output from modulator.
  end else begin
    DAC_out = -0.5;
    mod_out_r = 0; //Output from modulator.
  end
  diff1 = sum1 - DAC_out; //iv. Input to filter.

The filter

The next step, the fifth, is to create a new filter output, given the filter input diff1. The filter is easily converted into VerilogA code given the structure shown in figure (7.13).

In the code in_delayx corresponds to a delay element on the input side of the filter whereas out_delayx corresponds to a delay element on the output side of the filter. The constants a,b,c and A,B,C,D are defined as
parameters and equal the multiplication coefficient. The filter code should be self explanatory.

```plaintext
//------------FILTER------------------/
//Left (input) side:
in_mul1 = in_delay1 * a;
in_mul2 = in_delay2 * b;
in_mul3 = in_delay3 * c;
in_delay3 = in_delay2;
in_delay2 = in_delay1;
in_delay1 = diff1;
sum_input_nodes = in_mul1 + in_mul2 + in_mul3;

//Right (output) side:
out_mul1 = out_delay1 * A;
out_mul2 = out_delay2 * B;
out_mul3 = out_delay3 * C;
out_mul4 = out_delay4 * D;
sum_output_nodes = sum_input_nodes - out_mul1
                 - out_mul2
                 - out_mul3
                 - out_mul4;
out_delay4 = out_delay3;
out_delay3 = out_delay2;
out_delay2 = out_delay1;
out_delay1 = sum_output_nodes;
```

There is one difference with the filter given in figure (7.13) and the code above. In the code there is only three delay elements on the input side, whereas in the figure there is four (the first delay element consist of two cascaded). To get the correct results, in the correct order, we had to take away one delay element from the code. We found the solution, by analyzing in detail the values in every node at each time step in Simulink.

The full implementation of the ΔΣ-modulator can be found in appendix A.
7.3 Coding Efficiency

During our investigation of using a ΔΣ-modulated signal as input to a switching amplifier and now that we have a high-level model, we thought of how efficient it is to represent a signal the way a ΔΣ-modulator does. This chapter gives the result from this investigation.

7.3.1 Switching Activity

In high speed digital circuits high switching activity is a major contributor to power losses, often referred to as dynamic power dissipation:

\[ P_{\text{dyn}} = C_L V_{DD}^2 f_{0\rightarrow1} = C_L V_{DD}^2 P_{0\rightarrow1} f \]  \hspace{1cm} (7.7)

In the equation above \( P_{0\rightarrow1} \) is the probability for a zero to one transition and \( f \) is the clock frequency. (Lowering the supply voltage is of course not an option for a power amplifier, even though it would be the most efficient way to reduce the dynamic power.) The question is what switching activity that can be expected from a ΔΣ-modulated signal? A brief investigation of the number of transitions in the ΔΣ-modulated signal compared to the sampling clock frequency is presented in [12]. We have not redone this investigation, a picture providing the results is shown below:

![Figure 7.14: Average pulse frequency of ΔΣ-modulators [12].](image)

In the case of a \( f_s/4 \) modulator, which means that the sampling frequency is four times the signal frequency, the average pulse frequency is close to unity for a single tone. Compared to the frequency required in a tuned switching power amplifier the switching frequency is four times higher. As
could be expected. By using a $f_s/3$ modulator the situation gets a little bit better. This time the "optimal" average switching frequency is simply: $1/3f_s \approx 0.33f_s$ (or less). As can be seen, a single tone requires almost $0.6f_s$, i.e. $3 \times 0.6 = 1.8$ times more than the switching frequency required for standard tuned switching power amplifier.

In addition the power added efficiency will suffer because of high switching activity, as the input gate-source capacitance has to be driven.

### 7.3.2 Coding Efficiency

As the process of noise shaping also involves amplification of out of band quantisation noise, the output filter has to filter away a lot of energy. The signal power, compared to the total output power, is thus small. The term defining this relationship is named *coding-efficiency* and is given in equation (7.8) [12].

\[
\eta = \frac{P_{\text{before filter}}}{P_{\text{after filter}}} \quad (7.8)
\]

As a BP∆Σ-modulator should be avoided, because of high switching activity, we started to measure the coding efficiency for different low pass ∆Σ-modulators. The reason was to see if there are any advantages, in terms of coding efficiency, of using a certain modulator. The parameters we have changed are: order, OSR and input signal. The result is presented in table (7.1). (Actually, we are looking at the coding efficiency as a function of SNR.)

<table>
<thead>
<tr>
<th>Order</th>
<th>OSR</th>
<th>1 sin[%]</th>
<th>2 sin[%]</th>
<th>3 sin[%]</th>
<th>Q-data[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64</td>
<td>49</td>
<td>35</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>49</td>
<td>35</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>49</td>
<td>35</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>OF\textsuperscript{3}</td>
<td>35</td>
<td>27</td>
<td>22</td>
</tr>
</tbody>
</table>

**Table 7.1:** Coding efficiency for different low pass modulators. 2)-3dBFS 3)OF:Overflow

At first, after rounding the values, there seems to be no difference between the modulators. However, there is a very important difference between higher order modulators and lower order modulators, namely stability. Lower order modulators are more stable then higher order modulators.
This means that larger input signals can be used with lower order modulators, which increases the utilization of the dynamic range and the coding efficiency. By restricting the input signal to -6dB full scale, which makes all \( \Delta \Sigma \)-modulators stable, the coding efficiency stays the same. As the coding efficiency stays the same, for small enough input signals, we summarize the results for low pass \( \Delta \Sigma \)-modulators in table (7.2). Included, is also some statistics for the input signal used. As can be expected the absolute mean value, \( |\bar{x}| \), is a good indication of available coding efficiency.

| Input=x | \( |\bar{x}| \) | \( var(x) \) | Std. dev. | Coding eff. [%] |
|---------|-------------|-------------|-----------|--------------|
| 1 sin   -6dBFS | 0.159       | 0.031       | 0.177     | 35           |
| 2 sin   -6dBFS | 0.113       | 0.020       | 0.142     | 27           |
| 3 sin   -6dBFS | 0.090       | 0.010       | 0.123     | 22           |
| Q-data  -6dBFS | 0.067       | 0.015       | 0.084     | 13           |

Table 7.2: Coding efficiency for LP\( \Delta \Sigma \)-modulators.

