FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Self-Calibrated Current Reference

Bruno Miguel da Silva Teixeira

FOR JURY EVALUATION

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Vítor Grades Tavares

Synopsys Supervisor: Joaquim Machado

June 29, 2015

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

MIEEC - MESTRADO INTEGRADO EM ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES | **2014/2015**

A Dissertação intitulada

"Self Calibrated Current Reference"

foi aprovada em provas realizadas em 22-07-2015

o júri

Presidente Professor Doutor José Alberto Peixoto Machado da Silva
Professor Associado do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

Professor Doutor Paulo Mateus Mendes
Professor Auxiliar do Departamento Eletrónica Industrial da Universidade do Minho

Professor Doutor Vitor Manuel Grade Tavares
Professor Auxiliar do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.

Autor - Bruno Miguel da Silva Teixeira

Faculdade de Engenharia da Universidade do Porto

Resumo

Uma referência é um bloco essencial em muitas aplicações de sinal misto e radiofrequência, como conversores de informação, PLL e conversores de energia. A implementação mais usada em CMOS para referências de tensão é o circuito Bandgap, devido não só á sua previsibilidade, mas também á sua baixa dependência da temperatura e tensão de alimentação.

Correntes de referências estão também a ganhar um papel importante em sistemas elétricos devido á sua maior velocidade. O processamento de sinais em forma de corrente é executado de forma mais rápida do que sinais em forma de tensão. Portanto, para uma certa tecnologia, circuitos analógicos desenhados em modo de corrente operam mais rapidamente do que os de tensão.

Este trabalho estuda a referência bandgap. Uma das topologias mais utilizadas em ambientes de baixa tensão foi analisada e desenhada. Os problemas e métricas mais pertinentes associadas a este circuito, incluindo precisão, rejeição de ruido e comportamento da temperatura foram também discutidos. As limitações deste sistema devido a variações de processo foram estudadas e vários métodos de compensação foram sugeridos e aplicados.

Um segundo bloco foi também desenvolvido para efetivamente converter a tensão do circuito de bandgap para uma corrente de referência. Uma topologia baseada em reguladores de tensão foi proposta e desenvolvida. Devido a grandes variações de processo desta topologia, uma resistência programável e respetivo bloco de calibração foram desenvolvidos de modo a atingir a precisão desejada.

Todos estes circuitos foram implementados em tecnologia CMOS TSMC 40nm.

Abstract

A constant reference is a pivotal block in several mixed-signal and radio-frequency applications, like for instance, data converters, PPLs and power converters. The most used CMOS implementation for voltage references is the Bandgap circuit due to its high-predictability, and low dependence of the supply voltage and temperature of operation.

Current references are also gaining an important role in electric systems due to the increased speed. Processing current signals is done faster than voltage signals and, therefore, for a given technology, analog circuits designed in current mode operate faster than their voltage mode counter parts

This work studies the bandgap voltage reference. One of the most relevant topologies for low voltage environments is designed and analyzed. The most relevant issues and performance metrics for BGR, including accuracy, PSRR and temperature behavior are also discussed. The limitations of this system due to process variations are studied and several methods for correction are proposed and employed.

A second block was designed in order to effectively act as a converter between the voltage reference into a current reference. A topology based on low dropout regulator was proposed and developed. Due to heavy process variation, a programmable resistor coupled with a calibration block was designed in order to achieve the desired accuracy.

All of these circuits were implemented in TSMC 40nm CMOS technology.

Acknowledgments

I would like to express my gratitude to:

My family, especially my mother, for all the support they gave me. It was only through their help that I had the opportunity to study and invest in my career.

To my friends who turned my attention away from my thesis when I felt like hitting my head against the walls.

To all the people at Synopsys, especially my friends in the Analog Team: Bruno, Dénis, Hugo, Nuno and Patricio that always gave me good advice and of course, supplied the best tea and cookies.

My supervisor at FEUP, Vítor Grade Tavares for the great opportunity to work with him. He always gave me valuable guidance and support.

To my supervisor at Synopsis, Joaquim Machado that allowed me to developed my thesis in the great environment of Synopsys and opening a few moments in his busy schedule to share some of his knowledge. Without him I wouldn't be able to finish my work.

And finally to FEUP, for giving me some of the best years of my life.

Bruno Teixeira

"However strong you become, never seek to bear everything alone.
If you do, failure is certain."

Itachi Uchiha

Contents

List of Figures

List of Tables

Abbreviations and Symbols

$\beta = u_n C_{ox}$	Beta of the mosfet
C_{ov}	Overlap Capacitance per unit width
C_{ox}	Gate-Oxide Capacitance Per Unit Area
$CTAT$	Complementary to Absolute Temperature
E_{eff}	Average Transverse Electric Field
E_g	Energy Bandgap
DEM	Dynamic Element Matching
$fchop$	Chopping Frequency
G_m	Transconductance
γ	Body-Effect Constant
I_D	Mosfet Drain Current
I_C	BJT Collector Current
I_S	BJT Saturation Current
ϕ_F	Fermi Potential
ϕ_{ms}	Metal-Silicon Work Function
K_b	Boltzmann Constant
L	Channel Length
MC	MonteCarlo
N_A	Acceptor Concentration
N_c	Extrinsic Carrier Concentration
N_D	Donor Concentration
n_i	Intrinsic Carrier Concentration
$PTAT$	Proportional to Absolute Temperature
Q_{ss}	Surface-State Charge Density Per Unit Area
ρ	BJT process dependent temperature constant
$S_{LR,T(nom)}$	Line Regulation at nominal temperature
$T.C.$	Temperature Coefficient
t_{ox}	Gate-Oxide Thickness
μ_{cb}	Impurity Scattering
$u_{n/p}$	Carrier Mobility
μ_{ph}	Lattice Vibration
V_{BE}	Base-Emitter Voltage
V_{bg}	Bandgap reference voltage
V_{DS}	Drain-Source Voltage
V_{DSAT}	Mosfet Saturation Voltage
V_{G0}	Bandgap Voltage of the silicon at 0K
V_{GS}	NMOS Gate-Source Voltage
V_{OV}	Override Voltage
V_{SG}	PMOS Source-Gate Voltage
V_T	Thermal Voltage
V_{TH}	Threshold Voltage
W	Channel Width
z_{out}	Output Pole
ZTC	Zero Temperature Coefficient Point

Chapter 1

Introduction

1.1 Motivation

Most, if not all electrical circuits, use a reference, be it voltage or current. A reference in a circuit establishes a stable point used by other sub-circuits to generated predictable and repeatable results. This reference point should not change significantly under various operating conditions. Temperature is an important parameter which affects the performance of references. Special attention should therefore be paid by the designer to the temperature behavior of the reference.

Current references are used in most of the basic building blocks. Usually, the current in different basic blocks results from mirroring of one or more references. Therefore, it is important that the master current of the system to be PVT independent and designed with the required accuracy.

Voltage references have been used in various fields of application, for example in digital to analog (D/A) converters, the automotive industry and in battery-operated DRAMs. In D/A converters, depending on the digital input signal, the analog voltage is a fraction of the internal reference voltage. As for many applications this digital to analog conversion should not depend on temperature, so the reference voltage has to be temperature-independent. Nowadays, high resolution D/A converters are being used and consequently the reference voltage must be very stable as each variation in the reference voltage is directly sensed in the D/A-converter output.

In the automotive industry, electronic circuits are used to realize larger systems with more functions. However, the automotive environment is very extreme and the temperature variations can be in the range of -40^{o}C to 125^{o}C. Similarly, in battery-operated DRAMs, voltage references are used for power-supply voltage stabilization. In this case, the power consumption is of prime importance.

Bandgap voltage references are the most popular precise references used in various circuits. A bandgap voltage reference (BGR) has high power rejection and its output voltage is very stable against temperature and process variations. It can be implemented using available, vertical or lateral BJTs in any standard CMOS technology [11], [17]. However, when the supply voltage falls below 1 V, the performance of a conventional bandgap reference degrades.

As an alternative, voltage references can also be implemented in MOS technology using the threshold voltage difference [22]. But this solution requires multi-threshold transistors, often recurring to use the threshold voltages of a PMOS and NMOS.

1.2 Specifications and Objectives

Before we begin, it is necessary to understand in what technology and conditions we will be working with, but also the goals to achieve with this work.

This work was built with 40nm CMOS technology. During our work, we will be using high voltage mosfets with minimum dimensions of $W = 0.36um$ and $L = 0.27um$. Bipolar junction transistors will be of PNP-type and have a fixed area of $256um^2$. Process resistors were built with polysilicon material over P-type substrate.

We wish to achieve a precision of $\pm 5\%$ variation over PVT conditions. To reach this objective, we will be resorting to topologies robust to process variations, calibration blocks and dynamic process compensation methods.

Objective	Process	Voltage	Temperature
$\Delta < \pm 5\%$	40nm	1.62V to 3.63V	-40°C to 125°C

1.3 Structure of the Document

This document as the following structure:

- Chapter 2 provides a theoretical background regarding the basic properties of semiconductors, the temperature behavior of several parameters of the mosfets with particular focus on the bipolar transistors. It also includes an introduction to switched capacitors and other resistor architectures. All of these concepts are introduced in this chapter in order to familiarize the reader to some of the specific topics used in this work. This chapter also presents the bibliographical review on voltage and current references. This chapter includes several topologies and an overall comparison between all studied articles.

- In Chapter 3, the steps behind the design of a bandgap voltage reference are presented along with detailed analysis of the behavior of several parameters that constitute the used topology. The results of the simulation are also contained in this chapter.

- Chapter 4 presents the voltage to current converter stage and shows the steps and decisions made in the design of this module. The simulation results along with extensive analysis are also presented.

- Chapter 5 presents dynamic compensation techniques for random process deviations. This chapter provides the steps to designing a dynamically offset compensated amplifier along with detailed analysis of the topology and results obtained once this amplifier has been

integrated in the voltage reference. It also integrates the same techniques on the bandgap and shows the overall results.

- Chapter 6, which is the final chapter, presents the conclusion obtained in this work along with proposals for future improvements in the developed circuits.

Chapter 2

Background and State of the Art

In this chapter, we consider some basics properties of semiconductor materials with particular emphasis on temperature dependence. This will be followed by a focus on temperature dependency of the threshold voltage and carrier mobility of the MOS transistor, and particular focus on the temperature behavior of the bipolar transistor.

This will be followed by a brief description of process corners and MonteCarlo iterations which will be used extensively in this work. Several resistor architectures are also explored since these components will be required in this work as well.

Finalizing, this chapter will contain the current state of the art regarding both voltage and current references and their development through the years.

2.1 Mosfet Fundamentals

In this section we will be reviewing the basic properties of mosfet devices.

- **Energy Bandgap:** The difference in energy between the valence band and the conduction band is called energy bandgap or simply bandgap, and is represented by E_g. E_g is the necessary energy in eV to create an electron/hole and is calculated by [14]

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T+\beta} \tag{2.1}$$

where $E_g(0)$ is the energy bandgap at 0 K, T is the absolute temperature in K, α and β are material constants. For the particular case of silicon, these values are 1.17, 4.73×10^{-4} and 636 respectively, as shown below.

$$E_g(T) = 1.17 - \frac{4.73 \times 10^{-4} T^2}{T+636} \tag{2.2}$$

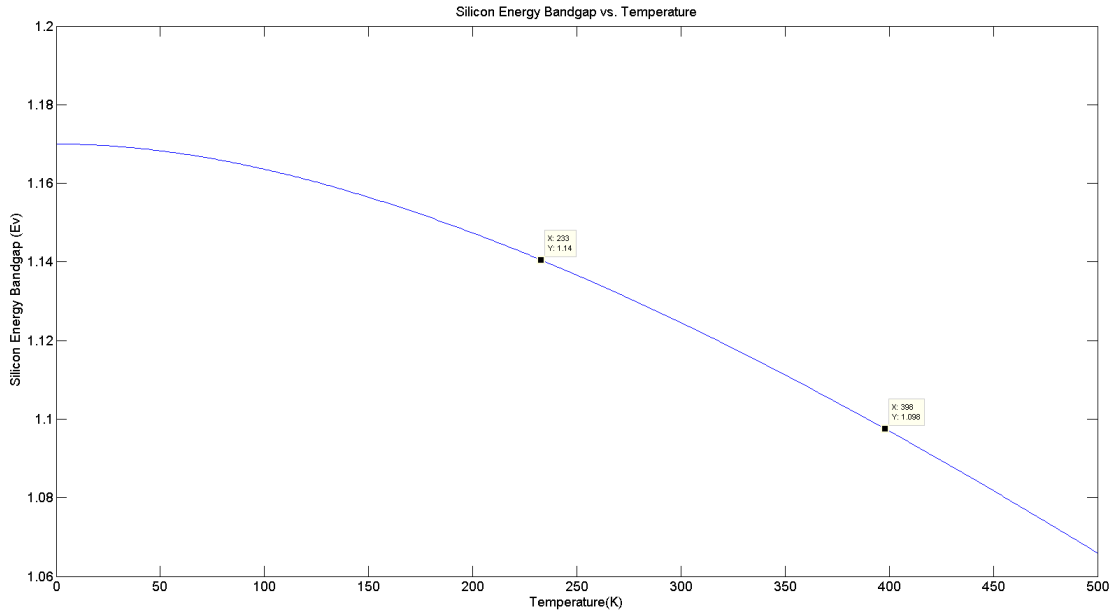


Figure 2.1: Silicon Energy Bandgap vs. Temperature

Figure 2.1 shows the energy bandgap variation with temperature. It can be seen from the plot that the energy value at 0K is roughly 1.17V, since the additional factor is small and can be neglected at this temperature. The two highlighted values correspond to the specific temperature range being considered in this work. As it can be seen, even for a broad temperature range -40°C to 125°C, E_g has a weak temperature dependence.

Intrinsic Carrier Concentration: The intrinsic carrier concentration, n_i, is the concentration of electrons in the conduction band per unit volume, at a given temperature T, in a semiconductor that is completely free of impurities or defects.

This concentration is given by

$$n_i = N_s e^{(-\frac{Eg}{2K_bT})} \tag{2.3}$$

where E_g is the energy bandgap, K_b is the Boltzmann constant and T is the temperature in Kelvin. It is worth noting that the concentration of electrons in the conduction band is equal to the concentration of holes in the valence band. For the particular case of silicon, a commonly accepted value is $9.65 * 10^9 cm^{-3}$ at room temperature (300 K). [1]

Extrinsic Carrier Concentration: The extrinsic carrier concentration, N_c, comes from the introduction of different atoms, called *dopant* atoms, into the previously pure material and can be expressed as

$$N_C = \begin{cases} \frac{N_D - N_A}{2} + \sqrt{(\frac{N_D - N_A}{2})^2 + n_i^2}, & \text{n-type (electron concentration)} \\ \frac{N_A - N_D}{2} + \sqrt{(\frac{N_A - N_D}{2})^2 + n_i^2}, & \text{p-type (hole concentration)} \end{cases} \tag{2.4}$$

where N_D and N_A are the donor and acceptor concentration, respectively. As long as the impurity concentration $|N_A - N_D|$ is much large than n_i, the intrinsic carrier concentration, will be approximately equal to the subtract doping.

Fermi Level: For an intrinsic semiconductor, the Fermi level lies around midway between the valence and conduction bands. For n-type materials, the Fermi level is closer to the conduction bands while for p-type material it is closer to the valence band. The value of the Fermi level depends on temperature due to the temperature dependence of n_i and the thermal voltage, $V_T = \frac{kT}{q}$ and is given by

$$\phi_F = \pm \frac{kT}{q} ln(\frac{n_c}{n_i(T)}) \tag{2.5}$$

where the positive or negative sign refers to the n-type of p-type material, respectively.

2.1.1 Threshold Voltage

A commonly used expression for the threshold voltage of the MOS transistor is given by

$$V_{th} = \phi MS \pm \frac{Q_{ss}}{C_{ox}} + 2\phi_F \pm \gamma(\sqrt{|V_{SB} + 2\phi_F|} - \sqrt{|2\phi_F|}) \tag{2.6}$$

where the positive and negative signal refers to n-channel or p-channel MOS devices, respectively. In this equation, ϕ_{ms} is the metal-silicon work function, Q_{ss} is the surface-state charge density per unit area, C_{ox} is the gate-oxide capacitance per unit area, ϕ_F is the Fermi potential and γ is the body-effect constant that depends on the subtract doping N_s, the gate-oxide thickness t_{ox}, the channel length L and the width W.

The gate-semiconductor work function, ϕ_{MS}, is expressed as

$$\phi_{MS}(T) = \begin{cases} -\frac{kT}{q} ln(\frac{N_s N_p}{n_i^2}), & (NMOS) \\ -\frac{kT}{q} ln(\frac{N_s}{N_p}), & (PMOS) \end{cases} \tag{2.7}$$

where N_p is the carrier concentration in the polysilicon gate. The temperature dependent terms are the intrinsic carrier concentration, n_i, and the thermal voltage, $\frac{kT}{q}$.

From a first look at the threshold voltage and taking into account both the Fermi level equation (2.5) and the ϕ_{MS} work function, we can see that the bigger contributors for the threshold voltage variation with temperature are the Fermi level, ϕ_F, and the gate-semiconductor work function, ϕ_{MS}.

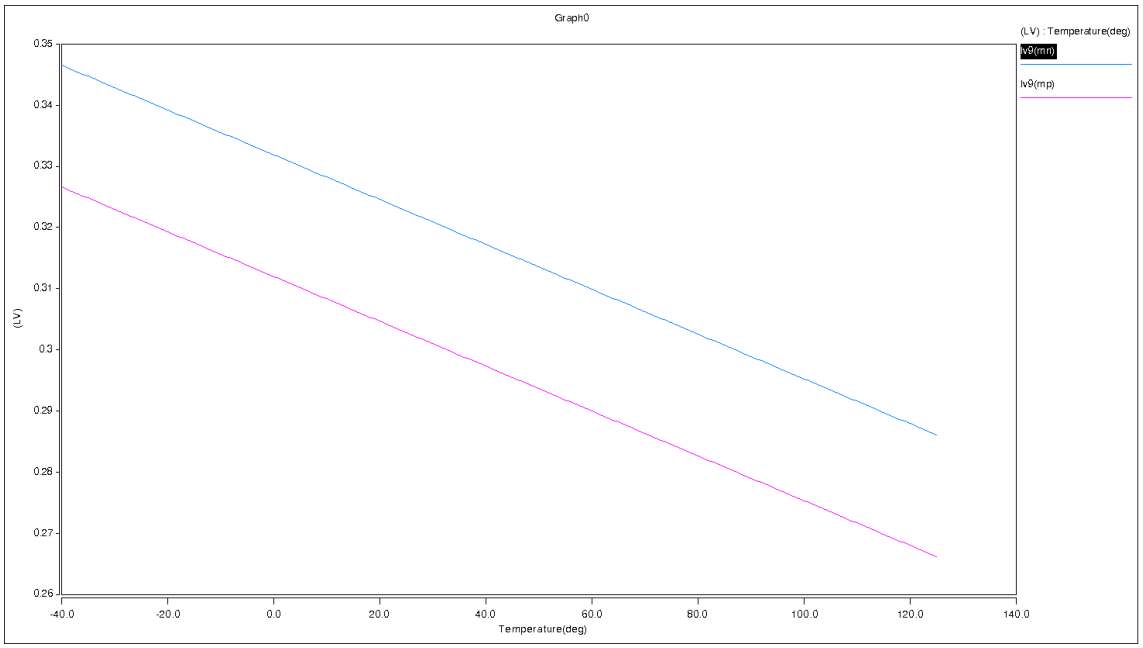


Figure 2.2: Absolute value of the threshold voltage variation of a typical PMOS and NMOS transistor (W/L=1u/1u). FreePDK 45nm CMOS Technology

It can be seen from figure 2.2 that the threshold voltage decreases almost linearly with temperature in both cases. Although this non-linearity is fairly small, and usually ignored in the first order approximations, is still noticeable. A careful analysis of the difference between both thresholds voltages (NMOS and PMOS) shows a PTAT behavior. There are in fact voltage and current references based on this very same behavior. [22] [18]

2.1.2 Carrier Mobility

We pay particular attention to the temperature dependence of the mobility as it is one of two main factors (the other is threshold voltage) in the temperature behavior of a mosfet. The carrier mobility, $\mu(cm^2/Vs)$, describes the drift velocity of a particle in an applied electrical field. The carrier mobility has a very complex temperature dependence, defined by the interplay of several parameters. The two major contributors are the lattice vibration, μ_{ph}, and impurity scattering, μ_{cb}. [7] [21]

The temperature behavior of the lattice vibration is given by

$$\frac{1}{\mu_{ph}} \propto T^{\frac{3}{2}} E_{eff}^{\frac{1}{3}} \tag{2.8}$$

and the impurity scattering is

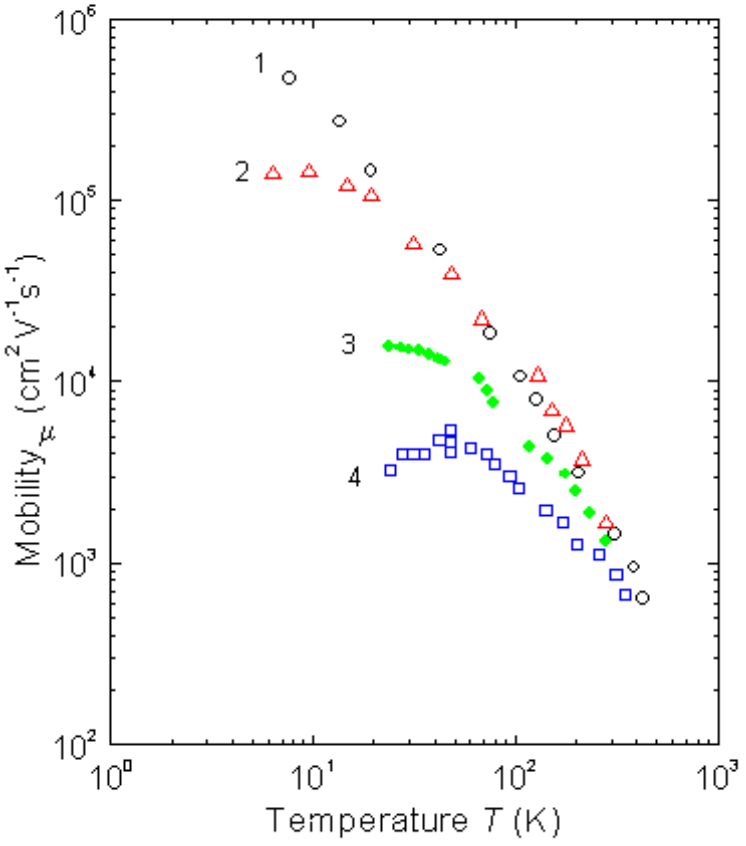
$$\frac{1}{\mu_{cb}} \propto T^{-1} E_{eff}^{-2} \tag{2.9}$$

where T is the absolute temperature in Kelvin and E_{eff} is the average transverse electric field.

Lattice vibration results from thermal vibrations of the atoms at any temperature higher than absolute zero (0 K). As the temperature increases, the carrier move faster and thus this factor dominates. Thus, at high temperatures, the mobility, μ, will be proportional to $T^{\frac{3}{2}}$. On the other side, at lower temperatures the atoms have low kinetic energy and thus impurity scattering dominates.

Based on 2.8 and 2.9 we can finally write the temperature behavior of the carrier mobility as the sum of both factors.

$$\frac{1}{\mu} \propto \frac{1}{\mu_{cb}} + \frac{1}{\mu_{ph}} \tag{2.10}$$



. http://www.ioffe.ru/SVA/NSM/Semicond/Si/electric.html#Hall

Figure 2.3: Electron mobility versus temperature for different doping levels. High purity to increased doping levels respectively (reprinted from the link above)

From figure 2.3, it is clearly visible the previous assumptions. We have four different doping levels represented. From high purity (1,2) to increased doping levels (3,4). They all converge to the

same value at high temperatures, proving that lattice vibration dominates at higher temperatures. Please note that the temperature behavior appears linear due to the use of a logarithmic scale.

2.1.3 Zero Temperature Coefficient Point

The Zero Temperature Coefficient Point or simply ZTC point is a certain voltage level that when applied to the gate of a MOS device will internally nullify temperature variation. Both the threshold voltage and the carrier mobility of a mosfet are dependent on temperature, as seen in previous sections [2]. For carefully sized MOS transistors with channel doping concentration on the vicinity of 10^{15} the output current becomes temperature independent in the ZTC point.

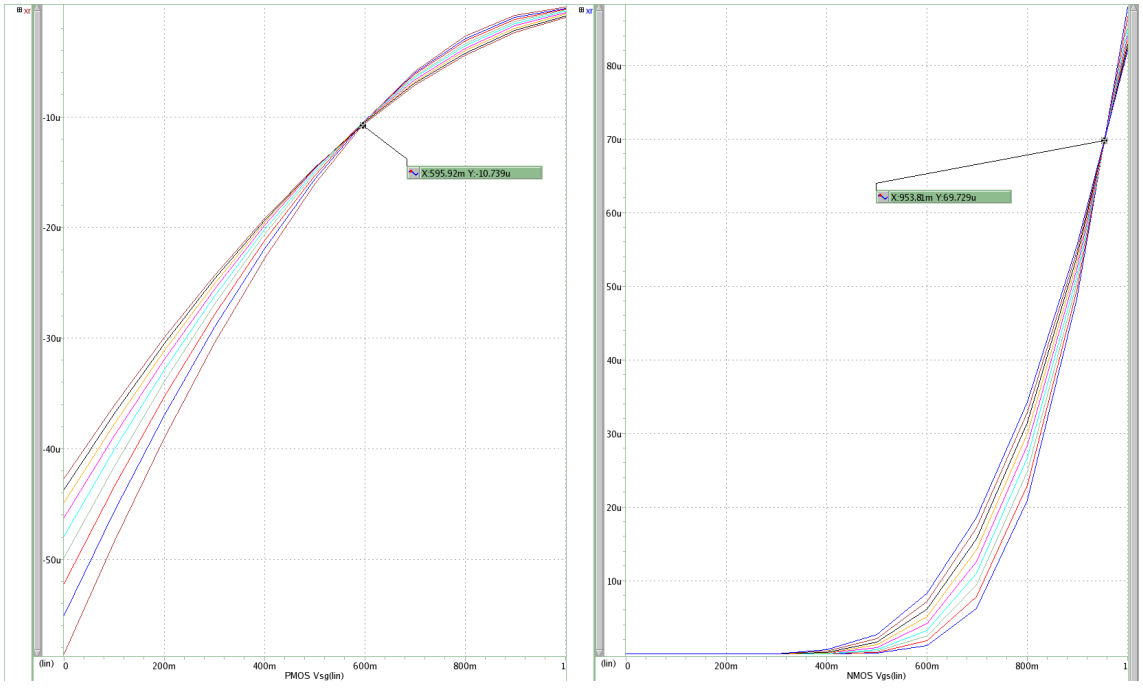


Figure 2.4: ZTC bias point for both PMOS and NMOS respectively in typical conditions. I_D vs V_{GS} at different temperatures

As we can see in the figure above, by biasing the gate voltage of a mosfet at approximately 0.605 V, the device will provide a current that is invariant to temperature shifts. This type of compensation works similarly to the PTAT and CTAT sum since with precise sized mosfets, the internal temperature dependences cancel each other out.

The mosfet current is modeled by

$$I_d = \frac{u_n C_{ox}}{2} \frac{W}{L} (V_{GS} - V_{TH})^2 \tag{2.11}$$

In equation 2.11, the carrier mobility, μ, and the threshold voltage, V_{TH}, are the major temperature dependent parameters. Both of these quantities diminish as the temperature goes up. However, they have an opposite effect on the drain current. As the mobility goes down, so does

the current. But, as the threshold voltage goes down, the drain current goes up due to the $V_{GS}-V_{TH}$ difference.

The ZTC bias point is a certain voltage level in which the changes in electron mobility and threshold voltage compensate each other. A transistor biased at this point will have minimal variation in its saturation current over temperature.

As it is visible in figure 2.4, a PMOS and NMOS biased at approximately 595mV and 953mV respectively, will be fairly independent of temperature due to mutual compensation.

2.2 BJT Fundamentals

In this section we will be reviewing the basic properties of BJT devices.

2.2.1 Base-Emitter Voltage

In voltage references, the BJT (Bipolar Junction Transistor) is usually connected in the diode configuration (i.e., the base terminal and the collector terminal are tied together), such that the Base-Emitter Voltage, V_{BE}, is used to provide a fixed junction voltage. However, the junction voltage is temperature dependent and cannot be used as a voltage reference by itself.

Neglecting the Early effect, the collector current density of a NPN transistor biased in the forward active region is given by,

$$J_C(T)A_E = J_S(T)A_E e^{(\frac{V_{BE}}{V_T})} \tag{2.12}$$

where A_E is the emitter area, J_S is the saturation current density, T is the absolute temperature in Kelvin, J_C is the current density and V_T is the thermal voltage, $V_T = \frac{kT}{q}$. Since the emitter area is constant we usually consider the collector current as,

$$I_C(T) = I_S(T)e^{(\frac{V_{BE}}{V_T})} \tag{2.13}$$

where I_C is the collector current and I_S is the saturation current.

Without going into further details of semiconductor physics we quote the base-emitter voltage from [20].

$$V_{BE}(T) = V_{G0}(1-\frac{T}{T_0}) + V_{BE}(T_0) - \frac{\rho kT}{q} ln(\frac{T}{T_0}) + \frac{kT}{q} ln(\frac{J_C(T)}{J_C(T_0)}) \tag{2.14}$$

where, V_{G0} is the bandgap voltage of the silicon at 0K, which is around 1.17V but subjected to the technology in use, V_{BE} is the bandgap voltage at temperature T_0 and ρ is a process dependent temperature constant.

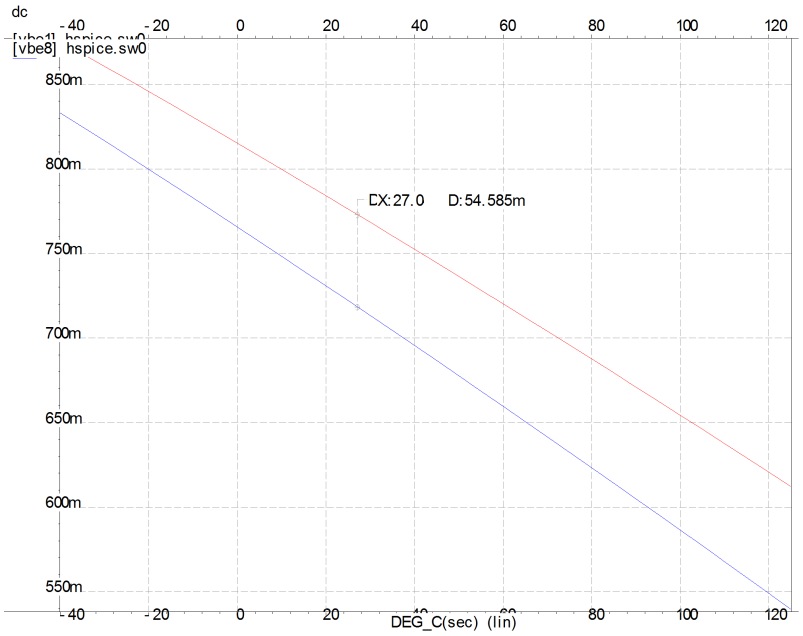


Figure 2.5: Base-Emitter voltage temperature variation of diode connected BJT biased at 5u. $256um^2$ BJT and $8x256um^2$ BJT

Figure 2.5 shows the behavior of the base-emitter voltage of two diode connected BJT biased at the same current but with different areas. It stands to reason that the BJT with higher area has a lower V_{BE}. Looking at the last term of 2.14, $\frac{kT}{q}ln(\frac{J_C(T)}{J_C(T_0)}$, the current density is the current divided by the area, and since both BJT are biased with the same current, the BJT with higher area would produce a smaller base-emitter voltage.

We can clearly see that the temperature variation is not linear since the difference between both BJT clearly increases. However, in the first order approximation we consider this voltage linear. To the process in cause, the slope is approximately -1.5mV/^{o}C.

Such temperature characteristic is called Complementary to Absolute Temperature (CTAT), where the rate of change of base-emitter voltage against the temperature is negative.

Differential Base-Emitter Voltage: This voltage is heavily used in bandgap voltage references. The key factor to understand is that while V_{BE} goes down with temperature (CTAT), the difference between both V_{BE} seems to increase with temperature. We call this characteristic, Proportional To Absolute Temperature (PTAT), since the rate increases with temperature.

Let's look at equation 2.14 and assume that both BJT are biased with same current but the second BJT has a higher emitter area. We will call this area difference N, such that $A_{E2} = NA_{E1}$

$$\begin{aligned} V_{BE1}(T) - V_{BE2}(T) = V_{G0}(1-\frac{T}{T_0}) + V_{BE}(T_0) - \frac{\rho kT}{q}ln(\frac{T}{T_0}) + \frac{kT}{q}ln(\frac{J_{C1}(T)}{J_{C1}(T_0)}) - \\ [V_{G0}(1-\frac{T}{T_0}) + V_{BE}(T_0) - \frac{\rho kT}{q}ln(\frac{T}{T_0}) + \frac{kT}{q}ln(\frac{J_{C2}(T)}{J_{C2}(T_0)})] \end{aligned} \tag{2.15}$$

Looking at 2.15 we can see that the only difference is the last term because the second BJT has a higher area than the first. If both BJT have the same current but one of them has higher area, it stands to reason that the bigger one will have smaller current density. Taking this into account we can narrow this difference to,

$$V_{BE1}(T) - V_{BE2}(T) = \frac{kT}{q} ln(\frac{J_{C1}(T)}{J_C(T_0)}) - \frac{kT}{q} ln(\frac{J_{C2}(T)}{J_C(T_0)}) \tag{2.16}$$

Now changing the current density into current using the BJT area we get,

$$V_{BE1}(T) - V_{BE2}(T) = \frac{kT}{q} [ln(\frac{I_{C1}}{A_{E1}}) - ln(\frac{I_{C2}}{A_{E2}})] \tag{2.17}$$

But both I_{C1} and I_{C2} are equal. Thus we get,

$$V_{BE1}(T) - V_{BE2}(T) = \frac{kT}{q} ln(\frac{A_{E2}}{A_{E1}}) = \frac{kT}{q} ln(N) \tag{2.18}$$

As we can see from 2.18, the difference between the V_{BE} of two BJT with different areas but biased with equal currents is clearly PTAT. If we take the derivate of 2.18 against temperature,

$$\frac{\delta \Delta V_{BE}}{\delta T} = \frac{k}{q} = \frac{1.3806 * 10^{-}23}{1.602 * 10^{-}19} \approx 0.09 mV/^{o}C \tag{2.19}$$

where k is the Boltzman constant and q is electrical charge on the electron. It we look back at 2.5 we see the highlighted difference between both V_{BE} at 27°C, which equals 54.585 mV.

Taking equation 2.18 and calculating with with $N = 8$ and $T = 300K(27^{o}C)$ we get,

$$V_{PTAT} = \frac{1.3806 * 10^{-}23}{1.602 * 10^{-}19} ln(8) \approx 53.76 mV \tag{2.20}$$

This theoretical value is very close to the simulated one.

The main idea behind bandgap voltage references is adding these two voltages in order to eliminate the temperature dependence. We will be seeing this in the next chapter.

2.3 Resistor Architectures

Either we wish or not, this type of circuits will always need some type of resistor to produce the desired current. There are three solutions for this problem. The first solution is using a MOSFET in ohmic mode, thus the name of this region. A MOSFET can operate as variable resistor although there are several drawbacks. The linearity is poor and is dependent of the biasing voltage.

If $V_{GS} > V_{TH}$ and $V_{DS} < V_{GS} - V_{TH}$, and the mosfet will operate in triode region, also called ohmic region. The resistance of the channel is modeled by (assuming $V_{DS} << V_{GS} - V_{TH}$),

$$r_{DS} = \frac{1}{\mu_n C_{ox}(\frac{W}{L})(V_{GS} - V_{TH})} \tag{2.21}$$

The second solution is simply using a physical resistor. There are a few disadvantages with using a resistor in high level integrations. Take for example a supply voltage of 1.25 V, like the standard bandgap reference, and you wish an output current of 1uA. This would require a $1.25M\Omega$ which would take a lot of area in the die. Not to mention the resistors are non-linear and the tolerances are large.

Process resistors have the advantage of not requiring any voltage level to create the resistance unlike mos resistors. Therefore, creating precision resistors using triode mosfet devices is more difficult.

Figure 2.6: Resistor variation versus temperature over typical, slow and fast corner.

Looking at figure 2.6, we see that the temperature variation of the resistor is fairly linear, only changing 0.8% over 165 degrees. However, the resistance varies a lot over process corners. The figure shows a $15k\Omega$ typical resistor at all three corners. It can be seen that at typical conditions the resistor maintains the desired value. However, there is nearly a 25% difference between the typical and fast corner.

2.3.1 Switched-Capacitor Resistor

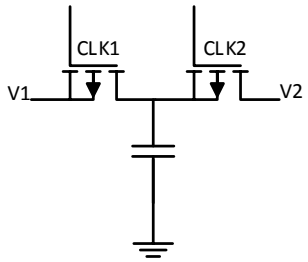


Figure 2.7: Switched Capacitor

The other option, and the most obvious solution to large resistors, is the use a switched-capacitor resistor, shown in figure 2.7. Instead of using a physical resistor, we can "simulate" one by controlling the charge and discharge of a capacitor. This circuit simulates the behavior of a normal resistor, with a few added benefits. [11] [4] [9]

The capacitor ratios can be tightly controlled, providing a more stable current overall, and by controlling the signal going into $\phi 1$ and $\phi 2$, we can control the value of the resistor and thus, change the current output.

Although this is a viable solution to higher levels of integration, it also brings a few more parameters to take into account during the development, like clock feedthrough and charge injection, not to mention the necessity of an external signal to clock the capacitors.

$$R_{eq} = \frac{1}{C_1 f} \tag{2.22}$$

The equivalent resistor value of the figure 2.7 is given by the inverse of capacitance times the switching frequency. If we want a $1.25M\Omega$, we would simply need a 10MHz clock and a capacitor with 80fF. Both MOSFETs used in the previous schematic are used as switches to control the charge and discharge of the capacitor. MOSFETs are considered good switches because of two main reasons:

• Off-resistance near $G\Omega$ range (subjected to technology). At lower nodes leakage increases so it must be taken into account.

• On-resistance in 100Ω to $5k\Omega$ range, depending on transistor sizing.

These type of circuits requires non-overlapping signal to both switches, requiring another component to the system. Figure 2.8 shows such component. Both $\phi 1$ and $\phi 2$ need to be non-overlapping to properly control that charge and discharge of the capacitor.

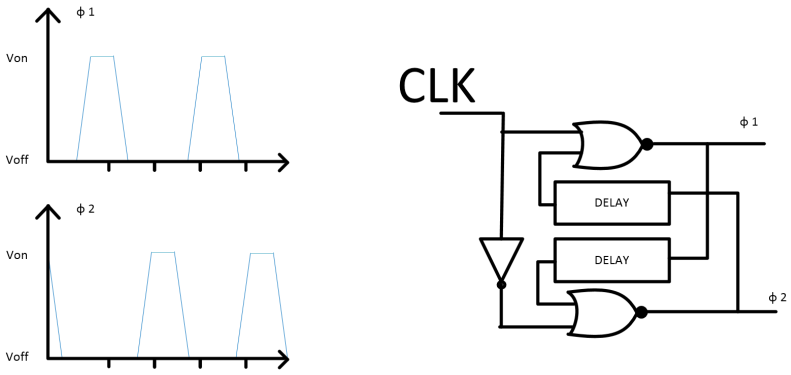


Figure 2.8: Non-overlapping clock generator

Clock Feedthrough

One of the main disadvantage of this resistor architecture are the parasitic capacitances between the gate-drain and gate-source. Depicted in figure 2.9, the effect introduces an error in the samples output voltage.

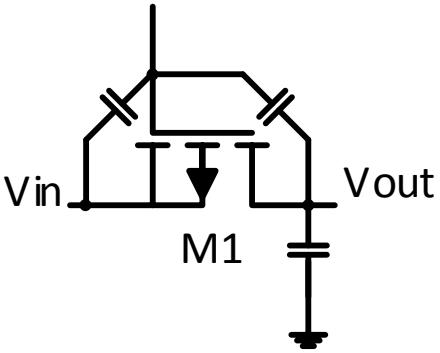


Figure 2.9: Parasitic Capacitances of Mosfet

Assuming that the overlap capacitance is constant, we have an error of,

$$\Delta V = V_{ck}\frac{WC_{ov}}{WC_{ov}+C_h} \tag{2.23}$$

where C_{ov} is the overlap capacitance per unit width. The error ΔV is independent of the input level, manifesting itself as a constant offset in the input/output characteristic.

Channel Charge Injection

Another problem we face with this architecture is the charge injection when the MOSFET is "turned off".

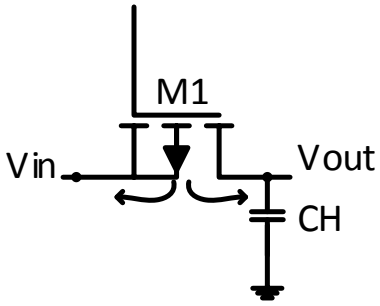


Figure 2.10: Channel Charge Injection

As depicted in figure 2.10, the charge injected to the left side is absorbed into the input source, creating no error. On the other hand, the charge injected to the right side will be deposited on Ch, introducing an error voltage stored on the capacitor. This error is modelled approximately by

$$\Delta V = \frac{WLC_{ox}(V_{dd} - V_{in} - V_{TH})}{2C_h} \tag{2.24}$$

A practical way to solve this problem would be using low input voltages, slow rising and falling edge switching, and larger capacitors. Another possible answer would be using complementary switches instead of a single MOS device to decrease the resistance.

Thermal Noise

Since one of main goals of this work is to minimize PVT variations we also need to take into account the thermal noise. The on-resistance of MOSFET switch will introduce thermal noise at the output and, when switch turns off, this noise is stored on the capacitor. This voltage is given by equation 2.25 and can be controlled by employing larger capacitors.

$$Noise = \sqrt{\frac{kT}{C}} \tag{2.25}$$

Like the previous effects, larger capacitors would require smaller frequencies, so it would be a trade-off between speed and precision.

2.4 Process Corners and MonteCarlo

In semiconductor manufacturing, a process corner is an example of a design-of-experiments technique that refers to a variation of fabrication parameters used in applying an integrated circuit design to a semiconductor wafer. Process corners represent the absolute extremes of these parameter variations within which a circuit that has been fabricated on the wafer will achieve. A circuit running on devices fabricated at these process corners may run slower or faster than specified and at lower or higher temperatures and voltages.

