Always-On Voltage Reference for Ultra Low Power Application

Always-On Voltage Reference

Jurgen Asko

Thesis submitted for the degree of
Master in Electrical Engineering, Information and Technology
60 credits

Department of Physics
The Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

Autumn 2022

Always-On Voltage Reference for Ultra Low Power Application

Always-On Voltage Reference

Jurgen Asko

Always-On Voltage Reference for Ultra Low Power Application

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

The objective of this thesis is to study and implement an ultra-low power, always-on voltage reference circuit for ultra-low power applications. The design studied and implemented is a 3 transistor (3T) topology operating in weak inversion. This circuit topology has been a central topic of scientific research in recent years. Along with the voltage reference circuit, a trimming technique was proposed. The objective for trimming was twofold: to achieve a stable temperature coefficient and to adjust the absolute voltage level of the voltage reference. Therefore two circuits were studied , the core cell 3T topology standing at its own and the 3T design along with the trimming circuits attached to it, with the latter being the one that was taped out. The voltage reference was designed, simulated and taped-out in a 180nm BCD Gen2 TSMC semiconductor process. Simulation results showed that the design can operate at a very low power consummation of $33.1pW$ (best) and $270pW$ (with trimming topology) while having an output voltage as high as $1.1V$. The design also achieves a low temperature coefficient of $7ppm/^oC$ in a temperature range from -20^oC to 85^oC. This performance, makes the design suitable for IOT and other ultra-low power applications, such as Bluetooth Low Energy(BLE), biomedical implants etc.

List of Figures

List of Tables

List of Abbreviations

3T	3-Transistors
ASIC	Application Specific Integrated Circuit
BGR	Bandgap Reference
BJT	Bipolar Junction Transistor
BLE	Bluetooth Low Power
CMOS	Complementary Metal Oxide Semiconductor
CTAT	Complementary to Absolute Temperature
DC	Direct Current
IC	Integrated Circuit
KVL	Kirchhoff's Voltage Law
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
OPAMP	Operational Amplifier
PCB	Printed Circuit Board
PSRR	Power Supply Rejection Ratio
PTAT	Proportional to Absolute Temperature
PVT	Process, Voltage, Temperature Variations
RT	Room Temperature
SOC	System On Chip
TC	Temperature Coefficient
VDD	Power Supply Voltage

List of Symbols

Symbol	Description	Units
C_d	Depletion Capacitance per Unit Area	F/cm^2
C_{ox}	Gate Oxide Capacitance per Unit Area	F/cm^2
V_f	Flat-band voltage	V
I_{gs}	Drain to source voltage	V
n_i	Intrinsic carrier concentration	cm^{-3}
t_{ox}	Gate txide thickness	$V/decade$
I$_{Sub}$	Sub-threshold current	A
K	Boltzmann constant	J/K
m	Subthreshold Slope Factor $1 + C_{ox}/C_d$	
S	Subthreshold Slope	$v/decade$
T	Absolute temperature	K
V$_{gs}$	Gate to Source Voltage	V
V$_{h,REF}$	Drain to Source current	V
V$_{REF}$	Reference voltage	V
V$_{TH}$	Threshold Voltage	V
V$_{T}$	Thermal voltage (kT/q)	V

Contents

Chapter 1

Introduction

In recent years, there has been a great interest in the Internet of Things (IOT), as well as other low power applications, such as biomedical implants and energy harvesting systems [5]. Maurizio et al. [10] reported that more than 100 billion electrical devices are expected to be able to connect to the internet by 2050 (figure 1.1).

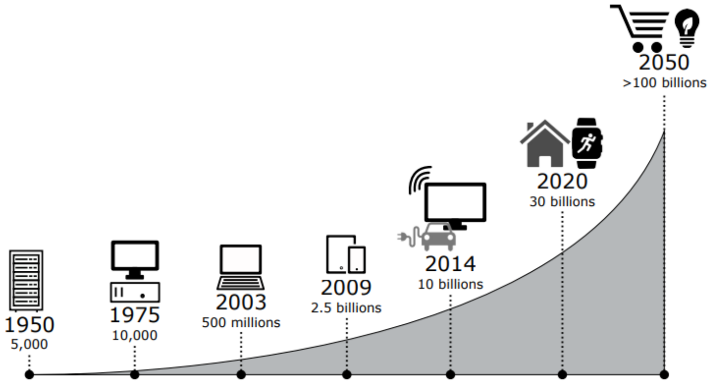


Figure 1.1: Expected IOT Growth
[10]

These devices are mostly battery powered, a fact that raises ethical and practical concerns about their energy profile and sustainability. Because battery replacement could become costly and most of the time impractical, there is an increasing need for IOT devices and other low power applications to operate on a low power budget. In order to operate, the majority of these systems use analog and mixed signal modules, such as analog to digital converters (ADCs) and radio frequency (RF) transmitters and receivers. These modules need to consume as little power as possible in order to optimize battery life [27]. Voltage reference circuits are essential building blocks for the proper operation of these modules and hence for the applications of IOT devices. Designing an always-on voltage reference circuit, which consumes a small amount of power is crucial for the overall power consumption of an IOT device. For example, for low-power chips/systems on chip(SoCs), used in IOT devices the sleep current dominates the power budget [13]. The reason for this is that SoCs stay in low power (sleep) mode 99% of the time [13]. Always-on voltage reference would reduce start-up time of the SoC when waking up from sleep, as it would not require additional time for the voltage reference to settle. Settling time of conventional non-always-on voltage references can dominate the waking-up time of the SoC and, hence, increase the total time that the chip stays in active power hungry state.

For the reason above, it is essential to reduce the overall power consumption of the IOT devices in order to make IOT more sustainable and easy to work with. In recent years, a considerable amount of research has been focused on voltage reference circuits, which consume power in the picowatt and/or even in the femtowatt regime [27] [34]. The main aim of this thesis is, at first, to investigate some of these, ultra-low power voltage reference circuit designs and, at second, to design and implement an ultra-low power, always-on voltage reference in a 180nm BCD2 GEN process along with a novel trimming circuit. The goal is to create a reproducible always-on voltage reference with a power consumption in the picowatt range.

1.1 Objectives

- Gain theoretical knowledge on complementary metal-oxide semiconductor (CMOS) voltage reference circuits and focus on ultra-low power topologies .
- Design a voltage reference circuit which is suitable for always-on operation.
- Design a trimming circuit that can trim the reference in the temperature and voltage domain.
- Design of the application specific integrated circuit and PCB.
- Perform measurements of the designed voltage reference to characterize the performance and compare it with the simulation results.

1.2 Thesis Structure

The project is divided into three main parts. The first part, focuses on the basics of voltage reference circuits, as well as their importance and usage in modern electronic systems. A brief history of voltage references along with the most common designs used on integrated circuits (ICs) are discussed. The second part explores recent recent ultra-low power topologies of voltage reference circuits and focuses on the design and implementation of the system (voltage reference circuit and trimming circuit). In the last part, results are presented and potential future improvements are discussed.

1.3 Tools Used

- Cadence Virtuoso is a simulation environment used widely for integrated circuits (IC) design.
- Calibre Tools for Layout Verification.
- Eagle PCB design software for the PCB design.
- Microsoft Visio and Adobe Photoshop for creating and editting the figures and diagrams used in the project.

Chapter 2

Theoretical Background

2.1 Voltage References

Voltage reference is a circuit that generates a precise DC voltage, which is not dependent on the supply voltage, temperature (or alternatively has a well defined dependence on temperature) and/or process variations. Voltage reference circuits are excessively used in electronic circuits and are of high importance for biasing modules and systems, especially voltage is needed for measurements to be done against [3, 25].They are used both in circuit board design and integrated circuits (ICs). In modern IC technologies, the voltage reference circuits are mainly implemented by combining MOSFETs and BJT or opamps in ICs. They are a key building block for almost any integrated system. Analog to digital (A/D) and digital to analog (D/A) converters, smart sensors, RF transmitters and receivers, and many other systems (varying from commercial to military applications) use voltage reference circuits in order to reach precise measurements and an efficient operation[18, 15]. The most common voltage reference circuit used in modern ICs is the bandgap voltage reference or BGR for short.

A block diagram of the BGR operation is presented in the next figure (figure **??**). BGR's operation is based on the bandgap energy of silicon, which is around 1.2V [14]. That is why the output voltage of a BGR is around 1.2V or an up/downscale version of 1.2V [14]. The BGR combines the complementary to absolute temperature response (CTAT) of the base-emitter voltage (V_{BE}) of a BJT transistor with a proportional to absolute temperature (PTAT) source (e.g., the thermal voltage V_T) as it is shown in figure 2.1. By adjusting and adding these two the opposite slopes, BGRs can produce an output stable in a specific temperature range.

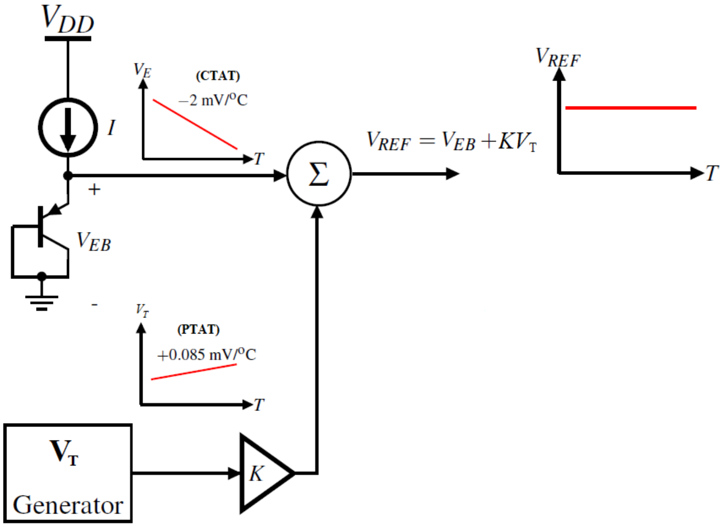


Figure 2.1: Bandgap Voltage Reference (BGR) Operation
[23]

Limitations of BGR circuits, such as minimum supply voltage, high output of around 1.2V and high power consumption, led to a drift towards designs that use only CMOS transistors.

