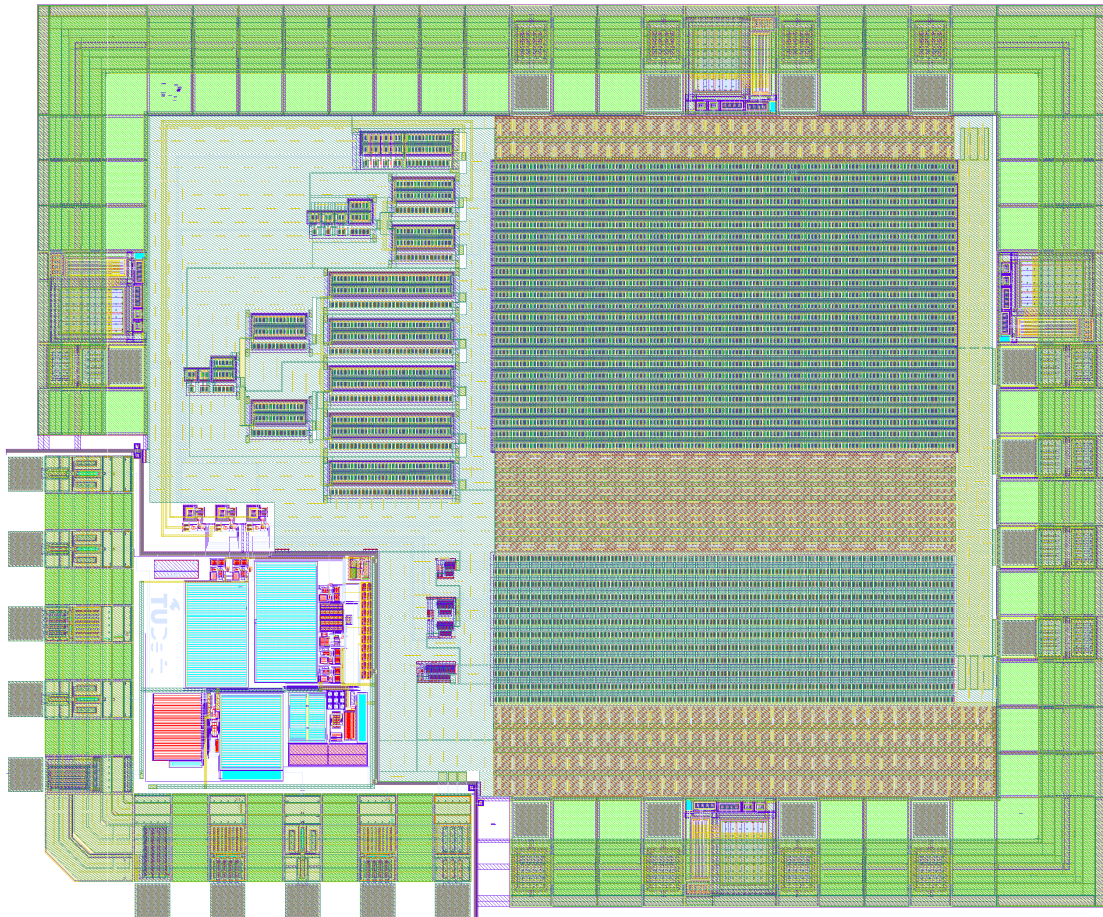# A high-efficiency switch-mode amplitude modulator for class E power amplifiers in nano-satellites

Thesis for a Master of Science degree in Microelectronics
Robin F. Kearey, B.Sc.
June 2010

TUDelft

Delft University of Technology

Delfi-n3Xt

Contents

Preface

This thesis concludes a year and a half of calculations, simulations, reading, drawing and writing. Its goal was to produce a working IC that meets its specifications, and hopefully, implement new ideas to show that they work. This goal has been reached, and a design has been produced that is ready for manufacturing.

Although the initial problem description seemed simple, it turned out to be a lot of work to actually make a circuit that performs well under all circumstances. Fortunately, modern computer-aided design tools can be a great help in making the right design choices, but only if one is able to operate the tools correctly and to interpret the results in the correct way.

I would like to thank Edin Wiek, Maurits Schaap, Wolter van der Kant, Sheng Li, Robin van Eijk, Ronald de Bock and Christiaan Hartman, with whom I had the pleasure of sharing an office at the 18th floor of the Electrical Engineering building, for their interesting discussions and for providing a pleasant working atmosphere. Thanks also to Eric Smit for frequently distracting us from our work. A big thank you to the secretary, Marion de Vlieger, who was always there to help with small and large problems. Finally, many thanks to my supervisors, Chris Verhoeven and Bert Monna, who taught me (almost) everything there is to know about IC design.

Robin F. Kearey
Delft, June 2010

Summary

This thesis describes the design, simulation and implementation of a supply modulator to be used in a VHF power amplifier on the Delfi-n3Xt nano-satellite. First, a set of specifications is defined that describe the required functionality. These are derived from earlier work on related systems and confirmed through discussions within the project team.

Secondly, a design strategy is developed that allows a structured and logical way of transforming the specifications to a working circuit. It is shown that power efficiency and distortion performance can be optimized separately and independently. Several novel solutions are found to optimize efficiency and reduce distortion.

The entire schematic is simulated and found to agree with calculations. Variations in supply voltage and temperature are taken into account, along with manufacturing spread in all components. Finally, a complete IC layout is produced that is ready for manufacturing.

1. Introduction

Delfi-n3Xt is a satellite built by Delft University of Technology with the purpose of educating students in all aspects of satellite technology.

One of the sub-projects within the Delfi programme is the Isis Transceiver, or ITRX. This is a UHF/VHF radio that will fly aboard Delfi-n3Xt to demonstrate new technologies in radio design. One of these technologies is a power amplifier (PA) that will be very power efficient.

The first design choices for the ITRX PA were made during the design of Delfi-C3, the predecessor of Delfi-n3Xt. Two documents detailing this work are [1] and [2].

The main objective of the ITRX PA project is to create a power amplifier operating in the VHF band (around 144 MHz) that is highly efficient and transparent to the modulation scheme used on the RF signal.

The specifications that the PA has to comply with have been discussed within the project team, and are summarized below.

Frequency range	145.8 – 146.0 MHz
Signal bandwidth	DC – 40 kHz
Output power	1.2 W
Out-of-band spurious signal level	-44 dBc
In band 3rd order intermodulation	-30 dB
Supply voltages	12 V, 3.3 V
Antenna impedance	50 Ω

Table 1-1: Power amplifier requirements

In order to obtain maximum power efficiency, a class E amplifier, introduced by [4], is used. The basic schematic of a class E amplifier is shown below.
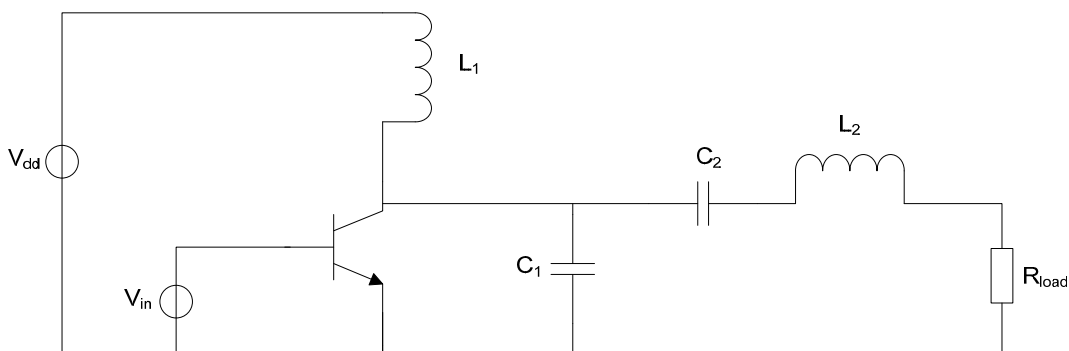
Figure 1-1: Basic circuit of a class E amplifier

A class E amplifier works by driving the transistor as a switch, which means that it is either fully turned on or fully turned off. L_1 and C_1 are tuned in such a way that the voltage across the transistor is zero when it turns on, and that the current through the

transistor is zero when it switches off. This ensures that the power dissipation in the switch is (ideally) zero at all times. The output filter consisting of C_2 and L_2 removes any harmonics caused by the switching action.

Although a class E amplifier is highly efficient, it is also highly nonlinear since it can only amplify the phase information of the input signal. An input signal given by $V_{in} = A(t)\cos(\omega t + \varphi(t))$ will appear at the output as $V_{out} = \dfrac{V_{dd}}{\sigma}\cos(\omega t + \varphi(t))$, where σ is a constant that depends on the details of the class E amplifier.

A straightforward way to fix this is to modulate V_{dd} with the information that was present in the amplitude of V_{in}. Figure 1-2 shows a circuit that implements this.
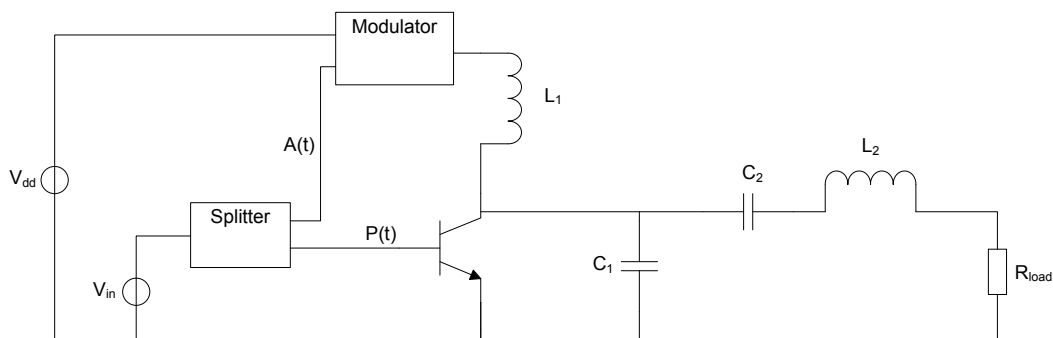
Figure 1-2: Class E amplifier with supply modulation

This system is called Envelope Elimination and Restoration (EER). By separating V_{in} into a phase component $P(t) = A\cos(\omega t + \varphi(t))$ and an amplitude component A(t) and combining them in the class E amplifier, it is possible to use any modulation scheme and still use the highly efficient class E amplifier.

The supply modulation has to be done efficiently. If an inefficient modulator is used, then the advantage gained by using the class E amplifier is lost. This means that the supply modulator will have to be a switch-mode amplifier as well.

The class E amplifier has been completely designed and simulated [1]. The splitter will be a digital circuit, implemented in an FPGA or an ASIC. This will also include a feedback loop to align the timing of the amplitude and phase paths, and thereby linearize the complete circuit. Details of this can be found in [2]. Pictured below is a block diagram of the system that is planned to be included in Delfi-n3Xt.
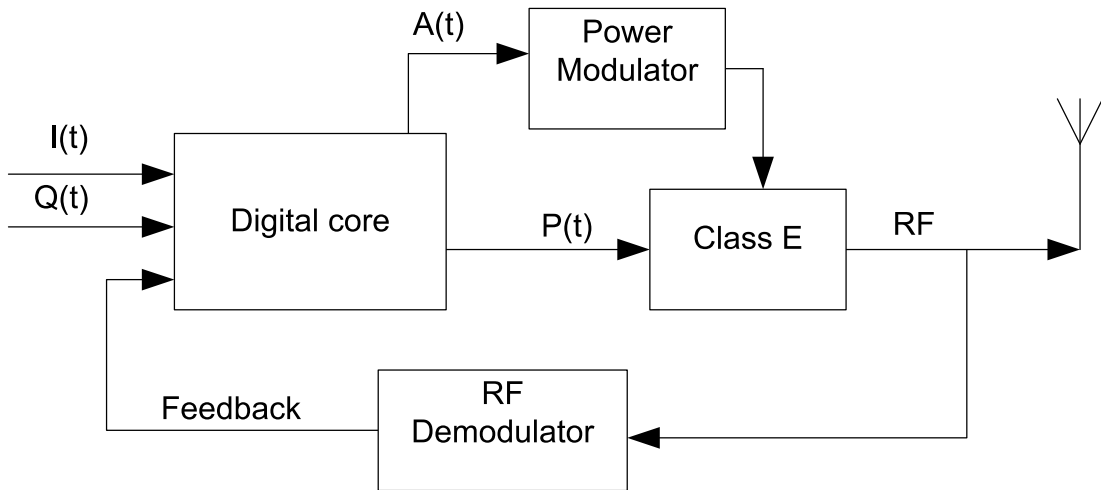
Figure 1-3: Complete circuit for Delfi-n3Xt ITRX power amplifier

The scope of this thesis is to design the power modulator. The relevant specifications are summarized below.

Frequency range	DC – 40 kHz
Load impedance	50 Ω
Supply voltages	12 V, 3.3 V
Maximum output voltage	10 V
Out-of-band spurious signal level	-44 dBc
In band 3rd order intermodulation	-30 dB

Table 1-2: Power modulator requirements

The maximum output voltage is derived from a remark in [1] that states that the maximum output power is 1.2 W. Higher power levels will lead to a voltage swing at the output transistor that exceeds its maximum voltage rating. Using the antenna impedance of 50 Ω and an efficiency of 70 %, this results in an input voltage of 10 V, or 83 % of the 12 V supply.

During the design phase, when it became clear that the power modulator would be implemented using a class D amplifier, the team expressed an interest in being able to drive it using a PWM signal directly. This feature will be treated as a "nice-to-have", to be implemented if possible without too much design effort.

2. Design strategy

A structured and hierarchical design strategy will be followed, inspired by [3], and adapted to the design of switch-mode amplifiers. The emphasis is on developing a design method that minimizes the number of iteration loops required, and that clearly relates design parameters to the specifications.

2.1 Basic principle

Several basic principles for switch-mode amplifiers are possible. These are normally called class D, class E and class F amplifiers. The principle of class E amplification was described in chapter 1. Class F is similar in operation to class E, the main difference being that a class F amplifier is tuned to make better use of the harmonics produced by the switching action.

Both class E and class F amplifiers have the drawback that they are relatively narrow-band amplifiers due to the resonant LC networks at their output. This also makes them hard to design for low frequencies, which would require large inductors and capacitors. Finally, they are unable to amplify DC signals. Therefore, class E and F are not the correct choice for a supply modulator.

A class D amplifier operates on a different principle: it first modulates the input signal to produce a pulse-width modulated signal that can be amplified by a set of switches. A lossless filter at the output then reconstructs the original waveform. Class D amplifiers can amplify signals at low frequencies, down to DC. They are less well-suited for high-frequency signals than class E or F amplifiers, but for the supply modulation this is not a problem. A class D amplifier is therefore chosen to implement the supply modulator. More details about its working principle are described in section 3.1.

2.2 Output stage

The most important specification of the supply modulator is its power efficiency. In fact, this is the only reason to choose a switch-mode amplifier. The output stage is the most critical part in determining the efficiency, so it will be designed first. This involves setting up a model of the output stage that relates the efficiency to the design parameters.

Once this model is in place, parameters from the available transistors can be filled in and optimal values can be calculated. The switching frequency can be calculated, which has to be optimized between low power dissipation and low spurious signal levels.

Section 3.2 describes the efficiency model in detail. Section 3.3 describes a method of reaching even higher efficiencies by dynamically adapting the output stage to the signal it is amplifying. In chapter 4 these calculations are implemented, and actual circuit parameters are derived. The gate drivers of the output transistors are also considered part of the output stage, and are designed in a similar way (section 4.5).

2.3 Input stage

The second most important specification is the distortion performance. This is mainly determined by the modulator that transforms the input signal into a pulse signal that drives the output stage. Any nonlinearity in this process leads to distortion in the output signal. It is possible to suppress distortion by using negative feedback around the amplifier. However, the benefit of this is limited, because class D amplifiers have relatively little gain. The circuit will therefore be deisgned to reach its distortion specifications without negative feedback. Chapter 5 describes the design of the input stage.

2.4 Support circuits

Finally, some support circuits will be designed, such as biasing, reference and start-up circuits. Some circuitry will also be included to facilitate testing of the IC after production. These circuits are described in chapter 6.
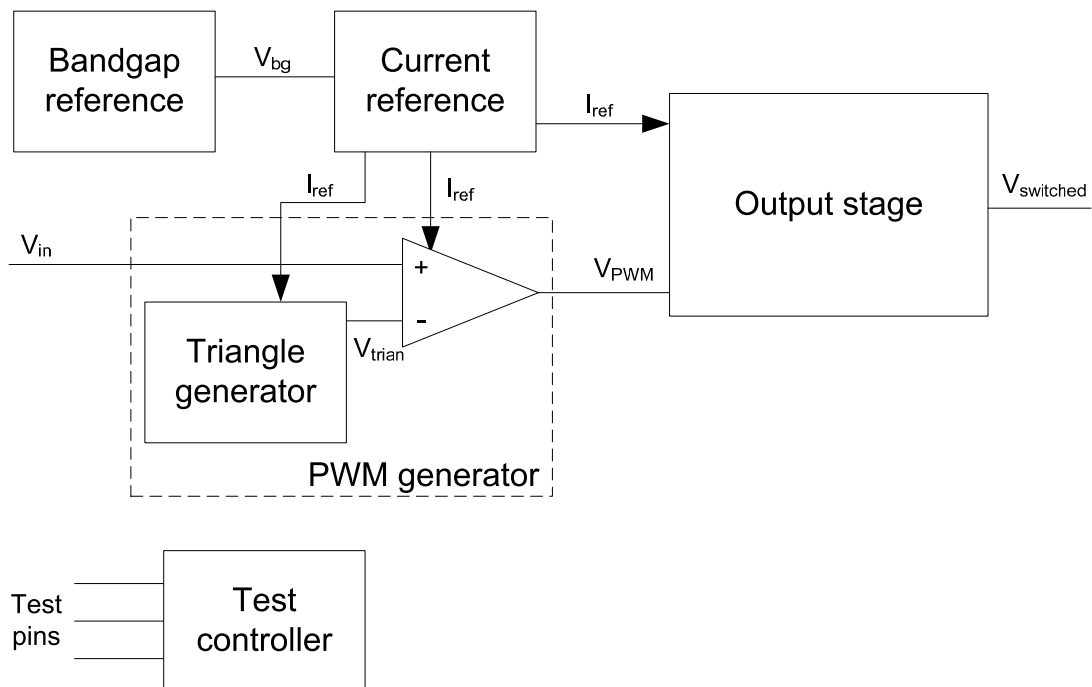

Figure 2-1: Schematic of all on-chip circuits

2.5 Technology and tools

In the interest of compatibility, the same IC process will be used as was used for the class E amplifier. This is the H35B4D3 process by Austriamicrosystems, which includes both high-speed 0.35 μm and high-voltage CMOS transistors. A few key specifications have been outlined below.

LVCMOS minimum channel length	0.35 μm
LVCMOS operating voltage	3.3 V
Number of masks	27
Number of metal layers	4
HVCMOS operating voltage	20 V, 50 V
Additional features	High-resistive poly, thick power metal

Table 2-1: Specifications of the H35B4D3 process

The IC will be designed using the foundry provided design kit for Cadence Custom IC Design System version 6.1.3 using the Spectre circuit simulator. It will be produced in a Multi Project Wafer (MPW) to reduce costs. Large-scale production is not a requirement.

The circuits will be simulated over all process corners included in the design kit, to make sure that the circuit will work on any wafer returned from the foundry. Furthermore, the circuits will be tested with supply voltage variations of +/- 10 % and over a temperature range of -40 to +85 °C.

The process documentation provided by Austriamicrosystems ([10] - [15]) describes in detail the available devices, their characteristics and performance, and the design rules that need to be followed to allow reliable circuit manufacturing.

3. Switch mode amplifiers

3.1 Class D principle

Linear amplifiers (class A, B, and AB) are relatively power inefficient by necessity. Because there is always a bias current flowing through the active device, they always dissipate a certain amount of power (namely $I_{bias} \cdot V_{bias}$) while amplifying.
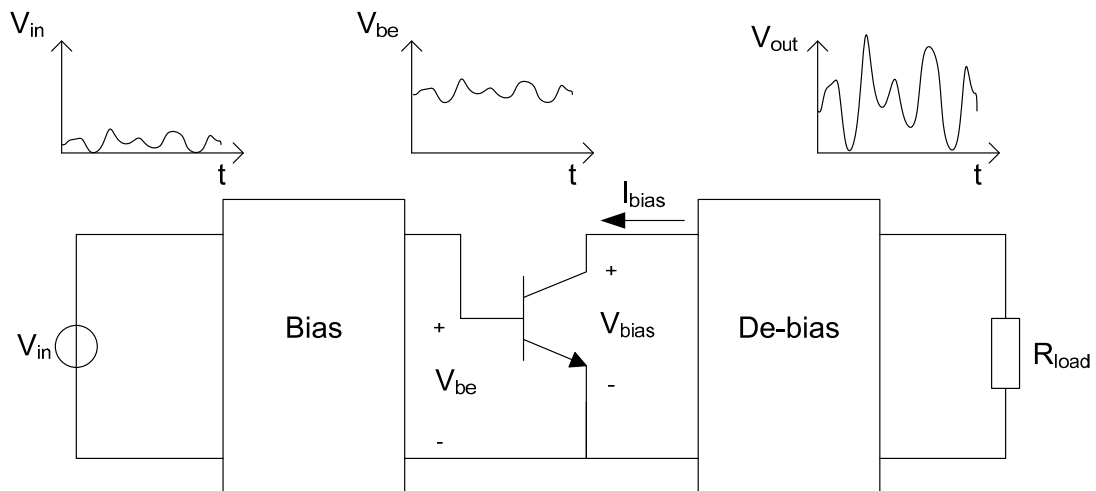

Figure 3-1: Generalized linear amplifier

Switch-mode amplifiers are specifically designed to be power efficient by allowing their transistors to be either completely on (in which case $V_{bias} = 0$) or completely off (in which case $I_{bias} = 0$). In both of these states, the power dissipated in the switch is zero.
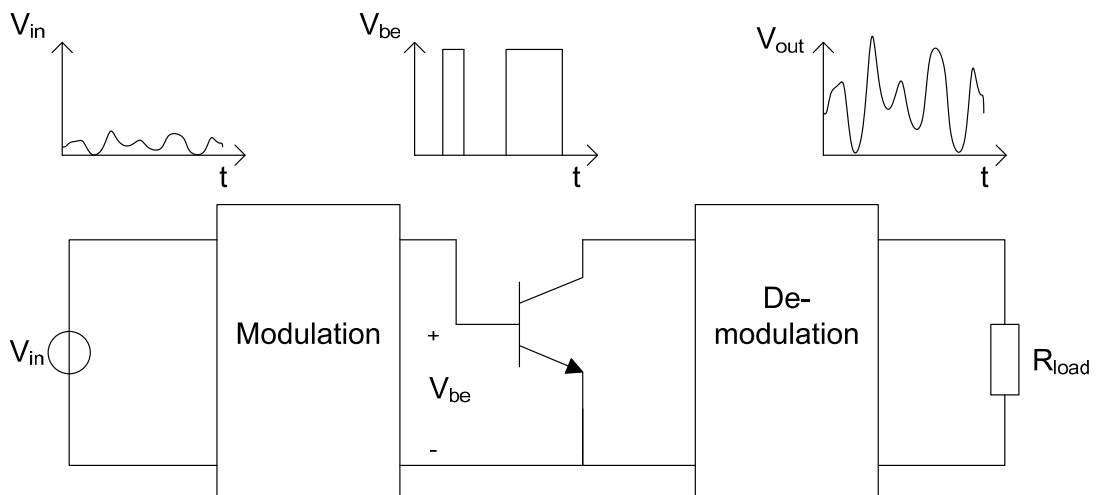


Figure 3-2: Generalized switch-mode amplifier

If the information in the signal can be efficiently modulated in such a way that it can be represented by the state of a switch, and the information can be efficiently retrieved after amplification, then the total amplifier will approach 100 % efficiency.

A switching amplifier only needs power to modulate the signal, drive the input of the switch, and demodulate the amplified signal. All of these actions can be completed using much less power than the biasing power required by a linear amplifier.

A suitable modulation system for low-frequency signals is pulse-width modulation (PWM). A pulse-width modulated signal consists of a series of pulses, each with a certain on-off ratio (called the duty cycle, denoted by δ) proportional to the amplitude of the input signal. In this way, the information that was encoded in the amplitude domain, is now encoded in the time domain. The amplitude of the PWM signal is either low or high, enabling it to be amplified by switches.

Figure 3-3 shows the time-domain PWM signal (blue) for a sinusoidal input (red). A simple way of generating this PWM signal is to compare the input signal with a triangle wave (green) of the same frequency as the desired PWM signal.
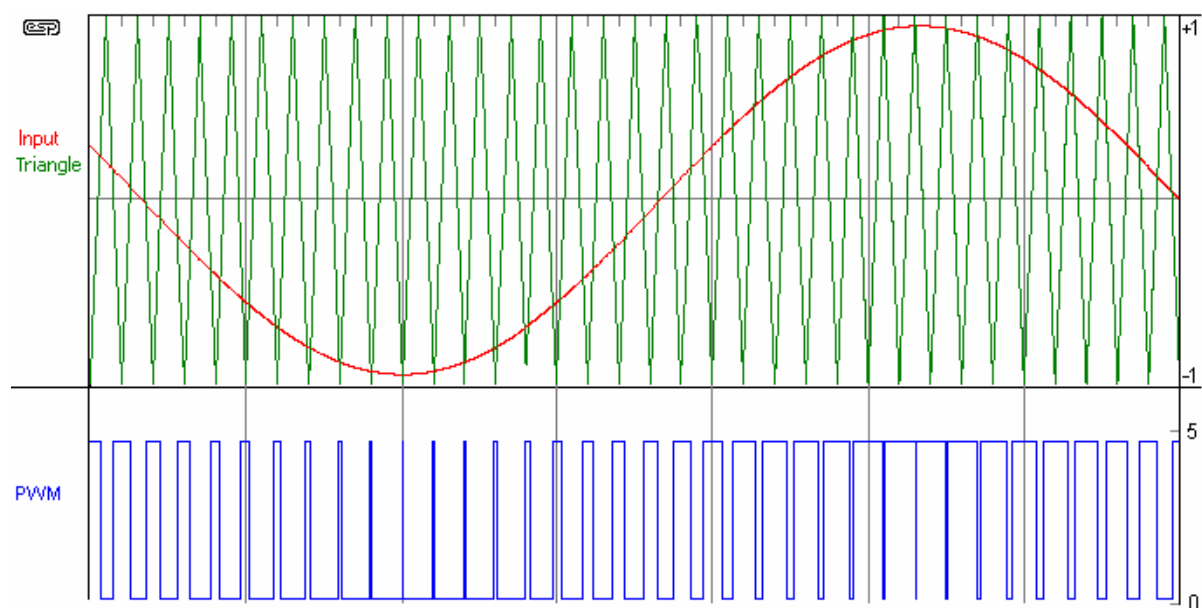
Figure 3-3: Illustration of PWM signal, from [16]

It is also very straightforward to retrieve the original information from the amplified PWM signal. A low-pass filter is sufficient to convert the information back into the amplitude domain. If this low-pass filter has no power loss, then the signal has been amplified with, ideally, no dissipation at all.

Several topologies are possible for PWM-based switch-mode amplifiers, which are shown in Appendix A. The required input and output quantities are both voltages. The output voltage has to have a smaller amplitude than the input voltage, which means that a voltage-to-voltage switching stage is a good choice. If used as a signal amplifier, this topology is usually called a class D amplifier in literature. This name will be used in this text as well. The basic schematic of a class D amplifier is shown in Figure 3-4.
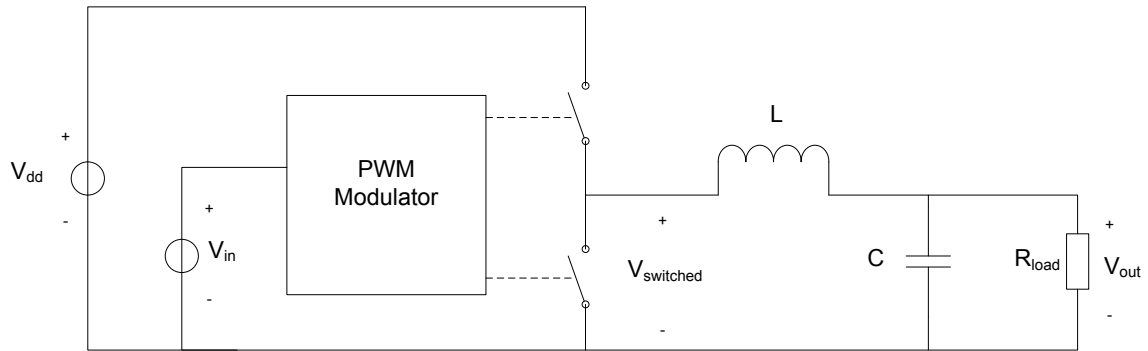
Figure 3-4: Basic circuit of a class D amplifier

The input signal is first modulated to create a PWM signal. This signal is used to turn the output switches on and off, creating an amplified version of the PWM signal at $V_{switched}$. The low-pass filter consisting of L and C blocks the high frequencies and thereby reconstructs the original input waveform.

The two switches are driven in a complementary way, meaning that when one is on, the other is off and vice versa. The two switches may never be on at the same time, since this would short circuit the supply voltage, nor should they ever be off at the same time, since this would interrupt the current flowing through the inductor, causing a large voltage spike at $V_{switched}$. In reality it is not trivial to prevent these two conditions, and special measures have to be taken in the driving circuitry (section 4.7).

Figure 3-5 shows the voltage and current waveforms. $V_{switched}$ is an amplified (and possibly inverted) version of the PWM signal. I_L is the current flowing through the inductor, ramping up and down as $V_{switched}$ alternates between high and low. The average value of the inductor current is equal to the output current (due to conservation of charge), and is given by $I_L = \dfrac{V_{out}}{R_{load}}$.

The current I flowing through an inductor L when a voltage V is applied to it, is given by $I(t) = I_0 + L\int V(t)dt$. In a class D amplifier, the voltage across the inductor is equal to $V_{dd} - V_{out}$. Since V_{out} is changing much more slowly than $V_{switched}$, it can be assumed to be constant during the charging of the inductor, and I(t) can be simplified to $I(t) = I_0 + L(V_{dd} - V_{out})t_{on}$.
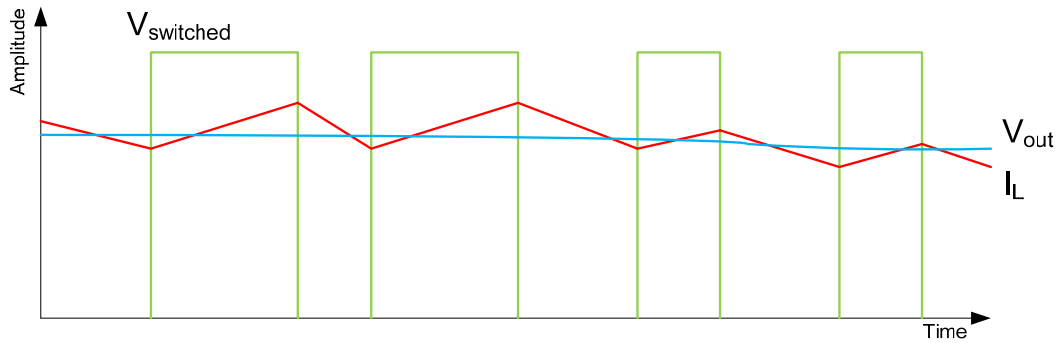
Figure 3-5: Waveforms at the output of a class D amplifier

The complete analogue circuit that will form the core of the power amplifier now consists of a class D and a class E amplifier, each amplifying that part of the spectrum that suits their mode of operation: the baseband (low frequencies) for the class D, and the carrier (high frequencies) for the class E. This is depicted in Figure 3-6.
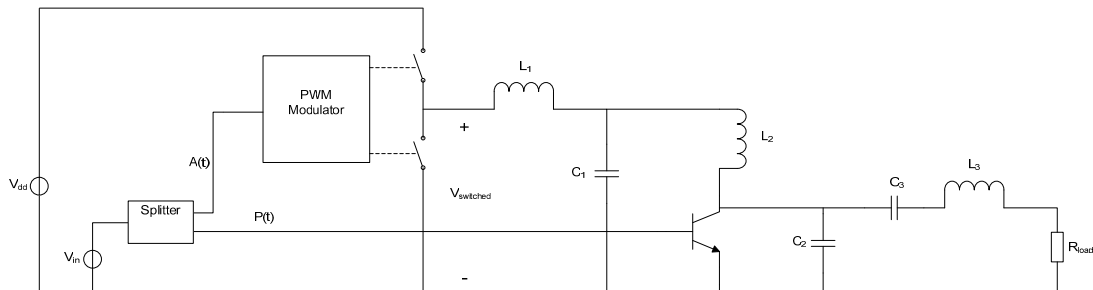

Figure 3-6: Combination class D and class E circuit

3.2 Efficiency calculations

The first step in the design of the class D amplifier is to set up a model that relates the design parameters to the efficiency. Since ideal switches consume no power, it is necessary to first define what kind of physical switches will be used. In CMOS technology, the most obvious choice is a MOSFET. The power efficiency is then limited by two main factors: dissipation in the on-resistance of the MOSFETs, and dissipation through the charging of capacitances (mainly the gate-source capacitance). The location of these parasitics is shown in Figure 3-7.

Since the MOSFETs are driven by a pulse waveform with a duty cycle δ, the circuit has two distinct phases. Figure 3-7 shows the parasitics in both the on-phase (δ) and the off-phase (1- δ).
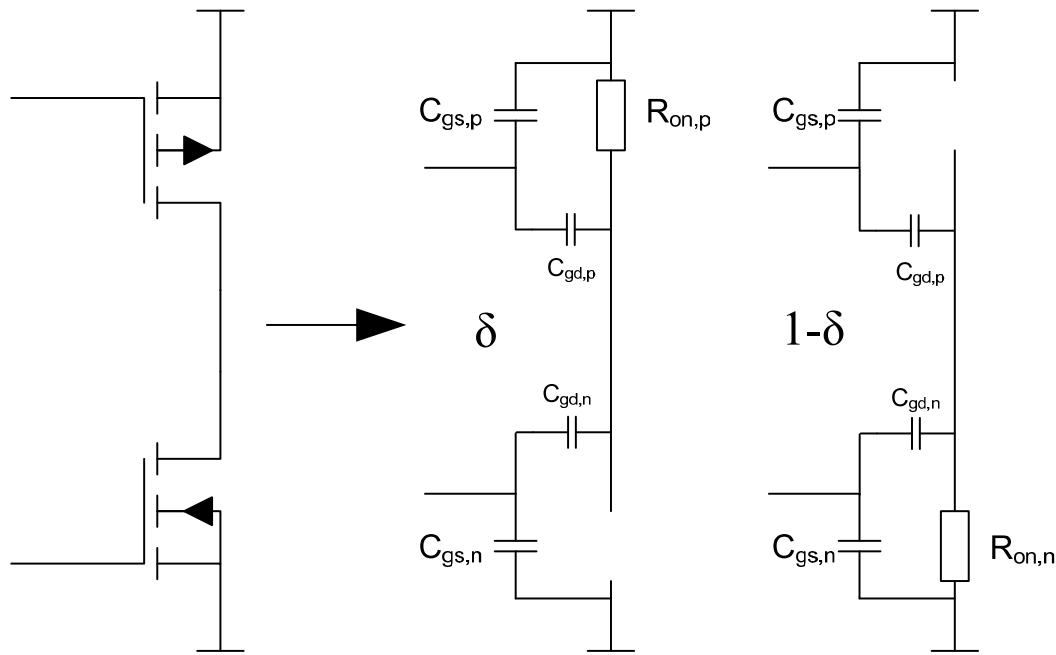
Figure 3-7: Parasitic components in CMOS inverter

The amount of power dissipated in the on-resistance of a transistor is equal to $P_{res} = I^2 R_{on}$, if I is a constant current. The on-resistance is given

by $R_{on} = \dfrac{1}{\mu_0 C_{ox} \dfrac{W}{L}\left(\left(V_{gs} - V_{th}\right) - 2V_{ds}\right)}$, which becomes smaller with increasing V_{gs}.

The amount of power needed to charge the gate-source capacitance C_{gs} from a voltage source V_{gs}, with a frequency f_{sw}, is equal to $P_{cap} = V_{gs}^2 C_{gs} \cdot f_{sw}$, in which C_{gs} = $C_{gate} \cdot W \cdot L$ with C_{gate} equal to the gate capacitance per unit area.

Of course, both of these two types of dissipation need to be as small as possible. However, they cannot be optimized independently, since both are functions of W, L and V_{gs}. Table 3-1 shows the requirements on these parameters for both types of dissipation.

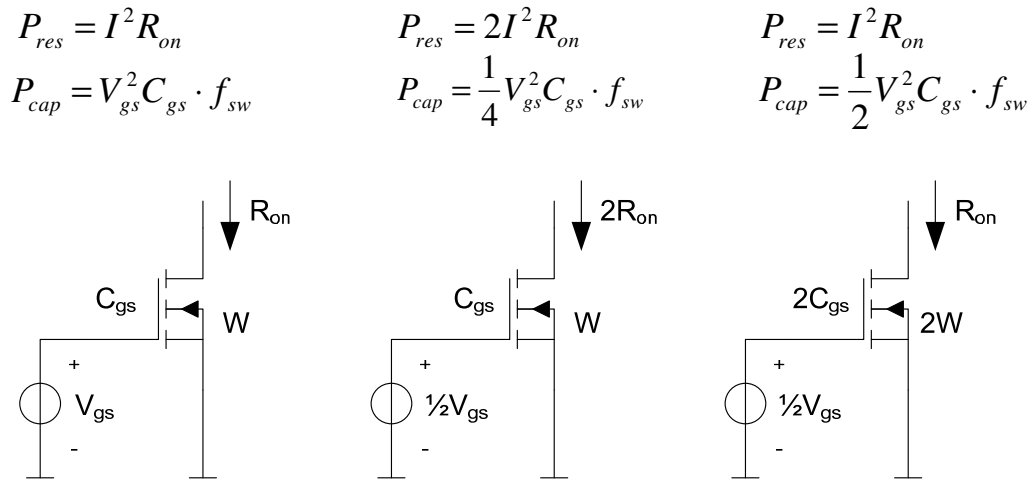|	W	L	V_{gs}
Low resistive losses	Large	Small	High
Low capacitive losses	Small	Small	Low

Table 3-1: Transistor requirements for low losses

One thing that is immediately obvious is that the L of the output transistors should be as small as possible. There is no reason to make it any larger, so in the rest of this discussion, L is assumed to be minimum size.

3.2.1 Optimum for V_{gs}

As shown in Table 3-1, a low V_{gs} is required to minimize resistive losses, while a high V_{gs} is required to minimize capacitive losses. This means that there is an optimal value for which the total dissipation is minimized.

As shown above, the R_{on} of a MOSFET goes down approximately linearly with increasing V_{gs}. The resistive losses are linearly related to R_{on}, so they too go down approximately linearly with increasing V_{gs}. The capacitive losses however, go up quadratically with increasing V_{gs}. Furthermore, both types of dissipation are linearly related to the width of the transistor. Figure 3-8 shows how an optimization for V_{gs} is performed.

$$P_{res} = I^2 R_{on}$$
$$P_{cap} = V_{gs}^2 C_{gs} \cdot f_{sw}$$

$$P_{res} = 2I^2 R_{on}$$
$$P_{cap} = \frac{1}{4} V_{gs}^2 C_{gs} \cdot f_{sw}$$

$$P_{res} = I^2 R_{on}$$
$$P_{cap} = \frac{1}{2} V_{gs}^2 C_{gs} \cdot f_{sw}$$

Figure 3-8: Optimization for V_{gs}

The left figure shows the initial situation. In the middle figure, V_{gs} is halved, causing R_{on} and therefore P_{res} to double, but P_{cap} to be reduced by 75%. Depending on the values of P_{res} and P_{cap}, this could be an improvement or a reduction in efficiency. Looking at the right figure however, it becomes clear that subsequently doubling the width of the transistor leads to an improvement in the efficiency in any case. This optimization could in theory be carried through until V_{gs} reaches the threshold voltage of the MOSFET, but is in practice limited by the nonlinearity of R_{on} versus V_{gs}.

It should be noted however that V_{gs} should be charged from an "efficient" voltage source. If there is a voltage source with a value V_{source} and a linear regulator is used to charge the gate to V_{gs} (where $V_{gs} < V_{source}$), then the energy required from the source is equal to $C \cdot V_{source} \cdot V_{gs}$, which decreases linearly with decreasing V_{gs} instead of quadratically, as shown in Figure 3-9. This means that the optimization shown above does not hold anymore.

$$P_{res} = I^2 R_{on}$$
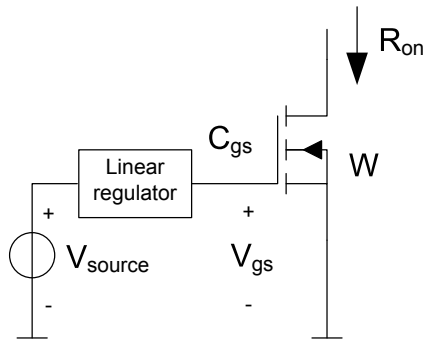$$P_{cap} = V_{source} V_{gs} C_{gs} \cdot f_{sw}$$

Figure 3-9: Charging a gate using a linear regulator

In practice, there is usually only one supply voltage, so it is not possible to do the optimization as shown above unless another switching regulator is used to generate the gate-charging voltage. This would add much more complexity, so in this design, the V_{gs} will have to be equal to one of the supply voltages.

3.2.2 Optimum for W

The resistive losses are inversely proportional to W, while the capacitive losses are proportional to W. Furthermore, the resistive losses are also related to the duty cycle δ since $I_{avg} = \dfrac{\delta \cdot V_{dd}}{R_{load}}$ (ideally; a more accurate calculation is shown below). The optimal W therefore depends on the duty cycle of the PWM signal.

3.2.3 Complete calculations

To simplify the efficiency calculations, an "effective on-resistance" of the MOSFETs can be defined as

$$R_{on,avg} = \delta \cdot \frac{R_{on,p}}{M_p} + (1 - \delta) \cdot \frac{R_{on,n}}{M_n},$$

in which M_p and M_n are the multiplicity of the PMOS transistor and the multiplicity of the NMOS transistor, respectively. R_{on} is the on-resistance of a unit transistor.

The current through this resistance is not constant, but ramping up and down (I_L from Figure 3-5). Calculating the exact power dissipated by this waveform makes the calculations very complex, because the current levels depend on the duty cycle and on the on-resistance. A simplification is to take the average value of the output current and calculate the power dissipated as

$$P_{res} = I_{L,avg}^2 \cdot R_{on,avg}$$

The simplification here assumes that the amplitude of the triangle component in the current is small compared to the DC component, and that its average value is close to

its RMS value. This last assumption is true, since the RMS value of a triangle wave is $\frac{A}{\sqrt{3}} \approx \frac{A}{1.732}$, in which A is the peak value, while the average is $\frac{A}{2}$.

It is convenient to include the losses in the DC resistance of the inductor (L_1 in Figure 3-6) in the resistance equation as well, because this resistance is directly in series with the on-resistance of the MOSFETs. The complete equation for the resistive power dissipation then becomes

$$P_{res,tot} = I_{L,avg}^2 \cdot \left(R_{on,avg} + R_{coil} \right)$$

The power dissipated in charging and discharging the gates is equal to

$$P_{gate} = \left(C_{gs,p} \cdot V_{gs,p}^2 \cdot M_p + C_{gs,n} \cdot V_{gs,n}^2 \cdot M_n \right) \cdot f_{sw}$$

The gate drivers also consume some power; this is accounted for as a fixed amount per gate area. It should be a relatively small amount, so more complicated calculations are unnecessary.

$$P_{driver} = P_{driver,p} \cdot C_{gs,p} \cdot M_p + P_{driver,n} \cdot C_{gs,n} \cdot M_n$$

Like the gate-source capacitance, the drain-source capacitance is charged and discharged, only this time to V_{dd} and back:

$$P_{drain} = \left(C_{gd,p} \cdot M_p + C_{gd,n} \cdot M_n \right) \cdot V_{dd}^2 \cdot f_{sw}$$

The total amount of power dissipated in the load impedance (the useful power) is

$$P_{out} = \frac{V_{out}^2}{R_{load}},$$

where V_{out} would ideally be equal to $\delta \cdot V_{dd}$. However, the voltage drop over the on-resistance also needs to be taken into account. This can be calculated, using the average on-resistance defined above, to be

$$V_{out} = \delta \cdot \frac{V_{dd}}{1 + \frac{\left(R_{on,avg} + R_{coil} \right)}{R_{load}}},$$

so that

$$I_{out} = \delta \cdot \frac{V_{dd}}{R_{load} + R_{on,avg} + R_{coil}}$$

The overall efficiency now becomes

$$Eff = \frac{P_{out}}{P_{out} + P_{diss}} \; ,$$

in which $P_{diss} = P_{res,tot} + P_{gate} + P_{driver} + P_{drain}$. Filling in all previous equations leads to the equation below.

$$Eff = \frac{P_{out}}{P_{out} + P_{diss}} =$$

$$\frac{\delta^2 V_{dd}^2}{\left((1+Q)^2 Rload \left(\frac{\delta^2 V_{dd}^2}{(1+Q)^2 R_{load}} + \frac{\delta^2 V_{dd}^2 Q}{(1+Q)^2 R_{load}^2} + C_{gate,p} V_{gate,p}^2 f_{sw} M_p + C_{gate,n} V_{gate,n}^2 f_{sw} M_n + P_{drv,p} C_{gate,p} M_p + C_{gd,tot} V_{dd}^2 f_{sw} \right) \right)}$$

in which $Q = \dfrac{\dfrac{\delta R_{on,p}}{M_p} + \dfrac{(1-\delta) R_{on,n}}{M_n} + R_{coil}}{R_{load}}$

This efficiency is now a function of the duty cycle, the switching frequency, the supply voltages, and the transistor parameters. After filling in values for the transistor parameters (in this case the values from the H35 IC process), the voltages (those used in this design) and the switching frequency (as derived in section 4.2), the efficiency can be plotted as a function of the duty cycle. In the figure below, this has been done for a number of different W/L ratios.
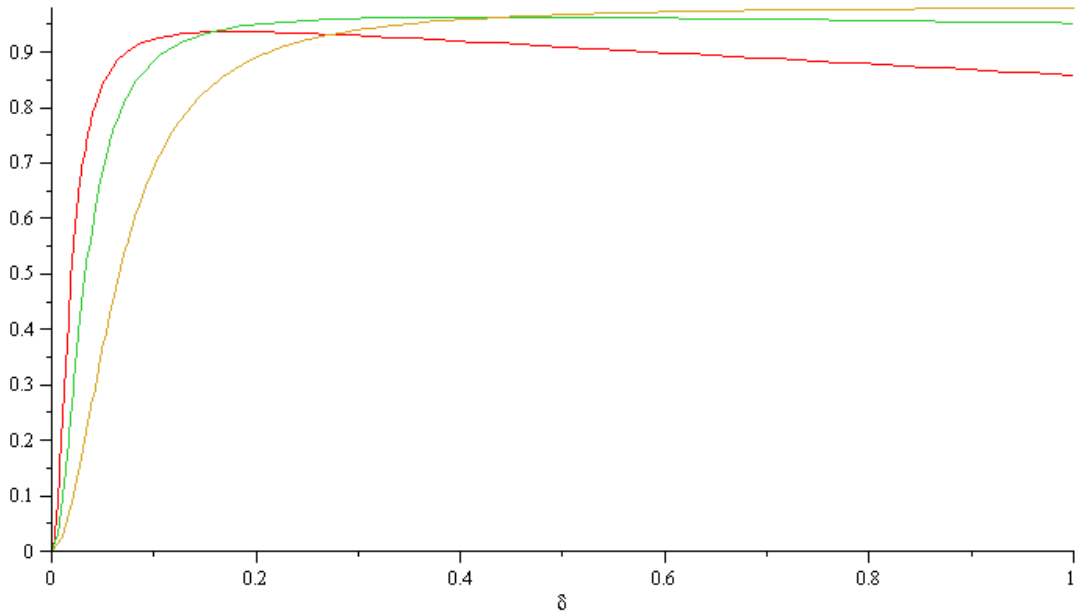
Figure 3-10: Efficiency versus duty cycle for Mn = 378, Mp = 148 (red), Mn = 861, Mp = 586 (green) and Mn = 1423, Mp = 3356 (yellow)

This graph shows that there is not one perfect W/L ratio that will give optimal efficiency over the entire range of duty cycles. If the duty cycle can be expected to be mostly concentrated around one value, then there is one W/L that gives the highest efficiency at that particular value. If there is a certain range of expected values, then it is possible to calculate a W/L that gives the best performance over this range (as shown, for example, in [7]).

3.3 Dynamic transistor sizing

An even better approach however, is to change the W/L dynamically according to the instantaneous value of the duty cycle. Of course it is not possible to change the physical size of the transistors, but placing several different transistors in parallel and switching on only the ones that are needed is a close approximation.
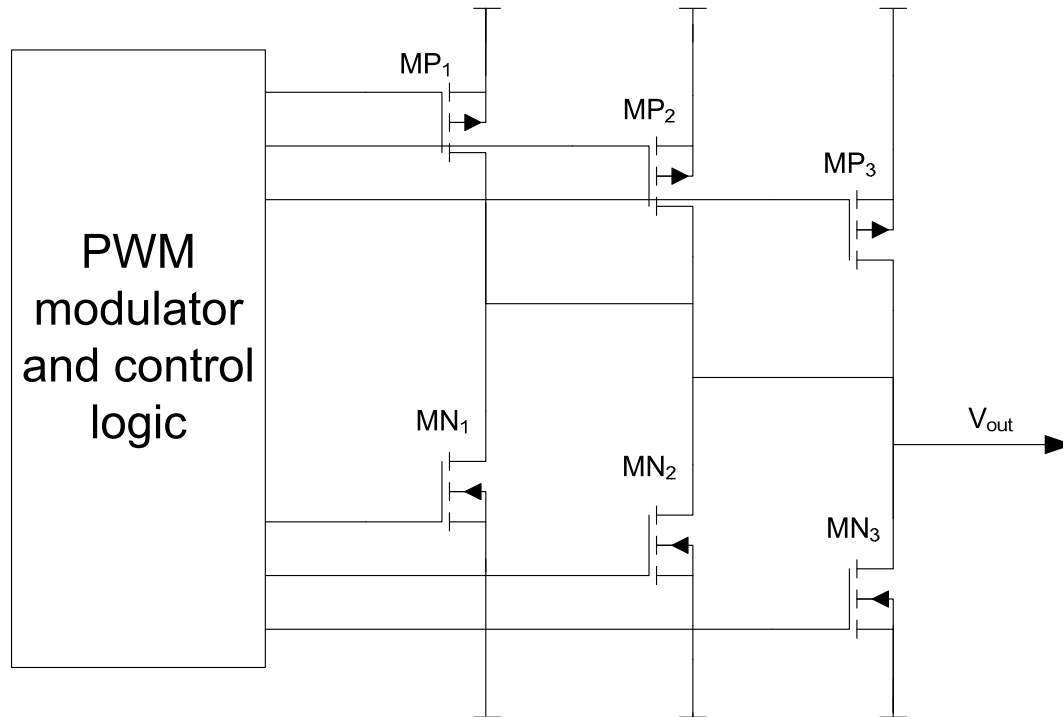

Figure 3-11: Principle of dynamic transistor sizing

The control logic in Figure 3-11 is responsible for deciding which transistors need to be switched on. If, for example, the input signal has a small momentary value for which the highest efficiency is achieved with only MN_1 and MP_1, then the control logic keeps the gates of MN_2 and MN_3 tied to ground, thus saving the current that would otherwise be needed to charge their gates. When the input signal rises to a high value, then additional transistors are activated to keep the on-resistance of the total output stage as low as necessary.

