



NTNU – Trondheim
Norwegian University of
Science and Technology

Design of Energy Efficient Low Noise Amplifiers in 40 nm bulk CMOS and 28 nm FD-SOI for Intravascular Ultrasound Imaging

Endre Rørlien Bjørsvik

Electronics System Design and Innovation

Submission date: July 2014

Supervisor: Trond Ytterdal, IET

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Problem Description

Continuous downscaling of CMOS technologies provides new challenges and opportunities for energy efficient analog circuits. The main goal for this thesis is to design energy efficient LNAs in nanoscale CMOS technologies for application in medical ultrasound imaging. These amplifiers should be used to sense signals from capacitive micromachined ultrasound transducers. A complete design attempt with specification, transistor schematic, layout and verification should be performed. Ultrasound specific features as single-ended to differential conversion and variable gain should also be implemented. Activities in the project include:

- Based on a specification, improve and expand earlier attempts on intravascular LNAs in 40 nm bulk CMOS.
- Design a similar LNA in 28 nm CMOS.
- Verify their performance.
- Compare the two technologies with each other and with previous attempts.

Assignment given: 17. January 2014. Supervisor: Trond Ytterdal, IET.

Abstract

Intravascular ultrasound imaging has during the last decades become an important tool for diagnosis and treatment of coronary diseases. A transition to three-dimensional imaging would drastically improve image quality of this technique, but this transition also comes with several complexity challenges. One way to meet these challenges is by integrating more functionality into the ultrasound catheter. This large scale integration requires a new generation of ultrasound circuitry with tiny power consumption and area footprint in order to fit everything into a tiny catheter. An important part of this circuitry is the analog front-end, which often limit performance in signal processing systems. The trade off between power consumption, speed and noise performance make front-end design a challenging task. In addition should capacitive micromachined ultrasound transducers be interfaced in an optimal way.

This thesis presents design of low noise amplifiers in two nanoscale CMOS technologies, aimed for use in integrated three-dimensional intravascular ultrasound systems at 5 MHz. Nanoscale technologies has tempting properties in the digital domain, but poses some challenges in analog applications. Focus has been put on designing robust and energy efficient amplifiers with small footprints and good enough performance. By using g_m/I_d as an optimization tool, a minimum power consumption was achieved. This resulted in a power consumption of $20.5 \mu\text{W}$, signal-to-noise-ratio of 40.7 dB and noise figure of 10.5 dB for a 40 nm bulk CMOS technology. A 28 nm FD-SOI technology achieved a power consumption of $16.0 \mu\text{W}$, signal-to-noise-ratio of 39.1 dB and noise figure of 9.26 dB. These amplifiers were fitted into layout areas of 88 and $150 (\mu\text{m})^2$ respectively, and parasitic extraction was performed to verify performance.

A key in this solution was to use a common-gate-common-source amplifier with feedback biasing. Feedback biasing was made possible through use of compact high resistance subthreshold diodes and area efficient moscap capacitors. This resulted in amplifiers with both small variability and small footprint. Variable gain was implemented in order to enable time-gain-compensation that handles the large dynamic range from ultrasound echoes. Single-ended to differential conversion in the amplifier enabled use of a differential analog-to-digital converter behind the amplifier, which is also an important part of the front-end.

Sammendrag

Intravaskulær ultralydabildning har i løpet av de siste tiårene blitt et viktig verktøy for å oppdage og behandle karsykdommer. En overgang til tredimensjonal avbildning vil føre til en markant forbedring av bildekvaliteten på denne teknikken, men denne overgangen innebærer også mange kompleksitetsutfordringer. En måte å løse disse utfordringene på er ved å integrere mer funksjonalitet i selve ultralydkateteret. Slik storskala integrasjon krever en ny generasjon av veldig små ultralydkretser med lavt strømforbruk hvis det skal være mulig å få plass til alt i et tynt kateter. En viktig del av disse kretsene er det analoge grensesnittet som ofte begrenser ytelsen i slike signalprosesseringsystemer. Avveiningen mellom strømforbruk, båndbredde og støyegenskaper gjør design av dette til en utfordrende oppgave. I tillegg må grensesnittet tilpasses de kapasitive ultralydtranduserene som brukes.

Denne avhandlingen presenterer utforming og simulering av lavstøyforsterkere i to CMOS teknologier med nanostørrelse transistorer. Disse forsterkerene er beregnet på bruk i et integrert ultralydsystem som gjør tredimensjonal intravaskulær avbildning på 5 MHz. Teknologier med nanostørrelse har fristende egenskaper i det digitale domenet, men innfører samtidig en del analoge utfordringer. Fokus i oppgaven har vært å utforme robuste og energieffektive forsterkere med liten størrelse og god nok ytelse. Ved å bruke g_m/I_d som et optimaliseringsverktøy var det mulig å oppnå minimalt strømforbruk. Dette resulterte i et strømforbruk på $20.5 \mu\text{W}$, signal-til-støy-forhold på 40.7 dB og støyfigur på 10.5 dB for en 40 nm bulk CMOS-teknologi. En 28 nm FD-SOI teknologi oppnådde et strømforbruk på $16.0 \mu\text{W}$, signal-til-støy-forhold på 39.1 dB og støyfigur på 9.26 dB. Disse forsterkerne oppnådde en fysisk størrelse på henholdsvis 88 og $150 (\mu\text{m})^2$, og ekstraksjon av parasitter fra utlegg ble brukt for å bekrefte de øvrige resultatene.

Kjernen i denne løsningen var å bruke en common-gate-common-source forsterker med tilbakekoblingsbiasering. Tilbakekoblingsbiasering ble muliggjort gjennom bruk av kompakte og høyohmige subterskel-dioder og arealeffektive moscap kondensatorer. Dette resulterte i forsterkere med liten variasjon som i tillegg brukte lite areal. Variabel forsterkning ble implementert for å muliggjøre tid-forsterkning-kompensasjon som håndterer det store dynamiske området fra typiske ultralyd ekko. Single-ended til differensiell konvertering i forsterkeren muliggjorde bruk av en differensiell analog-til-digital omformer bak forsterkeren.

Table of Contents

Problem Description

Abstract	i
Sammendrag	ii
Table of Contents	iii
List of Tables	v
List of Figures	vii
Abbreviations	xi
1 Introduction	1
1.1 Main contributions	2
1.2 Thesis outline	2
2 Background Theory	5
2.1 Intravascular ultrasound imaging	5
2.1.1 Imaging system	5
2.1.2 Ultrasound transducers	7
2.1.3 Ultrasound front-end electronics	8
2.1.4 Ultrasound digital post-processing	9
2.2 Amplifier design	9
2.2.1 Amplifier properties	9
2.2.2 Low noise amplifier	13
2.2.3 Amplifier topologies	13

2.2.4	Current mirror biasing	15
2.2.5	Feedback biasing	17
2.2.6	Noise analysis	20
2.2.7	Differential amplifiers	20
2.2.8	Frequency response	23
2.2.9	Variable gain	24
2.3	CMOS Technology	26
2.3.1	Transistor properties	26
2.3.2	Subthreshold operation	27
2.3.3	Transistor noise	28
2.3.4	Nanoscale analog design	29
2.3.5	Bulk and SOI CMOS processes	31
2.4	Figure-of-merit-based design	32
2.5	Computer models	33
2.6	Existing work	34
3	Design Methodology	35
3.1	Specifications	35
3.2	Amplifier topology	36
3.2.1	Biasing	38
3.2.2	Variable gain	39
3.3	Transistor sizing	41
3.4	Layout	45
4	Results	49
4.1	Frequency response	51
4.2	Noise simulation	52
4.3	Large-signal behavior	54
4.4	Transient behavior	56
4.5	Monte Carlo variations	58
5	Discussion	59
5.1	Technology differences	59
5.2	Noise performance	60
5.3	Frequency response	62
5.4	Linearity	63
5.5	Energy efficiency	64
5.6	Variable gain	65
5.7	Feedback biasing	66
5.8	Matching	67

6 Conclusion	69
6.1 Future work	70
Bibliography	71
A Additional analytic expressions	77
A.1 Feedback biased common-source amplifier	77
A.2 Feedback biased common-gate amplifier	77
A.3 Feedback biased common-gate-common-source amplifier	78
B Calculations	81
B.1 Input SNR	81
C Test bench	83
D Additional simulation plots	85
D.1 Input referred noise	85
D.2 Small-signal frequency response	87

List of Tables

3.1	LNA specifications.	37
3.2	40 nm transistor values and dimensions.	43
3.3	28 nm transistor values and dimensions.	44
4.1	Test bench parameters.	50
4.2	Final results for both amplifiers.	50
4.3	Summary of significant spot noise sources at f_c	53
4.4	Mean and standard deviation values from 200 monte carlo mismatch simulations	58
5.1	Energy efficiency for three different amplifiers	65
C.1	Component values for CMUT model in figure 2.4b.	83

List of Figures

2.1	Different transducer array arrangements.	6
2.2	Digitized ultrasound receiver signal chain	7
2.3	Illustration of CMUT structure on ASIC	8
2.5	Typical amplifier characteristic.	10
2.6	Typical amplifier frequency response	12
2.7	Test setup for input impedance.	12
2.8	Common-source stage	14
2.9	Common-gate stage	15
2.10	Common-source stage with active PMOS load	15
2.11	Current summing	16
2.12	Feedback biased common-source amplifier with ideal load	17
2.13	Feedback biased common-gate amplifier with ideal load	19
2.14	Ideal differential signaling.	21
2.15	Single-ended to differential conversion	22
2.16	Common-gate-common-source transistor circuit	22
2.17	Typical variable gain output characteristic with 8 log-scaled gain steps from 26 to 0 dB.	25
2.18	Simple gain stage	27
2.19	CMOS noise model	29
2.20	Illustrative cross-section of CMOS.	31
2.21	Illustration of fundamental trade offs in analog CMOS design	32
3.1	Amplifier schematic	38
3.2	Amplifier layout for 40 nm technology.	46
3.3	Amplifier layout for 28 nm technology.	47

4.1	Source independent small-signal frequency transfer functions for both amplifiers.	51
4.2	Input loss (v_{in}/v_{cmut}) and input impedance simulated for different gain settings at f_c	52
4.3	Small-signal transfer functions for 40 nm with different gain settings.	52
4.4	Output noise power for all rails, including an ideal noise sum for reference.	54
4.5	HD ₂ simulated using harmonic balance for various values of V_{cmut} .	55
4.6	V_{out} HD ₂ simulated using harmonic balance on both amplifiers for various values of V_{out}	56
4.7	Transient run on 28 nm with variable gain.	57
4.8	Transient run on 28 nm without variable gain.	57
B.1	CMUT equivalents for noise calculation	81
C.1	Amplifier test bench schematic.	84
D.1	Simulated $N_{in,eq}$ for 40 nm and 28 nm technology. Post-layout.	85
D.2	Simulated $N_{in,eq}$ for 40 nm technology with and without amplifier noise. Post-layout.	86
D.3	Simulated $N_{in,eq}$ for 28 nm technology with and without amplifier noise. Post-layout.	86
D.4	Simulated small-signal output and input signals in 40 nm with $v_{cmut} = 15.4$ mV. Post-layout.	87
D.5	Simulated small-signal output and input signals in 28 nm with $v_{cmut} = 12$ mV. Post-layout.	87
D.6	Signal loss from v_{cmut} to v_{in} for different gain settings in 40 nm. Post-layout.	88
D.7	Signal loss from v_{cmut} to v_{in} for different gain settings in 28 nm. Post-layout.	88

Abbreviations

LNA	=	Low Noise Amplifier
IVUS	=	IntraVascular UltraSound
MEMS	=	MicroElectroMechanical Systems
CMUT	=	Capacitive Micromachined Ultrasonic Transducer
CMOS	=	Complementary Metal Oxide Semiconductor
MOSFET	=	Metal Oxide Semiconductor Field Effect Transistor
ASIC	=	Application-specific Integrated Circuit
EDA	=	Electronic Design Automation
FOM	=	Figure Of Merit
KISS	=	Keep It Simple, Stupid
HD	=	Harmonic Distortion
SNR	=	Signal-to-Noise Ratio
DR	=	Dynamic Range
2D	=	Two-dimensional
3D	=	Three-dimensional
DC	=	Direct Current
RF	=	Radio Frequency
RMS	=	Root Mean Square
FD-SOI	=	Fully Depleted Silicon On Insulator
LVT	=	Low Threshold Voltage
RVT	=	Regular Threshold Voltage
ADC	=	Analog to Digital Converter
SAR	=	Successive Approximation Register
CS	=	Common-source
CG	=	Common-gate
CG-CS	=	Common-gate-common-source
DRC	=	Design Rule Checking
LVS	=	Layout Vs. Schematic
PEX	=	Parasitic EXtraction

Introduction

Ultrasound imaging for medical purposes is a safe and accurate method for diagnosis and therapy of various hidden faults and diseases in the body, often as an alternative to X-ray. Efficient localization and characterization of e.g. coronary diseases is possible by examining the inside of arteries using ultrasound imaging. Intravascular ultrasound systems consist of a catheter featuring ultrasound transducers which is inserted into the artery, and an external machine controlling the catheter. Systems for two-dimensional imaging have been available for years, but three-dimensional imaging would represent a major leap in terms of image and diagnosis quality. Three-dimensional systems pose several challenges due to a rapidly increasing complexity. These challenges include restricted area and power consumption, enormous data transport and complicated beam control. Implementation of three-dimensional imaging becomes increasingly feasible as new technologies emerge. Recent advances in microelectromechanical and CMOS technology may make it possible to implement in-probe digitization, and integrate a larger part of the signal chain into an in-probe integrated circuit. This would reduce many complexity problems and enable efficient three-dimensional imaging.

The key to obtain good performance in signal processing systems is first of all to have a good receiver front-end. In an integrated ultrasound system this would include a low noise amplifier and analog-to-digital-converter. Noise performance, bandwidth and linearity are particularly important features in a front-end. A low noise amplifier may be the limiting component in such a system, and sufficient effort should be put in making it as good as necessary.

Nanoscale CMOS technologies are tempting because of their ability to perform

energy and area efficient digital processing, but they introduce several challenges for analog design, as reduced supply voltage and gain. It is important to study if these challenges impact performance of the final system. Design attempts are great for such investigation. The main objective of this thesis is to investigate design and implementation of low noise amplifiers in modern nanoscale CMOS technologies. Two low noise amplifiers specialized for intravascular ultrasound are designed and verified in two commercially available nanoscale CMOS technologies. These technologies, 40 nm bulk and 28 nm FD-SOI, are evaluated based on design experiences and achieved performance. A complete design approach is performed, from specification to design, layout and verification in modern EDA tools.

Analog amplifier design includes several challenges such as already mentioned performance characteristics, but also size and robustness are important properties. One achievement is to obtain good performance from an ideal concept, but the concept must also be realizable in a real circuit. Implementing robust and realizable designs is one of the goals in this thesis.

1.1 Main contributions

Several techniques and mechanisms were analyzed during this design in order to make proper design choices. Feedback biasing of both common-gate and common-source stages was implemented in order to ensure robust biasing. Analytic expressions for these couplings were derived in order to analyze their behavior. Noise analysis of the common-gate-common-source amplifier topology was performed in order to understand its cancellation capabilities. These findings resulted in two realizable amplifier designs implemented in 40 nm bulk CMOS and 28 nm FD-SOI with very small footprints. Variable gain in these amplifiers was implemented using switched triode transistors in order to save space and keep the circuit simple. Very low power consumption was achieved by using a systematic design approach built on g_m/I_d .

1.2 Thesis outline

Chapter 2 presents background information that motivates the need for in-probe digitization, and circuit theory necessary to make good decisions during design. Circuit theory includes general amplifier knowledge and presents some useful amplifier topologies together with their characteristics. This is followed by CMOS

transistor equations and special considerations in nanoscale CMOS technologies. Based on this theory are specifications and amplifiers developed in chapter 3. Design methodology and decisions are explained in the same chapter. Verification of and results from these designs are presented in chapter 4. Chapter 5 analyzes these results and other relevant decision made during design. Results are also compared with previous work. All experiences are summed up in the conclusion in chapter 6, and suggestions for further work are presented.

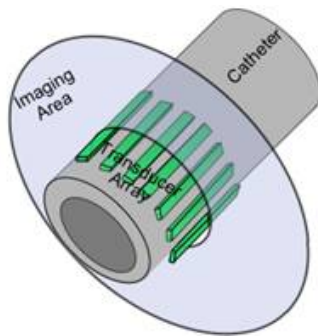
Background Theory

2.1 Intravascular ultrasound imaging

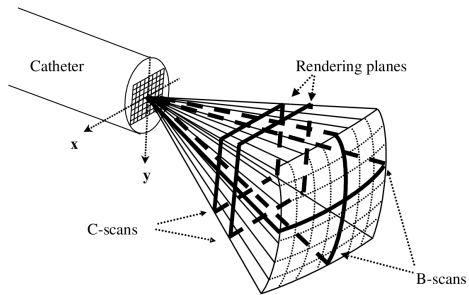
2.1.1 Imaging system

An efficient and accurate way of examine the inside of human arteries and vascular systems is by real-time intravascular ultrasound imaging (IVUS). A catheter is inserted into the artery and navigated to the destination requiring examination. The catheter features ultrasound transducers at or around the tip, which emits and receives ultrasound waves (high frequency pressure waves) to construct an image of the inside. This image can be observed on an external display during the entire examination. Two-dimensional (2D) side-looking imaging constructs a plane image of a cross section perpendicular to the artery, and has been commercially available for many years. This is usually achieved by placing a one-dimensional array of ultrasonic transducers around the end of a catheter, as depicted in figure 2.1a.

One problem with side-looking imaging is that artery plaque may loosen when the catheter touches it. A forward-looking solution is more preferable as it can observe blockages ahead of its path and avoid or fix these [2]. Three-dimensional (3D) imaging uses spatial beamforming to constructs a forward looking volumetric representation of the artery, and is currently subject for extensive research. A phased two-dimensional transducer array is able to do such imaging as depicted in figure 2.1b. Emitter beam is controlled by adjusting the phase of each transducer element, and received echo direction can be calculated from phase information in



(a) One-dimensional transducer array for side-looking imaging [1]



(b) Two-dimensional transducer array for volumetric imaging [2]

Figure 2.1: Different transducer array arrangements.

all channels. One challenge with 3D imaging is the huge number of transducers necessary to do spatial beamforming, and the complexity this causes. Transducers and front-end circuitry have to be small, integrable and consume very little power in order to place all of them on a millimeter catheter tip. On top of that is an enormous amount of wires necessary to carry the transducer signals in and out of the body. A proposed solution to these wire limitations is to do some signal processing inside the catheter and only transfer processed data out of the body [3]. This simplifies data transfer significantly, but puts greater demands on circuits within the catheter. At the same time does it relax requirements put on front-end circuitry, as it will no longer have to drive a long wire to an external system.

Figure 2.2 shows the receiver signal chain in a digitized ultrasound system with in-probe processing. Emitter chain has been omitted, even though it uses the same transducer elements as the receiver chain. Many capacitive micromachined ultrasound transducers (CMUTs) sense incoming pressure waves and generate time-varying current signals. Each CMUT has its own low noise amplifier (LNA) and analog-to-digital converter (ADC). Each LNA receives the electric signal and makes it possible for the ADC to sample it, as the signal from the CMUT is too weak to be sampled directly. After digitization, all signals are combined in a processing unit which performs different levels of digital processing before data is transferred out of the body.

Several advanced techniques are applied when an image is reconstructed from ultrasound echoes. One of the most successful techniques is to utilize nonlinear properties of human tissues which generate harmonic frequencies [4]. By listening for second harmonic echoes in addition to fundamental echoes, it becomes easier to characterize the type of tissue that is examined. Second harmonic echoes

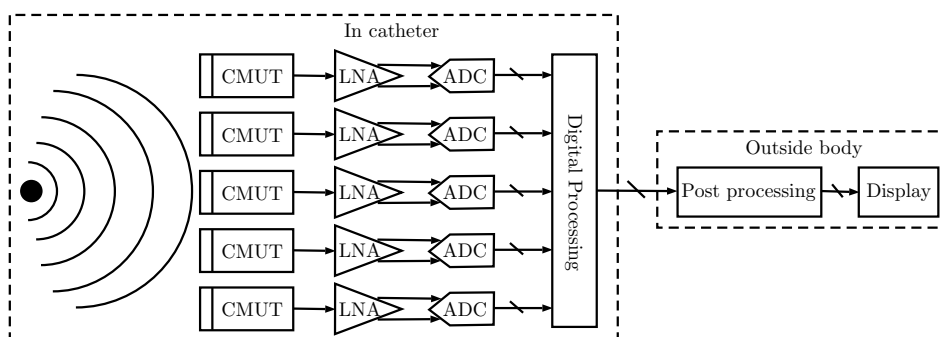


Figure 2.2: Digitized ultrasound receiver signal chain

also have lower side lobes than fundamental echoes due to absorption, which make the image sharper and more defined.

2.1.2 Ultrasound transducers

Traditionally, transducers based on piezoelectricity have dominated ultrasound applications [5]. During the last two decades, MEMS technology have made it possible to make capacitive micromachined ultrasound transducers (CMUT), which is manufactured directly on regular silicon wafers. This makes them highly integrable in arrayed CMOS applications [6]. Small size and simple manufacturing of CMUTs makes it possible to make large transducer arrays for three-dimensional imaging with a small footprint. CMUTs also offer wider bandwidth than piezoelectric transducers [6], making it easier to perform harmonic imaging.

CMUTs are micromachined three-dimensional devices where a thin silicon membrane and metal electrode is suspended on top of a substrate cavity. A typical structure is depicted in figure 2.3. This structure functions as a variable capacitor where the capacitance varies because of incoming pressure waves moving the membrane. When a sinusoidal sound wave hit the transducer, the membrane will start vibrating and create a sinusoidal current because of varying capacitance. The transducer is biased with a large DC voltage in order to increase current from the small capacitance variation.

In order to maximize transmitted and received signal, transducers are designed to resonate at their center frequency. Size and properties of transducers are hence dependent on the center frequency they are supposed to operate on. To achieve good acoustic coupling and beamforming capabilities, an effective element size of approximately half the acoustic wavelength is often chosen [8]. A common model

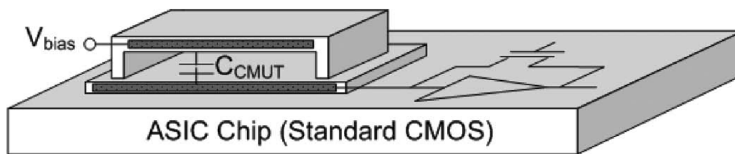
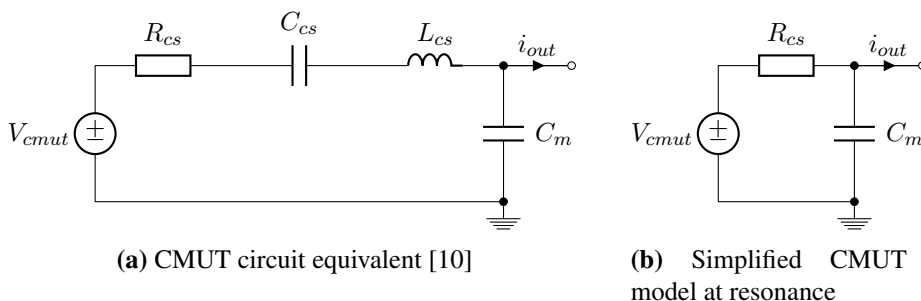


Figure 2.3: Illustration of CMUT structure on ASIC [7]

for electrical behavior of CMUTs is shown in figure 2.4a, and is based on Mason's circuit models for electromechanical transducers [9]. At resonance, reactance of L_{cs} and C_{cs} cancel each other and simplifies to the model shown in figure 2.4b.