2.2 Performance Parameters

As mentioned earlier, voltage references are fundamental building blocks of integrated circuits and they are used as reference. Processing data with analog to digital converters (ADCs) and digital to analog converters (DACs), as well as biasing other modules, requires voltage and currents references of a high accuracy in order to achieve the desired operation standards. Voltage references are crucial when defining the accuracy of these modules. For example, in an analog to digital converter a really precise voltage reference is needed in order to compute the least significant bit(LSB).

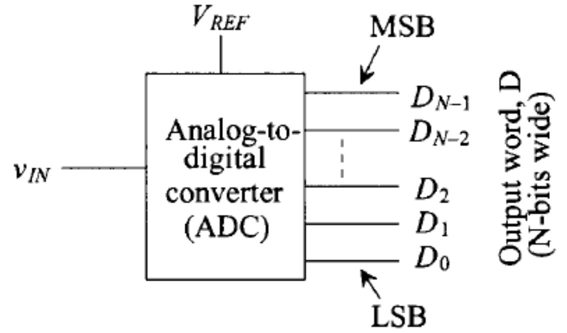

V_{REF}
MSB
Analog-to-digital converter (ADC)
v_{IN}
D_{N-1}
D_{N-2}
D_2
D_1
D_0
LSB
Output word, D (N-bits wide)

Figure 2.2: Block Diagram of An Analog to Digital Converter
[3]

Figure 2.2 shows a simple block diagram of an ADC. The LSB that defines the accuracy of the output data can be found as ($V_{LSB} = V_{REF}/2^N$), where N is the number of bits. The reference voltage is essential for that conversion. If the voltage reference is not stable for any reason, it results in a loss of information in the output word. It is then a need to set some performance parameters for voltage reference circuits, which define whether the circuits perform accurately enough or not. Some of the most common parameters that define the accuracy of a voltage reference are discussed below. Line regulation, temperature coefficient and power consumption are considered to be the most important[18]. Depending on the given application, one might be interested in achieving accuracy on some or only on one of these parameters.

2.2.1 Temperature Coefficient

It is important for a voltage reference to produce an output voltage level, which does not vary within a wide range of temperature. However, CMOS device parameters are heavily dependent on temperature [2]. Device parameters, such as threshold voltage (V_{th}) and carrier mobility (μ), can be influenced by the temperature. Table 2.1 demonstrates the desired operating conditions for certain temperatures of electronic systems in different fields. Temperature coefficient (TC) is defined as a change of voltage (typically in ppm) per degree Celsius, i.e., $ppm/^oC$. As shown in equation 2.1, TC takes into consideration the highest and the lowest voltage deviation across the full range of temperatures, in which the circuit is intended to be used. Small values of TC mean high stability of the voltage reference across temperature, and thus the voltage does not deviate from its nominal value over temperature [18]. Therefore, a voltage reference is critical to remain stable over temperature. It is important to mention that process variations can also affect the temperature response and stability of a system.

$$TC = \frac{V_{REF(max),V_{IN(nom)}} - V_{REF(min),V_{IN(nom)}}}{(T_{max} - T_{min})V_{REF(nom)}} \times 10^6 (ppm/^oC) \tag{2.1}$$

Standard	Minimum	Maximum
Commercial	0^oC	70^oC
Industrial	-40^oC	85^oC
Military	-55^oC	125^oC

Table 2.1: Operating Temperature Range Standards [35]

2.2.2 Line Sensitivity/Line Regulation

Line regulation or line sensitivity describes the ability of the voltage reference circuit to produce the same output voltage (V_{REF}) over a range of input voltages ((ΔV_{IN})) at a nominal temperature (25^oC). Line regulation is defined as the variations of the output voltage reference with respect to the input voltage range, and can be calculated by the equation 2.2 [18].

$$LS = \frac{\Delta V_{REF}}{\Delta V_{VDD} \times \Delta V_{REF\mu}} (\times 100(\%)) \tag{2.2}$$

Where ΔV_{REF} is the range of the output voltage measured within the range of the input voltage ΔV_{IN} and $\Delta V_{REF\mu}$ is the mean value of the output voltage. Small values of line sensitivity LS indicate small variations of the output voltage within a given range of input voltages and thus a less sensitive reference circuit in terms of input voltage variations. It should be noted that the voltage reference output can be also affected by the capacitive and resistive load, C_{Load} and R_{Load} respectively, the operating temperature, and the load current, I_{Load}. Thus, line regulation LS is measured with these operating parameters always specified.

2.2.3 Power Supply Rejection Ratio (PSRR)

Power supply rejection ratio (PSRR) describes the ability of a voltage reference circuit to reject noise of a specific frequency range. This noise is usually found in the power rails and it is caused by signal coupling, power surge and/or other noise sources. PSRR is expressed in decibels (dB) and can be described as:

$$PSRR(f) = 20 \log \frac{V_{REF,AC}}{V_{IN,AC}} (dB) \tag{2.3}$$

where $V_{IN,AC}$ is the AC component of the supply voltage i.e the noise coupled at the supply rails and $V_{REF,AC}$ is the AC component of the voltage output [34]. The frequency range of PSRR depends on the application of the voltage reference.

2.2.4 Output Noise

Output noise is a frequency dependent performance parameter, which defines the amount of undesirable components of the output signal. The output noise of a voltage reference circuit is measured with respect to its root mean square (RMS) value [34].Noise is random and thus the peak to peak noise voltage can be estimated by multiplying the RMS value by 6. For example, a voltage reference with $2\mu VRMS$ at $10Hz - 20kHz$ noise density will have a peak-to-peak noise voltage of approximately $12\mu V$.

2.2.5 Quiescent Current

The quiescent current, I_q is the current needed for the steady operation of a voltage reference circuit without a resistive load being connected to its output [18]. I_q is defined as the current drawn from the reference circuit under nominal conditions, i.e., $V_{IN(nom)}$, T_{nom} and $I_{Load} = 0A$. It results in a nominal power consumption ($V_{IN(nom)} \times I_q$) of the reference circuit. It is important to keep the quiescent current as low as possible in order to reduce overall power consumption and thus achieve long working hours for battery operated applications[18].

2.3 More Design Considerations

Some other design considerations, such as circuit area, power dissipation, device mismatch and ease of output trimming, need to be taken into account when designing voltage reference circuits [18]. As described by Chi-Wah Kok and Wing-Shan Tam [18], circuit size and power dissipation not only play a keu role in the market value of a product, but are also critical for the output noise of the reference circuit. Both device mismatch and ease of output trimming are related to process variations. All devices are sensitive to process variations [18], which consequently causes device parameters to deviate from each other and creates device mismatches, which can affect the performance of the reference circuit. To address these mismatch problems, trimming is used. Through trimming it is possible to compensate for these mismatches and bring the voltage reference output closer to the intended values. However, trimming can be somewhat challenging. Trimming not only requires more silicon area and hence it has a higher cost, but it also introduces more noise to the reference circuit. As both process variations and trimming methods play a crucial role on voltage reference circuits, they are further discussed in 2.3.1 and 2.3.2, respectively.

2.3.1 Process Variations

One of the challenges of designing integrated circuits is the so-called process variations. When manufacturing ICs and proceeding from a pure silicon to an actual CMOS device, there are several fabrication steps that are important to be completed. These particular steps give the silicon its electrical characteristics and final shape [3][2]. Process variations describe the devoid to totally control fabrication steps, such as the diffusion of dopants, etching, mask alignment etc[2]. These variations can lead to significant alterations of the electrical characteristics of the parameters of IC device, such as the threshold voltage, oxide thickness, transistor dimensions, sheet resistance etc[2]. Put that into perspective, in a $2\mu m$ CMOS process, a polysilicon line can vary up to 30% of its drawn value and a metal line up to 20% affecting the width to length ratio (W/L) of a device [2]. Process variations are usually divided into two categories, inter-die and intra-die variations [6].

Inter-die variations refer to the fluctuations of the device parameters that occur between different dies, wafers (i.e., "wafer to wafer") and can be expressed as a random variable, (see equation2.4) [28], [1].

$$P = P_{nom} + \Delta P_{inter} \tag{2.4}$$

In the above equation, P_{nom} is the nominal value of the process parameter under consideration and P_{inter} is a random variable with a zero mean value, which is usually represented by a Gaussian distribution with a given standard deviation. The P_{inter} has a single value for all components on the die. Inter-die variations tend to shift a parameter value equally across all devices on one die. The threshold voltage (V_{th}), for example, will deviate in the same direction (increasing or decreasing) across all transistors in the same die. As the chip is usually placed randomly in the wafer, it is assumed that each inter-die variation factor caused by different physical and independent sources. Materials and gas flow variation(linear variation), wafer spin process and exposure time(radial variation) are some of the sources of the inter-die variations.

On the other hand, intra-die variations cause device parameters to deviate from their designed values across different locations in the same die. Intra-die variations are responsible for mismatches to supposed identical devices and are described as a random variable as it is shown in the equation 2.5 [4][1],

$$P = P_{nom} + \Delta P_{inter} + \Delta P_{spatial}(\chi_i y_i) + \Delta P_{random,i} \tag{2.5}$$

where $\Delta P_{spatial}$ represents intra-die variation that consists of a spatially correlated component which is a function of the location on the die. P_{random} represents a random component which has no correlation with the other devices and is considered a random variable for each device.

The Inter-die variations have been considered more significant than the intra-die variations. However, as technology scales down, intra-die variations are becoming comparable, and in some cases, they can be more critical than the inter-die ones. In a 130nm process, intra-die variations can cause approximately 30% of the overall performance variations[1]. Wafer-level and layout-dependent variations are the two main sources of intra-die variations. Wafer-level variations arise as a result of lens aberrations and produce small trends, which represent the spatial range across the die[1]. Layout-dependent or die-pattern variations occur by cause of lithographic and etching techniques during the fabrication process. For example, photolithographic interactions and plasma etch micro-loading can induce significant alterations in two identical metal lines of the same die. Figure 2.3 demonstrates the differences between inte-die and intra-die variations.