A drawback of this system is that the drain-source capacitances of all transistors are always in parallel, and cannot be turned off. This leads to a smaller efficiency at low duty cycles. Figure 3-12 shows the effect of keeping the Cds fixed at the largest value (for Mn = 1423, Mp = 3356), with all other values identical to the ones used in creating Figure 3-10.
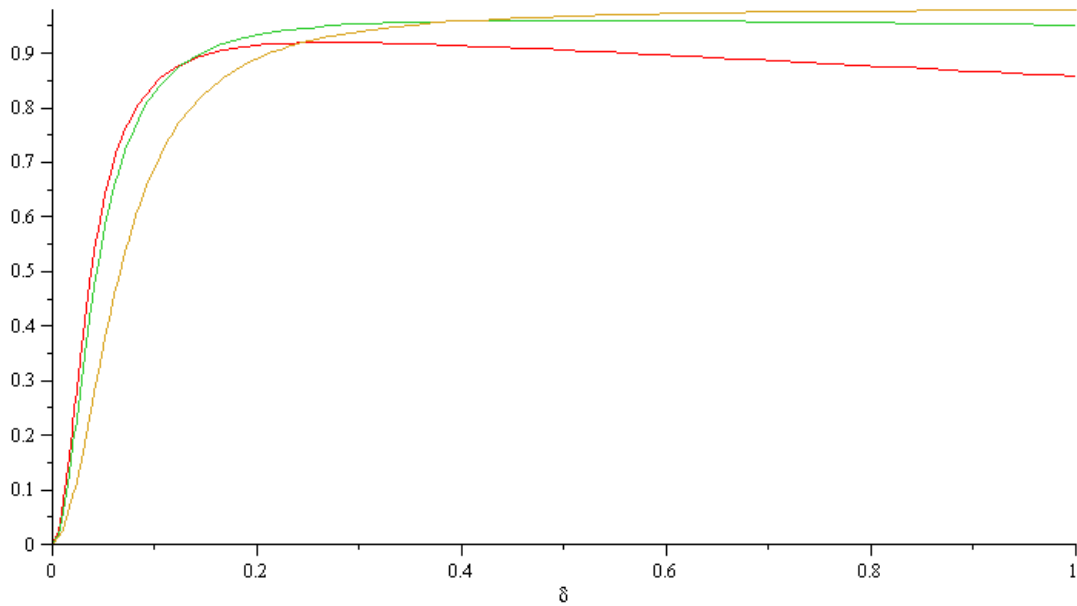
Figure 3-12: Efficiency versus duty cycle for Mn = 378, Mp = 148 (red), Mn = 861, Mp = 586 (green) and Mn = 1423, Mp = 3356 (yellow), with C_{ds} identical in all cases

The difference between the curves is now less dramatic, but still present. It is clear that using the yellow curve rather than the green curve for duty cycles above 0.4 will lead to an increase in efficiency of up to 3 %. At low duty cycles, the difference will grow up to 10 %. Depending on the distribution of the signal values, the increase in efficiency obtained from dynamic transistor sizing can be several percent. In a class D amplifier with more than 90 % efficiency, this is a significant improvement. It is therefore decided to implement dynamic transistor sizing in the current design.

Incidentally, it appears that a patent application [9] was granted describing this technique, just a few months before the calculations above were developed.
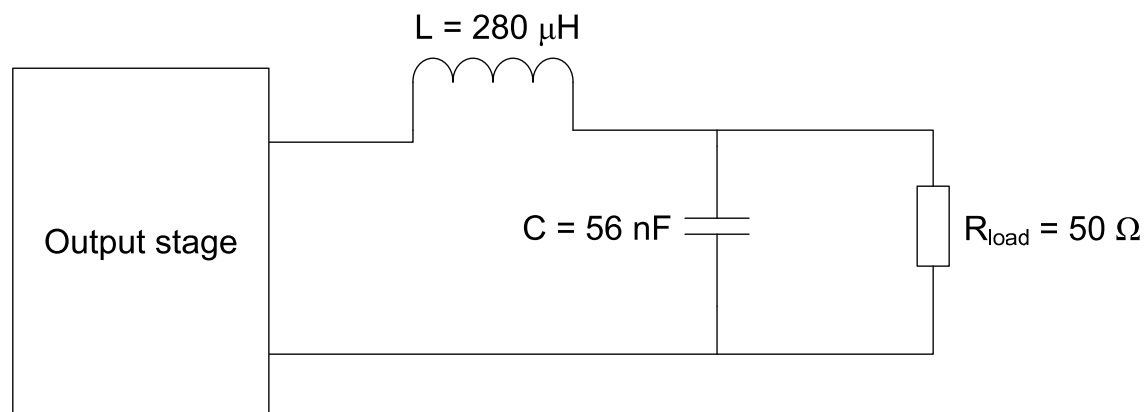
4. Design of power stage

4.1 Filter components

The filter at the output of a class-D amplifier should be a lossless voltage-to-voltage filter. This means that it should be at least a second-order low-pass filter consisting of an inductor L and a capacitor C. The schematic for this is shown below. The corner frequency of the filter should be placed at the edge of the input frequency bandwidth. Placing it any lower will cause it to suppress the signal bandwidth, and placing it any higher will only reduce the attenuation of the switching frequency.

The component values for a two-pole LC low pass filter with a corner frequency of 1 rad/s are given by $L_0 = \sqrt{2}$ and $C_0 = \dfrac{1}{\sqrt{2}}$. Transforming these to 40 kHz and 50 Ω using $L = \dfrac{R_{load} L_0}{\omega_0}$ and $C = \dfrac{C_0}{R_{load} \omega_0}$ results in L = 280 μH, C = 56 nF.

L = 280 μH

Output stage   C = 56 nF R_{load} = 50 Ω

Figure 4-1: Output filter

Commercially available inductors for this application often have a DC resistance of up to about 100 mΩ. The resistance of bondwires is of the same order of magnitude, as is the resistance of package leads and PCB traces. Adding these up, a total parasitic resistance of 300 mΩ will therefore be considered in the calculations.
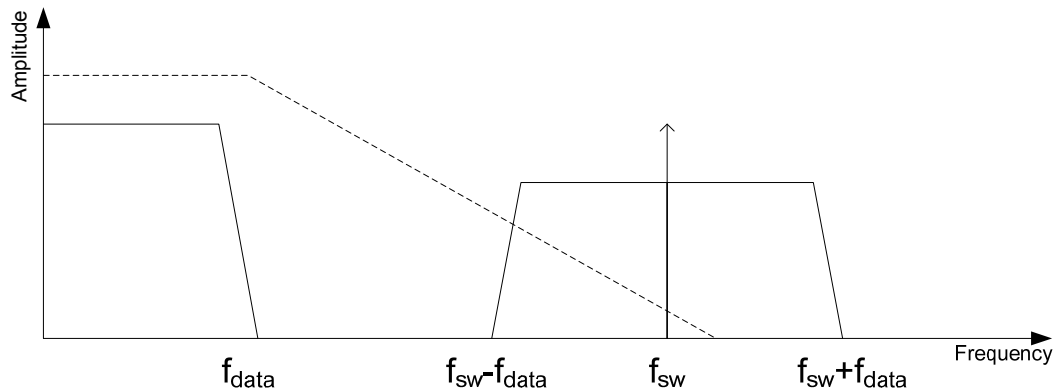
4.2 Switching frequency

From a signal processing point of view, there is a lower bound on the switching frequency: Shannon's theorem states that it should be at least twice as high as the highest signal frequency to prevent aliasing.

In section 3.2 it was shown that the gate charge losses are directly proportional to the switching frequency. Since the resistive losses are not affected by the switching frequency, it can be concluded that from a power efficiency point of view, the only requirement is that the switching frequency be as low as possible.

Another criterion for choosing the switching frequency could be power efficiency: any significant amount of power located at high frequencies ending up in the load

impedance is a source of unwanted dissipation. The specification that the spurious signals should be 44 dB below the carrier already shows that this is not a concern in this design: -44 dB corresponds to a power ratio of 0.004 %.

Figure 4-2 shows the spectrum at the output of the switching stage.

Figure 4-2: Spectrum at output of switching stage

Although the minimum switching frequency is twice the signal bandwidth, a more practical lower bound is the lowest frequency that can be filtered out sufficiently well. If this turns out to be unpractically high, then the filter order must be increased.

The corner frequency of the filter (represented by the dotted line) is placed at the signal bandwidth. If a second-order filter is used, then the attenuation increases by 40 dB/dec from that point. If 44 dB of attenuation is required at the switching frequency, then the switching frequency needs to be 1.1 decades, of 12.6 times higher than the corner frequency.

However, there are also sidebands located next to the switching frequency which need to be attenuated at the output. These sidebands extend down to $f_{sw} - f_{data}$. This requires the switching frequency to be placed at least f_{data} above the -44 dB point.

Taking into account the spread in the filter component values and the switching frequency leads to a further increase. Assuming that the inductor and capacitor have a spread in their values of 10%, and that the switching frequency can change by 5%, the worst-case required switching frequency becomes $(12.6 + 1) \cdot f_{data} \cdot 1.15 = 622$ kHz. This is not inconveniently high, so it will be used as the nominal switching frequency for this design.

4.3 Transistor selection

Now that the switching frequency is known, the only information needed to find the optimal transistor sizes for high efficiency, is the on-resistance and the C_{gs} and C_{gd} of the available transistors. These are calculated from a transient simulation, the results of which are in Table 4-1.

The H35B4D3 process includes several high-voltage transistors. The 20 V transistors are available with two different gate oxide thicknesses, designated as thin (3.3 V) and thick (20 V). The 50 V transistors will not be considered here, because they have a

higher R_{on} than a 20 V transistor with the same C_{gs}, and will consequently provide lower performance.

		NMOS20T	NMOS20H	NMOS20H
W/L		20/0.5	20/0.5	20/0.5
V_{gs} (V)		3.3	3.3	12
C_{gs} (fF)	Min	83.97	15.7	18.3
	Typ	108.5	21.5	24.2
	Max	136.5	27.3	30.2
C_{gd} (fF)	Min	6.84	7.76	8.30
	Typ	8.0	8.60	8.80
	Max	9.75	9.13	8.94
R_{on} (Ω)	Min	235.1	2035	231.4
	Typ	406.6	4268	388.8
	Max	626.5	30k	637.6

Table 4-1: Basic NMOS parameters

A simple calculation can now be performed to determine which transistor will be able to provide the highest efficiency. Since the conduction losses scale linearly with R_{on} and the capacitive losses scale linearly with $C_{gs} \cdot V_{gs}^2$, a Figure of Merit can be defined equal to $FOM = R_{on} \cdot C_{gs} \cdot V_{gs}^2$. The results are in the table below.

	NMOS20T	NMOS20H	NMOS20H
W/L	20/0.5	20/0.5	20/0.5
V_{gs} (V)	3.3	3.3	12
FOM	480k	1000k	1355k

Table 4-2: Figure of Merit of available NMOS transistors

A lower FOM means a lower total power dissipation and therefore a higher efficiency. It is clear that the NMOS20T has the lowest FOM and should therefore be the transistor of choice. The same calculation can be done for the PMOS transistors, the results of which are shown in Table 4-3.

		PMOS20T	PMOS20H
W/L		20/0.6	20/1.1
V_{gs} (V)		3.3	12
C_{gs} (fF)	Min	58.7	19.9
	Typ	76.0	26.0
	Max	95.5	32.3
C_{gd} (fF)	Min	11.15	6.34
	Typ	12.0	6.70
	Max	12.87	6.94
R_{on} (Ω)	Min	455.5	722.5
	Typ	769.5	1141
	Max	1139	1705
FOM		637k	4272k

Table 4-3: Basic PMOS parameters

This shows that the PMOS20T is the best choice. However, there is one important drawback to this transistor. Since the thickness of the gate oxide defines the maximum voltage between the gate and the source, not between the gate and V_{ss}, a thin-oxide PMOS can become difficult to drive. Looking at Figure 4-3, it is clear that $V_{gate,p}$ should never fall below $V_{dd} - V_{gs,p}$. Not only does this mean that an additional supply voltage $V_{gate,p}$ is required (which should be generated in an efficient way), but it also makes the circuit rather fragile. If, during power-up, the supply voltage reaches its nominal value before $V_{gate,p}$ does, then the maximum gate-source voltage is exceeded and the PMOS will likely be damaged.
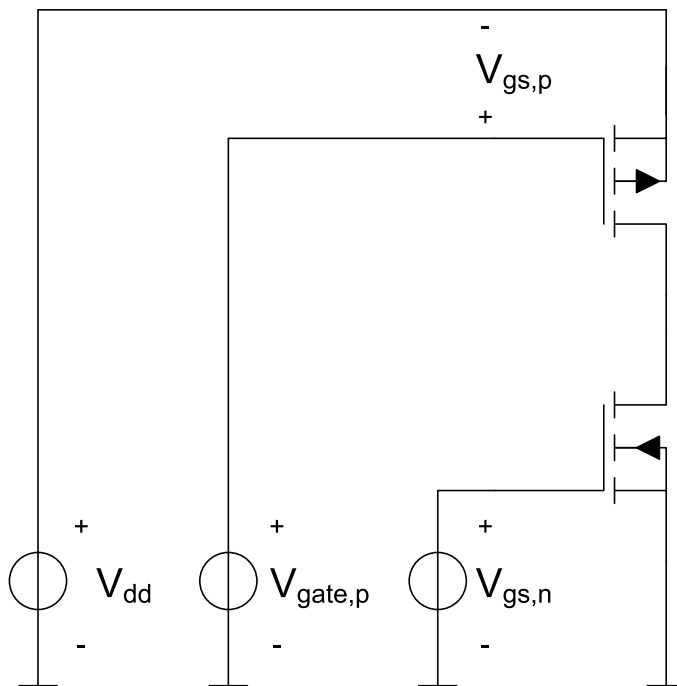
Figure 4-3: Gate drive voltages

Another option would be to use an NMOS instead of a PMOS for the high-side transistor. This is often done in class D designs, because NMOS transistors generally have a lower R_{on} for a given gate area. Unfortunately, this leads to the same difficulties with driving and reliability, because the high-side NMOS would require a gate voltage higher than V_{dd}.

Logically, it follows that there are a total of four possible combinations. Table 4-4 shows all the options and their advantages and drawbacks.

Low-side transistor	High-side transistor	Efficiency	Driveability
NMOS	NMOS	High	Difficult
NMOS	PMOS	Medium	Easy
PMOS	NMOS	Medium	Very difficult
PMOS	PMOS	Low	Difficult

Table 4-4: Selection table for transistor types

The easiest and safest option is therefore to choose a PMOS transistor for the high-side transistor that can withstand the full V_{dd} swing at its gate. The PMOS20H is the transistor of choice.

4.4 Transistor parameters

Now that the transistor types have been chosen, the optimal transistor sizes to reach maximum efficiency can be calculated. The transistor parameters derived in section 4.3 and the equations derived in section 3.2 are implemented in a computer algebra system (Appendix C), and the optimal transistor sizes are calculated using numerical optimization. The results for three different duty cycles are shown in Table 4-5.

Duty Cycle	Maximum efficiency	Optimal NMOS size (unit transistors)	Optimal PMOS size (unit transistors)
0.1	92.4 %	378	148
0.25	95.8 %	861	586
0.8	97.7 %	1423	3356

Table 4-5: Optimal transistor sizes for three different duty cycles

Plotting the efficiency as a function of the duty cycle leads to the graph below. It is clear that none of the solutions lead to maximum efficiency over the entire range. The red curve (optimized for $\delta = 0.1$) is efficient for low duty cycles, but drops down to about 80 % efficiency for high duty cycles. The yellow curve (optimized for $\delta = 0.1$) reaches more than 95 % efficiency at high duty cycles, but drops down quickly at low duty cycles.
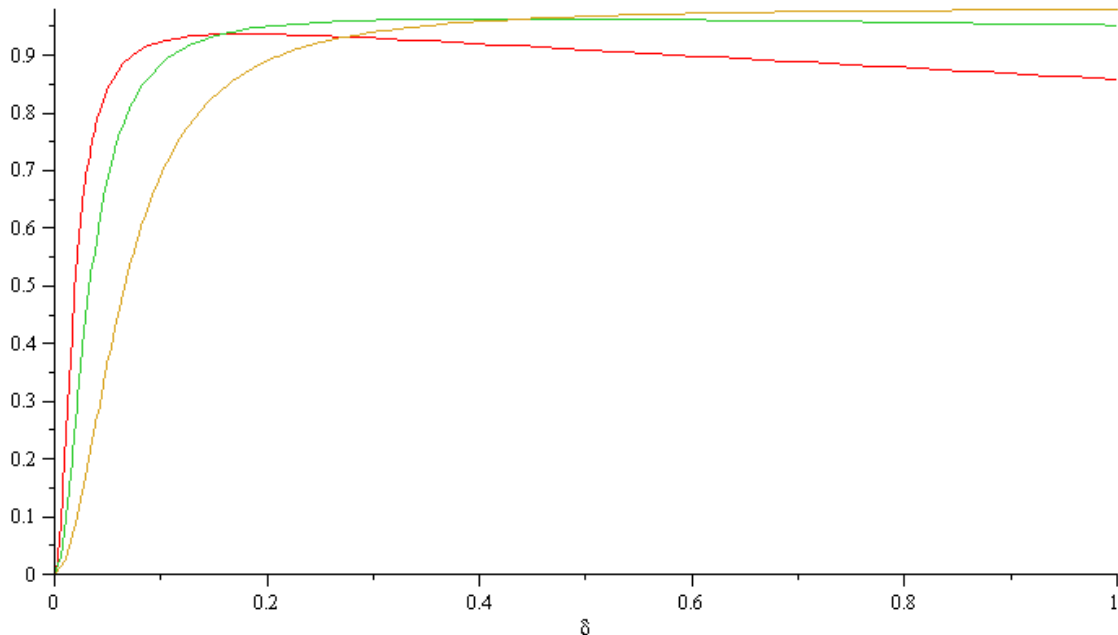
Figure 4-4: Efficiency as a function of the duty cycle for the three different transistor sizes calculated above. This is the same graph as Figure 3-10

As described in section 3.3, the overall efficiency can be increased by dynamically changing the transistor sizes as a function of the duty cycle. This does change the curves somewhat, because the C_{gd} of the largest transistor will be present with the smaller transistors as well.

A second optimization has to be performed, taking this into account. Since the expected maximum duty cycle is about 0.8 (as derived in chapter 1), the maximum transistor size is chosen for this value (NMOS size 1423, PMOS size 3356). The calculations can now be run again, as shown in Appendix C. This leads to the following results:

Duty Cycle	Maximum efficiency	Optimal NMOS size	Optimal PMOS size
0.1	84.8 %	585	184
0.25	94.7 %	1231	669
0.8	97.7 %	1423	3356

Table 4-6: Optimal transistor sizes when using dynamic transistor sizing

The efficiency as a function of duty cycle now becomes the graph below. Although the difference between the three curves is less dramatic than in Figure 4-4, it still makes sense to apply dynamic transistor sizing.
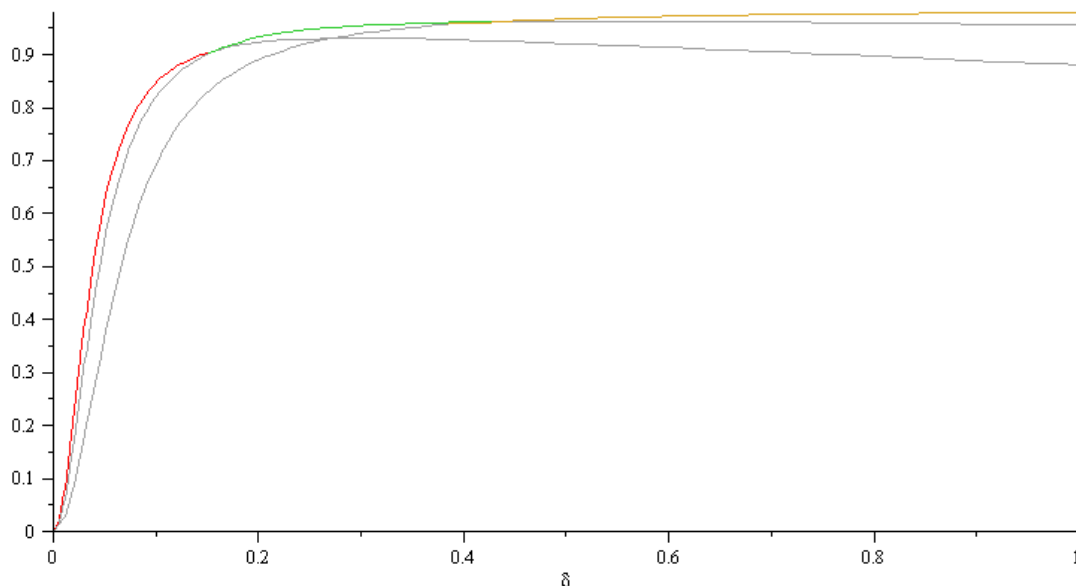
Figure 4-5: Efficiency as a function of duty cycle, using the values from Table 4-6 and dynamic transistor sizing

The coloured curves in Figure 4-5 show the sections that will be used. It is clear that this curve is above the gray curves, and therefore has a higher efficiency. Comparing Figure 4-5 with Figure 4-4, the expected increase in efficiency is about one or two percent, for an input signal that is distributed evenly across the input range. It is decided to use these three curves, since dividing the output stage into more than three sections (i.e. adding another curve to Figure 4-5) would not significantly increase the efficiency.

To get an idea for the accuracy required in the implementation of these transistors, the efficiency can be plotted as a function of transistor sizes, for a certain duty cycle. From Figure 4-6 it is clear that the optimum is rather flat, so it is permissible to deviate even a few hundred units from the calculated sizes if this is required for e.g.

layout reasons or to save chip area. In this design however, there is no need to be this frugal and the transistors are designed at their optimal sizes.
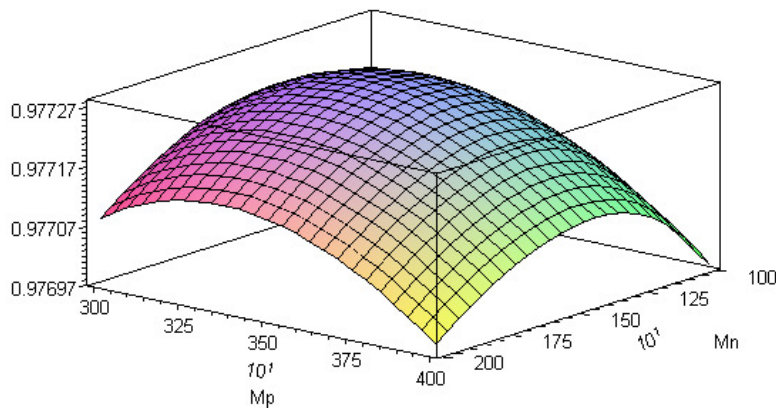

Figure 4-6: Efficiency as a function of Mp and Mn, for δ=0.8

The handover points can be read from Figure 4-5 to be around δ = 0.15 and δ = 0.45. For maximum accuracy, the exact values to be implemented in the circuit will be determined later from a Spectre simulation.

4.5 Gate drivers

Driver circuits are needed to make sure that the gate capacitance of the output transistors are driven quickly enough. Two problems are caused when the gates are driven too slowly. Firstly, this will cause the output transistors to spend time in a region where their on-resistance is higher than designed for, causing lower efficiency. Secondly, the width of the PWM pulses will be poorly defined, giving rise to distortion in the signal.
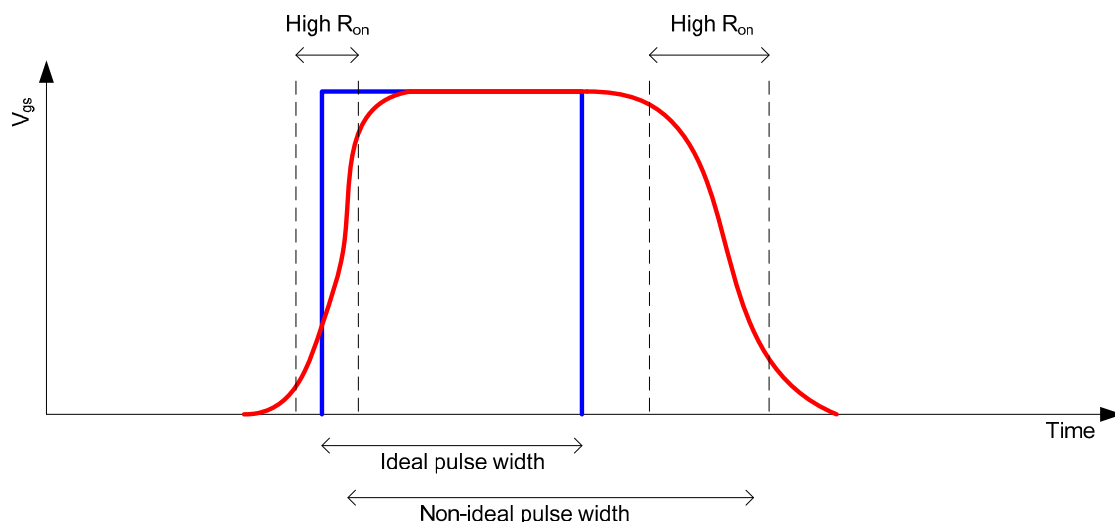
Figure 4-7: Effect of gate being driven too slowly

It is hard to make an accurate prediction of the amount of distortion produced by inaccuracies in the timing. The in-band distortion requirement is -30 dB, or about 3 % in voltage. Since the output voltage depends linearly on the duty cycle of the PWM signal, the duty cycle should be accurate to within this 3 %. The duty cycle is again

linearly dependent on the pulse width, so also the pulse width should be accurate to within 3 %.

Since $P_{res} = I_{L,avg}^2 \cdot R_{on,avg}$ and $I_{out} = \delta \cdot \dfrac{V_{dd}}{R_{load} + R_{on,avg} + R_{coil}}$, an inaccuracy in R_{on} will only cause a significant change in the power dissipation if it is a significant fraction of $R_{load} + R_{on} + R_{coil}$, which it is not. A 3 % inaccuracy in the pulse width is therefore more than accurate enough to keep the efficiency close to the calculated values.

Given a switching frequency of 622 kHz and a minimum duty cycle of 10 %, the smallest pulses are 0.16 µs wide. An accuracy of 3 % means that the pulse width should be accurate to within about 5 ns. This means that the rising and falling edges should each be less than 2.5 ns wide. Taking some margin for safety, the drivers will be designed to drive the gates within 1.5 ns (worst-case).
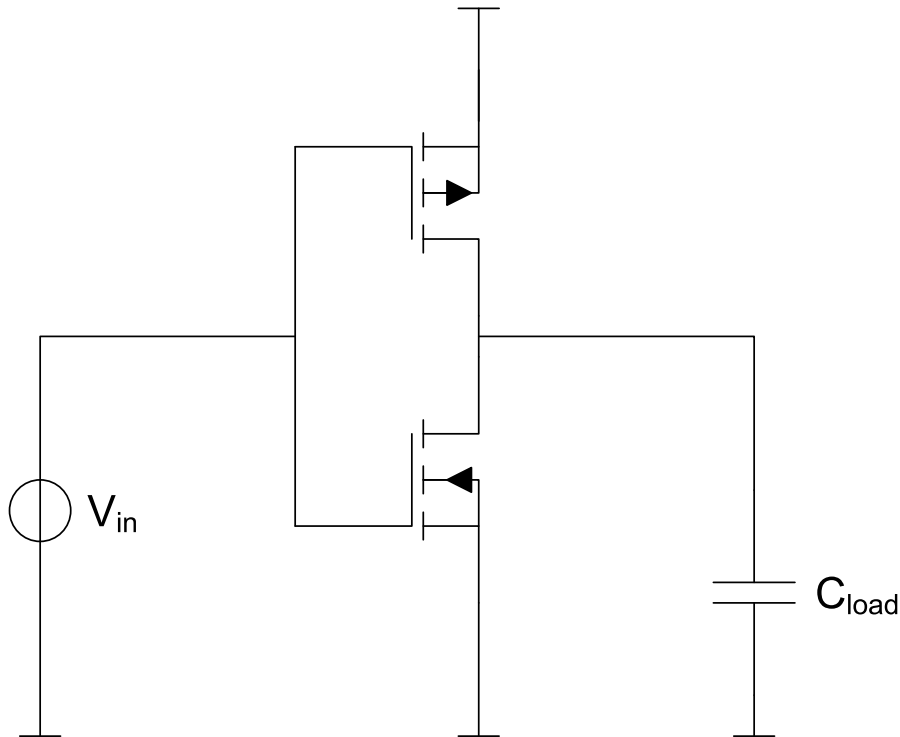
Figure 4-8: Gate driver

The driver consists of an inverter. Calculating the resulting rise time is not difficult (it should be about $R_{on} \cdot C_{load}$), but simulating is more accurate and rather straightforward. To make an easily scalable driver, a load $C_{load} = 10$ pF is applied, and the width of the transistors is tuned until the driver is able to drive C_{load} within 1.5 ns. This leads to a driver size that can later be scaled to drive the C_{gs} of the output transistors.

The driver circuit itself also has a gate capacitance that needs to be driven by the circuits that come before it. Therefore, the driver will be driven by another driver, driven by a third driver, until the gate capacitance is small enough to be driven by the preceding circuits.

The smaller the input capacitance of the driver, the shorter the driver chain will be. The driver circuit has a "charge gain" equal to $\dfrac{C_{load}}{C_{in}}$, which will be used to determine the length of the driver chain required.

The driver chain for the NMOS output transistor consists of standard 0.35 μm low-voltage transistors, because the V_{gs} of the output NMOS is only 3.3 V. The results are in Table 4-7.

	Min	Typ	Max
NMOS size		6	
PMOS size		17	
Rise time (ns)	0.82	1.0	1.4
Fall time (ns)	0.80	1.0	1.3
Input capacitance (fF)	310	310	347
Charge gain	29	32	32
Power consumption (μW/MHz)	2.23	2.23	2.52

Table 4-7: NMOS driver specifications

The PMOS output transistor is driven by a chain of NMOS20H and PMOS20H transistors, because these need to drive 12 V to the PMOS gate. The results in Table 4-8 show that the charge gain is significantly lower than for the NMOS drivers, which means that the PMOS driver chain will probably be longer than the NMOS driver chain.

	Min	Typ	Max
NMOS size		18	
PMOS size		38	
Rise time (ns)	0.83	1.0	1.3
Fall time (ns)	0.81	1.0	1.3
Input capacitance (pF)	1.3	1.6	1.9
Charge gain	5.2	6.1	7.5
Power consumption (μW/MHz)	147	164	175

Table 4-8: PMOS driver specifications

Since the output transistors are divided into three parts (which are dynamically enabled and disabled), the drivers will also need to be divided into three parts. This is schematically depicted in Figure 4-9.
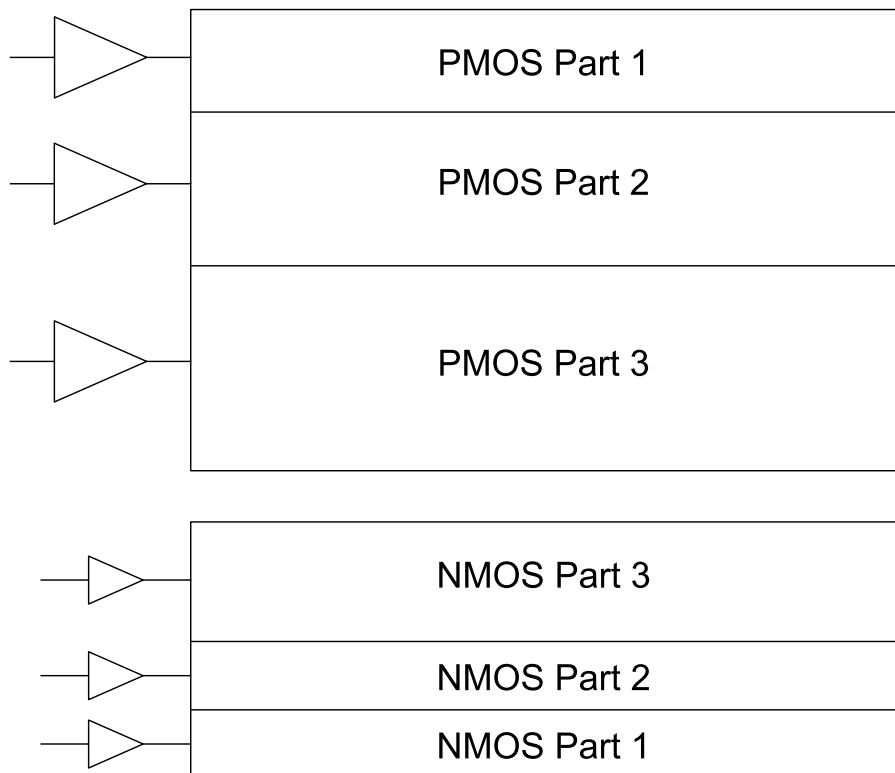
Figure 4-9: Block diagram of output transistors and drivers

A complete list of all transistor sizes is shown in Table 4-9. The driver chains are extended until one of the transistors reaches unit size.

NMOS total 1423	Output transistor parts	Driver 1	Driver 2	Driver 3
	585	116	4	1
		41	2	1
	646	128	4	1
		46	2	1
	192	39	2	1
		14	1	1

PMOS total 3356	Output transistor parts	Driver 1	Driver 2	Driver 3	Driver 4	Driver 5
	184	23	4	2	2	2
		11	2	1	1	1
	485	61	10	2	2	2
		29	5	1	1	1
	2687	334	55	10	2	2
		159	27	15	1	1

Table 4-9: Output transistor and driver sizes

4.6 Dynamic transistor size circuit

The three output stages need to be enabled or disabled according to the momentary duty cycle, or, equivalently, the momentary input voltage. This can be achieved with the circuit shown below. Note how the driver circuits now have a "disable" pin which causes the corresponding PMOS gate to be pulled high, and the NMOS gate to be pulled low, effectively disabling both transistors.
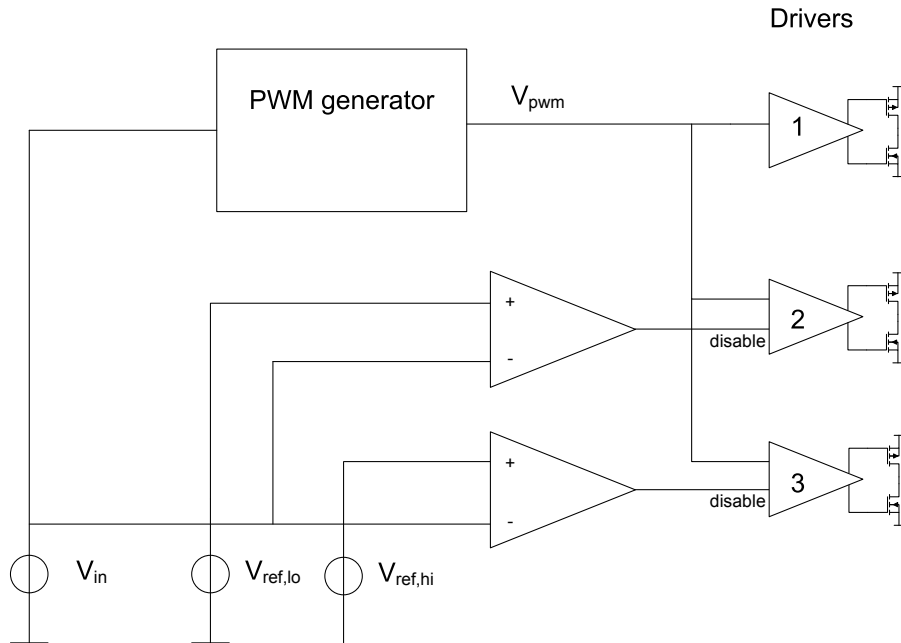
Figure 4-10: Dynamic transistor sizing circuit

The comparators do not need to be particularly fast, since they have to track the input signal which has a limited bandwidth. However, in order to maximize the efficiency gained by using dynamic transistor sizing, the comparators should respond within one cycle of the PWM signal, which means a reaction time of about 1 μs.

The reference voltages will be implemented by passing a reference current through a resistive divider. There is a certain amount of inaccuracy in the resulting voltage, mainly caused by mismatch in the components used to generate it from a stable reference voltage. These components are shown below.
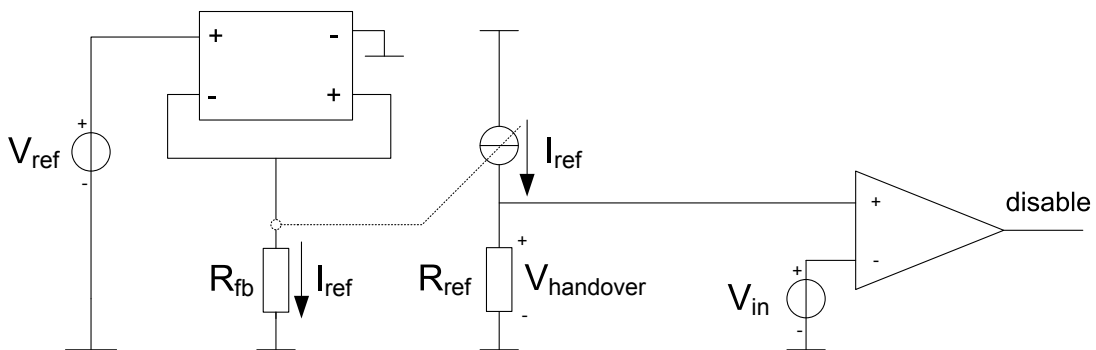
Figure 4-11: Practical implementation of handover voltage

As shown in Figure 4-6, the size of the transistors is not extremely critical. The maximum inaccuracy in the handover voltages is therefore chosen to be 1/20th of the complete range, or 50 mV. Looking at the intersection between the red and green curves in Figure 4-5, an inaccuracy of 50 mV will lead to a loss in efficiency of less than 1 %. Investing more effort in the accuracy of the implementation of the handover voltages will therefore not lead to a significantly higher efficiency.

Section 6.1.1 shows that the bandgap reference has an inaccuracy of +/- 25 mV. This means that only +/- 25 mV is left of the +/- 50 mV specified above. This has to be distributed over all sources of inaccuracy between the bandgap voltage and the comparator. These are:

- Mismatch of components in the bandgap reference
- Offset in the current reference
- Mismatch of the resistors
- Offset in the comparator

Since these are all statistical processes described by a normal distribution, their contributions have to be added quadratically. The total allowed $\sigma_{Voff,total}$ = 5 mV.

The offset voltage caused by resistor mismatch equals $\sigma_{Voff,res} = \dfrac{I_{ref} A_R}{\sqrt{WL}} = 790\ \mu V$.

Dividing the remaining offset over the other sources means that each can contribute 2.5 mV. Converting this to specifications for each circuit leads to the values in Table 4-10.

Mismatch of components in the bandgap reference	2 mV
Offset in the current reference	2 nA
Offset in the comparator	2.5 mV

Table 4-10: Mismatch values for each circuit

4.7 Avoiding clock overlap

As stated in section 3.1, it is important to ensure that both transistors will never be turned on at the same time, since this would short circuit the supply voltage. A straightforward way of preventing this is to add a circuit that only allows a transistor to turn on when the other is off, and vice versa. The circuit shown below achieves this.
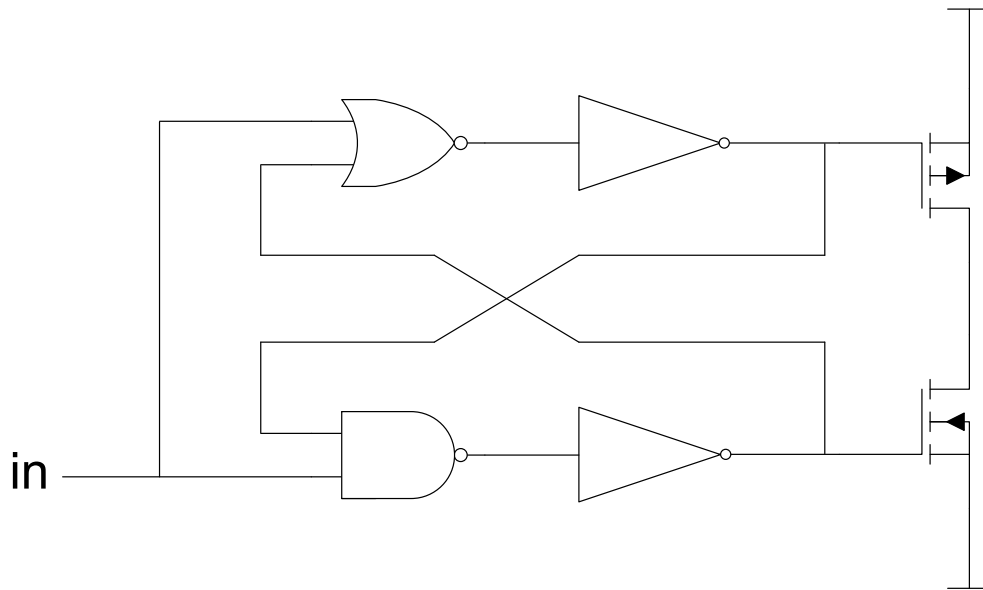
Figure 4-12: Non-overlap circuit

The two inverters shown represent the gate drivers, since they have an inverting transfer. Effectively then, Figure 4-12 consists of an AND gate and an OR gate driving the transistors' gates.

It works as follows: Suppose the input is "low", then both transistor gates are "low". If the input changes to "high", then the OR gate starts pulling the PMOS gate toward V_{dd}. The NMOS gate however, will not be driven "high" until both inputs of the AND gate are "high", in other words, until the PMOS gate has been charged to V_{dd} (turning off the PMOS). A similar reasoning holds for the high-to-low transition.

The body diodes of the MOSFETs prevent the output voltage from spiking when the current through the inductor is turned off. If both transistors are turned off, then the output node is pulled above V_{dd} or below V_{ss}, causing the respective body diode to become forward biased and clamp the output voltage to one diode drop above V_{dd} or below V_{ss}. Figure 4-13 shows how the body diode of the NMOS starts conducting when the NMOS has been turned off but the PMOS has not yet been turned on.
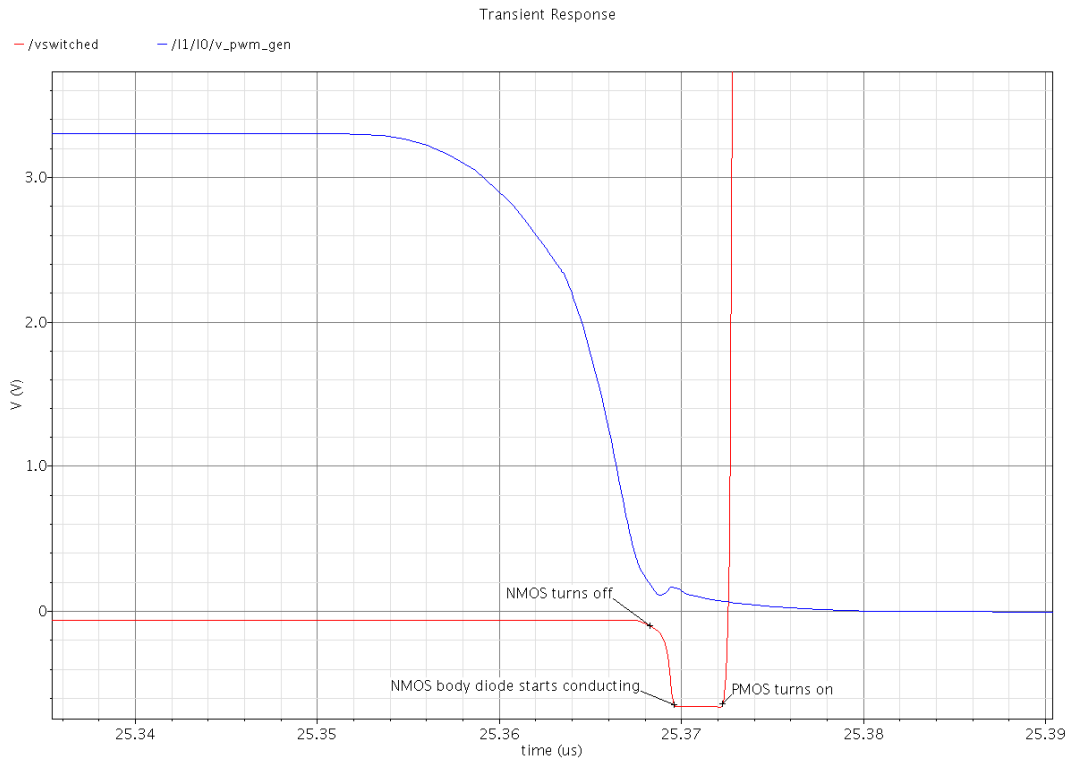
Figure 4-13: NMOS body diode conducting

4.8 Level shifters

The PMOS power transistors are driven with a gate voltage of 12 V, while the low-power circuits run on 3.3 V. It is therefore necessary to shift the voltage levels up to drive the gates and down to provide the feedback. The up-shifting circuit is shown below.
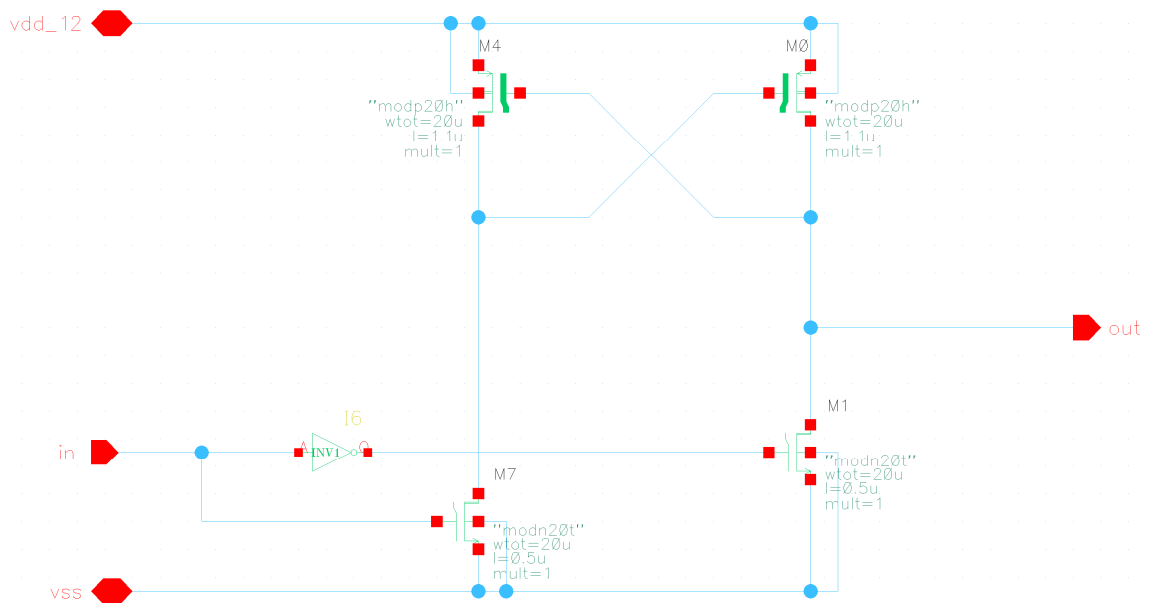
Figure 4-14: 3.3 V to 12 V level shifter

The down-shifting circuit consists of a 20 V resistant NMOS transistor (M0) with its gate connected to 3.3 V. This ensures that its source can never rise above 3.3 V - V_{th}, since this would cause the transistor to turn off.
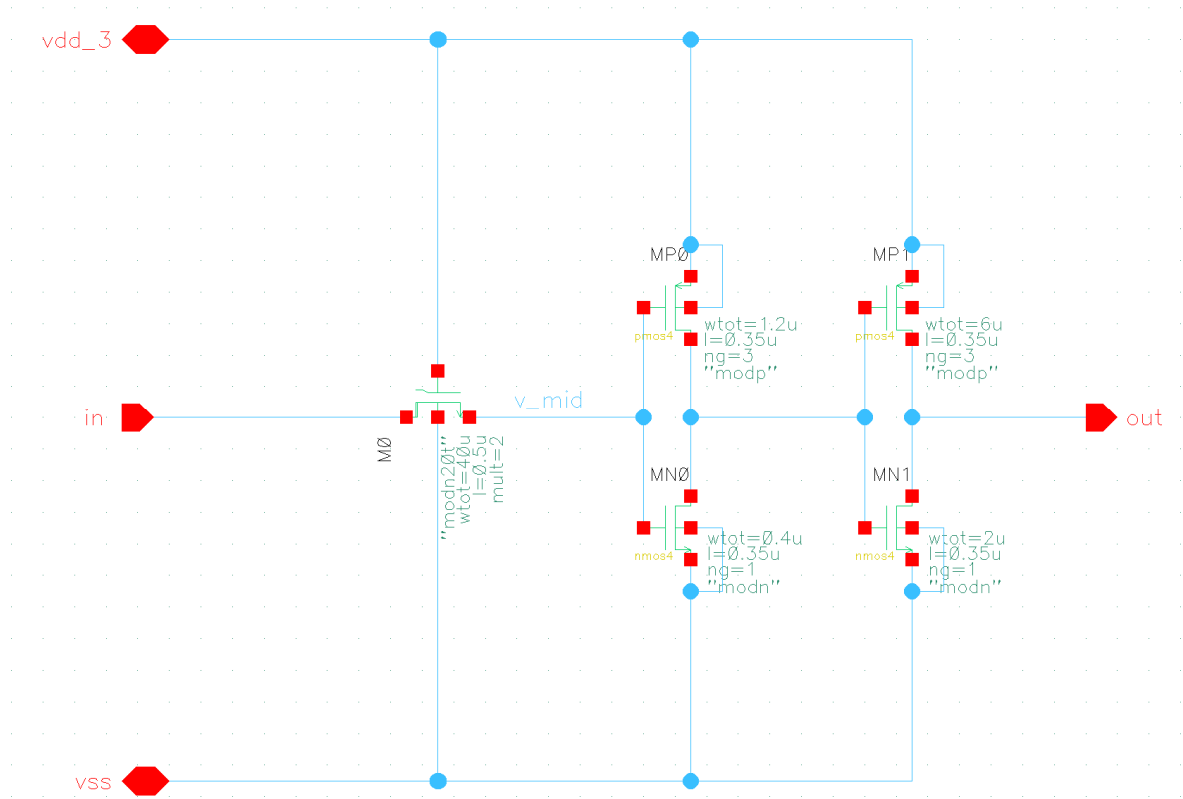
Figure 4-15: 12 V to 3.3 V level shifter

5. Design of input stage

The next stage in the design process is to design the low-power signal processing circuits. The main part of these circuits is the pulse-width modulator that converts the analogue input signal into a stream of pulses.

As stated in section 3.1, the simplest way of creating a PWM signal is to compare the input signal with a triangle wave. In this chapter, the design of a triangle wave generator and a comparator will be described, which together form a PWM generator. The main performance issue that needs to be considered is the linearity of the transfer from input voltage to output duty cycle. This is determined by the quality of the triangle wave and the speed and resolution of the comparator.

5.1 PWM generator

The schematic of the PWM generator is shown below.
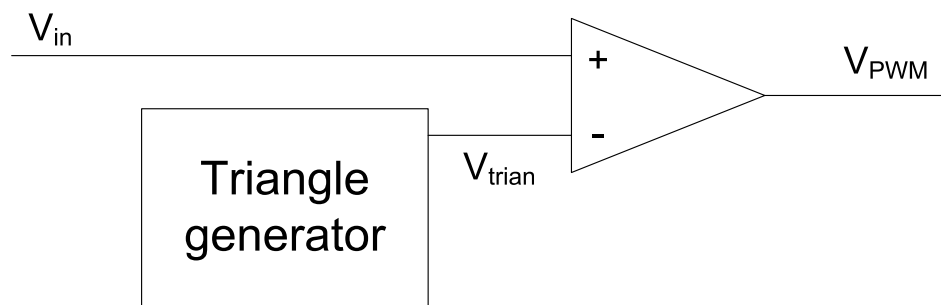
Figure 5-1: PWM generator

The triangle generator needs to provide a triangular wave with accurately straight edges. The voltage levels also need to be accurate, to provide the correct input voltage range.

The comparator needs to have similar delays when switching from low to high and when switching from high to low. This is because a difference in the delays Δ gives an offset in the duty cycle equal to $(\Delta_1 - \Delta_2) \cdot f_{sw}$. If this offset should be smaller than, for example, 5%, then the difference in switching delay should be smaller than 80 ns.

A more important requirement on the comparator is the resolution it can achieve. If the resolution is too low, this will lead to excessive quantization noise.

5.2 Triangle generator

The required frequency of the triangle wave was determined in section 4.2 to be 622 +/- 5 kHz. The voltage swing of the triangle wave determines the input voltage range. To allow some headroom for biasing (described below), the input range is chosen to be between 1.0 and 2.0 V. The peaks of the triangle wave should have an overshoot of less than 100 mV, to enable the maximum duty cycle of 80 % to be reached.