When multiple transducers are stacked densely together in a two-dimensional array, unwanted electric and acoustic coupling usually appear. Berg [11] studied what impact front-end circuitry had on these unwanted effects. Their results showed that coupling effects were minimized when front-end input impedance was kept low.

2.1.3 Ultrasound front-end electronics

The front-end consisting of a LNA and an ADC plays a key role in the complete IVUS system. An ADC makes it possible to process signals digitally, and a LNA is necessary in order to sample the weakest signals with good accuracy. Performance of signal processing systems is often limited by their front-end performance. In the proposed 3D IVUS architecture are signals digitized already in the front-end, which makes performance solely dependent of the front-end.

Echoes received by ultrasound transducers have greatly varying amplitudes. Close objects return large echoes, while more distant objects return weak echoes. Be-

cause sound has a finite speed, distant echoes will return at a later time than close echoes. Combining these properties implies that early echoes usually have larger amplitude than late echoes. This can be utilized by the front-end electronics to maximize signal-to-noise-ratio and minimize distortion. By varying the amplification during listening time, output amplitude can be kept constantly high and hence maximize signal-to-noise-ratio (SNR). And with reduced amplification in early periods, large input signals that would normally have saturated the amplifier are kept within reasonable limits. Ultrasound systems benefit greatly from such time-gain compensation (TGC) in the front-end, and especially volumetric systems since they sweep all tissue depths.

2.1.4 Ultrasound digital post-processing

One reason for making an IVUS system with in-probe digitization and processing is to simplify data transport out of the body. By performing enough signal processing inside the probe, most signal redundancy can be removed before data is transmitted out of the body. This is one of the steps required for dealing with the complexity of two-dimensional transducer arrays. As CMOS processes are shrunk, more digital transistors can be packed together in a small area. Power consumption for each transistor is also reduced. This makes modern nanoscale CMOS processes interesting for use in in-probe processing.

2.2 Amplifier design

2.2.1 Amplifier properties

Amplifiers are active electronic devices that increases signal strength of input signals without adding unnecessary interference. A good specification is the key to an efficient and sufficient design solution. In order to make such a specification we need to define universal and measurable amplifier parameters. An amplifier can be described by several performance parameters which will be defined here.

The gain of an amplifier can be given as either small-signal or large signal gain. Small-signal gain is given by (2.1), and describes the ideal gain for small signals around a defined operating point. Large-signal gain takes into account nonlinear compression that appears for large output signals, and is given by (2.2). Both

equations refer to figure 2.5.

$$A_v = \frac{dV_{out}}{dV_{in}} = \frac{v_{out}}{v_{in}} \quad (2.1)$$

$$A_{vl} = \frac{\Delta V_{out}}{\Delta V_{in}} \quad (2.2)$$

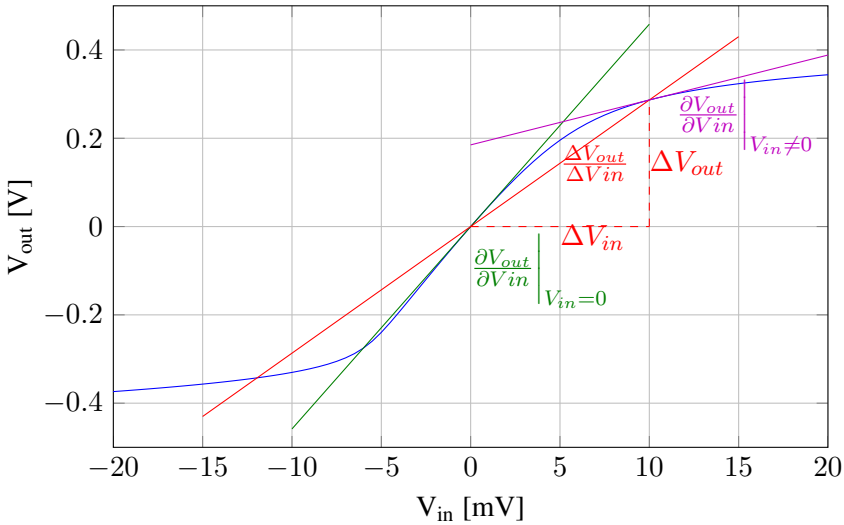


Figure 2.5: Typical amplifier characteristic shown in blue. Green slope illustrates small-signal gain around zero. Red slope illustrate large-signal gain at 10 mV. Purple slope illustrates small-signal gain around 10 mV.

Transistors are inherently nonlinear devices, which make transistor amplifiers non-linear as well. But amplifier gain can be considered approximately linear as long as the output signal is not too large compared to the desired operating point. Output signal may be approximated by a Taylor series around the operating point [12], as given in (2.3), which implies that nonlinear components will appear when the output signal deviates significantly from the operating point. nonlinear components generate harmonic frequencies when a single-tone sinusoidal signal is applied, as shown with four terms (a_0 to a_3) in (2.5).

$$V_{out} = a_0 + a_1 V_{in} + a_2 V_{in}^2 + a_3 V_{in}^2 + \dots \quad (2.3)$$

$$\begin{aligned}
 V_{out} \Big|_{V_{in}=V_i \sin(\omega t)} &= \frac{1}{2}a_2V_i^2 + a_0 + \left(\frac{3}{4}a_3V_i^3 + a_1V_i \right) \sin(\omega t) \\
 &\quad - \frac{1}{2}a_2V_i^2 \cos(2\omega t) - \frac{1}{4}a_3V_i^3 \sin(3\omega t) \quad (2.4)
 \end{aligned}$$

$$= v_0 + v_1 \sin(\omega t) + v_2 \sin(2\omega t) + v_3 \sin(3\omega t) \quad (2.5)$$

When gain deviates from the linear slope around an operating point, nonlinear components are generated in the output signal. A common way to describe these nonlinearities is by measuring the ratio between nonlinear and linear output signal content, as given by (2.6) and (2.7). v_1 is amplitude of desired linear content, while v_2 and v_3 describes amplitude of second and third harmonic content.

$$\text{HD}_2 = \left(\frac{|v_2|}{|v_1|} \right)^2 \approx \left(\frac{a_2V_i^2}{a_1V_i} \right)^2 = \left(\frac{a_2}{a_1}V_i \right)^2 \quad (2.6)$$

$$\text{HD}_3 = \left(\frac{|v_3|}{|v_1|} \right)^2 \quad (2.7)$$

Amplifiers do not have constant gain for all frequencies due to frequency dependent loads like capacitances. Such loads can either be intended or parasitic, and it is usually a goal to minimize such loads in order to reduce power consumption. The bandwidth of an amplifier refers its useful frequency band, and is often defined as the frequency range where small-signal gain is greater than or equal to -3 dB of maximum gain. An example of an amplifier frequency response is shown in figure 2.6. In some cases does not maximum gain appear at the desired center frequency due to design choices. Gain at center frequency would then be a more useful reference for bandwidth calculations than maximum gain. Bandwidth can then be considered the frequency range where gain is greater than or equal to -3 dB of gain at center frequency.

Input impedance is an important amplifier parameter, as already stated in [11]. It is a small-signal parameter that describes how much current flows into an amplifier when an input voltage is applied. Given the test setup in figure 2.7, input impedance, Z_{in} , is given by (2.8).

$$Z_{in} = \frac{v_{in}}{i_{in}} \quad (2.8)$$

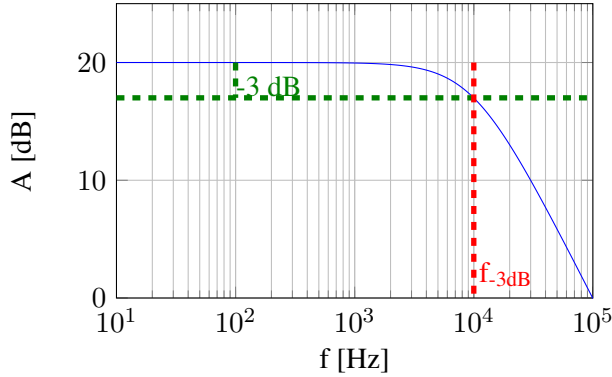


Figure 2.6: Typical amplifier frequency response (blue) with -3 dB bandwidth indicated (red).

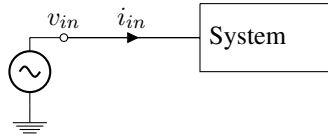


Figure 2.7: Test setup for input impedance.

Noise performance of low noise amplifiers are often described by noise factor. Noise factor compares output noise of a noisy amplifier to output noise of a noiseless amplifier, both with an input noise from a fixed source. In that way it can be measured how much the amplifier degrades the signal. A definition of noise factor is given in (2.9) along with a rewrite that includes input noise power (N_{in}) and output amplifier noise power (N_A) generated by the amplifier [13]. This expression shows that noise factor should be minimized, but it can not be less than unity. Noise *figure* (NF) measures noise *factor* (F) in decibels.

$$\begin{aligned}
 F &\equiv \frac{\text{SNR}_{in}}{\text{SNR}_{out}} = \frac{N_{out}|_{\text{noisy amplifier}}}{N_{out}|_{\text{noiseless amplifier}}} \\
 &= \frac{N_{in}A_v^2 + N_A}{A_v^2 N_{in}} = 1 + \frac{N_A}{A_v^2 N_{in}}
 \end{aligned} \tag{2.9}$$

Output SNR is given as

$$\text{SNR}_{out} = \frac{v_{out}^2}{N_{out}} = \frac{v_{out}^2}{A_v N_{in} + N_A} \tag{2.10}$$

Dynamic range is a measure of an amplifier's useful output amplitude range. An output signal that is smaller than noise generated by the amplifier can not easily

be distinguished from the noise, and is in general not particularly useful. When an output signal becomes so large that it violates the linearity specification, the signal can be considered absolute maximum. This squeezes the useful output amplitudes between a minimum, v_n , and maximum, v_m , level. Dynamic range is usually defined as given in (2.11) using these two extremes.

$$\text{DR} = \frac{v_m}{v_n} \quad (2.11)$$

2.2.2 Low noise amplifier

The reason for using a low noise amplifier (LNA) in a signal chain is mainly to minimize errors introduced by subsequent components in the signal chain. By increasing signal strength early in the signal chain, imperfections introduced by filters, ADCs and processing later in the chain get less impact on final signal-to-noise-ratio. As shown in [14, pp. 371-372], must all noise sources in a circuit be referred to a common node, often input or output of a chain, in order to do a fair comparison. (2.50) gives the input referred noise power $N_{in,eq}$ for a typical signal chain featuring an LNA as the first block. Noise powers in that expression is referred to *input* of each block. Other errors such as distortion and offset may also be evaluated in the same manner. From (2.12) it can be seen that errors made in the LNA will directly affect noise performance of the system. Errors made later in the signal chain will be divided by the LNA gain, which implies that the LNA should have sufficient gain to minimize the effect of errors in later blocks.

$$N_{in,eq} = N_{\text{LNA}} + \frac{N_1}{A_{\text{LNA}}^2} + \frac{N_2}{A_{\text{LNA}}^2 A_1^2} + \dots \quad (2.12)$$

2.2.3 Amplifier topologies

There are almost countless ways of using MOS-transistors in amplifiers. Many of them are based on basic topologies as the ones presented here. The principle behind most of these voltage amplifiers is to generate a voltage by controlling the current through a fixed resistive load. A MOS-transistor may be used as a controlled current source where small-signal transconductance (g_m) is an important property. The fixed load is usually implemented by exploiting small-signal output resistance (r_o) of MOS-transistors. In that way an amplifier may be made entirely out of transistors.

A NMOS-based inverting common-source (CS) stage is shown in figure 2.8a. By analyzing the linearized model in figure 2.8b, it can be seen that input impedance is ideally infinite due to ideally infinite gate resistance. Voltage gain from input to output is given as

$$A_v = -g_m (r_o \parallel R_L) = g_m r_{out} \quad (2.13)$$

where r_{out} can be considered the equivalent output resistance given by

$$r_{out} = (r_o \parallel R_L) \quad (2.14)$$

Small-signal models presented here are simplified models that omit tiny effects like transistor capacitances and body effect for simplicity reasons.

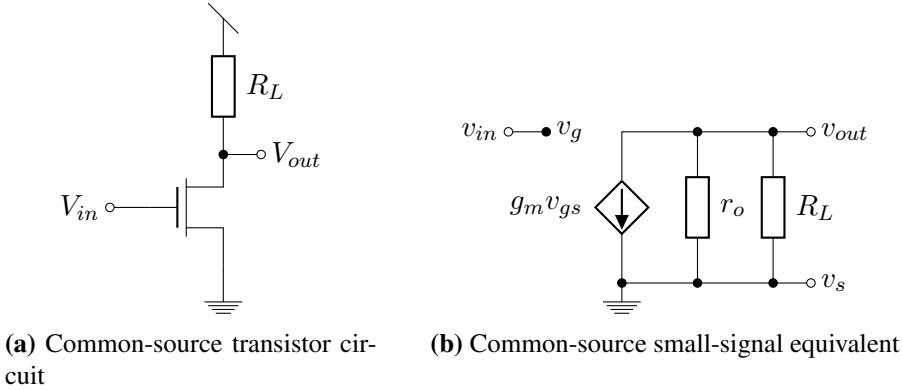


Figure 2.8: Common-source stage

A common-gate (CG) stage is shown in figure 2.9a. Analysis of the small-signal model in figure 2.9b shows a non-inverting behavior where gain is given by (2.15) and input impedance is given by (2.16). The last term in (2.15) is always less than unity, and where $A_v \gg 1$ can it usually be omitted [15].

$$A_v = g_m (r_o \parallel R_L) + \frac{R_L}{R_L + r_o} \quad (2.15)$$

$$r_{in} = \frac{(R_L + r_o) R_S}{g_m r_o R_S + R_L + R_S + r_o} \quad (2.16)$$

For $R_S \ll 1/g_m$, r_{in} can be approximated to R_S . With $R_S \gg 1/g_m$, it can be approximated to the impedance looking into source of the transistor:

$$r_{in} \approx \frac{1}{g_m} \left(1 + \frac{R_L}{r_o} \right) \quad (2.17)$$

Between these two extremes, r_{in} can be considered a parallel combination of these two contributions.

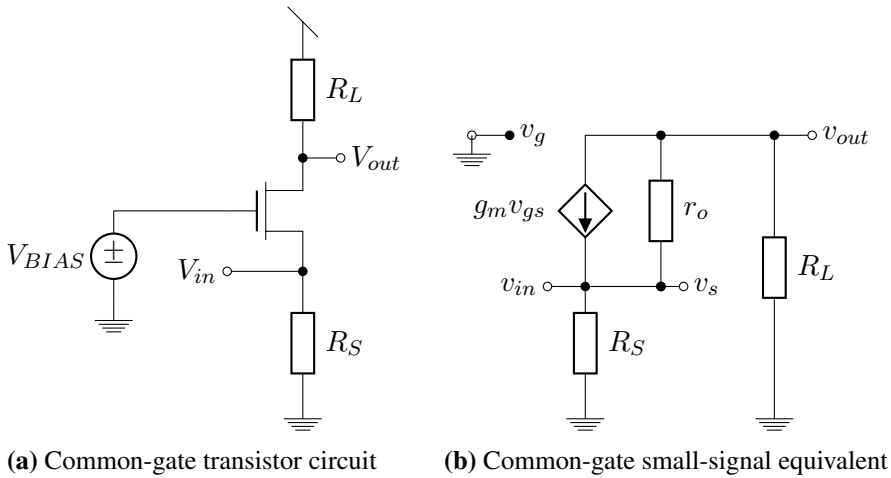


Figure 2.9: Common-gate stage

2.2.4 Current mirror biasing

The simplest way to bias a MOS-transistor is to use a current biased current mirror with a diode connected transistor. A proper gate voltage is created by the diode connected transistor, which can be used to bias the gate of a similar (or properly scaled) transistor. Such configurations are widely used as active loads in amplifiers, as shown in figure 2.10 [15]. The load resistance, R_L , will then be r_{o2} from the load transistor (Q_2).

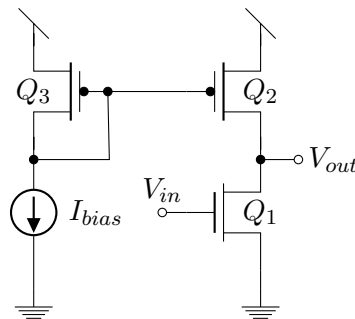


Figure 2.10: Common-source stage with active PMOS load. Biased using PMOS current mirror.

Transistors in a current mirror should in principle be physically equal in order to copy bias current directly. If a different current than the bias current is desired, several equal transistors with equal gate voltage could be laid out in parallel. This is called M-factor scaling and could save both bias current and transistor area by using a narrow diode-connected transistor that conducts less current as gate voltage generator. This gate voltage would make wider transistors conduct a larger current. A drawback with this method is that mismatch in the voltage generator would be amplified by the larger transistor, resulting in a larger current mismatch than necessary. Small transistors tend to have larger mismatch than large transistors, which make this issue even more relevant.

Simple current mirror biasing, as already shown in figure 2.10, can be used on any transistor configuration as long as gate voltage is supposed to be constant. If the gate voltage is not constant as i.e. in a common-source stage, some kind of voltage or current summing must be performed in order to mix bias and signal voltage. Voltage summing is not straightforward [16], but current summing may be performed by using resistors, as shown in figure 2.11. In this case, the sum of currents generates a gate voltage V_{in} , given by (2.18).

$$V_{in} = R_3 \frac{R_2 V_{bias} + R_1 v_{in}}{R_1 R_3 + R_2 R_3 + R_1 R_2} \quad (2.18)$$

In case of a large R_3 ($\gg R_1$ and $\gg R_2$), (2.18) simplifies to (2.19).

$$V_{in} = \frac{R_2 V_{bias} + R_1 v_{in}}{R_1 + R_2} \quad (2.19)$$

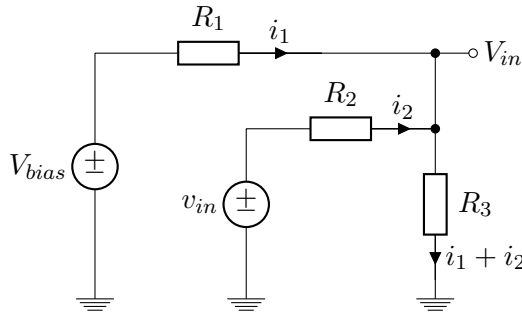


Figure 2.11: Current summing

2.2.5 Feedback biasing

An alternative way to bias a transistor is by using feedback biasing. A resistive feedback network between drain and gate will adjust gate voltage such that the transistor drives the current applied by a current source at drain. This feedback will also force the transistor into saturated operation. This configuration is shown on a common-source stage in figure 2.12a. An important detail in such a circuit is to isolate any external DC bias voltage from the gate so only the feedback can set the operating point. Gate-isolation could be done with a capacitor as shown with C_{ac} . The feedback works in principle as a diode-connection, except that the resistive network avoids reducing gain to unity. As would be shown, it is important that the feedback resistance is large compared to other resistances in the circuit if gain and other parameters should be kept unaffected. A small-signal model of figure 2.12a with finite load impedance, R_L , is given in 2.12b. Analysis of this model with infinitely large C_{ac} gives the small-signal gain in (2.20), which simplifies to the original (2.13) for small values of η . The ratio $\eta = R_L/R_f$ indicates how much the feedback affects different amplifier properties, and is minimized for $R_f \gg R_L$. Complete analysis is given in appendix A.

$$A_v = -\frac{\left(g_m R_L - \frac{R_L}{R_f}\right) r_o}{\left(1 + \frac{R_L}{R_f}\right) r_o + R_L} = -\frac{(g_m R_L - \eta) r_o}{(1 + \eta) r_o + R_L} \quad (2.20)$$

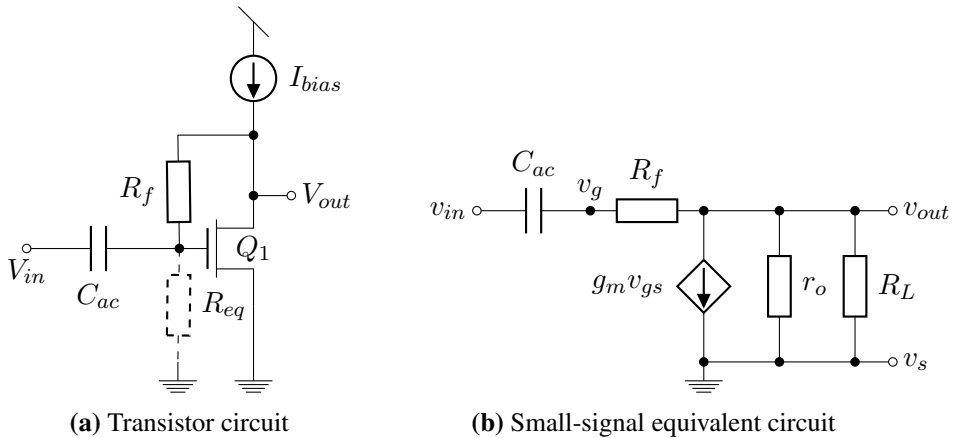


Figure 2.12: Feedback biased common-source amplifier with ideal load

Input resistance at gate is dependent on the feedback resistance R_f . Due to negative gain from gate to drain may this resistance be considered as a smaller equiv-

alent resistance R_{eq} from gate to ground according to the Miller theorem [17]. This equivalent resistance is dashed in figure 2.12a and its value is given by (2.21). From (2.21) can it be seen that input resistance at gate is dependent on both feedback resistance and small-signal gain.

$$R_{eq} = \frac{R_f}{1 - \frac{v_{out}}{v_g}} \quad (2.21)$$

By reordering the small-signal transfer function of v_g/v_{in} (given in (A.2)) without an infinitely large capacitor, it becomes obvious that this miller effect has a significant impact on high-pass response of the stage. This result is shown in (2.22).

$$\begin{aligned} \frac{v_g}{v_{in}} &= \frac{C_{ac} \frac{(1+\eta)r_o + R_L}{(g_m r_o + 1)\eta + r_o/R_f} s}{C_{ac} \frac{(1+\eta)r_o + R_L}{(g_m r_o + 1)\eta + r_o/R_f} s + 1} \\ &= \frac{C_{ac} \frac{R_f}{1 - v_{out}/v_g} s}{C_{ac} \frac{R_f}{1 - v_{out}/v_g} s + 1} = \frac{C_{ac} R_{eq} s}{C_{ac} R_{eq} s + 1} \end{aligned} \quad (2.22)$$

where v_{out}/v_g is given as:

$$\frac{v_{out}}{v_g} = -\frac{(g_m R_L - \eta)r_o}{(1 + \eta)r_o + R_L} \quad (2.23)$$

Another form of capacitive coupling is necessary when feedback biasing is applied to common-gate stages. A capacitor from gate to ground is necessary in order to keep the gate voltage constant when output voltage swings. A coupling like that is shown figure 2.13a together with an equivalent small-signal model in figure 2.13b. The small-signal equivalent contains a finite load resistance, R_L , from the I_{bias} current source.

