University of Twente

Faculty of Electrical Engineering,
Mathematics & Computer Science

Charge Steered RF PA

Erik Olieman
MSc. Thesis
August 2010

Supervisors:
prof. dr. ir. B. Nauta
dr. ir. A.J. Annema
dr. ing. E.A.M. Klumperink

Report number: 067.3372
Chair of Integrated Circuit Design
Faculty of Electrical Engineering,
Mathematics & Computer Science
University of Twente
P. O. Box 217
7500 AE Enschede
The Netherlands

Summary

This master thesis describes the development of a new type RF PA and upconverter combination. Instead of the normally used voltage steering the amplifier uses charge steering. The parasitic capacitances of the MOSFET in combination with using charge steering form a feedback network that linearizes the amplifier.

The developed device is a switching amplifier. For good behavior it uses a capacitor to be able to quickly inject high amounts of current. Derived equations show that feedback becomes better if the gate-drain capacitance is relative large compared to the gate-source capacitance. For this reason a capacitor is periodically placed between the gate and the drain terminals of the amplifying transistor using switches. Half of the time this capacitor is placed there, during this time the charge present on the capacitor determines the output voltage. The other half of the time the amplifying transistor is turned off.

The designed RF power amplifier and upconverter combination is designed in 65nm CMOS and operates at a center frequency of 1GHz. The maximum output power is 17.5dBm, its 1dB compression point is only 0.6dB below its maximum output. Up to this output power also two-tone tests perform well, the worst in-band IMD3 component is at -29dBc at an amplitude equal to the 1dB compression point. An efficiency of 21% is reached at the 1dB compression point.

Charge injection by the switches decreases performance for lower output values, the range of powers across which the amplifier performs well is around 20dB. The maximum is the 1dB compression point, so the minimum power, where charge injection becomes a significant problem, is 20dB below the 1dB compression point.

Another problem is the required capacitances; the total size of the capacitors used in the circuit is over 80pF, which would cost a lot of chip area.

Contents

1 Introduction

Low power consumption is critical for portable systems to achieve good battery life. Power consumption in wireless systems is typically dominated by the RF power amplifier (PA), so a PA with a high efficiency is critical.

Switch mode power amplifiers with constant envelope output can achieve high efficiencies, however they can only amplify constant envelope modulation schemes, while most modern modulation schemes do not have constant envelopes. This allows for more efficient use of the spectrum and high data-rates, but comes at the cost of high linearity requirements. Nonlinear behavior causes distortion products in adjacent channels, lowering the percentage of the spectrum that can be used by other devices, and in-band distortion, lowering the achievable data-rate.

Linearity and efficiency can be exchanged for one another, so the goal is to design a system with both a high linearity and good efficiency.

1.1 Amplifier types

RF power amplifiers can be divided into two groups, linear and switching amplifiers. They are further divided into different classes: A, B, C, D, etc.

Linear amplifiers have an output that is at least partly sinusoidal. Class A, AB, B and C belong to the linear amplifiers. They differ in the time a transistor conducts, the conduction angle, from 100% in class A amplifiers till less than 50% in class C amplifiers. A lower conduction angle increases the efficiency, but at the cost of linearity.

Linear amplifiers can reach an arbitrary linearity by decreasing the output swing. However again, this causes a steep decline in efficiency.

Switching amplifiers do not attempt to create a sinusoidal output, they create a square wave. The output filter removes higher harmonics while keeping the first harmonics. This allows for a theoretical efficiency of 100%. A fully switching amplifier can only be used in constant envelope modulation schemes. By modulating the supply voltage also the output amplitude is modulated. In theory it would still be possible to have 100% efficiency by using a class D amplifier to modulate the output amplitude, but in practice the efficiency will decrease. Time delay between the modulation of the supply and the switch signal, together with the finite bandwidth of the supply modulation gives rise to distortion products, even when everything else is ideal. Since both these effects are related to the bandwidth of the RF output, the performance is limited for high bandwidth systems.

In this project a switching amplifier is developed that uses direct feedback to obtain a good linearity. This way it should be able to combine the high efficiency of switching amplifiers with the good linearity of linear amplifiers.

1.2 Project description

RF power amplifiers generally use voltage control; the intrinsic capacitances of the transistor only decrease the efficiency since they have to be charged and discharged, while they provide no benefits. For this master thesis the possibility for using charge/current control instead is investigated. Here the intrinsic capacitances form a feedback network that can linearize the transistor. Since the capacitances are intrinsic to the transistors this feedback is (almost) instantaneous, meaning there should be no stability problems.

1.3 Contents

In chapter 2 the underlying theory is discussed and relations are derived which describe the behavior of the amplifier. With the use of these relations a top level design is step by step created in chapter 3. The building blocks used in the top level design are created on transistor level in chapter 4. This is followed by simulation results in chapter 5 and the conclusion and recommendations in chapter 6.

Appendix A goes into more detail regarding derivation of equations, while appendix B contains the designed schematics.

2 Principle

RF power amplifiers generally employ voltage steering. The schematic shown in Figure 2-1 is the general circuit usually used by voltage steered power amplifiers.

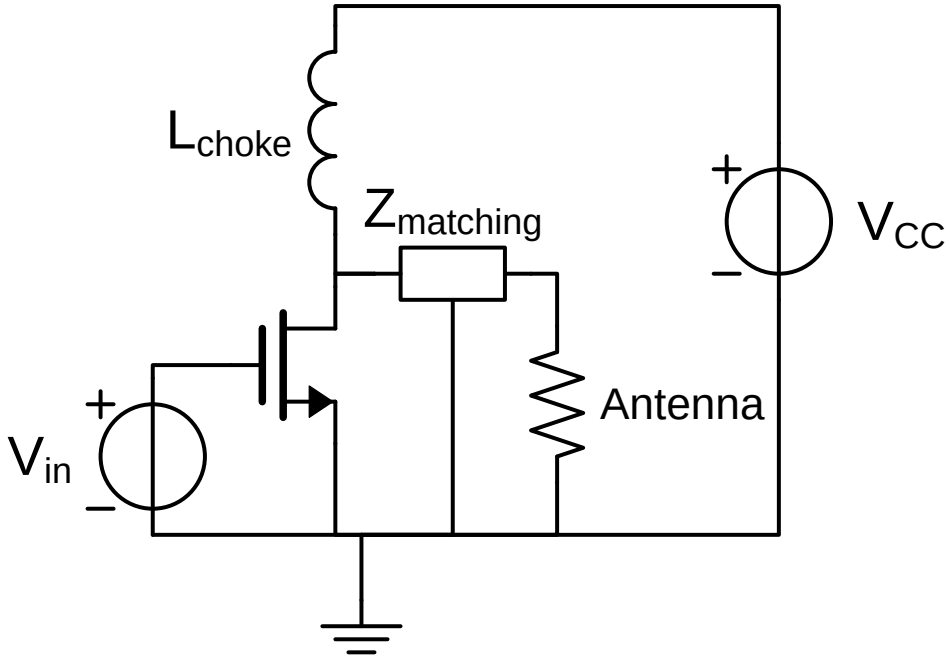

Figure 2-1: Voltage steered RF PA architecture

The gate is controlled by a voltage source. A transistor amplifies the input voltage and via a matching circuit an antenna is connected. This basic circuit can be used for many different operation modes, but some principles stay the same. The voltage source puts a voltage on the gate that is independent of everything else.

The basic principle of a charge steered PA instead of a voltage controlled PA is shown in figure 2-2.

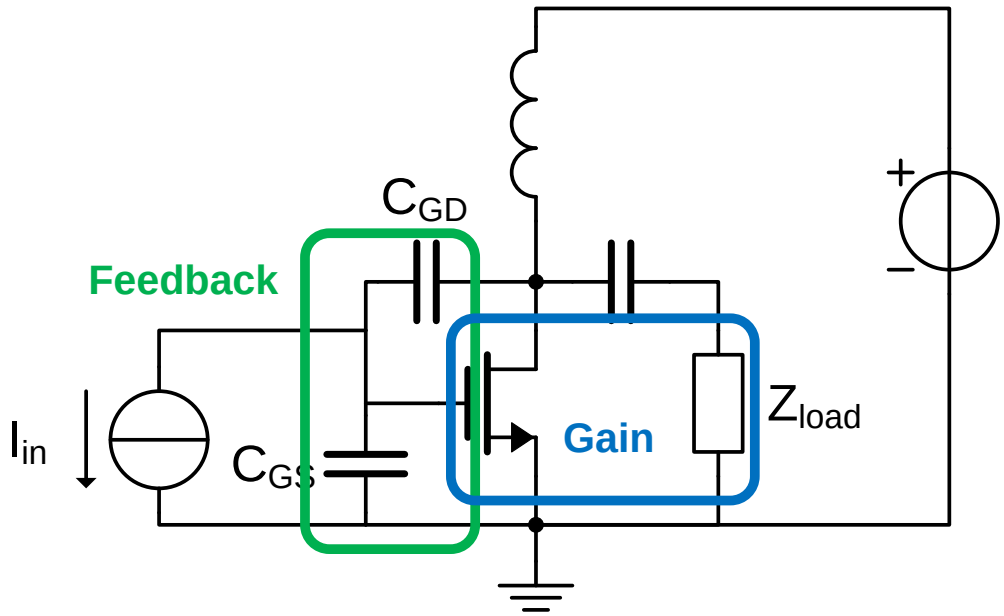

Figure 2-2: Principle of charge steering

Instead of a voltage at the gate, now a charge is applied on the parasitic capacitances of the transistor. When the drain voltage increases, the gate voltage will also increase via the voltage divider C_{GD} and C_{GS}. The increase in gate voltage results in an increase in drain current, which decreases the drain voltage. Hence this mechanism provides negative feedback and has as result that the output voltage is relative independent of the transconductance of the transistor, in contrast to voltage steered amplifiers that do not have this feedback.

2.1 Equations

To get a more accurate impression of a charge steered amplifier relations are derived with various levels of accuracy that describe that behavior of the amplifier. First a highly simplified version is derived, followed by more accurate ones.

2.1.1 Basic

First the small signal response of the amplifier is calculated. The individual steps are explained in the appendix, section A.1. The output voltage is given by:

$$V_{out} = -V_{in} R_L g_m \quad \textbf{(2-1)}$$

The input voltage can also be related to the charge present on the capacitors.

$$V_{in} = \frac{Q_{in} - Q_{GD}}{C_{GS}} \quad \textbf{(2-2)}$$

The charge in the gate-drain capacitance is given by the capacitance and the voltage across it.

$$Q_{GD} = C_{GD}(V_{in} - V_{out}) \quad \textbf{(2-3)}$$

Combining (2-2) and (2-3) and rewriting them gives:

$$V_{in} = \frac{Q_{in} + C_{GD} V_{out}}{C_{GS} + C_{GD}} \quad \textbf{(2-4)}$$

Combining (2-4) and (2-1) gives the output voltage as function of input charge.

$$V_{out} = -\frac{Q_{in} R_L g_m}{C_{GS} + C_{GD} + C_{GD} R_L g_m} \quad \textbf{(2-5)}$$

When the output resistance of the transistor is taken into account, (2-5) turns into:

$$V_{out} = -\frac{Q_{in} R_L r_{out} g_m}{(R_L + r_{out})(C_{GS} + C_{GD}) + C_{GD} R_L r_{out} g_m} \quad \textbf{(2-6)}$$

For an infinite transconductance, the voltage gain would be infinite, just like the ideal situation for an OpAmp, this reduces (2-6) to:

$$V_{out} = -\frac{Q_{in}}{C_{GD}} \quad \textbf{(2-7)}$$

In the ideal situation the output voltage depends linearly on the input charge and the gate-drain capacitance. In real situations the voltage gain is not infinite, but still it decreases dependency on the transconductance and output resistance of the transistor.

2.1.2 Output impedance

The output impedance of a circuit is given by (2-8).

$$Z_{out} = R_L * \frac{U_{open} - U_{loaded}}{U_{loaded}} \quad \textbf{(2-8)}$$

Combining this with (2-6) allows us to calculate the output impedance of the circuit.

$$Z_{out} = \frac{r_{out}(C_{GS} + C_{GD})}{C_{GS} + C_{GD} + C_{GD} r_{out} g_m} \quad \textbf{(2-9)}$$

A larger transconductance results in a lower output impedance. This was expected since feedback will lower the output impedance, and a higher transconductance implies a larger loop gain, so more feedback.

2.1.3 Quadratic analysis

It is interesting to analyze how the linearity of the system depends on how the transistor is controlled. For this the same method is used as in section 2.1.1. However now the drain current is assumed to have a quadratic dependency on the input voltage. This is a useful approximation since the basic MOST model is a quadratic relation between input voltage and output current.

$$I_D = V_{in} g_m + V_{in}^2 g_2 \quad \textbf{(2-10)}$$

Using the same methodology as in section 2.1.1 and some extra steps to simplify the result, the resulting output voltage becomes:

$$V_{out} = -\frac{\left(1 + g_m R_L C_r + \frac{2R_L Q_{in} C_r^2 g_2}{C_{GD}}\right) - \sqrt{(1 + g_m R_L C_r)^2 + \frac{4R_L Q_{in} C_r^2 g_2}{C_{GD}}}}{2 g_2 R_L C_r^2} \quad \textbf{(2-11)}$$

$$C_r = \frac{C_{GD}}{C_{GD} + C_{GS}} \quad \textbf{(2-12)}$$

To compare this to the voltage steered case a Taylor approximation is calculated. Since we want to compare the second order harmonics we need the corresponding component in the Taylor series.

The second order term in the Taylor series is noted as H_2. For both the voltage and charge steered situation H_2 is divided by the load impedance, so the second order component in the drain current can be compared directly.

For the charge steered version the input charge is substituted by the corresponding input voltage.

$$Q_{in} = V_{in}\left(C_{GS} + C_{GD}(1 + g_m R_L)\right) \quad \textbf{(2-13)}$$

Using (2-13) the second order coefficient of the drain current of the charge steered circuit is given by equation 2-14.

$$\frac{H_2}{R_L} = -\frac{(C_{GD} + C_{GS}) g_2}{(C_{GD} + C_{GS} + g_m R_L C_{GD})} \quad \textbf{(2-14)}$$

For the voltage steered version this is equal to equation 2-15.

$$\frac{H_2}{R_L} = -g_2 \qquad \textbf{(2-15)}$$

Compared to the voltage steered version, it is clear that the second order distortion of the charge steered circuit will always be lower. As expected the linearity increases with the loop gain.

When the same analysis is done for the first harmonic around $V_{in} = 0$ the result is that they are identical, this is expected since in the linear case the charge steered PA has an output equal to the voltage steered PA when (2-13) is substituted in the relation between input charge and output voltage (2-142-11).

Equation 2-11 shows a potential problem: The dependency between output voltage and input charge includes a square root, which has an infinite Taylor series. So while the voltage steered amplifier would only have second order distortion, and hence has no in-band intermodulation distortion, the charge steered amplifier does have intermodulation distortion. So in order to be able to compare intermodulation products a more accurate analysis is required where both the voltage and charge controlled versions have intermodulation distortion.

2.1.4 Mobility reduction

In order to find the effect of charge steering on (third order) intermodulation distortion, a transistor model that includes higher order components in the output current than the square law model previously used is required. It was found that mobility reduction is one of the major causes of non-ideal behavior in MOSFETs [**1**]. The square law relation for the drain current with mobility reduction added is shown in equation 2-16. This formula also covers velocity saturation, which only adds an extra factor to θ.

$$I_D = \frac{1}{2}\frac{K}{1+\theta(V_{in}-V_{TH})}(V_{in}-V_{TH})^2 \qquad \textbf{(2-16)}$$

This results in high order distortion in both the voltage and charge controlled versions, enabling a fair comparison between these two.

The output voltage when charge steering is used is equal to:

$$V_{out} = \frac{-1-(\theta+C_rR_LK)\left(\frac{Q_{in}C_r}{C_{GD}}-V_{TH}\right)+\sqrt{c_2Q_{in}^2+c_1Q_{in}+c_0}}{2C_r\left(\theta+\frac{1}{2}R_LKC_r\right)}$$

$$c_0 = 1-2(\theta+C_rR_LK)V_{TH}+\theta^2V_{TH}^2$$
$$c_1 = \frac{2R_LKC_r^2+2\theta C_r-2\theta^2C_rV_{TH}}{C_{GD}} \qquad \textbf{(2-17)}$$
$$c_2 = \frac{\theta^2C_r^2}{C_{GD}^2}$$

C_R is given in (2-12).
The Taylor expansion of (2-16) times R_L and of (2-17) are both calculated and used to compare the linearity. For a fair comparison a principle similar to the OIP_X is required, so (2-18) is used:

$$F_X = \frac{H_1^X}{H_X} \qquad \textbf{(2-18)}$$

This equation is, minus some scaling factors, the same as what is used to calculate the OIP_X, where X is the order of the distortion that we want to look at, while H_Y is the Y'th order behavior, which is calculated by the Taylor expansion.

As discussed earlier, the resulting equations are too large to be useful for symbolic comparisons. For that reason they are compared numerical. The used values are:

R_L	50Ω
C_{GS}	2pF
C_{GD}	1pF
V_{TH}	0V
K	2.6

These values correspond roughly to a thick oxide transistor in 65nm CMOS with dimensions of 50x40µm/0.28µm. Only the threshold voltage is set at zero for simplicity, since it only results in a shift it does not affect the comparison. The corresponding θ value is around 2. For comparison also higher and lower θ values are used.

First a θ of 2 is used, corresponding to a reasonable amount of third order distortion. The result for third order products, calculated from (2-18) is plotted below; higher corresponds to a better linearity.

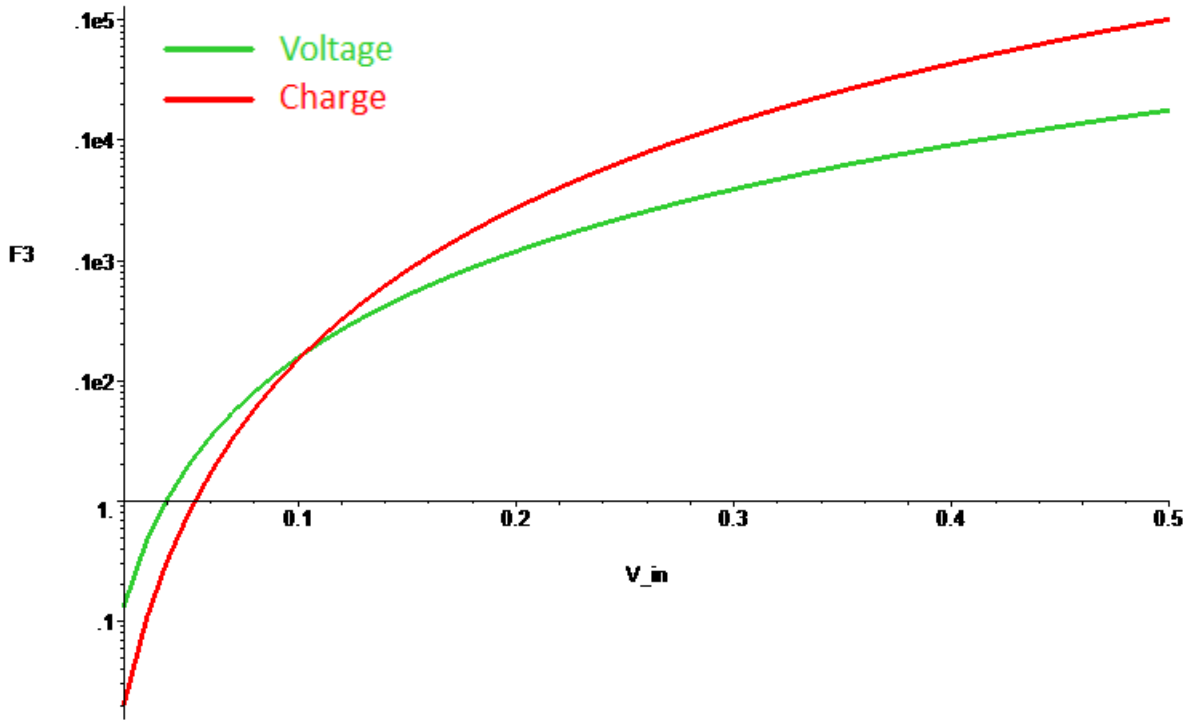

Figure 2-3: Linearity regarding third order intermodulation products for θ =2

For both voltage steering and charge steering the linearity improves for higher input voltages. Since the mobility reduction part of the transfer function starts to cancel the quadratic behavior this is expected. The charge steered situation increases its linearity much faster, because the increase in loop gain results in more feedback, improving linearity. However for low voltages the voltage controlled version has higher linearity. In the charge steered circuit there is little feedback at low

inputs, and a lot of second order distortion that is converted to third order distortion by intermodulation.

Figure 2-4 shows the same for second order distortion. Here the charge steered version is always better, but this is not as important as third order distortion.

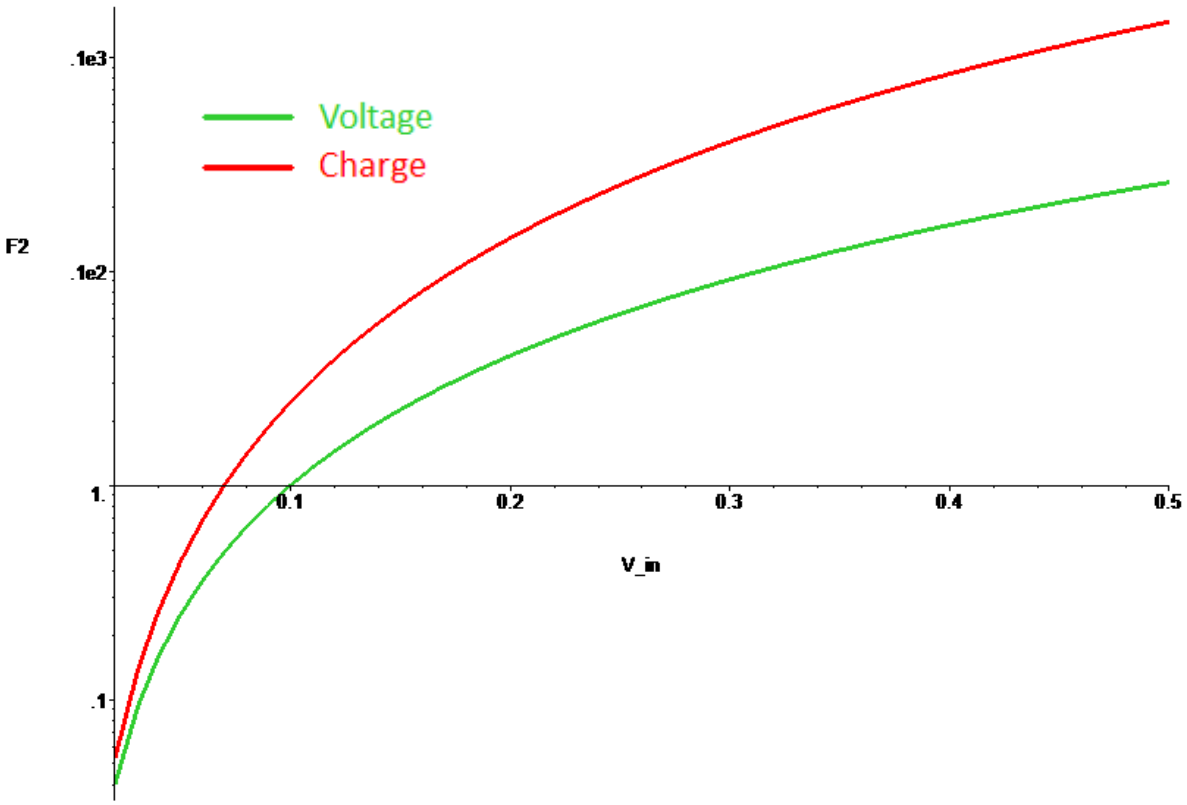

Figure 2-4: Linearity regarding second order intermodulation products for θ =2

When the third order products compared to the linear products are plotted again, but for a much lower θ value, the problem with second order distortion is clearer; figure 2-5 shows this.

For $V_{in} = 0.3V$ and a low value for θ the amount of third order components in the transconductance is much lower. The amount of second order distortion transformed into third order by the charge steered transistor stays roughly equal, so this is dominant and for low values of θ the third order distortion in the charge steered circuit approaches a minimum value due to this, while this does not happen in the voltage steered circuit.

When θ increases the feedback starts to suppress the third order distortion, and the charge steered circuit becomes more linear.

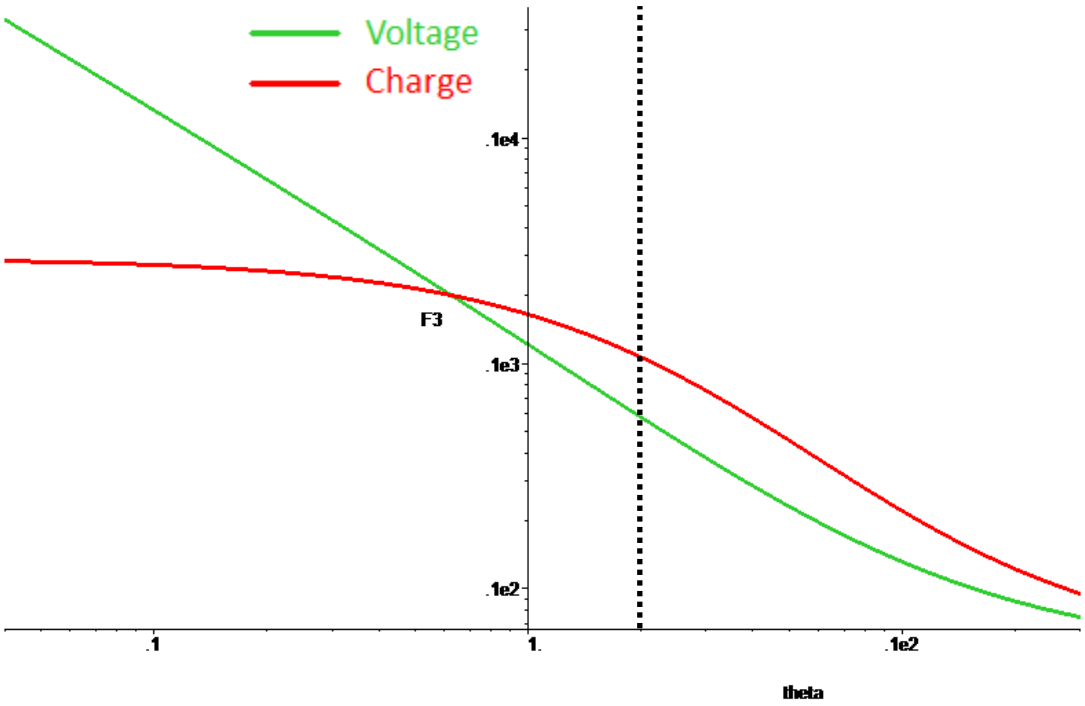

Figure 2-5: Linearity regarding third order intermodulation products for Vin=0.3V, dashed line represents nominal θ

It can be concluded that for devices that have a behavior close to the ideal square-law relation charge steering does not offer advantages. However when the behavior moves away from the square-law behavior with more non-idealities, the charge steering has better linearity, especially when the loop gain is high.

2.2 Capacitances

Until now it was assumed that the capacitances of a MOSFET have a fixed value, which is not the case. Since the transfer function in case of charge steering depends mainly on the capacitances it is important to know their behavior. In this paragraph the sources of the capacitances and their behaviors are discussed.

The parasitic capacitances of MOSFETs come from three sources [**2**][**3**]:

- Intrinsic capacitances
- Extrinsic capacitances
- Junction capacitances

The intrinsic capacitances are those due to the channel. They are highly bias dependant, since they depend on the depletion and inversion layer shape and size.

Extrinsic capacitances are relatively bias independent. They are for example the overlap capacitances between the gate and the heavily doped source and drain connections.

The junction capacitances are between the source/drain/channel areas and the bulk of the MOSFET. They are bias dependent, but they are not related to the gate capacitances.

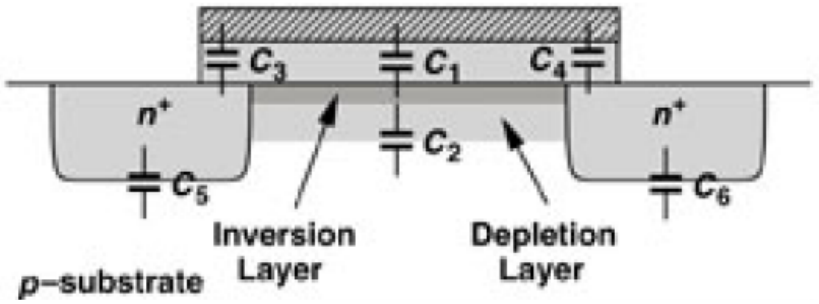

Figure 2-6: MOSFET capacitances

The two most important capacitances of a charge steered amplifier are the gate-source and especially the gate-drain capacitance: they directly affect the transfer function. The drain-bulk capacitance is also relevant since it needs to be charged and discharged by the amplifier continuously, so it affects the efficiency.

