Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

Muhammad Shafiq

Design of Energy Efficient LNAs for Medical Ultrasound Imaging Applications

Master's thesis in Electronic systems Design
Supervisor: Trond Ytterdal
August 2021

NTNU
Norwegian University of Science and Technology

Muhammad Shafiq

Design of Energy Efficient LNAs for Medical Ultrasound Imaging Applications

Master's thesis in Electronic systems Design
Supervisor: Trond Ytterdal
August 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

Problem Formulation

Thesis description:

The chip area and power efficiency are the two main concerns while designing an ultrasound medical imaging device by using a CMOS technology. Downscaling of this technology has provided energy efficient small chips at the size of nanometer level. Therefore, the modern analog circuit design demands high performance by utilizing a low supply voltage. The main objective of this thesis is to design a front-end LNA in any commersially available 130nm CMOS technology. The amplifier should be able to receive low amplitude ultrasound transducer signals. The amplifier should be designed by utilizing at least two different CMOS topologies.

The following specifications should be satisfied:

1. Implementation of two fully differential CMOS topologies in Cadence
2. DC gain about 500 with a supply voltage of 1.5 V
3. Unity gain frequency of 200 MHz
4. The SNR shouls be at least 50 dB and PM 60 dB
5. HD2 should be at least 40 dB less than the fundamental harmonic
6. Overall power consumption should be less than 30 μW
7. Make a comparison of the two implemented topologies

Supervisor: Prof. Trond Ytterdal, IET, NTNU, Trondheim.
Thesis assigned: 8th January, 2021.

Acknowledgement

This master's thesis is an individual, delimited research project and is carried out under the supervision of Prof. Trond Ytterdal by using the available means at the university campus. The thesis is an final requirement of the 2-years master's programme at the department of Electronic Systems Design offered by The Norwegian University of Science and Technology (NTNU), Trondheim, Norway. The most of the thesis work has been accomplished at the university campus. The physical appearence on the university campus was quite challenging due to the covid 19 epidemic. But the design process specially in the Cadence virtuoso tool, have been very exciting and great oppertunity to learn and explore the CMOS technologies. The credit of course goes to my supervisor Prof. Trond Ytterdal who provided me this great oppertunity to work on such an exciting task.

Therefore, i would like to thank my supervisor Prof. Trond Ytterdal for his availibility, guidence and valuable feedback throughout the thesis time period. With his smily face, his suggestions been always helpful for me in the process of learning and understanding. Specially, the knowledge transformation, even at the fundamental level, boosted my motivation to work. I would love to work with him in the coming future if i get any further oppertunity. I would also like to thank my family and friends to be there with their great support.

Trondheim, February 2021

Muhammad Shafiq

Abstract

The developement of modern intravascular ultrasound imaging has played a vital role in the field of medicine. These probe devices senses and amplifies the low frequency and low quality signals like ECG, EMG, EEG and EOG etc. All of these signals measure biopotentials or electrical output resulting by different activities of a human body. The challenge is to enhance the better quality, energy efficiency and minimized area which involves a complete complex design phase. Hence the LNA is the most critical fron-end element in a sensing probe. The trade-off among various parameters like speed, size, power consumption and noise performance, make it even more challenging for the analog designers. This thesis presents the design and simulations of a front-end low noise amplifier by adopting two different design architectures.

The two proposed architectures of fully differential Operational Transconductance Amplifier (OTA) are implemented by using a commersially available $130\,nm$ CMOS technology. The Cadence design tool has been utilized in order to complete the design phase. The gm/I_d methodology has been adopted while sizing the transistors.

The performance of the both topologies have been tested and verified by using corners and the Monte Carlo simulations. In the end, both of the topologies are compared to each with respect to the defined figure-of-merit.

Contents

I Introduction **1**

1 Introduction **3**
1.1 Motivation . . . 3
1.2 Main Contributions . . . 5
1.3 Thesis Outline . . . 5

II Theoretical Background **6**

2 Ultrasound Imaging Systems **8**
2.1 Basic ultrasound principle . . . 8
2.2 Two-dimensional Imaging . . . 8
2.3 Three-dimensional imaging . . . 9
2.4 Front-end electronics . . . 9

3 Amplifiers **11**
3.1 Low Noise Amplifiers . . . 11
3.2 Performance Parameters . . . 12
3.2.1 The Gain . . . 12
3.2.2 Frequency response . . . 12
3.2.3 Linearity . . . 13
3.2.4 Noise Analysis and Modeling . . . 14
3.2.5 Noise spectral density . . . 16
3.3 CMOS Technology . . . 17
3.4 Figure-of-Merit . . . 19
3.5 CMOS Architectures . . . 19
3.5.1 Single-ended versus differential operation . . . 20
3.5.2 Standard differential stage . . . 20
3.5.3 Two-stage OTA . . . 22
3.5.4 Two-stage Miller OTA . . . 24
3.5.5 Telescope cascoding-OTA . . . 25
3.5.6 Folded cascode OTA . . . 27
3.5.7 Current Mirror OTA . . . 29
3.5.8 Gain boosting . . . 31
3.5.9 Feedback Amplifiers . . . 31
3.5.10 Inverters . . . 32

III Method **34**

4 Design Methodology **36**

4.1 Selected Design Method . . . 36
4.2 Selected Architectures . . . 37
4.3 Selected FOM . . . 38
4.4 Technology characteristics . . . 38
4.5 Fully differential folded cascode OTA . . . 39
4.5.1 Hand calculations . . . 41
4.5.2 Transistors sizing . . . 42
4.6 Biasing circuit for the folded cascode OTA . . . 43
4.7 CMFB circuit for the folded cascode OTA . . . 44
4.8 Fully differential inverter-based current mirror OTA . . . 44
4.9 Transistors sizing . . . 47
4.10 CMFB circuit for the current mirror OTA . . . 48
4.11 Biasing circuit for the current mirror OTA . . . 48

IV Results & Discussion **51**

5 Simulation Results **53**
5.1 Simulation Setup . . . 53
5.2 DC Simulations . . . 53
5.3 Frequency response . . . 54
5.4 Transient response . . . 54
5.5 Noise Performance . . . 56
5.6 Large-signal analysis . . . 56
5.7 Mismatch Performance . . . 56
5.8 Results Summary . . . 62

6 Discussion **65**
6.1 Frequency Response . . . 65
6.2 Transient Response . . . 66
6.3 Noise . . . 66
6.4 Linearity . . . 67
6.5 Comparison . . . 67

7 Conclusion **70**
7.1 Future work . . . 70

Appendices **73**

A Transient Analysis **75**
A.1 Full spectrum for fully differential folded cascode OTA . . . 75
A.2 Full spectrum for fully differential inverter-based current mirror OTA . . . 76

B Testbenches and Schematics **77**
B.1 Schematic for fully differential folded-cascode OTA . . . 78
B.2 Schematic for fully differential inverter-based OTA . . . 79
B.3 Testbench for fully differential folded-cascode OTA . . . 80
B.4 Testbench for fully differential current mirror OTA . . . 81
B.5 DC operating points for fully differential folded-cascode OTA . . . 82
B.6 DC operating points for fully differential inverter-based OTA . . . 83

C Open loop gain curves and testbenches **84**
C.1 Open loop gain for fully differential folded-cascode OTA 85
C.2 Open loop gain for fully differential inverter-based OTA 86

Bibliography **88**

List of Figures

1.1 Cross-section ultrasound image of a fetus [2]. 4

2.1 Two-dimensional ultrasound array used in 3D ultrasound scanning [19]. 9

3.1 A typical amplifier characteristic curve [3]. 12
3.2 Single-tone spectrum with 1st, 2nd, 3rd and fifth order harmonics [5]. . 14
3.3 Spectral density [32]. 16
3.4 Flicker noise [32]. 17
3.5 CMOS technology scaling [32]. 18
3.6 Tdade-offs in CMOS IC. 19
3.7 Differential Amplifier. 21
3.8 Standard differential stage. 21
3.9 Two-stage OTA. 23
3.10 Two-stage Miller OTA. 24
3.11 Telescope cascoding OTA. 26
3.12 Fully Differential Folded Cascode OTA. 28
3.13 Fully differential Current Mirror OTA. 30
3.14 Feedback circuit. 31
3.15 Inverter-based differential amplifier. 32

4.1 Fully Differential Folded Cascode OTA. 40
4.2 Biasing circuit for the folded cascode OTA topology. 43
4.3 CMFB circuit for the folded cascode OTA. 44
4.4 Fully differential inverter-based current mirror OTA. 45
4.5 Small-signal model. 46
4.6 Commom-mode feedback circuit for current mirror OTA. 48
4.7 Biasing circuit for the current mirror OTA. 49

5.1 DC gain and phase margin for folded cascode OTA. 54
5.2 DC gain and phase margin for inverter-based current mirror OTA. . . . 55
5.3 Transient response for folded cascode OTA. 55
5.4 Transient response for current mirror OTA. 56
5.5 Output noise for folded cascode OTA. 57
5.6 Output noise for current mirror OTA. 57
5.7 Large-signal gain and f_{ug} for folded cascode OTA. 58
5.8 Large-signal gain and f_{ug} for current mirror OTA. 58
5.9 Fundamental and 2nd harmonics for folded cascode OTA. 59
5.10 Fundamental and 2nd harmonics for current mirror OTA. 59
5.11 Harmonic statistics for folded cascode OTA. 60
5.12 Harmonic statistics for current mirror OTA. 61

A.1 Transient response. 75

A.2 Transient response. 76

B.1 Virtuoso schematic. 78
B.2 Virtuoso schematic. 79
B.3 Virtuoso schematic. 80
B.4 Virtuoso schematic. 81
B.5 DC operating points. 82
B.6 DC operating points. 83

C.1 Open loop gain characteristic curve. 85
C.2 Testbench for open loop gain. 85
C.3 Open loop gain characteristic curve. 86
C.4 Open loop gain characteristic curve. 86

List of Tables

4.4.1 STD and HG NMOS transistor parameters. 38
4.4.2 STD and HG PMOS transistor parameters. 39
4.5.1 Assumed values for g_m/I_d, gm and g_m/g_{ds}. 42
4.5.2 Simulated dimensions for all transistors. 42
4.5.3 Optimized dimensions. 42
4.6.1 Optimized dimensions. 43
4.7.1 Optimized dimensions. 44
4.9.1 Assumed values for g_m/I_d, gm and g_m/g_{ds}. 47
4.9.2 Simulated dimensions for all the transistors. 47
4.9.3 Optimized dimensions for all the transistors. 47
4.10.1 Optimized transistors sizing for CMFB circuit. 48
4.11.1 Optimized transistors sizing for the biasing circuit. 48

5.7.1 HD1 and HD2 for folded cascode OTA. 60
5.7.2 HD1 and HD2 for current mirror OTA. 60
5.8.1 Specified parameter results summary. 62

List of Abbreviations

ADC	Analog to Dogital Converter
ADE	Analog Design Environment
CMFB	Common-mode feedback
CMOS	Complementary Metal-Oxide-Semiconductor
CMUT	Capacitive micromachined ultrasonic transducer
DC	Direct Current
ECG/EMG	Electrocardiogram/Electromyogram
EEG/EOG	Electroencephalogram/Electrooculogram
FOM	Figure-of-merit
HD	Harmonic Distortion
IVUS	Intravascular Ultrasound
LNA	Low Noise Amplifier
OTA	Operational Transconductance Amplifier
PVT	Process, Voltage and Temperature variations
ROI	Region-of-interest
SINAD	SIgnal-to-Noise And Distortion
SNR	Signal-to-Noise Ratio
VGA	Variable Gain Amplifier

Part I

Introduction

Chapter 1

Introduction

This master thesis investigates two well known CMOS circuit topologies, in order to design a low noise energy efficient front-end amplifier for medical ultrasound equipment. The output electrical potential from a human body is low quality and low amplitude which can be problematic for a sensing probe. In edition, the quality of these biopotentials will also depend upon the locations used as references for the measuring devices. In response to this, the thesis explores the energy efficient LNA design architectures which may be used to improve the quality of the detected biopotentials as well as amplifying those signals. Finally, the thesis also concerns the comparison based on a defined FOM as well as suggestions for selecting the most successful architecture out of the two presented architectures.

1.1 Motivation

In today's healthcare challenging environment, changes are happening very fast. This unprecedent emergency must be addressed quickly, precisely and with versatility. Furthermore, diagnosis and treatment of the hidden diseases in the human body, has been one of the most challenging part in the field of medical research. By using modern medical equipment like ultrasound imaging (sonography), today, it is possible to view inside the human body and therefore, making it possible for a physician to evaluate, diagnose and treat many of these hidden diseases.

An ultrasound examination is done by placing a probe (transducer) directly on the human body or inside an body opening. High-frequency sound waves are reflected back and captured in real-time which provide necessary information about the movements of the internal organs as well as blood flowing through the blood vessels. Ultrasound examinations are considered generally safe under certain acceptable conditions since they do not produce any ionizing radiation which may be harmful as in the case of x-rays [2]. Some common ultrasound imaging procedures include abdominal ultrasound, bone sonometry, doppler ultrasound and ECG etc. For example, ultrasound examination is widely used to monitor the health status of the mother and the fetus. The growth of the different body parts of the fetus is observed as shown in the figure 1.1. Thus, ultrasound imaging has played a vital role in the field of medicine and has successfully been used for over 20 years with an excellent safety record [7].

Traditionally, two-dimensional imaging systems, like doppler ultrasound or harmonic imaging suitable for different applications, have been used for many years. Three-dimensional

Figure 1.1: Cross-section ultrasound image of a fetus [2].

systems were complex and challenging for the designers interms of their cost (chip area), energy efficiency and data transmission. Recent developments in the field of integrated circuit technology, have made it possible to implement such complex three-dimensional systems. For example, Intravascular ultrasound, IVUS, has become quite trendy now a days where a plaque in the blood vessels can be detected by entering a catheter with a probe into the veins. The operation is very sophisticated and the probes need to be very small [3], [17].

As part of the ongoing development in the CMOS technologies, the technology downscaling has made it possible to implement such sophisticated operations at nanoscale levels. The main objective of the downscaling is to produce costless and robust integrated circuits. Simultaneously, the designers have to face many challegges like requirements for speed, accuracy, power consumption, linearity and stability of the integrated circuits.

The thesis problem relate to several interesting investigations by exploring two different CMOS architectures. Accuracy requires more power which becomes big challenge for the analog designers since IVUS demands very low power level. The transducer will detect a very weak ultrasound echo which needs to be amplified by adding limited noise or distortion. Therefore, the performance parameters for the LNA are of great interest and importance. Furthermore, the image quality sets limits for the second harmonic distortion (HD2) [17].

Finally, the thesis investigates the performance of two fully differential CMOS topologies, one with the folded cascode OTA and the second with the inverter-based current mirror OTA. Specially, the specified performance parameters like SNR, power consumption and phase margin of the two topologies are compared to each other. In addition, the second harmonic is compared with the fundamental harmonic for each topology. A commercially available 130 nm CMOS technology has been used in the entire design. The design schematics and testbenches are made in Virtuoso (Cadence). The design verifications and simulations are performed by using ADE (Analog Design Environment) tool in Virtuoso. On the basis of the simulation results, the current mirror OTA topology is suggested because of its great advantage of low power consumption.

1.2 Main Contributions

The main contributions in this thesis work, are the design and verification of energy efficient low noise LNA by using a 130nm CMOS technology. The design elements contributed in the design phase are as follows:

- a fully-differential folded cascode OTA
- a fully-differential inverter-based current mirror OTA
- a common-mode feedback circuit for both topologies
- a biasing circuit for both topologies
- a gain boosting configuration for the current mirror OTA

1.3 Thesis Outline

This thesis work includes 4 parts and 7 chapters. A short description of the thesis work is given below:

- Part 1 includes only chapter 1 which introduces the objective of the thesis.
- Part 2 presents the necessary theory to carry out the thesis work. This part includes the following chapters.
  - Chapter 2 presents the basic principles and background theory of the ultrasound imaging systems.
  - Chapter 3 presents the basic theory of amplifiers, particularly LNAs. Also, performance parameters, CMOS technology, and available architectures are studied in this chapter.
- Part 3 includes only chapter 4 which describes the methodology adopted for the design, and the proposed architectures used for the design are explained and analyzed.
- Part 4 includes the following three chapters.
  - Chapter 5 which shows the final results achieved from the simulations run for the different performance parameters for both of the topologies.
  - Chapter 6 discuss the final results obtained from the simulations.
  - Chapter 7 concludes on the basis of findings from the final results discussed in chapter 6. In addition, a future work is also proposed in order to improve the design.
- Appendix includes schematics, testbenches and resulted graphs obtained from the simulations.

Part II

Theoretical Background

Chapter 2

Ultrasound Imaging Systems

The ultrasound imaging systems are frequently used in clinical diagnosis, specially human arteries and vascular systems are examined in real-time since this imaging is non-invasive. The ultrasound imaging and its applications are being used successfully for more than 50 years. In the beginning, sonar and modified radar systems were used while conducting the experiments. Since then the advances in ultrasound imaging systems have been evolved rapidly and they are used almost in all the hospitals for diagnostic purposes [14]. Generally, ultrasound analysis are regarded as safe since it is non-invasive and no ionizing radiation is involved. Usually, the patients are not distressed due to ultrasound examinations [19].

2.1 Basic ultrasound principle

The desired location for the ultrasound examination is navigated by passing a catheter inside the artery of a human body. Center frequencies vary in between 2 and 15 MHz for medical ultrasound. Thus the speed of sound c at these frequencies is around 1540 m/s in a human body. The wavelength can be calculated as

$$\lambda = \frac{c}{f_c}$$

Where f_c denotes the center frequency. The expression gives a wavelength range from 0.77 to 0.10 mm. Thus a region of interest (ROI) may be investigated by focusing the sound at these frequencies. The width of the transducer will determine the resolution. The larger the transducer, the higher the resolution. A short ultrasound pulse can be used to determine the axial resolution. The signal power loss will be higher at higher frequencies since the dispersive attenuation of sound waves will limit the penetration depth of the ultrasound. Thus a typical system will have a specified dynamic range. Also, the depth of the region to be imaged will determine the probe selection [14].

2.2 Two-dimensional Imaging

For many years, traditional two-dimensional imaging has been practiced which shows a plane cross section image, perpendicular to the artery. A side-looking image is produced which may be problematic when the catheter touches the plaque in the artery. In order to observe the blockages ahead, a forward-looking method should be adopted [3].

2.3 Three-dimensional imaging

Three-dimensional ultrasound has got great attraction in the field of research and business for the last 10 years. The traditional freehand scanning systems are being replaced by three-dimensional real-time systems but still the speed is the major challenge. However, the data acquisition rate has been increased after the introduction of the parallel beam formation as shown in the figure 2.1.

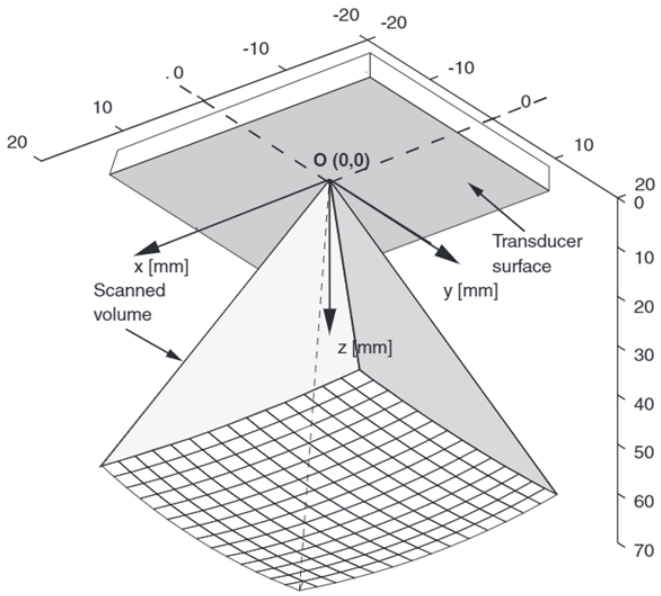


Figure 2.1: Two-dimensional ultrasound array used in 3D ultrasound scanning [19].

2.4 Front-end electronics

The overall system performance will depend on the analog signal processing components in the front-end. It is not possible to compensate for noise and distortion once they have been added in the input signal. The front-end electronics generally consists of an LNA and ADC. The LNA plays a key role in the front-end electronics of a ultrasound imaging syatem. The main objective of this LNA is to amplify a low amplitude weak signal and avoid the undesirable noise where as an ADC is used to analog to digital conversion. Thus, the overall role of the frond-end electronics will be to minimize the distortion and maximize the signal-to-noise ratio (SNR) by exploiting the early echo amplitudes since they are larger. Since, this thesis work is limited to the LNA design so focus will be more on the LNA in the rest of the thesis.

The pressure waves detected from the human body are converted to time varying signals by using several micromechanical ultrasound transducer (CMUTs). Each of the CMUT is connected to an LNA which amplifies the received signal to a certain level suitable for sampling. An image may be produced then by reconstruction of the ultrasound echoes. There are several advanced techniques available for this process. Human tissues generate harmonic frequencies due to their nonlinear behavior. The type of the tissue under examination may be characterized by listening the 2nd harmonic echoes as well as the fundamental harmonic echoes. Thus, the exploitation of these properties have made it possible to improve the image quality [3].

Chapter 3

Amplifiers

Amplifiers are used in almost every type of electronic devices. They act like active electronic devices in the circuits and the purpose of their use is to amplify the input signal to a certain desired level without any addition of undesired interference. A suitable design will depend particularly on the specifications which are further based on the several parameters of interest. These parameters are called the performance parameters and will be defined in this chapter.

3.1 Low Noise Amplifiers

There are several types and applications of amplifiers. A particular design of amplifier will generally depend on the application, it will be used for. As mentioned in section 2.4, the motivation of this thesis is to design an LNA capable of detecting and analysing ultrasound measurements that can appreciate many aspects of different measurement environments. Acoustic waves propagating in the air are actually the sound waves and acoustic waves having frequency higher than 20 kHz are known as ultrasound. In an ultrasound measurement, the signal amplitude is observed with respect to time. Three main techniques are used to observe these measurements which are known as standing wave measurements, transit-through and pulse-echo measurements [8]. Out of these three, the pulse-echo measuring technique is considered while the design phase. According to this technique, a single transducer is used to transmit and receive the signal. Transducer transmits a pulse of ultrasound energy which is then reflected back from the medium of interest and then the changes in the acoustic properties are analyzed [11]. Two ultrasound transducers are used in transit-through where as transmitter and receiver has continuous transmission in standing wave measuring technique.

The receiver part of the transducer is implemented by a Low Noise Amplifier (LNA). As mentioned earlier in section 2.4, the purpose of this LNA is to amplify the received signal and thus it will act as a buffer which is capable of proving some gain to the weak signal. If the amplified signal is strong enough, the total signal-to-noise (SNR) ratio will not be that much affected by the rest of the components in the chain. The attenuation in the signal path is random which is difficult to compensate and therefore, a Variable Gain Amplifier (VGA) is used to handle this difficulty. In this way, the image quality is also improved [9].

3.2 Performance Parameters

The LNA dominates the sensitivity factor in an front-end analog system and that is why an LNA is one of the most important component in a whole analog system. Its performance will play a key role in the overall system performance. There are many trade-offs among the different performance parameters while designing high-performance amplifiers. These trade-offs set many limits and challenges for the analog designers. Some of the important parameters are power consumption, gain, speed, bandwidth, linearity, output swing, input impedance, output impedance and noise. The design should provide a certain level of SNR even at very low input signals. The important performance parameters are described in the following sub-sections.

3.2.1 The Gain

The gain or DC gain is one of the most critical performance parameter, specially for the low-frequency applications. It can be defined both for small-signal and for large-signal. In general, the gain is ratio between the output signal power to the input signal. Both small-signal and large-signal expressions are given below.

$$A_v = \frac{dV_{out}}{dV_{in}} = \frac{v_{out}}{v_{in}} \tag{3.1}$$

$$A_{vl} = \frac{\triangle V_{out}}{\triangle V_{in}} \tag{3.2}$$

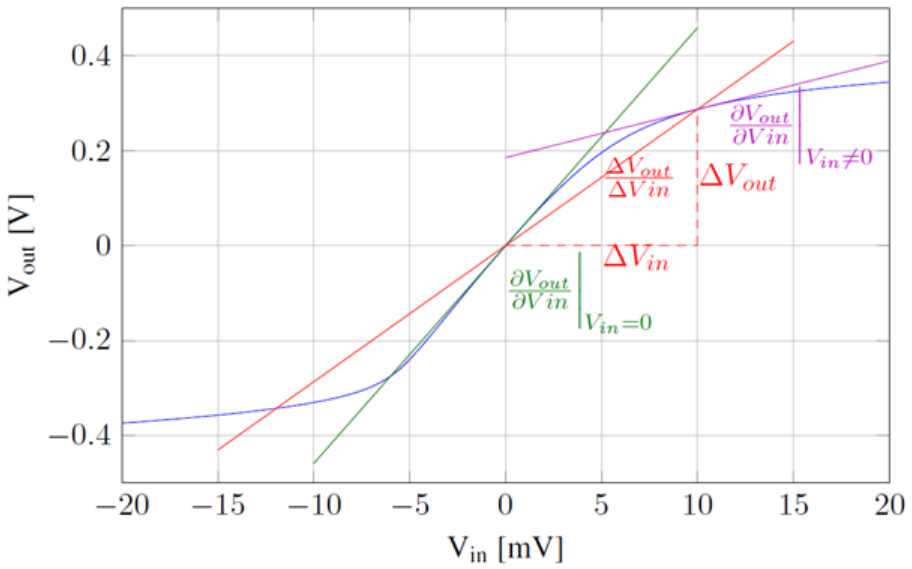


Figure 3.1: A typical amplifier characteristic curve [3].

The unity gain frequency of the amplifier determines the speed of the amplifier. The phase margin (PM) will determine the stability of the amplifier.

3.2.2 Frequency response

When we are dealing with real frequencies, we need to do s-plane analysis. For the voltage gain, the transfer function $A_v(s)$ becomes a function of s no where $s = j\omega$ (real frequency).

So,

$$A_v(j\omega) \Rightarrow Frequency\,response \tag{3.3}$$

In analog CMOS, there is only one frequency dependent element and that is a capacitor. CMOS circuits are typically limited to $\leq 1\,GHz$. Inductors could also be used but in that case, it would be RF CMOS rather than analog CMOS and the frequency is then limited by $\leq 30\,GHz$. For the analog CMOS, the inductors could be too large for the on-chip area because the frequency is low ($\leq 1\,GHz$) [32].