In our case, we will have the typical, fast and slow corners for each device in use, mosfets, BJT and resistors. This gives a total of 5x3x3 = 45 corner combination in our simulations. One thing to note is that the worst case corners neglect the within-die fluctuations of the individual devices. The corner variation will apply equally to all devices.

Even in the absence of any process variations, there are limits on the accuracy with which devices can be fabricated. MonteCarlo simulations on the other hand, account for the individual shift in the parameters of the devices. These simulations are random in nature and tend to model the variations and mismatch between individual devices.

Take for example the threshold voltage of a PMOS transistor. In the slow corner case, every single transistor in the circuit would have the same threshold voltage behavior, although they would all be in the slow corner. With MonteCarlo simulations, each device has its own behavior.

The combinations of worst case corners and MonteCarlo provides good models for process variations and the within-die variations and mismatch of process parameters.

2.5 State of the Art

In this section, we will be presenting the state of the art solutions to the problems explained in the introduction. Some of the solutions here exposed will sometimes sacrifice a precision in one element in favor of another. A high precision circuit like this should always have in mind the system in which it will be integrated.

2.5.1 Traditional Bandgap Voltage Reference

We begin our research with the simplest and one of the oldest model of BGR (Bandgap Reference), depicted in the figure below. It is worth noting that while this design was released a long time ago it still is one of the most used today or, at least, the general idea behind it. [3]

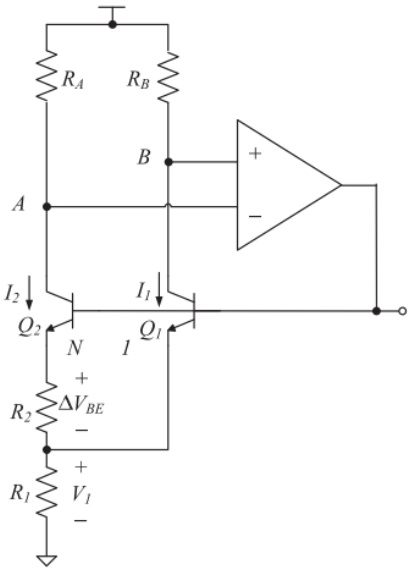


Figure 2.11: Simplified Brokaw bandgap voltage reference circuit (Brokaw, 1974)

This architecture uses the concept explained in section 2.2.1. To compensate the temperature it extracts the CTAT voltage from V_{BE} of the bipolar transistor, and the PTAT from the ΔV_{BE} of both BJTs. Thus, by adding these two voltages we arrived at a voltage level with very low dependency on temperature. The operational amplifier forces the voltages levels at V_A and V_B to be equal. If both resistors R_A and R_B are matched then the same current flows on both branches of the schematic. Since R_1 is the sum of two identical currents, we can write:

$$V_1 = 2IR_1 <=> V_1 = 2R_1 \frac{\Delta V_{BE1,2}}{R_2} \tag{2.26}$$

Since $V_{REF} = V_{BE1} + V_1$, thus:

$$V_{REF} = V_{BE1} + (\frac{2R_1 ln(N)}{R_2})V_T \tag{2.27}$$

Where N is the ratio between the bipolar transistors and V_T is the thermal voltage. By adjusting the relation between R1 and R2 but also the areas of both BJT we can adjust the slope of the PTAT voltage to better eliminate temperature dependency. This design is also highly insensitive to supply changes because we are extracting a physical parameter of the device.

2.5.2 MOS Biased Bandgap Voltage Reference

Another familiar model is presented in [19]. It replaced the bias resistors and amplifier with a self-biased current mirror as it is shown below. It works similar to that of figure 2.11 with the exception that the feedback loop was replaced by the transistors M_{P1}, M_{P2}, M_{N1} and M_{N2}.

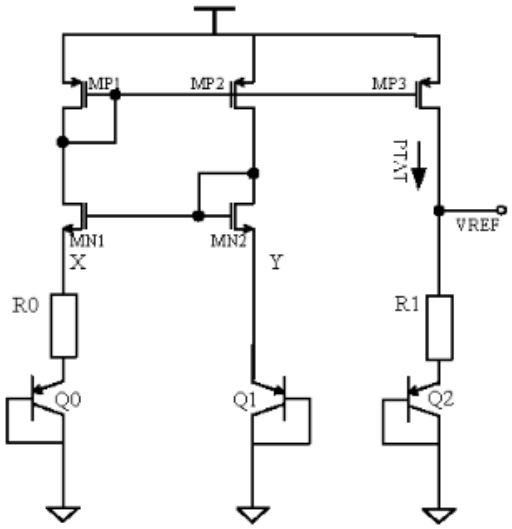


Figure 2.12: Self Biased BGR Architecture (From [19])

Since M_{P1} and M_{P2} have the same V_{GS}, by selecting $SP1 = SP2$, we shall obtain $ID_{MP1} = ID_{MP2}$. As a result, the currents flowing through M_{N1} and M_{N2} should be the same. If we do not consider the channel modulation effect, the drain voltages of M_{P1} and M_{P2} will be the same. Since the gates of M_{N1} and M_{N2} are connected together, by selecting $SM_{N1} = SM_{N2}$, we shall obtain $ID_{MP1} = ID_{MP2}$. As a result, $I_X = I_Y$ which yields,

$$V_{EB1} = V_{EB0} + R_0 I_X \tag{2.28}$$

$$I_X = \frac{V_{EB1} - V_{EB0}}{R_0} = \frac{\Delta V_{EB1,0}}{R_0} \tag{2.29}$$

Since, $I_X = I_{Q0}$ and $I_Y = I_{Q1}$, and the emitter area ratio between both BJT transistors is N then,

$$I_X = I_Y = \frac{V_T ln(N)}{R_0} \tag{2.30}$$

Finally, since I_{PTAT} = Ix we can write

$$V_{REF} = I_X R_1 + V_{EB2} \tag{2.31}$$

As we go into higher integration, the self-biased mirror would suffer from channel modulation, thus highly increasing supply dependency.

2.5.3 Low-Voltage Voltage Reference

A slight improvement over both the previous references is the one presented here. This improvement comes in the way of lowering the supply voltage necessary to produce the PVT voltage. Of course this also lowers the bandgap voltage itself to around half of the theoretical value. [17]

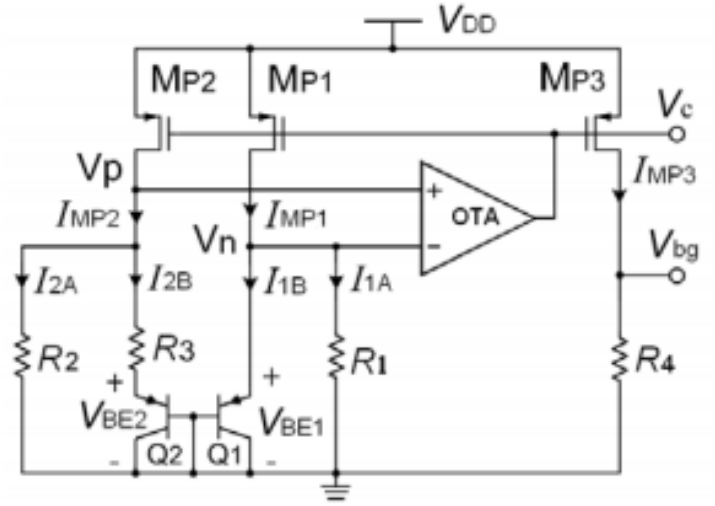


Figure 2.13: Low-Voltage Voltage Reference (From [17])

The output of this bandgap reference is 611.9 mV with a variation of 0.8mv. Using lower voltages in higher levels of integration will slightly lower the short channel effects of the MOS devices but also increase devices longevity by lowering their stress limit.

This improvement comes from folding the conventional BGR circuit to ground on nodes V_p and V_n. If resistors R1 and R2 are matched, then the same current will flow from both branches, $I_{2A} = I_{1A}$. The operational amplifier also clamps both V_p and V_n to equal voltage levels. Since,

$$V_p = V_n = V_{BE1}, V_{r3} = V_{BE1} - V_{BE2}, I_{MP2} = I_{2A} + I_{2B} \tag{2.32}$$

$$I_{MP2} = \frac{V_{BE1}}{R_2} + \frac{\Delta V_{BE}}{R_3} = \frac{V_{BE1}}{R_2} + \frac{V_T ln(N)}{R_3} \tag{2.33}$$

Thus we can write,

$$Vbg = m(\frac{V_{BE1}}{R_1} + \frac{V_T ln(N)}{R_3})R_4 \tag{2.34}$$

Where m is ratio between M_{P2} and M_{P3}. By properly adjusting the values of the resistors and BJT areas we can achieve 611.9 mV with an accuracy of 0.3%.

2.5.4 Current Reference Architecture

A current architecture to generate a current reference is shown below. This architecture refers only to the conversion between a voltage level and its current supply. This topology requires a bandgap reference and works by controlling a certain dump of energy in node 1. [12]

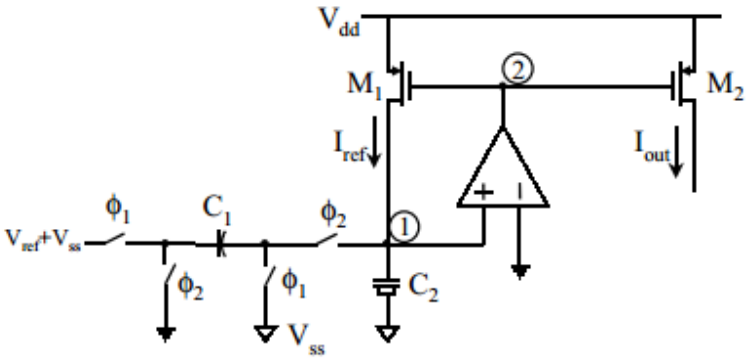


Figure 2.14: Current Reference Architecture (From [12])

During $\phi 1$, C_1 charges to V_{REF}. During $\phi 2$ this charge is dumped on node 1. Thus, the ripple voltage is

$$\Delta V = -(\frac{2C_1}{C_1 + 2C_2})V_{REF} \tag{2.35}$$

The ripple voltage can be lowered by increasing the value of C_2. However, a larger C_2 reduces the ripple at the expense of increased die area. The current delivered by the switched-capacitor is approximately

$$I_{REF} = C_1 V_{REF} F_{clk} \tag{2.36}$$

This architecture offers an accuracy of 0.029%. Further improvement to this topology can be made by placing an amplifier connected in M_2 to increase the output resistance. Another improvement can be achieved through the addition of a filter between the amplifier output, node 2, and the gate of M_2.

2.5.5 Switched-Capacitor Current Reference

A realization of a current source using the bandgap reference presented in section 2.13 and a modification of 2.14 is shown below. One of the latest available types of current references with low variation over PVT parameters uses the above described components. According to [11], a low dependence PVT circuit is presented. It was designed using 0.13um CMOS technology and provides an overall precision of $\pm 3\%$ over PVT conditions.

One factor that we must take into account when using switched-capacitors is the frequency response of the system. This model has an 8us settling time, meaning that once this circuit was powered, its output is only reliable after the settling time has been reached.

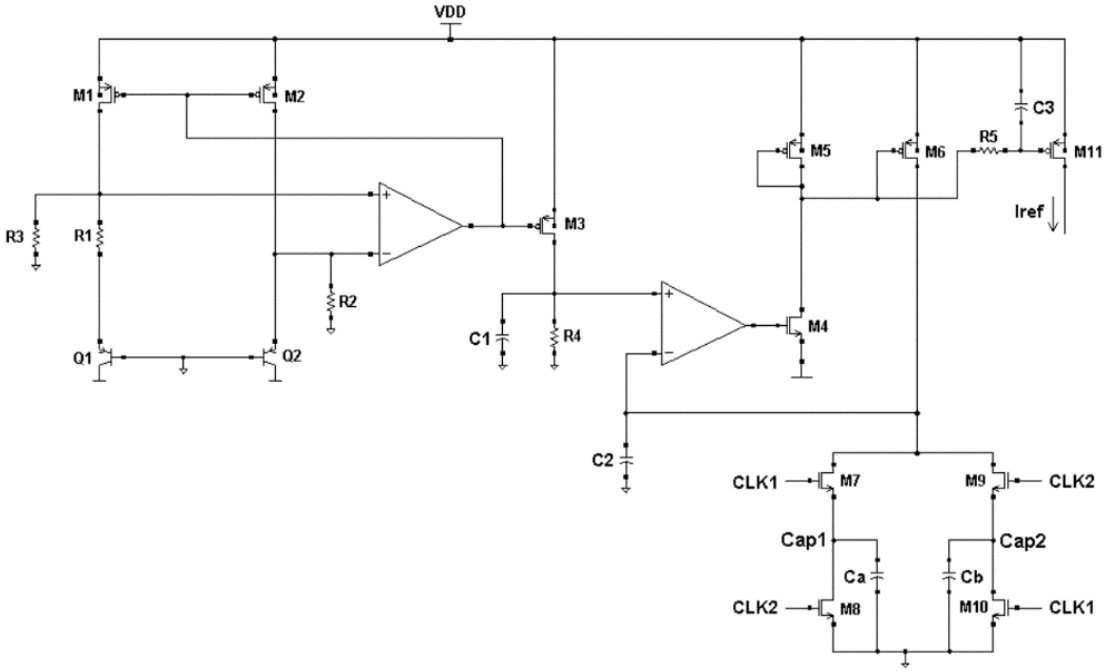


Figure 2.15: Switched-Capacitor Current Reference Schematic ((From [11]))

This architecture is depicted in figure 2.15 and we can easily identify the three general blocks that compose this system. At the top left we have the traditional bandgap voltage reference to provide PVT invariant voltage, this version is the low voltage topology discussed in section 2.13. This reference voltage is fed into a voltage to current converter on the top right. The final piece of this schematic is the switched-capacitor resistor on the bottom right. This resistor is controlled by the CLK1 and CLK2 signals used to adjust the resistor value and output current. This current reference uses the PTAT and CTAT currents from the bipolar transistors, generating a voltage insensitive to both temperature and supply voltage.

It is worth noting that the current source here applied uses a filter between M_6 and M_{11} to reduce the ripple voltage.

$$I_{REF} = 2C_a V_{bg} F_{ref} \tag{2.37}$$

The output current is given by the previous equation in which, V_{bg} is the bandgap voltage, C_a is the value of the capacitor and F_{ref} is the working frequency. The factor 2 comes from the simulated parallel resistors of C_a and C_b. If both capacitors have the same value, then the resistance is cut in half. This topology has a 3% variation over PVT. This error is mainly dependent of the offset voltage from the operational amplifiers and the ripple voltage.

2.5.6 All MOSFET, Two Resistor Current Reference

There have been many proposed current references through the years, and as we go into higher integrations, the resistors do not scale as well as MOSFETs. The next logical step in current reference construction was the idea of trying to eliminate virtually all resistors, resorting only to transistors.

To minimize production costs, this architecture uses no BJTs, external components or trimming procedures. This circuit was designed for 22nm technology and simulations results show a PVT tolerance of around 10%. The main advantages of this circuit is the low voltage operation, suitable for sub-micron applications due to reduced electrical stress limit. [13]

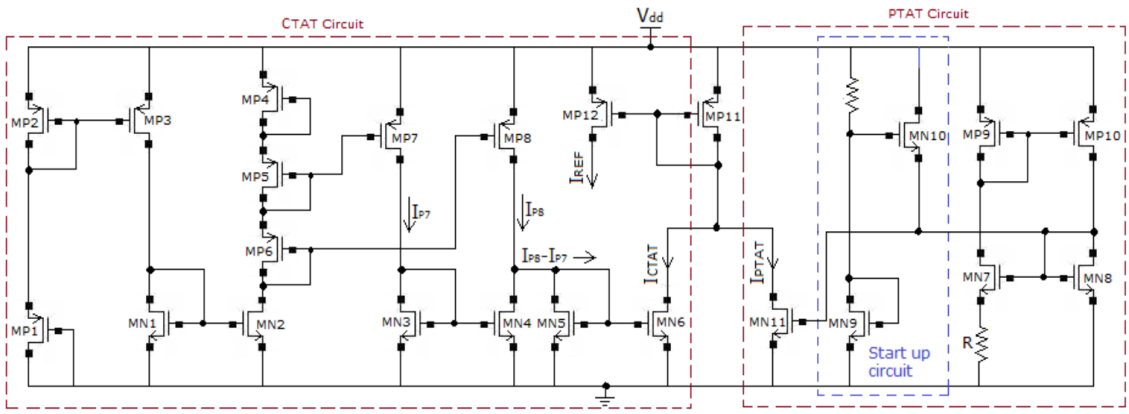


Figure 2.16: All MOSFET, Two Resistor Current Reference Schematic (From [13])

The architecture is shown in figure 2.16. In the schematic, we can see some of the basic modules like the PTAT circuit whose current is generated by a Beta multiplier reference with the exception of MN7 and MN8 being in weak-inversion and the startup circuit necessary for setting the operating point.

Since transistors MN7 and MN8 are working in weak-inversion, their current is governed by a different equation from before. In weak-inversion the current is given by,

$$I_{n8} = I_0 W_{n8}/L_{n8} e^{\frac{V_{gs8}-V_{th}}{nV_T}} \tag{2.38}$$

Using N8 = AN7, and Vgs8 = Vr+Vgs7 we can write the current in MN8 as,

$$I_{n8} = I_0 W_{n7}/L_{n7} e^{\frac{V_{gs7}-V_{th}}{nV_T}} <=> I_{n8} = A e^{\frac{V_r}{nV_T}} \tag{2.39}$$

$$V_r = nV_T(\frac{S_{n7}}{S_{n8}} * \frac{S_{p10}}{S_{p9}}) \tag{2.40}$$

Solving for Vr, we arrive at voltage value that is clearly PTAT and very similar to the differential base emitter voltage of two BJT. The slope of the voltage can be adjusted using the size ratios of the four transistors.

The CTAT current generated to eliminate the temperature dependency is the difference of two others. To eliminate the process dependency, the MOSFET seven and eight were properly scaled to arrive at a theoretical 0% process dependency.

This topology is viable to high integrations as this architecture is technology independent and allows integration into 22nm and beyond. However, it suffers from low accuracy by working in sub-threshold region of operation. In this region temperature dependence suffers from high non-linearity and thus reports only an accuracy of 10% over PVT.

A suggestion to be analyzed further in order to improve this circuit would be replacing the resistor R by a switched-capacitor resistor, lowering its variation and providing a more accurate PTAT current.

2.5.7 Resistorless Current Reference

The last step in current reference architecture would be a Resistorless current reference. A resistorless circuit allows for higher integrations while simplifying the circuit layout and area consumption. A resistorless current reference source is depicted in figure 2.17. [8]

This circuit uses cascode structures to improve the power supply rejection ratio. The reference current source has been designed in 65 nm technology. The presented circuit achieves 55 ppm/^{o}C temperature coefficient over range of -40 ^{o}C to 125 ^{o}C. Reference current susceptibility to process parameters variation is $\pm$3%. The power supply rejection ratio without any filtering capacitor at 100 Hz and 10 MHz is lower than -127 dB and -103 dB, respectively. The current reference source presented here consists of MOSFETs and vertical PNP bipolar transistors. It is designed for 3.3 V supply voltage with low sensitivity to process variation and small temperature coefficient.

Temperature independence is achieved by obtaining appropriate temperature coefficients of the summed currents I1, the PTAT current, and I2, the CTAT current. Low sensitivity to supply voltage bases on using cascode structures. Transistors MB1-MB13 form the bias block, which provides the bias voltage to cascode structures and bulk node of M8. Moreover, transistors MB9-MB13 work as start-up circuit.

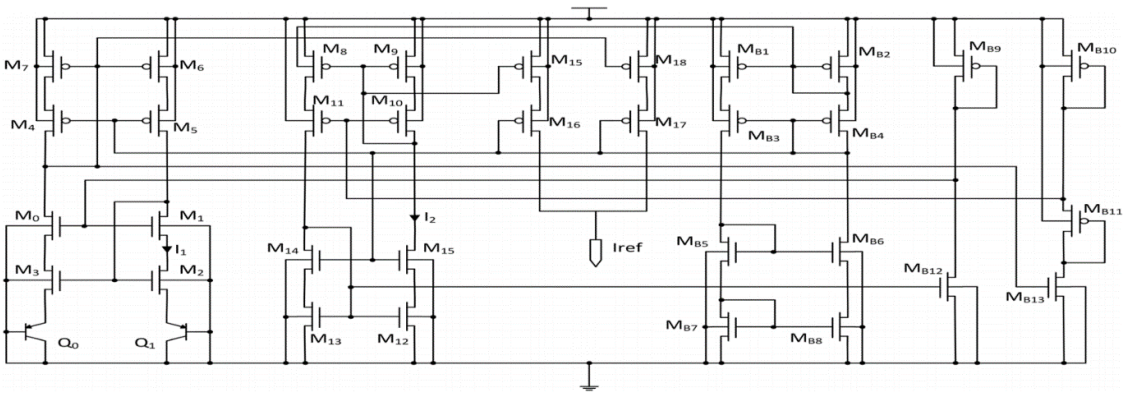


Figure 2.17: Resistorless Current Reference (From [8])

That PTAT current is achieved using the difference between the base emitter voltages of Q0 and Q1. This current can be easily adjusted be changing the sizes of both M3 and M2 but also the areas of both BJTs. This current is given by,

$$I_{ptat} = \frac{uC_{ox}}{2} * \frac{(\frac{kT}{q} ln(\frac{A_0}{A_1}))^2}{(\sqrt{\frac{L_3}{W_3}} - \sqrt{\frac{L_2}{W_2}})^2} \tag{2.41}$$

The CTAT current is formed by transistor M8-M15 and is based on threshold voltage difference. The difference is obtained by using the body effect.

$$I_{d8,9} = \frac{uC_{ox}}{2} * \frac{W_{8,9}}{L_{8,9}} (V_{gs} - V_{th})^2 \tag{2.42}$$

Since V_{gs}=$V_{gs}8$=$V_{gs}9$, and solving $V_{gs}9$ and substituting in $V_{gs}8$ we arrive at:

$$I_{ctat} = \frac{uC_{ox} W_8 W_9 (V_{th8} - V_{th9})^2}{2(\sqrt{L_8 W_9} - \sqrt{L_9 W_8})^2} \tag{2.43}$$

The summation of both currents is made in the transistors M15-M18. The variation in this architecture comes from the non-linearity of both currents.

2.5.8 4-Bits Trimmed CMOS Bandgap Reference

Every single architecture we have seen until now relied primarily on analog design. The matching of transistors with proper weighing, the sum of CTAT and PTAT currents and cascodes structures. However, there is another type of architectures. In these topologies, we use a more digital approach by adding trimming procedures.

Although these trimming procedures usually increase the accuracy of the architecture, they also have several disadvantages like the increase in die consumption and the need of external digital signals to drive the latches.

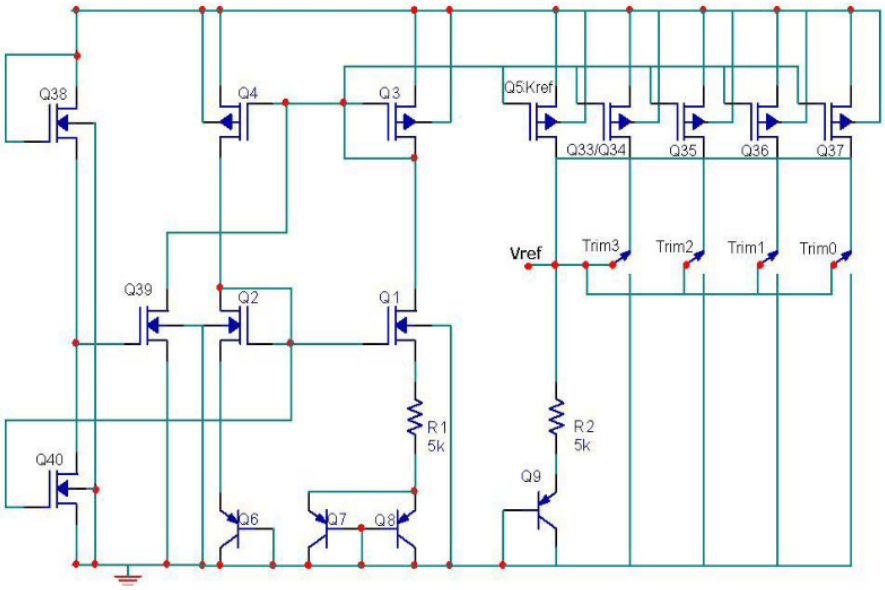


Figure 2.18: 4-bit trimmed CMOS Voltage Reference (From [6])

Before we head into the trimming procedures let us first observe the remaining circuit. The bandgap voltage core is identical to that of section 2.12. They extract CTAT from V_{BE} and PTAT from the differential base emitter voltage of BJT. A small difference is Q7 and Q8 are split into two, parallel connected BJTs instead of a single transistor, to slightly lower process dependency.

Trimming is the concept of shutting down parts of the circuit in order to benefit the overall goal. In this design, we use the standard bandgap voltage reference with Q1-Q8. Q38-Q40 represent the startup circuit. The last four mosfets (Q34-Q37) can be disconnected from the circuit in order to improve process independence.

Let us assume that we have an ideal output of 1.2V and the worst case of expected variation is 5%. The number of trimming bits necessary is given by,

$$n_{bits} = \log_2\left(\frac{1.2}{0.05 * 1.2} + 1\right) \geq 3.41 = 4bits \tag{2.44}$$

Transistors Q34-Q37 are parallel connected to Q5, the mosfet responsible by driving ΔV_{BE} into R2. By switching Q34-Q37 to on or off states we can adjust the current flow through R2. Since there are 4 available bits, this gives us 16 levels.

2.5.9 Overview

From the relevant architectures and compensation techniques we have seen, it should be pointed out that every circuit regarding insensitive temperature relies always on the same technique, that one being the sum of two, opposed scaling, temperature dependent currents to remove the variance.

Another technique is called the ZTC point. Instead of summing two currents, we try to bias the MOS device at a specific voltage in which its internal parameters automatically cancel each

other out, temperature wise of course. However, sometimes such bias point is difficult or even impossible to find.

Supply independence can be achieved mostly by two ways. The first is using the traditional bandgap reference. The voltage reference is achieved by extracting the bandgap silicon voltage, this value being independent of the supply itself seeing it is a physical constant of the silicon itself. The other solution is the use of cascode structures to increase the PSRR (Power Supply Rejection Rate), like the Wilson current mirror that offers very small supply dependence.

Process compensation is difficult to do. The techniques are usually based on optimal transistor matching and intelligent layout design. Prioritizing circuit symmetry is usually the best way to go.

Ref.	Tech	Techniques Used	Devices	Comp.	Output	Area mm^2	Accuracy (%)
[1]	0.13 um	PTAT+CTAT Switched-Capacitor	MOSFET BJT	PVT	16.07 uA	-	0.61
[12]	0.18 um	Switched-Capacitor	MOSFET	PVT	6.88 uA	-	0.029
[11]	0.35 um	PTAT+CTAT	MOSFET BJT	PVT	612 mV	0.117	0.327
[9]	22 nm	PTAT+CTAT	MOSFET	PVT	82 uA	-	5
[10]	65 nm	PTAT+CTAT Cascodes Structures	MOSFET BJT	PVT	6.46 uA	-	3
[2]	0.5 um	PTAT+CTAT	MOSFET	VT	2.670 V	0.01	1
[3]	0.25 um	PTAT+CTAT Cascodes Structures	MOSFET	VT	10.45 uA	0.002	1
[13]	0.18 um	PTAT+CTAT	MOSFET	PVT	58 uA	-	1,38
[14]	0.18 um	ZTC	MOSFET BJT	PT	144.3 uA	-	7
[18]	0.35 um	Trimming	BJT	PVT	1,23V	0.140	1

Table 2.1: Comparison between Current & Voltage References

Chapter 3

Bandgap Voltage Reference Design

Voltage references find applications in a variety of circuits and systems including linear and switching regulators, Analog to Digital (A/D) and Digital to Analog (D/A) converters, voltage to frequency converters, power supply supervisory circuits, power converters and other circuits requiring an accurate reference voltage. An ideal voltage reference must be, inherently, well-defined and its output voltage should be independent of temperature, power supply variations, load variations and other operating conditions.

This chapter discusses the theory and issues, with respect to temperature, that surround the design of a voltage reference. A voltage reference can be categorized into different performance levels (i.e., zero-order, first-order, or second-order). Zero-order references usually are designed using a Zener or a forward-biased diode and typically are not temperature compensated and will not be discussed here. For the first-order voltage references, the first-order term of the polynomial relationship of the output voltage with respect to temperature is canceled. Second-order as well as high-order voltage references, compensate one or more higher-order temperature dependent terms. These accurate voltage references are used for applications such as high-performance data converters and low-voltage power supply systems.

The chapter starts with a section on the design of the most popular voltage references, i.e., the bandgap voltage references. Bandgap voltage references work by summing two voltages levels with opposed scaling. We add a CTAT voltage (equation 2.14) to a properly weighed PTAT (equation 2.19).

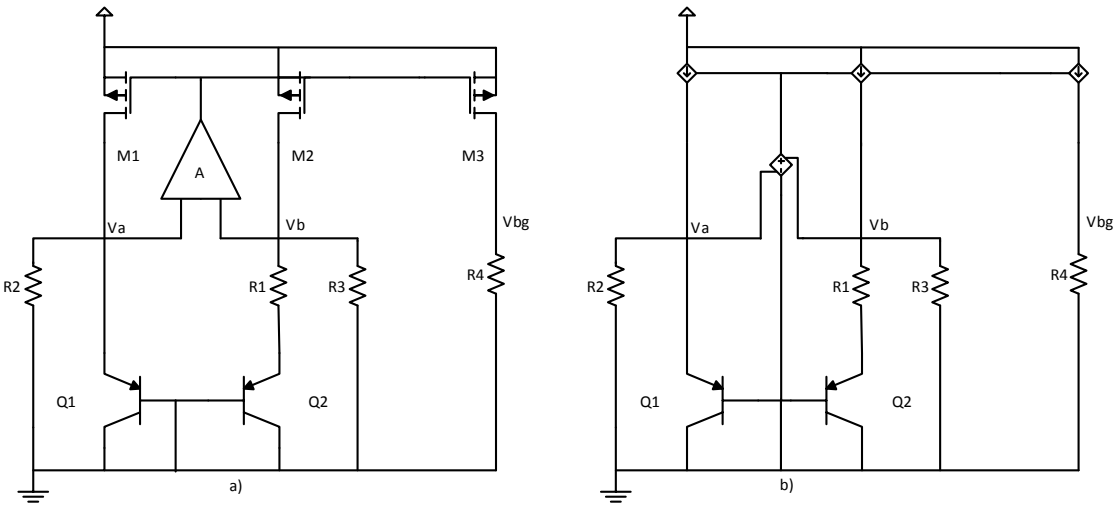


Figure 3.1: a) Current Mode Bandgap Voltage Reference b) Ideal MOSFETs and OPAMP

A standard topology for the current mode bandgap voltage reference is shown in figure 3.1 a). In this section, we will be providing a general understanding of this topology and deriving the output voltage, V_{bg}. In the first-pass analysis of the voltage reference, several simplifying assumptions are made:

1. All PMOS are identical and ideal. 3.1 b)
2. Q_2 is perfectly N times bigger then Q1.
3. The common mode gain of the opamp is zero and its differential gain is infinite such that Va = Vb with no residual offset.
4. Resistor R2 is perfectly matched to R3 such that the currents flowing for each branch are equal.
5. All resistors are ideal and show no temperature dependence or process variation.

3.1 Analysis of the current-mode bandgap voltage reference core

From 2.2.1 we know that temperature independence is achieved by adding the CTAT voltage, which is the standard base-emitter voltage, and a PTAT voltage which is the difference between two base-emitter voltages.

As a starting point, and assuming again an ideal opamp with infinite gain, we begin by saying that Va = Vb with no residual offset. The voltage at node V_a is the base-emitter voltage of Q_1 biased at a certain current. Due to the opamp effect, this voltage is mirrored on node V_b.

Since all PMOS are matched, the same current will flow through all three branches of the circuit. In particular, both BJT will be biased with the same current since $R2 = R3$.

Looking at resistor $R1$, we have the base-emitter voltage from Q_1 in its upper node (V_b) and the base-emitter voltage from Q_2 on the lower node (i.e., emitter node from Q_2).

Thus, the voltage across resistor $R1$ is in fact,

$$V_{R1} = V_{BE1} - V_{BE2} \tag{3.1}$$

We know that the difference between two base-emitter voltages biased by the same current but with different emitter areas yields,

$$V_{BE1}(T) - V_{BE2}(T) = \frac{kT}{q} ln(\frac{A_{E2}}{A_{E1}}) = \frac{kT}{q} ln(N) \tag{3.2}$$

where N is the ratio between emitter areas of both BJTs. Assuming an N=8, we have a voltage drop across $R1$ of

$$V_{PTAT} = \frac{1.3806 * 10^{-}23}{1.602 * 10^{-}19} ln(8) \approx 53.76mV \tag{3.3}$$

Summing up until now we have the PTAT voltage across $R1$, and the CTAT voltage across $R2 = R3$.

Thus we can say that the current flowing through M2 is equal to

$$I_{M2} = \frac{V_{CTAT}}{R_3} + \frac{V_{PTAT}}{R_1} \tag{3.4}$$

This current is mirrored to the output branch (M_3) and flows through resistor $R4$. Finally, we can write the bandgap voltage output as the sum of these two voltages.

$$V_{bg} = R_4[\frac{V_{BE1}}{R_3} + \frac{\frac{kT}{q} ln(8)}{R_1}] \tag{3.5}$$

or equally,

$$V_{bg} = \frac{R_4}{R_3} V_{BE1} + \frac{R_4}{R_1} \frac{kT}{q} ln(8) \tag{3.6}$$

The output voltage is, as intended, the sum of two voltages with opposed scaling with temperature. Since we know from 2.5 that the CTAT has a slope of approximately -1.5mV/^{o}C, we need to find the appropriate ratio of $\frac{R_4}{R_3}$ and $\frac{R_4}{R_1}$ in order to eliminate the temperature variation.

Taking into account the bandgap output voltage derivative against temperature we have

$$\frac{\delta \Delta V_{bg}}{\delta T} = -1.5mV \frac{R_4}{R_3} + 0.09 ln(8) \frac{R_4}{R_1} \tag{3.7}$$

and choosing $R1 = 10k\Omega$, $R2 = R3 = 80k\Omega$, $R4 = 50k\Omega$ we achieve

$$\frac{\delta \Delta V_{bg}}{\delta T} = -1.5mV \frac{50k\Omega}{80k\Omega} + 0.187mV \frac{50k\Omega}{10k\Omega} \approx 0 \tag{3.8}$$

While theoretically we could achieve zero temperature dependency, in the real world such thing is not possible. Especially since we neglected the second order effects. Looking at figure

3.2, which represents the CTAT and PTAT voltages and the output of the bandgap circuit with the calculated resistors ratios. We can see that the slope of both PTAT and CTAT is very close, with a difference of only 8.5uV. This stems from the fact that the base emitter voltage is not completely linear with temperature.

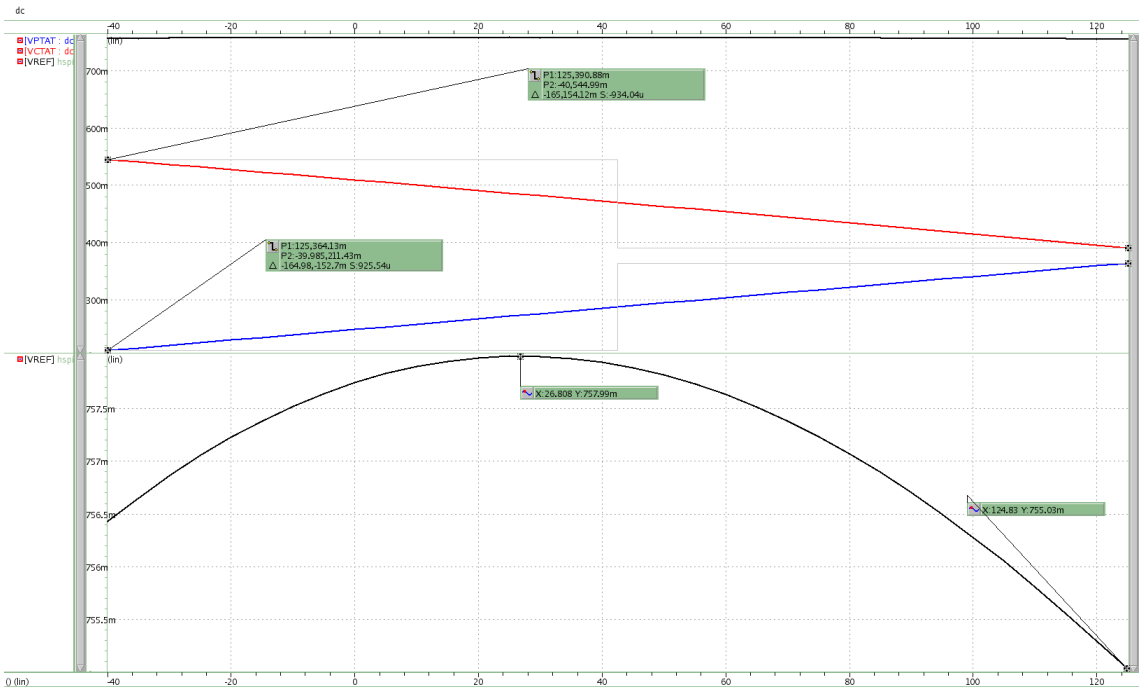


Figure 3.2: CTAT + PTAT Voltages at 1.62 supply voltage

From the highlighted values, the variation from -40°C to 125°C at 1.62V stands around 0.39 %. This variation comes from the neglected second order effect of the base-emitter voltage that was neglected in the first order approximation and the small offset of the operational amplifier. Not to mention that the resistor values were not optimized by simulations.

3.2 Design of the current-mode bandgap voltage reference core

In this section we will go through the necessary steps and decisions made in regards to the design of bandgap core. Note that we will no longer consider anything ideal and the circuit in mind will be that of 3.1 a). We will begin with the design from the top and will later on explain the design of the operational amplifier and other auxiliary circuits.

We begin by taking in mind one of the key factors of how this topology works. The feedback system composed by the operational amplifier and PMOS currents sources will change the current in both branches in order to make V_a and V_b equal. The first step is determining this current so we

can size the PMOS mirrors accordingly. The collector current density for the bipolar transistor is given by 2.13 and will be repeated here for convenience.

$$I_C(T) = I_S(T)e^{(\frac{V_{BE}}{V_T})} \tag{3.9}$$

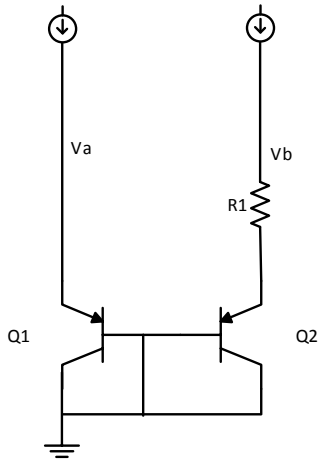


Figure 3.3: BJT bias current measurement scheme

We intend to bias both BJT with the same current, in order to achieve $V_a = V_b$. As shown in figure 3.3, the voltage on the left branch, V_a, is directly tied to emitter the of Q1 and thus will vary exponentially with the current, 3.9. On the right side branch, due to the resistor, the voltage at V_b will vary linearly with the current, taking into account an ideal resistor. Building from 3.9, we can write both voltages as

$$\begin{aligned} V_a &= V_{BE1} = V_T ln(\frac{I_{C1}}{I_S}) \\ V_b &= V_{BE2} + I_C R_1 = V_T ln(\frac{I_{C2}}{I_S}) + I_{C2} R_1 \end{aligned} \tag{3.10}$$

where V_T is the thermal voltage, I_C is the current and I_S is the saturation current.

Now we need to find the current that will make these two voltages identical. We can find this value by making $V_a = V_b$, taking into account that Q2 is 8 times larger than Q1 and solving for I_C.

$$V_T ln(\frac{I_C}{I_S}) = V_T ln(\frac{I_C}{8I_S}) + I_C R_1 \tag{3.11}$$

$$I_C = \frac{ln(8) * V_T}{R_1} = 5.4uA \tag{3.12}$$

The thermal voltage at room temperature which is approximately 26mV and using R1=10k as we did in the section before, we arrive at a current of 5.4uA. This is the current that needs to flow through both branches in order to make $V_a = V_b$ at 27°C and with $R_1 = 10k$.

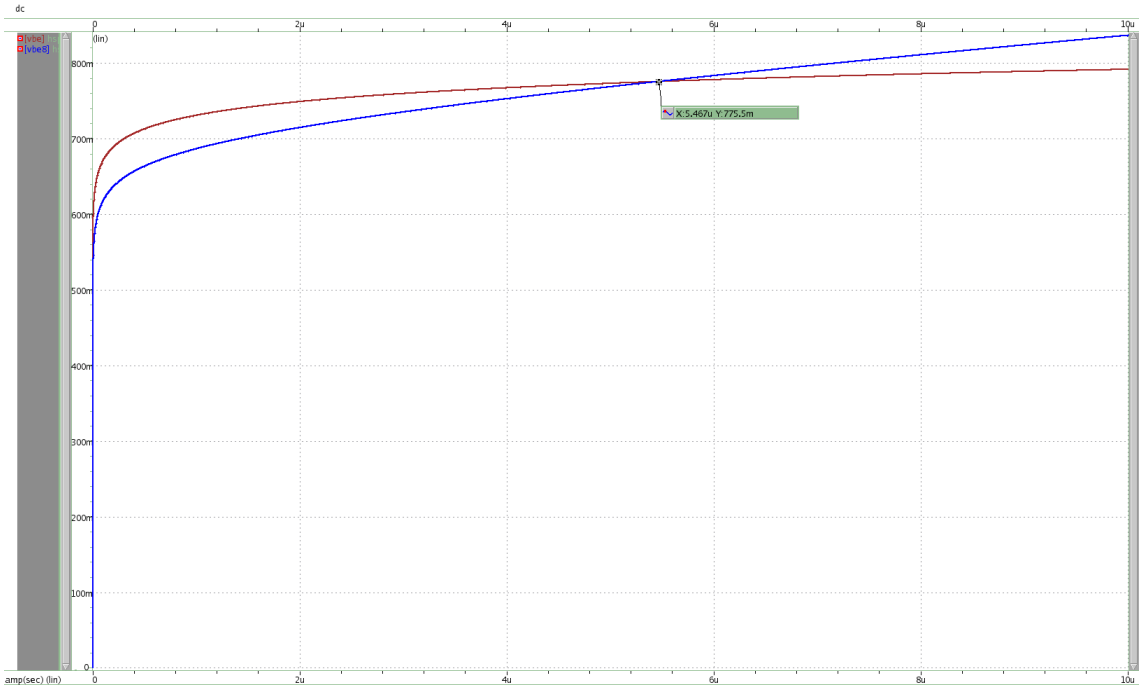


Figure 3.4: Base-Emitter Voltage biased with different currents

Figure 3.4 shows the variation of the V_a and V_b voltages with several currents. The highlighted value shows the bias current in order to achieve $V_a = V_b$, at 27°C.