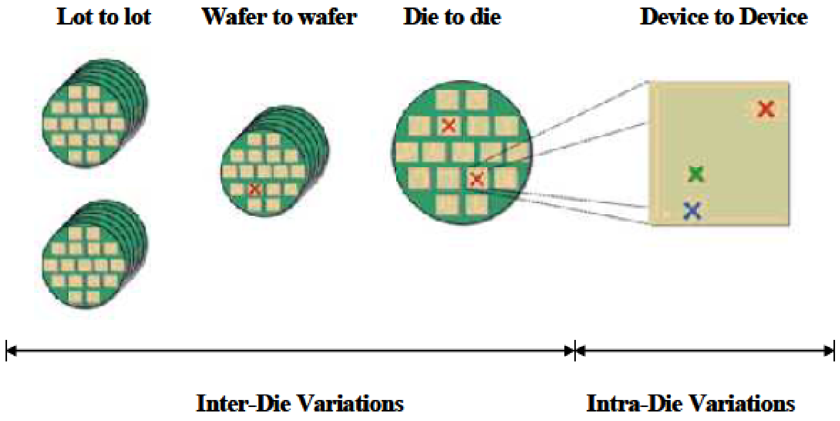


Figure 2.3: Inter-Die vs Intra-Die Variations
[1]

Intra-die variations are usually splited into two categories, the systematic and the random variations[1]. Systematic variations occur during fabrication steps and are highly predictable. Gate length variability would be an example of a systematic variation. These variations have been reported to affect to a lesser extent the electrical characteristics of the devices. Since the sources of process variations are predictable, they can be more sufficiently minimized by circuit designers. On the other hand, random variations are caused by random and unpredictable phenomena in the semiconductor fabrication process, such as channel doping fluctuation. The random variations are difficult to be identified, and hence can cause significant mismatches between the adjacent transistors [1].

Due to their unpredictability and lack of compensation, these variations create great challenges in achieving acceptable yields in sub-micrometer processes. This happens because the number of dopant atoms decreases largely in the nanometer scale and therefore even small variations in their number and location could lead to significant deviations in the performance of a certain device. Figure 2.4 illustrates how variations in the location of dopant atoms in a device channel can affect the threshold voltage (V_{th}). Sources of such variations could be lithography, etching, Chemical Mechanical Polishing(CMP) etc.

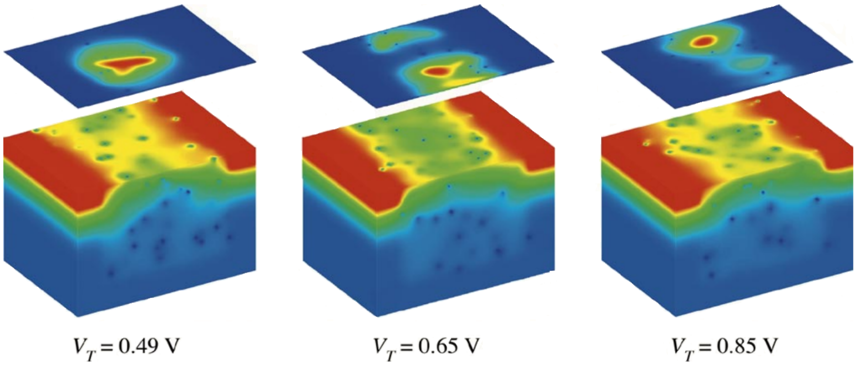


Figure 2.4: Random Variations of dopant atoms in a given device channel [1]

To ensure that a circuit will have an acceptable yield during manufacturing, designers give consideration to these process variations by carefully studying and simulating statistical models for the CMOS devices. The statistical analysis of these parameters is a complex process are various methods, which can be used to achieve it. The most common methods, which take into account the process variations before tape-out, are the generating of worst case file (WCF) or corner design parameters and by applying the application of the Monte Carlo simulations [2]. Corners simulation takes into account the worst case parameters of a device in order to ensure that a circuit will operate under the worst case scenario. This method counters the inter-die variations. There are five types of corners known as: TT (Typical p-channel, Typical n-channel), SS (Slow p-channel, Slow n-channel), FF (Fast p-channel, Fast n-channel), FS (Fast p-channel, Slow n-channel) and SF (Slow p-channel, Fast n-channel) [2]. All these cases simulate the extreme conditions for various parameters. For example, the SS corner for the I_{drive} current will produce a lower I_{drive} than the TT corner, which targets the typical operation, and this would be the worst-case scenario. On the other hand, FF corner will result in a higher I_{drive} current. Figure 2.5 shows the different types of corners for various model parameters.

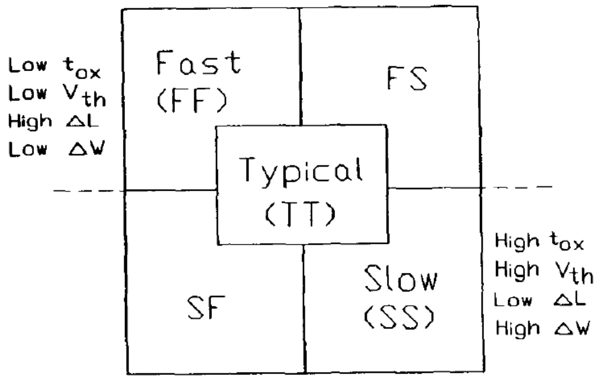


Figure 2.5: Worst Case Files (Corners) for Various Model Parameters [1]

Aside from corners, Monte Carlo is a technique, which can be used by designers to simulate the effects of both inter-die and intra-die variations [17][1]. The Monte Carlo is mainly used to simulate how the mismatches of individual devices affect the operation of the circuit. The method analyzes a large amount of circuit model parameters and generates random numbers for each input variable. Then, it simulates the mismatch effects of these parameters. For example, in a pair of identically drawn resistors a random component is added on each one of them (figure 2.6). This component is calculated based on a given statistical distribution which is determined by resistor size, technology etc [17].

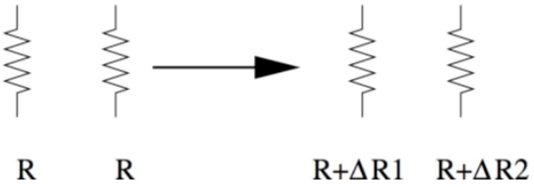


Figure 2.6: Simulating Mismatch Between Two Identical Resistors [17]

The pair will then be simulated with the added random component and produce an output parameter result. In this case, the only parameter changed is the resistance and hence the output will result in a pair of different value resistors. This method is very useful, but at the same time, it can be expensive and time consuming, especially when dealing with large parameter devices or nonlinear components such as MOSFETs [1].

2.3.2 Voltage Reference Trimming

In order to overcome the challenges of process variations in voltage reference circuits, trimming is used. Trimming is a calibration technique that is predominantly used to adjust variations in voltage reference circuits post-silicon (i.e., post fabrication). It is usually achieved via a separate circuit topology, which is used to correct abnormalities of the output voltage reference and achieve the required accuracy [18]. Trimming usually provides calibration to only one trimming element (i.e., adjusting only one of the performance parameters) depending mainly on the project budget and the specific application. There are several ways to calibrate a voltage reference circuit. However, the most commonly used are the laser trimming and the linked fuse resistor. As the name suggests, laser trimming regulates voltage reference parameters with the use of a laser. This method is highly accurate and area efficient, and it is largely used to trim high precision voltage references. On the other hand, the method can be insecure for the reason that the laser energy occasionally can cause damage to the substrate of the chip and reduce the overall performance [18]. Moreover, laser trimming is a high price method. On the contrary, linked fuse resistor trimming is less costly, and for this reason, it is implemented more often. However, it consumes a comparably larger area and it is harder to achieve high accuracy with it. Linked fuse resistor trimming basically selects the correction circuit by blowing up a resistor fuse with a high current. The method is primarily used to correct resistor values and is implemented by connecting a resistor fuse (R_{fuse}) in parallel with the trimming resistor (R_{trim}), both attached to trimming pads.This topology is presented in figure 2.7.

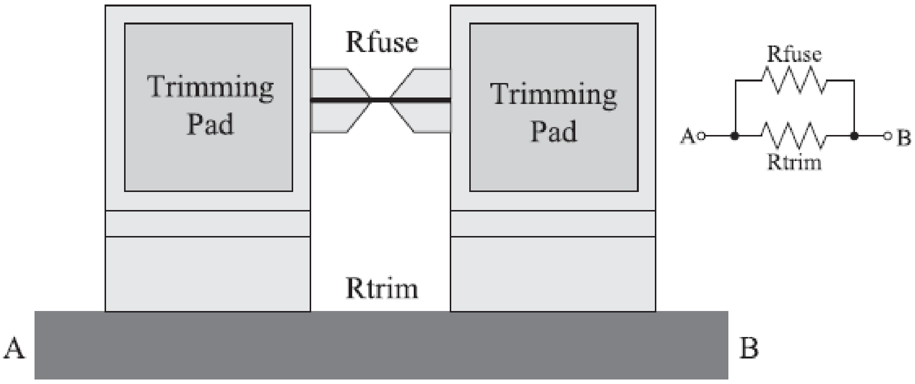


Figure 2.7: Typical Fuse Resistor Topology
[18]

When the fuse resistor (R_{fuse}) is in place, the resistance between points A and B is equal to "R_{fuse}" which is almost 0. When R_{fuse} is blown open by applying a high current, the resistance between A and B equals the value of R_{trim}. This resistor trimming is called the "modulated trimming" because it affects the induced voltage/current in a voltage reference circuit. Apart from resistor trimming, there are two other methods of trimming, namely, voltage trimming and current trimming[18]. These three methods, which are chiefly used to trim voltage reference circuits, are illustrated in the following figure.