The gate-drain capacitance is the sum of its intrinsic and extrinsic capacitances. The intrinsic capacitance is largely non-existent in saturation; there is not much of an inversion layer at the drain side of the transistor in saturation. When the device is turned off it is also not present, only when it enters triode the intrinsic capacitance becomes significant.

What remains is the overlap capacitance, the capacitance between the gate and the drain area. Since these are independent of the state of the MOSFET, this capacitance should be constant. In reality the electric field between the gate and the drain is affected by the state of the MOSFET, which means the capacitance is not entirely constant. However compared to for example the transconductance of a transistor, it is much more stable.

The gate-source capacitance is similar, however here the intrinsic capacitance plays a large role. Near the source there is a large inversion layer which results in a large gate-source capacitance. This causes the gate-source capacitance to be much more bias dependent than the gate-drain capacitance. However once it is fully saturated also the intrinsic capacitance becomes quite constant.

So especially the gate-drain capacitance is relative constant, which is required since the transfer function depends a lot on this capacitance.

3 Top level design

Charge/current steering can be used in two ways: linear and switching. In the case of a linear amplifier there would be a current source directly connected to the gate that puts the RF signal that needs to be transmitted directly on the transistor. The alternative is a switching amplifier, which switches the transistor between its off-state and a certain on-state.

The decision was made to create a switching PA instead of a linear class A PA. Both are possible, but a switching PA should achieve a higher efficiency and was considered to be more interesting. In this chapter the top level design is created step by step.

3.1 Location of charge injection

The main design problem is where to inject the charge. The switching behavior requires a lot of current in a short period, which is best done using capacitors. So it needs to be determined where the capacitor injects it charge; there are two possible locations where the injection capacitor can be placed: either between the gate and the source or the gate and the drain.

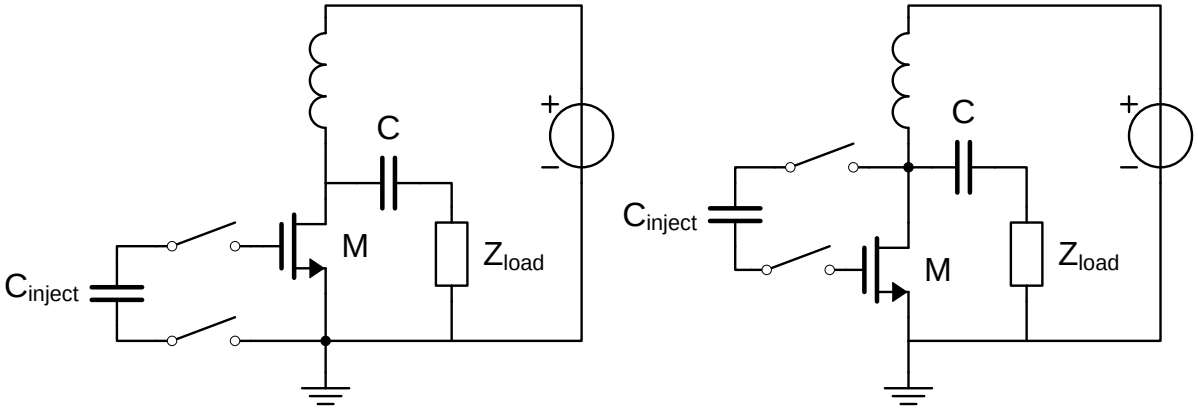

Figure 3-1: Injection place possibilities, between gate and source or gate and drain

Using the gate-source capacitance is the easiest solution; it is straight forward to charge the capacitor to the required value, and can be placed in parallel easily. The alternative, gate-drain, is harder to create. The capacitor needs to be able to be charged to a much larger range of possible voltages since the drain potential is not constant.

However putting the injection capacitor between the gate and the drain has also advantages compared to putting it between the gate and the source. Equation 2-5 gave the basic transfer function, which was:

$$V_{out} = -\frac{Q_{in} R_L g_m}{C_{GS} + C_{GD} + C_{GD} R_L g_m} \quad \textbf{(3-1)}$$

With an infinite transconductance (3-1) reduces to Q_{in}/C_{GD}. In practical situations the voltage gain will not be infinite, but only 3-4x voltage gain is a more realistic value. The gate-source capacitance will also be 2-3 times larger than the gate-drain capacitance. The injection capacitance needs to be a few times larger than the capacitance it is placed next to, so when the gate-source capacitance is used for injection, the C_{GS} term will be the dominant term in the denominator, since it is significantly larger than the C_{GD} and $C_{GD} R_L g_m$ terms.

From this it follows that (3-1) then can be simplified to:

$$V_{out} \approx -\frac{Q_{in} R_L g_m}{C_{GS}} = V_{in} R_L g_m \quad \textbf{(3-2)}$$

This is effectively equal to a voltage steered PA, and feedback does not play a role anymore. The only possibility to create enough feedback while using the gate-source capacitor is by placing a static capacitance parallel to the gate-drain capacitance. However this would have a large negative effect on the efficiency, since it would need to be charged and discharged every cycle.

When the injection capacitor is placed between the gate and drain terminals, the dominant factor will be $C_{GD} R_L g_m$, the feedback even becomes more effective due to the relative decrement of the C_{GS} term. Contrary to adding a static gate-drain capacitor this does not have a large impact on the efficiency since it is disconnected when the transistor is turned off, so it is not discharged.

For these reasons the gate-drain method is used. Since it also needs to be possible to turn the transistor off, a switch from the gate is added to achieve that. The switch between the injection capacitor and the drain is not necessarily needed, so that brings the design for now to:

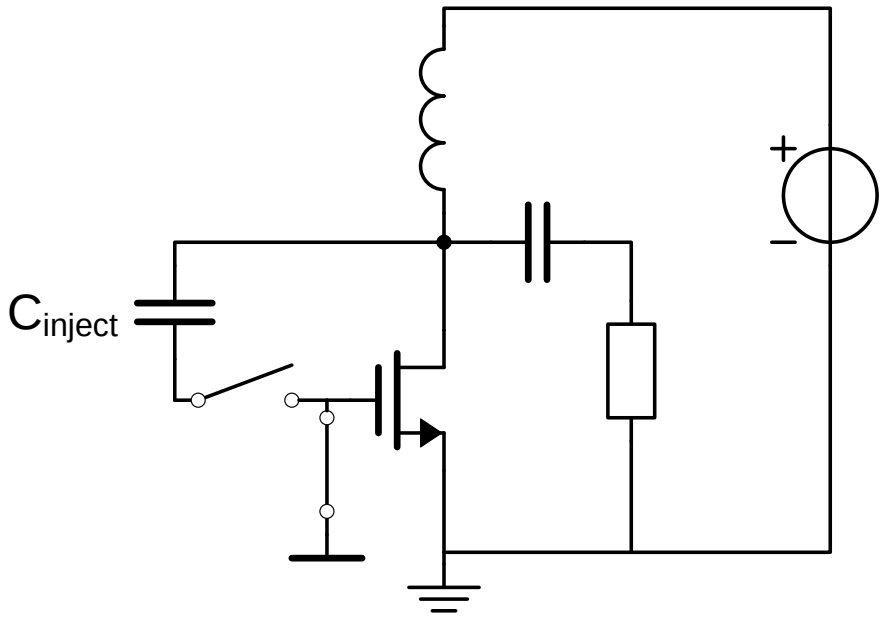

Figure 3-2: Design with gate-drain injection

3.2 Charging the capacitor

Because both the gate and the drain terminals have a quite large swing, and with opposite phases, the capacitance needs to be charged/discharged across a large range.

In 65nm CMOS, taking into account the 1.2V supply and that the amplifying transistor is a thick oxide device, the minimum voltage across C_{inject} is roughly:

$$V_{inject} = V_{TH} - V_{CC} \approx -0.7V \quad \textbf{(3-3)}$$

While the maximum voltage is:

$$V_{inject} = (V_{TH} + V_{GT}) - 0 \approx 0.9V \quad \textbf{(3-4)}$$

In principle the drain side of the injection capacitor does not need to be decoupled from the drain,

when a current source is used for charging and discharging the capacitor the drain terminal will be a ground for the low frequency input signal.

However this gives a problem since the drain voltage has a large swing, this is shown in figure 3-3. When the injection capacitance is charged to its maximum value, the output voltage is around zero volt while the corresponding gate voltage will be roughly 0.9V. When the transistor is switched to its off-state, the output voltage will quickly peak to far above 1.2V. The voltage across the capacitor will stay the same, so also the other side of the capacitor, where the current source is present, will rise quickly to high voltages, at which a realistic current source typically cannot function well.

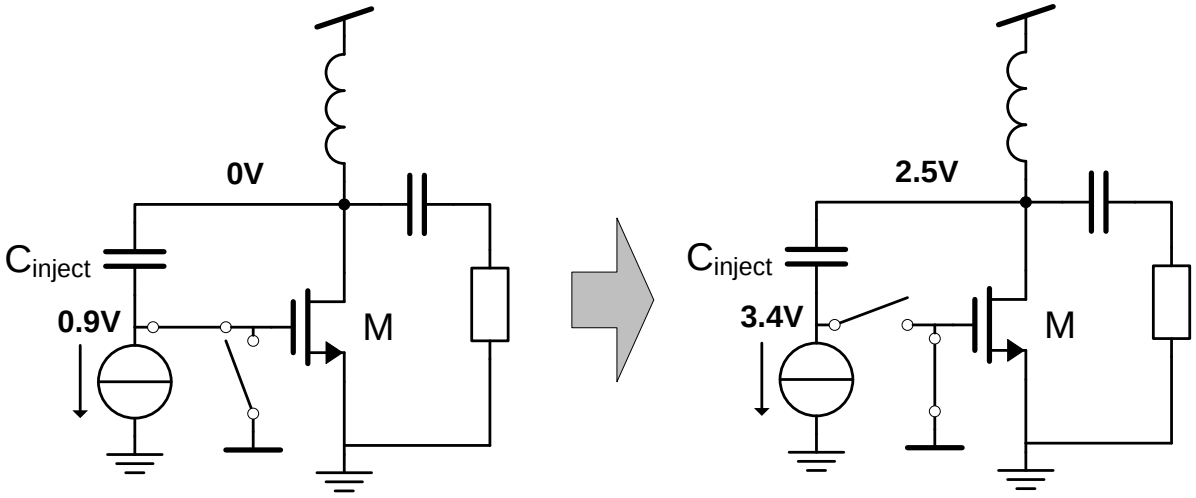

Figure 3-3: Voltages with injection capacitor fixed at drain

Even when the capacitor is not directly connected to the drain there is still a large swing across the capacitor, as given by (3-3) and (3-4). So to limit the voltage swing at the input to reasonable values differential steering of C_{inject} is required, so each source only requires a swing half of the total required swing.

Figure 3-4 shows the two possible ways to put the required charge on the capacitor, voltage steering and current steering. When the charge is put on the capacitor using a voltage source, two accurate voltage sources for differential steering are required. Also both voltage sources need to be disconnected from the amplifier itself when it is in its on-state. When a current source is used as input, an additional voltage source is required to keep the swing at the current source small, the current source itself may always be connected to the circuit.

So in both cases two sources are required. However in the case of the current source, the voltage source is only required to keep the swing at the current source sufficiently small: it has no direct influence on the charge on the capacitance, while in the case of two voltage sources both directly determine the charge on the capacitor, so both need to be of much higher quality than the voltage source that supplements the current source.

For this reason a current source is used as input. The extra voltage source should generate roughly the same voltage as the output voltage of the circuit when the transistor is conducting, in order for the current source to see a stable voltage.

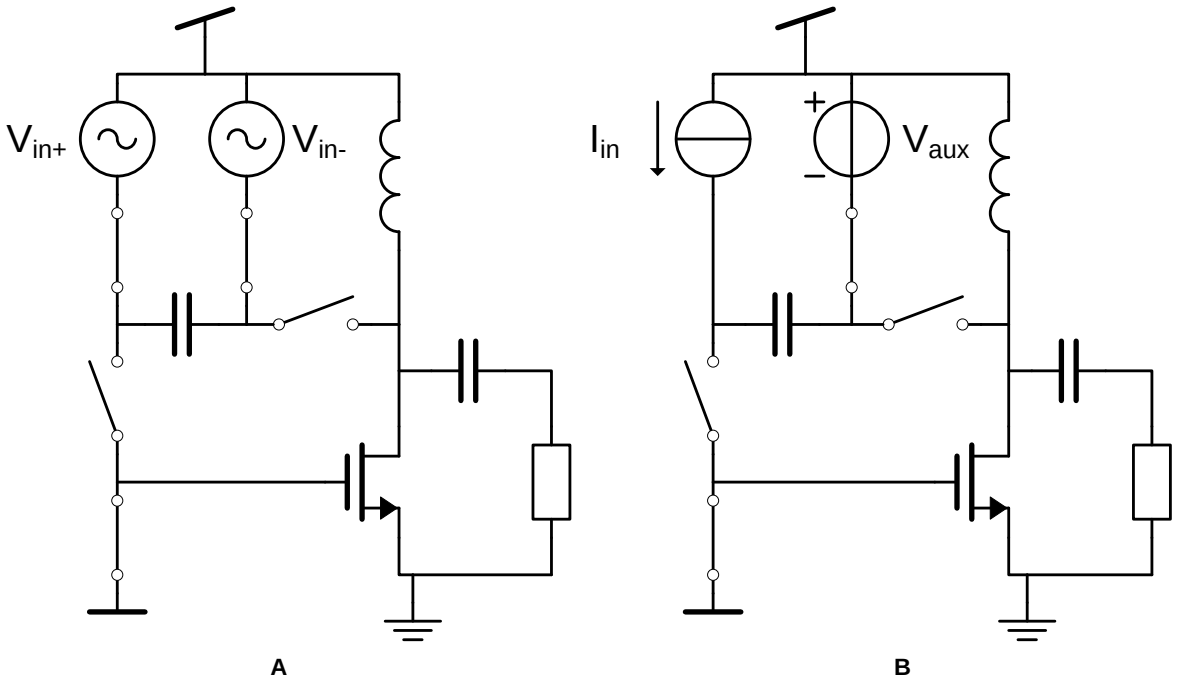

Figure 3-4: Design with differential steering; a) voltage steering, b) current steering

To turn the transistor of the gate can be switched to ground, to its threshold voltage or some other voltage below threshold. Switching to another voltage does not offer benefits compared to either switching to ground or to threshold. Switching to ground is easier and makes sure the transistor is fully turned off. However it also means that the parasitic capacitances of the amplifying transistor are fully discharged, so they have to be charged again the next cycle, decreasing the efficiency. Additionally if the input source is turned off, the voltage across the injection capacitor will end up such that the gate voltage is always zero volt. Now when a little bit of current is entered by the source, nothing will happen since it first needs to get the gate voltage above threshold.

The alternative is switching to threshold. This is a bit harder since a voltage source is required that represents the threshold voltage. This is a DC voltage source, which is easier to make than an AC source, however it does need to be able to handle the current spikes that will occur when the gate is switched to threshold. Also the transistor will be conducting more current than when it is switched to ground, since it is in weak inversion.

The advantages are the higher efficiency, since the parasitic capacitances of the amplifying transistor are not further discharged than required and with no input current the gate voltage will stay at threshold voltage, so with some input current the transistor will then immediately start conducting.

3.3 Carrier suppression

In the PA the charge on the injection capacitor represents the baseband signal, while the switches mixes this up to the required RF output. A problem arises when a simple sinus with no DC component is the baseband signal. Mixed up, this results in two sine waves at slightly different frequencies, as given by (3-5). The corresponding waveform is shown in figure 3-5.

$$y = \sin\left((\omega_{carrier} - \omega_{base})\, t\right) + \sin\left((\omega_{carrier} + \omega_{base})\, t\right) \quad \textbf{(3-5)}$$

Since the input sinus has no DC component, the modulus of the required output signal contains zero-crossings, which means the output transistor would both need to source and sink current, while it can only sink current. A number of solutions are available, which are discussed in the following sections, a choice is made in section 3.3.4.

In addition to these methods also a DC bias current could be added, however then it becomes a class A amplifier instead of the required switching amplifier.

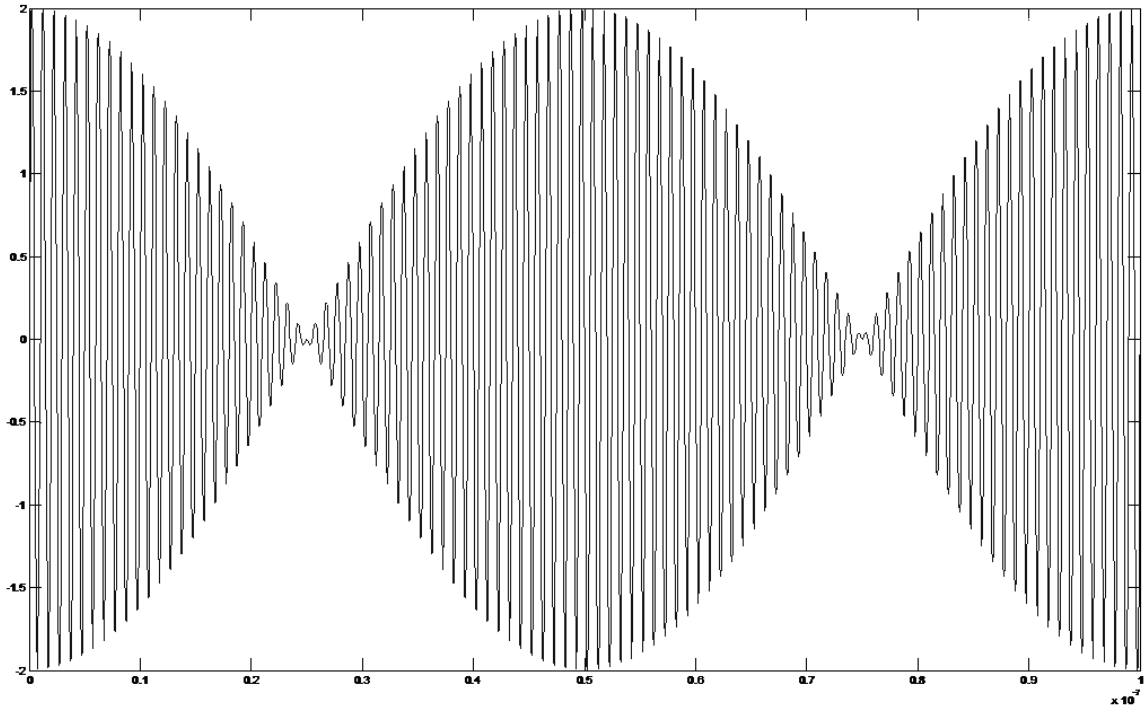

Figure 3-5: The sum of two sine waves with slightly different frequencies

3.3.1 Shifting up

A solution to amplifying (3-5) with devices that can only sink current is shifting the entire signal up by adding a signal component at the carrier frequency, such that the modulus is always positive. The corresponding equation is (3-6), figure 3-6 shows the waveform.

$$y = \sin\big((\omega_{carrier} - \omega_{base})\, t\big) + \sin\big((\omega_{carrier} + \omega_{base})\, t\big) + 2\sin(\omega_{carrier} t) \qquad \textbf{(3-6)}$$

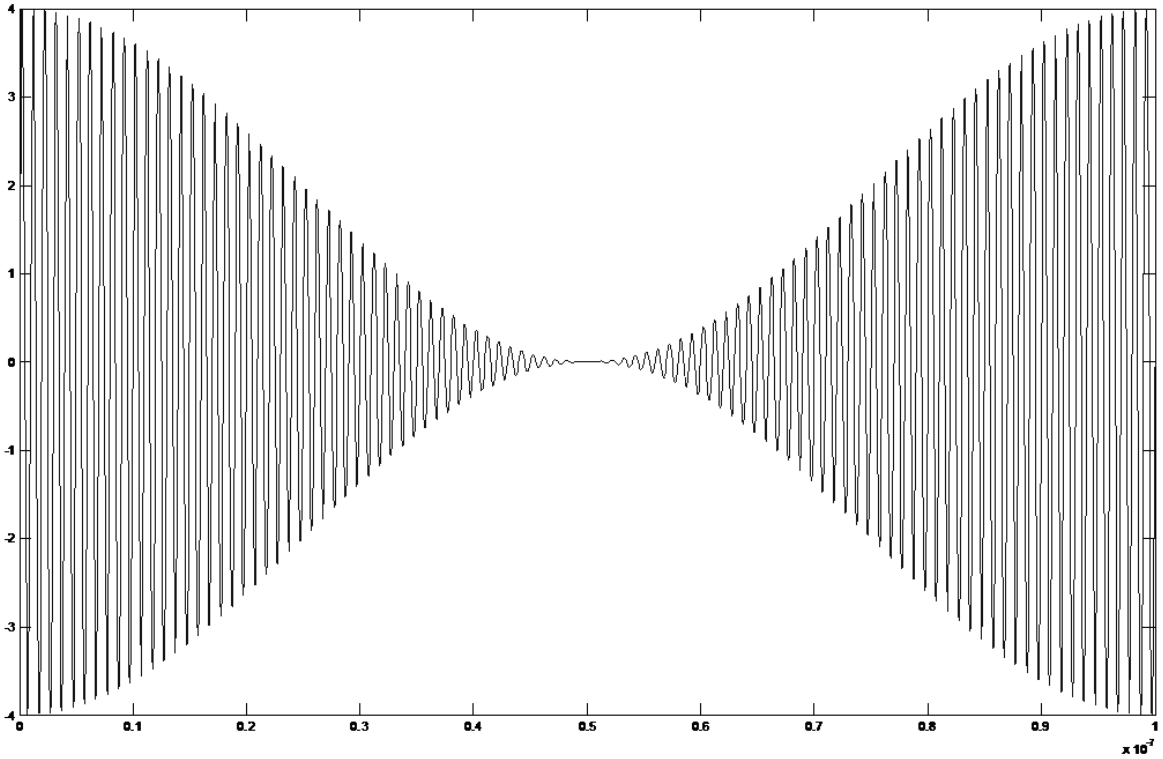

Figure 3-6: The sum of two sine waves with a carrier signal added

Now the modulus does not have zero-crossings anymore, and it can be created straight forward. However the PA now also transmits at the carrier frequency, which holds no information, so only waste energy. This component needs to be as large as the two original sine waves combined; this lowers efficiency with a factor two, which is unacceptable.

3.3.2 Differential

A solution to the problem from the previous section would be working with a differential setup, shown by figure 3-7. When the two halves of the differential circuit get fed an opposite modulus signal, the two sine waves will appear differential at the output while the carrier is a common mode signal.

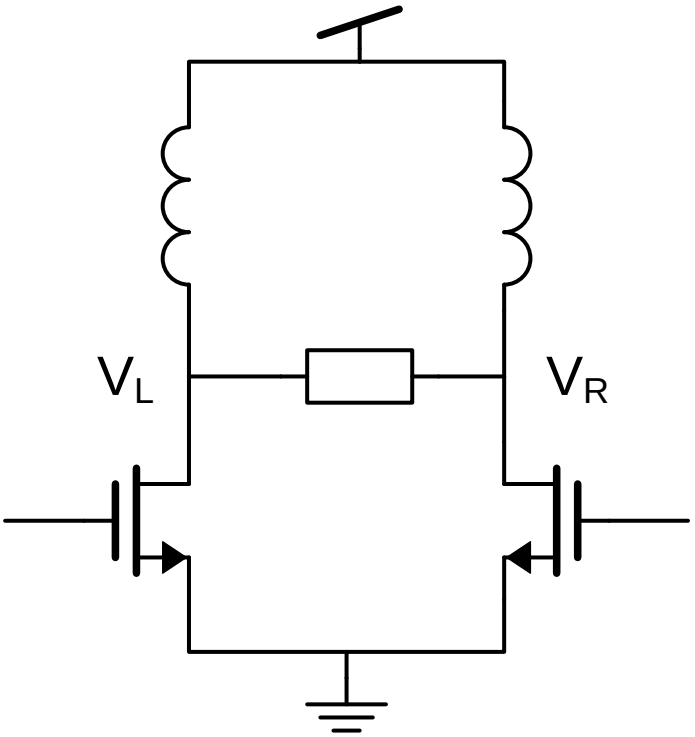

Figure 3-7: Differential circuit

The output voltages are still similar to equation 3-6. However now at one side the modulus signal is inverted, so the voltages become:

$$V_L = \sin\big((\omega_{carrier} - \omega_{base})\, t\big) + \sin\big((\omega_{carrier} + \omega_{base})\, t\big) + 2\sin(\omega_{carrier} t) \quad \textbf{(3-7)}$$

$$V_R = -\sin\big((\omega_{carrier} - \omega_{base})\, t\big) - \sin\big((\omega_{carrier} + \omega_{base})\, t\big) + 2\sin(\omega_{carrier} t) \quad \textbf{(3-8)}$$

The output voltage is the difference between those two, so the carrier component is removed while the wanted signal is kept.

$$V_{diff} = 2\sin\big((\omega_{carrier} - \omega_{base})\, t\big) + 2\sin\big((\omega_{carrier} + \omega_{base})\, t\big) \quad \textbf{(3-9)}$$

In principle this removes the carrier signal without dissipating it. However there is a fundamental problem. Since the carrier is a common mode signal, it does not see the load. So only the transistor and the inductor are relevant for the common mode signal.

As discussed before the transistor effectively operates as a voltage source in the on-state. The voltage always has to be lower than the supply voltage in its on-state, since the transistor can only sink current. This means the current through the inductor will increase when the transistor is on. In the off-state the transistor acts like an open switch, so the current in the inductor at the carrier frequency cannot go anywhere: it does not see a load. This would then result in a huge voltage spike, which is unwanted. Two solutions would prevent this.

- The easiest one would be making sure there is also a load that is seen by a common mode signal. While this would easily solve the problem, it would solve it by dissipating the carrier, which gives the same problem as transmitting the carrier.
- The other solution would be storing the energy on a capacitor while the transistor is in its off-state: simply by making an LC-tank. This works, for a short time period. It stores the energy, so in the beginning of the next cycle it has the same current going through the inductor as in the end of the last cycle. It will during the half of the cycle where the transistor is on

integrate current again. So every cycle the current through the inductor increases, which also means an increased voltage across the capacitor that forms the LC-tank with the inductor in the off-state. This will keep increasing until it becomes high find an alternative path, for example by exceeding the maximum voltage the transistor can survive.

3.3.3 Phase of the switches

The output signal of the circuit is the modulus signal, which is the charge on the injection capacitor, multiplied with the switch signal. So instead of making the modulus signal negative, it is also possible to obtain the same result by inverting the clock signal to the switches. This is similar to what is done in the envelope elimination and restoration (EER) technique [**4**], the difference is that EER uses an RF input and splits that into a phase and an envelope part, to recombine them later, while here we start with a phase and an envelope signal.

If the modulus signal is a sine wave, we can represent it by:

$$y_{mod} = A_{mod} * \cos(\omega_{mod} t + \phi_1) \quad \textbf{(3-10)}$$

This signal is multiplied by a square wave due to the switches. Since the output is filtered, and there are no high frequency components present in the baseband signal that could be mixed into the output band due to harmonics of the switching signal, the switching can be seen as:

$$y_{switch} = A_{switch} * \sin(\omega_{switch} + \phi_2) \quad \textbf{(3-11)}$$

There is also a DC component present in the square wave, but also this does not generate an output in the required band, so it can be ignored.

These two multiplied with each other give the (ideal) output. For simplicity ϕ_1 is set at zero. The resulting signal around the switching frequency, which is the part of the signal we are interested in, is given by (3-12).

$$y_{out} = A_{mod} A_{switch} * \frac{\sin(\omega_{switch} + \phi_2 + \omega_{mod}) + \sin(\omega_{switch} + \phi_2 - \omega_{mod})}{2} \quad \textbf{(3-12)}$$

When the modulus is positive ϕ_2 will be zero. If the modulus becomes negative the absolute value of the modulus is taken, so it is effectively multiplied by minus one. At the same time ϕ_2 is changed to 180 degrees. Any sine wave shifted 180 degrees is equal to minus one times the original [**5**]. So now both the modulus and the switch signal are multiplied by minus one. Since these two are again multiplied with each other, the resulting signal is the original signal. However now the modulus signal does not contain zero crossings anymore, since the absolute value is taken.