Although not of primary interest, we have also measured the coding efficiency for a fourth order \( f_s/4 \) BP\( \Delta \Sigma \)-modulator. A fourth order BP\( \Delta \Sigma \)-modulator has the same properties as a second order low pass \( \Delta \Sigma \)-modulator, in terms of SNR. The OSR for a 60MHz 3GPP \((f_c = 2.14GHz)\) transmission band becomes:

\[
OSR = \frac{f_s}{2f_b} = \frac{4 \times f_c}{2f_b} = \frac{4 \times 2.14 \cdot 10^9}{2 \cdot 60 \cdot 10^6} = 71.33 \ldots
\] (7.9)

From equations (5.20) and (5.22) we know the theoretical maximum available SNR for a second order and fourth order BP\( \Delta \Sigma \)-modulator respectively. The theoretical maximum available SNR for a second order BP\( \Delta \Sigma \)-modulator becomes:

\[
SNR_{2nd\text{order}} = 6.02 + 1.76 - 12.9 + 30 \log(71.33) \approx 50.5dB
\] (7.10)

and for a fourth order BP\( \Delta \Sigma \)-modulator:

\[
SNR_{4th\text{order}} = 6.02 + 1.76 - 12.9 + 50 \log(71.33) \approx 87.5dB
\] (7.11)

To guarantee an adjacent channel leakage power ratio (ACLR) limit of 50dBc [1] a fourth order BP\( \Delta \Sigma \)-modulator should be chosen. A higher order than this is should not be used, because of stability as explained previously.

Again a signal with an envelope having high absolute mean, gives higher coding efficiency.
### 7.3 Coding Efficiency

| Input=x | $|\bar{x}|$ | $\text{var}(x)$ | Std. dev. | Coding eff. |
|---------|----------|----------------|-----------|-------------|
| 1 sin -6dBFS | 0,156 | 0,031 | 0,177 | 38 |
| 2 sin -6dBFS | 0,101 | 0,016 | 0,125 | 21 |
| 3 sin -6dBFS | 0,076 | 0,010 | 0,102 | 16 |
| IQ-data -6dBFS | 0,067 | 0,007 | 0,084 | 13 |
| IQ-data -3dBFS | 0,099 | 0,014 | 0,119 | 19 |
| IQ-data -1dBFS | 0,124 | 0,022 | 0,150 | 24 |

Table 7.3: Coding efficiency for a fourth order BPΔΣ-modulator, designed for OSR=64.

#### 7.3.3 Conclusion

We have made three important observations from this investigation:

i. For both LP and BP modulators the order should be kept as low as possible.

ii. For BPΔΣ-modulators the switching activity should be minimized. This is done by lowering the sampling frequency. However, lowering the sampling frequency will of course lower the OSR, which implies lower SNR.

iii. Poor coding efficiency will degrade the overall performance, no matter how efficient the amplifier is. Researching different filter topologies and ways to utilize the filtered energy might be one solution to this problem.

As the solution based on direct RF-signal generation with a BPΔΣ-modulator suffers both from poor coding efficiency and high switching activity, we think that it should be avoided.
7.4 Class E Design Example in CMOS 90nm

In a previous section we stated that Class E amplifiers were not suitable for using together with a ΔΣ-modulator. Nevertheless we have built one and tested the concept to demonstrate the point further as well as show the outstanding properties these amplifiers after all do possess.

Even though the transistor is not needed to be very fast we have chosen to use a new CMOS technology for this design, an example of a design in the older 0.35μm process can be found in [6] and [24]. The schematic of the amplifier is shown in figure (7.15). By assuming a breakdown voltage of 1.4 volt (may be a little bit too high) and using no safety factor, \( V_{DD} \) is evaluated to 0.4 volt. Further, we chose the load resistance and the Q quite arbitrarily to 3Ω and 7 respectively. As the supply voltage is so low we decreased the load resistance far from 50Ω to be able to get any output power. Further the Q value is chosen to be a representative value; neither too large nor to small. However, a smaller Q could be beneficial for our intended application. Now, evaluating equations (4.23) to (4.25) gives us values for C1, C2 and L2. L1 is chosen to 20nH, which implies 250Ω at 2GHz. The design so far is summarized in table (7.4).

As we are working with full custom design the size of the transistor can be chosen. We swept the size and looked at the amplifier’s drain voltage and current. We also plotted the drain efficiency. We choose a very large transistor (4mm), where the drain efficiency reached a maximum and the waveforms looked tunable. We know from figure (4.5a) how a typical mistuned class E amplifier looks like. The drain current of this amplifier is about 150 mA. By analyzing the \( I_{DS} - V_{DS} \)-characteristics we found that this transistor can handle around 300mA to 400mA. The waveforms of the
untuned amplifier are shown in figure (7.16).

![Waveforms](image)

**Figure 7.16**: Waveforms for initial untuned design

Now we have to consult the tuning guide given in appendix (C). This tuning guide gives consultation of how to adjust C1 and C2 according to the $V_{DS}$ slope at transistor turn on. As can be seen in figure (7.16) the slope seems to be zero. According to the tuning guide we should thus decrease C1 and increase C2. A variable, x, is defined. This variable is subtracted from C1 and added to C2. The result of sweeping x from 0 to 1p is shown in figure (7.17). Choosing x to be 0.27p gives us the highest drain efficiency and another $V_{DS}$ slope.

Again examining the $V_{DS}$ slope at transistor turn on in detail gives us information what to do next. The slope is now negative. Further it seems as the rightmost and uppermost case, in the tuning guide, is the one corresponding to our case. This means that C1 should be decreased. This

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>7</td>
<td>“Reasonable”</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>0.4V</td>
<td>1.4/3.56 (Choosen large to increase P)</td>
</tr>
<tr>
<td>R</td>
<td>3Ω</td>
<td>Choosen small to increase P.</td>
</tr>
<tr>
<td>C1</td>
<td>5.532pF</td>
<td>Eq. (4.23)</td>
</tr>
<tr>
<td>C2</td>
<td>3.849pF</td>
<td>Eq. (4.24)</td>
</tr>
<tr>
<td>L1</td>
<td>20nH</td>
<td>$Large \ X_{L1}@f_c$ (2\pi f L_1 = 250)</td>
</tr>
<tr>
<td>L2</td>
<td>1.910nH</td>
<td>Eq. (4.25)</td>
</tr>
</tbody>
</table>

**Table 7.4**: Summary of design before tuning.
Investigation

Figure 7.17: Decreasing C1 and increasing C2

time the variable, deltaC1 is defined. This variable is of course subtracted from C1. The result of sweeping deltaC1 between 0 and 1p is shown in figure (7.18).

Figure 7.18: Decreasing C1

Finally, one more adjustment is needed to reach 88% drain efficiency, were we stop. This time we are back at the first case, with zero slope
at transistor turn on. Now we know that this implies that $C_1$ should be decreased and $C_2$ should be increased. A new variable, $y$, is defined. Sweeping the variable $y$ from 0 to 1p gives the results in figure (7.19) (only three values are presented in the figure for clarity). The chosen value is shown as the curve in the middle.

