RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

Sebastian Millner

An Integrated Operational Amplifier for a Large Scale Neuromorphic System

Diplomarbeit

HD-KIP-08-19

KIRCHHOFF-INSTITUT FÜR PHYSIK

Faculty of Physics and Astronomy
University of Heidelberg

Diploma thesis
in Physics

submitted by
Sebastian Millner
born in Witten
2007

An Integrated Operational Amplifier for a Large Scale Neuromorphic System

This diploma thesis has been carried out by Sebastian Millner at the
Kirchhoff-Institut für Physik
under the supervision of
Prof. Dr. K. Meier

Ein integrierter Operationsverstärker
für ein großskaliges neuromorphes System

In dieser Arbeit wird ein Operationsverstärker für die neuromorphe Hardware des FACETS-Projekts entworfen. Zwei Versionen dieses Verstärkers wurden entwickelt. Das Masken Layout wurde erstellt und zu Testzwecken wurde ein kleiner Prototyp-Chip in der Form eines ASICs designed und produziert. Um die Funktion des Verstärkers zu validieren wurden umfangreiche Messungen durchgeführt. Die Spezifikationen stellen eine große Herausforderung dar. So sollen sowohl Eingang als auch Ausgang beide Versorgungsspannungen erreichen koennen. Bei geringem Ruhestrohm und einer Versorgungsspannung von nur 1,8 V sollen eine hohe Bandbreite und unbegrenzte Stabilität möglich sein. Die entwickelten Verstärker erfüllen alle Spezifikationen.

Die Verstärker werden als Parameter-Impedanzwandler und als 50 Ω Ausgangstreiber im analogen ASIC HICANN des FACETS-Projekts verwendet werden. Dieses auf Waferebene integrierte, neuromorphe System stellt eine Experimentierplattform für die Neurowissenschaft dar.

An Integrated Operational Amplifier for a Large Scale Neuromorphic System

In this thesis, an operational amplifier for the FACETS neuromorphic Hardware device is designed. Two versions of the amplifier have been designed, the mask layout has been created and a full custom ASIC prototype has been produced. Extensive measurements have been done to validate the functionality and the performance. Challenging are features like rail-to-rail input and output voltage range, low quiescent current, 1.8 V power, a high bandwidth and unlimited stability. The designed amplifiers match the given benchmarks.

The Amplifiers will be used as voltage parameter buffer and 50 Ω output driver in the analog ASIC HICANN of the FACETS stage 2 wafer scale integrated neuromorphic hardware device, implementing an experimental platform for neuroscience.

Contents

Chapter 1

Introduction

Most physicists only know operational amplifiers as small IC's they used in experiments. The concept of operational amplifiers goes back to the early 40s of the last century, when John R. Ragazzini[19] used them for the analog computation of dynamical physical systems. Of course, these amplifiers where huge and build with tubes. The classical one-chip operational amplifier is the uA741. Thousands of different opamp ICs are available nowadays. Of course, the amplifiers designed in this work are no single chip opamp ICs although it would be possible to produce them like this.

The designed amplifiers will be integrated into a full custom mixed signal ASIC and are highly specialized circuits while still allowing a multi purpose use in the chip.

Why does a physicist design operational amplifiers? First of all, operational amplifier design is complex and a great challenge. Lots of physical background is needed to perform analog ASIC design and to understand it in addition.

The amplifiers are a tiny part in a huge project called FACETS, with the final goal of designing a neuromorphic hardware device allowing experiments in neuroscience that have never been possible so far. The physics of the brain is emulated by electronic circuits in a VLSI system. Operational amplifier design enables me to be part in this fascinating project and to gain the electronic back ground knowledge needed to be capable of doing neuromorpic circuit design.

Before I come to the structure of my thesis, I want to bring in some philosophical aspects about physics and the brain. When I first learned about atoms at school, I thought about the brain consisting of atoms, too. With the naive knowledge I had at this time, I simplified the brain to a deterministic system resulting in the conclusion that neither free will nor consciousness could exist. Switching to chaos physics later to find some antithesis did not help. Having no access to enough values to perform deterministic calculations does not mean the mathematically exact values do not exist in nature. Because no one can hear the tree fall, it does not mean that there was no noise. Some antithesis can be found by the comprehension of quantum mechanics. These philosophical aspects of classical physics are commonly known and have been discussed by R. P. Feynman in his lectures on Physics [7] for instance. The major error is the border crossing between natural science and arts and the abstraction between the system and its parts. Maybe the brain is fully deterministic - but who cares? Still the exploration of physics in neuroscience is a fascinating field of activity.

Additionally my personnel believe is that no "real" artificial intelligence is possible on a classic touring machine based digital computer with discrete states.

Outline

After a short insight into the FACETS project and the work done in our group, this thesis is separated into 4 parts.

Analog ASIC Design In this chapter, some basic background knowledge needed for ASIC design is given. This includes simulation techniques, design flow aspects, transistor modelling and some information on the used process.

Fundamentals of Circuit Design This chapter starts with a general introduction of operational amplifiers, next the specifications and functions of the designed amplifiers are discussed followed by some feedback issues. Basic circuits and design techniques used in the next chapter are introduced.

The Amplifier Design Here the design of the amplifiers of this thesis and their layout is described. Right at the beginning a switch between the theory above and the concrete process is done, and then the amplifier basic schematic is build up step by step, increasing complexity. The aspects of the final schematic are discussed for each flavour of the amplifier. At the end of this chapter, the layout is created. This section starts with some background information about matching and then explains the layout of important matched transistors.

Experimental Verification In this chapter, the amplifiers have to proof their functioning in realty. A prototype ASIC is developed. The experimental setup used during measurement is described, including a small custom PCB. Last but not least, experiments exploring the specifications of the amplifier are carried out and evaluated. Results are compared with simulations.

Chapter 2

The FACETS Project

This work has been carried out within the FACETS[1][1, 2] project. The FACETS project is a consortium of 15 European groups aiming at the exploration of new computational paradigms beyond the Touring Concept to understand the functioning of the brain. This involves biological measurements, modelling and simulations and finally the development of a neuromorphic hardware device based on the results.

The neuromorphic hardware device is necessary for neuroscience since even today's computational power of digital computers is limiting possible experiments with theoretical models. Simplifying the brain as a highly complex dynamical system, a huge number of differential equations has to be solved numerically using a digital system. Even the BlueGene super computer used in the Blue Brain project[3] is much slower than biological time for simulating systems of 10^6 neurons[2]. The visual cortex has about 10^9 neurons. Large scale long term learning experiments are not feasible. Limits will not disappear within the next 50 years, even if Moores Law would continue[2] which is not the case, as micro electronics approach physical boarders.

In the neuromorphic hardware, differential equations are not solved but rebuilt - the model is not simulated but emulated. Biology is implemented according to a model by the use of equivalent electronic circuits. VLSI implementation allows high synapse counts and a speed up factor of up to 10^5 in comparison to biological time[4]. The used model is a transconductance based model of spiking neural networks.

The hardware implementation is divided into two parts called "Stage 1" and "Stage 2". The major part of "Stage 1" is the ASIC "SPIKEY" [8] implementing 384 neurons and about 100,000 Synapses. 16 of these chips can be connected to each other via a backplane systems[9], allowing a total neuron number of 6144, but network traffic is a limit and the input count of each neuron does not scale. Stage 1 is used to validate the principle of the neuromorphic hardware and to allow first experiments. Some results can be seen in[5].

For larger scaled networks "Stage 2" comes into account. The concept is to build a large mixed signal chip using wafer-scale integration. Connections between the reticles of one wafer will be done by post processing, crossing the scribe line[2] by an additional metal layer. Synapse addresses and interconnectioning of neurons allow more synapses per neuron[6].

A prototype Wafer without any layers except metal 6 has already been produced and is currently in Berlin for post-processing to verify the feasibility of the concept. The post-processed prototype wafer will be tested in a setup designed by Dan Husmann [10], which is similar to the setup later to be used by the real wafer. Each reticle of the wafer consists of 8 analog network chips called HICANN (High Input Count Analog Neural Network). No full reticle is used to retain the possibility to produce prototypes in cheaper MPW (Multi Project Wafer) runs. The

[1] Fast analog Computing with Emergent Transit States

[2] Space between two reticles used for test structures by chip fabs

first prototype of the HICANN will be submitted in 2008. The amplifiers designed in this thesis are dedicated to the HICANN chip.

Chapter 3

Analog ASIC Design

The abbreviation "ASIC" stands for "Application Specified Integrated Circuit". In this work, operational amplifiers for the FACETS project have been designed and prototyped on a small ASIC. This chapter gives an insight into the techniques used for analog ASIC design and some background knowledge on CMOS models.

3.1 Simulation

Here we describe the functional principle of analog and digital simulations. A comparison is hardly possible as the concepts are different.

3.1.1 Digital

In the digital world, time is discrete. Signal changes are caused by the change of other signals like a clock for example. If we look at an idealized register, for instance, the stored value may be changed at the positive edge of the clock signal. There is nothing to simulate for this value during the rest of the cycle. The simulator can wait for the event positive clock edge and than start calculating.

If feedback is applied - the result of a circuit has influence on the input - a way has to be found to model this sort of continuous process. Time is discretized in so called delta times to simulate actions happening in parallel. All operations done in this cycle are executed on the system state at the beginning of the cycle. The system state is changed after all operations are done. This way a sequential working CPU is able to simulate parallel functioning.

Digital circuits can be described by hardware description languages like VHDL or VERILOG which are compiled into a net list.

3.1.2 Analog

Good explanations about analog simulators can be found in [29, 28] for instance. As sample of an analog simulator, I refer to SPICE[1][28]. An analog simulator is much more complicated. First of all, it is harder to trigger the simulation by events like in a digital simulator as there are devices like capacitors for example which have a time dependent voltage.

What an analog simulator does at the beginning of each simulation is the calculation of the DC working point. In a DC analysis, time-dependent components like capacitances or inductors are excluded. On the system left, Kirchhoff's laws are applied and non linearities are linearized, differential equations are numerically integrated. Solving the system of equations left is nothing more and nothing less than matrix inversion.

[1] Simulation Program with Integrated Circuit Emphasis

After the DC operating point has been found other simulations can start. For an AC analysis, the small signal model at the DC operating point is used for nonlinear components to calculate the transfer functions, which is again nothing more than finding the solution for a system of linear equations.

The transient analysis is more complicated, as the working point is a function of time, too. This is why the step size is really important here. If there are changes in the system, the time interval, the equations are solved has to be diminished. If you perform analysis, this can directly be seen in the simulation log. At each point of the analysis, the operating point of each device and so their transfer function might have changed - the linear system of equations has to be newly arranged.

In principle an analog simulator like SPICE can be used for any system based on a continuity law[2]. The complexity of an analog simulation is much higher than that of the digital as each step an matrix inversion has to be done.

Circuits are described by schematics or languages like *Verilog A*. In an analog hardware description language, differential equations can be assigned to parameters like current and voltage. Input of the simulator is a net list again. Models of special components like Transistors are usually included in the simulator and programmed in C. The transistor model used in this Work is the BSIM 3.3 v2[26] model from Berkeley.

One particular point about the BSIM 3.3 v2 model is that all operating zones of the MOST are implemented via one equation. In earlier models, there were some problems at the transitions of the zones. The resulting model is very unhandy.

3.1.3 Boundaries

Complexity is a boundary for both types of simulations, whereas logically much larger systems can be simulated by a digital simulator. Timing delays caused by routing can be modeled in digital simulations, but other analog effects like jitter or unclear clocks set limits. Boundaries of digital simulations are set by analog behaviour.

The analog simulation depends on the accuracy of the models used. With proceeding technological improvement, smaller transistors are developed and new effects have to be included in the models to be still an effigy of reality. Channel length modulation is hardly an issue for a transistor with $3\,\mu$m channel length for example, whilst the influence is drastic for a 180 nm device.

The accuracy of the convergence during the numeric solving of the system of linear equations [3] of the analog simulation can be user set. A too high convergence border could result in a wrong solution for the equation system. It is possible, that no convergence is found during the solving of the system. User interference by setting start values can help in this case.

The general border of simulations are process variations whilst here a digital system may not work at all, whilst an analog system works different. The real parameters are not exactly the ones given in the model, but include real world variations. Process variations can be modeled by doing worst case simulations with guaranteed boundary parameters given by the producer of the chip. Of course no continuous variations of these parameters are possible.

There are not only variations between different chips, but inside a single chip, too. On some analog circuits like differential stages, this can have really bad influence as it causes an offset not included in the simulations.

[2] Maybe it is a good idea to use a thread-able circuit simulator as simulator core for spiking neural networks using ideal capacitance and transconductance components.

[3] Iterative algorithms for this problem can be found in [11] for instance

3.2 Design Concepts

3.2.1 Engineering Techniques

To get an insight into engineering techniques in general, I took a course in software engineering[30]. These techniques are mainly usable for the design of a large scale system, whilst the design done in this work is only a small component. The major concept is to start lots of work in system design before the single components are designed. At first requirement engineering is done setting the requirements and constraints of the system. The next step is to design the system itself, using abstract schematics like class diagrams or block diagrams without specifying the concrete data types or bus sizes. Now the components or classes are substantiated with specific functions and interfaces. Finally, the components are designed. During this top-down method, test cases or test benches are developed for system level, integration level component level and for a final test according to the abstract requirements. The concrete design proceeds bottom-up then. The components have to pass their test are tested in the integration test are then assembled to work in the system which has to pass the system test. The described model is called W-model and is an extension of the V-model as tests are developed from the beginning on.

Of course the use of such a predefined process is very critical, when no standard products are produced, the requirements are at the limit of technical feasibility and creativity is needed. It could be obstructive in science, where understanding and knowledge are the major issue.

A strait forward design like described here was not possible in this work.

3.2.2 Analog Design Flow

The first step is the specification. At first the goal was just to design a rail-to-rail amplifier as voltage buffer, and then further specifications like the quiescent current, the needed bandwidth and the load have been defined. Some specifications changed during the design as they where physically impossible.

Now circuit design started dividing the problem logically into input stage, output stage and compensation. After a fitting circuit was found it has been adapted to the specifications. As start point for simulations, calculations have been done. After the specifications have been verified in the normal simulations, corner simulations have been carried out to prove the tolerance on process variations. In this work, there is a slow corner (s), a typical corner (t) and a fast corner (f) for each transistor type. Combinations of these corners have been simulated. snfp stands for slow NMOS and fast PMOS for instance.

Next step was the creation of the layout of the circuit. Afterwards, a parasitics extraction was done followed by more simulations. Now the circuit was prototyped.

The final verification of the circuit is the analog measurements and a comparison to simulations.

3.3 A brief Description of MOSFETs

MOSFET stands for Metal Oxide Semiconductor Field Effect Transistor. Actually todays metal oxide is mostly silicon oxide. We use PMOS for p-type metal oxide semiconductor and NMOS analogue. In C(omplementary)MOS technology, both transistor types are available.

Here we will give a short overview of the MOSFET model used for hand calculations. In simulations, the BSIM 3.3 v2 model [26] is used. The simplified model used here is explained in detail in [17] or in even more detail in [23] for example.

A standard CMOS transistor has four contacts. Charge carriers enter the device at the Source(S), are controlled by the Gate(G) and then drained to the Drain(D) connector. Additionally there is a contact for the substrate called Bulk(B).

3.3.1 Operating Zones

I_{DS} vs. V_{GS}

The characteristics of MOSFETs can be divided into several operating zones. We assume $V_{DS} \geq (V_{gs} - V_T)$ and $V_{BS} = 0$. The $I_{DS} - V_{GS}$-characteristic can be split into three zones: Weak Inversion, Strong Inversion and Velocity Saturation. In Weak Inversion,

$$I_{DS} \propto e^{\frac{V_{GS}}{nk_BT/q}} \tag{3.1}$$

where $n = 1 + C_D/C_{OX}$ is obtained by V_{BS} and never accurately known. The magnitude for the sample described [17] is 1.2 to 1.5. Weak inversion applies until $(V_{GS} - V_T) \approx 0.1\ V$ Now the Strong Inversion Area begins, where

$$I_{DS} = \frac{KP}{2n}\frac{W}{L}(V_{gs} - V_T)^2(1 + \lambda V_{DS}). \tag{3.2}$$

We will tell more $KP = \mu C_{OX}$ later. W is the channel width and L the length. During design, we usually use $K' = \frac{KP}{2n}$ and neglect λ in hand calculations. λ is caused by channel length modulation and can be reconsidered analogue to the inverse of the early voltage of a bipolar transistor. It is anti proportional to the channel length of course. Neglecting λ for small channel lengths is critical, as it is already 0.75 for 300 n channel length in our process[4]. Nevertheless, for $V_{DS} = 0.2\,\mathrm{V}$ which is the common choice, $\lambda * V_{DS} << 1$. The Strong Inversion area is the zone, transistors are normally biased in analog design as the transconductance characteristic is the most precipitous there. There is a problem with the transition to the next zone, as the transition happens at lower V_{GS} for smaller channel lengths. For a channel length of 350 nm the width of the strong inversion zone is only 0.3 V in [17]. For 180 nm it nearly disappears, but as the threshold voltage goes down here, the strong inversion area at 300 nm is larger in a $180\,rmnm$ process.

In velocity saturation, the transconductance does not change with $(V_{gs} - V_T)$ anymore, so I_{DS} is proportional to $(V_{gs} - V_T)$.

I_{DS} vs. V_{DS}

We assume a transistor biased in strong inversion. The characteristic is divided into two zones. The zone above $V_{DS} = V_{GS} - V_T$ is called saturation and the I_{DS} dependency is the same as described above then. For lower V_{DS} the transistor is biased in the so called linear region. Now the dependency is:

$$I_{DS} = KP\frac{W}{L}(V_{GS} - V_T)V_{DS} - \frac{V_{DS}}{2} \tag{3.3}$$

Of course, it is only linear for small V_{DS}. The resistance at zero V_{DS} is defined as R_{on}. Now $KP = \mu C_{OX}$ can easily be understood as $C_{OX} * (V_{GS} - V_T)WL$ is the conductor density in the channel, $\frac{V_{DS}}{L}$ is the applied field and L is the track of the conductors.

[4] This is a result of simulations done in this work

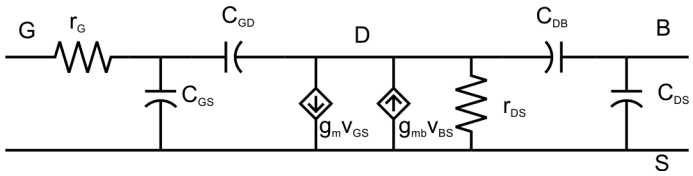


Figure 3.1: MOSFET small signal model, comp. [17]

3.3.2 Small Signal Model

Small signal parameters like the transconductance have already been used above. In the small signal model, only small changes at a certain point of operation are assumed. What is done is in principle nothing more than a Taylor series of first order - we look at the first derivative. All constant potentials can be set equal as what we are interested in is only the change. The major parameter of a CMOS transistor is its transconductance g_m which is defined as:

$$g_m := \frac{\partial I_{DS}}{\partial V_{GS}} \tag{3.4}$$

The next important term is the drain source resistance in saturation. It is the inverse of the increase of I_{DS} in saturation, and so λ is included.

$$r_{ds} = r_0 := (\frac{\partial I_{DS}}{\partial V_{GS}})^{-1} = \frac{1}{\lambda I_{DS0}} = \frac{V_E L}{I_{DS0}} \tag{3.5}$$

I_{DS0} is the drain source current without channel length modulation in this case.

An additional transconductance in the model is caused by the body effect - a parasitic JFET built via the substrate. This JFET is depending on V_{BS}. The resulting transconductance g_{mB} is working against g_m. The complete small signal model of a MOST transistor working in strong inversion[17], including all important capacitances, can be seen in Figure 3.1

The most important capacitor here is the gate source capacitance which takes the main part of the gateoxid capacitance.

3.4 The used Process

The process, this thesis is done for is a self aligning 180 nm mixed mode process with one poly silicon layer and 6 metal layers. Triple well technology allows the bulk isolation of NMOS devices. Additionally, a dielectric layer between metal 5 and metal 6 allows the use of precise metal capacitors with comparable small size. Discrete bipolar devices are possible, too but have not been used in this work. Resistors can be created with salicide technology, meaning self aligning siliciding of metal to perform contacs[27]. Resistors used in this work are poly silicon resistors.

Chapter 4

Fundamentals of Circuit Design

4.1 Operational Amplifier

Operational amplifiers were originally introduced by John R. Ragazzini in his work "Analysis of Problems in Dynamics by Eletronic Circuits" [19].

> "The term 'operational amplifier' is a generic term applied to amplifiers whose gain functions are such as to enable them to perform certain useful operations such as summation, integration, differentiation, or a combination of such operations."

The planned use of such an amplifier was the simultaneous solving of differential equations of dynamical problems in physics. An example given by Ragazzini is the control of the elevator of an airplane. The principle is similar to the idea of the FACETS project although the problem is less complex. A dynamical system is rebuilt by electronic circuits using equivalent observables, to gain access to the results of coupled differential equations without solving them analytically or numerically.

As an abbreviation for operational amplifiers, the term "opamp" is common. The output of an opamp is a voltage. If the output is a current, the term is "operational transconductance amplifier (OTA)".

4.1.1 Modeling

There are three golden rules for the ideal operational amplifier:

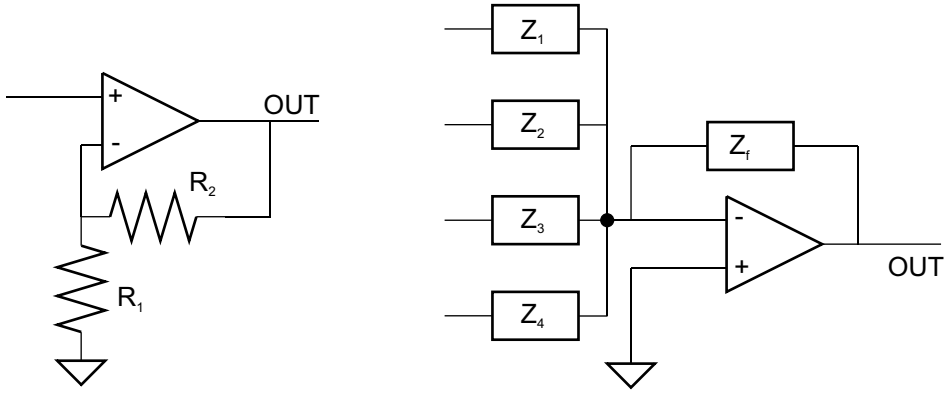
1. The input resistance is infinite.
2. The output resistance is zero.
3. The gain is infinite.

Using these rules, circuits can be constructed to perform analog computation. Two examples are shown in Figure 4.1. The first circuit is a non inverting amplifier. Calculation of the gain is easy if we take into account that the potential on both input nodes is equal because of the infinite gain. The resistors form a voltage divider and so the gain has to be:

$$A_V = \frac{R_1 + R_2}{R_1} = 1 + \frac{R_2}{R_1} \tag{4.1}$$

An amplification of one, or buffer mode is achieved by short cutting R_2 and removing R_1.

The second circuit, shown in Figure 4.1 b) is an inverting amplifier with some additional impedances. At first, we exclude $Z_{2,3,4}$. The amplification is



(a) Non inverting Amplifier (b) Inverting Amplifier

Figure 4.1: Basic Opamp Circuits

$$A_V = -\frac{Z_f}{Z_1}. \tag{4.2}$$

Then, as both inputs of the amplifier are on the same potential, the input resistance is infinite and Kirchhoff's current law applies at the negative input node of the amplifier. If we then include $Z_{2,3,4}$, a sum is evaluated:

$$V_{OUT} = -\sum_i \frac{Z_f}{Z_i} \tag{4.3}$$

Here again just Kirchhoff's current law has been applied at the input node. If complex impedances are used, the circuit can be integrating or differentiating analogue to low-pass and high-pass. Using several operational amplifiers, complex differential equations can be evaluated.

4.1.2 Specifications

Of course, no real amplifier is able to achieve the golden rules. A nearly[1] infinite input resistance for low frequencies is possible, using CMOS devices, but unlimited gain and a zero output resistance are impossible. As each amplifier has capacitances included, there is always some kind of low pass behavior resulting in a border frequency and a phase shift. Furthermore several specifications apply to real amplifiers. We will discuss some here. Most of them will be used in this thesis.

Bandwidth

The bandwidth of an amplifier is defined as the frequency range, where the voltage gain stays above $1/\sqrt{2}$ or $-3\,\mathrm{dB}$ of the max voltage gain - this is equal to a power loss of one half. The bandwidth is given by the cut off frequency of the nodes in an amplifier.

Gain Bandwidth Product

The gain bandwidth product or GBW is the nothing more than the name tells. Theoretically, the GBW stays constant when the gain is adjusted via feedback as the low passes are balanced by the intern gain. This way, the GBW is equal to the bandwidth in a closed loop configuration, where the amplification is one.

[1] There is always some leakage current via tunnel currents and impurities in the gate oxide

Common Mode Rejection Ratio

A real differential stage has a differential voltage gain A_{diff} and a common mode voltage gain A_{cm} The Common Mode Rejection Ratio(CMMR) is defined as:

$$\mathrm{CMMR} := \frac{\mathrm{A_{diff}}}{\mid \mathrm{A_{cm}} \mid} = 20\log(\frac{\mathrm{A_{diff}}}{\mid \mathrm{A_{cm}} \mid})\mathrm{dB} \tag{4.4}$$

The CMMR of the classic uA741 is $90dB$ or $32k$ for instance [27].

Slew Rate

The Slew Rate (SR) is defined as the maximum possible change velocity of the output voltage.

Equivalent Input Noise

The equivalent input noise is the noise needed at the input port to achieve the noise produced at the output port according to the gain.

Power Supply Rejection Ratio

The Power Supply Rejection Ratio (PSRR)[31] is defined as

$$\mathrm{PSRR} = \frac{\mathrm{A_V}}{\Delta \mathrm{V_{OUT}}/\Delta \mathrm{V_{Power}}} \tag{4.5}$$

and connects changes in power to changes in output. It should be as big as possible.

Figure of Merit

To evaluate a design, it is possible to define an abstract Figure of Merit (FOM). The FOM used in this work is the same used in [17]:

$$\mathrm{FOM} = \frac{\mathrm{GBW} \cdot \mathrm{C_{load}}}{\mathrm{I_{quiescent}}} \left[\frac{\mathrm{MHz} \cdot \mathrm{pF}}{\mathrm{mA}}\right] \tag{4.6}$$

In [17], the FOM achieved in sample circuits is in the order of magnitude of 600 MHzpF/mA.

4.2 The Designed Amplifiers

In this work, two amplifiers are designed, using the same schematic as base. The future job of the amplifiers and the resulting benchmarks to be reached in this work are presented in the following.

4.2.1 Voltage Parameter Buffer

In the FACETS stage 2 system, neurons and synapses have to be controllable by lots of parameters, namely control voltages and currents. Most of these parameters will be individually adjustable for each neuron by floating gate cells designed by André Srowig and Jan-Peter Loock [24]. Nevertheless, there are global parameters as well. Individual parameters are especially not possible in the synapse array. Global voltages have to be driven. As the set of parameters is not clear, the capacitance which has to be loaded cannot be defined. Additionally depends on the function of the control voltage. This is why the driving amplifier has to be unlimited stable meaning it should not be unstable for any load capacitance. Furthermore it should be able to

drive large capacitances which means it has to source and drain relatively high currents. The aimed load is 100 pF, which is connected via a 50 Ω resistor to facilitate stabilization[6]. The GBW should be the low-pass frequency of the given output RC circuit, which is round 32 MHz.

For some parameters, like the *leakage current control voltage* for instance, it is necessary to reach the ground rail. Amplifiers of stage 1 are not able to do so, so needed small voltages have to be plugged external. This is not possible for a huge multi-chip wafer-scale system like FACETS stage 2 as spacial expensive external pins would be needed. In FACETS Stage 2 these voltages have to be generated inside the chip. Hence, an output buffer is needed, which is capable of reaching both rails. As the amplifications is one, this ability has to be fulfilled by input and output.

In the FACETS stage 2 System, lots of chips are working in parallel on a very small area. Power has to be transported to the chip and absorbed by cooling devices in return. The total power of one Wafer should be in the magnitude of 1 kW which is around the power of a small stove top. A whole wafer has 60 entire reticles[10] and each consists of 8 HICANNs leaving $2W$ power for each HICANN. Every single microwatt costs, so the voltage buffer has to be small in power consumption. The benchmark for the quiescent current is set to 200 μA. In addition the current consumption needs to predictable. The quiescent current needs to be independent of the common mode DC input voltage and should be controllable by biasing. The whole amplifier should only have one biasing pin.

Lots of signals will be routed above the amplifier and the coupling between those routes and the amplifiers should be minimized. This is why metal layers 5 and 6 are not allowed in layout and metal layer 4 can only be used for constant potentials like ground and power. Nevertheless, fixed pattern noises can be a problem - no bad CMMR is allowed.

The DC-offset of the amplifier should be small, so there has to be a high gain at low frequencies. Influence of on chip threshold voltage - and doting gradients has to be minimized.