However there remains one major problem: the circuit operates correctly when it has a large voltage gain from input to output, when this gain is low it does not suppress nonlinearities. Around the zero crossings the current through the transistor is very low, which means that the transconductance is very low, so the voltage gain is also low.

The resulting distortion will be mainly second order distortion, since this is the dominant distortion for low output currents. However due to the nature of the distortion, it is symmetric for both halves, a lot of IMD products are generated even with only second order distortion present. The distortion products for a transistor with only second order distortion and with everything else ideal is shown in figure 3-8. In figure 3-8 f_c is the carrier frequency and f_b is the baseband frequency

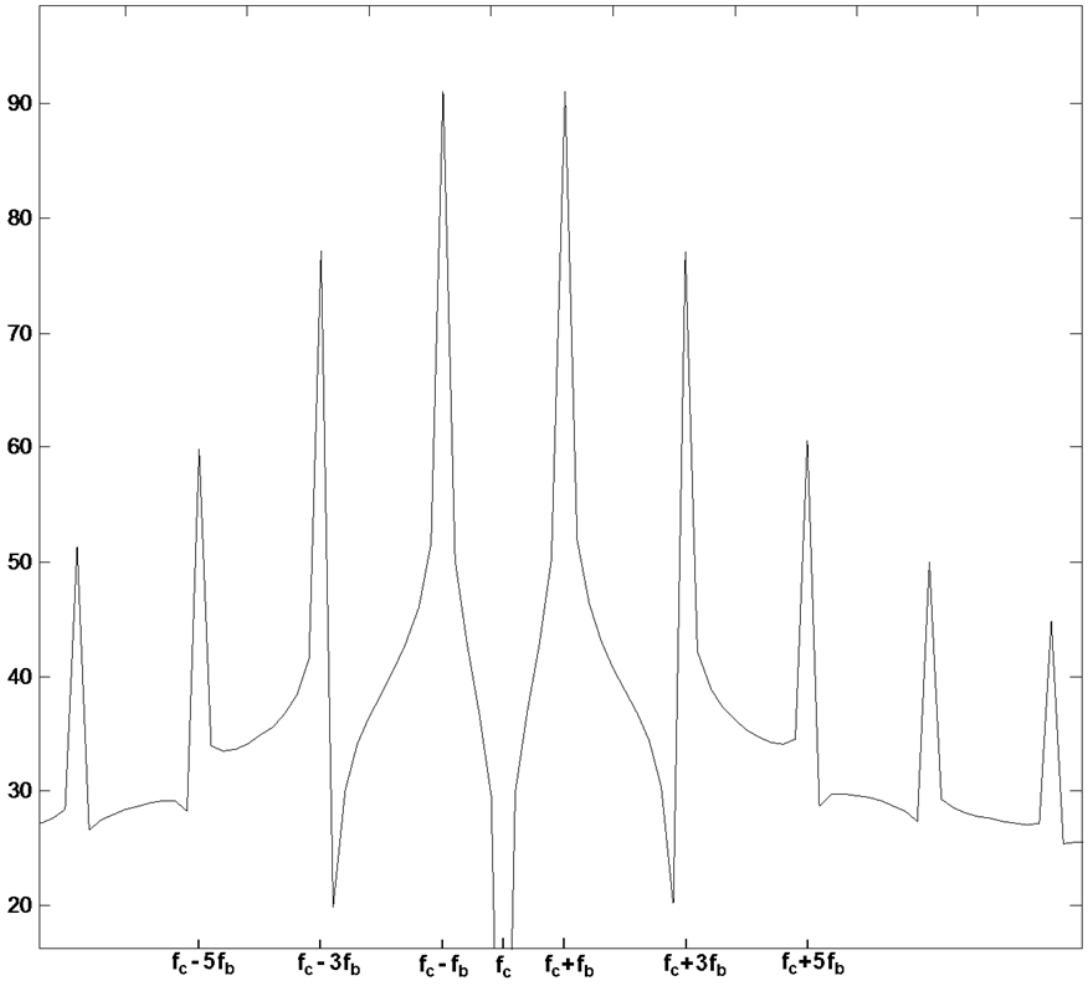

Figure 3-8: IMD products with zero crossings and only 2nd order distortion in the drain current

This problem could be solved using an additional amplifier in parallel to the main amplifier. This amplifier should behave as a current source, contrary to the main amplifier that ideally behaves as a voltage source.

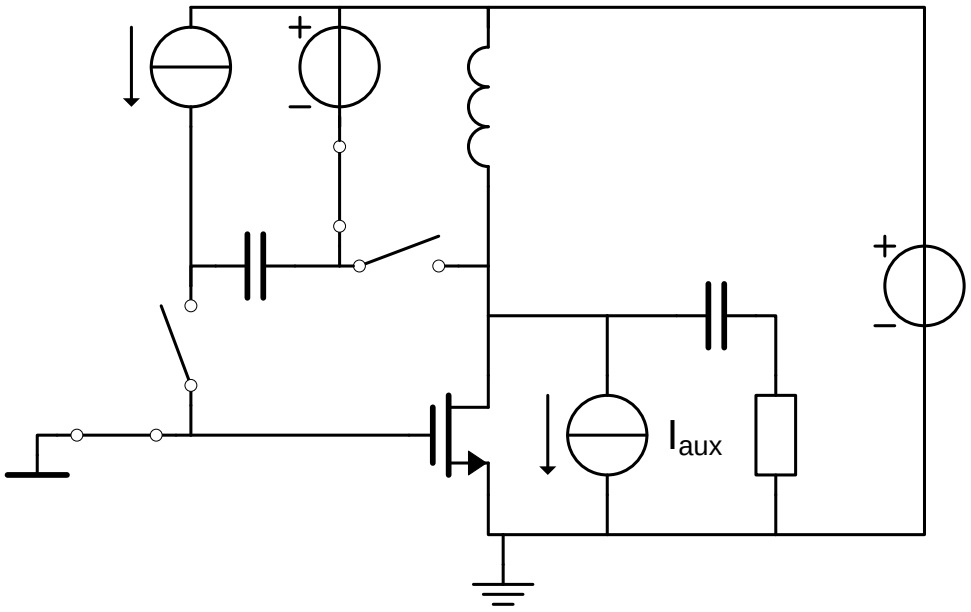

Figure 3-9: PA with auxiliary amplifier

For low output voltages the auxiliary source must produce a current similar to what the main amplifier would produce in the ideal situation. Since the output current is low, a conventional amplifier can reach a good linearity.

For higher output voltages it does not matter what the auxiliary amplifier does exactly, since with a high output voltage the main amplifier acts as voltage source, so the current source does not influence the output. So for higher output voltages the current source can be limited, so it has a low maximum current output.

Another solution would be increasing the effective transconductance. This can be done by placing another amplifier directly in front of the main transistor. This amplifier created additional voltage gain, resulting in effectively a higher transconductance.

However this removes the main advantage of the used method for feedback: being very fast because only intrinsic capacitances and one capacitor switched parallel are used. An extra amplifier will add at least one pole, and simulations show that it easily results in oscillations, so this is not a viable option.

3.3.4 Conclusion

Despite having disadvantages, the only real option for removing the carrier is using zero crossings of the modulus and inverting the switch signals when required. The main problem is the distortion that especially for low output signals will be present.

However while definitely present, simulations in section 4.5.4 indicate that another mechanic, charge injection, is the dominant cause for distortion for low output values. Since this is dominant and independent of feedback, it is not required to add extra elements to cancel problems due to low feedback at low output values.

3.4 Steering

To get the required output signal the injection capacitor needs to be at a certain level of stored charge. Two effects need to be taken into account to know how to steer the input current source.

- First there is the capacitive behavior. When the charge on the capacitor needs to be increased, a current is required to increase it, like on a regular capacitor.
- There is also a resistive factor. Every time the injection capacitor is placed parallel to the transistor it loses a part of it charge to the intrinsic capacitances of the MOSFET. This goes ideally linear with the amount of charge on the injection capacitor. So every cycle it loses a certain amount of charge. But since that happens at a very high frequency, looking from the low frequency current source it behaves like a resistor, just like a switched capacitor network can behave like a resistor.
  This is useful behavior, otherwise it would be hard to make sure the DC level cannot walk away, the resistive behavior makes sure the current source can also determine the DC level.

Effectively for the current source this means it is driving a capacitor and resistor in parallel. To get the correct voltage on them, it needs to differentiate the baseband input to get the current required to charge the capacitive part, added with a linear term for the resistive part.

Figure 3-10 shows all the required components for the steering. These are not part of the amplifier itself, so they are implemented using ideal components. In an actual implementation they are mainly digital components.

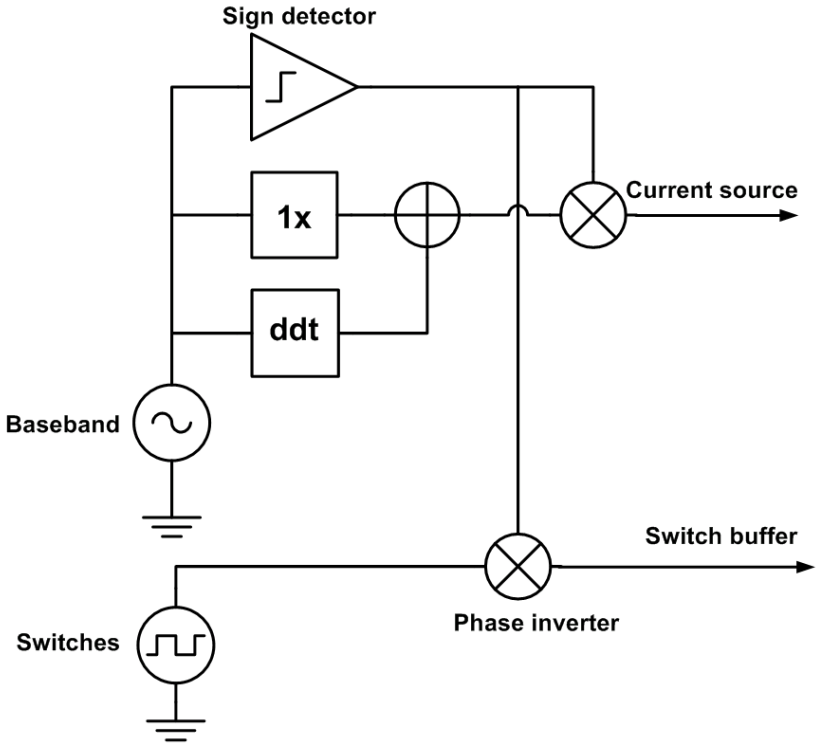

Figure 3-10: Required steering components

The sine source generates the baseband source. As explained there is a linear term and a differentiated part. A sign detector detects if the signal is above or below zero, if it is above both the switch signal and the baseband signal are sent unchanged to the amplifier. If it is below zero the baseband signal is inverted and the switch signal is shifted 180 degrees.

3.5 Complete amplifier

An output filter is added to the design, which is required to increase the performance; mainly efficiency. Also a buffer for the switching signal is required. With these components included the complete top level design is known.

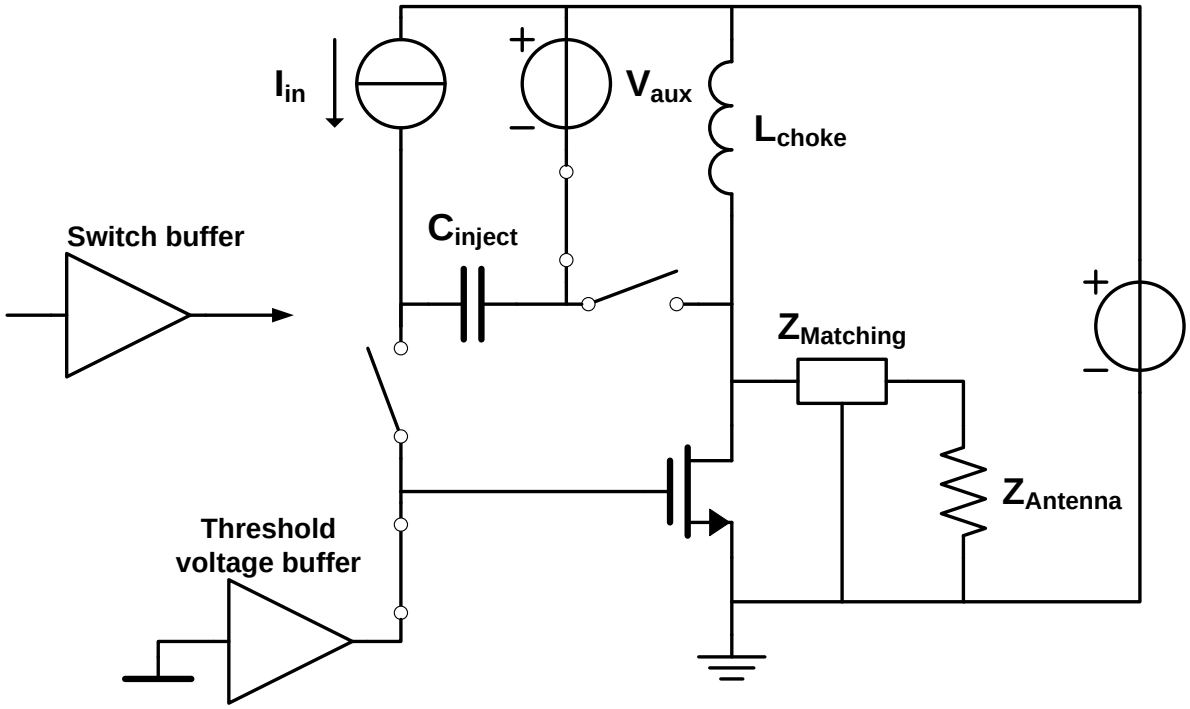

Figure 3-11: Complete design of the PA and upconverter combination

Section 3.1 discussed why the injection capacitor should be placed between the gate and the drain. The required differential current steering was determined in section 3.2, where also the need for the threshold voltage buffer was discussed.

4 Design

The design of the power amplifier consists of several sub-designs, determined in chapter 3. The schematic is shown below.

The different parts that are required follow from this schematic. First there is the amplifying transistor, the core of the design. The switches and injection capacitor take care of the actual steering. The current source is the input, together with the switch signals that need to be buffered. Also the threshold voltage requires a buffer. The output is connected via a load network. All these are designed and discussed in this chapter.

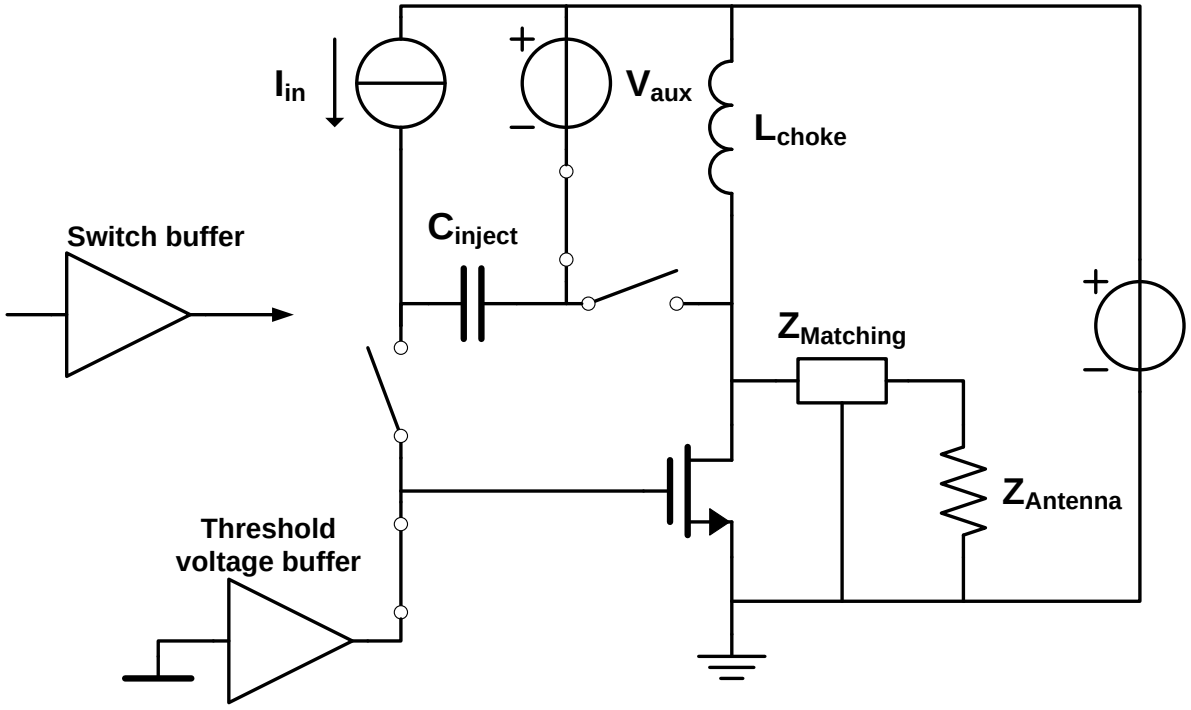

Figure 4-1: Complete design of the PA and upconverter combination

The amplifier is designed for a carrier frequency of 1GHz. This value has been chosen as a reasonable frequency, since it is not designed for a specific application.

4.1 Amplifying transistor

The amplifying transistor needs to be a thick-oxide device to handle the output voltages that can easily become more than 1.2V. To obtain high switching speeds the length should be minimum length, which is 0.28μm in our 65nm technology. The total width is set at 4mm, 100x40µm; there is not a specific power that needs to be delivered, so the width is not very important.

These dimensions are used combined with a gate-source voltage of 700mV and a gate-drain voltage of 600mV to find the resulting transistor properties. These voltages are chosen since they will be close to a real operating point. The exact values are not required, since normally they will have huge variations during operation of the amplifier. The resulting values are:

Table 1: Transistor properties

Threshold voltage	±500mV
Transconductance	450mS
Output impedance	34Ω
Gate-source capacitance	4pF
Gate-drain capacitance	2pF
Drain-source capacitance	1.6pF

4.1.1 Output impedance

Equation 2-9 gave the output impedance of the circuit depending on transistor properties:

$$Z_{out} = \frac{r_{out}(C_{GS} + C_{GD})}{C_{GS} + C_{GD} + C_{GD} r_{out} g_m} \quad \textbf{(4-1)}$$

This allows us to calculate the output impedance of the circuit given bias conditions. This is done for the used transistor at the previous described operating point.

The injector capacitor needs to be added to C_{GD}, since it is parallel when the transistor is in its on-state. In section 4.2 it is set at 10pF. Entering this in the equation for the output impedance yields:

$$Z_{out} = \frac{34 * (4 + 2 + 10)}{(4 + 2 + 10 + (2 + 10) * 34 * 0.45)} = 2.73\Omega \quad \textbf{(4-2)}$$

4.2 Injection capacitor

A larger size of the capacitance that is switched between the gate and the drain of the transistor results in more feedback and a more linear gate-drain capacitance. However it also means more current is required at the input, and a larger portion of the feedback goes via switches instead of directly through the intrinsic MOSFET capacitances. This decreases both the efficiency and indirectly also the linearity. Efficiency is obvious, since the larger steering current costs power. The linearity decreases because this current goes via the injection capacitor to the load, and this current contains many harmonics. The amplifier will try to compensate for this due to the feedback, but it will always add harmonics.

The capacitor, C_{inject}, is set at 10pF. This means the gate-drain capacitance becomes the dominant capacitance, and is linearized by the injection capacitor. Going higher than this will not give many benefits.

4.3 Input current source

The input current source steers the amplifier and hence must have a high linearity. As explained in section 0 the amplifier can be seen as a resistor and capacitor parallel from the point of view of the current source. It needs to be able to discharge this capacitor. This means the input source needs to be able to both source and sink current, otherwise discharging can only be done by the exponential RC behavior.

An easy way to get a well defined linear current source is by using a current mirror. This means the baseband input of the circuit is also a current and not a voltage, but since digital to analog converters also normally have a current output, it can be directly connected to the PA.

When the auxiliary voltage source works well the output voltage of the current source will be between threshold and the gate voltage required to pull the voltage at the output to ground. This is roughly between 0.5V and 0.9-1.0V. This would leave only very little headroom for a current source

that delivers current from a 1.2V supply. However since the design already requires a 2.5V supply, it might as well be used here also, creating ample of voltage headroom for a current mirror.

A current mirror can only source current and here it also needs to sink current. Adding a current source that sinks current at the output can work, but this always dissipates power from the 2.5V power supply. If an NMOS current mirror to ground is placed, the current source is both able to source current from the 2.5V supply and to sink current to the ground. How much the NMOS mirror sinks depends on their gate voltage, which increases if they need to sink more. Figure 4-2 shows this.

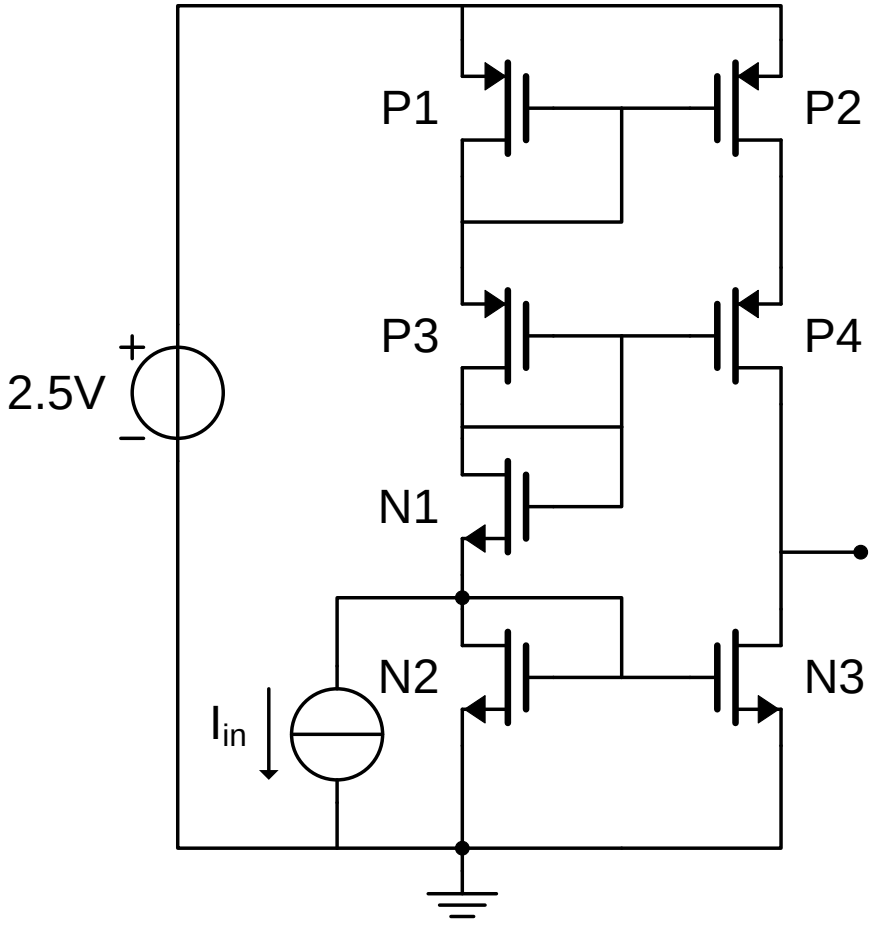

Figure 4-2: Complete input current source

Transistors N2 and N3 are able to sink current. This happens if the current from the input source becomes negative. When this happens the cascoded current mirror on top does not need to supply current anymore, resulting in an increase in gate voltage of N2 and N3, forcing them to start conducting. Transistor N1 is used to lower the voltage between the cascoded PMOS current mirror and the input source and the two NMOS transistors

W-scaling between the left and right half of the circuit is used to decrease the current flowing through the left half of the circuit. This has two advantages, it decreases the power consumption, of the amplifier itself, and it lowers the required steering power. The downside is that the smaller input current needs to charge and discharge the larger capacitors of the right half side. The signal is mainly baseband, so this is not a problem. However only at zero crossings of the modulus, the signal has a rapid change, the bandwidth must be sufficient to accommodate this.

4.3.1 Dimensions

The transistors at the left hand side of the current source have a multiplier half of those at the right hand side. So a 2x current gain is present, increasing the efficiency and decreasing the required input power. Larger current gains are possible, but then the linearity is affected, mainly at the fast current changes that include high frequency components, with a larger current gain a smaller input current has to charge all the parasitics.

The two bottom NMOS transistors are long; this gives the required high output impedance. The PMOS transistors can be shorter, since they are cascoded. They need to have a large enough width to provide a sufficiently small input impedance. Transistor N1 is dimensioned to provide the required voltage drop.

Transistors P1 and P3 are 5x20µm/0.1µm, while at the right side they have a two times larger multiplier, so 10x20µm/0.1µm. This results in an output impedance of 800Ω with a transconductance of 20mS for the right hand side, around their operating point of -600mV gate-source voltage. The cascode structure has a total output impedance of 13kΩ.

Transistor N1 has a size of 10x20µm/0.06µm; with these dimensions it has a voltage drop of 0.5V. N2 and N3 have sizes of respectively 4x5µm/1µm and 8x5µm/1µm. This results in a bias current of 1.2mA combined with an output impedance of 14kΩ for the right hand side of the circuit.

Top and bottom combined give an output impedance of the complete current source of 6.7kΩ.

4.3.2 Behavior

Figure 4-3 shows both the actual and the ideal output current for a sinusoid baseband signal. The steering circuit generates a waveform shown by the red ideal line. The simulated output current is very close to the ideal line. The simulated current does include a significant amount of small peaks due to switching in other parts of the amplifier, but since this is at a high frequency and the current is integrated by the injection capacitor this does not have a negative influence on the performance of the PA.

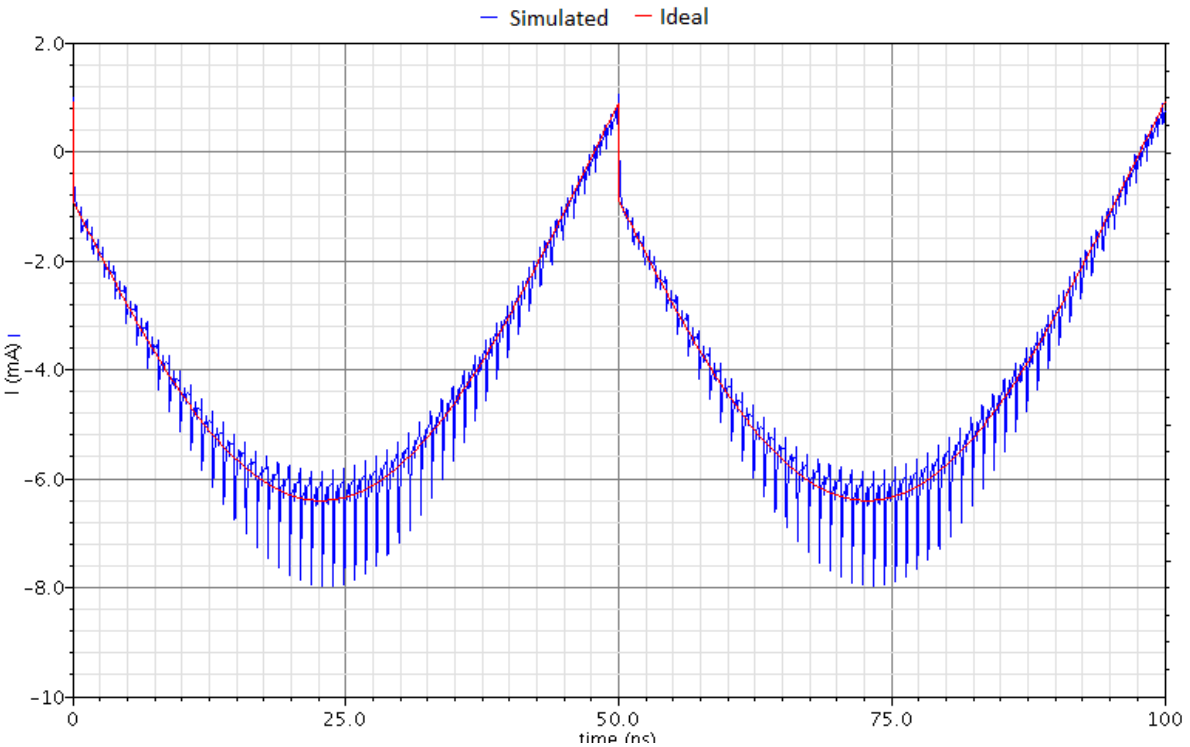

Figure 4-3: Simulated output current of the current source versus the ideal output current for sinusoid baseband signal

One of the main goals of the PA was that it needs to be easy to control. So in the case of a current input, the input impedance needs to be sufficiently low. Figure 4-4 shows the input current and input voltage. From this it follows that the input impedance is a bit under 150Ω. By increasing the transconductance of the transistors in the mirror this can be decreased further.