So, the impedance of a capacitor, when using s-plane analysis, is given by,

$$z = \frac{1}{sC} \tag{3.4}$$

There are lot of capacitors on a chip. First of all, there is a capacitor between every node and ground. So, every node in a CMOS circuit will have some capacitance to the ground. There will also be a capacitance between any two nodes. Of course, there will be a small capacitance between the two nodes that are far apart from each other but there will be a considerable capacitance between the two nearby located nodes. Since, the impedance of a capacitor goes down in frequency, there will be a signal loss as the frequency gets higher. So, analog CMOS circuits are typically, band pass or low pass circuits. So, too high pass circuits are not desired here because if frequency gets high enough, there will be a shunt capaciror to the ground which will eventually shunt the signal. So, always band limiting functions are appreciated [32].

3.2.3 Linearity

In an ideal linear system, the level of output signal is directly proportional to the level of input signal. But in a real world system, over an unlimited signal levels, this performance is not achievable. However, for real world practicle systems, an acceptible range of signal levels is defined, also known as dynamic range of a component [21]. Dynamic range is an important performance parameter for high accuracy applications. Linearity becomes important performance parameter as the signal swing becomes large and a non-linear circuit will leads to harmonic distortion. Harmonic distortion is simply the undesired multiples of the fundamental frequency. These are nonlinearities in a device which add overtones in the spectrum. Number of harmonics will depend on the bandwidth of the system. Linearization techniques are applied to the non-linear circuits. The linearity of a CMOS technology based circuit is generally determined from the input transistors [32].

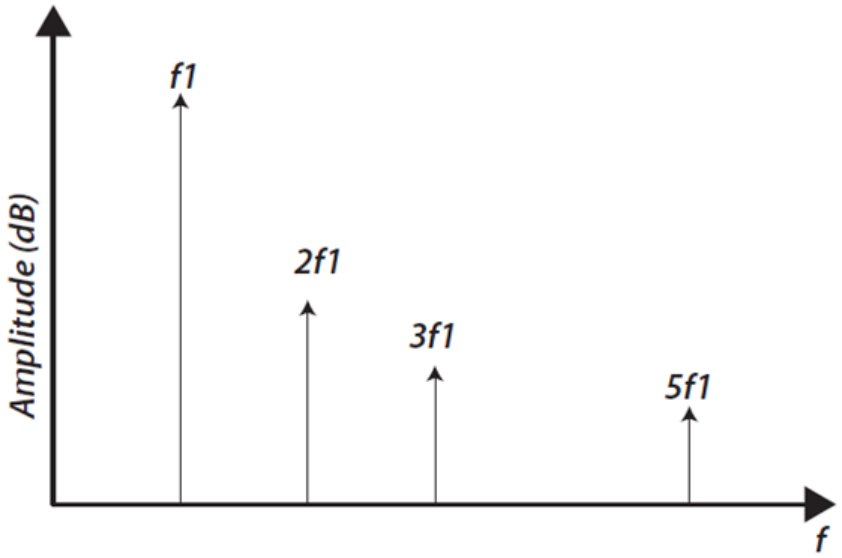


Figure 3.2: Single-tone spectrum with 1st, 2nd, 3rd and fifth order harmonics [5].

3.2.4 Noise Analysis and Modeling

Noise has been a well known problem for the analog designers since it involves trade-offs with other important performance parameters like speed, linearity and power consumption. It degrades the desired signal quality by limiting the minimum signal level [23]. Therefore, analog designers must have a good understanding of noise analysis and noise sources in their toolbox in order to be able to design a high performance analog integrated circuit.

Noise is not predictable at any time and therefore, it is a random process. There are many sources of noise. The focus will be on the noise that is generated inside the devices and is visible at the terminals, known as inherent device noise. For example, a resistor will typically generate thermal noise which is possible to measure by looking at the terminal voltage of the resistor or the current in the resistor. Transistors will also generate noise. So, these two are the main elements in CMOS circuits that will generate noise [32]. Thermal noise and flicker noise are the two dominate types of noise in a CMOS transistor. Of course, other circuit elements will also generate noise, for example, diodes will generate noise. Real capacitors and real inductors will generate noise because there will be parastic resistance and an inductor on-chip will have parastic resistance in series, so, it can be said that inductors will also generate noise. Ideal inductors and ideal capacirors will not generate noise. A circuit will also have interaction with the other parts of the circuit and with the outside world which will result in interference noise. So, there will be noise on power lines, rails will have noise typically from other circuits, specially from digital circuits. So, digital circuits generate a lot of switching noise which will influence analog circuits. But interference noise may or may not be random and is not considered in this thesis work.

A time-domain analysis of noise signals, frequency domain analysis of noise signals and noise models are needed for the analog designer's toolbox. Time-domain analysis include rms value, signal-to-noise ratio (SNR) and noise summation. For hand calculations, it is assumed that all the noise sources are uncorrelated otherwise, it would not be possible to do hand calculations.

RMS value

The root mean square or rms voltage will give an average value and is defined as,

$$V_{n(rms)} \equiv \left[\frac{1}{T}\int_0^T V_n^2 dt\right]^{\frac{1}{2}} \tag{3.5}$$

which is simply the square of the voltage integrated over a certain time and divided by the time T. Here the time interval is from 0 to T. It can also be done for other quantities for example rms current is defined as,

$$I_{n(rms)} \equiv \left[\frac{1}{T}\int_0^T i_n^2 dt\right]^{\frac{1}{2}} \tag{3.6}$$

A better estimate for rms value may be achieved by integrating for longer time interval. The square of the rms value presents the normalized noise power. So, in analog CMOS, real power is not under discussion since resistance is not included [32]. The traditional power is defined as,

$$P = VI = \frac{V^2}{R} \tag{3.7}$$

$$P = V^2 for R = 1 \tag{3.8}$$

Since the resistance is not considered, it is set to 1 in the expression which then gives the normalized noise power of the signal. So, the noise power is just the square of the rms values [4].

SNR

The signal-to-noise ratio or SNR of a signal is defined as,

$$SNR \equiv 10\, log_{10}\left[\frac{P_s}{P_n}\right] dB \tag{3.9}$$

where P_s is the signal power and P_n is the noise power. Both are normalized to 1 because typically, the ratio of signal power to the noise power is of interest since any resistor can be chosen then. The SNR for the normalized signal power and normalized noise power can be written as,

$$SNR = 20\, log\left[\frac{V_s}{V_n}\right] \tag{3.10}$$

Frequency domain analysis will present noise spectral density which tells how much noise power per Hertzs do we have. For example, we have a certain noise power from 1 to 2 Hertzs and we characterize this with noise spectral density. There is also white noise with flat spectrum and it contains all frequencies.

3.2.5 Noise spectral density

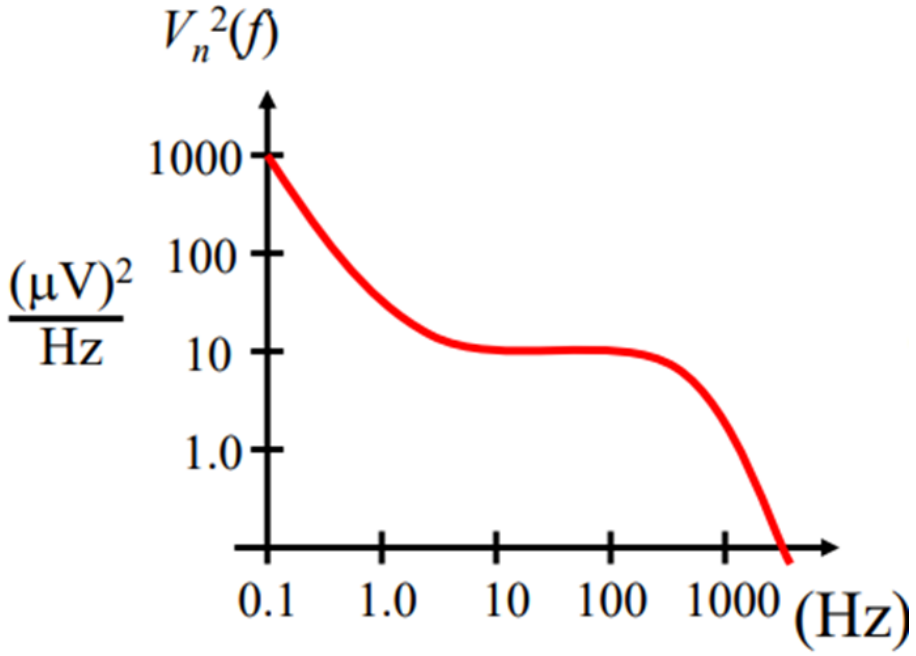


Figure 3.3: Spectral density [32].

In a noise spectral density, a noise voltage is squared and plotted versus frequency. This is a typical curve for a transistor. At low frequencies, there is flicker noise and noise spectral density decreases with frequencies. The flat part in the curve is typically the thermal noise. At high frequencies, there will be noise filtering because of the capacitors. The filters do not just filter the signals but also filter the noise. In the frequency domain, it is also possible to get rms values by integrating the noise spectral density over all frequencies. By taking the square root of noise spectral density, root spectral density can be achieved. White noise has flat spectrum and flicker noise is typically $1/f$-noise at low frequencies but it could be a different slope and the slope may depend on frequency in that frequency range. After the thermal noise, at a certain frequency, the flicker noise is equal to the thermal noise and that is called the noise corner. The noise corner moves with the size of transistor and biasing of transistor. In order to get rid off flicker noise, the signal should be at frequencies which are above the noise corner. This technique is called spectrum planning [32]. For example, mixers are used to mix the signals in radio receivers, if the carrier frequency is 2.4 GHz the signals are typically mixed very early in the signal processing chain. But it might not help in CMOS receivers because here one must take into account where the noise corner is. Typically, mixing is done down to a certain level, called the intermediate frequency. This frequency is not very low and it should typically be around 10 MHz in order to stay away from the flicker noise region. There is also telegraph noise, shot noise and lot of different types of noise [32].

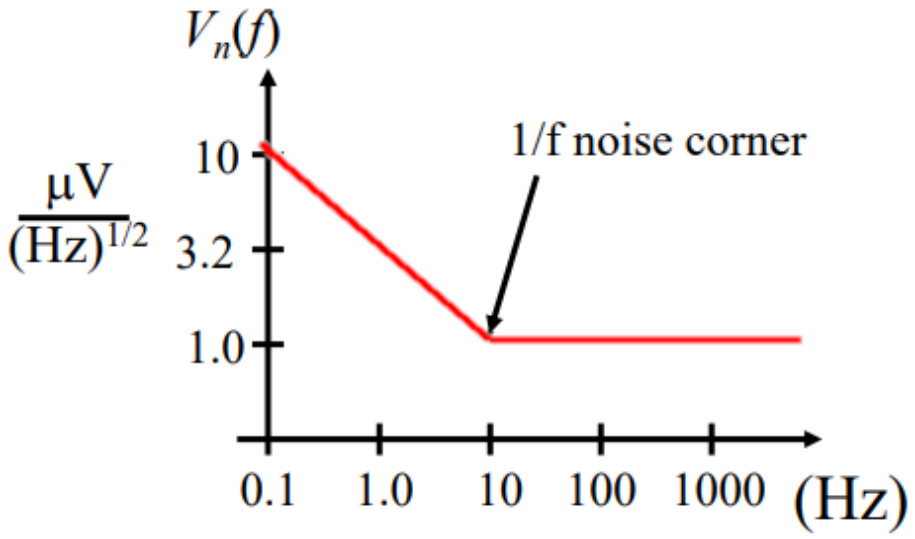


Figure 3.4: Flicker noise [32].

Thermal noise

Thermal noise, also known as Nyquist or Johnson noise, is the most basic type of noise and is generated by thermal agitation of bound charges since carriers have thermal energy or thermal movements [21]. Thermal noise has white spectral desnsity and is proportional to the absolute temperature. It also depends on the length of the transistor because of its sensitivity against the bias current. Thermal noise can be modelled as a noise current source [1].

Flicker noise

The source for flicker noise is not yet fully understood. Researches are having debates about the real cause of the flicker noise. So, flicker noise is typically just considered empirical. It is observed and then empirical fitting is applied to measure it. There is still no physics for that [32].

Shot noise

Shot noise was introduces while studying the vacuum-tube diodes. Shot noise often dominates over the thermal noise during the subthreshold region of operation [10].

3.3 CMOS Technology

The modern integrated circuit design is implemented by using both CMOS and analog circuits. CMOS implementation often dominates and becomes the best choice for the designers since it is cheaper and cost is the most important parameter for the companies to be considered. Analog design for printed circuit board (PCB) is different from analog design for integrated circuits. Discrete design of PCB is implemented by operational amplifiers (Opamps) while IC design is implemented by Operational Transconductance Amplifiers (OTA). The difference between the two is that Opamps have low output resistance and the OTAs have high output resistance. More power is required to implement low output resistance [32].

More over, digital circuits take the advantage of down scaled technologies due to their small area, low power supply and less parasitics. Low parasitics will also improve the energy efficiency of the device and higher bandwidths are possible to achieve [17]. Generally, in order to reduce the cost, both analog and digital circuits are desired to to be implemented on the same system-on-chip (SOC) by using a single available technology. A chip which is less than 50nm becomes transparent and flexible. 7nm is the smallest transistor size on-chip today. It seems to be moving to 5nm in the near future [32].

The main objective is to produce low price, energy efficient circuits with reliable quality and accuracy. With the increasing downscaling of CMOS technologies to the nanometer lavel, there are several new challegnes and nonidealities to be considered for the analog designers. Several nonidealities and their effects have been reported in [16]. For example, hot carrier injection and time-dependent dielectric breakdown limiting the power supply, lithographic and stress limiting the accuracy, mobility and leakage reduce the performance of the device and there are many more. Thus, it is very important to understand and minimize the effects of these physical phenomena in order to improve the reliability and performance of the circuit. To enhance the accuracy and bias current temperature dependency on-chip, some techniques have been estabilished in [16]. Of course, there will be more physical limits and design challenges in the future CMOS technology. According to Moore's law, the circuit complexity is almost doubling for every 18 months as shown in the upper part of the figure 3.5. The transistor gate length decreasing to one half for approximately every 5 years as shown in the lower part of the figure 3.5.

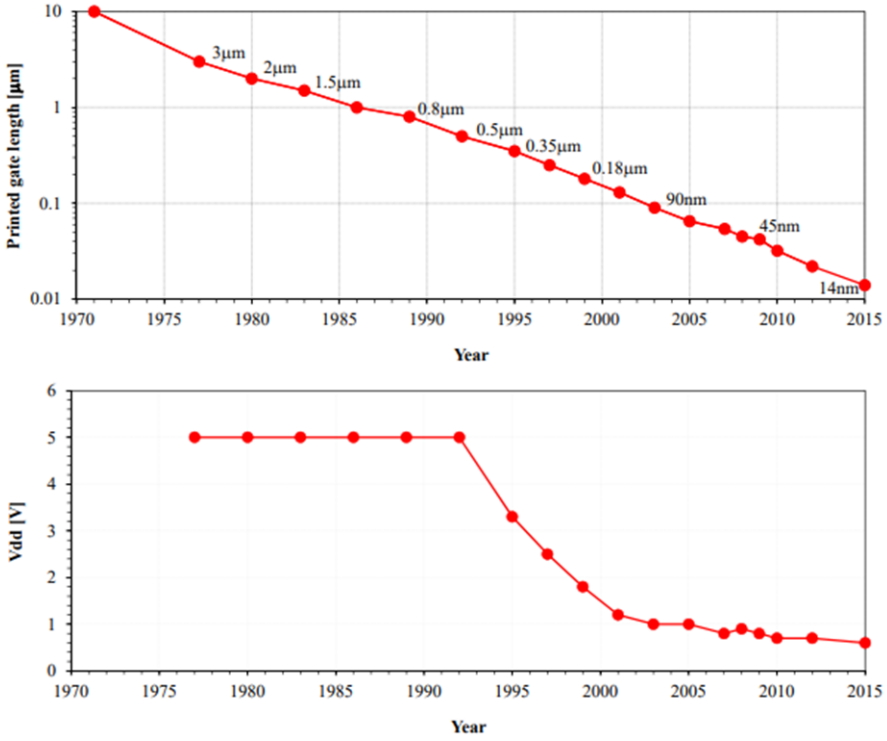


Figure 3.5: CMOS technology scaling [32].

3.4 Figure-of-Merit

A figure-of-merit (FOM) is a measure of energy efficiency of a device and is helpful in comparing different topologies [32]. There are several definitions of figure-of-merit (FOM) depending on the application. A well defined FOM ensures the design quality but in some cases, it may be difficult to define a meaningful figure-of-merit. Performance matrix of analog circuits is determined by the following three factors.

- Accuracy (Dynamic Range)
- Speed
- Power consumption

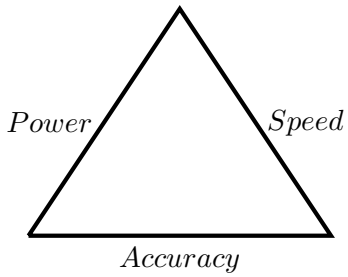


Figure 3.6: Tdade-offs in CMOS IC.

For high speed applications, high gm with higher bias current and small capacitors is desired but at the same time, lower bias current and lower capacitance is required for low power applications. On the other hand, higher capacitance and high gain are more appreciated for applications where high accuracy is required. So, there is always a trade-off among several performance parameters, therefore, a good figure-of-merit is needed to be assumed early in the design phase. The performance parameters of a MOSFET transistor changes with scaling down the CMOS process technologies. Downscaled feature sizes also limit the supply voltage requirements which are challenging for analog circuit designers [33]. A commonly used fogure-of-merit is defined as,

$$FOM = \left[\frac{Power}{DR^2.BW}\right] \tag{3.11}$$

where Power is actually the power consumption, DR represents the dynamic range and BW represents the bandwidth of signal which can be handled. Here the FOM is related to the three fundamental properties of analog circuits as accuracy is determined by the intrinsic gain of the transistor, speed of the transistyor is determined by the unity gain frequency and finally, power consumption is related to gm/I_d. The largest gm/I_d is in weak inversion and is typically monotonic. It tells how good the transistor is in translating a given bias current into transconductance [32]. Transductance and bias current will affect the FOM. The objective is to minimize the bias current to save the power but there are also requirements for the transconductance which should be satisfied. Therefore, one must be careful while looking at any figure-of-merit of interest.

3.5 CMOS Architectures

The use of MOS-transistors in amplifiers can be implemented in several different ways. They are also called as voltage amplifiers where a current is passed through a load resistor

in order to generate the desired voltage. Several CMOS topologies with their advantages and disadvantages are available. Some of the topologies have been reviewed in this section. In addition, biasing and feedback concepts are reviewed and the following topologies are studied for this purpose.

- Single-ended versus differential operation
- Standard differential stage
- Two-stage OTA
- Two-stage Miller OTA
- Telescope cascode OTA
- Folded Cascode OTA
- Current Mirrior OTA
- Feedback amplifiers
- Gain boosting
- Cross-coupled pair

3.5.1 Single-ended versus differential operation

Single-ended means that signal is referenced to the ground which is assumed to be at some constant potential. On the other side, a differential signal will have two equal and opposite paths and then the voltage will be defined as the difference between the two voltages. The sign can easily be flipped by just flipping the two wires. Both terminals should have the same impedance level with respect to the ground. The differential circuits are more robust and stable to the external noise [23].

The signal swing is also twiced when we move to differential. This means that the amplitude becomes two times and that is four times the signal power. So, by going to differential, four times signal power is achieved automatically. The noise at the two output is completely un-correlated. In linear case, one must get double SNR or 3dB extra SNR automatically when moving to differential. Second order harmonic will typically dominate in a single-ended output amplifier since that is the biggest harmonic. But second order harmonic will be removed in differential case and instead third order harmonic is dominated.

One big drawback with the differential circuit is that we need some extra circuitry since the power consumption will be increased. Generally, it is assumed that for an accuracy of 8-bits or higher, it is good practice to use differential. For example, for a requirement of 60 dB SNR, differential topology will be better since 60 dB SNR means 10-bits accuracy. For a 6-bits of accuracy requirement, a single-ended topology will be better to save power and have less complex circuit. Differential circuits are used a lot since 7 or 8-bits accuracy requirements are quite common today.

3.5.2 Standard differential stage

The differential input is very commonly used in integrated amplifiers where a pair of input transistors is used. The block diagram of a fully-differential amplifier is shown in the figure 3.7.

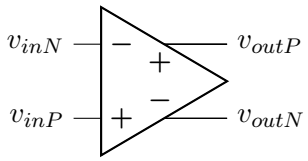


Figure 3.7: Differential Amplifier.

Such a stage is called differential stage which can be realized by implementing two common-source amplifiers as shown in the figure 3.8.

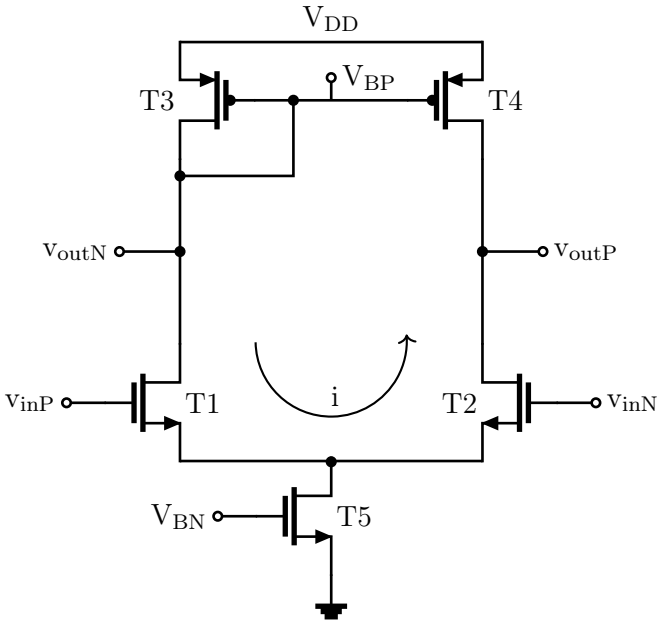


Figure 3.8: Standard differential stage.

Here are some important definitions for differential amplifiers.

$$v_{in} = v_{inP} - v_{inN} \tag{3.12}$$

$$v_{out} = v_{outP} - v_{outN} \tag{3.13}$$

The above two equations are called differential input and differential output voltages respectively. The potential in the center of the differential signal is known as the common-mode (CM) voltage level which can also be expressed for the two common mode voltages.

$$v_{CM_{out}} \equiv \frac{v_{outP} + v_{outN}}{2} \tag{3.14}$$

$$v_{CM_{in}} \equiv \frac{v_{inP} + v_{inN}}{2} \tag{3.15}$$

Signal processing is used to implement equal amplitudes but opposite phases which is difficult to implement perfectly. When analyzing these circuits, the signals are always assumed differential which means that all voltages will have the same amplitudes but with a 180^o phase shift between them. Average voltages do not change which is an approximation. Because in reality, even DC voltages will also move due to the noise. Small signal operation can be analyzed very easily since the averages do not move. This type of analysis technique

is called half-circuit analysis [32]. Here we have two common-source amplifiers and the gain and the output resistance is known. If we assume,

$$gm1 = gm2 = gm3 = gm4 = gm$$
$$rd_{s1} = rd_{s2} = rd_{s3} = rd_{s4} = rd_s$$

Then by diode connecting T3, the expressions for the gain and the output resistance can be written as,

$$A_v = gm\frac{rd_{s2}}{2} \tag{3.16}$$

$$R_{out} = \frac{rd_{s2}}{2} \tag{3.17}$$

The expressions are exactly the same as for the common-source amplifier. In case of negative gain, just the wires can be flipped to flip the design. Actually the designer needs to take more care about the following assumptions for the dimensions of the transistors [32].

$$\left(\frac{W}{L}\right)_{M1} = \left(\frac{W}{L}\right)_{M2}$$
$$\left(\frac{W}{L}\right)_{M3} = \left(\frac{W}{L}\right)_{M4}$$

So, it is possible to get high gain with the differential stage if the size of M1 is equal to the size of M2 and the size of M3 is equal to the size of M4. Of course, it is possible to get very high gain by cascoding. So, the designers try to make M1 and M2 identical and M3 and M4 identical by design. But when they come back after fabrication, they are never identical due to PVT variations.

When differential input stage is used then more power is needed since there will be two branches. If common-source topology is used, the signal current will be wasted and the gain will be halved. To solve this problem, a current mirror can be used in order to mirror the signal current or reuse the signal current. After the signal current is mirrored, one can get twice the signal current in the output resistance as twice the gain. Then we are back with the same gain. We also need bias voltages for the PMOS transistors. The bottom transistor is called a tail transistor with bias voltage and of course, we generate those bias voltages with current mirrors. This differential topology may be used with cascode or cascade gain stage depending upon the application it is used for.

3.5.3 Two-stage OTA

An OTA can only drive loads since it is basically an op-amp without the output buffer. This is the most power consuming block of an analog integrated circuit [15].