4.2.2 50 Ω Output Driver

Voltages like the *membrane potential of the neurons* can be directly measured on an oscilloscope via 50 Ω transmission lines in the FACETS stage 1 system. The output drivers have a high quiescent current, depending strongly on the input voltage. The amplifiers are hot spots on the chip when observed it with an infrared camera. The idea is now to reuse the schematic of the Parameter Buffer with its low current consumption to get a more power efficient 50 Ω output driver for the stage 2 system. These output drivers are primarily needed for the prototypes of the HICANN. The benchmarks for this amplifier are a GBW of 350 MHz when driving a 50 Ω transmission line and an input independent quiescent current of 2 mA. It is not necessary to reach the rails - there can be an offset of 0.5 V to the power rail. I aimed for a maximum output voltage of $1.6V$ which is equivalent to a current sourcing of 16 mA at a 100 Ω load (50 Ω series output resistance and 50 Ω at the end of the line). The 50 Ω output resistance is inherent part of the design again.

4.3 Feedback and Stability

Feedback has been introduced by H. S. Black in his classic paper "Stabilized Feed-Back Amplifiers" [13] and is a major concept in nowadays amplification. Here we will concentrate on what it means for our design, what problems are implied and how to solve them.

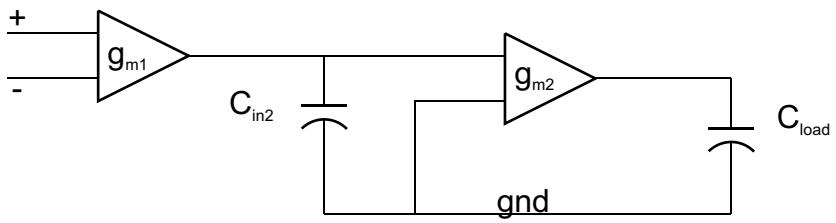


Figure 4.2: generic two stage amplifier

4.3.1 Why Feedback is necessary

Process variations cause each amplifier to be individual. The output offset and the total gain vary. To adjust a certain output voltage and a certain gain, feedback is needed. Our amplifier will be used as a voltage buffer with an amplification of one. The resulting circuit is a closed loop configuration - the output is directly connected to the negative differential input pin. This way, the amplifier works against the difference of input and output and so they are equal in the ideal case.

4.3.2 Poles, Peaking and Ringing

If a phase shift of 180 ° occurs at the output and the amplifier is connected in a closed loop, the negative feedback becomes positive. Phase shifts are created in an amplifier by low pass filters in the small signal model. Roughly speaking, the output node of each stage of an amplifier causes such a low pass if loaded by a capacitance. Each stage has an input capacitance, so looking at a two stage amplifier we simplified have a situation similar to the one shown in Figure 4.2. Of course, each low pass not only creates a phase shift, but also attenuation. The phase shift is only a problem, if the total amplification is still above one. This usually happens in a small frequency range. The cut-off frequencies of the low-passes are called "poles".

Normally, the amplification of an amplifier in a closed loop configuration is limited by a certain frequency f_c, where it decreases because of the first low pass. If the amplifier is not stable and some positive feedback is applied, the amplification increases massively at f_c before the low pass attenuation starts dominating. This increase and decrease looks like a peak in the frequency response and so this phenomenon is called "peaking".

If we put a transient square signal on the input of such an amplifier, all frequencies of the Fourier spectrum of the pulse below f_d are amplified by one whilst f_d is accentuated. An oscillating signal can be observed which is called "ringing".

4.3.3 Calculation of the Transfer Functions

The transfer function of a single stage amplifier with a capacitive load is just the gain multiplied by the low-pass transfer function:

$$A_V = \frac{A_{V0}}{i\frac{f}{f_c} + 1} \tag{4.7}$$

Where $f_c = \frac{g_m}{2\pi C_{load}}$ is the cut-off frequency of the low pass. If we add a second stage, we have to multiply the transfer function of the second low pass:

$$A_V = \frac{A_{V0}}{(i\frac{f}{f_{c1}} + 1)(i\frac{f}{f_{c2}} + 1)} \tag{4.8}$$

This is generally the open loop transfer function of a two stage amplifier as treated in this thesis.

A feedback system can be described if we just apply that the output of a system should have influence on the input and concentrate on first order terms. We want to get the transfer function C of a system A whose output is routed back to the input via the transfer function of a system B. Before the feedback comes into action, the output is just the input multiplied by A when leaded back to the input, so the new input is $1 + AB$ now. Finally we get:

$$C = \frac{A}{1 + AB}. \tag{4.9}$$

We apply this to get the transfer function of a closed loop two stage amplifier. As $B = 1$ in this case, things do not get too complicated.

$$A_{CL} = \frac{A_V}{1 + A_V} = \frac{1}{1 + \frac{(i\frac{f}{f_{c1}}+1)(i\frac{f}{f_{c2}}+1)}{A_{V0}}} \tag{4.10}$$

With $\mathrm{GBW} = \mathrm{f_{c1}} \cdot \mathrm{A_V}$ and negligence of terms with $1/A_V$ we get

$$A_{CL} = \frac{1}{1 - \frac{f^2}{\mathrm{GBW f_{c2}}} + i\frac{f}{\mathrm{GBW}}}, \tag{4.11}$$

which is similar to the transfer function of a damped harmonic oscillator. Clearly, the stability is depending on the ratio of f_{nd} and the GBW. From literature [17], we get a sufficient phase margin at a ratio of 3. The phase margin is 72° in this case. A phase margin of this magnitude is a design goal for stable amplifiers.

4.3.4 Compensation

The main aspect of compensation is to achieve the needed phase margin of 72°. This is done by directly designing the amplifier for a second pole greater than $3 \cdot \mathrm{GBW}$. This can be done easily, as the second pole can be adjusted for a certain capacitor C_{load} via the change of the transconductance of the second stage. This transconductance results in a certain input capacitance C_{i2} of the second stage which is the load capacitance of the first stage. Now the transconductance of the first stage can be aligned for the wished GBW. The resulting amplifier is stable for capacitors not bigger than C_{load}.

The transconductance needed for the second stage is huge, in comparison to the first stage as C_{load} is usually big. The appliance of a miller compensation can reduce the needed transconductance of the output stage. A capacitance C_{comp} is connected between the output nodes of both stages. This reduces the pole frequency of the second stage as C_{load} is conducting for high frequencies, enlarging the transconductance for our output low pass. Exact calculations lead to a factor[17]

$$\frac{1}{1 + \frac{C_{i2}}{C_{comp}}} \tag{4.12}$$

whilst C_{comp} is added to the load capacitance of the first stage enlarging the needed transconductance here. The usual choice is $C_{comp} = 3 \cdot C_{i2}$.

4.4 Input Stages

Here we will discuss important aspects of some input stages. We will need this knowledge later in the design chapter.

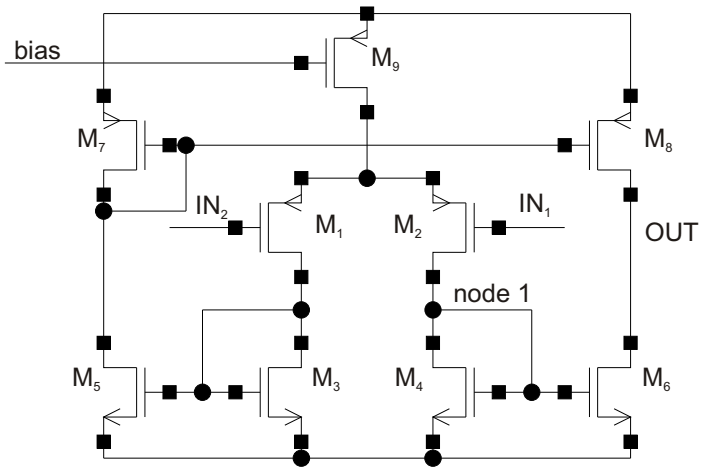


Figure 4.4: NMOS input stage

4.4.1 Simple CMOS OTA

The simple CMOS OTA is basically nothing more, than a differential pair with a current mirror load. The circuit can be seen Figure 4.3.

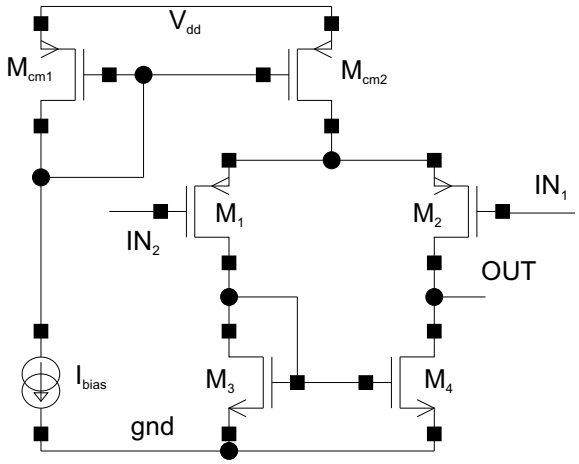


Figure 4.3: Simple CMOS OTA

The gain is given via $g_{m1} \cdot r_{out}$. If a resistive load is used instead of M3 and M4, the gain is bounded, as the resistance has to be chosen regarding the cross current to adjust the common mode output. Small signal and DC resistance are the same here. This is different if we apply a current mirror as done here. The common mode output can be adjusted by the transconductance of the mirror which creates the input resistance of the mirror. At the output node, the resistance is different, as the current through M_4 is not depending on the output potential in the small signal model. We have to use r_0 of M_4 which is much bigger - even that big that we have to include r_0 of the input transistor now, too. We get $r_{out} = r_{load} \parallel r_{in}$.

A very important note is, that it is not possible to use external biasing for the adjustment of the load cross current. This biasing would compete with the input current source. Without this constraint, even the simple CMOS OTA could be used for amplification down to one rail.

4.4.2 Symmetrical CMOS OTA

This circuit is introduced in [17] for instance. Here the load current of the simple OTA is mirrored as can be seen in Figure 4.4. This way, the load of both transistors of the input pair is exactly the same, resulting in a better symmetry which is the main advantage of this stage. CMMR and offset of the stage are improved. As the output current of the simple OTA is multiplied by the current mirrors (M4/M6 and M3/M5), the GBW is enhanced. Using B as current multiplication factor of the mirrors, the gain can be calculated as

$$A_V = g_{m1} B r_{out} = \frac{B g_{m1} V_{En} L_6}{B I_{DS1}} = \frac{2 V_{En} L_6}{V_{GS1} - V_T}, \tag{4.13}$$

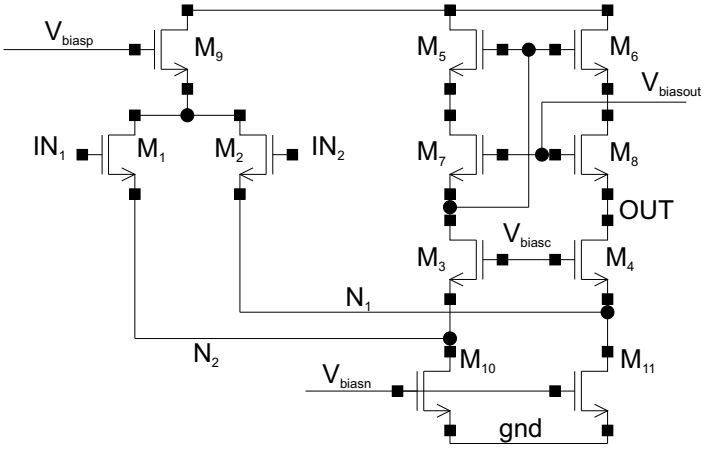


Figure 4.5: Folded Cascode OTA

where the channel length of M8 has to be big enough to neglect r_{DS} of M8. As the channel length of M4 and M6 are equal, the gain nearly matches the gain of the single stage amplifier above. The advantage is the exclusion of r_{DS} of M1. The bandwidth can be calculated as

$$BW = \frac{1}{2\Pi r_{out} C_L}, \tag{4.14}$$

which leads us to the enhanced

$$\mathrm{GBW} = \mathrm{B}\frac{\mathrm{g_{m1}}}{2\Pi \mathrm{C_L}}. \tag{4.15}$$

As the output node is an additional high ohmic node to node one and surely the leading one, we have to check the non dominant pole generated by node 1. The situation is similar to a two stage opamp. At node 1 several capacitors occur, whilst the resistance is given by the transconductance of M4. Appraising the maximum of the drain bulk capacities of M2 and M4 by C_{GS4} and implying $C_{GS6} = B \cdot C_{GS4}$, we finally get $C_{n4} = (3 + B)C_{GS4}$ which leads us to a non-dominant pole [17] at

$$f_{nd} = \frac{g_{m4}}{2\pi(B+3)C_{GS4}} \geq 3\mathrm{GBW}. \tag{4.16}$$

The B in the denominator sets a boarder to the maximum GBW reachable via the enlargement of B.

4.4.3 Folded Cascode

The folded cascode is one of the most common used input stages and many designers limit themselves to this design [17]. The gain bandwidth is equal to the GBW of a standard single stage OTA. Generally, there are two main advantages of this stage. The first one is the high frequency of secondary poles which simplifies the design and opens the GBW range to high frequencies. The second main advance is the capability of amplifying input voltages below one rail.[17] The circuit can be seen in Figure 4.5

The ability of crossing one rail at the input signal can be understood if we compare transistors M10 and M11 of this circuit to the transistors M3 and M4 of the symmetrical amplifier in Figure 4.4 or in analogous manner to the load transistor of the simple differential OTA. A diode connected most[2] needs at least $V_{GS} = V_{DS} = V_T + 0.2V$ to be working in strong inversion.

[2]most with drain and gate connected

Adding the saturation voltage of the input transistor, we get a voltage of around $V_S = V_T + 0.4V$. As the input transistor should be working in the middle of the strong inversion region, it needs $V_{GS} = V_T + 0.2V$, too. This is not possible though, if the gates of the input transistors are near ground. As M10 and M11 in the folded cascode are no diode connected most their V_{DS} can easily be below $V_T + 0.2V$ and so the rail is reachable for the input transistors. The current mirror load has to be used in the simple CMOS OTA, as a load like M10 and M11 in this circuit would cause additional biasing. It is only allowed for the folded cascode, as there is an additional branch through the cascode transistors.

To understand the DC operation, we assume a current of $50\,\mu$A for both input stages and the cascodes. Hence, the current source M9 drives $100\,\mu$A which are divided into the current for each input transistor. M11 and M10 have to drain $100\,\mu$A each as the current of the cascodes and the input transistors is summed. At last, the current of both cascode transistors is mirrored. The mirroring circuit used above has the advantage that the small signal resistance is very high as caused by the cascode transistors M7 and M8.

In the small signal model, the current difference at node 1 and 2, caused by a differential input voltage, is drained from the cascodes by M11 and M10. The small signal gain can be calculated using the resistance of the output node, r_{out} and the transconductance of the input transistors hence.

$$A_V = g_{m1} r_{out} \tag{4.17}$$

r_{out} is the parallel connection of the cascode amplified resistances of the current mirror M6/M5 and the parallel connection of the input transistor M1 and the load transistor M11:

$$r_{out} = r_{06}(g_m r_0)_8 \parallel (r_{01} \parallel r_{011})(g_m r_0)_4 \tag{4.18}$$

Calculating the gain bandwidth, which is the important characteristic in our design, r_{out} is truncated and we get the same result as for the single stage amplifier.

$$\mathrm{GBW} = \frac{\mathrm{g_{m1}}}{2\pi \mathrm{C_L}} \tag{4.19}$$

To obtain the second pole of the stage, we have to look at node 1 respectively 2. Again we have to look at the capacities on constant potential, which are the C_{GS} of the cascode transistors and the drain bulk capacitors of the input transistors and M11/M10. We estimate the maximum as $C_{n1} \approx 3 * C_{GS3}$. So the final frequency is

$$f_{nd} = \frac{g_{m3}}{2\pi 3 C_{GS3}}. \tag{4.20}$$

The current consumption is doubled, although the gain bandwidth stays the same. The gain is enhanced for small frequencies which is good to eliminate DC offsets in closed loop mode. The current consumption is higher than that of an equivalent symmetrical OTA, but the absence of a small frequency second pole allows a higher GBW. We need a rail-to-rail input swing, so we have to use the folded cascode.

4.5 Rail-to-Rail Input Stages

A rail-to-rail input stage has an operating range from the positive to the negative supply voltage. Using a single differential pair for instance, the operating range is bounded by the threshold voltage of the input transistors and the saturation voltage of the current source. Only one supply rail can be reached with this technique.

In this work, two principles of rail-to-rail input stages will be discussed. The focus is laid on the use of a PMOS and an NMOS folded cascode in parallel. Another solution is a signal

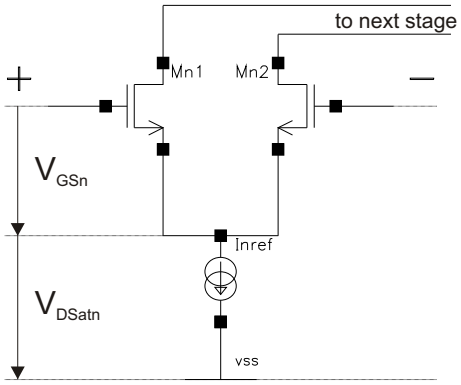


Figure 4.6: NMOS input stage

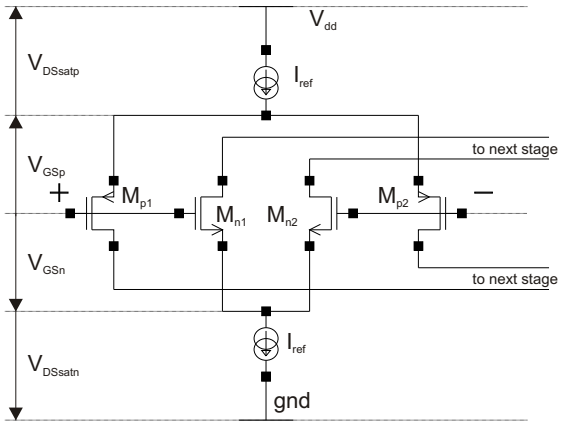


Figure 4.7: Two input stages in parallel

compression circuit which projects the input on a smaller input-range. We will come to this at the end this of section. At least it is also possible to use depletion transistors which will not be further illustrated as those devices are not available in our process.

4.5.1 Parallel Folded Cascodes

Switching two folded cascodes of opposite transistor types parallel(Figure 4.7), you will get an input stage which includes both rails. For operation in the strong inversion region in our process, the NMOS input stage needs a minimum gate-source voltage of $V_{GS_n} = V_{T_n} + 80\,\mathrm{mV} = 400\,\mathrm{mV}$ resulting in a minimal input voltage of $V_{GSDS_n} >= V_{DSSat_n} + V_{gs_n} = 0.6\,\mathrm{V}$. We used $V_{DSSat_n} = 0.2\,\mathrm{V}$ as typical saturation voltage of the NMOS current source. Figure 4.6 shows the NMOS stage together with the voltages. Reasoned by analogy we get for the PMOS stage $V_{GS_p} = -V_{T_p} - 80\,\mathrm{mV} = 540\,\mathrm{mV}$: $V_{GSDS_p} <= -V_{DSSat_p} + V_{gs_p} = -(0.74\,\mathrm{V})$ leading to $V_{dd_min} = V_{GSDS_n} - V_{GSDS_p} >= 1.34\,\mathrm{V}$ as minimal input voltage range for our input stage. Not to be working at the beginning of the strong inversion, but in the middle of it, we should keep a distance of 0.2 V to the threshold voltage and not only 80 mV as done before. With this harder constraint, we even get a minimal voltage of $V'_{dd_{min}} = 1.58\,\mathrm{V}$. As we can use a voltage of 1.8 V there is no problem.

If you look at the progress of the transconductances g_{mn} and g_{mp} as well as their sum g_{mges} (figure 4.8) you will realize, that there is an area in the middle of the voltage range, where both

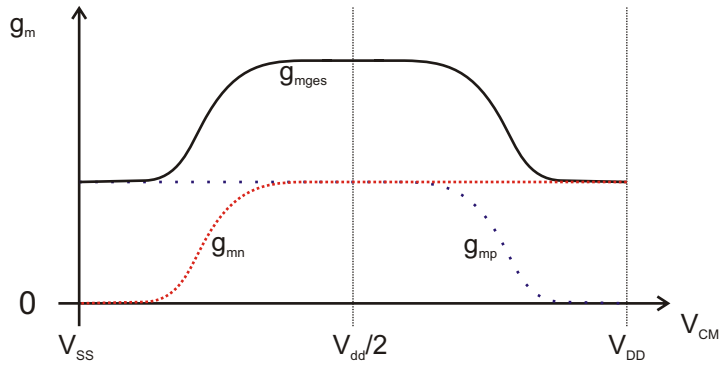


Figure 4.8: Progress of the transconductances over the whole common mode range for to cascode stages in parallel [17]

stages are active. This results in a doublet transconductance in the middle, when both stages are connected in parallel. As this behavior causes distortion, it must be avoided. Unfortunately this is only possible under forfeit of current efficiency. Different Techniques lead to a solution. Common for all discussed in this work is that they were only applicable for frequency in a magnitude of tens MHz's because operation in strong inversion or weak inversion is needed. High frequencies need small capacitors in the input stage, which leads to a smaller channel length. Velocity saturation can occur, deteriorating the g_m-control [15]. In our process, the compensation works for higher frequencies, too.

Single Current Mirrow

The sum of the transconductances, $g_n + g_p = g_{ges}$, has to be constant over the entire operating range. This circuit has been introduced in [14] for instance. Recapitulating the current dependence of the transconductance in the strong inversion region yields $g = \sqrt{2I_{DS}K'W/L}$. As both stages have to be concerted, the factor $\sqrt{2I_{DS}K'W/L}$ must be identical for PMOS and NMOS stages and can be set equal to one in the following. We get:

$$\sqrt{I_{DS_p}} + \sqrt{I_{DS_n}} = const. \tag{4.21}$$

The current in each single stage is given by the bias current from the current sources, as far as allowed by the operation region and the dimensioning of the stages. It is possible to control the bias of one stage by the current of the other stage using a current mirror. With this technique, the transconductance variation can be reduced to a factor of $\sqrt{2}$ whilst the total current is bisected.

For low voltage applications with a voltage range that low, the cross section of the active regions is only in the weak inversion region, it is possible to get a direct adjustment of the transconductance as transconductance and cross current are proportional in this region. This method has its origin in bipolar amplifiers [18], where the dependency of current and transconductance is linear for higher voltages, too. Applications can also be found in low voltage CMOS amplifiers [14]. The concrete circuit can be see in Figure 4.9. For smaller channel lengths, the ratio of the PMOS current mirror has to be changed.

To explain the circuit, we will start with a common-mode voltage $V_{cm} = V_{dd} = 1.8\,\mathrm{V}$. The control voltage V_{bias} shall be 1.2 V. The transconductance of the transistors Mn1 and Mn2 is given by the current source. The voltage at node 1, $V_{(1)}$, behaves in a way that the NMOS stage is operating in the middle of the strong inversion region, meaning $V_{GSn} = Vcm - V(1) =$

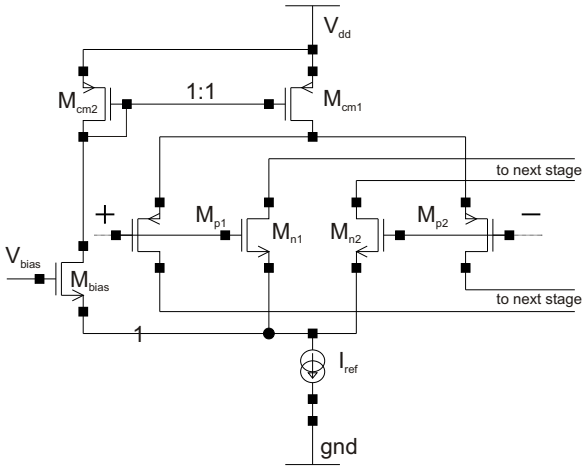


Figure 4.9: g_m-Compensation with single Current Mirror [14]

$V_{Tn} + 0.2\,\mathrm{V}$. This is possible as long as $V_{(1)}$ does not undercut the saturating voltage of the current source. Thus, with a threshold voltage $V_{Tn} = 0.33\,\mathrm{V}$, $V_{(1)}$ will be $1.27\,\mathrm{V}$ at the beginning. The biasing transistor is switched off.

With falling common-mode voltage $V_{(1)}$ decreases, finally resulting in $V_{GSBias} = V_{Bias} \rightarrow V_{Tn} = 0.33\,\mathrm{V}$ and a current in the PMOS branch. At $V_{cm} = V_{GSBias}$ the current is shared between the bias transistor and the NMOS input transistors, hence both stages - NMOS and PMOS - contribute equally to the total transconductance. With proceeding fall of the input voltage more current flows through the bias transistor. Finally the NMOS stage is switched off and the total current is mirrored into the PMOS stage.

Current Switch

The problem with the simple current mirror is that usually, both stages are working in strong inversion in the middle of the operating range. The idea is now to keep one stage off until the other goes down to weak inversion. When switching on, both circuits are working in weak inversion. The circuit is actually the same as shown in Figure 4.9. The problem with this technique is the dependency on the varying threshold voltage which can vary. Each amplifier would need its own individual biasing voltage. The ratio at the PMOS current mirror has to be adjusted to n_n/n_p as the diffusions capacitor is more important at small channel length. This modified version of the single cm can be found in [17].

3-times Current Mirror

Whilst the simple current mirror bisects the current in the middle of the common-mode area, we now aim at multiplying it by four at the boundaries where only one stage is active. Again we avail the drain source voltage of current sources falls when the appropriate stage gets inactive and redirect the current in the particular other stage. The difference is that each stage has its own current source and the current from the other stage is not only mirrored, but multiplied by three, resulting in a quadruplicated current at the boundaries. The associated schematic is shown in Figure 4.10

The quadruplication of the current realizes an adjustment in the strong inversion region, as the sum of the square roots is kept constant. A problem occurs at the intersections as there applies $g_{mges} \propto f(x) := \sqrt{4 - 3x} + \sqrt{x}$ [17] and $f(x)$ has a maximum for $x = \frac{1}{3}$. This adds up to a deviation of 15 %. More discrepancies are caused by eliding the weak inversion region in this

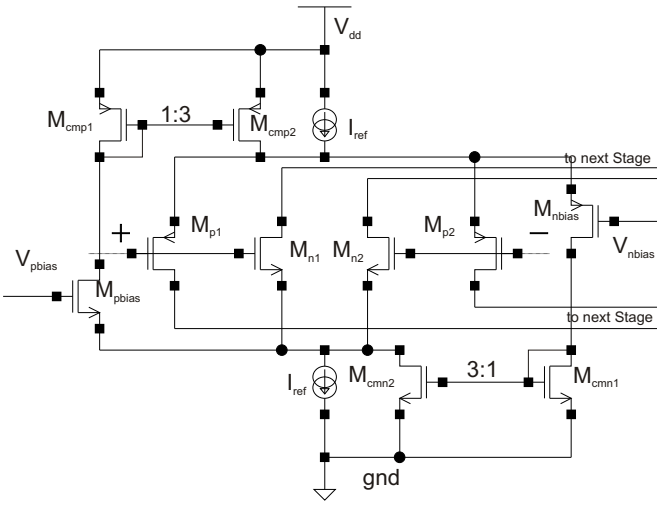


Figure 4.10: g_m-compensation with three times current mirrors [14]

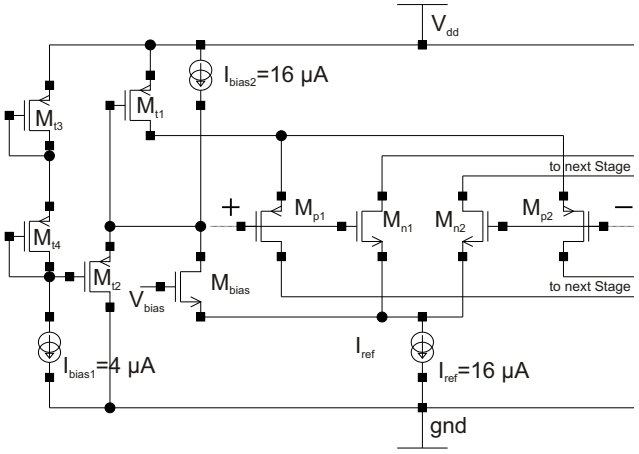


Figure 4.11: g_m-compensation with translinear circuits [14]

method. Ron Hogervorst achieves a deviation of 16 % with this method on an older process[14] which sticks more to the hand models used than today's common sub micron processes.

Control of the Sum of the Current Square Roots with translinear Circuits[14]

Similar to the 3-times current mirror, the sum of the square root is maintained here. The difference is this time the current is not quadruplicated at the tails, but reduced in the middle.

A translinear circuit guarantees the linearity of one value via non linear parameters. The best example is the simple current mirror, where the equation of the transconductances leads to a linear dependency of the currents. A similar construction can be used, to keep the sum of the gate source voltages of two pairs of transistors constant. Figure 4.11 shows the translinear circuit, consisting of transistors Mt1, Mt2, Mt3 and Mt4. It makes use of $V_{gsMt1} + V_{gsMt2} = V_{gsMt3} + V_{gsMt4}$ and therefore $\sqrt{I_{dsMt1}} + \sqrt{I_{dsMt2}} = \sqrt{I_{dsMt3}} + \sqrt{I_{gsMt4}}$ if all transistors are working in the strong inversion region and parameters have been chosen adequately. The sum of the square roots is kept constant which is exactly what we need for the biasing currents in our rail-to-rail input stage.