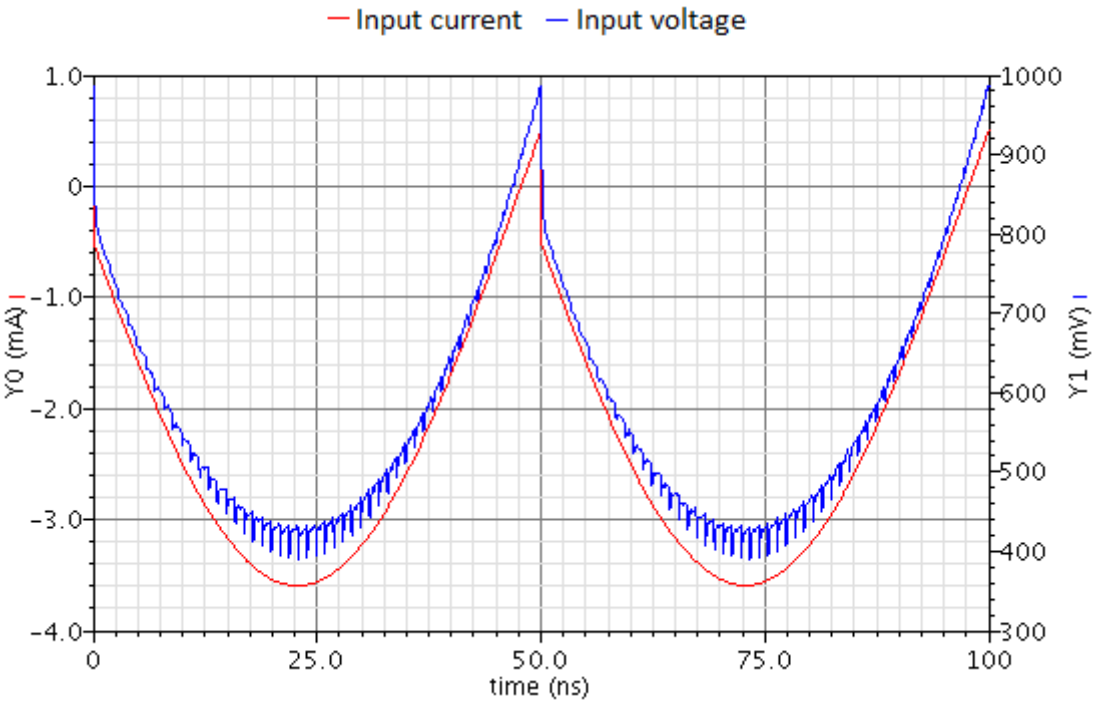

Figure 4-4: Input current and input voltage of the current source

The design effectively contains several transistors in diode configuration stacked on top of each other. So in theory this means a small increase in supply voltage can result in a large increase in bias current. Simulations show that the input current is barely affected by the supply voltage, changes from -20% to +20% from the nominal supply voltage of 2.5V have only very minor influences on the output of the complete amplifier. Figure 4-5 shows the current through the bottom N3, it is clear that it becomes larger with a larger supply voltage. However even at 3V supply it is still much smaller than the output current, which had for this simulation a peak-peak value of 7mA.

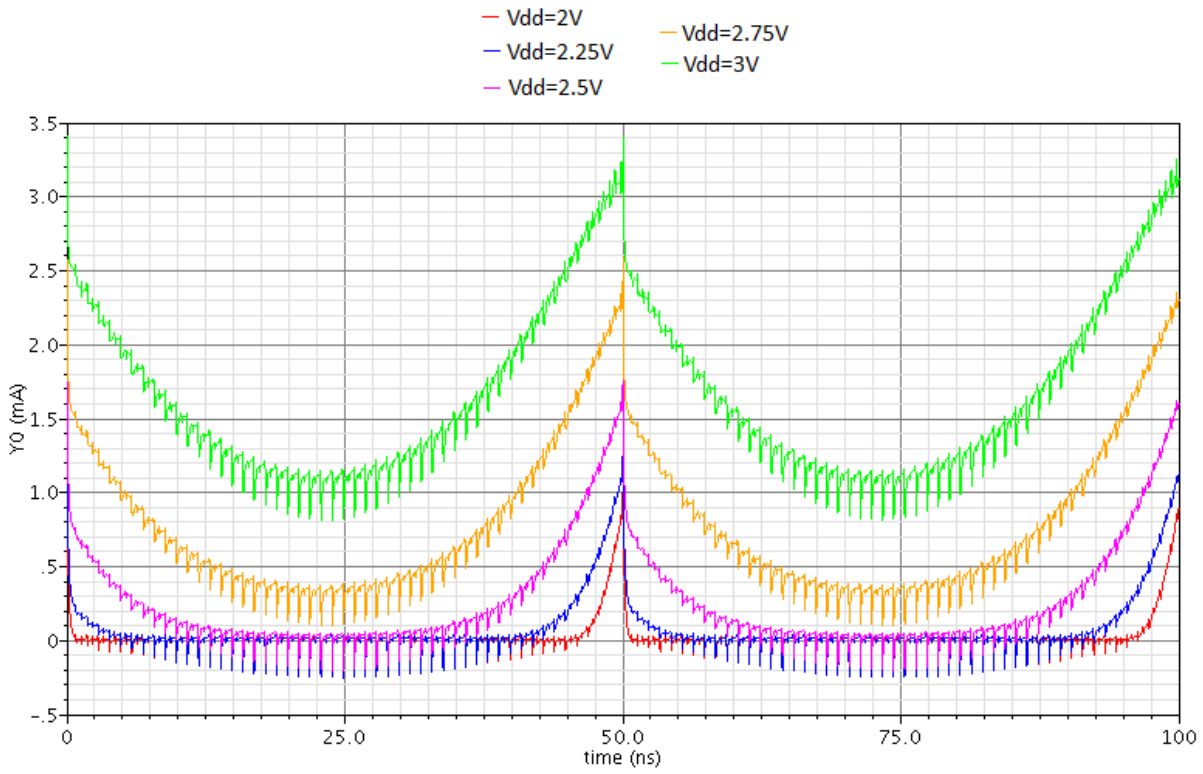

Figure 4-5: Current through N3 for varying supply voltages

4.4 Auxiliary voltage source

The task of the auxiliary voltage source, shown in Figure 4-1, is to make sure the voltage at the current source is stable, to compensate for its limited output impedance. So ideally this voltage source has a value exactly equal to the output of the amplifier when it is in its on-state. An additional demand is that it may not use special inputs, only the already available input signals can be used.

A simple voltage source can be made using a common-source amplifier. This does need to get a steering voltage though, and it is only allowed to use the already available signals, which is a problem. Another problem is that it does need to be able to handle at least a few milli-amperes, which requires a small drain resistor, which means high power consumption.

The alternative is using the output of the PA. This creates the required voltage half the time. A capacitor can store this voltage during the other half, and after this it can be charged to the right voltage again by the PA. Figure 4-6 shows this in the circuit. This also means one switch can be removed, since the capacitor always needs to be connected to the injection capacitor.

When the transistor is in its on-state this capacitor is connected in parallel with the output, so it is charged to exactly the required value. The other half of the cycle it is in series with the injection capacitor. From this it follows that to be a good voltage source this capacitor needs to be significantly larger than the injection capacitor. In principle it is counter-intuitive to put not only an extra capacitor parallel to the output, but also a very large one. However since it is only parallel one half of the cycle, it will barely affect the output, only the difference in output voltage between two cycles and the charge from the input source needs to be removed from the capacitor every cycle.

A capacitor of 10pF is used, equal to the injection capacitor. A larger value would make it a bit easier for the current source, but requires more chip area. A smaller value than this and the voltage at the current source will start to fluctuate too much, resulting in extra distortion at the output of the PA.

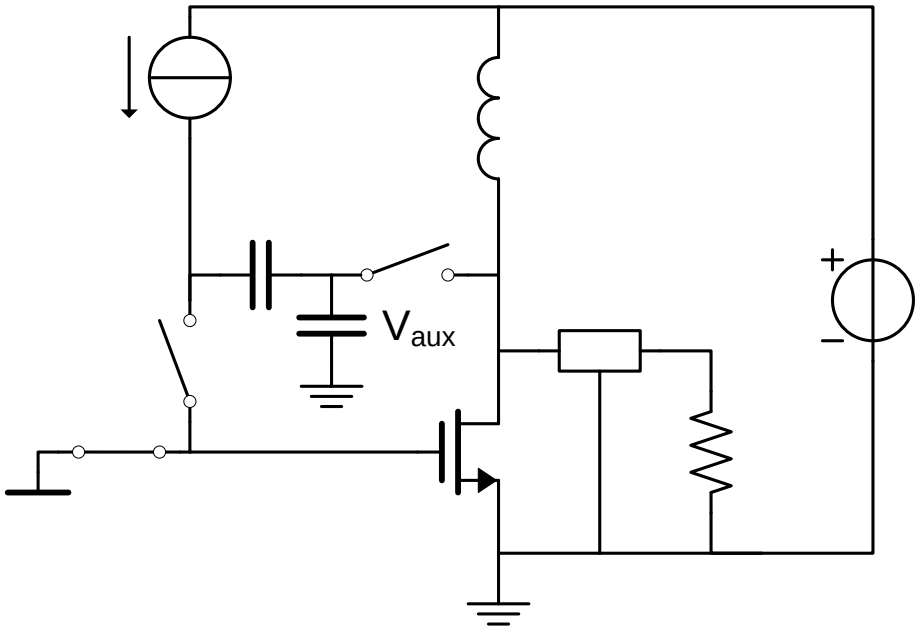

Figure 4-6: Auxiliary voltage source in the circuit

4.5 Switches

No switches should be any larger than necessary. Larger switches have more parasitics and higher power consumption than smaller switches. So where possible the switches need to be minimum length.

The width of the switches depends mainly on the current that needs to flow through them: when the switches are closed the voltage across them should be low, which requires wide transistors. However to limit parasitics they should not be wider than required for good switching. Depending on the voltage they have to switch they can be implemented using NMOST, PMOST or both.

Figure 4-7 shows the names of the switches, these are used in the design and optimizations in the following sections.

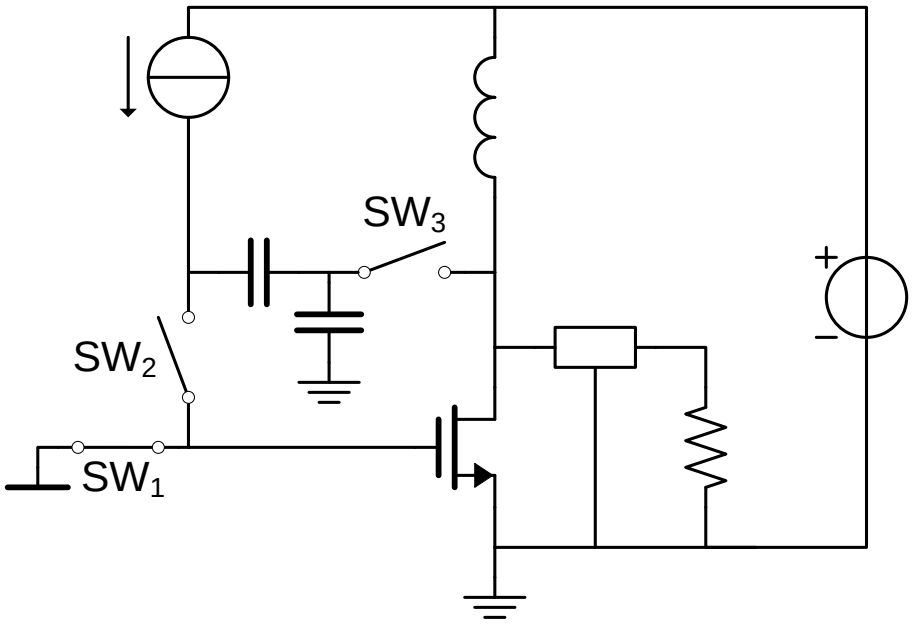

Figure 4-7: Switches

4.5.1 Switch 1

When switch one is closed it will always pull the voltage to a known value, roughly the threshold value. Since this voltage is relative low, an NMOS switch will work much better than a PMOS switch for this. This switch needs to be large enough to force the voltage at the gate quickly to threshold.

In figure 4-8 the gate voltage is shown for different widths of the transistor. Also it is shown with a maximum steering voltage of 1.2V and with a maximum of 1.7V. 1.2V is the normal available voltage, however the switching behavior is not good with this voltage. Since the minimum voltage the switch is connected to is 0.5V, and the maximum voltage that is allowed between two terminals is 1.2V, the switching voltage may be up to 1.7V, so it is also simulated with this voltage.

From figure 4-8 it is clear that a switch steered with 1.7V is much faster than a switch steered with 1.2V. For this reason it is steered with a voltage switching between 0.5V and 1.7V.

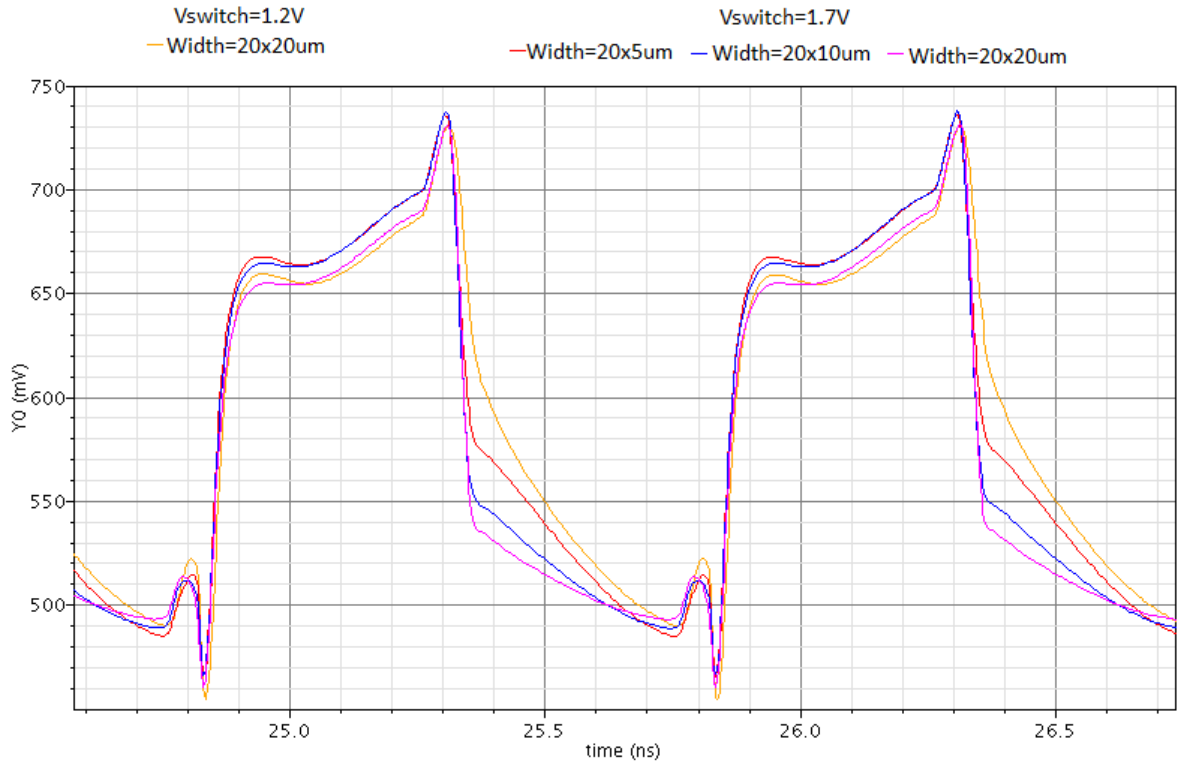

Figure 4-8: Gate voltage as function of the width of switch 1

The used switch is 10μm wide with a 20x multiplier, it is clear from the graph that the transistors turns off a bit slower with this width compared to a larger width, but simulations show that this has no negative effect on the performance. Slower is not wanted though, so this width gives a good mix between performance and power consumption.

4.5.2 Switch 2

Switch two has similar requirements as switch 1, in that is also needs to be able to quickly force the gate voltage to a certain value, dictated by the injection capacitor. However since this can be a wide range of values an NMOS alone will not be enough, so a PMOS needs to be parallel, to be able to operate as a good switch over a wide range of voltages.

The total width of these two transistors will be larger than that of the NMOS, partly because PMOS have the need for an inherent larger width, and partly because the gate switch always has a switch

voltage 1.2V above the voltage it switches to, while this switch will have average overdrive voltages of only 0.6V.

Figure 4-9 shows the voltage across the switch as function of its width, with a multiplier of 40, the width of the NMOS and PMOS devices is equal for now.

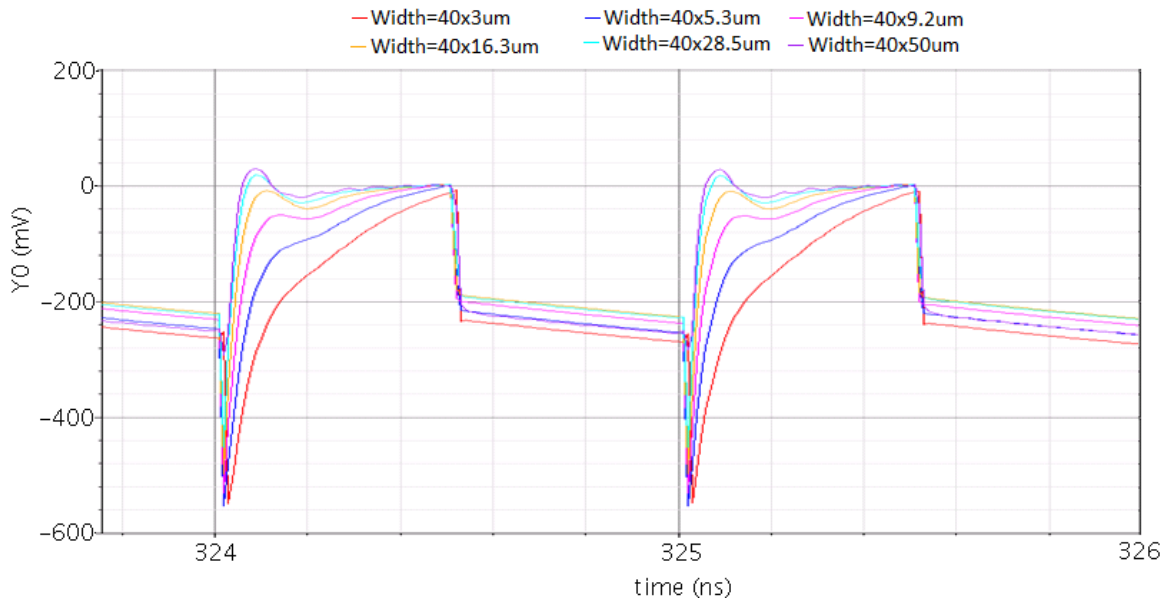

Figure 4-9: Voltage across switch 2

A width of $30\mu m$ is used, from these graphs it looks like a smaller width can also be used, but for these simulations switch 3 was ideal. Since switch 3 is in series with switch 2 and will have similar on-resistance, switch 2 needs to be larger than shown in these simulations to keep the total on-resistance of switch 2 and 3 combined sufficiently low.

4.5.3 Switch 3

Switch three has a large swing; it needs to be able to short voltages between 0V and 1.2V. This would indicate that an NMOS-PMOS pair is required. However considering the large voltages that can arise at the output, up to 2.5V, high voltage devices are used here. Since from the NMOS the gate voltage will then be switched to 2.5V, an NMOS could also do it without aid from a PMOS. However for reasons explained in section 4.5.4.1 the NMOS-PMOS combination is the best option.

The requirements for the on-resistance are similar to that of switch 2. The transistors will be $0.28\mu m$ long instead of $0.06\mu m$ that normal transistors are, so they need to be wider for similar behavior. However the higher overdrive voltage lowers again the required width.

Figure 4-10 shows the voltage across the switch with an NMOS-PMOS pair, the pair has a multiplier of 40 and equal widths.

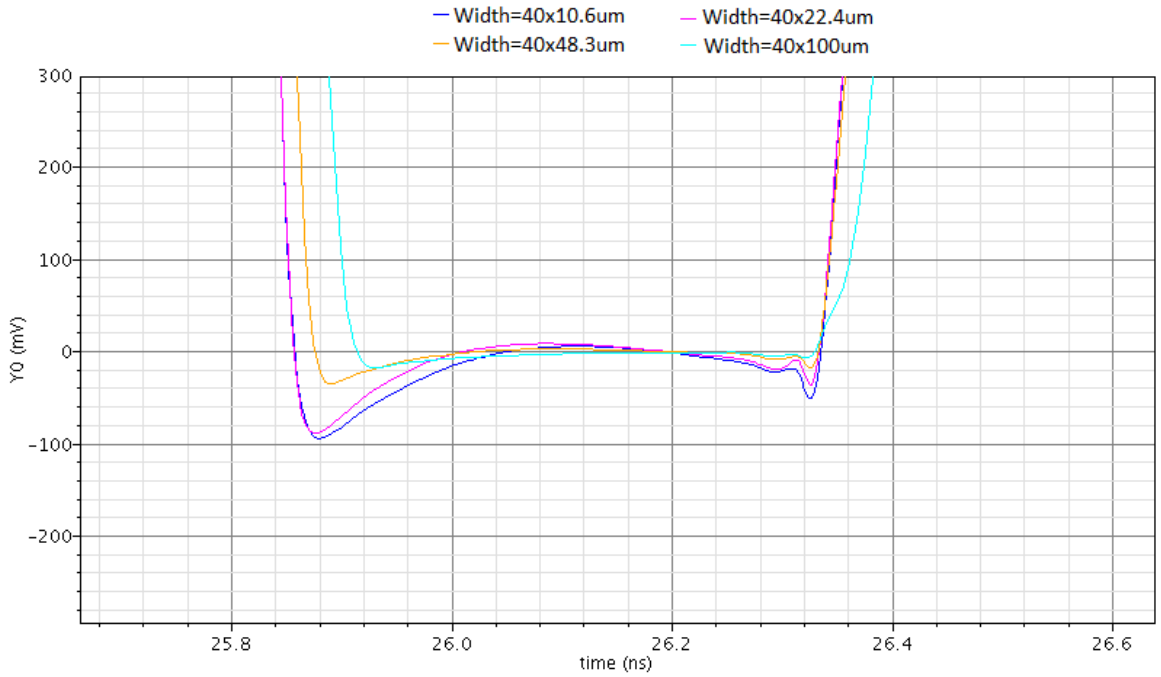

Figure 4-10: Voltage across switch 4 as function of the width, with an NMOS-PMOS pair

The differences in voltage do not seem to be that large between widely varying widths, however they do propagates directly to the gate, where those differences are significant. The transistors are dimensioned at 40x30µm/0.28µm. Simulations show that this is wide enough for a good linearity.

4.5.4 Charge injection due to switches

A problem is charge injections due to the switches. Even without a baseband input signal there is an RF output signal. The gate voltage of the amplifying transistor is shown in figure 4-11, the threshold voltage source here was set at a bit lower value than its default 500mV, but this does not affect the injection.

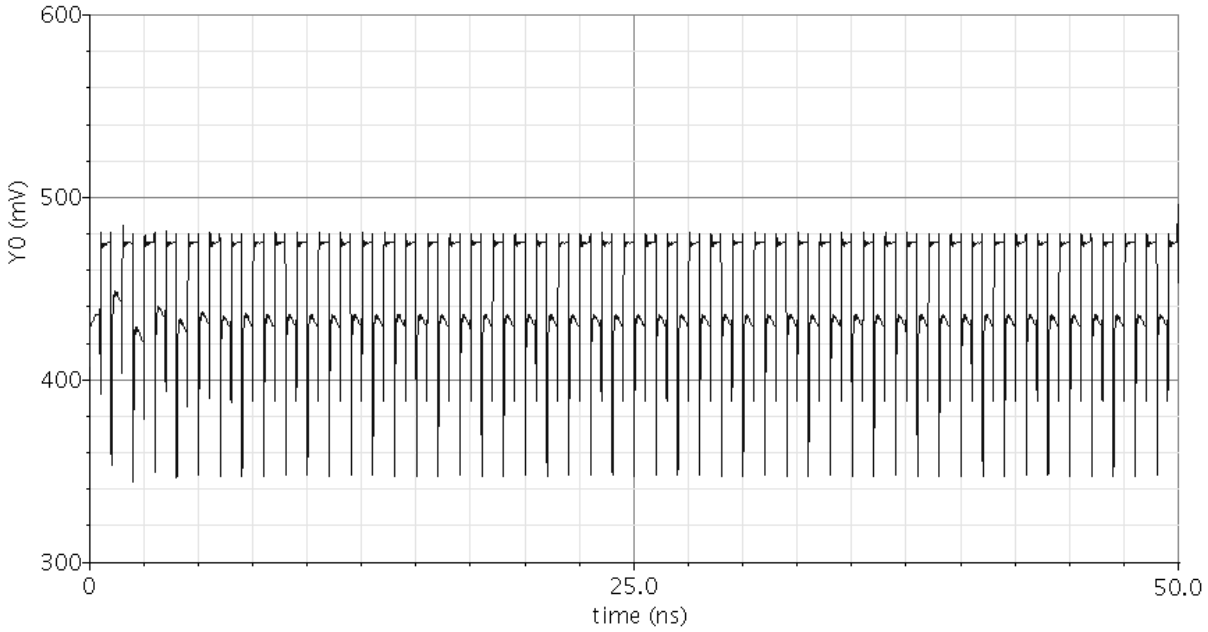

Figure 4-11: Gate voltage without input signal

In the ideal situation the gate voltage should be a flat line. This is important, since now a small input signal would get dominated by the much larger influence of charge injection. This is especially relevant when the input signal has zero-crossings. The phase of the switches will then be inverted, which in combination with the charge injection generates many harmonics.

4.5.4.1 Optimization techniques

A possible solution to counteract the effects of charge injection would be adding a DC current source besides the normal input source that negates the charge injected by the switches. However this would result in two problems: first it will not compensate it exactly, since the switches do not inject a constant current but a train of short peak currents, and secondly because it would be hard to get it at the correct value, independent of process spread.

A good practice is to make the switches as small as possible without influencing the performance for high input signals. This also increases efficiency; however this alone does not give sufficient results.

The problem can be solved by injecting an opposite charge at exactly the switching moment, there are two ways to do this [**6**]; either at both sides of the switching transistor a shorted transistor is of half width is placed that has an inverted control signal, or parallel a PMOS is placed with an inverted control signal. These methods are shown in figure 4-12.

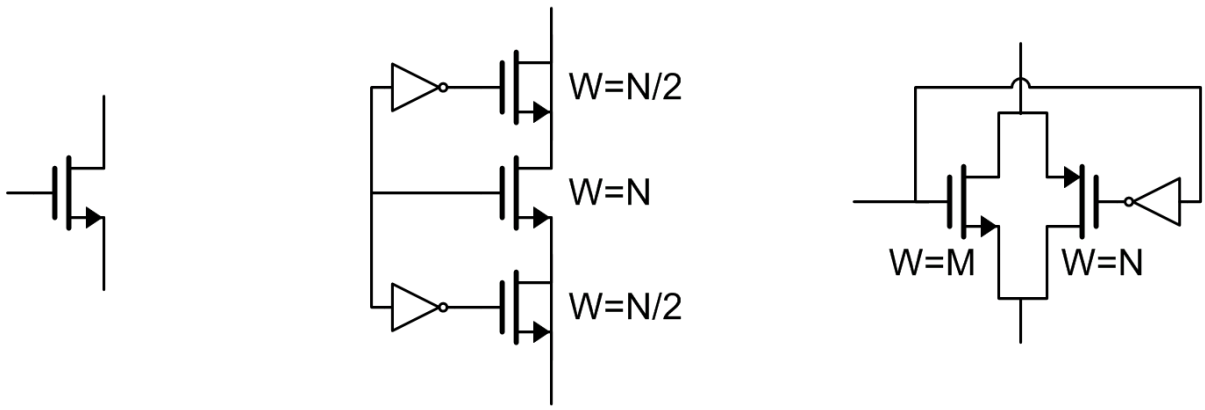

Figure 4-12: Charge injection optimization

The first method is straight forward to implement and does not affect the behavior besides cancelling the injected charge. The second method does require the PMOS to have the correct width for cancellation, the amount of charge injected depends on the gate-source voltage of the transistor, so equal width is not always the correct solution. However the PMOS does help with conducting, lowering the on-resistance.