The most straightforward way to generate a voltage ramp is to charge a capacitor with a constant current. The voltage is then given by $V(t) = V(0) + \dfrac{I \cdot t}{C}$. This ramp can be

transformed into a triangle wave by setting two thresholds and reversing the direction of current when the voltage reaches one of these thresholds. The basic circuit is shown in Figure 5-2.
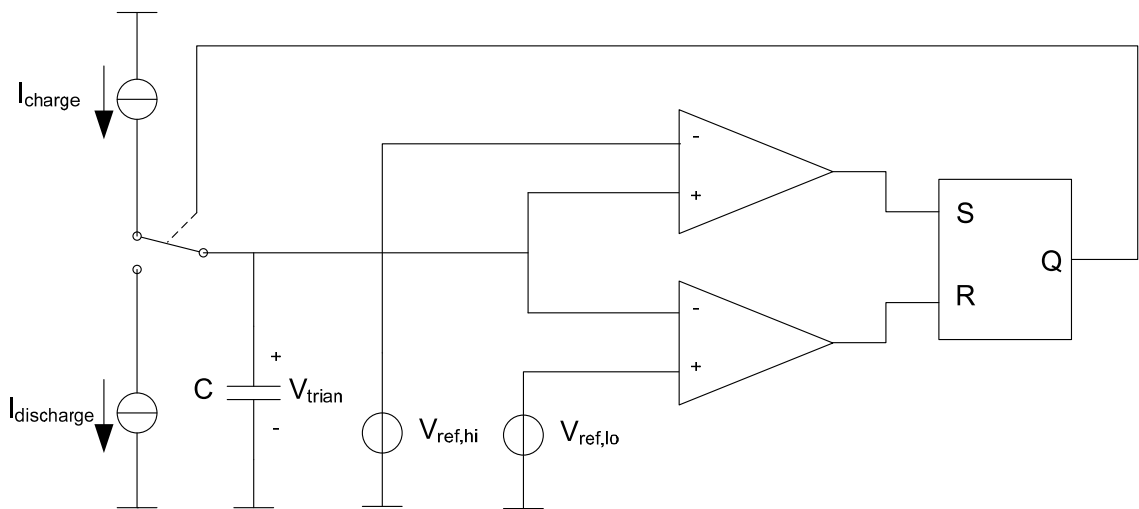
Figure 5-2: Basic circuit of the triangle generator
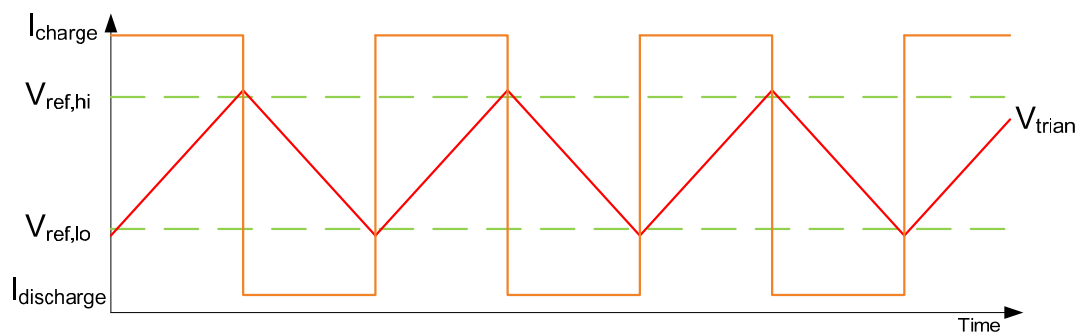
Figure 5-3: Waveforms in the triangle generator

Since on-chip capacitors have tolerances of about +/- 20 %, and the on-chip current reference is only accurate to within +/- 30 %, it is not possible to make an accurate frequency without using some external reference. One way to do this is to place the capacitor externally and make it adjustable. Trimming the capacitor value after production makes it possible to compensate for any inaccuracy in the charging current and thereby provide an accurate switching frequency.

The required capacitance can be calculated from $C = \dfrac{I_{cap}}{2V_{span}f_{sw}}$. A lower bound for the capacitance is given by the parasitic capacitances of the bond pad and package leads, which are on the order of 1 pF. Choosing a charging current of 10 µA leads to a capacitance of 8 pF, which can be implemented without too much disturbance from the parasitic capacitances.

Overshoot is caused by offset and delay in the comparators. Offset can work in both directions (overshoot and undershoot), but delay only makes the triangle overshoot its boundaries. Since the capacitor voltage changes by $\frac{V_{span}}{2f_{sw}}$ = 1.2 mV/ns, an overshoot of 100 mV is reached in 80 ns. The comparator therefore needs to respond no slower than this.

A good specification for the comparator would then be that it should have a reaction time of about 50 ns, so that it causes no more than 63 mV of overshoot. The offset can then cause another 37 mV in either direction without causing problems.

Another issue that affects the overshoot is the resolution of the comparator. The resolution needs to be substantially smaller than the allowable overshoot, to ensure that the comparator can reliably determine whether the signal has passed a threshold. An estimate for the required resolution is about 5 mV.

As mentioned before, it is difficult to determine an accurate relation between nonlinearity and distortion, so an order-of-magnitude approach is used. The distortion requirement is -30 dB, or 3 %. Taking some margin for safety, the ramp voltage is chosen to be linear to within 1 %. This means that the charging/discharging currents should also be constant to within 1 %.

This accuracy is achieved by using a cascoded current mirror. The mirror (MN1 and MN2 in Figure 5-4) provides the current, while the cascode (MN3 and MN4) ensure that the voltage over the mirror stays constant, reducing channel-length modulation. The cascode is designed with a large W/L to increase its gain, while the mirror is designed with a small W/L to increase its matching.

Figure 5-4: Cascoded 1:20 NMOS current mirror

The exact value of the charging and discharging currents is not critical, so a maximum error of +/- 10 % is chosen. Since the charging current delivered by the PMOS current mirror has to go through the NMOS current mirror as well, and since mismatches have to be added quadratically, this means that each current mirror should match its currents within 7 %. Using the matching calculations from Appendix C, the transistor sizes shown in Figure 5-5 are calculated.

It is also necessary to switch the current mirrors on and off. The most straightforward way is to insert another transistor in series with the mirror's output lead. Unfortunately, there is not enough voltage headroom in the PMOS current mirror to accommodate another transistor. Therefore the PMOS current mirror will be kept turned on at all times, and only the NMOS current mirror will be switched on and off. The NMOS current does need to be twice as large to make this work, but this is not a major issue since the power dissipation of the triangle generator is still much smaller than that of the output stage.

Completely blocking the NMOS current mirror causes the current mirror to start up too slowly, since it has to charge its gate-drain capacitance. Therefore, the current is diverted to V_{dd} during the charging phase. The complete charging and discharging circuit is shown below.

Figure 5-5: Complete charge/discharge circuit

The reference voltages that determine the upper and lower limits of the triangle wave are generated by passing a copy of the standard reference current through a resistive divider. Since the reference current from the current reference is 1 µA, the resistive divider has to consist of two 1 MΩ resistors.

The accuracy of the reference voltages is then determined by the matching of these resistors to the resistor in the current reference. Because the resistors are 4 µm wide and 3167 µm long, the three-sigma mismatch with the current reference resistor is $\frac{3 \cdot A_R}{\sqrt{WL}} = 0.17$ %, which translates to a voltage offset of 1.7 mV. This is far less than the maximum allowable offset derived above.

The latch is composed of two NOR gates from the Austriamicrosystems library. Finally, circuits are added to disable the triangle generator when needed. When the disable pin is high, all current mirrors are turned off, as are the comparators, while the

output voltage is left floating. This makes it possible to apply an external triangle wave for testing purposes.

The complete schematic can be found in Appendix D. A transient simulation shows that the circuit generates a high-quality triangle wave.
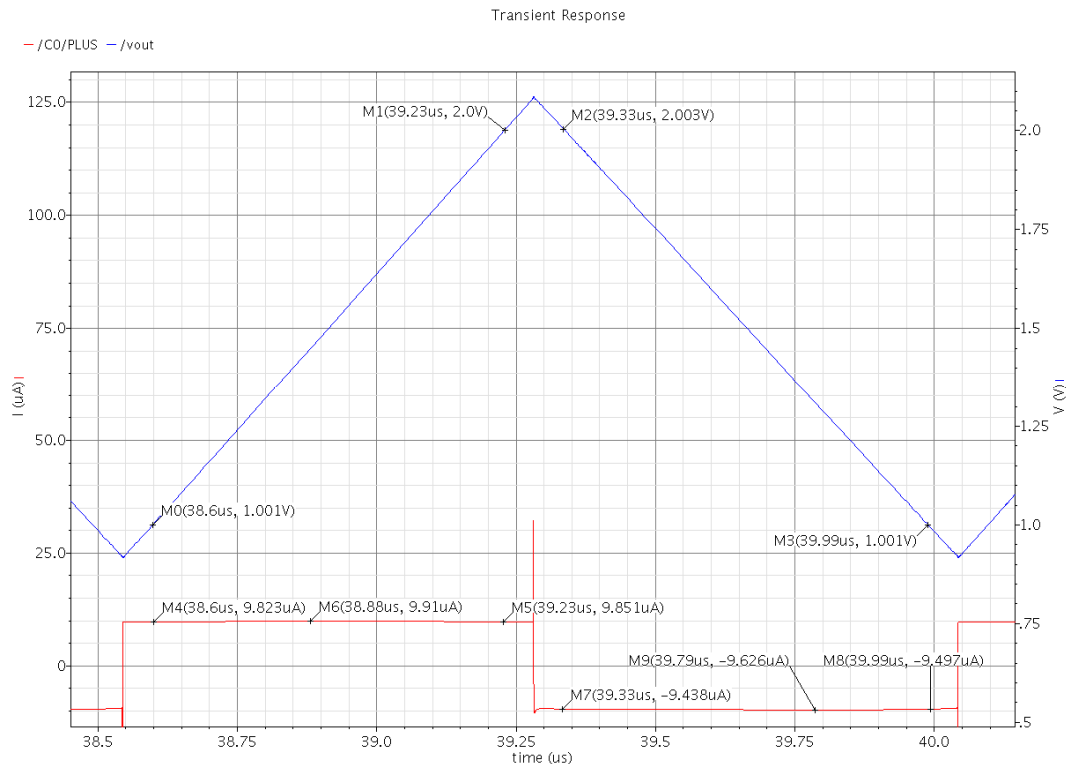
Figure 5-6: Output voltage of triangle generator (blue) and charge/discharge current of capacitor (red)

Figure 5-6 shows that the capacitor current is constant to within about 0.8 % during the charge phase, and to within about 2 % during the discharge phase. This is less accurate than specified before, but there is no voltage headroom left to further improve the output impedance of the charge/discharge currents. The specification was a conservative estimate however, so it is left to simulations of the complete circuit (section 7.2) to find out if the required distortion performance is achieved.

5.3 Comparator

Comparators are used in three different places in the complete class-D amplifier circuit:

- Triangle generator (2)
- PWM generator (1)
- Output stage, providing dynamic multiplicity (2)

It would save much design time if one circuit could be used for all these applications. The requirements for each of these applications are summarized in Table 5-1.

	Triangle	PWM gen	Output stage
Load capacitance	7 fF (NOR21)	141 fF (3x(NOR33 + NAND34))	28 fF (NOR24)
Maximum delay	50 ns	Difference < 80 ns	1 us
Resolution	5 mV	1 mV	1 mV
Sigma input offset voltage	10 mV	10 mV?	2.5 mV

Table 5-1: Requirements for the comparators

Combining the strictest specifications shows that a universal comparator should drive 141 fF within 50 ns, with a sigma input offset voltage of 2.5 mV. The resolution should be 1 mV.

The driving capacity of the comparator is determined by its output stage. The output stage consists of an NMOS and a PMOS transistor to discharge and charge the load capacitance. Because they should never be on at the same time, it is possible to connect their gates together and form an inverter. The W/L of the two transistors can be tuned until they can drive the load within the specified time. To allow for some additional delay in the first stage, the output stage should actually drive the load faster than required.
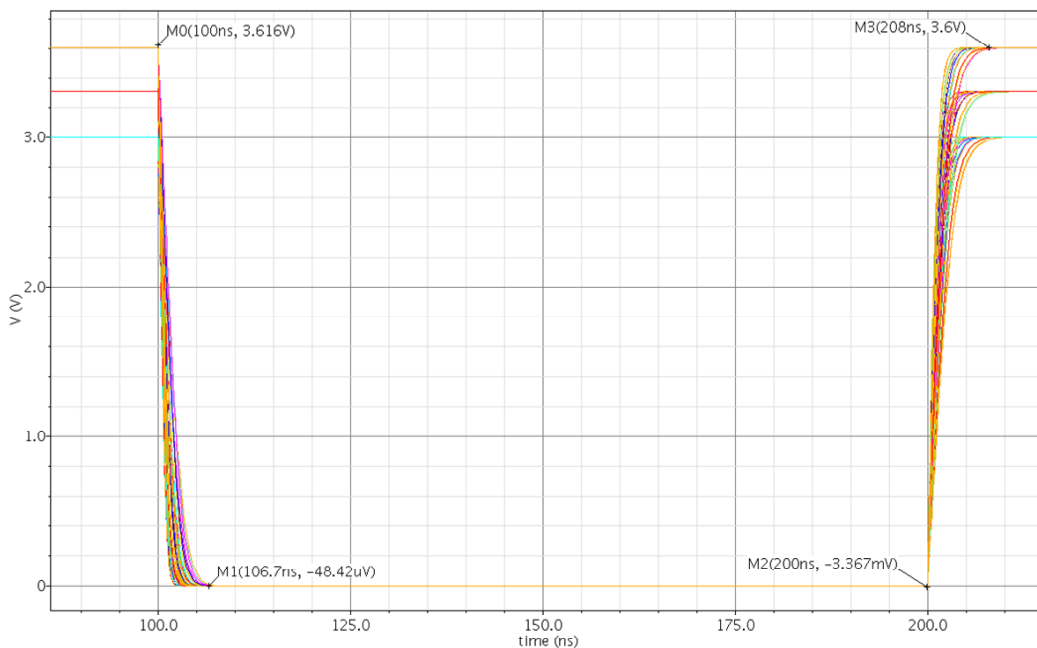
Figure 5-7: Rise and fall times over all corners

Tuning appears to be unnecessary, since the minimum W/L of 0.7/0.35 for the NMOS and a W/L of 1.5/0.35 for the PMOS is large enough to drive the load within about 8 ns.

The first stage of the comparator needs to drive the output stage. It is composed of the simplest circuit that can convert a differential input voltage into a single-ended, rail-to-rail output voltage. The complete circuit is depicted below.
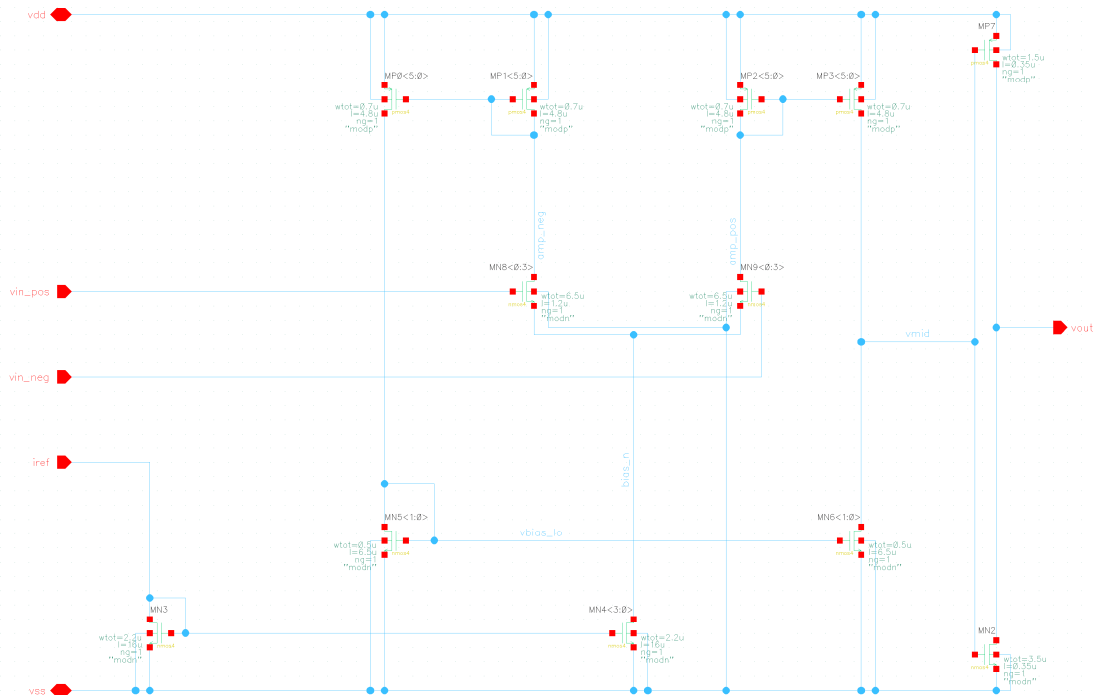
Figure 5-8: Complete circuit of comparator

The first stage consists of a differential pair that converts the input voltage into a current, plus a number of current mirrors that convert it to a single-ended signal. The output of the first stage must be able to supply enough current to charge the input capacitance of the second stage within the required time.

The W/L of the differential pair is optimized to provide as much gain as is necessary to achieve the required 1 mV resolution. The W/L of the current mirrors is optimized for current matching. The circuit is simulated over all corners to ensure correct biasing in all cases.

The transistors can now be sized to appropriate areas to ensure correct matching. The offset caused by the current mirrors is transformed to the input by dividing their current offset by the g_m of the differential pair. The calculations are performed in the same way as in section 6.1.3. The results are shown in Table 5-2.

	W/L (µm/µm)
Differential pair	26/1.2
PMOS current mirrors	3.5/4.8
NMOS current mirror	1.0/6.5

Table 5-2: W/L ratios for the comparator

The mirror that provides the biasing tail current is matched to 5 %.

Apart from the offset caused by statistic matching properties, the comparator also has a static offset caused by the asymmetry between the NMOS and PMOS transistors. When the two inputs are at the same voltage, the output of the first stage is not exactly halfway between the supply rails. Even if it were, the output stage still has its threshold voltage at some other value, causing offset again. It is therefore necessary to match these two thresholds and "straighten out" the comparator.

Tuning the output voltage of the first stage is not a good idea, since this would likely cause unequal loads on the differential pair. By tuning the output stage however, it is possible to match the two voltages without additional penalties. Increasing the W of one transistor actually increases the drive capability of the comparator. Care must be taken to ensure that the output stage can still be driven by the first stage.

The simulation results below show that the comparator achieves the correct propagation delay and resolution.
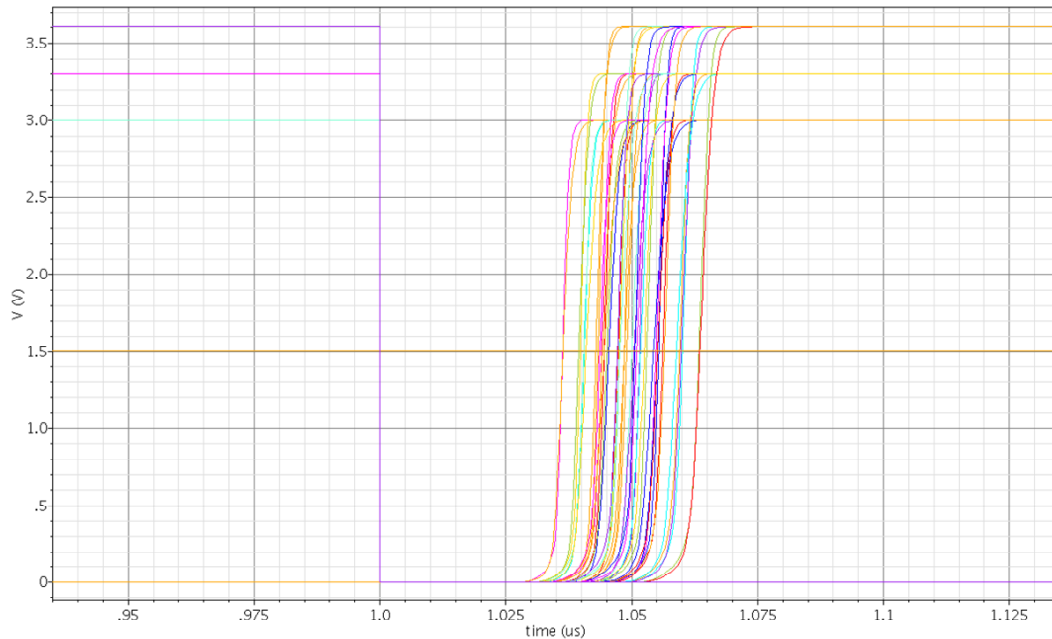
Figure 5-9: Rise time of the comparator
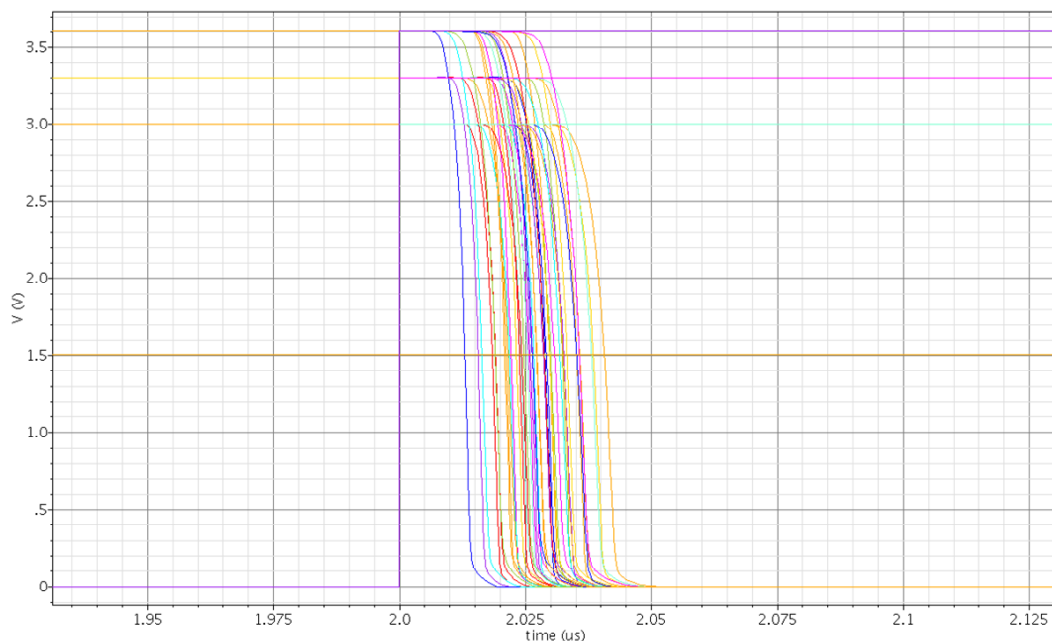
Figure 5-10: Fall time of the comparator

Figure 5-9 and Figure 5-10 show how the comparator reacts to a quickly-changing input signal, over all corner cases. The three different voltages at the high end of the curves are due to the minimum, typical and maximum supply voltage levels.
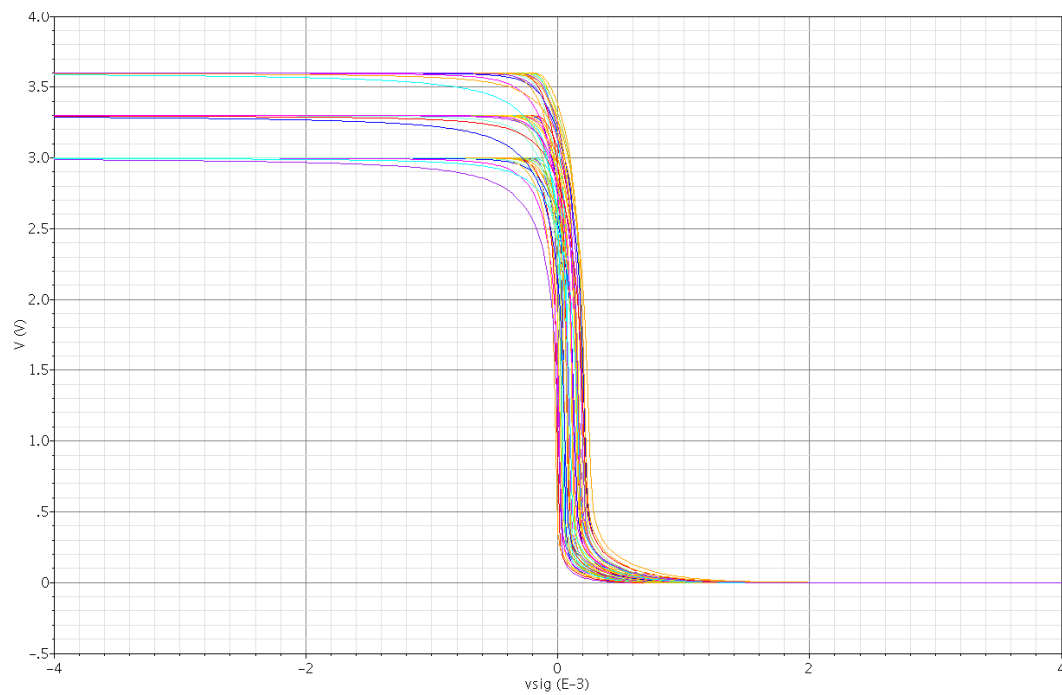
Figure 5-11: DC simulation showing resolution of the comparator

The resolution is shown to fall easily within 1 mV. The variation in the resolution due to process variations is smaller than the variation in the propagation delay shown above.

6. Design of support circuits

Three more circuits are required to achieve a complete IC design. These are the reference sources (for voltages and currents) and the test controller that will be used to validate the IC after fabrication.

The biasing circuits will provide a number of identical copies of a relatively stable reference current. Because the support circuits have to be low power, a standard reference current of 1 μA will be supplied. If a circuit requires more than this, then it will have to multiply the reference current through e.g. a current mirror.

The test controller has connections to useful nodes in all other sub-circuits, to facilitate testing and debugging after fabrication.

6.1 Bandgap reference

The purpose of the bandgap reference is to provide a stable reference voltage that can be used to fix voltages on the entire chip to a definite value. The reference voltage should vary as little as possible with temperature, supply voltage and process variations.

The required accuracy of the reference voltage is determined by the required accuracy of the reference voltages used elsewhere on the IC. It was determined in section 4.6 that the handover points for the dynamic transistor sizing should be accurate to within +/- 50 mV. The bandgap reference will be designed in such a way that the handover voltages will be accurate to within this specification.

The main idea behind a bandgap reference [5] is that it is possible to achieve a voltage that is constant with respect to temperature by adding one voltage that is proportional to temperature to another voltage that is negatively proportional to temperature. If these voltages are related to a physical constant, then also the absolute value of the voltage is accurately determined.

The base-emitter voltage of a bipolar transistor operating at a collector current I_c is negatively proportional to absolute temperature. The difference between the base-emitter voltages of two differently sized bipolar transistors is proportional to absolute temperature. If one of these two is multiplied by a factor A and added to the other, then the temperature dependence is cancelled. The exact value of the resulting voltage is (theoretically) equal to the bandgap voltage of silicon at 0 K. This principle is illustrated in Figure 6-1. It should be noted that this image is a simplification, and that the straight lines in reality are curved.

Manufacturing variations in the bipolar transistors will cause the exact slope of the temperature curves to differ. The starting point of the curves is determined by a physical constant, and is therefore a reliable reference to derive accurate voltages from.
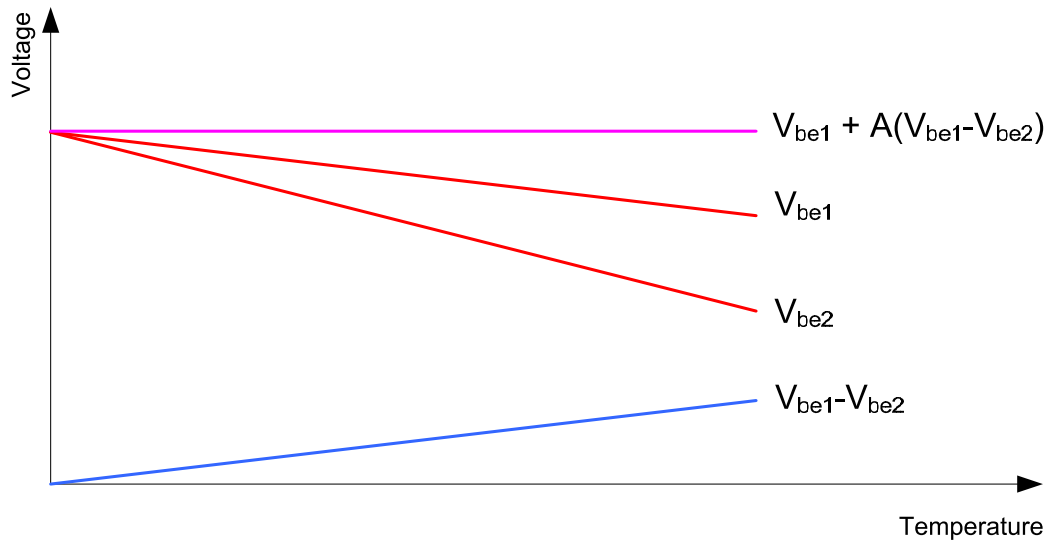
Figure 6-1: Principle of a bandgap reference

6.1.1 Temperature behaviour

Two types of bipolar transistors are available in the H35 process. These include 2 μm × 2 μm lateral PNP transistors and 10 μm × 10 μm vertical PNP transistors. Vertical transistors have smaller variations in their parameters and are therefore the better choice, at the expense of some chip area.

The first step is to determine the currents and voltages that are needed to correctly provide temperature compensation. A ratio of 8/1 is chosen because this will enable a compact IC layout consisting of a 3 × 3 grid with the single transistor in the middle. The basic schematic that implements the bandgap function is shown below.

The current through both branches of the circuit is equal because of the current mirror-like structure at the top. These two transistors are driven by the amplifier, which will output a voltage such that its two inputs achieve the same voltage. This voltage will be equal to V_{be1} (of the left bipolar transistor). The voltage across R_0 will therefore be equal to $V_{be1} - V_{be2}$. Finally, the ratio between R0 and R1 implements the scaling factor A from Figure 6-1.
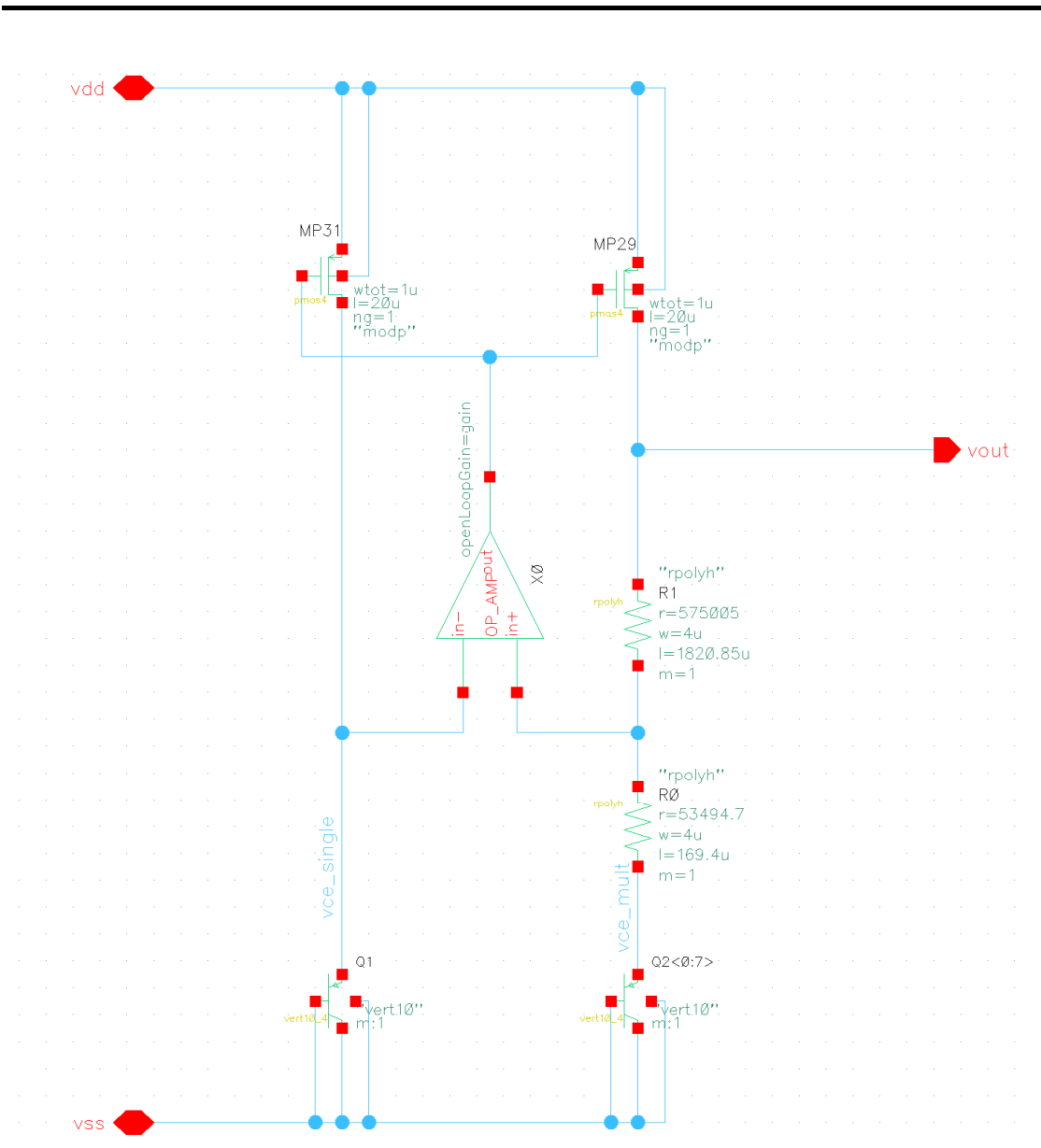
Figure 6-2: Circuit for determining the temperature behaviour of the bandgap reference

The two bipolar transistor sets are biased at 1 μA. The temperature behaviour is optimized by adjusting the resistors until the peak of the nominal output voltage versus temperature curve is located in the centre of the required temperature range. The result is shown below for all process corners.
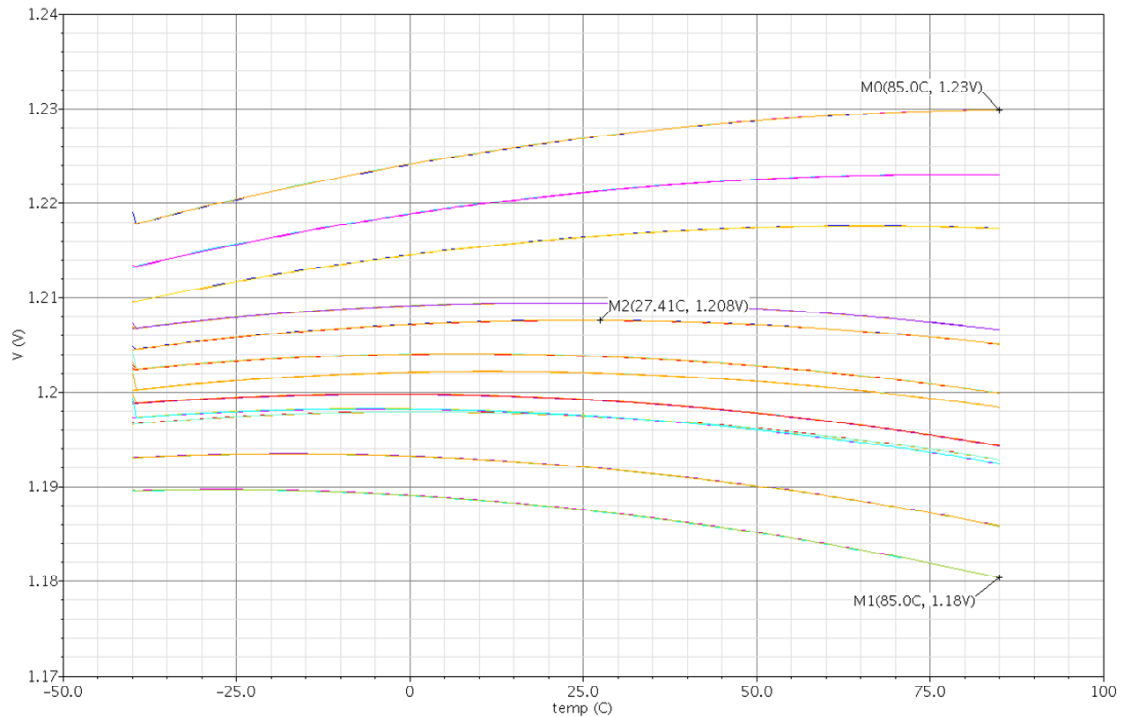
Figure 6-3: Output voltage as a function of temperature

Ideally, these curves should be straight lines, as shown in Figure 6-1. The fact that they are curved is due to higher-order temperature dependencies.

The nominal behaviour is shown by the curve indicated by marker M2. The other curves are determined by the manufacturing variations in the parameters of the bipolar transistors. They lead to a total variation in the output voltage of +/- 25 mV. Since these are corner cases, the chip has to work correctly within this entire range, and the only way to reduce this variation is post-manufacture trimming, which is not available. This inaccuracy is therefore unavoidable, and has to be accounted for in further calculations.

6.1.2 Amplifier design

The next step is to replace the ideal amplifier with a real circuit. An amplifier is needed that can work with an input voltage of about 600 mV and that is self-biased. The circuit is shown below.
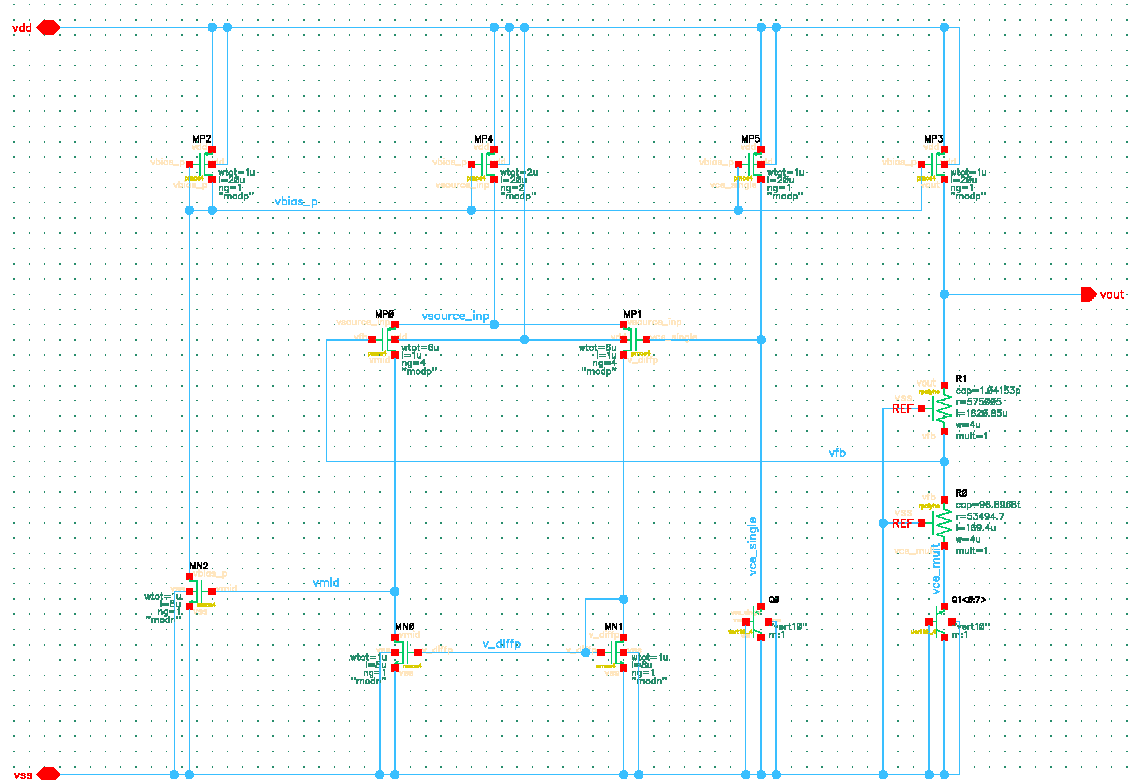
Figure 6-4: Bandgap reference with real amplifier

The input stage consists of a PMOS differential pair (MP0 and MP1), so that it can reliably amplify signals close to the V_{ss} rail. The second stage (MN2) provides additional gain and drives the output current mirror (MP2 – MP5). An extra copy of the output current is made (MP4) to provide the bias current for the amplifier, so that no external biasing is required.

The W/L of each transistor is tuned to achieve a correct operating point (transistors in saturation) across all corner cases. Current mirrors are tuned to maximize their overdrive voltage, and the differential pair is tuned to minimize its overdrive voltage. Both of these optimizations increase their respective matching parameters. The resulting W/L are shown below.

	W/L
Input differential pair	6/1
Input current mirror	1/8
Output current mirror	1/20

Table 6-1: W/L of each transistor in the bandgap reference

6.1.3 Matching

Now that all the W/L have been established, it is time to scale all transistor gate areas in order to obtain correct matching properties. The total allowed output voltage offset is $\sigma_{Voff,out} = 2$ mV, as determined in section 4.6.

If minimum sized resistors are used (4μm × 170μm), then the resistors cause a mismatch resistance of 0.25 %, or an offset resistance of 133 Ω. This leads to an offset voltage of 138 μV. This is far less than the other sources of mismatch, so there is no need to make the resistors any larger than they are.

The bipolar transistors cannot be changed in any way, so their mismatch is fixed at 0.04% of I_s. This means a voltage offset of $\sigma_{Voff,bipolar} = V_T \ln(A_{I_s}) = 10.6\ \mu V$.

The contribution of each source of mismatch is listed in Table 6-2.

	σ	σ^2
Input differential pair	$\dfrac{A_{V_{gt}}}{\sqrt{W_1 L_1}}$	$\dfrac{210}{W_1 L_1}\ \mu V^2$
Input current mirror	$\dfrac{I_{bias} A_{I_d}}{g_{m,1}\sqrt{W_2 L_2}}$	$\dfrac{51}{W_2 L_2}\ \mu V^2$
Output current mirror	$\dfrac{I_{bias} A_{I_d} r_{out}}{gain_{total}\sqrt{W_3 L_3}}$	$\dfrac{70}{W_3 L_3}\ \mu V^2$
Resistors	138 μV	19 nV^2
Bipolar transistors	10.6 μV	0.11 nV^2
Total allowed	2 mV	4 μV^2

Table 6-2: Sources of mismatch in the bandgap reference

The sum of the variances of all the mismatch sources added together should be no larger than the variance of the total allowed mismatch. Since the total sum contains three variables, there is no single solution. However, a numerical optimization becomes possible by adding the condition that the total transistor gate area should be as small as possible. Using a numerical optimization program (Appendix C), the following values are obtained.

	W·L	**W/L**
Differential pair	109 μm^2	26/4.3
Input current mirror	53.8 μm^2	2.6/21
Output current mirror	63.0 μm^2	1.8/35.5

Table 6-3: W and L of each transistor after matching optimization

6.1.4 Frequency behaviour

The next step is to make sure that the circuit is stable enough, i.e. that is does not oscillate or ring. This requires calculating the poles and zeroes of the circuit and, if necessary, perform a frequency compensation to move the poles so that the circuit will have a Butterworth transfer. First, the loop is cut open to allow an open-loop AC analysis to be performed. This circuit is shown in Figure 6-5.
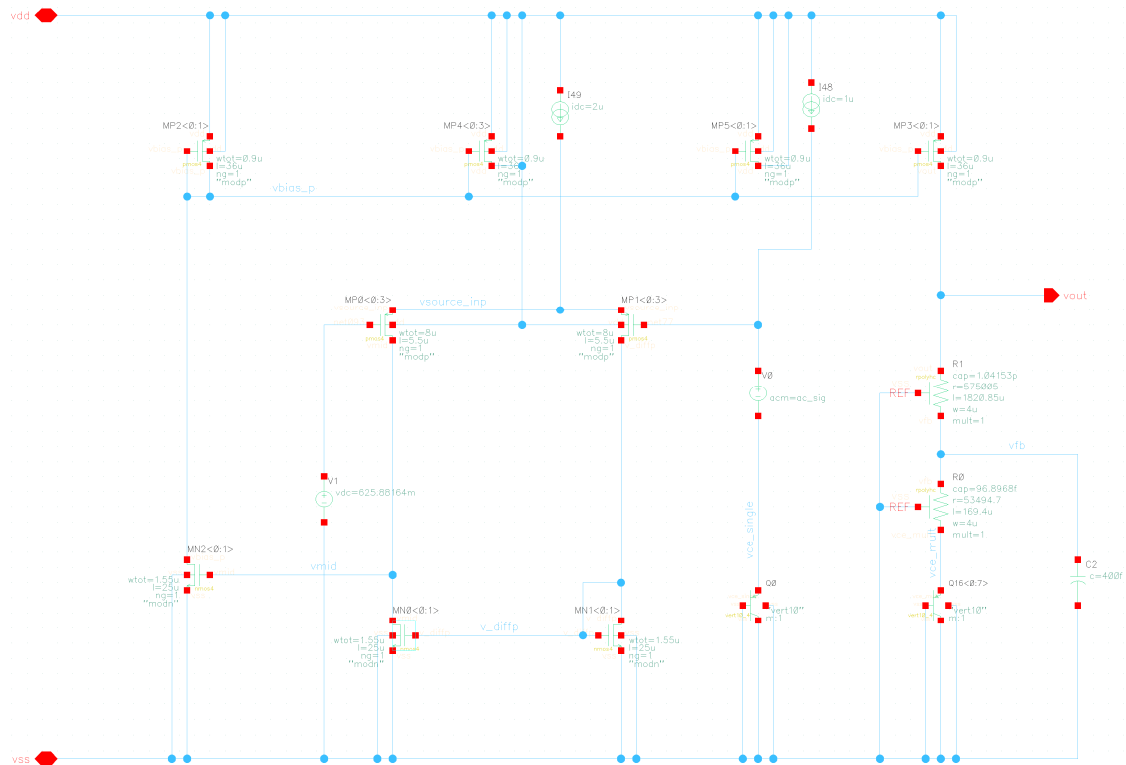
Figure 6-5: Open-loop circuit

The loop is broken at the feedback point, and the DC voltage present at the feedback input is added as a DC voltage source to maintain correct biasing.

The bode plot resulting from this AC analysis is shown in Figure 6-6.
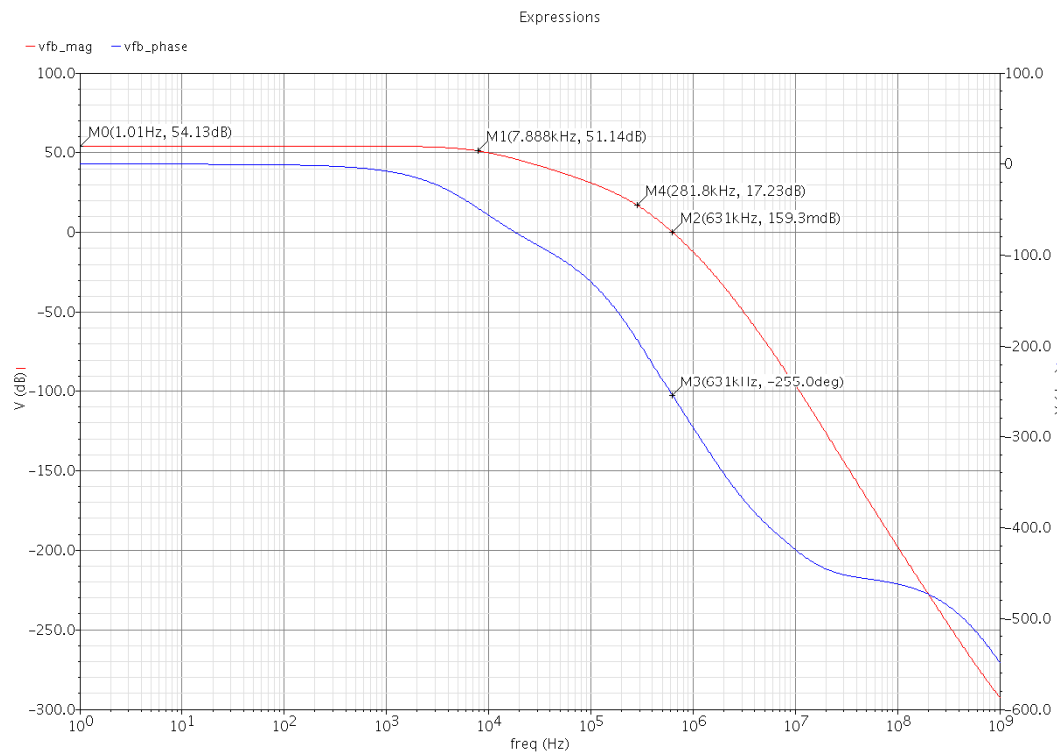
Figure 6-6: Open-loop bode plot

This plot shows that the circuit oscillates, since the phase shift at the unity gain point (given by M2 and M3) is more than 180 degrees. There are two dominant poles, one at about 8 kHz and one at about 280 kHz.

The first pole is caused by C_{gs} of the second stage (MN2), with the g_{ds} of the first stage and its current mirror. The second pole is caused by the C_{gs} of the entire output current mirror and its g_m.

The amplifier is stabilized according to the frequency compensation method outlined in [3].

Adding a phantom zero at the input is not effective, since the source impedance is quite low already ($1/g_m$ of the bipolar transistor). Adding one at the output is not very effective either, since the load is only a small capacitance and an effective phantom zero would therefore require a very large resistor (or even an inductor). A phantom zero in the feedback network is not effective either, since the feedback network only causes a small reduction in loop gain.