It is obvious, that the current of the PMOS-stage is the current through Mt1. The currents

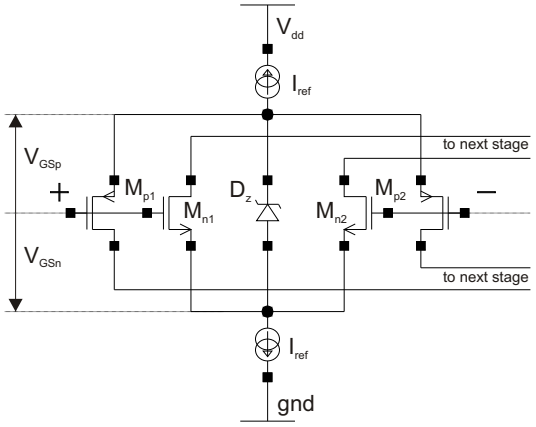


Figure 4.12: g_m-compensation with ideal zener diode [15]

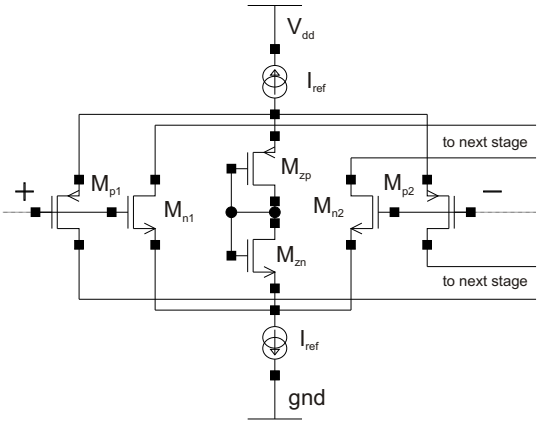


Figure 4.13: g_m-compensation with two diode connected mosts in series implementing a zener diode[15]

produced by the sources Ibias1 and Inref are equal and their root is just twice the root of the current from ibias2. The current of the NMOS stage is now $I_{Mn1+Mn2} = I_{nref} - I_{Mbias} = I_{bias2} - I_{Mbias} = I_{Mt2}$.

Electronic Zener

Another idea to keep the sum of the square roots of the currents constant is to assure the uniformity of the sum of the gate source voltages of the input stages. Which is $-V_{GSp} + V_{GSn}$ with the denominations from Figure 4.12. A zener diode can be switched between the two current sources Ipref and Inref with a zener voltage $V_{Zener} = -V_{GSpM} + V_{GSnM}$ where V_{GSxM} is the voltage the regarding input stage needs to work in strong inversion with a cross current of $\frac{1}{4} * I_{ref}$. The factor $\frac{1}{4}$ is there because the tail current of an input stage working at the respective rail of the operating range has to be four times bigger than in the middle. The circuit with an ideal zener can be seen in Figure 4.12. In the middle of the operating range, $\frac{3}{4}$ of the total current has to flow through the zener diode. This results in 50 % worse current efficiency in this area in comparison to the unbalanced stage. But after all the zener compensation is more compact and power-efficient than the stages above [15]

In a standard CMOS-process there is no zener diode available, so one has to solve the problem of implementation. A simple but not accurate method is the use of diode connected mosts. The

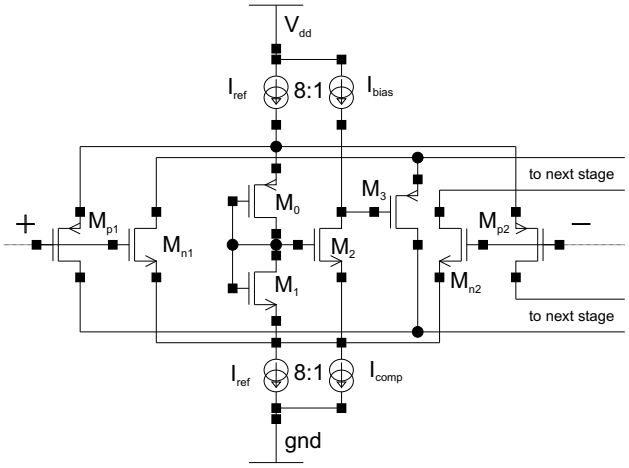


Figure 4.14: g_m-compensation with an electronic zener diode[15]

curve of a diode connected most is not as sharp as that of a zener diode but sufficient for a g_m-compensation with an error of 28 % [15]. For the implementation see Figure 4.13. Two most devices of opposite type are diode connected in series to construct the right zener voltage. The dimension of these devices has to be 6 times larger than the one of the input transistors because the current is 6 times that big. One great advantage of this method is that only two devices are needed and especially not additional biasing voltages at all.

A more precise method than the two diode connected mosts employs a device called "electronic zener" which has sharper break in the characteristic at the zener voltage. The circuit can be seen in Figure 4.14. When working near the rails where only one stage is active, the zener diode is not conducting. The current source Icomp is causing an error in the g_m of 6 % at the upper rail and is needless in this part of the working range. In the middle of the operating range, when both circuits are active, the electronic zener garantees the constancy of the sum of the gate source voltages. M1 and M0 have the same W/L as the respective input transistors to construct the right zener voltage. One eighth of the current flows through this diode. When the diode is conducting, the gate source voltage of M2, V_{gs2} is the same as that of M1 and as M2 is equally dimensioned to M1 it is capable to drain the same current which is Ibias or one eight of Ipref. The current through M2 is drained by the compensation currentsource icomp now. The voltage at the drain of M2 decreases, activating the source follower M3. M3 must be dimensioned properly to drive five eighth of Ipref now to make the g_m-compensation complete. With additional circuits to compensate icomp, Ron Hogervorst reaches a g_m which only varies 8 % over the whole common mode range with an electronic zener on a 1 μm BiCMOS process [15]. The electronic zener uses at least two more transistors in the input stage which depend on each other in matching. There is a strong dependence on the drain source voltage of M2 and the gate source voltage of M3 resulting in matching problems when smaller processes are used.

4.5.2 Signal Compression Circuit

Another method for reaching rail to rail input has been proposed by Yukizaki et al. [21]. The main amplifier is a folded cascode amplifier which is able to attain one supply rail. To get to the other supply rail, a special signal compression circuit with a nearly linear characteristic is used. If the whole amplifier is only used as voltage buffer additional noise created by inequalities of the characteristic of the compression circuits is suppressed.

In principle the compression circuit consists of 3 parts. One shifts the input signal to a higher voltage and compresses it. This is done using three source followers, one PMOS, one NMOS

and one PMOS in series. Voltages above the supply voltage minus the threshold voltage of the first source follower, $V_p = V_{THpmos}$ are projected to the supply rail by it, flattening the curve for higher voltages. The second part uses two source followers and an inverting circuit to project all signals below V_p to a constant value whilst signals above are compressed and inverted. The third circuit sums the outputs of the first two and by this creates the compressed version of the input signal.

This method has been simulated on spice with a 180 nm CMOS process by Yukizaki reaching supply voltages down to 0.7 V [21]. Problems could occur with matching in the compression circuit because V_p is a critical value which is connected to the variating threshold voltage of the PMOST. Further process variations could also effect gradient of the output of the first and the second circuit destroying the linear characteristic of the total compression circuit.

4.6 Output Stages

The output stage of an amplifier usually has to drive huge loads like a big capacitor or a small resistor, which means, that a high current must be sourced or drained. Furthermore the quiescent current of the stage should be minimal whilst the characteristic should be as linear as possible to avoid distortion. Sufficient gain should be provided. Classical, stages are divided into three classes: Class A, class AB and Class B amplifiers.

A class A amplifier has very low distortion whilst the maximum output current is smaller than the quiescent current, which is high over the whole operating range. The best example is a simple NMOS source follower with a resistive load. The characteristic is linear as long as only small currents have to be driven in comparison to the crosscurrent current.

Class B amplifiers have no biasing current at all. The Problem is that this kind of amplifier has to be switched on at certain voltage resulting in a characteristic that has a break at a certain point where for example negative swings of the input are ignored. This problem is called crossover distortion.

The trade off between theses classes is the class AB-amplifier, which has a low biasing current in comparison to the maximum output current and some distortion left. We will focus on this class in the following.

4.6.1 Basic Needs of Class AB Stages

Requirements and some solutions will be discussed here [17].

Rail-To-Rail Output

Both rails should be reached by the stage. This inhibits the use of source followers and push-pull stages. A push-pull CMOS stage consists of a NMOS and PMOS transistor, source connected to each other. The minimal distance to a rail is the $V_{DS} = V_{GS}$ of the corresponding transistor hence.

The simplest rail-to-rail output stage is the standard CMOS inverter. Of course, the transconductance of the regarding output transistor goes down at the rails, as saturation is left when V_{DS} goes below $V_{GS} - V_T$ nevertheless, the rail can almost be reached if no resistive load is applied. For a resistive load, we analyze wlog the case of a resistance R_{load} to ground when the power rail V_{dd} should be reached. The cross current is neglected. If $\frac{V_{dd}}{R_{load}}$ is bigger than I_{DSmax} of the PMOS output transistor, only $R_{load}I_{DSmax}$ is reached. Else the distance to the rail is given by the voltage divider consisting of $R_{on} = \frac{1}{KP(W/L)(V_{GS}-V_T)}$ of the PMOS and R_{out}. Without resistive load or for small ones, the biasing current is the limiting factor, as it causes a voltage drop at R_{on}

Controllable Quiescent Current

In contrast to the input stage, the cross current cannot be directly controlled by current mirroring here. Notwithstanding the quiescent current has to be controlled somehow and should not change during a common mode sweep of an upstream connected input stage. This means the operating point of the output stage should be fixed for constant quiescent current. To adjust the quiescent current the operating point is changed. Additionally, an offset voltage between the gates of NMOS and PMOS can be applied. Enlargement of the offset voltage lowers $\mid V_{GS} - V_T \mid$ for NMOS and PMOS and accordingly the quiescent current.

Large Drive Capability

The drive capability is defined as $\frac{I_{max}}{I_{quiescent}}$. The I_{DS} characteristic of the transistors has to be more than linear to be capable of sourcing and draining currents much higher than the quiescent current - this is called an expanding characteristic [17]. It is especially needed, as the input swing of the output stage is limited. For instance 20 mA are needed to drive a 50 Ω transmission line at 1 V. The advantage of the expanding characteristic is its problem, as the nonlinearity is causing distortion. Feedback is necessary for balancing. Achieving the expanding characteristic can be done by biasing the output transistors in strong inversion. This is done by the adjustment of the operating point and the applied offset voltage between the gates, introduced above. Biasing is impeded by the choice of smaller channel lengths as this reduces the strong inversion area. A flat common mode output characteristic of upstream stages is important.

Small Size

A small $W * L$ of the output transistors is indispensable not only to minimize area consumption but also for stability reasons. A higher C_{GS} at the output transistor results in a higher load capacitance of the upstream stage.

4.6.2 Quiescent Current Control

In this work, the quiescent current will be controlled by a translinear circuit similar to the one used in 4.5.1. The concrete circuit can be seen in Figure 4.15.

The difference here is that the dimensioning of the transistors is not equal as M_{out} is the output transistor. Again, the sum of the gate source voltages is kept constant: $V_{gsM1} + V_{gsM2} = V_{gsMout} + V_{gsMt}$. It is directly controlled by the biasing current. M_{T1} will be part of our input stage later and the input current I_{in} will be the crosscurrent of the cascode branch. According to the sum of the $V_{GS}s$, the sum of $\sqrt{I_{DS}L/W}$ is kept constant, so for constant I_{in}, I_{ds} of M_{out} is constant.

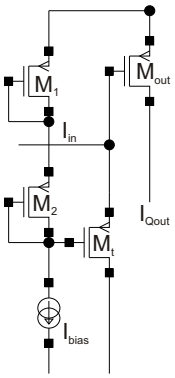


Figure 4.15: Translinear quiescent current control

Chapter 5

The Amplifier Design

In the previous chapters, the theoretical background and design techniques of circuits and operational amplifiers have been discussed. In this chapter the practical design of the amplifier will be contemplated. This means that we have to switch between the simple hand calculation model and the complex BSMIM 3.3 model of the used 180 nm process. Above considered circuits will give different results because of variations in the threshold voltage for example. After this important step we will come to practical examples of input stages resulting in the input stage finally used for the discussed amplifier. The output stage will be designed next. Compensation considerations will be done followed by a conclusion over the whole amplifier with three versions. Eventually the layout will be discussed, starting with deliberations about transistor layouts.

5.1 Connecting Hand Model with Simulations

Working in strong inversion we generally have

$$I_{DS} = K'_n \frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \tag{5.1}$$

for the drain source current of the NMOS transistor[1] as simple model. λ is set to zero for most hand calculations, so what we broadly need is the threshold voltage and

$$K'_n = \frac{KP}{2n} = \frac{\mu C_{ox}}{2n}. \tag{5.2}$$

The mobility of electrons in silicon is typically $\mu_n \approx 600\,\mathrm{cm}^2/\mathrm{Vs}$. From our process parameters, we get a gate oxide capacitance of $(820 \pm 50)10^{-5}\,\mathrm{pF/um}^2$. We get $KP_n \approx (490 \pm 30)\,\mu\mathrm{A/V}^2$. The parameter n depends on the biasing voltage and cannot be directly obtained[17]. Employing a K'_n in the dimension of 270 $\mu\mathrm{A/V}^2$ and a ratio $K'_n/K'_p \approx 2.4$ leads to good simulation results for channel lengths around 300 nm.

The measured threshold voltage for some W/L configurations is on hand and has a massive dependence on the channel length. Given values for the NMOS transistor in the typical region for a drain source voltage of 1.8 V vary from 0.28 V for 10 μm length and 240 nm width and 0.42 V for 180 nm channel length and 10 μm channel width which is a variation of more than one third. Incorporating process variations there could even be a deviation of 100 % and including a temperature range from $-55°$ C up to 125° C a total variation of more than 200 % is reached.

For the PMOS transistor, the values of the threshold voltage in the typical region are less effused. We get a range between -0.46 V for 10 μm length and width and -0.4 V for 240 nm

[1] see 3.3

width and 180 nm length, which is a variation of only 15 %. An important notice should be laid on the difference between the gate source voltages of NMOS and PMOS.

To have better control of the gate source voltages and to ease matching, it is a good idea to concentrate on one channel length.

5.1.1 Parameter Extraction

Since the model files were not available at the beginning of this work, several ways were developed to extract parameters from the simulation to produce circuits that behave the way, the calculations predict. The first idea was to use the value of the drain source current I_{DS} for $V_{GS} = V_{DS} = 1.8\,\mathrm{V}$ given in the electrical design rules. The transistor is working in velocity saturation in this region. Neglecting channel length modulation, the equation is:

$$IDS = \frac{K'_n \frac{W}{L}(V_{GS} - V_T)^2 V_{DS}}{1 + \theta(V_{GS} - V_T)} [17,\ 23] \tag{5.3}$$

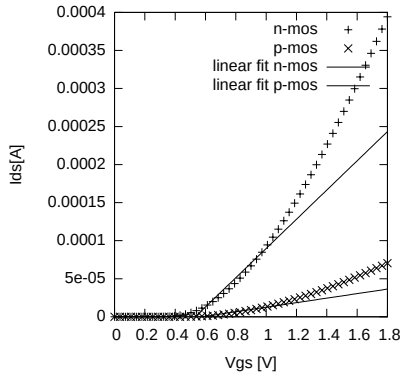
With $K'_n = KP/2n$, $\theta \approx$
$mu/2nLv_{sat}$, a maximum speed for the electrons of $v_{sat} \approx 10^7 \frac{cm}{s}$ and some transformations, we get a K'_n of approximately $166\,\mu\mathrm{A/V}^2$ and therefore $n \approx 1,47$. The problem with our calculation is that it does not include effects like channel length modulation which has a big influence at channel lengths of 180 nm at high drain source voltages.

Simulations did not fit accurately enough so a different way to obtain K'_n had to be found. The next idea was to leave the velocity saturation method using transistors with bigger channel lengths and gate source voltages that will surely lie in the strong inversion region. For this instance, simulations with PMOS and NMOS transistors where done, varying the channel length. This was done because it was observed that the dependency of the drain source current on L is not just $\frac{1}{L}$ For these measurements, a channel length independent threshold voltage hat been assumed, which explains the strong variation in the K's obtained. From the electrical design rules, we got the linear extrapolated threshold voltage for a NMOS transistor with a length of 180 nm and a channel width of $10\,\mu$m, which is around $V_{T0} \approx (0.51 \pm 0.07)\,\mathrm{V}$. As reference gate source voltage we choose $V_{GS} = 0.72\,\mathrm{V}$, certainly located in the strong inversion area. As the threshold voltage is not independent of the channel length, only the value for $L = 180\,\mathrm{nm}$ has a little relevance. The measured drain source current for this L is $I_{DS} = 673\,\mu\mathrm{A}$ and we calculate $K'_n \approx 275\,\mu\mathrm{A/V}^2$. With KP we get $n \approx 0.89$. The same can be done for the PMOS-transistor. Actually, even the relevance of this result is only little because channel length modulation results in a big λ for $L = 180\,\mathrm{nm}$, the strong inversion operation can not be guaranteed and the given threshold voltage is only a linear extrapolated one. n should not be smaller than one!

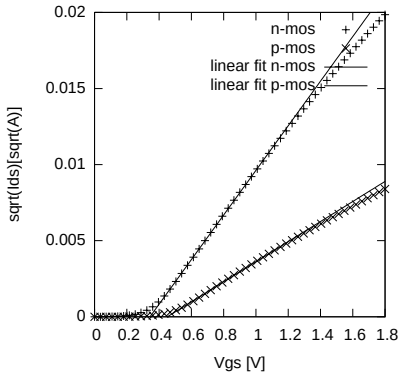
To illustrate the problem with the extrapolated V_{GS}, Figure 5.1 shows a linear fit to I_{DS} (a) and a linear fit to $\sqrt{I_{DS}}(b)$. While the threshold voltages obtained in a) fit to the ones from the electrical design rule file, the values for b) fit to the ones from the model files. The consequence of this calculation is that the extrapolated voltage cannot be used for calculations in strong inversion using our model.

The next measurement was the determination of the threshold voltage in dependency to the channel length. For this purpose, two measuring points in the strong inversion region for two different gate source voltages resulting in different drain source currents were used. As model

$$I_{DS} = \beta(V_{GS} - V_T)^2 \tag{5.4}$$



(a) Linear fit for I_{DS}. The threshold voltages are around 0.55 V respectively −0.55 V



(b) Linear fit to the square root of I_{DS} approximately obtaining much more realistic values of 0.35 V for the NMOS and −0.45 V for the PMOS

Figure 5.1: Linear fit to obtain the threshold voltage for NMOS and PMOS with $W = 2\,\mu$m and $L = 1\,\mu$m.

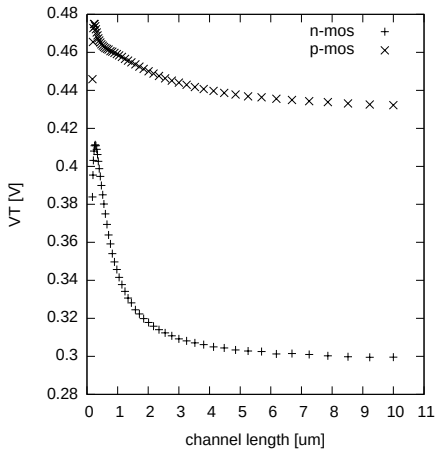


Figure 5.2: Calculated V_T

has been employed for the strong inversion with β and V_T as parameters to be found. The channel length has been varied for a width of 10 μm and a V_{DS} of 1.8 V. As can be seen in Figure 5.2 the model works quite well for NMOS at channel lengths above 300 nm. Extrapolating a K' this way, we get a variation of only 50 μA/V^2 The measured threshold voltages are in the range of the threshold voltages in the model files for higher channel length. For smaller channel length, less than 300 nm, V_T of the NMOS goes downs although it should still rise. The behaviour for the PMOS does not fit to the behaviour in the models files, as the threshold voltage should be smaller for smaller channel length. For the calculation of K' the threshold voltage for small channel length of the PMOS is corrected.

The extraction of K', which is equal to $\frac{\beta L}{W}$ can be seen in Figure 5.3. For small L, there is still a strong dependence on the channel length, which should not be the case. Channel length modulation arises the current and hence the K' for smaller channel length. In a simple model, we can include modulation by adding the factor $(1 - \lambda V_{DS})$ to the model where $\lambda = 1/V_E * L$ with the process parameter V_E. To get our final parameters, we fit the function

$$K' = \frac{K''}{1 - \frac{V_{DS}}{V_E L}} \tag{5.5}$$

to the K' from the last stage to the least sum of squares on variation of V_E. The NMOS threshold voltage has been corrected. For the NMOS transistor we get $V_{En} = (21 \pm 1)\,\mathrm{V}/\mu\mathrm{m}$ which is 5 times bigger than the sample V_E of $4\,\mathrm{V}/\mu\mathrm{m}$ given in [17]. The V_E or the PMOS is $V_{Ep} = (5.3 \pm 0.1)\,\mathrm{V}/\mu\mathrm{m}$. The fit can be seen in Figure 5.3, too.

There are other ways to determine V_E. Regarding the V_DS derivative of a mos with $L = 500\,\mathrm{nm}$, we get $V_{En} \approx 19\,\mathrm{V}/\mu\mathrm{m}$ which has the same order of magnitude and $V_{Ep} \approx 23\,\mathrm{V}/\mu\mathrm{m}$ which is much bigger than above. We will use these values, as the way of determination is much clearer here.

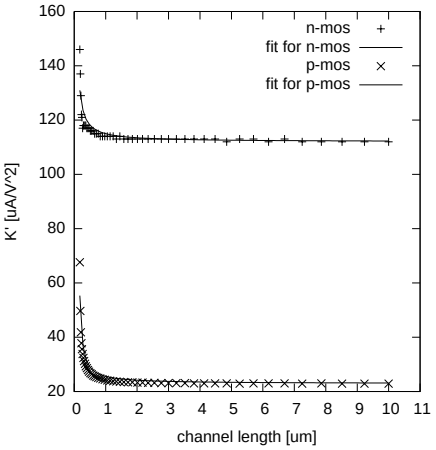


Figure 5.3: Correctec K' with fits

Final Work flow

It does not make sense, to exactly model the threshold voltage in our hand calculations, as they would not be handy anymore afterward and process variations would make any circuit useless which has critical dependency on the threshold voltage. What is important are the transconductances, and the capacities, which can be calculated easily. As K'_n we have chosen $275\,\mu\mathrm{A}/\mathrm{V}^2$ in the final circuit to get a starting point for the simulations. As factor between the transconductances of NMOS and PMOS-transistor, 2.4 is used, as this is nearly the factor between the charge carrier velocities. This way, rough start values for the simulations were obtained. As simulations have been used for optimization, the real K' is less important.

To get rid of the threshold voltages, the final design is done from a current perspective as this parameter can be controlled by current mirrors. Neither the actual gate source voltage nor the threshold voltage are used, but their difference $V_{GS} - V_T$. It will be assumed to have a certain value to control the working range of the transistors. The W/L is adjusted in a way, that right current flows at this point and the transconductance fits.

The length of the transistors is chosen to be preferably above 300 nm to keep the strong inversion area big enough. The same L is chosen, as this approach minimizes variations between the ratio of NMOS and PMOS.

In the final design, the finger width of each transistor type is kept constant and the total width is only modulated by the finger number. Of course these finger width and number have to be adapted during layout to avail the needed transistor structures.

5.2 Input Stage

Lots of input stages have been calculated, designed and simulated during the design phase, starting with a simple differential pair with ideal current sources and loads, switching the ideal parts to transistors. Next steps have been a symmetrical input stage and a folded cascode. As parameter extraction went on different, values for the transconductance parameter K' have been used during the design. This is no problem, as described above.

5.2.1 Stepwise Assembly

Simple CMOS OTA

The final circuit with real current sources and NMOS load can be seen in Figure 4.3.

We start using an ideal current source and ohmic load resistors. As current, we chose $200\,\mu\mathrm{A}$ which is flowing into both input transistors. The common mode output should be 0.9 V for

maximum output swing, so load resistors of 9 kΩ are chosen. With $V_{GS} - V_T = 0.2\,\mathrm{V}$ and a used K'_p of $26\,\mu\mathrm{A/V^2}$, we get a needed W/L of 100 and chose $W = 50\,\mu\mathrm{m}$ and $L = 500\,\mathrm{nm}$. The calculated g_m of the transistors is 1 mA/V, so we get a voltage amplification of $A_V = g_m * r_{out} = 9$ Simulations nearly exactly match with our calculations. The simulated small signal voltage amplification is 8.56. The source voltage of the PMOS is pulled a little bit down by the current source. This is caused by the different threshold voltage. K' seems to fit in the order of magnitude which is confirmed by Figure 5.3. The bandwidth is $10^7 Hz$

As current will be sourced by a current mirror later, we lower the source voltage of the PMOS by 200 mV choosing smaller load resistors of 7 kΩ. The resulting amplification is 7 as predicted.

The next step is the replacement of the resistive load by transistors, respectively a NMOS current mirror. Of course, this way feedback on one branch is applied. The diode connected most of the mirror is adjusted to drain a current of $100\,\mu\mathrm{A}$ at 0.7 mV $V_{DS} = V_{GS}$. As V_T is around 0.4 V and the transconductance parameter $K'n$ $120\,\mu\mathrm{A/V^2}$ for a channel length of 500 nm. The needed $\frac{W}{L}$ is 10 and we get a perfectly fitting common mode output voltage with $W = 5\,\mu\mathrm{m}$. The small signal gain goes up to more than 48.

Although, the current of the NMOS M1 is equal to the current of the resistance above, the small signal resistance of M1, r_{DS1} is much higher than the former load. With $V_{En} = 19\,\mathrm{V}/\mu\mathrm{m}$ as determined above, we get a resistance of $r_{DS1} = \frac{V_E L}{IDS} = 95\ k\Omega$ which is switched in parallel to the drain source resistance of the input transistor MI1, $r_{DS1} \approx 115\ k\Omega$ As final small signal load resistance, we get $r_load = 52\ k\Omega$ so the small signal gain in the simulation is no problem.

Now, we replace the current source by a real current mirror. The most important is to keep the L big enough, as λ gets higher for smaller channel length which would result in a strong dependence between the V_{DS} of the mirror and the current. Choosing $L = 500\,\mathrm{nm}$ again is convenient here. A multiplication factor can be used in the current mirror not to add to much cross current. We chose a factor of 5. The finger width has to be chosen in a way, that one finger alone is able to drive $40\,\mu\mathrm{A}$ current at 0.2 V $V_{GS} - V_T$. A finger width of $10\,\mu\mathrm{m}$ is chosen. The simulation results fit to the predictions. The gain nearly stays the same and $200\,\mu\mathrm{A}$ are nearly reached by the constructed current source. There is only a small dependency between the input signal and the current.

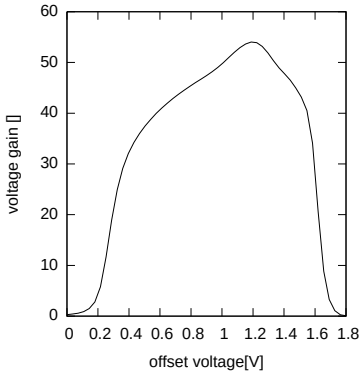


Figure 5.4: Gain characteristic of the designed simple OTA at 10 kHz

Figure 5.4 is showing the development of the gain over the whole input range. For small voltages, the amplifier is hardly working because the input transistors are biased in the linear region at relatively high $V_{GS} - V_T$ and accordingly low V_{DS}. Actually the V_{DS} is only 80 mV at 0 V input. With rising input voltage, saturation is finally reached for the input transistors at nearly 0.2 V.

Symmetrical Amplifier

The schematic of the symmetrical amplifier can be seen in Figure 4.4 of the last chapter. We try to achieve a GBW of 200 MHz at a load of 2 pF (To compare: The GBW of the last stage is only 70 MHz at this load). Therefore, using Equation 4.15 we get $g_{m1} = 2.5\,\mathrm{mS/B}$. The

second step is to calculate the maximum B using Equation 4.16. After insertion of the gate oxide capacitance and some modifications, we get

$$46.4 * 10^{-6} \frac{\mu A}{V nm^2} \leq K'_n \frac{(V_{GS4} - V_T)}{L_4^2} \frac{1}{(3+B)}. \tag{5.6}$$

A lot of interesting facts can be seen in this formula. We have to use a small L and or a high $(V_{GS} - V_T)$. The transistor, the calculations are made for is part of a current mirror. A higher V_{GS} means a higher drop out voltage is needed on the other side of the mirror to be in saturation. A smaller L results in a higher λ and hence a bigger derivative in strong inversion. We chose $(V_{GS} - V_T) = 0.2\,\mathrm{V}$ and $L = 300\,\mathrm{nm}$, imply K'_n $120\,\mu\mathrm{A/V^2}$ and get $B \leq 2.74$ We have to chose $B = 2$ and so 1.25 mS are left for the transconductance of M1 which is only a little bit more than in the last circuit. The W/L of the input stage pair has to be 125. We continue using 500 nm as channel length of the input stage. As W/L for M4, we chose 12.5 in analogy to above. The parameters of the input current mirror can be reused, too, but we have to enhance the finger number as now $250\,\mu\mathrm{A}$ are needed. The current mirror M7/M8 get the same W/L as M6 and M7 but a much higher L to warrant assumptions made above.

Simulations result in a GBW of 181.6 MHz with a phase margin of 76.55°, which is a little bit more than required. The accordance between simulations and calculations is very good. With the measured cross current of $750\,\mu\mathrm{A}$, this fits to a figure of merit of 514 MHzpF/mA. Using smaller channel length for the load and a higher B, even 800 MHzpF/mA have been reached.