4.5.4.2 Optimization

Switch one consists of a single NMOS, so it will inject charge. A PMOS can be placed parallel, however it will barely conduct with the voltage it needs to pull the gate to, 0.5V. To inject a charge similar to that of the NMOS it needs to be wider, since its overdrive voltage will be much lower.

The alternative is half width NMOS devices at both sides of the NMOS switch. One side of the switch is connected to a voltage source, in the ideal case this is not affected by charge injection, so the half width NMOS can be removed from this side. Since the choice is now between a PMOS with a large width, or an NMOS with half width, adding a half width NMOS device between the switch and the gate of the amplifying transistor with source and drain shorted is the best option.

Switch two already has an NMOS-PMOS pair. So this can be used to keep the injected charge small. The multiplier of the PMOS is decreased from 40 to 35, which gives for the used operating point a better performance regarding charge injection.

Switch three also has an NMOS-PMOS pair. Their sizes can stay the same; the operating point is such that they cancel each other's injected charge with equal width. This is also the reason why in the previous section the single NMOS variant was not used, to cancel the injected charge it would require two half width NMOS devices at both sides which would not help conducting, while the PMOS does decrease the on-resistance.

Figure 4-13 shows the gate voltage with the optimization added, which is roughly 18dB better than the non-optimized version.

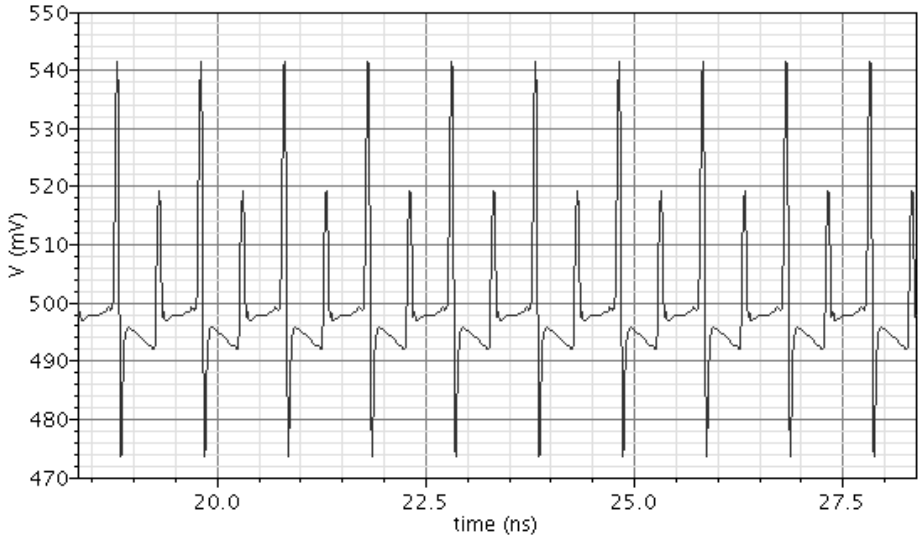

Figure 4-13: Gate voltage with optimized charge injection

4.6 Switch control voltages

The design has three switches; all three require different voltage levels. The control signals for each switch need to be created and they need to arrive at the same time.

	Minimum	Maximum
Switch 1	0.5V	1.7V
Switch 2	0V	1.2V
Switch 3	0V	2.5V

The switch control signals are created from a switch signal that switches between zero and 2.5V. From this the other voltages are made simply by using a thick-oxide inverter that is connected to the minimum and maximum required voltage. After this another inverter is placed for steep edges, since these can be normal thin-oxide devices for all except switch 3 they can be shorter and are faster.

The input of the PA will have a normal 1.2V switch signal, so they need to be amplified to 2.5V first. It seems counter-intuitive to first amplify it to 2.5V and then move it back to 1.2V again. However this is done since going from 2.5V to 1.2V is easier to implement than the other way around. So this way the paths for the different switches stay equal as long as possible, decreasing the amount of problems that can happen due to wrong timings.

Ideally all switches would switch at exactly the same time. However since they have different signal paths, with some slower than others, this will not be the case. For this reason there needs to be some headroom, to make sure it cannot happen that switch one can conduct at the same time as one of the other switches.

4.6.1 Implementation

The generation of the control signals is shown in figure 4-15. It is shown single ended, but in reality it is a differential.

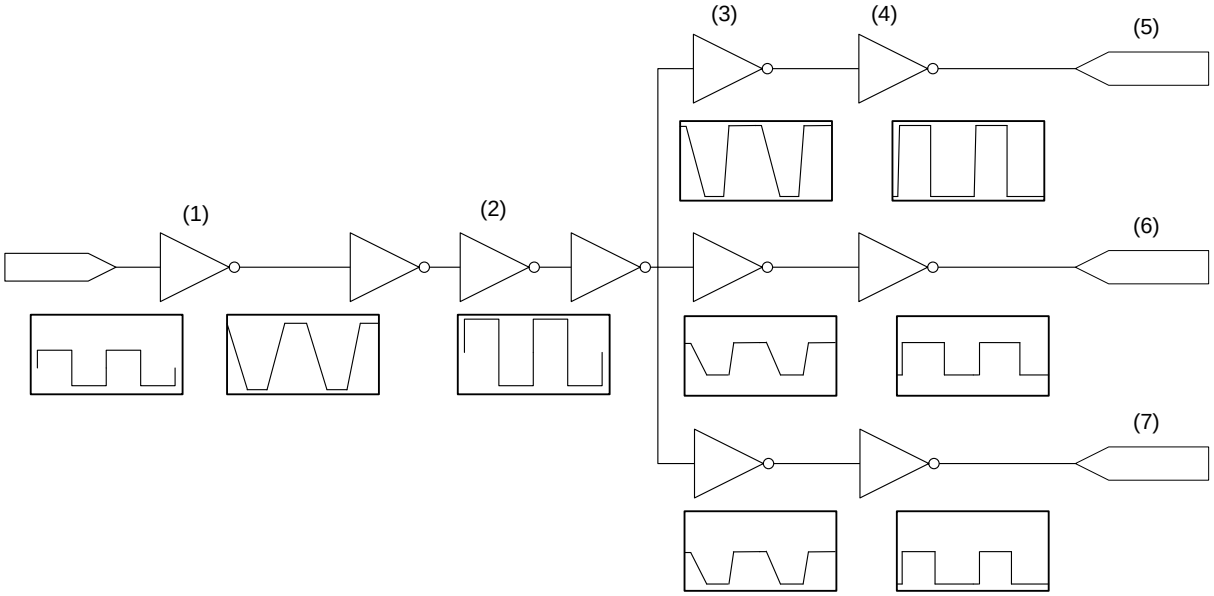

Figure 4-14: Switch control voltages generation

The input of the system is a 1.2V signal. This is immediately amplified to a 2.5V signal by (1). This is done using a normal common-source stage with resistive load, doing it later would require a smaller resistance to reach the required bandwidth, since the transistor dimensions become larger. A smaller resistance means a larger current goes through the common-source transistors when it is in its on-state, decreasing efficiency.

Now the switch signal is at 2.5V three inverters (3) with increasing widths amplify it to get a good block signal and allow for a larger load. The first inverter also is designed to make sure the duty cycle is 50%; this is not trivial since the flanks of the square wave after (1) are not steep.

After this the signal splits into three paths, every path goes to another inverter pair (3). These pairs have different supply voltages: 0V-2.5V, 0.5V-1.7V and 0V-1.2V. The widths of these transistors are such that the output duty cycle makes sure switch one never conducts at the same time as one of the other switches. For example as the width of a PMOS transistor in an inverter is decreased, the output will go slower to supply, decreasing the duty cycle. The required 1.7V is made with a resistive voltage divider between the 2.5V and 1.2V power supplies, followed by a capacitor to supply peak currents.

After this stage there is one more stage of inverters (4) for every signal, to create a signal with steep flanks from the previous signal, since transistors with relative small widths were used to decrease the speed at which a signal goes to ground or supply the edges after (3) are not very steep. This inverter stage has the same supply voltages as the previous ones, however except those switching between 0V and 2.5V they are implemented using regular transistors instead of thick oxide transistors.

This results in the required 0V-2.5V (5), 0.5V-1.7V (6) and 0V-1.2V (7) signals. Since there are only two inverters specific for each voltage level the timing is well defined.

The resulting switch signals are plotted in figure 4-15; all switches have differential inputs, so the differential signal is plotted. A high signal means the switch closed, so conducting.

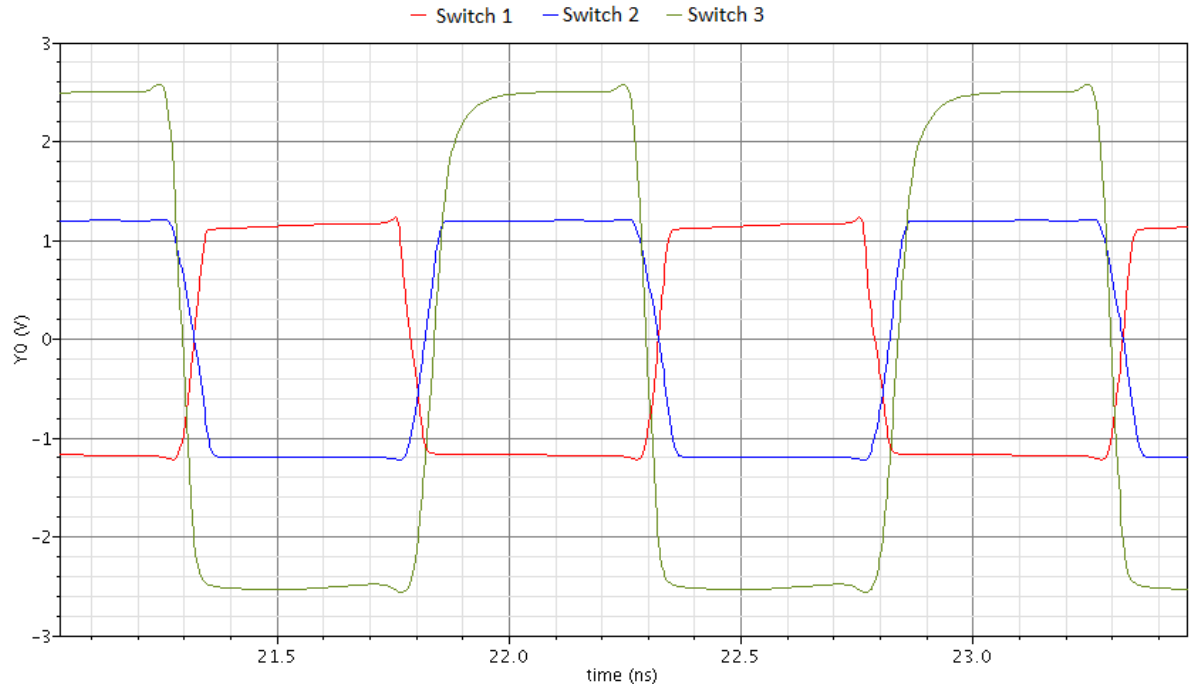

Figure 4-15: Switch control voltages

Switch one starts conducting at the same time when switch two stops conducting, there is very little time between the two. However this does not seem to have a negative effect on the performance. Everywhere else there is plenty of time between opening and closing of switches.

4.7 Output filter

For optimum efficiency an output filter is required. This removes higher harmonics from the output signal and decreases the amount of charge spent on charging and discharging capacitors.

A class E filter is used, this can achieve good efficiency and is used often, its components are shown in figure 4-16.

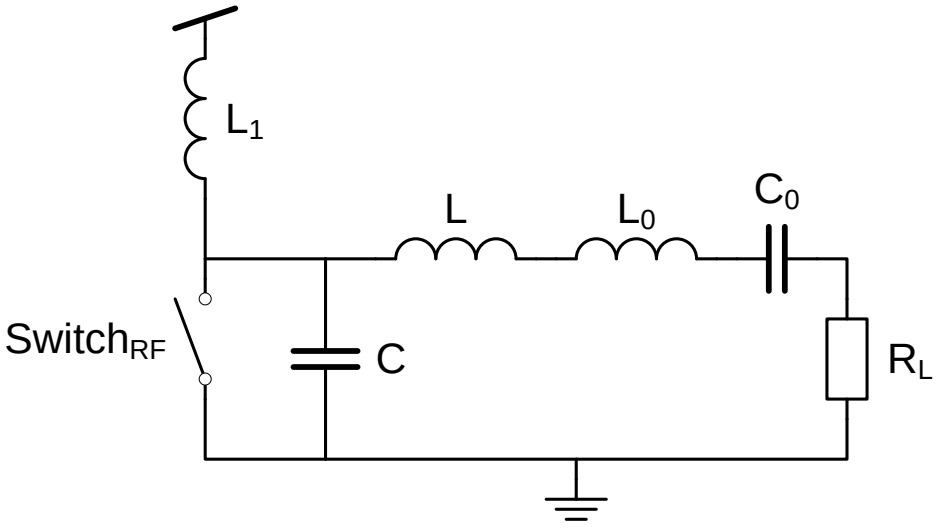

Figure 4-16: Class E output filter

The most important characteristic of a class E filter in regards to the efficiency is its drain voltage waveform, shown in figure 4-17. When the amplifying transistor is conducting the output voltage is determined by the circuit. When the switch is in the open position the voltage will behave as a damped sine wave. It is dimensioned such that when the PA starts to conduct again, which happens at T=1 in figure 4-17, the voltage across the transistor is again at the same value it was during the last on-state. This way the amplifier does not need spend energy charging/discharging its parasitics every cycle.

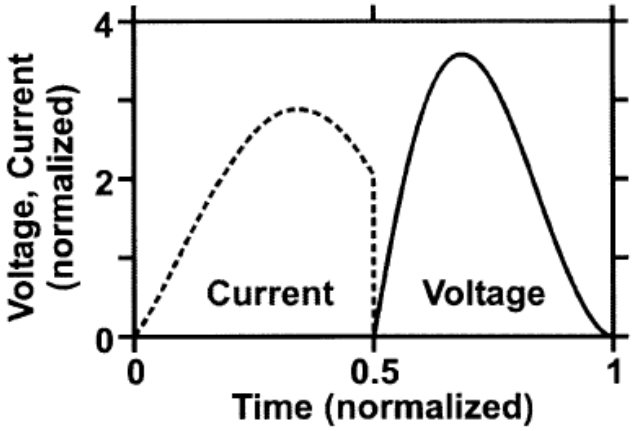

Figure 4-17: Normalized class-E waveform for one period

Inductor L_1 from figure 4-16 is a choke; it acts as current source and allows the voltage to increase above the power supply voltage. A significant part of the capacitance of capacitor C will come from the parasitic of the amplifier. L_0 and C_0 are a series resonance circuit tuned to w_0, which is $2\pi * 1GHz$. They make sure only the required harmonics pass to the load. Larger impedances of L_0 and C_0 provide better filtering, but also require a larger inductor.

Normal class E amplifiers have a transistor that in its off-state is an open connection between ground and the output, while in its on-state it is a short, so it functions as an ideal switch. This amplifier is in its off-state also an open connection, however in its on-state it acts as a voltage source. So for this amplifier figure 4-16 becomes figure 4-18. Since the envelope signal is a much lower frequency than the RF signal, it can be seen as a DC source from the point of view of the filter. All other connections to the ground in the output filter are either capacitors or via capacitors, so the DC value does not matter for them. So the amplifier is effectively modulating the ground voltage, which will have no effect on the behavior of the class E filter. For this reason the class E filter can be used for this amplifier.

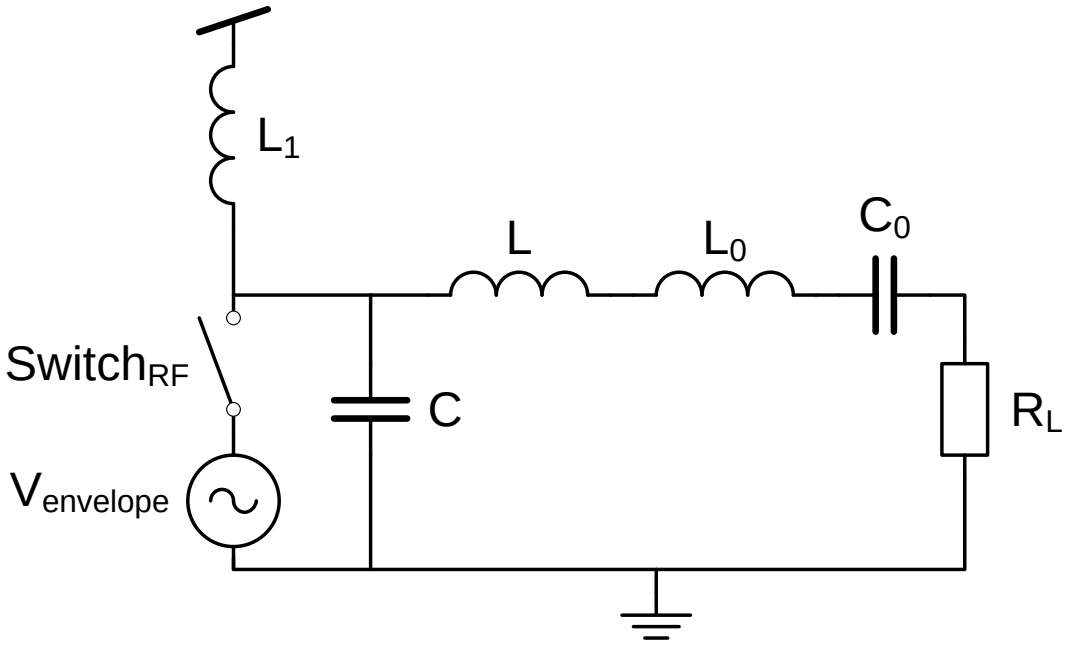

Figure 4-18: Class E output filter including ideal equivalent output circuit of the designed amplifier

4.7.1 Inductors

High quality inductors are required for good performance. There are three possible options for inductors: on-chip, off-chip and bond wires.

- Using on-chip inductors is a preferable option, since it means no off-chip elements are required. However at a frequency of 1GHz the inductors have a quality factor of only roughly 5-10. Simulations show that mainly efficiency degrades with a lower quality factor, so it is not preferred. It also costs a lot of chip area.
- Off-chip inductors can have a very high quality factor, where the series resistance is truly negligible. However they are expensive and a one-chip solution is preferred.
- While not primarily intended as inductors, bond wires do have a significant inductance and a low series resistance, so they are suitable as inductors. As rule of thumb they have an inductance of 1nH per mm combined with a series resistance of 0.125Ω per mm. This results at 1GHz in a quality factor of 50. Normal length will be from less than a mm up to a few mm, so inductors between roughly 0.5nH and 5nH can be created using bond wires. They do not require any off-chip components, and they are required anyway, so they are cheap. The main disadvantage is the range of possible inductances and a relative wide spread in their inductance.

Bond wires are the best option here. The spread in the choke coil will not be a problem, but in the series resonance tank it might be a problem, however this inductor is usually implemented off-chip. A bond wire for the choke coil is easy to implement, since it goes off-chip anyway. It only needs to be of the correct length.

However new techniques might result in future devices having no bond wires, so simulations are done for a quality factor of 10 for on-chip, 30 for bond wires and nearly infinite for off-chip.

4.7.2 Component values

The component values required for the class E filter can be determined using the equations from [**7**]. The equations given there are used to design the output filter. A *q*-factor of 1.1 is used. Together with the operating frequency, the power supply and an output power of 100mW.

The series tank has to suppress higher harmonics, both to help against unwanted radiation and for better class E behavior. A larger inductor gives a better suppression of the harmonics, but requires more space and has a larger parasitic series resistance. Figure 4-19 shows the 2nd and 3rd order harmonics with a 15Ω load as function of different inductor values.

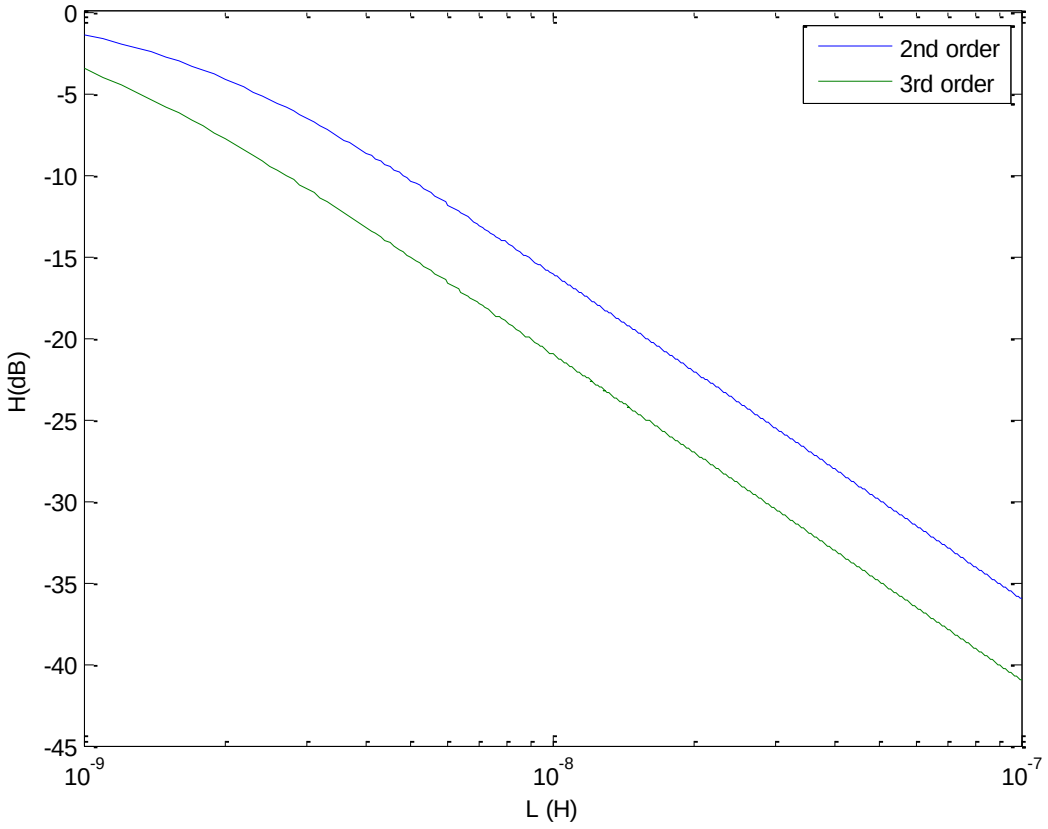

Figure 4-19: Harmonic suppression as function of inductor

Second order harmonics should be small compared to 3rd order harmonics, since they do not exist with a 50% duty cycle. An inductor of 10nH is used, the other values that were determined using the equations from [**7**] are:

Table 2: Output filter

L	1.4nH
L_0	10nH
L_1	4.7nH
C	3pF
C_0	2.5pF
R_L	15Ω

4.8 Threshold voltage source

The last component from figure 4-1 that needs to be implemented is a buffer that keeps the threshold voltage at the required level. This buffer only needs to sink current that comes from the parasitics of the main capacitor when it is discharged to threshold. Ideally it stays exactly at the threshold voltage, but this is not realistic and simulations indicated that variations of a few tens of millivolts are acceptable.

To do this an NMOS transistor can sink the current, while a capacitor parallel removes the current peaks. The NMOS is controlled by an OpAmp with feedback from the drain of the NMOS, so where should be the threshold voltage.

The OpAmp is created with a very basic differential input pair combined with a common-source PMOS amplifier. No compensation capacitors are present: the output, the NMOS, is a current source into a mainly capacitive load. This is the dominant pole in the system, so no other one should be added.

The circuit is shown in figure 4-20.

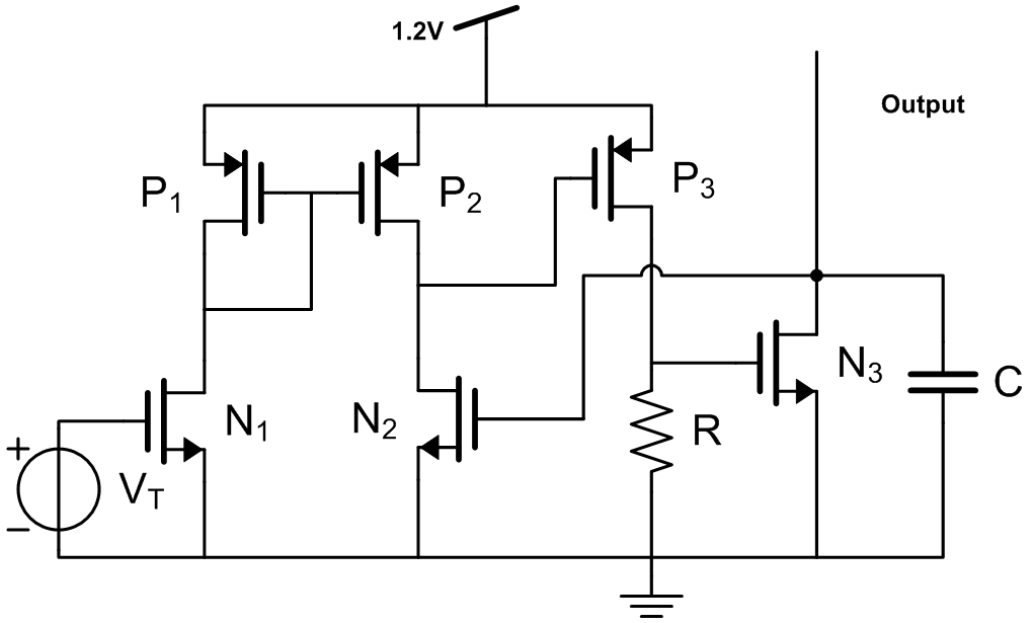

Figure 4-20: Threshold voltage source

The OpAmp is most likely not the best possible design. It is more a proof-of-concept that it works. Transistors N1 and N2 have dimensions of 3x5µm/0.06µm, while P1 and P2 have 2x1µm/0.06µm. P3 is 50% larger than that with 3x1µm/0.06µm. R is 4kΩ and N3 is 10x10µm/0.06µm.

The capacitor needs to be large for correct circuit behavior, since it needs to absorb the current peaks produced by the switching of the amplifier. Large capacitors are also expensive, so it should not be larger than absolutely necessary.

Figure 4-21 gives the resulting threshold voltage with different values for the capacitor. For a good overall linearity of the amplifier a capacitance of 60pF is required, but larger is better.

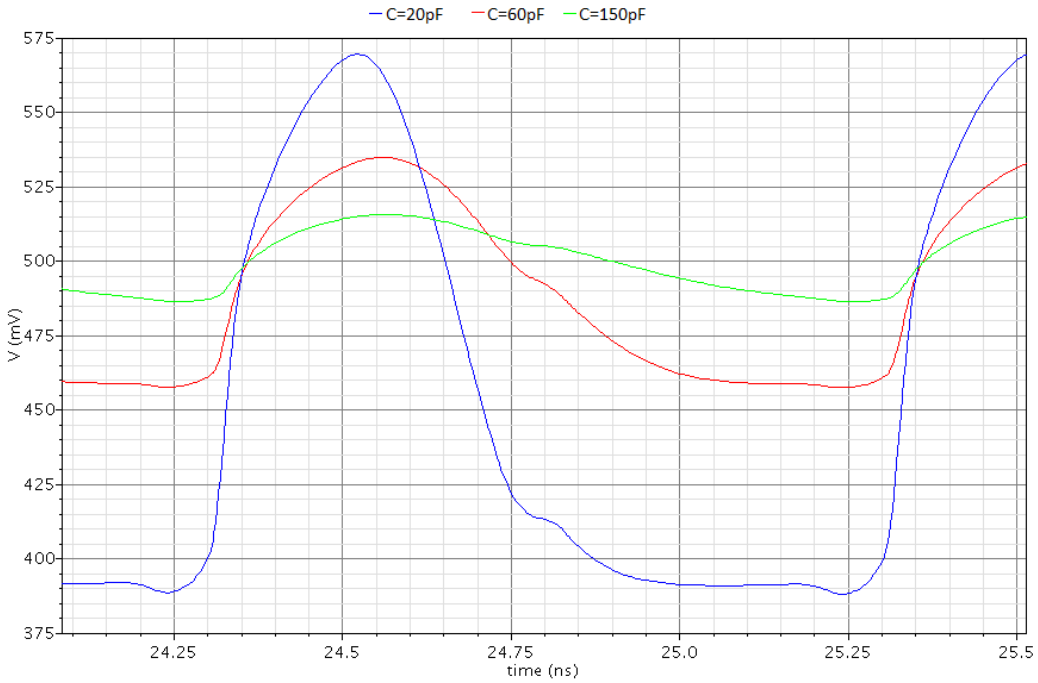

Figure 4-21: Threshold voltage with different capacitances

The total DC current consumption is 386µA, bringing the DC power consumption to less than 0.5mW.