The next most favourable method is pole-splitting. This involves adding a capacitor C_{split} between the gate and drain of MN2, to push the two poles away from each other. Figure 6-7 shows how the closed-loop poles move for various values of $C_{split.}$
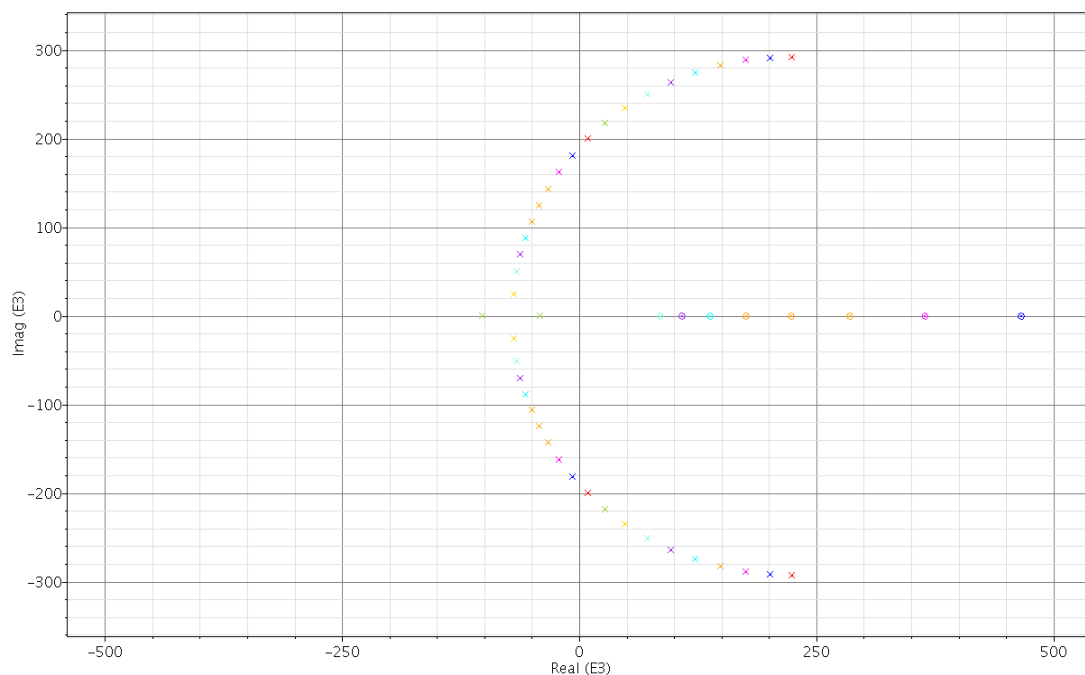
Figure 6-7: Root locus for several values of C$_{split}$

Figure 6-8 shows a close-up of the region around a Butterworth transfer. It shows that the optimal C_{split} to obtain this is equal to 5 pF. However, taking into account all variations in manufacturing, temperature and supply variations, a value of 10 pF is a safer choice to guarantee a stable, non-ringing amplifier. Figure 6-9 shows the location of the poles for all variations with a C_{split} of 10 pF.
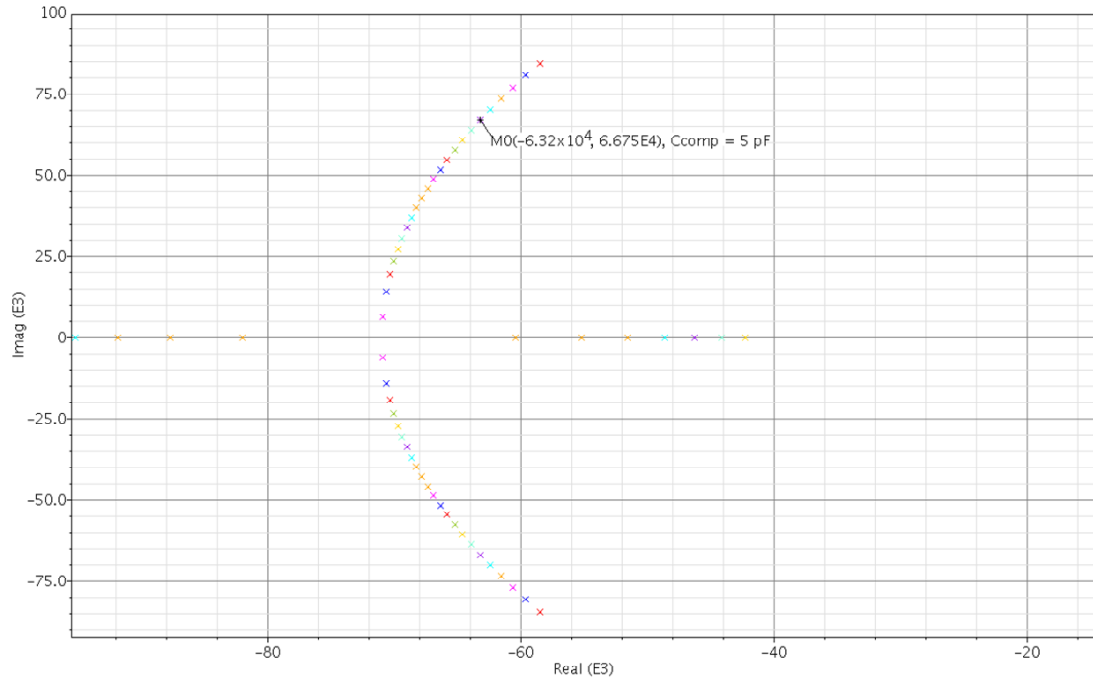
Figure 6-8: Close-up of region around Butterworth transfer
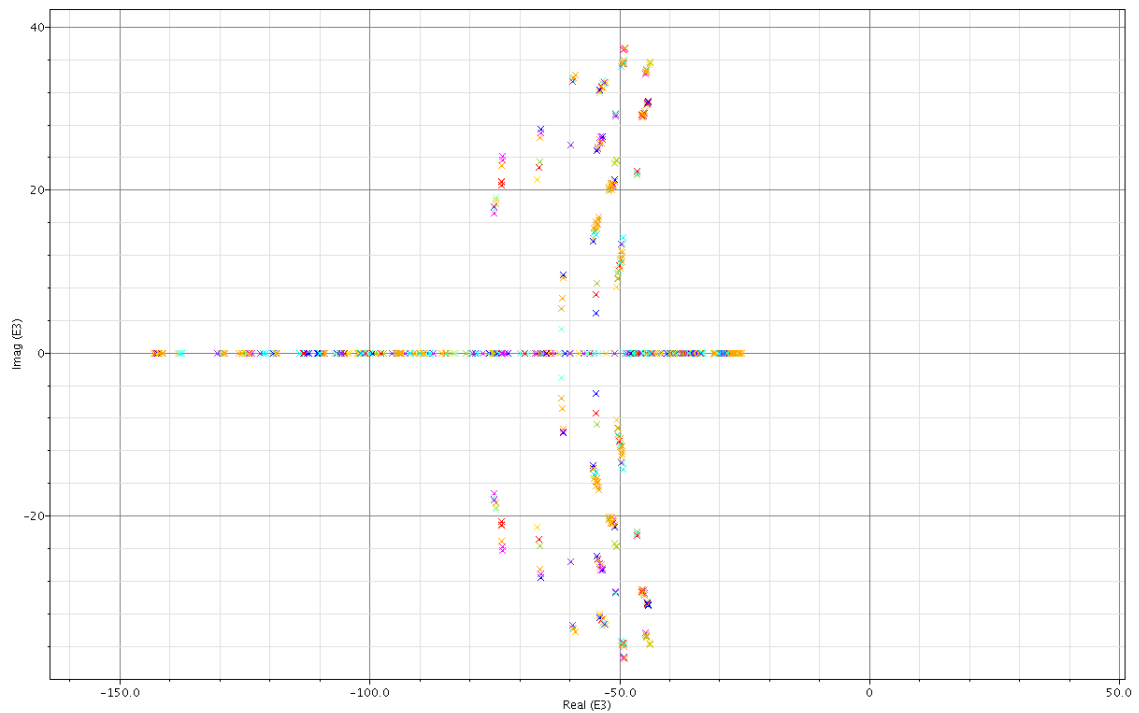
Figure 6-9: Pole-zezro plot for all corner cases, $C_{split} = 10$ pF

6.1.5 Start-up circuit

The bandgap reference is somewhat notorious among analogue circuit designers because of its tendency to become stuck in an undesired "zero" operating condition: a stable condition in which the output voltage is zero. The following analysis shows what happens in this case.

Looking at Figure 6-2, the current in both branches is equal and will be denoted by I_e. For the current through a bipolar diode it holds that

$$I_e = I_s \exp\left(\frac{V_{be}}{V_T} - 1\right)$$

So that the base-emitter voltage of Q_1 becomes

$$V_{be,1} = V_T \ln\left(\frac{I_e}{I_s} + 1\right)$$

For Q_2, with multiplicity M, this becomes

$$V_{be,2} = V_T \ln\left(\frac{I_e}{M \cdot I_s} + 1\right)$$

The amplifier sets the current I_e such that the voltage at the amplifier's inputs equals zero. This condition is fulfilled when

$$V_{be,1} = V_{be,2} + I_e \cdot R_2,$$

or,

$$V_T \ln\left(\frac{I_e}{I_s} + 1\right) = V_T \ln\left(\frac{I_e}{M \cdot I_s} + 1\right) + I_e \cdot R_2$$

Because $R_2 = \dfrac{V_T \ln(M)}{I_0}$, in which I_0 is the nominal operating current of the bandgap reference, this becomes

$$V_T \ln\left(\frac{I_e}{I_s} + 1\right) = V_T \ln\left(\frac{I_e}{M \cdot I_s} + 1\right) + I_e \cdot \frac{V_T \ln(M)}{I_0} \tag{1}$$

Plotting these voltages as a function of I_e, using numerical values, results in the following graph.
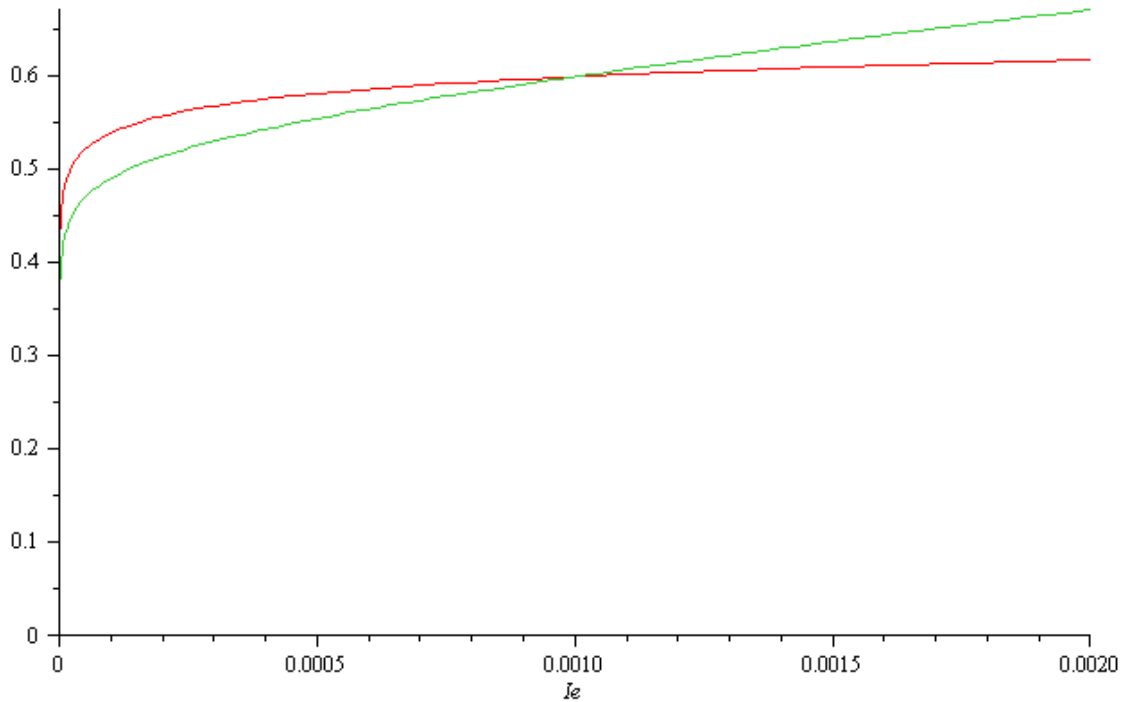
Figure 6-10: $V_{be,1}$ (red) and $V_{be,2}$ (green) as a function of I_e

Graphically, and analytically, it follows that there is a stable point when equation (1) is satisfied. If $I_e \gg I_s$, then this can be simplified to

$$\ln\left(\frac{I_e}{I_s}\right) = \ln\left(\frac{I_e}{M \cdot I_s}\right) + \frac{I_e}{I_0 \ln(M)}$$

which is satisfied for $I_e = I_0$ (in this case 1 µA).

However, there is also another solution. As expected, this is when $I_e = 0$. Theoretically this is not a stable point, since $V_{be,2} + I_e \cdot R_2$ is smaller than $V_{be,1}$ when $I_e < I_0$, and so any disturbance (such as noise) will tend to push the circuit out of the zero state and towards the desired operating point.

Practical experience shows that this often does not happen, however. Three main reasons for this can be seen: firstly, if the amplifier is biased by a copy of I_e, then the amplifier itself is effectively switched off in the zero state and will be unable to move the circuit to the normal operating point. Secondly, zero volts may be out of the range of the amplifier's input circuit, preventing correct operation again. Thirdly, a DC offset at the amplifier's input (as caused by mismatch) can cause $V_{be,1}$ to appear smaller than $V_{be,2} + I_e \cdot R_2$, causing the zero solution to become a stable point.

In order to ensure that the bandgap reference always ends up in the correct region, it is therefore necessary to completely remove the stable point at $I_e = 0$. Analytically, this can be done by adding a current I_{start} as a function of I_e to the left-hand side of equation (1):

$$V_T \ln\left(\frac{I_e + I_{start}}{I_s} + 1\right) = V_T \ln\left(\frac{I_e}{M \cdot I_s} + 1\right) + I_e \cdot \frac{V_T \ln(M)}{I_0}$$

If $I_{start}(0) > 0$ and $I_{start}(I_0) = 0$, then the zero solution is removed while the normal solution is unaltered. A simple function that satisfies these requirements is
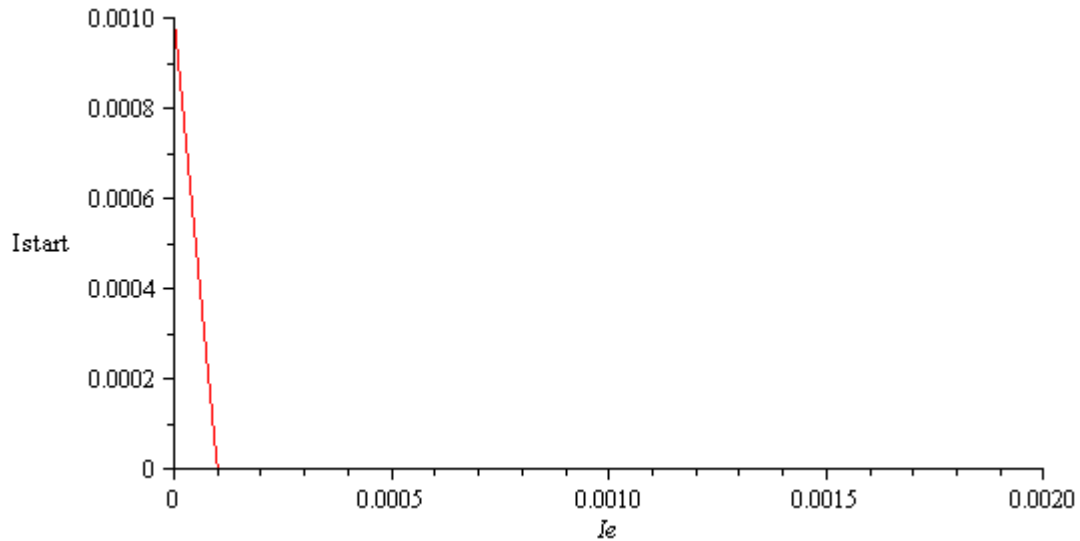
Figure 6-11: Example start-up current as a function of I_e

If this current is added to the current in the left branch (through Q_1), then Figure 6-10 becomes as follows.
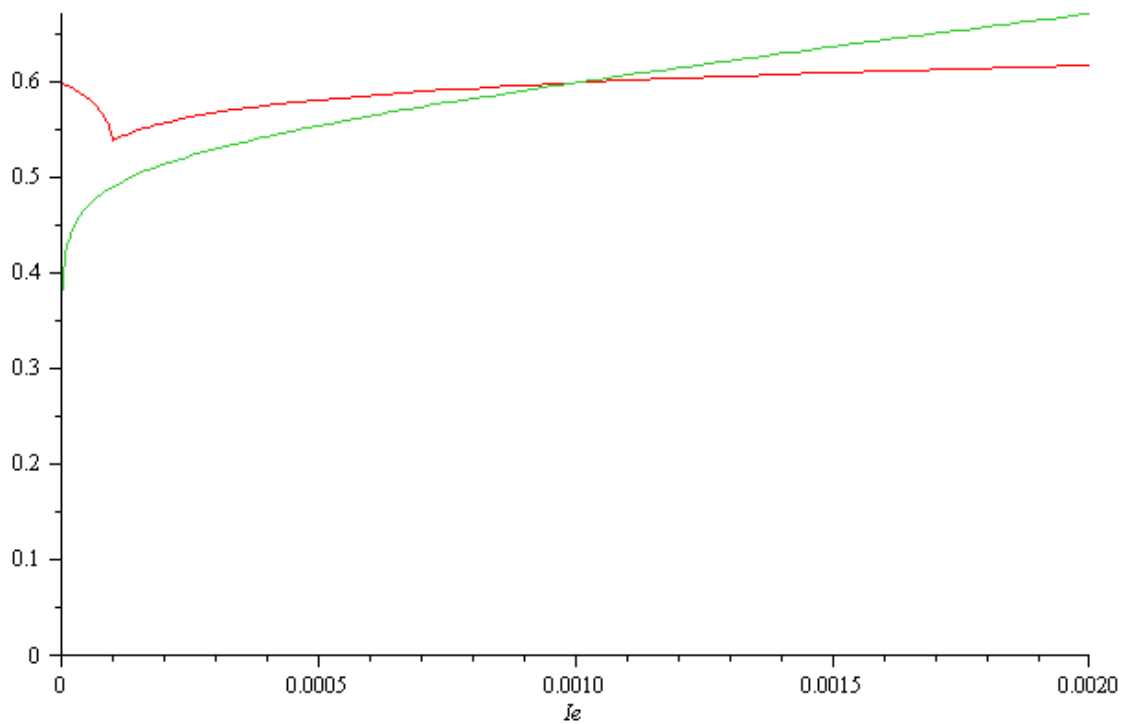
Figure 6-12: $V_{be,1} + I_{start}$ (red) and $V_{be,2}$ (green) as a function of I_e

Clearly, the zero solution has been eliminated, without affecting the nominal solution.

To make this work in practice, the amplifier also needs to be biased correctly for all values of I_e. Adding I_{start} to its bias current ensures that its bias current is never zero, and that the amplifier is always capable of moving the circuit to its correct operating point.
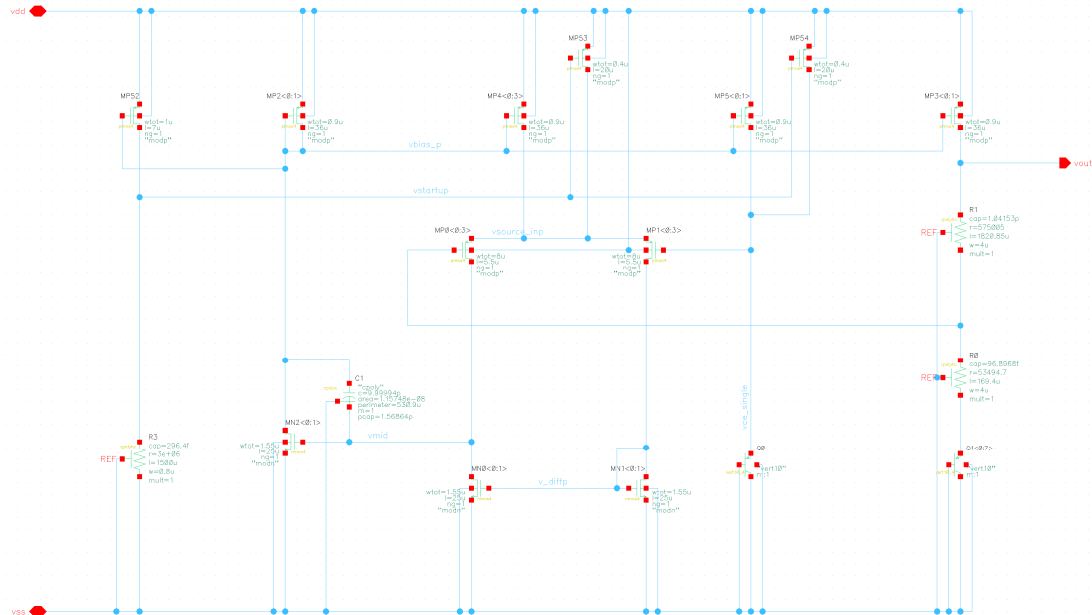
Figure 6-13: Bandgap reference with start-up circuit

The start-up circuit consists of an additional transistor (MP52), of the same W/L as the current mirror, that passes a copy of the mirrored current through a 3 MΩ resistor (R3). If the current (which is equal to the emitter current of the bipolar transistors) is below 1 μA, then the voltage $V_{startup}$ will be lower than 3 V, and turn on MP53 and MP54. When this happens, MP53 will provide a bias current to the amplifier, while MP54 injects a current I_{start} into Q0.

Simulating the circuit in Spectre leads to the following results, which are very similar to the mathematical derivation above.
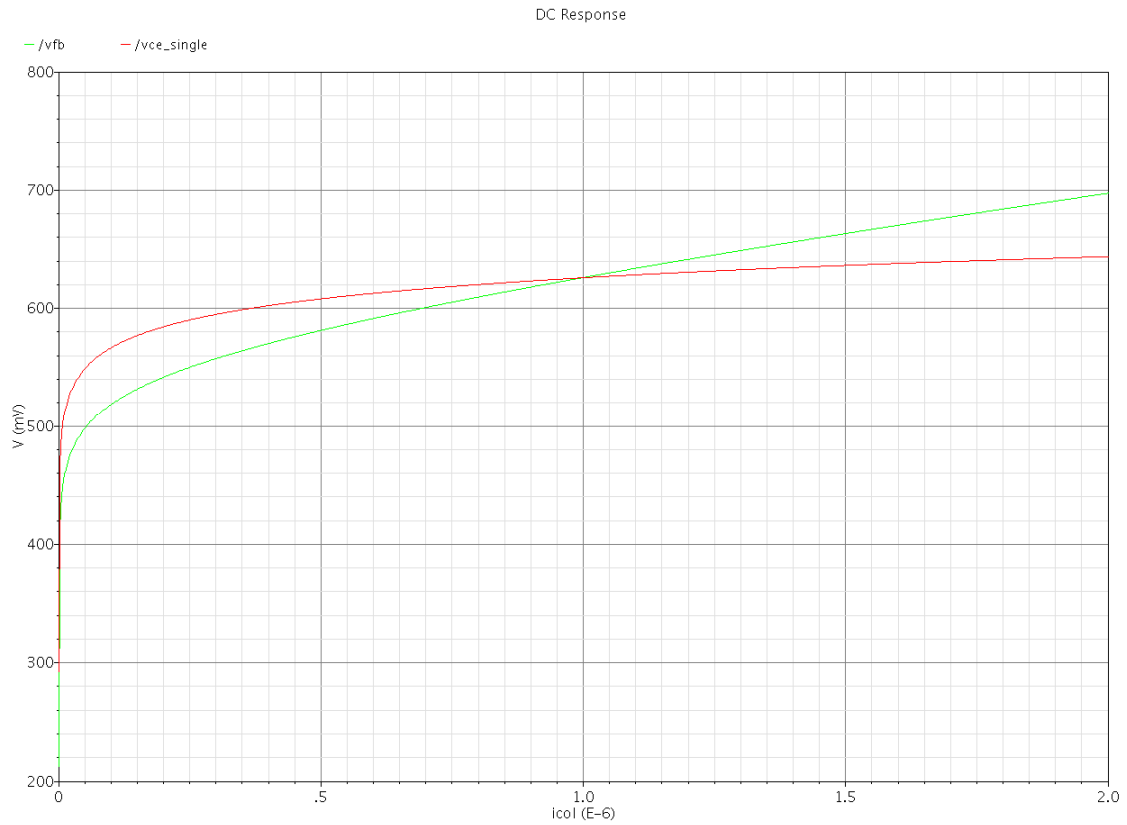
Figure 6-14: Simulated $V_{be,1}$ (red) and $V_{be,2}$ (green) as a function of I_e
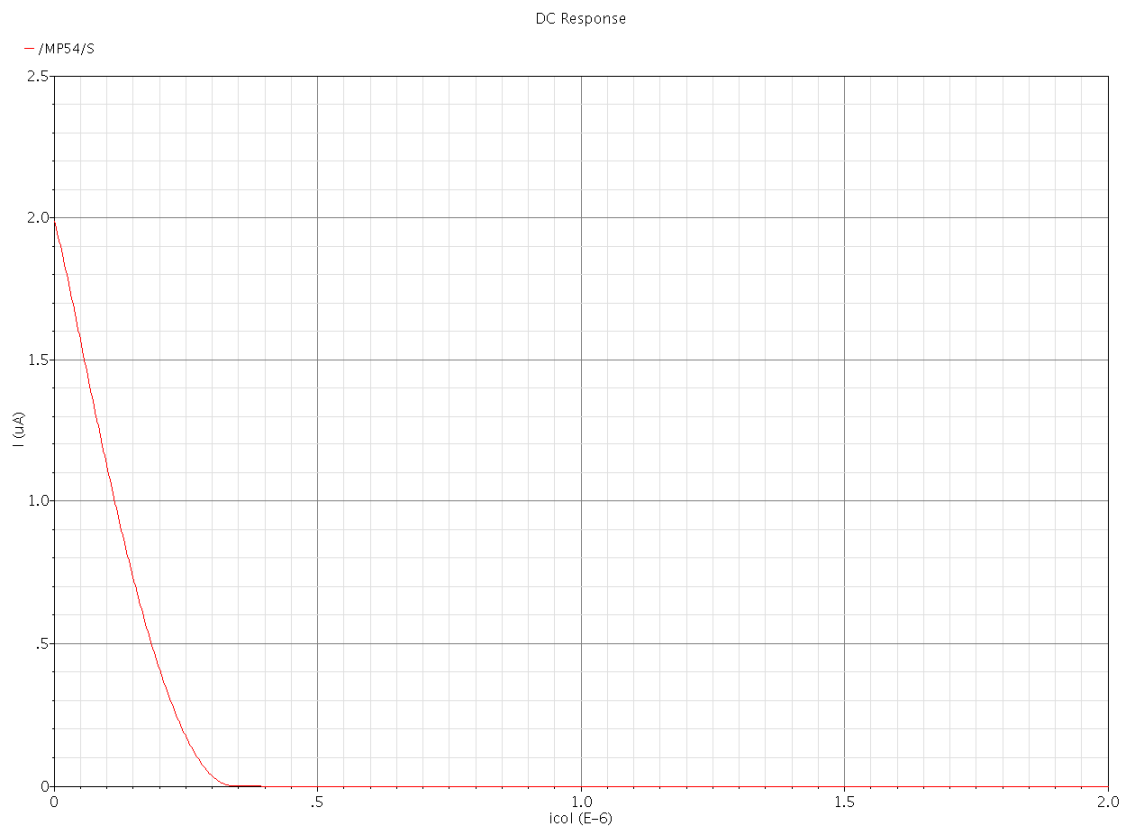
Figure 6-15: Simulated I_{start} as a function of I_e
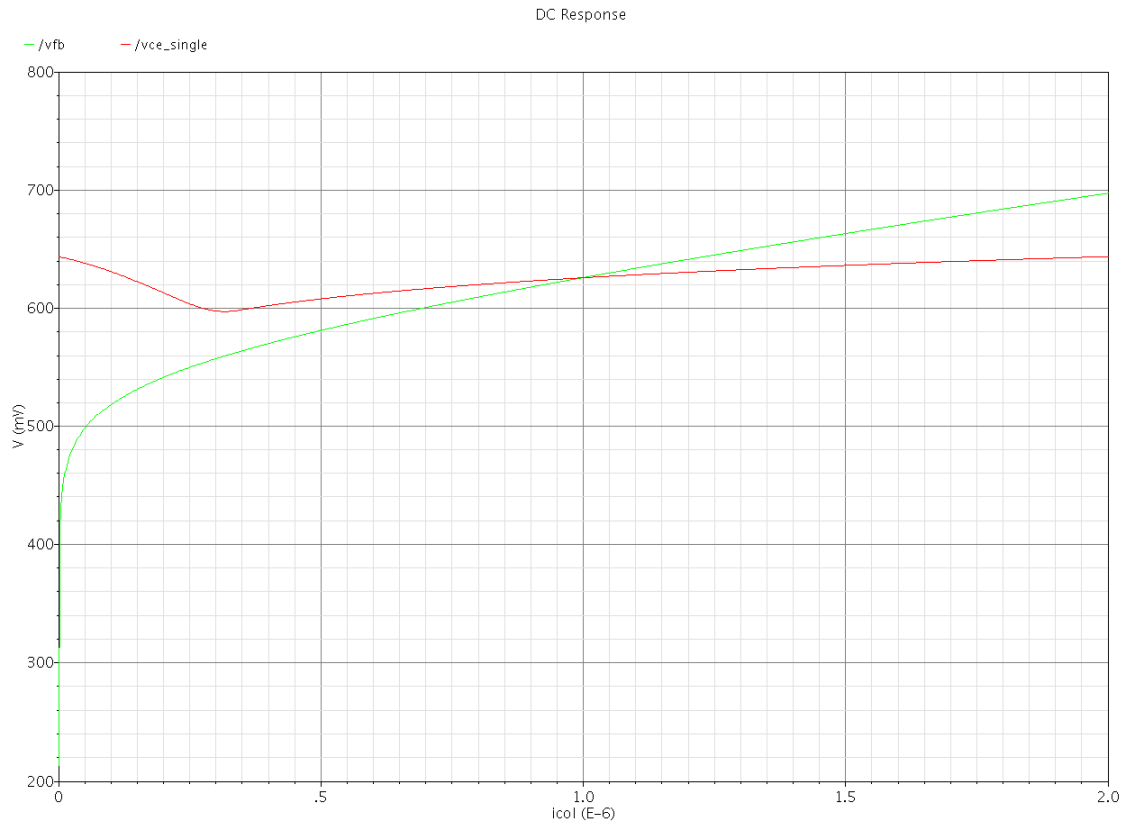
Figure 6-16: Simulated $V_{be,1} + I_{start}$ (red) and $V_{be,2}$ (green) as a function of I_e

6.1.6 Circuit finishing

A transient simulation is now run to test whether all calculations were correct. To test the start-up behaviour, the response of the output voltage to a start-up pulse is plotted over all corners.
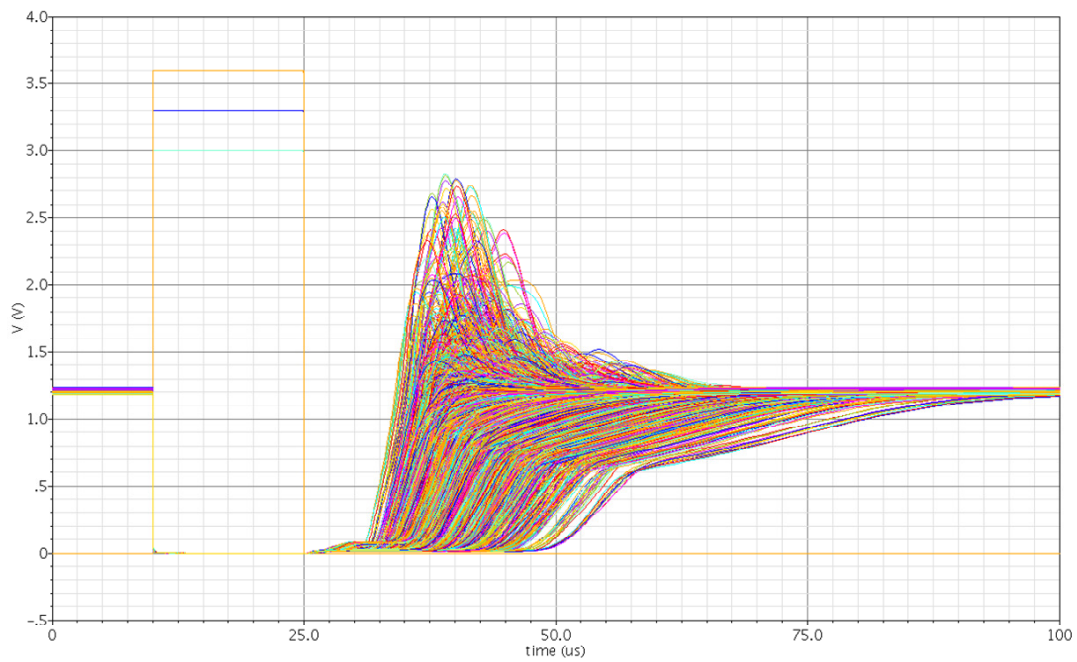
Figure 6-17: Start-up behaviour of the bandgap reference

The simulation shows that the circuit starts up correctly in all cases, although it does sometimes overshoot. This lasts only a few microseconds, and should not pose a serious problem to the rest of the circuits because the current reference (described below) also takes some time to start up.

6.2 Current reference

The current reference circuit should provide a number of identical copies of a stable reference current. It derives this stability from a stable voltage produced by the bandgap reference. The currents do not need to be very accurate, as long as they are accurately matched.

The basic function of the current reference is a voltage-to-current amplifier. The exact value of the current is dependent on the exact value of the resistor that sets the gain. Because on-chip resistors have a spread of +/- 20%, the reference current will vary by this much from wafer to wafer.

Voltages can be made accurately by passing a current from the current reference through a resistor that is matched with the resistor inside the current reference. Ideally, the voltage produced will then be an accurate copy of the bandgap voltage.

In the figure below, two currents are used for biasing other circuits (not shown), while a third is used for generating a voltage V_{out}. If there is an accurate ratio between I_{ref} and I_{out}, and R_{load} is matched with R_{fb} to produce an accurate resistance ratio, then V_{out} will be an accurately scaled copy of V_{ref}.
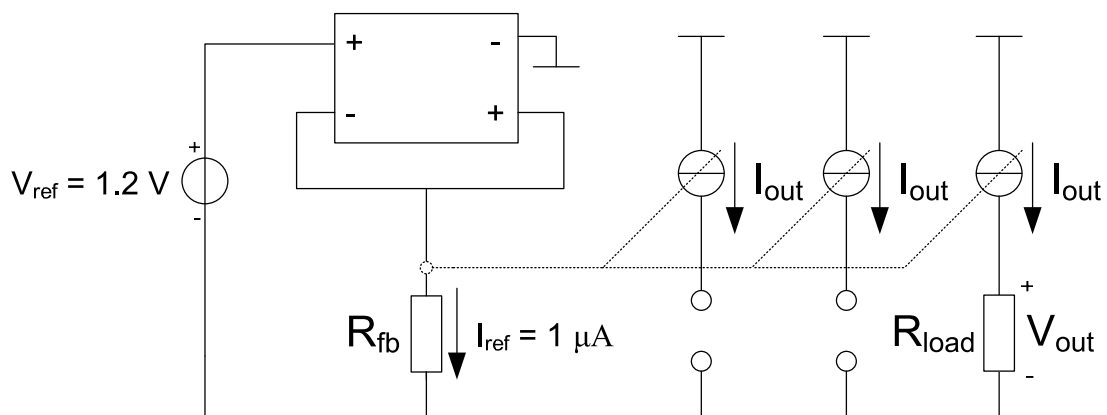
Figure 6-18: Basic current reference circuit

The following specifications are required for the current reference.

Nominal input voltage	1.2 V
Output current	1 µA +/- 30 %
Number of output currents	6
Total mismatch of output current	σ = 2 nA
PSRR	> 25 dB up to 1 MHz

Table 6-4: Requirements for the current reference

The mismatches of the output currents are measured relative to the current flowing through the reference resistor. The requirement for the PSRR is derived from the fact that the power supply for the IC will likely be a switching power regulator operating at a frequency of several hundred kHz.

According to Figure 6-18, the resistor should have a value of V_{ref}/I_{out} = 1.2 MΩ. It is possible to change this to a different value, and insert a scaling factor in the current mirror at the output. However, this will only complicate the matching of these transistors, so the resistor is chosen to be 1.2 MΩ.

6.2.1 Circuit topology

A straightforward transistor implementation of the current reference is shown in Figure 6-19. The input consists of an NMOS differential pair, feeding a second stage that drives a current through resistor R1. The second stage is used as the input of a current mirror that provides the output currents.

First, the W/Ls of all transistors are sized to obtain a convenient DC operating point. The gate overdrive voltage of the output transistors is chosen to be 0.9 V, because a high overdrive voltage improves current matching.

The input current mirror (MP7 and MP8) is sized identically to the second stage and biased at the same current. This ensures that the drain voltages of the differential pair are as closely matched as possible, to eliminate systematic offset due to channel-length modulation.

The input differential pair is sized with as large a W/L as possible, to provide maximum gain. The W/L is limited by the requirement that the transistors stay in saturation.
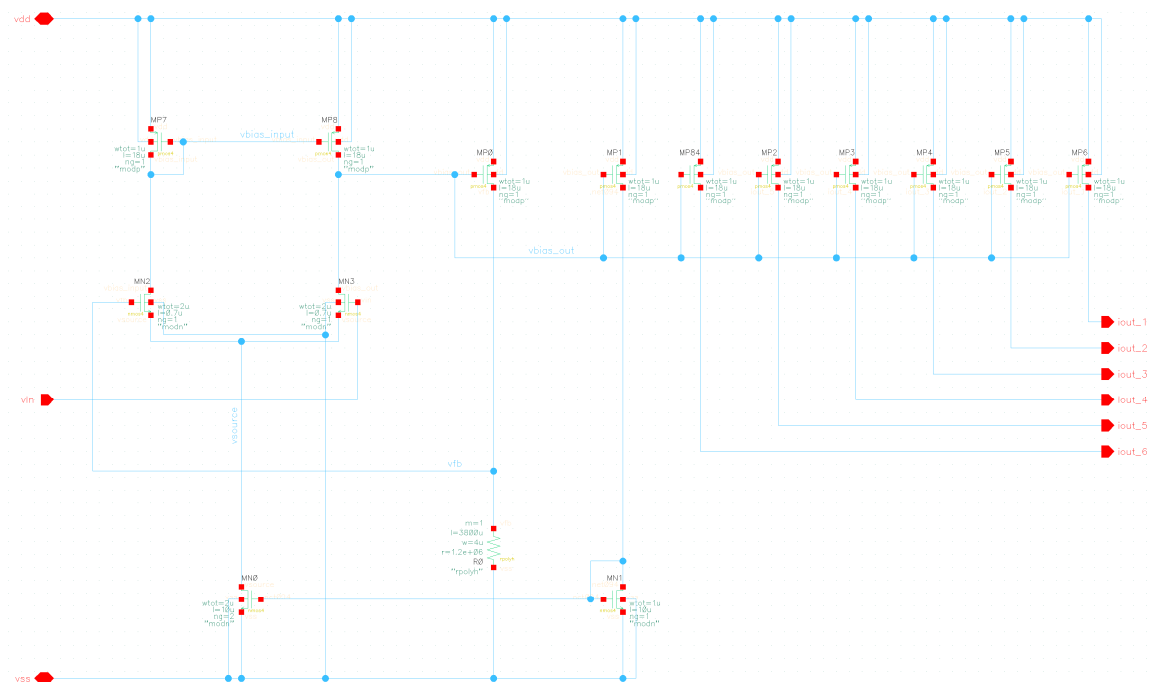
Figure 6-19: Basic implementation of the current reference

The input stage is biased by a copy of the output current. This ensures that the current through MP7 and MP8 is equal to the current through the second stage (as stated above, to improve matching). The disadvantage is that the circuit is now not necessarily self-starting anymore and needs a start-up circuit. This will be dealt with later. The biasing mirror is sized such that the gate-source overdrive voltage is high enough to provide for reasonable matching, but low enough to keep the output transistor in saturation.

The exact W/L ratios are tuned until the circuit is correctly biased across all process variations.

6.2.2 Calculating mismatch contributions

The allowed mismatch of the output current is given by $\sigma(I_{out})$ = 2 nA. This must be divided over all matched components, as described in Appendix B.

The mismatch contributed by the input differential pair is located at the input of the amplifier. Its contribution equals $\dfrac{A_{V_{gs}}}{\sqrt{W_1 L_1}}$ in which A_{Vgs} = 9.5 mVμm.

The mismatch from the input current mirror is scaled by $g_{m,1}$ =13.6 μA/V (worst-case). It contribution therefore becomes $\dfrac{I_{bias} A_{I_d}}{g_{m,1}\sqrt{W_2 L_2}}$. The overdrive voltage of this stage is 617 mV (worst-case), which means that A_{Id} becomes 4.8 %μm.

The mismatch from the output current mirror is directly in parallel with the output current. So, $V_{off} = \dfrac{A_{I_d} I_{out} R_{ref}}{\sqrt{W_3 L_3}}$.

An allowed output current mismatch of 2 nA becomes an input offset voltage of 2 nA · 1.2 MΩ = 2.4 mV. This has to be distributed over all mismatch sources.

	σ	σ²
Input differential pair	$\dfrac{A_{V_{gs}}}{\sqrt{W_1 L_1}}$	$\dfrac{90}{W_1 L}\,\mu V^2$
Input current mirror	$\dfrac{I_{bias} A_{I_d}}{g_{m,1}\sqrt{W_2 L_2}}$	$\dfrac{12.5}{W_2 L_2}\,\mu V^2$
Output current mirror	$\dfrac{A_{I_d} I_{out} R_{ref}}{\sqrt{W_3 L_3}}$	$\dfrac{3.3}{W_3 L_3}\,mV^2$
Total allowed	2.4 mV	5.76 μV²

Table 6-5: Mismatch contributions

In order to save chip area and to maximize bandwidth, the total gate area needs to be minimized. The optimal areas are calculated using numerical optimization. The transistors can then be scaled, keeping W/L constant, to the areas that have been calculated.

	Area	W/L
Input differential pair	117 μm^2	18 / 6.4
Input current mirror	43 μm^2	1.8 / 28
Output current mirror	706 μm^2	6.4 / 125

Table 6-6: Calculated gate areas

It should be noted that the output mirror consists of very large transistors, which might pose a problem for the next step. If the gate capacitance needs to be reduced, then it is possible to scale down the output transistors that are only used to provide the bias current for other circuits, because this does not require accurate matching. Only the output current that will be used to generate reference voltages will then be accurately matched.

The mirror used to bias the input stage (MN0 and MN1) is scaled to produce a three-sigma drain current mismatch of 5 %. This leads to a gate area of 19.3 μm^2.

6.2.3 Frequency behaviour

The circuit is now ready for an AC analysis. The first step is to determine the loop gain and phase as a function of frequency. This can be done using an AC analysis.

The loop gain can be determined by opening the loop. A convenient point to do so is at the output of the feedback network. The DC voltage that is present during normal operation is applied to the input of the differential pair, and its input capacitance is added to the feedback network.

The input of the amplifier is driven by a source with the same impedance as the output impedance of the bandgap reference (660 kΩ).
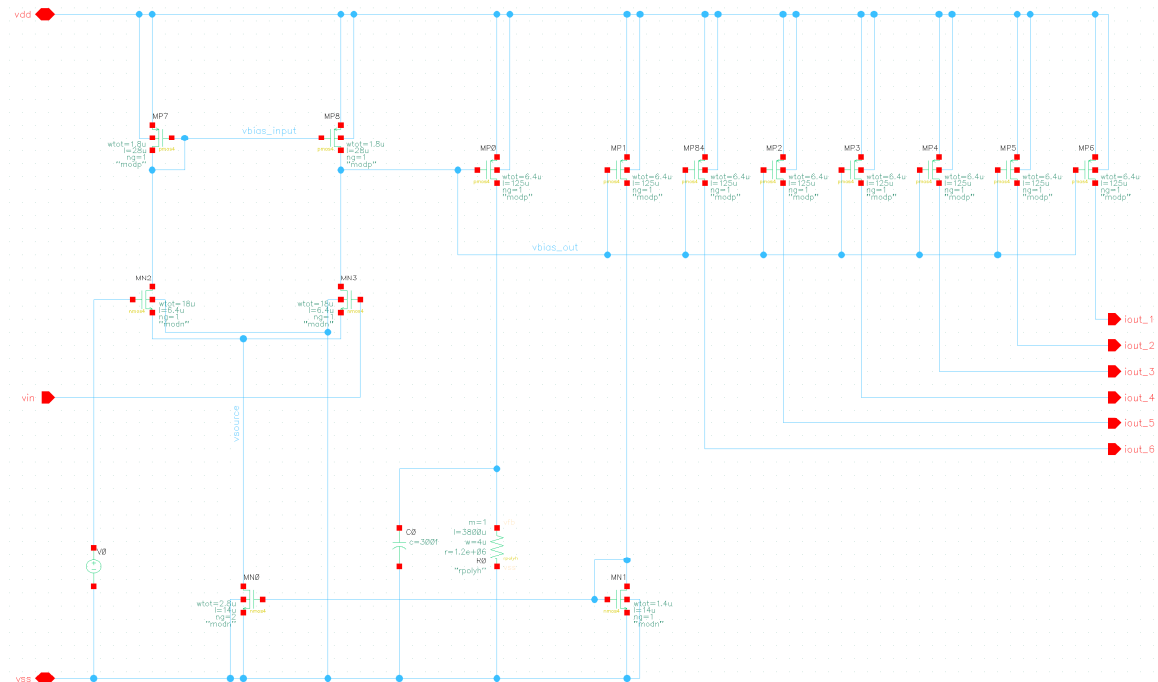
Figure 6-20: Open-loop equivalent circuit
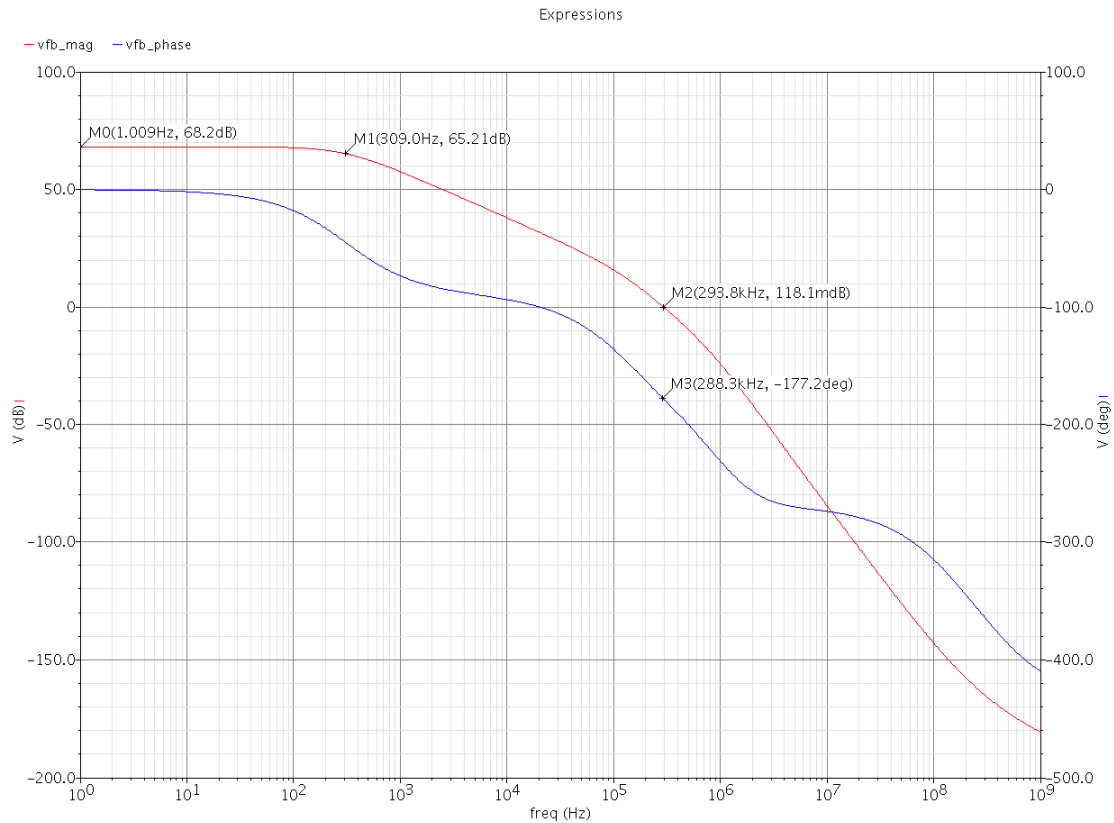
Figure 6-21: Open-loop bode plot

Because the phase at the unity-gain point is almost 180 degrees, the circuit is on the edge of instability. In order to do frequency compensation and make it stable again, it is necessary to extract the small signal parameters and calculate the poles and zeros of the circuit.

The relevant small-signal parameters are as follows.

R_{source}	660 kΩ
$g_{m,1}$	19.2 μA/V
$g_{m,2}$	2.2 μA/V
$g_{m,3}$	2.2 μA/V
$g_{ds,1}$	16.7 nS
$g_{ds,2}$	675 pS
$C_{gs,1}$	300 fF
$C_{gs,2}$	2.2 pF
$C_{gd,2}$	1.3 fF
C_{load}	9.4 pF
C_{Rfb}	450 fF
$C_{gs,bias\ mirror}$	130 fF
$g_{m,bias\ mirror}$	2.2 μA/V

Table 6-7: Small-signal parameters

Running a pole-zero analysis results in the following plot.
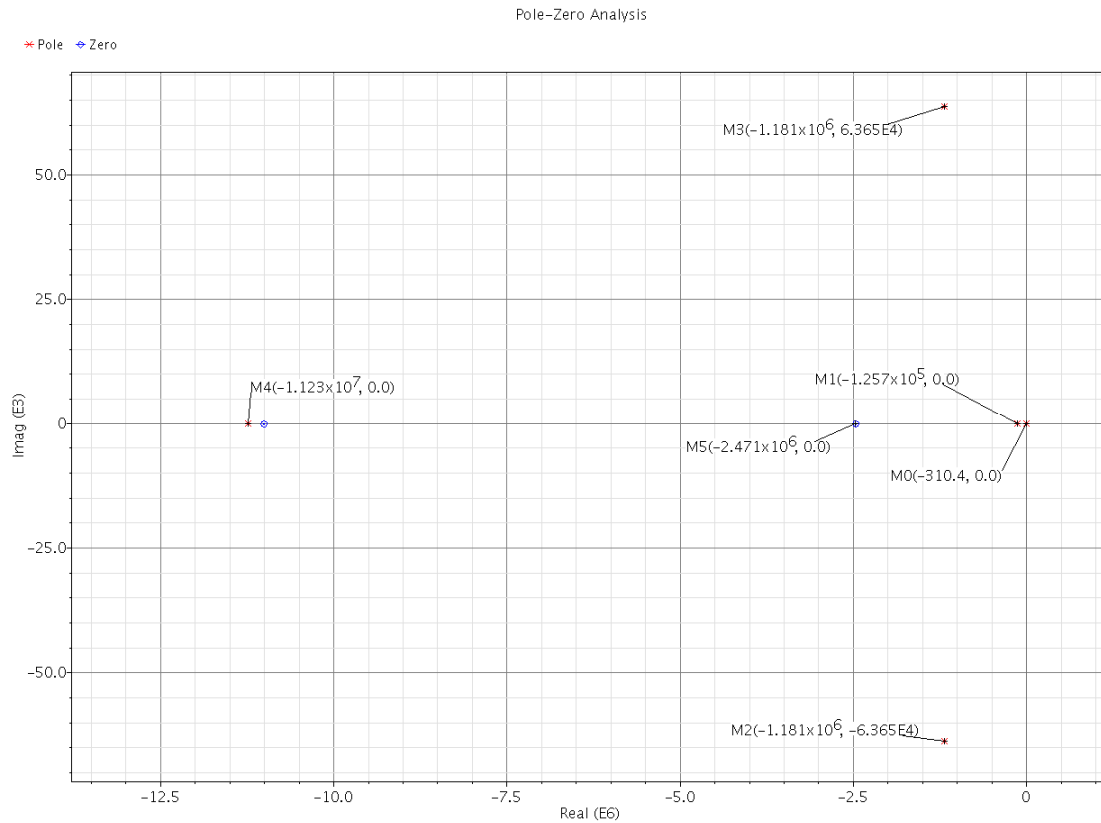
Figure 6-22: Open-loop pole-zero analysis

The open-loop poles should be directly related to circuit parameters (usually R-C combinations). The origins of the lowest four poles in Figure 6-22 are described in Table 6-8.

Pole	Frequency	Origin
p_1	310 Hz	$\dfrac{C_{load}}{\left(g_{ds,1} + g_{ds,2}\right)}$
p_2	126 kHz	$R_{fb} \cdot \left(C_{gs,1} + C_{R_{fb}}\right)$
p_3	1.2 MHz	$R_{source} \cdot C_{gs,1}$
p_4	1.4 MHz	$\dfrac{C_{gs,biasmirror}}{g_{m,biasmirror}}$

Table 6-8: Origin of poles

The third and fourth poles are very close together and are therefore shown to have small imaginary parts in the simulation.

When the loop is closed, the two low-frequency poles come together and then move up and down. The third and fourth poles do the same.
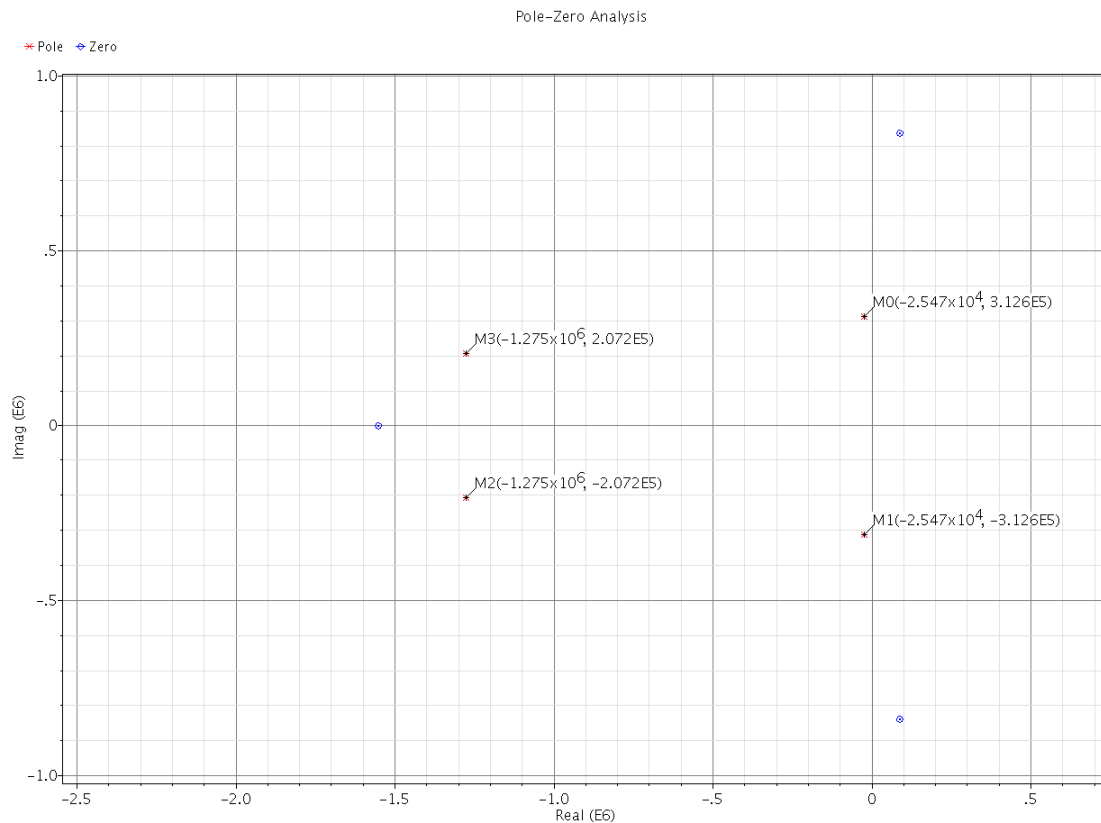
Figure 6-23: Closed-loop pole-zero analysis

Poles p_1 and p_2 need to be moved more to the left in order to make the system more stable. In its present state it will not oscillate, but it will ring, and it is dangerously close to oscillation. This is undesirable, and it would be better if the poles were more or less in Butterworth position.