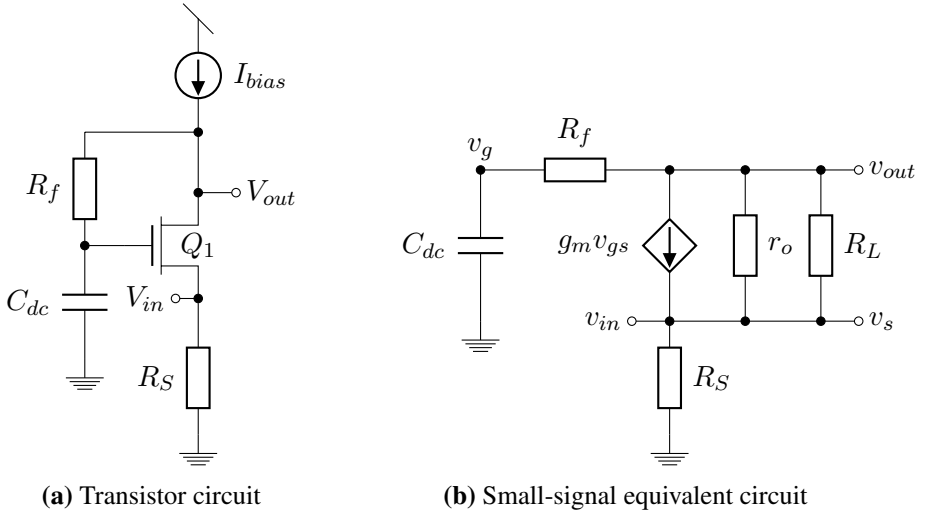


Figure 2.13: Feedback biased common-gate amplifier with ideal load

Voltage gain of this circuit can be found by analyzing figure 2.13b. Complete derivation and expressions for this exercise is found in appendix A. For an infinitely large C_{dc} , gain becomes as given in (2.24), which also simplifies to its original sibling (2.15) for small η .

$$A_v = \frac{(g_m r_o + 1) R_f R_L}{(R_f + R_L) r_o + R_f R_L} = \frac{(g_m r_o + 1) R_L}{(1 + \eta) r_o + R_L} \quad (2.24)$$

This circuit also features a high-pass response that is manipulated by the Miller effect, as seen from (2.25). It should be noted that this stage has a low frequency gain of one instead of zero.

$$\frac{v_{out}}{v_{in}} = \frac{C_{dc} R_f s + 1}{C_{dc} \frac{R_f}{1 - \frac{v_{out} - v_{in}}{v_g - v_{in}}} s + 1} = \frac{C_{dc} R_f s + 1}{C_{dc} R_{eq} s + 1} \quad (2.25)$$

Here the Miller gain is given by:

$$\frac{v_{out} - v_{in}}{v_g - v_{in}} = - \frac{(g_m R_L - 1 - \eta) r_o}{(1 + \eta) r_o + R_L} \quad (2.26)$$

Traditionally, feedback biasing has not been widely used due to its limited bias and output range. The diode connection forces gate and drain DC voltage to be approximately equal. In order to maximize output swing and linearity, output operating point has usually been placed at half the supply voltage, $V_{DD}/2$. In a

feedback biased stage this implies that also the gate have to be biased at $V_{DD}/2$. Section 2.3.1 will show that this result in a rather large overdrive if V_{DD} is large compared to the threshold voltage V_{th} , as was the situation for earlier technologies with high supply voltage. Such large overdrive is undesirable as overdrive also set minimum output voltage for saturated transistors. Because of this output swing limitation, feedback biasing has been considered a suboptimal solution compared to fixed biasing. However, in modern CMOS technologies, supply voltage have been reduced more than threshold voltage, which makes output swing from feedback biased stages more equal to fixed biased stages. This mostly eliminates the drawback of feedback biasing [18].

2.2.6 Noise analysis

Due to the random nature of noise, it is most common to treat noise as an averaged power. This implies that noise is often described as noise *power* rather than noise *voltage*. It is also important to distinguish between correlated and uncorrelated noise as these behaves differently when combined. (2.27) describes the sum of completely uncorrelated noise sources, which is rather different the sum of completely correlated noise sources described in (2.28). Uncorrelated noise is summed as a sum of squares, while correlated noise is summed before squaring.

$$V_{n,sum}^2 = \sum_k V_{n,k}^2 \quad (2.27)$$

$$V_{n,sum}^2 = \left(\sum_k V_{n,k} \right)^2 \quad (2.28)$$

Regular signals are usually correlated as well. This implies that also signals should be summed according to (2.28) when signal power is considered.

2.2.7 Differential amplifiers

Differential signaling is a technique where a signal is described as a *difference* between two complementary signals instead of an absolute signal compared to a fixed reference (i.e. ground). This technique is beneficial for several reasons. Noise that is common to both signals will be canceled and the effective amplifier output swing is doubled, as illustrated in figure 2.14. From (2.3) can it be observed that even order nonlinearities always have positive amplitudes due to squaring,

which make them common to both signals as long as distortion is equally large. Since all interference that is common to both signals is canceled will even order nonlinearities be canceled as well [15].

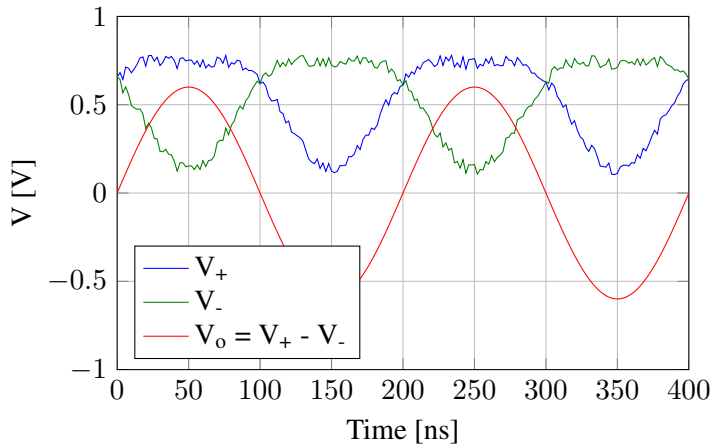


Figure 2.14: Ideal differential signaling. Two complementary signals (blue and green) with equal noise and distortion added, and the resulting differential signal (red).

When ignoring all kinds of interference cancellation, differential signaling also effectively doubles the signal-to-noise-ratio (SNR) due to increased signal swing. As described in section 2.2.6, will a doubling of signal voltage amplitude correspond to four times higher signal power. Uncorrelated noise power on each rail is only summed after squaring, according to (2.27). This may increase signal power twice as much as noise power, hence doubling SNR [15].

In case of a single-ended input signal, this signal needs to be converted to a differential signal if the remaining signal chain is differential. A conversion could be performed by applying the same signal to an inverting and a non-inverting amplifier, as shown in figure 2.15. Such a device is sometimes called a balun/unbal because it performs a balanced-unbalanced conversion.

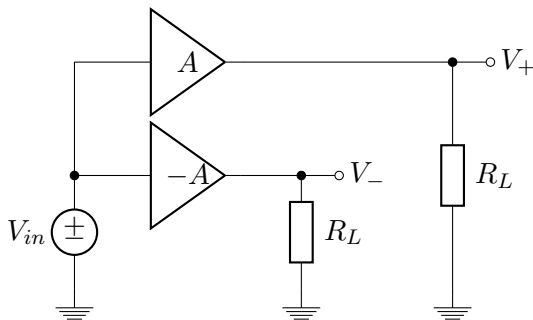


Figure 2.15: Single-ended to differential conversion

[19] and [20] has proved that the common-gate-common-source topology shown in figure 2.16 does single-ended to differential conversion, and is able to perform cancellation of noise and distortion from the common-gate transistor. Fluctuations from the common-gate transistor, modeled as a current source i_{nT} , generate a voltage over R_S . This voltage is input to the common-source stage and compensated for there due to negative gain.

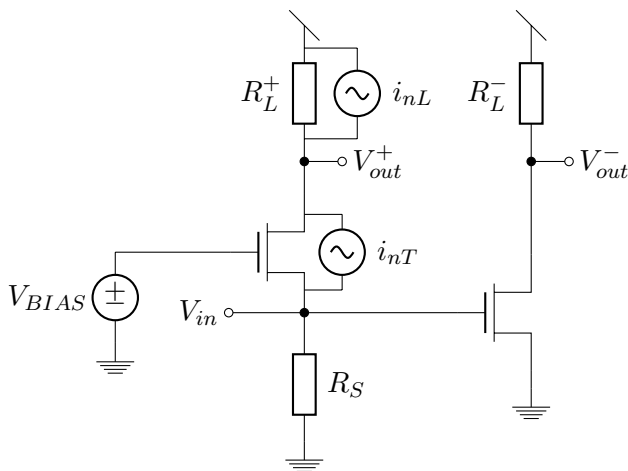


Figure 2.16: Common-gate-common-source transistor circuit

Noise and distortion from the *load* is unfortunately not canceled by this mechanism. Analysis of load fluctuations in this CG-CS topology was not found in any literature, which made it necessary to examine it here. Noise and distortion generated by the common-gate load was modeled as a current source, i_{nL} , in parallel with R_L as shown in figure 2.16. Since the noise on each rail in this case is correlated, it is sufficient to treat them as voltage signals rather than power signals.

Contributions on V_{out}^+ and V_{out}^- from this source were derived using the already presented small-signal models. Figure 2.9 was used to find contributions on V_{out}^+ and V_{in} from I_{nL} . Noise on V_{in} was input to common-source in figure 2.8 in order to find noise at V_{out}^- . This resulted in these three transfer functions:

$$\frac{v_{out,n}^+}{i_{nL}} = -R_L^+ \parallel ((g_m^+ r_o^+ + 1)R_S + r_o^+) \quad (2.29)$$

$$\frac{v_{in,n}}{i_{nL}} = -\frac{R_L^+ R_S}{(g_m^+ r_o^+ + 1)R_S + r_o^+ + R_L^+} \quad (2.30)$$

$$\begin{aligned} \frac{v_{out,n}^-}{i_{nL}} &= -g_m^- (r_o^- \parallel R_L^-) \frac{v_{in,n}}{i_{nL}} \\ &= g_m^- (r_o^- \parallel R_L^-) \frac{R_L^+ R_S}{(g_m^+ r_o^+ + 1)R_S + r_o^+ + R_L^+} \end{aligned} \quad (2.31)$$

It should be noted from these expressions that $v_{out,n}^+$ and $v_{out,n}^-$ ended up with opposite signs, which results in summation during subtraction instead of cancellation. This actually amplifies noise and distortion from the common-gate load. The amount of interference that is coupled to CS is given by:

$$\left| \frac{v_{out,n}^-}{v_{out,n}^+} \right| = \frac{\left| \frac{v_{out,n}^-}{i_{nL}} \right|}{\left| \frac{v_{out,n}^+}{i_{nL}} \right|} = g_m^- (r_o^- \parallel R_L^-) \frac{R_S}{(g_m^+ r_o^+ + 1)R_S + r_o^+} \quad (2.32)$$

This noise coupling can be minimized by reducing R_S , but that is not always desired. It can also be minimized by increasing r_o^+ . These modifications would at the same time reduce cancellation of interference from the common-gate transistor, according to [19].

2.2.8 Frequency response

As seen in section 2.2.3, do amplifiers have non-zero resistive output impedance due to i.e. r_o and R_L . In CMOS amplifiers, these resistances are usually large in order to create large gain from small transconductances. This makes output voltage sensitive to low-impedance loads. Internal loads are usually not resistive in CMOS circuits, but also capacitive loads can feature low impedance if the frequency is

high enough. This creates the frequency-dependent response mentioned in section 2.2.1. A low-pass first-order combination of resistance and capacitance will have a response as already shown in figure 2.6, which is given by (2.33).

$$|A_v(f)| = \frac{A_v(0)}{\sqrt{1 + \left(\frac{f}{f_{-3dB}}\right)^2}} \quad (2.33)$$

where the cutoff frequency, f_{-3dB} , is given in (2.34). This cutoff frequency is also valid for high-pass combinations. $A_v(0)$ is the low frequency gain given by g_m and r_{out} , according to (2.13).

$$f_{-3dB} = \frac{1}{2\pi RC} \quad (2.34)$$

Another frequency parameter of amplifiers is unity-gain frequency. A transconductor loaded with a capacitance will not be able generate a gain of more than one at unity-gain frequency. This frequency is given by (2.35).

$$f_{ug} = \frac{g_m}{2\pi C_L} \quad (2.35)$$

By reordering and substituting (2.13) and (2.35) into (2.34), can cutoff-frequency be given entirely by gain and unity-gain frequency:

$$f_{-3dB} = \frac{f_{ug}}{A_v(0)} \quad (2.36)$$

As the low-frequency gain is given by g_m and r_{out} , it is clear that only three parameters, g_m , r_{out} and C_L , are necessary to describe small-signal frequency-dependent gain of an amplifier. These three parameters also play a central role in other amplifier properties such as power consumption and noise.

2.2.9 Variable gain

Variable gain is a feature that is used to keep signal-to-noise-ratio (SNR) high for a wider input range than only peak input. Given a system with constant amplifier noise power (N_A) and varying input signal amplitude (v_{in}), SNR at output is given by (2.37). Gain is limited by the product of A and v_{in} , as large output amplitudes create distortion. By having large gain for small input signals and small gain for large input signals, could both v_{out} , distortion and SNR ideally be kept constant [21]. A typical output characteristic for a variable gain amplifier with coarse gain

steps is shown in figure 2.17. Even though this plot depicts coarse gain steps could variable gain be implemented with continuous gain control as well.

$$\text{SNR} = \frac{v_{out}^2}{N_A} = \frac{(A \cdot v_{in})^2}{N_A} \quad (2.37)$$

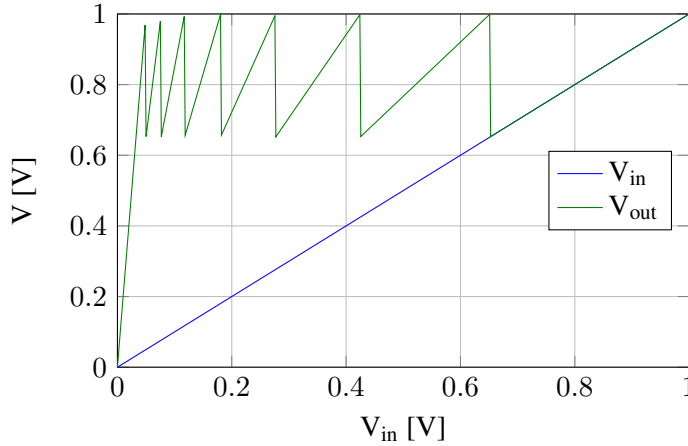


Figure 2.17: Typical variable gain output characteristic with 8 log-scaled gain steps from 26 to 0 dB.

When considering input noise, N_{in} , in addition to amplifier noise, the situation becomes less ideal, as stated in (2.38). If output SNR is dominated by noise from the source, as is the goal in LNAs, SNR will be higher for high input signals than small input signal, as less input noise is amplified with low gain. At the same time does (2.9) state that noise factor benefit significantly from high gain, which is important for low input signals. This makes variable gain a clever extension of the useful input level range.

$$\text{SNR} = \frac{v_{out}^2}{A^2 N_{in} + N_A} \quad (2.38)$$

There are several ways to implement variable gain in a variable gain amplifier (VGA). It is reasonable to start with a maximum gain and find clever ways to reduce this gain, as amplifier performance is usually most vulnerable at maximum gain. In principle, there are two main techniques to *reduce* overall gain. Either truncation of the input signal, or reducing gain in the amplifier itself. Input truncation involves some kind of voltage or current division at or before the amplifier input. The second method can be achieved by changing g_m , r_{out} or both, as seen from (2.13) and (2.15).

2.3 CMOS Technology

2.3.1 Transistor properties

When designing transistors it is useful to have analytic models for their behavior. Unfortunately, simple models for CMOS transistors are very approximate compared to their real behavior, especially when all operating regions should be considered. Even though simple models are too inaccurate for quantitative use, it is still possible to use them qualitatively in order to tweak parameters in the right direction. The Shichman-Hodges model [22] for a MOSFET in strong inversion gives drain current, I_d , as given in (2.39). $V_{\text{eff}} \equiv V_{gs} - V_{th}$ is effective voltage, which is also called overdrive. λ is the channel-length modulation constant.

$$I_d = \begin{cases} \mu_i C_{ox} \frac{W}{L} \left(V_{\text{eff}} V_{ds} - \frac{V_{ds}^2}{2} \right), & \text{if } V_{ds} < V_{\text{eff}} \\ \frac{1}{2} \mu_i C_{ox} \frac{W}{L} V_{\text{eff}}^2 (1 + \lambda V_{ds}), & \text{if } V_{ds} > V_{\text{eff}} \end{cases} \quad (2.39)$$

The first region is for linear/triode operation where V_{ds} is small, and the transistor behaves approximately as a resistor where V_{eff} controls the resistance. This is not particularly useful for amplifier transistors, but may be used as area-efficient high-resistance devices with approximately constant resistance. Its resistance and simplified resistance are given in (2.40).

$$R_{ds} = \frac{V_{ds}}{I_d} = \frac{L}{W} \frac{1}{\mu_i C_{ox} \left(V_{\text{eff}} - \frac{V_{ds}}{2} \right)} = \frac{L}{W} \frac{1}{\mu_i C_{ox} V_{\text{eff}}} \quad (2.40)$$

The second region in (2.39) is for a saturated transistor where drain current is nearly independent of V_{ds} . This makes the transistor behave approximately as a current source, but with a finite output resistance due to channel-length modulation. The small signal transconductance, g_m , of a MOSFET is derived from this equation:

$$g_m = \frac{\partial I_d}{\partial V_{gs}} = \sqrt{2\mu_i C_{ox} \frac{W}{L} I_d} \quad (2.41)$$

Under the same conditions, small signal output resistance, r_{ds} , is given by equation (2.42), where (λL) can be considered approximately constant [15].

$$r_o = r_{ds} = \frac{1}{\lambda I_d} = \frac{L}{(\lambda L) I_d} \quad (2.42)$$

From (2.41) we can find current efficiency, g_m/I_d , which normalizes transconductance with respect to current. (2.43) shows an approximate expression for a

saturated transistor in strong inversion. As seen from (2.43), current efficiency increases with lower drain current. For very low drain currents the transistor enters medium and weak inversion and makes the expression invalid. In weak inversion (subthreshold), g_m is approximately proportional to I_d which causes a constant g_m/I_d [23].

$$\frac{g_m}{I_d} = \sqrt{2\mu_i C_{ox} \frac{W}{L} \frac{1}{I_d}} \quad (2.43)$$

For the simple amplifier stage depicted in figure 2.18, low-frequency small-signal voltage gain, A_{vi} , is given by (2.44). This is also called intrinsic gain of a transistor.

$$A_{vi} = -g_m r_{ds} = -\frac{g_m}{g_{ds}} = -\frac{1}{(\lambda L)} \sqrt{2\mu_i C_{ox} W L} \frac{1}{I_d} \quad (2.44)$$

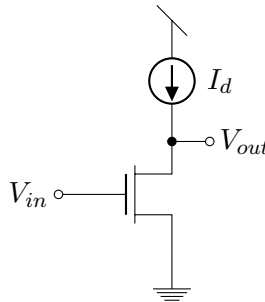


Figure 2.18: Simple gain stage

2.3.2 Subthreshold operation

Previously stated expressions for CMOS transistor behavior are based on the assumption that transistors are operated in strong inversion, where $V_{gs} > V_{th}$. This mode features a strongly inverted channel where excess charge create good conductivity. When $V_{gs} < V_{th}$ there is no excess charge, which make the transistor behave differently. This is called subthreshold operation since gate-source-voltage is significantly smaller than threshold voltage. A general rule of thumb for subthreshold is to keep V_{gs} at least 100 mV below V_{th} [24]. This lead to a weakly inverted channel, where the only way for charge to flow is by diffusion. Compared to excess charge conduction are diffusion currents very small and also highly dependent on temperature. An approximate model of subthreshold current is stated in (2.45), where V_T is thermal voltage, n is a slope factor and K are various constants [25]. V_T is approximately 26 mV at room temperature, while n often varies between 1.2 and 1.5.

$$I_d = K\mu C_{ox} \frac{W}{L} e^{\frac{V_{\text{eff}}}{nV_T}} \quad (2.45)$$

V_{eff} is negative for subthreshold operation, and since V_T is a very small voltage will this exponential relationship quickly lead to very small currents. Very low currents indicate very high resistance, which may be favorable in e.g. feedback bias configurations. In order to use this property could a transistor be diode connected ($V_g = V_d$). The drain current is then given by (2.46), which gives the resistance in (2.47).

$$I_d = K\mu C_{ox} \frac{W}{L} e^{\frac{V_{ds}-V_{th}}{nV_T}} \quad (2.46)$$

$$R_{ds} = \frac{V_{ds}}{I_d} = \frac{1}{K\mu C_{ox}} \cdot \frac{L}{W} \cdot \frac{V_{ds}}{e^{\frac{V_{ds}-V_{th}}{nV_T}}} \quad (2.47)$$

As the exponential function grows faster than a linear function is this resistance clearly nonlinear, and R_{ds} decreases when V_{ds} is increased. This also implies that resistance decreases when current increases. However, as long as V_{ds} is kept small enough may this resistance be so large that resistance variation does not matter much.

2.3.3 Transistor noise

MOSFETs generate various kinds of noise due to different mechanisms, and noise is also dependent on transistor operation region. White thermal noise (I_{nd}) is a noise source that emerges from channel resistance, like thermal noise in regular resistors. The channel behaves rather different in triode region than in saturation, which makes it useful to have different expressions for noise in these two regions. This noise is modeled as a current source between drain and source, as shown in figure 2.19. Noise current spectral density $I_{nd}^2(f)$ is given by (2.48) for a MOS transistor in strong inversion [26]. k is Boltzmann's constant and T is absolute temperature.

$$I_{nd}^2(f) = \begin{cases} 4kTg_{ds}, & \text{if } V_{ds} < V_{\text{eff}} \\ 4kT \left(\frac{2}{3}\right) g_m, & \text{if } V_{ds} > V_{\text{eff}} \end{cases} \quad (2.48)$$

Low frequency flicker noise, $V_{fg}^2(f)$, emerges from imperfections in the interface between channel and gate oxide. Equivalent gate noise voltage spectral density

may be approximated by (2.49), which refers to the noise source at gate in figure 2.19. K is a process- and bias-dependent parameter [27]. The $1/f$ relationship in noise spectral density can be a huge disadvantage in low frequency applications, especially if small transistors are preferred. This noise is also called $1/f$ -noise due to the $1/f$ relationship.

$$V_{fg}^2(f) = \frac{K}{WLfC_{ox}} \quad (2.49)$$

When referring both noise sources as an input voltage, the total expression becomes

$$V_{ni}^2(f) = \frac{I_{nd}^2}{g_m^2} + V_{fg}^2 = \frac{2}{3} \frac{4kT}{\sqrt{2\mu_i C_{ox} \frac{W}{L} I_D}} + \frac{K}{WLfC_{ox}} \quad (2.50)$$

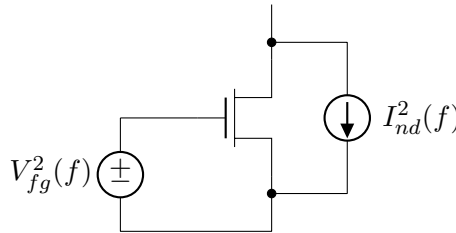


Figure 2.19: CMOS noise model

Capacitances in a circuit will, together with resistances, form low-pass filters which filter this noise. Since noise is generated by those resistances, the resulting noise power over various capacitances can be described by (2.51).

$$v_n^2 \propto \frac{kT}{C} \quad (2.51)$$

2.3.4 Nanoscale analog design

Most *digital* circuits benefit significantly from lithography process shrinking due to smaller footprint, less parasitic capacitances and lower supply voltages. CMOS processes are therefore constantly shrunk. This implies that analog circuits working together with digital circuits must be implemented in the same nanoscale processes. There are both challenges and benefits from designing analog circuits in nanoscale processes. When transistor sizes are scaled, higher bandwidths can be achieved, but several other transistor parameters are not scaling well with process

scaling. Threshold voltage is not scaling proportional to supply voltage, g_m is decreased and intrinsic gain is also reduced [28].