5 Simulation results

Simulations of the designed amplifier are discussed in this chapter. First the two most important results are discussed, those regarding the linearity and the efficiency. After that the effects of different quality factors of the choke, variation in the threshold voltage source and width-scaling of the amplifier on the performance of the amplifier are simulated and discussed.

Unless otherwise noted a quality factor of 30 is assumed for the choke. Another quality factor has mainly influence on the efficiency, and only a small influence on the linearity; a lower quality factor results in a bit higher impedance seen by the amplifying transistor due to the extra resistance in series with the choke.

The MOSFETs are TSMCs CLN65 devices. All other components, except the choke, are assumed ideal. The switch frequency is set at 1GHz for all the simulations.

5.1 Linearity

Linearity is both assessed using single tone outputs to find the 1dB compression point and two-tone outputs to find the intermodulation behavior.

5.1.1 1dB compression point

The 1dB compression point is simulated using a 1GHz output signal with increasing amplitude, created by using a DC current input. Since the designed system is both amplifier and upconverter, this results with a 1GHz switch signal in a 1GHz RF output. Figure 5-1 shows the resulting output power, the Q-factor of the choke coil was 30. Here a difference in Q-factor only results in a small vertical shift, but has no significant influence on the shape.

The vertical axis shows the output power of the 1GHz tone in dBm, while the horizontal axis is related to the logarithm of the input current, but does not directly represent a power. This result contains three interesting parts, low power, medium power and high output power.

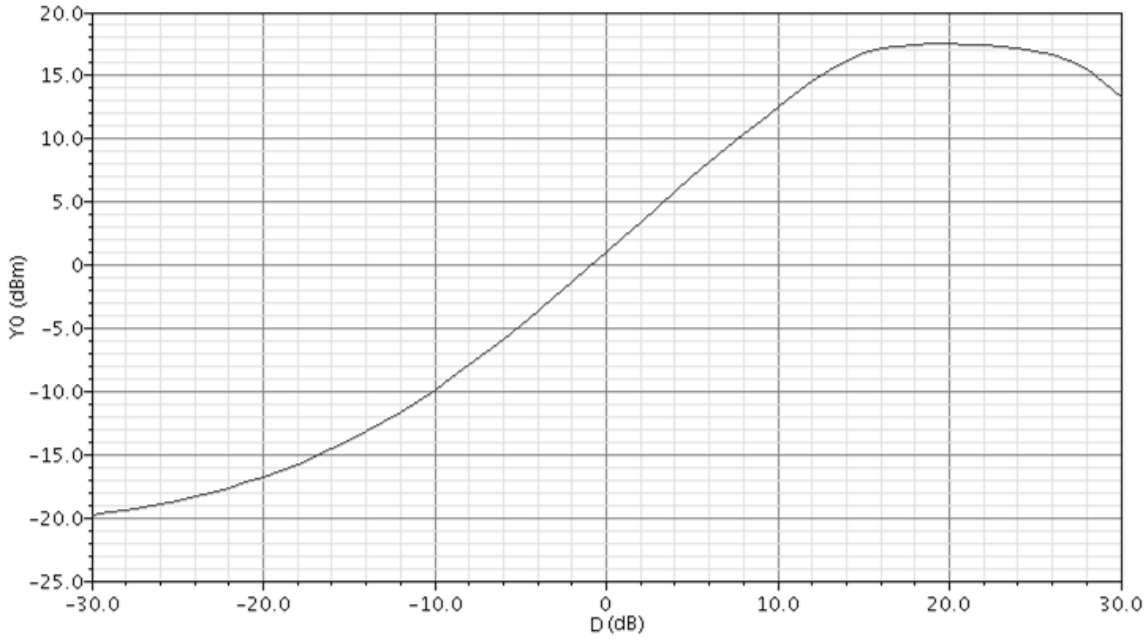

Figure 5-1: Single tone output power with increasing input power

Towards low input powers the power starts to saturate to a certain level. It can be shown that this is due to charge injection; with no input present the output power is still at -21.2dBm.

In the center region the output rises slightly faster than what should happen in the linear case. This happens due to the increasing gain at higher outputs, the feedback is not sufficient to completely eliminate this. Since both axes are logarithmic the linear line cannot have a different steepness to fit better.

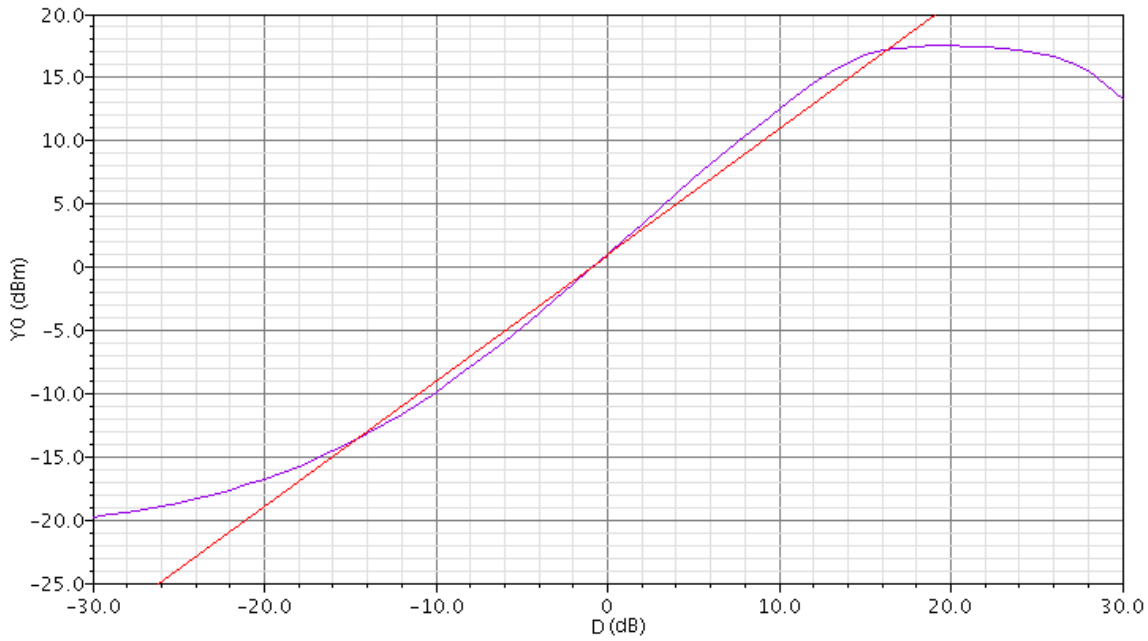

Figure 5-2: Single tone output power with ideal linear curve

This makes calculating the 1dB compression point from this curve tricky; generally the extrapolation can be started from a low input power, however in this case starting a low input power means the -1dB line is several dBs below the actual output power when the amplifier starts to saturate. However since the amplifier quickly goes from normal operation to clipping it does not matter a lot for the output referred 1dB compression point, the curve is almost flat at the 1dB compression point.

Extrapolated from an input of 10dB, the output referred 1dB compression is 16.9dBm, while when it is extrapolated from an input of 0 dB it is equal to 17.4dBm. The maximum reached output power is 17.5dBm, with a theoretical limit of 21.9dBm if an ideal square wave with an amplitude of 1.2V was present.

The final interesting area of this graph is where the input power is high; if the input power continues to increase the output power starts to decrease. This happens because the amplifier gets oversteered and leaves its operating region. Figure 5-3 shows the gate voltages for an input of 15dB, roughly the maximum usable value, and 30dB which is far too high.

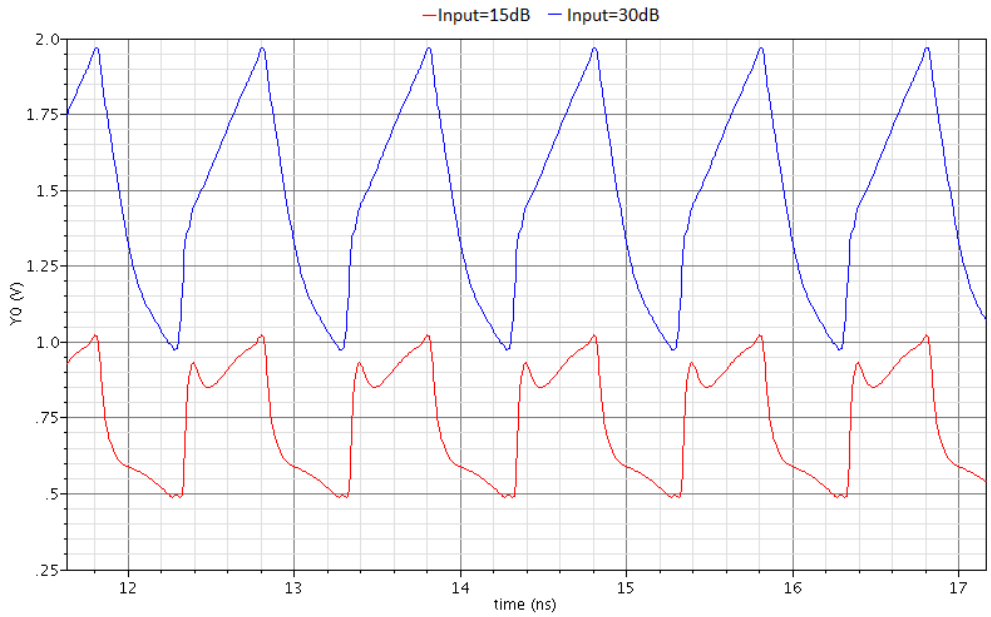

Figure 5-3: Gate voltage in normal operating region and when oversteered

At 15dB input the gate voltage is as expected, it has some issues pulling the gate to threshold while due to the input current the gate voltage increases in its on-state.

However with 30dB input the gate voltage reaches far too high voltages. The cause is a combination of too high input currents and the non-ideal switch that has to pull the gate voltage to threshold combined with the non-ideal threshold source. The high input current causes it to reach high voltages. At this point the amplifier is not capable anymore of pulling the gate to threshold. Since the gate does not reach threshold the capacitances of the amplifying transistor are not sufficiently far discharged, and the next cycle it reaches even higher voltages. This continues until periodic steady state is achieved.

5.1.2 Intermodulation distortion

Figure 5-4 shows the output power of a two-tone test. The baseband input is a 10MHz signal with 1GHz switch input. Due to the upconverter the first order responses are at 990MHz and 1010MHz, with third order IMD products at 970MHz and 1030MHz. An input of a certain amounts of dB here means an amplitude of the baseband signal equal to the DC input used for the single tone test, so the two-tone output will start to clip for the same input as the single tone output did.

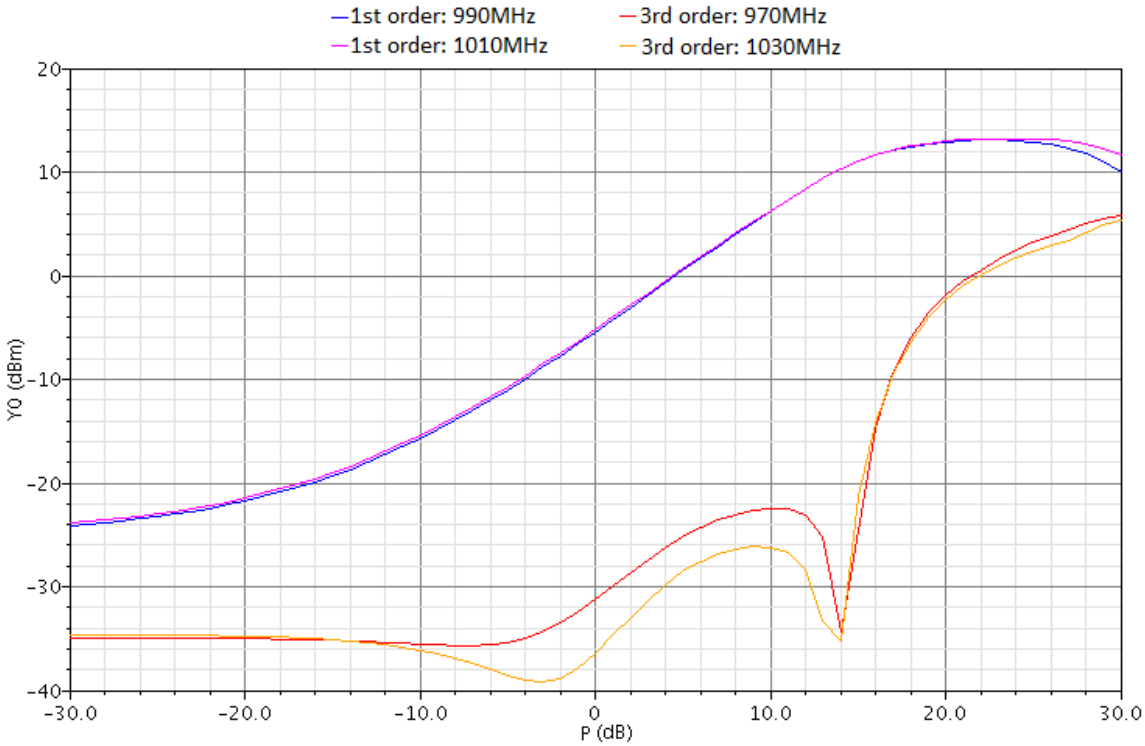

Figure 5-4: Two-tone test with 10MHz input

At low input powers the same behavior can be seen as in the single-tone test. Due to charge injection the signals approach a certain limit. For higher powers the IMD products stay roughly 25-30dB below the first order response. At an input of a bit over 10dB the IMD products rapidly decline, this is most likely a combination of the higher feedback that results in better linearity and a point where different mechanics that cause third order harmonics cancel each other. The difference between the first order response and the 970MHz IMD product is shown in figure 5-5.

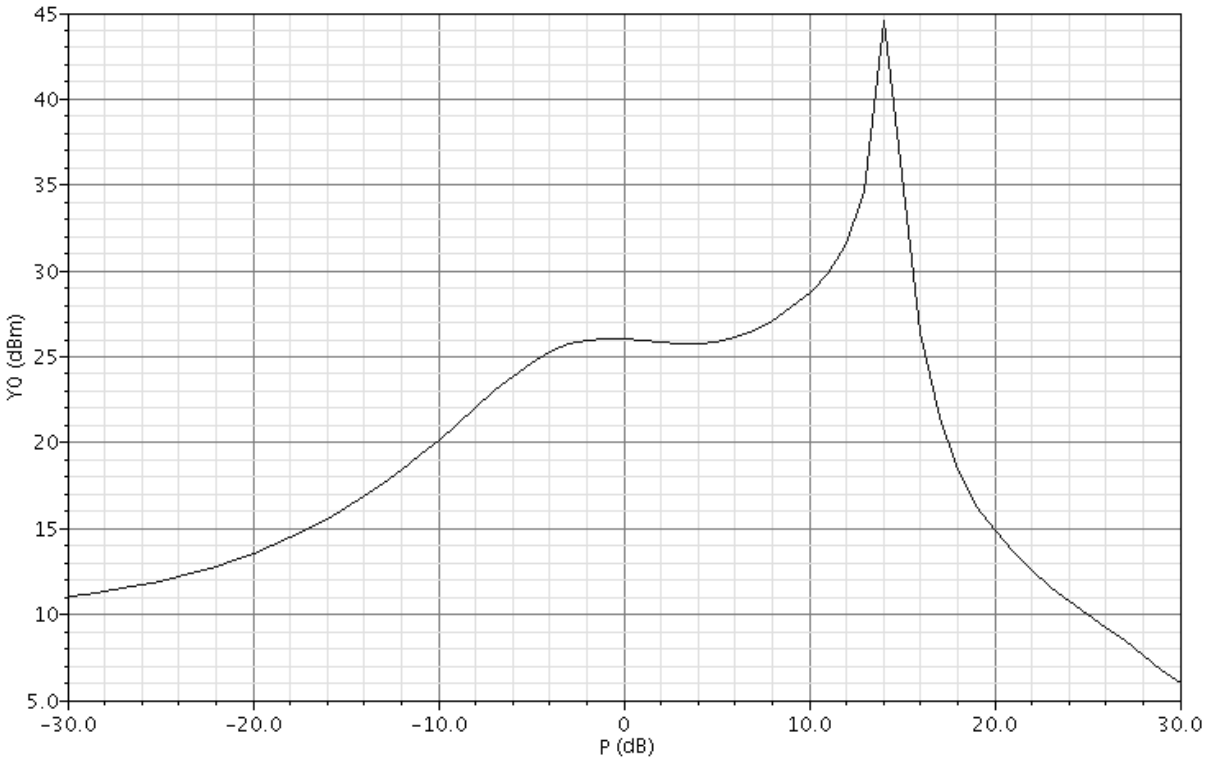

Figure 5-5: Difference between first order response and worst IMD3 product

The difference between the IMD products is because to the input current that flows through the injection capacitor also being mixed up. This also results in a noticeable, but much smaller, difference between the two first order responses. With a lower baseband frequency this problem becomes smaller, shown in figure 5-6.

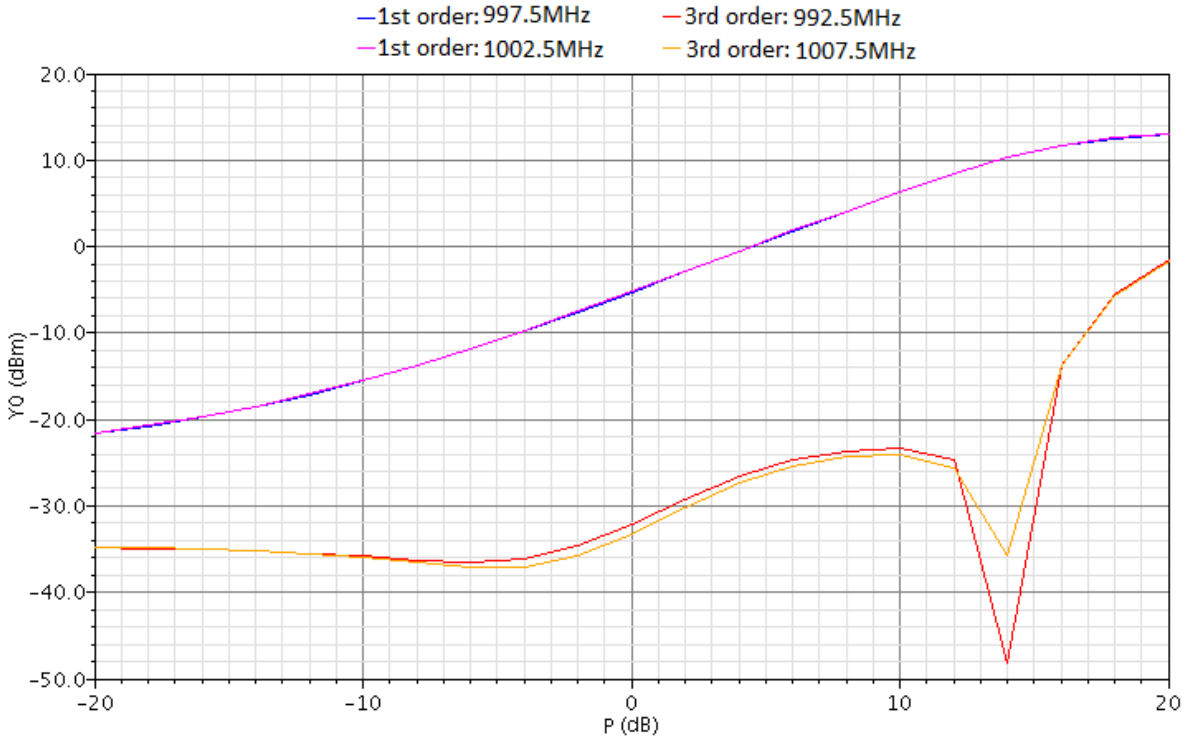

Figure 5-6: Two-tone test with 2.5MHz input

5.2 Efficiency

The efficiency depends on the output power delivered. A higher power means a higher efficiency, since most of the power consumption in the amplifier will be largely independent of the output power delivered. Figure 5-7 shows the efficiency and total power consumption versus the output power.

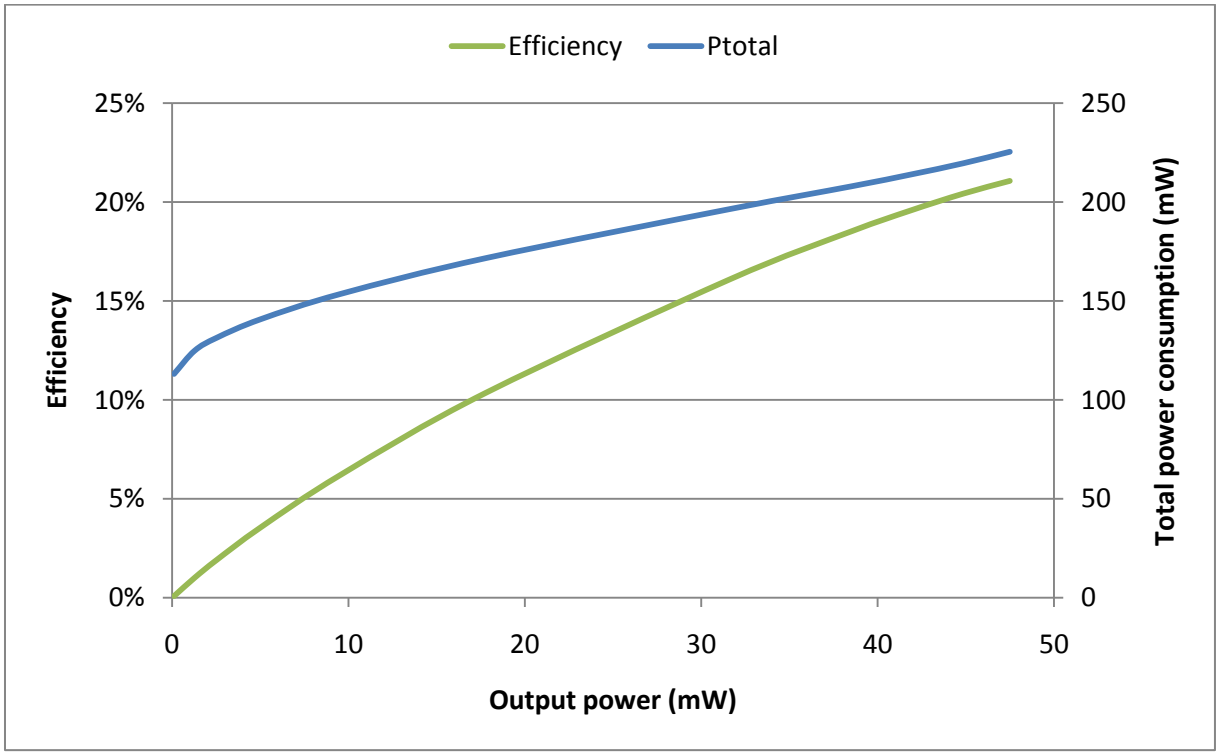

Figure 5-7: Efficiency as function of the output power

5.2.1 Power consumption per component

At a DC input of 15dB, which means the amplifier's output is 1GHz with maximum amplitude, the distribution of the power consumption is as follow:

Table 3: Power distribution of amplifier at maximum output

Component	Power consumption	Percentage
Total	225.4mW	100%
Amplifier	177.9mW	78.9%
Input current source	26.7mW	11.8%
Switch control	100.6mW	44.6%
Initial amplification	37.4mW	16.5%
Switch 1	6.9mW	3.1%
Switch 2	5.2mW	2.3%
Switch 3	51.1mW	22.7%
Threshold source	0.45mW	0.2%
Output stage (minus load)	50.1mW	22.2%
Load	47.5mW	21.1%
1.2V rail	99.7mW	44.2%
2.5V rail	125.6mW	55.7%

With no input, the power distribution is:

Table 4: Power distribution of amplifier with no input

Component	Power consumption	Percentage
Total	109.1mW	100%
Amplifier	109.1mW	100%
Input current source	4.6mW	4.2%
Switch control	99.2mW	90.9%
Initial amplification	37.4mW	34.3%
Switch 1	7mW	6.4%
Switch 2	5.1mW	4.7%
Switch 3	49.8mW	45.6%
Threshold source	0.44mW	0.4%
Output stage (minus load)	4.9mW	4.5%
Load	7.6µW	0%
1.2V rail	6.9mW	6.3%
2.5V rail	102.2mW	93.7%

The power distribution with a maximum 10MHz baseband input (a two-tone output) is:

Table 5: Power distribution of amplifier with full swing 10MHz input

Component	Power consumption	Percentage
Total	181.4mW	100%
Amplifier	155.4mW	85.7%
Input current source	18mW	9.9%
Switch control	99.1mW	54.6%
Initial amplification	37mW	20.4%
Switch 1	6.9mW	3.8%
Switch 2	5.1mW	2.8%
Switch 3	50mW	27.6%
Threshold source	0.44mW	0.2%
Output stage (minus load)	37.9mW	20.9%
Load	26mW	14.3%
1.2V rail	65.9mW	36.3%
2.5V rail	115.5mW	63.7%

An interesting phenomenon is the power consumption of switch three; it is far higher than that of the others. Since switch two and three are quite similar except the oxide-thickness and length, they are compared. The NMOS in switch two has as dimensions 40x30µm/0.06µm. With this the total gate capacitances come to 1.47pF, fully switched on with no drain-source voltage. The NMOS in switch three has as dimensions 40x30µm/0.28µm. This brings the total gate capacitance to 2.34pF.

The power dissipated in a switch per cycle is approximated by (5-1).

$$E_{switch} = \frac{1}{2} CV^2 \qquad \textbf{(5-1)}$$

The power dissipated by switch three is already 60% higher due to larger capacitance. The voltage

squared adds a factor 4.3; these two combined bring the difference to a factor 7. This is still less than the simulated difference. The remaining difference can be explained by different efficiencies of the amplifying inverters for both switches and that (5-1) is for these switches only an approximation. Also for the PMOS transistors in both switches the difference in capacitance is a bit larger.

It can be concluded that the switches use a lot of power. With no output over 90% of the power is dissipated by the switches and their control.

5.3 Inductor quality factor

Since the quality factor of the choke coil is not known, the efficiency and IMD3 products are now simulated for several different quality factors: 10, 30 and 1000. The efficiency at an input of 15dB with a single tone output is 17% for a quality factor of 10, 21% for a quality factor of 30 and with a quality factor of 1000 the efficiency increases to 23%.

Figure 5-8 shows one of the first order responses and the worst IMD3 product with a 10MHz baseband input for the different quality factors.