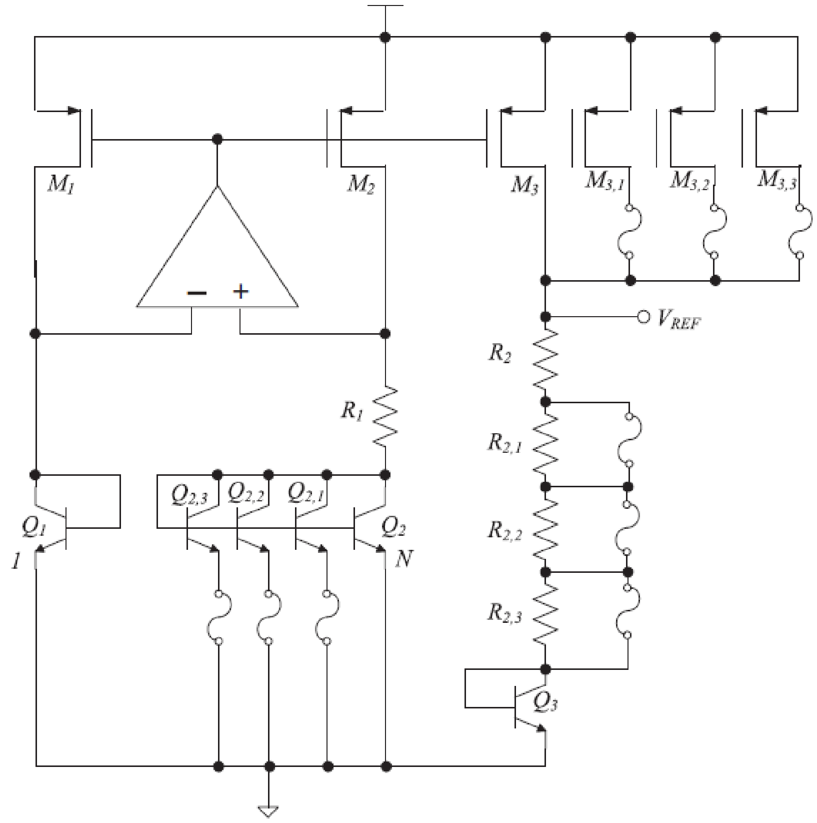
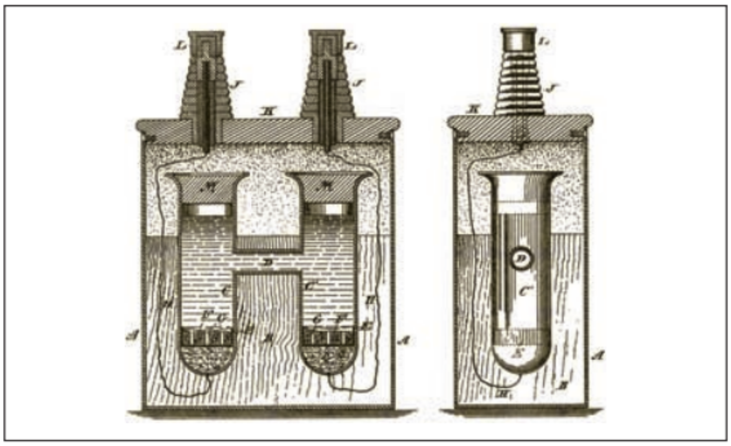


Figure 2.8: Modulated, Voltage and Current Trimming
[18]

This is an opamp-based β-multiplier BGR. As it is shown in this circuit, all trimming methods are implemented with blown fuses. By blowing any of these fuse, the physical parameters of the devices will change in order to get the required results. By changing the emitter area of Q_2 we can alter the output voltage. By blowing any of the M fuses in the current mirror it is possible to alter the W/L ratio of M_3 in order to conduct more current. Whereas, by blowing any fuses in the resistor array, we increase its resistance and hence the voltage/current over it.

2.4 History of Voltage References

Voltage references and current sources had been used long before the first integrated circuits started to emerge. As reported by Linden T Harrison in the book Current Sources and Voltage References, voltage references had the form of bulky and pricy laboratory standards before starting being used in integrated circuits or as separate IC products [15]. Back in the 1940s and 1950s, these standards were basically resistors combined with vacuum tubes or other instruments based on the tube. The most used of that period were the Weston cell and the Clark cell along with some kind of batteries, with the Weston cell being the most famous one. The Weston cell was a chemical cell, which could produce an output voltage of 1.019V at room temperature. Its architecture can be seen in figure 2.9. During the Second World War, the first mercury based cells, which were able to produce a voltage of 1.35V at certain mAs and had an operating life of around 1000 hours, were created. These mercury cells were small and cheap.



Weston voltaic cell (4 April 1893) in the U.S. patent 494,827 (courtesy NJIT).

Figure 2.9: The Weston Cell [21]

It was only until the 1950s, and in combination with the introduction of the zener diode by Clarence Zener, that a voltage reference was first created as a discrete semiconductor. The diode zener operation relies on the physics and characteristics of the reverse bias PN junction. It is widely used until today in commercial and industrial applications due to its accuracy, small size and low cost. An example of a voltage reference using the zener diode is illustrated in figure 2.10. The depicted circuit is from the 1970s and combines a zener diode and a normal diode with a precision opamp in order to create a 6.2V output voltage reference at 6mA.

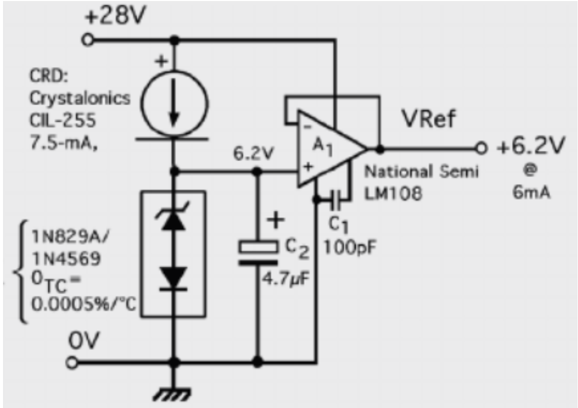


Figure 2.10: Precision Voltage Reference from early 1970s [15]

In 1969 Bob Widlar created the first integrated voltage reference based on the bandgap voltage of silicon. This voltage reference was part of the power regulator LM109, which was the first monolithic $5 - volts$, $1 - amp$ linear voltage regulator. Later, Widlar presented the first ever implementation of a bandgap voltage reference (BGR), which was eventually released by the National Semiconductor in 1971 [15, 23]. A simplified version of this novel circuit is presented in figure 2.11.

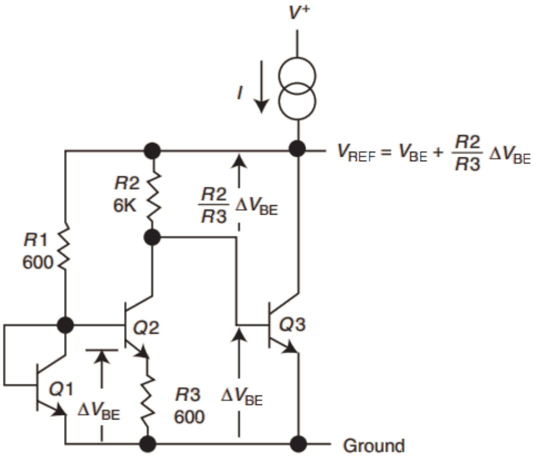

V+
I
$V_{REF} = V_{BE} + \frac{R2}{R3}\Delta V_{BE}$
R2
6K
$\frac{R2}{R3}\Delta V_{BE}$
R1
600
Q2
Q3
Q1
ΔV_{BE}
R3
600
ΔV_{BE}
Ground

Figure 2.11: First integrated BRG proposed by Robert J Widlar [36]

In figure **??** Q2 operates in 10 times lower current density than Q1, the emitter-base voltage differential ΔV_{BE} of the two BJTs appears across R2 and Q3 is a gain stage that regulates the output voltage. The output voltage is derived from equation 2.6.

$$V_{REF} = V_{BE} + \frac{R2}{R3}\Delta V_{BE} \tag{2.6}$$

The emitter–base voltage differential between two transistors operated at different current densities is given by equation 2.7 where k is

Boltzmann's constant, T is absolute temperature, q is the electron charge and J_1, J_2 are current density.

$$\Delta V_{BE} = \frac{kT}{q} \log_e \frac{J_1}{J_2} \tag{2.7}$$

By adjusting the current density one can compensate for the dependence of ΔV_{BE} in temperature. Widlar combined the CTAT response of the V_{BE} with the PTAT response of ΔV_{BE} To create the first practical implementation of the BGR.

This idea was already introduced and implemented with diodes back in 1964 by David Hilbiber [16]. Hilbiber presented the same concept of diodes operating with differnt current densities in order to get a PTAT response, similar to that of ΔV_{BE}. His proposed circuit can be seen in figure 2.12. However Hilbiber's design would not operate at lower voltages and it was not suited for monolithic construction i.e., it could not be constructed as an integrated circuit [36].