High-field and short-channel effects play an important role when transistors become short and narrow. The most important effects are mobility degradation (velocity saturation), hot-electron effects and drain-induced barrier lowering. All these effects have complicated models with many variables which make it impossible to calculate transistor sizes using simple analytic expressions such as the Shichman-Hodges model [29], especially when different operating regions are considered. More complex models such as EKV[30] have been developed in order to attempt nanoscale analytic expressions, but these expressions tend to become cumbersome because of their complexity. A new method based on physically measurable transistor properties together with optimization techniques has therefore been proposed. The g_m/I_d method takes advantage of today's computing power and good transistor simulation models, and optimizes transistor sizes based on measurable properties such as g_m , g_m/I_d , g_m/g_{ds} and V_{ds} . This approach also includes models for subthreshold, weak inversion, strong inversion and all the middle modes since it is based purely on simulation models [29].

The reason for using those four particular parameters is because they are all measurable sizes, and because they create a complete and unique description of the transistor behavior. Section 2.2.1 proved that g_m and g_m/g_{ds} fixed the gain-frequency response for a given C_L , and g_m/I_d decides current efficiency of a given stage. g_m/I_D has showed to be approximately constant when the process is shrunk, and can therefore be used as a global efficiency measure independent of technology. V_{ds} describe the surrounding condition for the transistor, and hence make the description unique.

Another important effect of process scaling is gate leakage. One of the big differences with 45/40 nm technologies compared to earlier technologies was the introduction of high- κ insulator. In earlier technologies, gate oxide was becoming so thin that electrons easily tunneled through the oxide and created significant gate leakage. By replacing traditional silicon dioxide (SiO_2) with an oxide featuring higher κ (i.e. hafnium-based), gate thickness could be increased without sacrificing C_{ox} , and hence reduce gate leakage [31].

Matching accuracy does fortunately scale with process shrinking, which makes it possible to make smaller circuits with equal robustness as older technologies. But reducing the transistor size is not always easy, as the effects mentioned above may hurt the performance. A common rule to avoid the worst nanoscale effects in analog design is to keep the gate lengths at least two times the minimum gate length [32]. It may also be wise to follow recommendations from the manufacturer's

process documentation.

2.3.5 Bulk and SOI CMOS processes

CMOS manufacturers constantly strive to reduce cost and complexity, and increase speed, yield and energy efficiency. Some short-channel effects make it difficult to scale process nodes further down in an efficient manner. Traditional bulk processes manufacture transistor implants directly on a doped silicon wafer. This technique leads to lack of control with the channel depth since the substrate is much thicker than desired channel thickness. In long channel devices this was not a big issue, as the channel was much longer than the channel depth and the channel could be considered approximately homogeneous. As channel length keep shrinking, electric fields spreading further down in the substrate create a non-homogeneous channel with undesired behavior. One of the most successful attempts to solve this problem is the Fully Depleted Silicon-on-Insulator (FD-SOI) technique. A thin oxide layer (insulator) is created on top of the substrate, and all transistors are grown on top of this insulator. Figure 2.20 illustrate the difference between bulk and FD-SOI process. FD-SOI form a very shallow (down to a few nanometers), homogeneous and controllable channel that does not even need to be doped. The result of this technique is improved speed, leakage current and variability [33]. Other properties as g_m/I_d and g_m/g_{ds} are also improved by FD-SOI [34], which makes it possible to achieve higher energy efficiency and gain.

SOI was earlier considered a more complex and expensive process than bulk. As process nodes are scaled down, bulk processes requires a greater and greater effort in order to offer good performance. This makes the complexity at today's smallest nodes, i.e. 28 nm, approximately equal for bulk and FD-SOI [35].

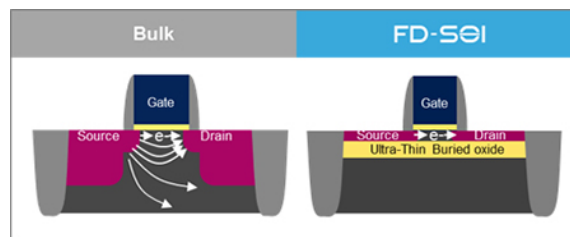


Figure 2.20: Illustrative cross-section of bulk process (left) and FD-SOI process (right). White arrows illustrate electric field lines.[36]

2.4 Figure-of-merit-based design

Design of energy efficient analog circuits poses several inevitable trade offs that must be considered. As stated in [37], it is a fundamental trade off between power, speed and accuracy, where good power implies low power, and good speed and accuracy implies high speed and accuracy. Figure 2.21 illustrate these trade offs where for instance high accuracy would require either low speed, high power or both. This relationship is the foundation of the commonly used figure-of-merit (FOM) stated in (2.52), which is a measure for energy efficiency. P is consumed power, DR is dynamic range as described in section 2.3.1, and f is bandwidth. This FOM is related to the consumed energy per pole, and should be minimized. [23] states a fundamental limit for analog processing given by (2.53), which can be used to compare the achieved performance with an absolute limit. A capacitive output load is used early in the derivation of (2.53), but is canceled when power and signal-to-noise-ratio is combined due to its role in both power consumption and noise filtering.

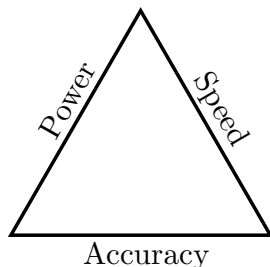


Figure 2.21: Illustration of fundamental trade offs in analog CMOS design [37].

$$\text{FOM} = \frac{P}{\text{DR}^2 \cdot f} \quad (2.52)$$

$$\text{FOM}_{\min} = 4kT \quad (2.53)$$

As shown in section 2.3.3, thermal noise may be reduced by increasing current, and flicker noise can be reduced by increasing transistor size. On the other hand, section 2.3.1 shows that gain and current efficiency decreases with increased current, and bandwidth decreases with increased length or decreased current. Amplifier design with all these trade offs in a power-constrained LNA is a challenge that requires a FOM to find the best solution.

A FOM could also be necessary when comparing an amplifier with existing solutions. Noise and distortion performance for LNAs are usually given as noise figure (NF) and harmonic distortion (HD_2), which are not directly compatible with the dynamic range defined in (2.11). When comparing existing work, it may therefore be useful to replace DR as given in (2.54). This also includes amplifier gain, as gain is included in noise factor according to (2.9).

$$DR_{LNA} = \frac{1}{F \cdot HD_2} \quad (2.54)$$

Since each ultrasound element has a constrained area ($A_{tot} = W_{tot} \cdot L_{tot}$), it may also be wise to include total area consumption in the FOM, as proposed in (2.55). When constructing a new FOM, it is important that its content reflect actual trade offs, and that size of these trade offs are evenly matched with the remaining expression. According to (2.49), flicker noise power may be halved by doubling area. This trade off is directly accounted for in FOM_{LNA} . As thermal noise is kT/C -filtered according to (2.51), it is possible to reduce thermal noise by increasing circuit capacitances. Increased capacitances could be done by increasing circuit element sizes, as is reflected in the area-term of FOM_{LNA} .

$$FOM_{LNA} = \frac{P \cdot A_{tot} \cdot F \cdot HD_2}{f} \quad (2.55)$$

2.5 Computer models

CMOS circuits are designed using sophisticated electronic design automation (EDA) software with several tools to ease development and assure high quality circuits after production. Component models may be more or less accurate depending on how well the environment is specified, as nanoscale effect may play an important role during e.g. layout. If effects are dependent on how the components are laid out, these effects would usually not appear during the schematic design phase. This raises the necessity of an iterative design flow where performance of smaller schematic solutions is checked against the corresponding layout solution.

Fabrication plants (fabs) supply several design rules and restrictions to avoid non-functional circuits and bad matching characteristics. Rules such as minimum and maximum transistor widths and lengths are usually implemented in the schematic tool, but there are also layout specific rules that can not be checked during schematic design. These rules are fed into the design tool that performs Design Rule Checking (DRC). This tool checks all dimensions in the entire layout against rules about

minimum and maximum dimensions such as length, width, area, distance, overlap, clearance and density.

2.6 Existing work

Most existing IVUS LNA work using CMUTs, such as [10, 38–43], focus on systems with ADCs and processing outside the body. Such amplifiers drive a long cable, which is a significantly larger load than an integrated ADC. They also have very different power and area budgets since the chip does not include either ADC or digital processing. Typically such amplifiers consume milliwatts rather than microwatts. The proposed FOM does not include load, which makes comparison with such amplifiers unfair as larger loads demands more current given the same bandwidth.

A comparable work was found in [20] which featured a g_m -boosted common-gate common-source amplifier in 65 nm technology with a noise figure of 2.98 dB, power consumption 67 μW and area of 375 $(\mu\text{m})^2$. That work was partly based on RF LNA work performed by Bruccoleri and Nauta in [44]. IVUS engineers tend to borrow some techniques from RF applications due to similarities of ultrasound and RF, and because the amount of work performed within RF is huge.

Design Methodology

The purpose of this thesis was to design LNAs for use in 3D IVUS systems using nanoscale CMOS technologies. This chapter focuses on design and implementation of these amplifiers. It describes design methodology used and design decisions made during development. This included development of specifications, topology and circuit decisions, as well as transistor sizing and layout creation. The design task was solved by first improving and extending an earlier design attempt made in [45]. This amplifier was designed in a commercially available general purpose 40 nm bulk CMOS technology. Next up was design of a similar LNA in low power 28 nm FD-SOI CMOS technology to see if further improvement was possible in another technology.

3.1 Specifications

A specification for amplifier and surrounding system was made in cooperation with prof. Ytterdal in order to make an amplifier that could function in a real system. In [45] it was demonstrated that most of the specifications there were reasonable and achievable, which made a good starting point for this thesis. Revised specifications are summarized in table 3.1.

Maximum supply voltage was given by technology requirements and was necessary in order to avoid transistor breakdown. Some extra supply voltage in 28 nm gave more voltage headroom than in 40 nm, which is usually positive from an analog point of view.

Transducers manufactured for this system were designed to resonate at a center frequency f_c of 5 MHz, which made it necessary to use the same frequency as center frequency for this LNA. At center frequency, their impedance was measured to $10\text{ k}\Omega \angle -60^\circ$. As stated in section 2.1, could undesired coupling between CMUTs be reduced by using low impedance front-end. An maximum input impedance of half the source impedance was considered appropriate, hence an input impedance of 0.5 - 5 k Ω .

Ultrasound systems are usually wide-band systems with at least 100% bandwidth, which translates to 5 MHz bandwidth around center frequency of 5 MHz. This frequency range from 2.5 to 7.5 MHz also decided the noise bandwidth used for SNR calculations. For an LNA it is desirable that the amplifier makes less noise than the source it is sensing, which translates to a noise figure of less than 3 dB. However, only noise figure was not sufficient to completely describe noise properties of a distortion-limited amplifier, as noise figure is independent of input signal. A SNR requirement was also necessary in order to have practical input and output signal amplitudes. A rule-of-thumb within IVUS applications has been to have at least 40 dB SNR in order to create good images. As the system may be used for harmonic imaging as well, it was required to have a harmonic distortion of less than -50 dB.

In order to receive the strongest ultrasound echoes, a low-gain setting of 0 dB was required to avoid compression. A single-ended to differential conversion was also desired in order to use a differential ADC behind it. Recent research on SAR ADCs has shown that these ADCs may be suitable for IVUS applications due to their resolution, speed and energy efficiency. Such an ADC may act as an output capacitance of approximately 250 fF at each output rail. In order to fit both LNA and ADC at the backside of a CMUT, it was proposed a maximum LNA size of $100 \times 100\ \mu\text{m}$.

3.2 Amplifier topology

No specific amplifier topology was required for this project in order to utilize all technology benefits and restrictions. One critical design specification was the medium-to-low input impedance required to avoid unwanted coupling between CMUTs. Another benefit would be to keep the amplifier simple (KISS) and robust, as this greatly simplifies design and layout, and usually saves both area and power. Both [20] and [45] demonstrated promising results for the CG-CS topology described in section 2.2.3, even though some serious issues were identified.

Table 3.1: LNA specifications. All parameters are given at f_c unless otherwise specified.

Parameter	40 nm bulk	28 nm FD-SOI
Supply voltage (V_{DD})	0.9 V	1.0 V
Center frequency (f_c)		5 MHz
Bandwidth		5 MHz
Lowest gain setting ($A_{v,min}$)		0 dB
Noise figure (NF)		< 3 dB
Output SNR (SNR_{out})		> 40 dB
Harmonic distortion (HD_2)		< -50 dB
CMUT impedance (Z_{cmut})		10 k Ω \angle -60 $^\circ$
Input impedance (Z_{in})		\approx 0.5 - 5 k Ω
Load capacitance on each rail (C_L)		250 fF
Max area (A_{tot})		100 \times 100 μ m

This topology also maintains the desired input impedance and features a simple behavior.

A differential pair with one fixed input was also considered as a possible topology. That topology consists basically of two common-source stages with a common current source. It features single-ended-to-differential conversion when one of the inputs is fixed, and is an efficient gain stage with good even harmonic canceling. Its input impedance is in principle very high, which was an undesired starting point for this application, even though input impedance could be modified with resistors and/or capacitors at the input. However, the largest drawback was a common-mode feedback circuit that was required in order to get good signal swing at both rails. Such a circuit adds extra complexity and power consumption, which was undesired on a chip that should feature several hundred or thousands of these amplifiers.

Based on the promising results and general simplicity in [20] and [45], a common-gate-common-source topology with modifications was chosen. Final topology is shown in figure 3.1. In the choice between NMOS or PMOS input, NMOS was chosen due to its higher mobility, which made it possible to make small transistors with high g_m . This also implied PMOS loads, which features higher output resistance than NMOS due to lower mobility and was therefore better suited as loads.

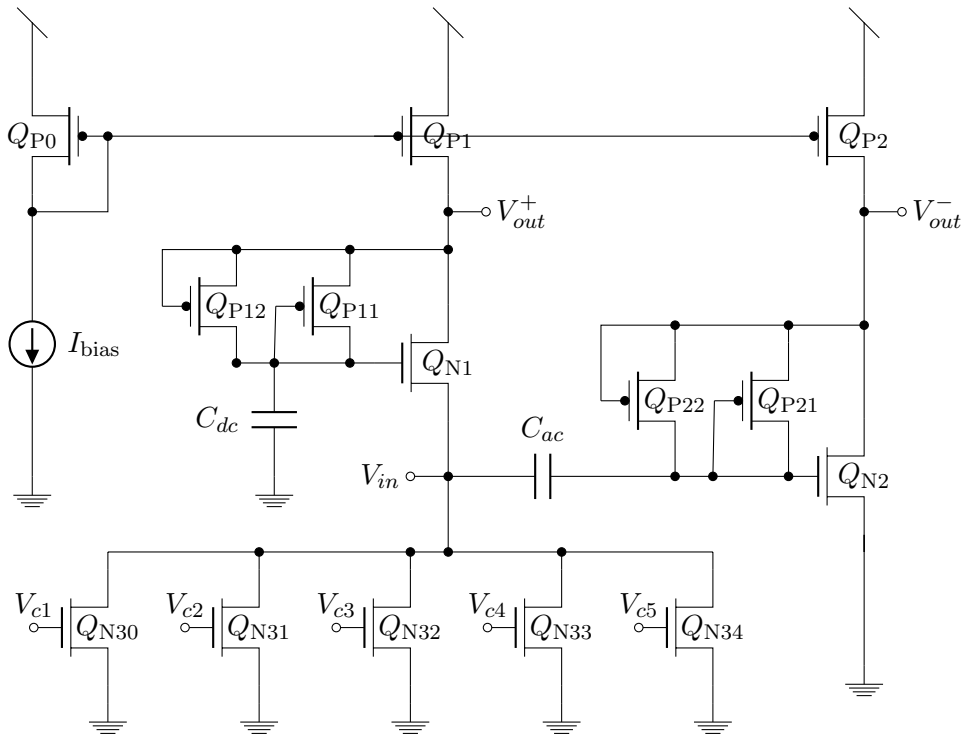


Figure 3.1: Amplifier schematic

3.2.1 Biasing

Both [20] and [45] suffered from bad bias robustness when mismatch between devices was simulated using Monte Carlo, which resulted in bad harmonic distortion due to sub-optimal bias points. This indicated that something clever had to be done with the biasing. Another interesting observation from [45] was that a feedback biased common-source-stage did not suffer from these matching issues. Such dynamic regulation is one of the strengths with feedback control. It was therefore decided to use this technique as bias for both common-source and common-gate in the chosen topology.

Feedback biasing of both stages was performed as described in section 2.2.5, using AC-coupling between V_{in} and CS gate, and a capacitor between CG gate and ground. In order to minimize area consumed by coupling capacitors, moscaps were chosen as capacitors. They offered highest capacitance per area, and the inaccurate and variable capacitance of moscaps did not affect the performance. The most important feature of coupling capacitors is that they are large enough for the

purpose. The necessary capacitance value was also minimized by choosing a large feedback resistance. According to (2.34), does the product of C and R control cutoff frequency. These large feedback resistances were implemented using diode connected subthreshold PMOS transistors. The exact value of these resistors was not important either, as long as they were large enough, and an approximate value of $1.4 \text{ G}\Omega$ turned out to be suitable. This would ensure a very low η even for mega-ohm amplifier loads. In order to ensure good noise cancellation for low frequencies, a high-pass cutoff frequency of 100 kHz was desired. According to (2.34) and (2.21), this frequency would require a coupling capacitance of approximately 20 fF given $1.4 \text{ G}\Omega$ feedback resistance and a gain of 20.

Two diode connected transistors in parallel, with opposite directions, were necessary as feedback resistors since diodes only conduct current one way. During steady-state analysis were all simulations successful with only one diode from drain to gate, but transient testing demonstrated that the feedback was not able to pull current from gate when transient steps were applied. Such steps would occur due to variable DC voltage at V_{in} when gain is varied. A second diode in the opposite direction fixed this issue.

In 40 nm technology, gate area of Q_{N1} and Q_{N2} transistors and their coupling capacitors turned out to be so large that they conducted a significant DC gate current. Parasitic leakage has an undesired effect on feedback biased transistors, especially with large feedback resistors as used here. Even small currents create a significant voltage drop across the resistor when resistance is high. Another issue was that resistance in diode-connected transistors decreases rapidly when current is increased, as stated in (2.47). This was avoided by using 1.8 V IO-transistors as moscaps, and 1.2 V IO-transistors as CS- and CG-transistors. IO-transistors use thicker gate oxide in order to handle higher voltages, and does not have the same amount of gate leakage as regular 0.9 V transistors. 28 nm technology did not show the same amount of gate leakage, even though gate areas was of comparable size, so thin oxide devices could be used for all transistors there.

3.2.2 Variable gain

As described in section 2.2.9, may gain control be implemented by either truncating the signal at input or manipulating g_m and R_{out} . We know from (2.41) that g_m can be manipulated by either changing transistor size or bias current. Transistor length is very hard to change during operation, while transistor width may be altered by switching on and off parallel transistors. A simpler way of changing g_m is usually to alter I_d . However, (2.44) show that increasing I_d actually increases

r_o more than g_m is reduced, so total gain may increase when I_d is reduced. However, according to (2.35), bandwidth is reduced when g_m is decreased. These side effects lead to the conclusion that gain control using variable current, and even variable g_m , was an undesirable solution. It is usually considered a waste to use more current for less gain, and a low-gain setting should be considered a relaxed state with less accuracy than high-gain setting.

Another possible solution was to alter r_{out} , which was formed by a parallel of two r_o 's in the chosen topology. r_o can be changed either by changing length of transistors, which have been proven difficult, or by changing the current according to (2.42), which had undesirable side effects. A third option was then to create controllable shunt resistances at the output node, either between CG and CS output, or from each output to ground. These resistances should have values equal to or lower than r_o in order to decrease gain in practical steps. r_o is usually hundreds of kilo-ohms or larger. Due to low sheet resistance for resistors in available technologies, large linear resistors would have consumed very large area. Such resistors would also have to be switched in and out by transistor switches, as resistor resistance could not be electrically controlled. Several publications were found on making floating variable resistors from transistors that could be connected between CG and CS output, but none was considered sufficient. The last attempt on this strategy was to use transistors in triode region between each output and ground. It turned out to be difficult to get good triode operation and linearity from such transistors due to the high output DC level of 0.5-0.6 V. Triode transistors can only be considered approximately linear as long as V_{ds} is small, according to (2.40).

Truncation of the input signal seemed like a more viable solution. The concept in figure 2.16 already had a resistor, R_S , from input to ground, which featured at least ten times less resistance than r_o in [45]. By decreasing resistance in this resistor, more signal current was sent to ground instead of into the amplifier. DC level at input was also significantly lower than DC at output, which made triode transistors behave more as linear resistors at input. In order to simplify control of resistance in R_S , it was decided to use a control signal that was either on (V_{DD}) or off (ground), instead of a continuously varying bias voltage. Such a large overdrive would maximize transistor linearity according to (2.40). This resulted in a parallel array of binary-scaled triode transistors with separate control signals which had a total resistance of 13 k Ω to 800 Ω for 40 nm, and 20 k Ω to 1 k Ω for 28 nm.

Supply voltage and bias current in CG limited maximum source resistance values. During transistor sizing, 12 and 8 μ A were found to be suitable bias currents for 40 and 28 nm. This current flows through the source resistor and generates a DC voltage at V_{in} . In order to achieve maximum 0.15 V DC voltage at V_{in} , source

resistance was limited to 13 and 20 k Ω . A higher DC voltage at V_{in} would have stolen valuable voltage headroom from the common-gate stage, which was not desirable.

A variable gain strategy that decreases input impedance when signals are strong may be particularly beneficial for CMUTs, as coupling between elements are worse for stronger signals than small signals, according to [11]. By additionally decreasing input impedance for these strong signals, larger amount of coupling are avoided.

3.3 Transistor sizing

In order to minimize FOM_{LNA} , it was necessary to maximize both area efficiency and energy efficiency. A great amount of area savings was achieved in the previous section by using transistors as both resistors and capacitors, even though their size was still relatively large. This made it less crucial to focus on the size of the remaining transistors. As stated in section 2.2.8, may the entire amplifier response be described by g_m , g_m/g_{ds} and C_L . This amplifier had a fixed load capacitance which made it possible to use g_m/I_d and V_{ds} to optimize transistor sizes. g_m/I_d , the current efficiency, was a key to maximize energy efficiency using this approach.