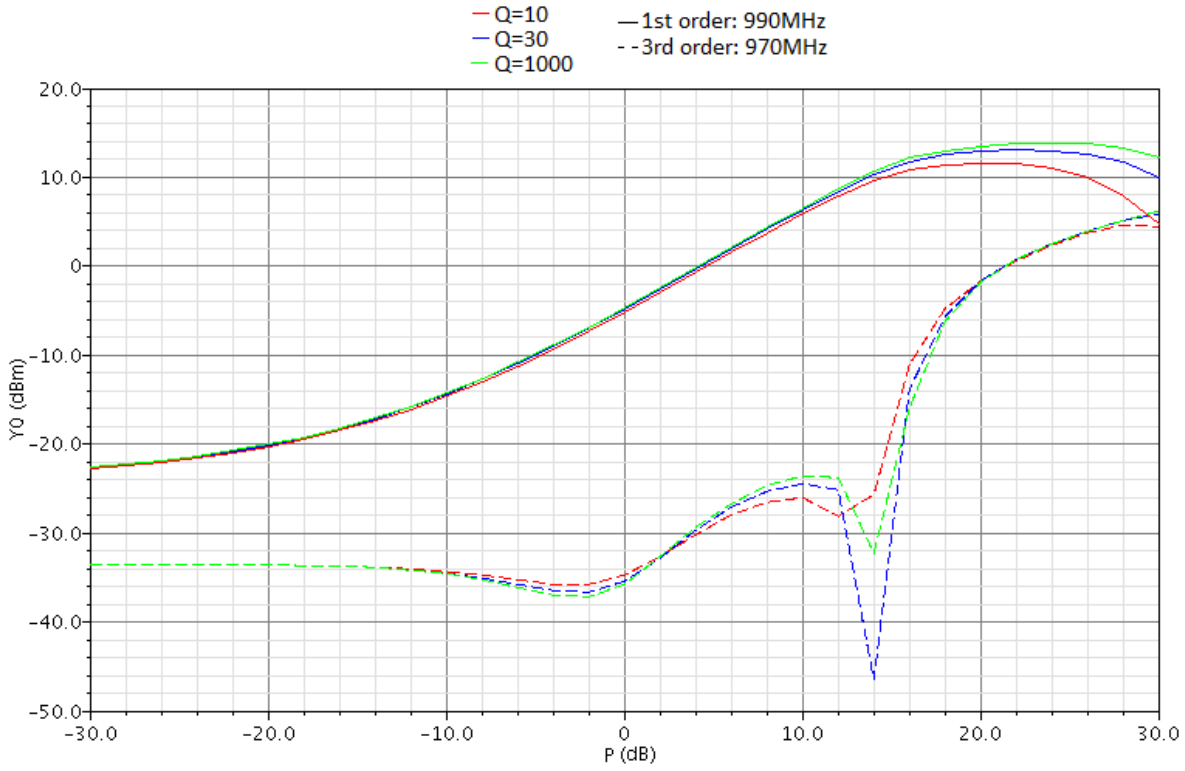

Figure 5-8: Third order intermodulation distortion as function of input power and inductor quality factor

The quality factor barely has an influence on the output harmonics. The IMD3 curves have a bit different shape, but there is not one clearly better than the others. The step size in input power can explain the different sizes of the notches around P=13dB, the resolution is not sufficiently high so the notch can be between two points.

For high input powers the amplifier starts to saturate at somewhat lower output power with an inductor with a low quality factor, which is to be expected with the extra resistive part present in the inductor.

5.4 W-scaling

With larger widths in the amplifier the efficiency will decrease, but the transconductance will increase, resulting in a higher loop gain.

The load and matching network are now scaled to represent width-scaling. For the general behavior it does not matter if the amplifier's width is increased or the width of the load is decreased, what is important is the ratio. To do this the impedances of the load and the matching network are multiplied by a factor *W*, so their width is divided by *W*. Since the ratio between the widths in the load and in the amplifier is important, this is for the behavior equivalent to increasing the width of the components in the amplifier by a factor *W*.

The minimum size of the W-factor is roughly 0.35, at this value the capacitor to ground in the matching network needs to be around 1.5pF, the total parasitic capacitances from the drain to the ground are also almost this large, so for an even smaller W-factor either the complete matching network needs to be redesigned or a negative capacitance is required.

Besides finding the efficiency and linearity depending on W-scaling, it also allows us to calculate the output impedance of the amplifier.

5.4.1 Efficiency

Figure 5-9 gives the efficiency and output voltage as function of the load impedance. As one would expect with a smaller impedance of the load, so effectively smaller width of the amplifier, the efficiency increases.

Since this was simulated using a constant input power, the output voltage decreases with smaller load impedance. So if also for lower loads the output would be at its maximum the efficiency there would be higher. Using a 5Ω load in combination with an input of 20dB results in an efficiency near 35%. With this input the output is still not full swing. Full swing is also not possible to reach; in section 5.1.1 showed that for inputs above roughly 20dB, the output starts to decrease. This effect is relative independent of the load, so increasing the input further does not help.

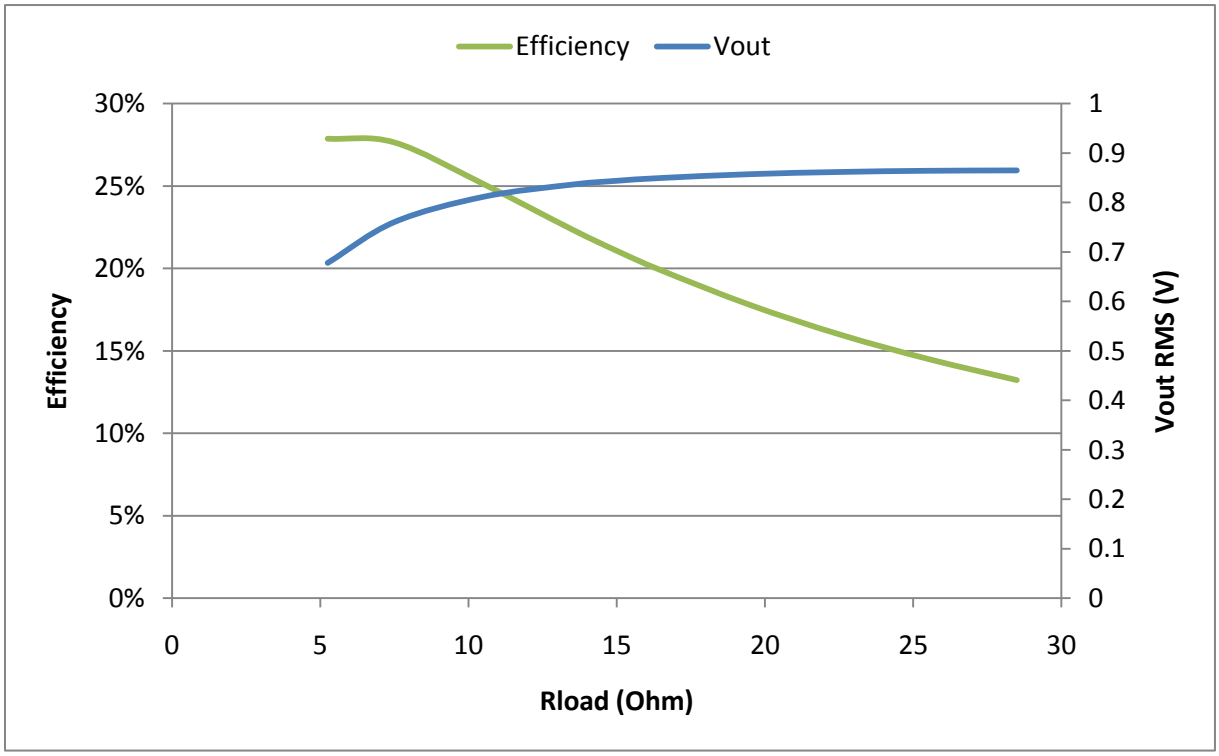

Figure 5-9: Efficiency depending on the output impedance, single tone output with 15dB input

5.4.2 Linearity

For the linearity simulations the input is set at 10dB. At 15dB input the output voltage clips for most values above 15Ω, there is more headroom with 10dB input.

Figure 5-10 shows the result from the simulations.

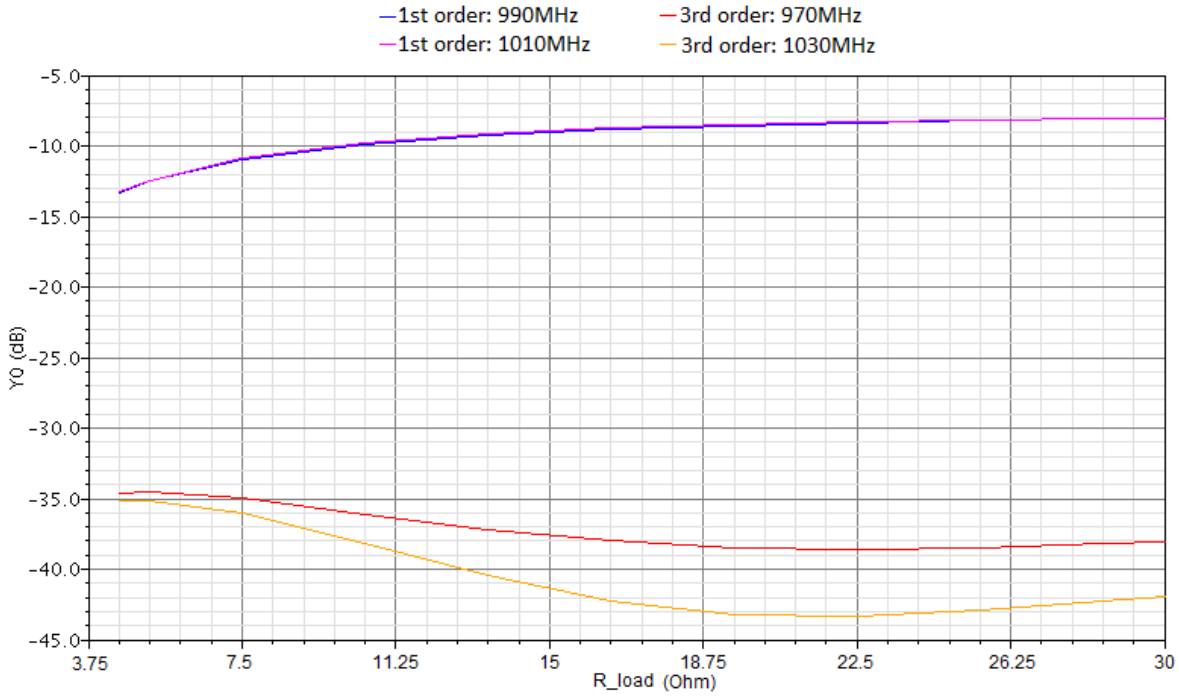

Figure 5-10: First order response and IMD3 products as function of W-scaling parameter

For large load impedances there is less distortion than for small impedances. This is because the loop gain is higher, resulting in more feedback, that in turn results in less distortion. Together with figure 5-9 this shows the trade-off between linearity and efficiency.

5.4.3 Output impedance

The input is also set at 10dB for output impedance calculations. The output impedance is calculated using the previously used equation 2-8:

$$Z_{out} = R_L * \frac{U_{open} - U_{loaded}}{U_{loaded}} \quad \textbf{(5-2)}$$

It is simulated for several load impedances, the loaded output voltage is measured for different loads and the open voltage is roughly measured using a load with a scaling factor of 10, so 150Ω. Figure 5-11 shows the resulting calculated impedances.

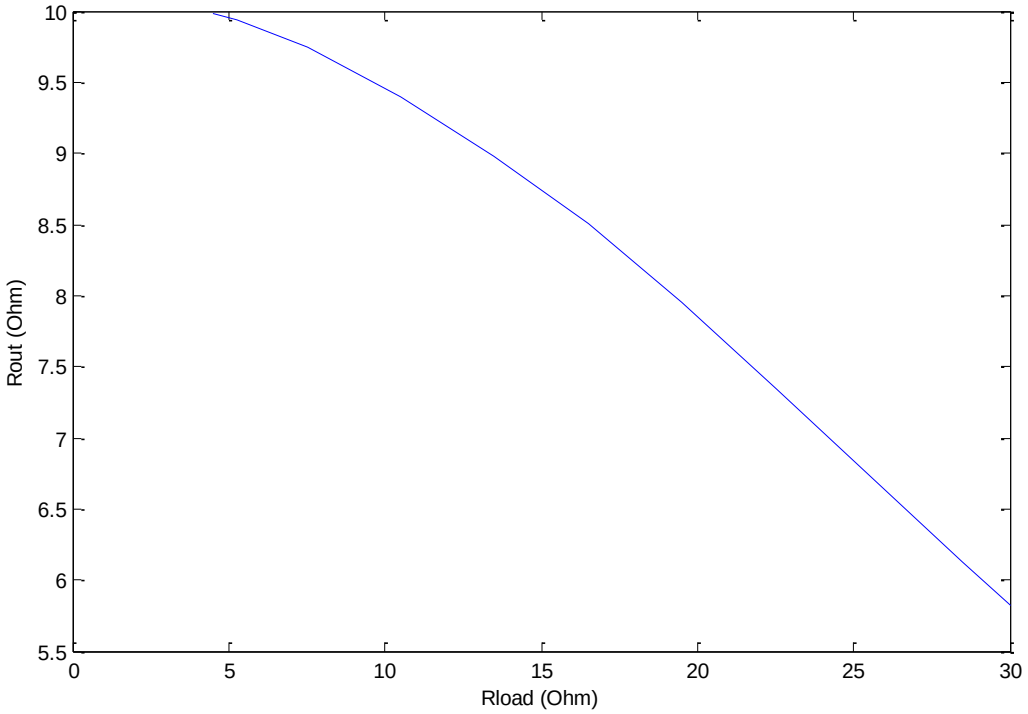

Figure 5-11: Simulated output impedance

The output impedance drops for increasing load impedances. This is mainly due to the increase in loop gain: a higher load yields a larger gain, while the output impedance decreases with an increase in loop gain.

5.5 Threshold voltage

The threshold voltage source does not need to be exactly at the threshold voltage of the transistor. For an ideal transistor it would be preferred to be very close to the threshold voltage, however due to sub-threshold conduction a higher or lower voltage is also possible. A lower voltage decreases the current flowing in the off-state, resulting in a higher efficiency. With a higher voltage it starts to operate more like a class-A amplifier, the efficiency decreases but the linearity increases.

Figure 5-12 shows the efficiency and how far below the wanted signal the worst IMD3 component is. The efficiency is at a single tone output with an input of 15dB, while the IMD3 products were simulated at an input of 10dB. The nominal value of the threshold voltage is 500mV.

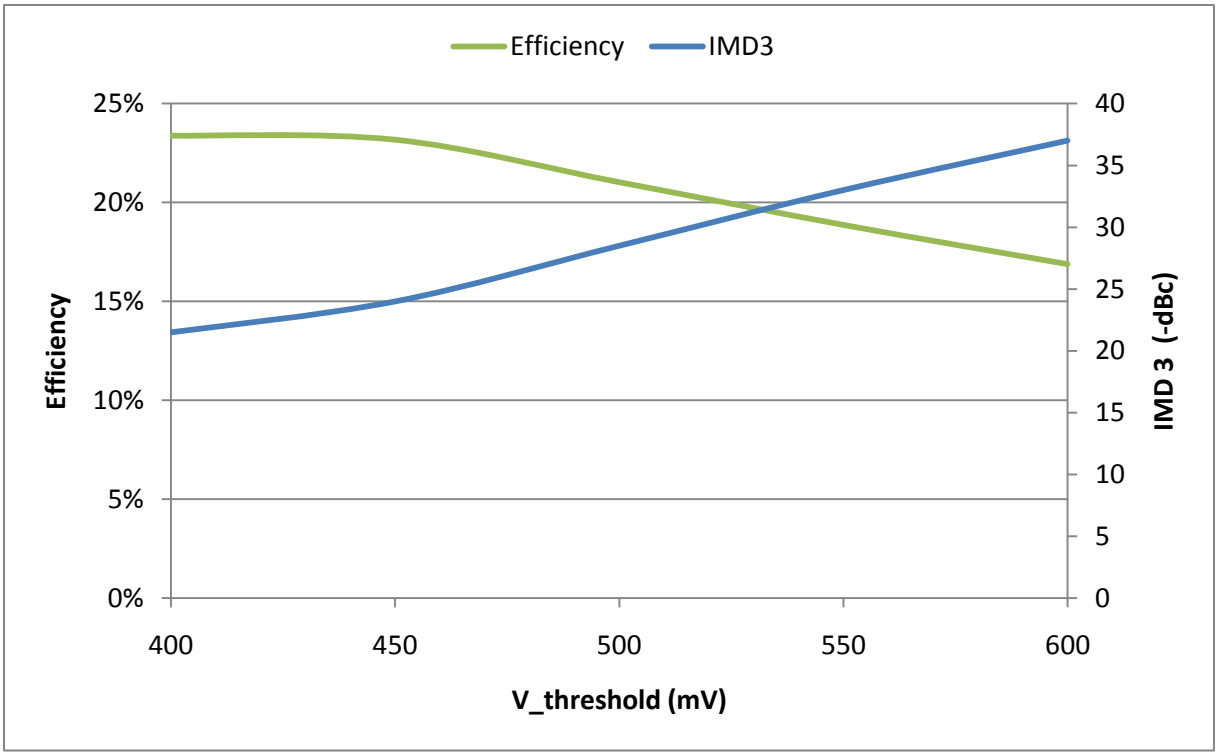

Figure 5-12: Effect of threshold voltage source on efficiency and linearity

5.6 Comparison with published Pas

To be able to assess the performance of the design it needs to be compared to other designs. Table 6 shows a comparison with PAs published in the last few years.

Both the efficiency (η) and the IMD3 products are measured at the 1dB compression point. In normal operation this will be the maximum useful output power, so it is a good point to measure the efficiency. The same is true for IMD3 products, especially since the problem they form in adjacent channels is largest at the maximum output power. The OIP3 would be a better number for comparison; however the designed amplifier, and many other, especially switching, amplifiers, do not have IMD3 products that follow the theoretical line.

All other designs present in the table are only PAs; they do not include an upconverter. However there are not many publications that only include an upconverter and a PA with the required data for comparisons.

Table 6: Comparison with other designs

Process	Freq	Pmax	P1dB	η	IMD3	Inductors	Notes	Ref
65nm	1GHz	17.5dBm	16.9dBm	21%	-28dBc	Bond wire	Incl upconverter	*This*
0.13µm	2.4GHz	27.5dBm	24dBm	25%	-29dBc	On-chip		[**8**]
65nm	1.95GHz	31dBm	25.6dBm	10%	-	On-chip		[**9**]
0.18µm	915MHz	20dBm	14.5dBm	13%	-	Bond wire		[**10**]
0.18µm	-	22dBm	20.5dBm	36%	-24dBc	On-chip	Excl driver	[**11**]
0.13µm	350MHz	9dBm	-	11%	-	Off-chip	f_{max}=2.4GHz	[**12**]

[**11**] does not include a driver stage, which explains the high efficiency. The operating frequency is not mentioned in the publication.

The comparison should only be used as an indication; most of the data from the other amplifiers are realized designs instead of only simulations. However at the same time they require a mixer to create the necessary inputs, which will also consume power and add nonlinearity.

Compared to other amplifiers the 1dB compression point of the designed PA and upconverter combination is very close to the maximum output. The efficiency is also quite good, which will be at least partly due to the high 1dB compression point.

The maximum efficiency of for example [**10**] is above 30%, significantly higher than the efficiency reached here. However at its 1dB compression point this efficiency is more than two times lower at 13%.

6 Conclusion

A charge steered combination of an RF PA and an upconverter was designed. This design allows for feedback at a frequency of 1GHz to linearize the output of the PA, in contrast with regular PAs where direct feedback is not an option due to stability concerns.

Since the gain of a transistor is higher at larger input signal levels, the design performs better with large inputs than with small inputs. At high outputs the linearity is very good, but at low values the linearity degrades due to lower loop gain and due to problems with charge injections.

The design has its IMD3 products almost at -30dBc at the 1dB compression point. The reached efficiency at the 1dB compression point is 21%. Conventional RF power amplifiers can usually reach high efficiencies, but they do need to be steered to their maximum output for this to happen. When their output power is lowered until where their IMD products are sufficiently low, their efficiency is often significantly worse. This design has its 1dB compression point only 0.6dB below its maximum output, which allows for a good efficiency despite requiring quite some power in other components.

The useful input range of the designed system is roughly 20dB: from a minimum of roughly -5dB until a maximum of 15dB. With higher values it starts to clip, while with lower values charge injection becomes dominant.

There are three major problems present. Firstly, the limited gain of one stage at gigahertz frequencies limits the loop gain. However while more gain would be preferred, it is not a requirement for a working design, as shown in this thesis. This is mainly a problem at low output powers, the gain is much lower when this is the case, resulting in more distortion products.

Secondly the required switch size is a problem: the switches that control the amplifier need to be large to be able to make the amplifier switch fast enough, but large switches require a lot of charge to control, which gets injected into the system. This both results in a lower efficiency and a worse performance for especially low input powers. The switch connecting the drain and the injection capacitor consumes ten times more power than the other switches individually, making it the most power hungry component in the design.

Thirdly the injection capacitor and the capacitor creating the auxiliary voltage source are both quite big, 10pF, which is large but still doable. However the threshold voltage source requires a capacitor of at least 60pF, this would cost a very large area on a chip.

6.1 Recommendations

The required size of the switches can be severely decreased when bootstrap circuits are introduced. A PMOS-NMOS combination would still be required against charge injection, but when both are equipped with their own bootstrap circuits they can both be fully conducting in the on-state. It would require extra bootstrap circuits though, which also consume power.

The width of the switches can also be decreased depending on the input signal levels, with large inputs use more parallel transistors for switching, while for small input signal levels some of the switches can be deactivated, decreasing power consumption and charge injection, which is mainly relevant for low input powers. However the switches would still need to be wide enough for correct performance.

The largest capacitor is required for the threshold voltage source. This is needed to be able to quickly sink the current spike coming from the amplifying transistor's parasitic capacitances that are discharged. However in theory the only requirements are that in the off-state the transistor does not conduct, and before it is turned on again the voltage at the gate is at the threshold voltage. So the initial current spike could be quickly routed to the ground terminal, and then relative slowly the gate voltage can be charged again to threshold. This can eliminate the required 60pF capacitor that is now used in the design. However due to subthreshold conduction this will be visible at the output, so it is an option to look into, but with problems that need to be overcome.

7 Bibliography

[1] B. Razavi, "Mobility Degradation with Vertical Field," in *Design of Analog CMOS Integrated Circuits*.: McGraw-Hill, 2000, ch. 16.2.2, pp. 585-589.

[2] B. Razavi, "MOS Device Capacitances," in *Design of Analog CMOS Integrated Circuits*.: McGraw-Hill, 2000, ch. 2.4.2, pp. 29-33.

[3] Agilent Technologies, Inc., "Capacitance Model ," in *Nonlinear Device Models, Volume 1*., 2007, ch. 4, pp. 326-342.

[4] L. Khan, "Single-sided transmission by envelope elimination and restoration," in *Proc. IRE*, 1952, pp. 803-806.

[5] M. Abramowitz and I.A. Stegun, "Functions of Angles in Any Quadrant in Terms of Angles in the First Quadrant," in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*.: Dover, 1972, ch. 4.3.44, p. 73.

[6] B. Razavi, "Charge Injection Cancellation," in *Design of Analog CMOS Integrated circuits*.: McGraw-Hill, 2000, ch. 12.2.4, pp. 421-423.

[7] M. Acar, A.J. Annema, and B. Nauta, "Generalized Design Equations for Class-E Power Amplifiers with Finite DC Feed Inductance," in *36th European Microwave Conference*, Manchester, UK, 2006, pp. 1308-1311.

[8] G. Liu, T.-J. King, and A.M. Niknejad, "A 1.2V, 2.4GHz Fully Integrated Linear CMOS Power Amplifier with Efficiency Enhancement," in *Proceedings of CICC*, 2006, pp. 141-144.

[9] Y. Luque, E. Kerherve, N. Deltimple, and D. Belot, "CMOS PA design dedicated to UMTS (3G) applications in 65 nm technology," in *Circuits and Systems for Communications, 2008. ECCSC 2008. 4th European Conference on*, 2008, pp. 101-104.

[10] G. Tongqiang, C. Baoyong, Z. Chun, and W. Zhihua, "Design and analysis of a highly integrated CMOS power amplifier for RFID reader," in *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, 2008, pp. 1480-1483.

[11] Y. Ding and R. Harjani, "A high-efficiency CMOS +22-dBm linear power amplifier," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1895-1900, September 2005.

[12] R. Shrestha, E. Mensink, E.A.M. Klumperink, G.J.M. Wienk, and B. Nauta, "A multipath technique for canceling harmonics and sidebands in a wideband power upconverter," in *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International*, 2006, pp. 1800-1809.

Appendix A: Equations

A.1 Basic

First the small signal response of the amplifier is calculated. Also it is assumed that the impedance of the capacitors is much larger than that of the load, so the output voltage is given by:

$$V_{out} = -I_D R_L \quad \text{(A-1)}$$

$$I_D = V_{in} g_m \quad \text{(A-2)}$$

$$V_{out} = -V_{in} R_L g_m \quad \text{(A-3)}$$

The input voltage depends on how the input charge is distributed across the two capacitors. The input voltage is determined by the charge on the gate-source capacitance, the difference between the input and output voltage and the gate-drain capacitance determines how much charge is stored there.

$$Q_{in} = Q_{GS} + Q_{GD} \quad \text{(A-4)}$$

$$V_{in} = \frac{Q_{GS}}{C_{GS}} \quad \text{(A-5)}$$

$$V_{in} = \frac{Q_{in} - Q_{GD}}{C_{GS}} \quad \text{(A-6)}$$

The charge in the gate-drain capacitance is given by the capacitance and the voltage across it.

$$Q_{GD} = C_{GD}(V_{in} - V_{out}) \quad \text{(A-7)}$$

Combining (A-6) and (A-7) and rewriting them gives:

$$V_{in} = \frac{Q_{in} - C_{GD}(V_{in} - V_{out})}{C_{GS}} \quad \text{(A-8)}$$

$$V_{in}\left(1 + \frac{C_{GD}}{C_{GS}}\right) = \frac{Q_{in} + C_{GD} V_{out}}{C_{GS}} \quad \text{(A-9)}$$

$$V_{in} = \frac{Q_{in} + C_{GD} V_{out}}{C_{GS} + C_{GD}} \quad \text{(A-10)}$$

(A-10) combined with (A-3) gives the output voltage as function of input charge.

$$V_{out} = -\frac{Q_{in} + C_{GD} V_{out}}{C_{GS} + C_{GD}} R_L g_m \quad \text{(A-11)}$$

$$V_{out}\left(1 + \frac{C_{GD} R_L g_m}{C_{GS} + C_{GD}}\right) = -\frac{Q_{in}}{C_{GS} + C_{GD}} R_L g_m \quad \text{(A-12)}$$

$$V_{out} = -\frac{Q_{in}}{\left(1 + \frac{C_{GD} R_L g_m}{C_{GS} + C_{GD}}\right)(C_{GS} + C_{GD})} R_L g_m \quad \text{(A-13)}$$

$$V_{out} = -\frac{Q_{in} R_L g_m}{C_{GS} + C_{GD} + C_{GD} R_L g_m} \quad \text{(A-14)}$$

When the output resistance of the transistor is taken into account, (A-14) turns into:

$$V_{out} = -\frac{Q_{in}(R_L//r_{out})g_m}{C_{GS} + C_{GD} + C_{GD}(R_L//r_{out})g_m} \quad \text{(A-15)}$$

$$V_{out} = -\frac{Q_{in}R_L r_{out} g_m}{(R_L + r_{out})(C_{GS} + C_{GD}) + C_{GD}R_L r_{out} g_m} \quad \text{(A-16)}$$

A.2 Output impedance

Equation A-16 is substituted in equation 2-8, where for U_{open} the load impedance is taken to be infinite. The result from this is:

$$Z_{out} = R_L \frac{\frac{Q_{in}r_{out}g_m}{C_{GS} + C_{GD} + C_{GD}r_{out}g_m} - \frac{Q_{in}R_L r_{out} g_m}{(R_L + r_{out})(C_{GS} + C_{GD}) + C_{GD}R_L r_{out} g_m}}{\frac{Q_{in}R_L r_{out} g_m}{(R_L + r_{out})(C_{GS} + C_{GD}) + C_{GD}R_L r_{out} g_m}} \quad \text{(A-17)}$$

This can be simplified to:

$$Z_{out} = \frac{(R_L + r_{out})(C_{GS} + C_{GD}) + C_{GD}R_L r_{out} g_m}{C_{GS} + C_{GD} + C_{GD}r_{out}g_m} - R_L \quad \text{(A-18)}$$

$$Z_{out} = \frac{r_{out}(C_{GS} + C_{GD})}{C_{GS} + C_{GD} + C_{GD}r_{out}g_m} \quad \text{(A-19)}$$

A.3 Quadratic analysis

For the behavior when the transistor has a second order behavior equation 2-10 is used. From this it can be calculated using the same principle as was used in section A.1.