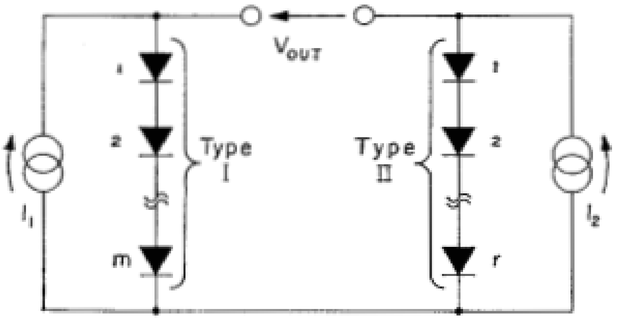


Figure 2.12: Diode based scheme presented by David Hilbiber [16]

The National Semiconductor made the breakthrough again in the middle 70s by introducing the LM199/399, which was the first subsurface (buried) zener. Specifically, it was R. Dobkin from National Semiconductor who in 1976 introduced the so-called legendary voltage reference [12, 15]. Unlike previous breakdown references, which used the emitter-base junction as a Zener diode, the buried zener moved the breakdown voltage deeper inside the substrate of the monolithic IC circuit, which means that the buried zener was simply placed deeper under the oxide making the zener immune to surface effects. The output voltage of the buried zener proposed by R.Dobkin in one of his papers [12] was $6.9V$ with a temperature coefficient of $0.5ppm/^oC$, and RMS noise of $7\mu V$ and could provide 0.5 to $10mA$ of current. Importantly, that design outperformed anything at that time, and remains until today one of the most stable devices that are being used. However, the buried zener had two major drawbacks. One is that it was too expensive and the other is that its output voltage is relatively high for many of the applications that are out today [15]. A cross section of the Dobkin's buried zenere is depicted in figure 2.13

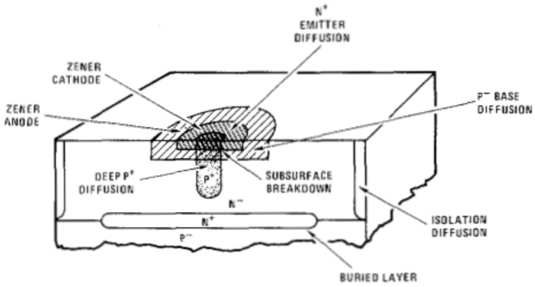


Figure 2.13: Cross Section of the Buried Zener Proposed by R Dobkin [12]

The next milestone was reached in December 1974, when A Paul Brokaw presented his bandgap voltage reference, which is broadly known as the Brokow cell. This was the first precision voltage reference, which was capable of outputting around $1.23V$ and had a temperature coefficient of 5 $ppm/^{o}C$ over the military temperature range [7]. The voltage reference circuit proposed from A Paul Brokaw can be seen in figure 2.14

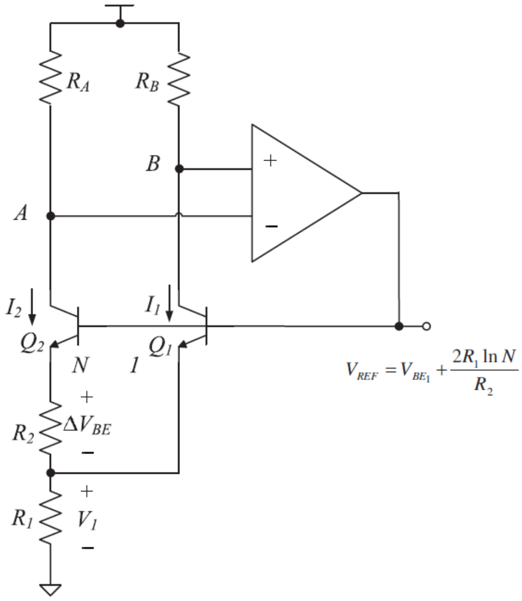


Figure 2.14: Simplified version of the the voltage reference proposed by Brokaw [7, 11]

The Brokaw cell utilized the same principle that Widlar had implemented in his design back in 1971, but improved it by adding a high gain operational amplifier (opamp). More specifically, and as it is shown in figure 2.14, due to the large opamp gain, both the inverting and non inverting inputs act like virtual ground and, thus, the opamp brings V_A at the same potential as V_B. By making $R_A = R_B$ it is ensured that the collector current of the Q_2 and Q_1 are equal i.e., $I_2 = I_1 = I$. To achieve the ΔV_{BE} as in Widlar's case, the area of the emitter of Q_2 is made several times (N) larger than the area of the Q_1. This results in a lower current density in Q_2 and thus creates a lower V_{BE}. The current flowing from R_1 and R_2 will be I and $2I$. From the Kirchhoff's voltage law or KVL we obtain that the current flowing through R_2 will be equal to:

$$I = \frac{\Delta V_{BE1,2}}{R_2} = \frac{V_T ln(N)}{R_2} \tag{2.8}$$

where N is the emitter size difference of Q_2,Q_1 and V_T is the thermal voltage. V_1 can be expressed by

$$V_1 = 2IR_1 = 2\frac{R_1}{R_2} ln(N) V_T \tag{2.9}$$

Then the reference voltage V_{REF} is described as

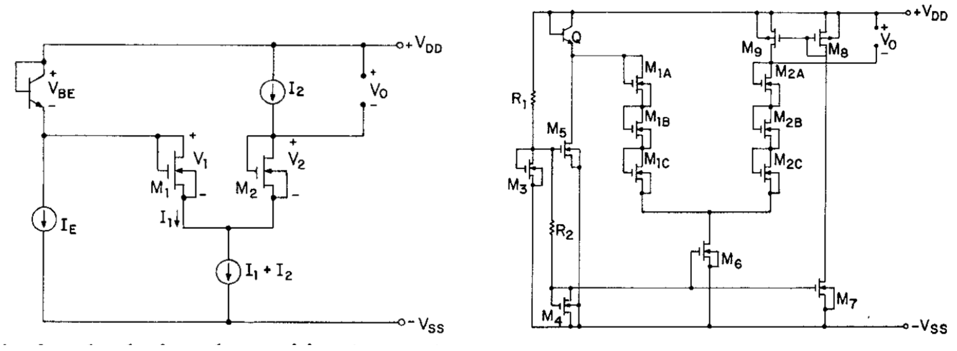
$$V_{REF} = V_{BE1} + 2\frac{R_1}{R_2} ln(N) V_T \tag{2.10}$$

From 2.10 we can observe that V_{REF} is the sum of the CTAT base-emitter voltage V_{BE1} and the scaled PTAT voltage V_1.

While trying to develop low power circuits for an electronic watch E. Vittoz and J. Fellrath published a model describing the DC behavior of MOSFETs operating in weak inversion in 1977. [32, 23]. In particular, in their paper they presented a number of current reference circuits along with a quartz oscillator scheme, all operating in weak inversion and consuming power in the nanowatt regime. It was in 1978 when Y.P. Tsividis and R.W. Ulmer [30] introduced the first voltage reference operating in weak inversion Figure 2.7. They used two diode connected NMOS transistors biased with I_1 and I_2 in weak inversion. The difference of the gate to source voltage (V_{gs}), which is equal to the drain to source voltage since the devices are saturated, is proportional to absolute temperature. The circuit generating the PTAT response is illustrated in figure 2.15(a), as well. Additionally, they combine the well known CTAT response of the V_{BE} of a BJT transistor to cancel out the effects and achieve a temperature insensitivity. The temperature response of the voltages of the two transistors operating in weak inversion can be expressed as

$$A = \frac{nk}{q} ln \left[\frac{I_1 (Z/L)_2}{I_2 (Z/L)_1} \right] \tag{2.11}$$

where A is the temperature coefficient, n is a process parameter, k is the Boltzmann constant, q is the electron charge, I_1 and I_2 are the bias currents and $(Z/L)_{1,2}$ are the width over length ratio of the 2 transistors.

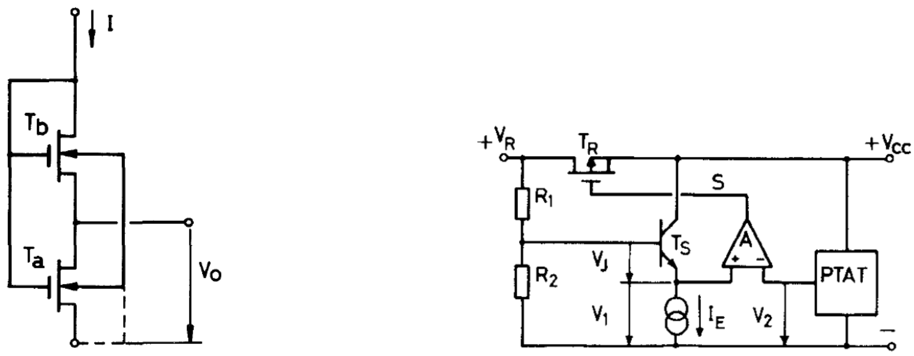


(a) PTAT Circuit

(b) Voltage Reference Circuit

Figure 2.15: Voltage reference proposed by Y.P. Tsividis and R.W. Ulmer [30]

But even while using the weak inversion region this reference by Tsividis and Ulmer operates with a $10V$ supply voltage and consumes $16\mu W$ of power. These specifications were not ideal for the electronic watch and thus a low voltage reference was proposed by E.A. Vittoz and O. Neyroud in 1979 [33, 23]. In the same principle Vittoz and Neyroud used the CTAT response of a BJT and combined it with a PTAT voltage achieved by NMOS devices operating in weak inversion. The proposed circuit presented in Figure 2.12, operates at $1.3V$ supply voltage, consumes only $200nW$ of power and achieves a TC of around $30ppm/^oC$.



(a) PTAT Circuit

(b) Voltage Reference Circuit

Figure 2.16: Voltage reference proposed by E.A. Vittoz and O. Neyroud [33]

With Vittoz and Neyroud designing a voltage reference based on the needs of the electronic watch, that is designing a voltage reference operating at a low supply and low power, it had become evident that supply voltage for electronic circuits was scaling down along with the shrinking technology. For that reason researchers started investigating sub-1V voltage references, which is a voltage reference that can operate with a supply voltage smaller than $1V$ [14]. Bandgap references, until that point, were based on the bandgap energy of silicon, which is around $1.23V$. This fact introduces a great challenge since the output of the voltage reference should be lower than its supply voltage. The first sub-1V voltage reference was presented in 1997 by Harry Neuteboom, Ben MJ Kup and Mark Janssens while they were designing a DSP-based hearing instrument IC. They were limited to work with a supply of $0.9V$, which made it impossible to work with a conventional bandgap reference (see figure 2.13) [22, 14].