A maximum gain was not defined in the specifications, but SNR requirements demanded a minimum source SNR since SNR can not be *enhanced* through an amplifier. With minimum 40 dB output SNR and maximum 3 dB noise factor, it was necessary to have a source SNR of at least 43 dB. An input SNR of 50 dB was chosen in order to come closer to [20] and make SNR and HD_2 approximately equal. By using the CMUT source model from section 2.1 with the specified impedance and an amplifier input impedance of 5 k Ω , it was possible to calculate a minimum transducer voltage. These calculations are given in appendix B and resulted in minimum transducer voltage, v_{cmut} , of 12.7 mV, which corresponds to $v_{in} = 2.4$ mV. A conservative choice of 50 mV output amplitude per rail was made in order to get good distortion performance. This corresponds to 100 mV swing per rail, and 200 mV differential swing. With an input amplitude of 2.4 mV, this required a gain of 20 at each rail.

In order to make the amplifier as energy efficient as possible, it was decided that upper cutoff frequency should be placed at maximum 10 MHz. With a cutoff frequency of 10 MHz and low frequency gain of 20, the unity-gain frequency became 200 MHz, according to (2.36). With a fixed load capacitance of 250 fF, this required a transconductance of 300 μ S according to (2.35). However, common-gate

input impedance was given partly by g_m , so 200 μS was chosen in order to keep input impedance near 5 $\text{k}\Omega$. This would also ensure that a sufficient signal amount was input to CS, which was a voltage amplifier.

From a LNA point-of-view, (2.48) and (2.50) showed that it was important to have high g_m on input transistors (NMOS) in order to minimize input-referred noise, and low g_m on load transistors (PMOS) in order to minimize noise current. In order to maximize energy efficiency, g_m/I_d was maximized for the input transistors. With a fixed g_m , bias current, I_d , was decided by the final g_m/I_d . The same bias current flowed through the load transistors, and with a lower g_m they got a lower g_m/I_d .

The available technologies suffered from a significant amount of flicker noise, which made it necessary to aim for large transistors areas as well. A high g_m/I_d usually involves a large width, and hence made the input transistors automatically large. The length is dependent on r_o , according to (2.42), which made it natural to specify a high r_o for load transistors in order for them to get a long length and low flicker noise. With a sufficiently high r_o in load transistors, r_{out} depended mostly on r_o in the input transistors. Hence became gain only dependent on g_m/g_{ds} in the input transistors.

Width- and length-optimization was performed using the *mosdesigner* application made by prof. Ytterdal. This application is a front-end for an Eldo optimizer test bench. g_m , g_m/g_{ds} , g_m/I_d , V_{ds} and V_{sb} are taken as input and used as goals in a DC optimization on the appropriate transistor model. This drastically decreased the time used for finding appropriate transistor widths and lengths, and g_m/I_d could be tweaked in order to maximize the current efficiency. At some point got the specified g_m/I_d higher than what the technology was able to perform, and the optimizer failed. This was considered as maximum current efficiency.

It turned out that different 28 nm devices featured substantially different noise performance. Low-threshold (LVT) NMOS devices featured up to twice as much flicker noise than regular-threshold (RVT) NMOS devices when g_m and I_d were equal. PMOS devices demonstrated opposite behavior. This made it necessary to use RVT NMOS devices and LVT PMOS devices throughout the circuit.

Final properties and dimensions for all transistors are given in table 3.2 and 3.3. R_S , R_f and $C_{ac/dc}$ transistors were not dimensioned using g_m/I_d -method, but using a manual DC-sweep test bench which modeled their regular environment. This task was performed with a focus on low footprint and noise.

A low V_{ds} of 0.35 V was used on CG and CS transistors to ensure at least some signal swing. With only 0.9 V supply voltage in 40 nm, 0.35 V was not far from

Table 3.2: 40 nm transistor values and dimensions. Total width is $W_{tot} = M \cdot W$. Names refer to components in figure 3.1.

Name	Device	g_m [μS]	g_m/g_{ds} []	g_m/I_d [1/V]	V_{ds} [mV]	V_{sb} [mV]	W [μm]	L [μm]	M
Q_{N1}	NMOS 1.2V	200	20	18.2	0.35	0.15	1.5	0.12	2
Q_{N2}	NMOS 1.2V	200	20	18.2	0.35	0	1.5	0.12	2
Q_{P1}	PMOS	200	20	18.2	0.35	0	0.81	0.21	2
Q_{P2}	PMOS	200	20	18.2	0.35	0	0.81	0.21	2
Q_{P0}	PMOS	100	10	18.2	0.35	0	0.81	0.21	1
Q_{P11}	PMOS			$R = 1.4 \text{ G}\Omega$			0.24	0.08	1
Q_{P12}	PMOS			$R = 1.4 \text{ G}\Omega$			0.24	0.08	1
Q_{P21}	PMOS			$R = 1.4 \text{ G}\Omega$			0.24	0.08	1
Q_{P22}	PMOS			$R = 1.4 \text{ G}\Omega$			0.24	0.08	1
Q_{N30}	NMOS			$R_{on} = 13 \text{ k}\Omega$			0.36	0.9	1
Q_{N31}	NMOS			$R_{on} = 6.5 \text{ k}\Omega$			0.36	0.45	1
Q_{N32}	NMOS			$R_{on} = 3.25 \text{ k}\Omega$			0.36	0.225	1
Q_{N33}	NMOS			$R_{on} = 1.63 \text{ k}\Omega$			0.36	0.11	1
Q_{N34}	NMOS			$R_{on} = 1.63 \text{ k}\Omega$			0.36	0.11	1
C_{ac}	NCAP 1.8 V			$C = 34 \text{ fF}$			2	3	1
C_{dc}	NCAP 1.8 V			$C = 34 \text{ fF}$			2	3	1

what was necessary to only bias the CG transistors. Source resistance stole 0.15 V of these 0.9 V, leaving 0.75 V to the amplifier stage. A bias point centered in this range would require minimum $0.75/2 = 0.375 \text{ V}$ over each transistor, which is not far from 0.35 V. With 1 V supply voltage in 28 nm, this headroom was increased to 0.425 V, which enabled larger signal swing. To make CS and CG stages as equal as possible was 0.35 V chosen as V_{ds} for all transistors.

Some transistors were not designed according to the optimal procedure previously described. In 40 nm it was difficult to design PMOS loads (Q_{P1} and Q_{P2}) with high g_m/g_{ds} and low g_m . This forced g_m to be increased and g_m/g_{ds} to be decreased. A low g_m implied low g_m/I_d , which resulted in a high overdrive that was difficult to use with low supply voltage. Because of these trade offs, it was decided to use equal parameters for both driver and load transistors in 40 nm, as indicated in table 3.2. Another discrepancy was the CS driver transistor in 28 nm (Q_{N2}), which had to be made eight times larger than Q_{N1} because of flicker noise. This was done by increasing both width and length, albeit width was increased twice as much as length.

Table 3.3: 28 nm transistor values and dimensions. Total width is $W_{tot} = M \cdot W$. Names refer to components in figure 3.1.

Name	Device	g_m [μS]	g_m/g_{ds} []	g_m/I_d [1/V]	V_{ds} [mV]	V_{sb} [mV]	W [μm]	L [μm]	M
Q_{N1}	NMOS	200	40	25	0.35	0.15	1.92	0.066	1
Q_{N2}	NMOS	200	40	25	0.35	0	1.91	0.132	4
Q_{P1}	LVT PMOS	40	80	5	0.35	0	1.80	2.0	1
Q_{P2}	LVT PMOS	40	80	5	0.35	0	1.80	2.0	1
Q_{P0}	LVT PMOS	40	80	5	0.35	0	1.80	2.0	1
Q_{P11}	LVT PMOS			$R = 1.4 \text{ G}\Omega$			0.20	0.038	1
Q_{P12}	LVT PMOS			$R = 1.4 \text{ G}\Omega$			0.20	0.038	1
Q_{P21}	LVT PMOS			$R = 1.4 \text{ G}\Omega$			0.20	0.038	1
Q_{P22}	LVT PMOS			$R = 1.4 \text{ G}\Omega$			0.20	0.038	1
Q_{N30}	NMOS			$R_{on} = 20 \text{ k}\Omega$			0.40	1.1	1
Q_{N31}	NMOS			$R_{on} = 10 \text{ k}\Omega$			0.40	1.1	2
Q_{N32}	NMOS			$R_{on} = 5 \text{ k}\Omega$			0.40	1.1	4
Q_{N33}	NMOS			$R_{on} = 2.5 \text{ k}\Omega$			0.40	0.55	4
Q_{N34}	NMOS			$R_{on} = 1.25 \text{ k}\Omega$			0.40	0.275	4
C_{ac}	NCAP			$C = 104 \text{ fF}$			4	4	1
C_{dc}	NCAP			$C = 80 \text{ fF}$			4	3	1

3.4 Layout

Layouts were made for both amplifiers in order to extract parasitics and have an extra performance verification. In [45] it turned out that models behaved quite differently in schematic and layout. This time surprises were avoided by verifying smaller pieces of layout during schematic development. The development kit for 40 nm also had Design-for-Manufacturing-options which were used to ensure that standard devices would behave well in an analog application. These options increased surrounding limits around the transistors in order to avoid some nanoscale effects.

The layouts were made according to design rules given by the manufacturer, and passed both Design Rule Checking (DRC) and Schematic vs Layout (LVS). When placing and routing the components, focus was put on making a compact and robust circuit with little routing parasitics. Regularity was also considered important in order to achieve good matching. Resulting layout is shown in figure 3.2 and 3.3. Calibre PEX from Mentor Graphics and Assura QRC from Cadence were used to extract parasitics from these layouts.

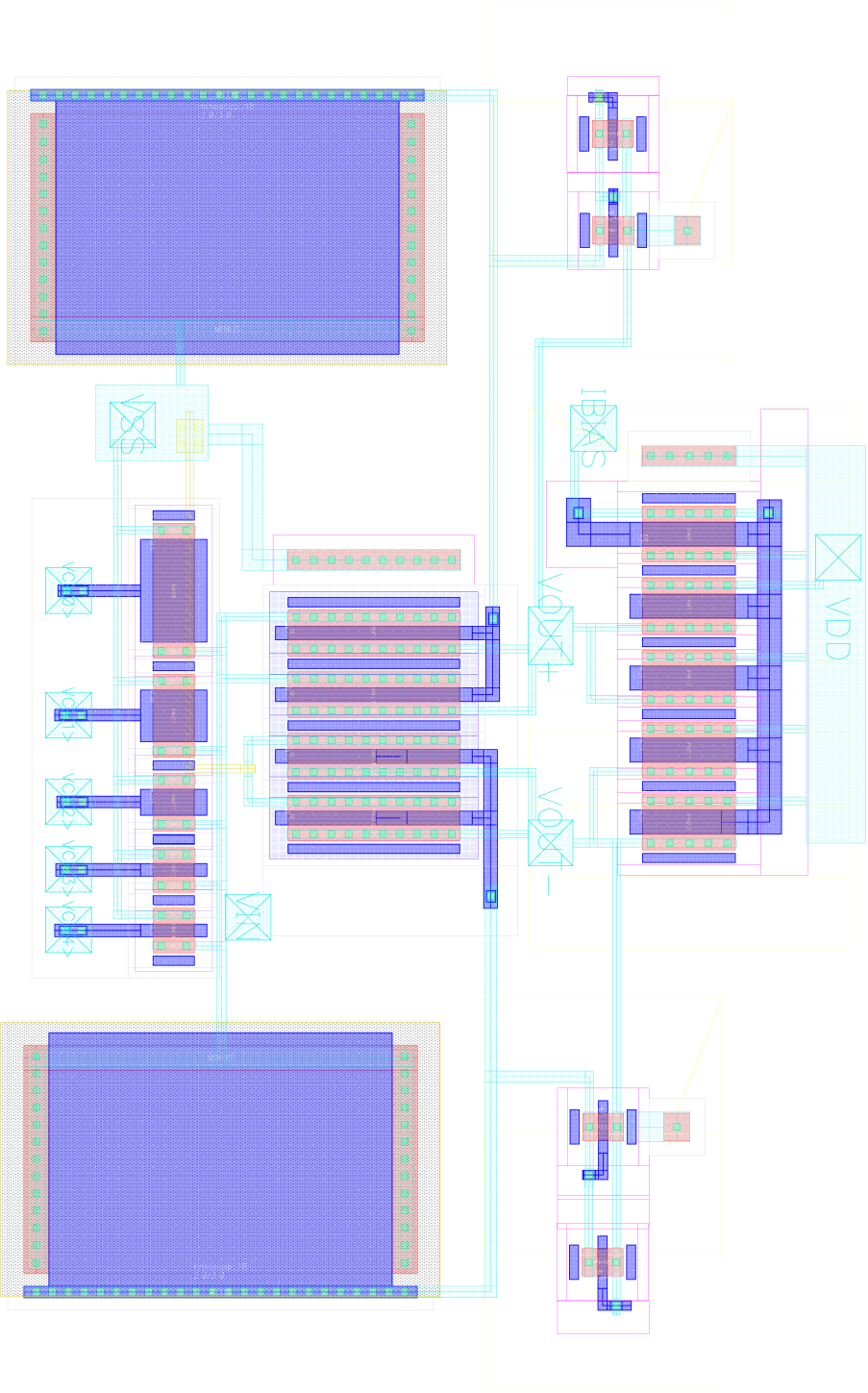


Figure 3.2: Amplifier layout for 40 nm technology. Blue = Poly, Cyan = Metal 1, Yellow = Metal2, Red = OD, Pink = PP, Grey = NP

Chapter 4

Results

This chapter presents achieved simulation results and the methods used to verify these results. Design was performed using Cadence Virtuoso and simulated with Spectre. Final results were obtained from post-layout circuits, in order to present as realistic performance as possible. To make sure the design was robust against mismatch, all results were obtained from 200 Monte Carlo mismatch-iterations where this was possible. Results presented here is an average of these iterations. Every result was verified by comparing the result to calculations on an appropriate transient run.

A test bench featuring CMUT source, supply voltage, bias current, input- DC-block, gain logic and output load was created in order to perform simulation on different amplifiers in an efficient manner. V_{out}^+ and V_{out}^- was subtracted using an ideal component in order to analyze the differential signal $V_{out} = V_{out}^+ - V_{out}^-$. The simplified CMUT model in figure 2.4b was used as source for the amplifier, and two 250 fF capacitors were used as output loads. This test bench is shown in figure C.1. Component values for the CMUT model is given in table C.1. Different CMUT voltages were used with the two amplifiers due to their different gain and distortion. v_{cmut} was adjusted to a good compromise between SNR and HD_2 . Test bench parameters are given in table 4.1. All tests were run at maximum gain since this gave best noise figure.

Table 4.1: Test bench parameters.

Parameter	40 nm bulk	28 nm FD-SOI
Supply voltage (V_{DD})	0.9 V	1.0 V
Bias current (I_{bias})	6 μ A	8 μ A
CMUT signal (v_{cmut})	15.4 mV	12.0 mV
Center frequency (f_c)	5 MHz	
Load capacitance (C_L)	2 \times 250 fF	
Gain control signal ($V_c<1:5>$)	V_{DD} , Code = (1 0 0 0 0)	
CMUT impedance (Z_s)	10 k Ω \angle - 60 $^\circ$	

An overview of final results is given in table 4.2. The following sections describe these parameters and how they were simulated.

Table 4.2: Final results for both amplifiers.

Parameter	40 nm bulk	28 nm FD-SOI	Specification
Amplifier bandwidth	12.8 MHz	8.4 MHz	5 MHz
Small-signal gain (A_v)	12.7 dB	16.8 dB	-
Lowest gain setting ($A_{v,min}$)	-5.1 dB	-1.4 dB	< 0 dB
Output SNR (SNR_{out})	40.66 dB	39.1 dB	> 40 dB
Noise figure (NF)	10.54 dB	9.26 dB	< 3 dB
Harmonic distortion (HD_2)	-47.9 dB	-48.7 dB	< -50 dB
Max input impedance (Z_{in})	5.6 k Ω	4.7 k Ω	\approx 0.5 - 5 k Ω
Physical size (A_{tot})	11 \times 8 μ m	15 \times 10 μ m	< 100 \times 100 μ m
Power consumption	20.5 + 5.4 μ W	16.0 + 8 μ W	Minimize

Power consumption was simulated using DC and transient RMS analysis, and distinguished between amplifier power and bias branch power. This decision was made because bias voltage generated by the current mirror was always constant and could in principle be shared by multiple amplifiers in an array. The necessary current required for biasing could also be significantly reduced using M-factor scaling, which was only performed in 40 nm here. Hence, it was more interesting to measure power consumed by only the amplifier. DC and transient simulation results were very equal, which made it possible to use only DC simulation as a reliable result.

Physical size was measured on final layout shown in figures 3.2 and 3.3, and included everything in these figures.

4.1 Frequency response

Bandwidth was simulated using both small-signal (`ac`) and periodic large-signal frequency analysis (`pss` with `pac`), which gave approximately equal results. Since the source model was only valid at f_c , it was not possible to use the entire system for bandwidth calculation, as v_{in} would be incorrect for all other frequencies than f_c . Bandwidth calculation was therefore restricted to the amplifier itself, which was independent of source impedance. Gain at f_c was used as bandwidth reference point, and the transfer function v_{out}/v_{in} was used to find the -3 dB range around this reference. Figure 4.1 compares small-signal transfer functions of 40 and 28 nm. Included in the figure are also plots of ideal transfer functions from section 2.2.3 and appendix A using intended transistor properties from table 3.2 and 3.3.

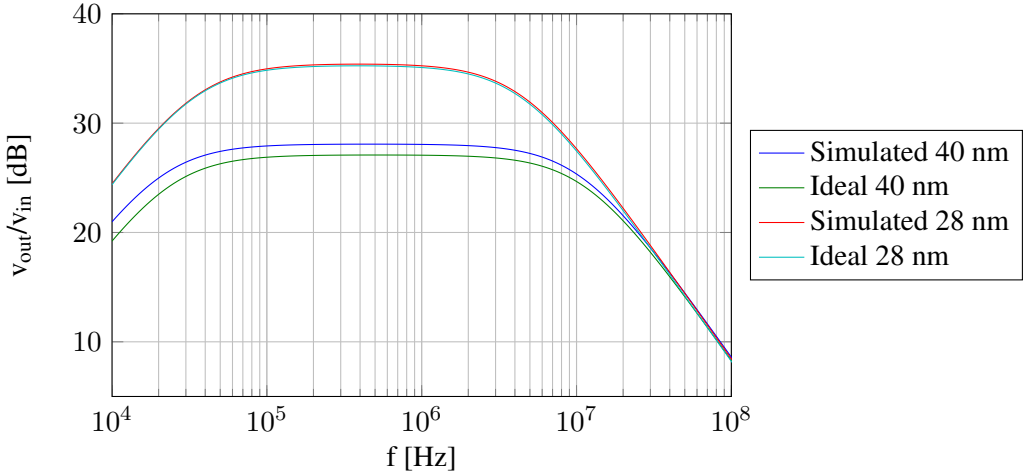


Figure 4.1: Source independent small-signal frequency transfer functions for both amplifiers. Ideal plots are derived from ideal models in section 2.2.3 and appendix A.

With a valid source impedance at f_c it was, possible to simulate gain at this frequency. The gains presented in table 4.2 were therefore calculated from v_{out}/v_{cmut} , and describes gain for the overall system. This was also performed using both small-signal and large-signal AC-analysis. From these simulations, input impedance was calculated using (2.8). Input impedance was evaluated for five different gain settings and is shown in figure 4.2. How this varying Z_{in} affected v_{in}/v_{cmut} and v_{out}/v_{in} are shown in figure 4.3. Only 40 nm is shown in figure 4.3, as 40 and 28 nm behaved equally.

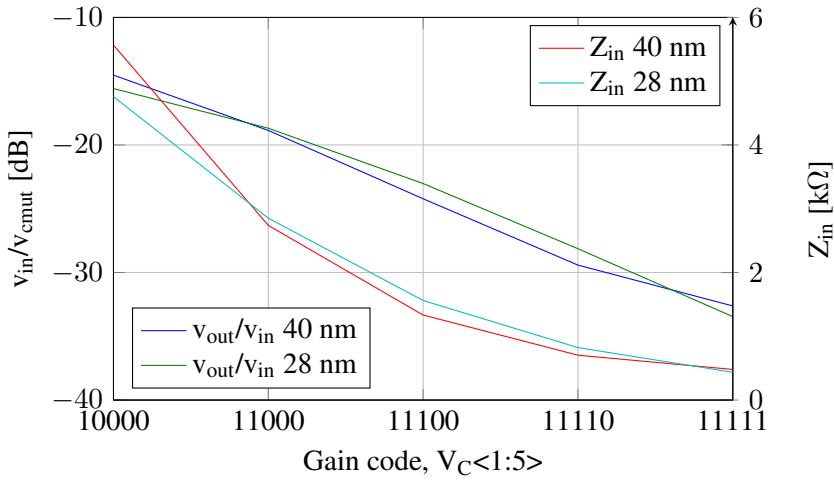


Figure 4.2: Input loss (v_{in}/v_{cmut}) and input impedance simulated for different gain settings at f_c .

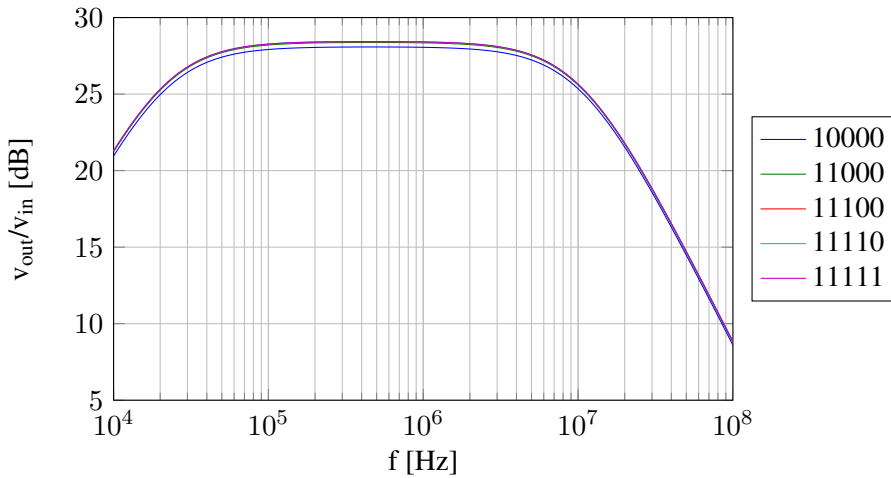


Figure 4.3: Small-signal transfer functions for 40 nm with different gain settings.

4.2 Noise simulation

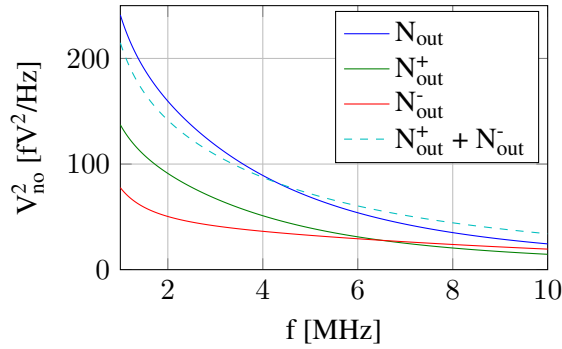
Output SNR was calculated using output signal level at f_c and integrated noise in the specified bandwidth of 5 MHz around f_c , as stated in (2.10). Small-signal AC analysis and small-signal noise analysis (`noise`) were used to find signal and

noise values respectively. Noise analysis also was able to extract a summary of each transistor's noise contribution to the final output noise. The most significant noise sources, measured at f_c , are shown in table 4.3.