A two-stage OTA is composed of input and output stages. The first stage is, generally, a differential input stage and the second stage is a common-source output stage here. So, in two stage OTA topology, there are two rails. In design, two separate pins are used for V_{DD} and V_{SS} in order to make it possible to change the connections afterwords. Here we have two input voltages and one output voltage and one bias voltage. This can be changed to bias current if the diode connected PMOS is included. The resulted interface will be then bias current instead of bias voltage [32]. V_{DD} and V_{SS} is also a part of interface. Of course, small signal gain is of interest which may be found as,

$$A_v = \frac{v_{out}}{v_{in}} \tag{3.18}$$

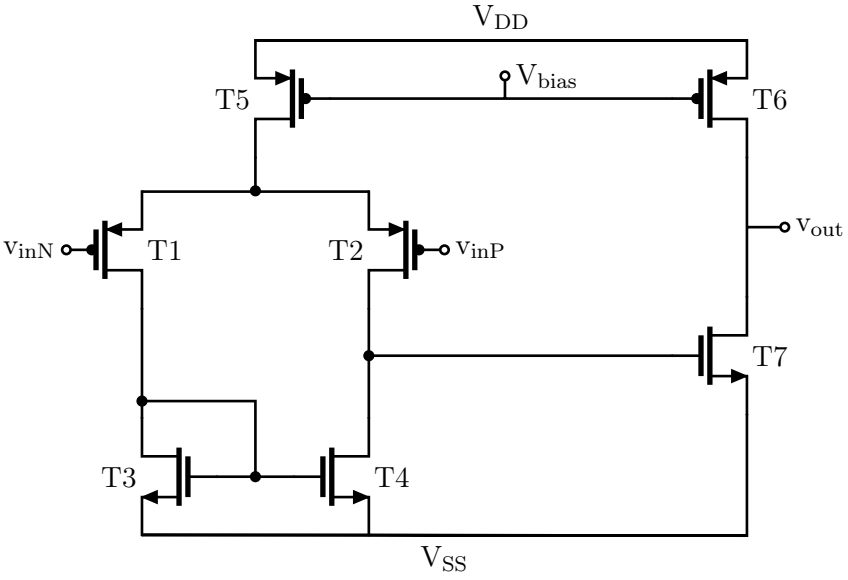


Figure 3.9: Two-stage OTA.

$$= \left[-gm7.(rd_{s7} \parallel rd_{s6}\right] . \left[gm1.(rd_{s4} \parallel rd_{s2}\right]$$

Assuming that,

$$gm1 = gm2 = gm3 = gm4 = gm5 = gm6 = gm7 = gm$$
$$rd_{s1} = rd_{s2} = rd_{s3} = rd_{s4} = rd_{s5} = rd_{s6} = rd_{s7} = rd_s$$

Finally, the expression for the gain becomes,

$$A_v = \left(gm\frac{rd_s}{2}\right)^2 \tag{3.19}$$

So, the gain is very high since it is proportional to $(rd_s)^2$ and is called intrinsic gain which is denoted by A_I. Intrinsic gain is a transistor property and is used as design equation during the design phase [32]. There are always several design equations in the design phase.

$$A_v = \frac{1}{4}\left(\frac{gm}{gd_s}\right)^2$$
$$= \frac{1}{4}\left(A_I\right)^2$$

$$A_I = 2\sqrt{A_v} \tag{3.20}$$

The gain distribution for multi-stage signal processing circuitry can be found by Friis formula. To have low noise, one should have high gain as soon as possible in the chip. If focus is more on noise then one may have more gain at one stage than at the second stage but typically, gain is same in both stages in the beginning [32]. All inputs are set to zero to find R_{out}. Thus all the control currents disappear as,

$$v_{inN} = v_{inP} = 0$$

The other property we can looked at is the input-reffered offset voltage. If you have different amplifiers and you connect the inputs together means DC operating point of the average. You expect zero at the output because the differential input voltage is zero. But in real world, you will get something else than zero. Therefore, there is a need to model the offset, means we need to apply a difference V_{OS}.

V_{OS} is typically separated into a systematic offset voltage and a random offset voltage. The random offset voltage is often dominated. In CMOS design, typical standard deviation for this voltage is $1\,mV$. One must use extremely large transistors if one would like to make it greater than $1\,mV$.

Also, M6 is different from M7 in terms of multipliers. In some books, it is mentioned that they will have same dimensions but they may have different multipliers since the current in the output stage is often scaled [32].

$$M6.I_{bias} = M7.\frac{I_{bias}}{2}$$
$$M7 = 2.M6$$

Most important is the range of common-mode voltages we can have. This can be found by moving the common-mode voltage to the lowest and the highest values at the input and observe the region of operation of the transistor before it dies. Low voltage design, for example 0.5 V design, is very challenging. g_{sg} is interested in case of PMOS since g_{sg} is +ve for PMOS. g_{sg} does not change much when we are forcing high bias. In triode region, the current source will drop the current and amplifier performance will be bad [32].

3.5.4 Two-stage Miller OTA

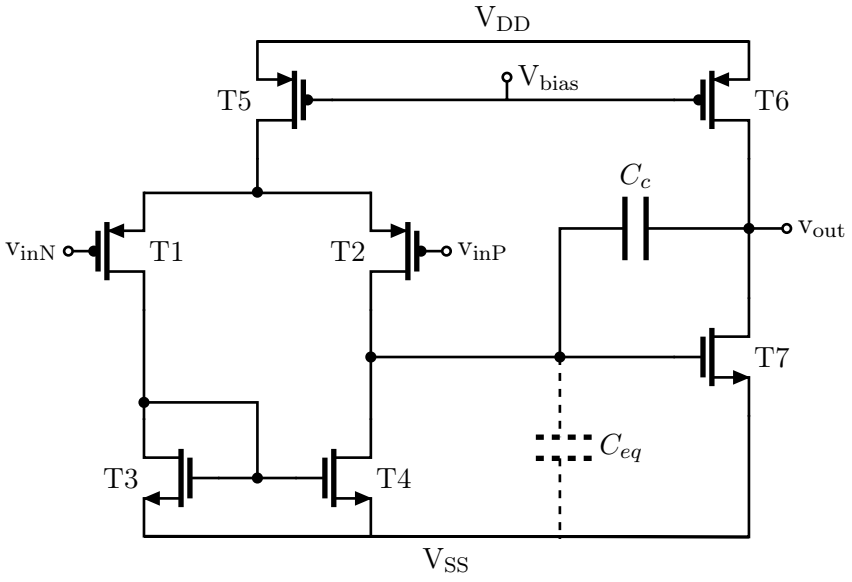


Figure 3.10: Two-stage Miller OTA.

Now we can analyze the two-stage OTA at high frequencies and applied feedback. For example, if we have very high gain like 500, 1000 or 2000, it is not to be used in open loop, so, we always add -ve feedback. When we apply -ve feedback, we have to check for the stability. We have to quikely find the expressions for the pole frequencies. An

efficient analog designer will look at the nodes and will quickly be able to write down the estimates of pole frequencies. Typically we have three nodes for each transistor to turn and every node will have a pole associated with it and every pole have a certain pole frequency associated with it. Voltages and resistances may be found by just looking into the nodes. We are only interested in the impedance level. So, high impedance means low pole frequency because pole frequency is inversely proportional to the some R and some C.

$$\omega_p = \frac{1}{RC} \tag{3.21}$$

So, the poles with high pole frequencies are not that much interesting in the beginning of the design phase at least. We are always interested in finding the poles with the lowest frequencies because those are the band-limiting or speed-limiting poles. So, for example, ω_{p1} and ω_{p2} will have the same value because both are associated with nodes that are at high impedance level. One might be very unlucky if ω_{p1} and ω_{p2} are very similar because then one might get a -180^o phase shift over a very short range of frequency since there are two poles. One pole will shift by -90^o and two poles will together will shift the phase by -180^o. This will create a trouble if loop gain is still larger than 1. So, for a two-stage amplifier, when the output is connected to the -ve input, the result will appear in an unstable amplifier. The reason is that as the phase shift becomes too large, the phase margin (PM) becomes extremely small, even may be -ve. This will be really bad since a +ve PM is always desired [32]. Then it comes two-stage Miller OTA into the picture where frequency compensation is applied. Frequency compensation is only related to the feedback systems. This can be done in several ways as follows.

- One of the two poles may be moved up or down. We have two poles at about the same frequency and we have to move them further apart. This technique is called splitting the poles.
- We can also apply pole-splitting but then we move both of the poles. One is moved up in frequency and other is moved down in frequency.
- We can take a zero and try to cancel pole 2 by putting the zero at the same frequency as the pole.

So, zeros can also be used to cancel the poles. But this is very difficult to implement in integrated circuits because PVT variations will move poles and zeros differently. So, at the nominal case, if you cancel a pole with zero, the pole will not be cancelled if the temperature is high or low supply voltage or slow or fast process. Therefore, typically we apply the first two techniques. We take the advantage of zeros to increase the PM but not to canceling the poles. We can chose the dominant pole at any of the two nodes and select it at the node 1 as we are using two-stage Miller OTA. We can take advantage of the Miller effect to implement the capacitor with a certain size with a smaller physical size because this stage has -ve gain. We connect it from the output of the first stage to the output of the second stage and call it compensation capacitor (C_c). The effective capacitance from this node to the ground is much larger than C_c, and is approximately C_c times the magnitude of the gain of the 2nd stage [32].

$$C_{eq} \approx C_c.A_2 \tag{3.22}$$

3.5.5 Telescope cascoding-OTA

Now we move to some more advanced type of OTAs which is an alternative of the two-stage Miller. These are load compensated means that the dominant pole is at the output node.

So, if one want to increase PM, one would typically increase the load capacitance. The more you load your amplifier, the more stable it is. It was opposite in case of two-stage Miller OTA. If you load Miller OTA more at the output, you will have less PM until the two poles swap places. So, you can also have two-stage Miller OTA with the dominant pole at the output but then you need to minimize your load capacitance. They use cascoding instead of cascading to get very high gain. In two-stage Miller, cascading was used. We took two stages with high gain and connect them in series to get very high gain. But here we can stack them on top of each other instead of stacking them in sideways and that is called cascoding [32].

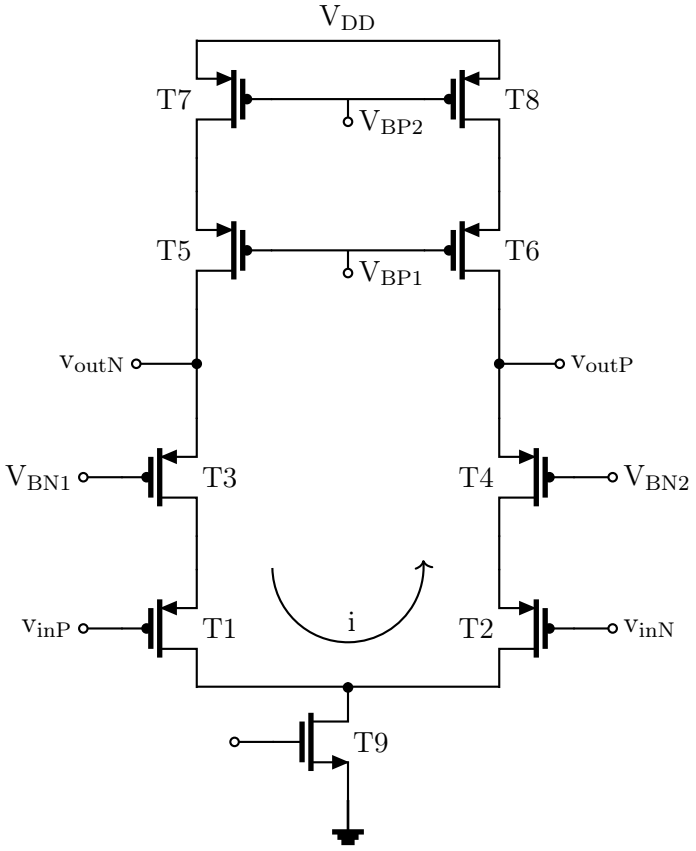


Figure 3.11: Telescope cascoding OTA.

So, we start with the standard differential stage discussed earlier in this chapter which has high gain. We call it high gain since it is proportional to $gm.rd_s$. Since we would like to have it proportional to the $(gm.rd_s)^2$, then we can cascode each of these transistors and that is called telescopic cascoding-OTA. First the NMOS transistors are cascoded and then the PMOS transistors are cascoded. So, 5 transistors are stacked on top of each other and the expression for the gain can be written just by inspecting the circuit by using half circuit analysis technique. A small current i may be viewed as the differential signal is apllied. We assume that the node connecting the input pair and the tail transistor is virtual ground. So, all the current that will goes through M1 will have the -ve counter part in M2. This current will give the output resistance at the output and will convert back to voltage. So, we have a small signal current that flows in the output resistor and generate voltage. Since it is small signal, so M1 will have $v_{in}/2$ and M2 will get -ve half $v_{in}/2$. It is not just happening but is the definition of the differential input [32]. This current is given

as,

$$i = gm\frac{v_{in}}{2} \tag{3.23}$$

We always assume that,

$$gm1 = gm2,\ gm3 = gm4,\ gm5 = gm6\ and\ gm7 = gm8$$

So, we always assume that the circuit is completely symmetric. Finally, this small signal current will flow out of v_{outp}. So, we have,

$$v_{out} = gm1.v_{in}.R_{out} \tag{3.24}$$

And the gain,

$$A_{v\phi} = \frac{v_{in}.gm1.R_{out}}{v_{in}} A_{v\phi} = gm1.R_{out}$$

The small signal current is typically the same for different architectures but R_{out} will vary for each architecture and it is the resistance looking in and out. We call it R_{out} up in parallel with R_{out} down because both get to AC ground in both directions. So, the output resistance of the cascode stages is given as,

$$\begin{aligned} R_{out} &= R_{out}up \parallel R_{out}down \\ &= gm6.r_{ds6}.r_{ds8} \parallel gm4.r_{ds4}.r_{ds2} \end{aligned}$$

We can see that we get output resistance proportional to $(r_{ds})^2$ and that will give us very high gain. For low supply voltages, it might be difficult to have all the transistors in the active region.

3.5.6 Folded cascode OTA

If we have low supply voltage then we can go for what we call a folded cascode OTA. As we know that the small signal equivalent circuits for the PMOS and NMOS transistors are identical. Therefore, one can fold the upper four transistors of the telescopic architecture and one should be able to implement the same very high resistance. Then we need to feed bias currents to all the branches. When you look down here, you are looking actually into the source over the transistor where you have very high impedance on the gate on the other side. So, this is like a common-gate. So, the resistance one can see down is the resistance of a common-gate amplifier and common-gate is an impedance transformer since we see higher and we have very high resistance connected to the drain of the common-gate. Now four transistors are stacked on top of each other, therefore, more swing is possible. Another benefit of this one is that the DC level at the output can be made equal to the DC level at the input. The expression for the gain is exactly the same as for the telescopic. One disadvantage as compared to the telescopic architecture is that folded architecture has four branches of current instead of two which means it will use more power than telescopic [32]. But folded architecture with its differential output, is extremely popular and is used widely in analog design.

The most challenging parameter to meet with this topology is linearity since there are still four transistors in stack. So, it is hard to get high linearity for high speed. But there is a possibility to add one more stage. Its single-ended version is also possible by tracing the power as been analyzed for the standard differential stage. But it might not be optimal since there will be a very high gate-source voltage.

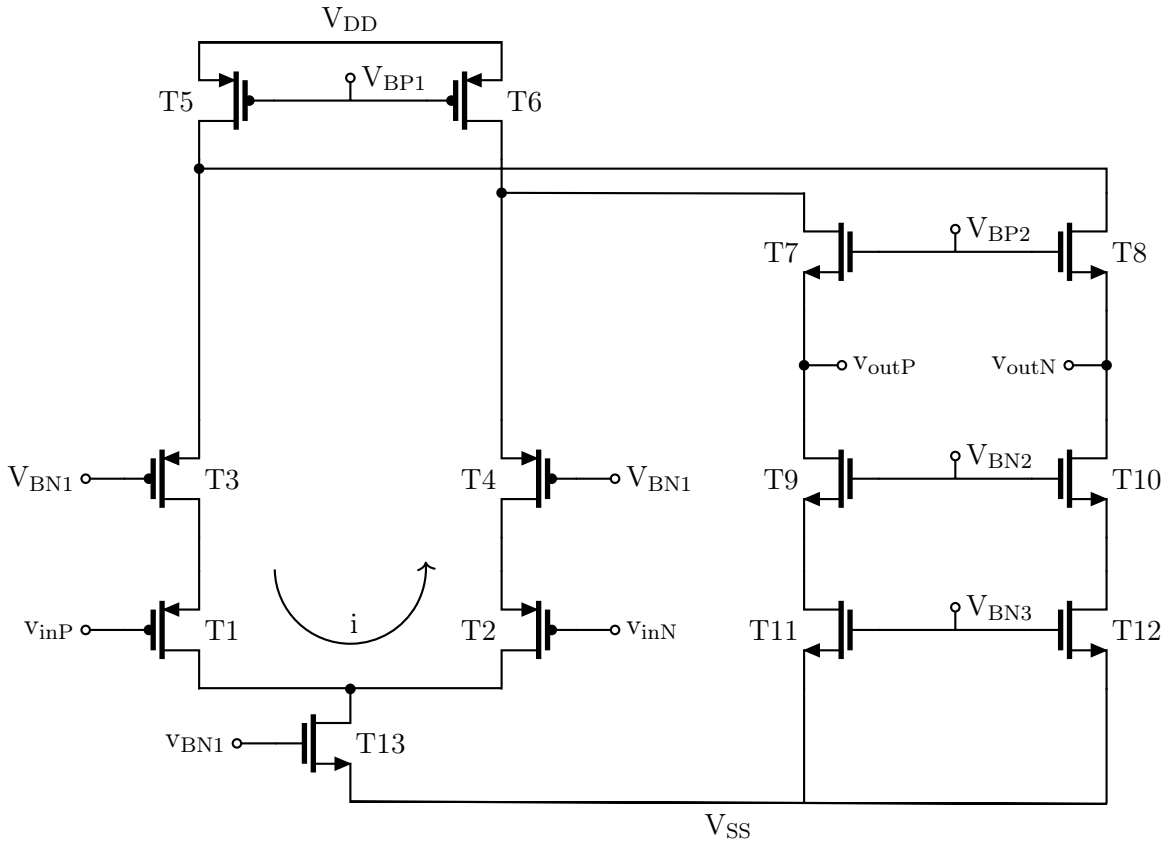


Figure 3.12: Fully Differential Folded Cascode OTA.

Telescopic has only one node with VH but folded has differential output, therefore, it has two nodes with VH. Due to the diode connection, the dominant pole is at the output and compensation is performed by adding some load capacitance. Lead compensation may also be applied by adding a resistor in series as was the case for two-stage Miller OTA. Adding capacitance will increase the stability. Generally, addition of extra capacitance is avoided because adding extra capacitance will slow down the amplifier. So, compensation is bad for speed and actually, a trade-off between speed and stability. Folding cascode is comparatively faster than the two-stage Miller. In case of two-stage Miller OTA, one has to add Miller capacitance in order to move one pole a lot because there were two poles at about the same frequency before compensation [32].

In order to find the expression for the unity gain frequency, we just need to find the node that will have the dominant pole. It can be found from the pole frequency expression. We also assume that we have first order response all the way to the unity gain frequency, so that we have to handle just one pole. Thus just one pole is active at the output and there is resistance and capacitance for this pole.

$$A(s) = \frac{gm1.R_{out}}{1 + \frac{s}{\omega p1}} \tag{3.25}$$

We also assume that load capacitance C_L completely dominates.

$$\omega p1 = \frac{1}{R_{out}.C_L} \tag{3.26}$$

So, expression for the unity gain frequency will be,

$$f_{ug} = \frac{gm1}{2\pi C_L} \tag{3.27}$$

We have the same expression as for the two-stage Miller OTA. The only difference is that we have C_L instead of C_c. We assume that all the other poles are above the unity gain frequency. They can easily be below as well but it does not matter here because if this is the case, the hand calculations become quite hard.

Typically, more transistors are used to implement the bias currents for this type of circuits that is why these are called advanced OTAs. So, biasing is much more difficult than for a two-stage Miller OTA. A better noise imunity is possible when implementing differential or fully-differential topologies. So, if there is equal noise at the two outputs, they will cancel each other when taking the difference of the two. Noise on the ground will be cancelled because it is a common-mode noise and same with V_{DD}. It is never fully cancellation in real circuits but always a partial cancellation. Different transistors are generating different noise but the M9 is the only transistor that will generate noise on both outputs. So, there is no reason for the noise power to be different. So, the noise power will also be twiced. The SNR will increase with the twice noise power and four time the signal power.

So, folded cascode topology with proper biasing, offers more freedom to set the input DC levels in order to get higher output swing. But circuit complexity is increased with this topology because of the common mode feedback (CMFB) implementation [22]. The high frequency behavior can be characterized by frequency response of the fully differential folded cascode OTA. A first-order MOS model can be used for design purpose [12]. At the cost of area and power, both speed and gain requirements are possible to achieve by implementing gain enhancement to a single-stage cascode topology. Also, area and power consumption may be reduced by replacing four single-ended amplifiers with two fully differential OTAs [20].

3.5.7 Current Mirror OTA

The current mirror OTA is also a load compensated OTA. The higher the load capacitance, the more stable the amplifier. The transistor gain is very low in deep-sub-micrometer processes because of the weak output impedance. Very high gain is possible with cascoding. Highest impedance level is only at the output here. Cascoding helps in boosting the gain and raising the circuit impedance but the output swing is then reduced. Thus a multi-stage current mirror can be a possible alternative [28].

As the factor K is increased, the pole at the output node will move down in frequency. So, it is not possible to use too high value of K. In order to find the gain expression, the expression for v_{out} can be found as,

$$v_{out} = v_{outp} - v_{outn}$$
$$v_{outp} = K.gm1.\frac{v_{in}}{2}.Z_1$$

Similarly,

$$v_{outn} = -K.gm1.\frac{v_{in}}{2}.Z_1$$

Also,

$$\Rightarrow v_{out} = K.gm1.v_{in}.Z_1$$

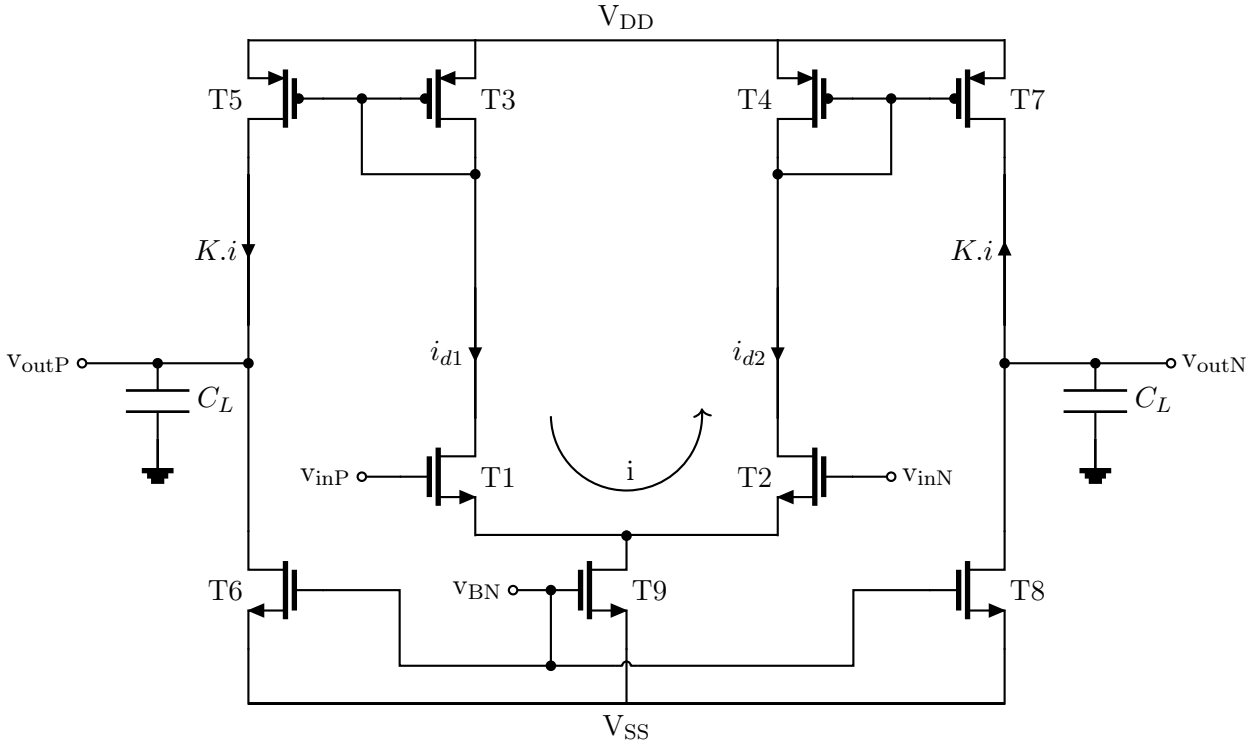


Figure 3.13: Fully differential Current Mirror OTA.

$$\Rightarrow \frac{v_{out}}{v_{in}} = K.gm1.Z_1 \tag{3.28}$$

Where Z_1 is not the the output resistance but a combination of C_L and the output resistance. In order to find a nice first order expression for the frequency response, we assume that C_L completely dominates over all the other capacitances. Z_1 can be found as,

$$Z_1 = \frac{R_{out}\frac{1}{sC_L}}{R_{out} + \frac{1}{sC_L}} \tag{3.29}$$

Bu subsituting 3.28 in 3.27, we get expression for the gain.

$$\frac{v_{out}}{v_{in}} = \frac{K.gm1.R_{out}}{1 + \frac{s}{\omega p1}} \tag{3.30}$$

Where,

$$\omega p1 = \frac{1}{R_{out}.C_L} \tag{3.31}$$

So, equation 3.29 shows that the DC gain i $K.gm1.R_{out}$ which means that DC gain can be increased by increasing the value of K. But when K is increased, the R_{out} goes down which is a parallel connection at the output. Thus, the output resistance is inversely proportional to K.

$$R_{out} \propto \frac{1}{K} \tag{3.32}$$

Thus the DC gain is almost independent of K. The expression for the unity gain frequency can be written as,

$$f_{ug} = \frac{K.gm1}{2\pi C_L} \tag{3.33}$$

So, here it is possible to crank up the unity gain frequency without changing the gain. With this topology, the gain and unity gain frequency are decoupled [32]. So, this architecture is also very popular and may be used for high input impedance amplifiers. Input impedance may be different since it depends on the electrode type. Compatibility is very important in order to reduce the attenuation which means the high input impedance is desired. Chopping reduces the flicker noise but low input impedance remains the problem. The input impedance of a capacitively-coupled neural amplifier is boosted by implementing the effect of a negative capacitor and is reported in [26]. Current-reuse amplifier topology is also popular for their low input-referred thermal noise but gain is limited due to low output swing. A two stage rail to rail amplifier structure is reported in [18]. The design boosted the gain than a usual current-reuse amplifier.

3.5.8 Gain boosting

Cascoding is one way to boost the gain but positive feedback can also be used which is much more controversial. For a feedback system, the expression for the closed loop gain is written as,

$$A_{CL} = \frac{A}{1 + A\beta} \tag{3.34}$$

Higher loop gain is possible if the denominator becomes less than 1. One technique based on positive feedback can be applied at M3 and M4. We see $1/gm4$ looking into the transistor when it is diode connected. The effective transconductance of this node can be reduced by subtracting another transconductance. So, $gm4 - gm$ will produce higher resistance. The negative transconductance can be produced by picking up the the same voltage for the gate of M4 as well as for the drain of M4. But we can also connect a PMOS in parallel to M4. The same can be done on the other side and is called cross-coupled. So, with the positive feedback, , one can get larger impedance or resistance and the gain will also be increased. But one must make sure that this difference should not be negative or zero. This technique is also very popular and is used a lot. This difference will change over PVT variations and mismatch of M3 and M4 will also change this difference. Another drawback of course, is that the impedance level goes up which means that the pole at this node will go down in frequency. On the other hand, the PM will also be increased. For this, a differential circuit is needed which is completely symmetric [32].