Folded Cascode

The circuit can be seen in Figure 4.5. The non dominant pole of the folded cascode is less critical and therefore, we go for a much higher gain bandwidth now. To remain the magnitude of the transconductances, we chose 1 GHz GBW at a load capacity of 200 fF. Equation 4.20 is used to calculate the frequency of the second pole. At a channel length of 300 nm and $V_{GS} - V_T = 0.2\,\mathrm{V}$ for the cascode transistor, we get $f_{nd} \approx 310\,\mathrm{MHz} * \mu\mathrm{m/L^2} \approx 3.5\,\mathrm{GHz}$ which is sufficient as it is bigger than 3 times the planned GBW.

We calculate the g_m of the input stage using equation 4.19 and get 1.25 mS so we can reuse the input stage and the biasing circuits above and use a biasing current of $250\,\mu\mathrm{A}$. Transistors M11 and M10 will be biased via current mirroring and dimensioned according to M9 with a factor of 1/2.4 as they are NMOS devices, but driving the same current as M9.

The cascode transistors should have an I_{DS} of $125\,\mu\mathrm{A}$ at $V_{GS} - V_T = 0.2\,\mathrm{V}$ so we get $W/L = 26$ and with $L = 300\,\mathrm{nm}$, $W = 7.8\,\mathrm{nm}$. The proportion of the current mirror transistors M5, M6, M7 and M8 is done using the same constraints resulting in $W/L = 120$ or $W = 36\,\mu\mathrm{m}$ and $L = 300\,\mathrm{nm}$. As realistic control voltages, we chose 0.9 V for the NMOS cascode and 1.1 V for the PMOS current mirror cascode.

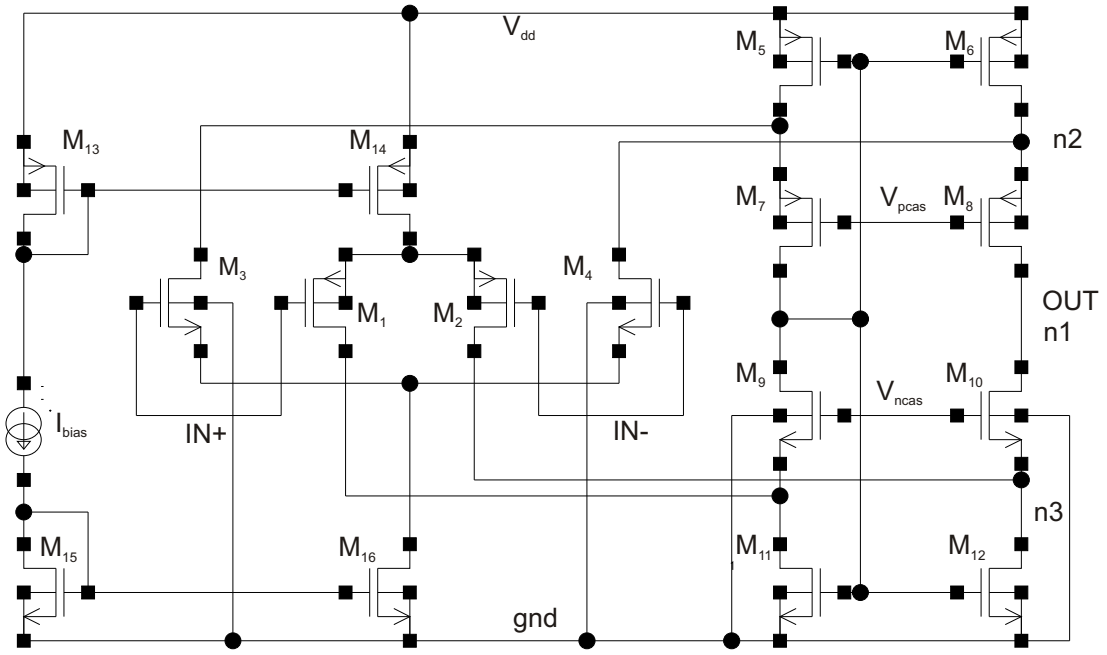


Figure 5.6: first parallel folded cascode OTA

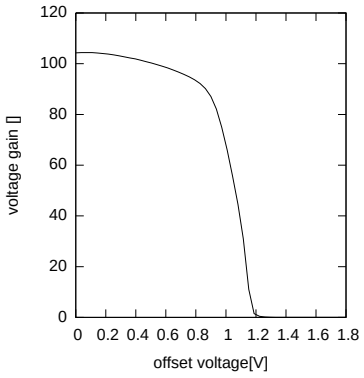


Figure 5.5: Gain characteristic of the designed folded cascode OTA at 10 kHz

Simulations result in a GBW of 685.3 MHz at a phase margin of 54.27° which is both worse than the predictions. The reason here lays in the problem of the switch between the simple calculation model and the complex BSIM model. One GBW reducing effect is the source bulk voltage of the cascode transistors, which can be eliminated by the use of triple well transistors. The resulting gain characteristic can be seen in Figure 5.5. Clearly the ground rail is easily reached. At 1.1 V the gain goes down as the bias current mirror leaves saturation and the input pair enters week inversion, finally cutting off, when the threshold voltage is reached. The designed OTA is a good start point for further simulations and parameter optimizations.

Parallel folded Cascode

The next step is the design of the parallel folded cascode. The circuit can be seen in Figure 5.6. The problem in connecting the cascodes of an NMOS and a PMOS stage lies in the biasing which has been done by M10 and M11 and the current mirroring, which has been done by M5 - M8 in the last circuit. Both need a constant potential on their sources. Two biasing circuits would be opponent; one would cause an asymmetry in the circuit. The found solution here is to leave the biasing and use two current mirrors which are connected to each other.

The circuit looks very symmetrical, but actually, there is a problem caused by the current mirroring. The circuits M9 - M11 and M5 - M8 form a current differential amplifier[17], which leads to different impedances seen by the input transistors. This has only an effect on higher order poles. [17].

Another problem of this circuit is the absent of biasing. The crosscurrent of the cascodes is not depended on the biasing current so far. A solution has to be found for this problem later.

The way, the current mirroring is done has an advantage in DC operation. The common mode output is kept nearly constant as the current mirror is applying some feedback here. When the voltage at node 1 goes up, V_{GS} of M11 and M12 is enlarged. The potential at node 1 is pulled down again this way. The common mode output is self stabilized.

That calculation of the impedances nearly stays the same. The small signal low pass frequency at node 8 will surely be smaller than the one at node 3, as the K' of the PMOS is smaller. As in the middle of the operating range, both input stages are operating, we have to use the sum of the transconductances there.

Assuming a constant ratio between K'_n and K'_p, it is sufficient to calculate the proportions of one stage and than use a constant scaling factor. The $K'_n = 270\,\mu\mathrm{A/V}^2$ obtained with the linear extrapolated threshold voltage given by the producer has been used in this calculation assuming a ratio of 2.4 for the PMOS stage. The target again is 1 GHz GBW which is wrong here, as the cut off frequency at node 8 is $f =\approx 67.2\,\mathrm{MHz} * \mu\mathrm{m/L}^2$, calculated with equation 4.20 for the PMOS cascode.

Again, we have to achieve a single stage transconductance of 1.25 mS, as we want the GBW to be 1 GHz for each stage. With the high K'_n above, we get a W/L of only 8.6 for the NMOS input stage. The biasing current is chosen to be $250\,\mu$A again. With $B = 25$, this is $10\,\mu$A for each finger of the current mirrors. This results in W/L=1 for the NMOS and we chose $W = L = 300\,\mathrm{nm}$ for each finger. As each cascode transistor has an I_{DS} of $125\,\mu$A we get W/L=10.4.. We chose $W = 5.2\,\mu$m and $L = 500\,$nm. For an I_{DS} of $250\,\mu$A, M11 and M12 need $W/L = 20.8$, which is realized by choosing $L = 300\,$nm and $W = 6\,\mu$A.

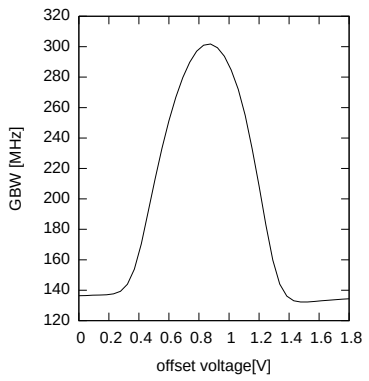


Figure 5.7: Gain-Bandwidth-Product of first double cascode OTA

With these parameters, simulation results in a GBW of 301 MHz and a phase margin of 68.37° at an offset voltage of 900 mV. The characteristic of the GBW can be seen in Figure 5.7. If we compare to the g_m-progress in Figure 4.8 this is a really exemplary result as the GBW is directly connected to the transconductance. The difference of the threshold voltages between NMOS and PMOS can be seen very nicely. The NMOS stage starts operating at 0.3 V whilst the PMOS stages stops operating at around 0.4 V.

Of course, there is a problem with the calculated transconductances, so this is only a qualitative but adequate result. The current in each cascode branch is much smaller than calculated. Only $80\,\mu$A are reached. The effect of the transconductances is much higher than in the previous circuits, as the current is given directly by the transconductances here and not via current mirroring, which is a clear disadvantage of the circuit. Currents of input and cascode stage can be aligned by the reduction of the bias current to $125\,\mu$A.

Notwithstanding, the important result is that we can attain both rails with an adequate GBW. Differences in the ratio between NMOS and PMOS in the input stage are balanced, as the transconductance is mainly given by the current, if the W/L is not too small. Current control of the cascode branch will be added, when we come to the design of the output stage.

To compensate the difference between NMOS and PMOS, simulations have been done to obtain the ratio between K' of NMOS and PMOS and after purge of the KP ratio, the ratio of the bias depending n. This has been done during the K' extraction with constant V_T. The determined ratios are included in the width of the PMOS transistors. After these adjustments, the GBW goes up to around 300 MHz for each stage. This restores the reference to the folded cascode amplifier designed above as the K'_n used here is twice the K'_n used above. This amplifier will be used for analysis of g_m-compensation.

5.2.2 g_m-Compensation

As basic circuit, the parallel folded cascode designed above is used. The functioning of the g_m-compensation has been described in the last chapter. We will analyze the results of the current switch, the 3-times current mirror and the zener-diode implemented by diode connected mosts.

Current Switch

We remember the circuit of Figure 4.9. The determined n-ratio is 1.05 but we get the best simulation results using a ratio of 1.4. The optimum compensation is reached for a bias voltage of 1.3 V The resulting progress of the GBW can be seen in Figure 5.8. At low voltages only the PMOS stage is working. The current of the NMOS mirror flows through the biasing transistor. The biasing voltage adjusts the point, when the PMOS stage is switched off and the NMOS stage is switched on. This transition should be at a point, where the PMOS stage is going down to strong inversion, which is encouraged by the reduction of its biasing current. Finally, the NMOS stage is taking over.

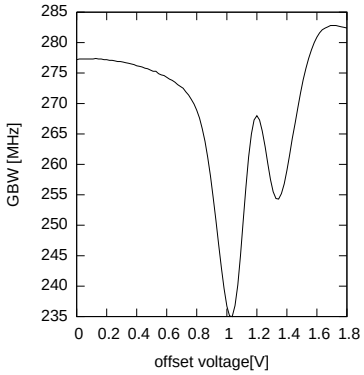


Figure 5.8: Gain-Bandwidth-Product with current switch gm compensation

There is still a variation of 50 MHz or one fifth. Optimization has been used to get the best biasing voltage. Even a small variation of 10 mV leads to much worse results. Actually, the maximum in the middle should not be there if the gm compensation was optimal. There is still some overlap between the strong inversion areas of both stages. This overlap can be reduced by the choice of a higher biasing voltage, but this leads to a smaller minimum indeed.

An optimization process like done here would not be possible for the final physical amplifier and variations of $10mV$ are easily reached by mis-match. This way of compensation is not practical for our design.

3-times Current Mirror

The 3-times current mirror can be seen in Figure 4.10. We now have to determine two bias voltages - one for the PMOS stage and one for the NMOS stage. The bias voltages have to be chosen close to the transistor type accordant $V_T + V_{DSsat}$ so the biasing transistors will be conducting, when one stage is off. V_{DSsat} is the saturation voltage of the input current sources.

It is very important not to choose the biasing voltages not to high as it should never occur, that both biasing transistors are conducting at the same time. Currents of both stages would boost each other until a maximum is reached.

A problem in the design is, that we cannot just leave the rest of the circuit as it has been to have a reference. The biasing current of the whole amplifier has to be quartered to get the same biasing current, but even afterward the cross current of each input stage is higher than in the designs before. This is why we get a generally higher GBW than above, which is no problem for the analysis of the compensation. The exemplary results can be seen in Figure 5.9. The deviation of the g_m is smaller than 15 % which is an even better result as expected by theory or achieved in by Hogervorst in [14].

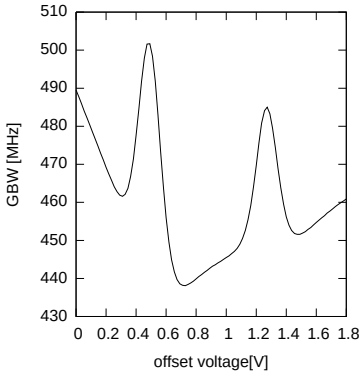


Figure 5.9: Gain-Bandwidth-Product with 3 times current mirror

Analysis of the circuit leads to the idea of using V_{DS} of the corresponding input current mirror as biasing voltage as it is just $V_T + V_{DSsat}$ as needed. The advantage would be that no additional biasing voltages were needed and the compensation was self-adjusting. The deviation determined by simulation is 16 %, which is still very good. Basically, two problems occur here. One is the effect that a situation where both bias transistors are conducting is easily reached by the use of high biasing currents. This results in very high cross current of the total stage at this bias. Of course, this problem it qualified as the amplifier is only designed for a certain bias range and the high current does not destroy the circuit. There could be some difficulties for the power supply line. The second problem is matching of the threshold voltages as the biasing transitors nbias and pbias would have to be matched with the regarding current mirror. The source potential of nbias is not ground, so we have two choices: We can apply a bulk source voltage by setting the bulk to ground or put the transistor into an own well and set the bulk on the source potential. The first solution would arise the threshold voltage dramatically but still results in a deviation of 16 %. The second would constrict matching further. Matching is generally hard, as neither the drain potential, nor the source potential are common. Further analysis would have to be done with this circuit.

Zener Diode

The zener diode is implemented as described above in Figure 4.13. As soon, as the difference of the source potential exceeds $V_{Tn} + V_{Tp}$ the diode connected mosts start conducting. As three fourth of the biasing current has to pass the diode, they will have to be dimensioned relatively big to specify 6 times larger than the input stage at common channel length.

The resulting characteristic of the gain-bandwidth can be seen in Figure 5.10. The deviation is 32 % which is a little bit worse than the 28 % from theory. No additional biasing has to be applied and the resulting stage is very compact as only the diode mosts are needed. Their size is big, but still small in comparison to the size of the transistors needed for the 3 times current mirror in sum. It is still possible to use smaller channel length here to reduce the total size. With a channel length of 180 nm, the deviation nearly stays the same. Additionally, this shows, that there is no strong dependency on the threshold voltage, as it is different for this channel length. Some variations of the width of the mosts have been done, which results in only small changes. Founded on these observations we chose to employ this kind of compensation. The compensation is worse, but stable without any critical additional biasing

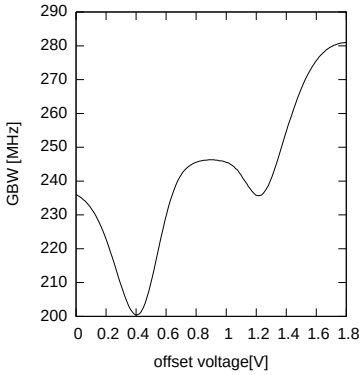


Figure 5.10: Gain-Bandwidth-Product with zener diode

voltages.

5.3 Output Stage

To get an insight into output stages, several have been constructed and simulated. Rail-to-Rail output with current control has only been reached with one. We focus on this stage. In principal it is similar to the CMOS inverter with voltage offset between the gates. Current control is done via translinear circuits as described in 4.5.1. The output stage is merged with the input stage, so we will start with modification of the cascode branch of the input stage. A similar output stage can be found in [16] for example.

5.3.1 Modification of the Cascode Branch

If we look at the DC common mode output of our input stage in the middle of Figure 5.11, the characteristic is already quite flat, which is necessary as the point of operation of the class AB output stage should be as constant as possible to reduce the DC offset dependency of the cross current. The flatness of the characteristic is mainly given via the feedback in current mirroring. What is needed to be done now is to split the output of the first stage into one voltage for the NMOS output transistor and one for the PMOS.

This is done by the insertion of parallel connected NMOS and PMOS transistors between the NMOS and PMOS cascode at the output node. Both transistor types are needed here because the current has to be kept constant. As the voltage of our new output nodes is going down, I_{DS} of the PMOS would decrease, while I_{DS} of the NMOS increases, keeping the potential difference ΔV_{out} constant. The current flow in the cascode branch can now be controlled by the biasing voltages V_n and V_p of the inserted transistors. A higher V_p and a lower V_n enlarges ΔV_{12} which decreases the gate voltages of the current mirrors. The DC voltage characteristic of our new output nodes is shown in Figure 5.11, too.

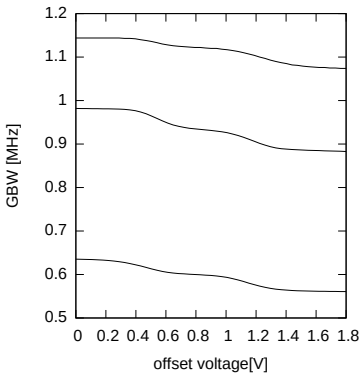


Figure 5.11: DC output of folded cascode intputstage without (middle) and with class ab control circuit

Figure 5.13 displays a schematic of the complete cascode branch with output transistors. Additionally, triple well transistors have been used for transistors M_{Tn} to allow to allow to set the bulk potential on the source potential. Else V_{BS} of the transistors M_t would be high causing a strong body effect. Generally, the choice of a floating well potential is critical, as it impedes matching. If we have a closer look at the generation of the two biasing voltages by two serial diode connected mosts each, we can find some nice translinear circuits in this schematic. Actually we are applying translinear biasing to the cascode branch. Generally we get two equations:

$$V_{GSM20} + V_{GSM19} = V_{GSM11} + V_{GSMn1} \tag{5.7}$$

$$V_{GSM17} + V_{GSM18} = V_{GSM5} + V_{GSMp1} \tag{5.8}$$

Setting all $K'n = K'p$ =1 and all $(W/L) = 1$, we can find equations for the currents:

$$\sqrt{I_{dsM11}} + \sqrt{I_{dsMn1}} = 2\sqrt{I_{Bias}} \tag{5.9}$$

$$\sqrt{I_{sdM5}} + \sqrt{I_{sdMp1}} = 2\sqrt{I_{Bias}} \tag{5.10}$$

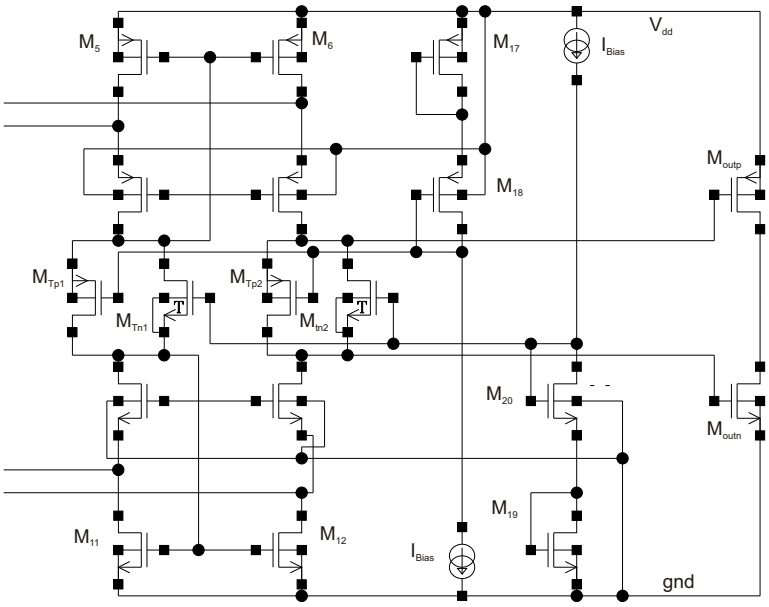


Figure 5.13: cascodebranch and outputstage

Well, what we actually want is to keep the sum of I_{dsMn1} and I_{sdMp1} constant. In our circuit, there are three important points in DC operation depending on whether both input stages are operating, or one. We examine I_{dsM11} and call it I_{ds0M11}. When the NMOS stage starts operating, both stages are operating and one fourth of the maximum current is sourced by the input stage I_{inmax} as the zener diode is consuming 3/4. I_{dsM11} is $I_{ds0M11}(1+c/4)$, where c is the ratio between I_{inmax} and I_{ds0M11}. Finally, when the PMOS stage switches off, we have $I_{ds0M11}(1+c/4)$. Analogue for I_{sdM5} - with the same c for perfect adjustment, what we assume here.

According to Equation 5.9, I_{dsMn1} has to change when I_{dsM11}. Applying Kirchhoff's current law, we assume $I_{dsMn1} = I_{sdMp1} = I_{ds0M11}/2 := I_{crossh}$ when both stages are operating, so $I_{dsMn1} + I_{sdMp1} = I_{crossh}$ at this point. When the PMOS stage is off, $I_{dsMn1} = (1 + c/4)I_{crossh}$ and $I_{sdMp1} = (\sqrt{1 + c/4} + 1 - \sqrt{1 + c})^2 I_{crossh}$. We get an analogue result for the case when the NMOS stage is off. The results are a little bit unhandy, so we set $c = 2$ as a common choice and the design goal in our design. As ratio between the cross current in the middle and near the rails, we get: 0.996, which is actually a very good result including still some potential via the choosing of W/L Of course some assumptions, like the perfect adjustment between NMOS and PMOS stage are not very practical. Problems caused by the g_m compensation have been ignored. A simulation result can be seen in Figure 5.12. The critical points are obviously the switching voltages of the input stages. The decline is caused by an asymmetry between PMOS and NMOS stage. The calculated deviation is overruled by these effects. The deviation resulting from simulations is below 10%.

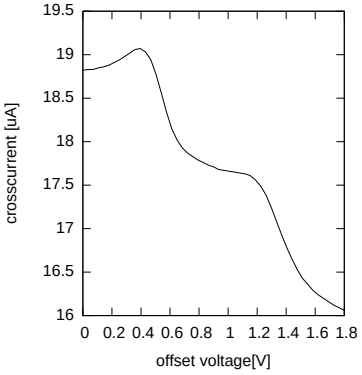


Figure 5.12: cross current of cascode branch adjusted by translinear biasing

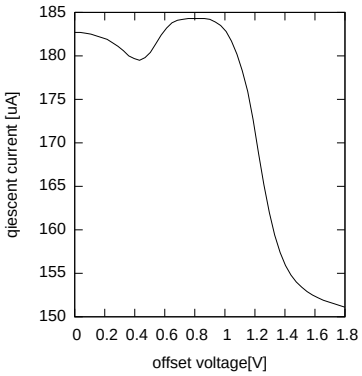


Figure 5.14: quiescent current of the outputstage

5.3.2 Output Transistors

The new output nodes of our input stages are connected to the gates of a NMOS and a PMOS transistor whose drains are connected to each other. Current control is done via translinear circuits again, as described in Figure 4.15. Transistor M_T in Figure 4.15 is equivalent to M_{Tn} respectively M_{Tp} here. The characteristic of the quiescent current of the output stage can be seen in Figure 5.14. It is not constant, but the deviation is not crucial. Best comprehension can be reached by comparison to Figure 5.13. At the beginning, the quiescent current is limited by the PMOS output transistor. With falling input of the output stage, the quiescent current is rising until NMOS and PMOS transistor are balanced. Afterwards, the NMOS transistor is limiting the current.

5.4 Compensation

Compensation is a challenge in our design as only small Miller capacitors are allowed by the small currents of the input stage and the small layout area. Further, our amplifier should not only be stable up to a certain load capacity, but unlimited. Actually, this is only possible including the additional 50 Ω output resistor.

5.4.1 The Output Resistor

Usually, the output low pass without compensation is defined by the output transconductance g_{mout} and the load capacitor C_load resulting in a cut-off frequency of $f_c = \frac{1}{2\pi C_{load} r_{mout}}$ with $r_{mout} = 1/g_{mout}$. At this frequency two effects occur as $\mid Z_{load} \mid = r_{mout}$ at this point. In the voltage divider formed by C_{load} and r_{mout}, C_{load} stops taking the leading role with a voltage drop of barely 1/2. Furthermore, as the dominant role of C_{load} in the total impedance is decreasing, a phase shift of 45° can be observed, which goes up two 90° for higher frequencies. If we put a resistance R between C_{load} and r_{mout} things get a little bit more complicated. Of course, as we want a voltage drop of 1/2 at $Z_{load} + R$, R is not allowed to be bigger than r_{mout}. Now at $f_c = \frac{1}{2\pi C_{load}(r_{mout}-R)}$, the voltage drop is on half. This is a lot bigger, but carefully treaded, as the limit of the impedance is R and not 0 for high frequencies. An important point of the addition of R is its effect on the phase shift. We calculate it as:

$$\varphi = arctan\left(\frac{\frac{-r_{mout}}{2\pi f C_{load}}}{R^2 + \frac{1}{(2\pi f C_{load})^2} + R r_{mout}}\right) \tag{5.11}$$

Directly, we realize, the phase shift is zero for infinite f. This can be seen more easily, if we have a look at the constructed voltage divider - at infinite frequencies, only real parts of the impedance are left. In addition, the phase shift has a minimum frequency different to $-90\,^\circ$ if the ratio between r_{mout} and R is not to big. This is still the case for a ratio of 100 for example. As $R = 50\,\Omega$ and $r_{mout} << 5\,\mathrm{k}\Omega$ no phase shift of $-90°$ is created by the output low pass. A way to see this phase shift without shifts created by other poles, is to apply a very big C_{load} - $1\mu F$ for example. This has been done in Figure 5.15. A pole causing a phase shift like this is called a negative zero. Still compensation is necessary for positioning of the pole.

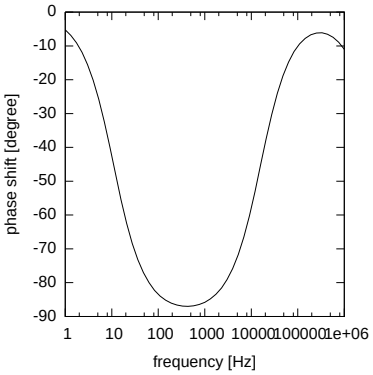


Figure 5.15: output AC phase shift for 1 μF load via 50 Ω

5.4.2 Choice of the Compensation Capacitor

A compensated standard Miller amplifier according to the design plan given in [17] has been designed to get an impression of the order of magnitude of the necessary compensation capacitance. Calculations in the folded designed cascode amplifier are much more complicated due to of the additional output resistance. We will give a short inside into possible calculations without R. As the amplifier should be unlimited stable, we chose a big output load of 1 nF. As GBW, we want to achieve 5 MHz - this results in an equivalent output stage as for 100 pf at 50 MHz. First, we calculate the transconductance g_{mout} as the low-pass frequency of the output stage, corrected by a factor 1.3 should be 3 times the GBW. We get 0.123 S. This results in a W/L of just about 2500 for NMOS transistors. If we choose the smallest channel length, W still has to be 450 um. This results in $C_{GSout} = 640\,\mathrm{fF}$ and hence $C_comp \approx 2\,\mathrm{pF}$. As transconductance of the input stage, we get $63\mu S$ which is no problem at all. The size of the compensation capacitor is a problem if we look at the physical size - metal caps reach only $150\,\mathrm{fF}/\mu\mathrm{m}^2$. The area would have to be more than $1300\,\mu\mathrm{m}^2$ or a square with a length of $36\,\mu\mathrm{m}$.

Generally: Enlarging the GBW at a fixed output capacitor by 10 leads to an input transconductance enlarged by a factor 100 and a compensation capacitor by a factor 10.

The start goal for our amplifier is a GBW of 50 MHz at a load of 100 pF connected via a $50\,\Omega$ resistor. An equal capacitor directly connected would result in a compensation capacitor of 2 pF as above and an analogue transconductance of $250\mu S$ at the input stage. Including the $50\,\Omega$ resistor, this GBW is impossible for a stable amplifier closed loop mode as the low pass frequency of a $50\,\Omega$ 100 pF low pass is 32 MHz. This frequency would result in a compensation capacitor of 1.28 pF.

As start values for the compensation capacitor, 300 fF is selected. With a normal Miller compensation, these capacitors would be placed between the output node and the gates of the output transistors. Here it works better to connect them before the cascodes. This way of compensation is called cascode-Miller compensation [16]. The placement of the compensation capacitors can be seen in the final circuit in Figure 5.18. After many simulation runs and area optimization, the finally used capacitance is 150 fF for each compensation capacitor in the metal cap version of the amplifier.

5.4.3 How the Compensation works

The critical capacitance is the one, which causes the low pass at the output node and that at the input node have the same cut off frequency as then the maximum phase shift created by these poles occurs. My theory is that for small capacitors, both poles, dominant and second, are

working in addition until the second pole has overtaken the first. As the first pole is given via the compensation capacitor, it is possible to adjust the load capacitor, when first and second pole are crossing. This crossing area is actually quite big as the overlap bandwidth can be calculated by with the first pole frequency as $f_{st} * 3 - f_{st}/3$.