![Figure 7.19: Decreasing C1 and increasing C2](image)

The different steps in the design are summarized in table (7.5). For this CMOS process the difference from the calculated values are quite substantial. $C_1$ is only 38% of the calculated value and $C_2$ is 125% of the calculated value. The output is, however, sinusoidal and the efficiency is very close to 88%.

<table>
<thead>
<tr>
<th>Design</th>
<th>$C_1$ [F]</th>
<th>$C_2$ [F]</th>
<th>$L_1$ [H]</th>
<th>$L_2$ [H]</th>
<th>$R$ [Ω]</th>
<th>$W_n$ [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.532p</td>
<td>3.849p</td>
<td>20n</td>
<td>1.910n</td>
<td>3</td>
<td>4m</td>
</tr>
<tr>
<td>2</td>
<td><strong>5.262p</strong></td>
<td><strong>4.119p</strong></td>
<td>20n</td>
<td>1.910n</td>
<td>3</td>
<td>4m</td>
</tr>
<tr>
<td>3</td>
<td><strong>2.662p</strong></td>
<td><strong>4.119p</strong></td>
<td>20n</td>
<td>1.910n</td>
<td>3</td>
<td>4m</td>
</tr>
<tr>
<td>4</td>
<td><strong>1.978p</strong></td>
<td><strong>4.803p</strong></td>
<td>20n</td>
<td>1.910n</td>
<td>3</td>
<td>4m</td>
</tr>
</tbody>
</table>

Table 7.5: Summary of design procedure. Changes are bold fonted.

### 7.4.1 Stationarity and Quasi-Stationarity

We have here chosen to evaluate two properties of the designed Class-E amplifier. The first is stationarity, which we call the behaviour of changing
the input signal and waiting for the stationary response. More specifically, we have swept the input signal with different duty cycles and evaluated the results at the end of each simulation. The second property, which we call quasi-stationarity, is referred to as evaluating the transient response directly when changing the duty cycle of the input signal.

**Stationarity**

To examine the ability to perform amplitude modulation by varying the duty cycle of the input signal we have swept the input duty cycle from 10% to 90%. The result is presented in figure (7.20). As can be seen additional predistortion needs to be done to achieve a linear relationship between the width of the pulse and the amplitude. This test is, however, not representative for a real input signal. In a real case stationarity can not be assumed.

![Figure 7.20: Stationary response when sweeping duty cycle from 10% to 90%](image)

**Quasi-Stationarity**

Instead of changing the duty cycle and examining the stationary response, we look at the transient response when changing the duty cycle. As can be seen in figure (7.21) some 50ns is needed to reach the stationary amplitude after changing the duty cycle. As the carrier frequency is 2GHz this means that about 100 RF cycles are needed to reach stationarity. If we were to change the duty cycle every RF cycle we would be in trouble. This amplifier is tuned to an operating frequency. A tuned circuit implies resonance or
close to resonance with a lot of energy going back and forth. Intuitively, we should not try to change this oscillation of energy in every RF cycle. Unfortunately, this often happens with $\Delta\Sigma$-modulation.

![Figure 7.21](image.png)

**Figure 7.21**: Quasi stationary response of 10% and 25% duty cycle.

**Application to $\Delta\Sigma$-modulation**

As was indicated in the preceding section the way a $\Delta\Sigma$-modulator represents a signal, is not ideal for a tuned circuit like a class E amplifier. In figure (7.21) we saw an indication of how many RF cycles that are needed for each new duty cycle. Just applying the $\Delta\Sigma$-modulated output to the amplifier results in more or less useless output, as can be seen in figure (7.22).
Investigation

Figure 7.22: Result of applying a $\Delta\Sigma$-modulated pulse train

7.5 Digital amplification

After making a survey of previously published approaches of linearising a switching amplifier, we gained some insight in the basic problem of amplifying a digital RF signal. The major advantage of a tuned switching amplifier, namely the LC network that stores energy between cycles, is also a large disadvantage. This fundamental energy property restricts the type of signal that can be amplified. In figure (7.23) is a simple demonstration of how a rapid change in amplitude is resisted due to the built up energy of the LC tank.

In addition to the problem of amplification, also creation of a digital representation of the analogue signal is a problem, simply because the clock speeds of the necessary circuitry exceeds those manageable today. For class D and E the continuous signal has to be represented (or created) by only a two level digital signal, while a general class S amplifier could be implemented using several levels (at the cost of complexity and power leaks).

7.5.1 PWM

All amplitude modulated RF signals have, compared to the switching frequency, a slowly changing envelope. This is the reason why a PWM-signal could (theoretically) be applied to a tuned switching amplifier. However,
Figure 7.23: The fast envelope changes in the upper example cause the output to stagger behind the signal, in the slower example the tank copes better but not perfectly.

a problem arises when both amplitude and phase modulation is desired, as in QPSK used in WCDMA. Figure (7.24) (from chapter (6.3.1)) illustrates the problem: the resulting pulse train’s repetition cycle does not always match the resonant frequency of the LCR circuit (required for e.g. class D and class E operation). However, the changes will occur much more seldom, decided by the modulation scheme. BP-PWM will at least be more stationary than a $\Delta\Sigma$-modulator.

The main reasons for not investigating BP-PWM further was that the realization, proposed in figure (7.24), requires circuitry at $8f_c = 8 \cdot 2.14GHz = 17.12GHz$ and we where very much inspired by all papers using $\Delta\Sigma$-modulation.

7.5.2 $\Delta\Sigma$-modulation
At first glance $\Delta\Sigma$-modulation is in all regards at least as bad as PWM, it is necessary to over-sample the signal, which implies high switching activity, and the generated digital pulse train is not periodic. However, $\Delta\Sigma$-modulation has the nice property of keeping a spectrally correct represen-
We were well aware of the great drawback with ΔΣ-modulator based design, namely the substantial filtering that has to be done of the switched signal to retrieve the desired output. Even if we could find a way to adapt the signal to a class E or class D amplifier the efficiency would suffer because of the output filter. Regardless of this, the design is more feasible as it is realizable.

Suggested Architectures

Two architectures are suggested below. The common blocks for both architectures are the output filter and the Class-S amplifier. The architectures have been modeled in Simulink, where the Class-S amplifier have been replaced by an ideal multiplication. Further, the output filter is in both cases an ideal 6th-order bandpass Chebyshev-II filter, with a bandwidth of 5MHz and a stop band attenuation of 40dB. The filter is chosen because of its steep characteristic at the cut of frequency. Both the LPΔΣ-modulator and the BPΔΣ-modulator should exhibit the same performance.