3.5.9 Feedback Amplifiers

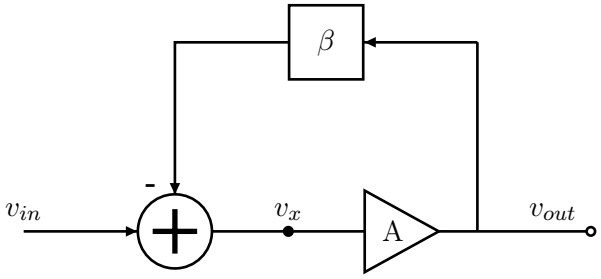


Figure 3.14: Feedback circuit.

The analog designer's job is to design amplifiers with accurate gain means very high open loop gain with negative feedback. This has been a traditional way where very accurate gain can be achieved after the feedback is applied.

This is called negative feedback since something is subtracted from the input. The transfer function from the input to the output is of interest and is called closed loop gain. So, expression for the closed loop gain as also mentioned in the previous sub-section, can be written as,

$$A_{CL} \equiv \frac{v_{out}}{v_{in}}$$

$$A_{CL} = \frac{A}{1 + A\beta} \tag{3.35}$$

This expression is often modified in the following form,

$$A_{CL} = \frac{A}{1 + A\beta} = \frac{1}{\beta}.\frac{A\beta}{1 + A\beta} \tag{3.36}$$

The reason for that is we can call $A\beta$ the loop gain. The loop gain is the most important transfer function when one analyze feedback circuits. From the equation 3.35, we can see that if loop gain approaches to infinity, the closed loop gain is then $1/\infty$. So, in this case, all the sensitivity of A is removed and A is no longer a part of the transfer function and that was the idea here [32]. Typically, β is fixed and A is increased until the closed loop sensitivity on A is less than a certain value. For example, if $\beta = 1$, then the entire output signal is feedbacked to the input signal and closed loop gain is 1. If $\beta = 0.5$ and A is extremely high then the closed loop gain will be 2.

3.5.10 Inverters

Invertors are more efficient than common-source amplifiers. So, the total transconductance is the sum of the individual transconductances of the PMOS and NMOS transistors. Moreover, with almost the same bias current, inverter-based amplifier provides 6dB more gain than a traditional common-source amplifier [30].

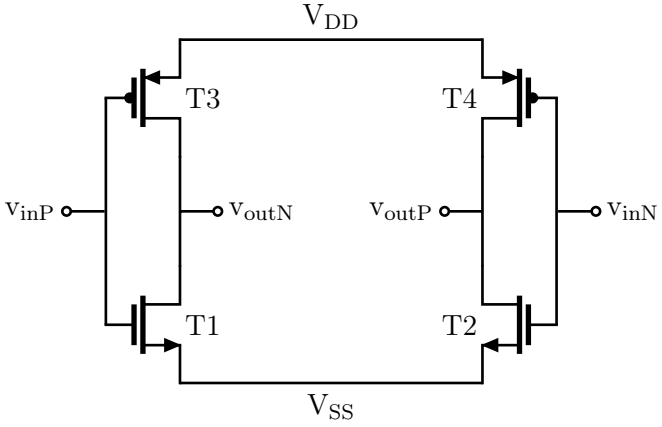


Figure 3.15: Inverter-based differential amplifier.

$$G_m = g_{mn} + g_{mp} \tag{3.37}$$

The gain is directly proportional to the gm of the amplifying transistor (NMOS) for a conventional common-source amplifier with a load resistor. But the gain of amplifier based on CMOS invertor is relatively large since it is directly proportional to the total

transconductance of the NMOS and PMOS transistors [27]. Output resistance is the same as for the common-source but you get twice the gain with the same current and same power consumption. If the load capacitance is not increased, the speed will also be twiced. With suitable transistor sizing, the total transconductance can be increased.

Part III

Method

Chapter 4

Design Methodology

This part of the thesis will propose the design methodology and architectures for two different OTAs. OTA stands for Operational Transconductance Amplifier where operational means high gain and transconductance means that the output resistance is high [32]. The design procedure will focus on on-chip design where we typically drive capacitive loads since extra power is needed to implement low output resistance. The procedure will explain the integrated amplifier design which is different from discrete amplifier design. Different methodology and tools are used to design integrated amplifiers. Integrated in this work means only the CMOS technology will be considered.

All integrated circuits contain several amplifiers which exhibit small-signal and large-signal properties. Therefore, it is very important to analyse the circuit behavior during different design phases. Small-signal properties are easier to analyze since then the linear circuit is of interest. Small-signal properties will include voltage gain, input impedance, output impedance and noise while large-signal properties will include large-signal gain, power, linearity etc. When working with large signals then circuit can not be linearized and things become messy. Non-linear equations are difficult to handle.

In this section, after selecting two topologies, a FOM will be defined. The dsigner's job is to translate given circuit specifications into transistor sizes and bias currents in order to optimize the defined FOM. Sizing of transistors means chip area which is straight forward the cost. So, it is important to have good understanding of transistor operation, semiconductor physics, equivalent circuit theory, transistor characteristic curves, operational modes and ability to quickly analyse and inspect the circuit diagram in order to write expressions for hand calculations.

4.1 Selected Design Method

The chip size has been scaled down to a few nano meters. The traditional square-law model for design does not include effects of scaling and it can not model weak and moderate inversions. Therefore, the square-law model can not be implemented [32]. We need to rely on simulator and implement g_m/I_d besed design methodology. The g_m/I_d acts as universal characteristic for same type (NMOS or PMOS) of transistors used in the same process. The relationship between the drain current and the normalized drain current is assumed as fundamental design tool in this methodology. A marginal reduction in the power consumption and low parasitic capacitances with the same bandwidth is achievable

by increased g_m/I_d values [6].

The g_m/I_d is a normalized quantity which means that g_m/I_d does not change that much by changing the physical dimensions (width and length) of the transistor. Of course, one would always try to be in the active region of operation since weak inversion operation leads to triode region where the transconductance is small because of the small bias currents. For example, if v_{gs} is 0.7, the transistor is still operating in the active region but as soon as gate voltage is reduced to a certain point, the region of operation is diverted from active to triode. So, g_m/I_d is the first FOM and is called the current efficiency that tells how much bias current is needed to get a certain transconductance. Highest g_m/I_d value for a MOSFET is about 30 [32]. BJTs have higher g_m/I_d. So, MOSFETs are quite bad in efficiency. So, the limit 30 is set by physics since it is set by fundamental properties. We can not get better for MOSFETs. Of course, this value will also depend on the temperature [32]. Today's research is beyond CMOS devices and one thing the researchers are working on is to get better g_m/I_d. So, there are several promising techniques to get better transistors. The second figure of merit is the intrinsic gain that tells how much gain can we have out of a single transistor. This is typically, a transistor property and is largest in weak inversion. The problem in weak inversion is that the speed is slow. The third FOM is the f_T (cut off frequency). By having these definitions and facts in mind, the g_m/I_d based methodology can be outlined in the following simple way.

- First of all, an estimate of g_m is required which may easily be calculated from the design specifications.
- Secondly, a suitable value for g_m/I_d is selected which is generally based on the gain and speed specifications.
- Then intrinsic gain is estimated which is based on the gain and output resistance.
- Finally, transistor sizing is done. Several steps are involved here. First of all, several plots for intrinsic gain are generated by sweeping gate lengths of the transistors. Then a suitable gate length is chosen which can provide enough intrinsic gain for a selected value of g_m/I_d. In the end, the width is changed until a certain level of transconductance is achieved which should almost be the same as calculated in the first step.

4.2 Selected Architectures

Different architectures with their advantages and disadvantages, are reviewed in the amplifier theory chapter. Two promising architectures, a fully differential folded cascode OTA and an inverter-based current mirror OTA, are selected for design. Telescopic cascode OTA is better tnan the folded cascode OTA in terms of power. Folded cascode OTA will use more power than the telescopic cascode for the same gain because there are more branches in the folded cascode topology. Folded cascode is better in terms of output swing because it has same DC levels at the input and output. In addition, a CMFB circuit is also needed to design a fully differential topology for the folded cascode. The performance of the two topologies may be compared even with the same g_m/I_d. Threshold voltage has many definitions. It varies and depends on v_{ds} which means it is man made [32]. Also the expression for the effective voltage is written as,

$$v_{eff} = v_{gs} - v_{th} \tag{4.1}$$

The expression shows that v_{eff} is also man made since v_{th} is man made. But g_m/I_d is not man made because it is something which can be measured at the terminals of the transistor [32]. Both topologies are implemented separately in the following sections.

4.3 Selected FOM

In this thesis work, a figure of merit in terms of energy efficiency is defined as,

$$FOM = \left[\frac{P.V_{n,in}^2}{10^{\frac{A_\phi}{20}}.f_{ug}.10^{\frac{HD2}{10}}} \right] \tag{4.2}$$

Input referred noise in included in the numerator of the expression since it should be a small quantity. Harmonic distortion is included in the denominator of the expression since higher value of HD2 is desired. To have a good figure-of-merit, the large values are desired in the bottom part of the expression and small in the top part. The unit of this FOM is energy efficiency or amount of power per performance. Energy efficiency means that the smaller the FOM, the better the energy efficiency [32]. A similar FOM may also be used for ADCs but then one have to use sampling frequency instead of unity gain frequency in the expression.

4.4 Technology characteristics

As mentioned earlier that a commercially available 130nm CMOS technology is utilized in implementing the both architectures. Two different types, standard (STD) and high gain (HG), of each NMOS and PMOS transistors from the chosen CMOS technology are compared first. The simulation results for some important transistor parameters for standard and high gain NMOS transistors is done in cadence and the values are listed in the table 4.4.1.

Parameters	**NMOS standard**	**NMOS high gain**
f_T	$7.8\,G(@1.8\,mA)$	$7\,G$
g_m	$\approx 235\,\mu$	$\approx 157\,\mu$
g_m/I_d	$\approx 2.8\,(max. \approx 32)$	$\approx 2.5\,(max. \approx 32)$
g_m/g_{ds}	$\approx 72\,(max. \approx 168)$	$\approx 71\,(max. \approx 470)$
I_d	$81.54\,\mu A$	$62\,\mu A$
r_{out}	$\approx 306\,K$	$\approx 453\,K$
v_{sat}	$\approx 452\,mV$	$\approx 711\,mV$
v_{eff}	$\approx 566\,mV$	$\approx 675\,mV$
v_{th}	$\approx 434\,mV$	$\approx 324\,mV$
v_{gs}	$1\,mV$	$1\,mV$
v_{ds}	$1.6\,mV$	$1.6\,mV$

Table 4.4.1: STD and HG NMOS transistor parameters.

Similarly, a simulation for standard and high gain PMOS transistors is done and the resulted values for important parameters are listed in the table 4.4.2.

Parameters	PMOS standard	PMOS high gain
f_T	$2.3\,G$	$3\,G$
g_m	$\approx 64\,\mu$	$\approx 78\,\mu$
g_m/I_d	≈ 3	≈ 3
g_m/g_{ds}	≈ 36	≈ 91
I_d	$-21.2585\,\mu A$	$-30.0018\,\mu A$
r_{out}	$\approx 563\,K$	$\approx 1\,M$
v_{sat}	$\approx -479\,mV$	$\approx -638\,mV$
v_{eff}	$\approx -586\,mV$	$\approx -657\,mV$
v_{th}	$\approx -415\,mV$	$\approx -344\,mV$
v_{gs}	$1\,mV$	$1\,mV$
v_{ds}	$1.6\,mV$	$1.6\,mV$

Table 4.4.2: STD and HG PMOS transistor parameters.

While comparison, the main focus has been on the three FOMs, g_m/I_d, unity gain frequency and intrinsic gain. It can be seen from the 4.4.1 and 4.4.2 that the standard NMOS has greater unity gain frequency than the high gain NMOS but standard PMOS has less unity gain frequency than the high gain PMOS. g_m/I_d is same for both pmos types but NMOS standard has little greater value than NMOS high gain. Intrinsic gain is almost the same for both NMOS types but PMOS high gain has much greater value than PMOS standard. Finally, when comparing NMOS with PMOS, NMOS has greater unity gain frequency and much more transconductance than the PMOS transistor. Also, both standard NMOS and PMOS have higher threshold voltages as compared to their high gain versions but threshold voltage will also depend on the drain-source voltage. Perhaps, the best way to compare the performance is to compare with the same gm/I_d. In the end, due to the above mentioned reasons, the standard type of both transistors is adopted during the design phase.

4.5 Fully differential folded cascode OTA

As discussed in section 3.5.6, a fully differential folded cascode OTA is quite suitable for low supply voltage design. So, a standard fully differential folded cascode topology with some modifications, is adopted and the block diagram of the topology is shown in the figure 4.1.

As it is named, fully differential means both input and output signals are differential. As mentioned in section 3.5.1, the main advantage of using fully differential signals is to reject the noise from the power lines and the substrate. This idea works because the noise is distributed on the both signal paths and rejected afterwords since the difference between the signals is considered then [23]. Of course, there is no absolute rejection of noise because the noise is a nonlinear phenomenon that will have a random change with any fluctuation in the voltage level. On the other hand, one main disadvantage of differential topology is that a common-mode feedback circuit is required in order to establish the common-mode voltage. This additional circuit results in more complex circuitry and more power consumption.

The input stage is a standard differential stage with two NMOS input transistors T1 and T2. The reason why the NMOS transistors are chosen as input pair is that the mobility of NMOS transistors is higher than PMOS transistors [34]. During the design process, it is

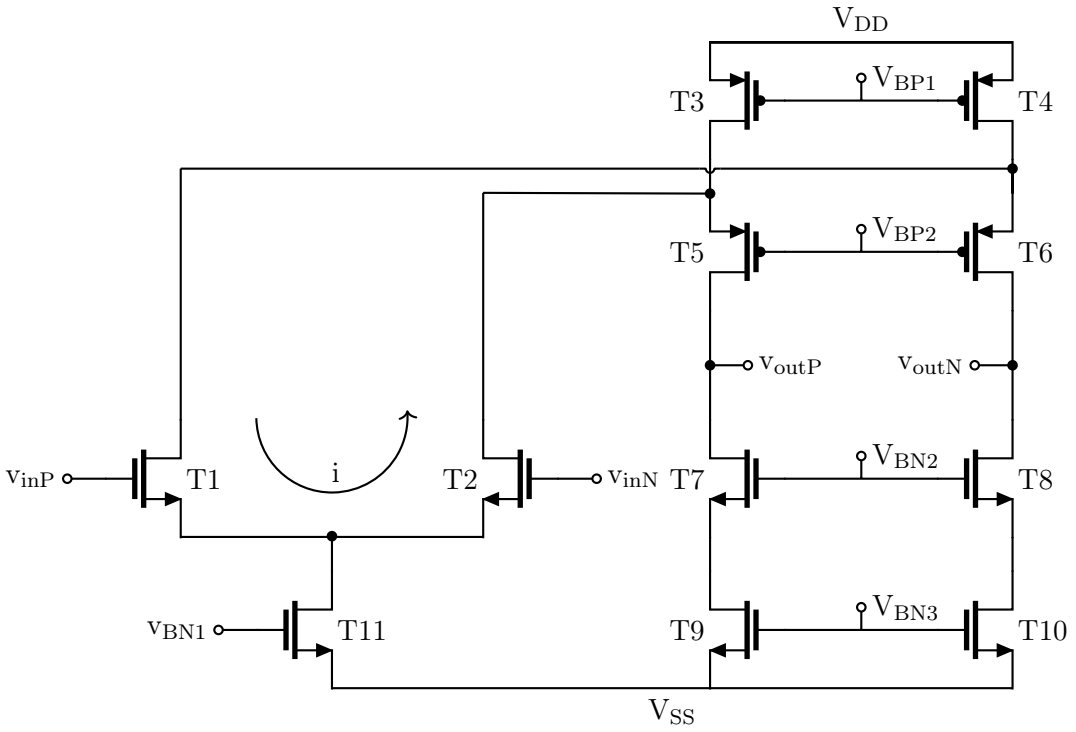


Figure 4.1: Fully Differential Folded Cascode OTA.

possible to achieve four times more transconductance with NMOS transistors as compared to PMOS transistors of the same dimensions which helps in reducing the number of PMOS transistors. It is easier to reduce the size when the transconductance of the transistor is very high [31]. Also, by considering input common-mode range limitations, NMOS input pair produces a larger gain than a PMOS input pair [13]. The two transistors T3 and and T4 on the top are making a current mirror source load. T5, T6, T7 and T8 are the cascode transistors while T9 and T10 are making the cascode current mirror or also called the load or driver transistors. The gate voltage on the driver transistors is determined by the output of the CMFB circuit. T11 is called the tail transistor and with its bias voltage , it is used to provide the bias currents to the input pair. No inductors are used for this design since inductors become too large on-chip because of the low frequency [32].

$$R_{out} = Resistance\,looking\,up \parallel Resistance\,looking\,down$$
$$R_{out} = R_{out}up \parallel R_{out}down$$
$$= gm6.r_{ds6}.r_{ds4} \parallel gm8.r_{ds8}.r_{ds10}$$

If we assume all the gm_s and rds_s are the same then the simplified expression for the amplifier's output resistance can be written as,

$$R_{out} = gm.r_{ds}.r_{ds} \parallel gm.r_{ds}.r_{ds}$$
$$= gm.(r_{ds})^2 \parallel gm.(r_{ds})^2$$

$$R_{out} = gm.\frac{(r_{ds})^2}{2} \tag{4.3}$$

$$\begin{aligned} A_v &= gm.R_{out} \\ &= gm.gm.\frac{(r_{ds})^2}{2} \end{aligned}$$

$$A_v = gm^2.\frac{(r_{ds})^2}{2} \tag{4.4}$$

4.5.1 Hand calculations

In analog design procedure, some hand calculations are done first in order to have some rough estimate of design. In all specifications, one must have a line which is set by the requirement of transconductance of the input transistors. The idea here is to find the required transconductance for the input transistors and also have an estimates of the output resistance which can provide enough gain. In order to do hand calculations, important equations for the selected topology are required.

The transconductance for the input pair, can be calculated from the unity gain frequency expression.

$$g_m > 2\pi f_T C_L \tag{4.5}$$

Load capacitance is assumed to be 50 fF where as unity gain frequency is 200 MHz according to the specifications. By substituting these values in the equation 4.2, we get,

$$g_m > 62.83\,\mu S$$

By including some design margin, it is assumed that $g_m = 100\,\mu S$.

From the specifications, the power consumption should not exceed $30\mu W$. Then the requirement for the total current can be found as,

$$\begin{aligned} P &= V.I \\ \Rightarrow I &= \frac{P}{V} \\ \Rightarrow I &= \frac{30\,\mu W}{1.5\,V} \\ I &= 20\,\mu A \end{aligned}$$

The target R_{out} is calculated as,

$$\begin{aligned} R_{out} &= \frac{A_v}{gm} \\ &= \frac{500}{100\,\mu S} \\ R_{out} &= 5\,M\Omega \end{aligned}$$

4.5.2 Transistors sizing

In analog design phase, width, length and gate-source voltage are the three design variables or nodes which need to be determined. g_m/I_d provides the gate-source voltage, gm provides the width and g_m/g_{ds} sets the length requirements. There is freedom to chose these nodes but there is a fourth node, threshold voltage, which is assumed to be fixed for this work in order to obtain certain properties. The FOM is important to be considered while sizing the dimensions of the transistors. As the mission is to minimize the FOM, the more focus here will be to minimize the area and maximize the energy efficiency. That is why the transistors are used to generate the required biasing currents. The capacitors and resistors are avoided on the chip and instead the transistors are used as capacitors and resistors on the chip. Only a few transistors are doing the job on the chip and rest are used for biasing purposes. Therefore, transconductance, intrinsic gain and unity gain frequency will play a vital role in sizing the transistors. The sizes are optimized by utilizing the g_m/I_d and drain-source voltages since the load capacitance is assumed to be fixed. The values assumed for all the three important figure of merits are listed in the table 4.5.1.

Parameters	**Input transistors**	**Cascode transistors**	**Current mirrors**
g_m/I_d	16	12	8
g_m	$100\,\mu S$	$75\,\mu S$	$50\,\mu S$
g_m/g_{ds}	55	40	30

Table 4.5.1: Assumed values for g_m/I_d, gm and g_m/g_{ds}.

The g_m/I_d for the input transistor should be high and is selected as 16 because with high g_m/I_d, one can use small current to get a certain transconductance. For current mirros, a low g_m/I_d is desired since the current mirrors should have high overdrives because of noise and mismatch and high overdrive means low g_m/I_d. But if g_m/I_d becomes too low then a long transistors should be required in order to get high enough output resistance in the output branch [32]. The simulated values for all the transistors are listed in the table 4.5.2.

Transistors	**W**	**L**	g_m/g_{ds}	g_m	I_d	g_m/I_d
Input NMOS	$1.35\,\mu m$	$210\,nm$	54.9	$102.6\,\mu S$	$6.5\,\mu A$	15.9
Cascode PMOS	$1.8\,\mu m$	$230\,nm$	40	$75.4\,\mu S$	$-6.7\,\mu A$	11.2
Cascode NMOS	$470\,nm$	$200\,nm$	40.3	$76.4\,\mu S$	$6.8\,\mu A$	11.2
Current mirror NMOS	$200\,nm$	$190\,nm$	30	$52.9\,\mu S$	$6.4\,\mu A$	8.2
Current mirror PMOS	$800\,nm$	$220\,nm$	31	$50.7uS$	$-6.7\,\mu A$	7.5

Table 4.5.2: Simulated dimensions for all transistors.

The final optimized sizes are listed in the table 4.5.3.

Transistors	**W**	**L**
Input NMOS	$2.8\,\mu m$	$350\,nm$
Cascode PMOS	$1.6\,\mu m$	$250\,nm$
Cascode NMOS	$800\,nm$	$900\,nm$
Current mirror NMOS	$1.2\,\mu m$	$250\,nm$
Current mirror PMOS	$3.5\,\mu m$	$580\,nm$

Table 4.5.3: Optimized dimensions.

4.6 Biasing circuit for the folded cascode OTA

An analog integrated circuit is composed of several subcircuits working together. Each subcircuit also consists of many transistors. As mentioned earlier in section 4.5.2, only few of the transistors are doing the actual job, the rest of the transistors are used for biasing. The purpose of a bias circuit is to generate a dc voltage to keep a transistor near to its desired operating point. Naturally, any change in the temperature or transistor parameter will also change the bias requirement [4]. In analog CMOS, only transistors are preferred for biasing. The quality of dc voltage and current sources is a key factor when determining the overall performance of analog circuits. Therefore, the proper biasing of amplifier is very important and is quite a challenging task for analog designers. The sharing and distribution of dc voltages is very difficult task since a bias circuit provides voltages to several subcircuits. Biasing can be done in several ways. As a rule of thumb, PMOS transistors are biased with PMOS transistors and NMOS transistors are biased with NMOS transistors in order to avoid detracking of the process corners [32]. A single current source is used in the testbench and rest of the desired bias currents are generated by using current mirrors. The bias circuit designed for the fully differential folded cascode topology is shown in the figure 4.2.

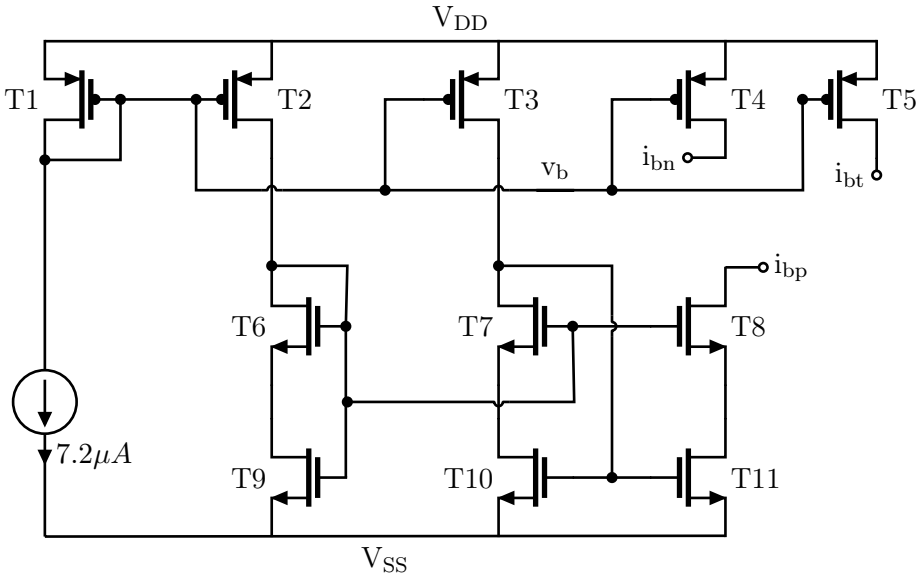


Figure 4.2: Biasing circuit for the folded cascode OTA topology.

The optimized sizes of the transistors are given in the table 4.6.1.

Transistors	**W**	**L**
T1 - T5	$4.1\,\mu m$	$220\,nm$
T6 - T8	$1.2\,\mu m$	$300\,nm$
T9 - T11	$550\,nm$	$130\,nm$

Table 4.6.1: Optimized dimensions.

4.7 CMFB circuit for the folded cascode OTA

The CMFB should be applied when designing differential topologies. The purpose of the CMFB circuit here is to detect the common-mode outputs of the fully differential folded cascode OTA and then compare it to the reference voltage and send the difference through negative feedback to the bias network of the amplifier in order to set the common-mode output level [24].