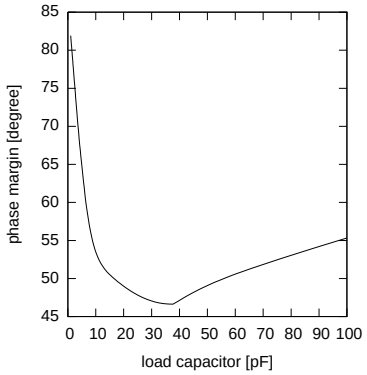


Figure 5.16: Phasemargin

As both, the derivative of the phase shift created by the second pole - calculated in equation 5.11 and the derivative of the attenuation at the output are depending on load capacitance, there can be a minimum final phase shift for a certain capacity. Further analysis would have to be done to proof this. Figure 5.16 shows the phase margin of the final amplifier. The critical capacitor is around 40 pF. The phase margin goes down to less than 70 ° which causes a little peaking in closed loop mode.

5.4.4 Compensation with Gate Capacitors

In the HICANN of the FACETS project, the small version of this amplifier will be placed in an area, where no metal 5 and metal 6 is allowed, as these layers will be used for network routing. This is why we are not allowed to use the process's metal caps there. Gate capacitors are the alternative. The advantage is that the capacitance density of gate capacitors is much higher than that of metal caps. The problem is the dependence on the gate voltage - V_T has to be reached to build a conducting channel below the gate and activate the gate oxide capacitance. Figure 5.17 shows the characteristic of a NMOS gate capacitor(ncap) with $1pF$ maximum capacitance. The threshold voltage can be enlarged further by the body effect. The break down of the capacitance for low voltages is a real problem in our design as for low output voltages, for the PMOS input stage switches off and for high voltages, the one for the NMOS input stage switches off whilst in the middle of output range, both capacitors are at maximum value. The choice is now, to use ncaps for the PMOS branch and pcaps for the NMOS branch and to use values big enough to reach a sufficient compensation even when one stage is switched off. The resulting capacitance is 407 fF each.

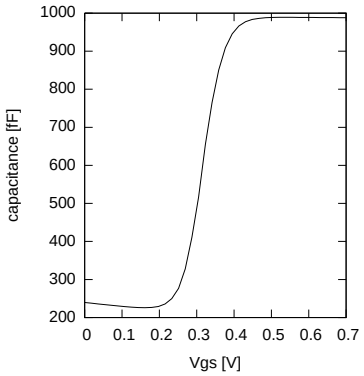


Figure 5.17: Nmos gate capacitor Ω

5.5 The final Design

Here, we will give an overview about the finally designed amplifiers. A last change has been made and will be presented here. The final circuit is a result off a long simulation phase. Conspicuities of the final parameter choices will be discussed.

5.5.1 Final Schematic

The final schematic can be seen in Figure 5.18. It is similar to amplifiers designed by Ron Hogervorst [14, 16] except for some important details. Compensation works completely different

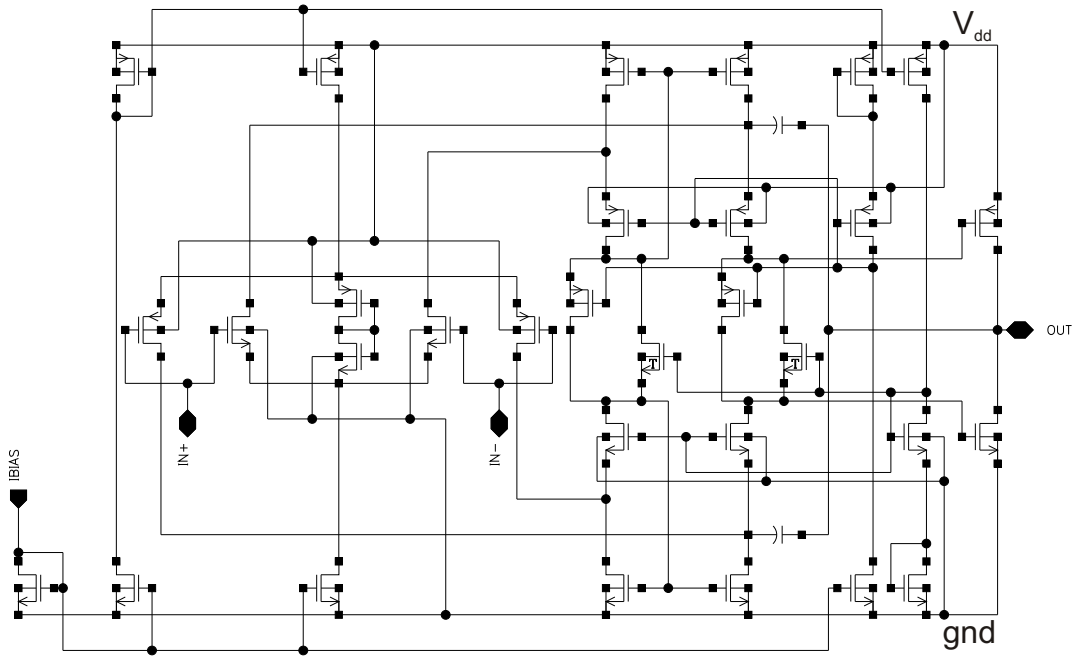


Figure 5.18: Final Circuit of the Small Amplifier

here, as described above and triple well transistors had to be used as our supply range is very small - only the half of the 3.3V used by Hogervorst - and we need to keep the threshold voltages low. Further while, the biasing of the cascodes is done differently which is one of our last design changes, and crosscurrent of the output stage is very small in our design. The connection between the cascode gates and the two serial connected mosts, used for the translinear biasing of the cascode branch and output stage, reduces the number of extern biasing signals to one. This is extraordinary as usually a high number of biasing voltages is needed for parallel folded cascode amplifiers. It is allowed, as the strong inversion area is big enough for the chosen channel length of the cascode transistors.

In comparison to the W/L needed for the output stage of a compensated amplifier with a load of 100 pF at 50 MHz, this output stage is very small. The size of the output stage is a simulation result. For the small amplifier, the W/L is only 44 for the NMOS transistor. It has to be that small because of the small cross current allowed. Nevertheless, the needed GBW is reached. It is engaged by a small peaking allowed in closed loop and the "more than linear" characteristic of the class AB stage. Further more, the GBW of the first stage is higher than that of the second one, so the gain of the first stage shifts the cut-off frequency of the second one and the low pass at the output.

We will do some estimation to get a comparison. Usually a g_m of 20 mS would be needed to get 35 MHz GBW at 100 pF directly connected, using the low pass frequency. This would result in a current of 2 mA at $V_{GS} - V_T = 0.2V$, in strong inversion, hence the figure of merit is $FOM = 1750\,\text{pfMHz/mA}$ at best. From our simulations, we get $FOM \approx 70\,\text{MHz} * 100\,\text{pF}/0.2\,\text{mA} = 35,000\,\text{pfMHz/mA}$!.

5.5.2 Small Amplifier

This Amplifier is just named "Small Amplifier" Some aspects of the small amplifier have already been discussed. At a bias current of 2.5 μA a GBW of 70 MHz is reached with a DC crosscurrent below 200 μA and a phase margin of 55.2 ° at 900 mV input offset in the typical corner. In the

slow corner, the GBW goes down to 64.28 MHz with a phase margin of 50.5°. Simulating the fast corner, a GBW of 77MHz is reached with a phase margin of 58.78 ° The critical corner is the slow one. In closed loop, the maximum peaking has been determined. As described, there is a critical capacitance, when the poles of first and second stage overtake each other. Corner simulations sweeping offset voltage and load capacitor have been done to confirm the stability. No excessive peaking has been observed so far. Some results will be shown in the experimental part. The unity gain bandwidth in closed loop configuration matches the benchmark. The same applies for the current consumption.

5.5.3 Small Amplifier with Gate Capacitors

We call this amplifier "NC amplifier". Most effort has been spend on the design of the normal small amplifier, switching to the alternative compensation capacitor during the layout phase. Except the compensating capacitor, the small amplifier and this NC amplifier are equal. The abbreviation "NC" stands for NMOS capacitor although NMOS and PMOS have been used here. The achieved GBW is around 60 MHz here at 100 pF load. It is lower, as bigger compensation capacitors had to be chosen. The phase margin is 54°. At a critical offset voltage, when output reaches the voltage where the NMOS blocking cap is switched on, the phase margin reaches a minimum. It is around 0.5 V. At 30 pF, the phase margin goes down to 44 ° at this point. This is getting even worse in the slow corner. In closed loop configuration, this amplifier has more peaking than the small amplifier. The worst case voltage peaking is 2.2. As this could be a problem, the ringing of the amplifier has been observed via a transient analysis with a square pulse at the critical voltage. It can be seen in Figure 5.19. A short look at the frequency of the ringing leads to just about 50 MHz which fits to the peak frequency of 57 MHz in the frequency response. The ringing is not problematic. Sufficient constancy of the output is reached after one ringing period.

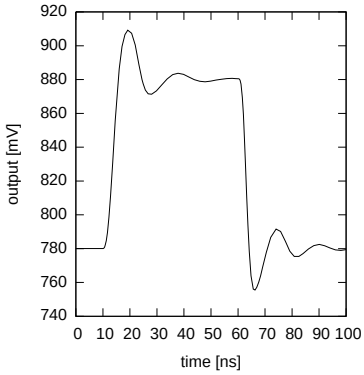


Figure 5.19: Square Pulse Output of NC amplifier

5.5.4 big amplifier

The Big Amplifier has to drive a 50 Ω load at a closed loop bandwidth of 350 MHz with a maximum cross current of 2 mA. This is generally reached by enlarging W/L of the transistors of the small amplifier. It is not sufficient to use the same factor for all transistors. The input scaling factor of the input stage is chosen smaller than that of the output stage, as a lower output resistance is needed. To increase the gain in the middle of the operating range, the scaling of the zener diode is much smaller - we cannot afford to supply so much current here. The current mirrors stay equal to minimize layout complexity - of course, it is necessary, to add a small input current mirroring stage to scale the current this way. The scaling factors resulting for simulations are 6 for the PMOS input stage and 7.5 for the NMOS input stage. The NMOS stage is bigger to allow a more symmetric layout later. The same values have been used in the cascode branch. The zener diode has been scaled by only a factor of 2. The modification of the output stage is more complicated. The scaling factor here is 11.25 for the NMOS transistor. The ratio between PMOS and NMOS has been enlarged from 2.4 to 3to be capable of sourcing higher currents and to simplify the layout process. The channel length of the output stage has been increased to broaden the strong inversion range. As high currents will have to flow through the output stage, ground and power are separated from the input stage. Both, NMOS stage and

PMOS transistor are placed reclusive in a well to insulate their bulks to reduce substrate noise.

Simulated results are an open loop GBW of 177.4 MHz with a phase margin of 71.13° at an offset of 900 mV with a bias current of 3 μA. In closed loop configuration, a GBW of 342.2 MHz is reached. The worst case quiescent current is below 3 mA and can be reduced by the choice of a different bias. The quiescent current does not rise with the input voltage. Figure 5.20 shows the quiescent current in closed loop connection. Near the rails, it goes down as one of the output transistors is biased at low $V_{gs} - V_T$ then. The bump in the middle is caused by the smaller g_m-compensation.

The maximum source able current is 28 mA, 50 Ω output resistance included. The maximum drain able current is 31 mA. The maximum peak in the closed loop frequency response is below 1.13 and hence uncritical.

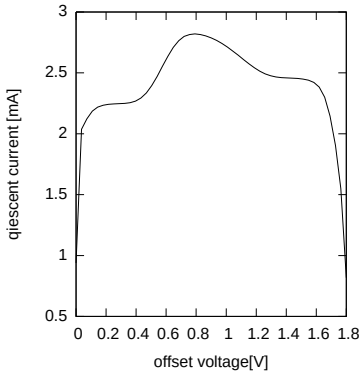


Figure 5.20: Quiescent current of the big amplifier in closed loop configuration

5.6 Layout

In this section, the layout of the amplifiers will be described. We will start with the introduction of some general aspects of layout. Then we will have a look at the layout of some important transistors in the designs. Last but not least we will give some words about the total topology.

5.6.1 General Aspects

In simulations, the complex BSIM 3.3 v3 model has been used. Despite its complexity, it is no copy of reality as not all effects can be included and discrete approximations have to be done. Additional parasitic resistances are caused by the real layout of the transistors.

Mismatch of transistors is caused by gradients in the threshold voltage for example. This mismatch has to be counterbalanced by special layout techniques like common-centroid layout.

The best companion in the layout phase has been the book: "The Art of Analog Layout" by Alan Hastings[25]. We will describe some general aspects here.

Reduction of Resistances

The resistance of poly silicon is high in comparison to the resistance of the aluminium used in the metal layers(The ratio is around 100). This is why the poly silicon should not be used for routing. A gate should be connected to metal directly. A transistor can be divided into several smaller transistors, placed on the same diffusion these are called fingers. Division can be used to reduce the lengths of the gate poly and so the resistance. A sample transistor with fingers can be seen in Figure 5.21. Poly silicon is less critical, when very small constant currents have to be driven, as only a small voltage drop is caused. This has been done for some biasing circuits. In our process, contacts between the metal layers and contacts to poly silicon and diffusion have a very high resistances in a magnitude of several Ohms. Contacts should be placed in parallel and a change between metal layers should be avoided if possible.

Fringing effects

Etching of poly silicon works better if the etching area is big. As etching is done for a certain time, this difference can cause under etching and over etching resulting in smaller or wider gates.

This is why the etching area of matched transistors should be equal. For this purpose, dummy gates have been added in Figure 5.21. Of course, a higher finger number is necessary, as here, etching effects would be the same for both gates even without dummies. Dummy gates should not be left floating. In our design, we connect them to back gate potential which is the common choice.

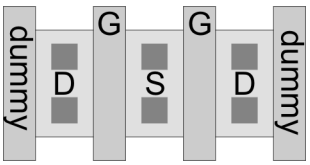


Figure 5.21: transistor with dummy gates

It should be avoided to connect gates with poly silicon, although this makes rooting a lot easier. It causes small niches, it has a strong effect on etching as the poly silicon is used as etching mask in a self aligned process.

Another fringe effect at the poly silicon can be caused by the direction of the ion implant, which may not be perfectly orthogonal. This causes a different implant on both sides of a gate. To reduce influence of this effect, current directions of matched transistors should be equal.

Gradients

Gradients in doping concentrations will occur on a real chip, which has strong influences on the threshold voltage and the transconductance. For better matching, matched transistors should be placed as close to each other as possible, preferably on the same diffusion. The gate area should be big to cancel gradients - the standard deviation is proportional to $1/WL$.

Common centroid layout is another technique to cancel gradients as the deviations compensate each other around the centroid if the gradient is constant which can be expected for small clearances. Hence, the layout should be kept as quadratic as possible to minimize clearances in addition in addition. The simplest quadratic common centroid layout can be done for two transistors with two fingers each and common source. Two diffusions, are used. The arrangement is:

$$_{D}A_{S}B_{D}$$
$$_{D}B_{S}A_{D}$$

Here A and B are fingers of the first respectively the second transistor. As long as the gate potentials of both transistors are equal, design of this layout is relatively easy. If they are not, there has to be a conductor crossing for the drain contacts and for the gates which lead to some compromises in the balancing of the resistors. As more fingers emerge, layout can get arbitrarily complicated. The highest finger number in our amplifiers is 16 per transistor for the PMOS input stage of the big amplifier.

Substrate Noise

A varying bulk potential, caused by hot electron injection for example has direct influence on the threshold voltage. High current devices should be placed far away from matched devices and be surrounded by a ring of bulk contacts to absorb currents. The matched devices should be surrounded by contacts, too, to be protected from bulk intern and extern bulk variations. Bulk contacts should be placed, where ever possible.

Signal Coupling

The gate of matched devices should not be crossed by signal lines as coupling could influence the potential. Crossing should be shifted to higher metal layers.

5.6.2 Input Stage

Matching is very important for differential input stages, as a mismatch would directly lead to an offset voltage between both stages which is different from amplifier to amplifier. Common centroid layout has been applied for all input transistors in the small amplifier and in the big amplifier. We start with a detailed description of the NMOS input stage of the Small amplifier which is the easiest one.

The layout can be seen in Figure 5.22. The outer gates are dummy gates. The dummy gates are surrounded by diffusion to be recognized as real transistor gates by the LVS[2] tool. In principal the design rules allow a diffusion ending below a dummy gate which had been done at first to save area.

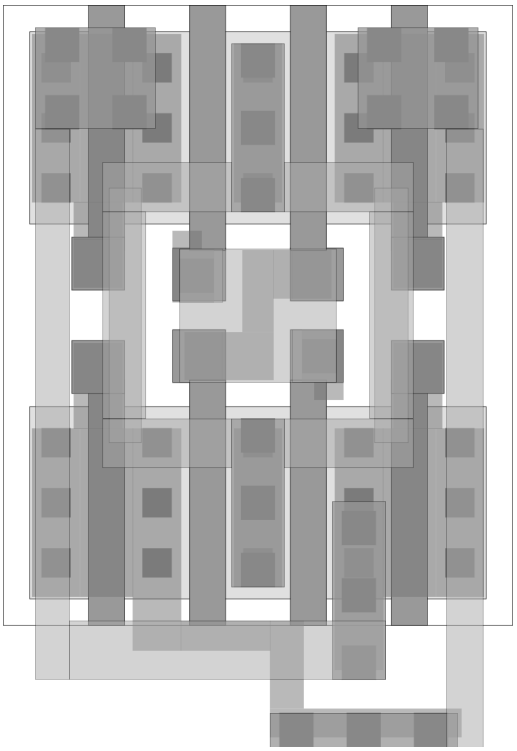


Figure 5.22: NMOS input stage of the small amplifier

The finger arrangement is the same as described in 5.6.1. The source of the transistors is in the middle. Source contacts on both diffusions are connected via a ring of Metal 2 for best symmetry. Source is connected via Metal 3 outside the transistor. The drain contacts are routed via metal 2, too, using the same number of vias for both transistors. The critical area is the crossing of the gates in the middle where only very little space is available. The final solution is a routing in metal layers one and two. The great disadvantage is the additional resistance included by the vias for one of the transistors. This causes some asymmetry. There is no real alternative indeed. Routing both connections on metal one and two cannot be done as there is not enough space for the vias. Routing with poly silicon for one transistor could lead to a similar resistance as two vias, but is not possible as the minimum poly to poly spacing is much bigger than the metal to metal distance. No problems caused by the different gate resistance have been found in simulations including parasitics. The input pins of the gates are on the top of the layout.

The common centroid layout of the PMOS stage is:

$$_{D}A_{S}B_{D}B_{S}A_{D}$$
$$_{D}B_{S}A_{D}B_{S}A_{D}$$

Input pins have chosen to be on different sides in this layout reducing the difference in the gate resistance as no crossing in the middle is needed this way. The crossing has been done on the top and downside, where there is enough place. This way of connection is possible without losing much symmetry as respectively two gates of the same transistor are next to each other.

The layout of the NMOS input stage of the big amplifier is:

$$_{D}A_{S}B_{D}B_{S}A_{D}A_{S}B_{D}B_{S}A_{D}$$
$$_{D}B_{S}A_{D}B_{S}A_{D}B_{S}A_{D}B_{S}A_{D}$$

Regular structures have been used here, to speed up the layout development. Usually the choice would be not to build two rows of 8 fingers, but 4 rows of 4 fingers. This has not been done here, as the finger width is too big and a topology with four rows would be very long. Of course, it is

[2]Layout versus Schematic

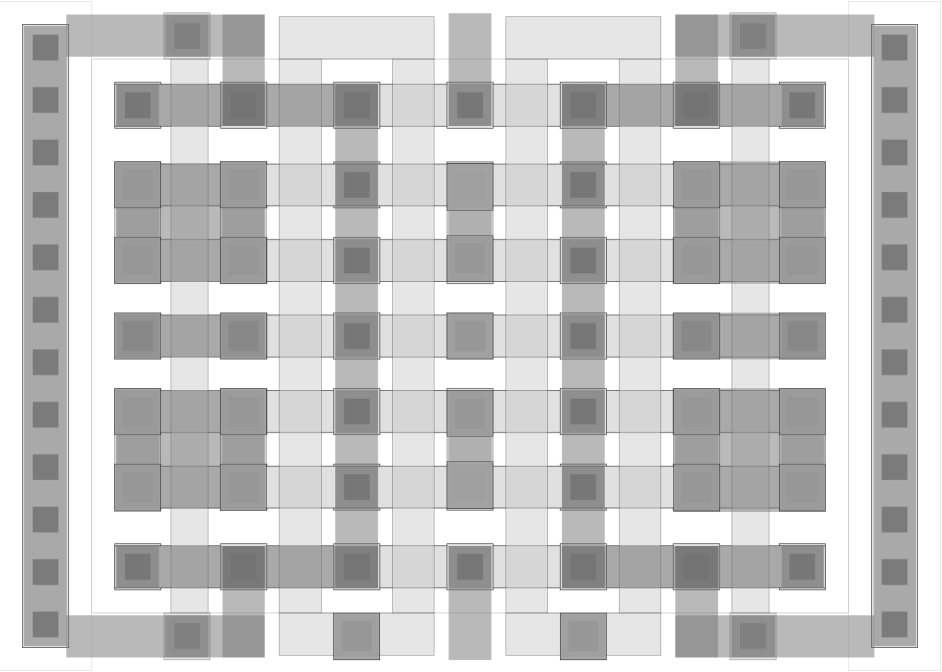


Figure 5.23: NMOS current mirror

possible to divide each finger and layout 16 finger transistors, but this would be area inefficient here, too.

16 finger transistors have been used in for the PMOS input stage of the big amplifier where the arrangement is common centroid, too:

$$
\begin{array}{c}
{}_{D}A_{S}B_{D}B_{S}A_{D}A_{S}B_{D}B_{S}A_{D} \\
{}_{D}B_{S}A_{D}B_{S}A_{D}B_{S}A_{D}B_{S}A_{D} \\
{}_{D}B_{S}A_{D}B_{S}A_{D}B_{S}A_{D}B_{S}A_{D} \\
{}_{D}A_{S}B_{D}B_{S}A_{D}A_{S}B_{D}B_{S}A_{D}
\end{array}
$$

Here, one gate port is in the centre of the layout, whilst the other is on the bottom and the top, keeping the gate resistance as equal as possible at minimum total area consumption.

5.6.3 Biasing

Matching of the current mirror transistors in the biasing circuit is a challenge in our design, as the current has to be mirrored into three different transistors and finally 3 different transistors at the NMOS side. 4 Transistors with 2 different finger numbers have to be matched. The NMOS current mirror, used in all amplifiers will be described here.

Matched current mirror transistors have the advantage, that gate and source are equal, which makes the layout a lot simpler. Nevertheless, the finger number had to be adjusted to avail a symmetric layout. As for the input transistor at least two fingers are needed, to get some kind off symmetry and particularly surround the not common port drain, a high finger number is needed for the biasing transistor of the input stage to achieve some current multiplication. The number two is a problem, as the current multiplier for the output stage biasing and the PMOS stage biasing should be one and hence the finger number of these transistors must be two, too. At least four are needed to build a symmetric, common centroid layout. The compromise is to match the PMOS bias in common centroid layout, whist the output biasing transistor is matched worse. With a finger number of two of the input transistor, the input bias transistor needs 32 fingers. A topology with 5 rows is chosen, whilst one additional row is added on top

and bottom to build the output biasing transistor. Redundant fingers in these rows get source and drain short cut and work as dummies. Using A for a finger of the input transistor, B for the input stage biasing, C for the PMOS stage, D for dummy fingers and E for the output biasing, we get the following topology:

$$
\begin{array}{c}
{}_{D}D_{S}E_{D}E_{S}D_{D} \\
{}_{D}B_{S}B_{D}B_{S}B_{D} \\
{}_{D}B_{S}B_{D}B_{S}B_{D} \\
{}_{D}C_{S}A_{D}A_{S}C_{D} \\
{}_{D}B_{S}B_{D}B_{S}B_{D} \\
{}_{D}B_{S}B_{D}B_{S}B_{D} \\
{}_{D}D_{S}D_{D}D_{S}D_{D}
\end{array}
$$

The usual dummy gates at the end of each diffusion are still not included here. The layout of the bias NMOS bias current mirror can be seen in Figure 5.23

5.6.4 Output Stage

The output stage is less complicated than the ones described above. Nevertheless, the output stage of the big amplifier is very interesting, as it has to be capable of driving high currents of more than 30 mA and the general size is quite big. A single device has been designed, which is used one time for the NMOS stage and 3 times for the PMOS stage. This unit is a transistor with a finger width of 6 μm and a finger number of 20. To reduce the gate resistance, the whole transistor is surrounded by a ring of metal 1 which connects all gates. For higher currents and less resistance, the source and drain areas of the transistor are widened to allow two contacts next to each other. The whole device is surrounded by a closed ring of bulk contacts. Putting four units of the corresponding doting one below the other and connecting the drains orthogonal by wide Metal 2 stripes forms our final output stage. Between the units, the Metal 2 stripes are connected by wider horizontal stripes, leading to a huge metal 6 plane. Wide metal connections are needed to fulfill electro migration rules.

Other Transistors

Matching with common centroid layout has been done for the current mirrors in the cascode branch, too. Looking at the cascodes, matching is a little bit problematic as only the gate potential is common. For the class AB and current control transistors, even the bulk potential is not common which means that some distance between the transistors is already given by the minimum distance of wells on different potential.

As matching of the cascodes is expected to be uncritical, no further matching techniques have been applied here. Both transistors have been placed in the same orientation, but to save total layout area, the distance between the two PMOS cascodes is 15 μm for example in the big amplifier. With our used lengths and widths, the deviation in threshold voltages is 2 mV according to the matching report. Later experiments will show that this matching problem is no issue for the big amplifier. In the small amplifier, the cascode transistors have been put close to one another to improve matching. Nevertheless matching is a problem for the small amplifier as we will see later. The mismatch problem here could lay in the class AB control as here the distance is around 10 μm which causes a variation of 6 mV in the threshold voltage if we apply the $1/\sqrt{WL}$ proportionality of the deviation on the matching report. Matching will be even worse as the used transistors are triple well transistors and placed in different wells.

5.6.5 General Topology

The total dimensions of the small amplifier are 30 μm X 25 μm - around 1/4 of the width of human hair[27]. The big amplifier is around 100 μm X 40 μm in size. The topology of both amplifiers has been chosen close to the schematic - inputs on the left, outputs on the right. The layout of the big amplifier is clearly dominated by the huge output stage, whilst the output stage does not stick out in the layout of the small amplifier.

Chapter 6

Experimental Verification

To verify the functionality and the performance of the amplifiers, they have been produced on a chip, as even corner simulations might not cover the whole operation. Especially for the chosen parallel input stage matching can be a real problem. As the used standard BSIM 3.3v2 transistormodels do not include noise modelling, particularly flicker noise can be a problem, too.

6.1 Prototype Design

The prototype chip is a $1.56 \cdot 1.56\,\text{mm}$ mini ASIC. Mini ASICS are good for prototyping, as there can be nine different chips on a single $5 \cdot 5\,\text{mm}$ slot of a multi project reticle. The effort is small in comparison to a standard MPW run.

6.1.1 Chip Topology

The chip is packaged in a "PLCC44" package as it is easy to handle and Andr Srowig, already made good experience with these packages when testing his Floating Gate chips[24] for the FACETS project in the Electronic Vision(s) group. Another advantage is that the pad ring of the Floating Gate chip can be reused here. There are eleven pads on each side of the chip.

There are three different types of amplifiers and each needs at least five dedicated connections, whilst the power and ground nets will be shared. The output stage of the Big Amplifier needs an independent power supply and ground to avoid influence on the rest of the chip. As most current will flow through these nets, each of them gets two pins to achieve a low-resistance connection to the outside world. With one amplifier of each kind in closed loop configuration, 15 pads are used for signals and power so far. The output of each amplifier is connected to the corresponding pad via a $50\,\Omega$ resistor. These amplifiers will be called independent in the following, as bias, input and output are not shared.

For an efficient disposition of the remaining, a bias multiplication circuit is used, whilst input ports are shared for different amplifiers. This way, six additional amplifiers of each kind are realized. Two at a time are connected in open loop mode, whilst the rest is connected in closed loop feedback. Only two of the closed loop connected ones have a $50\,\Omega$ resistor at the output. This retains the feasibility to perform measurements without internal series termination. There are two identical circuits of each kind to enable direct comparison. The left pins where planed to be used for another layout version of the NC Amplifier which has been abandoned. Now they are used by two additional open connected NC amplifiers.

The area in the middle of the chip is used for 23 additional Small Amplifiers which are connected to probe pads - pads, too small for bonding, but big enough to allow measurements with wafer probes. They are regularly arranged to facilitate automatic measurement. No feedback is

applied to retain all possible circuits. For half of the amplifiers, the input parallel termination is included, realized by a 50 Ω resistance between positive and negative input.

6.1.2 Supply Structures

For the sake of completeness, a short description of the bias multiplier and the padring, including ESD[1]-structures is given in the following.

Bias Multiplier

The Bias Multiplier is a circuit consisting only of PMOS and NMOS current mirrors. The bias current is quadrupled using NMOS mirrors and then doubled with PMOS mirrors as we need current sources. The main problem of this component is the layout, as two identical amplifiers could not be compared if the bias currents do not match.

A star-shaped topology has been chosen, placing the NMOS transistors in the middle and one PMOS circuit on each side. Matching between the PMOS circuits is not important as each of them is an independent current mirror circuit.