Suggestion I:
The first architecture investigated, could be used to achieve double sideband amplitude modulation, by using a LPΔΣ-modulator with a switching amplifier. The architecture is illustrated in figure (7.25 a). The baseband data, made up of two sinusoidal of 1MHz and 2MHz respectively, is ΔΣ-modulated into a two level representation. Ideally, the operation of the amplifier, is to multiply the ΔΣ-modulated signal to achieve a frequency
7.5 Digital amplification

shift. With this ideal behaviour the output from the switching amplifier looks like the subplot Switched RF in figure (7.26). In this case the ∆Σ-modulator output does not contain a DC-component as the two output levels are positive and negative. A DC-component would be up-converted, by the multiplication, to the carrier frequency. However, to achieve multiplication with ±1 is not trivial (see section (7.6)).

![Diagram](image)

**Figure 7.25:** DSB AM by power supply switching.

The modulator operates at 32 times the required sampling frequency, i.e. \(5 \cdot 2 \cdot 32 = 320\text{MHz}\). (The bandwidth is chosen quite arbitrarily to 5MHz, according to a single WCDMA channel.) The theoretical maximum SNR according to equation (5.22) becomes approximately 70dB. This is 25dB more than the required 45dB ACLR according to 3G specifications, to allow for lower peak to average power ratio signals than a single sinusoidal. Figure (7.27) shows the frequency response of the architecture proposed in figure (7.25). The noise shaping of the ∆Σ-modulator is clearly visible.

In figure (7.25 b) an idea is given on how to achieve SSB AM modulation. This idea is not further investigated, as it required knowledge of power combiners. However, it has the great advantage of not creating an image. **Suggestion II:**

Our suggestion, to be able to use quadrature signaling, is to up-convert the in-phase and quadrature components of the baseband to an intermediate frequency. The intermediate frequency is then BP∆Σ-modulated. The architecture is illustrated in figure (7.28).

One drawback of this architecture, is the fact that the RF-band, which we want to transmit, is located at \(f_c \pm f_{\text{intermediate}}\). This means that we have to filter away half of the signal energy. One possible advantage is the
Investigation

ability to use “ordinary” Gnd and Vdd logic. This, unfortunately, means that a large DC-component will be up-converted to $f_c$, but as the RF-band is located at $f_c \pm f_{\text{intermediate}}$ this term can be filtered. Additionally it can cause problems with images, as explain in (7.5.3). No matter, much of the output power will be lost in the filter.

To make the signals more easily visible in the frequency domain, the intermediate frequency is chosen to 100MHz, which implies that the sampling
rate is 400MHz for the BPΔΣ-modulator. With 5MHz signal bandwidth, the OSR becomes \( \frac{400}{2f_b} = \frac{400}{2 \cdot 5} = 40 \) instead of 32 as in the previous example. The frequency response can be seen in figure (7.29).

Finally, to excite the system with some realistic waveforms we have done a simulation with WCDMA-data. We have used Matlab to produce the in-phase and quadrature waveforms. In this example only 50µs of data have been used. The resulting frequency response is shown in figure (7.30).
7.5.3 Signal Investigation

With the architecture given in figure (7.28) in mind. Depending on the choice of carrier and intermediate frequencies it is possible to arrive at different end signal spectrum. Not all combinations are suitable due to the placement of the signal mirrors. After choosing an intermediate frequency the carrier has to be chosen so that the mirrors are prevented from causing distortion of the signal. Let us look at a specific example with an intermediate frequency of 200MHz, and a carrier frequency of 2GHz. The baseband data signal, seen in figure (7.31), is chosen to 5 sine waves to be easily recognizable in the frequency domain.

After the first up-mixing and SSB processing as well as $\Delta \Sigma$-modulation the signal spectrum looks like figure (7.32). The major reason for SSB processing is in this case to easier visualize the problem of images. According to equation (6.10), repeated here for convenience, images of the $\Delta \Sigma$-signal are found every 400MHz ($f_s = 800MHz$). It should also be noted that the spectrum is mirrored in $f_{image}$ for $N = 1, 3, 5, \ldots$, i.e. odd $N$.

$$f_{image} = \frac{f_s}{4} \pm N \cdot \frac{f_s}{2} = (1 \pm 2N) \cdot \frac{f_s}{4}, \quad N = 1, 2, 3, \ldots$$

First we simply try to “see” what happens. The carrier wave, being a square wave lacking DC component, is seen in figure (7.33 a). A square wave with a DC component is seen in figure (7.33 b). The result of mixing the square wave with the $\Delta \Sigma$-output can be predicted by simply superimposing instances of the $\Delta \Sigma$-signal over each of the components of the square wave. Since the mirrors of the $\Delta \Sigma$-signal can be found at regular intervals, it
Figure 7.31: This is the original signal in (a) time and (b) in frequency domain.

Figure 7.32: A broad view of the intermediate signal’s spectrum
Figure 7.33: Square Carrier, (a) without a DC component, (b) with a DC component.
is possible to choose the interval between two components of the square carrier so that the mirror images are superimposed on each other. To be fully correct an infinite number should be considered, but for a clearer image, only a few have been included as in figure (7.34).

![Figure 7.34: One copy of the intermediate spectrum is superimposed over each of the carrier’s components, in (a) only the major two, in (b) only the major three.](image)

A more formal investigation can be done by listing the images from the $\Delta\Sigma$-signal and the images created by the harmonics in the square wave. There will be in band distortion cause by either up-mixing or down-mixing, depending on which signal frequency is wanted. Figure (7.35) shows the result of our example. In the leftmost column the images in the $\Delta\Sigma$-signal are listed, i.e. the images a DC-component in the carrier would cause. In the column labeled $f_c = 6\,GHz$ are the images caused by mixing the first harmonic in the square wave with the $\Delta\Sigma$-signal. As already been mention, the $\Delta\Sigma$-images are mirrored in $f_{image}$ for $N = 1, 3, 5, \ldots$, i.e. odd $N$. This is the reason for the up-mixing distortion, in the case of a DC-component in the RF carrier as can be seen in figure (7.36 a). Further it is also the
reason for not causing down-mixing distortion as can be seen in both figure (7.36 a) and (7.36 b). (Should 1.8GHz be the final frequency, down-mixing distortion would occur.)