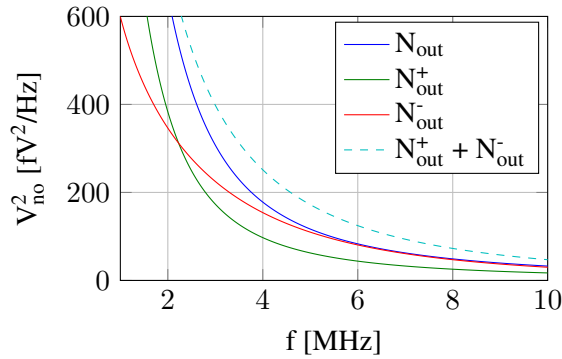
Table 4.3: Summary of significant spot noise sources at f_c . Noise types: R = resistive noise, T = transistor thermal noise, F = transistor flicker noise

Device	Type	$V_{no}^2(f_c)$ [fV ² /Hz]					
		40 nm			28 nm		
		V_{out}	V_{out}^+	V_{out}^-	V_{out}	V_{out}^+	V_{out}^-
R_{cs}	R	6.2	1.9	1.3	14.3	3.8	3.3
Q_{N31}	F	0.4	0.1	0.1	3.2	0.9	0.7
Q_{N31}	T	10.4	3.1	2.1	24.3	6.5	5.5
Q_{N1}	F	0.3	1.6	0.5	0	24.6	29.4
Q_{N1}	T	0	1.8	0.6	0	5.5	6.6
Q_{P1}	F	1.6	0.9	0.1	1.3	0.9	0
Q_{P1}	T	18.0	10.2	1.5	14.0	9.4	1.3
Q_{N2}	F	5.6	0	5.4	16.4	0	16.3
Q_{N2}	T	6.0	0	5.8	25.8	0	25.8
Q_{P2}	F	0.6	0	0.6	0.6	0	0.6
Q_{P2}	T	5.4	0	5.3	7.1	0	7.0
P_0	F	1.3	1.9	0.8	0.3	0.9	0.7
P_0	T	11.6	17.5	7.5	6.0	9.2	8.0
Total		68.7	39.5	32.5	116.8	62.5	109.7

Figure 4.4 show output noise power for different rails as a function of frequency. The ideal noise signal $N_{out}^+ + N_{out}^-$ was added as a comparison to N_{out} .



(a) 40 nm



(b) 28 nm

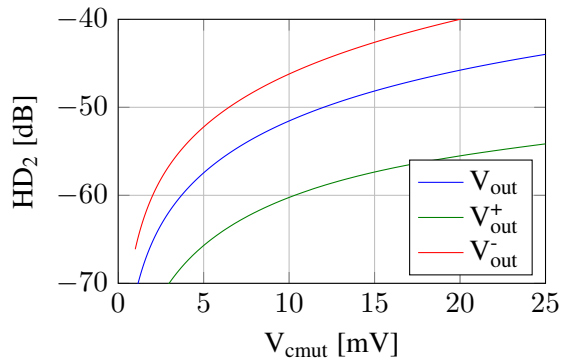
Figure 4.4: Output noise power for all rails, including an ideal noise sum for reference.

According to (2.9), may noise factor be calculated by comparing output noise of a noiseless amplifier to output noise of a noisy amplifier. This was done by performing the previously mentioned SNR calculations with and without amplifier noise. Spectre has the ability to disable noise in specific components, which made it easy to disable amplifier noise in the test bench.

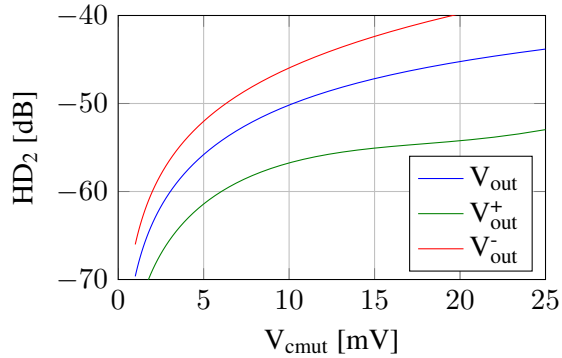
4.3 Large-signal behavior

Harmonic balance (hb) is an efficient tool to simulate large-signal steady-state behavior of a circuit. A specified number of harmonic components is fitted to the circuit in order to obtain a steady-state solution, and hence is harmonic content extracted automatically. This was used to simulate HD₂ which, according to (2.6), is given by first- and second-harmonic content. Harmonic balance was run with a

single-tone input as specified in table 4.1. In order to make the simulation more accurate were four harmonic frequencies extracted with an oversample rate of two. Figure 4.5 show HD_2 as V_{cmut} was swept over a reasonable range.



(a) 40 nm



(b) 28 nm

Figure 4.5: HD_2 simulated using harmonic balance for various values of V_{cmut} .

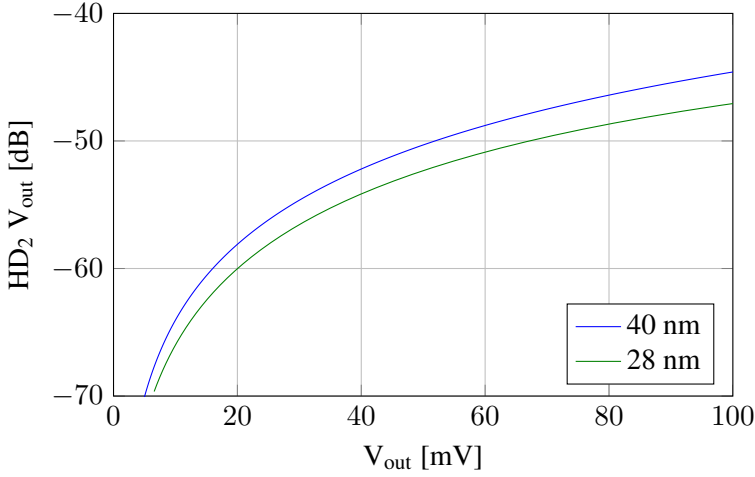


Figure 4.6: V_{out} HD_2 simulated using harmonic balance on both amplifiers for various values of V_{out} .

4.4 Transient behavior

In order to simulate real world behavior, it was created a transient test using varying input amplitude and varying amplifier gain. This gave a visual impression of how these amplifiers would behave during real operation. A received pulse was modeled as

$$V_{cmut}(t) = V_0 e^{-\frac{t}{\tau}} \sin(2\pi f_c t) \quad (4.1)$$

where $V_0 = 240$ mV and $\tau = \frac{20}{f_c} = 4$ μ s. This signal was an approximation to a received ultrasound echo. Gain was increased every two microsecond by using the five gain settings: 11111, 11110, 11100, 11000 and 10000. Transient runs with and without variable gain in 28 nm is given in figure 4.7 and 4.8 respectively. 40 nm is not shown since it behaved equally.

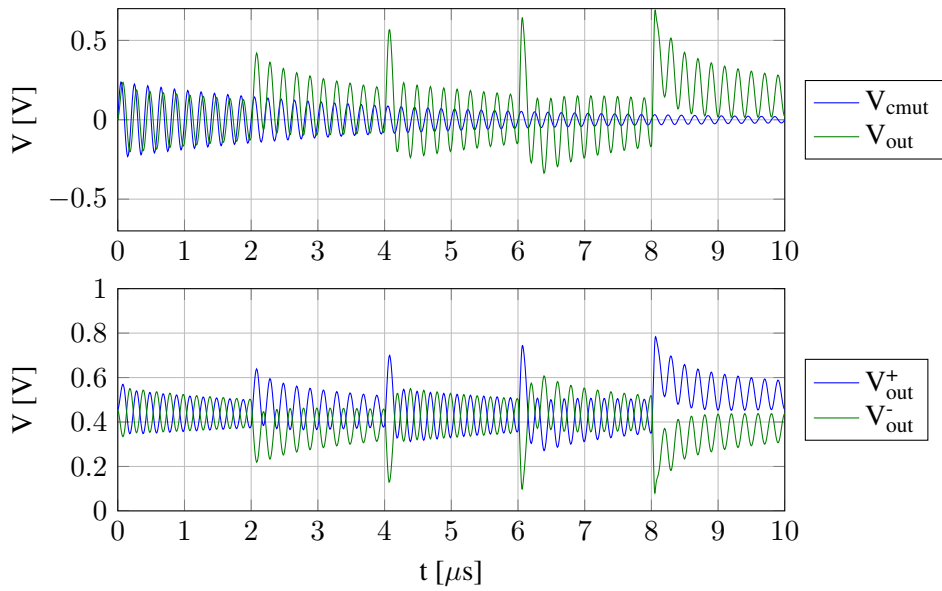


Figure 4.7: Transient run on 28 nm with variable gain.

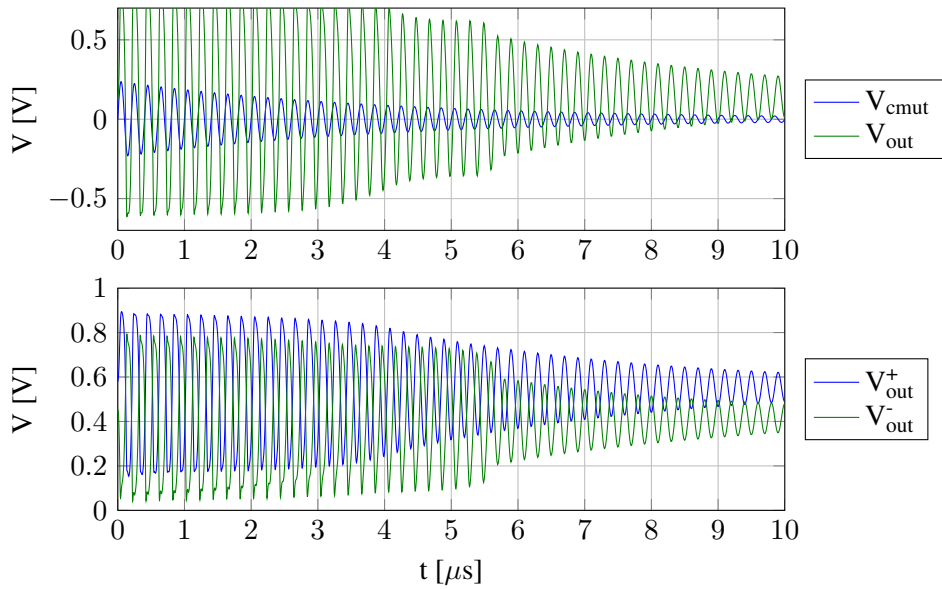


Figure 4.8: Transient run on 28 nm without variable gain.

4.5 Monte Carlo variations

Variability of the two technologies was analyzed using monte carlo simulations. Only mismatch variations were included in the analysis in order to evaluate biasing robustness. Table 4.4 summarizes mean (μ) and standard deviation (σ) values for the most important properties. The ratio σ/μ normalizes standard deviation in terms of μ , which make it possible to compare variability on parameters featuring different μ 's.

Table 4.4: Mean and standard deviation values from 200 monte carlo mismatch simulations

Parameter	40 nm			28 nm		
	μ	σ	$\frac{\sigma}{\mu}$ [%]	μ	σ	$\frac{\sigma}{\mu}$ [%]
v_{out}/v_{cmut} [dB]	12.7	0.1	0.6	16.4	0.2	1.1
SNR_{out} [dB]	40.7	0.1	0.1	39.1	0.0	0.1
HD_2, V_{out} [dB]	-47.9	2.2	-4.5	-48.7	0.7	-1.5
HD_2, V_{out}^+ [dB]	-55.1	4.4	-8.0	-55.8	1.9	-3.5
HD_2, V_{out}^- [dB]	-42.4	0.3	-0.8	-44.2	0.3	-0.7
DC, V_{out}^+ [mV]	576	20	3.4	581	5	0.8
DC, V_{out}^- [mV]	432	64	1.5	456	3	0.7
I_d^+ [μ A]	10.8	0.9	8.2	8.0	0.1	1.1
I_d^- [μ A]	12.0	1.1	9.1	8.0	0.1	1.1

Chapter 5

Discussion

This chapter sums up and discusses notable observations and experiences from the previously described design phase and simulated results.

5.1 Technology differences

A g_m/I_d -based design approach proved to be an efficient and exact way of designing transistors. It significantly reduced time spent on sizing transistor widths and lengths, and also brought power consumption down to a minimum level. The two amplifiers were designed slightly different as it was difficult to achieve low- g_m/I_d loads in 40 nm. Actually, 40 nm bulk had an overall tendency to be less manageable than 28 nm in terms of biasing. One evident observation was 40 nm CG which conducted 10% less current than intended. This was caused by the limited supply voltage which did not give enough headroom for a sufficient V_{ds} . CS lacked the source resistor which stole approximately 0.15 V in CG, and hence had more headroom which made biasing more accurate than in CG. This indicated that 40 nm transistors were driven close to their saturation limit and that supply voltage may have been close to the practical limit for simple analog designs.

FD-SOI on the other hand, demonstrated excellent bias accuracy. 28 nm current mirrors conducted the desired current in almost all configurations. Monte carlo results showed minimum and maximum currents within $\pm 3\%$ from the desired current. A standard deviation of 0.1 μA in FD-SOI current can be considered way better than 1 μA which was demonstrated in 40 nm bulk. An extra 0.1 V supply

voltage may have contributed to this, as it gave more voltage headroom. But the process in itself was probably the most significant difference. As described in section 2.3 was FD-SOI an attempt to avoid most unwanted nanoscale effects that appears in bulk processes. Especially the homogeneous and ultra-thin body makes the channel easily saturated. At least in terms of bias accuracy may it be safe to conclude that FD-SOI actually is more accurate than bulk. This in turn led to overall better variability on 28 nm for all simulated properties. Such results are a good basis for robust designs.

Also in terms of other transistor properties such as g_m , g_m/g_{ds} and g_m/I_d , FD-SOI seemed a more flexible technology than bulk. g_m/I_d could be pushed higher for a given g_m , leading to lower power consumption. It was easier to achieve high g_m/g_{ds} which directly affected gain. Especially achievable g_m/g_{ds} is a property that has been decreasing with smaller technology nodes, and it is therefore remarkable to achieve better g_m/g_{ds} in a smaller technology node. Most of these property differences are probably due to better channel control in FD-SOI.

As mentioned in chapter 3, regular 40 nm transistors suffered from a significant gate leakage. Gate conductance was estimated to approximately $30 \text{ n}\mathcal{U}/(\mu\text{m})^2$ on 40 nm NMOS transistors during normal operation. Compared to 28 nm and 1.2V 40 nm transistors which were approximated to $15 \text{ p}\mathcal{U}/(\mu\text{m})^2$, regular 40 nm was 2000 times worse than these. This ratio applied to both NMOS and PMOS, and may indicate that this 40 nm technology did not feature proper high- κ gate oxide or that these oxides have been significantly improved since 40 nm was released.

5.2 Noise performance

As noise contributions in table 4.3 showed, CG-CS demonstrated topology excellent noise cancellation of noise from Q_{N1} in both technologies. Significant amounts of flicker and thermal noise from Q_{N1} appeared on both rails and were completely canceled after subtraction. This was consistent with results presented in [44], even though that solution was not based on AC-coupling between the stages. However, it may be interesting to point out that noise from Q_{P1} was constructively subtracted the same way as input signal and input noise was. This mechanism caused Q_{P1} to contribute with more noise than it actually produced, as explained in section 2.2.7. Fortunately, the coupling between Q_{P1} and CS was not as strong as between Q_{N1} and CS due to the resistance in Q_{N1} . Noise contribution from Q_{P1} at V_{out}^- was therefore only 10% of what it was at V_{out}^+ . Nevertheless, this emphasized the necessity of minimizing load noise. It was also observed that low- g_m loads in the 28 nm amplifier dominated total noise power less than high- g_m

loads in 40 nm. In 40 nm, PMOS noise represented 56% of total noise, compared to only 25% in 28 nm. Because 40 nm was dominated by PMOS noise, figure 4.4 showed only small signs of cancellation in 40 nm compared to 28 nm, which was dominated by NMOS noise. At f_c , 28 nm V_{out} noise was approximately equal to V_{out}^- , which indicated that almost all V_{out}^+ noise was canceled.

For input noise sources, coupling to outputs was good, which made noise from Q_{N3x} generate a fair amount of noise that was also constructively subtracted. Such noise sources at the *input* of a LNA are usually considered bad practice as they degrade the signal independent of LNA gain and noise. In this case, 16 and 23% (40 and 28 nm respectively) of total output noise was generated by this resistance. But the chosen topology was completely dependent on its presence. In RF-applications, this source resistance is usually replaced with an inductor which are ideally noiseless and still feature high impedance at high frequencies. In CMOS is this difficult to achieve at low frequencies, and can be considered almost impossible when the amplifier should be made as compact as the one proposed here.

Unfortunately, 28 nm suffered from large amounts of flicker noise at these frequencies. At V_{out}^+ , flicker noise from Q_{N1} accounted for 40% of total noise power, but was fortunately canceled by CG-CS. Noise sources in CS were however not canceled in any way. Hence, CS required better noise performance than CG in order to not dominate the final result. This was the reason Q_{N2} in 28 nm was made eight times larger than its original size. Gate area of Q_{N2} in 28 nm was 2.8 times larger than in 40 nm, and its flicker noise still accounted for 14% of total noise. In comparison, flicker noise from Q_{N2} accounted for 8% of total noise in 40 nm. The increased transistor width in 28 nm also contributed to higher thermal noise due to higher g_m . In total, Q_{N2} became responsible for 36% of total output noise power in 28 nm, compared to only 17% in 40 nm. It could also be observed from figure 4.4 that noise corner for flicker noise appeared near the desired frequency band in both technologies.

A literature study was performed in order to locate any hints about flicker noise in FD-SOI processes compared to bulk. Some papers pointed on the back gate and isolated channel as an extra source for flicker-like noise, but no measurements concluded that FD-SOI should be worse than bulk [46]. This may indicate that noise models in 28 nm are a bit pessimistic, as it is a fairly new technology for analog purposes.

Even though 28 nm made 70% more noise than 40 nm, extra gain in 28 nm resulted in 1.28 dB better noise figure. Noise figures were albeit far from the desired specification of 3 dB. Several issues contributed to this result. Figure 4.2 shows how a combination of high source impedance and low input impedance resulted

in a significant voltage loss from v_{cmut} to v_{in} . The chosen topology required an input voltage in order to generate signal swing at CS, which was a voltage amplifier. Such a signal loss made amplifier noise more prominent than CMUT noise in final SNR and noise figure measurements. It was a decision to have approximately equal signal swing at both rails, and voltage loss was the consequence of this choice when it had to be combined with low fairly low input impedance. Another strategy could have been to decrease input impedance in CG further, and let the amplifier be a transimpedance amplifier which converts source current to an output voltage. As stated in (2.16), input impedance could have been decreased by increasing g_m . This would have severely decreased V_{out}^- signal swing, but the rail could still have been used for noise cancellation and been optimized for cancellation instead of signal amplification.

Large noise sources like Q_{N2} in 28 nm, PMOS loads in 40 nm and Q_{N3x} in general also contributed to make noise figure additionally worse. Both Q_{N2} and PMOS load noise could probably be reduced by further amplifier tweaking. 28 nm demonstrated that low-noise PMOS loads were possible, and 40 nm transistors featured significantly less flicker noise than 28 nm. One possibility could be to swap NMOS and PMOS transistors. This would increase input transistor sizes due to PMOS' lower mobility, and probably reduce flicker noise due to increased area. NMOS load transistors would also have to be larger than PMOS if same output resistance should be maintained, as NMOS has higher mobility than PMOS. Q_{N3x} noise however, may be difficult to reduce without changing topology or mode of operation, and cannot be replaced by an inductor either.

5.3 Frequency response

As the two amplifiers were designed using slightly different parameters, their frequency responses became different as well. The main difference was output resistance which became higher in 28 nm than 40 nm. A higher output resistance with equal transconductance led to higher low frequency gain, but also lower cut-off frequency, as stated in (2.13) and (2.34). As the frequency sweep continued far beyond cutoff frequency, the response was only dependent on g_m and C_L . This was clearly demonstrated in figure 4.1 where 28 nm got higher maximum v_{out}/v_{in} than 40 nm, but from 30 MHz and beyond the difference was negligible. A disadvantage with this 28 nm response was that 5 MHz ended up far into the roll-off region, which resulted in poorer gain flatness and bandwidth. Despite this little drawback, both amplifiers achieved sufficient bandwidth according to the specification. 40 nm had sufficient gain even beyond 10 MHz which is used for harmonic

imaging.

Another distinctive feature of these frequency responses was the high-pass characteristic at lower frequencies. Lower cutoff frequency was decided to appear around 100 kHz, and R_f , $C_{ac/dc}$ and gain was dimensioned according to this using (2.25), (2.21) and (2.34). Simulations seemed to match well with these calculations. The ideal model plotted together with simulated results did also match very well, even though it was only based on simplified linear models described in section 2.2.5. A small deviation was found at 40 nm, which probably came from an inaccurate bias point. 40 nm achieved a little higher r_o than projected due to lower bias current, hence achieving a little higher low frequency gain than the ideal response. Another mechanism that could lead to differences is body effect which was not modeled in the ideal case, and has a tendency to increase gain in common-gate stages. CG was subject to body effect due to its non-zero source voltage. FD-SOI is less affected by non-zero source voltages due to its isolated channel. Extra gain in 28 nm made it necessary to make 28 nm coupling capacitors larger than 40 nm due to the Miller effect. This resulted in extra area consumption in 28 nm.

From figure 4.1 it also seemed clear that these amplifiers had no parasitic poles anywhere near the dominant pole created by r_{out} and C_L . This indicated very small parasitics in both transistors and routing, and was further confirmed by the good match with ideal models. Small parasitics was one of the goals during layout creation and these results may indicate that this goal was achieved. Unity-gain frequency of 120 MHz at each rail was also consistent with theoretical assumptions, and matched the designed g_m and C_L .

5.4 Linearity

Even though the chosen topology boasted good opportunities for second harmonic cancellation, this mechanism seemed absent in HD₂ results. HD₂ on V_{out}^+ became 11 to 13 dB better than on V_{out}^- . Final V_{out} was obviously limited by V_{out}^- distortion, and got only 5-6 dB better than that rail. 5-6 dB improvement was primarily a result of doubled output amplitude from differential signaling. More distortion at V_{out}^- than V_{out}^+ was expected since V_{out}^- received some distortion from V_{out}^+ in addition to its own distortion. As stated in section 2.2.7, this common distortion was supposed to be canceled if it originated from Q_{N1} . It turned out that a significant amount of this distortion originated from Q_{P1} instead. As proved in section 2.2.7, noise and distortion from Q_{P1} are summed when subtracted, and hence not canceled. This was confirmed by replacing Q_{P1} with an ideal load, which resulted in better V_{out} HD₂ than both V_{out}^+ and V_{out}^- , even though V_{out}^- was still 10 dB

worse than V_{out}^+ . This result indicated that more effort should have been put in minimizing load distortion.

Gain and voltage swing in the two stages was designed to be approximately equal in order to generate equal amounts of harmonic distortion. Since even harmonics always have similar signs due to squaring, as described in section 2.2.1, it was expected to observe some cancellation from this mechanism as well. This mechanism was difficult to distinguish from CG-CS-cancellation, but the excellent cancellation when using an ideal CG-load may indicate that also this mechanism contributed.