Looking back to figure 3.1 a), we see that there is an additional resistor connected in parallel with each branch. Having a resistor connected in parallel will "eat" part of the current flowing through the branch, so we need to compensate such that the current going into the BJT is enough to achieve equal voltages.

From figure 3.4 we see that the voltage has a value of 775mV and using the resistor value from the previous section $R_2 = R_3 = 80k$, we conclude that this resistor will have $\frac{0.775}{80k} = 9.7uA$ flowing through it. Adding these two currents, $9.7uA + 5.4uA \approx 15uA$, gives us the necessary current flowing through each of the PMOS in order to achieve our virtual short circuit.

If all three PMOS have equal W/L ratios, we can say that the total current consumption of our circuit will be 45uA. We can lower the consumption by increasing the resistor values, and thus make less current flow through them, but this would translate into increased area in the final circuit.

Now that we have the current for each branch, we can size the PMOS accordingly. Remember that out circuit works with feedback. The operational amplifier will strive to make $V_a = V_b$ by changing the currents. These currents variations are produced by changing the gate voltage of the PMOS. In an ideal world we would not need proper sizing of the PMOS since the amplifier would

place whatever voltage level was necessary at the gates. However, in a real world implementation, the amplifiers are limited by their output swing, which represents the limits of their signal excursion.

Observing our circuit, the current is controlled by the V_{SG} of the PMOS. We need to choose an appropriate gate voltage that is higher than the threshold voltage in the worst case scenario, $|V_{TH}| > 640mV$, but still remains within 0V and our supply voltage, $0 < V_G < SupplyVoltage$. Choosing $V_G = 0.7V$ fulfills the requirements.

Now we can size our PMOS accordingly. Considering the mosfet are working in saturation region and ignoring second-order effects, the current is given by

$$I_d = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{SG} - |V_{TH}|)^2 \tag{3.13}$$

where the factor $\mu_n C_{ox}$ is called Beta (β), W is width of the transistor, L stands for the length, V_{SG} is the source-gate voltage and V_{TH} is the threshold voltage.

Solving for $I_d = 15uA$,

$$15uA = \frac{\beta}{2} \frac{W}{L} (0.92 - 0.64)^2 \tag{3.14}$$

$$\frac{W}{L} \approx 7.36 \tag{3.15}$$

By choosing a W/L ratio close to 7.36 we achieve the desired current, all the while having a decent gate voltage at the PMOS. It is recommended in multiple books and various articles using at least 4 to 5 times the minimum length of the technology at work, in order to lower the short-channel effects.

We are working with 40nm but the high voltage mosfets will be used, with minimum length of 0.27um, we should size the PMOS with W=8u and L=1u, but to lower device mismatches we need to use slightly large devices and thus we choose W=40u and L=5u. The reasons will be explored in later sections.

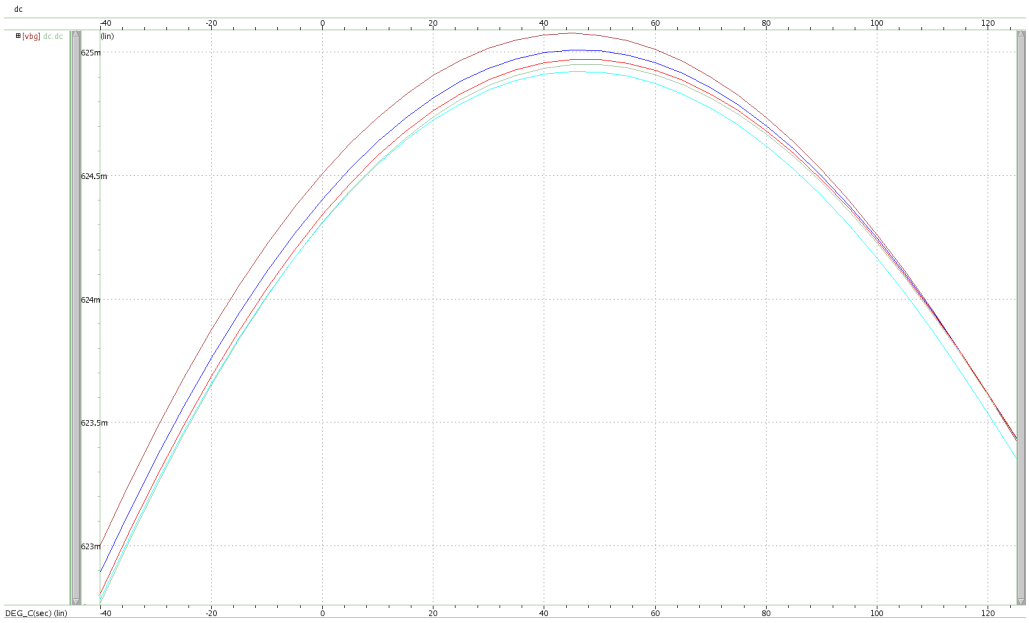


Figure 3.5: Output Bandgap Voltage Reference

Figure 3.5 shows the bandgap voltage reference at several supply voltages. The resistors were optimized by simulation and the new values are R1=11.2k, R2=R3=92K and R4=47K. After running the simulations and optimizing the resistors, we achieve a variation of 0.39% over the temperature range considered and supply voltage from 1.62 to 3.63V at typical corners. Please note that these results were conducted with process resistors and not ideal ones.

3.2.1 Increasing Supply Rejection

The bandgap voltage shows the expected curvature but the voltage seems to increase slightly with the increase of the supply voltage. This is due to the channel-modulation effect that causes an increased current proportional to $\lambda(V_{DS} - V_{DSAT})$, and as the supply voltage increases, V_{DS} also increases. One way to minimize this effect is to lower V_{DS} variation over the considered supply range.

We can do this by using cascodes that also improve PSRR (Power Supply Rejection Ratio) as well. We add these between our PMOS current sources and the BJT as seen below, figure 3.6.

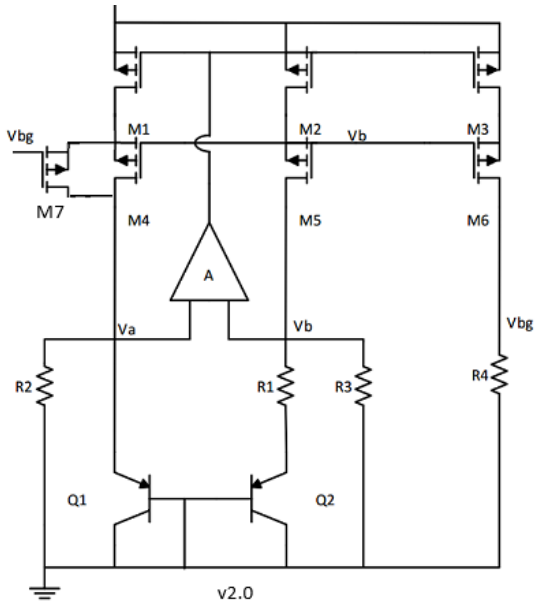


Figure 3.6: Final Topology of the Current-Mode Bandgap Voltage Reference

The cascodes correspond to M4-M6. In the worst case scenario we have 0.87V at Va and a supply voltage of 1.62V. So M4 needs to have appropriate bias voltage so that they do not move M1-M3 out of saturation region. With a supply voltage of 1.62 and a gate voltage of 0.7 we have

$$V_{SDSAT} = V_{SG} - |V_{TH}| = 1.62 - 0.7 - 0.6 = 0.32V \tag{3.16}$$

By giving a saturation margin close to 100 mV, we need a maximum of 1.2V at the drain of M1. It means that we need a bias voltage, V_b, of 0.6V. In regards to the sizing of the transistors themselves, they need a high W/L ratio as well. Simulations show that a W/L ratio of 74 works well so they were sized with W=20u and L=0.27u.

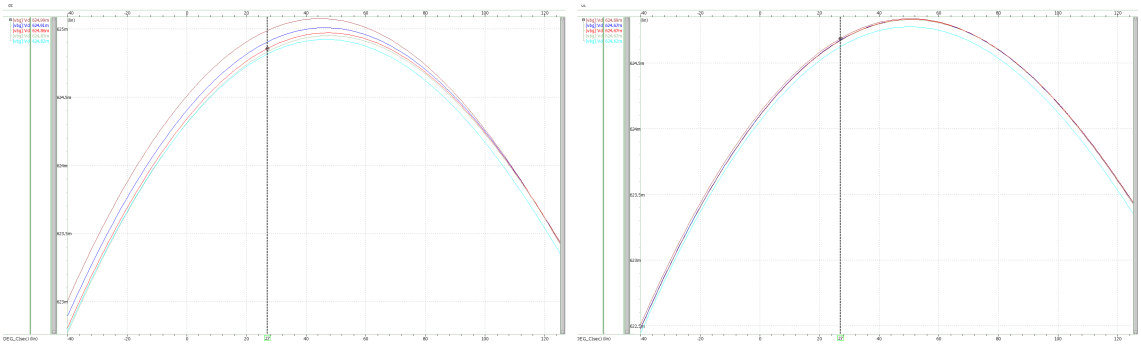


Figure 3.7: Output voltage with no cascode transistors (left) and with cascode transistors (right)

Figure 3.7 shows the effect of the cascode mosfets in the output voltage variation. It is clear that the curves are more packed together with the addition of the cascode transistors, (image on

the right).

The additional transistors lower the overall variation of the drain-source voltage of transistors M1-M3, lowering the impact of channel modulation due to power supply variations.

The figure shows the values for the output voltage at 27°C due to supply voltage shift. Adding the cascodes brings the variation down to 0.011% from the 0.027% with no cascodes, showing nearly a 2.5x increase in supply rejection, at 27°C.

3.3 Operational Amplifier

The operational amplifier is perhaps the most import part of the bandgap circuit, and most of the times the biggest source of errors in the bandgap voltage itself [16]. Up until this point, we assumed that the operational amplifier of figure 3.1 was ideal, but a practical implementation of an amplifier introduces offset between the input pairs, voltage noise, finite power-supply rejection, finite gain and limited output swing.

There are many topologies, each one more different than the other but they have one thing in common and that is a differential input pair. A PMOS differential input pair is almost always the best choice because they offer lower noise [15].

A single PMOS input differential pair with an active NMOS current mirror is a simple example of an amplifier, however it does not have enough output gain for our application. A solution is adding a common-source second stage. This configuration (i.e., input differential pair and common-source output stage) is the most common that provides both high gain and high output swing.

However, it comes with cost of being potentially unstable and thus needing compensation. The compensation is usually done by coupling a capacitor between the output of the first stage and the output of the second stage.

Instead, it was chosen the folded cascode topology since it provides a decent enough gain, and has enough output swing for our needs.

3.3.1 Bias Circuit

Before we move into the folded cascode topology and its design, we will be discussing the bias circuit for the amplifier. One of the most important parameters in analog amplifiers such our own is the transistor transconductance.

$$A_v = G_m R_{out} \tag{3.17}$$

An amplifier's gain is calculated by 3.17, where G_m is the amplifiers transconductance, usually equal to the transistor transconductance of the input transistor, and R_{out}, the output resistance. Unfortunately, the transistor transconductance is a function of the mobility, capacitance of the

oxide layer, the dimensions and the current flowing through it, $g_m(u, C_{ox}, W, L, I_d)$. All of these parameters change with process, affecting the transconductance and thus, the gain.

$$g_m = \sqrt{2u_{n/p}C_{ox}(W/L)I_d} \tag{3.18}$$

In order to achieve a more stable gain across PVT we resort to a simple auxiliary circuit, where the transistor transconductance is matched to that of a resistor, thus being fairly independent of voltage and temperature.

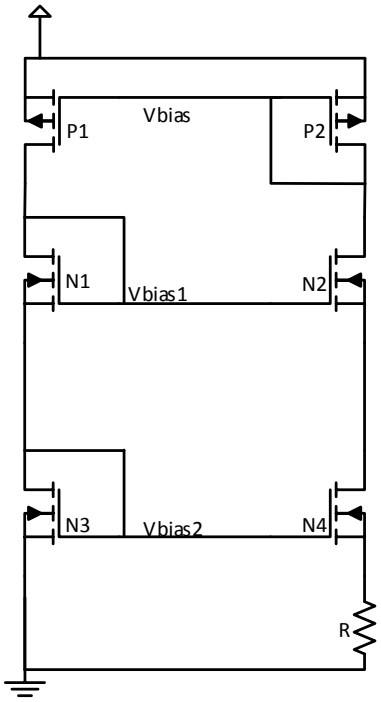


Figure 3.8: Constant-Transconductance Bias Circuit (startup not shown)

The bias circuit is shown in 3.8. First we assumed that P1 and P2 are matched, $(W/L)_{P1} = (W/L)_{P2}$. This results in both branches having the same current, $I_{d1} = I_{d2}$, due to the mirror effect of P1 and P2.

Looking at the bottom two transistor, we can write

$$V_{gs3} = V_{gs4} + I_{d2}R \tag{3.19}$$

where V_{gs} is the gate-source voltage for either N3 or N4.

If we subtract V_{th} from both sides of the equation we arrive at

$$V_{ov3} = V_{ov4} + I_{d2}R \tag{3.20}$$

where V_{ov} is the overdrive voltage of the transistor, $V_{ov} = V_{gs} - V_{th}$.

From 3.13 we can also write this equation as

$$\sqrt{\frac{2I_{d1}}{u_n C_{ox}(W/L)_3}} = \sqrt{\frac{2I_{d2}}{u_n C_{ox}(W/L)_4}} + I_{d2}R \tag{3.21}$$

But since $I_{d1} = I_{d2}$ and rearranging the expression we arrive at

$$\frac{2}{\sqrt{2u_{n/p}C_{ox}(W/L)I_d}}[1 - \sqrt{\frac{(W/L)_3}{(W/L)_4}}] = R \tag{3.22}$$

Recalling g_m from 3.17 we finally arrive at

$$g_m = \frac{2[1 - \sqrt{\frac{(W/L)_3}{(W/L)_4}}]}{R} \tag{3.23}$$

Thus, the transconductance of N3 is determined solely by R and geometric ratios. For the special case of $\frac{(W/L)_3}{(W/L)_4} = 1/4$, g_m is given solely by R

$$g_m = \frac{1}{R} \tag{3.24}$$

Not only is g_{m3} stabilized, but all other transconductance are also stabilized since all transistors are derived from the same biasing current. The table below shows the selected values for the transistors and the single resistor.

	W	L	Ω
P1,2	2u	1u	-
N1,2,3	2u	2u	-
N4	12u	2u	-
R	-	-	114k

Table 3.1: Constant-Transconductance Bias Circuit Element Values

3.3.2 Folded Cascode Amplifier

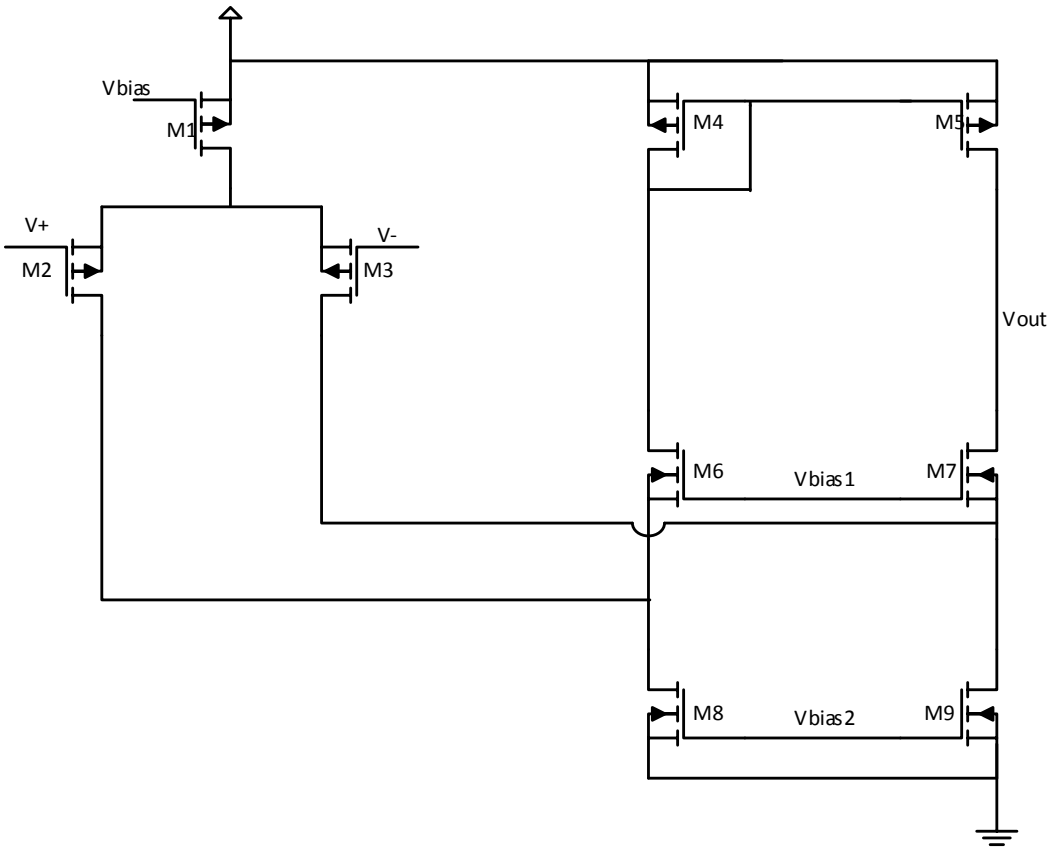


Figure 3.9: Folded Cascode Amplifier

The figure above shows the folded cascode amplifier topology used in this work. Transistor M1 provides the bias current mirrored from the constant-gm bias circuit. M2 and M3 are the PMOS input differential pair, M8 and M9 are current sources, M4 and M5 represent the active current mirror and finally, M6 and M7 are cascode, thus the name of the topology.

Since we want our amplifier to have a constant transconductance over PVT we need to use the bias circuit explained in the previous section, 3.3.1.

We begin with transistor M1 that provides tail current for the amplifier. A simple current mirror towards our bias circuit is exactly what we need. Please note that Vbias, Vbias1-2 correspond to current mirrors taken from our bias circuit.

Transistors M8 and M9 should be sized such the current I_4 and I_5 is never zero, in our case $I_{8/9} = 3I_1$. These transistors are current sources and required large areas to lower the offset due to process mismatches.

Transistors M2 and M3 represent our input differential pair and need particularly big sizes since most of the offset will stem from them.

The remaining transistors, M4 and M5 are the current mirror for single ended conversation.

	W	L
M1	2u	1u
M2,3	40u	1u
M4,5	4u	5u
M6,7	1u	1u
M8,9	12u	2u

Table 3.2: Folded Cascode Amplifier Element Values

Table 3.2 shows the size of the transistors used in the amplifier.

3.3.2.1 Output Swing

The output swing refers to the limits of the amplifier's signal excursion. With proper choice of V_{b1}, the lower end of the output swing is given by $V_{OV7} + V_{OV9}$ and the upper end is given by $V_{DD} - |V_{OV5}|$. Thus, the peak-to-peak swing swing of our amplifier is $V_{DD} - (V_{OV7} + V_{OV9} + |V_{OV5}|)$.

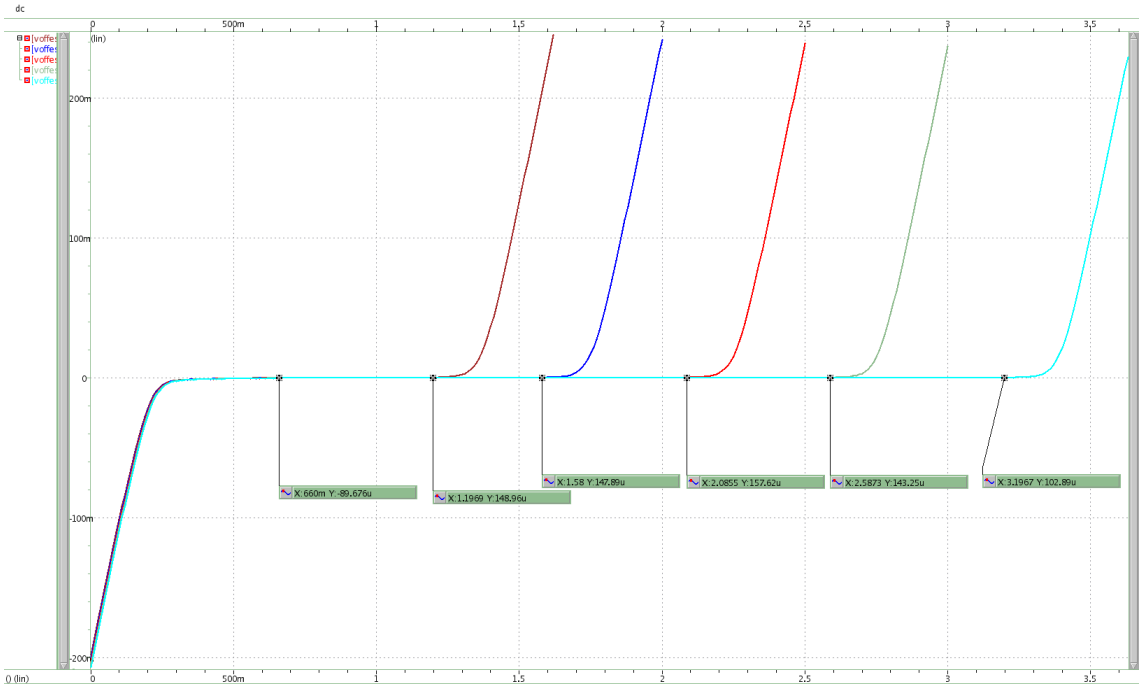


Figure 3.10: Differential input voltage across several supply voltages at typical corners and 27°C

In our case, simulations show that a swing from 0.6 to 1.2 at 1.62V is acceptable. Outside of this range, the systematic offset is too large as the devices start entering triode region.

Figure 3.10 shows the offset voltage between the two input terminals. The offset voltage lowers significantly once the input voltage surpasses 0.6V. The offset voltage remains fairly constant over the supply range and begins increasing once more close to the $V_{DD} - 0.42V$

3.3.2.2 Open-Loop Voltage Gain & Phase Margin

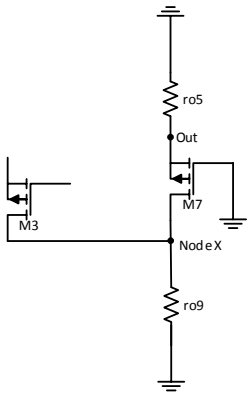


Figure 3.11: Half circuit for voltage gain analysis

Figure 3.11 shows the right half of the amplifier prepared for small signal analysis. Transistor M5 has been replaced by its equivalent output resistance, R_u.

The gain of the amplifier is given by

$$A_v = G_m R_o \tag{3.25}$$

where G_m is the transconductance of transistors M3 and R_o is the output resistance of the amplifier.

R_o of the amplifier is the parallel between the resistances looking up and looking down from the output node. The resistance looking up, R_u, is determined by the output resistance of transistor M5, $R_u = r_{o5}$.

The resistance looking down is approximately equal to $g_{m7} r_{o7} (r_{o9} || r_{o3})$, where r_{o9} is the output resistance of transistor M9. Replacing the calculated values in the gain equation we arrive at

$$A_v = g_{m3}([g_{m7} r_{o7} (r_{o9} || r_{o3})] || r_{o5}) \tag{3.26}$$

By resorting to the .OP command of HSPICE, we can extract the necessary parameters for our calculations. The open loop gain is then given by

$$A_v = 10.9u([3.48u \times 56M \times 50M || 74M)] || 847M) \approx 8480 = 78.56dB \tag{3.27}$$

Please note that the values were kept in their respective position in order to easily match them to the previous equation.

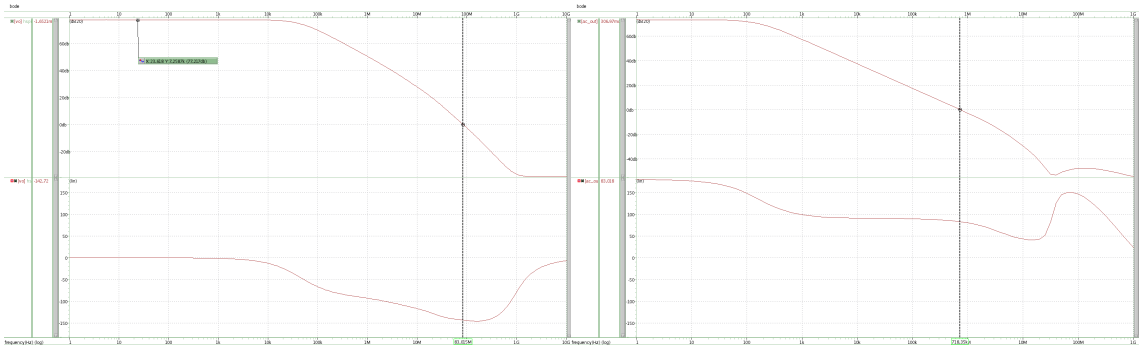
The voltage gain of our amplifier is 78 dB at typical corners and room temperature. This gain is more than sufficient for our needs since the offset due to the limited again is very small.

The frequency response of the folded cascode amplifier is determined primarily by the output pole which is given by

$$z_{out} = \frac{1}{2\pi R_o C_o} \approx 68kHz \tag{3.28}$$

where R_o is the output resistance determined previous and equal to $780M\Omega$, and C_o which represents the total capacitance associated with the output node, which is 3fF if the output node is floating.

Looking at 3.12a, we can clearly see that the calculated gain is very close to the simulated one.



(a) Frequency response of the cascode amplifier with 3fF

(b) Frequency response of the cascode amplifier with equivalent mosfet load

Figure 3.12: Frequency response of the cascode amplifier

Figure 3.12 shows the frequency response of the folded cascode amplifier with different capacitance loads in typical corners and room temperature. Figure 3.12a represents the frequency response of the amplifier with nothing connected to its output (i.e., the output terminal of the amplifier is floating). The response is determined primarily by the output resistance and the parasitic capacitances associated with the output node.

However, the operational amplifier was not designed to operate in standalone mode. This device will be integrated in our bandgap circuit which introduces a capacitance load on its output. Figure 3.12b shows the frequency response of the amplifier once it has been introduced in our bandgap circuit.

The amplifier's output will be attacking the three top PMOS current mirrors which regulate the overall current in the system, (figure 3.6). The transistors were designed with fairly big area in order to reduced mismatch between themselves. Increasing the transistors area implies increasing the parasitic capacitances, which in our case, increases the capacitance load at the output of the amplifier.

Looking at equation 3.28, we see that if the output resistance remains constant, the pole will be determined solely by the capacitance connected to the output node. The capacitance associated with the three PMOS dominates over the intrinsic 3fF of the output node, making the pole dominated by the output resistance of the amplifier and the capacitance of the current mirrors.

Looking at 3.12b we see the expected response. The amplifier's pole was pulled towards the low frequency area due to the high capacitance on its output. This has the benefit of increasing the overall stability but comes with a cost of losing bandwidth.

Since our amplifier is not required for any high speed operations, bandwidth is something we can easily sacrifice for increased stability.

3.3.2.3 Summary

The following table provides an overview of the folded cascode amplifier's characteristics.

Parameter	Test Conditions	Worst	Typical	Best
Gain (dB)	PVT	76	78	84.5
Phase Margin (^{o})	PVT	27.5	32	41
	PVT With Mosfet Load	74.1	82	86.09
Voltage Output Swing (V)	Vdd = 1.62 @ 27^{o}C	-	0.6 to 1.2	-
Input Offset Voltage(uV)	PVT	4000	110	80
Current Consumption (uA)	PVT	5.11	2.82	1.6
PSRR (dB) DC	PVT	-74	-78	-84
Slew Rate (V/us)	Vdd = 1.62 @ 27^{o}C Typical	-	5.43	-

Table 3.3: Folded Cascode Amplifier Parameters

The table below shows the same characteristics after layout design.

Parameter	Test Conditions	Worst	Typical	Best
Gain (dB)	PVT	66	82	98
Phase Margin (^{o})	PVT	-1	13	23
	PVT With Mosfet Load	52	80	88
Voltage Output Swing (V)	Vdd = 1.62 @ 27^{o}C	-	0.6 to 1.1	-
Input Offset Voltage(uV)	PVT	7509	800	-
Current Consumption (uA)	PVT	4	2.03	1
PSRR (dB) DC	PVT	-40	-48	-126
Slew Rate (V/us)	Vdd = 1.62 @ 27^{o}C Typical	-	2.12	-

Table 3.4: Folded Cascode Amplifier Parameters (Post-Layout Simulations)

After layout validation and post-simulations we can see that several parameters were affected. The biggest impact was the systematic input offset voltage increasing nearly 8 times. The other characteristic was the phase-margin actually being smaller than 0 and making the amplifier unstable in the worst case. Fortunately for us, the mosfet load is enough to give a decent phase margin even in the worst case scenario.

3.4 Startup Circuit

One point that should be carefully analyzed is the operating point of our circuit. Please refer to figure 3.4, repeated here for convenience.

Looking at this picture, we can clearly see that our circuit has two stable operating points. When all voltages are 0V, our circuit is in a stable operating point with all currents being zero as well. For obvious reasons we do not want our circuit operating at 0V, so there is a need for what is called a startup circuit.

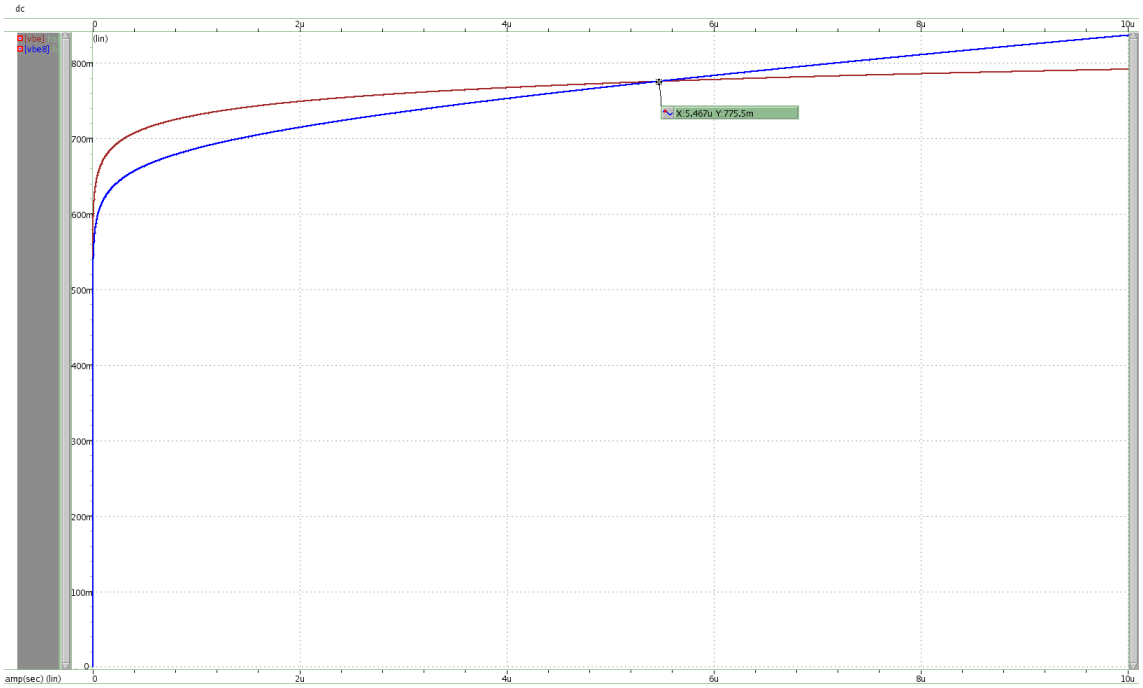


Figure 3.13: Base-Emitter Voltage biased with different currents

Looking at figure 3.6, the startup circuits is represented by transistor M7. In the 0V operating point we have $V_{bg} = 0$, $V_a = 0$ and $V_b = 0$. However, at any point in time, the source voltage of the M7 transistor is close to 660mV, in the worst temperature scenario, -40^oC.

The startup circuit works as follows. Suppose the circuit is operating with biasing current equal to zero, even though a 1.62, or bigger, supply voltage is applied - the wrong operating point. Thus, it is possible that the gate voltage of M1,2,3 is at a high enough voltage that causes V_{SGM1} to be lower than one threshold voltage and thus, be in cut-off region in which there is not enough current flowing through the branch.

In such situation, the bandgap voltage, V_{bg}, is equal to 0V. Since the source voltage of transistor M7 is at 660mV and the gate voltage is at 0V, V_{bg}, this transistor enters in conduction with $V_{SG} > V_{TH}$.

Transistor M7 will thus inject current into both the resistor R2 and the bipolar transistor Q1. This current will force the BJT into forward operation, generating a voltage level at node V_a. The

operational amplifier, seeing an offset voltage between nodes V_a and V_b, will strive to make them equal by forcing current into the second branch. Thus, the circuit will eventually converge towards the correct operating point.

With the reference in operation, the bandgap voltage is now equal to 624mV. Thus, the source-gate voltage of transistor M7, V_{SG}, will be approximately 37mV, much lower than the threshold voltage and thus forcing transistor M7 into cutoff, effectively disconnecting the startup circuit from the main circuit.

This circuit needs to be strong enough to force our circuit into the correct operating point but needs to turn off once the operating point has been achieved.

Figure 3.14: Startup response

The figure above shows the transient response of our circuit. The supply voltage was ramped slowly, reaching 1.62V at 10us.

We can clearly see that the amplifier's bias circuit would operate in cut off region without proper startup. The signal 'vbias' shows the voltage level is very close to the supply voltage. This signal is applied to a PMOS mosfet which translates into the device being in cutoff region, and thus, no significant amount current flows through the circuit.

We can see that startup circuit forces the bias circuit into the correct operating point and once it has been achieved, the bandgap circuit follows quite closely – signal 'vbg'. The bandgap voltage is achieved just close to 50us. Once the startup is finished, the current flowing through the startup circuit is negligible and around 569fA.

3.5 Stability analysis

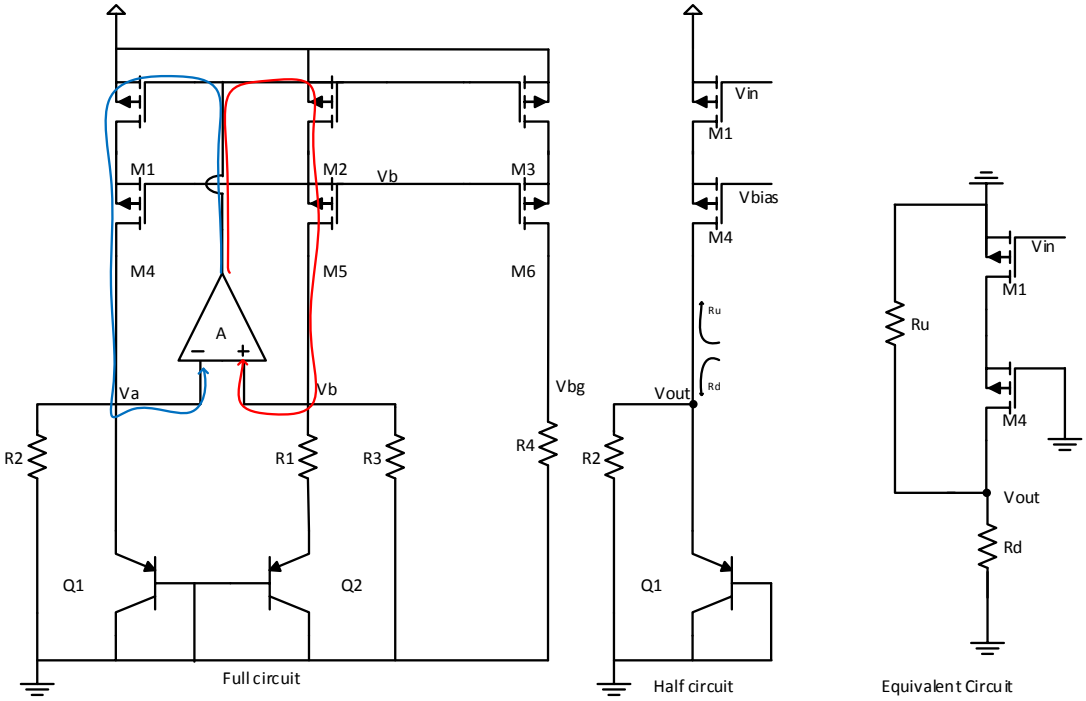


Figure 3.15: Bandgap circuit feedback loops

It is pertinent to note that the feedback signal produced by the operational amplifier returns to both its inputs.

The figure above shows our bandgap circuit with both feedback loops highlighted. In order to make sure amplifier is working as intended, the negative feedback must dominate over positive feedback, usually by a factor of at least two.

Both feedback loops can be thought as common source amplifiers. The input signal enters through the gate of the p-type transistors M1 and M2 and exits through their drain, respectively. The remaining terminal is known as common, thus the name, common source.

The voltage gain for common source amplifier is given by

$$A_v = -\frac{g_m R_L}{1 + g_m R_S} \tag{3.29}$$

where g_m is the device's transconductance, R_L is the load connected to the drain terminal and R_S is the load connected to the source terminal. Since our amplifier is a common source circuit with no source degeneration (i.e., $R_S = 0$), the voltage gain is simply given by

$$A_v = -g_m R_L \tag{3.30}$$

Figure 3.15 shows the left loop and its equivalent circuit for analysis. Looking from the output into our circuit we see the resistance going up, R_u, and the resistance going down, R_d.

R_d is given by

$$R_d = R_2 || \frac{1}{g_{mQ1}} || r_\pi \tag{3.31}$$

where g_{m1} refers to the transconductance of transistor Q1 and r_π is internal resistor from the base to the emitter.

R_u looking up is given by

$$R_u = g_{m4} r_{o4} r_{o1} \tag{3.32}$$

where r_{o1} and r_{o4} are the output resistances of transistors M1 and M4, respectively. Replacing both resistances into 3.30 we arrive at (check figure 3.15)

$$A_v = -g_m(r_u || r_d) \tag{3.33}$$

Since $r_u \gg r_d$, then $r_u || r_d \approx r_d$ and so the gain is given solely by the transconductance of M1 times its load, ignoring the cascode transistor M4.

Now writing the gain equations for both loops we arrive at

$$\begin{cases} A_p = g_{m1}(R_2 || \frac{1}{g_{mQ1}} || r_{\pi 1}), & \text{Positive feedback gain} \\ A_n = g_{m2}(R_3 || (\frac{1}{g_{mQ2}} || r_{\pi 2} + R_1)), & \text{Negative feedback gain} \end{cases} \tag{3.34}$$

Where $g_{m1} = g_{m2}$, $R_2 = R_3$, $g_{mQ1} = g_{mQ4}$ and $r_{\pi 1} = r_{\pi 2}$. It can be seen that the only different term between both equations is the addition of resistor R_1.

Using a operating point statement in HSPICE simulations (i.e., .OP) gives us the parameters of our circuit elements.

$$\begin{cases} 20log(A_p) = 101u(92k || \frac{1}{109u} || 13k) \approx -5.79dB, & \text{Positive feedback gain} \\ 20log(A_n) = 101u(92k || (\frac{1}{109u} || 13k + 11.2k)) \approx 4.45dB, & \text{Negative feedback gain} \end{cases} \tag{3.35}$$

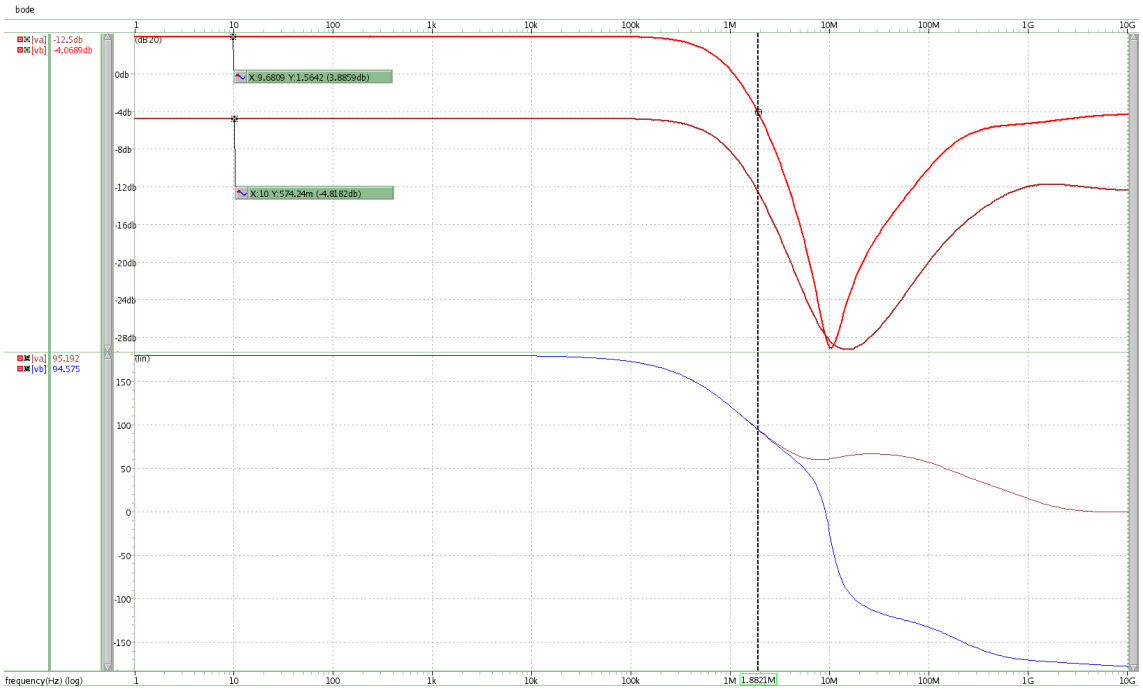


Figure 3.16: Bandgap circuit feedback loops

Figure 3.16 shows the simulated gain under typical conditions. It is visible that the calculated values are close to the simulated ones. However, the key note is that the negative feedback gain is higher than the positive feedback.

$$\frac{A_n}{A_p} = 3.88 - (-4.81) = 8.69dB \approx 2.72 \tag{3.36}$$

It is clear that the negative dominates over positive, assuring the correct operating conditions for our circuit. It is worth noting that the negative feedback is in fact connected to the positive terminal of the amplifier. This is because transistors M1 and M2 are in common source configuration, thus inverting the signal.