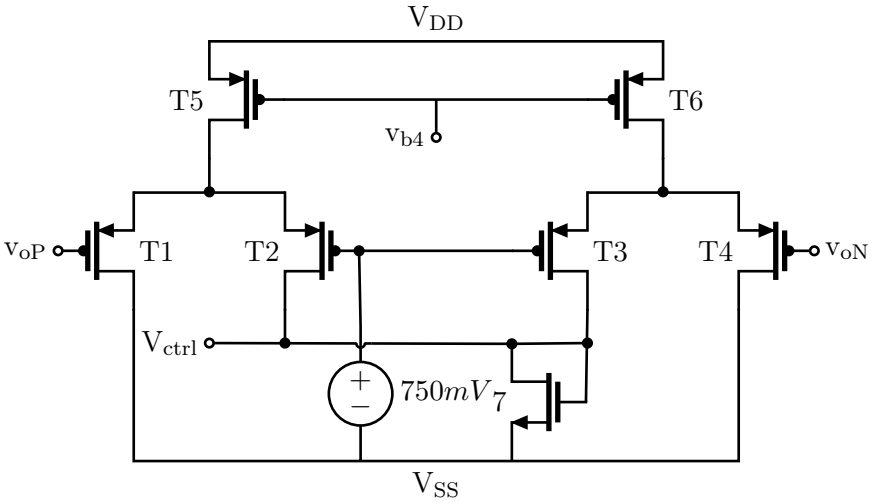


Figure 4.3: CMFB circuit for the folded cascode OTA.

The two PMOS transistors T5 and T6 are identical by design and four NMOS transistors T1-T4 are identical by design. A target common-mode voltage of 750 mV is applied to the gate of T2 and T3 since a 1.5 V power supply is used. The CMFB loop will adjust the common-mode voltage at the output to be close to the target value. CMFB circuit is also biased with v_{b4} because these transistors are in parallel with the v_{b4} transistors in the bias circuit. The optimized dimensions for the transistors are given in the table 4.7.1.

Transistors	**W**	**L**
T1 - T4	$800\,nm$	$700\,nm$
T5/T6	$1.2\,\mu m$	$190\,nm$
T7	$800\,nm$	$800\,nm$

Table 4.7.1: Optimized dimensions.

4.8 Fully differential inverter-based current mirror OTA

Generally, transistors are mostly operating in the saturation region in traditional operational amplifiers and one must provide at least one DC bias current to carry out the operation. Sometimes, the folded topologies are not possible to implement with low power supplies because the threshold voltage of the transistors does not reduces as fast as the size or power supply reduces [30]. This problem can be handled by implementing an energy efficient circuit topology with current starving approach, known as inverter-based differential amplifier. Secondly, the leakage can be reduced rather by having stable higher threshold voltages which is possible by proper controlling of bulk to source voltages. A CMOS inverter may be thought as a transconductor, just like a single MOS transistor, which may

also be used as a gain stage [29]. Two CMOS inverters are used at the both inputs. This approach results in a combined transconductance of PMOS and NMOS transistors with an increament of 6dB gain, as mentioned in section 3.5.10. The expression for the output resistance can be found by considering the small-signal model as shown in the figure 4.5.

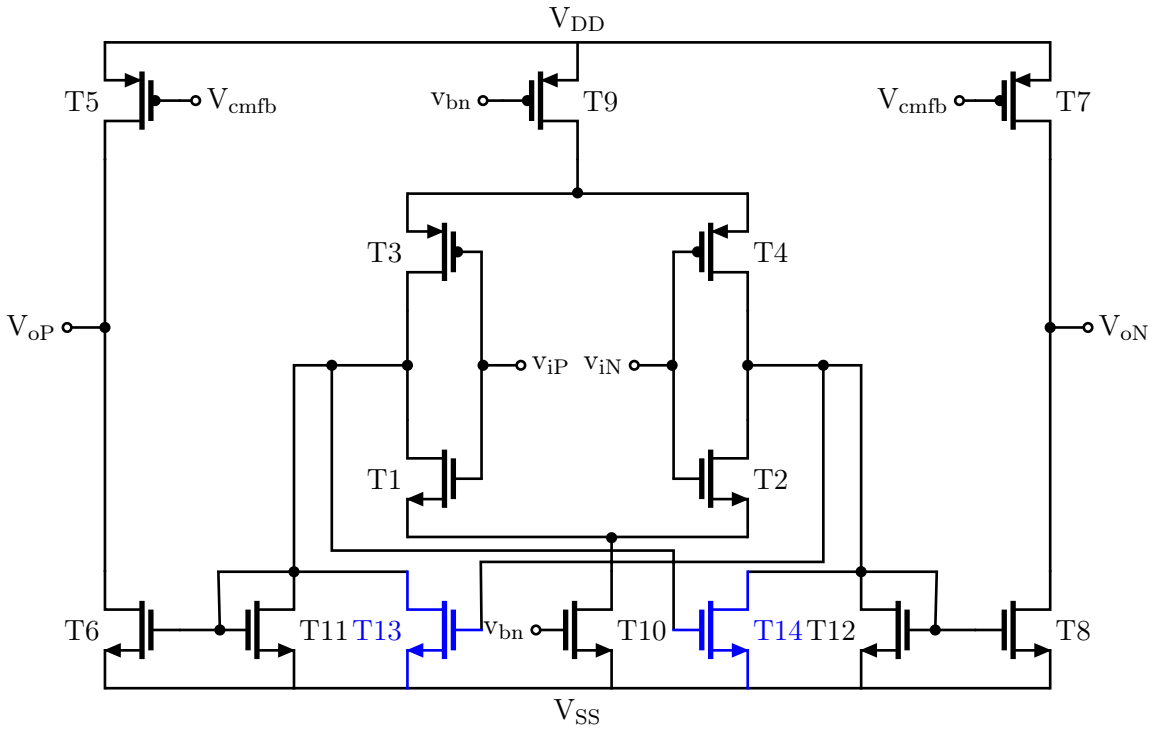


Figure 4.4: Fully differential inverter-based current mirror OTA.

A cross-coupled structure is adopted in the topology as shown in the figure 4.4. This modification helps in boosting the DC gain as well as achieving higher closed-loop gain as also reported in [25]. As mentioned earlier in section 3.5.7, the DC gain is almost independent of the value of K because as K goes up, R_{out} goes down since R_{out} is inversely proportional to K.

It is assumed that all the gm and r_{ds} are equal. To find a target R_{out}, a small signal is applied and the current is measured. Also, we set,

$$gm_n.v_{in} = 0$$
$$gm_p.v_{in} = 0$$
$$\frac{v_x}{i_x} = r_{ds} \parallel r_{ds}$$

So, the output resistance is just the parallel combination of the two resistances of T5 and T6 of the output branch.

$$\Rightarrow R_{out} = \frac{rds}{2} \tag{4.6}$$

The combined transconductance of the two input transistors T1 and T3 can be written as,

$$G_m = gm_{T1} + gm_{T2} \tag{4.7}$$

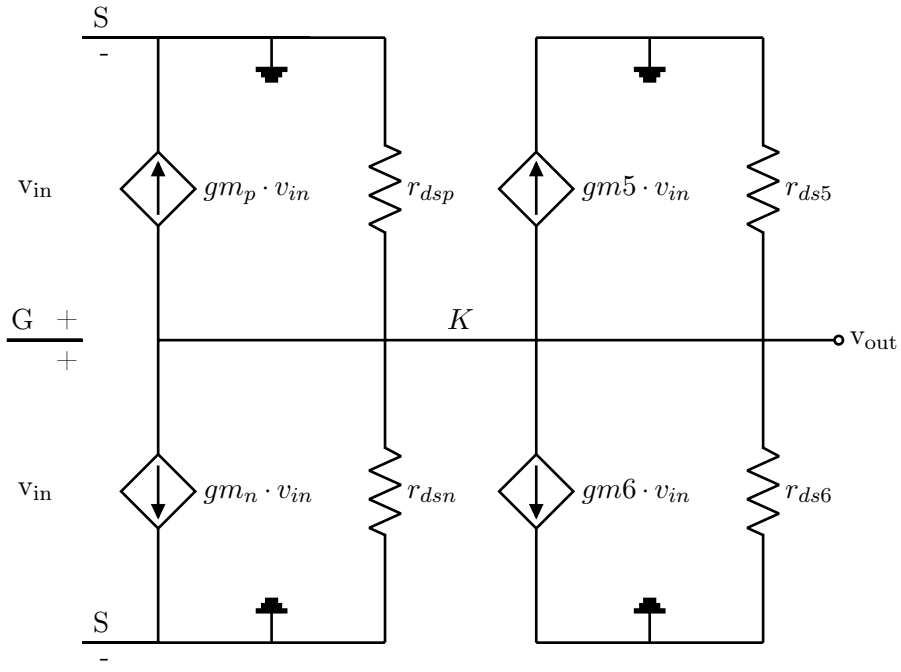


Figure 4.5: Small-signal model.

The total voltage gain can be found by calculating the gain at each node of the circuit. We have three nodes in the circuit, one is the input node, the second is the node for the diode connected NMOS and the third one is the node at the output. Another way to find the expression for the gain is to calculate the current that runs in the output resistance, multiply it by the output resistance, and divide it by the input voltage. This method is selected for simplicity,

$$v_{out} = v_o^+ - v_o^-$$
$$v_o^+ = K.G_m.\frac{v_{in}}{2}.R_{out}$$

Similarly,

$$v_o^- = -K.G_m.\frac{v_{in}}{2}.R_{out}$$

The total output voltage can be calculated by adding both v_o^+ and v_o^-,

$$v_{out} = [K.G_m.\frac{v_{in}}{2}.R_{out}] - [-K.G_m.\frac{v_{in}}{2}.R_{out}]$$
$$v_{out} = K.G_m.v_{in}.R_{out}$$
$$\Rightarrow \frac{v_{out}}{v_{in}} = K.G_m.R_{out}$$

By substituting the expression for R_{out}, we get,

$$A_v = K.G_m.\frac{rds}{2} \tag{4.8}$$

We are just interested in the signal current and the output resistance. We dont rely on instrinsic gain for the inverter here because we connect a low impedance at the output of the inverter. Since we are connecting the output of the inverter to a diode connected transistor T11, $1/gm$ of T11 will be in parallel with resistance of the inverter. The inverter resistance is much larger than then $1/gm$, so it disappears from the expression [32]. Here the gm of the input transistors and R_{out} are the two most important quantities.

4.9 Transistors sizing

The transistors dimensions are found in the same way as for the fully differential folded cascode topology by using the gm/I_d methodology. The assumptions made for the parameters are listed in the table 4.9.1 and the simulated values for dimensions are given in the table 4.9.2.

Parameters	**Input transistors**	**Cascode transistors**	**Current mirrors**
g_m/I_d	16	12	9
g_m	$80\,\mu S$	$60\,\mu S$	$45\,\mu S$
g_m/g_{ds}	40	30	23

Table 4.9.1: Assumed values for g_m/I_d, gm and g_m/g_{ds}.

Transistors	**W**	**L**	g_m/g_{ds}	g_m	I_d	g_m/I_d
Input NMOS	$1.1\,\mu m$	$200\,nm$	50.9	$83.15\,\mu S$	$5.2\,\mu A$	16
Cascode PMOS	$1.3\,\mu m$	$140\,nm$	27.4	$64.35\,\mu S$	$-5.2\,\mu A$	12.47
Cascode NMOS	$350\,nm$	$150\,nm$	23	$61.4\,\mu S$	$5.3\,\mu A$	11.6
Current mirror NMOS	$180\,nm$	$150\,nm$	23	$50.8\,\mu S$	$5.3\,\mu A$	9.5
Current mirror PMOS	$600\,nm$	$150\,nm$	22.6	$45.7uS$	$-5.2\,\mu A$	8.8

Table 4.9.2: Simulated dimensions for all the transistors.

The final optimized dimentions for all the transistors are given in the table 4.9.3.

Transistors	**W**	**L**
T1/T2	$1.6\,\mu m$	$800\,nm$
T3/T4	$4.5\,\mu m$	$400\,nm$
T5/T7	$700\,nm$	$500\,nm$
T6/T11/T12/T8	$350\,nm$	$500\,nm$
T13/T14	$280\,nm$	$500\,nm$
T10	$800\,nm$	$600\,nm$
T11	$1\,\mu m$	$300\,nm$

Table 4.9.3: Optimized dimensions for all the transistors.

The transistors T5, T6, T7, T8 and T10 are used with multipliers of 3 and T9 is used with a multiplier of 7.

4.10 CMFB circuit for the current mirror OTA

The CMFB circuit designed for the fully differential inverter-based OTA is a flipped version of the CMFB circuit used for the fully differential folded cascode OTA. The reason for that is the gate voltage of a PMOS should come from the gate voltage of another PMOS.

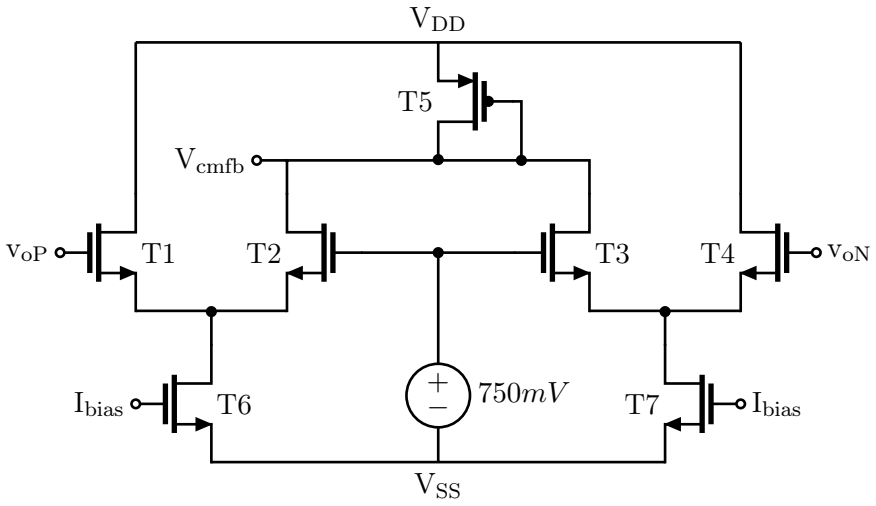


Figure 4.6: Commom-mode feedback circuit for current mirror OTA.

The dimensions for the transistors used in CMFB circuit are optimized and are given in the table 4.10.1.

Transistors	**W**	**L**
T1 - T4	$150\,nm$	$1\,\mu m$
T6/T7	$800\,nm$	$600\,nm$
T5	$800\,nm$	$500\,nm$

Table 4.10.1: Optimized transistors sizing for CMFB circuit.

4.11 Biasing circuit for the current mirror OTA

A current mirror topology based configuration is used to design the biasing circuit for the fully differential inverter-based current mirror OTA. The biasing circuit is shown in the figure 4.7.

The dimensions for the transistors used in the bias circuit are optimized and are given in the table 4.11.1.

Transistors	**W**	**L**
T1/T2	$800\,nm$	$600\,nm$
T3	$1\,\mu m$	$500\,nm$

Table 4.11.1: Optimized transistors sizing for the biasing circuit.

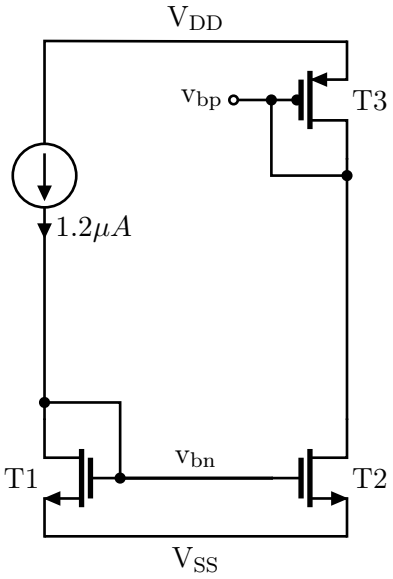


Figure 4.7: Biasing circuit for the current mirror OTA.

Part IV

Results & Discussion

Chapter 5

Simulation Results

This chapter will present the simulation setup and results. The simulations are performed by using the Virtuoso tool for design from Cadence. The simulation environment ADE-L and ADE-XL is utilized. The purpose of the simulations is to verify the small-signal and large-signal properties of the amplifier. Most of the interesting performance parameters are investigated through the simulations and verified with the calculations. This practice will essentially, proves the robustness and reliability of the design.

5.1 Simulation Setup

An analog circuit has to go through several phases. After the design phase, the simulation phase begins. This phase is accomplished typically, by using a simulator which uses much more complex models. Thousands of simulations are needed to be performed prior to sending the design for fabrication. In addition, the simulations are never guaranteed to work but at least the simulations are performed for all the process corners. Process corners are maximum variations one may expect during manufacturing. Also, Monte Carlo statistical analysis with mismatch, is adopted for extracting the second harmonics. After examining the design architecture carefully and verifying different design aspects with the theory, different testbenches are prepared in order to run the simulations. The design schematics and testbenches used for both architectures are shown in the appendices B.1, B.2, B.3 and B.4. All the simulations are performed at $27\circ$.

5.2 DC Simulations

First of all, the DC simulations are performed in order to determine the operating point of the transistors, also known as the bias point, and verify whether the region of operation of the transistors is in accordance with the expected one. After confirming that the design is up and running with all the transistors in the saturation region, the AC simulations are performed in order to find the frequency response of the amplifier which is mentioned in the next section. Moreover, a more visible and zoomed version of DC operating points for both topologies is also shown in appendices B.5 and B.6.

5.3 Frequency response

Here the response of the system for higher or real frequencies is investigated that is why it is called the frequency response. In analog CMOS, typically capacitors are considered as frequency dependent impedances. When dealing with CMOS circuits, the typical limit for the frequency is $\leq 1GHz$ [30]. The simulations are performed for both of the designed architectures. The plots for the DC gain and phase for both of the architectures are shown in the figure 5.1 and figure 5.2.

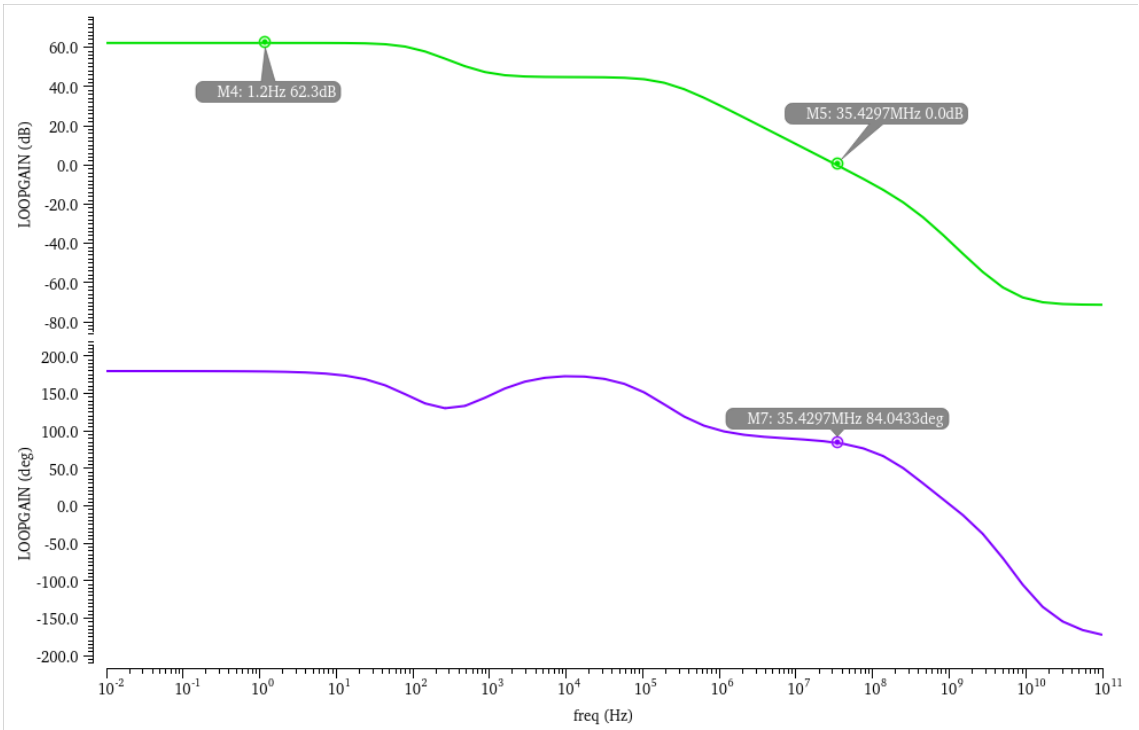


Figure 5.1: DC gain and phase margin for folded cascode OTA.

The dc gain for the fully differential folded cascode architecture is 62.3 dB and 46.9 dB for fully differential inverter-based current mirror OTA. The phase margin is $\approx 84°$ for folded cascode and $\approx 84°$ for the current mirror topology.

5.4 Transient response

The transient response of a circuit over a specified time interval, is computed by performing the transient analysis. The effects of large-signal noise can be examined by the transient analysis. One advantage is that it provides an opportunity to inspect the effects of noise in the time domain. The calculator utility in ADE is used to calculate multiple types of measurements and evaluation of expressions etc. The simulations time is chosen as $50\,\mu S$ in order to assure the steady state response of the signal. The transient noise is also included in the simulations. The bandwidth of the pseudorandom noise source is set to a maximum value of $10\,GHz$. The full spectrum for the transient analysis for the both architectures, is shown in the appendix A.1 and A.2. In order to verify the output swing, transient simulations are also run for $2\,\mu S$. The output swings for folded cascode OTA and current mirror are shown in the figures 5.3 and 5.4 respectively.

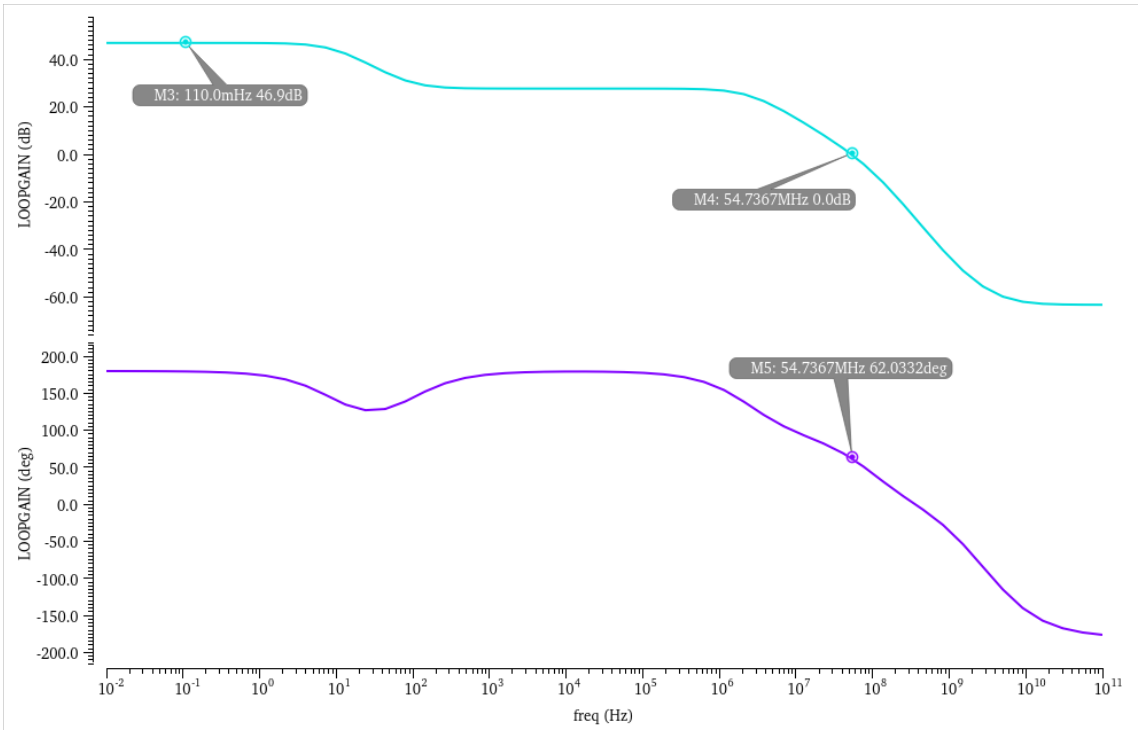


Figure 5.2: DC gain and phase margin for inverter-based current mirror OTA.

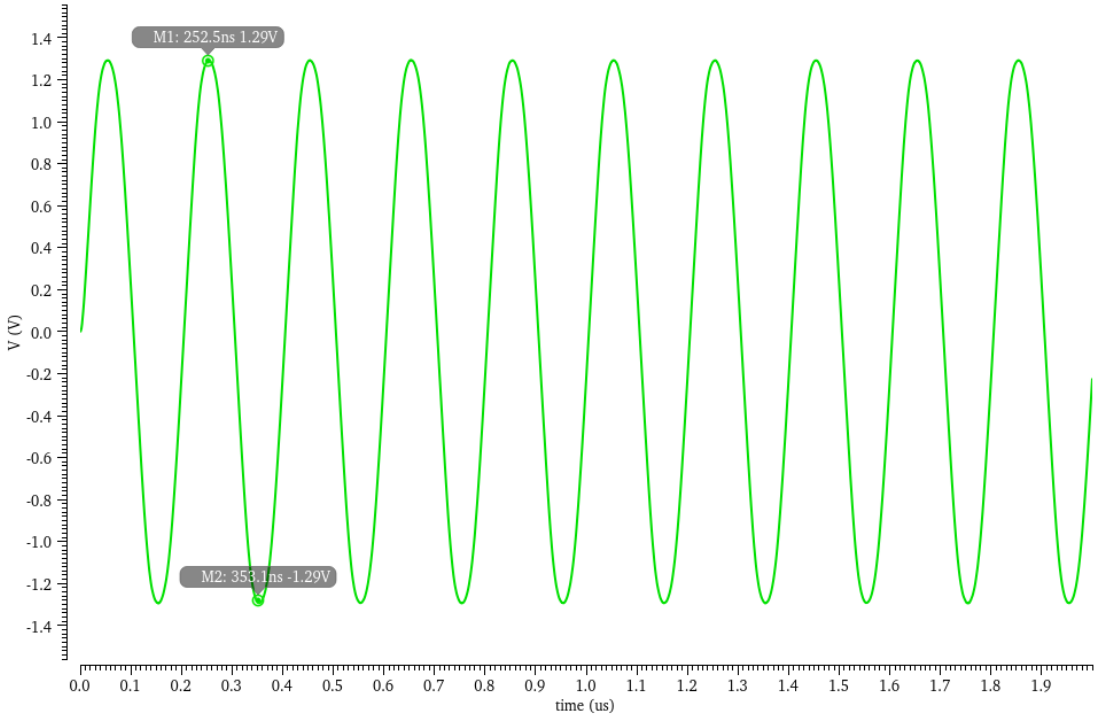


Figure 5.3: Transient response for folded cascode OTA.