Figure 4.6 showed a constant HD_2 difference between 40 and 28 nm when measured at equal output levels, with 28 nm having 2 dB less distortion. According to theory, low- g_m/I_d loads would have higher overdrive and hence higher saturation voltage, $V_{d,sat}$, than high- g_m/I_d loads. 40 nm loads was implemented with higher g_m/I_d than 28 nm which, in principle, would make 40 nm better suited as high-linearity loads. When we assumed that both amplifiers were limited by distortion from PMOS loads and still observe that 40 nm get 2 dB worse distortion, this correlate poorly with theoretical assumptions. An explanation may be that 28 nm demonstrated overall better bias and saturation properties than 40 nm, which make it difficult to use g_m/I_d as a direct comparison. Lower supply voltage may also play a role.

Despite the lack of cancellation, it was still achieved good harmonic distortion levels combined with an acceptable SNR. By improving CG load distortion could this result be made even better.

5.5 Energy efficiency

Both amplifiers achieved lower power consumption than any known IVUS LNA attempts, but as described in section 2.4, energy efficiency must be considered in order to make a fair comparison. Table 5.1 compares energy efficiency for amplifiers designed here with previous work in [20]. The functionality of these three amplifiers was approximately equal, which made it possible to compare them without any modifications. FOM and FOM_{LNA} were calculated using (2.52) and (2.55). Dynamic range in (2.55) was calculated by combining output SNR and HD_2 according to (5.1).

$$DR^2 = \frac{1}{\frac{1}{SNR} + HD_2} \quad (5.1)$$

Table 5.1: Energy efficiency for three different amplifiers

	[This work]		[20]
Lithography	40 nm	28 nm	65 nm
Power consumption [μW]	20.5	16.0	67.2
Bandwidth [MHz]	12.7	8.3	11.3
Area [$(\mu\text{m})^2$]	88	150	375
Noise figure [dB]	10.54	9.26	2.98
HD ₂ [dB]	-47.9	-48.7	-56.6
Dynamic range [dB]	39.9	38.6	50.4
FOM [aJ]	163.8	261.4	67.2
FOM _{LNA} [fJ $\cdot(\mu\text{m})^2$]	26.0	32.9	9.7

Even though both power consumption and area was better at 40 and 28 nm, 65 nm achieved better FOM's due to noise and linearity performance. There were some indications of bad robustness in [20], which may imply that performance of the 65 nm amplifier could be difficult to reach in practice. Nevertheless, both FOM and FOM_{LNA} showed worse numbers as process technology was shrunk. This may be an indicator of what can be expected from nanoscale analog circuits. Sub-micron technologies have speed advantages compared to older technologies, but in medium frequency applications like this project may the advantages not be quite as clear. Even though FD-SOI demonstrated excellent robustness and flexibility, it still ended up last in the efficiency race.

Maximum current efficiency was ensured through g_m/I_d optimization, and this contributed to the low power consumption achieved. This current efficiency optimization ensured that the amplifier got minimum current consumption for the given combination of g_m , g_m/g_{ds} and V_{ds} . If a lower current consumption should be achieved, this would require e.g. lower g_m . However, current efficiency did not directly take noise performance into consideration. g_m values were chosen with noise in mind. As stated in (2.50), input referred flicker noise is independent of g_m , so there might be more efficient optimization approaches for these flicker noise limited amplifiers.

5.6 Variable gain

The variable gain strategy with variable input impedance worked well, as depicted in figure 4.3. v_{in}/v_{cmut} was modified by R_S in reasonable steps, and v_{out}/v_{in} was

left completely unaffected. Distortion analysis performed with transistor-resistors and ideal resistors as source resistances showed very little difference in distortion. This made this solution both simple and area efficient. As gain control signals were implemented as simple digital signals can these gain control signals be made common for the entire chip. An additional feature is that the entire CG stage could be disabled by tying all gain signals to ground. This could become useful if it is desired to test performance with high input impedance and a pure voltage amplifier.

Figure 4.7 and 4.8 demonstrated the necessity of variable gain pretty clear. Without variable gain, both rails were severely limited by supply voltage, and the signal was heavily distorted due to this. With variable gain, a decent amplitude was maintained during the entire run. Variable gain was implemented as coarse gain steps, which introduced switching transients at each gain transition. Such transients also introduce distortion in the signal, but with the switching frequency as fundamental tone. With much lower switching frequency than f_c , only negligible amounts of distortion would appear in the signal band.

5.7 Feedback biasing

All results indicated that the implemented feedback biasing was a viable solution which was both robust and energy efficient. One drawback was the necessary coupling capacitors which together, according to figure 3.2 and 3.3, consumed 25 and 47% of total amplifier area for 40 and 28 nm respectively. These could have been reduced by increasing R_f or increasing lower cutoff frequency. R_f was already very high compared to other resistances in the circuit, and it was not desirable to let it get higher due to e.g. gate resistance. Lower cutoff frequency could have been increased, but this would have reduced low-frequency cancellation between CG and CS.

Using subthreshold diodes as feedback resistances also proved to be a clever choice. Linear resistors with such resistance would have become huge compared to the achieved amplifier sizes and also introduced significant parasitics. One pitfall was that all simulations except transient analysis only required a diode from drain to gate in order to complete with good performance. However, transient analysis revealed that also a diode from gate to drain was required in order to handle large switching transients. This emphasizes the point of always verifying performance using several tests and at least one test that is not based on steady-state simulation.

Due to fairly small output amplitudes of 30-40 mV per rail, there were no signs of

distortion from feedback diodes. A large signal simulation with ideal $1.4\text{ G}\Omega$ resistors instead of diodes showed no distortion differences between the two solutions, mainly due to other distortion sources. This indicates that diodes could easily be used for higher output amplitudes. At least 100-200 mV amplitude should be possible with threshold voltages of 400-500 mV.

5.8 Matching

As already discussed, FD-SOI matching was excellent during monte carlo variations, and also 40 nm bulk demonstrated reasonable values. One exception was HD_2 of V_{out}^+ which got a σ of 4.4 dB in 40 nm. An interval of $\pm 3\sigma$ contains 99.7% of all circuits and may be considered a rule of thumb within reliability. When looking at 3σ variations of V_{out}^+ , this would result in an interval of -41.8 to -68.2 dB. The worst limit was equal to V_{out}^- average, which may indicate that low V_{out}^+ distortion was more of a lucky coincidence. V_{out}^- , for its part, got a σ of 0.3 dB, which implied a maximum deviation of ± 1 dB. The same pattern was recognized in FD-SOI, but with smaller variations. All this indicate that HD_2 of V_{out}^+ may be worse than what average suggests, and that distortion in the two rails could end up approximately equal. Significant variations in V_{out}^+ also implied significant variations in V_{out} , although not as large as in V_{out}^+ . Worst case performance of V_{out} in 40 nm was estimated to -41.3 dB, which would make distortion contributions approximately equal to noise contributions at 40.7 dB.

Conclusion

This thesis has described design of two nanoscale low noise amplifiers for intravascular ultrasound imaging. They were based on a common-gate-common-source topology and implemented in 40 nm bulk CMOS and 28 nm FD-SOI. Through the use of g_m/I_d -based energy efficiency optimization, a low power consumption of 20.5 and 16.0 μW was achieved. These amplifiers were designed for sensing signal from capacitive micromachined transducers at 5 MHz with 100% bandwidth, which resulted in a noise figure of 10.5 and 9.3 dB. Even though noise figure was not impressive, SNRs of 40.7 and 39.1 dB was achieved with second harmonic distortion of -47.9 and -48.7 dB. Both amplifiers included variable gain, which was implemented using switched triode transistors. Everything was fitted into a tiny area of 88 and 150 (μm)². Feedback bias was used to bias both amplifier stages, which resulted in excellent robustness and a simple design. Such compact biasing would not have been possible without the combination of moscap capacitors and subthreshold diode resistors. The chosen topology proved to be a reasonable choice that fulfilled most specifications, e.g. single-to-differential signal conversion and low input impedance.

The investigated technologies featured both benefits and drawbacks. 28 nm FD-SOI demonstrated excellent variability and flexibility which resulted in really robust performance. It did however suffer from large amounts of flicker noise, which affected its noise performance. 40 nm bulk was less accurate and manageable, which limited the design freedom. These properties affected performance for both technologies, but it should probably be possible to avoid most issues by further transistor tweaking. Small process parasitics made it possible to accurately model small-signal performance using ideal small-signal models.

6.1 Future work

Despite decent overall performance several sub-optimal areas were revealed in the discussion. Most of them were related to noise and distortion, as indicated by the FOM summary where an earlier work proved to be more efficient. An investigation of how to extract more signal from the transducer without increasing input impedance would probably be an efficient way of increasing noise figure. This could for instance entail a transition to pure transimpedance operation. The most significant noise and distortion sources were also identified in the discussion, and further tweaking on these transistors would be necessary in order to reduce noise and distortion. Especially distortion in the common-gate load, Q_{P1} , should be investigated further as it prevented valuable distortion cancellation. Distortion effectively limits output swing, which is a sparse resource in modern technologies with low supply voltage. It may also be useful to investigate a swap of NMOS and PMOS transistors for noise reasons, even though this will probably increase amplifier size.

Even though a functional circuit was extracted from layout, these layouts were not ready for direct tape-out. Several production-specific DRC rules such as dummy fillings, dummy layers, ESD protection and guard were ignored during layout design since they usually have little impact on performance. These devices and layers should be implemented according to manufacturing rules before the circuit could be sent to production. In the case of final production could results also be verified through measurements, which are more credible than only simulation.

Bibliography

- [1] A. Sisman, J. Zahorian, G. Gurun, Mustafa Karaman, M. Balantekin, F.L. Degertekin, and P. Hasler. Evaluation of cmut annular arrays for side-looking ivus. In *Ultrasonics Symposium (IUS), 2009 IEEE International*, pages 2774–2777, Sept 2009.
- [2] Edward D. Light and Stephen W. Smith. Two dimensional arrays for real time 3d intravascular ultrasound. *Ultrasonic Imaging*, 26(2):115–128, 2004.
- [3] K. Kaviani, O. Oralkan, P. Khuri-Yakub, and B.A. Wooley. A multichannel pipeline analog-to-digital converter for an integrated 3-d ultrasound imaging system. *Solid-State Circuits, IEEE Journal of*, 38(7):1266–1270, July 2003.
- [4] M.A. Averkiou, D.N. Roundhill, and J.E. Powers. A new imaging technique based on the nonlinear properties of tissues. In *Ultrasonics Symposium, 1997. Proceedings., 1997 IEEE*, volume 2, pages 1561–1566 vol.2, 1997.
- [5] W.A. Smith. The role of piezocomposites in ultrasonic transducers. In *Ultrasonics Symposium, 1989. Proceedings., IEEE 1989*, pages 755–766 vol.2, 1989.
- [6] Butrus T. Khuri-Yakub, Ching-Hsiang Cheng, Fahrettin-Levent Degertekin, Sanli Ergun, Sean Hansen, Xue-Cheng Jin, and Omer Oralkan. Silicon micromachined ultrasonic transducers. *Japanese Journal of Applied Physics*, 39(Part 1, No. 5B):2883–2887, 2000.
- [7] G. Gurun, P. Hasler, and F.L. Degertekin. Front-end receiver electronics for high-frequency monolithic cmut-on-cmos imaging arrays. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 58(8):1658–1668, August 2011.

-
- [8] A.S. Ergun, C.-H. Cheng, U. Demirci, and B.T. Khuri-Yakub. Fabrication and characterization of 1-dimensional and 2-dimensional capacitive micro-machined ultrasonic transducer (cmut) arrays for 2-dimensional and volumetric ultrasonic imaging. In *OCEANS '02 MTS/IEEE*, volume 4, pages 2361–2367 vol.4, 2002.
- [9] H. Kagan Oguz, S. Olcum, M.N. Senlik, A. Atalar, and H. Koymen. A novel equivalent circuit model for cmuts. In *Ultrasonics Symposium (IUS), 2009 IEEE International*, pages 2193–2196, Sept 2009.
- [10] S. Sharma and T. Ytterdal. Low noise front-end amplifier design for medical ultrasound imaging applications. In *VLSI and System-on-Chip (VLSI-SoC), 2012 IEEE/IFIP 20th International Conference on*, pages 12–17, 2012.
- [11] S. Berg, T. Ytterdal, and A. Ronnekleiv. Co-optimization of cmut and receive amplifiers to suppress effects of neighbor coupling between cmut elements. In *Ultrasonics Symposium, 2008. IUS 2008. IEEE*, pages 2103–2106, Nov 2008.
- [12] R.T. Howe and C.G. Sodini. *Microelectronics : an integrated approach*. Upper Saddle River, N.J. : Prentice Hall, 1997.
- [13] D.M. Pozar. *Microwave engineering*. Wiley, 1997.
- [14] T.C. Carusone, D.A. Johns, and K.W. Martin. *Analog Integrated Circuit Design, second edition*. Analog Integrated Circuit Design. Wiley, 2011.
- [15] D. Johns and K.W. Martin. *Analog integrated circuit design*. John Wiley & Sons, 1997.
- [16] A.S. Sedra and K.C. Smith. *Microelectronic circuits / Adel S. Sedra, Kenneth C. Smith*. New York : Oxford University Press, 2011.
- [17] J.M. Miller. *Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit*. Scientific papers of the Bureau of Standards. Govt. Print. Off., 1919.
- [18] T. Singh, T. Saether, and T. Ytterdal. Feedback biasing in nanoscale cmos technologies. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 56(5):349–353, May 2009.
- [19] S.C. Blaakmeer, E.A.M. Klumperink, B. Nauta, and D.M.W. Leenaerts. Wideband cmos receivers exploiting simultaneous output balancing and noise/distortion canceling. In *Microwave Integrated Circuit Conference, 2008. EuMIC 2008. European*, pages 163–166, Oct 2008.

-
- [20] J.H. Eriksrod and T. Ytterdal. A 65nm cmos front-end lna for medical ultrasound imaging with feedback employing noise and distortion cancellation. In *Circuit Theory and Design (ECCTD), 2013 European Conference on*, pages 1–4, 2013.
- [21] H.H. Hansen. A 33 uw sub-3db noise figure low noise amplifier for medical ultrasound applications. Master’s thesis, Norwegian University of Science and Technology, 2011.
- [22] H. Shichman and D. Hodges. Modeling and simulation of insulated-gate field-effect transistor switching circuits. *Solid-State Circuits, IEEE Journal of*, 3(3):285–289, 1968.
- [23] E.A. Vittoz. Future of analog in the vlsi environment. In *Circuits and Systems, 1990., IEEE International Symposium on*, pages 1372–1375 vol.2, 1990.
- [24] D.J. Comer and D.T. Comer. Operation of analog mos circuits in the weak or moderate inversion region. *Education, IEEE Transactions on*, 47(4):430–435, Nov 2004.
- [25] Eric A. Vittoz. Design of analog-digital vlsi circuits for telecommunications and signal processing. chapter Micropower Techniques, pages 53–96. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1994.
- [26] R. Kasser. A new noise equivalent circuit for the junction fet with uncorrelated noise sources. *Proceedings of the IEEE*, 58(7):1171–1172, July 1970.
- [27] A. van der Ziel. Noise in solid-state devices and lasers. *Proceedings of the IEEE*, 58(8):1178–1206, Aug 1970.
- [28] A.-J. Annema, B. Nauta, R. van Langevelde, and H. Tuinhout. Analog circuits in ultra-deep-submicron cmos. *Solid-State Circuits, IEEE Journal of*, 40(1):132–143, 2005.
- [29] P.G. Jespers. *The gm/ID design methodology, a sizing tool for low-voltage analog CMOS circuits : the semi-empirical and compact model approaches / by Paul G. A. Jespers*. New York : Springer, 2010.
- [30] M. Bucher, C. Lallement, C. Enz, and F. Krummenacher. Accurate mos modelling for analog circuit simulation using the ekv model. In *Circuits and Systems, 1996. ISCAS '96., Connecting the World., 1996 IEEE International Symposium on*, volume 4, pages 703–706 vol.4, May 1996.
-

-
- [31] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neiryneck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki. A 45nm logic technology with high-k+metal gate transistors, strained silicon, 9 cu interconnect layers, 193nm dry patterning, and 100packaging. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 247–250, 2007.
- [32] Y. Chiu, B. Nikolic, and P.R. Gray. Scaling of analog-to-digital converters into ultra-deep-submicron cmos. In *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005*, pages 375–382, 2005.
- [33] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Buf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J-P. Schoellkopf, E. Perea, R. Ferrant, and H. Mingam. Innovative materials, devices, and cmos technologies for low-power mobile multimedia. *Electron Devices, IEEE Transactions on*, 55(1):96–130, Jan 2008.
- [34] J. Colinge. Fully-depleted soi cmos for analog applications. *Electron Devices, IEEE Transactions on*, 45(5):1010–1016, May 1998.
- [35] B. Pelloux-Prayer, M. Blagojevic, O. Thomas, A. Amara, A. Vladimirescu, B. Nikolic, G. Cesana, and P. Flatresse. Planar fully depleted soi technology: The convergence of high performance and low power towards multimedia mobile applications. In *Faible Tension Faible Consommation (FTFC), 2012 IEEE*, pages 1–4, June 2012.
- [36] STMicroelectronics. The fs-soi innovation. http://www.st.com/web/en/about_st/learn_fd-soi.html. [Online; accessed 2014-06-13].
- [37] T. Ytterdal and C. Wulff. On the energy efficiency of analog circuits in nanoscale cmos technologies. In *Microelectronics, 2008. ICM 2008. International Conference on*, pages 240–243, Dec 2008.
- [38] I.O. Wygant, Xuefeng Zhuang, D.T. Yeh, O. Oralkan, A.S. Ergun, Mustafa Karaman, and B.T. Khuri-Yakub. Integration of 2d cmut arrays with front-end electronics for volumetric ultrasound imaging. *Ultrasonics, Ferro-*

electronics and Frequency Control, IEEE Transactions on, 55(2):327–342, 2008.

- [39] A. Bhuyan, J.W. Choe, Byung Chul Lee, I. Wygant, A. Nikoozadeh, O. Oralkan, and B.T. Khuri-Yakub. 3d volumetric ultrasound imaging with a 32 x 32 cmut array integrated with front-end ics using flip-chip bonding technology. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, pages 396–397, 2013.
- [40] P. Behnamfar and S. Mirabbasi. Design of a high-voltage analog front-end circuit for integration with cmut arrays. In *Biomedical Circuits and Systems Conference (BioCAS), 2010 IEEE*, pages 298–301, 2010.
- [41] C. Tekes, G. Gurun, M.W. Rashid, J. Zahorian, T. Xu, J. Hasler, and F.L. Degertekin. 3-d real-time volumetric imaging using 20 mhz 1.5-mm diameter single-chip cmut-on-cmos array. In *Ultrasonics Symposium (IUS), 2012 IEEE International*, pages 1–4, 2012.
- [42] G. Gurun, P. Hasler, and F. Degertekin. A 1.5-mm diameter single-chip cmos front-end system with transmit-receive capability for cmut-on-cmos forward-looking ivus. In *Ultrasonics Symposium (IUS), 2011 IEEE International*, pages 478–481, Oct 2011.
- [43] D. Reda, E. Hegazi, K.N. Salama, and H.F. Ragai. Design of low noise transimpedance amplifier for intravascular ultrasound. In *Biomedical Circuits and Systems Conference, 2009. BioCAS 2009. IEEE*, pages 57–60, Nov 2009.
- [44] F. Bruccoleri, E.A.M. Klumperink, and B. Nauta. Wide-band cmos low-noise amplifier exploiting thermal noise canceling. *Solid-State Circuits, IEEE Journal of*, 39(2):275–282, Feb 2004.
- [45] E. Bjørsvik. Design of low power lna in 40 nm cmos for intravascular ultrasound applications. Technical report, Norwegian University of Technology and Science, 2014.
- [46] E.G. Ioannidis, S. Haendler, A. Bajolet, T. Pahron, N. Planes, F. Arnaud, R-A Bianchi, M. Haond, D. Golanski, J. Rosal, C. Fenouillet-Beranger, P. Perreau, C.A. Dimitriadis, and G. Ghibaudo. Low frequency noise variability in high-k/metal gate stack 28nm bulk and fd-soi cmos transistors. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 18.6.1–18.6.4, Dec 2011.