For optimal symmetry and to facilitate a common centroid layout5.6, each transistor needs to have four fingers. The transistors of one mirroring circuit are placed on the same diffusion. The input transistor is placed in the middle with its source pins on the exterior side. As the transistors of the current mirrors have common source connections, now two fingers of the first output transistor can follow on each side leaving exterior sources again. We continue with the other output transistors resulting in a layout where each output transistor has got the input transistor in the middle. The diffusion is completed by a dummy gate.

Pad Ring

The Pad Ring, designed by Andr Srowig, has the ability of providing four supply voltages whilst one is ground. These voltages are spread between the pads via broad metal lines on layer one to four. RF-transistors are used for the ESD-structures to reduce noise. Generally, the ESD-structures consist of negative biased diode connected NMOS and PMOS devices between the input pin and power and ground respectively. When the input signal exceeds the power voltage, the NMOS diode is conducting. For input signals below ground, the PMOS diode is conducting. This way, the circuit is protected against high input voltages.

6.2 Experimental Setup in General

A PCB[2] has been designed to accommodate the chip and peripheral components. As high speed signals will have to be measured, special instruments are necessary. Those will be described here. The setup itself has been measured to exclude errors in the later measurement of the amplifiers. During measurements, some gaps in the design of the setup have been uncovered which have to be compensated.

6.2.1 Circuit Board

The assembled printed circuit board can be seen in Figure 6.1. A PLCC44 SMD[3]-socket is used to hold the chip.

[1] Electrostatic Discharge
[2] Printed Circuit Board
[3] surface mounted device

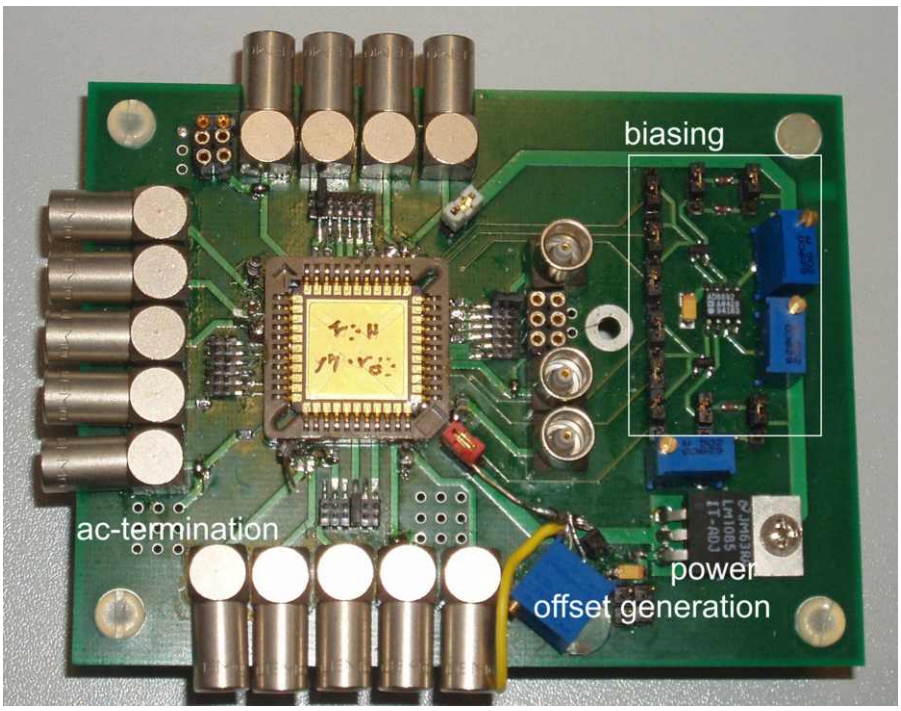


Figure 6.1: Measurement PCB including some fixes

Inside the chip the distances a signal has to cross, are short in comparison to the signal wave length. This is not the case for the measurement setup, so 50 Ω transmission lines have to be used. The continuity of the line impedance on board is realized by a ground plane and a certain line width[4]. As for normal board thickness of about 1.4 mm very wide lines would have to be used, a thinner stack-up with a thickness of 0.5 mm is chosen. Now the required transmission line width is around 0.9 mm.

The termination of the transmission lines is done as close to the chip as possible using a 51 Ω SMD resistor. For the connections to the instruments and signal sources, lemo jacks are used.

The outputs of the amplifiers have to be switched, as it would be very inefficient to host far more than 23 jacks on board. The switching has to be done with only little influence on the impedance of the transmission lines. This is why normal jumpers using vias are not practical. Special SMD-jumpers with a pitch of 1.27 mm are cut off on one side and soldered on the transmission line[6].

The input signals, the output signals of the independent amplifiers and the output signal of one open loop configured big amplifier are not switched but directly connected to lemo jacks to retain the possibility of measurements without jumpers. Each of these output signals can be connected to a field of nine through-whole pins, one connected to ground, to allow the plug of additional components like capacitors. The switching of these fields is done via an empty 0603 SMD-footprint as it fits on the transmission line without changing the width. This footprint can be closed using a drop of solder, when connection is needed. The same technique is used to allow the short cut of the positive and negative input ports for common mode measurements.

A star shaped arrangement has been chosen to minimize the length and the bends of the strip lines on the PCB. Right angled bends are avoided as each would create an impedance discontinuity and hence an antenna and reflections.

Small lines are used for the biasing currents, which are generated in distance to the high

[4]The impedance depends on the width of the line, its distance to the ground plane and the dielectric constant of the PCB base material.

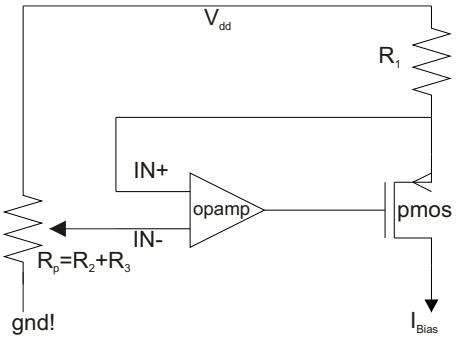


Figure 6.2: Biasing circuit

frequency signals, as space and length are uncritical here. The ohmic resistance of the biasing lines is not important as the currents are only very small and so is the voltage drop.

All ground pins of the chip, including the ground of the output stage of the big amplifier, are directly connected to the ground plane on the PCB. This way the ground resistance is minimized and no additional jumpers consume the worthwhile area near the chip. The disadvantage is, that it is not possible to place the ground of the chip on a different potential than the ground plane.

Power is driven to the chip via sick lines which can be disconnected via jumpers. The power of the small amplifier and the big amplifier can be switched off separately. The power supply is done by a 3 A adjustable linear regulator[32].

The AC blocking of the DC signals is done using 10 nF ceramic capacitors and 4.7 μF tantalum capacitors mostly located on the backside of the board directly under the package. The smaller capacitors are directly connected to the input pins by vias.

Biasing Circuit

Each bias connection can be switched between ground and the bias current source using jumpers. There is one current source for the independent amplifiers and one for the others to allow simultaneous operation at different bias currents.

At first, the biasing current has been generated using a voltage source driving a resistor which could be modified with a potentiometer. As the current in this setup would have a strong dependency on the input resistance of the biasing circuit of each amplifier, this solution has been abandoned.

The chosen solution is a current source with an operational amplifier and a PMOS transistor. A schematic of the circuit can be seen in Figure 6.2. The operational amplifier tries to keep the voltages at $IN+$ and $IN-$ equal, so the current through R_1 is kept constant. R_p is realized by a 5 kΩ potentiometer. R_1 can be switched between 22 $k\Omega$ and 220 kΩ using jumpers to achieve an adjustable bias current up to 50 μA with higher precision in adjustment for smaller currents by switching to the bigger resistor.

6.2.2 Instruments

During simulations, ideal signal sources and measurement probes are available, imposing no borders. The following real-world instruments have been used during the measurements.

Network/Spectrum Analyzer

The HP 4396A from Hewlett Packard has not only been used for analyzing but also as a signal source. It provides a HF output between 100 kHz and 1.8 GHz with a resolution smaller than 1 mHz and possible full span bandwidth. The power is adjustable between −60 dBm and 20 dBm. The analyzing functions will be described next.

Network Analyzer In general, the network analyzer compares magnitude and phase of two 50 Ω signals. In connection with a "T/R-Testset" it can be used to analyze the transmission and reflection coefficients of a system. This test set splits the HF-signal. One half is measured at an input port, whilst the other can be connected to a device under test. A second port is used to measure reflections in the signal, whilst the third port of the analyzer can be used to measure the output of the device under test. Calibration functions allow the compensation of setup parasitics.

The network analyzer can be used to measure the frequency response of our amplifiers in one single measurement as the HF-source is capable of providing a continuous spectrum. Usually lots of measurements with a sinus generator and an oscilloscope would have to be done. The measurement of the reflection coefficient can be used to determine the impedance of a transmission line.

Spectrum Analyzer A frequency range between 2 Hz and 1.8 GHz can be measured with the spectrum analyzer. The available resolution bandwidth is between 1 Hz and 3 MHz. A measurement averages over the corresponding resolution bandwidth. For further averaging, the time resolution can be adjusted between 3 mHz and 3 MHz to exclude fluctuations. The internal noise level is below 135 $\frac{\text{dBm}}{\text{Hz}}$ for frequencies above 10 kHz

Arbitrary Waveform Generator

The AWG2041 from TEKTRONICS is used as a function generator here. In this mode, it can generate the needed sinus and square signals for frequencies between 1 Hz and 10 MHz with an amplitude between 20 mV and 2 V into 50 Ω. The offset can be set between −1 V and 1 V with a resolution of 1 mV. This instrument will primarily be used during gain measurements and slew rate determinations.

Data Generator

The DG2020 A is a Data Generator from Sony/TEKTRONICS. This instrument has been used at the beginning of the measurements, as the AWG was not available. It is able to generate parallel digital data, for the analysis of digital circuits, with an amplitude between 1.2 V and 5 V with a maximum frequency of 200 MHz. This instrument has been used to obtain the slew rate until the AWG was on-hand. The generated data has only been a digital clock used as replacement for a perfect square pulse. As the frequency of the data generator is to low, the Fourier series of the square pulse is abandoned to early resulting in a signal which is only sufficient, if nothing more than a digital clock is needed. The etches are precipitous but there are some small oscillations at the beginning of each pulse. A sample shot of the generated clock signal, together with the response of the designed amplifier can be seen in Figure 6.3. Clearly, the slew rate is already limited by the slew rate of the input.

Oscilloscope

The used TEKTRONICS digitizing oscilloscope "TDS 784 A" has an analog bandwidth of 1 GHz and a digital sampling rate of 4 GS/s. The input ports can be terminated with 50 Ω or 1 MΩ.

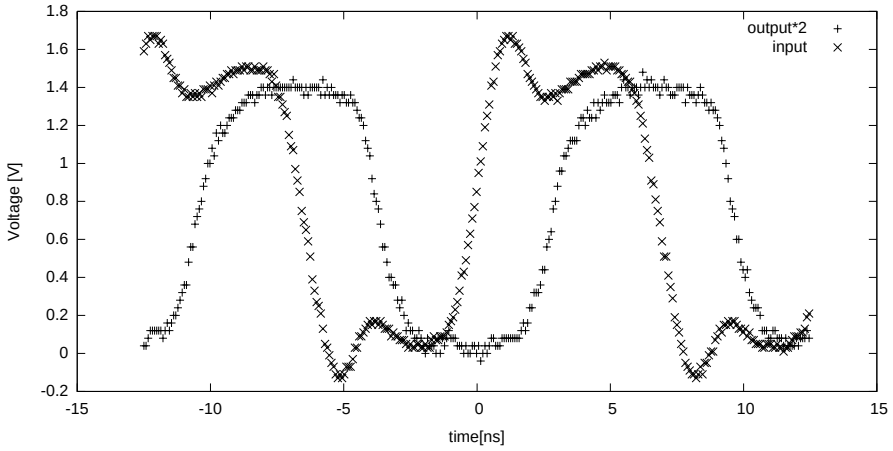


Figure 6.3: The input signal is the signal created with the data generator

Also AC termination or grounding is possible. Low capacitance active probes are available, too.

Further Instruments

A lot more instruments have been used:

- a laboratory DC power supply NGL 20 from Rhode & Scholz
- a PHILIPS multimeter PM 2525 called precision multimeter multimeter in the following
- a Fluke 73 III multimeter, which has a lower voltage resolution
- a HP 34401 multimeter
- a FLUKE PM 6306 RLC meter to measure board parasitics

6.2.3 Problems and Changes

The HF-frequency output has no DC offset, but a DC offset is needed to perform sweeps. The AC signal signal has to be coupled and an AC termination has to be used. For this purpose the DC termination is removed and an upright capacitor is soldered on the ground pad of its footprint. It is connected to an upright $50\,\Omega$ resistor on the signal pad. This way, the termination works only for AC signals and a DC offset can be applied without huge currents. To ac-couple the signal, two connectors with different capacitors are constructed. The values will be described in the corresponding experiment.

6.2.4 Measurements

A lot of measurements have been done to prove the functionality of the setup. Figure 6.4 is showing the results of a reflection measurement with direct termination via a $50\,\Omega$ end cap, AC coupling and AC termination. As can be seen, there are some bumps in the curve. These are caused be resonances in the cable. They could be used to calculate the ϵ_r of the cable dielectric. This has been done in another experiment resulting in $\epsilon_r \approx 0.6$ which is a little bit less than the typical values for a coaxial cable[27]. The difference might be caused by the cable connectors.

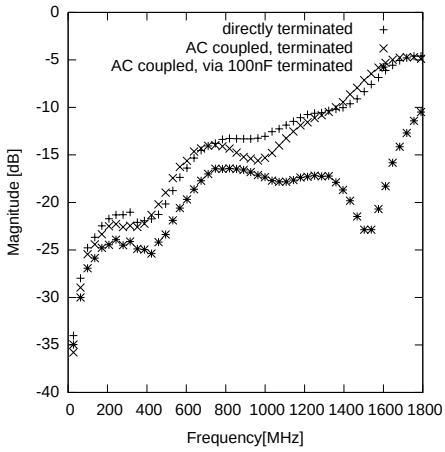


Figure 6.4: Reflection coefficient

6.3 Experiments

6.3.1 Initial Measurements

To check basic functionality of the chip, the independent small amplifier of the first chip (chip I) has been examined. It has independent input, biasing and output pins, so the influence of parasitics is minimized. The first check was to connect any bias to the biasing pin and a DC voltage, generated by the offset circuit, to the input pin. The result was that the amplifier is able to reach both rails within an offset of a few millivolts even with zero bias.

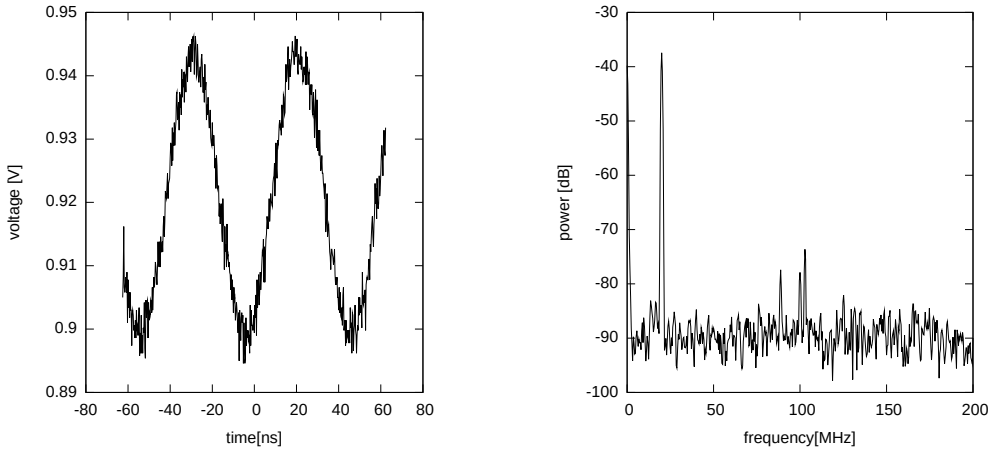
The next step was a short verification of the AC operation of the amplifier. At first, the output was measured using a 50 Ω cable connected to a high impedance input terminal of the oscilloscope. Even for frequencies below 20 Mhz, reflections still damage the signal in a way, no precise measurement is possible. A high impedance 1 GHz probe, which is directly connected to the lemo output connector on the board solves this problem. The output of the amplifier driven with a simple 20 MHz sinus signal together with its FFT[5] can be seen in Figure 6.5. Increasing of the frequency leads us to cut-off frequency of roughly 120 MHz for an offset of 900 mV. No critical peaking is seen during frequency sweeps even if a large capacitor to ground is put on the output. Large signals, oscillating between both rails can be reproduced nearly undistorted. A sweep of the offset between both rails shows, that the operating range is rail-to-rail with an accuracy of millivolts. Clearly, the gain-bandwidth-product goes down near the rails.

The first measurement showed that the tested amplifier on chip I seems to work well. More detailed and precise measurements will have to verify this, but still these results are very promising.

6.3.2 Biasing Circuits

For faster measurement, an equivalent biasing voltage has been used during the experiments instead of the biasing current. This is the voltage dropping out at the corresponding biasing circuit of chip I, when a certain bias current is flowing. Here the dependency between current and voltage is examined for all bias circuits of the independent amplifiers of chip I and for two bias multipliers of chip I.

[5]Fast Fourier Transformation



(a) 20 MHz sinus output

(b) FFT with square window, 50 point average. First peak is the signal, next are radio stations

Figure 6.5: First sketch sinus measurement

Setup

The setup is uncomplicated here. The jumper, connecting to the current source is replaced by a high precision multimeter(Section 6.2.2) which is capable of measuring micro amperes, whilst the voltage is measured with the probe of another multimeter directly at the chip pin. The measured internal resistance of the multimeter for the voltage measurement has been estimated by current measurements with the other to $R_i \approx 11.1\ M\Omega$. The resolution of the voltage measurement is 1 mV

Measurement and Evaluation

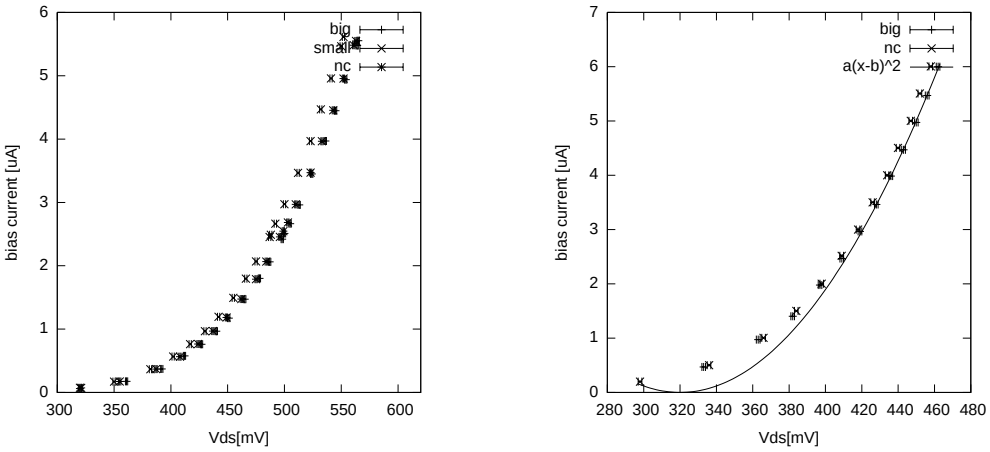
Figure 6.6 visualizes the result. The curves in each plot should be ideal. All amplifiers have identical bias input transistors and in the second plot even the whole circuit is alike. Differences are caused by process variations. We see the characteristic of a diode connect most. Figure 6.6(b) includes a fit to $f(x) = a * (x-b)^2$ which has been fitted for voltages above 420 mV which should be in the strong inversion area. The fit's results are $a = 298 \pm 6 \frac{\mu A}{V^2}$ and $b = (320 \pm 1)$ mV. b fits quite well to the threshold voltage of the used NMOS, but c is a factor of 4 too small to fit on K_n' as $\frac{W}{L} = 8$ here.

6.3.3 Current Draw

As the smaller version of the amplifier has been designed to work in a large scale system, power consumption is critical. The specification was a maximum of 200 μA for the small amplifier and 2 mA for the big amplifier.

Setup and Measurement Plan

We will examine the dependency on the DC offset and the equivalent biasing voltage. The offset voltage is given via the 5 kΩ potentiometer connected over a 100 kΩ resistor. Voltage measurement is done with a multimeter directly at the input pin. For current measurement, the power connection jumpers are replaced by the precision multimeter. During the measurements with the small amplifier, the power supply for the output stage of the big amplifier is shunt off. Measurement will be done with the independent versions of the small amplifier and the big



(a) independent amplifiers (b) Bias multiplier

Figure 6.6: Dependency between Bias currents and drain-source voltage

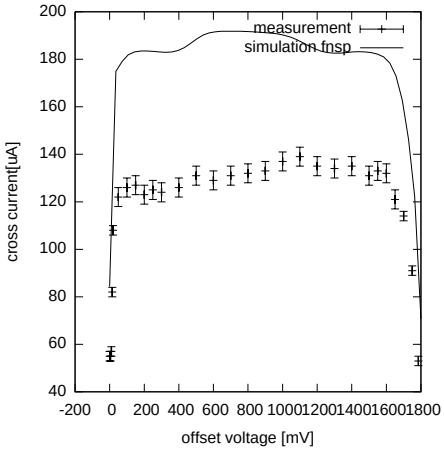
amplifier, not connected to any load except the measurement instrument, to distinguish between load current and cross current. As the Small Amplifier and the NC Amplifier are identical except for the compensation capacitors, there is no difference for DC measurements.

Measurement and Evaluation

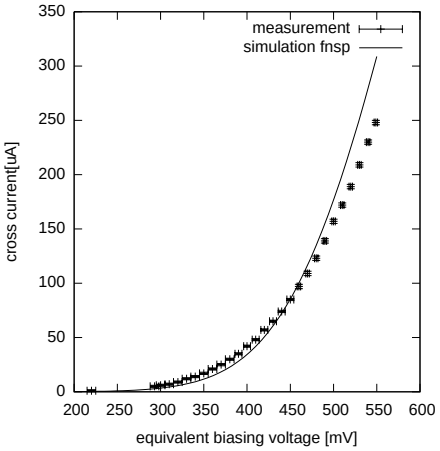
Small Amplifier The measurement is done with an equivalent biasing voltage of 500 mV which corresponds to a bias current of $(2.8 \pm 0.5)\,\mu$A according to Figure 6.6. Figure 6.7 displays the results compared to simulation results. Simulations have been done with parasitics. For the DC sweep, simulation and measurement only fit qualitatively - there is an offset of $50\,\mu$A. This offset can be explained by differences in the biasing current. The sweep of the biasing voltage demonstrates this problem. Nevertheless, the important results are that the offset curve is nearly flat except near the rails and that the current consumption can easily be controlled by the biasing voltage or current. The flat offset curve is the result of the g_m-compensation and the class AB-control(see Chapter 5). The switching of the PMOS and NMOS stage is clearly visible. In the middle of the sweep, both stages are working in strong inversion whilst near the rails only one input stage is operating. Another confirmation we get from the interpretation of this result is that the amplifier is switched off, when the bias is set to ground. This is very important, as without this information, the measured crosscurrent could be influenced by other amplifiers than our device under test.

Big Amplifier The results for the Big Amplifier are as nice as that for the small one. For the DC sweep, measurement and simulation quantitatively nearly fit which is a very good result as the problems with the equivalent biasing voltage are the same as above. The bias voltage was 527 mV here, resulting in a current of $2.8\,\mu$A. Figure 6.8 illustrates the results. The bump in the middle of the offset-sweep is much more distinct than for the small amplifier. This behavior is intended as the big amplifier is a scaled version of the small one with less g_m-compensation to achieve a higher GBW.

Finally the current sourcing and sinking capability of the amplifiers has been checked. The big amplifier is capable of draining and sourcing currents with a magnitude of 25 mA. Both versions of the small Amplifier reach 4.5 mA.

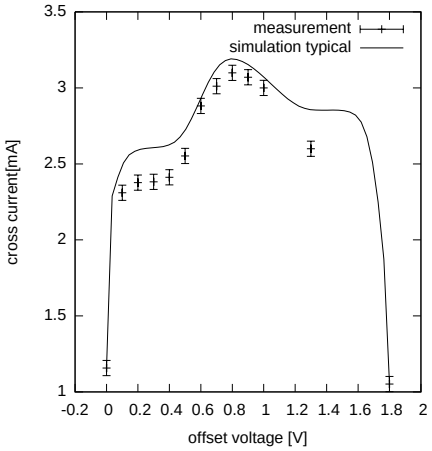


(a) dependency on offset voltage

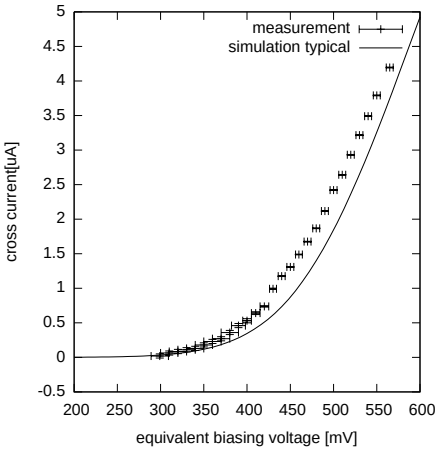


(b) dependency on equivalent biasing voltage

Figure 6.7: Cross current of the small amplifier



(a) dependency on offset voltage



(b) dependency on equivalent biasing voltage

Figure 6.8: Cross current of the big amplifier

Conclusion

Measurements very well approved the functionality of the g_m-compensation and fit to the predictions. The current consumptions are within the requirements for the small amplifier and a little bit above for the big one. The overstepping is tolerable. Sourced and sinked currents are big enough.

6.3.4 Stability

Stability is the most critical value of our amplifiers as the blocking capacitor has to be small to save chip area and current in the input stage. Even the small capacitors chosen for the small amplifier nearly cover the entire layout area. During simulations, maximum peaking has been observed for a certain load capacitor. This has to be checked. The top peaking allowed in simulations was 1.5. Measurements will have to prove this.

Setup and Measurement Plan

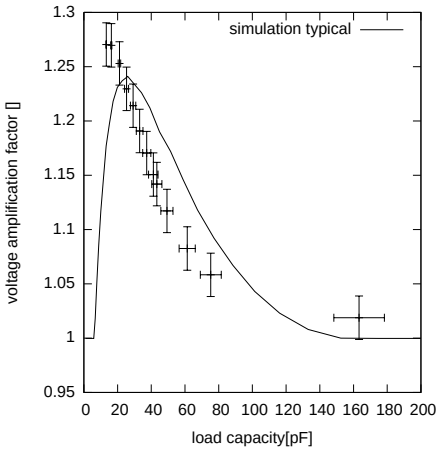
The network analyzer will be used to measure the frequency response of the amplifiers with independent input and biasing pins, which are closed loop connected. The signal is coupled in via a 22 μF capacitor and the AC termination uses a capacitor of this value, too. For the small amplifiers the output is measured with the active probe of the analyzer whilst the 50 Ω port is used for the big amplifier. Discrete values of load capacitors are used to sweep the load. A problem of our method is the huge relative error of the test capacitors which is 10 %. As biasing current we choose 2.5 μA. The DC offset voltage is given via a potentiometer connected to the line via a 100 kΩ resistor, to leave the impedance nearly unchanged. At a certain offset voltage and at a typical load capacitor, the chips will be swapped, to get an overview of process variations. At a certain load capacitance, the bias current and the offset voltage will be swept.

Measurement and Evaluation

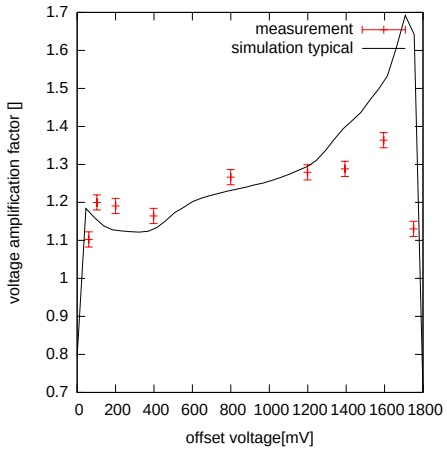
Small Amplifier Our first issue is the load dependency of the maximum peak. There is supposed to be a maximum peak at a certain capacity. This peak is not observed in the measurement. This raises the suspicion that the parasitic capacitors created by bond pads, board and lemo connector and probe connector are much bigger than predicted. Gauging the capacitor with power supply turned off with a RCL-meter gives (12.7 ± 0.1) pF and the capacitor of the probe is 0.6 pF . Some pF more might occur when the chip is turned on caused by the ESD structures in the bonding pads. Figure 6.9(a) shows the sweep results at 13.3 pF together with a simulation for the same bias and offset in the typical case. The amplitude is not going down for small capacitors, but there is a small plateau at the beginning indicating a maximum. The shift of the peak and the larger amplitude are caused by additional parasitics and process variations. Errors are caused by the relative error of the test capacitors and by fluctuations of the output peak during the measurement.

We continue varying the offset voltage with a load capacitor of 12 pF and respectively 25.3 pF with the parasitics described above. Figure 6.9(b) shows the measured amplification together with the simulation with typical transistors. Our results fit qualitatively to the simulation. There are peaks at low and high voltages connected via a plateau. Due to the limited resolution, the maximum at higher offset voltages is not included.

Finally a specific measurement at an offset voltage of 890 mV with a total load capacitor of 25.3 pF is done. We get an average peaking of 1.30 ± 0.06. This is worse than the simulations, but still tolerable for the intended application of the amplifier.