**Figure 7.35:** Images created by mixing a $\Delta \Sigma$-signal with a square wave.

In the case where the square carrier has a DC component it is apparent there will be in band distortion of the final signal, hence the requirement of no DC component in the carrier. We suggest doing this simple analysis before choosing carrier and intermediate frequencies. It is also possible to manipulate the frequencies to put the mirrors of the DC component in another part of the spectrum, but we believe it simplest just to avoid having such a component in the first place.

The mixing of the square wave is as such not an amplification or a real amplifier, but by sizing the transistors we can create a larger current capacity. The efficiency will not exceed that of an inverter and will not have the benefits of class E or D operation, but it will work (which the others will not).
Figure 7.36: In a) Display of how the DC component in the carrier causes in band distortion by misplacing a mirror. In b) no DC component is present.
7.6 Switching PA Stage

The signal requirements have turned the PA into a signal multiplier or mixer. Based on the desired behaviour a combinatorial net is constructed.

The signals A and B can experience binary values while the signal U can be either $V_{DD}$, $V_{SS}$ or GND. Since the recommended voltage of the 90nm transistors is 1.2V, $V_{DD}$ and $V_{SS}$ is +/-0.6V respectively. Please note that the 90nm process is in no way optimal for PA implementation, normally PAs or any power intense devices are constructed in a GaAs process capable of sustaining higher voltages, resulting in higher output power. The reason for using the 90nm process is mere convenience as that design kit was available without further ado.

A LPΔΣ-output
B Square wave
U Output

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>GND</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>GND</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>VSS</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>VDD</td>
</tr>
</tbody>
</table>

To our knowledge there is no official systematic way of realizing a multiple level CMOS net, there are also different considerations for which transistor type to use, we offer this implementation with the motivation that it works.

Added to the gate are capacitors of the same size as the transistor gate capacitances, but with the inverse of the signal of the gate applied over them. This attempts to compensate for signal leak effects (feed through) through the gates of the transistors. The necessity of these capacitors can be discussed but since we attempt to size the transistors to increase the current throughput, the gate capacitances can be noticeable (hundreds of femto Farads).

The final stage of the output chain has been determined by parameter sweeps. The “optimal” size (all other ideal), resulting in the highest power output, was found to be around 1000$\mu$ for PMOS and around 400$\mu$ for NMOS. To drive transistors of that size the preceding transistor sizes are increased by a factor 3 for each step, buffers are used where the transistors can not be sized.

The linearity of the system simulated in Cadance displays some deviation compared to Simulink simulations but still viable results. It is safe to assume further improvements to the transistor network would produce better linearity.
Normalized spectrums before and after filtering for two tones and a WCDMA signal follows, figures 7.40 and 7.41.

Further evaluation can be found in chapter 7.3.
Figure 7.38: Capacitances added matching the size of the gate capacitances of the transistors

### 7.7 Comparison

By comparing the performance of the digital suggestion to a Class-A amplifier it is possible to get some feedback on how well we have done. The implementation of the proposed up-mixing method in chapter 7.6 is not optimally implemented in a 90nm process and neither is the Class-A amplifier. However, they still demonstrate the two concepts to satisfaction and the comparison of the two speaks of their individual merits.

Since the guiding criteria has been linearity, power and efficiency those three has been measured. We have tried to set a realistic yet challenging goal for each of the three.

To define a realistic power goal we looked at the digital setup where 0.6V is driving a load of 50 Ohms, the expected coding efficiency for a single sinus has been shown in chapter 7 to be around 50% yielding $0.6^2/50 \times 0.5 = 3.6mW$. It is reasonable to expect some losses due to the non-ideal nature of the transistors we decided that 3mW dual tone input was suitable.
We have chosen our linearity goal as part of those dictated by the 3GPP, an ACLR of -45dB and -50dB at 5MHz and 10MHz outside the channel respectively. Note that this is in no way meant as a complete proof of the devices ability to function as 3GPP transmitters, but rather as a relevant measurement. The specification of 3GPP is much more detailed and testing is more rigorous than permitted by the scope of this thesis.

Efficiency is something that has been continuously sacrificed to achieve foremost linearity but also output power. It is possible to argue that the efficiency goal has been relaxed or secondary to that of linearity. This may seem rather peculiar since it was the benefits of increased efficiency that led to the thesis in the first place. However, without sufficient linearity, efficiency lacks meaning (a distorted signal is useless no matter how efficiently it was transmitted), hence the priority of the linearity with efficiency as a secondary goal. The remaining goal is of course that the digital device should be more efficient of the two.

### 7.7.1 Output Power

It was found necessary to slightly increase the voltage over the gate transistors to reach 3mW which of course reduce the safety margins and could even shorten the life expectancy of the transistors if taken to extremes. The modest voltage increase of 150mV is not expected to cause any such
Figure 7.40: Two tone spectrum, filtered/unfiltered, normalized and plotted

effects however.

At 3mW output power the architectures yielded:

**Analogue**

i. ACLR: -44.55dBc (5MHz)

ii. Efficiency: 1%

iii. IIP3: 10dBm

**Digital**

i. ACLR: -45dBc (5MHz)

ii. Efficiency: 25%

iii. IIP3: Not applicable

While not all that impressive, the digital amplifier shows a higher efficiency than the analogue amplifier. However, the complexity of the digital circuitry might not be justified by such a moderate increase.

### 7.7.2 One Tone Efficiency

The theoretical maximal efficiency of a class A amplifier is a straightforward 50% while the maximal theoretical efficiency of signal generated by the $\Delta\Sigma$-modulator varies depending on the signal contents. Hence we look at the efficiency of a single tone as well.

A single maximum power sine wave yields the following:

**Analogue**
7.7 Comparison

![WCDMA spectrum](image)

**Figure 7.41:** WCDMA spectrum, filtered/unfiltered, normalized and plotted

- Power: 3mW
- Efficiency: 1%
- ACLR: -44.55dBc (5MHz)
- IIP3: Not applicable

**Digital**
- Power: 3.3mW
- Efficiency: 39%
- ACLR: -48dBc (5MHz)
- IIP3: Not applicable

### 7.7.3 WCDMA Signal
Linearity has been central to the discussion throughout the thesis and most interesting is of course the result of a complex modern communications signal. The two amplifiers have been supplied with a 2GHz WCDMA signal with a 5MHz bandwidth.

**Analogue**
- Power: 5.7mW
- Efficiency: 1.98%
- ACLR: -48.55dBc (5MHz), -55dB (10MHz)
iv. IIP3: Not applicable

**Digital**

i. Power: 1.4mW

ii. Efficiency: 10%

iii. ACLR: -50dBc (5MHz), -50dBc (5MHz)

iv. IIP3: Not applicable
Chapter 8

Conclusions

In the previous chapters various ways of amplifying or creating an RF signal have been discussed. This chapter provides a few final comments on the technologies presented and attempts to predict where to direct future efforts in this field.