There are three places where a phantom zero can be added (at least in principle):

- At the input of the amplifier. This would require adding a capacitor from the input terminal to ground. However, this capacitor would be in parallel with $C_{gs,1}$, and it would therefore only lower p_3.
- At output of the amplifier. This is not possible either, since the load impedance is already zero: the output current flows only through the feedback resistor and not through some load impedance.
- In the feedback network. Unfortunately this cannot be done either, because a resistive I-V feedback network would require an inductor, which is not possible in this technology.

The second option is pole splitting. This has to be done between the first and second pole, i.e. between C_{load} and $C_{gs,1}$. This might be possible, since $p_2 < p_1$. It will probably not be very effective, since the voltage gain of the second stage is low (8 dB or 2.5x). An attempt at pole splitting shows that adding a splitting capacitor will only cause *both* poles to shift to a lower frequency, so there is no actual splitting taking place. This is shown in Figure 6-24. Apparently, the voltage gain needs to be higher to make pole splitting possible.
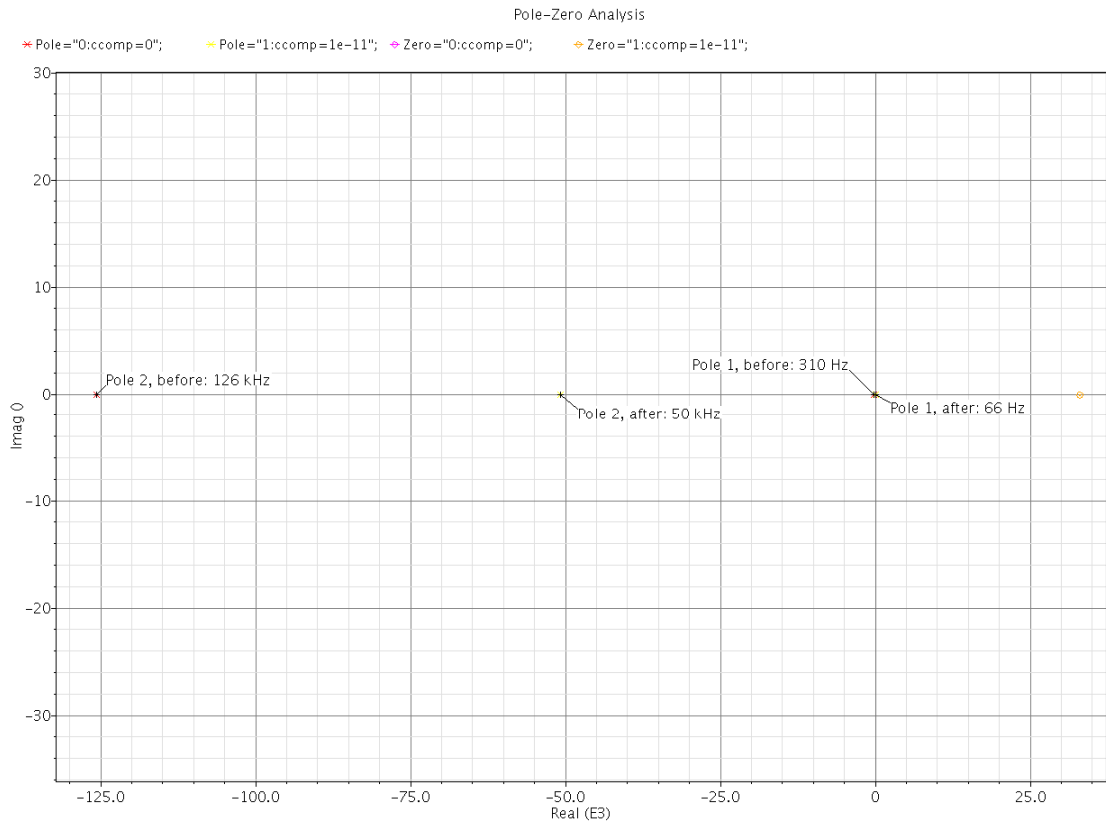
Figure 6-24: Pole-zero plot before and after attempted pole splitting

The next option is pole-zero cancellation. This can be done by adding an RC-section from the gate of MP0 to ground, causing a zero z_0 at $R_{comp} \cdot C_{comp}$ and a pole p_{comp} at $R \cdot C_{load}$. The zero z_0 should be located at the same frequency as p_2, so that they cancel, and p_{comp} should be higher in frequency than p_2. This is the case if $R_{comp} < R_{fb}$.
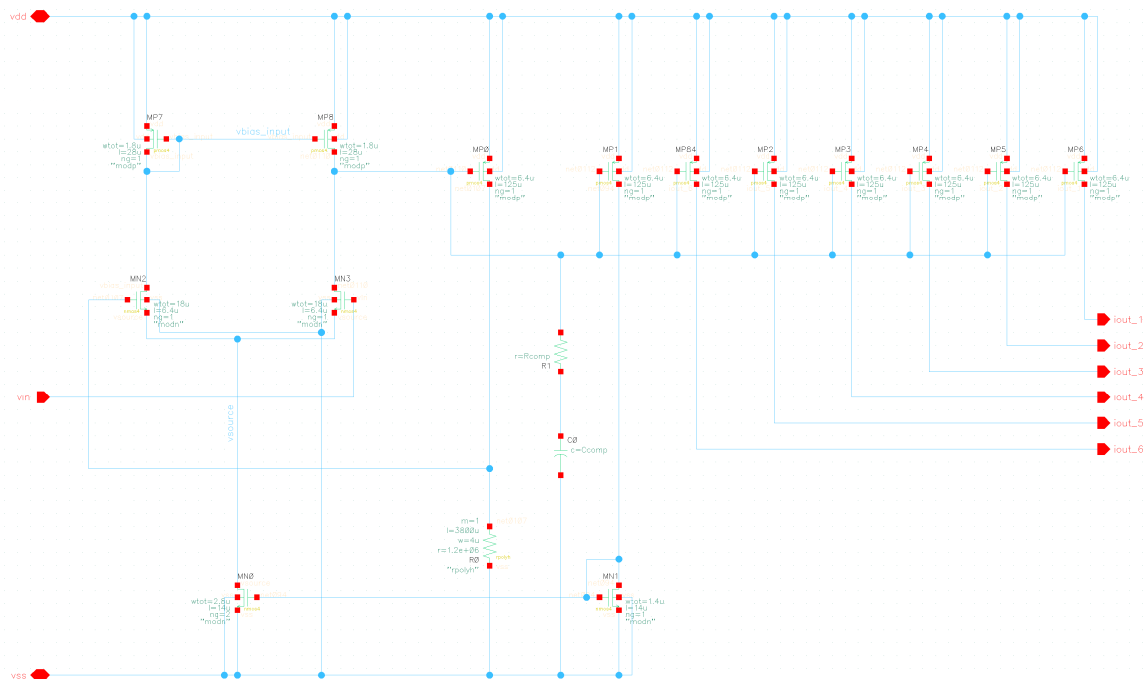
Figure 6-25: Current reference including pole-zero cancellation

The correct compensation resistance and capacitance to obtain a Butterworth transfer can be found by sweeping R_{comp} from R_{fb} downwards, while keeping $C_{comp} = p_2/R_{comp}$. Plotting the swept closed-loop poles and zeros results in the picture below.
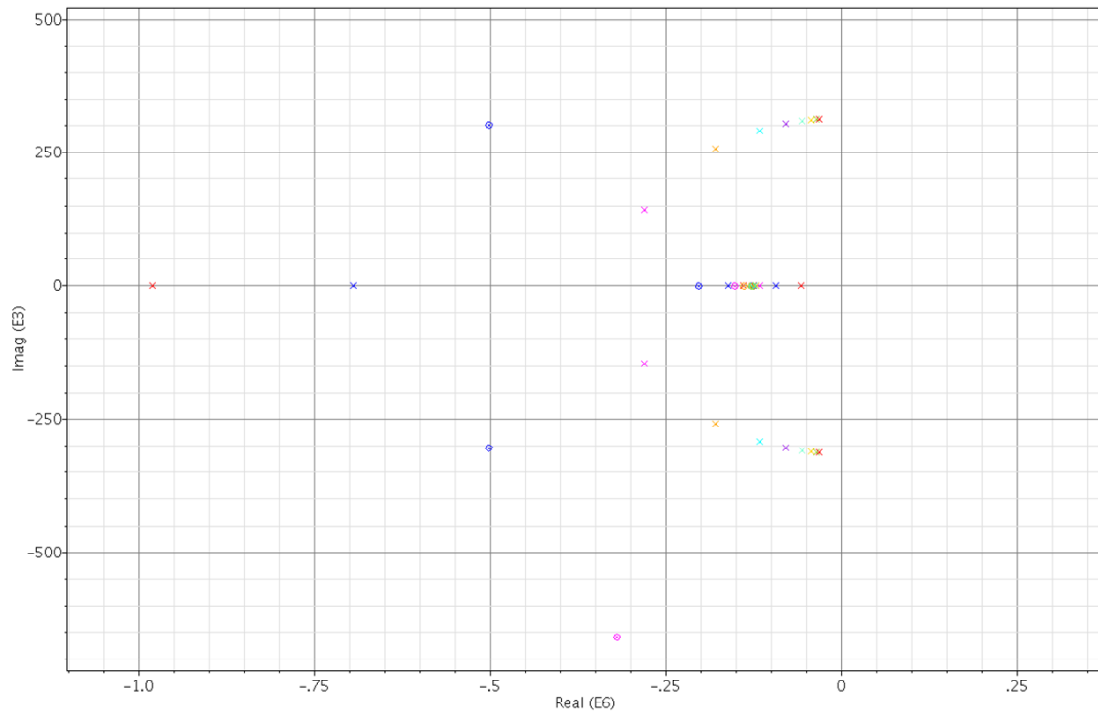
Figure 6-26: Closed-loop root locus for swept R_{comp} and C_{comp}

The exact value needed to obtain a Butterworth transfer can be found by zooming in (i.e., decreasing the sweep range). Note that the pole and zero on the real axis are close enough to cancel, as intended.
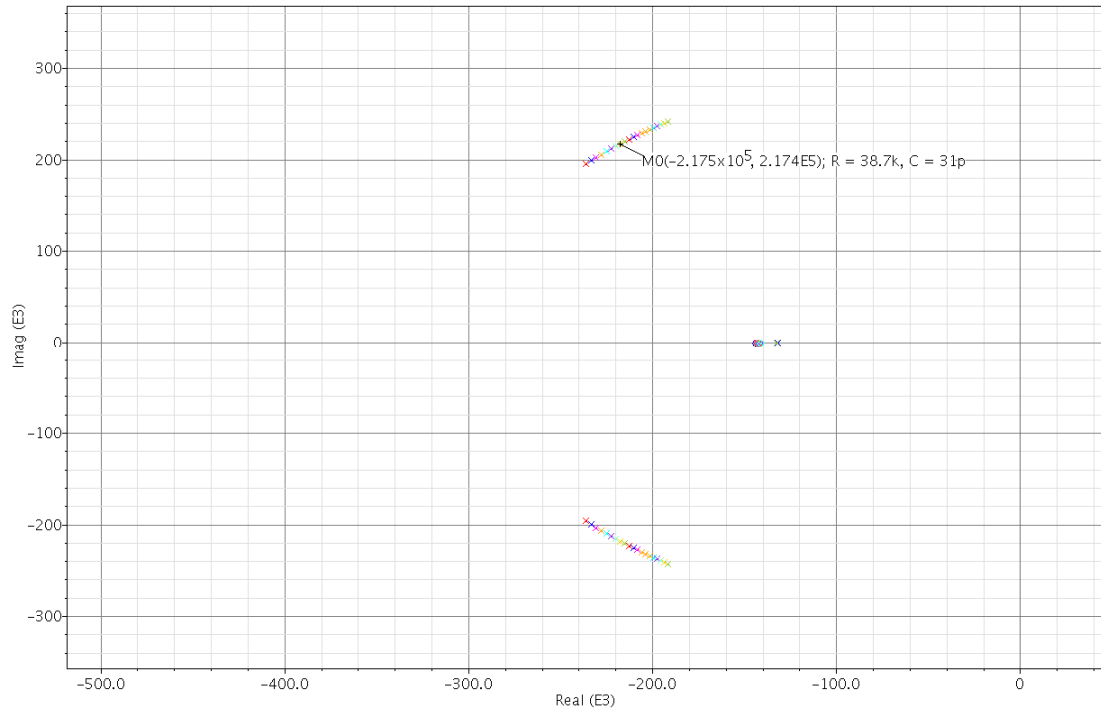
Figure 6-27: Close-up of root locus around Butterworth position

This is a workable implementation. However, it is possible to save chip area by using the capacitance that is already present at the output instead of adding another capacitor. This has the added benefit of moving the low-frequency pole to a higher starting frequency. Figure 6-28 shows how this is implemented.
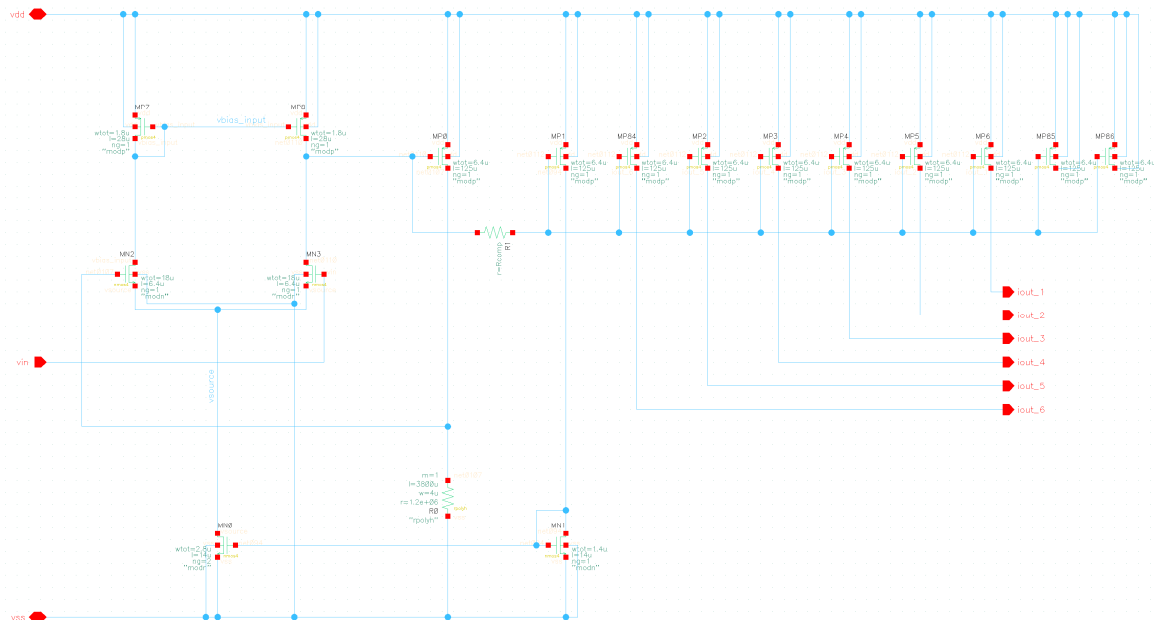
Figure 6-28: Compensation saving chip area

The additional capacitance needed to stabilize the amplifier is implemented by adding two additional output transistors. For now, they will act as dummies, but if it later appears that more copies of the output current are needed, they can be easily converted to active devices.

6.2.4 Circuit finishing

Now the compensation resistor is implemented with a real resistor from the Austriamicrosystems library. Its parasitic capacitance is connected to V_{dd}, to improve the PSRR. Simulating over all process corners leads to the following pole-zero diagram:
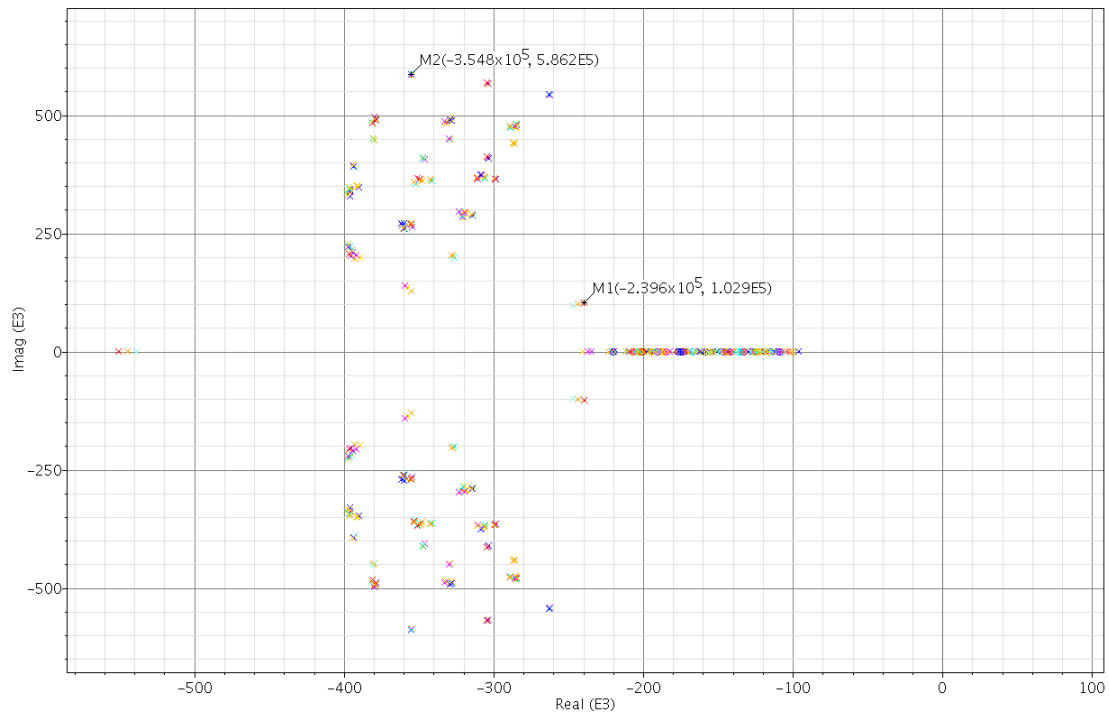
Figure 6-29: Pole-zero plot of the compensated circuit over all corners

Marker M1 and M2 are at high and low temperatures, respectively. The three branches of the root locus correspond to the three resistor corners. Although the circuit is stable at M1, it also has a very limited bandwidth and will therefore be rather slow. This can cause problems during start-up of the circuit. Some small adjustments in the resistor and capacitor values result in the plot below.
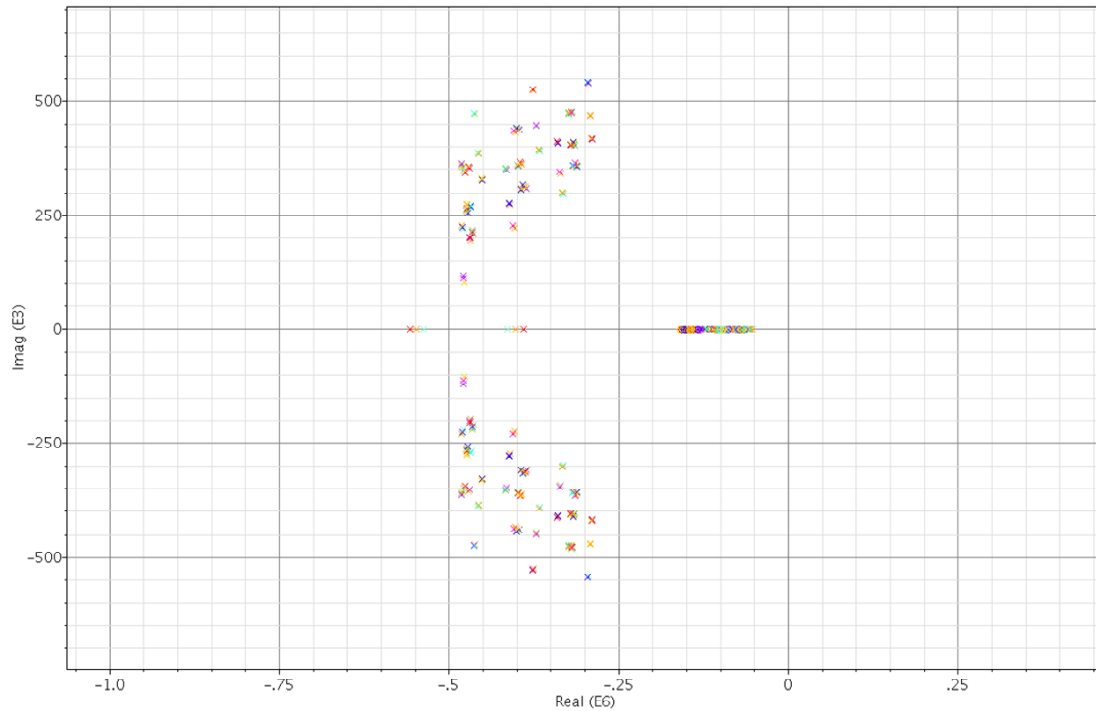
Figure 6-30: Pole-zero plot after adjustments

The last step is to do a transient simulation to verify that the circuit does not oscillate, and that it starts up correctly. For this to work reliably, the disable circuit is implemented, together with a start-up circuit.

The start-up circuit is necessary because the amplifier is biased by its output, in the same way as the bandgap reference. The start-up circuit is also implemented in the same way, although it only needs to start up the input differential pair.

The start-up circuit works by passing a copy of the output current through a resistor to create a voltage V_{start}. If the current reference is working correctly, this voltage will approach V_{dd}, causing the start-up transistor (MP104) to turn off. If the output current is zero, then the gate of MP104 is pulled towards V_{ss}, and MP104 will turn on, providing a "jump start" to the biasing mirror.

For clarity, the disable circuit is omitted from the following schematic. The disable circuit is described in more detail in section 6.3.
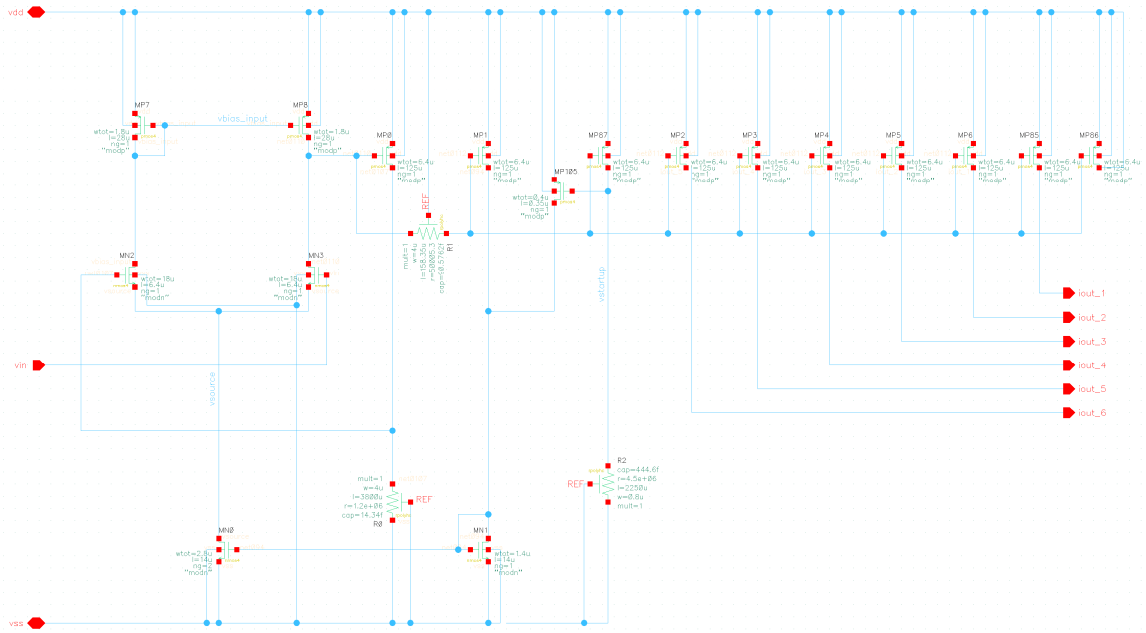
Figure 6-31: Complete current reference circuit without disable circuit
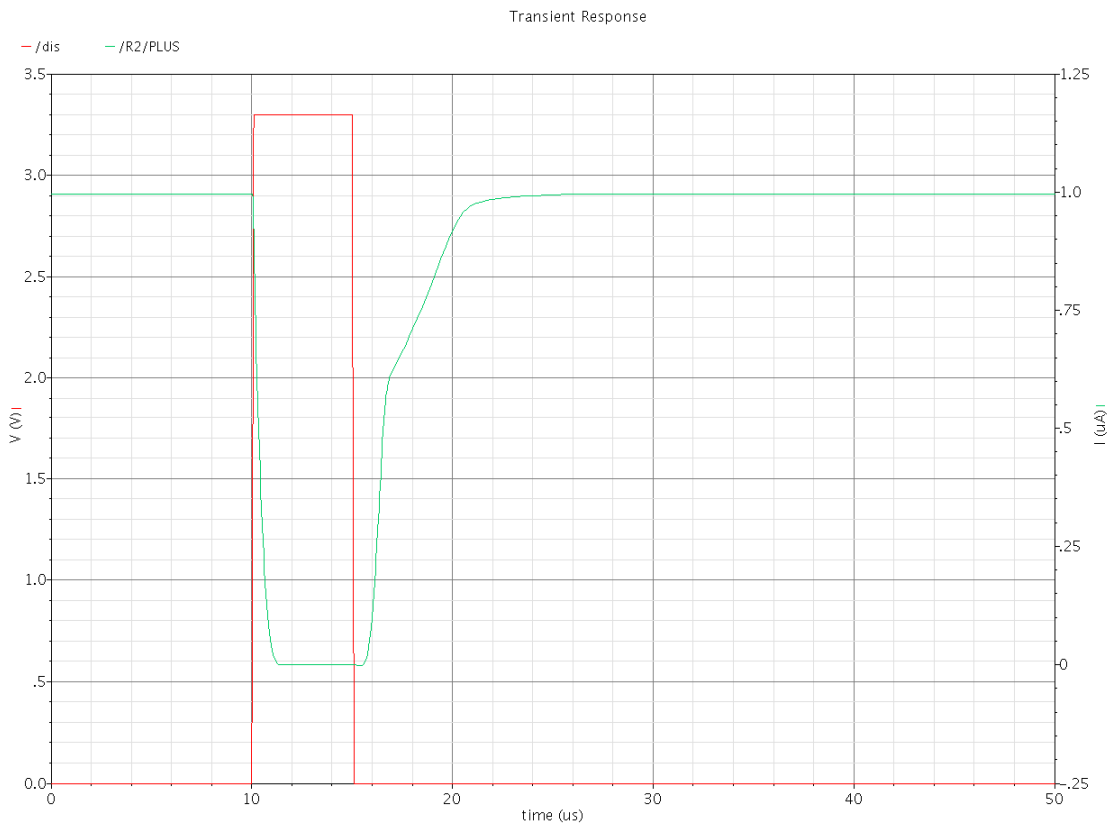
Figure 6-32: Nominal transient start-up behaviour. Red curve is disable signal

The current reference keeps working over all process corners. As expected, the absolute value of the output current is not very accurate. It does however remain within the specified 30% of its nominal value.
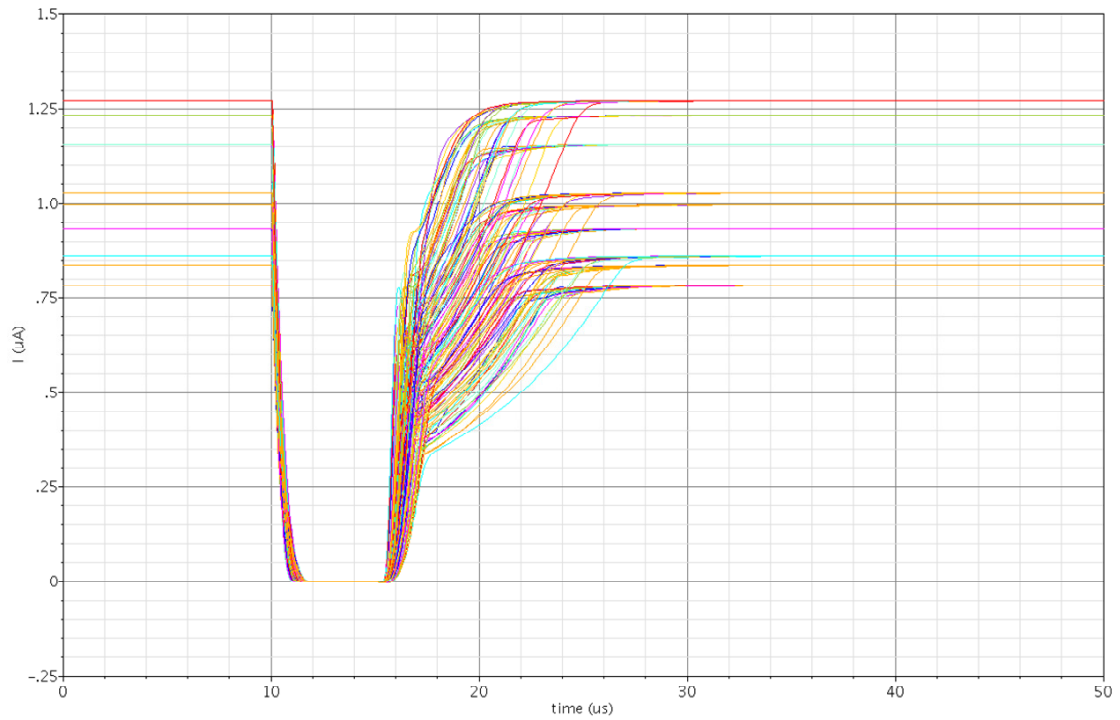
Figure 6-33: Start-up behaviour over all process corners

Next, an AC simulation shows that the power supply rejection ratio is below 25 dB up to 1 MHz over all corners.
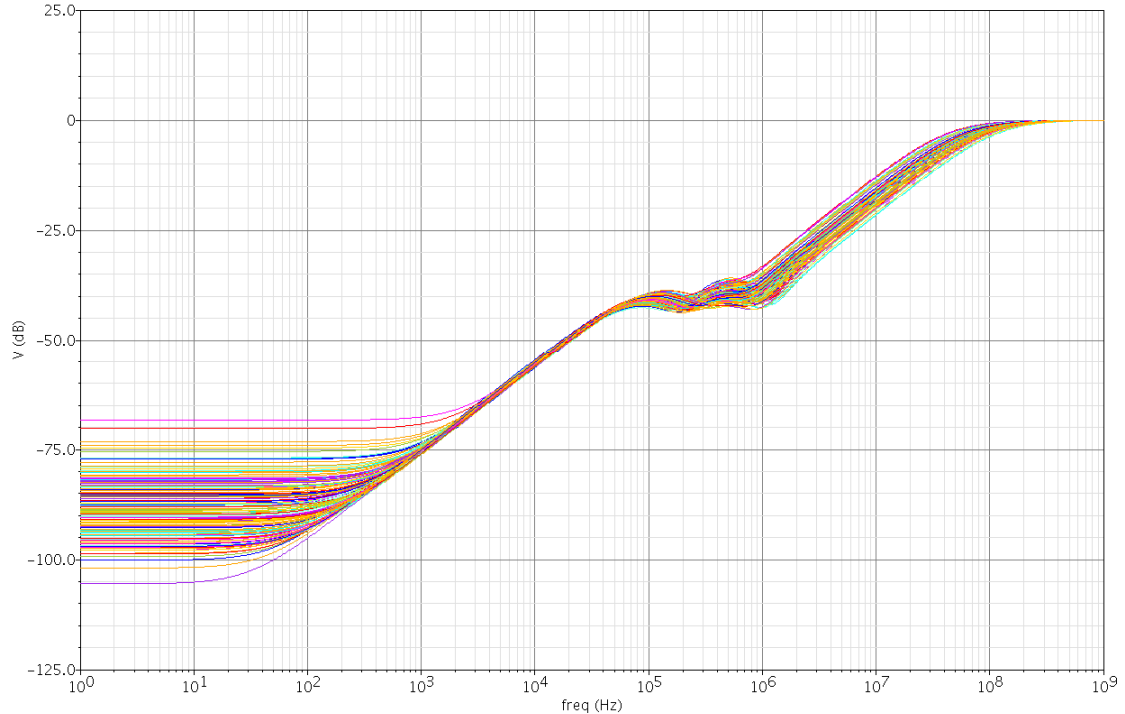
Figure 6-34: AC transfer from power supply to output

6.3 Test controller

The function of the test controller is to provide easy access to signals within sub-circuits that provide information on their functioning. It can also provide a way to

inject signals into the circuits to affect their functioning. All of this is necessary because it is almost impossible to probe for signals directly inside a working IC.

It is not deemed necessary to be able to probe into more than one place at a time, so the test controller is designed to allow one input/output pin to be connected to one place in the circuit at a time. It is however possible to set several switches to selectively enable or disable certain sub-circuits. This selection is made through a simple serial digital interface.

The implementation of this serial interface is shown below. It consists of two strings of D-flip-flops in a daisy-chain. A power-on reset circuit makes sure that it always powers up in the same state. To select a certain mode, a series of bits is clocked into the lower row of flip-flops using the test_clk and test_data pins, and confirmed by a pulse on the test_set pin. The bits are then transferred to the top row of flip-flops, whose outputs are connected to switches that select certain functions in the sub-circuits. This two-stage approach ensures that the switches are only driven after all bits have been set.
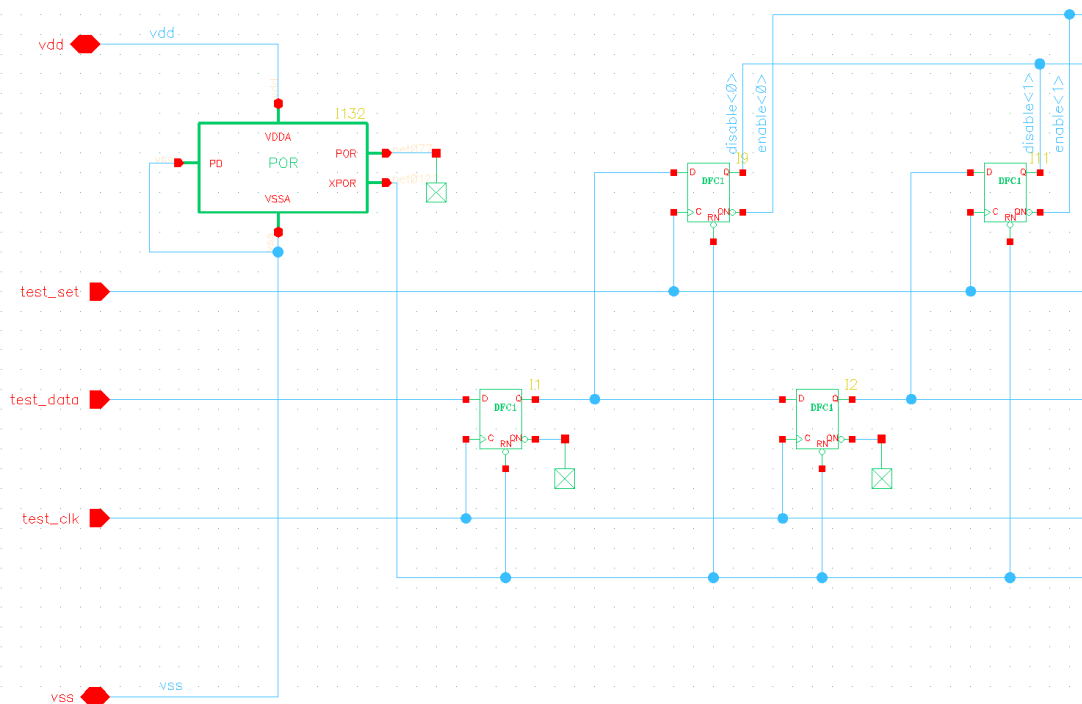
Figure 6-35: Basic schematic of test controller, showing two sets of flip-flops

The power-on reset and the flip-flops are taken from the Austriamicrosystems standard logic library, because there are no special requirements for these blocks.

The test controller is also a convenient place to implement the "straight PWM" function that was listed in chapter 1 as a "nice to have" feature. For this purpose, another pin called v_pwm is controlled by the test controller.

The full schematic in Appendix D shows how the test_probe pin is connected to the bandgap reference or to the current reference. In a similar way, the v_pwm pin is connected to or disconnected from the output of the PWM generator.

The spare transistor in the current reference that normally acts as a compensation capacitor is used to probe for the current delivered by the current reference. The current reference also has a "bypass" pin that can be enabled or disabled by the test controller. This reconnects the dummy transistor in the current mirror in such a way that it becomes the input of the current mirror, allowing the test operator to vary the output current of the current reference.

In order to facilitate testing, each sub-circuit has a "disable" pin that completely turns off the circuit. When a circuit is disabled, all current mirrors have their gates pulled to their source, outputs are fixed to a certain state, and any start-up circuits are disabled. The exact implementations can be found in the full schematics in Appendix D.

Table 6-9 below shows the function of all the bits that can be set in the test controller. The power-on state of each bit is zero.

Bit	Result if set
0	Disable and disconnect bandgap reference
1	Connect test_probe to bandgap voltage
2	Disable current reference
3	Connect test_probe to current reference If current reference is enabled (bit 2 low), allows measuring the current If current reference is disabled (bit 2 high), allows overriding the current
4	Disable and disconnect comparator in PWM generator
5	Disable triangle generator (leaving output floating)
6	Connect v_pwm to PWM input of output stage
7	Disable dynamic transistor sizing (leaving control to bits 8-10)
8	Enable output stage part 1
9	Disable output stage part 2
10	Disable output stage part 3

Table 6-9: Function of bits in test controller

7. Simulation results

Now that the complete circuit has been designed, it is time to test it in the simulator to see whether it performs as desired.

7.1 Efficiency

First, the input of the class D amplifier is fed a DC voltage and allowed to settle. This DC voltage is swept across the entire input range, and for each step the power efficiency of the amplifier is measured. The results are plotted below.
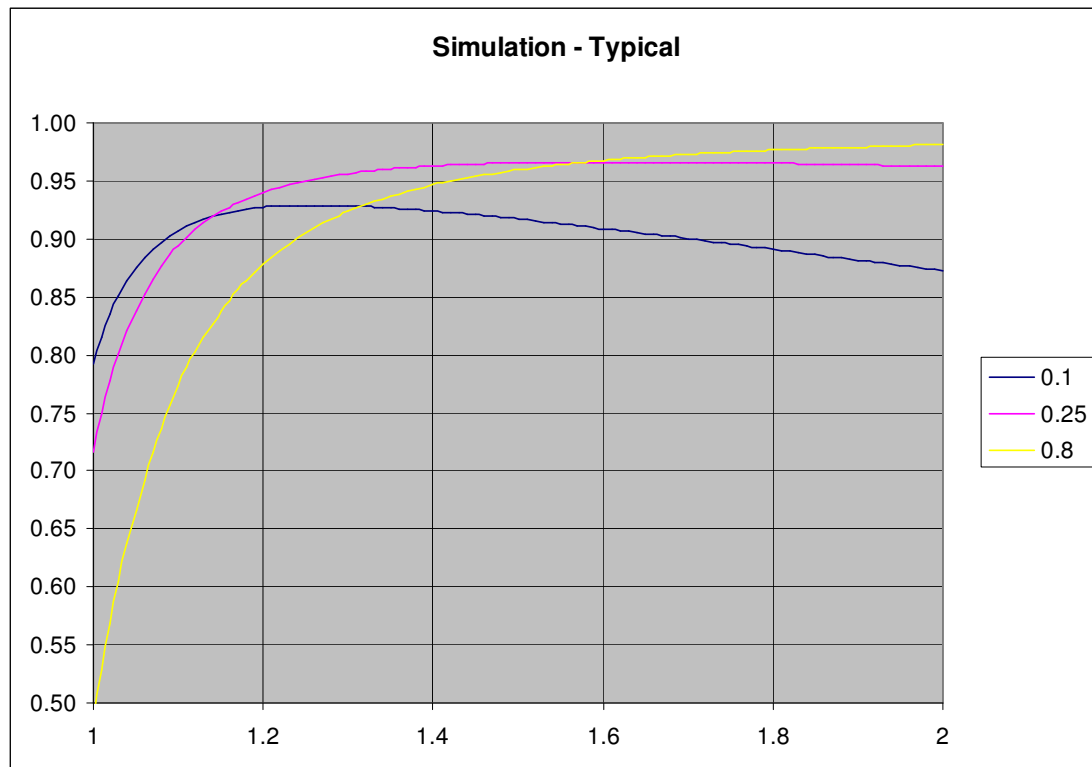
Figure 7-1: Efficiency as a function of input voltage for each transistor size

This clearly shows that dynamic transistor sizing can be beneficial. It also shows the exact location of the handover points. These are 1.14 V for the first transition and 1.58 V for the second. The respective resistor sizes to generate these voltages can now be implemented. A second simulation can then be run to verify that the dynamic transistor sizing circuit works.
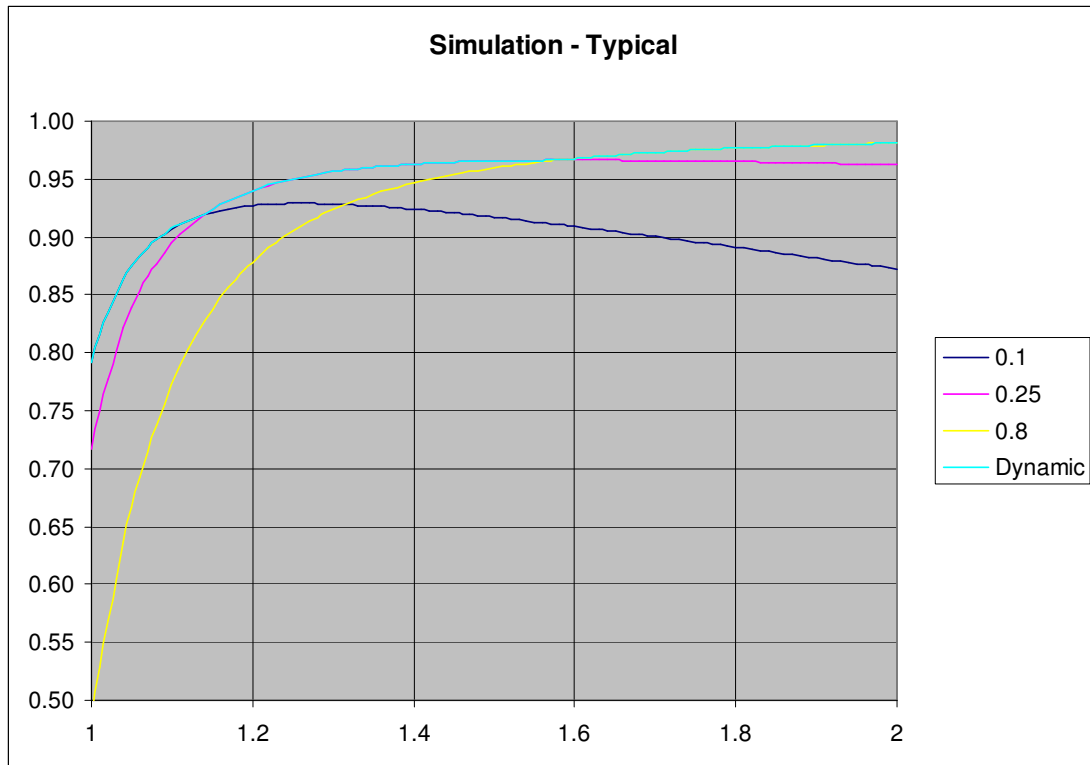
Figure 7-2: Efficiency as a function of input voltage using dynamic transistor sizing

The blue curve in Figure 7-2 correctly takes over from the three separate curves at the pre-determined duty cycles. Efficiency improvements of several percent can be achieved compared to a single-sized output stage.

7.2 Distortion

Next, a ramp voltage that runs through the entire input range is applied, to give a first idea about the linearity of the amplifier. The resulting waveforms are shown in Figure 7-3.
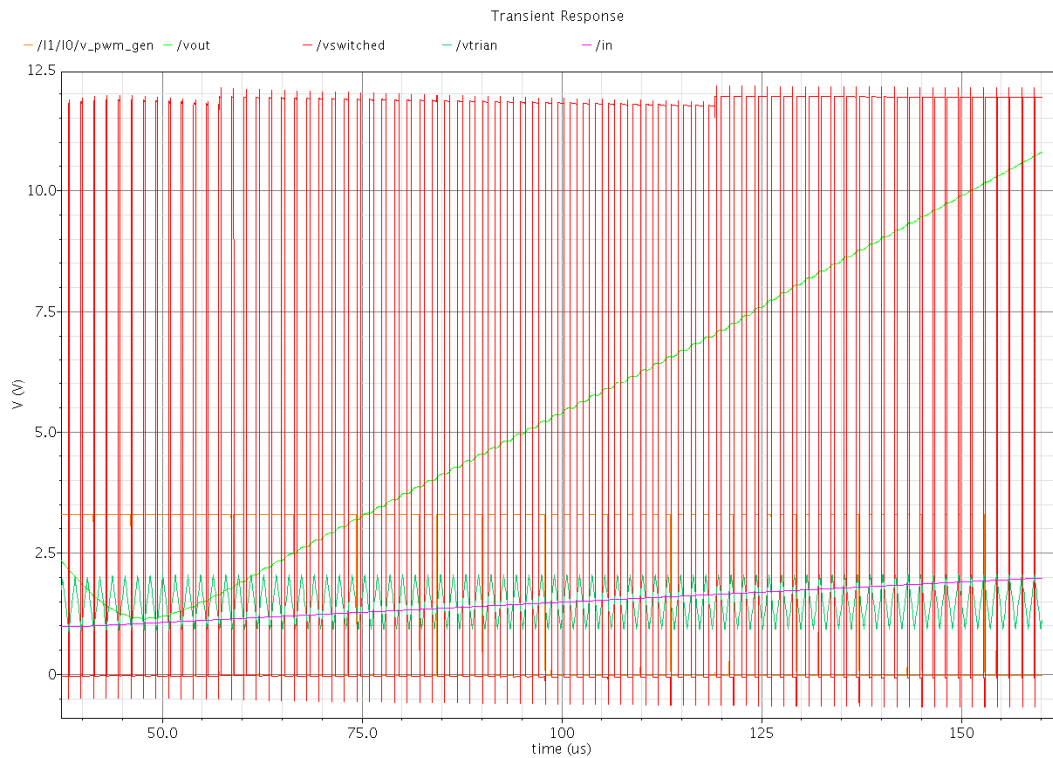
Figure 7-3: Waveforms for a 1 – 2 V ramp input

It is clear from V$_{switched}$ (red) that the dynamic transistor sizing is working. It appears that the output voltage correctly tracks the input voltage (the hook at the far left is due to the start-up swing). If severe distortion were present, the output voltage would be curved instead of straight. This is not an accurate measure of distortion however, and another simulation has to be run to test whether the amplifier meets its distortion requirements. This involves amplifying a large amplitude sine wave and measuring the harmonic content of the output signal.

First, this is done with a constant transistor size (the largest). A 10 kHz sine wave with maximum amplitude is applied at the input. The time-domain signals and the corresponding frequency-domain spectrum are shown below.
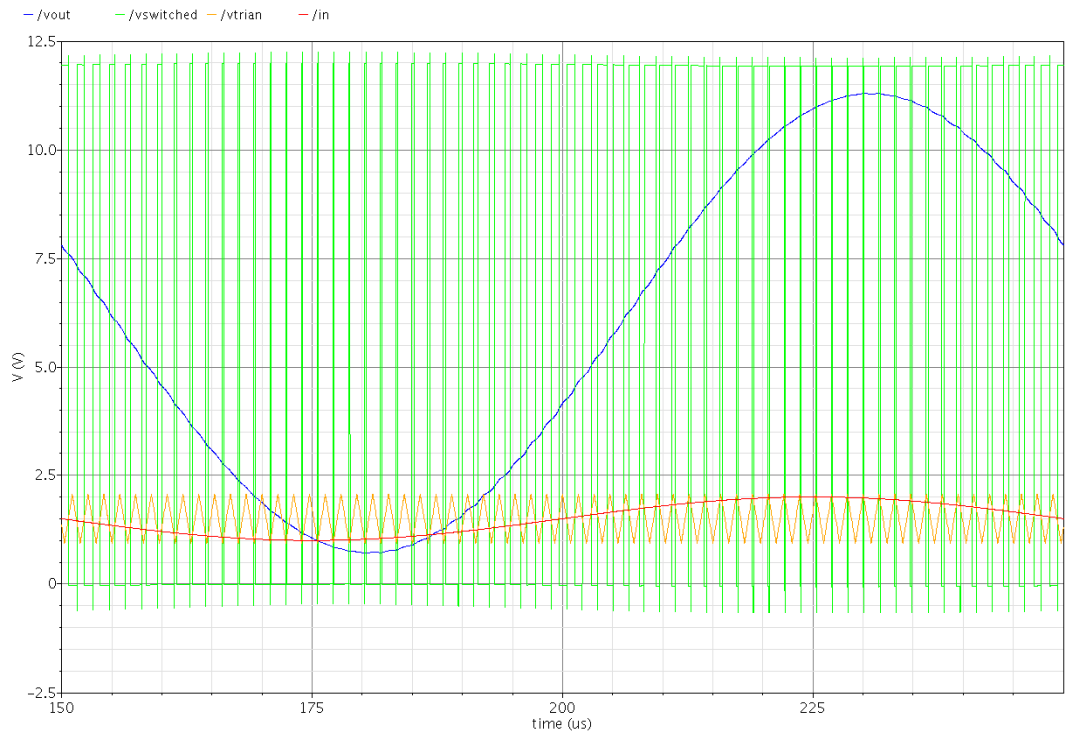
Figure 7-4: Time-domain signals for a 10 kHz sine wave, at maximum transistor size
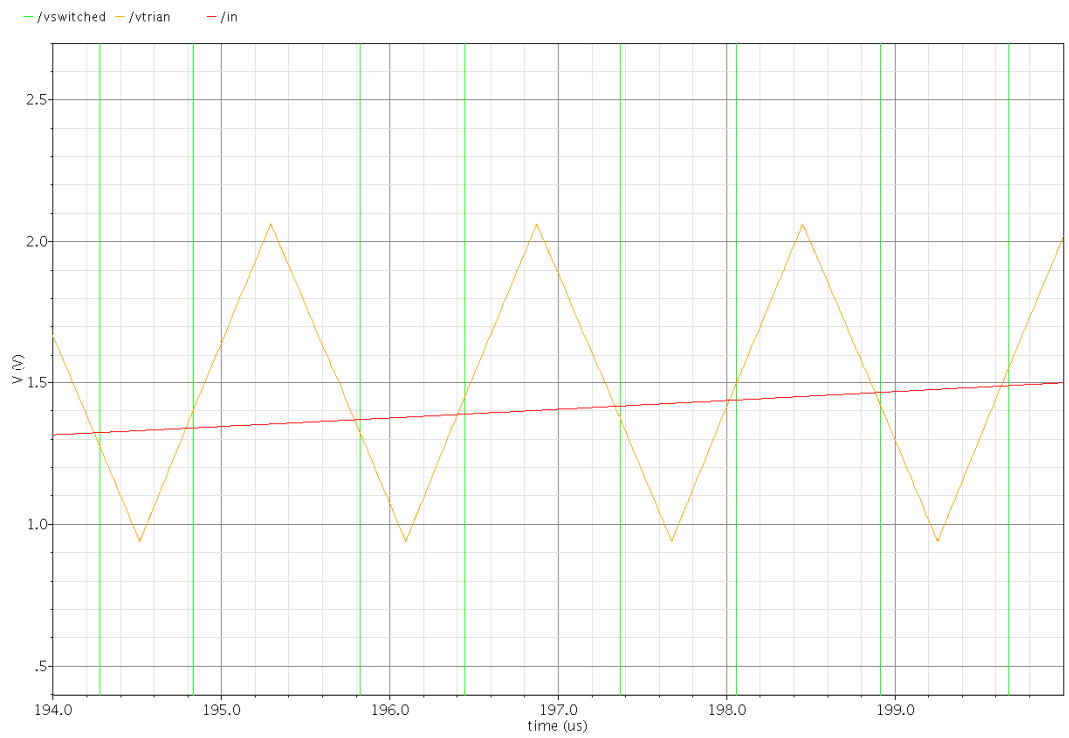
Figure 7-5: Close-up of input, triangle and switched voltage
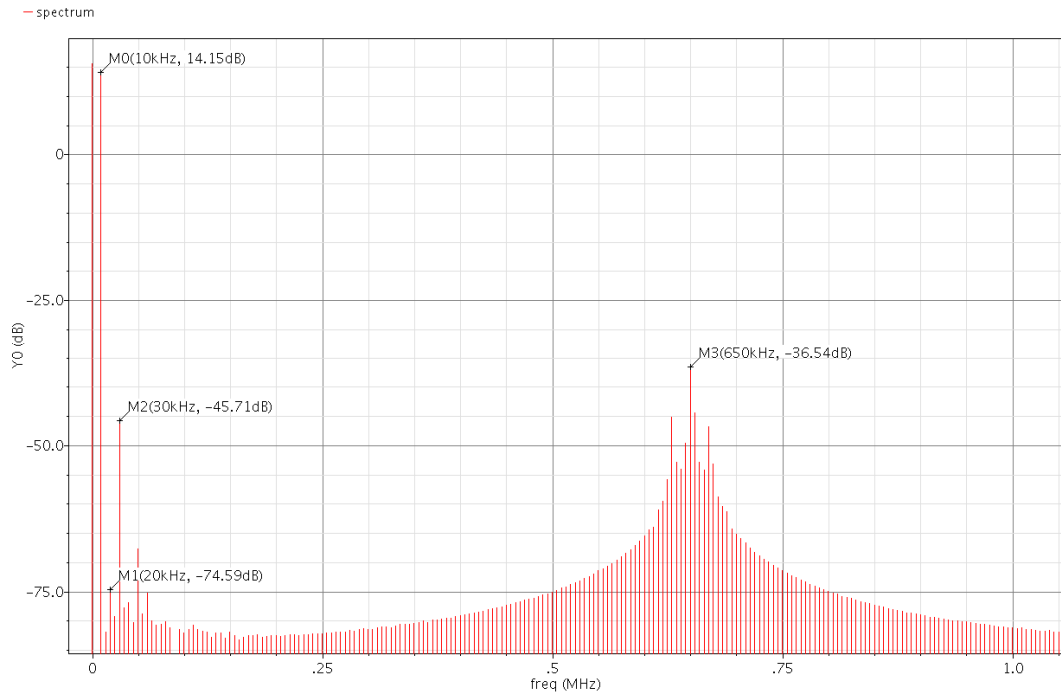
Figure 7-6: Frequency-domain spectrum for a 10 kHz sine wave, at maximum transistor size

The spectrum shows that the second harmonic at 30 kHz is almost 60 dB below the fundamental. This is much better than required. The overall efficiency for the entire signal swing is 96.8 %.