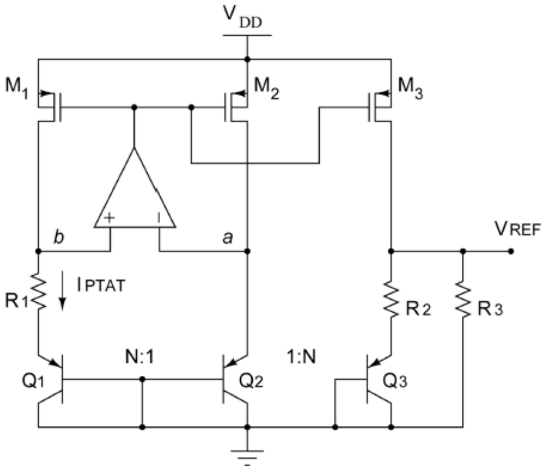


Figure 2.17: Sub-1V voltage reference proposed by Harry Neuteboom, Ben MJ Kup and Mark Janssens [14]

In their design they used a resistive division technique implemented by R_2 and R_3 to achieve a voltage output lower than $1.23V$ as presented in Figure 2.17. Specifically, the opamp is used to force the same current in the two branches, and by sizing Q_1 and Q_2 it is possible to achieve a ΔV_{BE} across R_1 that generates a PTAT current. This current is then replicated through the current mirror for M_1, M_1 and M_3 in the 3rd branch of the circuit. Therefore the output voltage is a result of a PTAT current and the CTAT V_{BE} of Q_3. By adding R_3 it is possible to divide the reference output and achieve a lower voltage level. The output of the proposed voltage reference can be described in equation 2.12.

$$V_{REF} = \frac{R_3}{R_3 + R_2}\left(V_{BE} + I_{PTAT}R_2\right) \tag{2.12}$$

The equation indicates that by adjusting R_2 and R_3 one can scale down the voltage reference. The proposed circuit operates with a supply voltage of $0.9V$ and achieves an output voltage of $0.67V$. until the early 2000s all designs for both bandgap and sub-1V bandgap voltage references were implemented by using resistors, which was contributing to a larger die area and therefore to a higher cost [23]. This issue was solved when Arne E Buck et al. proposed a resistor-less voltage reference consisted only of MOSFETS and diodes as illustrated in figure2.18 [8]

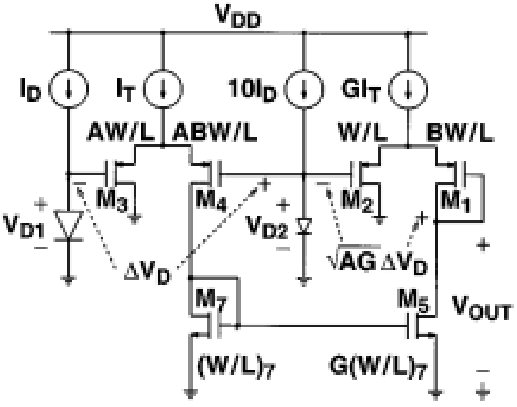


Figure 2.18: Resistorless Reference proposed by Arne E Buck et al [8]

As it is described by Arthur Campos de Oliveira [23], the proposed design compensates the temperature behaviour of $D1$ with the current gain $\sqrt{AG}$, which is a PTAT term.

Since 1995, low supply voltage and ultra low power consumption have been the main focus for voltage reference circuits[14]. As MOSFETs operating in subthreshold region mimic the BJT behaviour, designers started focusing on that region of operation as well. Many circuit designs emerged focusing on low power and low supply voltage since then. The voltage reference proposed by Giuseppe Vita and Giuseppe Iannaccone in 2007 [31] (Figure 2.19) with all transistors operating in strong inversion could achieve a TC of $10ppm/^oC$ with a minimum supply voltage of $0.9V$ while only consuming $3.6nW$ of power.

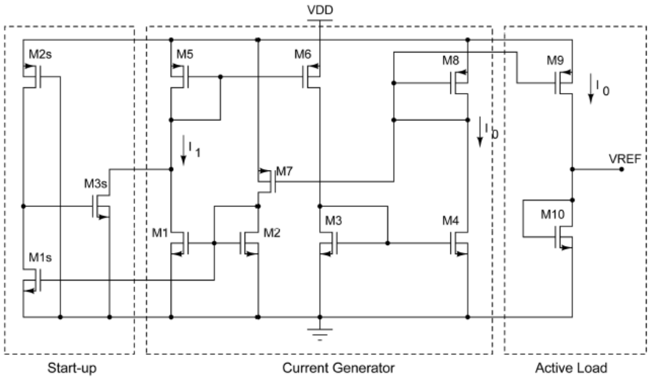


Figure 2.19: Voltage Reference proposed by Giuseppe Vita and Giuseppe Iannaccone in 2007 [31, 14]

Designs focused on MOSFETs operating in weak inversion could achieve even lower supply voltages. The reference topology proposed by T Ytterdal (figure 2.20) in 2003 [38] had a minimum supply voltage of $0.6V$ and the one proposed by Luca Magnelli et al [20] in 2011 (Figure 2.21) had a output voltage of around $263.5mV$ while operating at a minimum supply voltage of $0.45V$. The design had a TC of $42ppm/^oC$ in a wide temperature range from 0^oC to 125^oC and due to its low supply voltage and all transistors operating in weak inversion consumes only $2.6nW$.

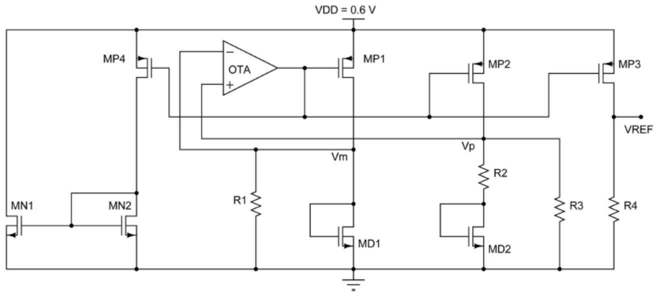


Figure 2.20: Voltage Reference proposed by T Ytterdal in 2003 [38, 14]

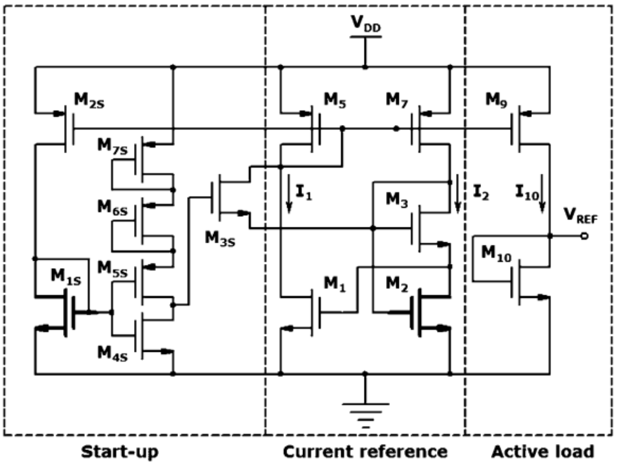


Figure 2.21: Voltage Reference proposed by Luca Magnelli et al in 2011 [20]

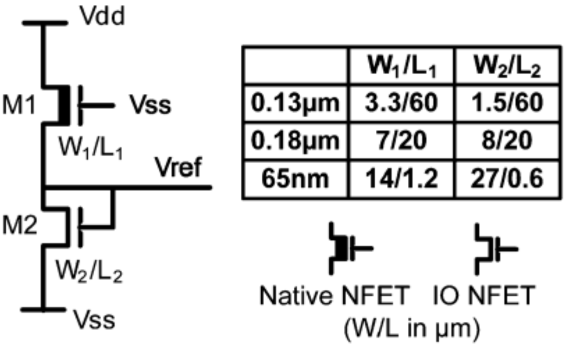
In recent years the designs are still focusing on ultra low power consummation focusing on applications, in which the battery life is critical. These designs can go down to the picowatt range when it comes to power consumption while being really area effective. In the next chapter we are going to present and analyze some of these designs, how they operate as well as some drawbacks and challenges related to them.

Chapter 3

Ultra Low Power Designs

3.1 A 2-Transistor Picowatt Voltage Reference

In 2012, Mingo Seok et al, [27] proposed a 2-transistor (2-T) voltage reference which consists of one native (i.e., near $0V$ threshold voltage) device acting as a current source and one diode connected nominal NFET device. The proposed design can be seen in figure (3.1).



	W_1/L_1	W_2/L_2
0.13μm	3.3/60	1.5/60
0.18μm	7/20	8/20
65nm	14/1.2	27/0.6

Figure 3.1: 2-T design proposed and W/L ratios in different process technologies [27]

This 2-T design is ideal for applications, in which the power budget is constrained. Some examples are IOT, biomedical and military applications. These applications usually operate in a region of pico to nanowatt power consumption in order to optimize battery life. Seok's design, which is fabricated at 130 nm, has been reported to consume only 2.22 pW at $0.5V$ and room temperature. Furthermore, they have achieved a temperature coefficient of 16.9 $ppm/^oC$ (best case) and 231 $ppm/^oC$ (worst case) with a line sensitivity of $0.033\%/V$ at an area of $1350\mu m^2$.

In figure 3.1, both $M1$ and $M2$ are biased in the sub-threshold region and have $L_1 = L_1 = 60m$ to minimize power consumption. The current of $M1$ is equal to that of $M2$ and thus the output voltage of the reference can be derived from the equation (**??**), where m_1, m_2 are the subthreshold slope factor of each device, V_T is the thermal voltage, W and L are the transistor width and length, μ is the mobility and C_{ox} is the oxide capacitance. This proposed design presented a breakthrough since it achieves a ×1000 and ×10 improvement in power consumption and area respectively comparing to prior work as it is shown in Figure 3.2

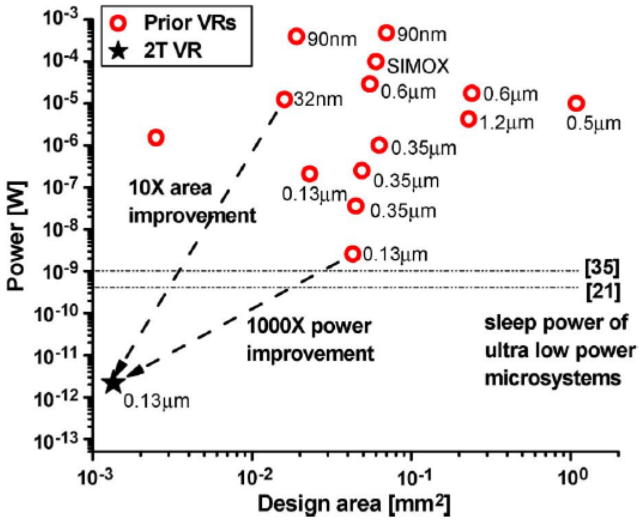


Figure 3.2: Power Consumption and area of the 2-T voltage reference proposed by Mingo Seok et al [27]

$$I_{Sub} = \mu C_{ox}\frac{W}{L}(m-1)V_T^2 exp\left(\frac{V_{gs}-V_{th}}{mV_T}\right)\left[1-exp\left(\frac{-V_{ds}}{V_T}\right)\right] \tag{3.1}$$

$$\begin{aligned} I &= \mu_1 C_{ox1}\frac{W1}{L1}(m_1-1)V_T^2 exp\left(\frac{0-V_{ref}-V_{th1}}{m_1 V_T}\right) \\ &= \mu_2 C_{ox1}\frac{W2}{L2}(m_2-1)V_T^2 exp\left(\frac{V_{ref}-V_{th2}}{m_2 V_T}\right) \end{aligned} \tag{3.2}$$

$$V_{REF} = \frac{m_1 m_2}{m_1+m_2}(V_{th2}-V_{th1}) + \frac{m_1 m_2}{m_1+m_2}V_T ln\left(\frac{\mu_1 C_{ox} W_1 L_2}{\mu_2 C_{ox} W_2 L_1}\right) \tag{3.3}$$