8.1 Analogue or Digital Amplification

Two things has been made apparent to us during writing this thesis, a simple analogue class A amplifier goes a long way and attempts at digital amplification requires a lot of fast circuitry as well as complicated filters. As previously mentioned in the thesis some suggestions for digital amplifiers require circuits working at eight times the intended carrier frequency or more, such methods are not plausible for RF applications today. Furthermore, while the super efficient switching amplifiers have become common in audio systems their limited bandwidth renders them useless for direct application to many RF systems even if a sufficiently fast circuit can be constructed. Hence the idea of up-mixing (wideband) ΔΣ-modulated signals from more manageable frequencies. Unfortunately the result is not very exciting since it sacrifices a lot of the intended benefits, namely efficiency.

8.2 Where to go from here

If we were to do another thesis today, we would look closer at the possibility of using a highly efficient Class-E amplifier together with a pre-distortion model. The model would have to keep track of the energy build up in the resonating LC-tank of the Class-E amplifier and then respond by increasing or decrease that energy so that the intended signal is reconstructed at the output of the amplifier. Possibly this requires additional control over supply voltage compared to the standard class E model and it is likely to either be less efficient than the original class E amplifier model or still very much
limited to small bandwidths. However it is certainly worth investigating further, side stepping the bandwidth limitation would open a broad field of application for such an amplifier.

The ΔΣ-modulation techniques, as used in the thesis, are all limited by their extensive filtering that has been shown to thoroughly ruin efficiency, as well as by the complicated digital filtering that restricts the modulator to comparably low frequencies. As such it really should be taken as proven that the modulators have no room in RF applications, at least not until processes capable of sustaining much higher clock frequencies are available.

Should there be interest to further investigate the mixing technique used in this thesis an attempt should be made to implement a high linearity mixer which could later be tweaked to handle larger effects. Preferably in a process such as GaAs where higher voltages can be used and higher effects reached.
Appendix A

4th order BP $\Delta\Sigma$

VerilogA-implementation

```verilog
// VerilogA for exjobb, fourth_order_bp_modulator, veriloga

'include "constants.vams"
'include "disciplines.vams"

module fourth_order_bp_modulator(signal_in, mod_out, mod_out_c, clk, noise);
output mod_out;
electrical mod_out;
output mod_out_c;
electrical mod_out_c;
input clk;
electrical clk;
input signal_in;
electrical signal_in;
input noise;
electrical noise;

//Process parameters:
parameter Vdd = 1.2;

//Filter parameters:
parameter a = -0.774000000000003463;
parameter b = -4.72621941582931272e-16;
parameter c = -0.558700000000003971;
parameter d = 0;
parameter A = -5.5511151231257827e-17;
parameter B = 1.2249999999999992;
parameter C = -1.38777878078144568e-16;
parameter D = 0.441299999999999304;

//Filter variables:
real in_mul1, in_mul2, in_mul3, in_mul4;
real in_delay1, in_delay2, in_delay3, in_delay4;
real sum_input_nodes;
real out_mul1, out_mul2, out_mul3, out_mul4;
real out_delay1, out_delay2, out_delay3, out_delay4;
real sum_output_nodes;
```
//Modulator variables:
real in_real;
real sum1;
real mod_out_r;
real DAC_out;
real diff1;
real dither;

analog begin
  @(initial_step) begin
    in_mul1 = 0;
in_mul2 = 0;
in_mul3 = 0;
in_mul4 = 0;
    out_mul1 = 0;
    out_mul2 = 0;
    out_mul3 = 0;
    out_mul4 = 0;
    in_delay1 = 0;
in_delay2 = 0;
in_delay3 = 0;
in_delay4 = 0;
    out_delay1 = 0;
    out_delay2 = 0;
    out_delay3 = 0;
    out_delay4 = 0;
    sum_input_nodes = 0;
    sum_output_nodes = 0;
  end//initial_step
  @(cross(V(clk)-Vdd/2, +1)) begin
    in_real = V(signal_in);
    dither = V(noise);
    sum1 = in_real+sum_output_nodes; //Add input and filter output.
    sum1 = sum1+dither; //Add white noise before quantizer.

    //2-level DAC:
    if (sum1>=0) begin
      DAC_out = 0.5;
      mod_out_r = 1;
    end else begin
      DAC_out = -0.5;
      mod_out_r = 0;
    end

    diff1 = sum1 - DAC_out;
//------------FILTER--------------------//
//Left side:
in_mul1 = in_delay1 * a;
in_mul2 = in_delay2 * b;
in_mul3 = in_delay3 * c;

in_delay3 = in_delay2;
in_delay2 = in_delay1;
in_delay1 = diff1;

sum_input_nodes = in_mul1 + in_mul2 + in_mul3;

//Right side:
out_mul1 = out_delay1 * A;
out_mul2 = out_delay2 * B;
out_mul3 = out_delay3 * C;
out_mul4 = out_delay4 * D;

sum_output_nodes = sum_input_nodes -
                      out_mul1 - out_mul2 -
                      out_mul3 - out_mul4;

out_delay4 = out_delay3;
out_delay3 = out_delay2;
out_delay2 = out_delay1;
out_delay1 = sum_output_nodes;

//---END-OF-_FILTER---------------------//
end //@cross end
V(mod_out) <+ transition(mod_out_r,10p,10p,10p);
V(mod_out_c) <+ transition(!mod_out_r,10p,10p,10p);
end //analog end
endmodule
Appendix B

Transistor Analysis; a Small Detour

This appendix was created as a result of our extensive analysis of the class A RF amplifier. The main reason for our big interest in this type of amplifier was two fold. First, of course, we wanted to learn the basics in RF power amplification. We chose to start with a class A RF power amplifier and we wanted to do the fundamental analysis as good as possible. We did not want readers of this report to feel that we did not actually understand the class A RF amplifier and its capabilities. Second, we had to learn the simulator and simulation techniques. We thought the work with the simulator would pay off and make the rest of our project continue smoother.

When beginning this project we were familiar with three types of simulations: AC, DC and transient. To evaluate a linear RF power amplifier we learned that another type of simulation, called PSS for periodic steady state, is beneficial. With this type of simulation, it is possible to measure for example output power, efficiency and linearity. As a comment, we also learned to perform an envelope following analysis, which should be used to fasten the analysis when applying slowly varying input signals (baseband data) to a RF circuitary.