Secondly, the same simulation is done with dynamic transistor sizing turned on. The results are shown below.
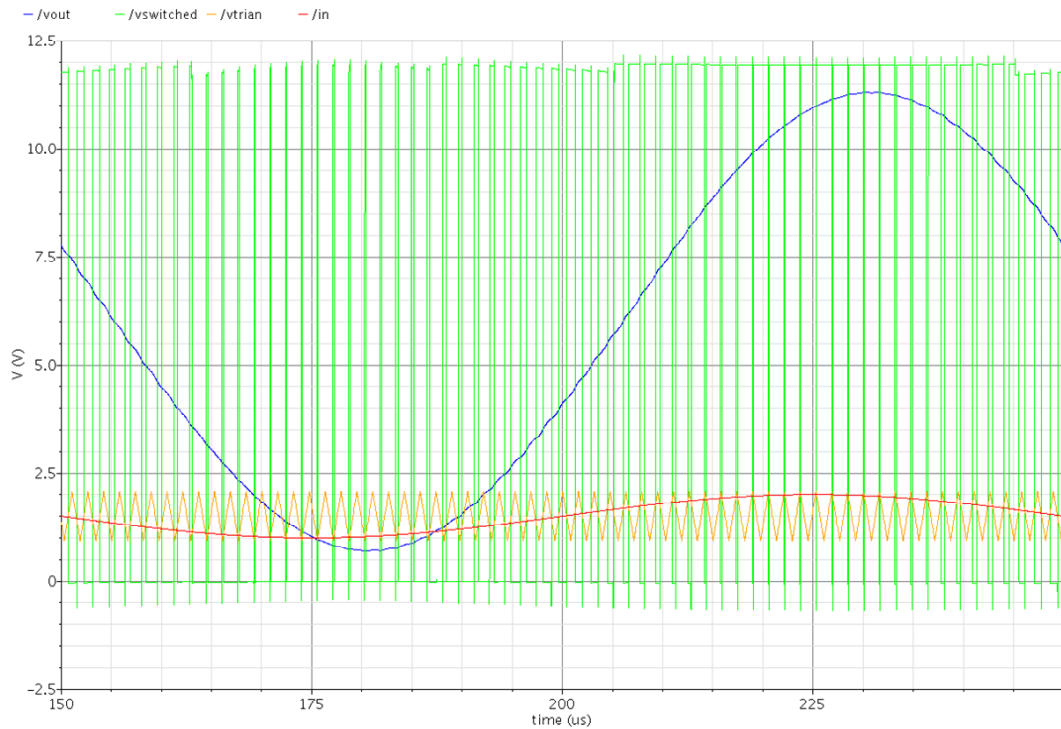
Figure 7-7: Time-domain signals for a 10 kHz sine wave, using dynamic transistor sizing
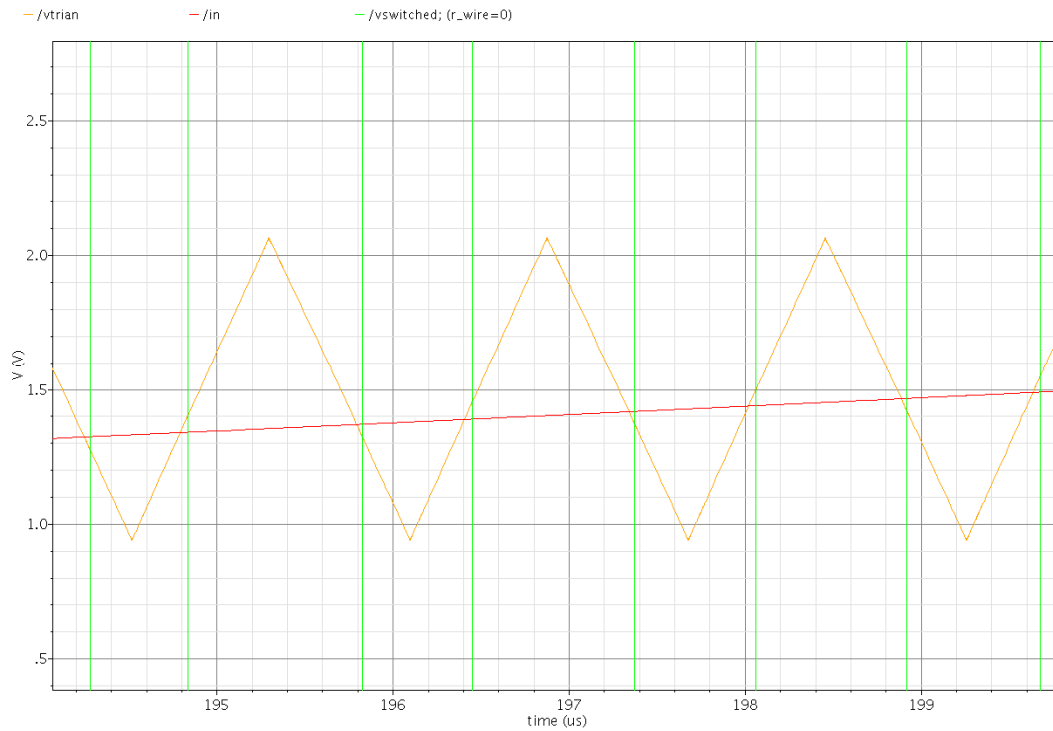
Figure 7-8: Close-up of input, triangle and switched voltage
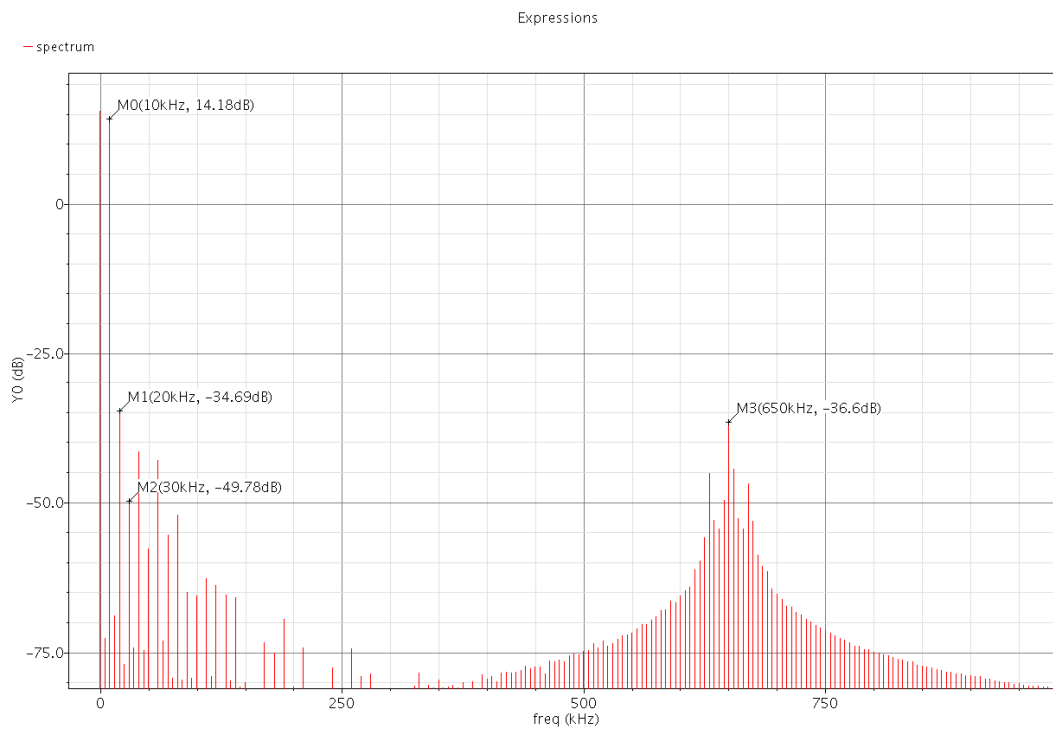
Figure 7-9: Frequency-domain spectrum for a 10 kHz sine wave, using dynamic transistor sizing

Figure 7-9 shows that using dynamic transistor sizing causes a change in the distortion performance. The second harmonic is even lower than in the previous case, while the first harmonic at 20 kHz is much stronger. In either case however, the circuit easily meets its distortion specifications. When using dynamic transistor sizing, the efficiency is increased to 97.5 %.

Finally, simulations are performed to determine the amount of intermodulation distortion. An input signal consisting of two sine waves at 10 kHz and 15 kHz is applied, and the spectrum of the output signal is plotted. Intermodulation products appear at 5 kHz and 20 kHz.
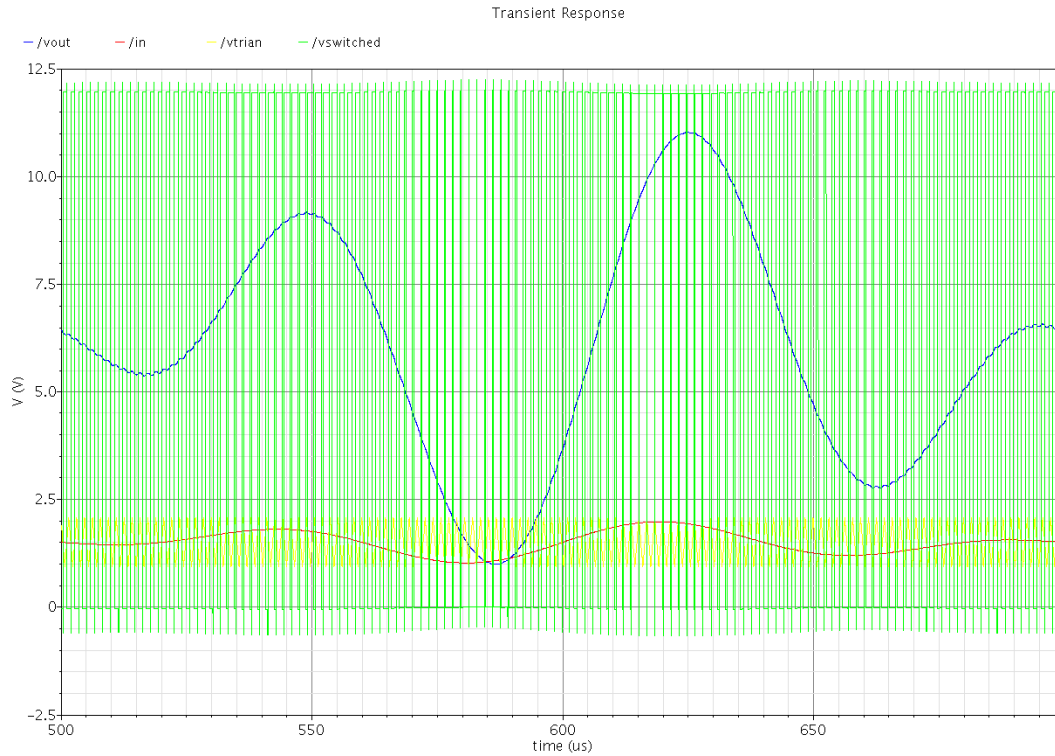
Figure 7-10: Time-domain signals for 10 khz + 15 kHz sine waves, at maximum transistor size
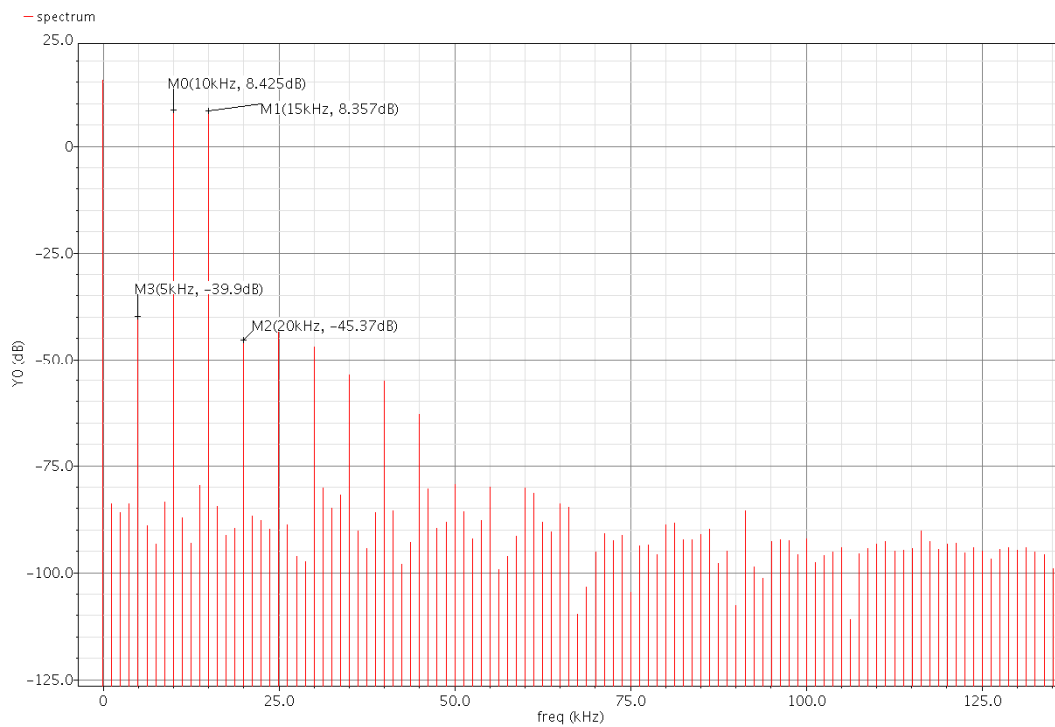
Figure 7-11: Frequency-domain spectrum for 10 kHz + 15 kHz sine waves, at maximum transistor size

The spectral analysis shows that the third-order intermodulation products are located about 48 dB below the signal frequencies. This is well within the -30 dB specification.

Turning on dynamic transistor sizing results in the plots below.
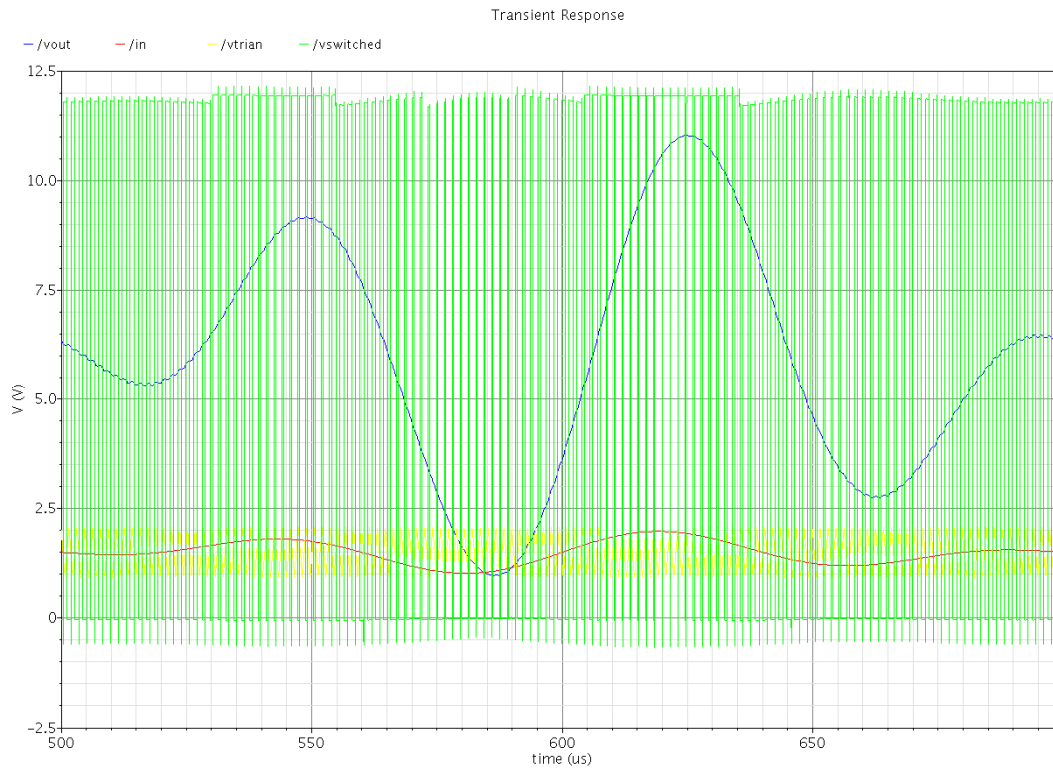
Figure 7-12: Time-domain signals for 10 khz + 15 kHz sine waves, using dynamic transistor sizing
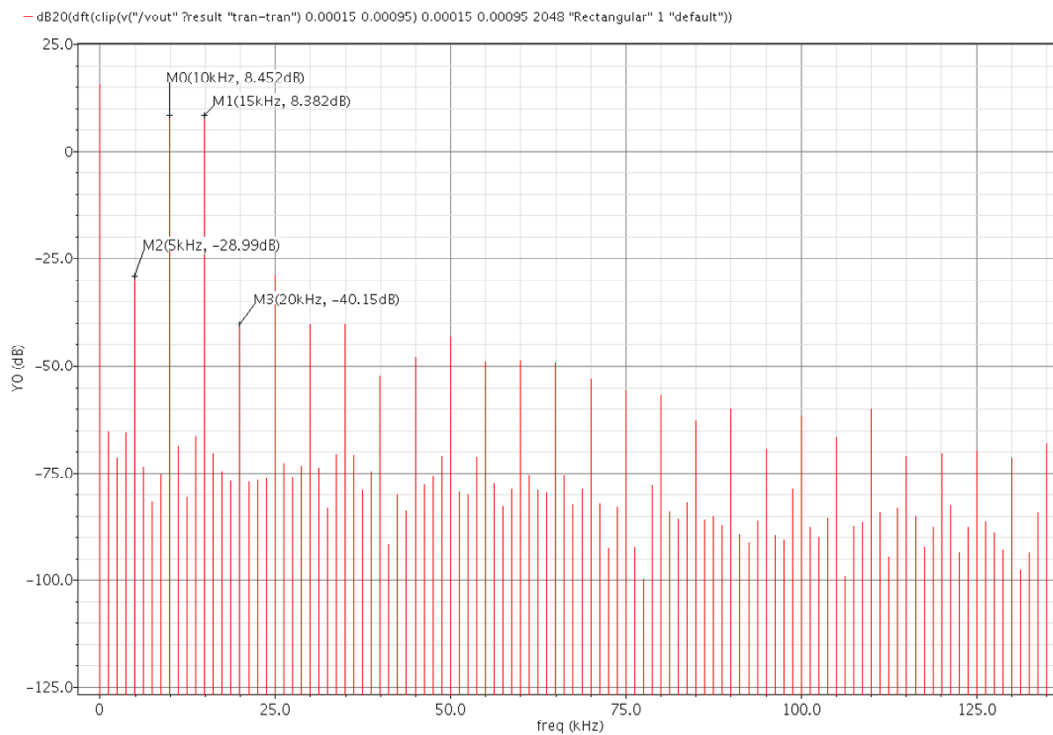
Figure 7-13: Frequency-domain spectrum for 10 kHz + 15 kHz sine waves, using dynamic transistor sizing

Figure 7-13 shows that the intermodulation distortion is increased by dynamic transistor sizing to -37.4 dB. This is still well within the specifications.

One problem that often plagues fast switching IC designs is the parasitic inductance of the bondwires that connect the IC to the outside world. The inductance and resistance of a bondwire are around 1 nH/mm and 150 mΩ/mm [17]. A simulation with 1 nH at both nodes of the 12 V power supply ($V_{dd,p}$ and $V_{ss,p}$) shows that spikes of more than 2 V appear on the supply lines.
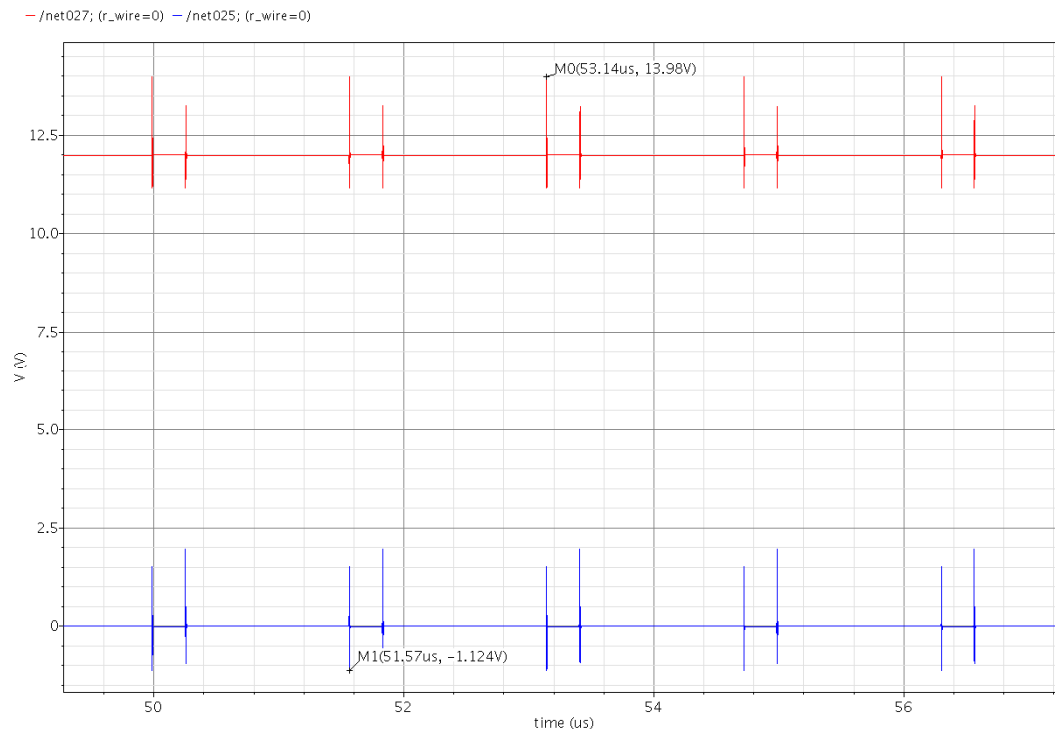
Figure 7-14: Voltage at supply nodes with 1 nH of parasitic inductance

These spikes are small enough to not damage the transistors, but it is still preferable to limit them as much as possible, since they will likely cause the IC to emit noise on the supply lines. If the parasitic inductance is reduced by a factor of four (i.e., using four bondwires in parallel), then the spikes on the V_{ss} line are reduced to about 875 mV.
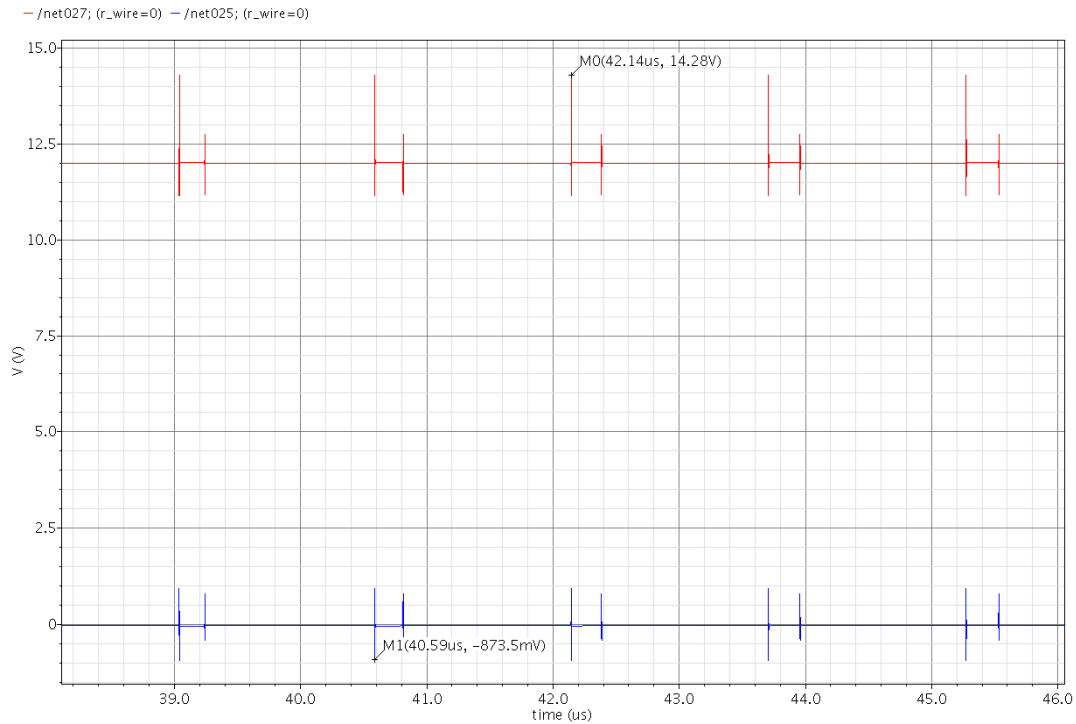
Figure 7-15: Voltage at supply nodes with 0.25 nH of parasitic inductance

While this is still not ideal, it is better than before, and using more bondwires would lead to problems with layout and packaging. It will be necessary to assess the amount of power supply noise caused by the chip during testing, and to take appropriate measures to protect other circuits from it. This will include adding supply filtering and optimizing the PCB layout to minimize noise injection and interference.

7.3 Conclusions

The results of the simulations can now be compared to the specifications.

Specification	Required	Achieved
Frequency range	DC – 40 kHz	DC – 40 kHz
Maximum output voltage	10 V	11 V
Out-of-band spurious signal level	-44 dBc	-50 dBc
In band 3rd order intermodulation	-30 dB	-37.4 dB
Power efficiency	As high as possible	97.5 %

Table 7-1: Simulation results compared to specifications

Table 7-1 shows that the circuit meets or exceeds all specifications. Power efficiency, which is the most important specification, is very close to the theoretical maximum. The distortion performance exceeds the specifications by several dB.

8. Layout

After the complete design and simulation cycle has been completed, it is time to make a layout of the circuit that can be used to make the actual IC in the foundry. This chapter will show the parts of the layout that required special attention. The layout of the complete IC can be found in Appendix E.

8.1 Output transistors

The layout of the output transistors is of key importance in the performance of the complete circuit. Parasitic resistances and capacitances can affect the power efficiency. Also, the output transistors are by far the largest component on the IC, and therefore largely determine the amount of chip area required.

The idea for the output transistors is to shape them in such a way that they occupy a rectangle that is somewhat taller than it is wide. This allows for room to place the gate drivers and support circuits to the left, with the goal of making the complete die approximately square.
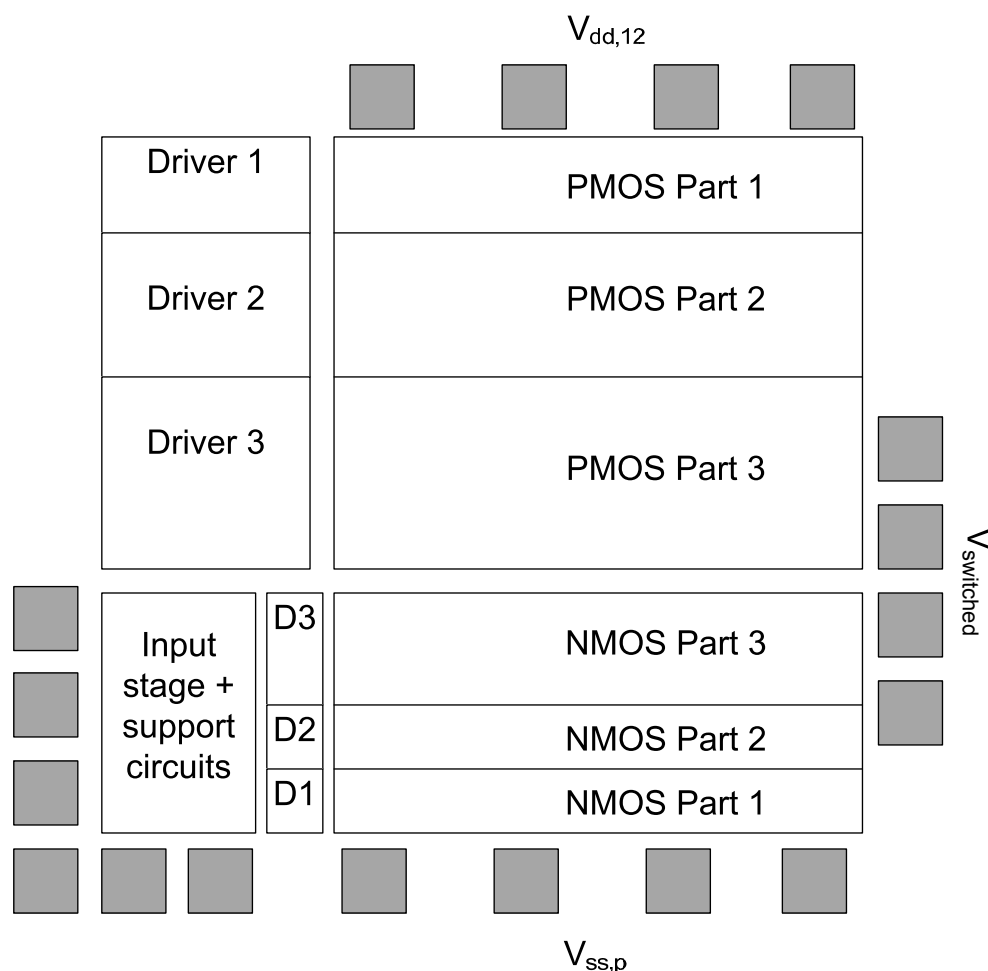
Figure 8-1: Floor plan of complete die

All output transistors will have their drain connected to the output node and their source to $V_{dd,p}$ (for PMOS) or $V_{ss,p}$ (for NMOS). As calculated in chapter 7, these

three nodes will be connected to the outside world with at least four bondwires, to minimize parasitic resistance and inductance.

A total of 300 mΩ was allocated in section 4.1 for parasitic resistances. If the metal resistance is allowed to take no more than 100 mΩ of this, then there should be no more than 7.5 squares of power metal between V_{dd} or V_{ss} and the output node, since the power metal layer on the chip has a worst-case sheet resistance of 13 mΩ/□.

Dividing the PMOS transistors into rows of 134 transistors and the NMOS transistors in rows of 126 makes for a conveniently sized area. The most straightforward way of connecting them is by simply running wires in the vertical direction, connecting all sources or drains together. Figure 8-2 shows this concept, with five rows of five PMOS transistors.
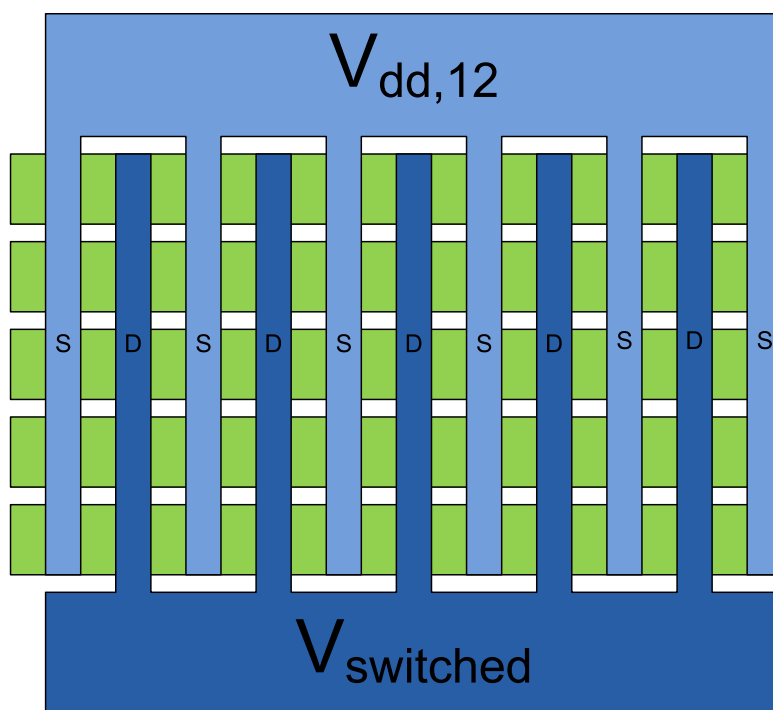
Figure 8-2: Basic layout of PMOS output transistors

The wires are 2.65 μm wide, which is the maximum possible width without violating clearance rules. The total length is 760 μm, which means that the strip has a length of 287 squares. Since each transistor has two of these wires connecting its source leads, and one connecting its drain lead, the total resistance in series with each transistor is related to its vertical grid position. On average however, each transistor has ¾ of a strip in series. Since all these strips are again in parallel, this means that the resistance between V_{dd} and the output pin is 1.7 squares.

The bar connecting the PMOS and the NMOS sections to the output pins is about four squares wide, which means an average of two squares for each transistor. The total metal resistance then becomes 3.7 squares for the PMOS section. Since the NMOS section is shorter, its resistance is even less than this. The on-chip wiring resistance therefore meets its resistance specifications easily.

Each row of PMOS transistors is surrounded by a P-type guard ring connected to V_{ss} to minimize substrate debiasing. The same is done for the NMOS transistors, with an N-type guard ring connected to V_{dd}. Finally, the transistor gates are all connected to a connection on the left edge of each row, to provide a convenient contact place for the gate drivers.
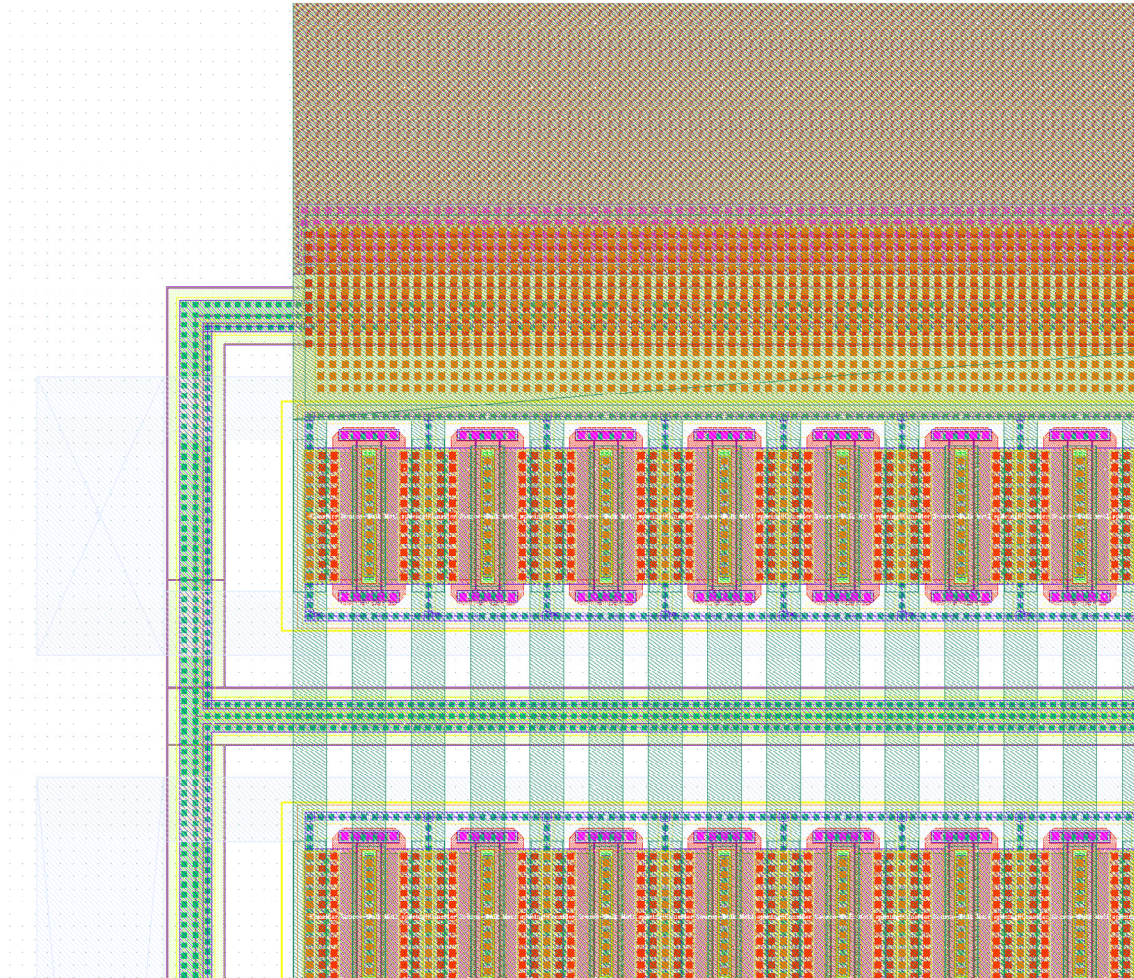
Figure 8-3: Detail of PMOS layout

8.2 Drivers

The drivers are actually the same circuit as the output transistors, and will be laid out in a similar way. After that, they will be connected as shown in Figure 8-1.

The drivers also include the level shifters that convert between 3.3 V and 12 V (section 4.8). These are combined with the circuits that provide the non-overlapping clock function. The exact location of these circuits relative to the location of the actual drivers will be chosen during the layout of the complete die (section 8.7).

8.3 Matched transistors

One of the main issues in making the layout of the support circuits is matching of transistors. All transistors that need to be closely matched are placed closely together, in a common-centroid layout if possible.

Differential pairs are usually the critical components in a matching chain. The transistors are divided into four identical segments, arranged in an ABBA-BAAB structure with abutted source and drain regions. Dummy transistors at the end of each row ensure that each transistor looks identical from each side, minimizing mismatch by underetching of polysilicon. Finally, a guard ring is placed around the entire area to minimize the influence of substrate noise.
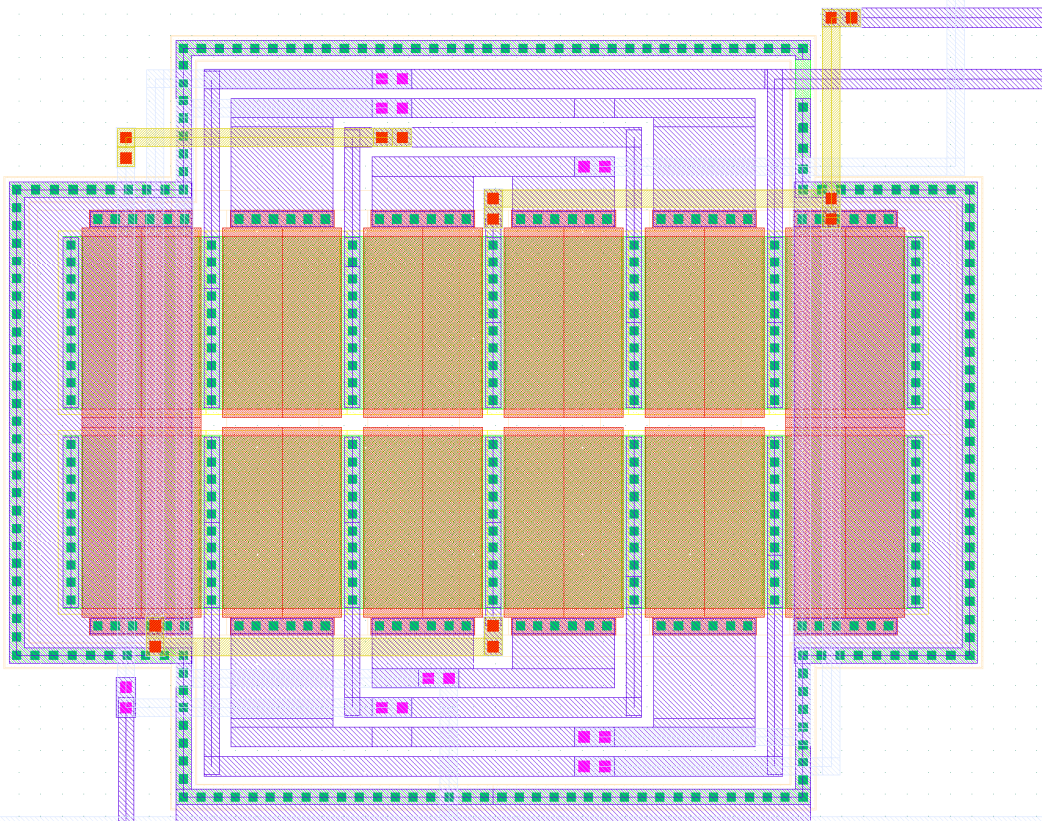
Figure 8-4: Layout of a PMOS differential pair

Current mirrors usually also need to be matched accurately. The layout is done in a slightly different way however, because these transistors are generally long and thin. In order to end up with a roughly square layout, the transistors are aligned as shown in Figure 8-5. Again, the transistors are arranged in a common-centroid layout and surrounded by a guard ring.
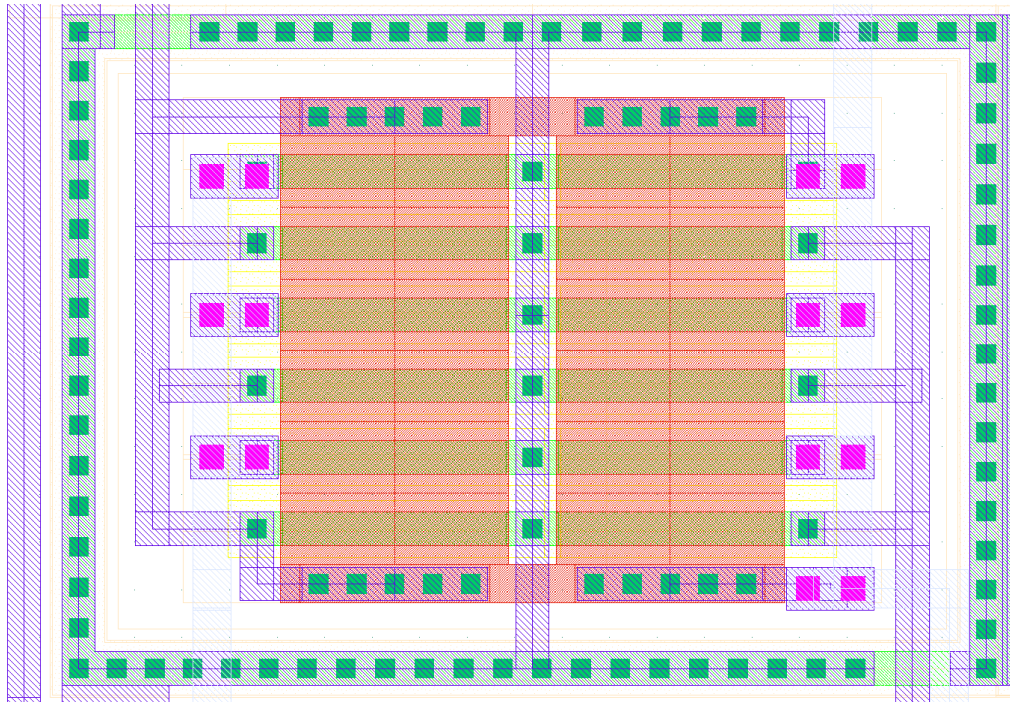
Figure 8-5: Example layout of a PMOS current mirror

8.4 Bandgap reference

The layout of the bandgap reference is focused on obtaining accurate matching. The most important components to be matched are the bipolar transistors, because they determine the temperature stability and the exact value of the output voltage. The design kit only contains one standard version of the vertical PNP transistor that cannot be changed, so the only freedom is in the placement of the transistors. It has already been decided in the circuit design phase to use a one-to-eight ratio, so that the transistors can be laid out in a square, with the single transistor in the middle. This should cancel out gradients of temperature, diffusion depths et cetera [6].
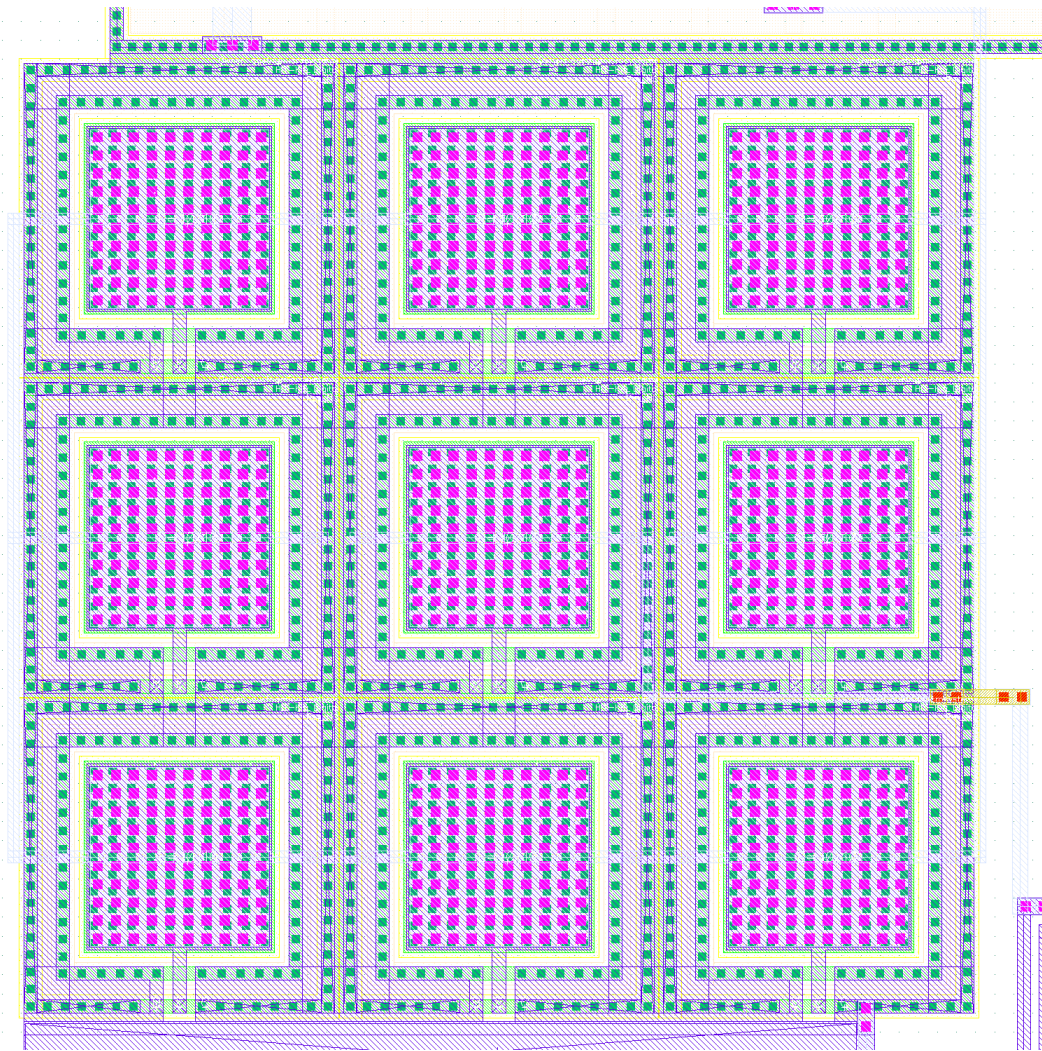
Figure 8-6: Layout of bipolar transistors

Another component that requires accurate matching is the resistive divider. First, the resistors are divided into a number of identical unit resistors. Since the ratio of the resistor values is $575 \text{ k}\Omega / 54 \text{ k}\Omega \approx 43/4$, unit resistors of 13.5 kΩ can be used to build the resistors from 43 and 4 units, respectively. The 43 resistors are arranged in two rows, with the four remaining resistors divided evenly and symmetrically among them. Dummy resistors are added to the end of each row, to reduce the effects of overetching and underetching of the polysilicon.
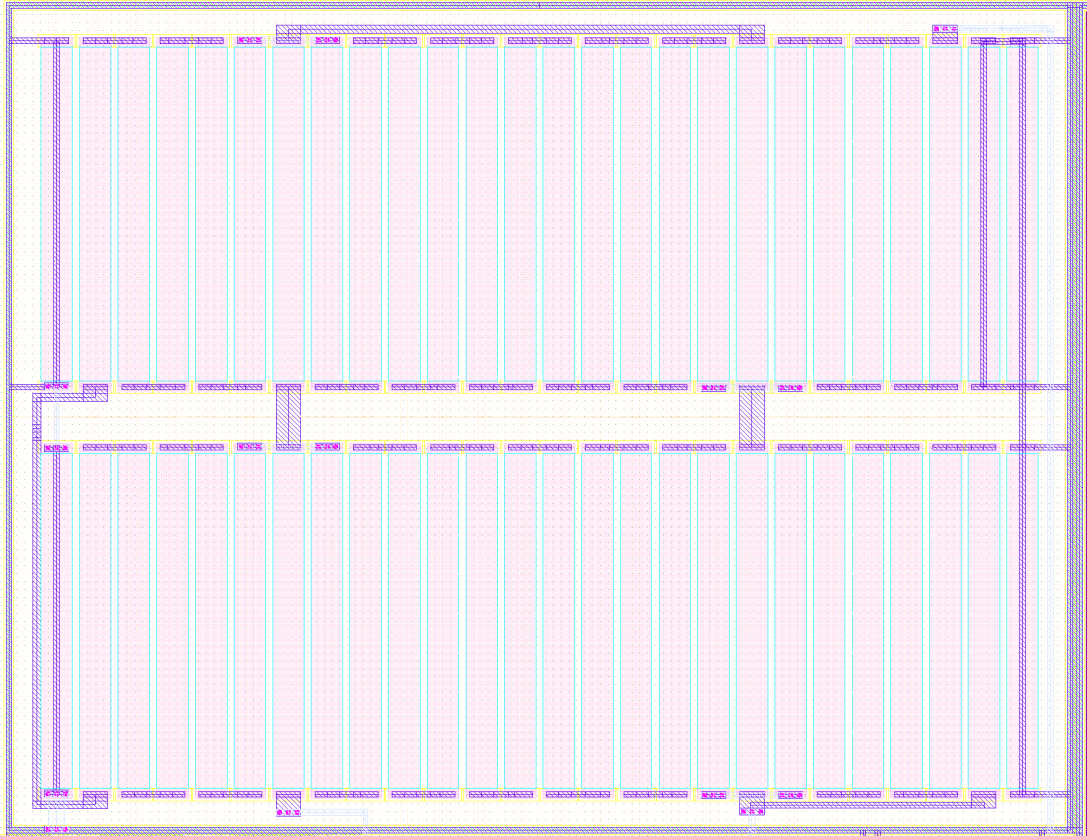
Figure 8-7: Layout of bandgap reference resistor

The resistor in the start-up circuit does not need to be accurate, so it is made from a strip of minimum width polysilicon in a snake structure to save space.