As the equation suggests, the reference voltage can be derived from the threshold voltage difference of $M1$ and $M2$. Hence any devices with significant difference in threshold voltage can be use and the sizing of the two devices.To minimize the temperature compensation the optimal size for $M1$ and $M2$ can be found from equation 4.18

$$\frac{dV_{REF}}{dT} = 0 \longrightarrow \left(\frac{W_1}{W_2}\right)_{opt} = \frac{\mu_2 C_{ox2} L_2}{\mu_1 C_{ox1} L_1} exp\left(\frac{q}{k}(C_{Vth2}-C_{Vth1})\right) \tag{3.4}$$

The design can offer a good line sensitivity (LS) as far as long channel devices are used. Because short channel effects become negligible when long channel devices are used, the terms in equation 3.3 become insensitive to power supply. Moreover, a good temperature response is achieved. It has been shown that the threshold voltage V_{th} and thus the first term in 3.3 is complementary to temperature [29], while the second term, V_T and m are proportional to temperature. Consequently, by the correct sizing of the transistors the two terms cancel each other to achieve a very low $ppm/^oC$ temperature coefficient.

Because the voltage output is related to the sensitive of process variations parameters, such as V_{th}, C_{ox}, width and length of the devices, the authors proposed a trimming scheme, which allows to adjust the width to length ratio (W/L) of the devices post-silicon, as it is depicted in figure 3.3.

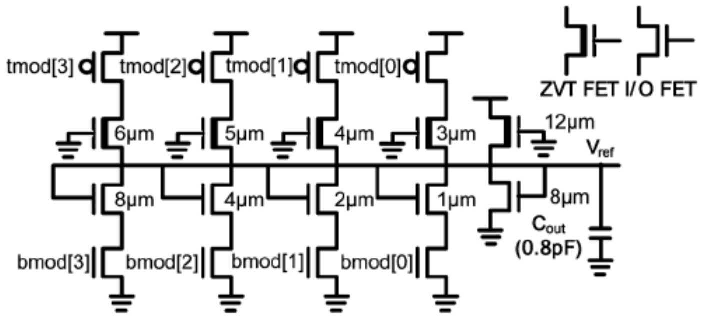


Figure 3.3: Trimming Scheme proposed by Mingo Seok et al [27]

With the proposed trimming scheme the authors target and achieve a below $50ppm/^oC$ temperature compensation. They reported that the spread of TC of the output voltage is improved almost by a factor of 10 compared to the non trimmable measurements across 25 dies. They also investigated a one temperature point (i.e., 80^oC) trimming technique to reduce trimming cost and time. With their proposed trimming circuit, they were able to adjust the width to length ration of the devices by sending signals on the top and bottom transistors.

3.2 420 fW Self-Regulated 3T Voltage Reference

Based on the same principle that was presented by Mingo Seok et al. [27], Hui Wang and Patrick P Mercier, in 2017, proposed a voltage reference topology that only consumes 420 femto Watt of power at a minimum operating supply voltage of $0.4V$ [34]. However, instead of using native devices like in [27], they only used nominal devices in standard CMOS technology. Their proposed design is presented in figure 3.4

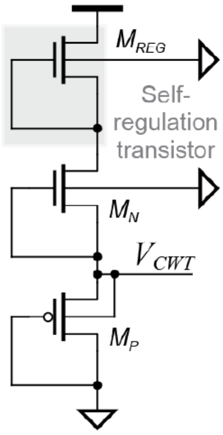


Figure 3.4: Voltage Reference Topology proposed by Hui Wang and Patrick P Mercier [34]

In more detail, they have used a NMOS device as a current source (M_N in figure 3.4) to drive a diode connected PMOS device (M_P) that serves as a reference resistor. A principle of operation can be seen in the figure (3.5). The output voltage reference is a function of the difference of the threshold voltages of M_P and M_N. Transistor M_{REG} is added to improve the line sensitivity of the reference.

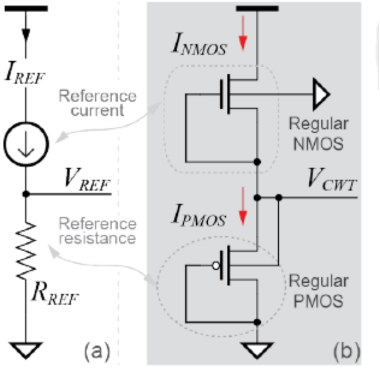


Figure 3.5: Voltage Reference Topology proposed by Hui Wang and Patrick P Mercier [34]

The design was manufactured in a nominal $65nm$ process and could produce an average output voltage of $342.8mV$. That was measured across 38 samples. Furthermore, this design achieves a line regulation of $0.47\%/V$ from $0.4V$ to $1.2V$, which together with its ultra low power consumption ($420fWatt$) makes it ideal for IoT applications, which consume highly small amounts of power, and for applications in which low-energy harvesters are used. In addition, the design achieves an average TC of $252ppm/^oC$ in a temperature range of -40^oC to 60^oC.

The voltage reference (V_{CWT}) can be derived from the fact that the currents I_{NMOS} and I_{PMOS} flowing through the NMOS and PMOS devices, respectively, are the same. These currents can be expressed as follows :

$$I_{NMOS} = \mu_1 C_{OX1} \frac{W_1}{L_1} (n_1 - 1) \phi_t^2 e^{\frac{0 - V_{th1}}{n_1 \phi_t}} \tag{3.5a}$$

$$I_{PMOS} = \mu_2 C_{OX2} \frac{W_2}{L_2} (n_2 - 1) \phi_t^2 e^{\frac{V_{CWT} - V_{th2}}{n_2 \phi_t}} \tag{3.5b}$$

and because these two currents are equal, an expression of the output voltage reference V_{CWT} is described by the following equation.

$$V_{CWT} = n_2 \phi_t ln \frac{m_1 C_{ox1} (n_1 - 1) W_1 L2}{m_2 C_{ox2} (n_2 - 1) W_2 L1} + \frac{n_1 V_{th2} - n_2 V_{th1}}{n_1} \tag{3.6}$$

The results from the measurements of the voltage reference and the power consumption of the design over 38 different samples can be seen in the following figure. These results mirror the design performance when operating from $0.4V$ at 20^oC.

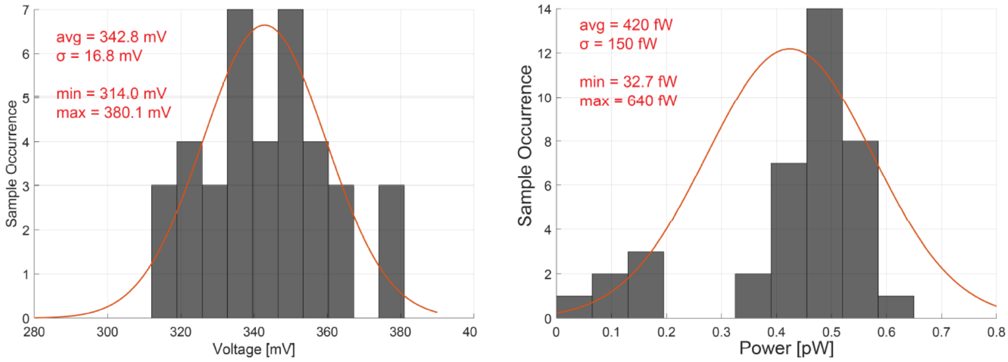


Figure 3.6: Measured Output Voltage and Power Consumption from supply of $0.4V$ at 20^oC

[34]

3.3 Subthreshold Voltage Reference With Scalable Output Voltage

In 2017, Inhee Lee, Dennis Sylvester and David Blaauw [19], proposed an ultra-low power voltage reference circuit that compared to the previous work could achieve an output voltage as high as that of a BGR. Their design was capable of achieving an output voltage of around $1.2V$ while only consuming tens of picowatts. Their proposed topology, which also consists of a trimming circuit, is depicted on the next figure. The output voltage V_{REF}, which is higher compared to the other works discussed in this thesis, can be achieved by stacking diode-connected PMOS transistors.

That means that the output voltage can be scaled with the number of PMOS devices.