The main result of this chapter is simulation results of linearity as a function of bias voltage. We begin with some theory of transistor modeling to try to explain the way our thoughts wandered. Simply, we wanted to verify the results predicted when looking at the transistor characteristics as a mathematical polynomial. Now, of course, the conclusion is obvious. But, at the time we were somewhat blinded by formulas and advanced simulation techniques.

B.1 Transistor Modelling

In the chapter (2) “Linearity and Units of Measurements” we introduced a way to model the transistor. With this model we were able to analyse and
understand some important properties of the power amplifier. For example, we looked at gain compression and intermodulation distortion. With a simple third degree polynomial at hand, we could explain these phenomena from a mathematical point of view. However, we have encountered different types of models during our investigation. For example, gain compression and intermodulation distortion is amplitude distortion. A polynomial with only real coefficients can only model amplitude distortion. If we let the coefficients be complex, we can model both amplitude and phase distortion. This model (polynomial) is often called Volterra series \[5\]. A Volterra series is shown in the equation below, note that the coefficients \( z_i \) are complex.

\[
y(t) = z_1 x(t) + z_2 x(t)^2 + z_3 x(t)^3 + \cdots + z_n x(t)^n \quad (B.1)
\]

Even though Volterra series can be fitted to I-V curves and model both amplitude and phase distortions, the model is still only valid for weakly nonlinear devices. One way to keep the device somewhat weakly nonlinear is by assuming it to be well backed off from the compression point. In other words, we should not use too large signals. Just to make clear the complexity and the amount of research done in the area of transistor modeling, presented below is a list of some models for GaAs MESFETs and HEMTs\(^1\). All models are summarized in \[5\].

i. Curtice quadratic nonlinear model

ii. Curtice-Ettenberg cubic nonlinear model

iii. Materka-Kacprzak nonlinear model

iv. Raytheon (Statz et al.) nonlinear model

v. TriQuint nonlinear model

vi. Chalmers (Angelov) nonlinear model

vii. IAF (Berroth) nonlinear model

Of course, we have not studied any of these models, but we became very much aware of the problem with a nonlinear transistor. Unfortunately, we focused a little bit too much on this problem before going on to switching amplifiers. Let us go on with a small signal model of the transistor, to be able to reason about causes for nonlinearities in the MOS-transistor.

\(^1\)GaAs MESFETs and HEMTs are the dominant transistor technologies for high frequency power amplifiers.
B.2 Small-Signal Modelling

If we examine the transistor for small input signals, the model to use is the “small-signal equivalent” [10]. Such a model assumes small variations around a bias point. This means that neither the gate source voltage nor the drain source voltage is varied much. A first order approximation is shown in figure (B.1). As the model is valid only for small signal variations, we must be careful when using it to state things for large signal behaviour. It can, anyway, be a good starting point.

![Small signal model of NMOS transistor](image)

**Figure B.1:** Small signal model of NMOS transistor [5]

A MOS-transistor is intended to work as a current source, which is steered by a voltage. This behaviour can be identified in the model by the current source having a current equal to $gm \cdot V_{gs}$. Ideally the gate-source resistance is zero and hence the input voltage will be the gate-source voltage, which steers the current source. However, if for example $R_{gs}$ varies as a nonlinear function of the input signal. Then the input-output relationship becomes non-linear [5] (as the current source will not change as linearly). Unfortunately, all components in the model varies, more or less, as a function of for example bias voltage, frequency and temperature. Very simplified we can make a list of sources of nonlinearity.

i. Nonlinear I-V
   The main contributor to nonlinearity. Fortunately, today, it seems as BSIM3v3 shows good agreement with measured curves. This means that we at least can, to some extent, depend on the simulator.

ii. Nonlinear C-V
    In the microwave region $C_{gs}$ influences the intermodulation distortion level. Both $C_{gs}$ and $C_{gd}$ should be included in the model.
iii. Gate-source resistance
The effect of $R_{gs}$ becomes significant only at higher frequencies close to $f_t$.

iv. Temperature dependence
Silicon MOSFET devices are very sensitive to the operation temperature. Mainly the effective carrier mobility $\mu$ and the threshold voltage $V_{th}$ are sensitive to variation in temperature.

B.3 Nonlinear (I-V) $i(v, v)$

If we assume the transistor to be operating in a small signal regime and that the large signal behaviour will degrade the performance. Then we can state that: if the small signal model does not fulfill our requirements, neither will the large signal behaviour (real case).

One small signal model, with only dominant terms, for the I-V characteristics is given in equation (B.3) [4]. $V_{GS}$ and $V_{DS}$ are fixed DC-voltages (the bias point). Hence, the term $i_{DS}(V_{GS}, V_{DS})$ represents the bias current.

$$i_{DS}(v_{GS}, v_{DS}) \approx$$

$$i_{DS}(V_{GS}, V_{DS}) + \frac{\partial i_{DS}}{\partial v_{GS}} v_{gs} + \frac{\partial i_{DS}}{\partial v_{DS}} v_{ds} + \frac{1}{2} \frac{\partial^2 i_{DS}}{\partial v_{GS}^2} v_{gs}^2 + \frac{1}{6} \frac{\partial^3 i_{DS}}{\partial v_{GS}^3} v_{gs}^3$$

By introducing commonly used variables, this relation can also be stated as:

$$i_{DS}(v_{GS}, v_{DS}) \approx$$

$$i_{DS}(V_{GS}, V_{DS}) + g_m v_{gs} + g_d v_{ds} + g_m^2 v_{gs}^2 + g_m^3 v_{gs}^3$$

In the chapter (2) “Linearity and Units of Measurements” various effects of nonlinearity are discussed. Specifically, the source for intermodulation distortion is identified as the third order term in the polynomial modelling the transistor. If we identify the same term in this model, then $g_{m3}$ is “responsible” for intermodulation distortion [4].

A DC simulation showing the drain current, $i_{DS}$, as a function of the gate voltage, $v_{GS}$, and its derivatives is shown in the figure (B.2). $V_{DS}$ is held constant 3.3 V, according to the supply voltage of the 0.35 $\mu$m process. The lower left plot displaying $g_{m3}$, is the plot we are expecting to reveal some of the reasons for intermodulation distortion.

As can be seen $g_{m3}$ is approaching zero as $V_{gs}$ is increased. However it equals zero in one point, close to the transistor turn on when $V_{gs} \approx$
$V_{\text{threshold}} \approx 0.6$ volt for this (300µm wide) transistor. This is an indication of that the lowest intermodulation distortion is possible for an ideal class B amplifier [4]. (The term power in class B amplifier is left out with purpose as it is small signal behaviour we are analysing.)