3.6 Results

In this section, we will be presenting the results of several important parameters of our circuit. Each result will be presented in a sub section, providing the theoretical calculation and the simulated results for comparison, in the worst, typical and best case scenarios. Please note that these results do not include MonteCarlo simulations.Those simulations are included in section 3.6.5.

3.6.1 Current Consumption

Current consumption is one of the most important parameters in any circuit. We want any circuit designed to consume as less energy as possible. Motivated by emerging battery-operated appli-

cations that demand intensive computation in portable environments, we will strive to make our power consumption as small as possible without sacrificing other specifications.

There are some simple techniques to use in low power design like operating in low voltages, using reduced w/l ratio types cmos, low threshold voltages and high resistances.

Looking at our circuit, figure 3.6, we can clearly see that our power consumption comes from the top three PMOS mirrors, that regulate the current going into each branch, our operational amplifier and its bias circuit.

The total current consumption is given by

$$I_T = 3I_{PMOS} + I_{AMP} + 2I_{BIAS} \tag{3.37}$$

where I_{PMOS} is the current in each branch and, ignoring second order effects and eventual mismatches, is equal between all branches, I_{AMP} is the current used by the amplifier and I_{BIAS} is the current consumed by the amplifier's bias circuit in each branch. The current from the startup circuit is considered irrelevant as explained in the previous section.

The current consumed by the PMOS mirror can be calculated by using 3.11 and factoring in the side resistors.

$$I_{PMOS} = \frac{ln(8) * V_T}{R_1} + \frac{V_T ln(\frac{I_{C1}}{I_S})}{R_2} \tag{3.38}$$

$$I_{PMOS} = \frac{ln(8) * 26mV}{11.2k\Omega} + \frac{26mV ln(\frac{\frac{ln(8)*26mV}{11.2k\Omega}}{I_S})}{92k\Omega} \approx 13.25uA \tag{3.39}$$

The current used by the bias circuit of the operational amplifier, I_{BIAS}, can be calculated by 3.23 and factoring in the resistor value and size of the transistors.

$$g_m = \frac{2[1 - \sqrt{\frac{1}{6}}]}{114k} \approx 10.4uS \tag{3.40}$$

Using equation 3.17, repeated here for convenience, and solving for I_d

$$g_m = \sqrt{2u_{n/p}C_{ox}(W/L)I_d} \tag{3.41}$$

$$I_d = \frac{(g_m + g_{mb})^2}{2u_{n/p}C_{ox}(W/L)} = \frac{(10.4u + 3u)^2}{2 * 52u * 2} \approx 860nA \tag{3.42}$$

Please note that it was included the transconductance due to the body effect, g_{mb}. The transistor's subtract was connected to the highest potential, V_{DD}, for PMOS and the lowest potential, GND, for a NMOS. When the source bulk voltages are different, $V_{SB} \neq 0$, the body effect appears, causing an unwanted change in our transconductance.

The last component is the operational amplifier consumption. The current use is regulated by the bias circuitry and since our amplifier is biased through a current mirror with the same W/L ratio, we can say that

$$I_{AMPOP} = I_{BIAS} \tag{3.43}$$

The total current usage is easily computed by adding all three components.

$$I_T = 3I_{PMOS} + I_{AMP} + 2I_{BIAS} = 3 * 13.25u + 863nA + 2 * 863nA \approx 43.15uA \tag{3.44}$$

The following table shows a comparison between the calculated values and the simulated ones at typical corners, $V_{DD} = 1.62V$ and $T = 27^oC$.

	Calculated	Simulated
I_{PMOS}	13.25uA	13.252uA
I_{AMP}	863nA	815.45nA
I_{BIAS}	863nA	811nA
Total	43.15uA	42.578uA

Table 3.5: Calculated current usage vs. Simulated values

The small shifts in the calculated currents against the simulated ones come from unaccounted secondary effects of the transistors. The overall current usage is good and similar the main article we based the early topology on [17], which reported a total consumption of 45uA at typical conditions as well.

	Current	Temperature	Voltage
Worst Case	43.03uA	T=100°C	3.63V
Typical Case	42.578uA	T=27°C	1.62V
Best Case	41.79uA	T=-40°C	1.62V

Table 3.6: Current usage for worst and best scenarios for typical corners

Table 3.6 shows the worst and best scenarios in typical conditions. It is important to study the typical case in particular since most of the manufactured circuits will fall under these conditions.

The following table shows the worst and best cases overall, between all process corners, temperature and supply range.

	Current	Resistor Corner	Mosfet Corner	Bjt Corner	Temperature	Voltage
Worst Overall Case	54.806uA	Fast	Slow-Slow	Slow	125°C	3.63V
Typical Case	42.578uA	Typical	Typical	Typical	27°C	1.62V
Best Overall Case	34.924uA	Slow	Slow-Slow	Fast	-40°C	1.62V

It is interesting to see that the resistor corners dominate the worst and best cases. This is due to the fact that resistors vary heavily under process and most of the current usage stems from them.

If we neglect the current use of the amplifier and its bias circuit, and consider only the bandgap core, $I_{total} = I_{PMOS}$, we can write

$$I_{TOTAL} = I_{BJT} + I_R \tag{3.45}$$

I_{BJT} remains fairly constant over process but I_R changes a lot due to the resistor effect, that at process corners can reach up to 20% variation. Considering a 20% shift in the resistors corners of R_2 and R_3 from the nominal value 92kΩ, we get $R_{2fast} = 73.6k\Omega$ and $R_{2slow} = 110k\Omega$, and $R_{1fast} = 9k\Omega$ and $R_{1slow} = 13.4k\Omega$

$$I_{PMOS} = \frac{ln(8) * V_T}{R_{1FAST}} + \frac{V_T ln(\frac{I_{C1}}{I_S})}{R_{2FAST}} = 51uA \tag{3.46}$$

$$I_{PMOS} = \frac{ln(8) * V_T}{R_{1SLOW}} + \frac{V_T ln(\frac{I_{C1}}{I_S})}{R_{2SLOW}} = 35uA \tag{3.47}$$

These results are very similar to the measure ones. As expected, resistors corners dominate the current use as proved and observed.

3.6.2 Temperature Behaviour

Most of the times we are more interested in the temperature behaviour of the system than the overall variation. Many current sources used in today's electronics are called PTAT current sources. These references do not offer precision in regards to never changing over PVT, but what they offer is a predictable behaviour. Very similar to the PTAT voltage that increases with temperature, this current has the same behaviour.

In this section, we will be discussing the temperature behaviour of our topology, including the inflection point, the PPM of our circuit and finally, the worst possible case.

The inflection point can be determined by taking the derivative of 3.6 with respect to temperature. It is assumed that all resistors share the same temperature coefficient and thus, is eliminated due to division.

$$\frac{\delta \Delta V_{bg}}{\delta T} = \frac{R_4}{R_3} V_{BE1} + \frac{R_4}{R_1} \frac{k}{q} ln(8) \tag{3.48}$$

Since V_{BE} is very complex we will be recurring to simulations. The equation above includes several circuits parameters, such the ratio between resistors and area difference between both BJT. Then, we can conclude that the inflection point will move according to those parameters.

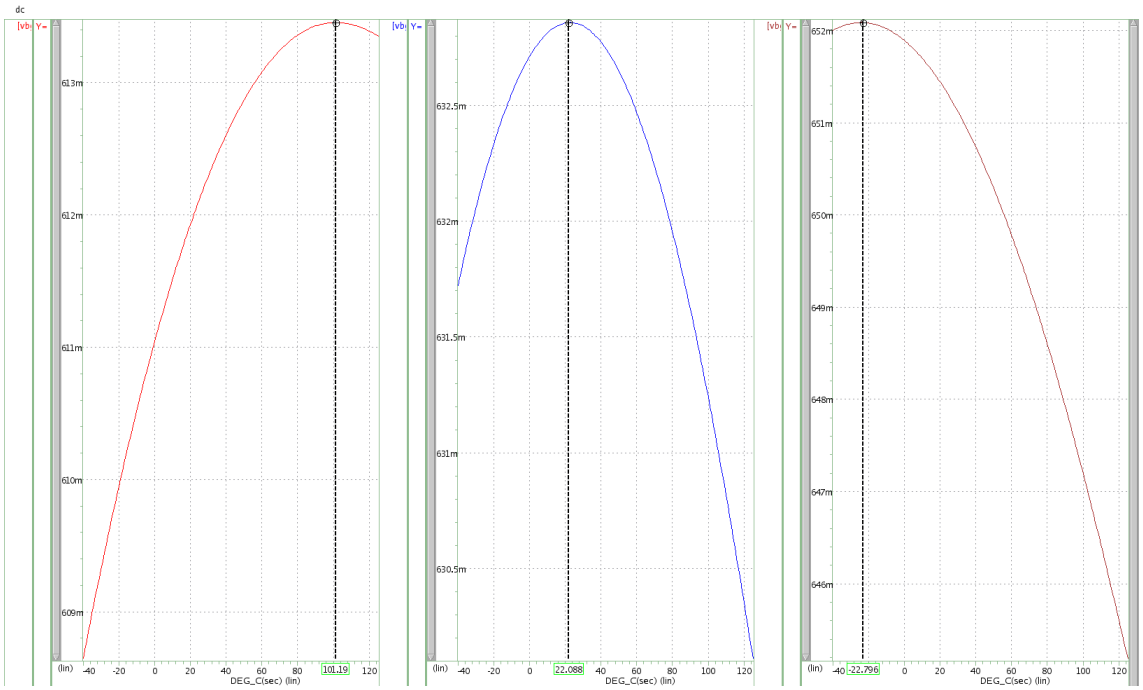


Figure 3.17: Multiple inflection points

Figure 3.17 shows the temperature behavior over several inflection points. This behavior was simulated by changing the value of R_2 to $94k\Omega$, $92k\Omega$ and $90k\Omega$, respectively. As expected, we can clearly see a shift in the inflection point of our curve. Lowering the resistance value of resistors $R_2 = R_3$, pulls the curve to the left, towards the lower temperatures.

It is also clearly visible that the ideal inflection point is somewhere in the middle of the temperature range. Since placing the inflection point close to the edges of our temperature shows high variation. Simulations showed that the ideal inflection point in typical corners was situated at 50°C, very close to the middle of the temperature range, 42.5°C.

The temperature coefficient is calculated by

$$T.C. = \frac{\Delta V * 10^6}{\Delta T} \tag{3.49}$$

ΔT is our temperature range, from -40°C to 125°C, which yields a global of 165 degrees, and ΔV is our voltage drift, calculated by $V_{MAX} - V_{MIN}$.

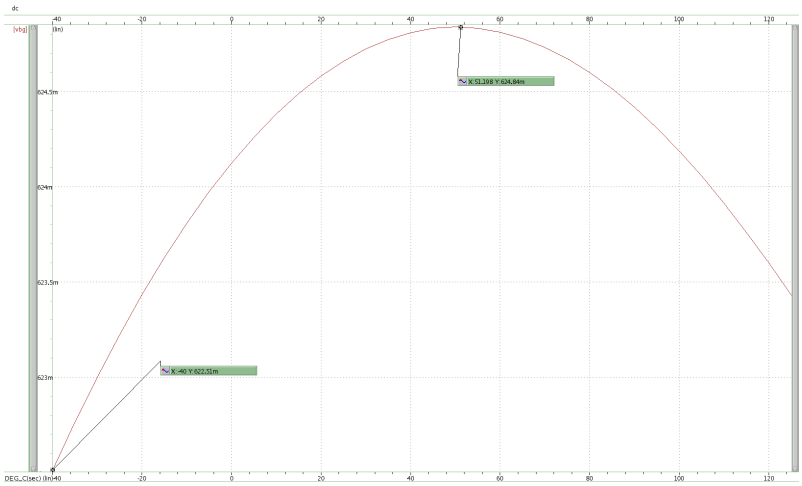


Figure 3.18: Temperature behavior at typical corners, Vdd=1.62

Looking at the figure above we calculate our temperature coefficient by

$$T.C. = \frac{(0.62484 - 0.62251) * 10^6}{165} \approx 14.2 ppm/^oC \tag{3.50}$$

Which is a very good value for first order correction. Lower temperature coefficients would require curvature compensation, or second order correction, and is not explored in this work. This work shows an improvement over [17] and one of very lowest T.C.s seen with first order temperature correction. Even including our supply range, and the temperature coefficient increasing slightly to 15.7 ppm/^{o}C, it still dominates the lowest values over the various articles explored for this work (although some of the results are difficult to compare because our results are not from an actual fabricated circuit).

The lowest T.C found was [5] which showed a T.C. of 17.6 ppm/^{o}C over a 120 degree range, in typical conditions.

	TC	Resistor Corner	Mosfet Corner	Bjt Corner	Temperature	Supply
Worst Overall Case	36.4	Fast	Slow-Fast	Slow	-40^{o}C to 125^{o}C	1.62V
Typical Case	14.2	Typical	Typical	Typical	-40^{o}C to 125^{o}C	1.62V
Best Overall Case	12	Fast	Fast-Fast	Typical	-40^{o}C to 125^{o}C	1.62V

Table 3.7: Temperature Coefficients for worst, typical and best cases

The table above shows the typical and extreme results of the temperature coefficient variation with process corners. It is interesting to note that the best case scenario is not, in fact, the typical corner as one would have imagined.

This particular combination of corners provides a better matching of the temperature coefficients CTAT and PTAT, since our circuit was optimized to the typical corner, which is where the bigger majority of the devices manufactured will fall into.

3.6.3 Line Regulation

The line regulation, $S_{LR,T(nom)}$, specifies the variation of the output voltage reference circuit, ΔV_{bg}, with respect to the input voltage variation, $V_{MAX} - V_{MIN}$, at a nominal temperature. Line regulation is specified as uV/V.

$$S_{LR,T(nom)} = \frac{\Delta V_{bg}(\Delta V_{IN})}{\Delta V_{IN}} \tag{3.51}$$

where ΔV_{bg} is the variation of the reference voltage simulated within the input voltage variation in the range of [VIN(min), VIN(max)], and ΔV_{IN} is the supply range.

The test bench used for this measurement is presented below. A supply voltage is connected to our bandgap voltage reference who possesses a single output signal, the bandgap voltage.

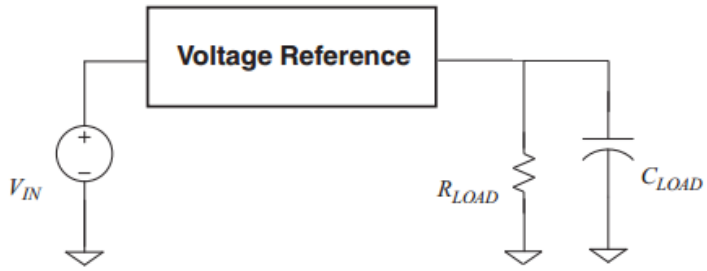


Figure 3.19: Test-bench for bandgap voltage reference line regulation

Note that line regulation can be affected by R_{LOAD} and C_{LOAD}. In the particular case of our reference, the resistive load connected to our voltage reference needs to be high so that the current drawn from the circuit is negligible and does not affect the circuit overall performance.

If R_{LOAD} is comparable to R_4 of the internal circuit, and let us assume $R_4 = R_{LOAD}$, then R_{LOAD} would draw about half of the current flowing through the third branch of our bandgap, which would lower the bandgap voltage to half of the expected value. This is undesired, so R_{LOAD} needs to be in the high $M\Omega$ region.

In CMOS technology, bandgap voltage references are typically connected to mosfet gates, which have very high resistance.

The following results were conducted with $R_{LOAD} = 100M\Omega$ and $C_{LOAD} = 200fF$.

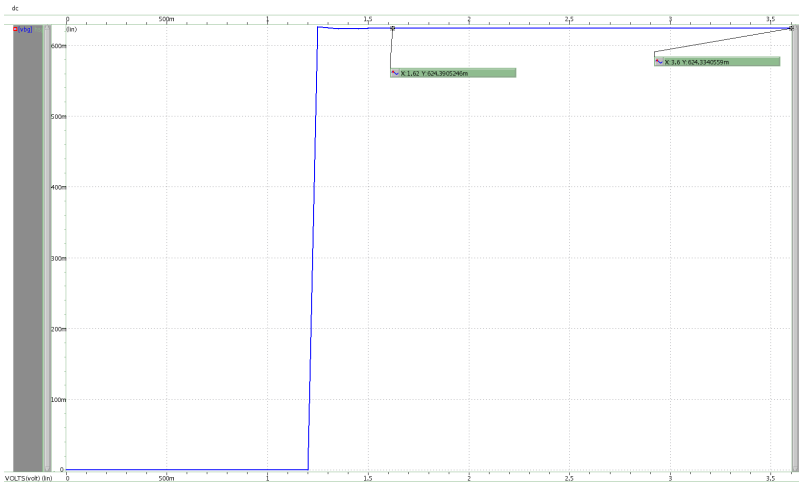


Figure 3.20: Line regulation for typical corners at 27°C

Figure 3.20 shows the output voltage variation over the supply voltage range. It can be seen that the circuit starts operating at a voltage level close to 1.3V as the minimum supply voltage has been achieved. The output voltage remains very constant, almost independent of the input voltage.

From the previous equation we calculate the line regulation coefficient of our circuit as

$$S_{LR,T(nom)} = \frac{0.624334 - 0.624390}{3.63 - 1.62} \approx 27.8uV/V \tag{3.52}$$

Unfortunately, very few articles are used for comparison since most of them did not provide line regulation as a performance parameter for the bandgap circuit. Of the few available, the lowest value found was 0.15% [5], which equals to $750uV/V$ and is much larger than our results.

	$S_{LR,T(nom)}$	$S_{LR,T(nom)}$ (NO CASCODE)	Temperature	Supply Range (V)
Worst Case	40.43 uV/V	-	125°C	1.62 to 3.63
Typical Case	27.8 uV/V	84 uV/V	27°C	1.62 to 3.63
Best Case	27.27 uV/V	-	48°C	1.62 to 3.63

Table 3.8: Line regulation coefficients for worst, typical and best case scenarios at typical corners

Table 3.8 above shows the worst, best and typical scenarios for line regulation under typical corners. It is interesting to note that the line regulation is the lowest close to the inflection point of the temperature curve at 50°C. The table also contains with line regulation with the addition of the cascode transistors, M4-M6, from figure 3.6. It is clear that adding the transistors increases supply rejection as explained before but now proved with an objective measurement term.

Table 3.9 shows the line regulation coefficients under all process corners. Yet again the best case is not the typical case, but like in the previous section, the results are very close. Even the worst case scenario is significantly better than [5] in typical conditions.

	$S_{LR,T(nom)}$	Resistor Corner	Mosfet Corner	Bjt Corner	Temperature
Worst Case	152 uV/V	Fast	Slow-Slow	Fast	20°C
Typical Case	27.8 uV/V	Typical	Typical	Typical	27°C
Best Case	25.6 uV/V	Fast	Slow-Fast	Fast	50°C

Table 3.9: Temperature Coefficients for worst, typical and best cases

3.6.4 PSRR

In a real world implementation, the power rail for our circuit on the silicon is corrupted by the high frequency noise due several effects like signal coupling, feedback, power surges, switching lines and so on. The ability for our bandgap to reject this noise and other undesired signals is referred as power supply rejection ratio, PSRR.

The noise for PSRR measurement can be modeled by using the test bench of the previous section, and simply coupling a signal source on top of our main supply voltage source. The output will be the same voltage reference as expected, but with a small ripple from the signal source. The ratio between the amplitudes of both signals defines the PSRR.

$$PSRR(f) = \frac{V_{bg,f}}{V_{IN,f}} \tag{3.53}$$

Where $V_{bg,f}$ and $V_{IN,f}$ is the output and supply voltage, respectively at a specific frequency. PSRR is usually simulated in dB.

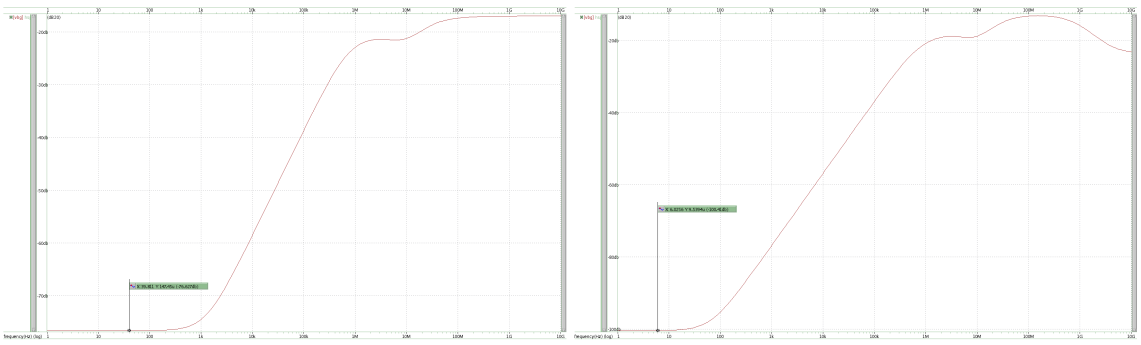


Figure 3.21: PSRR at 27°C and Supply voltage 2V

The figure above shows the magnitude of the power supply rejection ratio before and after the addition of the cascodes transistors M4-M6, respectively. We can clearly see an increase of 25dB with the addition of the cascodes transistors but with the downside of losing bandwidth in terms of rejection.

The PSRR with no cascodes starts to become more viable as we reach higher frequencies, close and beyond 100 kHz. Depending on the applications and the overall system in which this bandgap would be integrated, the cascodes could be changed or even down right removed if the supply noise dominated at high frequencies.

As a term of comparison please take in mind article [19], which states a PSRR of -93dB at DC and -78dB at 1 kHz. The following tables will show the worst, typical and best case scenarios at DC.

	PSRR(dB)	Resistor Corner	Mosfet Corner	Bjt Corner	Temperature	Supply
Worst Case	-45dB	Fast	Slow-Slow	Fast	60°C	1.62V
Typical Case	-100dB	Typical	Typical	Typical	27°C	1.62V
Best Case	-138dB	Typical	Typical	Slow	40°C	2.5V

Table 3.10: PSRR at DC

The following table represents the worst and best cases at 1 kHz.

	PSRR(dB)	Resistor Corner	Mosfet Corner	Bjt Corner	Temperature	Supply
Worst Case	-26dB	Fast	Slow-Slow	Fast	-40°C	1.62V
Typical Case	-40dB	Typical	Typical	Typical	27°C	1.62V
Best Case	-47dB	Fast	Slow-Slow	Fast	-40°C	3.63V

Table 3.11: PSRR at 1 kHz

3.6.5 Overview

The following table provides a summed overview of the parameters of our bandgap, including worst, typical and best cases scenarios. The following results also includes MonteCarlo simulations and will be presented with mean and standard deviation when available.

These results represent the full complete set of simulations and give a total coverage of 99.73%. Temperature range from -40°C to 125°C, supply voltage from 1.62V to 3.63V, all corners and 100 MonteCarlo iterations.

Parameter	Typical	Worst	Best	Mean (u)	Deviation (σ)
Output Voltage (mV)	624.26	600 / 658.7	624.26	624.26	9.76
Temperature Coefficient (ppm/°C)	14.2	60	8.25	-	-
Line Regulation (uV/V)	27.8	204	25	-	-
Current Consumption (uA)	42.578	56.7	33.7	*	*
PSRR (dB) DC	-100	-39	-141	-	-
PSRR (dB) 1kHz	-40	-39	-83	-	-

Table 3.12: Bandgap Voltage Reference Parameters

Current Consumption is dominated by the resistor variation and, as such, the mean will change according to its variation as explained in section 3.6.1. For fast and slow corners, the mean for power consumption is 53.8uA and 35.5uA, respectively.

The total variation of the output voltage featuring the complete set of simulations is $\pm$4.89%.

Parameter	Typical	Worst	Best
Output Voltage (mV)	623.26	589.5 / 657.7	623.26
Temperature Coefficient (ppm/^{o}C)	16.2	60	10.2
Line Regulation (uV/V)	160	1004	75
Current Consumption (uA)	40.47	53	32.5
PSRR (dB) DC	-70	-	-
PSRR (dB) 1kHz	-69	-	-

Table 3.13: Bandgap Voltage Reference Parameters (Post-Layout Simulations)

3.7 Error Sources

The study of the sources of error that introduce inaccuracies in references is extremely important in an environment in which the precision of a system's bandgap reference often dictates its overall accuracy performance. Analyzing these various error sources allows a designer to assess their relative impact on the accuracy of the reference voltage and thereby make important decisions regarding all aspects of the design, such as process technology, circuit topology, trim network and layout.

A number of factors degrade the accuracy of CMOS bandgap reference circuits, including process variations and mismatch, supply fluctuations and load variations. Next sections discuss these various error sources and quantifies their impact on accuracy, focusing on process variation and mismatch.

3.7.1 Process Variation and Mismatch

Process variation and mismatch is often referred as the sources of error that designers have no control over. Their unwanted effects have been mitigated primarily through careful layout, trimming procedures and a few advanced switching techniques.

The analysis of the error sources for our circuit will follow the topology of figure 3.6. This topology is a building block for many other references. The reference voltage generated is given by,

$$V_{bg} = \frac{R_4}{R_3} V_{BE1} + \frac{R_4}{R_1} \frac{kT}{q} ln(8) \tag{3.54}$$

Looking at our circuit, and taking into account the previous equation, there are a few apparent sources of error.

1. Mosfet Mismatch
2. BJT Variation and Mismatch
3. Resistor Mismatch
4. Operational Amplifier Offset

3.7.2 Mosfet Mismatch

This error arises from the a mismatch between the current mirrors of M1-3, which in turn leads to an unwanted deviation from the current ratios between the branches. The mismatch may occur due to variation of the size ratio (W/L), the carrier mobility, μ_p, or the threshold voltage, V_{TH}.

Luckily for us, mosfet variation is not a problem because the feedback system places the same currents between all branches, regardless of the process corner we are. However, the feedback system has no control over the mosfet mismatch and, as such, we only consider this effect for mosfets.

A 3σ current mismatch of 2% is common for smaller devices. Due to the mismatch for the current mirrors, the current flowing through the BJT branches will be slightly different, 2% to exact. This causes an unwanted factor in our reference voltage. Mosfet current mismatch can be modeled by

$$\Delta V_{bg} = \frac{R_4}{R_1}\frac{kT}{q} ln(8)\delta_M \tag{3.55}$$

where δ_M is the variation over the nominal value and ΔV_{bg} is the shift in the output voltage created by this error. The 2% mirror mismatch is directly translated into the output voltage. Matching performance can be improved by increasing the active area and overdrive voltages or by employing dynamic element matching techniques (DEM).

Increasing the active area of mosfets incurs the penalty of increased parasitic capacitances and for dynamic element matching there is the downside of increased noise due to the switching lines.

3.7.3 Resistor Mismatch

Once more we see why the bandgap voltage topology is one of the most used. The output voltage is only based on ratio of resistors and thus, highly tolerant to resistor variation, being only sensitive to resistor mismatch. Fortunately, resistors can be matched to a good degree of accuracy, typically 1% with careful layout [15].

Resistor mismatch directly influences the reference voltage, and is particularly important to study since our topology relies on two ratios between four resistors, three of them with different values.

The output voltage variation due to resistor mismatch can be modeled by

$$\Delta V_{bg} = \frac{\delta_{R4}R_4}{\delta_{R3}R_3}V_{BE1} + \frac{\delta_{R4}R_4}{\delta_{R1}R_1}\frac{kT}{q} ln(8) \tag{3.56}$$

where δ_{R3}, δ_{R4} and δ_{R1} represent the variation against nominal value of the resistor in cause. While absolute variation can go as far as 20% over corners, mismatch values are smaller.

A 3σ resistor mismatch of 2% in R_1 introduces an error of 4.8mV which is fairly significant. This alone causes an error close to 0.6% with only a mismatch between resistors R4 and R1.

3.7.4 BJT Mismatch

BJT mismatch errors result from a deviation in the desired ratio of the saturation current density J_S of transistors Q1 and Q2 [10]. If δ_Q is the variation in the ratio, the error in the reference voltage is given by

$$\Delta V_{bg} = \frac{R_4}{R_3} V_{BE1} + \frac{R_4}{R_1} \frac{kT}{q} ln(\delta_Q 8) \tag{3.57}$$

Since bipolar transistors can be matched to a high degree of accuracy (e.g. 0.1-1%), BJT mismatch has a small effect on the accuracy of the reference voltage. The error due to a mismatch of 1% is only 1mV, or roughly 0.2% for our reference, even after being multiplied by the $\frac{R4}{R1}$ ratio.

3.7.5 BJT Variation

In our case, BJT variation is translated into a shift in the base-emitter voltage, directly affecting the output voltage. The spread in the CTAT can be a considerable amount of error because is dictated by the process in use. The total error introduced can be modeled by

$$\Delta V_{bg} = \frac{R_4}{R_3} \delta V_{BE1} + \frac{R_4}{R_1} \frac{kT}{q} ln(8) \tag{3.58}$$

Simulations show a total variation of 20mV of the base emitter voltage. Our topology has the advantage of multiplying this eventual shift by the ratio $\frac{R4}{R3}$, which reduces this variation to about half since $\frac{R4}{R3} = 0.51$. Nonetheless, this 10.2 mV shift at the output represents a 17.6% overall variation exclusively to BJT variation.

3.7.6 Operational Amplifier Offset

The amplifier's offset is dictated by the non-infinite gain and eventual mismatches between the transistors. And unfortunately for us, the amplifier's offset is directly translated to the output voltage and multiplied by a factor of $\frac{R4}{R1} \approx 4.2$. The error caused by the amplifier's offset can be modeled by

$$V_{bg} = \frac{R_4}{R_3} V_{BE1} + \frac{R_4}{R_1} (\frac{kT}{q} ln(8) + V_{off}) \tag{3.59}$$

The amplifier's offset voltage can be easily simulated by adding an ideal voltage source between the non-inverting terminal of the amplifier and node V_b. Simulations are executed by ramping up the voltage source slowly up to 4 mV and watching the output for variation.

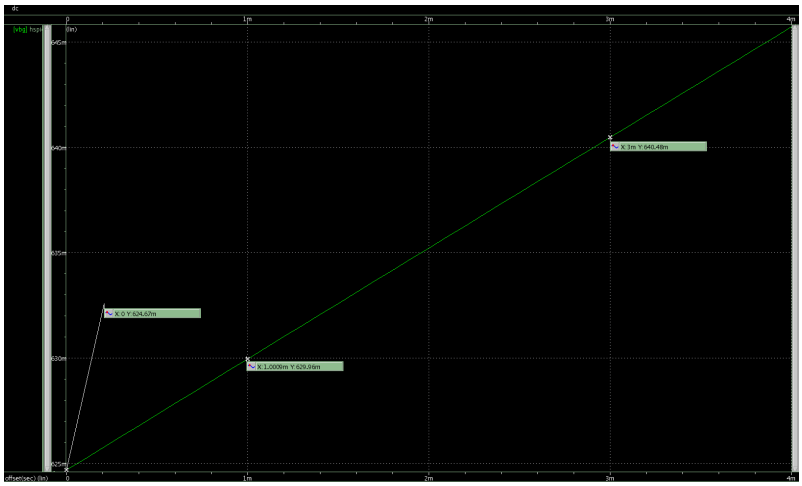


Figure 3.22: Output variation due to offset voltage

Figure 3.22 shows the simulated values. The output voltage varies linearly with the increase of the amplifier's offset voltage, and follows the predetermined slope fairly accurately. Extracting the values from the figure we see a slope close to 5.2 against the theorized 4.2 of the resistor ratio.

Our amplifier has $\pm 4mV$ offset in the worst case scenario. This 4 mV would be multiplied by the $\frac{R4}{R1}$ ratio, leading up to 16.8 mV overall error at the output. This represents 2.7% error at the output and is the dominating factor in the overall error sources.

There are several techniques to minimize and even reduce the offset to insignificant values. Techniques such as auto-zero and chopping techniques are regularly applied in integrated circuits for offset cancellation. Due to timing constraints, these were not applied but recommended for future work.

3.7.7 Overview

Type/Source of Error	Error Introduced	Analytical $\Delta V_{BG}(mV)$	Simulated $\Delta V_{BG}(mV)$
MOS Mismatch	2%	4.53 mV	6.92 mV
Resistor Mismatch (R1)	1%	2.28 mV	2.3 mV
Resistor Mismatch (R2/R3)	1%	3.9 mV	3.89 mV
Resistor Mismatch (R4)	1%	6.15 mV	7.06 mV
BJT Variation	20 mV	10.2 mV	10.24 mV
BJT Mismatch	1 mV	4.2 mV	4.49 mV
Amplifier Offset	2 mV	8.4 mV	10.55 mV

Table 3.14: Sources of Error. Analytical Values vs. Simulated Values

Table 3.14 shows a comparison between the analytical values and the simulated ones for manually introduced sources of error. It is clear that the presented equations for the analytical

values can model the impact of the several sources of error with very good approximation, the larger discrepancies belonging to the MOS mismatch modeling equation and Amplifier offset.

Type/Source of Error	Worst Case Deviation	Error Produced	Overall Error Contribution
MOS Mismatch	100 nA	4.7 mV	8.1%
Resistor Mismatch (R1)	216Ω	4.35 mV	7.5%
Resistor Mismatch (R2=R3)	621Ω	2.65 mV	4.5%
Resistor Mismatch (R4)	441Ω	5.65 mV	9.7%
BJT Variation	20 mV	10.2 mV	17.6%
BJT Mismatch	265 uV	1.13 mV	1.9%
Amplifier Offset	4 mV	21 mV	36.2%
Other Sources	-	-	10%

Table 3.15: Relative magnitude of the various sources in the overall error

Table 3.15 lists the biggest error occurred in each of the sources, the variation provoked at the output and the general contribution to the overall error of the output voltage. Analyzing the table is clear that the voltage variation is mostly due to the amplifier offset, that represents over a third of the total error generated over PVT, followed by the variation of the bipolar transistors.

The total error introduced by the several sources is given by

$$Error = \delta_{MOS} + \delta_{R1} + 2 * \delta_{R2} + \delta_{R4} + \delta_{Q} + \delta_{Q1,2} + \delta_{OFFSET} = 90\% \tag{3.60}$$

The total error sources listed equal to 90% of the total error introduced in the output voltage. The remaining 10% of error accounts for the noise of the circuit, and the eventual changes with temperature and voltage. Even a bandgap shows variation over temperature and supply, however small. Not to mention that in this work it was only done first order corrections.

3.8 Layout Considerations

When one designs analog circuits, several important layout issues should be considered to realize high-quality circuits. Firstly, the layout should be designed to avoid sources of systematic variation which will cause device parameters to deviate from their expected values. Second, good layout practices will minimize noise and outside interference in the circuit. And third, special layout procedures can and should be applied to reduce device mismatch between critical components.

Several techniques were adopted during the layout design.

1. All transistors were lined up in the same direction.
2. Big contacts were used over the active areas to ensure good connections.
3. Multi-gate finger mosfets for smaller parasitic capacitances.
4. Dummy polysilicon material at the terminations of the mosfets.

5. Interdigitated currents mirrors and resistors for 1d matching.

6. Common-centroid configuration for bjt devices and PMOS input differential pair for 2d matching.

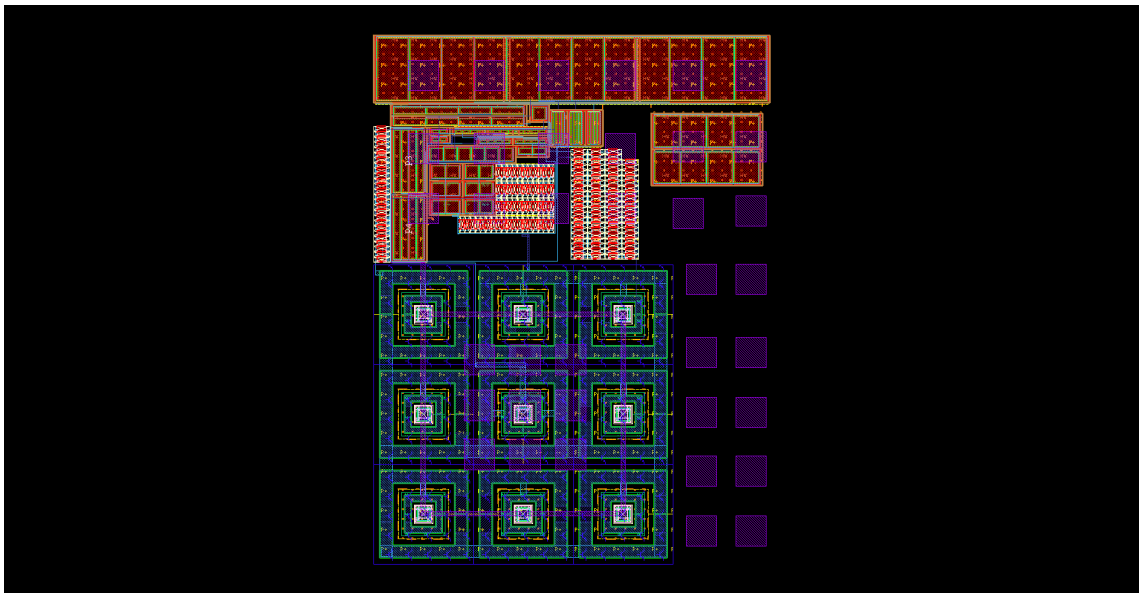


Figure 3.23: Complete Layout of the Bandgap Voltage Reference

This figure shows the complete Layout of the Bandgap Voltage Reference. Dummy transistors and metal layers added due to minimum density constraints.

Individual layout of several cells for the folded cascode amplifier can be found in appendix D and for the bandgap voltage reference in appendix E.

3.9 Summary

In this chapter, a voltage reference based on the bandgap topology was presented along with constraints and the implementation results. The proposed topology was analyzed both theoretically and backed up with results based on extensive simulations. Finally, the various results and error sources are discussed and analyzed in detail for future work.

The used topology is based on the sum of opposing temperature scaling currents in order to eliminate the temperature dependency of the output voltage. It features the addition of mosfet cascodes for higher noise reduction and increased supply rejection, showing a greater stability over supply fluctuations.

Simulations under typical conditions shows a very low temperature coefficient for a standard first order correction of temperature.

It was shown why the bandgap topology is heavily used in microelectronics. This topology shows a high insensitivity to mosfet and resistor variations. The results showed a maximum of

2.6% variation over process corners, proving that the topology is very resilient to process variation. The dominant factor in process variation if the BJT drift.

Unfortunately, it was demonstrated that this topology is heavily affected by device mismatch. It was shown that the biggest source of error was the random amplifier offset, responsible for nearly 40% of the overall voltage drift.

Due to the precision achieved in this chapter, this voltage reference would mostly be suited as voltage reference for low-dropout regulators that supply voltage to other applications. Usage in high precision would require additional improvements.

Chapter 4

Current Reference Design

Analog signals can be represented in either voltages or currents. The current mode approach is chosen mainly for the implementation of analog circuitry for the reason that many functions are easier to implement in current mode. For example, arithmetic operations such as addition, subtraction and scaling require operational amplifiers and resistors in voltage mode and so, voltage mode usually consumes more power and area. The current mode approach has advantages that it requires less hardware than its counterpart. It uses no operational amplifiers or resistors, which are very difficult to implement accurately in CMOS technology, but resorts to simple current mirrors.

Another advantage is speed. Processing current signals is done faster than voltage signals and, therefore, for a given technology, analog circuits designed in current mode operate faster than their voltage mode counter parts. Also, the performance of the current sources influences the power consumption, which is an important parameter in many applications.

Current references are needed for many major building blocks in analog circuits, such as operational amplifiers, analog buffers, oscillators, phase-locked loops, analog-to-digital converters and vice-versa. The performance of these systems is closely tied to accuracy and stability of the bias current.

The most common way for generating a stable voltage across PVT conditions is through the use of the bandgap topology as detailed in the previous chapter.

Current references, however, are specially difficult to produce with accuracy due to the nature of the output. Current is not an intrinsic parameter of any material contrary to the bandgap voltage of the silicon.

There are several types of current references available and are commonly used in circuit designs. One type of current references is a PTAT reference. It produces a current which is linearly proportional to temperature, much like the PTAT voltage for our bandgap. There is also their equivalent, the CTAT current source that exhibits the opposite behaviour and when summed provide a temperature independent current.

Square PTAT current sources are another useful type of current references and usually used for second order compensation.

Another current reference is based on the ZTC point of the mosfet as explained in section 2.1.3. However, the ZTC point is difficult to find with accuracy and is heavily impacted by process variations.

In this chapter, it will be presented a current reference based on the bandgap voltage reference from the previous chapter. The topology chosen will described and compensating techniques for process variation will be discussed.

4.1 Analisys of the voltage to current converter topology

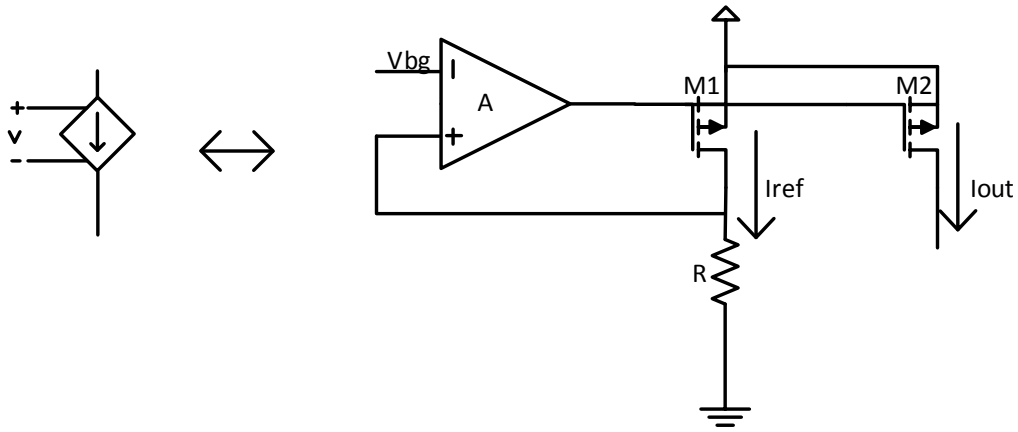


Figure 4.1: Voltage to Current Converter

The topology used for the voltage to current converter is shown in figure 4.1. This topology utilizes an operational amplifier and a p-type transistor to provide a voltage across a resistor R. A reference voltage, in our case the bandgap voltage reference from previous chapter, is applied to the inverting terminal of the amplifier.