Appendix A

Additional analytic expressions

A.1 Feedback biased common-source amplifier

A test voltage, v_{in} , is applied to v_{in} in the small-signal model in figure 2.12b. This gives the following expressions.

$$v_{out} = -v_{in} \frac{C_{ac} r_o R_L (g_m R_f - 1) s}{C_{ac} [(R_f + R_L) r_o + R_f R_L] s + (g_m r_o + 1) R_L + r_o} \quad (\text{A.1})$$

$$v_g = v_{in} \frac{C_{ac} [(R_f + R_L) r_o + R_f R_L] s}{C_{ac} [(R_f + R_L) r_o + R_f R_L] s + (g_m r_o + 1) R_L + r_o} \quad (\text{A.2})$$

$$A_v = \frac{v_{in}}{v_{out}} = - \frac{C_{ac} r_o R_L (g_m R_f - 1) s}{C_{ac} [(R_f + R_L) r_o + R_f R_L] s + (g_m r_o + 1) R_L + r_o} \quad (\text{A.3})$$

$$i_{in} = \frac{v_{in} - v_g}{1/(sC_{ac})} = v_{in} \frac{C_{ac} ([g_m r_o + 1] R_L + r_o) s}{C_{ac} [(R_f + R_L) r_o + R_f R_L] s + (g_m r_o + 1) R_L + r_o} \quad (\text{A.4})$$

$$r_{in} = \frac{v_{in}}{i_{in}} = \frac{C_{ac} [(R_f + R_L) r_o + R_f R_L] s + (g_m r_o + 1) R_L + r_o}{C_{ac} ([g_m r_o + 1] R_L + r_o) s} \quad (\text{A.5})$$

A.2 Feedback biased common-gate amplifier

A test current, i_{in} , is applied to v_{in} in the small-signal model in figure 2.13b. This gives the following expressions.

$$v_{in} = i_{in} \frac{[C_{dc}((R_f + R_L)r_o + R_f R_L)s + (g_m r_o + 1)R_L + r_o] R_S}{C_{dc}[(R_f + R_L)\{r_o + (g_m r_o + 1)R_S\} + R_f R_L]s + (g_m r_o + 1)(R_L + R_S) + r_o} \quad (\text{A.6})$$

$$v_{out} = i_{in} \frac{(C_{dc} R_f s + 1)(g_m r_o + 1)R_L R_S}{C_{dc}[(R_f + R_L)\{r_o + (g_m r_o + 1)R_S\} + R_f R_L]s + (g_m r_o + 1)(R_L + R_S) + r_o} \quad (\text{A.7})$$

$$v_g = i_{in} \frac{(g_m r_o + 1)R_L R_S}{C_{dc}[(R_f + R_L)\{r_o + (g_m r_o + 1)R_S\} + R_f R_L]s + (g_m r_o + 1)(R_L + R_S) + r_o} \quad (\text{A.8})$$

$$A_v = \frac{v_{out}}{v_{in}} = \frac{C_{dc} R_f R_L (g_m r_o + 1)s + (g_m r_o + 1)R_L}{C_{dc}((R_f + R_L)r_o + R_f R_L)s + (g_m r_o + 1)R_L + r_o} \quad (\text{A.9})$$

$$r_{in} = \frac{[C_{dc}((R_f + R_L)r_o + R_f R_L)s + (g_m r_o + 1)R_L + r_o] R_S}{C_{dc}[(R_f + R_L)\{r_o + (g_m r_o + 1)R_S\} + R_f R_L]s + (g_m r_o + 1)(R_L + R_S) + r_o} \quad (\text{A.10})$$

A.3 Feedback biased common-gate-common-source amplifier

A complete small-signal model of the common-gate-common-source topology. Loaded with C_L at both rails.

$$\frac{v_{out}}{v_{in}} = R_L \frac{g(s)}{h(s)} \quad (\text{A.11})$$

$$\begin{aligned} g(s) = & R_L + r_o + r_o^2 g_m + R_L r_o^2 g_m^2 + 2R_L r_o g_m - C_{ac} r_o^2 s + C_L R_L r_o s \\ & + C_{ac} R_L R_{f2} s + C_{dc} R_L R_{f1} s + C_{ac} R_{f2} r_o s + C_{dc} R_{f1} r_o s + C_L R_L r_o^2 g_m s \\ & + 2C_{ac} R_{f2} r_o^2 g_m s + C_{dc} R_{f1} r_o^2 g_m s - C_L C_{ac} R_L r_o^2 s^2 \\ & - C_{ac} C_{dc} R_L r_o^2 s^2 - C_{ac} C_{dc} R_{f1} r_o^2 s^2 + C_{ac} R_L R_{f2} r_o^2 g_m^2 s \\ & + C_{dc} R_L R_{f1} r_o^2 g_m^2 s + 2C_{ac} R_L R_{f2} r_o g_m s + 2C_{dc} R_L R_{f1} r_o g_m s \\ & + C_L C_{ac} R_L R_{f2} r_o s^2 + C_L C_{dc} R_L R_{f1} r_o s^2 + C_{ac} C_{dc} R_L R_{f1} R_{f2} s^2 \\ & + C_{ac} C_{dc} R_{f1} R_{f2} r_o s^2 - C_L C_{ac} C_{dc} R_L R_{f1} r_o^2 s^3 \\ & + 2C_L C_{ac} R_L R_{f2} r_o^2 g_m s^2 + C_L C_{dc} R_L R_{f1} r_o^2 g_m s^2 \\ & + C_{ac} C_{dc} R_L R_{f1} r_o^2 g_m s^2 + C_{ac} C_{dc} R_L R_{f2} r_o^2 g_m s^2 \\ & + 2C_{ac} C_{dc} R_{f1} R_{f2} r_o^2 g_m s^2 + C_L C_{ac} C_{dc} R_L R_{f1} R_{f2} r_o s^3 \\ & + 2C_{ac} C_{dc} R_L R_{f1} R_{f2} r_o g_m s^2 + 2C_L C_{ac} C_{dc} R_L R_{f1} R_{f2} r_o^2 g_m s^3 \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned}
h(s) = & 2R_L r_o + R_L^2 + r_o^2 + R_L^2 r_o^2 g_m^2 + 2R_L r_o^2 g_m \\
& + 2R_L^2 r_o g_m + 2C_L R_L r_o^2 s + 2C_L R_L^2 r_o s + C_{ac} R_L^2 R_{f2} s \\
& + C_{dc} R_L^2 R_{f1} s + C_{ac} R_L r_o^2 s + C_{ac} R_L^2 r_o s + C_{dc} R_L r_o^2 s \\
& + C_{dc} R_L^2 r_o s + C_{ac} R_{f2} r_o^2 s + C_{dc} R_{f1} r_o^2 s + C_L^2 R_L^2 r_o^2 s^2 \\
& + 2C_{ac} R_L R_{f2} r_o s + 2C_{dc} R_L R_{f1} r_o s + C_L C_{ac} R_L^2 r_o^2 s^2 \\
& + C_L C_{dc} R_L^2 r_o^2 s^2 + C_{ac} C_{dc} R_L^2 r_o^2 s^2 + 2C_L R_L^2 r_o^2 g_m s \\
& + C_{ac} R_L^2 r_o^2 g_m s + C_{dc} R_L^2 r_o^2 g_m s + 2C_L C_{ac} R_L R_{f2} r_o^2 s^2 \\
& + 2C_L C_{ac} R_L^2 R_{f2} r_o s^2 + 2C_L C_{dc} R_L R_{f1} r_o^2 s^2 \\
& + 2C_L C_{dc} R_L^2 R_{f1} r_o s^2 + C_{ac} C_{dc} R_L^2 R_{f1} R_{f2} s^2 \\
& + C_{ac} C_{dc} R_L R_{f1} r_o^2 s^2 + C_{ac} C_{dc} R_L^2 R_{f1} r_o s^2 \\
& + C_{ac} C_{dc} R_L R_{f2} r_o^2 s^2 + C_{ac} C_{dc} R_L^2 R_{f2} r_o s^2 \\
& + C_{ac} C_{dc} R_{f1} R_{f2} r_o^2 s^2 + C_L^2 C_{ac} R_L^2 R_{f2} r_o^2 s^3 \\
& + C_L^2 C_{dc} R_L^2 R_{f1} r_o^2 s^3 + C_{ac} R_L R_{f2} r_o^2 g_m s \\
& + C_{ac} R_L^2 R_{f2} r_o g_m s + C_{dc} R_L R_{f1} r_o^2 g_m s + C_{dc} R_L^2 R_{f1} r_o g_m s \\
& + 2C_{ac} C_{dc} R_L R_{f1} R_{f2} r_o s^2 + C_L C_{ac} C_{dc} R_L^2 R_{f1} r_o^2 s^3 \\
& + C_L C_{ac} C_{dc} R_L^2 R_{f2} r_o^2 s^3 + C_L C_{ac} R_L^2 R_{f2} r_o^2 g_m s^2 \\
& + C_L C_{dc} R_L^2 R_{f1} r_o^2 g_m s^2 + C_L^2 C_{ac} C_{dc} R_L^2 R_{f1} R_{f2} r_o^2 s^4 \\
& + 2C_L C_{ac} C_{dc} R_L R_{f1} R_{f2} r_o^2 s^3 + 2C_L C_{ac} C_{dc} R_L^2 R_{f1} R_{f2} r_o s^3 \tag{A.13}
\end{aligned}$$

Appendix B

Calculations

B.1 Input SNR

Given the CMUT model from figure 2.4a, we add a noiseless input resistance R_{in} termination to find input noise, V_{ni}^2 .

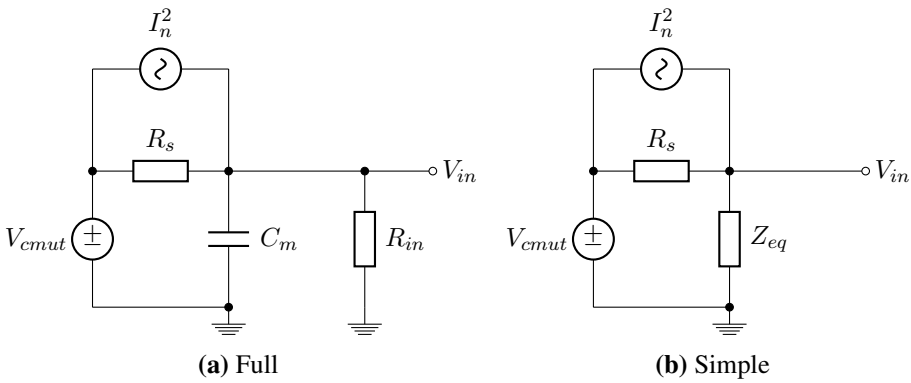


Figure B.1: CMUT equivalents for noise calculation

$R_s = 20 \text{ k}\Omega$, $C_m = 2.76 \text{ pF}$. Using an approximate input impedance, R_{in} , of $5 \text{ k}\Omega$, $f_c = 5 \text{ MHz}$, $f_{min} = 2.5 \text{ MHz}$ and $f_{max} = 7.5 \text{ MHz}$.

$$I_n^2(f) = \frac{4kT}{R_s} \tag{B.1}$$

$$Z_{eq}(f) = \left(\frac{1}{R_{in}} + j2\pi fC \right)^{-1} \quad (\text{B.2})$$

$$v_{ni}^2(f) = I_n^2(f) \left(\frac{R_s Z_{eq}(f)}{R_s + Z_{eq}(f)} \right)^2 \quad (\text{B.3})$$

$$\begin{aligned} v_{ni}^2 &= \int_{f_{min}}^{f_{max}} v_{ni}^2(f_c) df & (\text{B.4}) \\ &= 4kT \frac{R_s R_{in}^2 (f_{max} - f_{min})}{(j2\pi f_{min} C_m R_{in} R_s + R_{in} + R_s) (j2\pi f_{max} C_m R_{in} R_s + R_{in} + R_s)} \\ &= 57.9 \text{ pV}^2 \end{aligned}$$

$$v_{in} = v_{cmut} \frac{Z_{eq}(f_c)}{R_s + Z_{eq}(f_c)} = 0.1890 \cdot v_{cmut} \quad (\text{B.5})$$

$$\text{SNR} = \frac{v_{in}^2}{v_{ni}^2} = 6.164 \cdot 10^8 \cdot v_{cmut}^2 \quad (\text{B.6})$$

SNR can also be approximated by assuming a flat transfer function around f_c , with a bandwidth of $\Delta f = 5$ MHz:

$$\begin{aligned} v_{ni}^2 &= \int_{\Delta f} v_{ni}^2(f_c) df & (\text{B.7}) \\ &= I_n^2(f_c) \left(\frac{R_s Z_{eq}(f_c)}{R_s + Z_{eq}(f_c)} \right)^2 \Delta f \end{aligned}$$

$$\text{SNR}_{\text{approx}} = \frac{v_{in}^2}{v_{ni}^2} = \frac{v_{cmut}^2 \left(\frac{Z_{eq}(f_c)}{R_s + Z_{eq}(f_c)} \right)^2}{\frac{4kT}{R_s} \left(\frac{R_s Z_{eq}(f_c)}{R_s + Z_{eq}(f_c)} \right)^2 \Delta f} = \frac{v_{cmut}^2}{4kT R_s \Delta f} \quad (\text{B.8})$$

Appendix C

Test bench

Table C.1: Component values for CMUT model in figure 2.4b.

Component	Value
R_{cs}	20 k Ω
C_m	2.76 pF

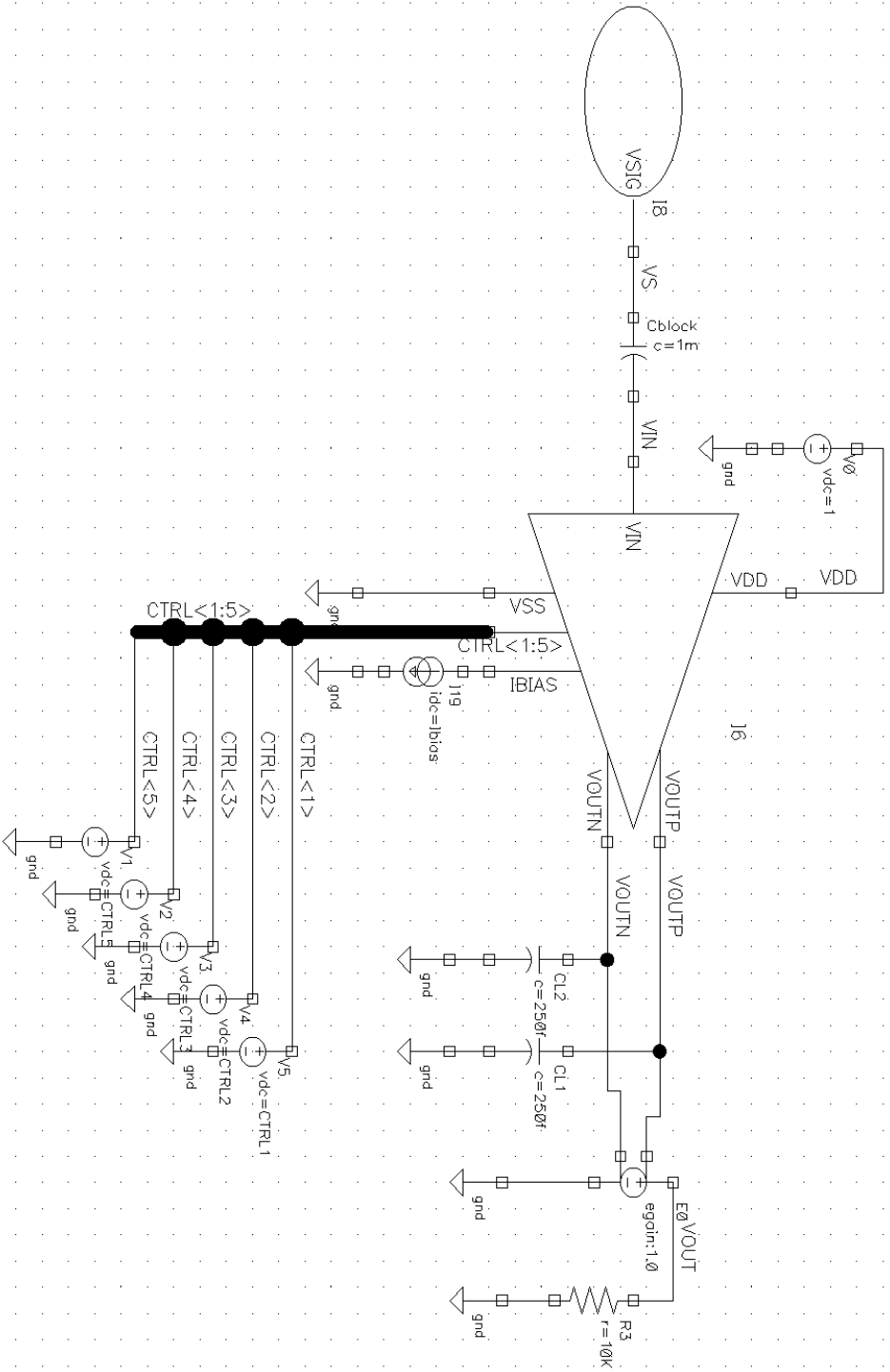


Figure C.1: Amplifier test bench schematic.

Appendix D

Additional simulation plots

D.1 Input referred noise

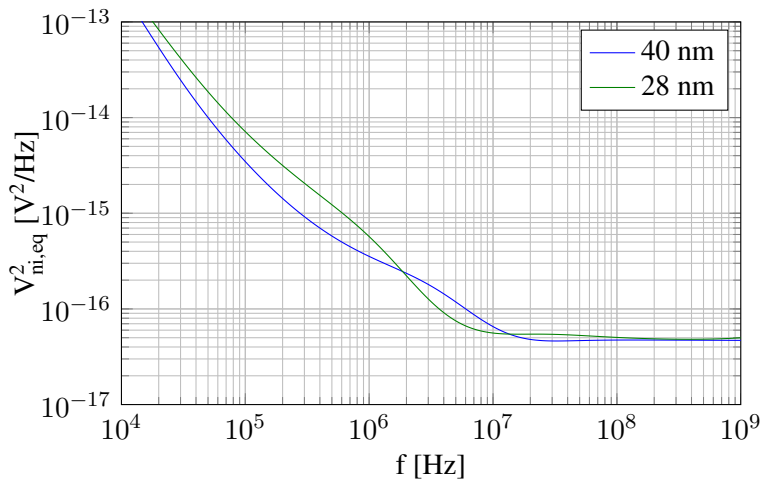


Figure D.1: Simulated $N_{in,eq}$ for 40 nm and 28 nm technology. Post-layout.

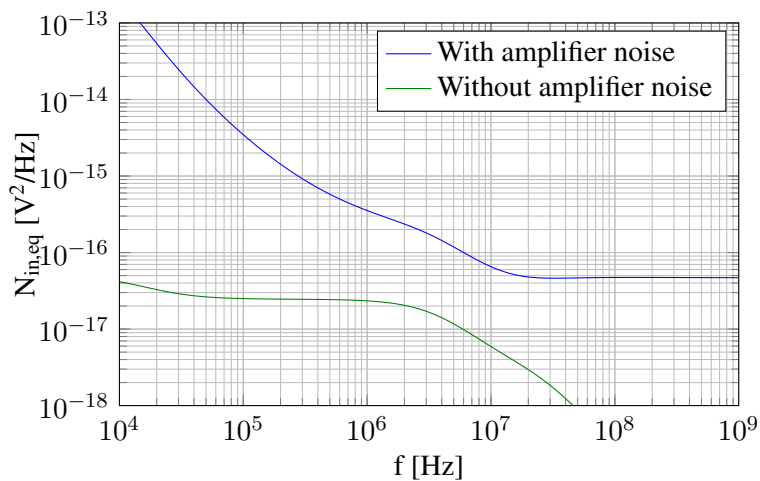


Figure D.2: Simulated $N_{in,eq}$ for 40 nm technology with and without amplifier noise. Post-layout.

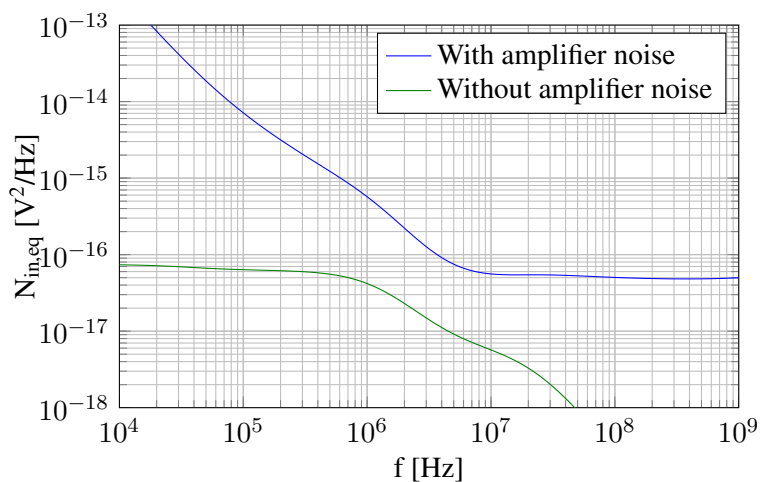


Figure D.3: Simulated $N_{in,eq}$ for 28 nm technology with and without amplifier noise. Post-layout.

D.2 Small-signal frequency response

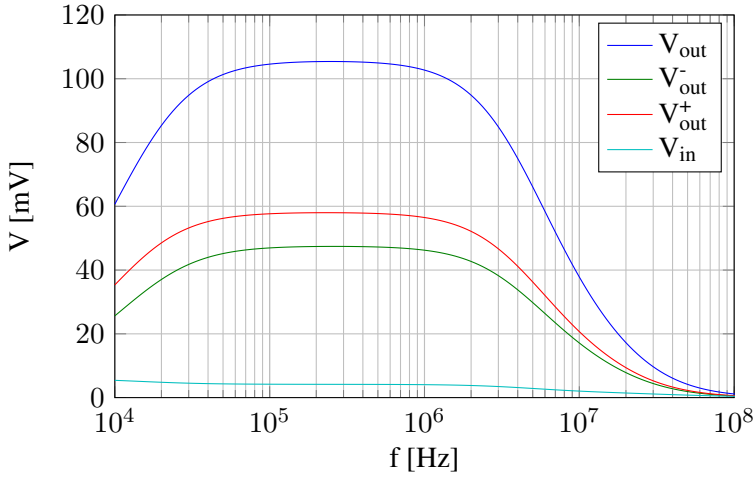


Figure D.4: Simulated small-signal output and input signals in 40 nm with $v_{cmut} = 15.4$ mV. Post-layout.

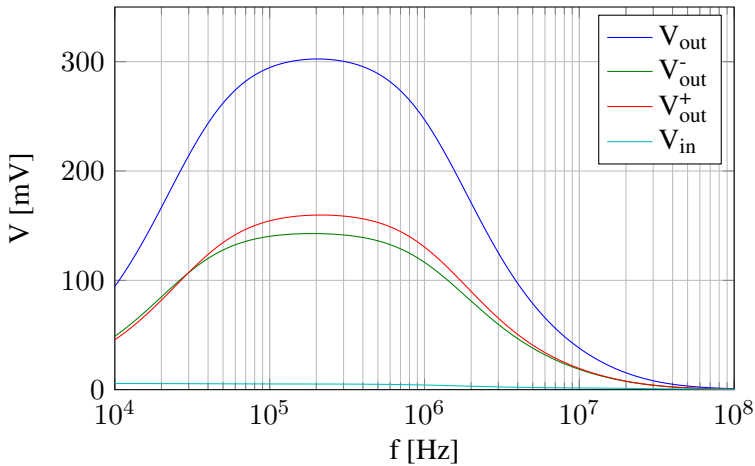


Figure D.5: Simulated small-signal output and input signals in 28 nm with $v_{cmut} = 12$ mV. Post-layout.

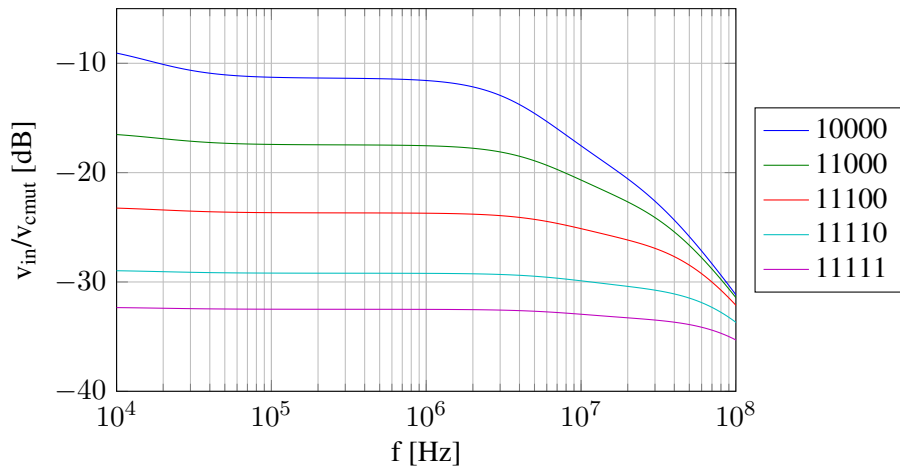


Figure D.6: Signal loss from v_{cmut} to v_{in} for different gain settings in 40 nm. Post-layout.

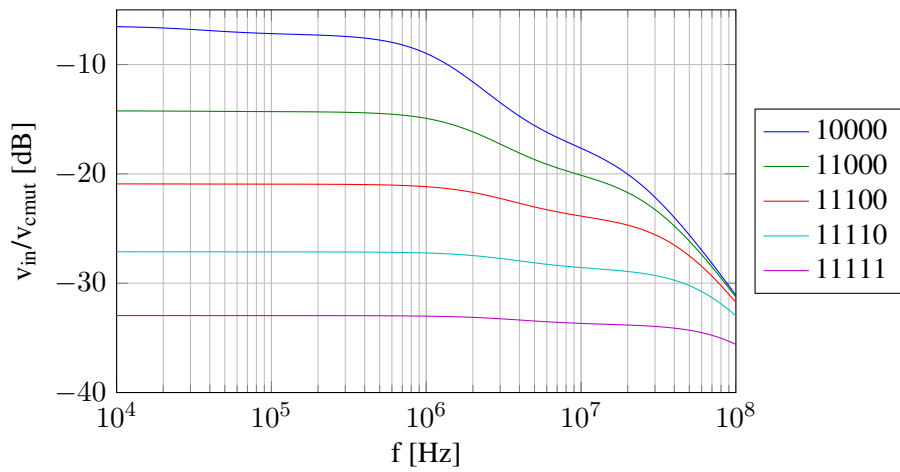


Figure D.7: Signal loss from v_{cmut} to v_{in} for different gain settings in 28 nm. Post-layout.