To further analyse the “class B phenomena” we performed a two-tone test. We used a periodic steady state analysis and swept the input power. On top of the periodic steady state analysis we parameterized the bias voltage. After the simulation we looked at the suppression of the intermodulation distortion component. It could be seen that there is, indeed, a sweet spot around the transistor turn on. An amplifier working in this area is usually regarded as a class B or AB amplifier. A class AB power amplifier could thus be a strong candidate for high linearity and efficiency. It can even outclass class A due to this sweet spot [4]. It requires, however, very careful biasing [4]. However, to be able to say if the class B configuration really outperforms the class A, in terms of intermodulation distortion, we must also compare the power gain. This reasoning is done in the end of this appendix.

We continued to investigate the dependency upon the bias point, or - from our viewpoint, the dependency upon the region of operation. In the first DC-analysis, the gate-source voltage (bias voltage) was swept, but we kept the drain-source voltage constant. This simulation thus implies that we are assuming very small variations around the bias point. For larger variations the assumption of constant drain-source voltage no longer holds. To be able to better model the behaviour of the transistor operating in a
large-signal regime, we let our imagination flow. The result was an idea to let the drain-source voltage be linearly dependent of the gate-source voltage. In other words we let a change in gate-source voltage also change the drain-source voltage. Using power amplifier terminology we can explain the idea as modelling ideal load lines as depicted in figure (B.3).

Figure B.3: $I_{ds}(V_{gs}, V_{ds})$ with examples of loadlines.

With this new model we once again extract the DC-characteristics. This time, however, both the gate-source voltage and drain-source voltage are swept, according to the chosen load line. The result is seen in figure (B.4). Every line in each of the four subplots represents a specific bias point, where $V_{gs}$ and $V_{ds}$ have been swept according to the chosen load line. In other words we are trying to simulate that the transistor is operating in a certain region. The left upper plot, which is displaying the first derivative of the load lines, should be fairly constant. If we try to understand the meaning of this first derivative it will be something like: the relation of change in gate-source voltage to change in drain current. Again it seems to be a good idea to increase the bias voltage.it is constant seems to increase with increasing bias voltage.

The next plot, in the upper right corner, is the second derivative of the DC-characteristics for this simulation. The scale is now in dB and it is easily seen that choosing a region of operation with higher bias voltage reduces this derivative. Putting this result in the light of a polynomial model, this should mean that increasing the bias point will decrease the harmonic distortion.

Let us continue with the lower left plot, which shows the third deriva-
derivatives of the DC-characteristics. With this plot it is harder to state anything at a first glance. Before we go on, there are at least two things to be considered. First, we can not interpret this parameter to any physical behaviour. Second, we do not know what to expect from the simulation model when performing three successive numerical derivations on its DC-data. However, let us assume that the simulation data is accurate enough and try to make some observations. For $V_{\text{bias}}$ equalling 1, 1.25 and 1.5 volt there exists singularities, so we therefore neglect these. If we focus on $V_{\text{bias}}$ equalling 1.75, 2, 2.25 and 2.5 volts it can be seen that the third derivative is again decreasing as we are increasing the bias point. Once again we predict that the linearity performance will gain from increasing the bias point.

As a next step we now tried to simulate the actual large signal behaviour for different bias points. Once again we performed a two tone test by using a periodic steady state analysis with swept input power and, on top of that, parameterized bias point. This time we included enough harmonics, in the simulation setup, to be able to measure the second order harmonic distortion component. The result of the simulation is shown in figure (B.5).

First we make the most obvious observation: by increasing the bias point the suppression of the second order harmonic is increased, independent of power gain. Increasing the bias point is a win-win situation with respect to harmonic distortion suppression. Let us continue with the intermodulation distortion.

To be able to say anything about the intermodulation distortion suppression, we have to be a bit careful. When analysing the intermodulation distortion suppression plot, we see that a lower bias point implies higher
suppression. Now we also have to take the output power in consideration. From the power gain plot we see that the power gain increases with increasing bias point. The question is then: is the difference in power gain larger than the difference in intermodulation distortion suppression? Or stated another way: if the difference in power gain is XdB then we need to input another XdB to get the same output power. By increasing the input power by XdB how much will the intermodulation distortion increase?

In one specific example, as seen by the markers A and B in figure (B.5), the difference in power gain is 2.46dB for the bias points 1.25 volt and 2 volt. Now we look at the intermodulation distortion suppression plot. We also see that, being forced to increase the input power by 2.47dB the suppression stays the same. However, there is one difference with the two cases. The efficiency will of course decrease with higher bias voltage; this can be seen in the subplot displaying drain efficiency.

Let us take a second example. By looking at the intermodulation distortion suppression curves for the bias points 2.0 and 2.25 volts it can seen that they are almost the same. Further, by looking at the power gain it can be seen that there is a difference. Increasing the bias voltage from 2.0 volt to 2.25 volt thus increases the intermodulation distortion suppression (and harmonic distortion suppression). It can be commented again that we, at the same time, loose efficiency.
B.3.1 Conclusion

During our previous investigation of the class A amplifier and during this more specific analysis of the linearity depending on the region of operation we have learned that there does not exists a perfectly linear class A amplifier. The nature of the transistor makes it impossible to build a perfectly linear amplifier. However, one recipe could be to use a very small region of operation and bias the transistor quite high. This recede has the large drawback of decreased efficiency (often to impractical levels). Today, linearity is increased by different linearization techniques.
Appendix C

C1 and C2 Tuning Procedure


**Figure C.1:** C1 and C2 tuning procedure. The vertical arrow indicates transistor turn-on. [23]
Bibliography


Appendix D

Terminology

3GPP: 3rd Generation Partnership Project
ACLR: Adjacent Channel Leakage Ratio
CDMA: Code Division Multiple Access
IP3: Third-order intersect point
NTF: Noise Transfer Function
OSR: Over Sampling Ratio
PAE: Power Added Efficiency
PSS: Period Steady State
QPSK: Quadrature Phase Shift Keying
RF: Radio Frequency
RFC: Radio Frequency Choke (big inductor)
SNR: Signal to Noise Ratio
STF: Signal Transfer Function
WCDMA: Wideband Code Division Multiple Access
Digital RF: Two Level Radio Frequency Signal
Upphovsrätt

Svenska

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under en längre tid från publiceringsdatum under förutsättning att inga extra-ordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmanens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnad som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är krävande för upphovsmannens litterära eller konstnärliga anseende eller egenart. För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/.

Copyright

English

The publishers will keep this document online on the Internet - or its possible replacement - for a considerable time from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/.

© Tobias Andersson, Johan Wahlsten
Linköping, 2007