The non-inverting terminal of the amplifier is tied to the resistor. The feedback action of the amplifier will force both terminals to equal voltages. If we take into account the eventual offset of the amplifier, then the voltage across the resistor R will be

$$V_R = V_{bg} \pm V_{offset} \tag{4.1}$$

This voltage produces a current that will be determined solely by the reference voltage, the value of the resistors and the eventual offset of the amplifier. If we discard the offset of the amplifier, the total current flowing through both the transistor M1 and the resistor R will be

$$I_{REF} = \frac{V_{bg}}{R} \tag{4.2}$$

The feedback system guarantees independence to process variation in regards to the transistors. However, the current will be dependent on process variations of the resistor R, that in CMOS technology can go up to 25% at process corners.

Since the bandgap voltage is fairly independent of PVT we need to focus on compensating the resistor R. Resistor R is independent of the supply voltage and shows a small temperature coefficient.

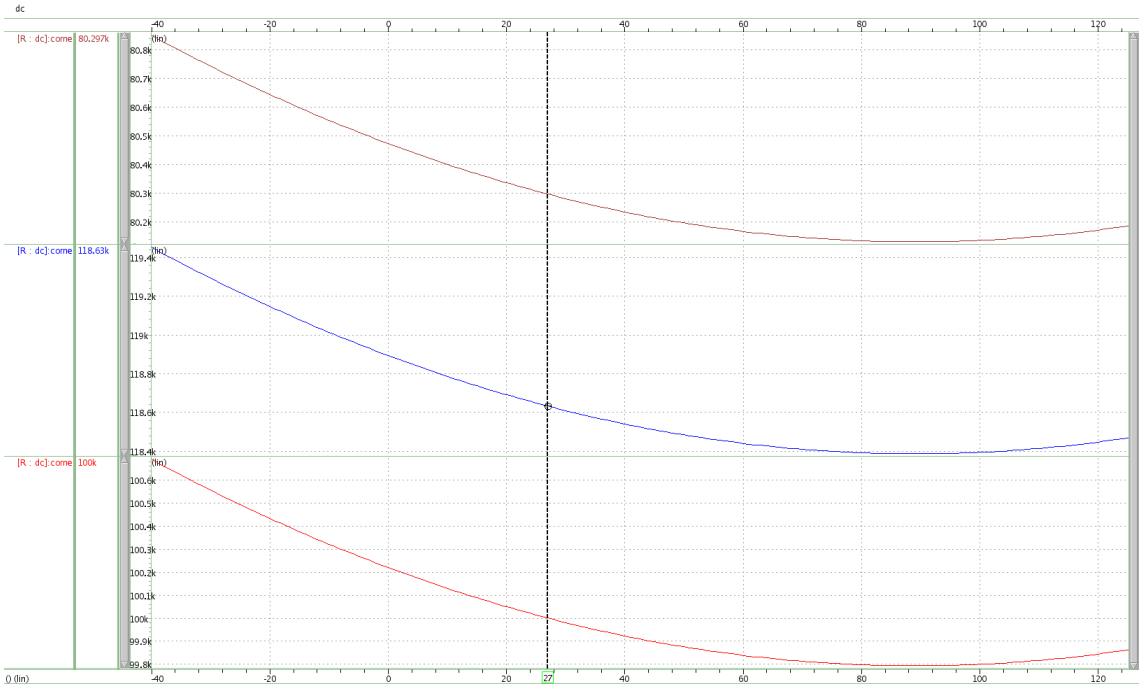


Figure 4.2: 100$k\Omega$ resistor temperature variation at fast, slow and typical corners

Figure 4.2 shows the temperature behavior of the resistors at all corners. It is clear that the variation with temperature of the polysilicon type resistors is small, around 0.898%, and is independent of process variations. However, the absolute value of the resistor itself varies heavily with process, reaching nearly 50% between fast and slow corners.

Another thing to note is that the resistance value diminishes with temperature, showing an opposite behavior to that of our bandgap voltage which increases with temperature. Although the inflection point of the resistor and bandgap voltage are different, they might eliminate temperature curvature to some degree.

4.2 Design of the voltage to current converter topology

In this section we will go through the necessary steps and decisions made in regards to the design of the voltage to current converter. We will be reusing the operational amplifier designed for the bandgap core in the previous chapter. For the amplifier's parameters check section 3.3.

This topology has a straight forward design procedure for initial simulations. We first need to choose the value of the output current. Since the output current is extrapolated through the ohm's law and we have a 'fixed' reference voltage we need to choose the value of the resistor.

Designing circuits with low power consumption is a priority, but without compromising area. Since we do not have a specified output current in mind, it was chosen an output current of $5uA$, which is small. Following ohm's law, the equivalent resistor is then given by

$$I_{REF} = \frac{0.625mV}{5uA} = 125k\Omega \tag{4.3}$$

where 0.625 mV corresponds to the typical value of the bandgap voltage.

Now that we have an output current in mind, we can size transistors M1 and M2 accordingly. Remember that out circuit works with feedback. The operational amplifier will strive to make $V_R = V_{bg}$ by changing the current across resistor R. These currents variations are produced by changing the gate voltage of the PMOS.

Observing our circuit, the current is controlled by the V_{SG} of the PMOS. We need to choose an appropriate gate voltage that is higher than the threshold voltage, $|V_{TH}| > 600mV$, as seen in figure 2.2, but still remains within 0V and our supply voltage, $0 < V_G < SupplyVoltage$. Choosing $V_G = 0.7V$ fulfills the requirements.

Now we can size our PMOS accordingly. Considering the mosfet are working in saturation region and ignoring second-order effects, the current is given by

$$I_d = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{SG} - |V_{TH}|)^2 \tag{4.4}$$

where the factor $\mu_n C_{ox}$ is called Beta (β), W is width of the transistor, L stands for the length, V_{SG} is the source-gate voltage and V_{TH} is the threshold voltage.

Solving for $I_d = 5uA$,

$$5uA = \frac{\beta}{2}\frac{W}{L}(1.62 - 0.7 - 0.6)^2 \tag{4.5}$$

$$\frac{W}{L} \approx 1.88 \tag{4.6}$$

By choosing a W/L ratio close to 1.88 we achieve the desired current all the while having a decent gate voltage at the PMOS. As discussed in the previous chapter, we will be using non-minimal dimension to avoid short-channel effects. These mosfets were sized with 10u/5u for lower mismatch error and increased circuit stability due to increased parasitic capacitances. 3.3.2.2

4.2.1 Programmable Resistor Network

As we discussed in the previous section, most of the error introduced in the output current will be from resistor variation over process. We have seen that temperature variation is small and supply variation is negligible.

The solution chosen for this problem was using a programmable resistor network. Like the names states, this resistor can have its overall value changed according to the necessary value.

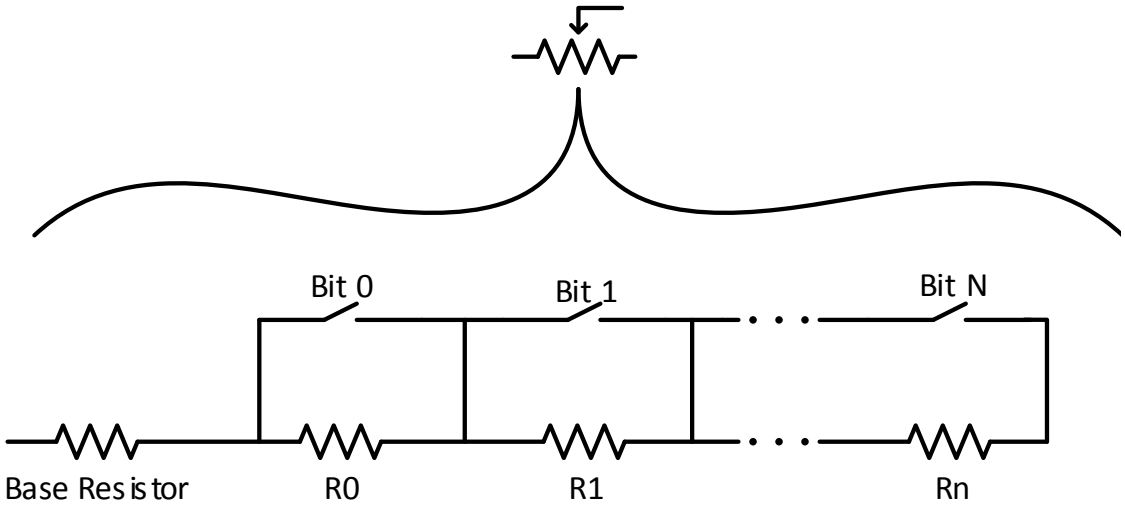


Figure 4.3: Programmable resistor topology

Figure 4.3 shows the topology for a programmable resistor network. This topology works by closing or opening the switches. These switches, under ideal conditions, have 0 resistance once closed and infinite resistance when open.

Depending on the combination of the bits, the overall resistance can be the base resistor, if all switches are closed, or the sum of the base resistors plus the resistor whose switch is open. The overall resistance value is programmable and the precision is limited by the number of bits and the resistance associated with the less significant bit, bit 0.

Take for example a base resistor of $100k\Omega$ with ten bits, each bit having an associated resistor with the equal value, $5k\Omega$. If the switches are ideal and all the switches are closed, then the current will flow through the path of least resistance (i.e., ideal switches have 0Ω resistance once closed) and thus the overall resistance will be equal to the base resistor, $100k\Omega$.

However, if half of the switches are open, the current will alternate between passing through the resistor, if the switch is open, or through the switch if it is closed. Thus, with half the switches closed, the overall resistance is equal to $100k\Omega + 5 * 5k\Omega = 125k\Omega$.

Now that we understand how the programmable resistor works, we can determine the number of bits, smaller resistor value and calculate the overall precision of the system.

The overall variation over PVT is dominated by the resistor variation and is close to 50% under extreme conditions. It was chosen a precision of 1% over the PVT range. The number of bits necessary for reaching the required precision is determined by

$$2^N = \frac{\Delta_{MAX}}{\Delta_{WISHED}} \tag{4.7}$$

where N is the number of bits, Δ_{MAX} is the overall variation of the system and Δ_{WISHED} is desired final variation.

Replacing the values, solving for N and performing change-of-base of the logarithmic to the natural logarithm we get

$$N = \log_2(\frac{50}{1}) \approx 5.64 = 6bits \tag{4.8}$$

With 5.64 bits we would be capable of achieving the desired 1% precision. However, the number of bits needs to be an integer value and in order to reach the minimum of 1% precision the number of bits was rounded up. With 6 bits the expected precision is 0.78%.

With 6 bits we have 64 levels between all possible combinations. We need special care in choosing both the base resistor since we must ensure that the base resistor will never, in any process variation, be able to surpass the desired value of $125k\Omega$.

Luckily, the process corners represents the absolute, extreme process variation that a device can reach. For a $100k\Omega$ resistor, the slow corner can go up to a maximum of $119.4k\Omega$, at -40°C, as seen in 4.2, proving that this base resistor has an adequate value.

The next step would be sizing the individual resistors themselves, each belonging to one bit. The final desired value for our resistor is $125k\Omega$ and our base resistor can, in the absolute worst possible case, hit a minimum of $80.1k\Omega$, close to 90°C at the fast corner.

This would give a range of $125k\Omega - 80.1k\Omega \approx 45k\Omega$ total resistance for our 6 bit resistors. However, not only will the base resistor change with process variations, but so will the individual resistors we use, so we need to compensate for that as well.

Using a 25% upper bound of corner variation, the total resistance of the combined 6 bits needs to be $1.25 * 45k\Omega \approx 56k\Omega$. Giving a small margin for eventual device mismatch and the total resistance was chosen as $100k\Omega$ base resistor plus $60k\Omega$ programmable resistor.

The smaller resistance, associated with the less significant bit, is given by

$$R_0 = \frac{60k\Omega}{2^6} \approx 940\Omega \tag{4.9}$$

Every single successive bit will be the double of the previous bit. The following table represents the bit and their corresponding resistance.

	Resistance (Ω)
Base Resistor	100k
Bit 0	940
Bit 1	1.88k
Bit 2	3.76k
Bit 3	7.52k
Bit 4	15k
Bit 5	30k

4.2.2 Non-Ideal Switches

In a practical implementation of the resistor network, our switches will not be ideal. And by ideal we consider having 0Ω resistance when closed and infinite resistance when open.

In CMOS technology, switches are implemented with mosfets operating in triode region. And while the off resistance is very high, near the $G\Omega$ range, the resistance once the switch is closed is not 0Ω. Typical values for minimum sized mosfets have an on resistance in the order of the $k\Omega$ which is very detrimental if our resistances have similar values. If the closed switch has a resistance of $10k\Omega$ then, for example bit 4, would be affected heavily. The total resistance associated with R4 and the closed switch would be

$$R_T = R4||10k\Omega \approx 6k\Omega \tag{4.10}$$

which would severely degrade the precision of the planned resistor network.

The triode region of a mosfet is characterized by the channel being formed, $V_{GS} > V_{TH}$, but the the drain-source voltage is inferior to the override voltage, $V_{DS} < V_{GS} - V_{TH}$. When a mosfet operates in this region, it behaves similarly to a voltage controlled resistor.

In triode region, the behavior of a mosfet can be modeled by

$$R_{DS} = \frac{1}{u_{n/p} C_{OX} (\frac{W}{L})(V_{OV} - V_{DS})} \tag{4.11}$$

where V_{OV} is the override voltage, $V_{GS} - V_{TH}$.

The total resistance is then given by the parameters of the device and the voltage levels. Just from looking at equation 4.11, it can be inferred that it is better if n-type mosfets are used since their $\mu_n C_{OX}$ is about four times bigger than p-type mosfets, giving a on resistance 4 times smaller under the same size and voltage levels.

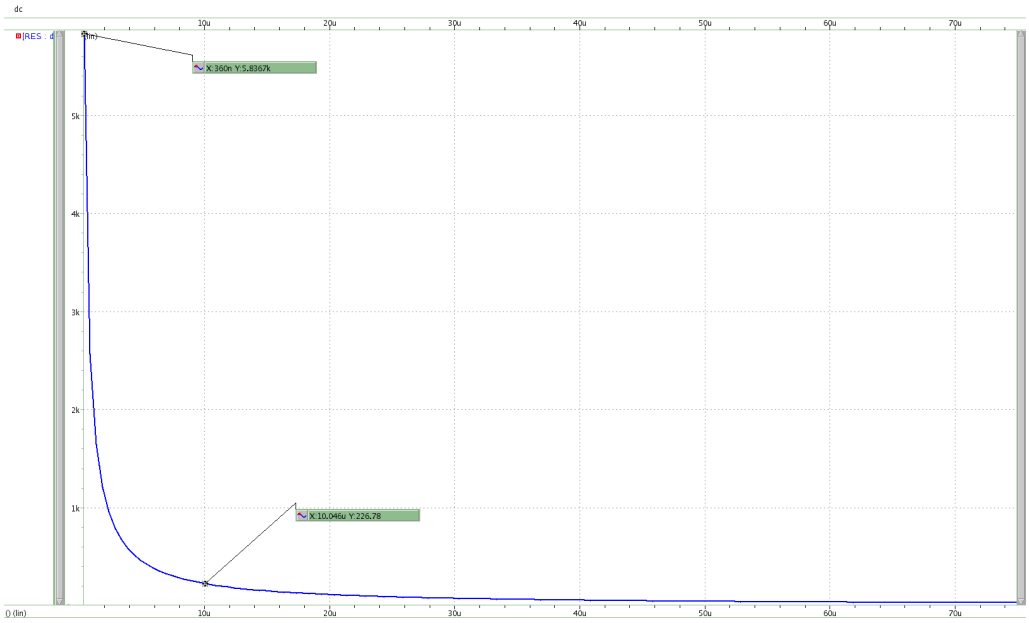


Figure 4.4: On Resistance variation with increasing W at typical corners

Figure 4.4 shows the on resistance variation as the width of the transistor is increased but the length remains fixed at his lowest value, 0.27um. The figure proves the validity of the model used.

$$R_{DS} = \frac{1}{\beta(\frac{10u}{0.27u})(1.62-0.54-0.6)} \approx 287\Omega \tag{4.12}$$

which is in the same order of magnitude but with a 27% relative difference to the highlighted value of a n-type transistor with W = 10u and L=0.27u since we are not considering short channel effects and the device's length is the lowest possible. In order to minimize area consumption, our switches were sized with the sames values. The on resistance is now small enough for the impact to be negligible.

Regarding process, temperature and supply variations of our switches. These variations would manifest themselves in the resistance of the device. However, these variations are not impactful to the point of becoming significant in regards to the nominal value of the resistance in typical conditions.

4.2.3 Calibration System

In this section we will be discussing how our calibration is performed. Calibration refers to determining the ideal combination of bits that would provide the lowest variation at the output in regards to a known magnitude of reference.

Our calibration system overview can be seen in the figure below.

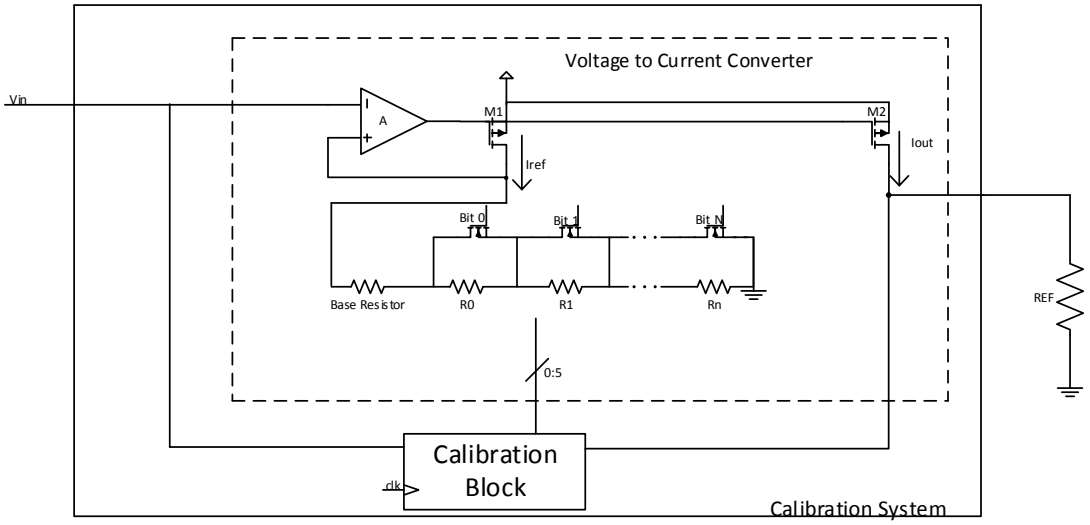


Figure 4.5: Voltage to Current Converter System

Looking from the top view, the complete current reference could be considered a black box with only two signals. We have the input voltage from our bandgap reference and the system provides a single terminal for the current output. With no load connected to it, the output terminal is floating and no current flows through it.

Our calibration systems work by comparing the input voltage against the voltage generated at the output node when a reference resistor of known value is connected to it. The current flowing through the output branch is

$$I_{OUT} = \frac{V_{in}}{R_{CAL}} \tag{4.13}$$

This current would flow through the reference resistor and generate a voltage given by

$$V_{OUT} = \frac{V_{in}}{R_{CAL}} * R_{REF} \tag{4.14}$$

Since the reference resistor is known and assumed to be fixed, the output voltage variations would be solely based on the PVT drift of the internal resistor, R_{CAL}, we wish to calibrate. By comparing these two voltages we can actuate on the internal resistor until this difference is zero, completing our calibration procedure.

There are two alternatives to the calibration procedure that we can employ in our system.

- Post-Fabrication Calibration
- Continuous Calibration

Post-Fabrication Calibration is done only once. As soon as the device has been produced, the reference resistor is attached and the calibration is executed under certain predefined conditions, like supply voltage 2.5V and temperature 25^{o}C.

This method has the advantage that after the calibration is done, the reference resistor is removed and will no longer be necessary thus reducing area. Another benefit is the lower power consumption since the calibration is not running all the time.

However, this method comes with the cost of lower accuracy regarding non process variation. Even after calibration, the internal resistor will change with temperature. This change, albeit very small, is still present and adds a small variation to the output result.

The other alternative is the continuous calibration. In this setup, the calibration algorithm is running all the time and continuously checking both voltage levels and producing the corresponding binary code for the resistor.

While having the positive side of high accuracy, due to even temperature changes being calibrated, it comes to the downside of needing a permanent reference resistor. Not to mention that the power consumption is also higher due to the calibration system not being shut down.

	Post-Fabrication Calibration	Continuous Calibration
Area	-	+
Current Consumption	-	+
Overall Variation	+	-

Table 4.1: Comparison between calibration methods

Table 4.1 shows a summary of the best characteristics of each calibration approach.

It was decided to choose the Post-Fabrication Calibration not only because of the reduced area and power consumption but also for the precision as well. The intended calibration was designed for a maximum precision of 0.78%. However, the temperature variation of the resistor is fairly small, close to 0.9% as seen in figure 4.2.

Since the temperature variation is not much bigger than the precision of we can attain, the overall benefits of the Post-Fabrication Calibration seem the best option to take.

4.2.4 Calibration Algorithm

An important part of the whole calibration system is the algorithm that rules how this calibration is done. The calibration system itself is represented in figure 4.5 by the block "Calibration Block". This block compares both voltage levels and acts on the calibration resistor until a lock state has been achieved.

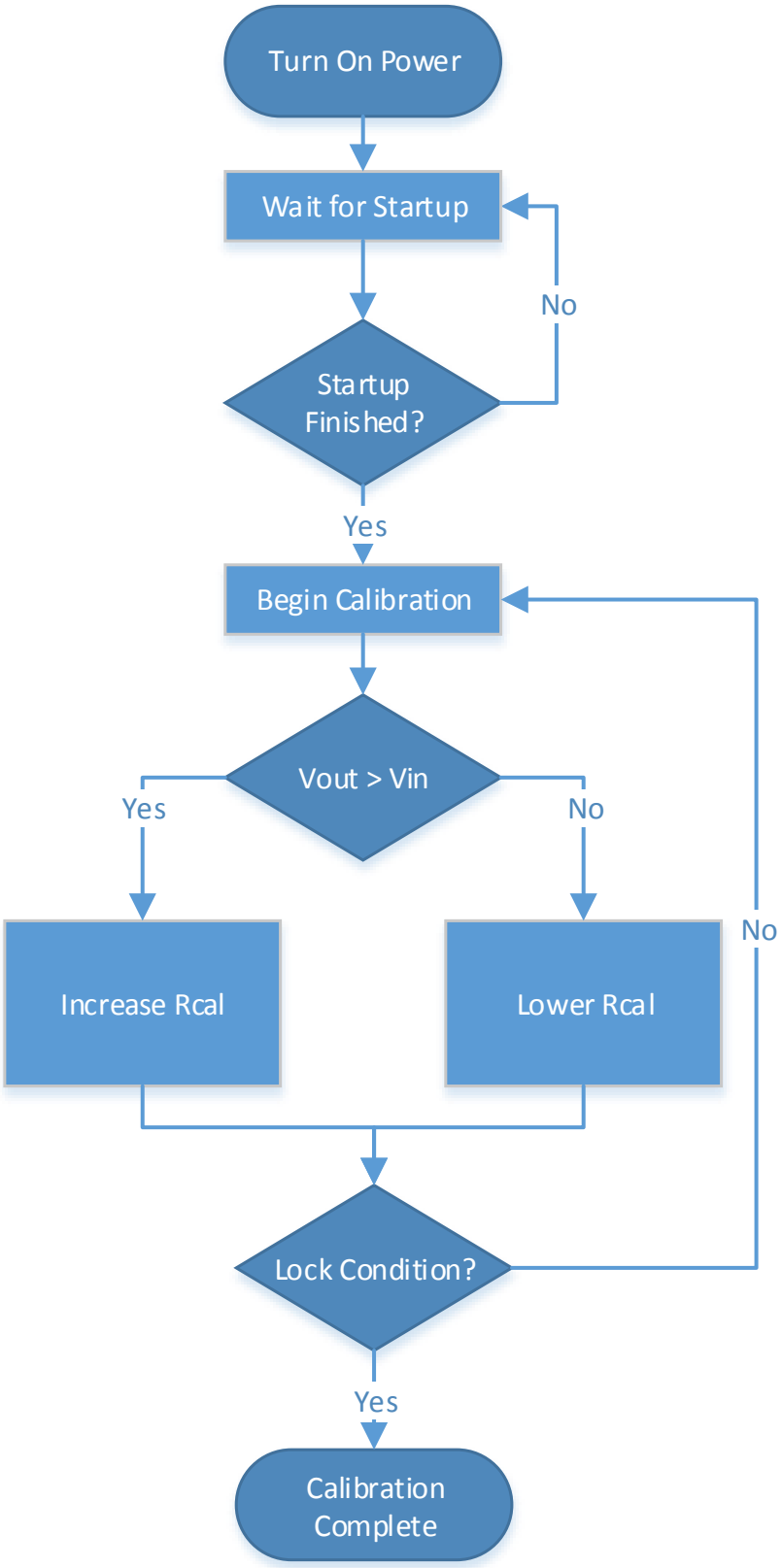


Figure 4.6: Calibration Algorithm

Figure 4.6 shows the calibration algorithm "inside" the calibration logic block shown in figure 4.5.

In a practical application, the power supply does not start at the minimum voltage considered in this work, 1.62V. As explained in section 3.4, the device requires a minimum time for the output voltage to settle in the correct operating point. Looking back at figure 3.14, we will consider a startup time of 60us.

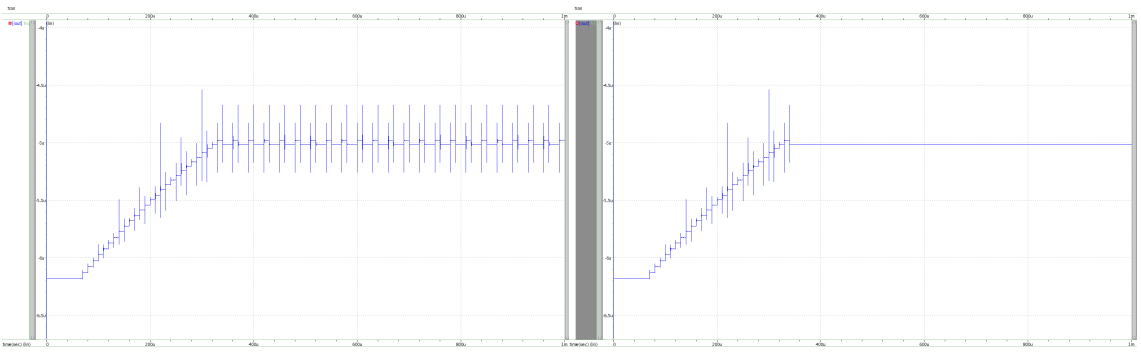
Once the output voltage of the bandgap as settled we can begin the calibration process. With 6 bits, we have a total of 64 combinations. Since most the devices produced will be in the typical region, the calibration code will begin in the middle of the total range which is 32. A calibration code of 32 is 100000 in binary, which correspond to turning on the most significant bit, bit 5, on as default.

The system now begins the effective comparison between the two voltages levels. For simplistic reasons, for the eventual translation of the verilog-A module into physical devices, the comparison was done by simple logic values instead of measuring the absolute values of the voltage levels.

If the output voltage is bigger than our reference voltage, we lower the calibration bits. Which in turn translates into increasing the calibration resistor, causing less current to flow to the reference resistor and thus lowering the voltage.

If the opposite is true, the output voltage being smaller than the reference voltage, we increase the calibration code, effectively lowering the calibration resistor and increasing the current. The increased current will increase the voltage across the reference resistor.

A problem with this type of calibration method is that the system will never settle into one correct value. The output voltage will never be equal to the reference voltage which will cause the output current to oscillate between two values.



(a) Transient response without lock condition (b) Transient response with lock condition

Figure 4.7: Transient response at typical corner, supply voltage 2.5V and room temperature.

Figure 4.7 shows the transient response of the calibration system with the implementation of the lock condition. It is clear that the plot proves the algorithm detailed before.

We have a small window of time in which the system does nothing and waits for the startup time to be complete. After that, the calibration begins and the current starts lowering until it has reached the desired value, close to 5uA, which represents the desired output.

Figure 4.7a shows the output oscillation due the lack of a locking mechanism that shuts down the calibration cycle. Without the locking condition, the output current would fluctuated between the two closest calibration codes.

Figure 4.7b shows the output oscillation with the lock condition implemented. It is clear that the output current locks into a particular calibration code of the two determined in the oscillating state.

The locking condition is achieved by storing the two previous calibration codes. If the third iteration of the calibration mechanism is equal to the first of the two stored codes, then a lock state has been achieved.

In order to simplify the eventual translation of the Verilog-A module into hardware, the effective calibration code chosen is the previous value. Once the final calibration code is chosen, the calibration mechanism shuts down.

The verilog-A module created can be found in appendix B.1.

4.2.5 Calibration Block

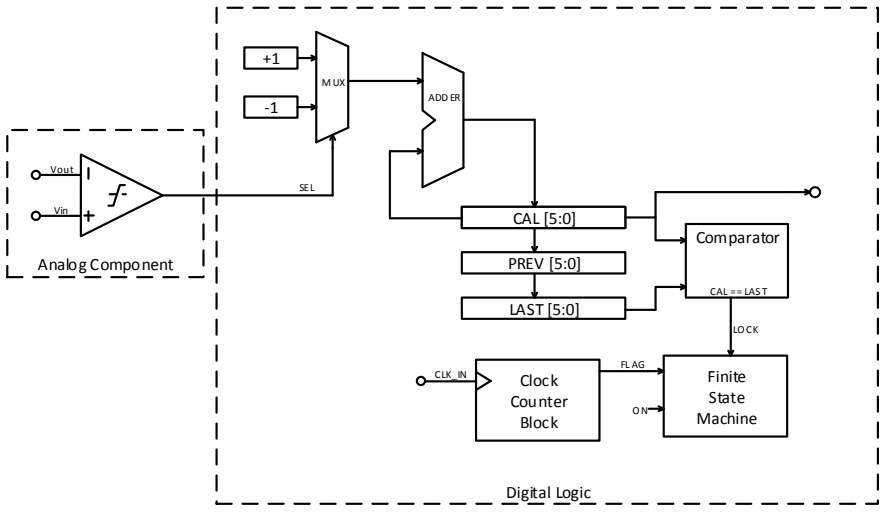


Figure 4.8: Calibration Block Module

In this section, we will be showing a translation of the verilog-A calibration module into hardware blocks. Figure 4.8 show the block diagram of our calibration module. This is a zoomed in view of the Calibration Block in figure 4.5.

The block view can be split into an analog component, which is a single voltage comparator between the input and output voltages, and the digital logic blocks. The output of the voltage comparator controls a multiplexer which chooses which value to add to the register CAL.

If the output voltage is bigger than the reference we add -1, but if the opposite is true, we add +1. The six bit register CAL stores the effective calibration code which will target our programmable resistor network. The values of the two previous iterations are also stored in the PREV and LAST registers.

A digital comparator is used to compare the two registers, CAL and LAST. If they have the same value then the lock condition has been achieved and the state machines locks our system in the desired calibration code.

Looking into the clock counter, the clock signal can have any frequency, providing the system actually works with the chosen frequency. We use this clock to measure the time we need for the bandgap voltage reference to startup, but also for the output current to settle every time the calibration code has been changed. If the clock signal is slow enough, then a single cycle is enough for the current to settle and we no longer require the block counter.

The clock signal can be realized with several topologies and does not need any particular accuracy. By identifying the worst case of the frequency (i.e., the fastest frequency achievable), we can upper bound the amount of cycles we need to wait. This clock can easily be designed using simple ring oscillator topologies with the downside of needing a register to count the cycles, or it can be done by using a mono-stable oscillator which triggers after a certain known time as passed.

4.3 Results

In this section, we will be presenting the results of the voltage to current converter with both an ideal voltage source to estimate the amount of error introduced by the converter alone, and with our bandgap voltage source as a reference voltage for the converter.

In this section it will also be included the calibration codes for several corners and the system's behavior after the calibration has been finished.

4.3.1 Current Consumption

Much like the power consumption of the bandgap voltage reference, the power consumption of the voltage to current converter can be calculated by the following sum.

$$I_{TOTAL} = I_{AMP} + I_{BIAS} + 2I_{PMOS} \tag{4.15}$$

where I_{AMP} is the current consumed by the amplifier, I_{BIAS} represents the current used by the bias circuit of the amplifier and finally, I_{PMOS} represents the current reference we wish to generate at the output.

Since we use the same amplifier, the same bias circuit and both PMOS have a calculated current of 5uA, after calibration, the total current consumption is given by

$$I_{VI} = 815n + 2 \times 811n + 2 \times 5u \approx 12.4uA \tag{4.16}$$

which is a very small current consumption. Lowering this value would be possible by simply increasing the resistors values with the downside of spending more area.

If we take into account our bandgap voltage reference, the total current consumption is achieved by adding the two values, taking special care to make sure the bias circuit is only taken into account once. This is because the same bias circuit is used for both amplifiers.

$$I_{TOTAL} = I_{BANDGAP} + I_{CONVERTER} \tag{4.17}$$

$$I_{TOTAL} = 42.578u + 815n + 2 \times 5u \approx 53.4uA \tag{4.18}$$

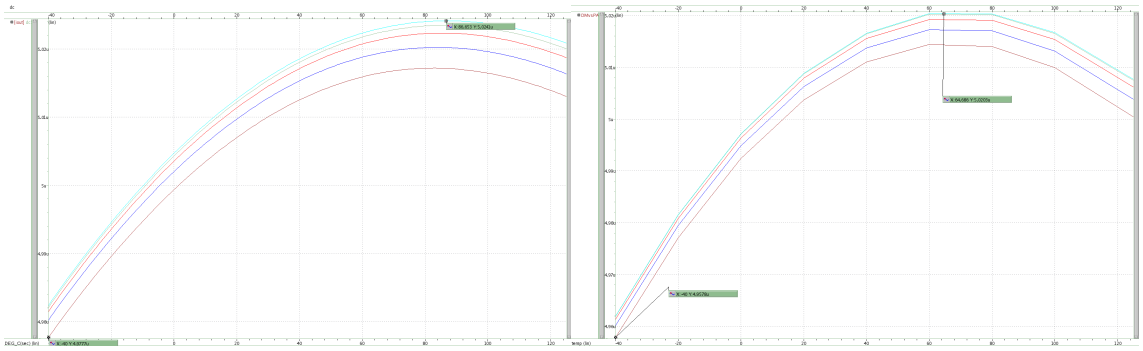
Total Current (After Calibration)	Calculated	Simulated
Without Bandgap	12.4uA	11.268uA
With Bandgap	53.4uA	52.25uA

Table 4.2: Calculated current usage vs. Simulated value

Since the calibrated current dominates over the smaller values (i.e., 10uA > 800nA), and process variations of the amplifier and its bias circuit are small in regards to power consumption, we can say that the total power consumption of voltage to current converter will always be close to 11uA.

4.3.2 Post-Calibration Variation

As explained before, we perform the calibration only once under certain conditions. This means that the calibrated resistor will change with temperature, and the output current will vary slightly due the changing output voltage of the load and the supply voltage itself, due to the channel modulation effect on the output transistor.



(a) With ideal voltage source (b) With bandgap source

Figure 4.9: Post-Calibration variation with temperature and supply

Figure 4.9a shows the variation of the output current with both temperature and supply after calibration has been achieved. The temperature variation is as expected and the current also shows a small shift over supply variation.

The total variation post calibration is under 0.95% from the converter alone. These measurements were performed with an ideal voltage source, since our bandgap also changes with temperature and supply, albeit a smaller variation.

Looking at the side figure, 4.9b, which shows the same variation but with the bandgap source as a reference voltage. We note a small increase from 0.95% of the converter to 1.26% due to the small shift of the bandgap reference. The bandgap variation "piles" on top of the converter variation in typical conditions.

	Variation (%)	Resistor Corner	MOS Corner	BJT Corner	Temp (^{o}C)	Supply (V)
Worst Case	2.40	Fast	Any	Any	-40 to 125	1.62 to 3.63
Typical Case	2.10	Typical	Any	Any	-40 to 125	1.62 to 3.63
Best Case	1.87	Slow	Any	Any	-40 to 125	1.62 to 3.63

Table 4.3: Variation with bandgap reference for worst, typical and best corners

Table 4.3 shows the overall behavior under the worst, typical and best conditions. The different variation from the previous value in typical stems from the bandgap reference voltage. The voltage to current converter has no control over the variation of the bandgap voltage and that is clearly seen in the table. The variation in each case is centered on the corners of the resistors, since it is the only variation we are calibrating in the converter.

The additional variation comes from the corner variation of the bandgap reference.

	Current Variation(uA)	Resistor Corner	MOS Corner	BJT Corner	MC Seed	MC Iteration	Temp (°C)	Supply (V)
Higher V	5,0739 to 5,1673	Fast	Any	Any	1	93	-40 to 125	1.62 to 3.63
Lower V	4,7141 to 4,7702	Slow	Any	Any	1	56	-40 to 125	1.62 to 3.63

Table 4.4: Variation with bandgap reference for worst, typical and best corners

Table 4.4 shows the output current variation with the worst possible scenarios of the bandgap voltage reference. Taking into account the highest and lowest value of the current variation we reach a total of $\pm 4.8\%$.

It is worth nothing that this variation is in fact smaller than the bandgap reference. This stems from temperature behavior of both systems. The bandgap temperature behavior assumes a concave style while the temperature drift of converter is a convex style. Both forms end up somewhat neutralizing the temperature variation at the output current.

However, we must also include the variation of the converter due to process mismatches. Simulations showed a total variation over process of 0.98% at the calibrating conditions, 2.5V and 25°C. Since we have a calibrating precision of 0.78%, the total error due to process mismatch is the difference.

The total accuracy of our system is then given by 9.61%+(0.98%-0.78%), which equals to $\pm 4.9\%$ global variation over PVT.

4.3.3 Load Regulation

Load regulation refers to the capability of our voltage-to-current converter of maintaining the output current despite changes in the load connected to it.

Our converter will have a steady output current of 5uA. Which means that theoretically, any load connected to the output will draw this current. The problem begins to rise as the load increases in size as well.

The output current is driven by a p-type mosfet. Which means that with a steady current, and a large load, the device could end up in triode region. If the load is large enough the current could even drop to zero.

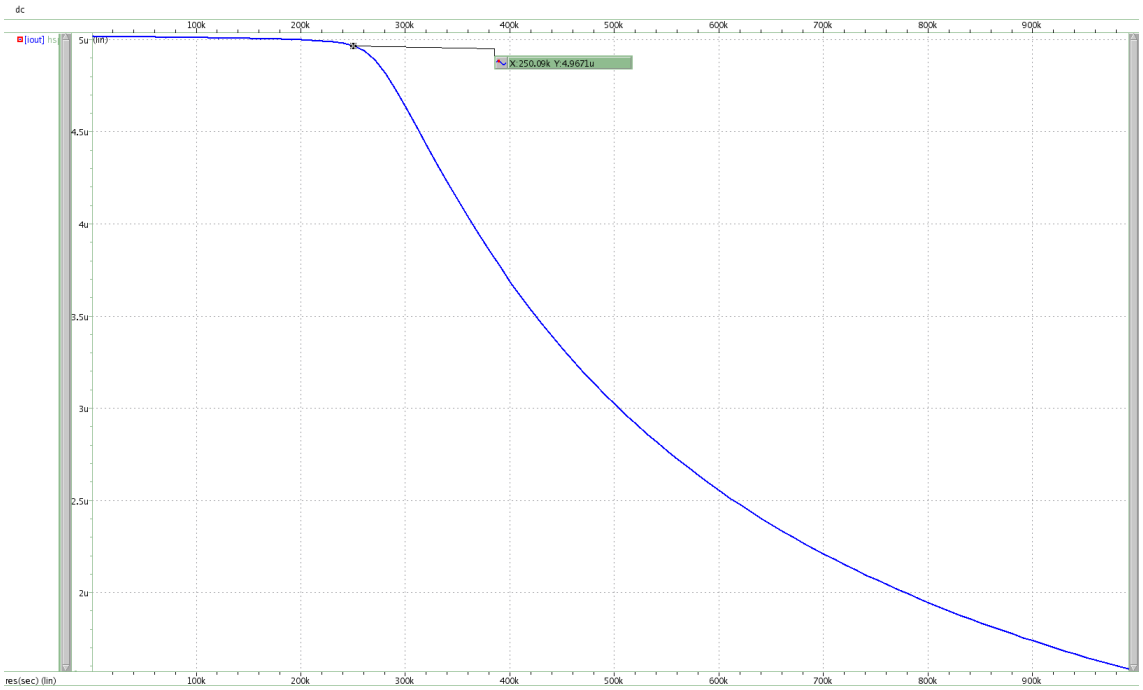


Figure 4.10: Output current shift due to load variation

Figure 4.10 shows the current shift with increasing output load. It is clear that the output current severely drops once the load connected to the output exceeds $250k\Omega$.

In the worst case, we have a source voltage of 1.62V at the output mosfet. With a 5uA current and a $250k\Omega$ load resistor, the drain voltage of the mosfet is 1.25V. Under these conditions the mosfet starts entering triode region and no longer behaving as desired.

Load regulation is given by

$$\%LoadRegulation = 100 * \frac{I_{minLoad} - I_{maxLoad}}{I_{nomLoad}} \tag{4.19}$$

where $I_{minLoad}$ is the current under the lowest load, $I_{maxLoad}$ is the current under the highest possible load in the worst supply voltage case and $I_{nomLoad}$ is the current under the nominal load value, in this case is $125k\Omega$, which was the reference resistor used for calibration.

$$\%LoadRegulation = 100 * \frac{5.0141u - 4.9653u}{5.0059u} \approx 0.97\% \tag{4.20}$$

where the considered lowest load was 0Ω and the output was shorted into ground.

4.3.4 Overview

The following table provides a summed overview of the parameters of our current source, including worst, typical and best cases scenarios. For some of the parameters in the following table, the calculations were not shown since they were subject of study in section 3.6.

The following results were simulated with our bandgap voltage source as a voltage reference for the voltage-to-current converter.

Parameter	Typical Case	Worst Case	Best Case
Output Current (uA)	5	5.1673	4.7141
Temperature and Supply Shift (%)	0.898	0.898	0.898
PSRR (dB) @ DC	-150	-136	-189
PSRR (dB) @ 10kHz	-148	-135	-165
Line Regulation (nA/V)	2.6	13.4	1.74
Load Regulation (%)	-	0.97	-

Table 4.5: Characteristics of the Voltage to Current converter with bandgap voltage reference

The temperature shift can be seen as independent of process variation as expected. The temperature curve remains the same regardless of the resistor corners.

The voltage to current converter shows a total of $\pm 4.9\%$ variation, considering the worst cases of the bandgap voltage reference, including the non-calibrated random error of the converter itself.