(a) varying the load capacity



(b) varying the offset voltage at 13.3 pF total load

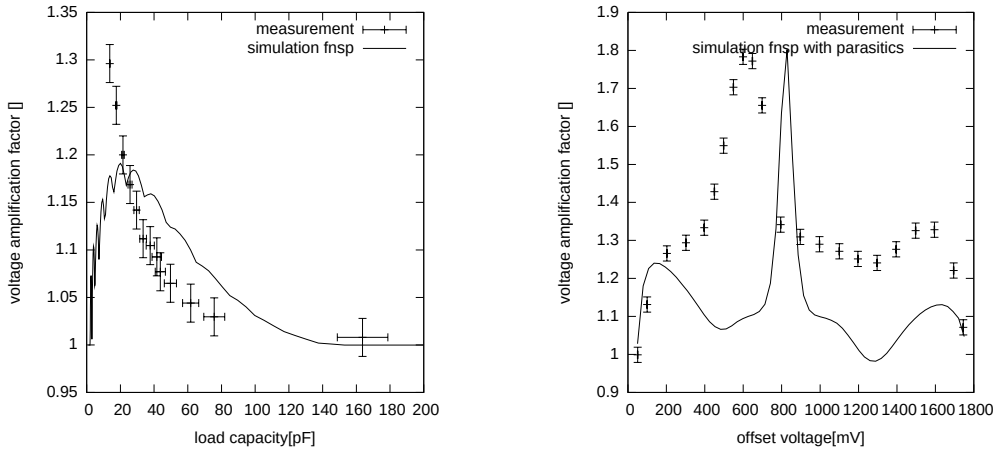
Figure 6.9: Maximum peak amplification of small amplifier in closed loop configuration

NC Amplifier Again, we start with the load dependency. The total of the setup parasitic capacity is $(13.7 \pm 0.2)\,\mathrm{pF}$ in this case. Comparing outcomes of simulation and measurement in Figure 6.10(a) reveals a drastic difference. No reduction for small loads could be observed so far and the maximum peaking amplitude is much higher. Variations by adding possible parasitic resistors or reducing the compensation capacitors in the circuit and corner simulations do not indicate any similar behavior. Measurements with several chips result in an average amplification factor of 1.17 ± 0.03 at $25.7\,\mathrm{pF}$ real load, so the amplifier characterized initially seems to be average. Measurements on setup with less parasitics would have to be done to further examine the behavior for low capacitances and to look for a decline of the maximum peak. Looking at the value of the maximum in the measurement and in the simulations leads to the prediction that the parasitic capacitance introduced by the setup might be much bigger than measured.

In a next step we vary the offset voltage. As can be seen in Figure 6.10(b) the measured result fits qualitatively to the snfp(see Section 3.2.2) simulation with parasitics. The peak in the middle of the voltage range is caused by the offset dependence of the pcap and ncap blocking capacitors (see Figure 5.17 as reference). The position of the peak has a direct dependence on the threshold voltages. A lower NMOS threshold voltage shifts the peak to the left, whilst a lower PMOS threshold voltage causes a shift to the right. This is why a snfp simulation has been chosen as comparison to the measurement to find an explanation for the peak position. The simulation is close to one aloof the peak because very high blocking capacitors had to be chosen to compensate the peaking in the middle.

Condensed, the offset dependency is qualitatively verified, whilst the dependency on the load capacity would still need more measurements.

Big Amplifier As peaking is less critical for the big amplifier, driving a $50\,\Omega$ transmission line, it is not necessary to examine the peaking in dependency to the capacitive load. Simulations showed no jeopardized points. The independent Big Amplifier on Chip I has been examined at specific offset voltages and an equivalent bias voltage of $500\,\mathrm{mV}$. Plots can be seen in the next subsection in Figure 6.13. Peaking remains below a factor of 1.1.



(a) varying the load capacity at 890 mV Offset voltage

(b) varying the offset voltage at 25.6 pF total load

Figure 6.10: maximum peak amplification of small amplifier with ncaps and pcaps in closed loop configuration

Conclusion

Measurements approved the stability of the amplifiers for the desired applications. Results were not perfectly reproduce able using the simulations, but this is no problem as only discrete models for process variations exist and not all parasitics are known.

6.3.5 Gain Bandwidth Product

The goal for the GBW has been 35 MHz for the small amplifier for a load of 100 pF via a 50 Ω resistor. For the big amplifier, 350 MHz had to be reached at a 50 Ω transmission line.

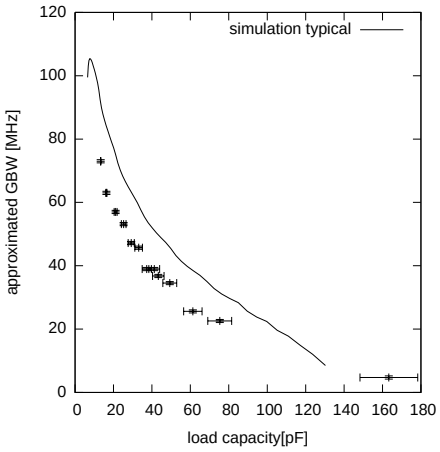
Setup and Measurement Plan

The GBW has been observed in closed loop mode during the same experiments as the stability, so setup and sweeps are equal. For the determination of the GBW, the last crossing to the line of amplification factor one is used. This is only an approximation as the −3 dB bandwidth is needed for the correct definition. The simulations, done in the design part of this work, use the right definition for the bandwidth. During verifications, simulations will be done with the approximated GBW, too.

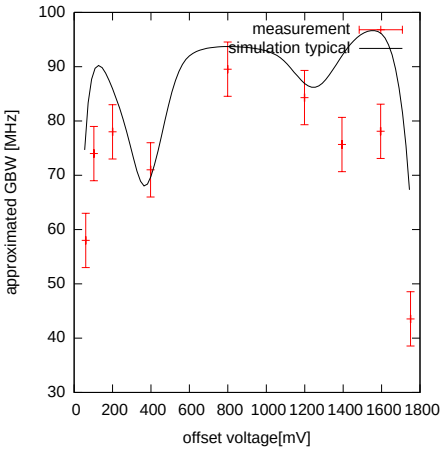
Measurement and Evaluation

We are using the knowledge from the subsection above for the interpretations of the results regarding parasitics et cetera. The error of the approximated GBW is caused by fluctuations which are mainly a result of antenna effects. The position of the operator relative to the coupling between HF-source and T/R-Testset has significant influence on the measurement results as environment influence is isolated. This causes an estimated error of 0.5 MHz on the approximated GBW.

Small Amplifier The connection between the approximated GBW and the load capacity for measurement and typical simulation with parasitics can be seen in Figure 6.11(a). The result fits to the simulations except a shift of 10 pF which could be caused by the bond pad. Hence,

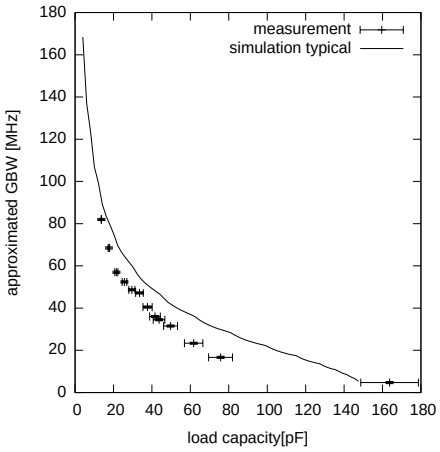


(a) varying the load capacity at 890 mV Offset voltage

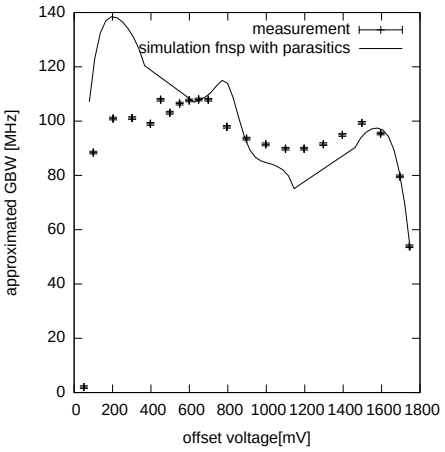


(b) varying the offset voltage at 13.1 pF total load

Figure 6.11: approximated GBW of small amplifier in closed loop connection



(a) varying the load capacity at 890 mV Offset voltage



(b) varying the offset voltage at 25.6 pF total load

Figure 6.12: approximated GBW of small amplifier with ncaps and pcaps in closed loop configuration

including this additional parasitic and switching to the −3 dB bandwidth, the desired GBW of 35 MHz at 100 pF is reached.

During the measurement of the offset voltage sweep, some problems with the biasing circuit occurred resulting in a higher fluctuation of the values. Measurements with a bias sweep showed, that the change of the approximated GBW is ±5 MHz for a variation of ±10 mV equivalent biasing voltage at the chosen voltage of 486 mV. Results are visualized in Figure 6.11(b). The measurement nearly fits to the simulation which is a very good result considering all possible error sources.

NC Amplifier Figure 6.12(b) displays the dependency between approximated GBW and offset voltage for the measured amplifier and a simulation with fnsp model, at 25.6 pF total load. The matching is accurate for voltages above 800 mV whilst for smaller inputs the simulations provides higher results. This behavior is could be caused by even smaller threshold voltages as in the corner model. The stability check above gives a hint for this theory, too.

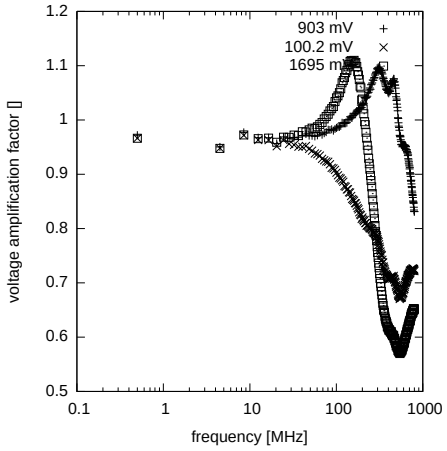


Figure 6.13: frequency response of the big amplifier at 500 mV biasing Voltage for different DC offset voltages.

Figure 6.12(a) shows the results of the load capacity sweep. The results fit to the simulations. Only a small additional capacitor would be needed to shift the curve. The target GBW is not reached here, a is below 50 MHz for 50 pF yet. With additional parasitics caused by the ESD structures and a shift upwards using the correct determination of the GBW, the target can be fulfilled.

Big Amplifier Figure 6.13 is showing the frequency response for the different offsets. The raise of the GBW in the middle of the operating range can clearly be seen. The GBW for higher voltages is bigger because of the lager PMOS output transistor. A GBW of 350 MHz is easily reached. For higher frequencies, the amplitude is increasing again. The reason for this behavior are setup parasitics as examined before.

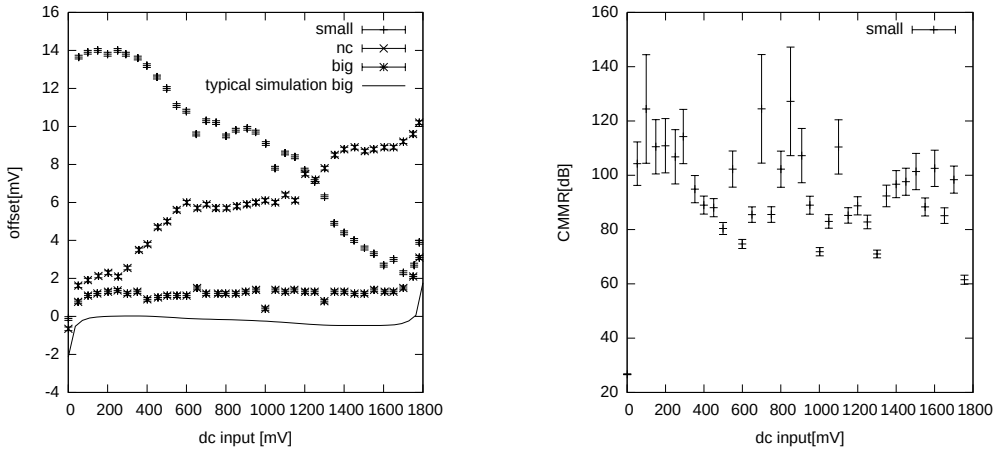
6.3.6 Closed Loop DC Offset and Common Mode Rejection Ratio

Generally there are two sources causing an offset between input and output in closed loop operation. The first one is the difference between the common mode output and the closed loop input which has to be compensated by the open loop gain. This generates an offset of $\frac{\Delta U}{gain}$. Hence a variation of the common mode output during a common mode sweep has a strong influence.

The second source is mismatch between the input transistors of the input stages. This effect is much more important as gain minimizes the effect of the first cause to less than one millivolt if we stay in distance to the rails. As we have two input stages working in parallel each will have its own mismatch offset.

DC-offset will be examined in detail for one chip and at three points for all chips chips to get some kind of statistic.

The Common Mode Rejection Ratio (CMMR) is defined as $\mathrm{CMMR} = 20\mathrm{lg}\frac{\mathrm{A_{differential}}}{\mathrm{A_{commonmode}}}$, where A_x is the voltage gain. As simulations showed that the common mode output is not constant over the whole input range, a snap shot of the worst case simulation CMMR for the Small Amplifier has been taken. At the highest slope of the output during a common mode sweep, simulations result in a common mode amplification of $A_{commonmode} = 5.7$ at 2.5 μA bias. The open loop gain at this point is in the magnitude of 3 k, so we get an estimated ratio of $CMMR \approx 50\,\mathrm{dB}$ for low frequencies.



(a) closed loop offset

(b) CMMR of small amplifier

Figure 6.14: Results of the offset measurement

The characteristic of the closed loop offset can be used to calculate the common mode gain created by the deviation of the offset which is the same as a differential input.

Setup

The DC input voltage is generated by the 5 kΩ potentiometer. Measurements will be done with precision multimeters on input and output port. A blocking capacitor of 4.7 μF is used at the input ports.

The error for the output measurement is 0.2 mV caused by fluctuations. The input error is caused by the limits of adjustment, in statistic measurements where one equal voltage is needed several times. The error here is 0.1 mV.

Results and Evaluation

Figure 6.14(a) expresses the offset between input and output for the three different amplifiers of the first chip. The simulation for the big amplifier can be seen as an example as simulations with the other amplifiers lead to similar results which is no surprise as the input stages are similar except a scaling factor, which should not effect the offset in simulations. Nevertheless, there is an offset in measurements. It is observed, that this offset is significantly smaller for the Big Amplifier. The difference to the big amplifier can be explained by the bigger input stage, which gives a better statistic for matching. The systematic shift of the curve in comparison to simulation is most probably caused by parasitic resistors in the setup.

Statistic The statistic can be seen in table 6.3.6. The values for the small amplifier and the NC amplifier have to be treated with care, as the variation is huge. The variation is a quantum for the matching of the input stages. A higher variation means worse matching. The three test points have been chosen, as at 200 mV only the PMOS stage is operating, whilst at 1500 mV only NMOSt stage is on. At 800 mV both stages are operating. The results fit to the expectations in the fact, that the deviation is much smaller for the big amplifier, as its input stage is bigger. In the middle of the operating range, the offset of NMOS and PMOS stage is averaged and so the resulting deviation is smaller. We would expect a worse matching for the NMOS stage, which is not the case in our statistic. An explanation for these results could be significant worse PMOS matching during the whole production run.

	200 mV	Δ	800 mV	Δ	1500 mV	Δ
small	-1	11	-1	8	-1	9
nc	-2	11	-2	9	-3	10
big	-2	5	-1	2	0	3

Table 6.1: offset and standard deviation at different DC input voltages measured with twenty samples

As we used special layout techniques to dimish matching problems via averaging and area minimizing, we would expect better matching results. Looking at the matching report we get a matching deviation of around 4 mV for the threshold voltage of the two NMOS transistors with the same size we used. Averaging should have even lowered this number.

The average value of the measured offset is in the region of one ore two millivolts which is the result, we expect as this is the gain induced offset seen in simulations. A systematic error of one millivolt can easily be included here, created by the ohmic resistance of the input lines.

More samples are necessary to get a better statistic.

CMMR We use the curve of the offset for the small amplifier of the first chip to get an impression of the effect, the offset can have on the common mode rejection ratio. We chose the small amplifier as the measured derivative is the biggest here. All we need to do is to calculate the equivalent differential voltage at one point. We do not need the differential gain at all, as it is truncated in the definition of the CMMR. As formula we get:

$$\mathrm{CMMR} = 20\log\left|\frac{\mathrm{dV_{Offset}}}{\mathrm{dV_{in}}}\right|, \tag{6.1}$$

which is simplified with a differential coefficient. Results can be seen in Figure 6.14(b). With negligence of obvious measurement errors (compare to Figure 6.14(a)), the CMMR is above 80 dB There are three chips with a bigger difference between the NMOS and PMOS stage, which will result in a worse CMMR. For these amplifiers, the worst case CMMR created by mismatch can be estimated to be above 70 dB. The mismatch CMMR is still better than the CMMR caused by the circuit output change of the big opamp in simulations.

Typical CMMRS are 90 dB for the 741 classic opamp or even 120 dB for high end devices [27]. A CMMR of 70 dB is tolerable for uncritical systems. The designed amplifiers are not supposed to work as differential amplifiers, but as voltage buffers in closed loop feedback mode, so the CMMR has no deep impact on the later system.

6.3.7 Slew Rate

The Slew Rate is defined as the maximum time derivation of the output for any input signal. It is measured for the closed loop configuration as this is the purpose of the amplifiers. As input signal, a rail-to-rail square pulse is used, generated by the AWG. The oscilloscope is used to ascertain the derivation. The output is terminated with 50 Ω DC termination for the big amplifier and 1 MΩ for the others.

We get a slew rate of 243 $\frac{\mathrm{mV}}{\mathrm{ns}}$ for the big amplifier, (72±2) $\frac{\mathrm{mV}}{\mathrm{ns}}$ for the small one and (58±2) $\frac{\mathrm{mV}}{\mathrm{ns}}$ for the NC Amplifier. Results fit to the simulation regarding the order of magnitude. Simulation results in $(77 \pm 1)\ \frac{\mathrm{mV}}{\mathrm{ns}}$ for the Nc Amplifier, $(112 \pm 1)\ \frac{\mathrm{mV}}{\mathrm{ns}}$ for the small and $(232 \pm 1)\ \frac{\mathrm{mV}}{\mathrm{ns}}$ for the big one. To get a feeling for these numbers we have a look at the slew rate of industry rail-to-rail amplifiers. The highest Slew Rate for Texas Instruments devices is $975\frac{mV}{ns}$ whilst the smallest offered is 32 $\frac{\mathrm{mV}}{\mathrm{ns}}$.

6.3.8 Open Loop Gain

The Open Loop Gain is very hard to measure, as it is in the magnitude of 1k. We need a nearly noise free input signal with an amplitude of less than 1 mV. The signal amplitude is realized via an attenuator of 1 k and AC coupled via a 47 μF capacitor. Two potentiometers are used to adjust the offset between the negative and positive port. As test frequency we chose 10 kHz as this frequency lies ahead the first pole. The bias current is 3 μA. The big amplifier is measured via a 50 Ω port of the oscilloscope, whilst a probe is used for the small and NC amplifier.

With a DC input input voltage of 550 mV I measured a gain of 237.5 ± 20 for the big amplifier and 2320 ± 20 for the mall and NC amplifier. The difference might seem to be wrong, but the small load of the big amplifier is reducing the gain. The measured gain is even bigger than the simulated one which lies in the magnitude of 150, but the curve is very steep at this point of the characteristic. The measured gain of the small amplifier is in the correct order of magnitude, too.

6.3.9 Noise Measurement

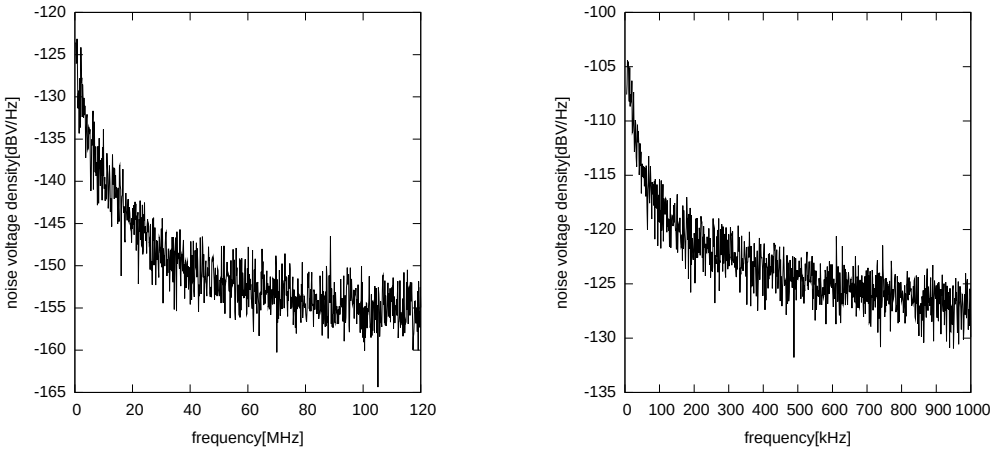
In the intended application of the amplifiers, they will be configured in closed loop with an amplification factor of 1 and therefore noise is negligible. Nevertheless noise is an important characteristic of operational amplifiers and a challenge in design. The designed amplifiers are not optimized for low noise but for low current and small area consumption. This is in contrast to low noise design as the flicker noise of CMOS transistors is proportional to $\frac{1}{WL}$. To get a sketch of the noise behavior, measurements with the big amplifier will be performed as this has the ability to drive small loads, particularly a 50 Ω transmission line.

Setup

As the power of noise should be very small (the power of thermal noise at a room temperature of $20°C$ in on a bandwidth of 1 Hz is $k_b T \approx 4 * 10^{-21}$ W $= -174$ dBm) a very precise measurement is needed.

The spectral analyzer is able to measure a minimum noise power density in the area of $-150 \frac{\text{dBm}}{\text{Hz}}$. If we look at the input impedance of the analyzer which is 50 Ω this does not look that bad. A 50 Ω resistor has a thermal noise voltage spectral density at room temperature of $\sqrt{4k_B TR} \approx 0.9 \frac{\text{nV}}{\sqrt{\text{Hz}}}$, which is the equivalent noise bandwidth the analyzer is normalized to. Regarding this noise voltage density we get a spectral power density of $4k_B TR = 0.2 \frac{\text{aW}}{\text{Hz}}$ or -156 dBm/Hz. Direct measurements of power and voltage density have demonstrated that the relation between power and voltage density is calculated this way internally.

As device under test, the amplifier outopen2 is chosen, because this one has an independent output line without any SMD jumper connectors which could work as antennas. To calculate the equivalent input noise density, a specific amplification is needed which is realized via negative feedback with a 750 Ω resistor and a ground resistor of 51 Ω at the negative input. These resistors build a voltage divider. As the negative input of the amplifier should be driven to the same as the positive input V_{in+}, the output voltage is $V_{out} = V_{in} * \frac{800\,\Omega}{51\,\Omega} = V_{in} * 15.71$. As comparison see Figure 4.1. The accuracy of the used resistors is 1 %, so we get $A_V = 15.7 \pm 0.2$ as final result for the voltage amplification. A more accurate amplification will be measured. A DC voltage is put on the positive input and blocked by a 100 nF ceramic capacitor. A 50 Ω which used to be the ac-termination in other experiments is left as its noise is negligible in comparison to that of the feedback circuitry. It is possible to disconnect the voltage source during measurements as the parasitic input resistance is small enough to allow the capacitor to hold the voltage. The spectral noise density of the feedback resistors and the output resistor is $\sqrt{4k_B T 850\,\Omega} \approx 13 \frac{\text{nV}}{\sqrt{\text{Hz}}}$.



(a) measured with $10kHz$ filter bandwidth (b) measured with $300Hz$ filter bandwidth

Figure 6.15: noise figure for measurement with feedback resistors

The maximum output of the amplifier lies in the area of 1.79 V if no DC current has to be sourced, which is nearly the case in here because the used port of the spectral analyzer works with AC coupling and the feedback resistors are relatively high. To be in the middle of the output operating range the DC input voltage should not lay above 50 mV. With this voltage we are still far enough from the rails to allow normal operation. This is why the amplification of 16 has been chosen.

To get a rough overview on the amplification and to have an alternative for the noise characterization, an AC signal from the arbitrary wave form generator (AWG) will be driven to the positive input with an offset of 50 mV, and an amplitude of 12.5 mW. The DC source used above will be used to generate the ground for the termination via the previous unnecessary termination resistance. This method is not very accurate as the digitally generated output signal of the AWG is very noisy and partly not well defined as the amplitude varies in a range of more than 1 mV.

Finally to get a comparison without feedback resistors, the amplifier is switched to open mode. The DC input offset is given via a capacitor to reduce noise from voltage dividers. The capacitor is loaded and than the voltage source will be switched off.

Measurement

To get a first check if the feedback is working, an input Voltage of $V_{in+} = (49.9 \pm 0.1)\,\mathrm{mV}$ is applied. The bias voltage is adjusted to $V_B = 450\,\mathrm{mV}$ which corresponts to a biasing current of $5\,\mu\mathrm{A}$ at the concerning bias multiplier circuit as measured above. Without any load the negative output is measured to be $V_- = 49.2\,\mathrm{mV}$ so there is an offset of 0.7 mV which fits to the measurements in the buffer mode.

When the analyzer is connected to the output, the DC output voltage stays constant as the load of the AC coupled analyzer is only capacitive. The measured output voltage is $(751 \pm 0.5)\,\mathrm{mV}$. Figure 6.15(a) shows the spectral density of the amplifier output. A first qualitative analysis predicts that the boost of noise for low frequencies might be caused by $\frac{1}{f}$-noise of the amplifier input stage. The small peak at 88.86 Mhz is a local radio station which transmits its signal from a radio tower in 5 km distance to the measurement with a power of 100 kW. A zoom to lower frequencies with a smaller measurement bandwidth can be seen in Figure 6.15(b). The difference at lower frequencies is caused by the different measurement bandwidths.

In the next step, the dependency of the noise and the bias current is examined. For this

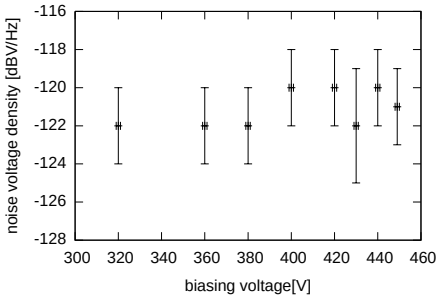


Figure 6.16: noise voltage density at 200 kHz versus biasing voltage. No dependency can be observed in the measuring range.

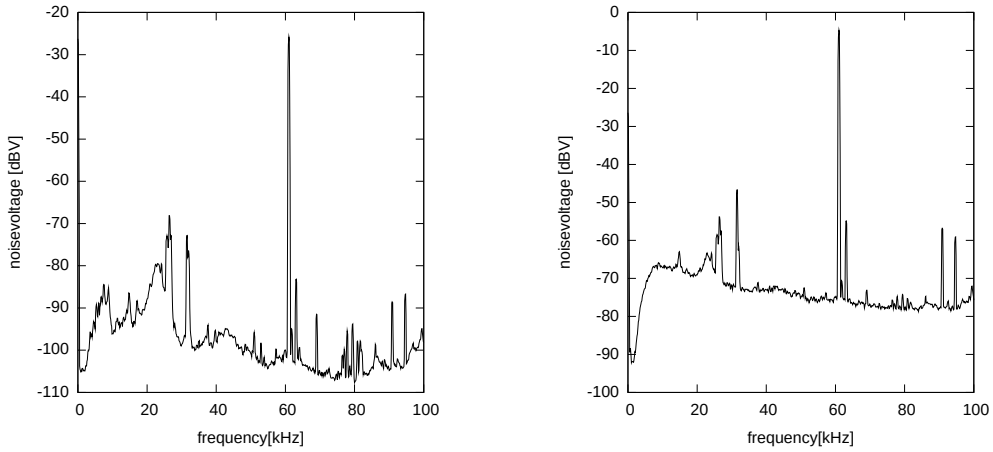
purpose we look at a certain frequency, namely 200 kHz and zoom into a display bandwidth of 2 kHz If we now change the measuring bandwidth to 10 Hz to get a reasonable resolution the output is flattering with an amplitude of 12 dBV or more. A change in the video bandwidth to 1 Hz solves this problem as now the output is an average over some measurements. Figure 6.16 visualizes the result. In the measured range, there is no dependence between noise and biasing voltage.

Now, the noise of different chips is compared. As bias reference, the generator is adjusted to a biasing voltage of 450 mV for the first chip. This voltage will be different for other chips as there are process variations in the biasing circuit of the chip and the bias generator on board keeps the current constant. For the three observed chips, the noise remains constant at a level of $(121 \pm 2)\,\frac{\mathrm{dBV}}{\mathrm{Hz}}$

Finally, the signal to noise ratio of the input signal and that of the output signal is measured with a signal frequency of 61 kHz. As predicted above, our signal source, the AWG is not very accurate. With an amplitude of 50 mV adjusted at the AWG we get an amplitude of (50 ± 3) mV at the oscilloscope and $36.18 * \sqrt{2} \approx 51.2$ mV at a peak voltage frequency of 61.8 MHz. As we do not know the source of this difference and we realize the error in the amplitude at the oscilloscope we see, that a setup with the AWG can only give a rough overview in noise measurement. Figure 6.17 shows the spectrum of the input and output signal. Board induced noise is not included in the input signal as it is measured via a direct connection between AWG and the spectrum analyzer. For the input, the signal to noise ratio (SNR) is roughly (73 ± 2) dB while it is (68 ± 2) dB leaving a noise factor of (5 ± 3) dB or 3 ± 3 respectively. The peaks can be used to get an impression of the gain. The input peak has a power of 25.8 dBm and the output of 4.7 dBm, therefore the voltage amplification is 10.5 dB or 11.2 respectively, which is much too small concerning our feedback.