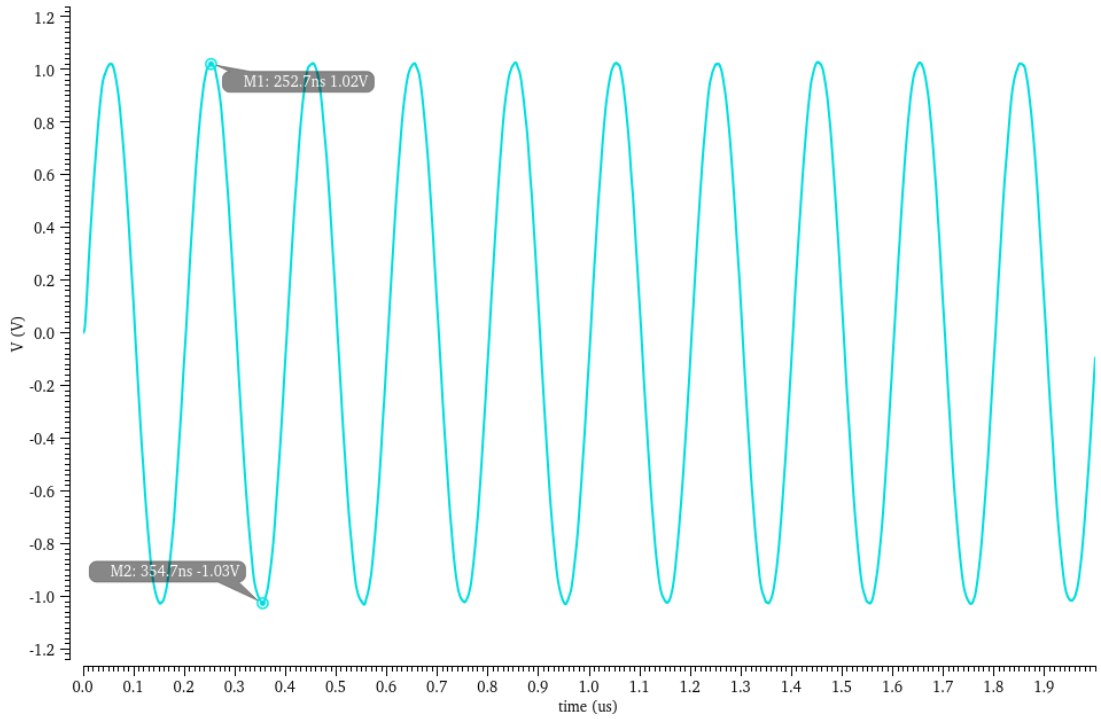


Figure 5.4: Transient response for current mirror OTA.

5.5 Noise Performance

The output noise is simulated through noise analysis with a frequency sweep up til 50 MHz. Logarithmic sweep type is selected with 2000 points per decade. The input-referred noise is then calculated by dividing the resulted output noise by the open-loop gain. The open-loop gain for both architectures are shown in the appendix C.1 and C.3. Two markers are created in order to point out the spots at 10 MHz. The output noise results from simulations are shown in the figures 5.5 and 5.6 for folded cascode and current mirror OTAs respectively. The values are given in the results summary.

5.6 Large-signal analysis

The linearity and large signal gain is verified by performing the transient analysis which is used to analyze the large-signal performance of the amplifier. The large-signal gain and unity gain frequency are extracted for both architectures and shown in the figure 5.7 and figure 5.7.

The large-signal gain and unity gain frequency for the fully differential folded cascode OTA is achieved as 4.97 and 190 MHz. Similarly, the large-signal gain and unity gain frequency for the fully differential current mirror OTA is achieved as 4.8 and 200.45 MHz respectively.

5.7 Mismatch Performance

The mismatch performance for the both architectures is verified by extracting the difference between the fundamental and second harmonic. A Monte Carlo analysis for transient response is performed with 300 iterations. A Monte Carlo is a statistical analysis to

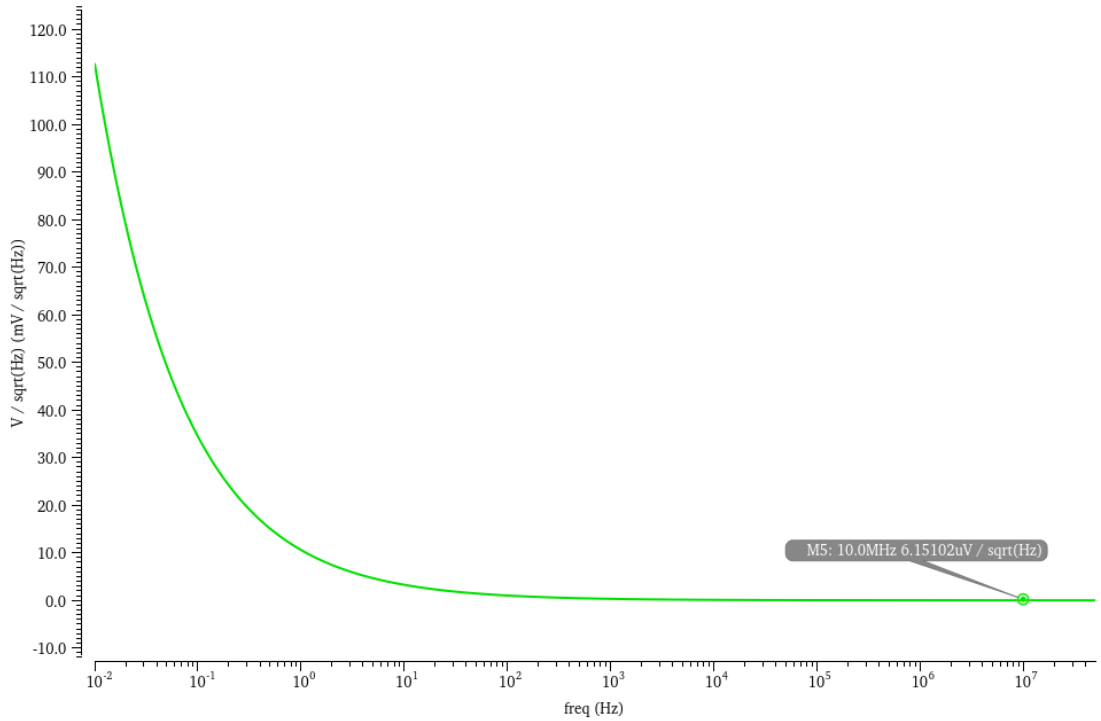


Figure 5.5: Output noise for folded cascode OTA.

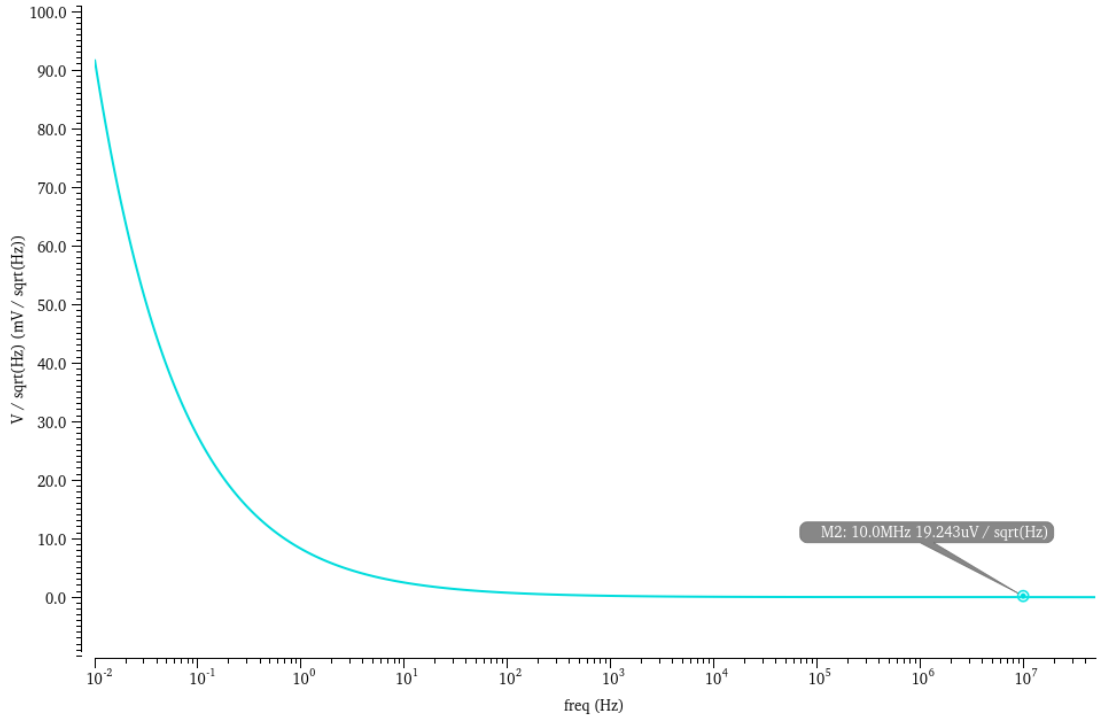


Figure 5.6: Output noise for current mirror OTA.

estimate parametric yields and generate information about performance characteristics. It is used for multiple tests. This method actually simulates the process variations by selecting random points in between the corners. More the number of simulations, more

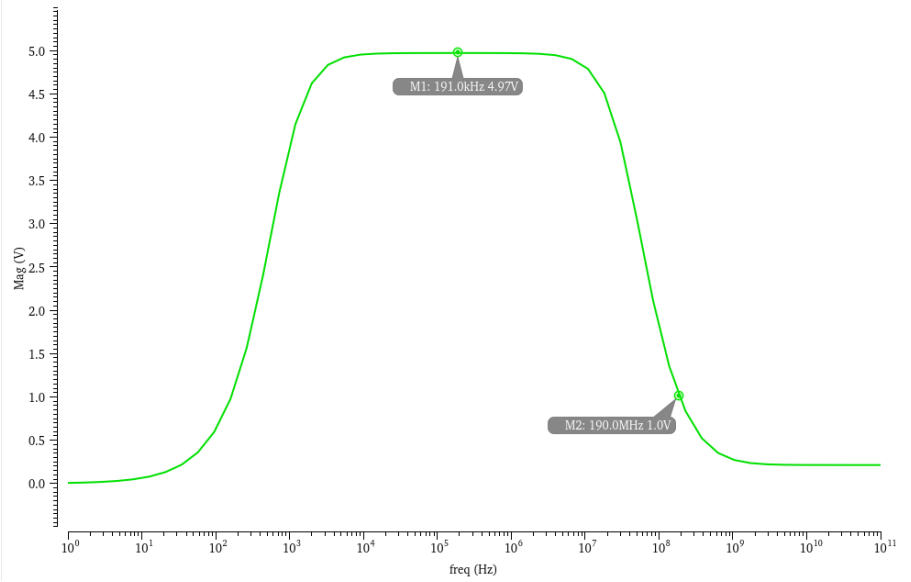


Figure 5.7: Large-signal gain and f_{ug} for folded cascode OTA.

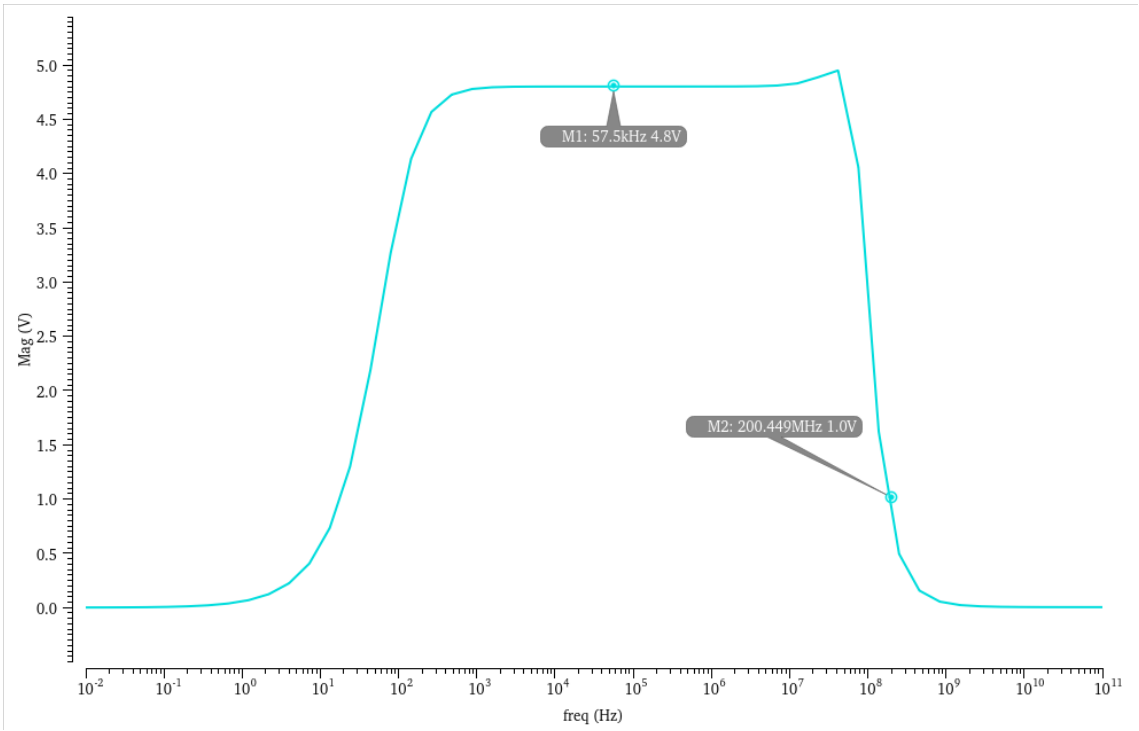


Figure 5.8: Large-signal gain and f_{ug} for current mirror OTA.

the area will be covered. Second option is, one can chose mismatch which is different from process variations and is not desired. Process variations are between two different ICs but mismatch is considered as process variations inside a single IC for example. Both the fundamental and the second harmonic for both architectures are shown in the following figures.

The fundamental, second harmonic and difference between the two is extracted and the values are given in the table 5.7.1 and 5.7.2.

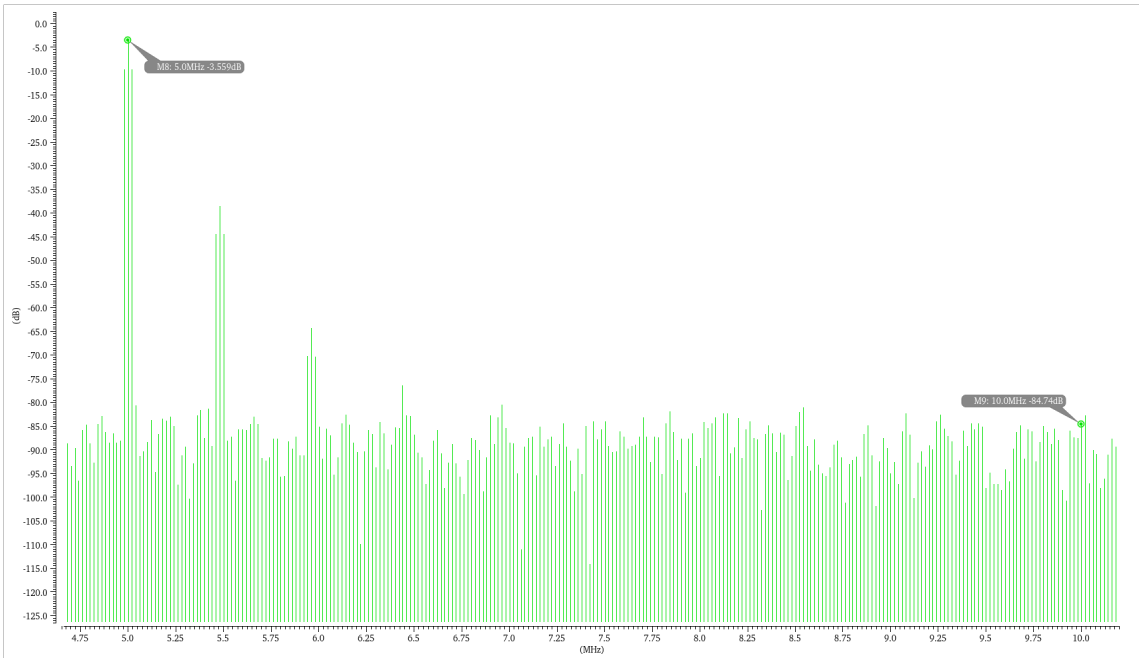


Figure 5.9: Fundamental and 2nd harmonics for folded cascode OTA.

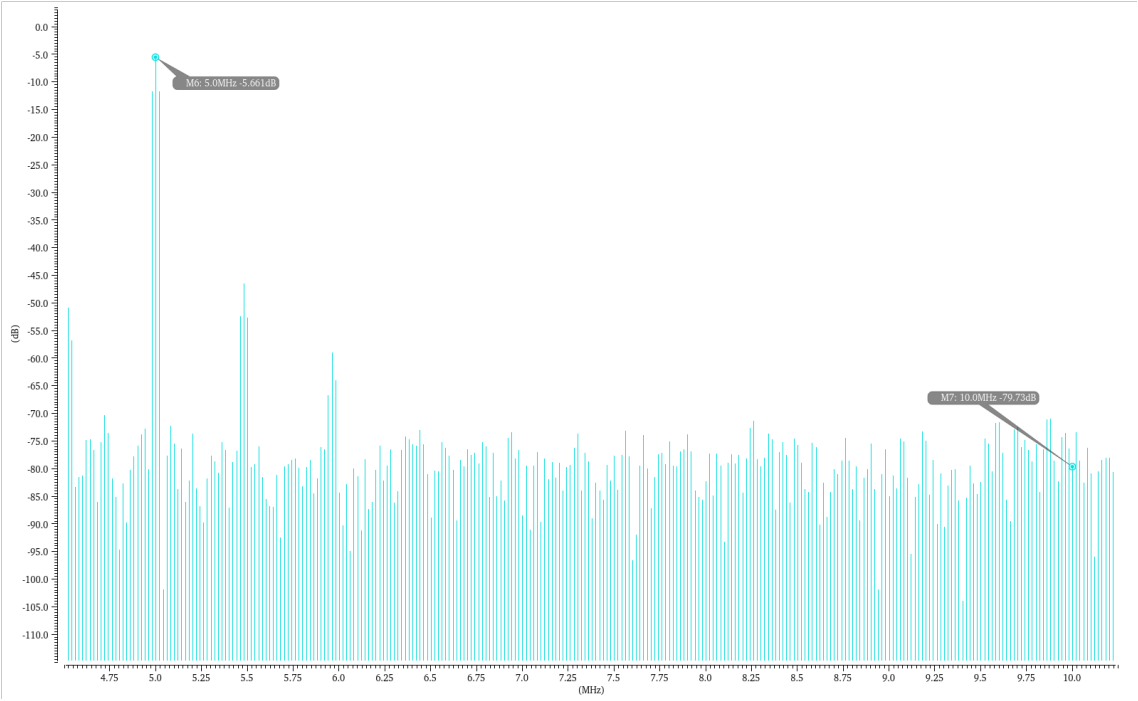


Figure 5.10: Fundamental and 2nd harmonics for current mirror OTA.

The statistical analysis are plotted as histograms and are shown in the figures 5.11 and 5.12.

Parameters	Minimum	Maximum	μ	σ
HD1	-7.745	-3.509	-4.273	619.4m
HD2	-90.54	-28.75	-46.5	8.407
HD1-HD2	22.48	86.99	42.23	8.56

Table 5.7.1: HD1 and HD2 for folded cascode OTA.

Parameters	Minimum	Maximum	μ	σ
HD1	-90.45	-28.08	-49.82	9.15
HD2	-13.99	-5.339	-6.563	1.213
HD1-HD2	21.59	84.87	43.25	9.434

Table 5.7.2: HD1 and HD2 for current mirror OTA.

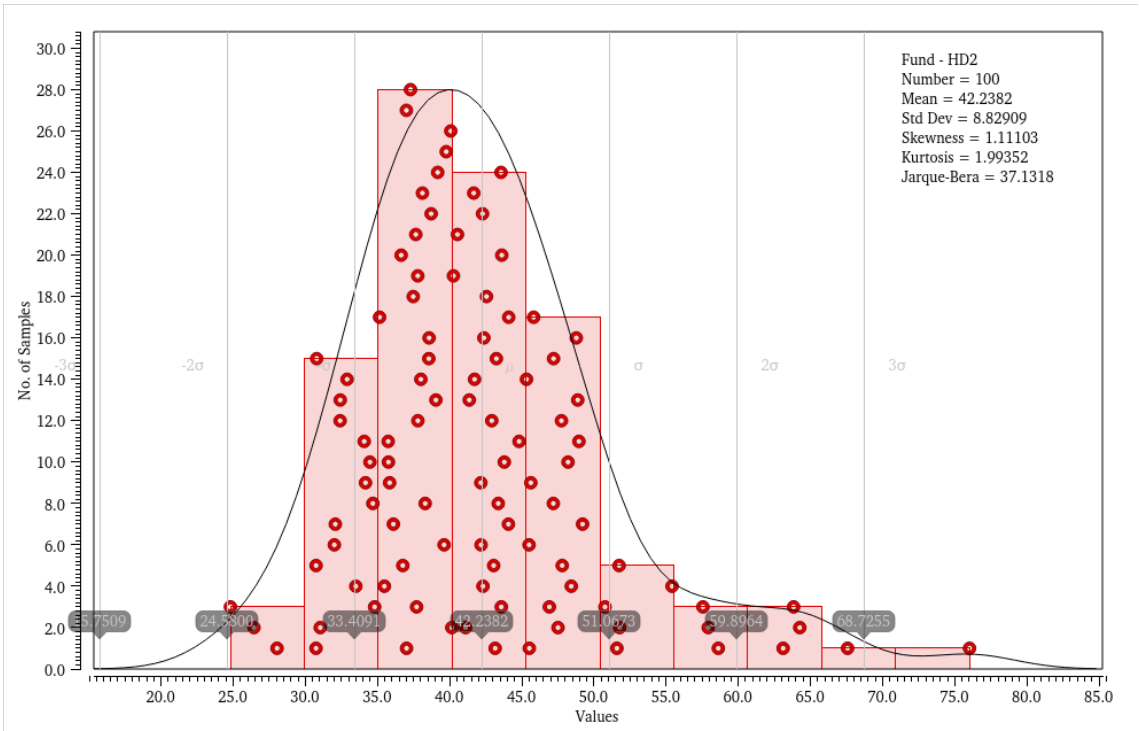


Figure 5.11: Harmonic statistics for folded cascode OTA.

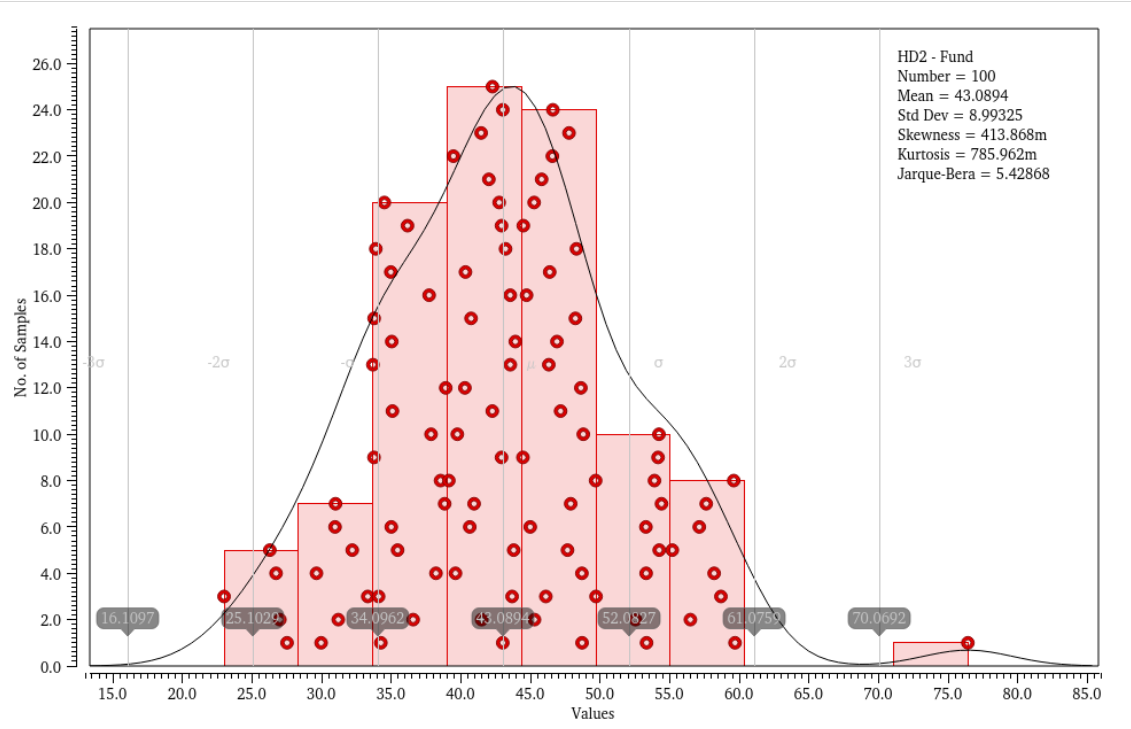


Figure 5.12: Harmonic statistics for current mirror OTA.

5.8 Results Summary

All the important results are summarized in the table 5.8.1. All the simulations are run for all the process corners. In addition, Monte Carlo statistical simulations with mismatch are performed for the second harmonic distortion extractions. The mean values of HD2 are listed in the table. In order to compare the two selected architectures, the desired values according to their specifications are also listed in the table 5.8.1.

Parameters	**Folded Cascode OTA**	**Current Mirror OTA**	**Specifications**
Supply voltage	1.5 V	1.63 V	Max. 1.5 V
DC gain (A_v)	62.3 dB	46.9 dB	$>$ 50 dB
Unity gain frequency	190 MHz	200 MHz	$>$ 200 MHz
Output SNR	58.08 dB	47.25 dB	$>$ 50 dB
Input referred noise	$21.4\,nV/\sqrt{Hz}$	$18.6\,nV/\sqrt{Hz}$	
Phase Margin (PM)	84.3°	$\approx 62°$	At least $45° - 60°$
HD1-HD2	42.23 dB	43.25 dB	$>$ 40 dB
Power consumption	$99.06\,\mu W$	$27.1351\,\mu W$	$< 30\,\mu W$
Current	$66.04\,\mu A$	$18.0901\,\mu A$	$< 20\,\mu A$
Closed loop gain	4.97	4.8	5
FOM	$5.1.10^{-13}$	$5.4.10^{-13}$	

Table 5.8.1: Specified parameter results summary.