8.5 Current reference

The current reference contains a differential pair, which is laid out as described above, and a large current mirror that requires accurate matching. The transistors of this current mirror are divided into four identical parts each, for a total of 40 identical transistors. These are interdigitated in a pattern like ABCD-DCBA-ABCD-DCBA. The transistor used for frequency compensation and testing (MP86 in Figure 6-31) is placed completely on the outside, because it does not need to be matched very accurately. This also removes the need for a dummy transistor.
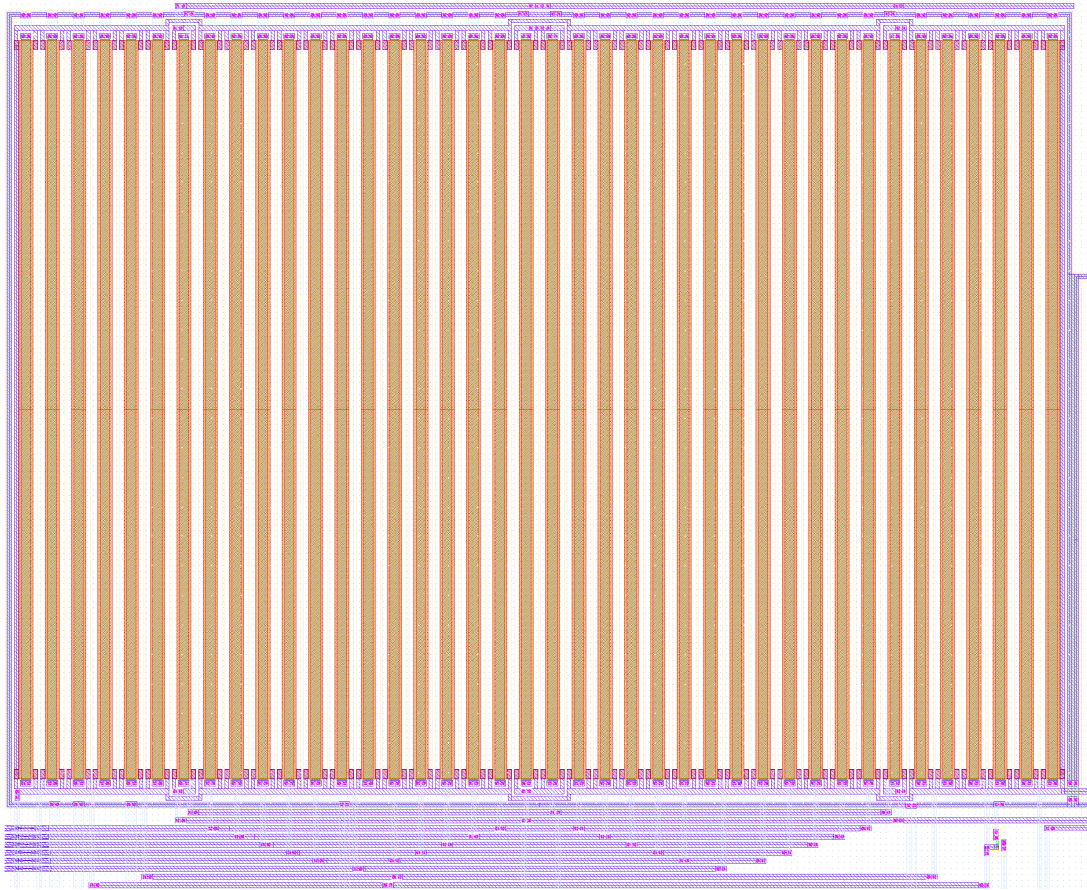
Figure 8-8: Layout of output current mirror

The reference resistor is a rather large resistor that needs to be matched with other resistors, so it is built up of several identical segments. Choosing the unit resistors to have a resistance of 40 kΩ, which corresponds to a width and length of 5 μm × 160 μm, will cause the total resistor to be approximately square. Any resistors in other sub-circuits that need to match with this resistor will be constructed out of the same unit resistors, oriented in the same direction.

8.6 Other support circuits

The comparator contains a differential pair and three current mirrors which require accurate matching. They are laid out as described above.

The triangle generator contains a resistive divider that has to match with the current reference resistor. It is therefore made up of the same unit resistors. There are also several current mirrors that need matching.

Another resistor that needs to match the reference resistor is located in the dynamic output stage. The same procedure is followed as in the triangle generator.

8.7 Complete circuit

Now that all sub-circuits have been laid out, it is time to connect everything together. The drivers are connected to the gates of the output transistors, and the level shifters are located near the low-voltage circuits.

A large guard bar is inserted to shield the precision circuits from the power circuits. This bar is built up from metallized diffusion layers in a P-N-P structure, to provide a large depletion region between the two areas which can collect any stray currents coming from the power devices.
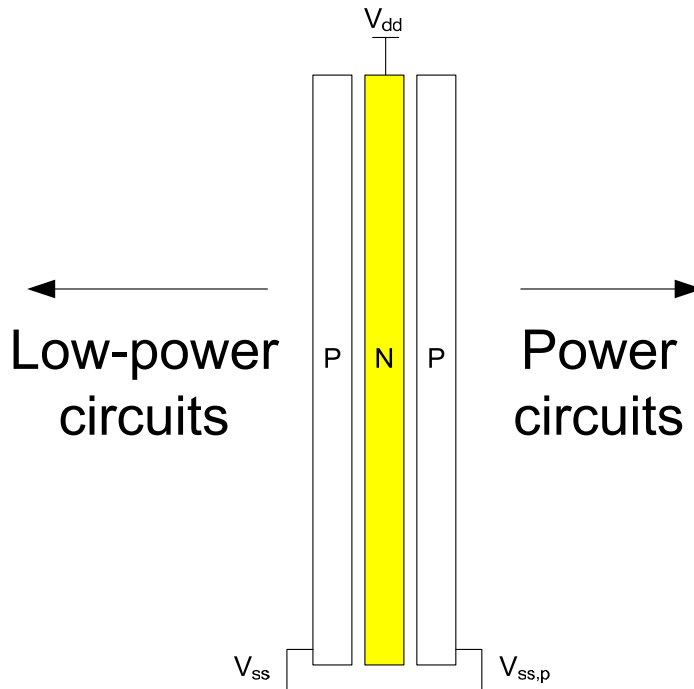
Figure 8-9: Guard bar between power domains

The circuit is therefore divided into two power domains. The low power circuits are powered by $V_{dd,3}$ and V_{ss}, while the high power circuit are powered from $V_{dd,3,p}$, $V_{dd,12,p}$ and $V_{ss,p}$. For the level shifters, another pin called $V_{dd,12}$ is added, to prevent the level shifters from receiving too much noise from the 12 V power supply. Only outside the chip can supplies with the same voltage from different domains be connected together. The PCB layout must also ensure that supply noise is minimized.

To make sure that the two grounds cannot "float away" from each other and cause large currents to flow through the substrate, they are connected by two anti-parallel diodes. This ensures that the voltage difference between the two ground nodes is never more than one diode drop. These diodes can conveniently be placed next to the guard bar.
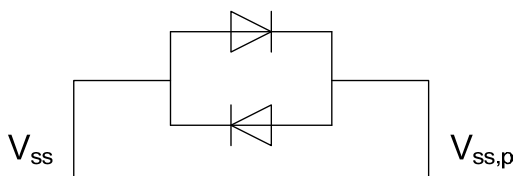
Figure 8-10: Connection between normal V_{ss} and power V_{ss}

The support circuits are now placed in the remaining area. As stated above, all matched resistors are placed in the same direction, and are located as closely together as possible.
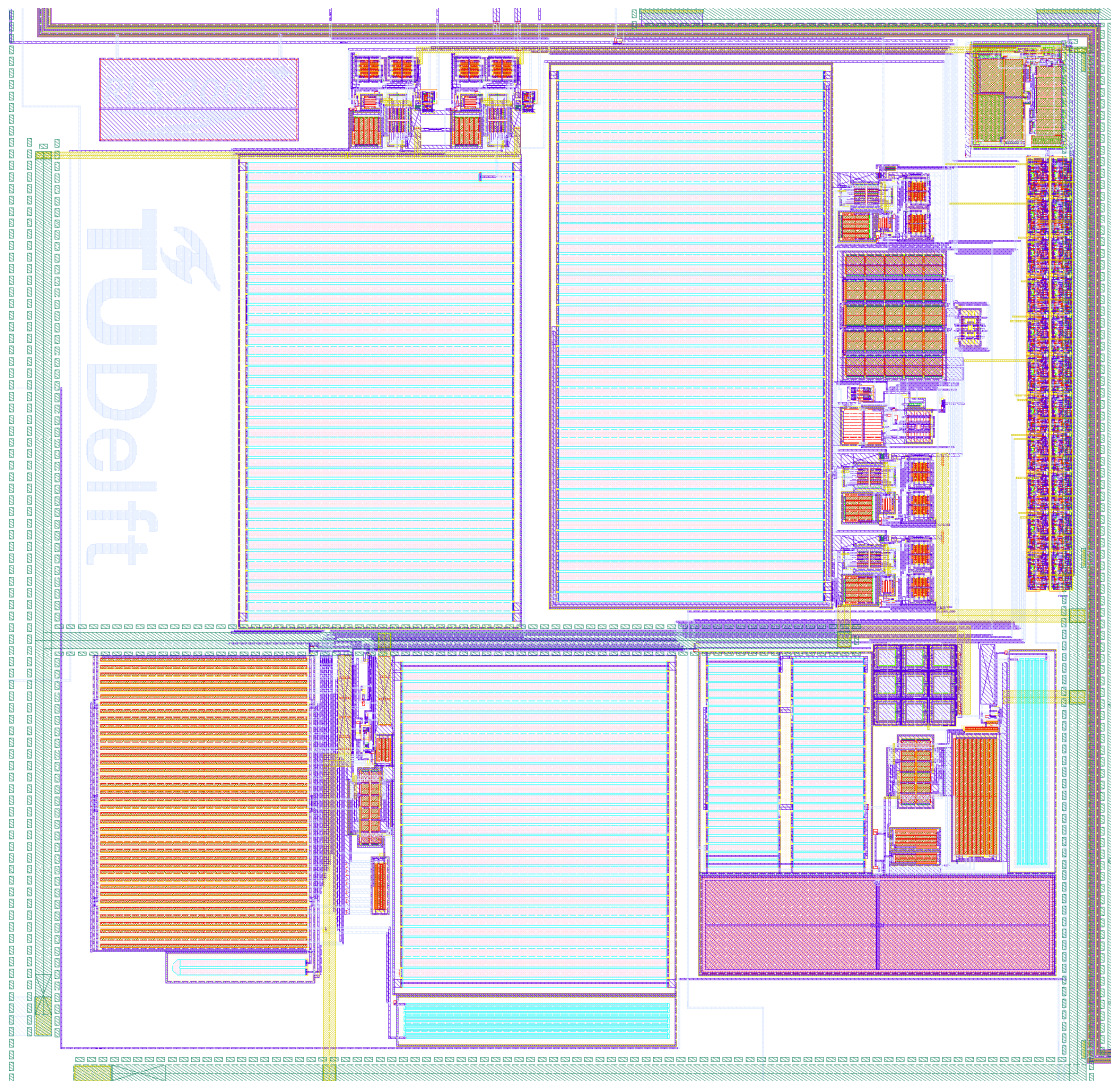
Figure 8-11: Layout of all support circuits together

Several different bondpads with ESD protection circuits are available in the Austriamicrosystems library. These also include spacers and connectors to build a complete pad ring. Since there are two power domains on this IC, the pad ring should also be divided into two sections.

The standard 3.3 V I/O and power pads are used for the low power circuits, while the 20 V pads are used for the high-voltage circuits. The guard bar is extended to separate the two ring sections from each other. The 3.3 V power pads include supply clamps, but the 20 V pads need them to be added manually. The location of the supply clamps can be seen in the complete IC layout (Appendix E).

8.8 Layout verification

Finally, the complete layout is checked for design rule violations using the Calibre toolset. Two examples of modifications to the layout, necessary to comply with the design rules, are shown below.
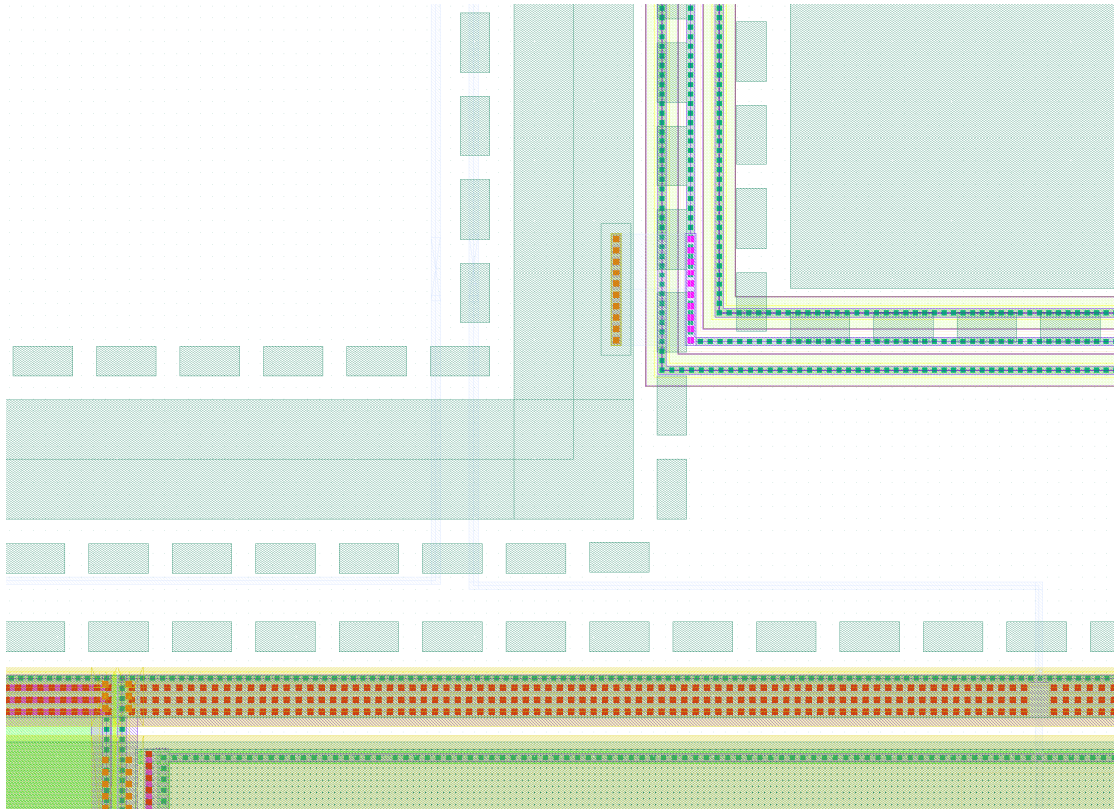
Figure 8-12: Dummy metal structures

Dummy blocks are added around the edges of the top metal layer to protect against assembly stress. The design rules stipulate that these dummies should be 2.5 μm × 5 μm wide and spaced 2 μm from the edge of the metal they need to protect.
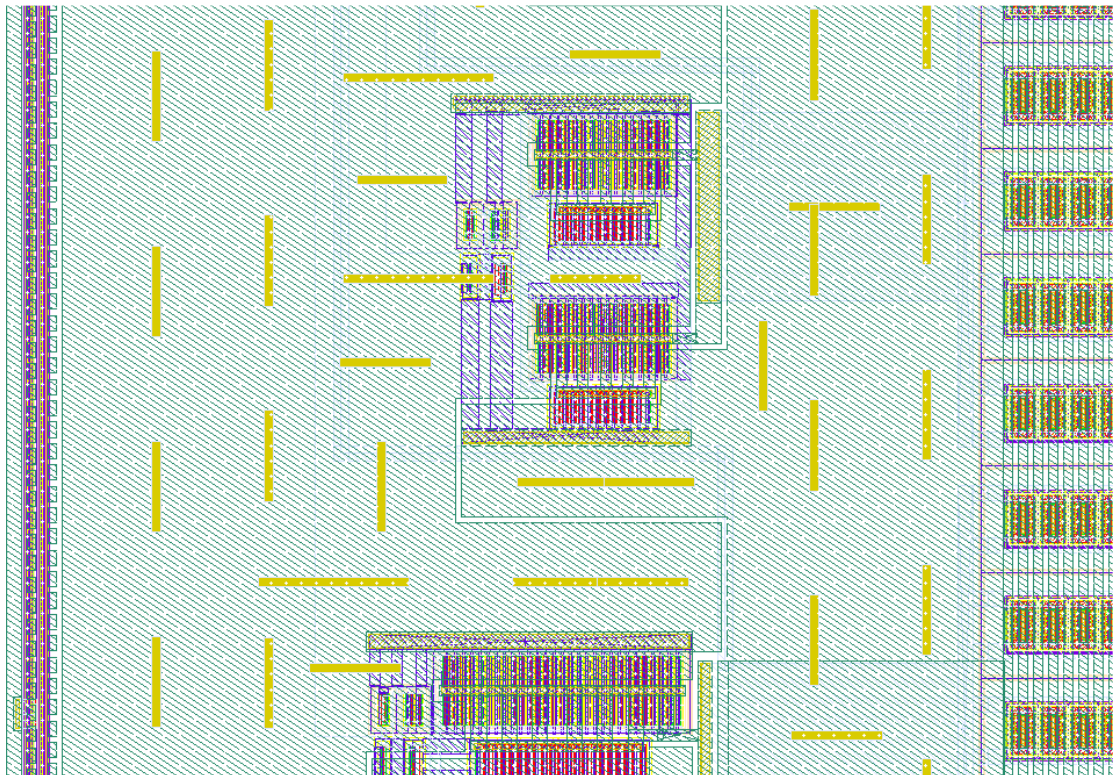
Figure 8-13: Stress relieve slots

Metal structures that are more than 35 μm wide need to be perforated with slots to release stress. A convenient "slot" layer is available for each of the metal layers that can be used to draw these slots and move them around as if they were actual structures. The slots are always placed in the same direction as the intended current flow, so that the influence on the wire's resistance is minimal.

Finally, a layout versus schematic (LVS) check is performed to verify whether the layout corresponds to the circuit that it represents. One problem with the layout is that there are two different ground nets (V_{ss} and $V_{ss,p}$) which are both connected to the substrate. The LVS tool is therefore unable to distinguish between the two ground nets, and gives several warnings. Upon close inspection, this turns out to be the only "discrepancy" between the layout and the schematic, and the layout can therefore be considered correct.

9. Conclusions and recommendations

A class D amplifier circuit has been designed that, according to simulations, is able to reach the specifications set in chapter 1. It reaches 97.5 % overall efficiency for a sine-wave input signal, for an output voltage up to 11 V. Harmonic distortion levels are below -45 dBc, while third-order intermodulation products are at -37.4 dB. Only three external components are required.

The circuit not only reaches the specifications under standard conditions, but also across all corner cases caused by manufacturing variations and environmental factors such as temperature and supply voltage variations. This is a definite requirement for any IC that will be produced and used in practice.

A structured and hierarchical design method was followed that provides a clear way of transforming a set of requirements into a circuit that can fulfil these requirements. Furthermore, an adaptive output stage has been implemented that increases the power efficiency from 96.8 % to 97.5 %.

Finally, the circuit has been laid out and checked to conform to manufacturing requirements. The circuit is now ready for production and testing. Additional circuits have been implemented to facilitate testing and troubleshooting.

The power efficiency could be further improved by using an NMOS transistor for the high-side output transistor (section 4.3). Care must be taken to ensure that the required gate drive circuits do not consume more power than is gained by using the NMOS transistor. If still more efficiency is required, then it will be necessary to reimplement the circuit in a different technology that provides a lower on-resistance per amount of gate charge.

If even better distortion performance is required, then the modulator can be improved or feedback can be added. Advanced modulation techniques such as sigma-delta modulation or digital PWM synthesis can all be used, and they can even be tested on the current circuit by using its straight-PWM input. If negative feedback is to be used, it may be necessary to increase the gain of the amplifier, i.e. by adding an additional gain stage at the input.

10. References

[1]   F. Stelwagen, *A Highly Efficient Space Qualified Class E Power Amplifier for μ-satellite Delfi-C³*, M.Sc. thesis, Delft University of Technology, 2007

[2]   H. Lan, *The Linearization and Power Control of Delfi-C3 Transmitter*, M.Sc. thesis, Delft University of Technology, 2008

[3]   C.J.M. Verhoeven, A. Van Staveren, G.L.E. Monna, M.H.L. Kouwenhoven, and E. Yildiz, *Structured Electronic Design: Negative-Feedback Amplifiers*. Boston: Kluwer Academic Publishers, 2003

[4]   N.O. Sokal and A.D. Sokal, "Class E – A new class of high-efficiency tuned single-ended switching power amplifiers", *IEEE Journal of Solid-State Circuits*, vol. SC-10, no. 3, pp. 168-176, June 1975.

[5]   A. van Staveren, "Integrable DC sources and references", in *Analog IC Techniques for Low-Voltage Low-Power Electronics*, W.A. Serdijn, C.J.M. Verhoeven, A.H.M. van Roermund, Eds. Delft: Delft University Press, 1995, pp. 120-141.

[6]   A. Hastings, *The Art of Analog Layout*, 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.

[7]   J.S. Chang, M. Tan, Z, Cheng, and Y Tong, "Analysis and design of power efficient class D amplifier output stages", *IEEE Transactions on Circuits and Systems*, vol. 47, no. 6, pp 897-902, June 2000

[8]   M.J.M. Pelgrom, A.C.J. Duinmaijer, and A.P.G. Welbers, "Matching properties of MOS transistors", *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1440, October 1989.

[9]   N. Shimizu, "Class-D amplifier circuit", U.S. Patent Application Publication no. US 2009/0066412 A1, March 12, 2009.

[10] Austriamicrosystems AG, *0.35 μm 50V CMOS Process Parameters*, rev. 5.0, 2007.

[11] Austriamicrosystems AG, *0.35 μm 20V CMOS Process Parameters*, rev. 4.0, 2008.

[12] Austriamicrosystems AG, *0.35 μm 50V CMOS Design Rules*, rev. 6.0, 2007.

[13] Austriamicrosystems AG, *0.35 μm 20V CMOS Module Design Rules*, rev. 4.0, 2008

[14] Austriamicrosystems AG, *0.35 μm C35 Matching Parameters*, rev. 2.0, 2006

[15] Austriamicrosystems AG, *H35 Layout and Design Guidelines*, rev. 1.0, 2006

[16] S. Sánchez Moreno, "Class D Audio Amplifiers – Theory and Design", *Elliott Sound Products*, June 2005. [Online]. Available: http://sound.westhost.com/articles/pwm.htm [Accessed: June, 2010].

[17] Elec. Pkg. Char. Group, "3D Bondwire Electrical Modeling Results", WPI Analog Lab Resources Page, 2001. [Online]. Available: http://ece.wpi.edu/analog/resources/1mil_bwire_RLC.pdf [Accessed: June, 2010].

Appendix A Hierarchy of switch-mode power converters

This section attempts to provide a structured, hierarchical classification of switch-mode power converter topologies. It appears that all possible switch-mode power converters can be constructed from a limited set of basic building blocks.

The design of switch-mode power converters is often divided among two different disciplines. Power engineers usually design "DC/DC converters", while signal processing engineers design "Switch-mode amplifiers". In fact, they are often referring to the same circuits: what a power engineer might call a buck converter, is called a class D amplifier by an audio engineer. The only difference is that a buck converter is optimized to deliver a constant voltage to a (usually) varying load, while a class D amplifier is optimized to deliver a varying voltage to a (usually) constant load.

It should be noted that the signal frequency and possibly the load impedance are the *only* differences between a buck converter and a class D amplifier. The design equations are identical, and both circuits can benefit from negative feedback.

This also means that all circuits described in this section can be used equally as DC power supplies, or as AC signal amplifiers. Some circuits (such as the boost converter) do have the disadvantage of being nonlinear, which makes amplifier design more difficult. This effectively means that these circuits *require* negative feedback, while for the linear circuits (such as the buck converter) it can be optional.

In all circuits shown below, the switches are assumed to be driven with an information-carrying signal such as a PWM signal. The switches should always be driven in a complementary way: when one is on, the other is off, and vice versa.

A switch-mode topology consists of three basic ingredients: a switching stage, an input stage, and an output stage.

A.1 Switching stages

The simplest switching stage consists of two switches that alternately pass and block the input voltage or current.
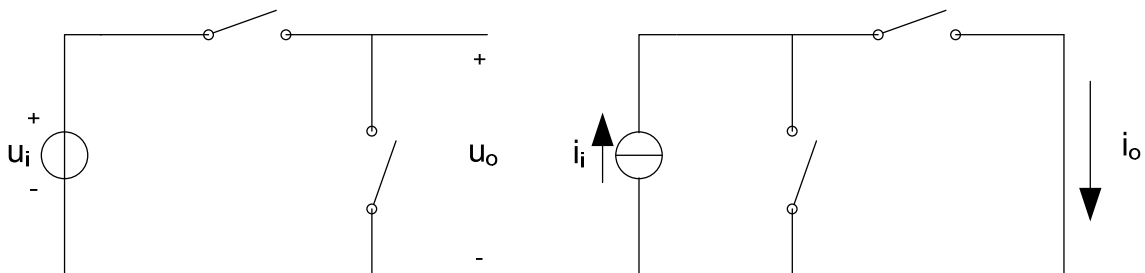
Figure 10-1: Basic voltage-to-voltage and current-to-current switching stages

These are building blocks for linear circuits: $u_o(t) = P(t) \cdot u_i$, and $i_o(t) = P(t) \cdot i_i$, where $P(t)$ is the modulation signal. The output quantity has the same sign as the input quantity. If linearity is a priority, and the output signal needs to be smaller in amplitude than the input signal, then these stages are the best choice.

A more complicated switching stage also contains a reactive element that converts current to voltage or voltage to current. These necessarily change the direction of the voltage or current.
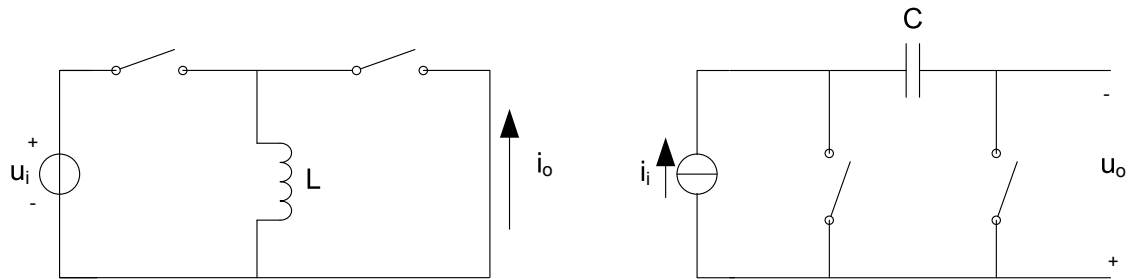
Figure 10-2: Single-ended voltage-to-current and current-to-voltage switching stages

These circuits are nonlinear, since the magnitude of the output quantity depends more or less quadratically on the duty cycle. Although these simple versions perform a sign inversion, this can be avoided by constructing a balanced version in which the output quantity can be used in either direction.
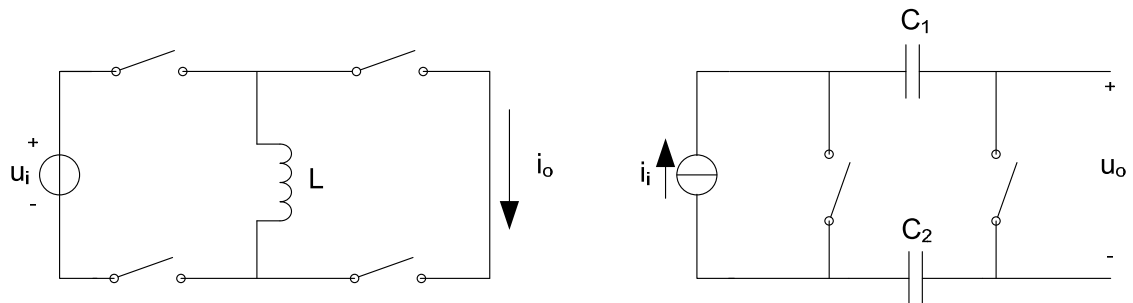
Figure 10-3: Balanced voltage-to-current and current-to-voltage switching stages

Note that these are actually more general than the single-ended circuits shown above. If the bottom terminals at the input and output are grounded, the circuits become equal to the single-ended versions.

In the U-I stage, the switches are synchronized with their neighbours above or below (in other words, the left switches always move together, as do the right switches). The transfer functions are identical to the single-ended stages.

If the direction of the input quantity needs to be reversed, and/or if the output quantity has to be both smaller and larger than the input quantity, then these stages should be used. They do have the disadvantage of being nonlinear, so in practice they require negative feedback.

A.2 Input stages

Two different types of input stages are available. The first type is a normal voltage or current source, as shown in the examples above. The second type consists of a source with a reactive component. This can be used to change the effective source type.
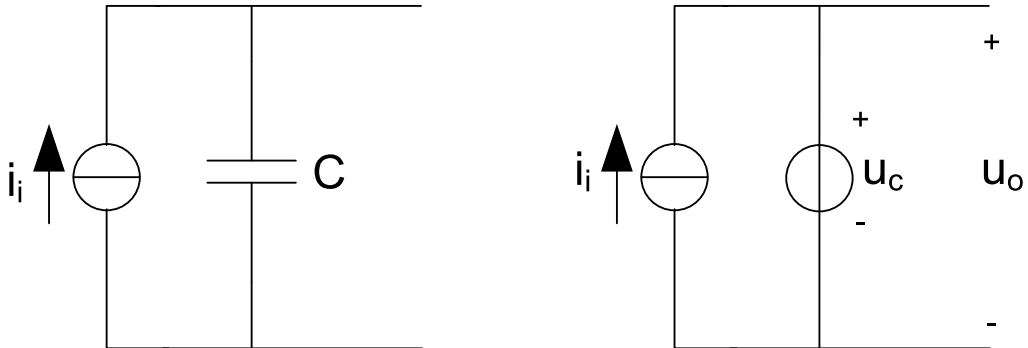
Figure 10-4: Reactive current-to-voltage source and its equivalent schematic
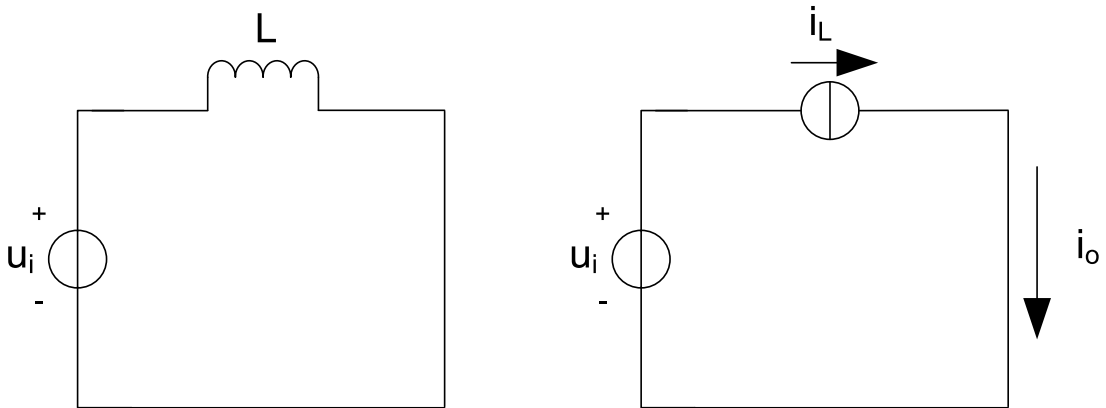
Figure 10-5: Reactive voltage-to-current source and its equivalent schematic

As shown in Figure 10-4, the current source will charge the capacitor C to a certain voltage during the off- phase of the switches. During the on-phase, the charged capacitor will act as a voltage source in parallel with the current source. This means that a current source connected in parallel with a capacitor can be used as if it were a voltage source (the value of which depends on the duty cycle), although with a current offset equal to the value of the current source. This also means that a power converter with this input stage has a non-linear transfer. A similar reasoning can be used for a voltage source with a series inductor, as shown in Figure 10-5.

A.3 Output stages

There are four basic output stages, divided according to their input and output quantities.
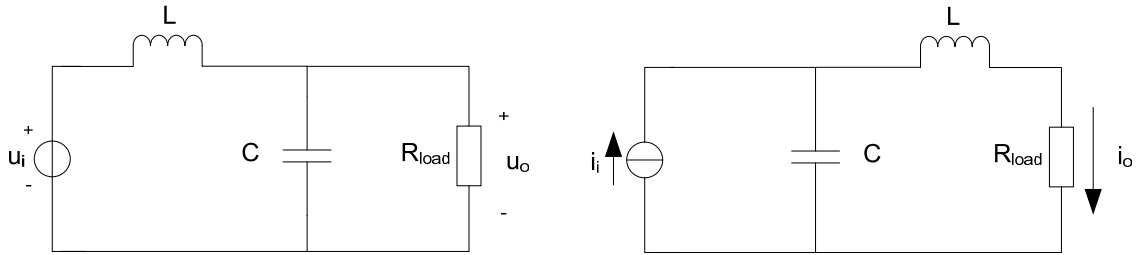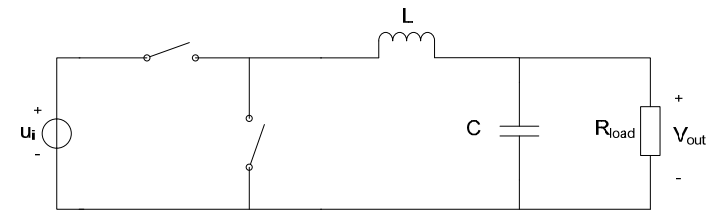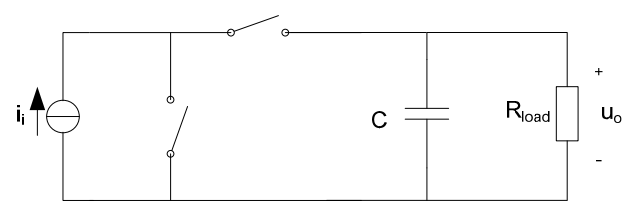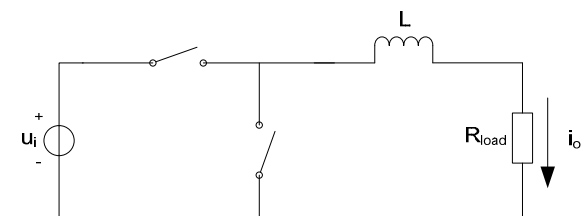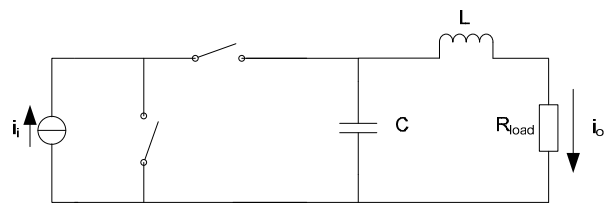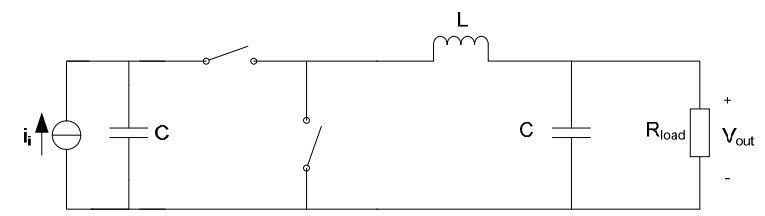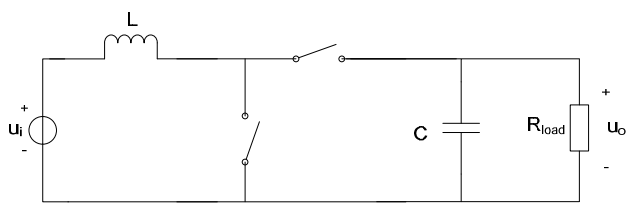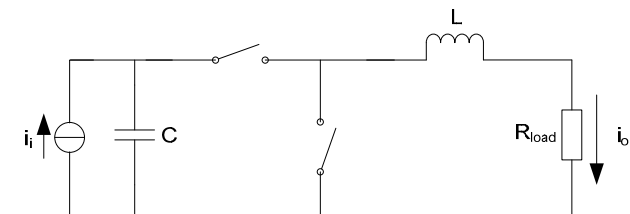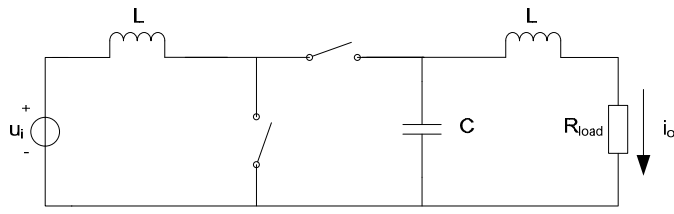
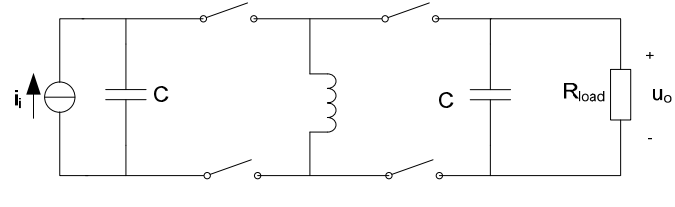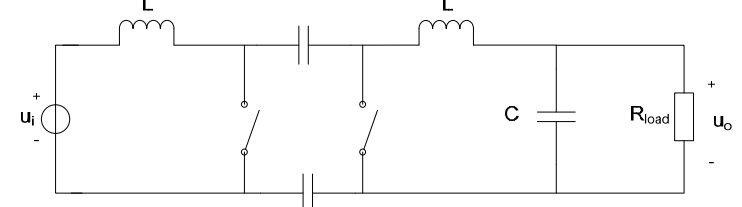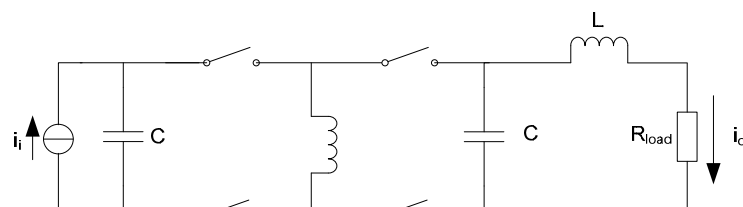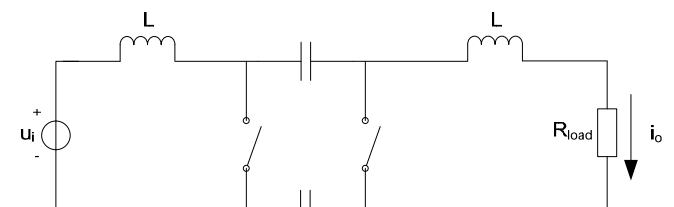**Figure 10-6: Voltage-to-voltage and current-to-current output filters**
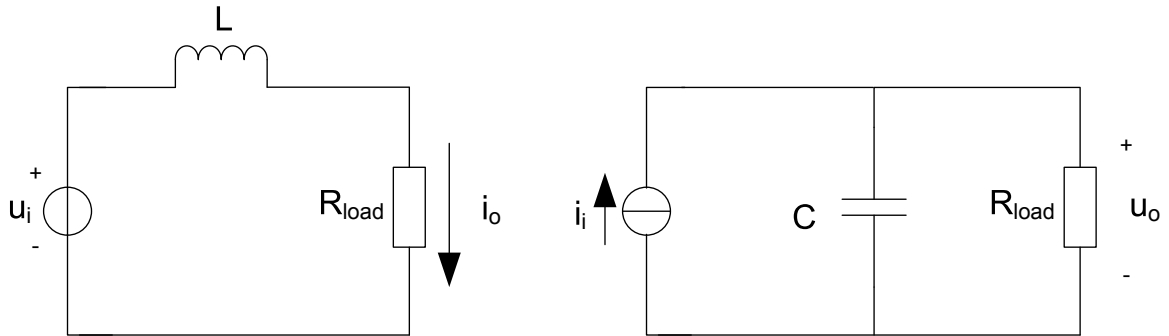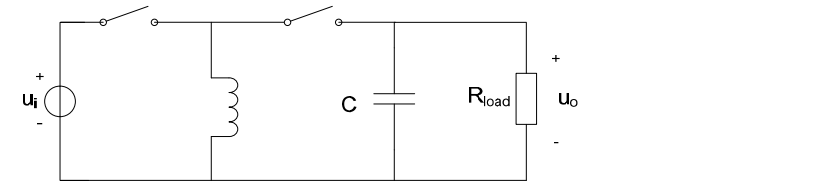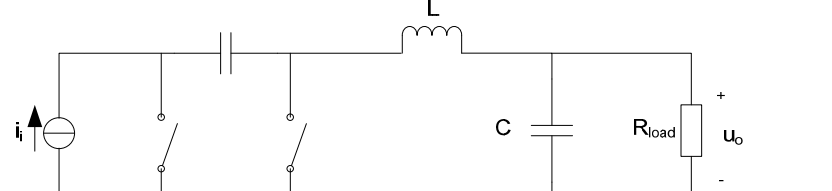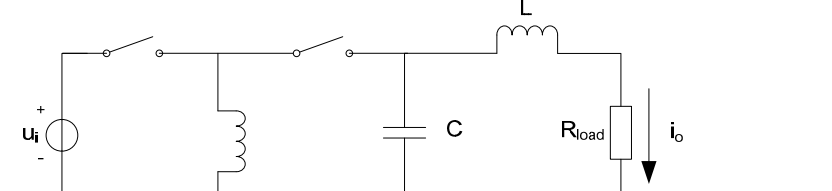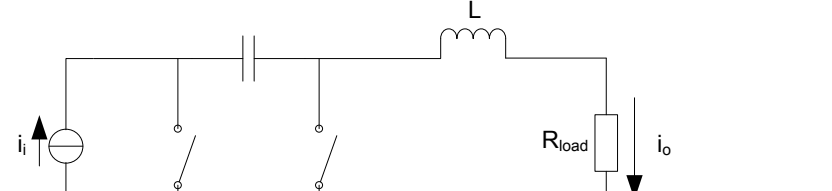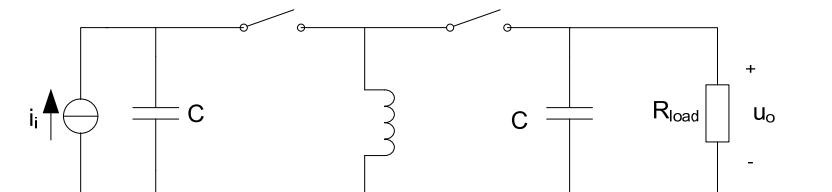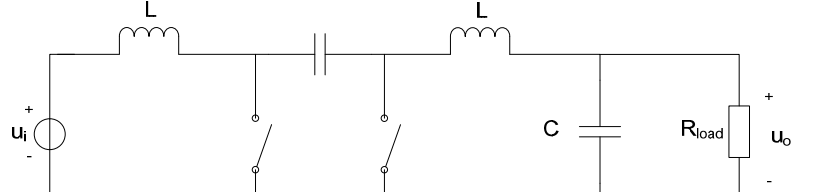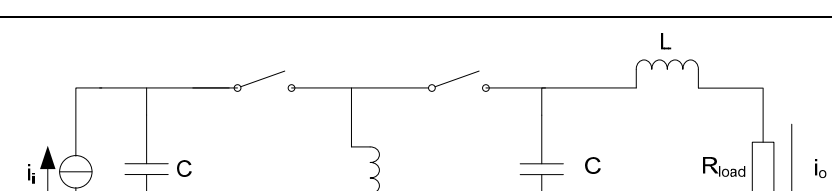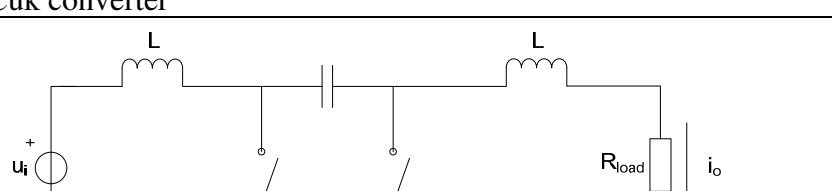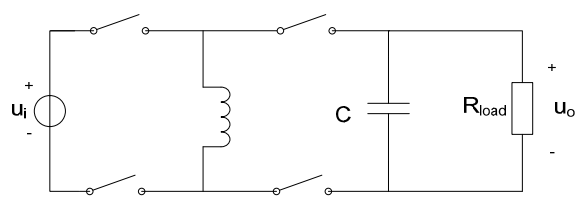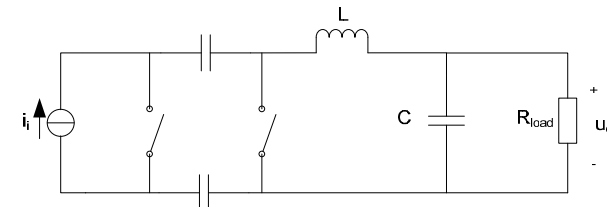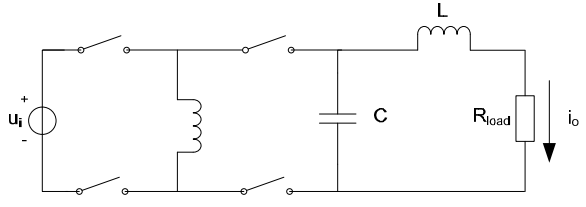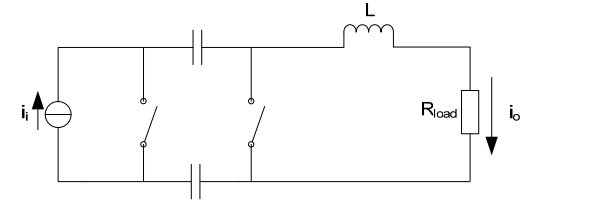
Figure 10-7: Voltage-to-current and current-to-voltage output filters

A disadvantage of the U-I and I-U demodulators (Figure 10-7) is that they only consist of a first-order filter, and therefore provide 20 dB less attenuation of the switching frequency compared to the U-U or I-I demodulators (Figure 10-6). Of course, all of these stages can be extended with additional filter sections to provide additional filtering.

Using these fundamental building blocks, all possible switch-mode power converters can be built in a logical way. The following pages show all possible configurations. When a circuit has a commonly used name, this has been shown in the table as well.

	U-U switch	I-I switch	
Simple source	L u_i +/- C R_load V_out +/- Class D amplifier / Buck converter	i_i C R_load u_o +/-	Voltage output
Simple source	L u_i +/- R_load i_o Current output buck converter	L i_i C R_load i_o	Current output
Reactive source	L i_i C C R_load V_out +/-	L u_i +/- C R_load u_o +/- Boost converter	Voltage output
Reactive source	L i_i C R_load i_o	L L u_i +/- C R_load i_o	Current output

	U-I switch	I-U switch	
Simple source	Buck-boost converter / Inverting converter		**Voltage output**
			Current output
Reactive source		Ćuk converter	**Voltage output**
			Current output

	U-I switch	I-U switch	
Simple source			**Voltage output**
			Current output
Reactive source		SEPIC converter	**Voltage output**
			Current output

Appendix B   Matching calculations

The matching of transistor parameters between two identically designed devices is described by the Pelgrom model [8]. This states that the variance of a parameter P consists of a term inversely proportional to the area of the devices and a term proportional to the square of the distance between them. Mathematically,

$$\sigma^2(\Delta P) = \frac{A_P^2}{WL} + S_P^2 D_x^2, \tag{2}$$

In which A_P is an empirically determined constant specific to the parameter in question, WL is the area of the device, S_P is an empirically determined constant that describes the mismatch as a function of the distance between two devices, and D_x is the distance between the devices. For closely spaced devices this simplifies to

$$\sigma^2(\Delta P) = \frac{A_P^2}{WL}, \tag{3}$$

which is used in the Austriamicrosystems design manual [14], i.e. the parameter S_P is not given and devices to be matched are expected to be located closely together. For MOS devices, the constants A_P are given for the threshold voltage V_{th} and for the current gain factor k. These values are shown in Table B-1.

In designing a circuit however, these values are not directly useable. As shown in [6], circuits can be optimized for either precisely matched gate-source voltages, or precisely matched drain currents.

The difference in gate-source voltage between two transistors is given by

$$\Delta V_{gs} = \Delta V_{th} - \frac{1}{2} V_{gt} \left(\frac{\Delta k}{k} \right) \tag{4}$$

If the variances of V_{th} and k are known, then it is possible to find the variance of ΔV_{gs} from

$$\sigma^2(\Delta V_{gs}) = \sigma^2(\Delta V_{th}) + \frac{1}{4} V_{gt}^2 \cdot \sigma^2 \left(\frac{\Delta k}{k} \right) \tag{5}$$

Because both $\sigma^2(\Delta V_{th})$ and $\sigma^2 \left(\frac{\Delta k}{k} \right)$ are of the form $\frac{A_P^2}{WL}$, also $\sigma^2(\Delta V_{gs})$ can be written in this form. A constant $A_{V_{gs}}$ can now be defined, equal to

$$A_{V_{gs}}^2 = A_{V_{th}}^2 + \frac{1}{4} V_{gt}^2 \cdot A_k^2 \tag{6}$$

Similarly, if the drain currents are to be matched, then the equation

$$\frac{I_{D1}}{I_{D2}} = \frac{k_2}{k_1}\left(1 + \frac{2\Delta V_{th}}{V_{gt}}\right) \tag{7}$$

can be written as

$$\frac{\Delta I_D}{I_{D1}} = \left(\frac{\Delta k}{k_1} + 1\right)\left(1 + \frac{2\Delta V_{th}}{V_{gt}}\right) - 1 = \frac{\Delta k}{k_1} + \left(\frac{\Delta k}{k_1}\frac{2\Delta V_{th}}{V_{gt}}\right) + \frac{2\Delta V_{th}}{V_{gt}} \tag{8}$$

so that

$$\sigma^2\left(\frac{\Delta I_D}{I_{D1}}\right) = \sigma^2\left(\frac{\Delta k}{k_1}\right) + \sigma^2\left(\frac{\Delta k}{k_1}\frac{2\Delta V_{th}}{V_{gt}}\right) + \frac{4}{V_{gt}^2}\sigma^2\left(\Delta V_{th}\right) \tag{9}$$

However, in this equation there is a term that cannot directly be derived from the two known parameters. If this term is assumed to be negligible, then an expression for A_{I_D} can be derived that looks very similar to (6).

$$A_{I_D}^2 = A_k^2 + \frac{4}{V_{gt}^2} \cdot A_{V_{th}}^2 \tag{10}$$

Comparing (6) and (10), it is found that for good voltage matching V_{gt} should be small, while for good current matching V_{gt} should be large. From (3) it is clear that in either case the gate area should be large for good matching (and that the transistors should be close together).

When designing a differential amplifier, it therefore makes sense to use transistors with a large *W/L* ratio, because these have a lower V_{gt} than transistors with a small *W/L* biased at the same drain current. There is however a practical limit, following from (6), at which reducing V_{gt} provides little additional benefit: if the term containing V_{gt} becomes much smaller than $A_{V_{th}}^2$, then it makes no sense to reduce V_{gt} any further. Table B-1 contains the values of V_{gt} for which the second term contributes 1% to the total.