4.4 Summary

In this chapter, a current source was presented based on the topology of a dropout regulator along with the implementation results. The proposed topology was analyzed both theoretically and backed up with results based on extensive simulations.

The used topology makes use of an operational amplifier in order to force a reference voltage across a resistive load. The feedback action guarantees insensitivity to both process variation and mismatch, barring any deviations on the amplifier itself, of course.

It was shown that, while the resistive load had a good behavior under temperature variations, the same could not be applied with process drift. Of all the devices used in this work, polysilicon resistors show the highest amount of drift over process corners, nearly reaching a total of 50% variation between fast and slow corners.

To overcome this high variation, a programmable resistor was created. To compensate process drift, a calibration mechanism and algorithm was implemented to act upon our programmable resistor with the help of a reference resistor.

The implemented system reduced the overall variation of 50% down to just 0.78% with 6 bits of calibration. Due to the small nature of the temperature drift of the resistor, a one time only calibration process was adopted. This process carried a lower power consumption and lower area usage compared to other alternatives.

Chapter 5

Dynamic Device Mismatch Compensation

One of the most notable features of nanometer scale CMOS technology is the increasing magnitude of variability of several key parameters affecting the performance of integrated circuits. We saw in section 3.7 that the biggest sources of error of the bandgap voltage reference were the random process variations of the devices. However, designers have little to no control on these undesired shifts.

Designers have to resort to laser trimmings, digital calibration, large devices and careful layout. But even with these solutions, random variations still spun several percent. In this chapter, we will be discussing and implementing dynamic techniques as a way to lower these variations when compared to the conventional, static bandgaps.

5.1 Dynamic Offset Cancellation Techniques

We saw in section 3.7.6 that the greatest source of error in the bandgap circuit was caused by the random amplifier offset, reaching almost 40% of the total error of the output voltage. In this section, we will explore several techniques for reducing this offset and implement one of them as a means to reduce this error.

Dynamic methods come with the benefit of neutralizing offset due to process drift, allowing more margin for designers to work with.

There are three types of techniques that can be applied in order to reduce the random offset of the amplifier. These techniques are trimming, auto-zeroing and chopping.

Trimming involves measuring and then reducing the offset during production, very similarly to the trimming operation we perform in chapter 5. While this approach can be used to obtain a reduction of the offset voltage, the process is more costly.

5.1.1 Auto-Zeroing Technique

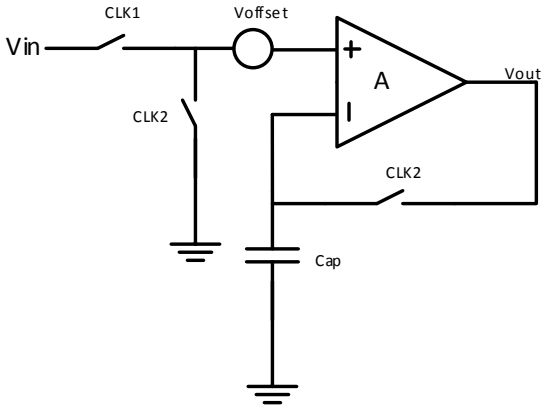


Figure 5.1: Auto-zeroing with offset storage

The idea behind the auto-zeroing can be seen in figure 5.1. Auto-zeroing is a discrete time sampling technique. It involves sampling the offset voltage of the amplifier in one phase, and then subtracting it from the input signal in the other clock phase.

When CLK2 is high, the amplifier is in auto-zeroing phase in which the offset voltage is sampled into the capacitor Cap. Once the nodes are settled, the voltage across the capacitor will be the offset voltage.

When CLK1 is high, the device will amplify the difference between both terminals. Since the capacitor contains the offset voltage, then this value will be subtracted from V_{in}, effectively neutralizing the offset.

This technique is fairly simple to apply but comes with the downside of the output value only being available half the time, since we require one of the clock cycles to sample the offset voltage.

5.1.2 Chopping

Unlike auto-zeroing, chopping is a modulation technique . Figure 5.2 shows the basic set up for the chopping technique on an operational amplifier followed by a low-pass filter.

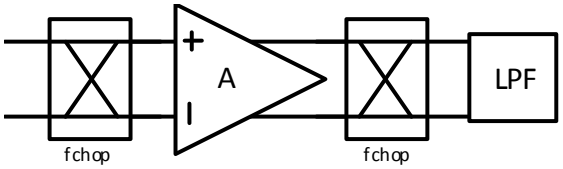


Figure 5.2: Basic Chopper Technique

The input signal first passes through a chopper driven by a clock at frequency $fchop$, thus it is converted to a square wave voltage at $fchop$. Next, the modulated signal is amplified together with its own input offset. The second chopper then demodulates the amplified input signal back to DC, and at the same time modulates the offset to the odd harmonics of $fchop$, where they are filtered out by a low-pass filter (LPF).

This results in an amplified input signal without offset. Low-frequency errors, such as 1/f noise and drift will be modulated and filtered out along with offset if the chopping frequency is higher than the 1/f noise corner frequency.

5.1.3 Chopped Operational Amplifier

In this section, we will be employing the chopping technique to our folded cascode amplifier. We will be analyzing the signal behavior in the frequency domain and the overall impact this technique has on the random offset.

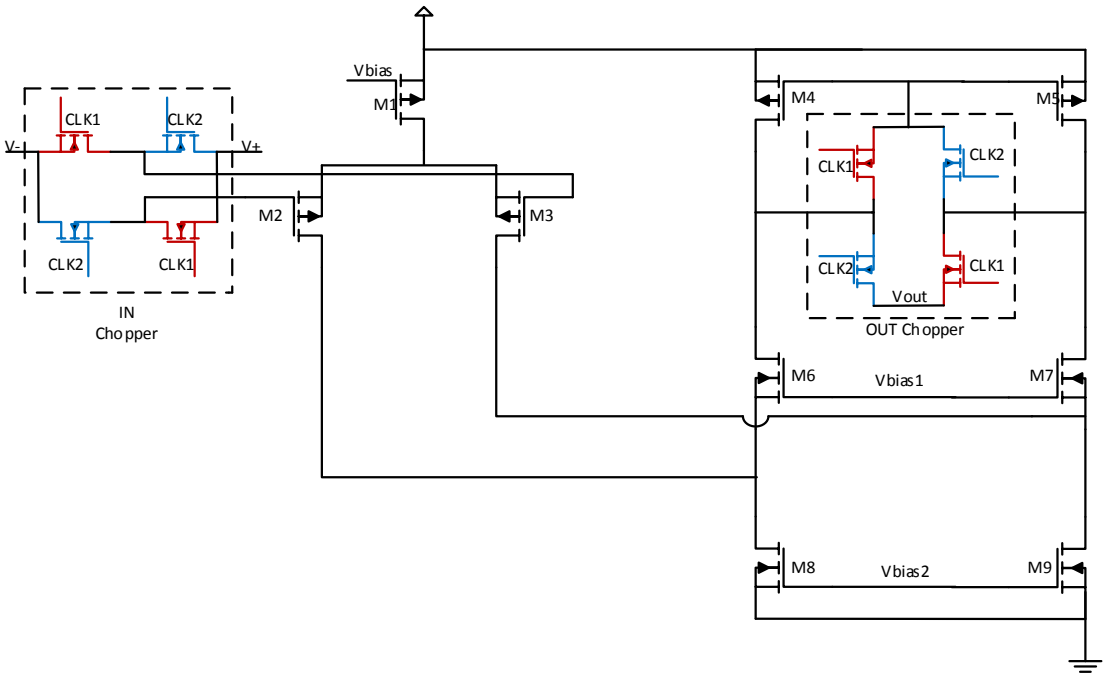


Figure 5.3: Folded cascode amplifier with choppers

Figure 5.3 shows the folded cascode amplifier with choppers before the input differential pair and at the output. The choppers are surrounded by a dashed line box. Inside each chopper module, the mosfets are colored in red and blue to show which devices are turned on in the same clock phase.

The use of choppers requires an external clock signal with two non-overlapping phases, so that the red and blue mosfets are never on simultaneously. The non-overlapping clock phases were generated using a similar scheme to figure 2.8.

The folded cascode amplifier was set up in unit gain configuration (i.e., the output node fed back to the inverting terminal of the amplifier) in order to measure the offset using this technique. An ideal voltage source of 0.7V was fed into the non-inverting terminal of the amplifier and the signal behavior was observed and analyzed in several nodes.

The basic steps of the chopping technique can be seen in the figure below.

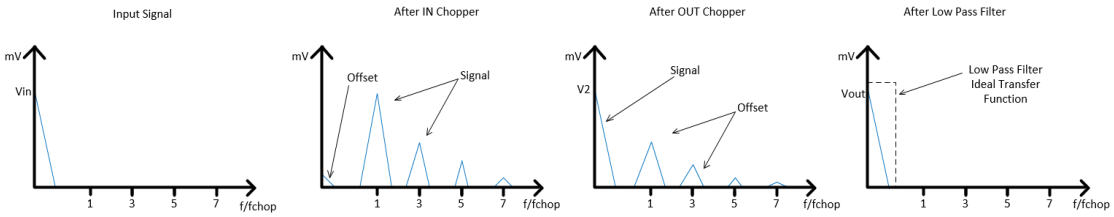
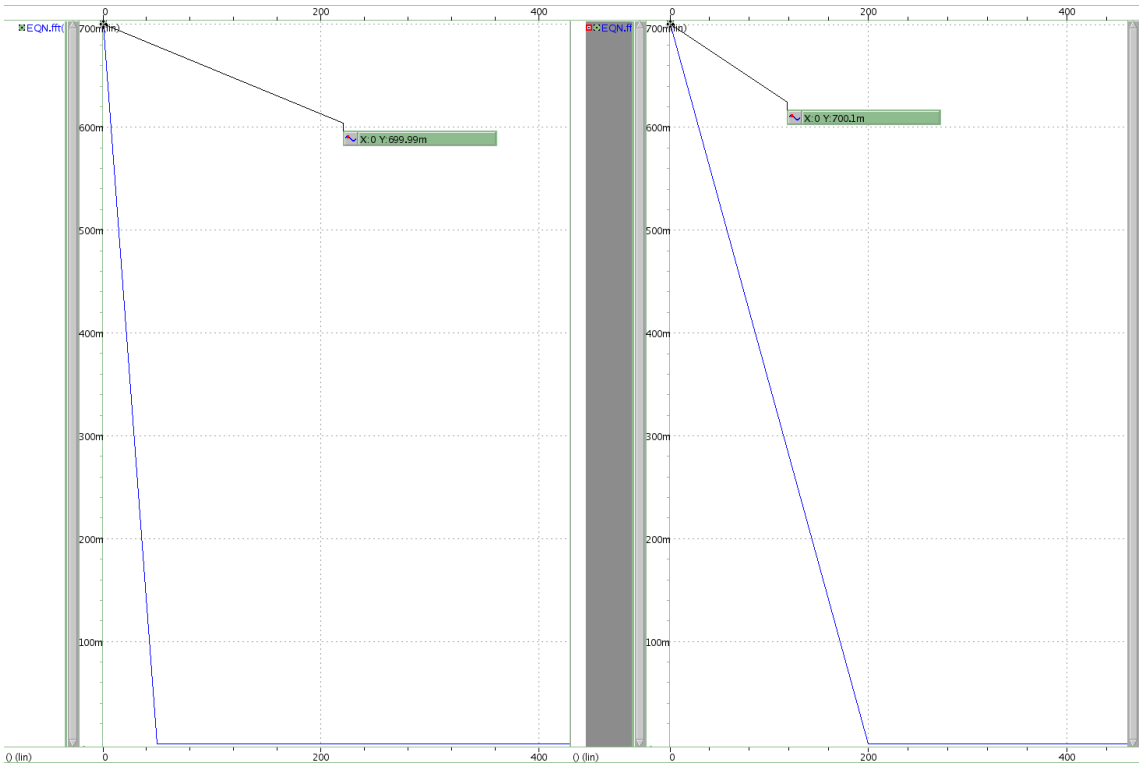


Figure 5.4: Chopping Stages

First, the input signal is modulated into the chopping frequency. In the second stage, the offset is modulated into the chopping frequency while the input signal is demodulated back into the low frequencies. Lastly, the offset, now in the high frequencies, is removed with the low pass filter.

The remaining offset is due to the non-infinite gain of the amplifier and several other non-ideal modules of the chopping system, such as clock skew between both phases, impedance mismatch between two input nodes, clock feedthrough, channel charge in the switches used for implementing the choppers and the eventual ripple voltage due to the low pass filter.



(a) Input Voltage (0.7V DC) (b) After Low Pass Filter

(c) After OUT Chopper

Figure 5.5: Signal behavior in the frequency domain

Figure 5.5 shows the signal behavior in several nodes of the chopped amplifier. In 5.5a we the see frequency equivalent of the ideal voltage supply. As expected, the dc source shows a 0.7 voltage level at the low frequencies, being completely 0V before reaching 50 Hz.

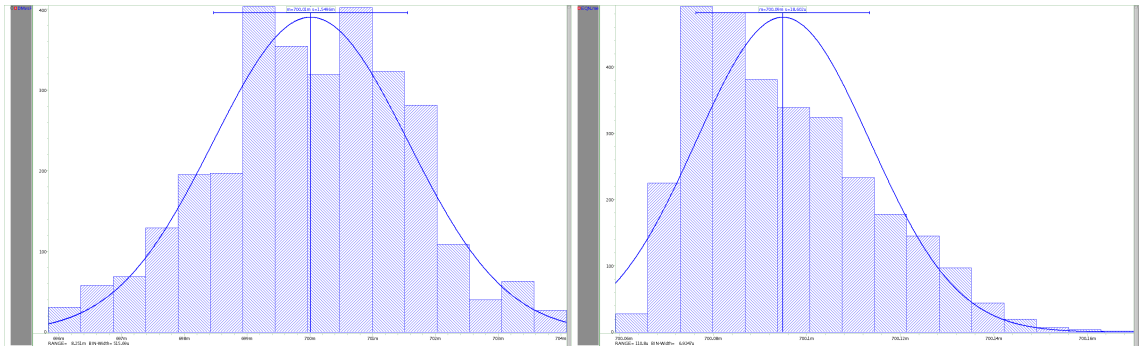
After the first chopper (i.e., IN Chopper), we would expect to see the input signal modulated into the chopping frequency, residual values in the odd harmonics, and the offset close to the low frequencies. However, due to the test configuration implemented, the expected value would be difficult to extract with accuracy.

Figure 5.5c shows the signal after the last chopper and before passing through the low pass filter. As expected, we can see our signal at the low frequencies, having already being modulated by the IN Chopper and demodulated back into DC after the OUT Chopper. The figure also shows the amplifier's offset modulated into the chopping frequency, 10 kHz, and the odd harmonics.

Lastly, figure 5.5b shows the amplifier output after the signal has been filtered through the low pass filter. The low pass filter used has a cutoff frequency of 100 Hz and attenuates the amplifier offset by a factor of 100 (i.e., 40 dB) as it reaches the chopping frequency. The output voltage is now 0.7001 mV, having reduced the random offset from 2.6mV to 100uV in this particular case of study.

5.1.4 Chopped Amplifier Results

In this section, we will be presenting the results of the offset simulations after integrating chopping techniques into our amplifier.



(a) Drift without chopping (b) Drift with chopping

Figure 5.6: Voltage drift from nominal value (0.7V) at 27^{o}C, supply = 1.62 V, all corners and 200 MC

Figure 5.6 shows a comparison between two histograms that represent the output voltage drift from the nominal value of 0.7 V. It is clear the impact of the chopping technique has on the voltage offset in voltage follower configuration.

Our previous amplifier, without any type of compensating techniques, had a worst case scenario of 4 mV of offset. With the chopping techniques, this offset is drastically reduced to a maximum of 170 uV, showing the efficacy of the technique applied.

The application of the chopping technique represents the reduction of the offset voltage by a factor of 23 in the simulated scenario.

The following table shows the characteristics of the chopped amplifier.

Parameter	Test Conditions	Worst	Typical	Best
Offset Voltage(uV) - Without Chopping	PVT	4000	110	80
Offset Voltage(uV) - With Chopping	PVT	308	90	23.9
Current Consumption (uA)	PVT	5.174	2.826	1.65

Table 5.1: Chopped Folded Cascode Amplifier Parameters

The offset reduction is very significant. The worst case scenario of the standard operational amplifier was an offset of 4 mV. With the addition of the chopping techniques, this offset was reduced to a maximum case of 308 uV, showing a reduction by a factor of 13.

In the bandgap voltage reference, we expect a 19 mV reduction of the overall variation at the output. The use of chopping is expected to reduce the bandgap voltage error by approximately 30%.

It was expected a slight increased in current consumption due to the added hardware in the non-overlapping clock generator of the switches used in the choppers. Nonetheless, this increased in current is vastly compensated for the reduced offset voltage.

However, chopping also requires more die usage for the realization of the low-pass filter, that in deep CMOS integrated circuits can be problematic due to the size of the components. To alleviate this problem the switching frequency should be increased.

5.1.5 Bandgap Voltage Reference with Chopped Amplifier

In this section, we will be presenting the improvements in the bandgap reference voltage due to the addition of the chopped amplifier instead of the regular one. The following results will only be in regards to the overall variation of the output voltage and not other parameters of this module. For other parameters of the bandgap reference see section 3.6.

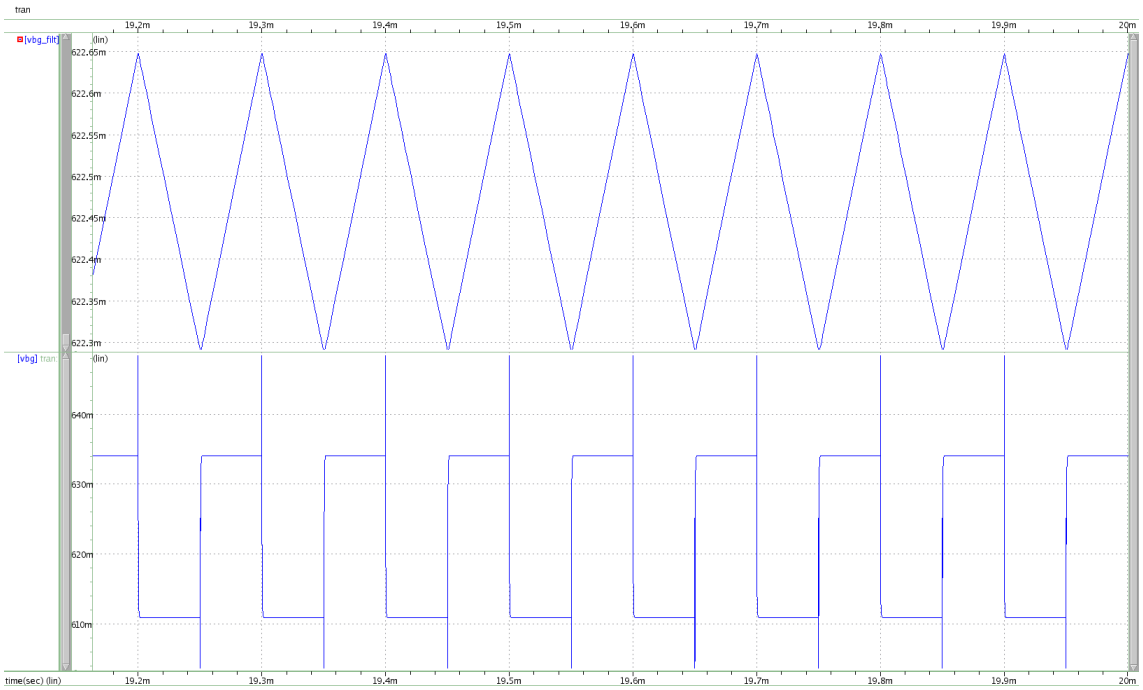


Figure 5.7: Output voltage behavior due to chopping amplifier

Figure 5.7 shows the output signals of the bandgap reference voltage before and after the low pass filter. Paying close attention to the Y scale of the presented plots, we can see the switching behavior expected to due the chopped amplifier.

The output voltage voltage fluctuates between the two extreme points because of the effect of the chopped amplifier. Another way to look at this is what would happen if the amplifier offset oscillated between positive and negative values.

When the offset is positive, the output voltage would be increased by a certain proportion. But when the offset is negative, the output would be decreased. In the time domain, a chopped amplifier forces the offset to oscillate between the nominal value of 0 V, assuming a 0V offset is possible. In the case of the bandgap, the output voltage fluctuates between the nominal value of 624 mV and the extremes caused by a $\pm 2.4mV$ of offset, in this particular scenario.

The other plot reveals the "real" output voltage after being filtered by a low pass filter. The residual rippled comes from the filter itself and is small. This particular scenario shows a ripple of 350 uV. Reducing this ripple is possible through the use of filter with lower cut off frequency or the application of 2nd order filters for higher attenuation.

Parameter	Test Conditions	Variation
Bandgap Voltage Without Chopping (%)	PVT	$\pm 4.89\%$
Bandgap Voltage With Chopping (%)	PVT	$\pm 3.4\%$

Table 5.2: Bandgap Voltage Reference with Chopped Amplifier Results

The table above shows a comparison between the output voltage variation before and with the addition of the chopped amplifier to out bandgap voltage reference. It is clear the error reduction with the addition of the chopped amplifier.

The previous variation over PVT range was $\pm 4.89\%$. However, by attacking the biggest source of error which was the amplifier, we managed to reduced the overall variation to $\pm 3.4\%$ over the same conditions.

In absolute terms, the reduction of the random offset voltage of the amplifier reduced the bandgap overall error by 17.7 mV against the expected 19 mV. The slight difference between the expected results and the simulated ones are the simple fact that the worst case scenario of the amplifier's offset was simulated when the device was in standalone mode (i.e., not integrated in the bandgap reference).

5.2 Dynamic Matching of the Bandgap Core

In this section, we will be employing the previous used techniques into the bandgap voltage core. Figure 5.8 shows the jumping bandgap voltage reference with a chopper between the branches connected to the BJTs and the chopped amplifier.

This works similarly to the chopper of the amplifier explained in the previous chapter. The random variations both due to mosfets variations but also the eventual load shifts of the BJTs and resistors are corrected. The switching network forces the output voltage to oscillate between the nominal value and the minimum and maximum value caused by the eventual errors.

By passing the signal through a low pass filter, the ripple error is heavily attenuated. One benefit of using these techniques is only a single low pass filter is required, even though multiple choppers are used.

By looking at the schematic, we see the chopper placed between the branches connected to the BJTs. When CLK1 is high, then M4 is connected to V_a and M5 is connected to V_b, like the static bandgap. When CLK2 is high, these networks switch, forcing M4 to V_b and M5 to V_a, effectively swapping the signal of the error.

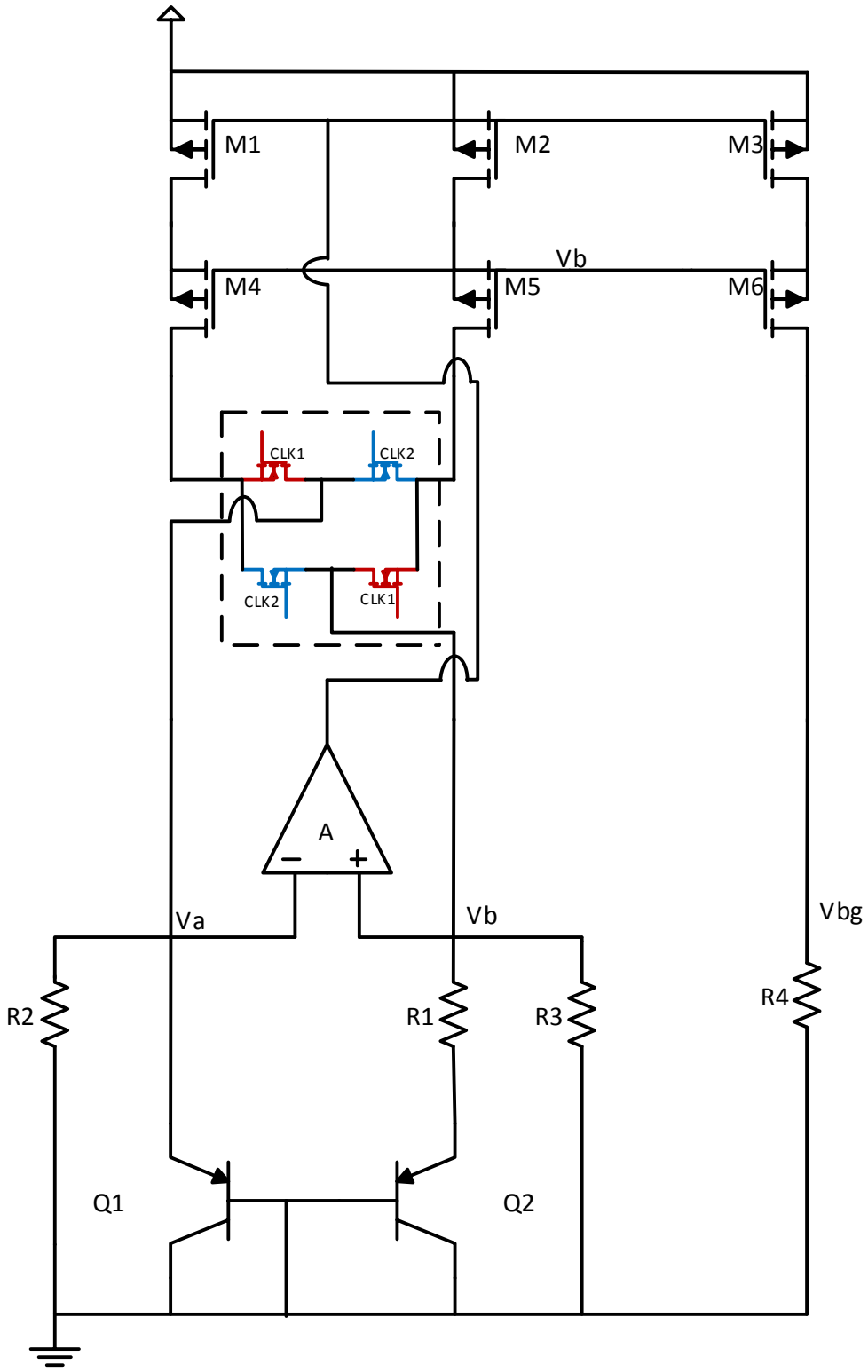


Figure 5.8: Jumping Bandgap Voltage Core with Chopped Amplifier

5.2.1 Jumping Bandgap Core with Chopped Amplifier Results

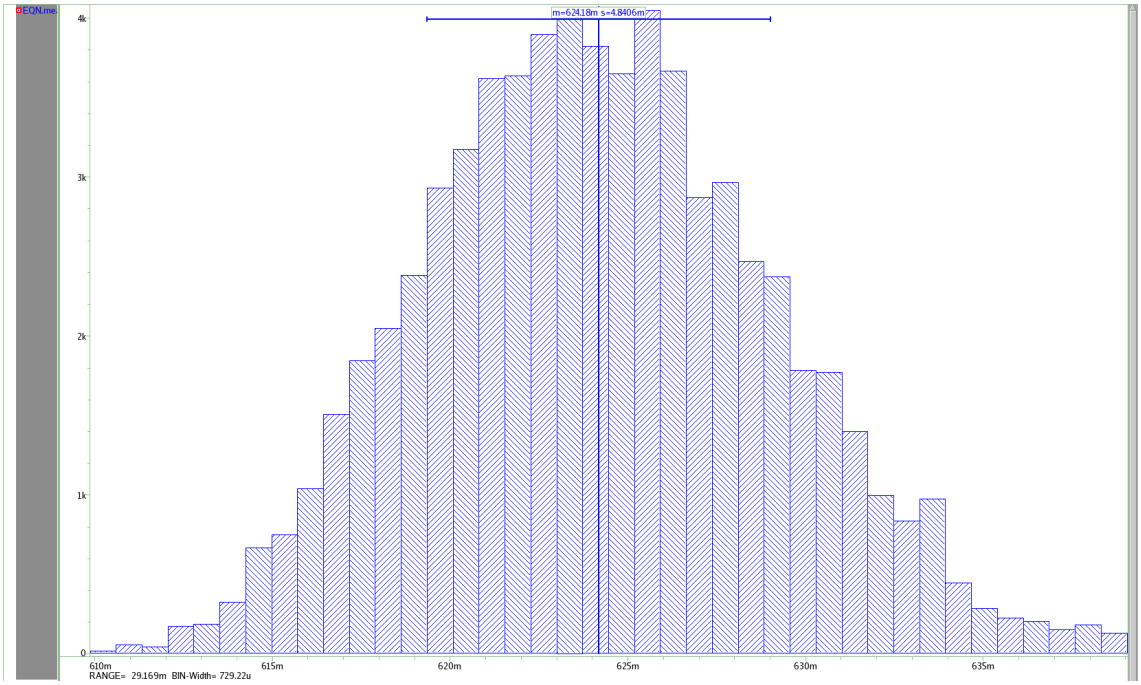


Figure 5.9: Output Voltage Variation of the Jumping Bandgap under PVT

Figure 5.9 shows the variation of the output voltage of the jumping bandgap under PVT conditions. The overall variation from the addition of the chopper in the bandgap core caused a reduction of the error close to 12 mV which represents a 30% error reduction.

Parameter	Test Conditions	Variation
Bandgap Voltage without Chopped Amplifier (%)	PVT	$\pm$4.89%
Bandgap Voltage with Chopped Amplifier (%)	PVT	$\pm$3.4%
Jumping Bandgap Core with Chopped Amplifier (%)	PVT	$\pm$2.38%

Table 5.3: Jumping Bandgap Core with Chopped Amplifier Results

If we use the jumping bandgap core with chopped amplifier as a voltage reference in our voltage to current converter we achieve a total variation of $\pm$2.88% at the calibration conditions. By factoring in the post-calibration variation considering both the bandgap reference and the converter, we can upper bind the total variation by $\pm$4%.

5.3 Summary

In this chapter, techniques to reduced random amplifier offset were presented. Of the three presented, the chopping provided the best results and so it was implemented in our folded cascode

amplifier.

The chopped amplifier shows a drastic reduction in random offset, reaching approximately 1300% improvement.

The chopped amplifier replaced the previous amplifier in our bandgap voltage reference and the device was simulated once more under PVT conditions. The results were as expected and the overall variation was reduced by 30.3%. The addition of the new amplifier translated into the removal of 17.7 mV of error caused the amplifier offset.

The same principles were applied to the bandgap core and improvements were significantly noted. Through the application of dynamic matchings techniques the overall error was reduced from $\pm$4.89% to $\pm$2.38%.

Chapter 6

Conclusions and Future Work

6.1 Summary of the work developed

This thesis was focused on the development of a voltage and current reference. Voltage and current references find applications in a variety of both analog and digital systems. The design of such systems requires the scrutiny and extensive detailed analysis of several factors. Temperature variation is usually the focal issue we tend to face in references.

In order to achieve the thesis goal various phases were carried out. The work done and the conclusions arrived at in each phase are presented below.

- An initial research was conducted to understand how these voltage and current references worked and the different topologies. The performance and other parameters were evaluated and a general topology was chosen for this work.

- The next phase was designing and simulating the chosen topology extensively with the usage of an ideal operational amplifier. This phase provided a throughout understanding of how this topology worked.

- After understanding how our circuit works, we moved on to the development of the operational amplifier. Several topologies were researched and simulated until the folded cascode topology was chosen due to the stability being mostly determined by the capacitive load at the output. The bias circuit was also developed in conjunction and the transconductance was matched to that of a resistor which has better behavior versus temperature.

- The next step was integrating the bias circuit and amplifier into the bandgap voltage reference. By using the bias circuit of the amplifier, it was introduced a cascode transistor in each branch of the bandgap reference, increasing supply rejection.

- With the voltage reference finished, we moved on to the current reference where extensive research and work was done in order to locate a suitable topology. Initial tests using a switched capacitor topology with capacitor detection by ring oscillator failed, forcing us to

moved on to the voltage to current converter with trimming and the corresponding calibration mechanism.

- The last item in this work is an improvement to the operational amplifier. By recurring to dynamic offset cancellation techniques we managed to virtually remove the random offset of the amplifier to a maximum of 309 uV.

- The chopped amplifier was integrated in the bandgap voltage reference where extensive tests were performed to measure to the impact of the new amplifier. Simulations shows the expected results and the overall error was reduced by around 30%. Finally, the same dynamic technique was applied to the bandgap core and the overall precision dynamic bandgap plus chopped amplifier showed a total of $\pm 2.38\%$.

6.2 Future Work

Even today, the development of references is still a laborious job. As technology improves and we delve into higher integrations, the development of high accuracy voltage or current references becomes more difficult.

Regarding this work, there is still some room for improvement in the overall accuracy of the bandgap voltage reference. Resistors could be replaced by switched capacitors. Capacitors show better temperature behavior and can be matched with higher degree of accuracy than resistors.

In the chosen current reference, the room for improvement would be small if not negligible. Increasing the number of bits would be counter productive. One possible improvement would be compensating the temperature variation of the resistor, which however small, is still there and participates to the overall error.

Appendix A

BGR Simulation Results

This section will contain the various plots and histograms from the extensive simulations performed. We will begin with the results of the bandgap voltage reference and then move into the analysis of the error sources.

A.1 BGR Parameter Results - Pre-Layout Simulations

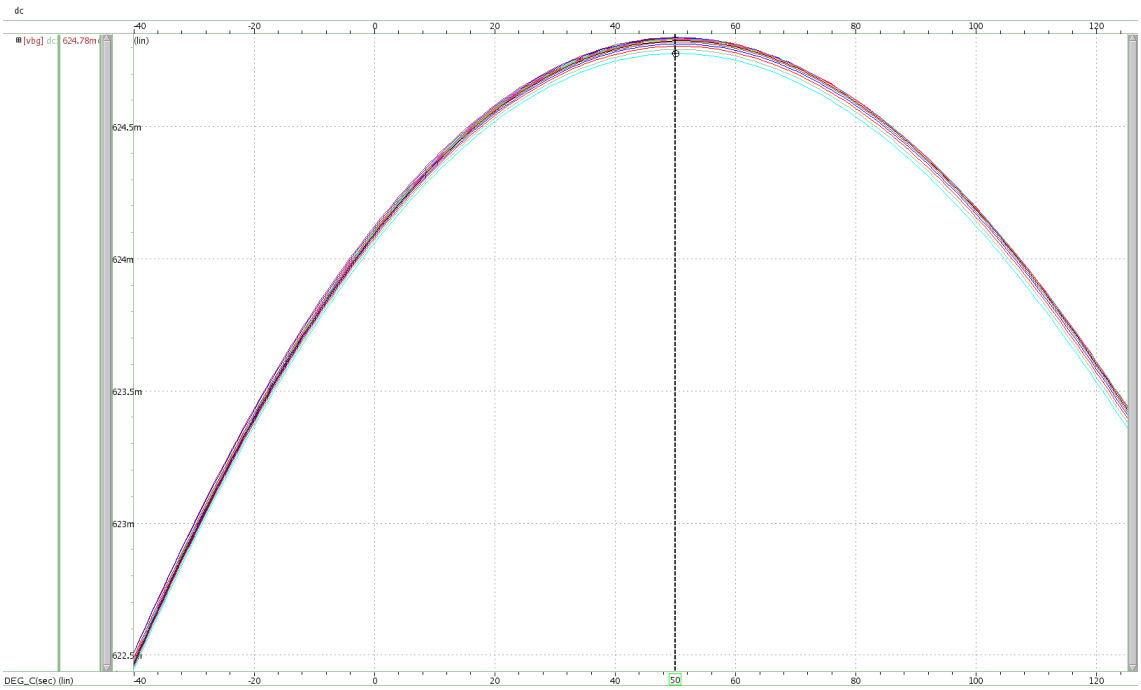


Figure A.1: Output Voltage Variation with Temperature and Supply Voltage at Typical Corners