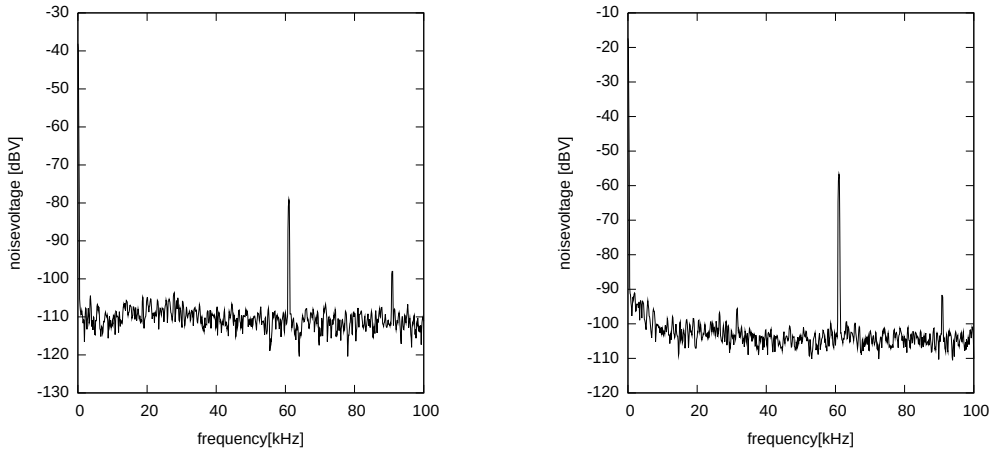
To get a control reference for the AC gain, the input and the output are measured on the oscilloscope. Unfortunately we have to switch to a lower input offset, as the DC-termination of the oscilloscope limits the output swing. The input of the amplifier has to be switched to DC-termination again. For an offset voltage of 30 mV we get an input amplitude of (12.7 ± 2) mV and an output of (185 ± 10) mV.

An idea to measure the AC gain with the spectral analyzer without switching the input signal between the amplifier and the analyzer terminal is to use an active probe with an high ohmic impedance and a small capacitance. To eliminate an error that may have been caused the small amplitude of the AWG output signal a 30 dB attenuator is used at the input. Figure 6.18 shows the input and output spectrum measured with a probe. The signal to noise ratio of the input is worse than that of the output and worse than that of previous measurements. The reason is the 10 dB attenuation of the probe as it damps the signal whilst noise from the analyzer stays the same. A short calculation of the gain gives us $79.1\,\mathrm{dBV} - 56.7\,\mathrm{dBV} = 22.4\,\mathrm{dBV} = 11.2\,\mathrm{dB}$



(a) input signal (b) output signal

Figure 6.17: spectrum to compare singal to noise ratios

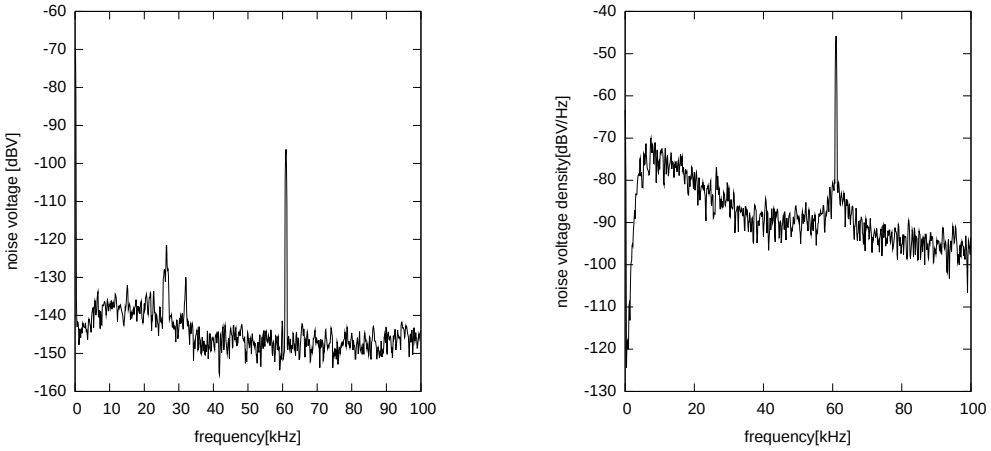


(a) input signal (b) output signal

Figure 6.18: Spectra measuered with an active probe

or 13.2 which is still less than the adjusted.

The noise measurement in open mode is a little bit tricky as a certain offset between both ports is needed for maximum gain and leak currents vary the offset voltage. Another problem is that we cannot measure the offset voltage once the source is switched off, as the intern resistor of our multimeter would drain the charge lowering the voltage. As it is observed, that the leak current at the positive input is much bigger and the voltage is decreasing faster despite the bigger capacitor. We load the negative port first and switch to the positive port then. Now the source is shunt leaving the system in a state, where the negative input voltage is a little bit below the positive one. As the positive port voltage is falling faster, the difference between both ports is changing. On the spectrum analyzer we observe an amplitude which is rising up to a certain level of maximum gain and falling again. A snap shot is taken at maximum gain which can be seen in Figure 6.19. Input and output signal have been measured without probe because its internal noise may destroy the accuracy.



(a) input signal (b) output signal

Figure 6.19: open loop noise figure

Evaluation

Gain Looking at the measured output voltage we get a gain of $A_{V_{DC}} = \frac{V_{in}}{V_{out}} = \frac{751}{49.9} = 15.05 \pm 0.03$. The error is caused by the inaccuracy at the input. The gain is not in the error margin of the adjusted gain. If we look at the offset between the negative and the positive input, this is nothing surprising, but as $\frac{751}{49,2}$ is still 15.264 and not within the error range, there has to be another problem as the simple voltage divider between V_{out}, V_{in-} and ground should be working definitely. Measuring the ground potential gives us the solution. The ground at the $51\,\Omega$ Resistor is 1 mV higher than that of our gauge. As $\frac{750}{48,2} = 15.56$ we finally have a result within the deviations of the adjusted amplification.

For the AC case, we use the results from the measurement of the sinus signal with the oscilloscope to calculate the gain. We get $A_{V_{61\,\mathrm{kHz}}} = \frac{185 \pm 10}{12.7 \pm 2} = 15 \pm 3$ so again the result is within the error margin, but the huge error relativizes this. The amplification of 11.2 measured with the spectral analyzer without probe is wrong as the AC termination of the analyzer and that of the setup are not equal and so we do not know the correct amplitude of the signal at the input. The measurement with the probe gives us a better result. For further calculations we will use the voltage amplification measured in the DC case as there are the least error sources included.

Noise with Feedback If we correct the noise output spectrum of Figure 6.15(b) with the port intern attenuation of 6 dB and divide by the determined gain we get to the equivalent input noise voltage of our amplifier. A plot in linear scale is shown in Figure 6.20. Noise created by the $50\,\Omega$ circuitry at the output can be neglected because of the amplification. After the strong noise at the beginning of the spectrum it falls down to approximately $18\,\frac{\mathrm{nV}}{\sqrt{\mathrm{Hz}}}$ which is not too bad as the classical uA741 has an equivalent input noise density of $20\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$[22] and there is still $4\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ included from our feedback circuit. Of course, our amplifier cannot win against low noise amplifiers reaching input noise levels of $0.9\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$, but those are not designed to drive $50\,\Omega$ lines and consume more power.

The boost at the beginning of the spectrum is clearly $\frac{1}{f}$-noise. The problem with $\frac{1}{f}$-noise is that there is no clear theory behind it which can pass on empirical constances. One theory is that defects and impurities randomly trap charge at the surface. The gate capacitor is said to flatten the noise[22], but a capacitor would lower the noise only down to a certain frequency. It lowers the frequency the $\frac{1}{f}$-noise of the device would be significant in a measurement. The empirical formula is

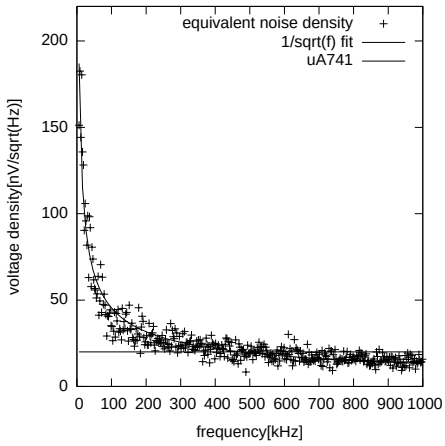


Figure 6.20: equivalent input noise density calculated with the results from noise experiments with feedback.

$$\bar{u_n^2} = \frac{K}{f} * \frac{1}{WLC_{OX}^2} * \Delta f. \tag{6.2}$$

The formula is usually given for the noise current which is connected to $\bar{u_n^2}$ via the transconductance g_m of the device. K is a process dependent constant which is typically in the magnitude of $10^{(-28)}$ for PMOS devices, whereas it is 50 times lager for NMOS[22, 23]. With a channel area of 14.4 μm^2 we get a capacitor of $(118 \pm 7)\ fF$ for one of our two PMOS input transistors. For the total noise density we get $\bar{u_n^2} = \frac{1}{f} * (1.0 \pm 0.1) * 10^{-13}\,\mathrm{V}^2$, or $\sqrt{\bar{u_n^2}} = (320 \pm 100)\,\mathrm{nV}\sqrt{\mathrm{f}}$ For both input transistors, we get $(450 \pm 150)\,\mathrm{nV}\sqrt{\mathrm{f}}$

Looking at our plots, we realize, that our actually measured noise is about a factor $\sqrt{1000}$ bigger than the predicted. It is no surprise, that it does not fit, as the sample value of K is from 1996, but a factor of 1000 is bigger than expected. Fitting the curve $f(x) = a + b/\sqrt{f}$ on the results leads to $a = 14909.9 \pm 1\%$ and $b = -1.66205 \pm 25\%$ with a correlation coefficient of 0.8. The fit is only useful to verify that the curve falls with $1/\sqrt{f}$.

The standard BSIM 3.3v2 model used in the simulations does not include flicker noise, but special RF-transistors with a fixed layout use an extended version of BSIM 3.3v2 with the switch $noimod = 2$ realizing flicker noise. To visualize the noise, Figure 6.21 shows the noise density at the drain of a PMOS in common source connection with a load of $150\,\Omega$ and $v_{supply} = 1.8\,\mathrm{V}$ whilst the gate is plugged to ground. The $\frac{W}{L}$ is 11.6 and the channel length is the same as in our amplifier. Comparing with the measured noise, we see that our measurement is in the correct order of magnitude.

Open Loop Noise For the open loop noise, we use the same plot(input noise in Figure 6.22) for noise and gain measurement. The difference between the peaks is $-20.2\,\mathrm{dB} + 70.0\,\mathrm{dB} = 49.8\,\mathrm{dB}$ power gain which is 24.9 dB or 309 voltage gain. We remember that this way of gain measurement is not accurate; nevertheless it fits to the measurements done before. The noise figure from Figure 6.19 is converted to the noise voltage density again and the gain is used to get the equivalent noise density. Figure 6.22 shows the result together with a fit to $1/\sqrt{f}$ and the input noise density. The output signal for small frequencies below 10 kHz is a measuring error caused by the AC termination of the analyzer. The peaks left of the signal in the input are not in the output because the signals have been measured separately and the input impedance of the analyzer and our system are different. Besides, the correlation between the input noise

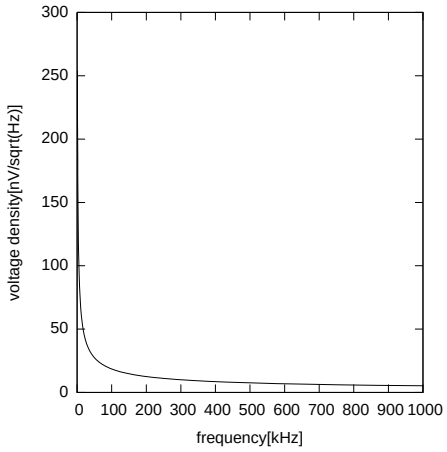


Figure 6.21: noise voltage density for an RF_PMOS-device with W = 35 μm and L = 300 nm.

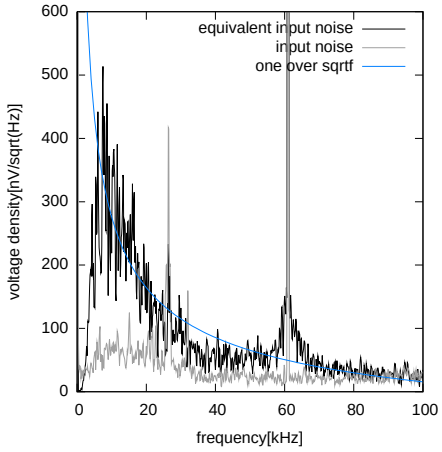


Figure 6.22: equivalent noise voltage density for the open loop connection

density and the equivalent input noise density can be seen very clearly The fit has been done for frequencies between 10 and 50 MHz and shows the $\frac{1}{\sqrt{f}}$ behavior.

Conclusion

Recapitulating we see that noise is no problem for our amplifier. The behavior is not extraordinary good or bad, but average. An amplifier with less noise could be produced using bigger input transistors.

6.3.10 Frequency Response

The measured frequency response will be discussed qualitatively for all three amplifiers. Measurements are done using the network analyzer which is a little bit problematic in open mode, as the first node lays ahead the lowest measured frequency.

Closed Loop

Big amplifier The results can be seen in Figure 6.23. Two different equivalent biasing voltages and a power off plot are displayed. The peaking is actually smaller than it looks like here as the

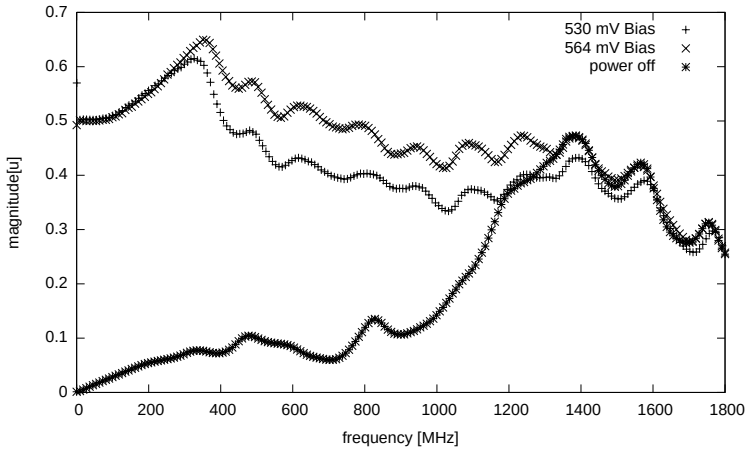


Figure 6.23: frequency response of the big amplifier in closed loop configuration

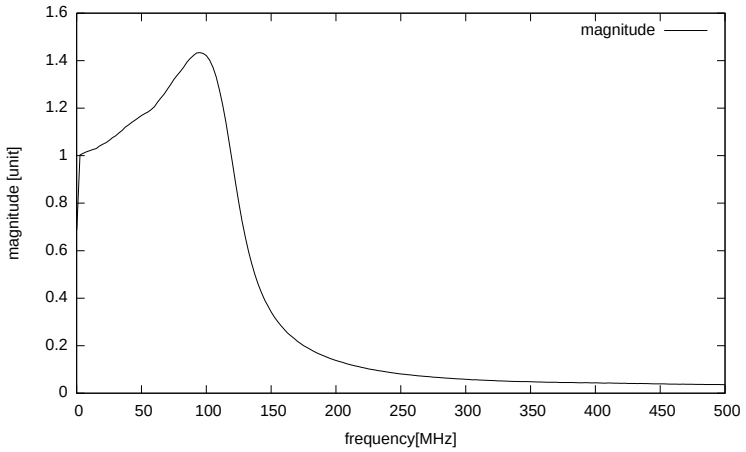


Figure 6.24: frequency response of the small amplifier in closed loop configuration

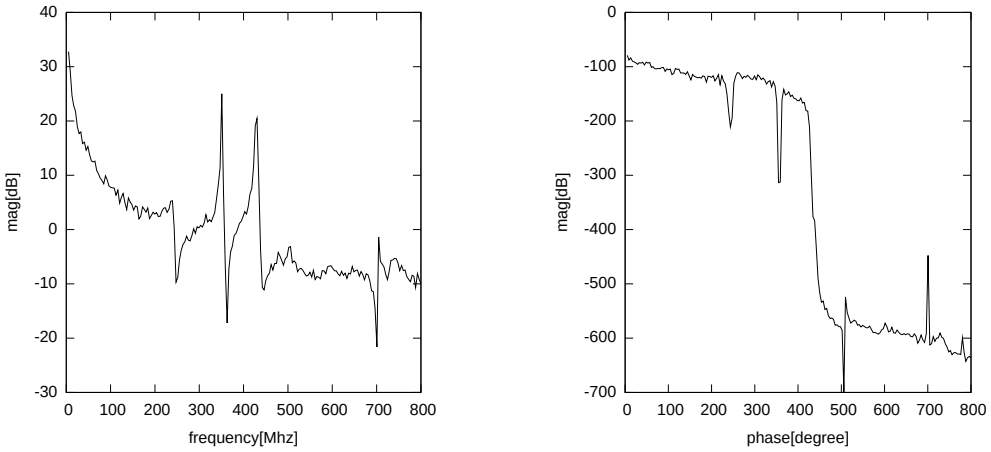
power of the signal is measured, which is proportional to the square of the voltage. The output is 0.5 because of the 50 Ω voltage divider at the output. At higher frequencies we see crosstalk, as input and output line are next to each other. The bumps are caused by resonances in the setup.

Small Amplifier Here, the signal has to be measured with a probe, as the maximum output current is too small. Bumps and crosstalk are gone as can be seen in 6.24, leading to the suspicion that they might have been caused by bad adjustment of the output transmission line. The phase response is not displayed, but there is a change from zero to $-180°$ beginning at the -3 dB frequency just as expected.

Open Loop

As bias current, 3 μA have been chosen, whilst the DC offset voltage is 900 mV.

Big Amplifier Figure 6.25(a) shows the phase and magnitude frequency response of the big amplifier. The big peaks in both plots are errors in measurement as the input signal had to be very small no to excess the maximum input power of the analyzer. The input power had to be reduced to -35 dBm and the output at 0 dB magnitude is just the same. 35 dBm equals a



(a) magnitude (b) phase

Figure 6.25: open loop frequency response of the big amplifier

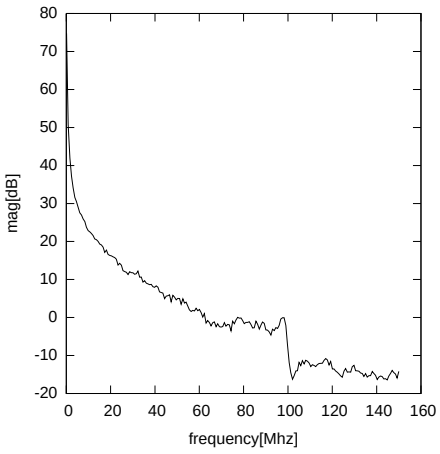


Figure 6.26: frequency response of the small amplifier

voltage amplitude of circa 100 μV at 50 Ω. A systematic error of 360 degree is included in the second half of the phase shift diagram. Except for these errors, the figures are exemplary. The zero intersection of the magnitude is in the rage of 350 MHz which is just where it should be. After the first pole, the phase shift is 90° whilst it is more than 180° after the second. Inasmuch as anything can be said about the phase margin considering these results, it can be estimated to 60°.

Small Amplifier The magnitude frequency response can be viewed in Figure 6.26. In difference to the response of the big amplifier, the magnitude covers 0 dB before the second pole. This is a result of the narrow optimization of compensation capacity and current consumption. In closed loop operation, the amplification is still one at these frequencies which is just how it should be. Higher gain is only needed at lower frequency to compensate the DC offset between input and common mode output.

Chapter 7

Conclusion

In this thesis, integrated operational amplifiers for a large scale neuromorphic system have been designed, laid out, prototyped and measured. As I am a physicist, I had only little experience in the design of operational amplifiers and VLSI circuits before this thesis. Operational amplifier circuits are well known and can get arbitrarily complicated. Lots of explorations have been done and are done.

My personal goal was to completely understand how the circuits work. My design approach was to calculate a circuit and then get verification by simulations. This has been possible for smaller designs but for more complex schematics, matching between theory and simulation was only qualitative, which is still sufficient as parameters can be changed during simulations. The handicap is the simplified hand model used in calculations, and the right parameter extraction.

Recapitulating, theoretical background needed for the design has been presented compactly in chapter 4. The knowledge has been used to design the amplifiers in the next chapter after some considerations about the switch between hand-model and the "real world" have been done. Starting with a simple CMOS OTA, the circuit has been built up enlarging the complexity. The resulting amplifier is very compact, merging input and output stage. The final topology is similar to amplifiers designed by Ron Hogervorst twelve years ago[14, 15, 16] which gives some validation as the design is established on an older process at lower frequencies with higher supply voltage. After the schematics of both amplifiers were done and verified, the layout has been developed, using common centroid layout to improve matching.

To allow experimental verification, the amplifiers have been prototyped on an ASIC and a small setup PCB has been designed. The principle functionality of the amplifiers is evident as measured. Noise has been observed in detail resulting in an average behaviour directly revealing 1/f noise. g_m-compensation, class AB current control and compensation work well. A small problem has been unfolded in the offset and the CMMR of the Small Amplifier. The matching here is worse than expected.

Generally, simulations differ a little bit from measurements, which can mainly be explained by process variations and matching issues. A qualitative matching between simulations and measurements could be found in nearly all measurements. The BSIM 3.3 v2 model seems to be accurate enough for parameters where matching is no problem - simulation should be sufficient in this case. Prototyping is needed, when matching or noise can be an issue.

The designed amplifiers are ready to be used in the HICANN prototype and will massively reduce the current consumption needed for buffering compared to FACETS stage 1. The offset problem of the Small Amplifier can be compensated by setting the input accordingly. Furthermore there is a random offset at the input produced by floating gate variations anyway. Matching of the cascode branch of the Small Amplifier could be improved by setting the Class AB control transistors of both branches on the same bulk potential to allow a closer positioning in layout. The output driver could be enhanced by enable transistors, allowing a complete power

off of the output stage[6] with a high ohmic resistance.

A first prototype of the HICANN will be submitted midyear 2008. Design studies showed the realizability of the wafer scale system so far - this will have to be proven during mechanical and post processing prototyping[10]. When the whole system is running finally, it will help humanity to get a little step further in disclosing the secrets of the human brain.

Bibliography

[1] *FACETS - Fast analog Computing with emergen Transitstates*, www.facets-project.org

[2] Meier, K. *FACETS Project Presentation*, www.facets-project.org

[3] *Blue Brain Project*, http://bluebrain.epfl.ch/

[4] Johannes Schemmel, Andreas Gruebl, Karlheinz Meier and Eilif Mueller, *Implementing Synaptic Plasticity in a VLSI Spiking Neural Network Model*, Proc of IJCNN 2006, 1-6, IEEE Press

[5] Daniel Bruederle, Andreas Gruebl, Karlheinz Meier, Eilif Mueller and Johannes Schemmel, *A Software Framework for Tuning the Dynamics of Neuromorphic Silicon Towards Biology*, Proc of IWANN 2007, Springer LNCS 4507, pp 479 - 486

[6] Conversation with Johannes Schemmel

[7] Richard P. FeynmN, Robert B. Leighton, Matthew Sands, *The Feynman Lectures on Physics - Quantum Mechanics*, California Institute of Technology, Addison-Wesley

[8] Andreas Grübl, *VLSI Implementation of a Spiking Neural Network*, Dissertation, Kirchhoff-Institut fr Physik Heidelberg

[9] Andreas Grübl, *Eine FPGA-basierte Plattform fr neuronale Netze*, Diploma Thesis, Kirchhoff-Institut fr Physik Heidelberg

[10] Dan Husmann de Oliveira, *A prototype system to test the reliability of connections between board and wafer via elastomeric contacs*, FACETS intern study

[11] Rolf Rannacher *Einführung in die numerische Mathematik* Vorlesungsskriptum SS 05, IWR Universitt Heidelberg (http://numerik.iwr.uni-heidelberg.de/)

[12] Kenneth S. Kundert, Olaf Zinke, *The Designer's Guide to Verilog AMS*, 2004 Springer Science+Business Media, Inc.

[13] H. S. Black, *Stabilized Feedback Amplifier*, Electrical Engerneering, vo. 53 pp 114-120, Jan 1934

[14] Ron Hogervorst, Remco J. wiegerink, Peter A. L. De Jong, Jeroen Fonderie, Roelof F. Wassenaar, Johan H. Huising, *CMOS low-voltage operational amplifiers with constant-gm rail-to-rail input stage*, Analog Integrated Circuits and Signal Processing, pp. 135 - 146,1994-03-01

[15] Ron Hogervorst, John P. Tero, and Johan H. Huijsing, *Compact CMOS constant-g_m Rail-to-Rail Input Stage with g_m-Control b an Electronic Zener Diode*, IEEE Journal of Solid-State Circuits, vol 31. No. 7, July 1996

[16] Ron Hogervorst, John P. Tero, Ruud G. H. Eschauzier, Johan H. Huijsing, *A Compact Power-Efficient 3 V CMOS Rail-to-Rail Inoutput/Output Operational Amplifier for VLSI Cell Libaries*, Solid-State Circuits Conference, 1994. Digest of Technical Papers. 41st ISSCC., 1994 IEEE International

[17] Willi M. C. Sansen, *Alanlog design essentials* Springerverlag 2006

[18] Johan H. Huijsing, Daniel Linebarger, *Low-Voltage Operational Amplifier with Rail-to-Rail Input and Output Ranges*, IEEE Journal of solid-sate circuits, Vol. sc-20, NO. 6. December 1985

[19] John R. Ragazzini, Robert H. Randall, Frederick A. Russel, *Analysis of Problems in Dynamics by Electronic Circuits*, Proceedings of the I.R.E., may 1947

[20] Tony van Roon, *741 Op-Amp Tutorial*, http://www.uoguelph.ca/˜antoon/gadgets/741/741.html

[21] Yutaka Yukizaki, Haruo Kobayashi, Takao Myono, Tasuya Suzuki and Nana Zhao, *Low-voltage Rail-to-Rail CMOS Operational Amplifier Design*, Electronics and Communications in Japan, Part 2 Vol. 89, No. 12, 2006

[22] Thomas H. Lee, *The Design of CMOS Radio-Frequency Intergrated Circuits, Second Edition*, Cambridge University Press 2004

[23] Kenneth R. Laker, Willy M. C. Sansen, *Desing of Analog Integrated Circuits and Systems*, McGraw-Hill 1994

[24] André Srowig, Jan-Peter Loock, Karlheinz Meier, Johannes Schemmel et al, *Analog Floating Gate Memory in a 0.180 μm Single-Poly CMOS Process*

[25] Hastings, Alan ,*The art of analog layout*, Prentice-Hall, 2001

[26] http://www-device.eecs.berkeley.edu/ bsim3/ ,*BSIM Homepage - official site*

[27] www.wikipedia.com

[28] T. Quarles, D. Pederson, R. Newton, A. Sangiovanni-Vincentelli, and Christopher Wayne, *The Spice Page*, http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE/

[29] *eCircuit Center*, www.ecircuitcenter.com

[30] Babara Paech, Jürgen Rückert, Software Engineering I, Lecture SS 2006, University of Heidelberg

[31] Randall L. Geiger, Philipp E. Allen, Noel R. Strader, *VLSI DESIGN TECHNIQUES FOR ANALOG AND DIGITAL CIRCUITS*, McGraw-Hill 1990

[32] National Semiconductor,*LM1085 - 3A Low Dropout Positive Regulators*

Acknowledgements

Finally I want to take this opportunity to express my gratitude to everyone who supported me in accomplishing this thesis.

- Prof. Karlheinz Meier for the friendly admission in the Electronic Vision(s) group, the launching of the FACETS project and the supervision of this thesis.
- Prof. Volker Lindenstruth for the second opinion
- Johannes Schemmel for the group management, lots of motivating conversations,
- Andreas Grübl, André Srwowig, Daniel Bruederle, Bernhard Kaplan, Johannes Bill and Eric Müller for proof-reading parts of this thesis
- Andreas Grübl who helped when ever new tools where needed to be understood and especially for the "+"-Button in Virtuoso
- Johannes Fieres and Daniel Bruederle for admitting me as a "Hardboy" in the "Softy Room"
- Tillmann Schmitz for giving me some introduction to VERILOG AMS at the begining of my thesis
- Dan Husmann for supportive conservations
- Ralf Achenbach for giving support with measurement instruments
- Stefan Philipp and Prof. Karlheinz Meier for activating my interest in neuromorphic engineering and hardware.
- all other members of the Electronic Vision(s) group.

Furthermore I would like to thank

- My parents for enabling me to study
- My friends especially Anna and Stefan
- My WG for listening to me before I left home and after I returned from work.
- My Scouting Clan esp. Phillip Gottschalk who alone took the lead of 16 Jungspfadfinder the last four weeks which is even hard for two, and Stefan who tolerated my abstinence in StaVo work the last few months.

Erklärung:

Ich versichere, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 20.12.2007

.....
(Unterschrift)