However the quadratic term requires more calculations to get the required result. After the same calculations as in section A.1 are applied and some reshuffling, the result is:

$$V_{out}^2 * g_2 R_L \left(\frac{C_{GD}}{C_{GD} + C_{GS}}\right)^2 + V_{out}\left(1 + \frac{R_L C_{GD} g_m}{C_{GS} + C_{GD}} + \frac{2R_L Q_{in} C_{GD} g_2}{(C_{GS} + C_{GD})^2}\right) + \frac{Q_{in}R_L}{C_{GS} + C_{GD}}\left(g_m + \frac{g_2 Q_{in}}{C_{GS} + C_{GD}}\right) = 0 \quad \text{(A-20)}$$

The quadratic formula can be applied here to find the output voltage. The formula and its components are shown below.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad \text{(A-21)}$$

$$a = g_2 R_L C_r^2$$

$$b = 1 + g_m R_L C_r + \frac{2R_L Q_{in} C_r^2 g_2}{C_{GD}} \quad \text{(A-22)}$$

$$c = \frac{Q_{in}R_L C_r}{C_{GD}}\left(g_m + \frac{g_2 Q_{in} C_r}{C_{Gd}}\right)$$

$$C_r = \frac{C_{GD}}{C_{GD} + C_{GS}} \quad \text{(A-23)}$$

Combining equations A-21, A-22 and A-23 gives the following result:

$$V_{out} = \frac{-\left(1 + g_m R_L C_r + \frac{2R_L Q_{in} C_r^2 g_2}{C_{GD}}\right) \pm \sqrt{1 + 2g_m R_L C_r + g_m^2 R_L^2 C_r^2 + \frac{4R_L Q_{in} C_r^2 g_2}{C_{GD}}}}{2g_2 R_L C_r^2} \quad \textbf{(A-24)}$$

This results in two possible outcomes, due to the $\pm$ sign. Filling in $Q_{in} = 0$ will give the correct one, since it is known that without charge in the output voltage should be zero:

$$V_{out} = \frac{-(1 + g_m R_L C_r) \pm \sqrt{1 + 2g_m R_L C_r + g_m^2 R_L^2 C_r^2}}{2g_2 R_L C_r^2} \quad \textbf{(A-25)}$$

So the correct one to use is plus, since that results in zero volt output without charge in. With some reshuffling, the output voltage is equal to:

$$V_{out} = -\frac{\left(1 + g_m R_L C_r + \frac{2R_L Q_{in} C_r^2 g_2}{C_{GD}}\right) - \sqrt{(1 + g_m R_L C_r)^2 + \frac{4R_L Q_{in} C_r^2 g_2}{C_{GD}}}}{2g_2 R_L C_r^2} \quad \textbf{(A-26)}$$

The Taylor approximation on this equation is used to find the nonlinear behavior. It is done around zero input charge. The Taylor expansion is given below.

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(x - a)^n \quad \textbf{(A-27)}$$

From this the second order coefficient can be calculated, it is equal to:

$$H_2 = -\frac{R_L C_r^2 g_2}{C_{GD}^2 (1 + g_m R_L C_r)^3} \quad \textbf{(A-28)}$$

The input charge as function of the input voltage is for a first order approximation given by:

$$Q_{in} = V_{in}\left(C_{GS} + C_{GD}(1 + g_m R_L)\right) \quad \textbf{(A-29)}$$

Combining these two equations gives a result for the second order behavior of the output voltage, which can easily be compared for the voltage steered case. When it is divided by the load impedance the result is the output current its second order behavior, which for the voltage steered version is g_2, and for the charge steered is then equal to the equation shown below.

$$\frac{H_2}{-g_2 R_L} = \frac{C_{GD} + C_{GS}}{C_{GD} + C_{GS} + g_m R_L C_{GD}} \quad \textbf{(A-30)}$$

A.4 Mobility reduction

For the behavior when mobility reduction is taken into account, equation 2-16 is used. From this it again can be calculated using the same principle as was used in section 2.1.1.

Solving this makes us arrive at the expression given below.

$$V_{out}^2 C_r \left(\theta + \frac{1}{2} R_L K C_r\right) + V_{out}\left(1 + \theta \frac{Q_{in} C_r}{C_{GD}} - \theta V_{TH} + \left(\frac{Q_{in} C_r}{C_{GD}} - V_{TH}\right) C_r R_L K\right) + \left(\frac{Q_{in} C_r}{C_{GD}} - V_{TH}\right)^2 \frac{1}{2} R_L K = 0 \quad \text{(A-31)}$$

This can be solved again using the ABC-equation. The same method as described in section A.3 is used to find out if a minus or plus is required in the ABC equation, but now this is done for the point where $V_{in} = V_{TH}$, at which the current should still be zero.

From this the resulting output voltage becomes:

$$V_{out} = \frac{-1 - (\theta + C_r R_L K)\left(\frac{Q_{in} C_r}{C_{GD}} - V_{TH}\right) + \sqrt{c_2 Q_{in}^2 + c_1 Q_{in} + c_0}}{2C_r\left(\theta + \frac{1}{2} R_L K C_r\right)}$$

$$c_0 = 1 - 2(\theta + C_r R_L K) V_{TH} + \theta^2 V_{TH}^2 \quad \text{(A-32)}$$

$$c_1 = \frac{2 R_L K C_r^2 + 2\theta C_r - 2\theta^2 C_r V_{TH}}{C_{GD}}$$

$$c_2 = \frac{\theta^2 C_r^2}{C_{GD}^2}$$

A Taylor expansion of equation A-32 gives the following results for the second and third order coefficient:

$$H_2 = \frac{4 c_2 c_0 - c_1^2}{16 C_r \left(\theta + \frac{1}{2} R_L K C_r\right)\left(c_2 Q_{in}^2 + c_1 Q_{in} + c_0\right)^{1.5}} \quad \text{(A-33)}$$

$$H_3 = -\frac{(4 c_2 c_0 - c_1^2)(2 c_2 Q_{in} + c_1)}{32 C_r \left(\theta + \frac{1}{2} R_L K C_r\right)\left(c_2 Q_{in}^2 + c_1 Q_{in} + c_0\right)^{2.5}} \quad \text{(A-34)}$$

Appendix B: Schematics

The designed schematics are displayed here. Most components have their dimensions shown, some have only parameter names. Those dimensions and values are all in chapter 4.

Figure B-1 shows the top level design. On top there are the input sources. Below that is the generation of the actual input signals. The *input_stage* block performs the phase shifting of the switch signal when the baseband input has a zero-crossing. It also functions as the required sign detector.

The amplifier is on the bottom, with the output filter and the load impedance (*PORT3*).

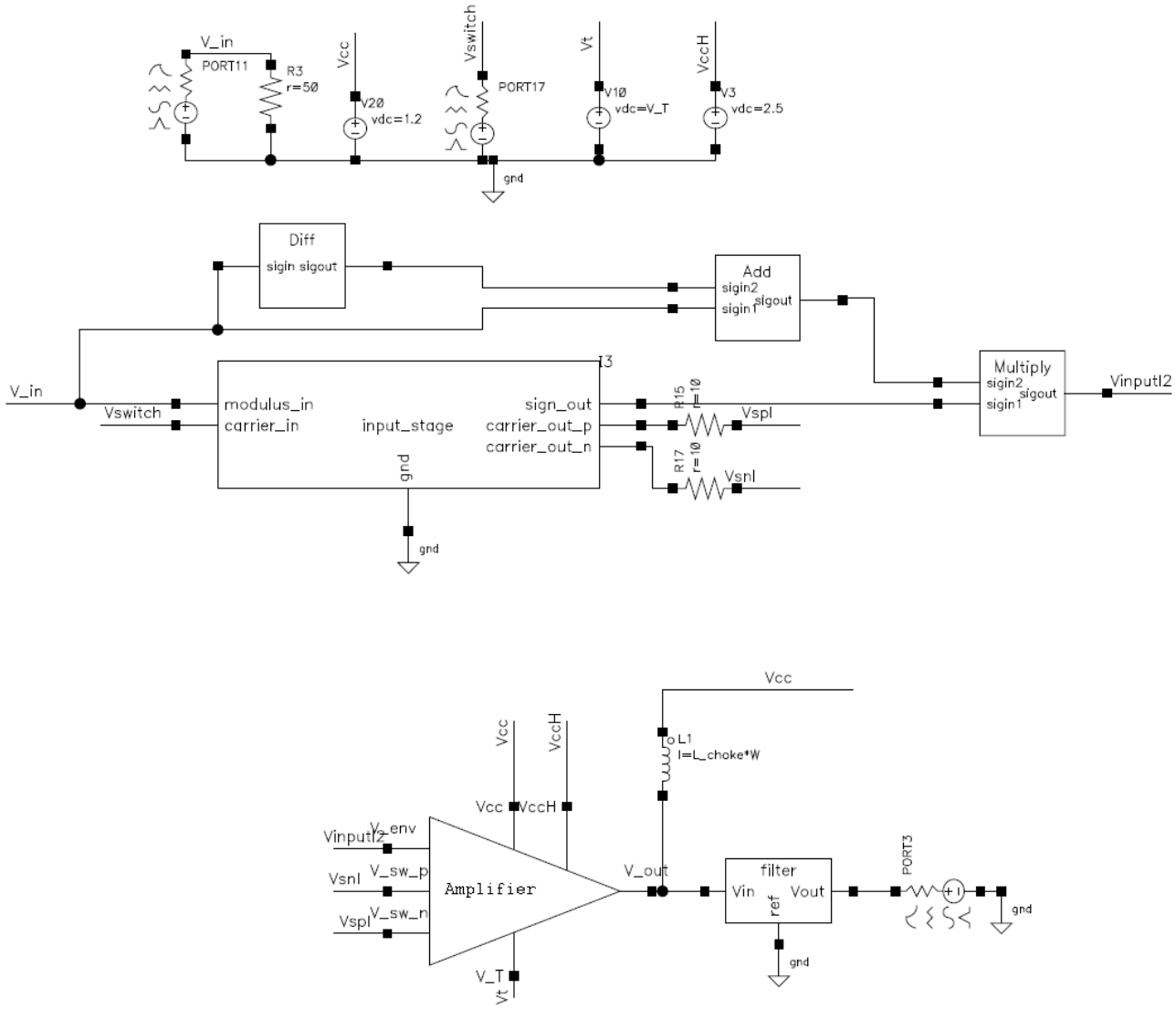

Figure B-1: Top level schematic

Figure B-2 shows the amplifier schematic. At the left top is the switch control buffer, which buffers the input signal and generated the required switch signals. Directly at the right hand side of it is the input current source, while the threshold voltage source is at the bottom.

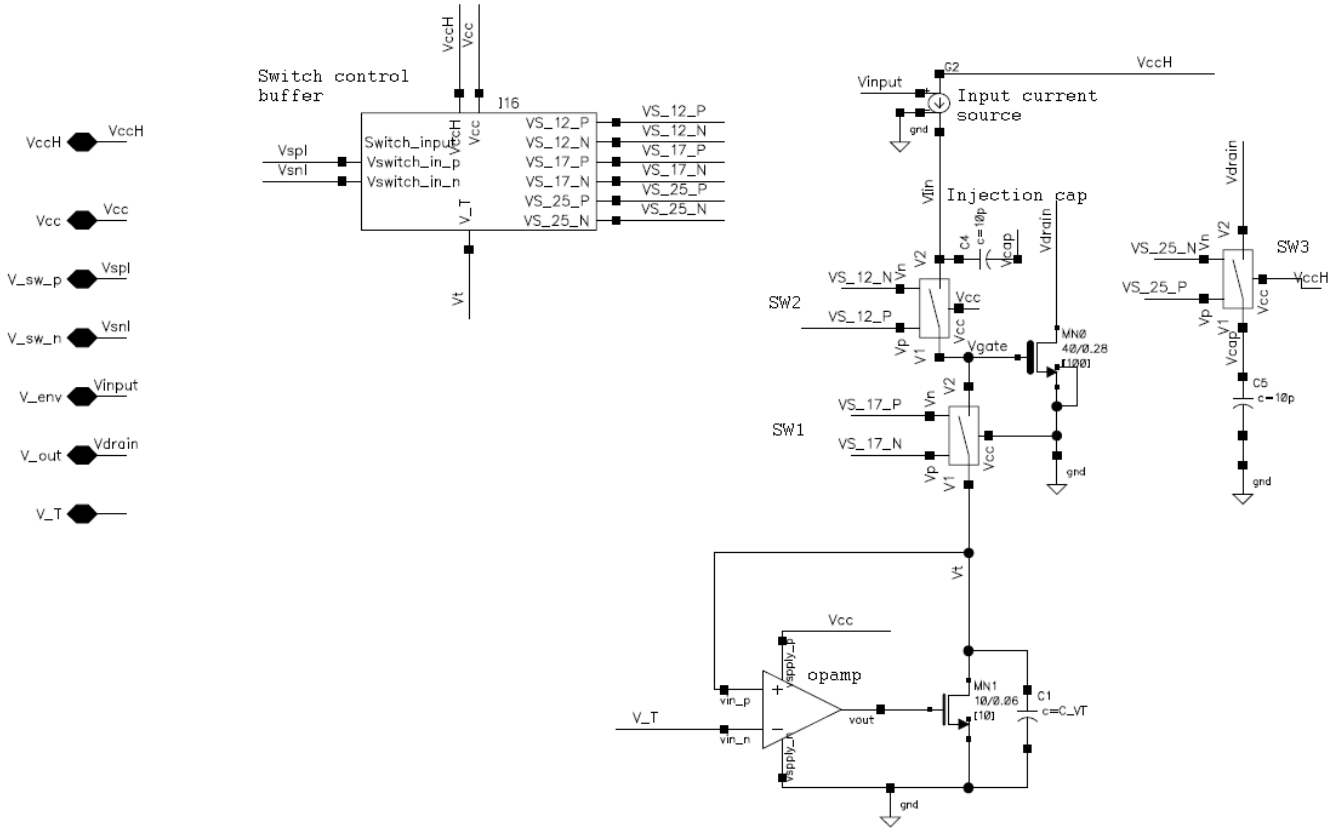

Figure B-2: Amplifier schematic

The switch buffer, the part with most transistors, is shown in Figure B-3. The voltage sources only are used to measure current, they do not influence the circuit. At the left top the required 1.7V is created. Below that is the initial amplification and buffering of the input. The right side is used to generate the signals with different voltage levels.

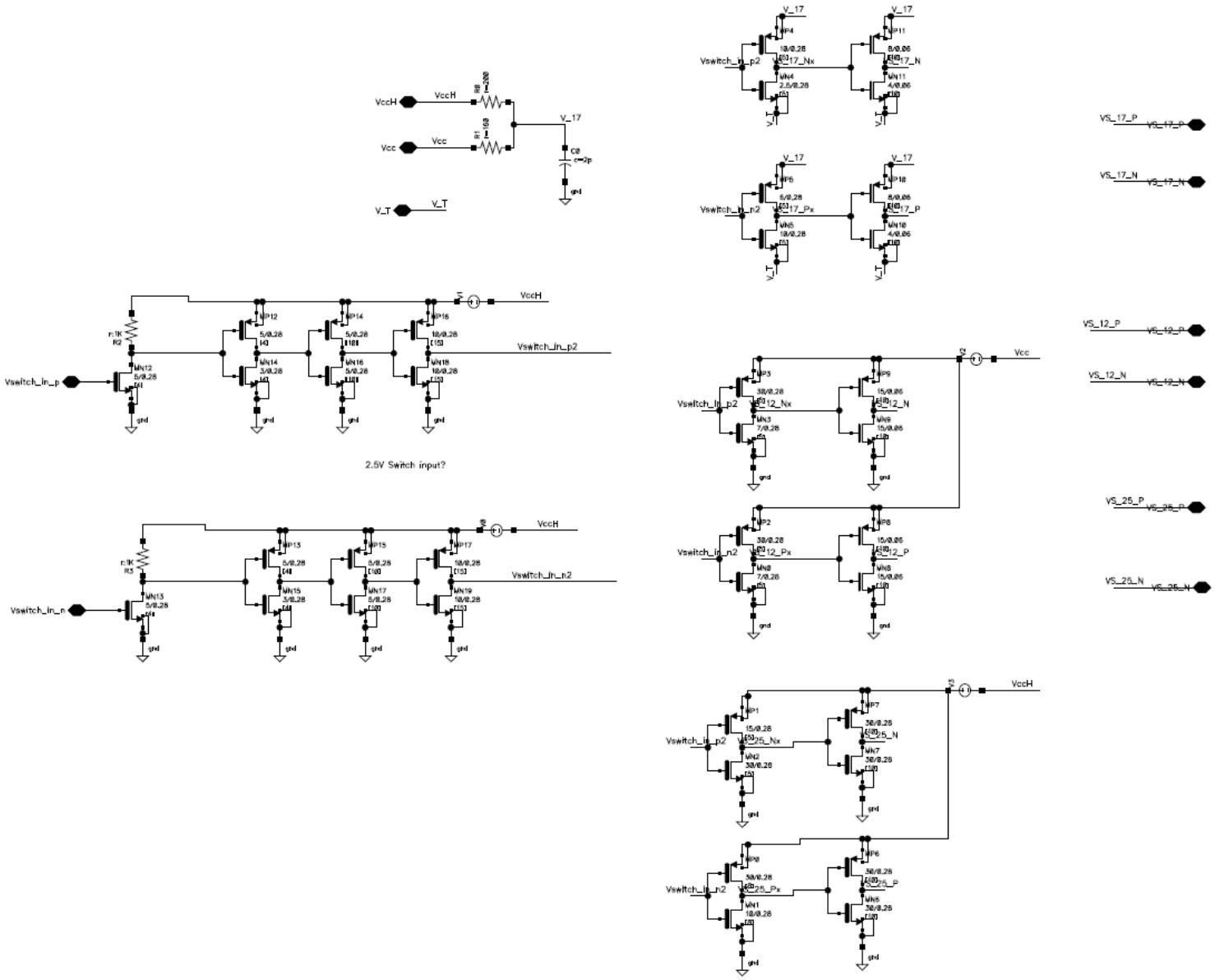

Figure B-3: Switch buffer schematic

Figure B-4 shows the input current source.

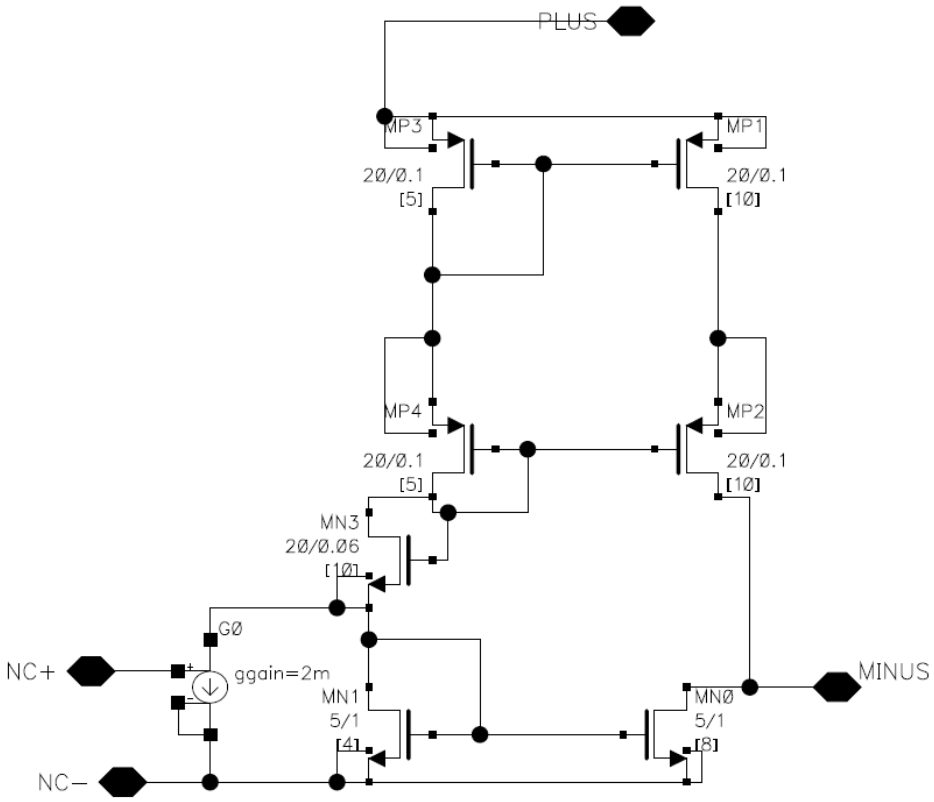

Figure B-4: Input current source schematic

Figure B-5 shows the OpAmp schematic.

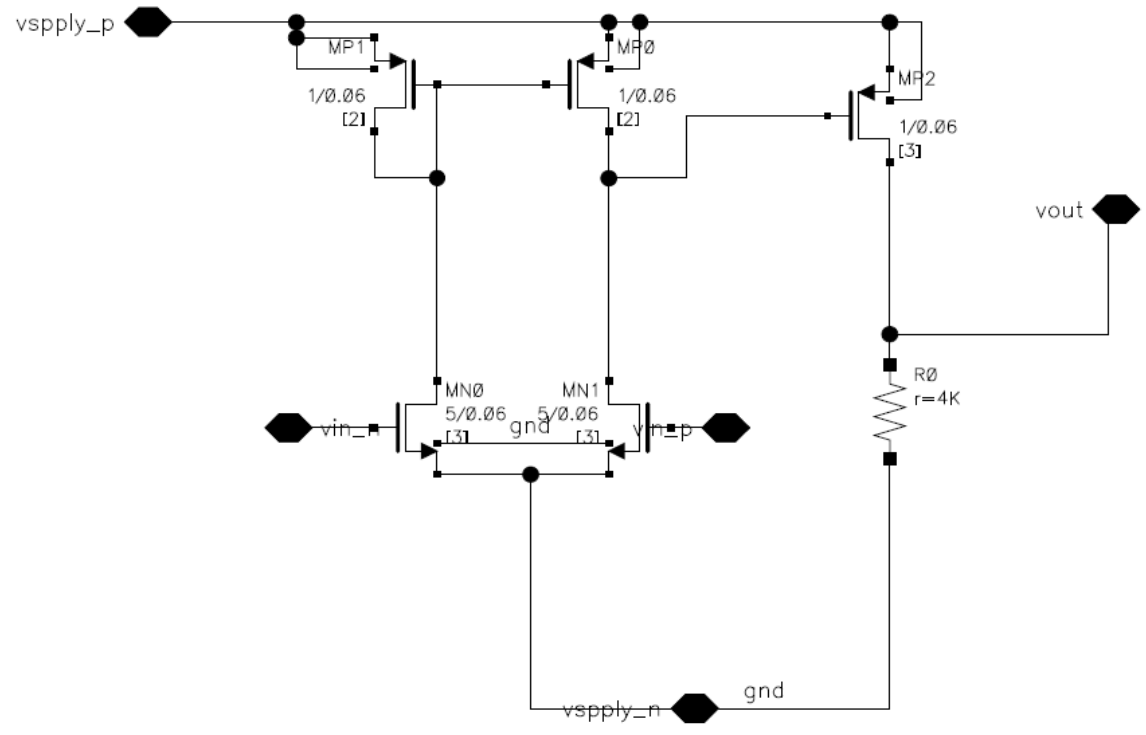

Figure B-5: OpAmp schematic

The switches are shown in Figure B-6, Figure B-7 and Figure B-8. The resistor in switch 1 is present to make sure the Vcc input is connected to something. This is only used for the bulk of PMOS devices; the connector is present to make sure all switches use equal connections, so a switch can be directly changed for another one.

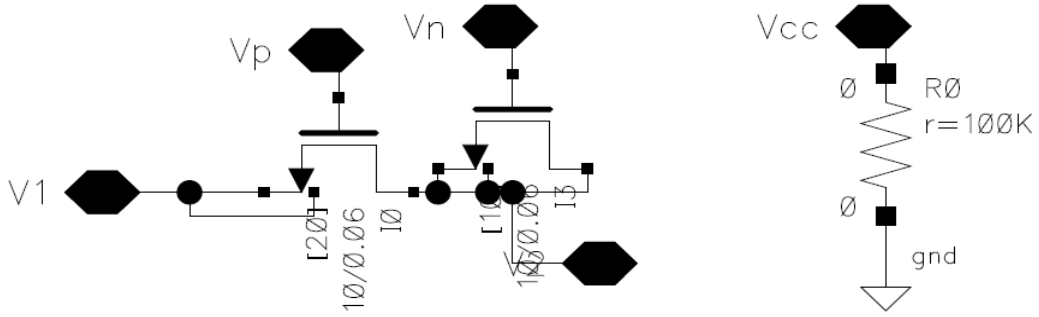

Figure B-6: Switch 1 schematic

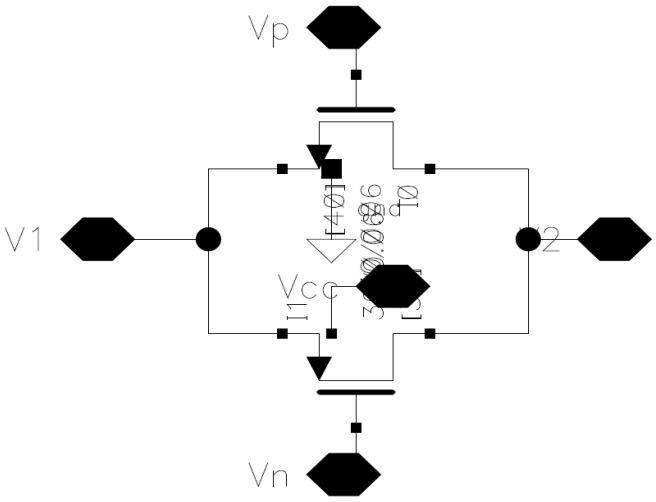

Figure B-7: Switch 2 schematic

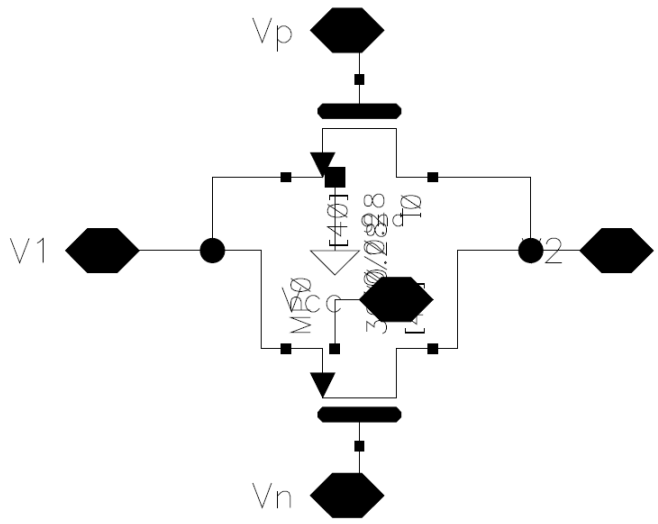

Figure B-8: Switch 3 schematic

Figure B-9 shows the output filter schematic, including the W-parameter. The '-1.4p' factor of the left-hand side capacitor is to compensate for the parasitic capacitances at the drain side of the amplifying transistor.

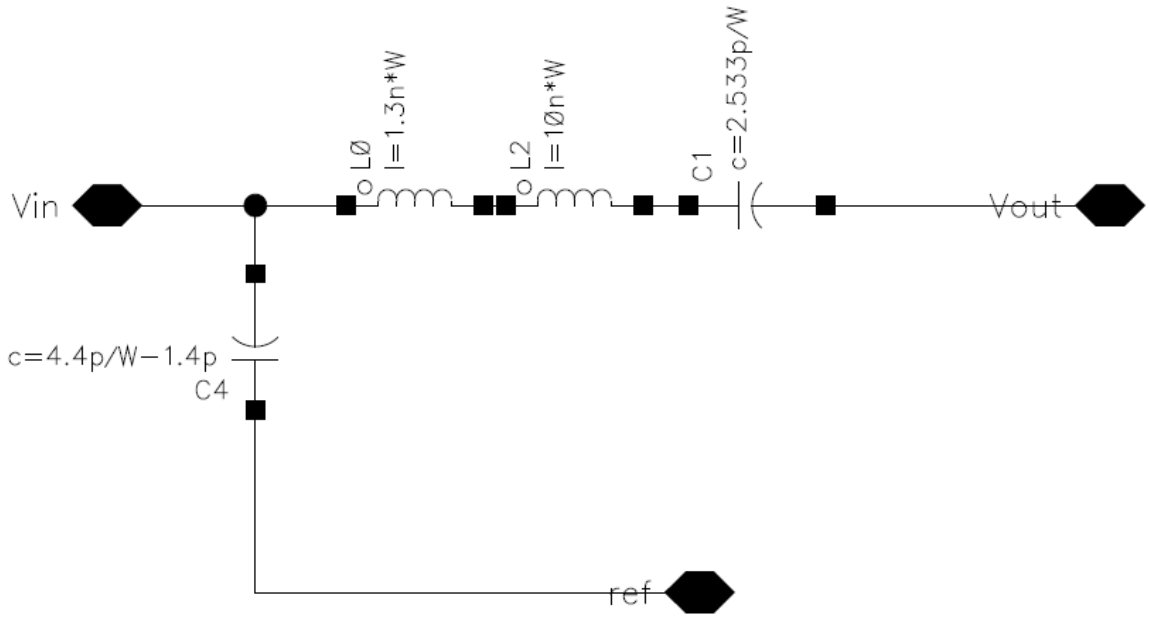

Figure B-9: Output filter schematic