Overall Output Voltage Variation: 0.4%

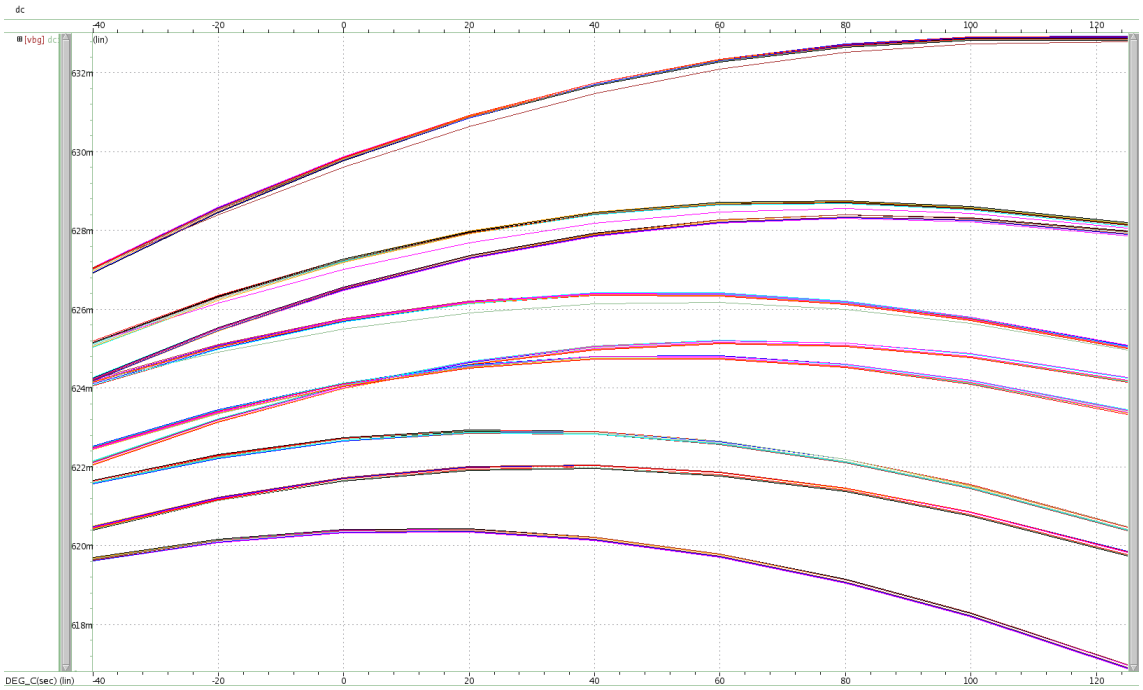


Figure A.2: Output Voltage Variation at all corners

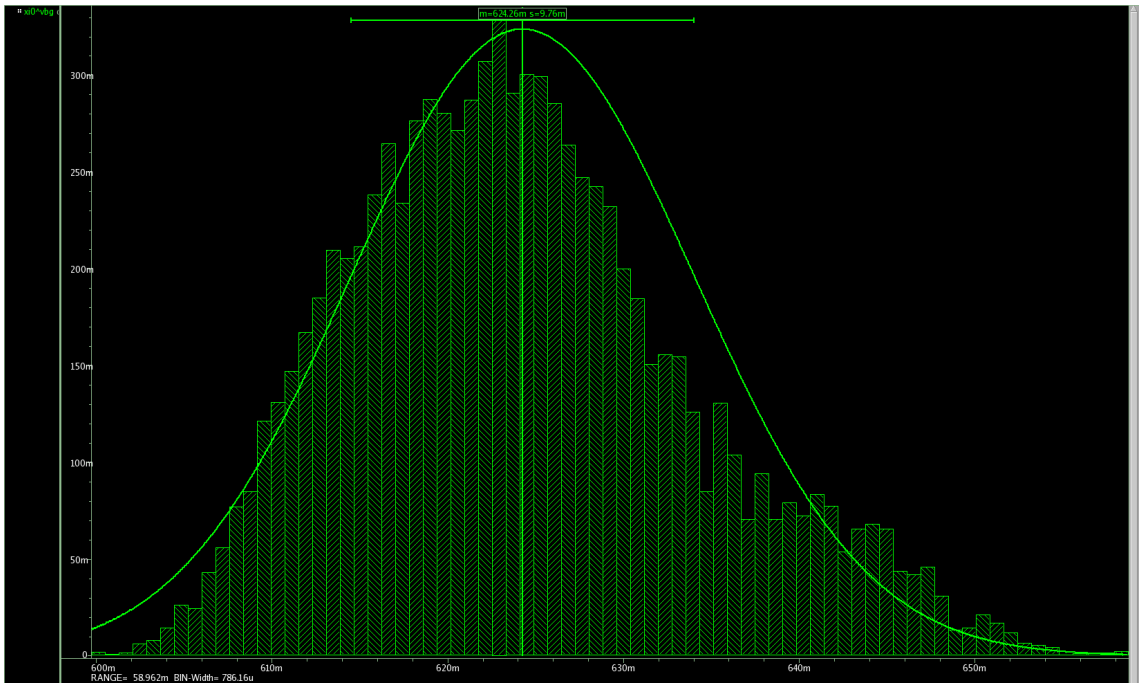


Figure A.3: Output Voltage Variation under PVT

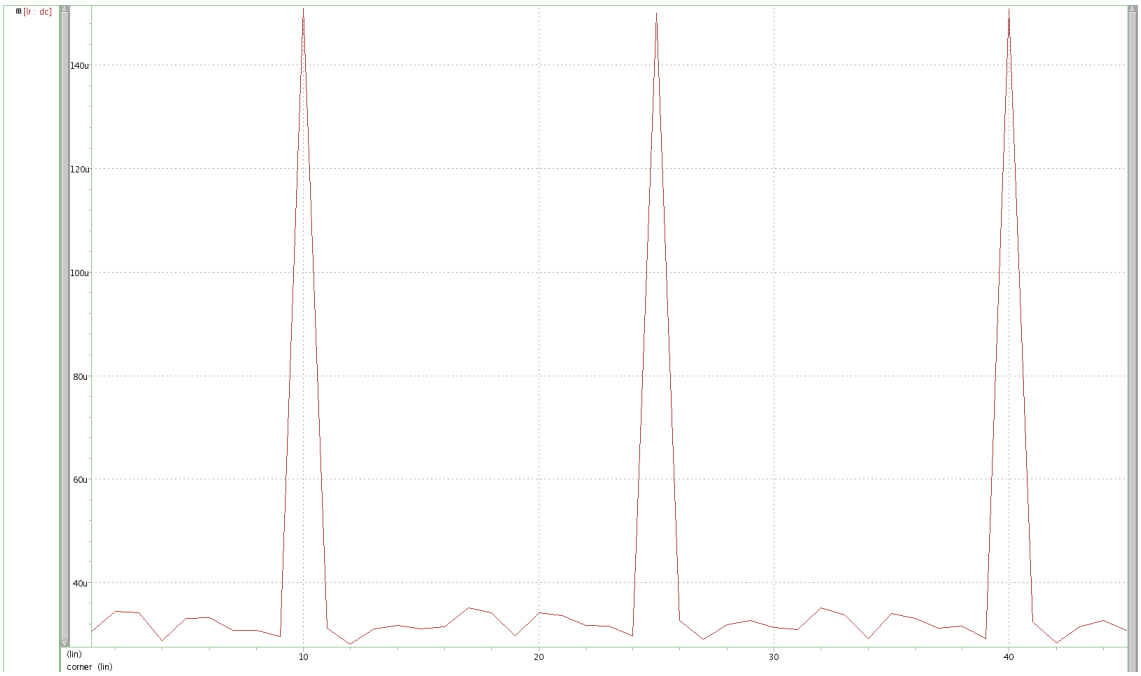


Figure A.4: Line Regulation at all corners

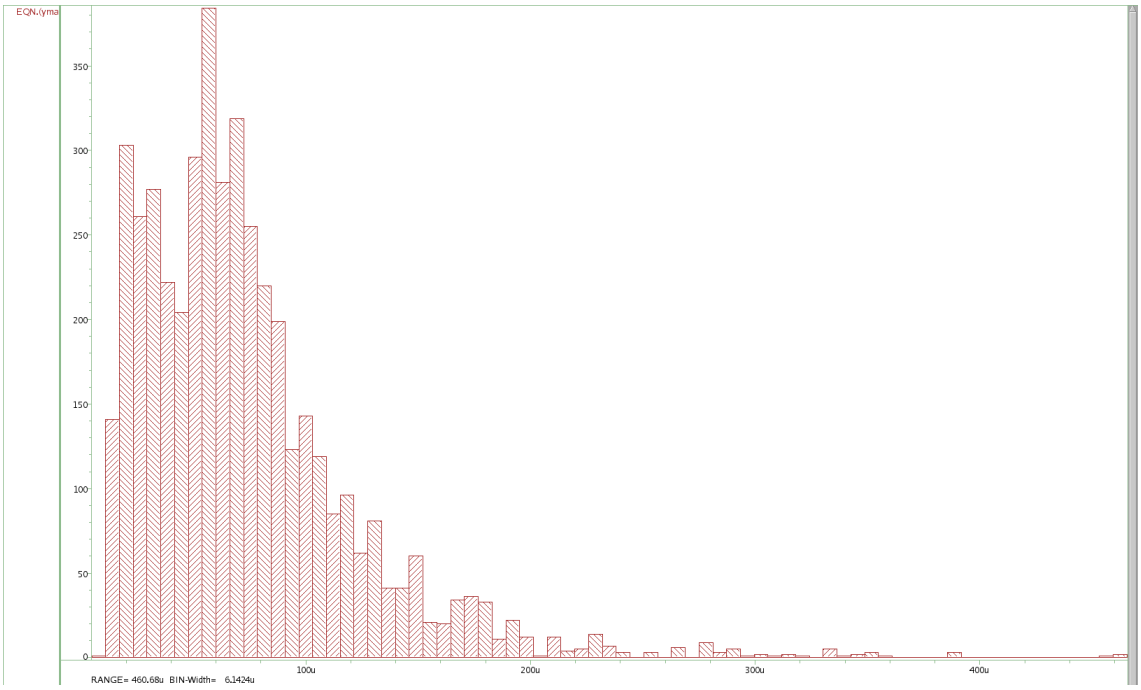


Figure A.5: Line Regulation under PVT

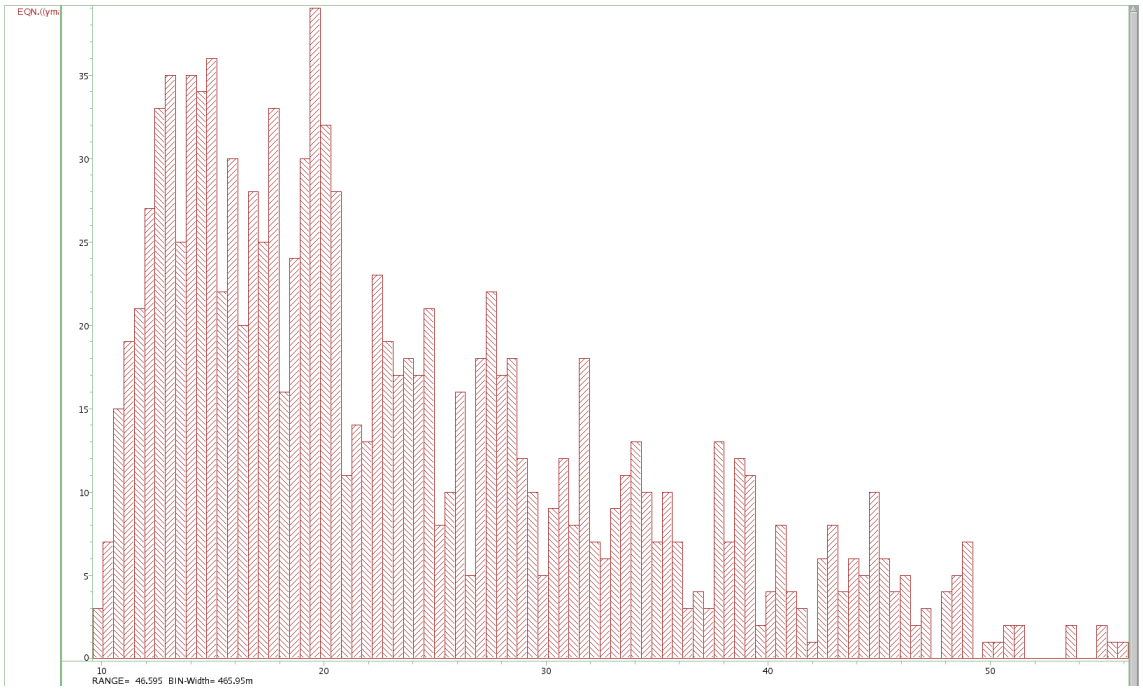


Figure A.6: Line Regulation under PVT

A.2 BGR Parameter Results - Post-Layout Simulations

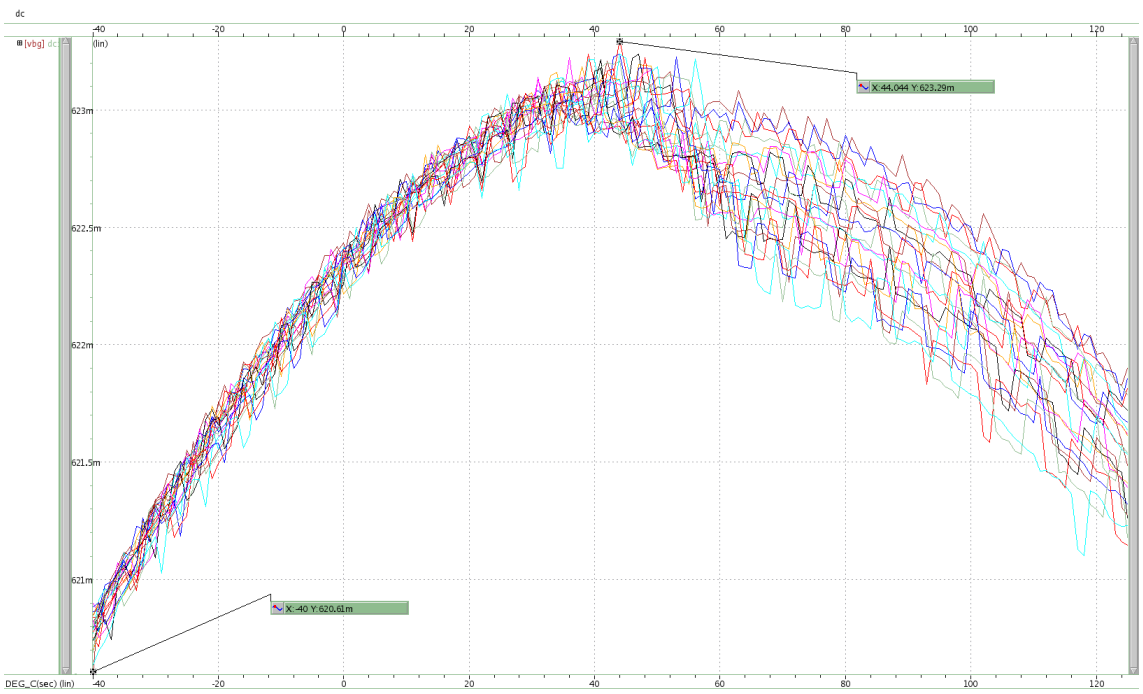


Figure A.7: Output Voltage Variation with Temperature and Supply Voltage at Typical Corners

Overall Output Voltage Variation: 0.43%

A.3 Folded Cascode Amplifier

Figure A.8: Power Consumption under PVT

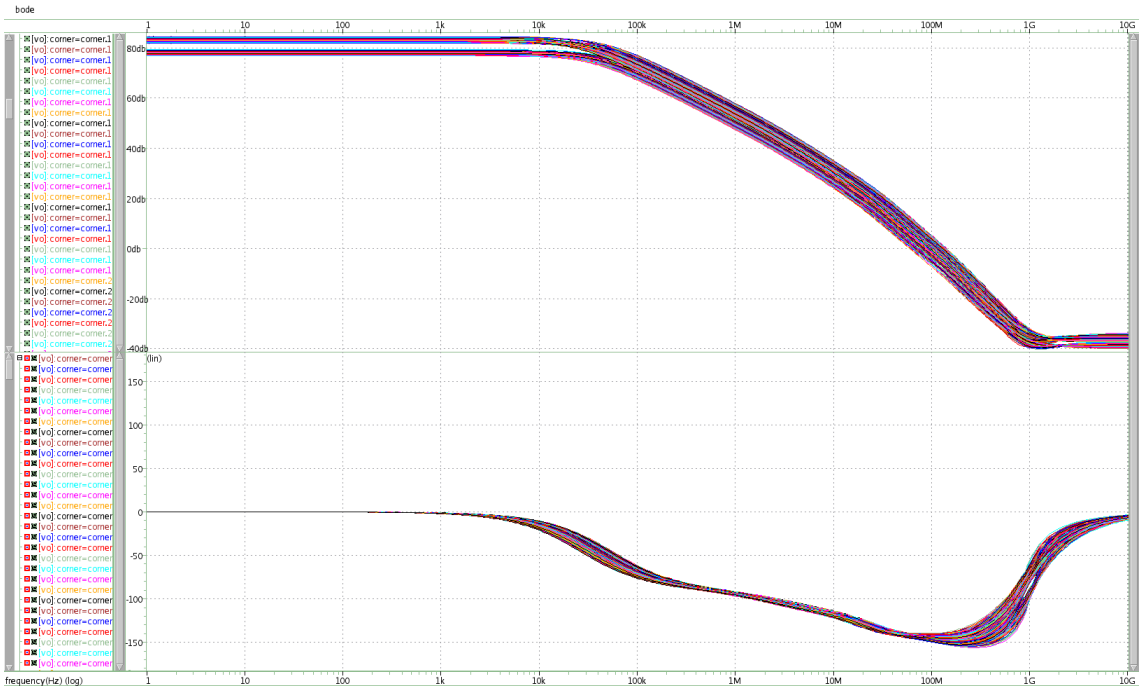


Figure A.9: Gain and Phase Margin under PVT

A.4 Folded Cascode Amplifier (Post-Layout)

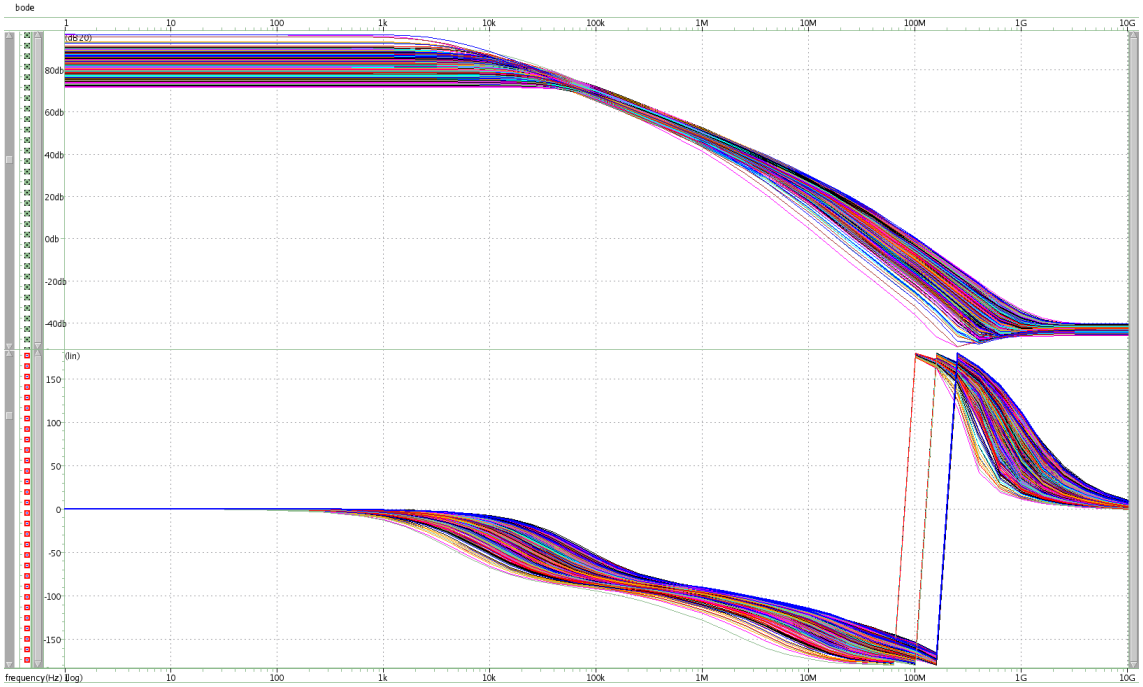


Figure A.10: Gain and Phase Margin under PVT

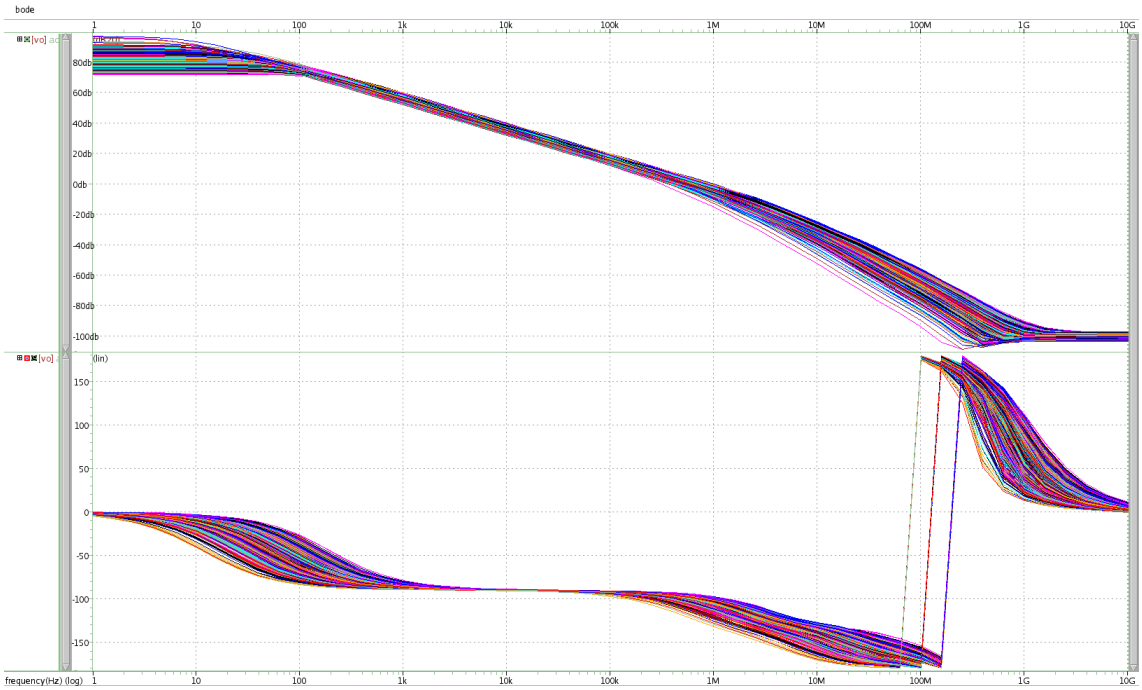


Figure A.11: Gain and Phase Margin under PVT with Mosfet Load

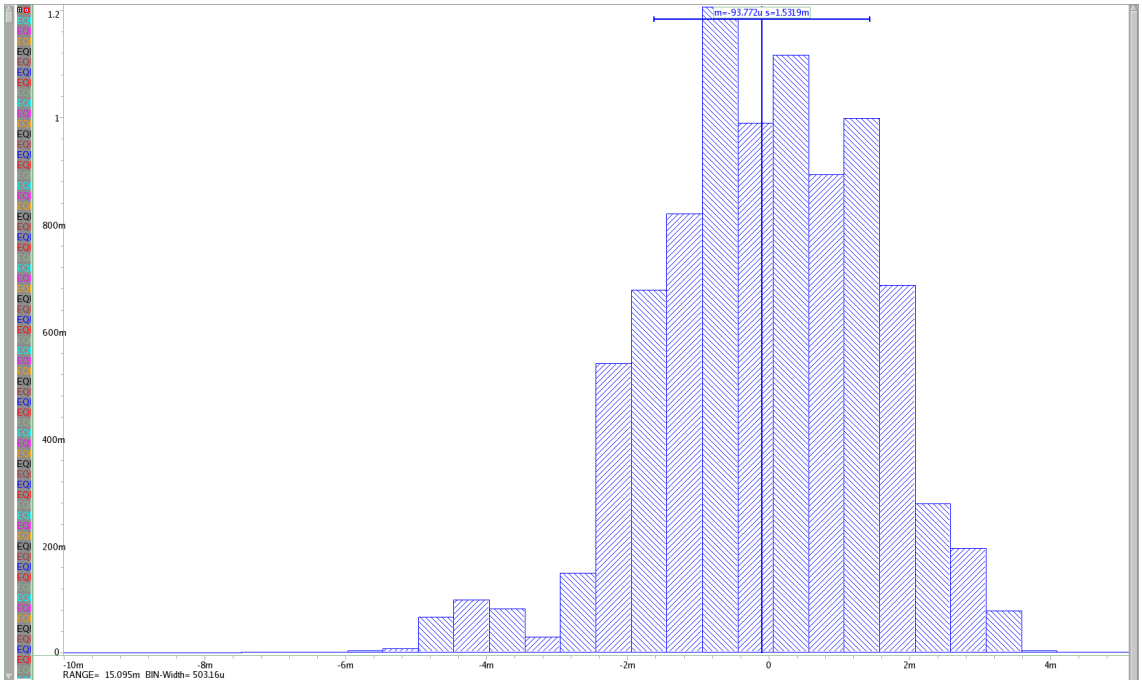


Figure A.12: Input Offset under PVT

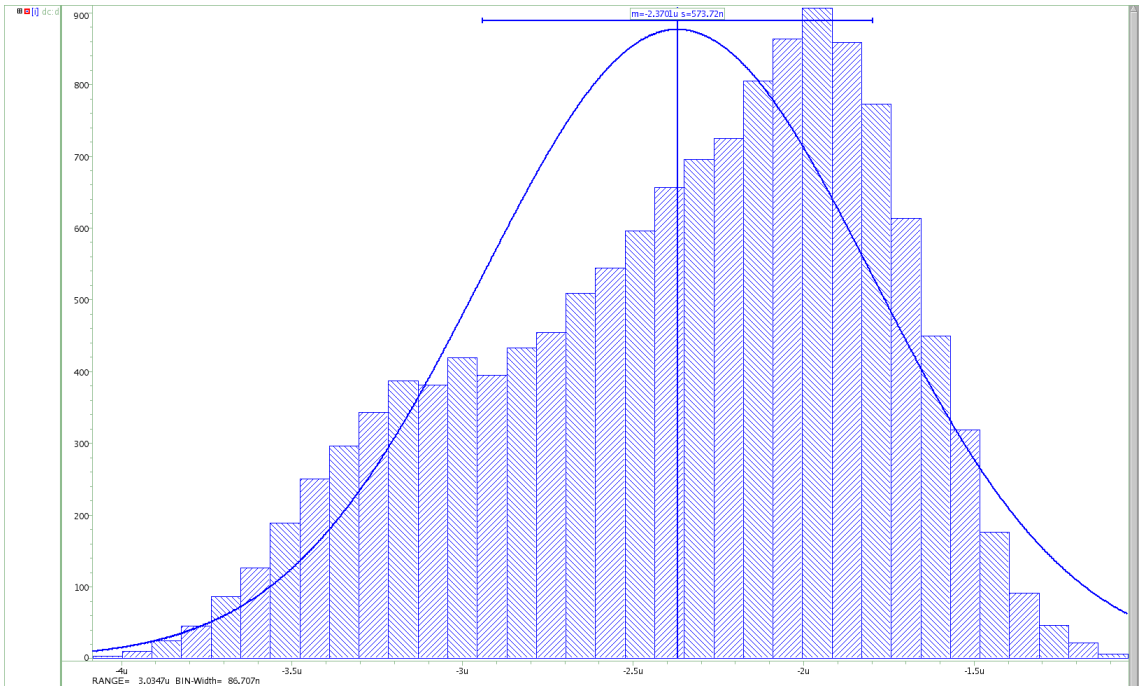


Figure A.13: Power Consumption under PVT

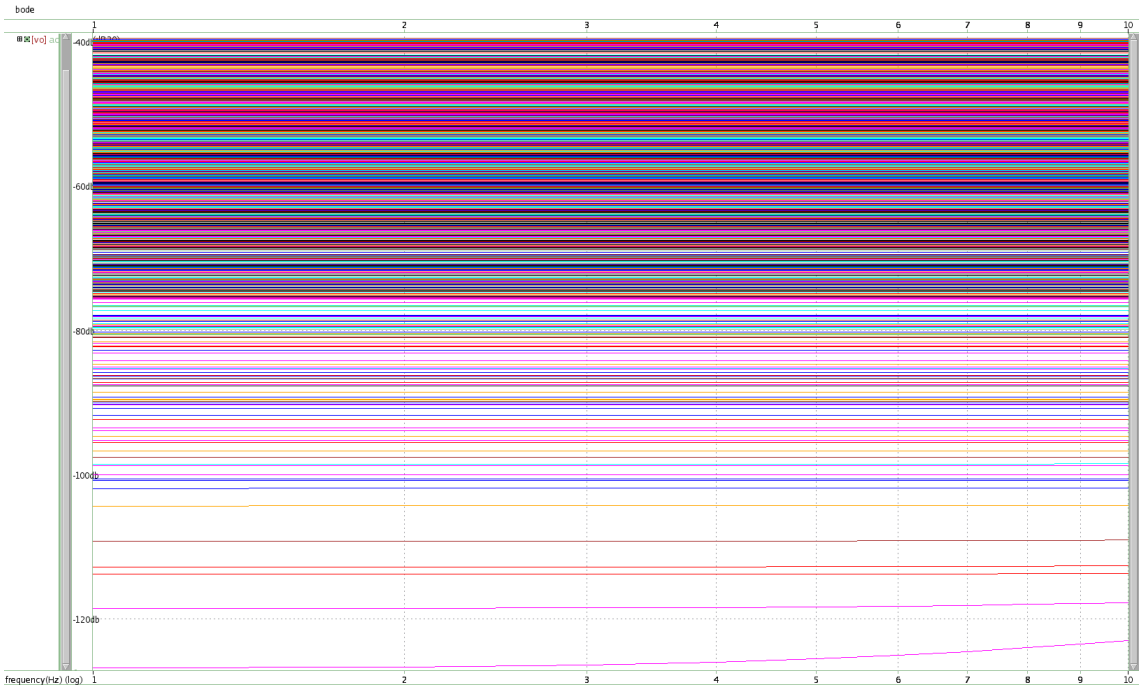


Figure A.14: PSRR under PVT

A.5 Sources of Error of the Bandgap Voltage Reference

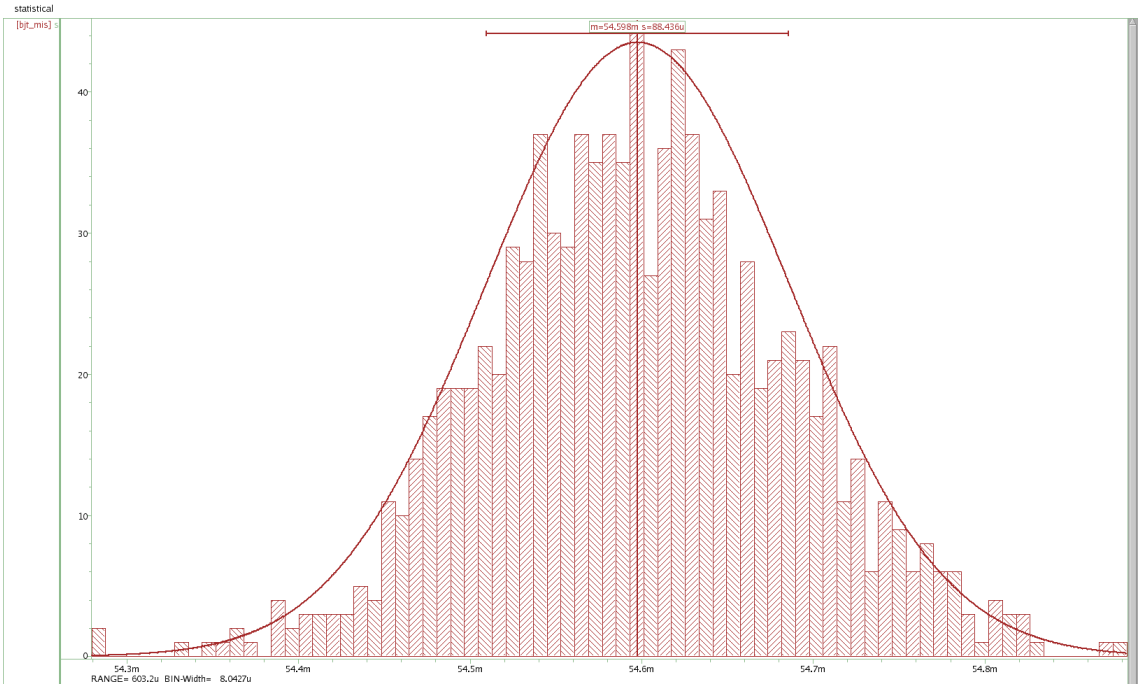


Figure A.15: BJT Mismatch under PVT

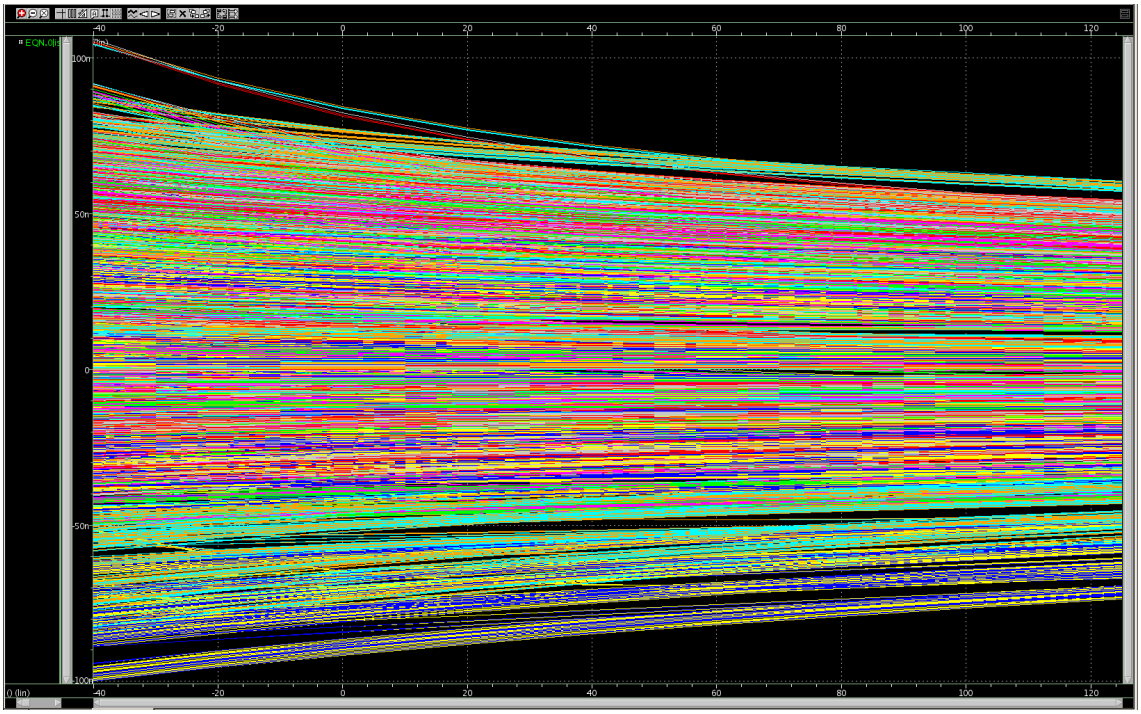


Figure A.16: Mosfet Mismatch under PVT

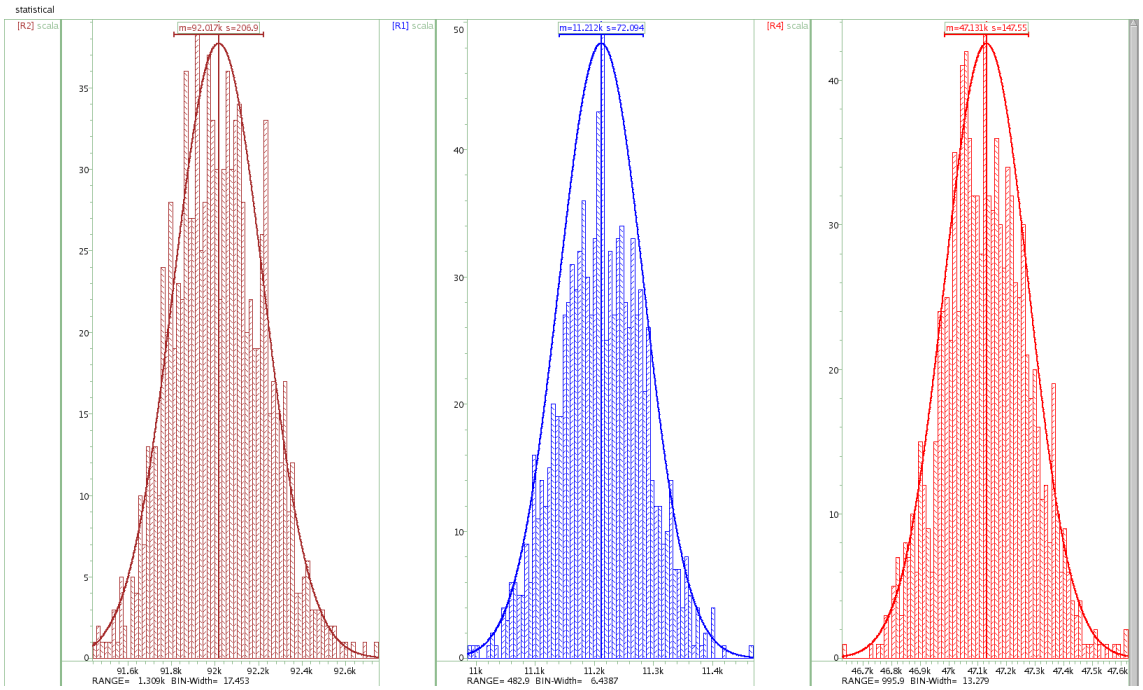


Figure A.17: Resistor Mismatch under PVT

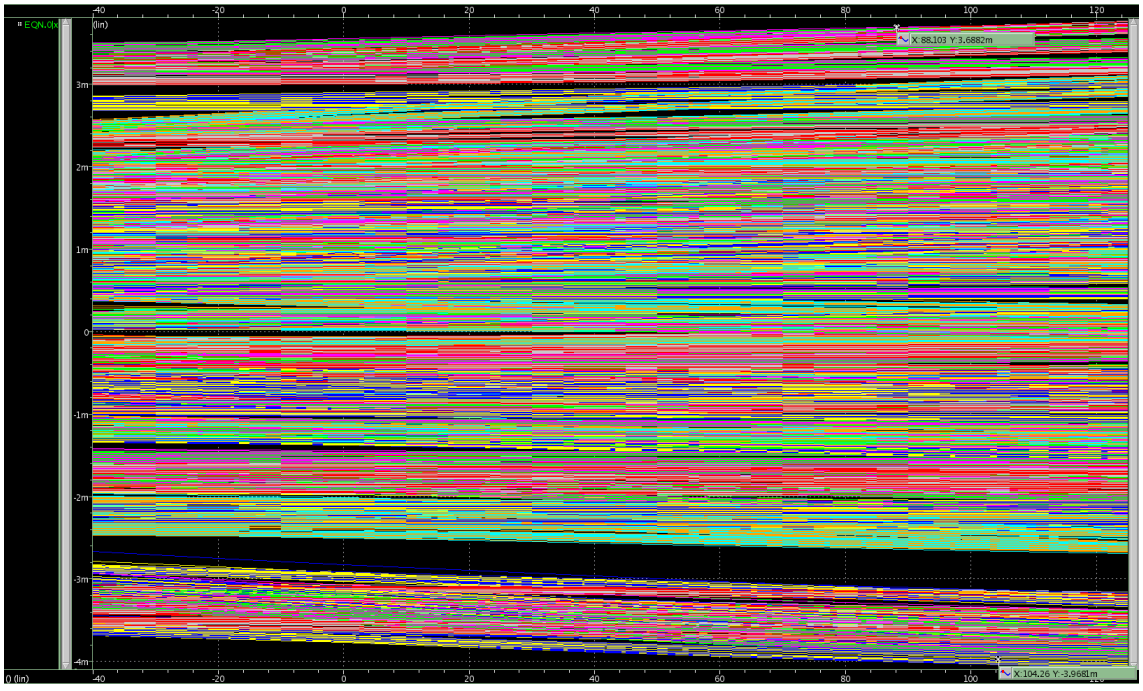


Figure A.18: Amplifier Offset under PVT

Appendix B

V-to-I Simulation Results

B.1 Calibration Module

```
// Created Tue May 26 14:11:33 2015

`include "constants.vams"
`include "disciplines.vams"

module digital_curr (input clk,input v_meas, input v_ref, output cal);
input v_meas,clk,v_ref;
output [5:0]cal;

parameter real vdd = 1.62;
parameter real clk_threshhold = vdd/2;
electrical v_meas,cal, clk,v_ref;
electrical cal[5:0];

integer result[5:0]=0;
integer en=1;
integer cali=32;
integer i=0;
integer last=0,counter1=0,prev=63;
integer state=-1;

real lastV=0, prevV=0;


analog begin
  @(cross(V(clk)-clk_threshhold,+1))
  begin
    //bandgap startup wait 50us min//
```

```
    if(state==-1 && counter1>2)
    begin
      state=0;
    end
    ////////////////////calibration check and correction/////////////
    if(state==0)
    begin
      last=prev;
      prev=cali;
      if((V(v_meas))<V(v_ref))
      begin
        cali=cali+1;
      end
      else
      begin
        cali=cali-1;
      end
      state=1;
    end
  ///////////////check stop condition/////////////////
  if(state==1)
  begin
    if(last==cali)
    begin
      state=2;
      en=0;
    end
    else
    begin
      state=0;
    end
  end

  ////////////end state machine
  ////////////bit logic conversion
  for(i=0;i<6;i++)
  begin
    if((cali >> i) & 1)
    begin
      result[i]=vdd;
    end
    else
    begin
      result[i]=0;
    end
```

```
  end
end


  //drive output signals
  for(i=0;i<6;i++)
  begin
    V(cal[i]) <+ result[i];
  end

end

////////////counter block
analog
begin
  @(cross(V(clk)-clk_threshhold,+1))
  begin
    if(en==1)
    begin
      counter1=counter1+1;
    end
  end
end

endmodule
```

B.2 Calibration Codes

CORNER	BJT	MOS	RES	2,5V @25ºC - 125k Calibration Code (Binary)	Bandgap + V-to-I Converter Post-Calibration Variation
1	FAST	FF	FAST	000010	2,40%
2	FAST	FF	SLOW	111010	1,87%
3	FAST	FF	TYP	100101	2,10%
4	FAST	FS	FAST	000010	2,40%
5	FAST	FS	SLOW	111010	1,87%
6	FAST	FS	TYP	100101	2,10%
7	FAST	SF	FAST	000011	2,38%
8	FAST	SF	SLOW	111010	1,87%
9	FAST	SF	TYP	100101	2,10%
10	FAST	SS	FAST	000011	2,38%
11	FAST	SS	SLOW	111010	1,87%
12	FAST	SS	TYP	100101	2,10%
13	FAST	TT	FAST	000010	2,40%
14	FAST	TT	SLOW	111010	1,87%
15	FAST	TT	TYP	100101	2,10%
16	SLOW	FF	FAST	000010	2,40%
17	SLOW	FF	SLOW	111010	1,87%
18	SLOW	FF	TYP	100101	2,10%
19	SLOW	FS	FAST	000010	2,40%
20	SLOW	FS	SLOW	111010	1,87%
21	SLOW	FS	TYP	100101	2,10%
22	SLOW	SF	FAST	000011	2,38%
23	SLOW	SF	SLOW	111010	1,87%
24	SLOW	SF	TYP	100101	2,10%
25	SLOW	SS	FAST	000011	2,38%
26	SLOW	SS	SLOW	111010	1,87%
27	SLOW	SS	TYP	100101	2,10%
28	SLOW	TT	FAST	000010	2,40%
29	SLOW	TT	SLOW	111010	1,87%
30	SLOW	TT	TYP	100101	2,10%
31	TYP	FF	FAST	000010	2,40%
32	TYP	FF	SLOW	111010	1,87%
33	TYP	FF	TYP	100101	2,10%
34	TYP	FS	FAST	000010	2,40%
35	TYP	FS	SLOW	111010	1,87%
36	TYP	FS	TYP	100101	2,10%
37	TYP	SF	FAST	000011	2,38%
38	TYP	SF	SLOW	111010	1,87%
39	TYP	SF	TYP	100101	2,10%
40	TYP	SS	FAST	000011	2,38%
41	TYP	SS	SLOW	111010	1,87%
42	TYP	SS	TYP	100101	2,10%
43	TYP	TT	FAST	000010	2,40%
44	TYP	TT	SLOW	111010	1,87%
45	TYP	TT	TYP	100101	2,10%