Chapter 6

Discussion

In this chapter, the findings of the thesis will be discussed. A summary of all the relevant results is shown in the table 5.8.1. Moreover, the names of the transistors used in the discussion chapter, are taken from the Virtuoso schematics and not from the design chapter. The operating points are care fully observed during the stability and transient analysis and the same value for operating point is used in all the testbenches for both cases.

6.1 Frequency Response

The gm/I_d values were almost the same for both of the selected architectures which justifies the design comparison between them. The load capacitors are set to the value of 50 fF for the fully differential folded cascose OTA and 65 fF for the fully differential inverter-based current mirror OTA. The frequency response is different for the two topologies since the design parameters are different. The value of loop gain achieved for the fully differential inverter-based current mirror OTA, is 46.9 dB which is little less than the desired value of 50dB. MN40 is a diode connected and that is why MN39 is of the same dimensions but with different multiplier of 3. The dimensions for the two PMOS inverter transistors are optimized to approximately double than that of the two NMOS inverter transistors. The reason for that is that the desired amount of bias current in the diode connected MN40 transistor. So, MP40 needs to be wider since the same current will also go through MP38. So, there is definitely a certain ratio between MP38 and MN37. Without gain boosting, the gain was even much lower. The gain boosting, not only boosted the gain but also the unity gain frequency. The two diode connected NMOS transistors, MN56 and MN57, are used in parallel to the MN40 and MN42 respectively. The diode connections for both are modified since both pick their gates from opposite output sides and hence making a cross-coupled topology. The cross-coupled pair increased the impedance level at the node having the diode connection. So, for a certain current comming from the output of the inverter, will get a large voltage variation at the gate which will further increase the current in the output branch. Thus, both DC gain and unity gain frequency are boosted in this way. The width factor of 8 is just experimented. The transconductance of the parallel transistors is achieved by subtracting the gm of MN56 from gm of MN40 since equal transconductances will result in zero transconductance which is not desired. The width of the new transistor, MN56 could not be increased further because the PM starts decreasing on the other side. This is what we call a positive feedback. Both have same lengths as MN40 and MN42 but widths are taken as a factor of 8th of the widths of MN40 and MN42. Multipliers are used

in the circuit with only a $1.2\,\mu A$ current source. But this is not the case for MP40 and MP41 since they do not have same gate voltages. Two resistors each of $1\,T\Omega$ are used in the testbench for current mirror OTA. The purpose of the resistors is to copy the DC level from the output of the inverters to the input of the inverters. The load capacitor used for current mirror OTA is optimized to 65 fF in order to set the noise level and increase the PM since it is a load compensated OTA with the dominent pole at the output. Adding of more capacitance drops the unity gain frequency which is already at the boundary. On the other side, the loop gain for the fully differential folded cascode is 62.3 dB which is well above the desired value. The phase marigin (PM) is also simulated for both topologies and is confirmed to be good enough prior to investigate the other performance parameters. But here we are interested only in the PM for the loop gain. The folded cascode OTA shows a good PM of 84.3° but the current mirror OTA also shows $\approx 62°$ which of course, also meets the requirement. In addition, while performing the stability analysis, the phase margin is also simulated for the CMFB loop. All the loops are checked during the stability simulations.

6.2 Transient Response

The DC level is zero in both plots shown in the figure 5.3 and 5.4 because of the differential output. The input differential amplitude for the folded cascode is 278 mV and the expected output swing should be 1390 mV for a gain of 5. But the resulted value is 1290 mV which is quite close to the expected swing. Similarly, for the current mirror OTA, the input differential amplitude is 220 mV and the expected output swing should be should be 1100 mV for a gain of 5. But the simulation results show 1020 mV which is also quite close to the expected output swing. So, both topologies perform nice transient responses without any clipping. For the current mirror OTA, SNR is 47.24 dB and is detected from the frequency spectrum by using hanning window with 1024 samples. SINAD is 38.6 dB but it is not relevant for low power ultrasound applications since it includes all the harmonics. Input amplitude is increased to 110 mV in order to increase SNR but it could not be increased more because then it would be difficult to satisfy the HD2 requirements since achieved HD2 values are already on their limits with respect to the specifications. So, there is always a trade-off between distortion and SNR.

6.3 Noise

In this thesis work, the noise generated only inside the device is considered, so the resistors and transistors are the two main contributor of the noise. In frequency domain, the noise spectral density is obtained. Noise spectral density is decresing with the frequency. Flicker noise is neglected here since it is just at the low frequencies but then the flat spectrum shows typically, the thermal noise. At high frequencies, the noise is filtered because of the capacitors. The input-referred noise is $18.6\,nV/\sqrt{Hz}$ for current mirror OTA and $21.4\,nV/\sqrt{Hz}$ for the folded cascode OTA. Current mirror has lower value, so it is better in terms of input-referred noise. The thermal noise is also proportional to the gm of a transistor. So, there are only two ways to reduce the noise, one to increase the load capacitance and the other to reduce the ratio between the gm_s of the MP42 and MN43 because the temperature can not be changed [32]. Because of this reason, generally, the KT/C_L noise is added in the end of the design phase.

6.4 Linearity

The second harmonic extractions are 43.25 dB and 42.23 dB for current mirror OTA and folded cascode OTA respectively. There is not that much difference between the two values in terms of linearity requirements. Both values just passed the linearity requirements. In addition, the simulations are performed with mismatch because the circuit is symmetrical. Otherwise, without mismatch, the harmonic distortion could artificially be very good which means harmonic affect is very low relative to the noise affect which further results in close values for SINAD and SNR. There are three spikes both for the fundamental and second harmonic. These are created by taking the fast fourier transfer (FFT) of the signal. The fundamental or main spike is at 5 MHz and the other two are closer to the fundamental. The total power of the signal is calculated by summing all the three spikes. These values are achieved for high input amplitudes since distortion is very less and is dumped for the low amplitudes. So, the power in the second harmonic should be at least 40 dB lower than the power in the fundamental harmonic.

6.5 Comparison

The most important equation is the FOM for comparison than the individual parameters. The FOM is calculated according to the equation 4.2. The value obtained for the folded cascode is 5.1 e-13 and for the current mirror is 5.4 e-13. This means that the fully differential folded cascode is better in terms of overall performance since it has a low FOM. But one can see that the current mirror is much better in terms of power consumption and input-referred noise since both values are on the numerator of the expression for the FOM. The folded cascode OTA uses almost four times more power than the current mirror OTA. While comparison, the power supplies should also be the same for both of the topologies. Unfortunately, the power supply used for the current mirror OTA is 1.63 V and the reason for that is the current mirror OTA did not pass the slow-slow process corner simulations. On the other side, the lower power supplies would also have effects on the SNR requirements because of the reduction in the signal power.

Chapter 7

Conclusion

In this thesis work, a low noise amplifier for medical ultrasound applications is designed in Cadence by using a $130\,nm$ commercially available CMOS technology. Two different architectures, a fully differential folded cascode OTA and a fully differential inverter-based current mirror OTA, are designed with their biasing and common-mode feedback circuits. The purpose of designing the two different topologies, is to compare the two topologies and find out the better one. The main focus is to investigate and compare SNR, PM and HD2 performance parameters during the comparison. Both architecture designs are based on gm/I_d methodology.

Both architectures under investigation, show some advantages and disadvantages. As mentioned in the discussion section, the fully differential inverter-based current mirror OTA, achieved a low power consumption level of $\approx 27\,\mu W$ with a second harmonic distortion level of $\approx 43.3\,dB$. This topology achieved very promising results in terms of power consumption and input referred-noise. On the other hand, the fully differential folded cascode OTA achieved a power consumption of $\approx 99\,\mu W$ with a second harmonic distortion level of $\approx 42.3\,dB$. This topology fulfills the requirement given in the design specifications. The closed loop gain is almost the same for both of the architectures. Even though the folded cascode seems to be much better in terms of PM, SNR and, DC gain, supply voltage and of course, the overall performance is also better according to the FOM but it also uses a lot of power which is of course, a disadvantage of this topology. Finally, because of the extra low power consumption, the inverter-based current mirror OTA should be a wise selection for the industry since the low power consumption is the most important and a key parameter for a low power LNA design.

7.1 Future work

The design schematics for two different architectures along with their simulations are presented in this thesis work. The overall performance of both architectures have been decent and reliable. But despite of some good results, there are some deficiencies in the design phase and most of them are already discussed in the discussion section. For example, the fully differential inverter-based current mirror may be further tweaked in order to improve the DC gain and SNR requirements since these requirements are not achieved according to the specifications. In addition, the Layout and post layout simulations may also be performed for the two architectures which could also help in proving the robustness and reliability of the design. In the end, the fully differential inverter-based current mirror

OTA should also be tweaked with a lower power supply. Thus, the mentioned deficiencies could be a potential task for the future work. On the other hand, specially the folded cascode OTA can also be further optimized in terms of power consumption since it used a lot more power then the inverter-based current mirror OTA.

Appendices

Appendix A

Transient Analysis

A.1 Full spectrum for fully differential folded cascode OTA

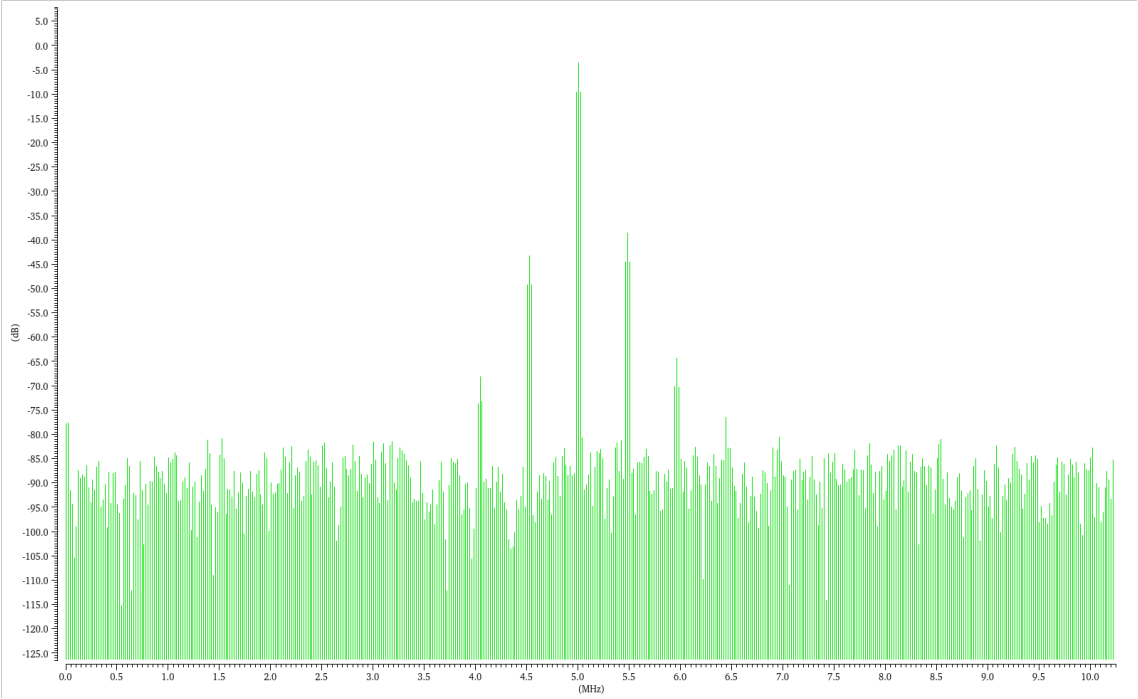


Figure A.1: Transient response.

A.2 Full spectrum for fully differential inverter-based current mirror OTA

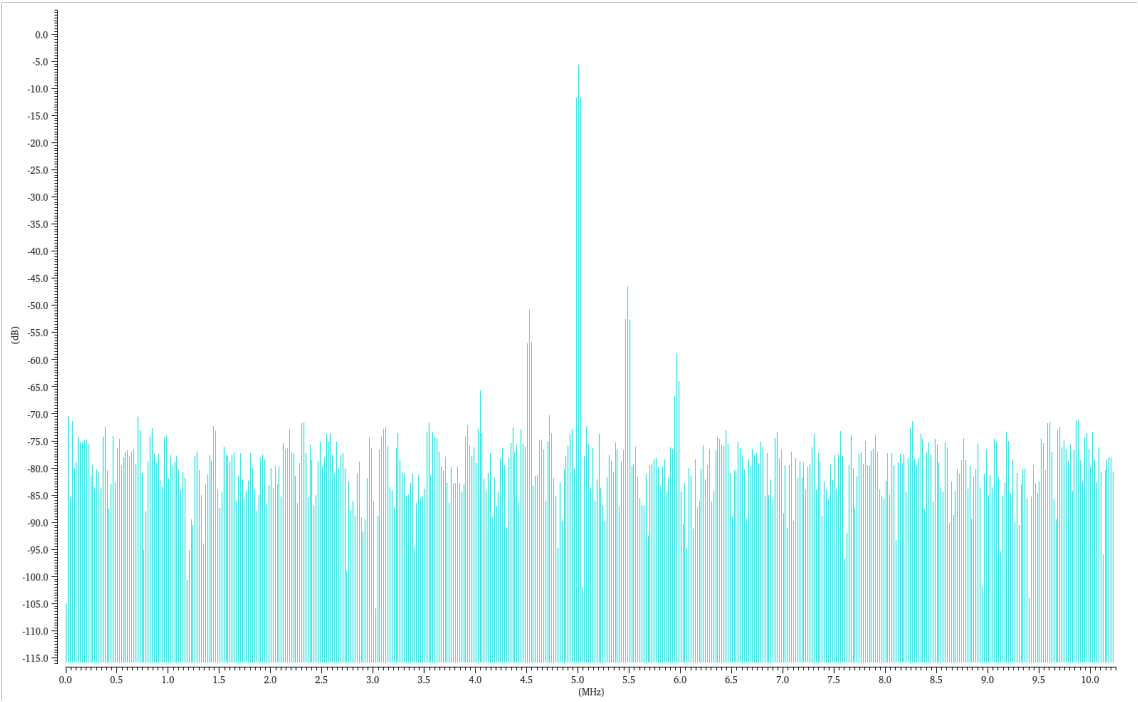


Figure A.2: Transient response.

Appendix B

Testbenches and Schematics

B.1 Schematic for fully differential folded-cascode OTA

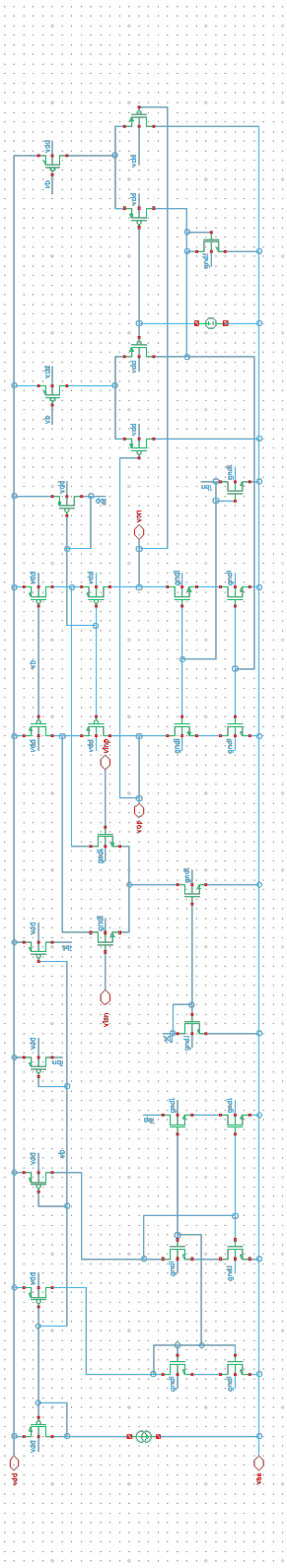


Figure B.1: Virtuoso schematic.

B.2 Schematic for fully differential inverter-based OTA

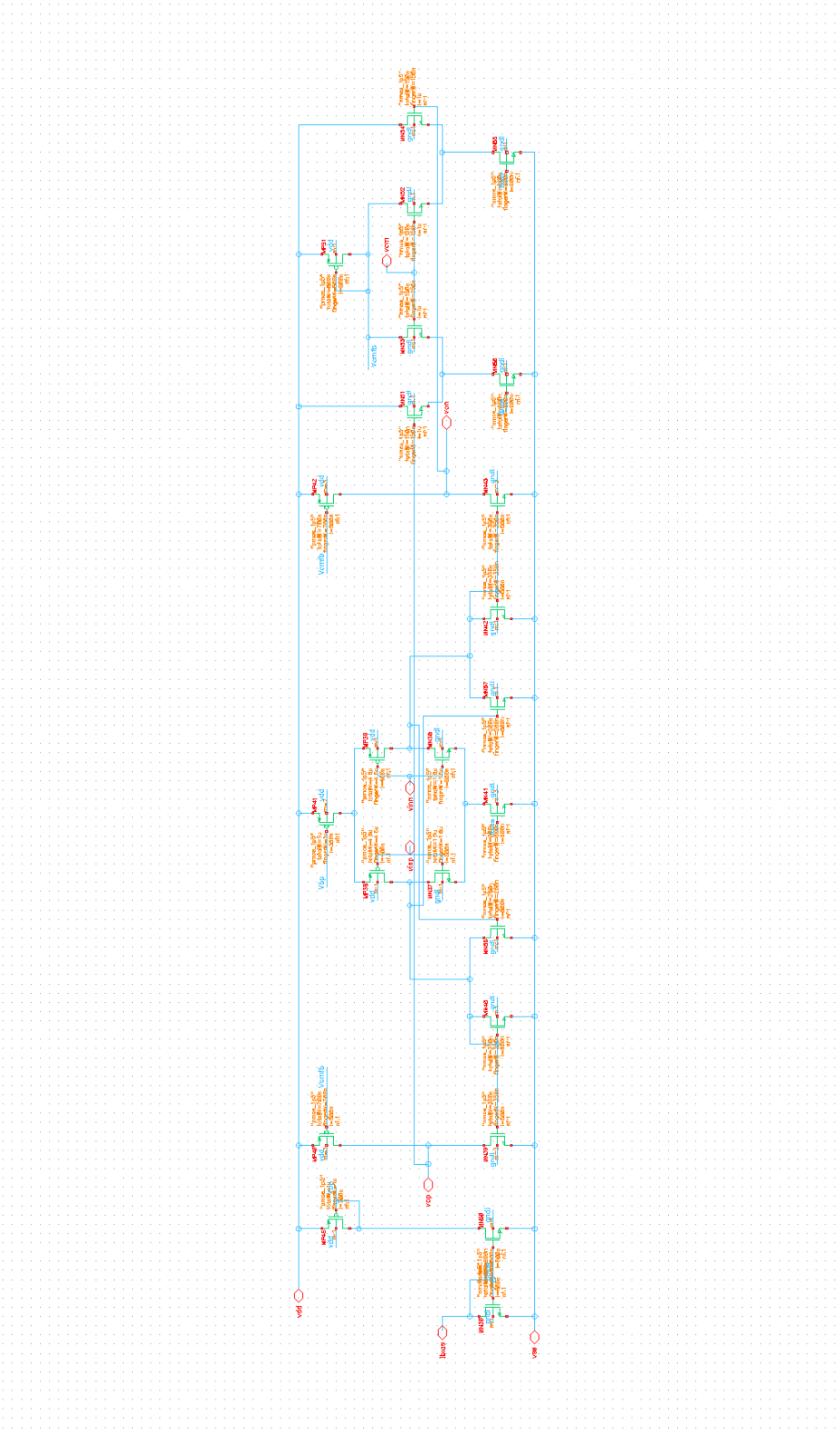


Figure B.2: Virtuoso schematic.

B.3 Testbench for fully differential folded-cascode OTA

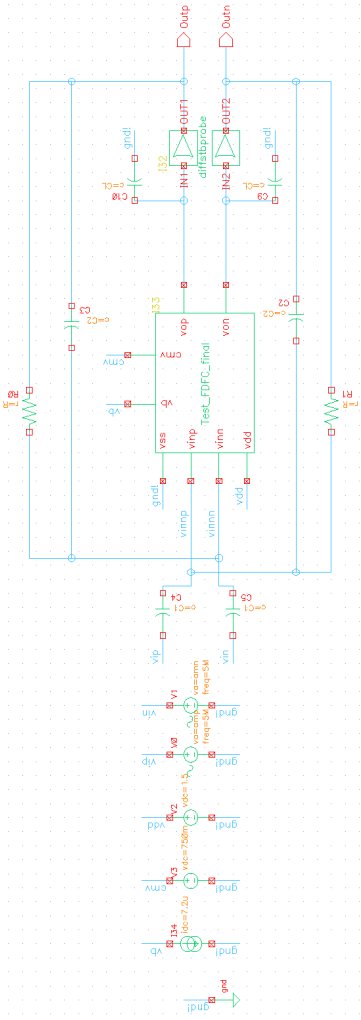


Figure B.3: Virtuoso schematic.

B.4 Testbench for fully differential current mirror OTA

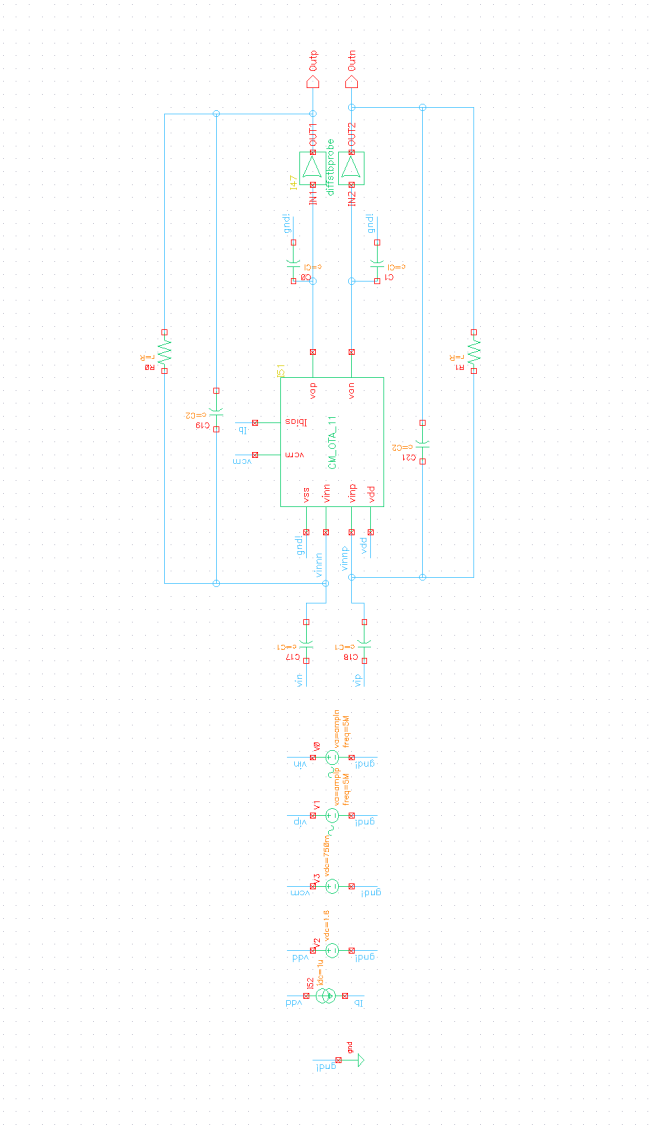


Figure B.4: Virtuoso schematic.

B.5 DC operating points for fully differential folded-cascode OTA

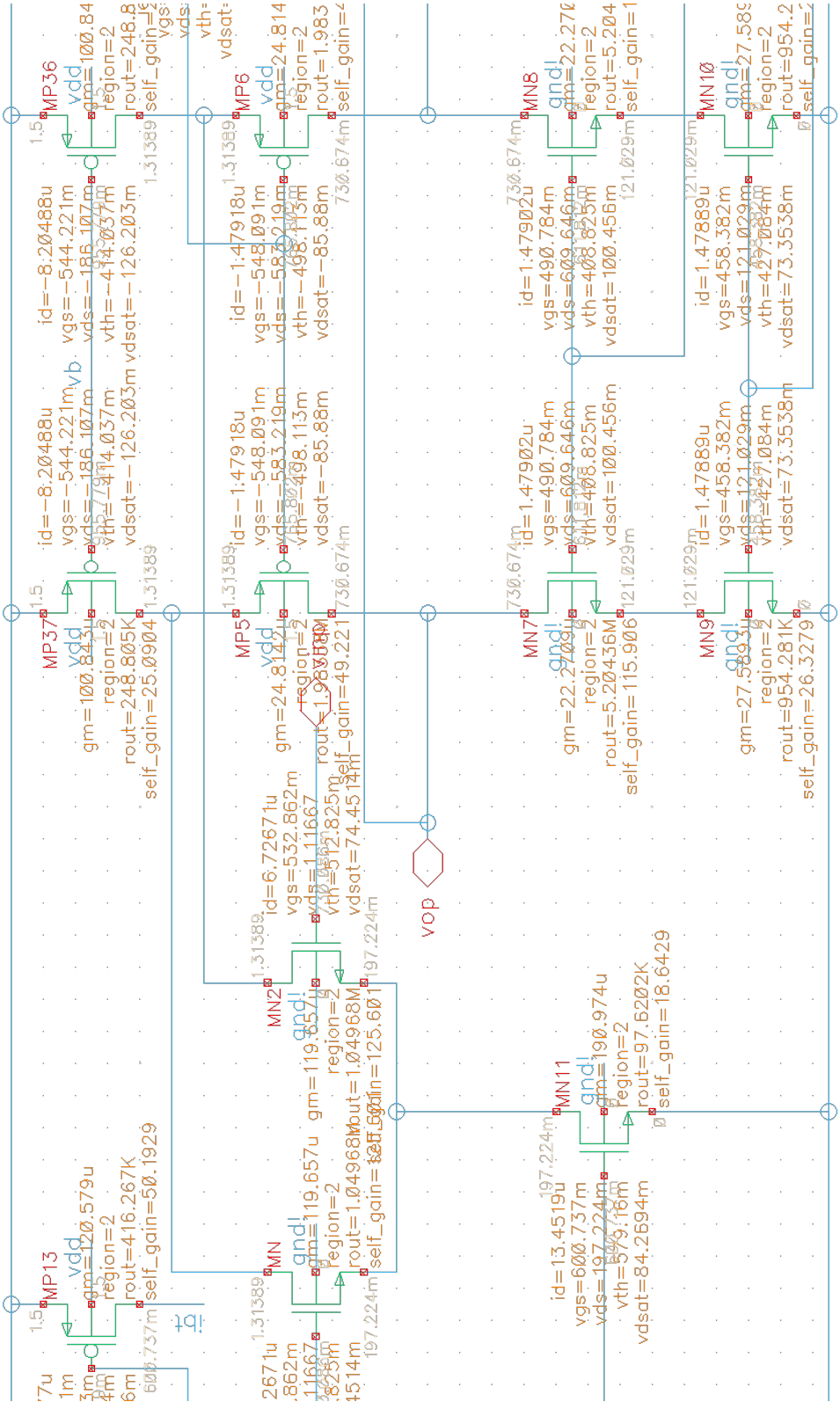


Figure B.5: DC operating points.