Conversely, when designing a current mirror, it is better to use transistors with a small *W/L*, to ensure a relatively large V_{gt}. Because the second term in (10) becomes negligible only for very large voltages, it is advisable to design current mirrors with as high a V_{gt} as possible. Table B-1 contains A_{Id} for two values of V_{gt}.

Device	A_{Vth} (mV μm)	A_k (% μm)	A_{Vgs} (mV μm) @ V_{gt} = 0.4 V	A_{Id} (% μm) @ V_{gt} = 0.5 V	A_{Id} (% μm) @ V_{gt} = 1.0 V
NMOS4	9.5	0.7	9.6	3.86	2.02
PMOS4	14.5	1.0	14.6	5.89	3.07

Table B-1: Matching parameters for MOS devices

Appendix C   Maple scripts

C.1 Calculating output stage transistor sizes

restart;
#Set up constants
#Supply voltage
Vsup := 12 :
#Switching frequency
f := 622e3 :
#Inductor resistance
Rcoil := 0.5 :
#On -resistance of MOSFETs
Rdsp := 1141 :
Rdsn := 406.6 :
#Load resistance
Rload := 50 :
#Gate-source capacitance
Cgatep := 26e−15 :
Cgaten := 108.5e−15 :
#Gate-drain capacitance
Cgdp := 6.7e−15 :
Cgdn := 8e−15 :
#Gate charge voltage
Ugatep := 12 :
Ugaten := 3.3 :
#Dissipation of PMOS driver
$$Pdrvp := \frac{2.23\mathrm{e}{-6} \cdot f}{1\mathrm{e}6} :$$

Set up equations
#Resistive losses

$Rdsavg := \dfrac{\delta \cdot Rdsp}{Mp} + \dfrac{(1 - \delta) \cdot Rdsn}{Mn} :$

$Rparavg := Rdsavg + Rcoil :$

$Vout := \dfrac{\delta \cdot Vsup}{1 + \dfrac{Rparavg}{Rload}} :$

$Iload := \dfrac{Vout}{Rload} :$

$Pout := Vout \cdot Iload :$

$Prpar := Iload^2 \cdot Rparavg :$

#Gate charge losses

$Pgate := Cgatep \cdot Ugatep^2 \cdot f \cdot Mp + Cgaten \cdot Ugaten^2 \cdot f \cdot Mn :$

#Gate driver power consumption. Note NMOS driver ignored

$Pdriver := Pdrvp \cdot Cgatep \cdot Mp :$

#Drain charge losses

$Pdrain := (Cgdp \cdot Mp + Cgdn \cdot Mn) \cdot Vsup^2 \cdot f :$

$Pdrainmax := (Cgdp \cdot Mpmax + Cgdn \cdot Mnmax) \cdot Vsup^2 \cdot f :$

#Total power and efficiency

$Pdiss := Prpar + Pgate + Pdriver + Pdrain :$

$Pdissdyn := Prpar + Pgate + Pdriver + Pdrainmax :$

$Eff := \dfrac{Pout}{Pout + Pdiss} :$

$Effdyn := \dfrac{Pout}{Pout + Pdissdyn} :$

#Find optimal transistor sizes for different duty cycles
$with(Optimization) :$
$Maximize(eval(Eff, \delta = 0.1), iterationlimit = 10000);$
$Maximize(eval(Eff, \delta = 0.25), iterationlimit = 10000);$
$Maximize(eval(Eff, \delta = 0.8), iterationlimit = 10000);$

#Plot efficiency as a function of duty cycle
$plot([eval(Eff, [Mn = 378, Mp = 148]), eval(Eff, [Mn = 861, Mp = 586]), eval(Eff, [Mn = 1423, Mp = 3356])], \delta = 0..1);$

#Now to take into account Cgd when using dynamic transistor sizing
$Mnmax := 1423 :$
$Mpmax := 3356 :$
$Maximize(eval(Effdyn, \delta = 0.1), iterationlimit = 10000);$
$Maximize(eval(Effdyn, \delta = 0.25), iterationlimit = 10000);$

#Plot efficiency again
$plot([eval(Effdyn, [Mn = 585, Mp = 184]), eval(Effdyn, [Mn = 1231, Mp = 669]), eval(Eff, [Mn = Mnmax, Mp = Mpmax]),], \delta = 0..1);$

#Plot to determine sensitivity of efficiency to transistor sizes
$plot3d\left(eval\left(\text{Eff}, \delta = 0.8\right), Mn = 1000..2000, Mp = 3000..4000, axes\right.$
$\left.= boxed, orientation = [-62, 70]\right);$

C.2 Calculating transistor sizes for matching
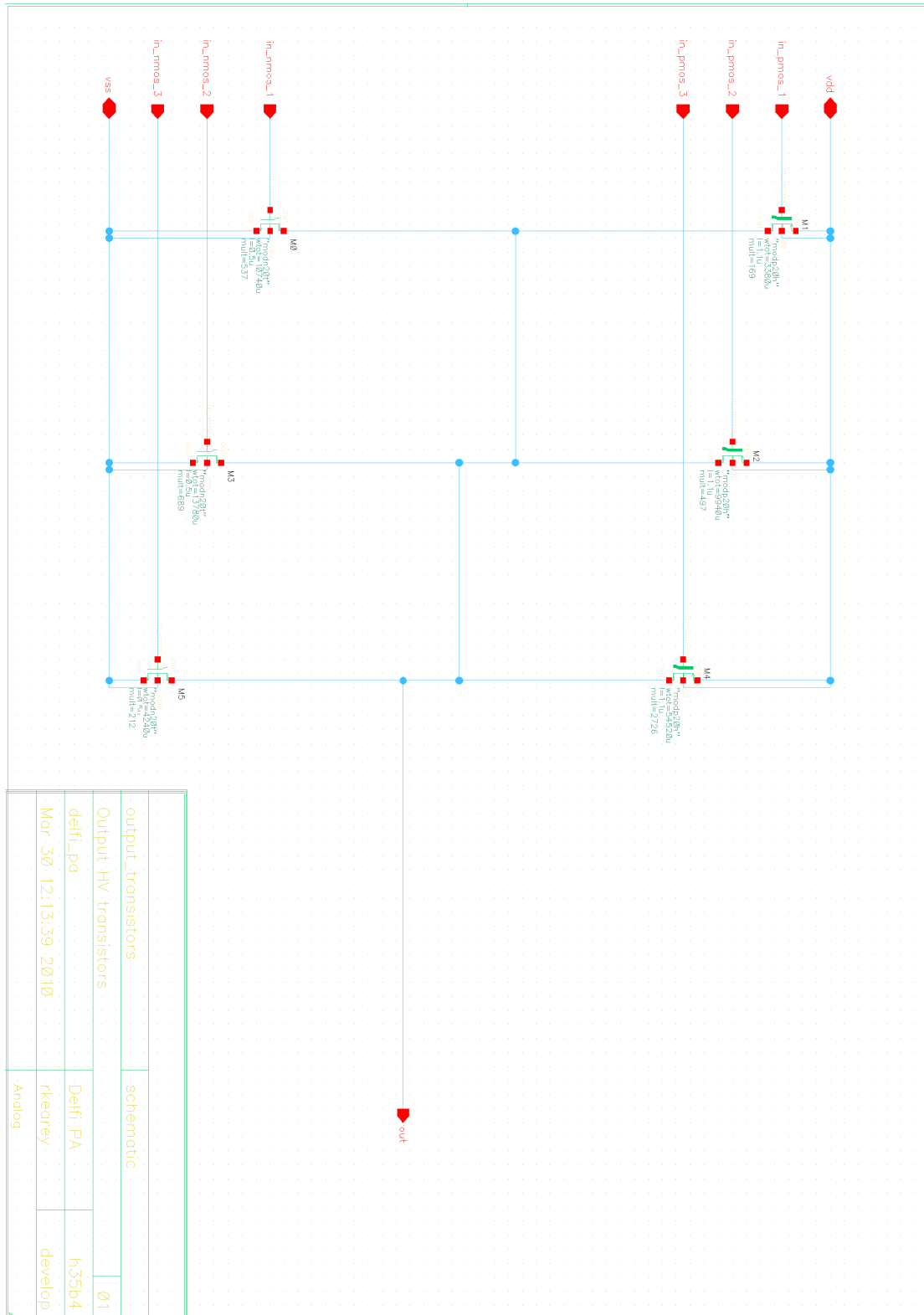
Calculation from bandgap reference shown.

restart;
with(*Optimization*) :
$Minimize\left(A1 + A2 + A3, \left\{\dfrac{210\text{e}-6}{A1} + \dfrac{51\text{e}-6}{A2} + \dfrac{70\text{e}-6}{A3}\right.\right.$
$\left.\left.+ 19\text{e}-9 + 0.11\text{e}-9 = 4\text{e}-6\right\}, assume = nonnegative\right);$

$[226.071136544267176, [A1 = 109.204897903211146, A2$
$= 53.816761469381070, A3 = 63.049477171674965]]$

Appendix D Schematics

D.1 Output transistors

D.2 Drivers
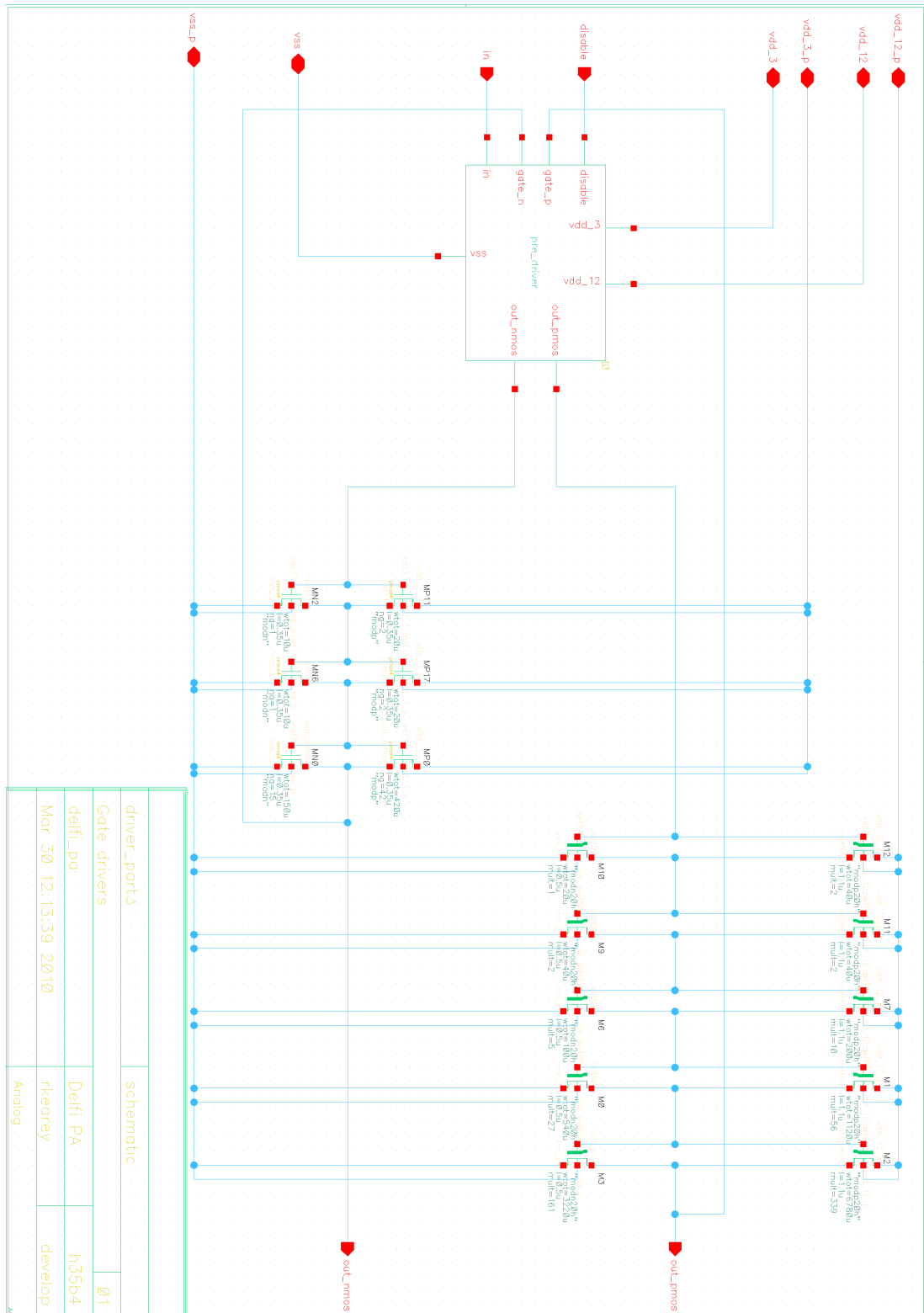
D.3 Non-overlap circuit
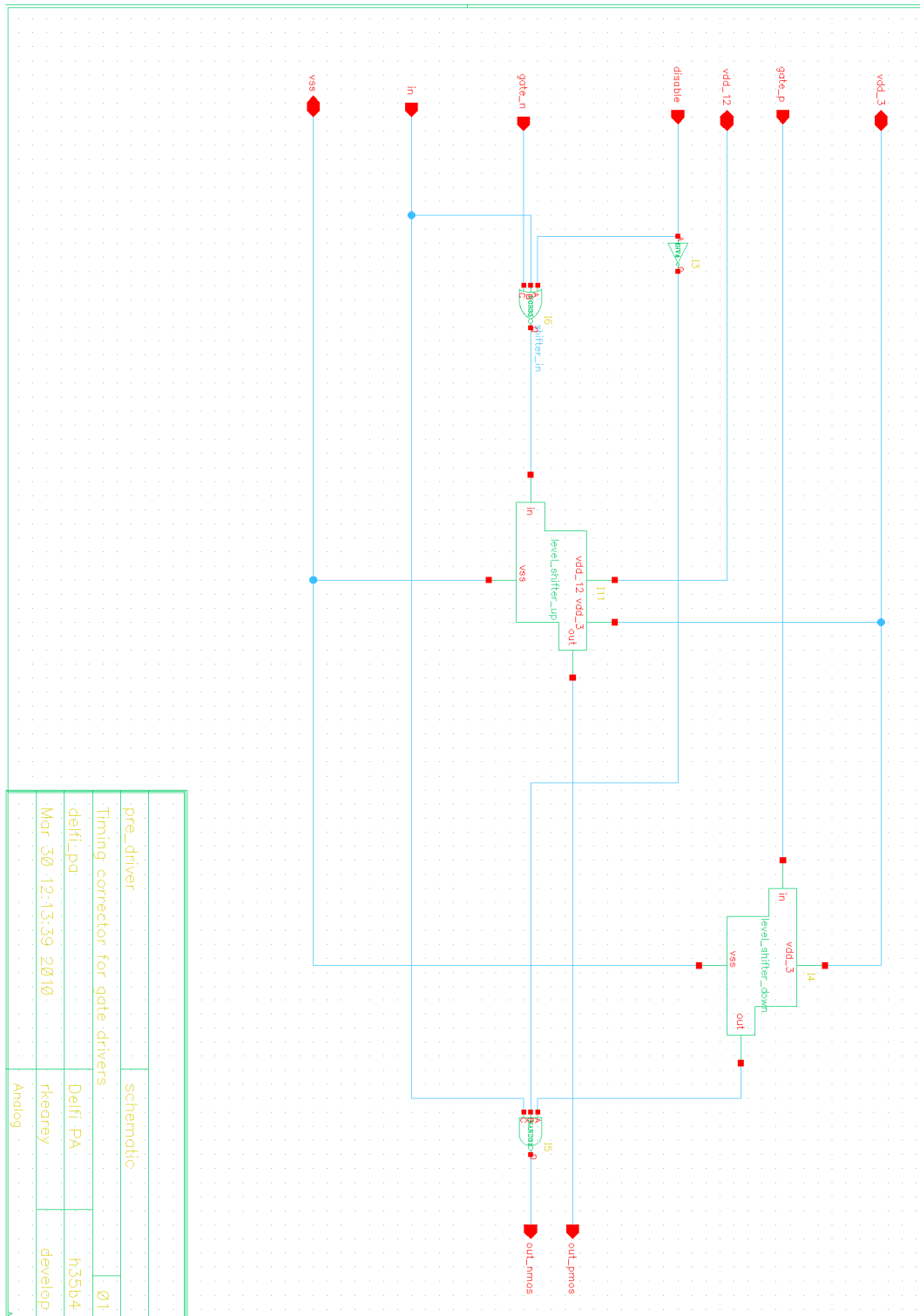
D.4 Level shifter down
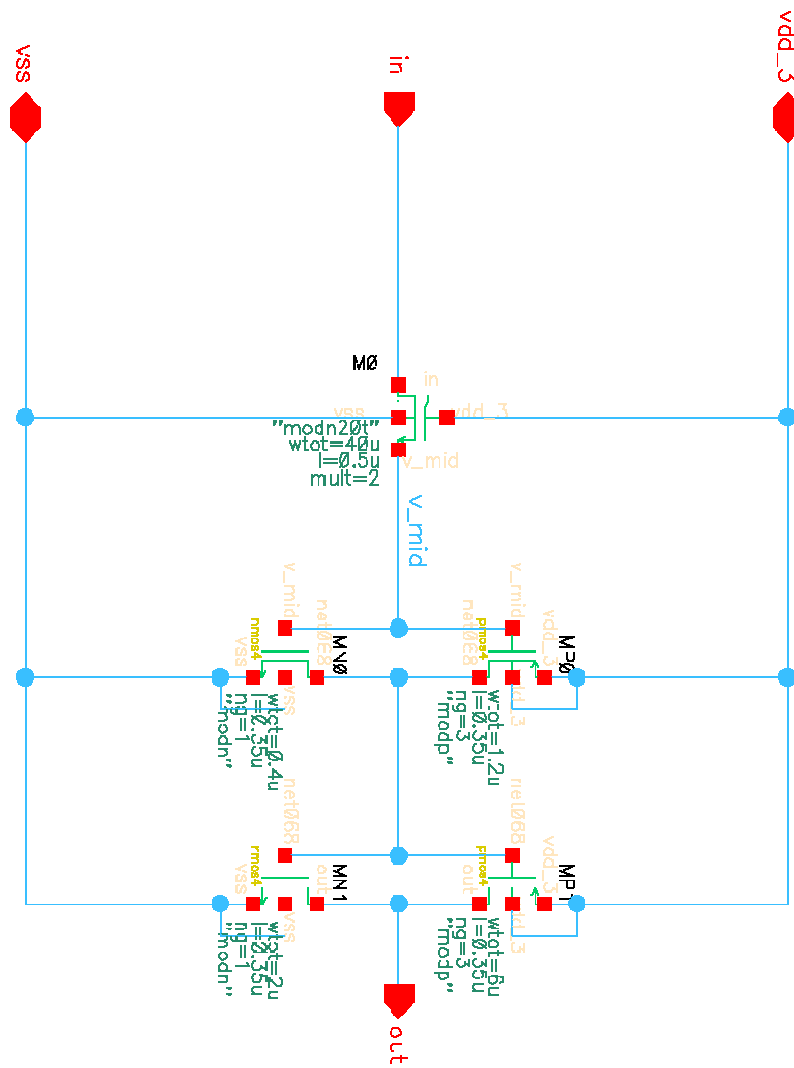
D.5 Level shifter up
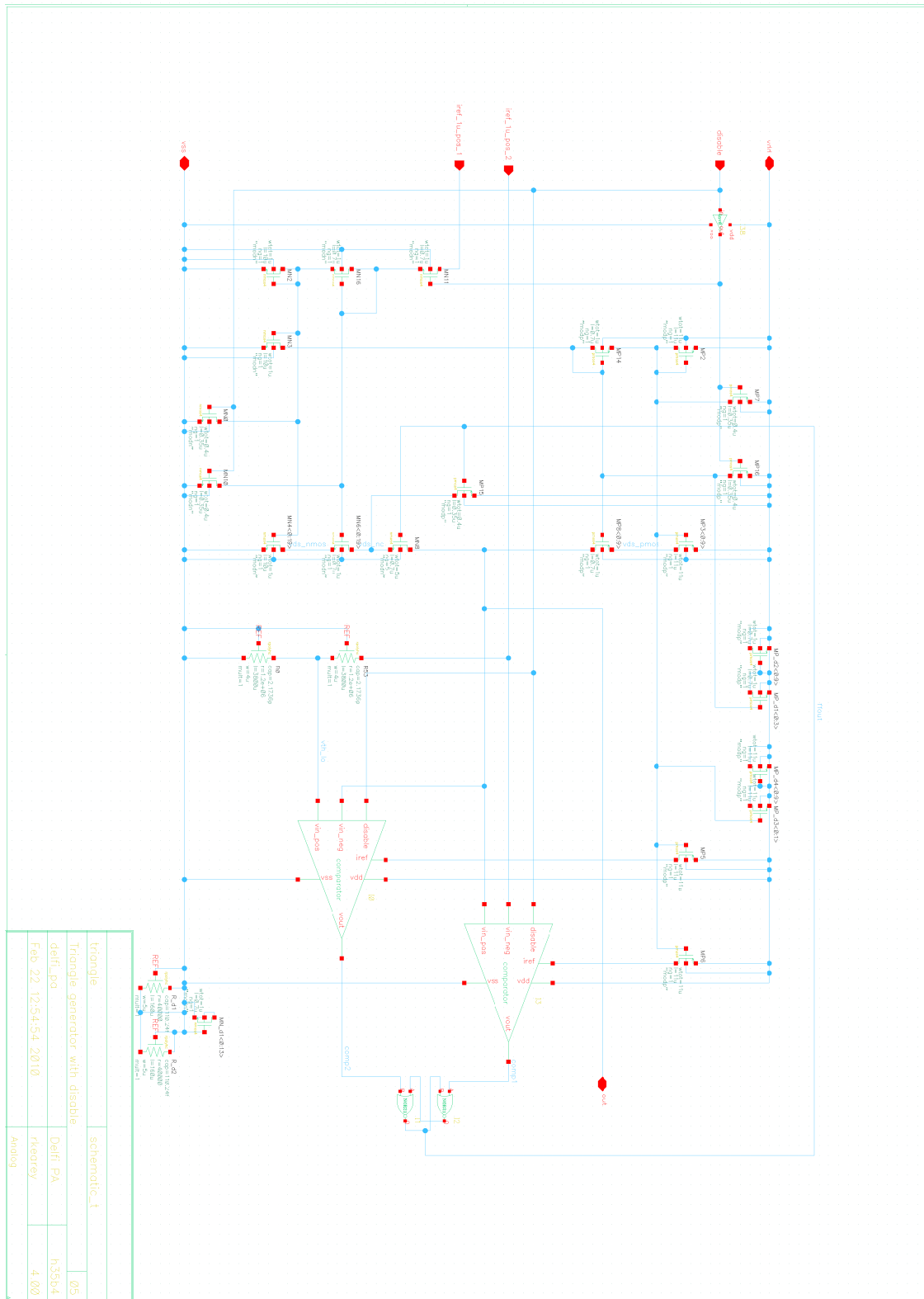
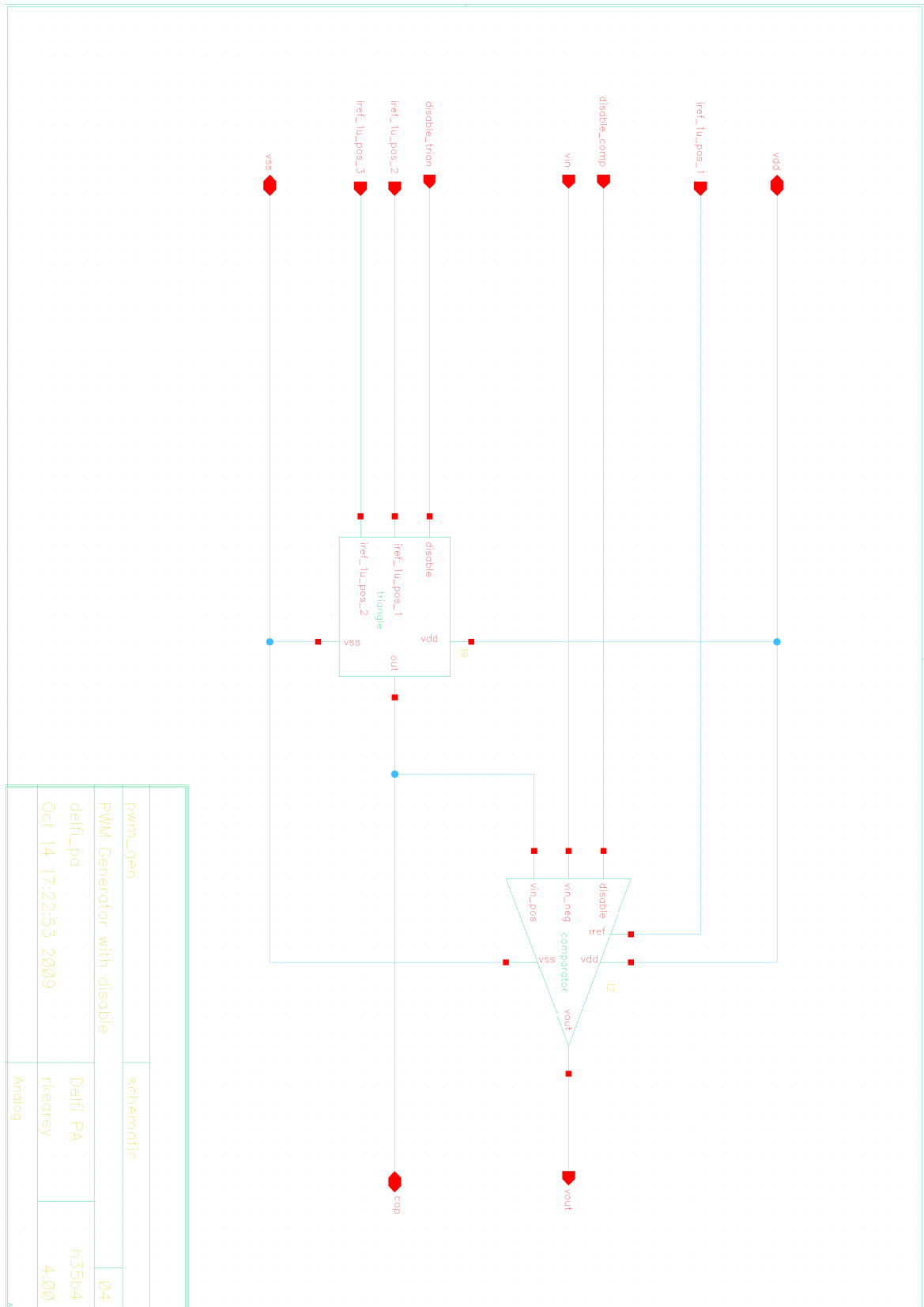## D.6 Dynamic transistor sizing circuit
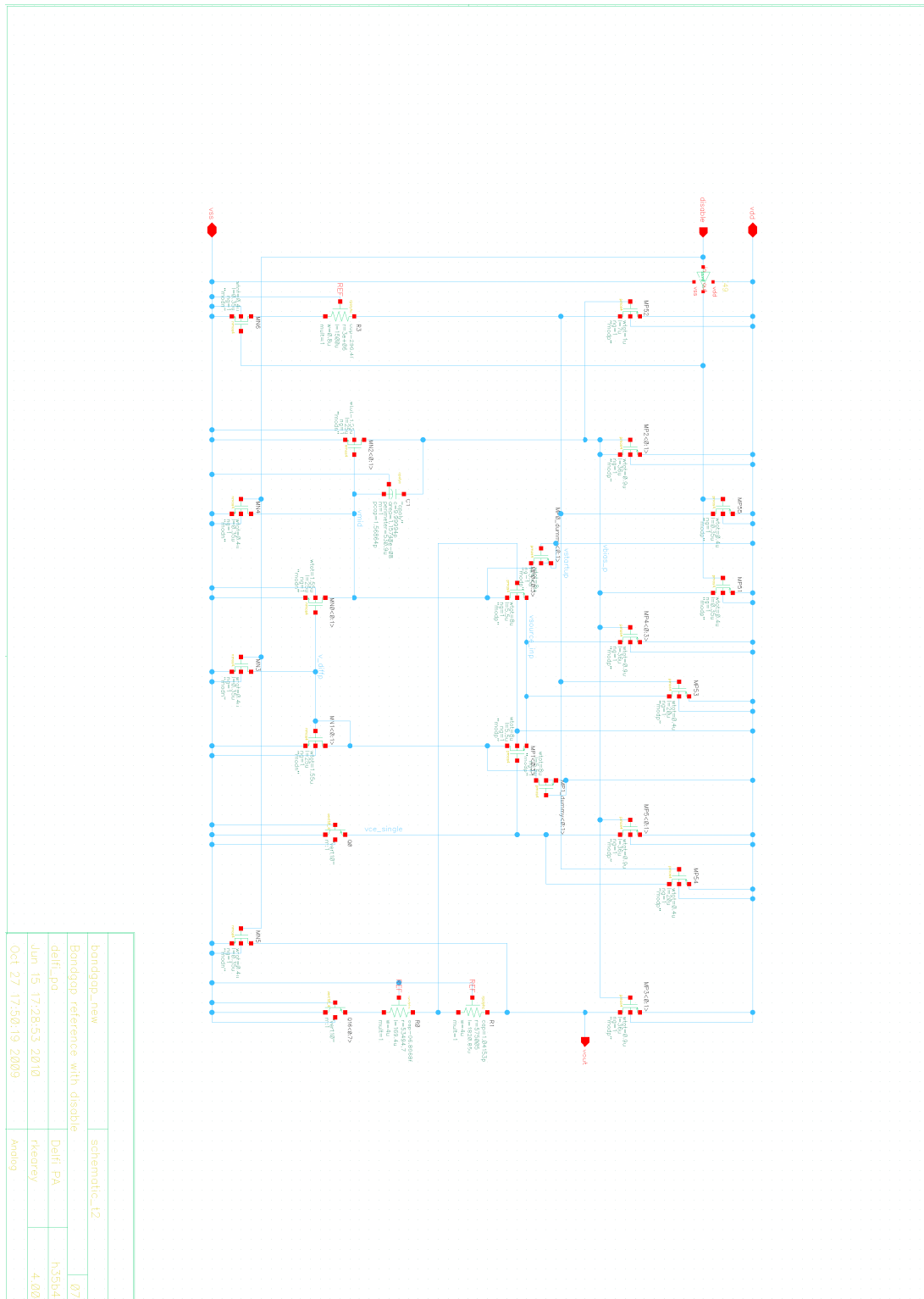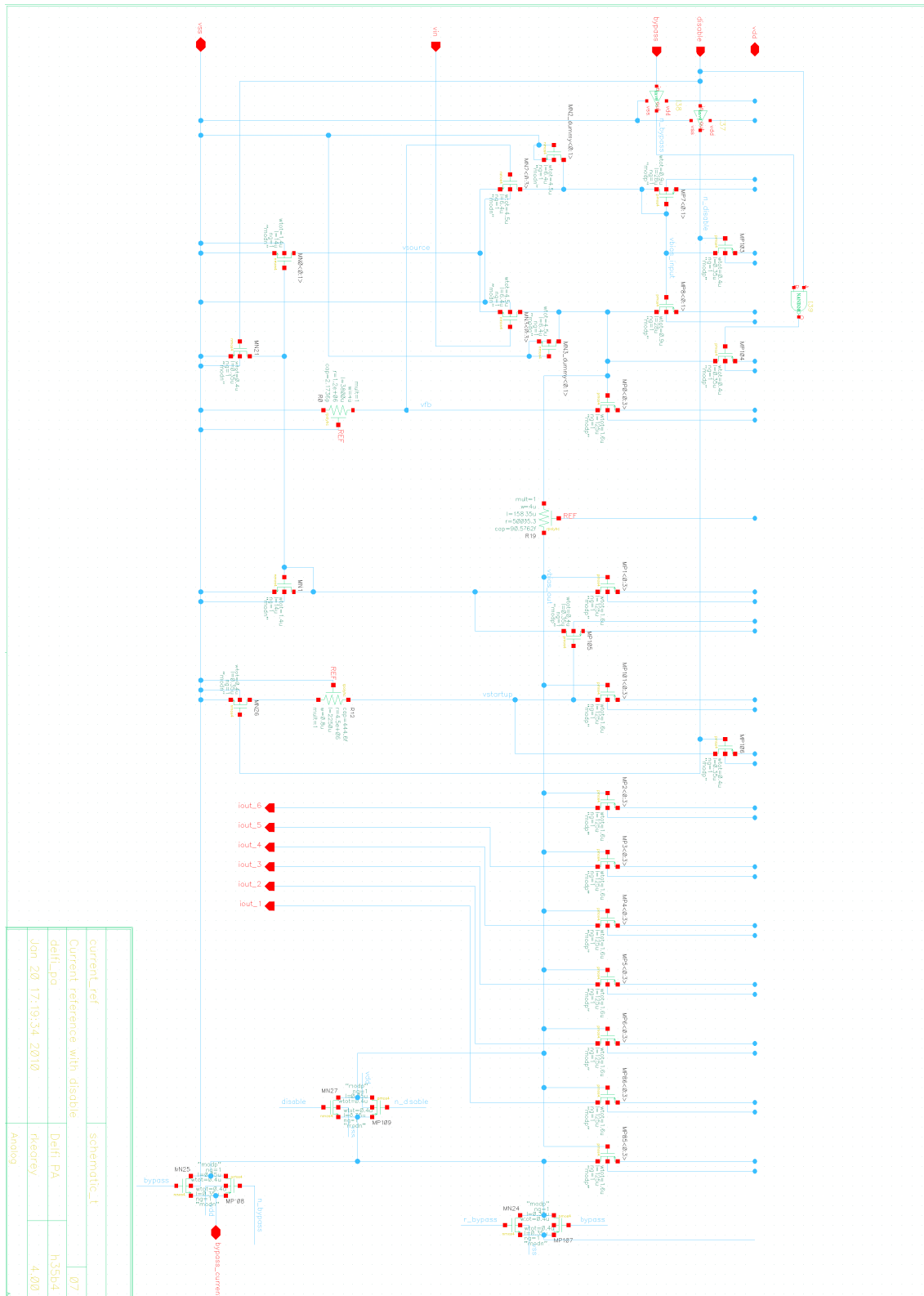
D.7 Complete class D amplifier
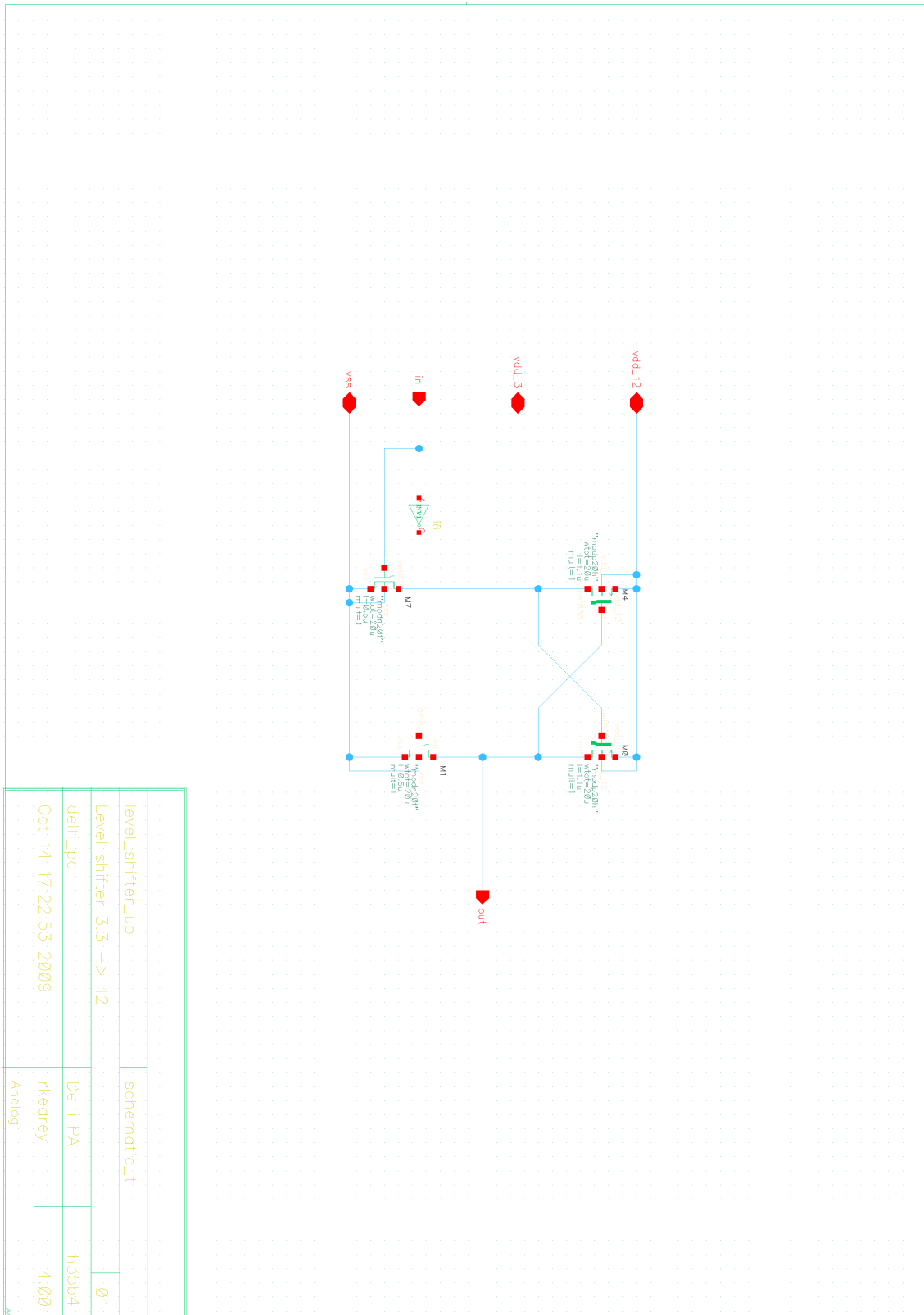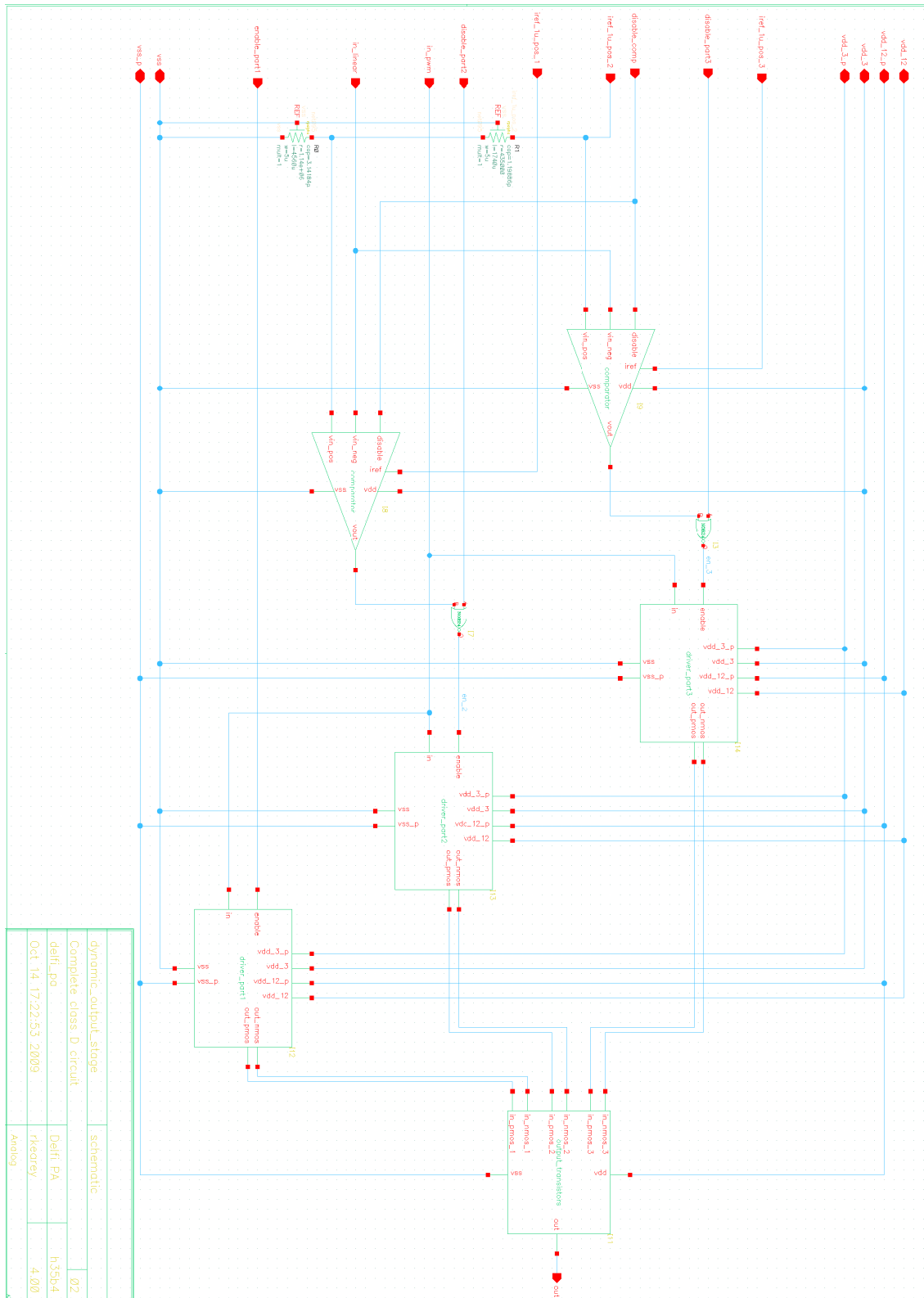
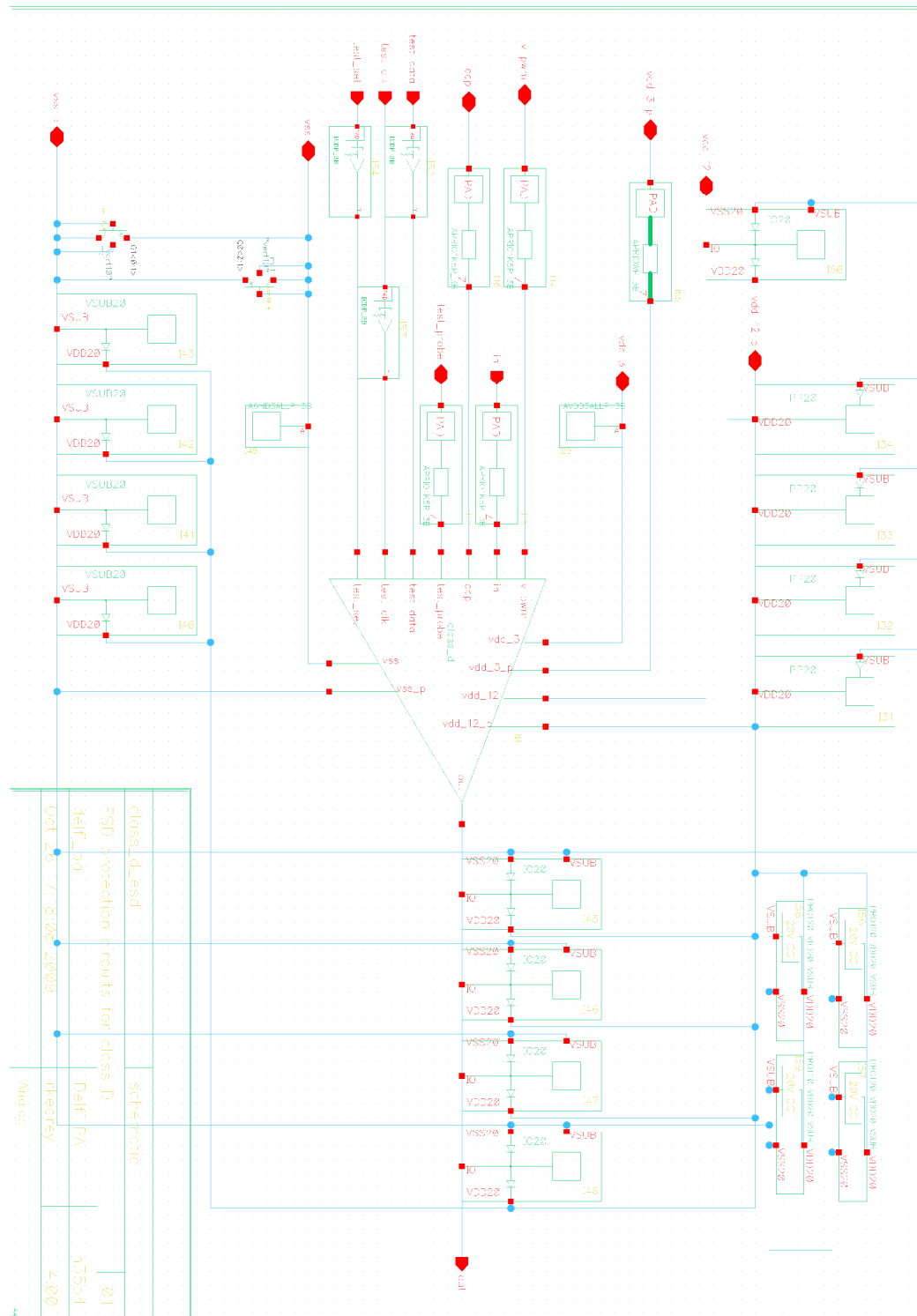# D.8 Comparator
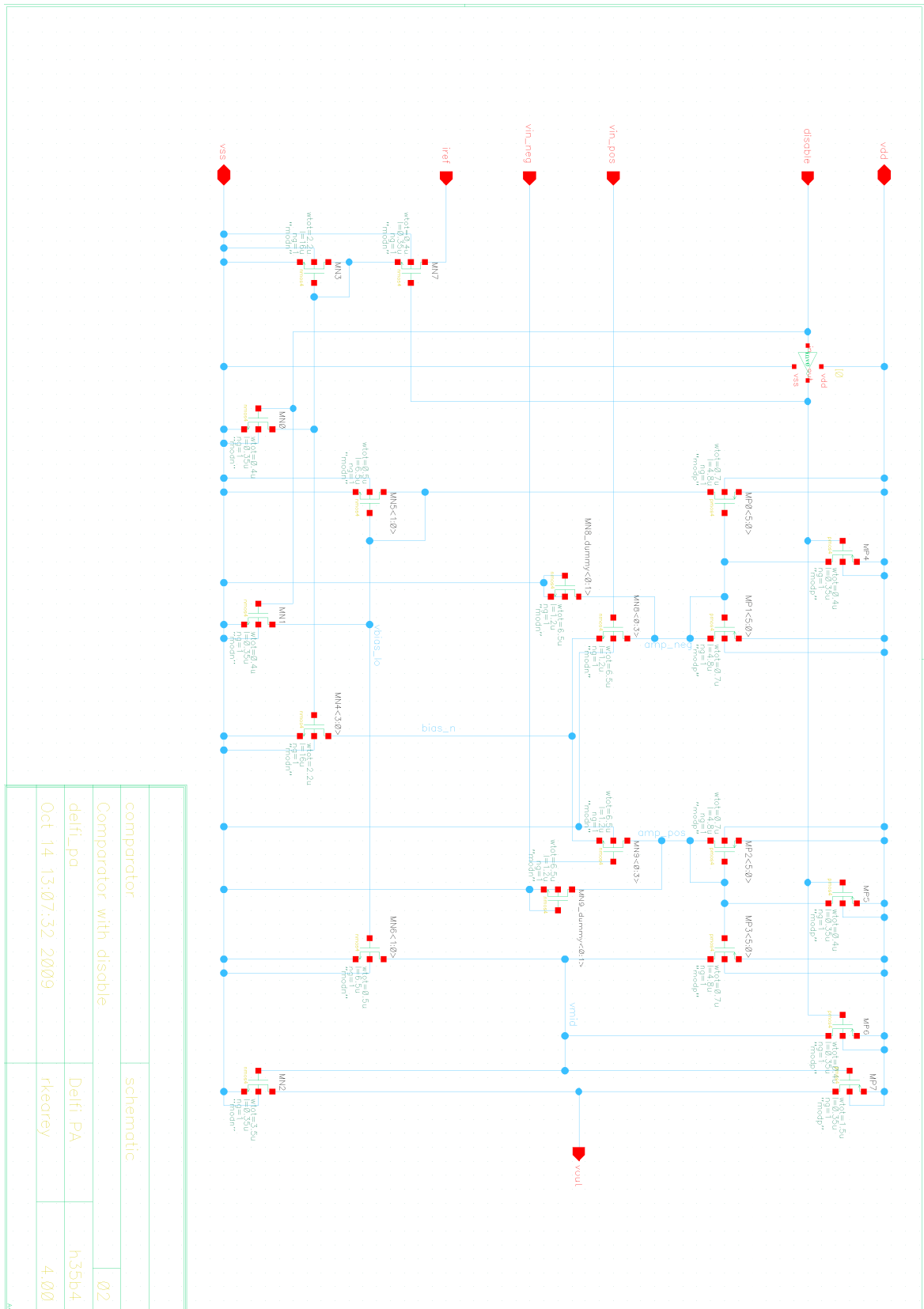
D.9 Triangle generator

D.10 PWM generator

D.11 Bandgap reference

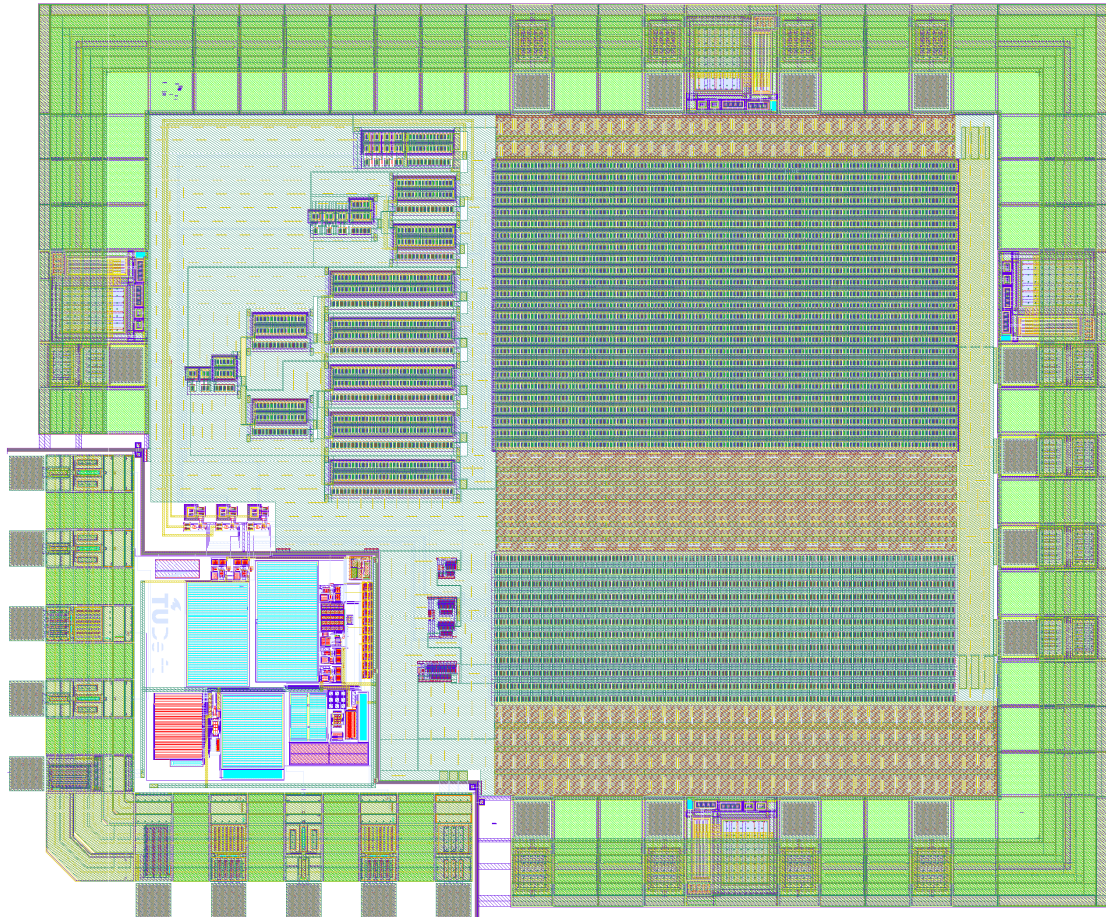## D.12 Current reference

D.13 ESD protection circuits

D.14 Test controller

Appendix E   Complete IC layout

Pad layout