Table B.1: Calibration Codes for all corners

B.3 Results

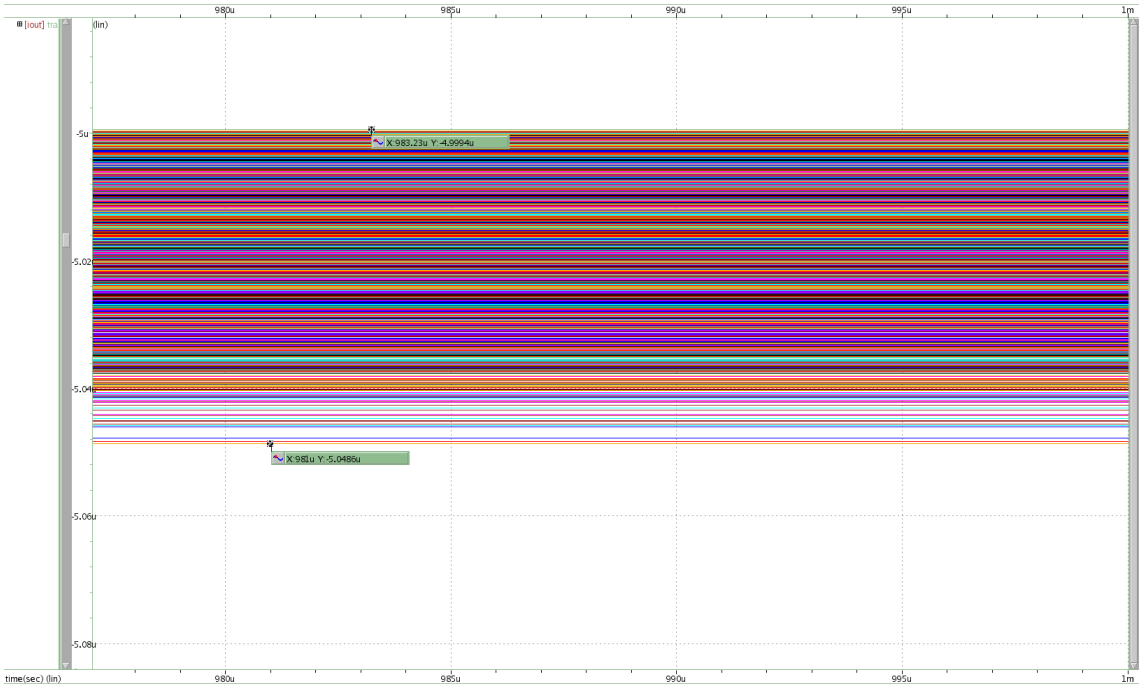


Figure B.1: Overall Precision under Device Mismatch

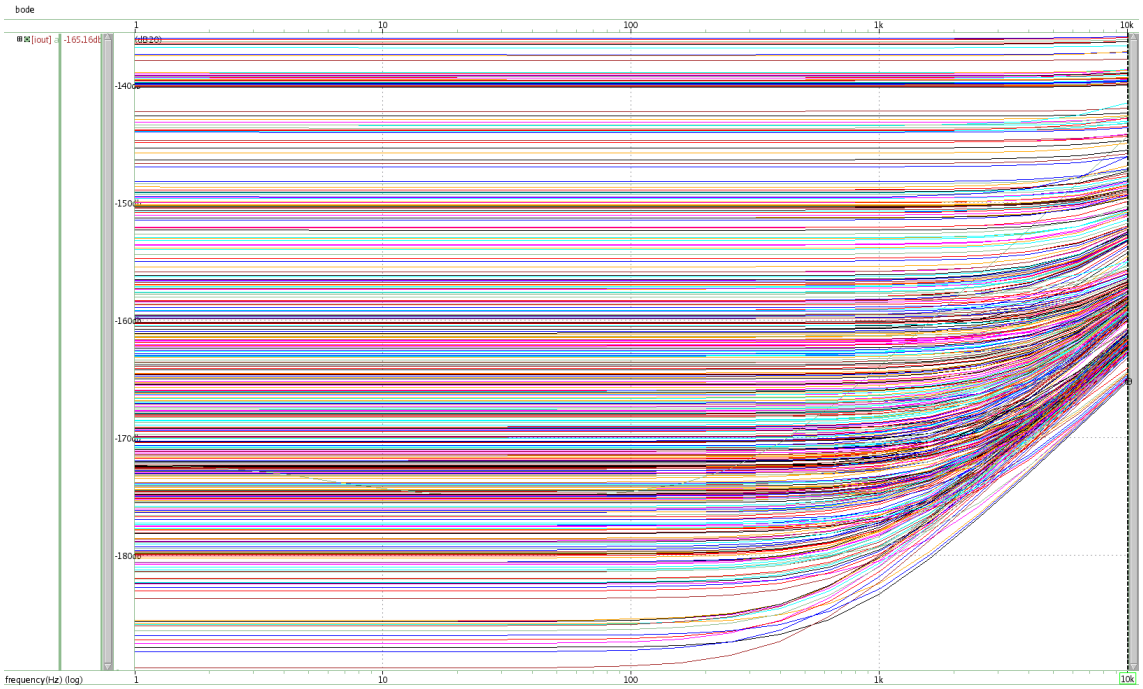


Figure B.2: PSRR under PVT

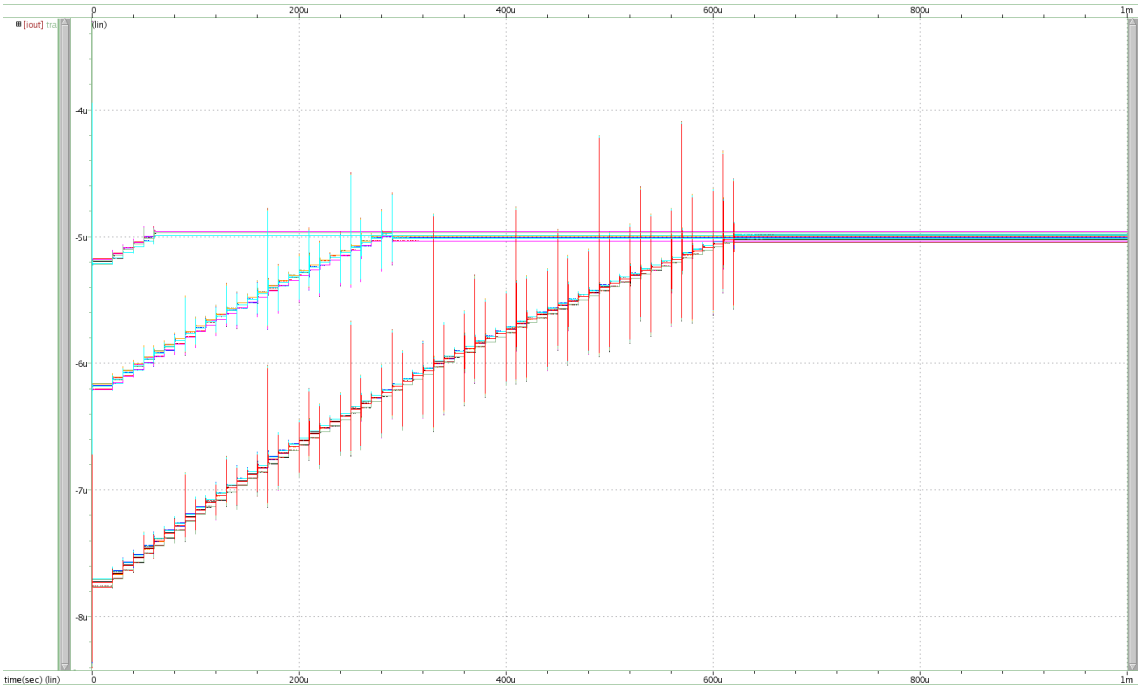


Figure B.3: Calibration under All Corners

Appendix C

Chopping Simulation Results

C.1 Offset Behavior

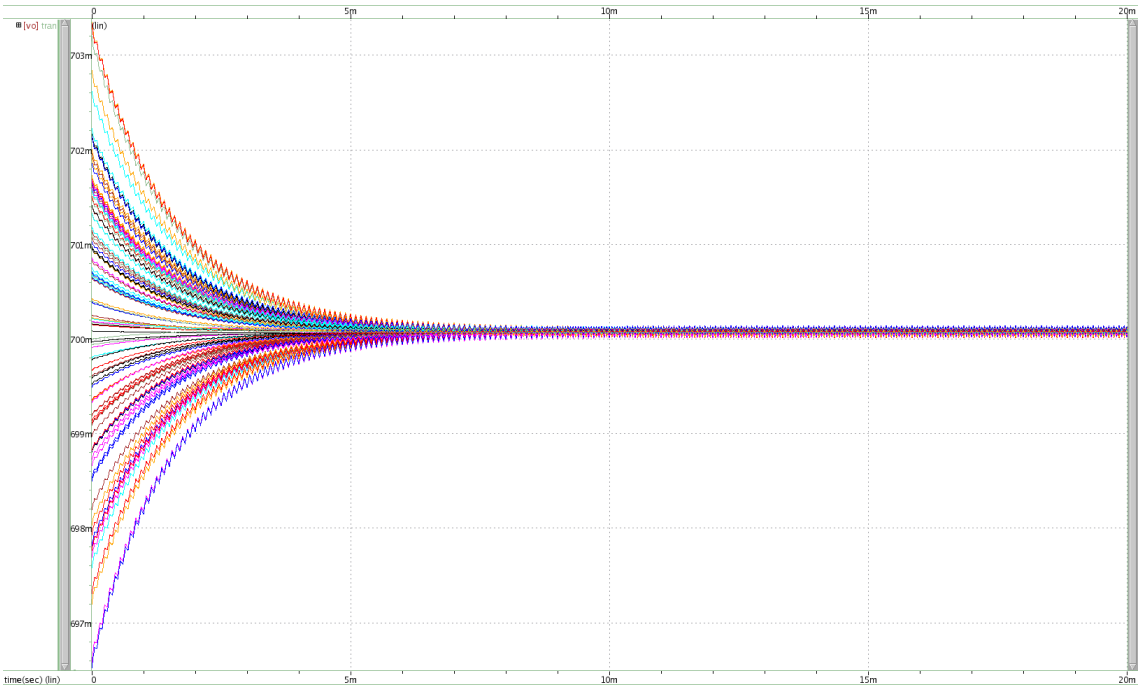


Figure C.1: Offset Behavior after Low Pass Filter

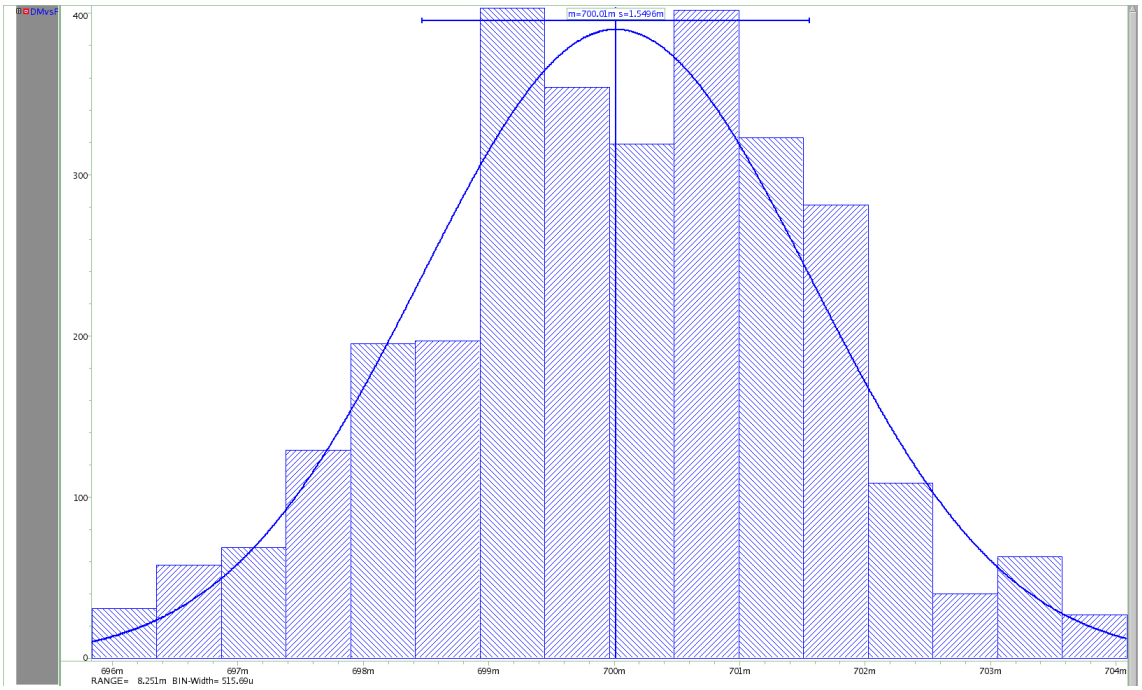


Figure C.2: Offset before chopping under PVT 1000 MC

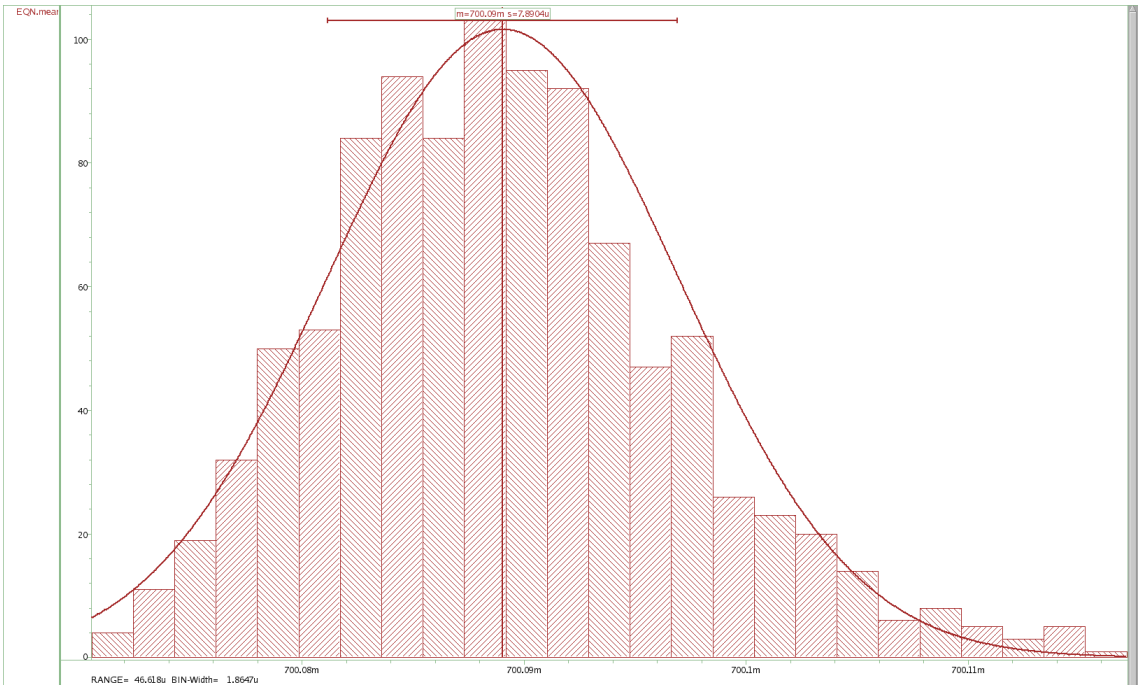


Figure C.3: Offset after chopping under PVT 1000 MC

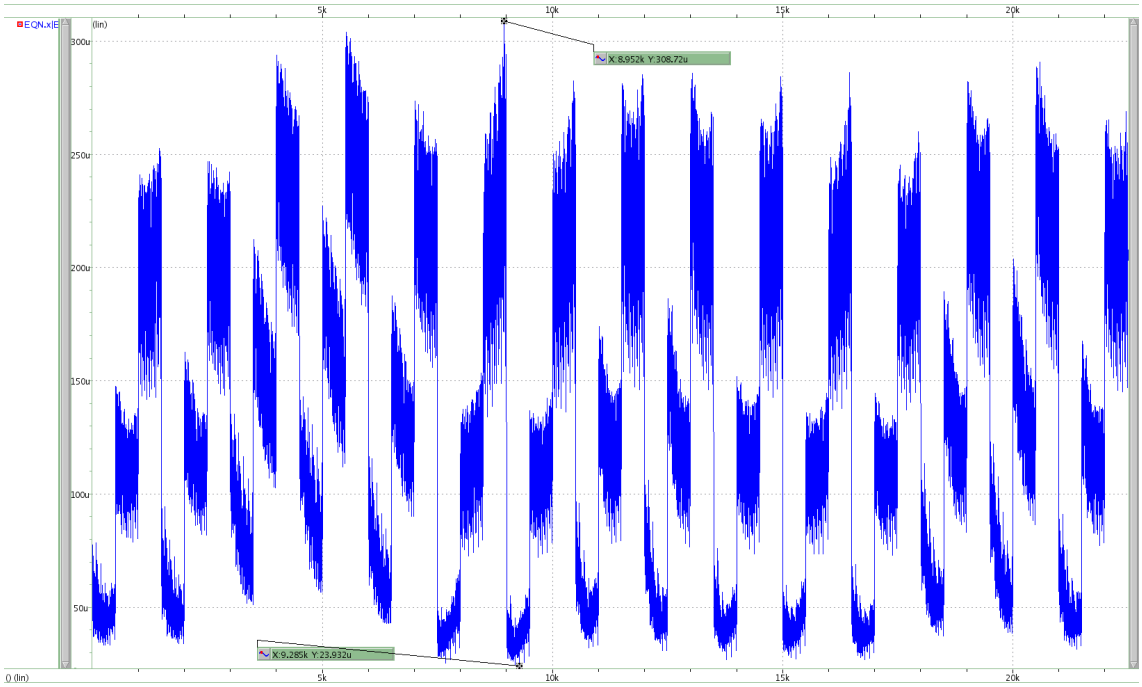


Figure C.4: Offset under PVT

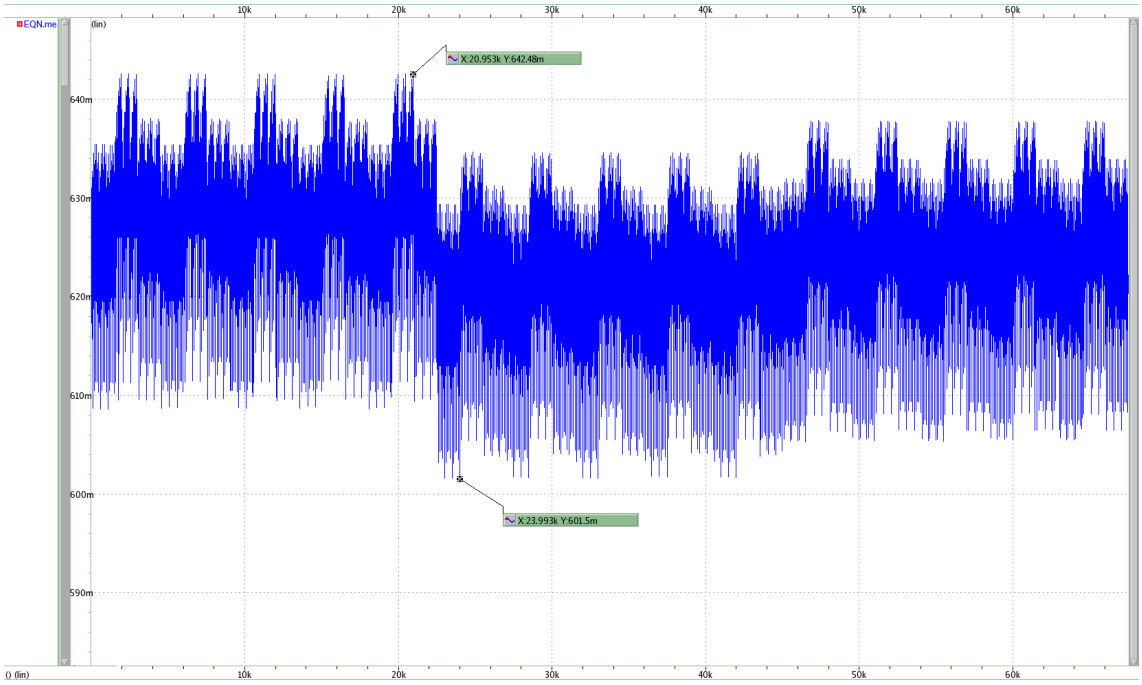


Figure C.5: Bandgap Voltage Variation with Chopped Amplifier under PVT

C.2 Dynamic Bandgap Voltage Reference

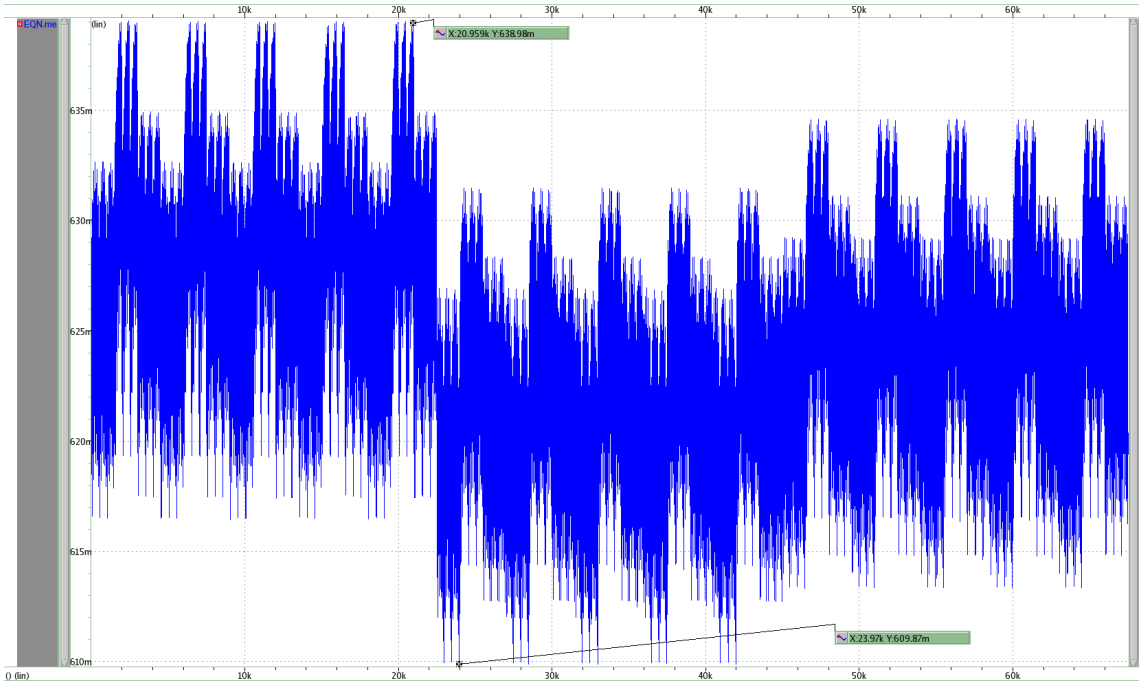


Figure C.6: Dynamic Bandgap Voltage Variation with Chopped Amplifier under PVT

Appendix D

Layout of the Folded Cascode Amplifier

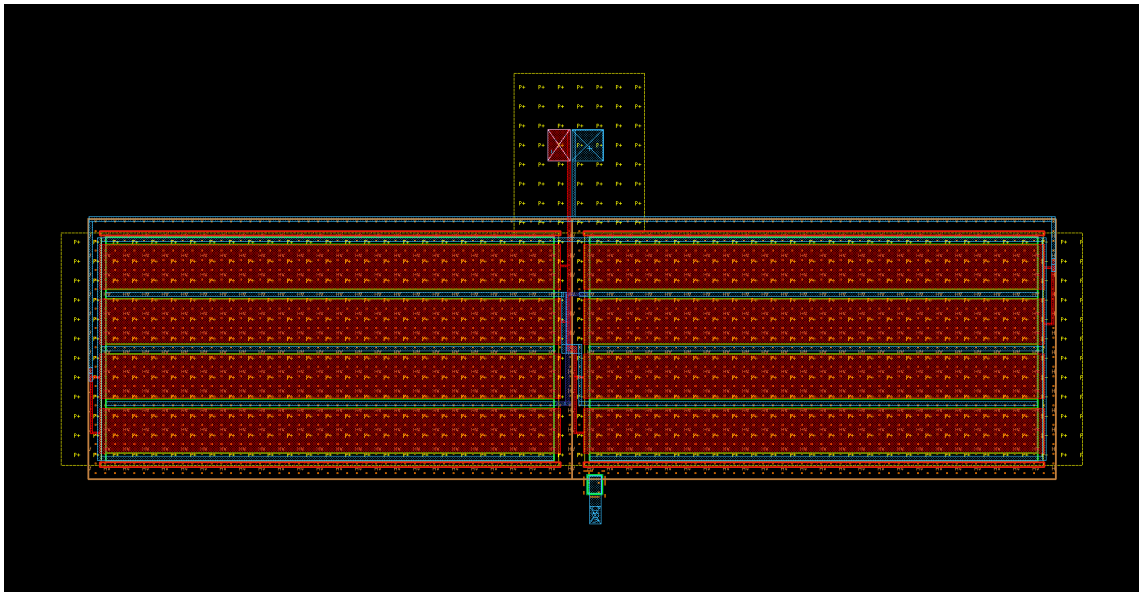


Figure D.1: Layout of the PMOS input differential pair

Layout of the PMOS input differential pair. Common-centroid configuration for 2d matching. Pattern used:
AABB
BBAA

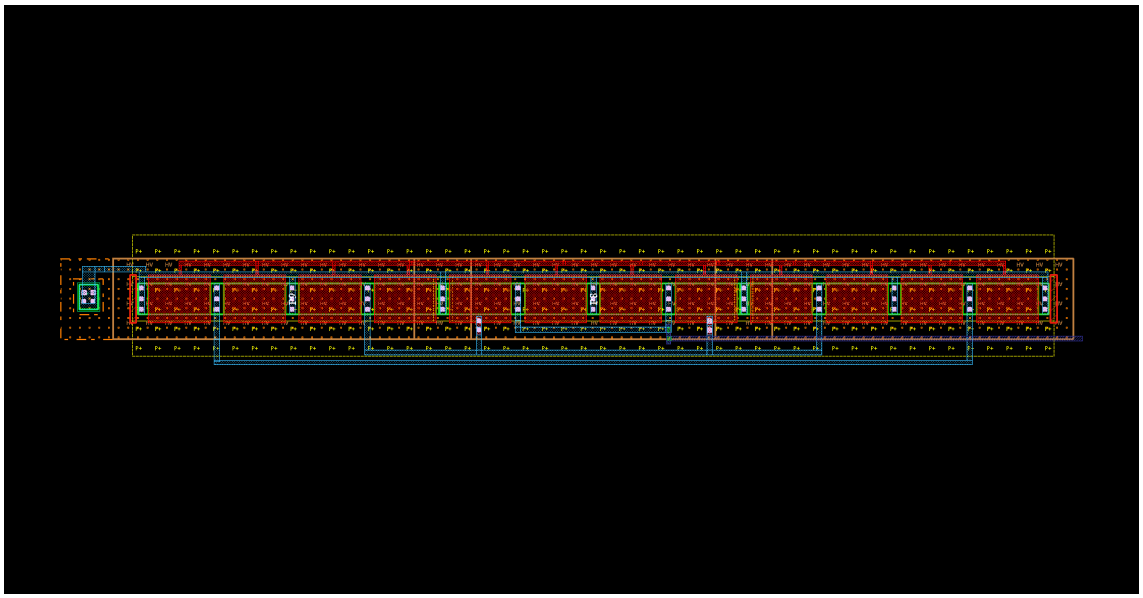


Figure D.2: Layout of the bias current mirror

Layout of the 3 pmos bias current mirrors. Interdigitated configuration for 1d matching. Pattern used: AABBCCCCBBAA

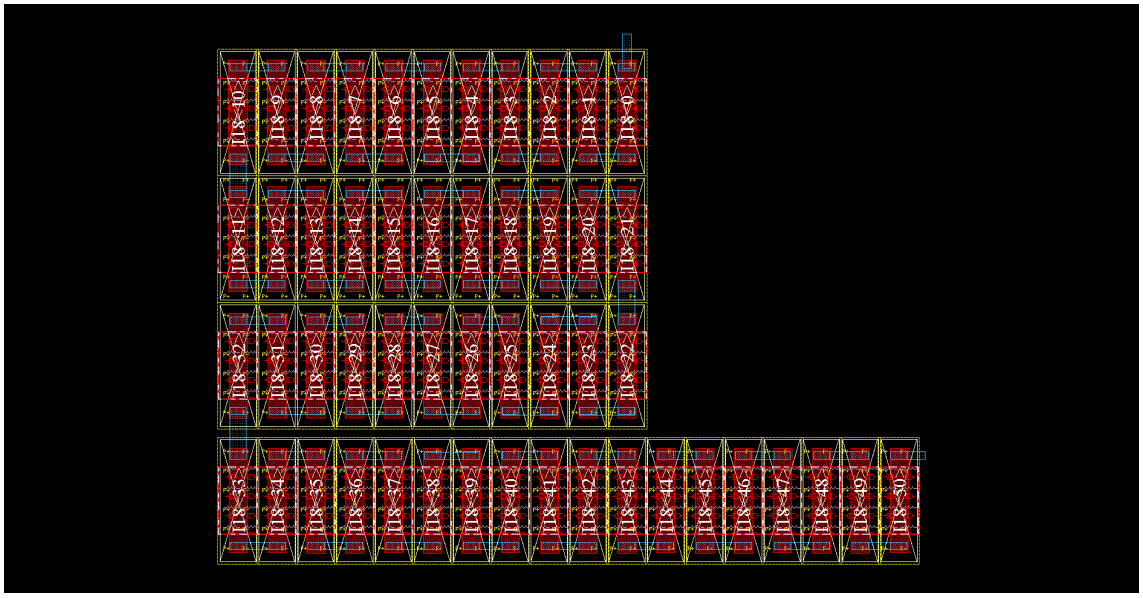


Figure D.3: Layout of the bias resistor

Layout configuration of the bias resistor. Serpentine configuration for thermoelectric cancellation.

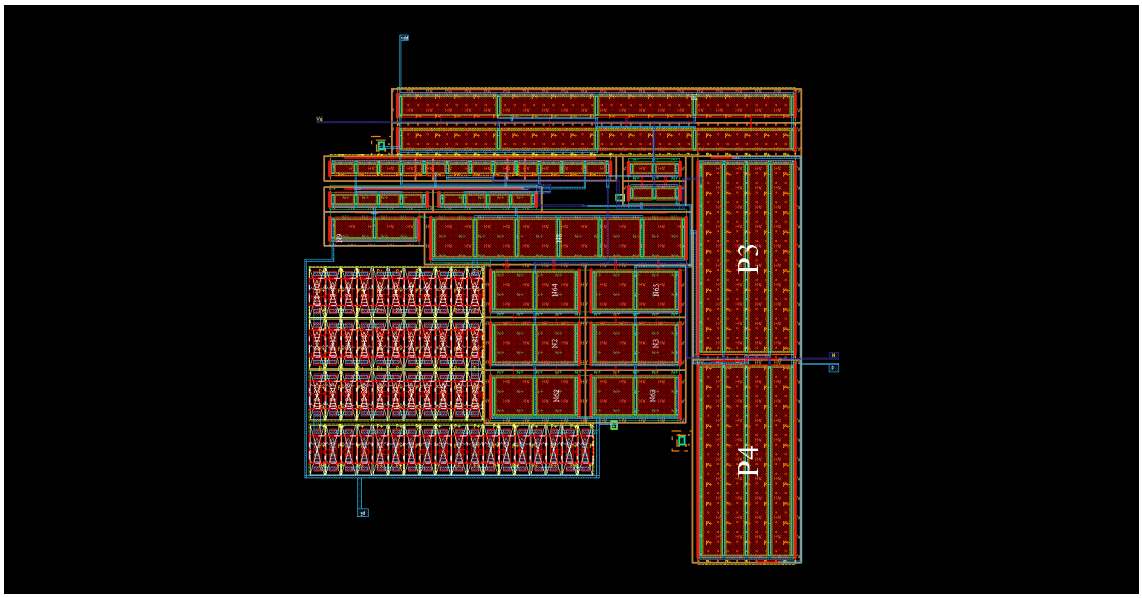


Figure D.4: Layout of the Folded Cascode Amplifier

Complete layout of the folded cascode amplifier. Input and Output signals extended with metals for better visibility.

Appendix E

Layout of the Bandgap Voltage Reference

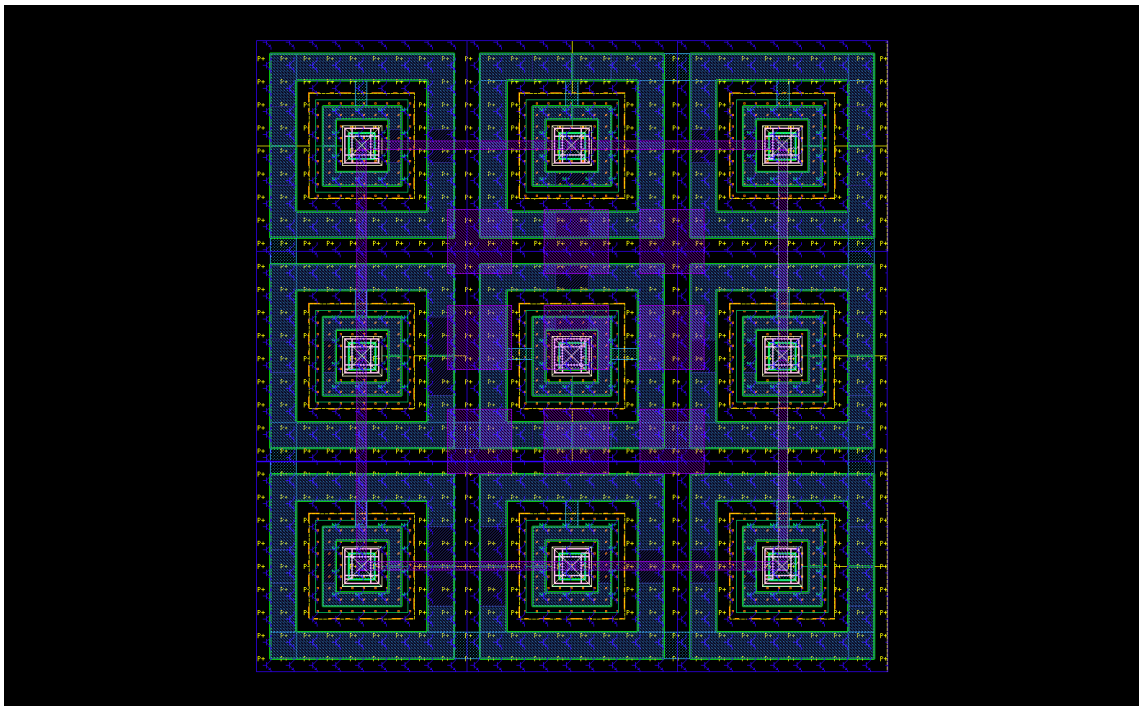


Figure E.1: Layout of the BJT devices

Layout of the BJT devices, 8-to-1 ratio. Common-centroid configuration for 2d matching. Pattern used:
BBB
BAB
BBB

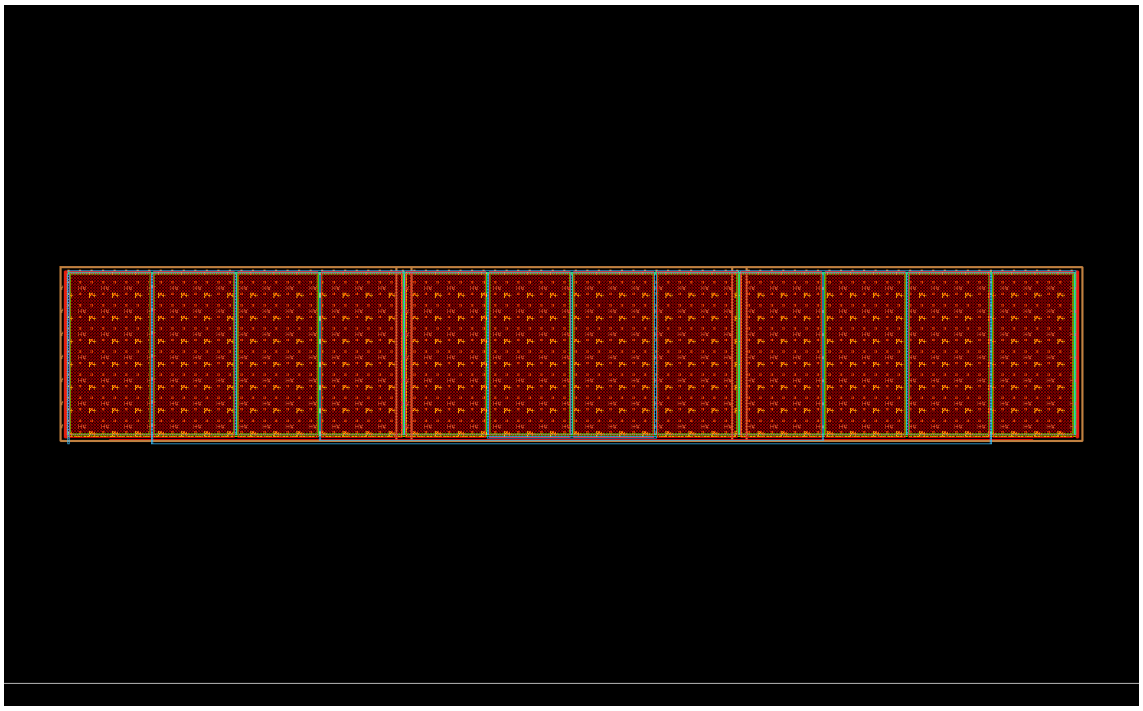


Figure E.2: Layout of the PMOS current mirrors

Layout of the PMOS current mirrors. Interdigitated configuration for 1d matching. Pattern used: AABBCCCCBBAA

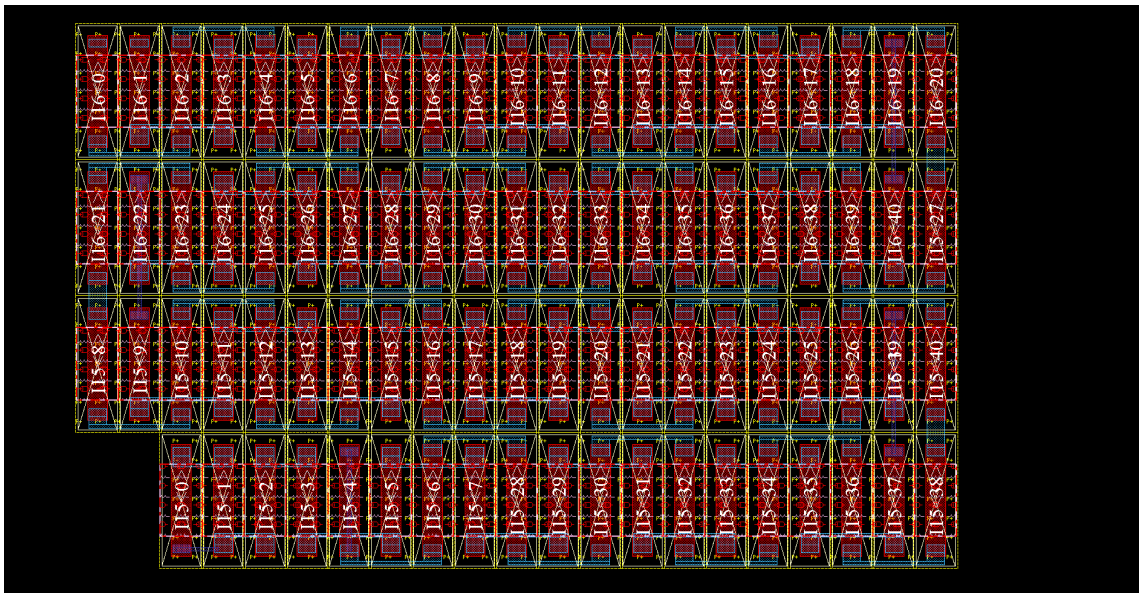


Figure E.3: Layout of the resistors R2 and R3

Layout of the resistors R2 and R3. Interdigitated configuration for matching.
Pattern used:ABABABABABABABABABAB.....

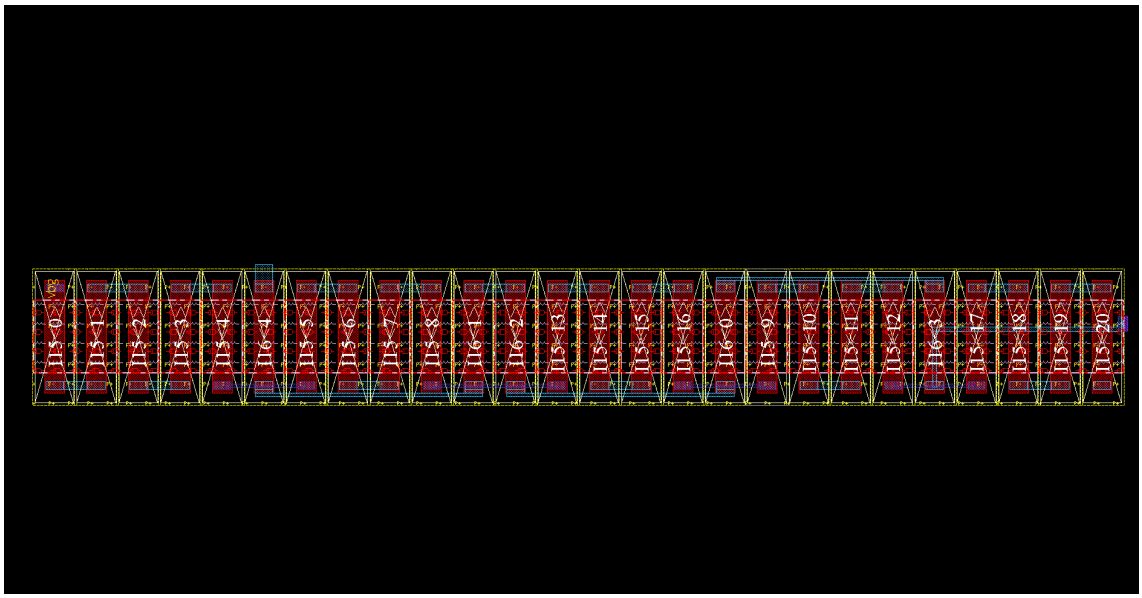


Figure E.4: Layout of the resistors R1 and R4

Layout of the resistors R1 and R4. Interdigitated configuration for matching.
Pattern used:AAAABAAAABAAAABBAAAABAAAAA

Bibliography

[1] Geelhaar F-Heiser G. Altermatt PP, Schenk A. Reassessment of the intrinsic carrier density in crystalline silicon in view of band-gap narrowing. *Journal of Applied Physics*, 2003.

[2] Abdelhalim Bendali and Yves Audet. A 1-v cmos current reference with temperature and process compensation. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS*, Vol. 54(No. 7), 2007.

[3] Abhisek Dey and Tarun Kanti Bhattacharyya. Design of a cmos bandgap reference with low temperature coefficient and high power supply rejection performance. *International Journal of VLSI design & Communication Systems*, Vol. 2.

[4] A.de la Plaza G. Torelli. Tracking switched-capacitor cmos current reference. *IEE Proc. -Circuits Devices Syst.*, 145(1), 1998.

[5] Yi-Meng Zhang Hao Zhang, Meng-Shu Huang and Tsutomu Yoshihara. A nano-power switched-capacitor voltage reference using mos body effect for applications in subthreshold lsi. *JOURNAL OF SEMICONDUCTOR TECHNOLOGY AND SCIENCE*, 14, 2014.

[6] Sergio Bampi Juan Pablo Martinez Brito. A 4-bits trimmed cmos bandgap reference with an improved matching modeling design. *Circuits and Systems, 2007*, 2007.

[7] Jon Duster Ping K Ko Kenneth Chain, Jian-hui Huang and Chenming Hu. A mosfet electron mobility model of wide temperature range (77–400 k) for ic simulation. *Semiconductor Science and Technology*, November 1996.

[8] Tomasz Borejko Michał Łukaszewicz and Witold A. Pleskacz. A resistorless current reference source for 65 nm cmos technology with low sensitivity to process, supply voltage and temperature variations. *Design and Diagnostics of Electronic Circuits & Systems (DDECS)*, 2011.

[9] A. Olesin, K.K.L. Luke, and R.D. Lee. Switched capacitor precision current source. 1983. US Patent 4,374,357.

[10] S. Lewis P. Gray, P. Hurst and R. Meyer. *Analisys and Design of Analog Integrated Circuits*. Fifth edition edition.

[11] Kulbhushan Misri Qadeer Ahmad Khan, Sanjay Kumar Wadhwa. A low voltage switched-capacitor current reference circuit with low dependence on process, voltage and temperature. *Proceedings of the 16th International Conference on VLSI Design*, 2003.

[12] M. E. Schlarmann S. Q. Malik and R. L. Geiger. A low temperature sensitivity switched-capacitor current reference. *European Conference on Circuit Theory and Design*, 2001.

[13] Suhas Vishwasrao Shinde. Pvt insensitive reference current generation. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. II, 2014.

[14] C. D. Thurmond. The standard thermodynamic functions for the formation of electrons and holes in ge, si, gaas and gap. *J. Electrochem. Soc.*, 122, 1975.

[15] David A.Johns Tony C. Carusone and Kenneth W. Martin. *Analog Integrated Circuit Design*. Wiley, second edition edition, 2011. ISBN: 978-0-470-77010-8.

[16] Gabriel A. Rincon–Mora Vishal Gupta. Predicting the effects of error sources in bandgap reference circuits and evaluating their design implications. *Georgia Tech Analog Consortium*.

[17] Zhao Qi Wang Xing-hua, Zhong Shun-an and Qu Ruoyuan. A low-variation, low-voltage cmos bangap reference circuit. *ICSE 2010*.

[18] Wilmar Bueno de Moraes Wellington Avelino do Amaral, José António Siqueira Dias. Cmos voltage reference based on threshold voltage summation. *Proceedings of the 8th WSEAS International Conference on Microelectronics, Nanoelectronics, Optoelectronics*.

[19] Wen Zhiping Wen Wu and Zhang Yongxue. An improved cmos bandgap reference with self-biased cascoded current mirrors. *Electron Devices and Solid-State Circuits*, 2007.

[20] R. J. Widlar. An exact expression for the thermal variation of the emitter base voltage of bi-polar transistors. *Proceedings of the IEEE 55*, January 1967.

[21] D. Wolpert and P. Ampadu. *Managing Temperature Effects in Nanoscale Adaptive Systems*. 2012. ISBN:978-1-4614-0747-8.

[22] D.J. Comer Y. Dai, D.T. Comer and C.S. Petrie. Threshold voltage based cmos voltage reference. *Circuits, Devices and Systems, IEE Proceedings*, 151, 2004.