B.6 DC operating points for fully differential inverter-based OTA

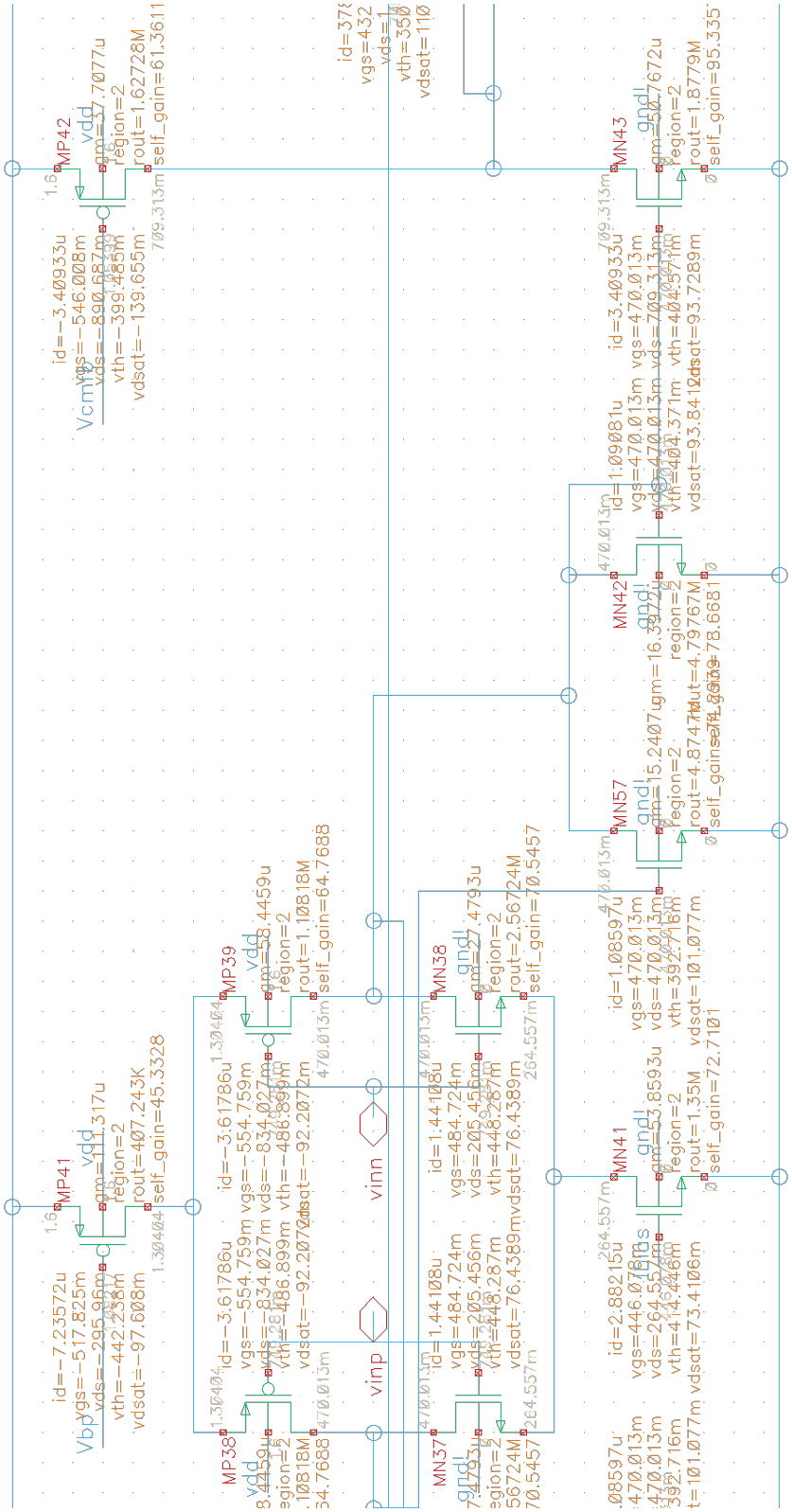


Figure B.6: DC operating points.

Appendix C

Open loop gain curves and testbenches

C.1 Open loop gain for fully differential folded-cascode OTA

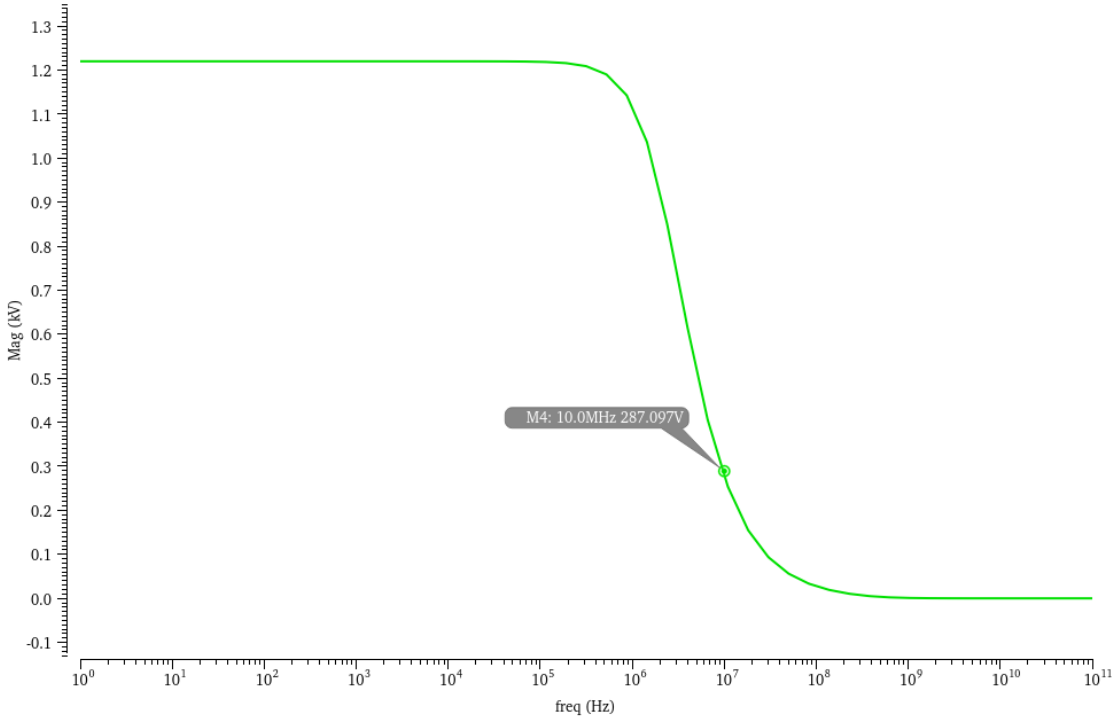


Figure C.1: Open loop gain characteristic curve.

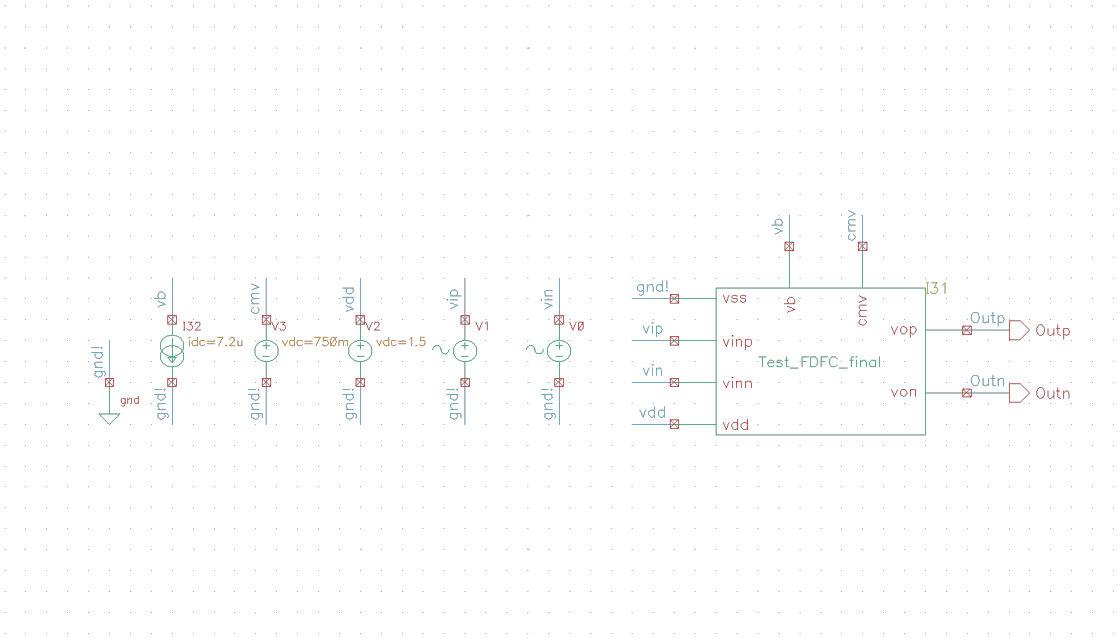


Figure C.2: Testbench for open loop gain.

C.2 Open loop gain for fully differential inverter-based OTA

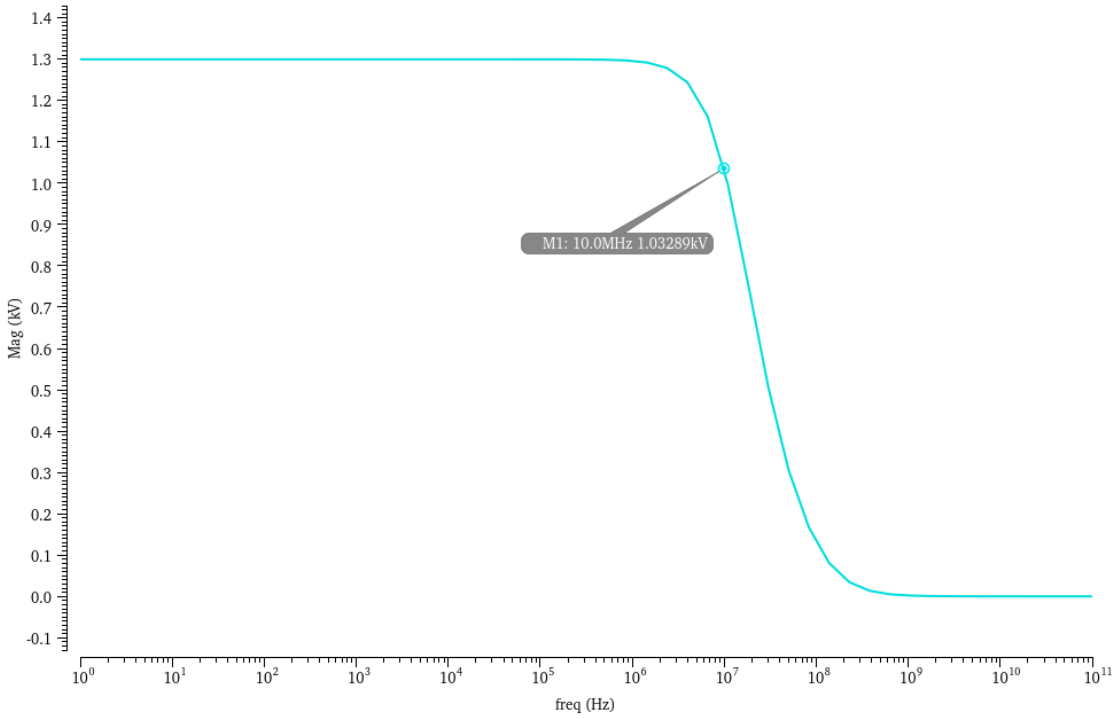


Figure C.3: Open loop gain characteristic curve.

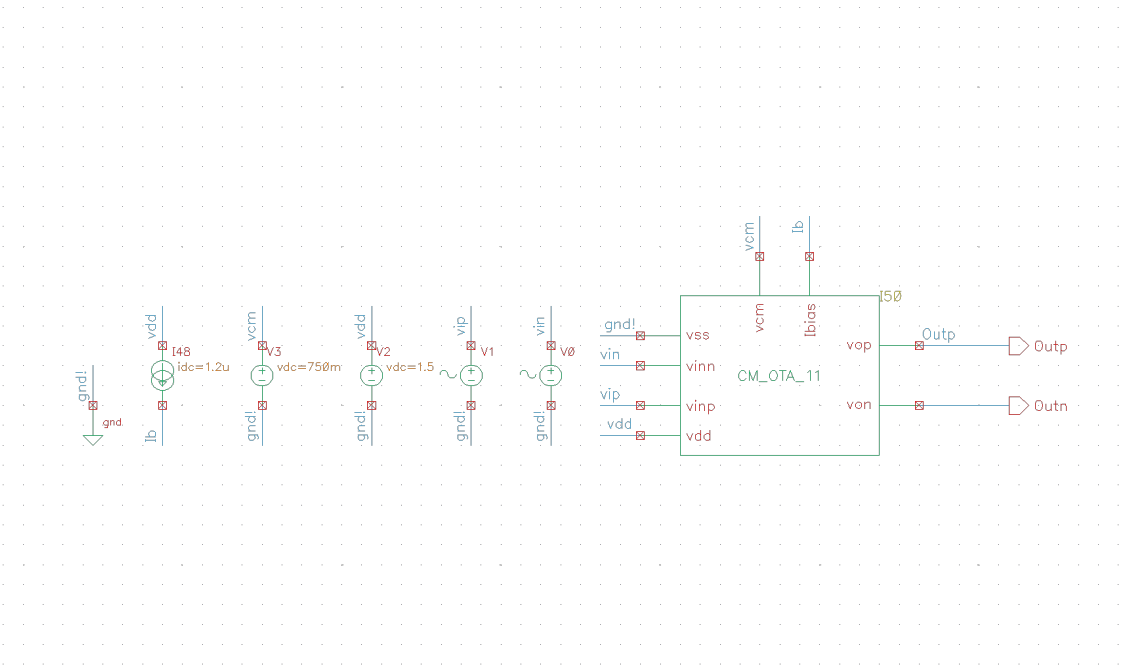


Figure C.4: Open loop gain characteristic curve.

Bibliography

[1] Simon Kristoffer Relling Berg. *A Chopper Offset-Stabilized Operational Amplifier in 22nm SOI*. NTNU, Trondheim, Norway, June 2019.

[2] National Institute of Biomedical Imaging and Bioengineering. https://www.nibib.nih.gov/science-education/science-topics/ultrasound.html. **Accessed: 2021-07-05**.

[3] Endre Rørlien Bjørsvik. *Design of Energy Efficient Low Noise Amplifiers in 40 nm bulk CMOS and 28 nm FD-SOI for Intravascular Ultrasound Imaging*. NTNU, Trondheim, Norway, July 2014.

[4] T.C. Carusone D.A. Johns and K.W. Martin. *Analog Integrated Circuit Design, second edition*. John Wiley & Sons, Ltd., 2011.

[5] Jon Håvard Eriksrød. *A 65nm CMOS Front-end LNA for Medical Ultrasound Imaging with Feedback Employing Noise and Distortion Cancellation*. NTNU, Trondheim, Norway, June 2013.

[6] D. Flandre F Silveira and P. G. A. Jespers. "A gm/I_d Based Methodology for the Design of CMOS Analog Circuits and Its Application to the Synthesis of a Silicon-on-Insulator Micropower OTA." In: *IEEE Journal Of Solid-State Circuits* 31 (9 Sep. 1996), pp. 1314–1319.

[7] U.S. FOOD and DRUG Administration. https://www.fda.gov/radiation-emitting-products/medical-imaging/ultrasound-imaging.html. **Accessed: 2021-07-05**.

[8] M. Gaibotti G. Palumbo D. Pappalardo. "Charge-pump circuits: Power consumption optimization." In: *IEEE Tran. Circ. Systems* 49 (Nov. 2002), pp. 1535–1541.

[9] Martin Gustafsson. "Integrated Low Power Ultrasound Sensor Interfaces." PhD thesis. Department of Computer Science and Electrical Engineering, Luleå University of Technology, 2005.

[10] Hans Herman Hansen. *A 33 μW Sub-3 dB Noise Figure Low Noise Amplifier for Medical Ultrasound Applications*. NTNU, Trondheim, Norway, July 2011.

[11] J. V. Hatfield and K. S. Chai. "A beam-forming transmit ASIC for driving ultrasonic arrays." In: *Sensors and Actuators A* A92 (Aug. 200), pp. 273–279.

[12] Sudhir M. Mallya (member IEEE) and Joseph H. Nevin (member IEEE). "Design Procedures for a Fully Differential Folded-Cascode CMOS Operational Amplifier." In: *IEEE Journal Of Solid-State Circuits* 24 (6 Dec. 1989), pp. 1737–1740.

[13] Mohd Fairus Ahmad Izatul Syafina Ishak Sohiful Anuar Zainol Murad and Neoh Siew Chin, eds. *Design of Folded Cascode Operational (Op-Amp) with Common-mode Feedback (CMFB) for Pipeline ADC* (School of Microelectronic Engineering, University Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia.). Malaysian Technical Universities Conference on Engineering & Technology (MUCET), Dec. 2013.

[14] Jørgen Arendt Jensen. "Medical ultrasound imaging." In: *Progress in Biophysics and Molecular Biology* 93 (2007), pp. 153–165.

[15] Kehul A. Shah Kalpesh B. Pandya. "Performance Evaluation of Different Types of CMOS Operational Transconductance Amplifier." In: *International Journal of Science and Research (IJSR)* 2 (Gujarat, India, March. 2013).

[16] Carsten Wulff Lanny L. Lewyn Trond Ytterdal and Kenneth Martin. "Analog Circuit Design in Nanoscale CMOS Technology." In: *Proceedings of the IEEE* 97 (Oct. 2009), pp. 1687–1714.

[17] Endre Larsen. *A fully Differential Front End Charge Sampling Amplifier for Medical Ultrasound Imaging in 28nm FD-SOI Technology.* NTNU, Trondheim, Norway, June 2015.

[18] Kebria Naderi, Erwin H. T. Shad, Marta Molinas, Ali Heidari, and Trond Ytterdal. "A Very Low SEF Neural Amplifier by Utilizing a High Swing Current-Reuse Amplifier." In: *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS).* 2020, pp. 1–4.

[19] Svetoslav Nikolov. "Synthetic aperture tissue and flow ultrasound imaging." PhD thesis. Department of Electrical Engineering, Technical University of Denmark, 2002.

[20] Carsten Wulff Øyvind Berntsen and Trond Ytterdal. *High Speed, High Gain OTA in a Digital 90nm CMOS Technology.* Department of Electronics and Telecommunication, NTNU, N-7491 Trondheim, Norway.

[21] David M. Pozar. *Microwave Engineering, fourth edition.* John Wiley & Sons, Inc., 2012.

[22] P. R. Bajaj & A. Y. Deshmukh Ratnaprabha W. Jasutkr. "Design of 1V, 018 μ Folded Cascode Operational Amplifier For Switch Capacitor Sigma Delta Modulator." In: *International Journal of Electrical and Electronics Engineering Research (IJEEER)* 3 (4 Oct. 2013), pp. 223–232.

[23] Behzad Razavi. *Design of Analog CMOS Integrated Circuits.* The McGraw-Hill Companies, Inc., 2001.

[24] Ghous Bakhsh Narejo Saleha Bano and S. M. Usman Ali Shah. "Low Power Fully Differential Folded Cascode OTA with CMFB Circuit." In: *Journal of Multidisciplinary Engineering Science and Technology (JMEST)* 4 (Electronic Engineering Department, NED University of Engineering and Technology, Karachi, Pakistan 7, 2017), pp. 1314–1319.

[25] Erwin H. T. Shad, Tania Moeinfard, Marta Molinas, and Trond Ytterdal. "A Power Efficient, High Gain and High Input Impedance Capacitively-coupled Neural Amplifier." In: *2020 IEEE Nordic Circuits and Systems Conference (NorCAS).* 2020, pp. 1–5.

[26] Erwin H. T. Shad, Kebria Naderi, and Marta Molinas. "High Input Impedance Capacitively-coupled Neural Amplifier and Its Boosting Principle." In: *2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS).* 2020, pp. 1–4.

[27] Sherif M. Sharroush, ed. *Design of CMOS inverter-based amplifier: A quantitative approach* (Department of Electrical Engineering, Faculty of Engineering, Port-Said, 42523 Port-Fouad, Egypt). International Journal of Circuit Theory and Applications, 2011.

[28] Chin-Kung Wu Tsung-Hsien Lin (member IEEE) and Ming-Chung Tsai. "A 0.8-V 0.25-mW Current-Mirror OTA With 160-MHz GBW in 0.18-μm CMOS." In: *IEEE Transactions on Circuits and Systems* 54 (Feb. 2007), pp. 131–135.

[29] Razvan G. Vieru and Romeo Ghinea, eds. *Inverter-Based Ultra Low Voltage Differential Amplifier* (Faculty of Electronics, Telecommunications & Information Technology, Technical University "Gh. Asachi", Iasi, Romania). IEEE, 2011.

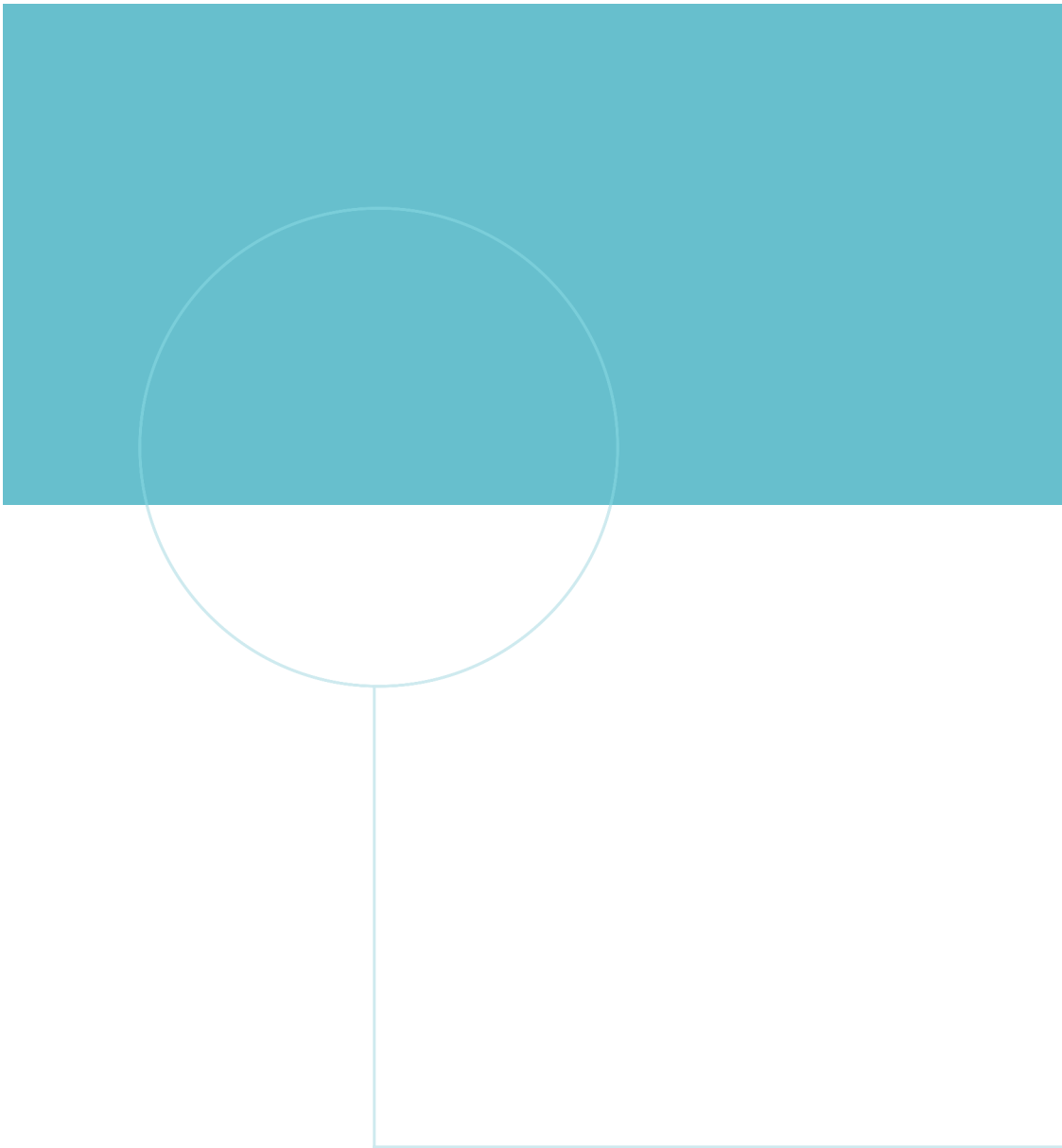
[30] Ryan Selby. William Wilson Tom Chen, ed. *A Current-Starved Inverter-Based Differential Amplifier for Ultra-Low Power Applications* (Fort Collins, USA). Department of Electrical and Computer Engineering, Colorado State University, USA, 2013.

[31] Shubhara Yewale and R. S. Gamad, eds. *Analysis and Design of High gain Low Power Fully Differential Gain-Boosted Folded-Cascode Op-amp with Settling time optimization* (Department of Electronics and Instrumentation Engineering, SGSITS, 23, Park Road, Indore, M.P. India-452003). International Journal of Engineering Research ans Applications (IJERA).

[32] Trond Ytterdal. *Analog CMOS1*. Course lectures at NTNU, Trondheim, Norway, 2020.

[33] Trond Ytterdal and Carsten Wulff. "On the Energy Efficiency of Analog Circuits in Nanoscale CMOS Technologies." In: *2008 International Conference on Microelectronics* (NTNU, Trondheim. 2008).

[34] Houda Daoud Dammak S. Bensalem S. Zouari and M. Loulou. "Design of Folded Cascode OTA in Different Regions of Operation through gm/I_d Methodology." In: *World Academy of Science, Engineering and Technology* 45 (2008), pp. 28–33.



NTNU

Norwegian University of
Science and Technology