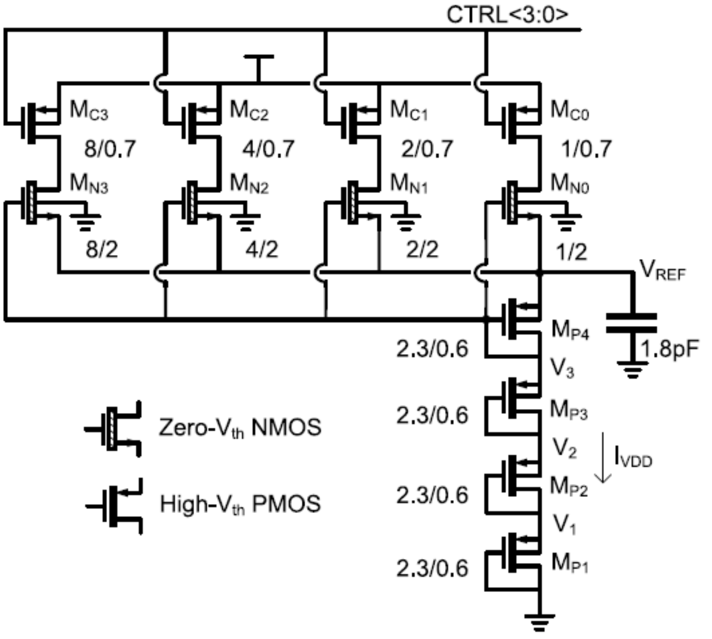


Figure 3.7: Voltage reference topology proposed by Inhee Lee, Dennis Sylvester and David Blaauw (2017) [19]

Here, M_{NX} are zero-V_{th} transistors, which are used as a current source for the PMOS devices. M_{CX} are the devices of the trimming circuit. They can change the W/L by sending digital signals. The M_{PX} composes the PMOS stacked transistors. As usual, the design is operating in the weak inversion and because the same current flows through the M_{NX} and M_{PX} devices, the output voltage is expressed as:

$$V_{REF} = N\left[\left(\frac{m_1|V_{th2}| - m_2|V_{th1}|}{m_1+m_2}\right) + \left(\frac{m_1 m_2 V_T}{m_1+m_2}\right) ln\left(\frac{\mu_1 C_{OX1}\frac{W_1}{L_1}(m_1-1)}{\mu_2 C_{OX2}\frac{W_2}{L_2}(m_2-1)}\right)\right] \text{[19]} \tag{3.7}$$

Here N is the number of the diode connected PMOS. By properly sizing the devices, a near-zero TC can be achieved. Their voltage reference showed an average voltage of $1.2V$ with a minimum supply of $1.4V$ while only consuming $35pWatts$ at room temperature. In addition, the reference showed a TC of $22ppm/^oC$ from 0^oC to 100^oC.

Chapter 4

Design Implementation

4.1 Core Cell

The design studied and implemented in this thesis is based on previous scientific work, which has been quoted in chapter 3. In the 2T voltage reference configuration proposed by Mingoo Seok et al. [27] we added one more diode connected NMOS device to increase the output voltage level, as shown in figure 4.1. A similar configuration was introduced by Inhee Lee, Dennis Sylvester and David Blaauw in 2017 [19], but implemented with PMOS devices instead.

Our final 3-T configuration (figure 4.1) was implemented using a depletion mode transistor ($M1$) as a current source and two identical high-threshold voltage diode connected NMOS devices ($M2$ and $M3$). The design was taped out with a BCD Gen 2 process from TSMC, which provides a deep nwell solution that allows body biasing. Therefore, the design was simulated both with and without the body effect for transistors $M0$ and $M1$.

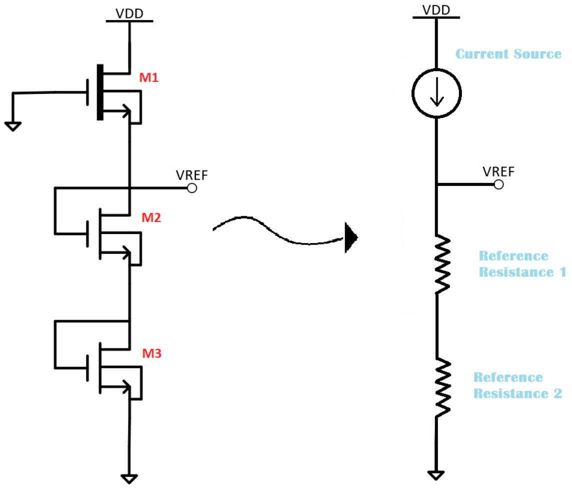


Figure 4.1: Core 3T Design

4.1.1 Principle of Operation

The design operates in the same way as the voltage references [19, 27] studied in chapter3. That is, the output voltage is based on the difference in threshold voltages of $M1$ and $M2$. As $M1$ is biased in the subthreshold region, the drain current can be defined as:

$$I_d = \mu C_{OX}\frac{W}{L}(m-1)V_T^2 e^{\frac{V_{gs}-V_{th}}{mV_T}}\left(1-e^{-\frac{V_{ds}}{V_T}}\right) \text{ [19]} \tag{4.1}$$

where μ is the mobility, C_{OX} is the oxide capacitance, W and L are the transistor size, m is the subthreshold slope factor and V_T is the thermal voltage. For $V_{ds} > 150mV$ the term "$(1-e^{-\frac{V_{ds}}{V_T}})$" is ≈ 1 and therefore, it can be neglected [19].

Then, when $M2$ and $M3$ are identical, we assume that μ_2 and μ_3 as well as m_2 and m_3 are equal. And because the same current flows through all three transistors, the equation shown below can be derived [19].

$$\begin{aligned} I &= \mu_1 C_{OX1}\tfrac{W_1}{L_1}(m_1-1)V_T^2 e^{\left(\frac{-V_{ref}/N-V_{th1}}{m_1V_T}\right)} \\ &= \mu_2 C_{OX2}\tfrac{W_2}{L_2}(m_2-1)V_T^2 e^{\left(\frac{V_{ref}/N-|V_{th2}|}{m_2V_T}\right)} \end{aligned} \quad \text{[19]} \tag{4.2}$$

where N is the number of diode connected NMOS devices. For $V_{ds} > 150mV$, the term $1-exp(-V_{ds}/V_T)$ can be ignored. Therefore the output voltage can be expressed as:

$$V_{REF} = N\left[\left(\frac{m_1|V_{th2}|-m_2|V_{th1}|}{m_1+m_2}\right)+\left(\frac{m_1m_2V_T}{m_1+m_2}\right)ln\left(\frac{\mu_1C_{OX1}\frac{W_1}{L_1}(m_1-1)}{\mu_2C_{OX2}\frac{W_2}{L_2}(m_2-1)}\right)\right] \text{ [19]} \tag{4.3}$$

All devices used, are $5V$ devices. The reason for this is because the process was offering only $5V$ depletion mode transistors.

4.1.2 Power Consumption

By looking at equation 4.1, it is expected that longer devices will result in lower current and hence in lower power consumption. In theory, when sizing the transistors, if the term W/L in equation 4.1 is drawn as small as possible, the lowest power consumption point could be achieved. However, that would come with a trade off in the chip area because the length of the transistors might need to be increased significantly.

To find out the power consumption of the design, different simulations were applied by sweeping the length of the transistors (L), while keeping the smallest device's width (W). The results of the simulations at room temperature and at a supply voltage of 1.2V, which is the minimum operating, are presented in the following figure.

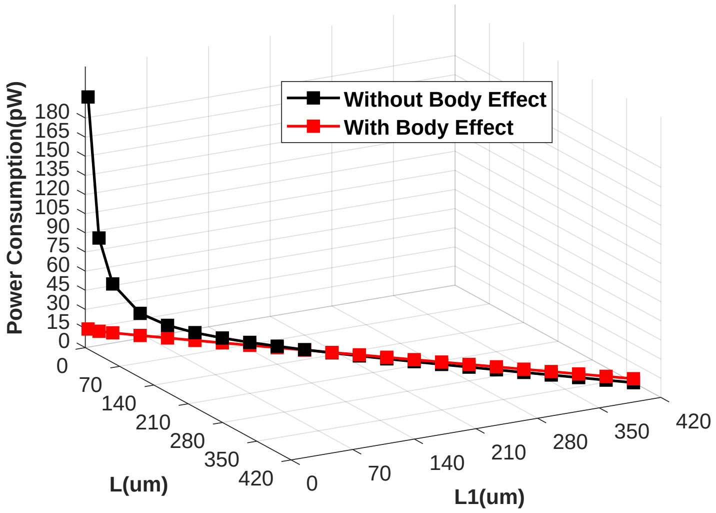


Figure 4.2: Power Consumption at 1.2V and Room Temperature With and Without the Body Effect, while Scaling the Length of $M1$, $M2$ and $M3$

Here, L is the length of $M1$ and $L1$ is the length of $M2$ and $M3$. Indeed, as the length of the devices increases, the power consumption decreases. However it seems that after 50μm power consumption it does not keep decreasing as expected. In spite of increasing the transistor length up to 400μm, the power consumption hardly reduces after 100μm. Therefore, in order to have a balance between power consumption and chip area, all the devices were chosen to have a length of 40μm with the minimum width possible. The device dimensions are shown in table 4.1.

Device	**M0**	**M1**	**M2**
Width (μm)	2	1	1
Length (μm)	40	40	40

Table 4.1: Transistor Sizing for low power consumption

In figure 4.2 the red line represents the power consumption for the body effect of transistors $M1$ and $M2$. When there is a body effect, the power consumption is lower even for shorter transistor lengths. The circuit has a power consumption of 15pW with all devices having a length of 2μm. The reason for that is that the threshold voltage V_{th} of $M1$ and $M2$ increases significantly with the body effect i.e., the source to body potential $V_{sb} \neq 0$, which obligates the transistors to conduct less current. In this way the chip area can be reduced by a factor of 20. However, it should be noted that these simulations were done post fabrication, and hence the tapped out design had the transistor sizes shown in table 4.1.

4.1.3 Thermal Effects & Temperature Compensation

As it was presented in chapter 2, it is crucial for the output of a voltage reference circuit not to deviate with variations in temperature. The drain to source current (I_{ds}) of a CMOS device is highly temperature dependent and can have a positive, negative or zero temperature coefficient depending on the bias region. It is convenient enough to rely only on the temperature coefficient of the threshold voltage (V_{th}) and the carrier mobility (μ) when modeling temperature effects of I_{ds} in long channel devices. Since long devices are used in this design, temperature effects from parameters such as carrier saturation velocity u_{SAT} and carrier field $\mathscr{E}_c$ can be neglected [2]. In the proposed design, the same current is flowing through all transistors and the output of the voltage reference is taken from the drain of $M1$. We can then assume that the temperature coefficient of the output voltage will follow the one of the current flowing through the transistors which is depended on V_{th} and μ. Indeed, the output voltage, which is described by equation 4.3, is a function of both the threshold voltage and the carrier mobility of $M1$ and $M2$. Therefore, the temperature coefficient of the voltage reference can be found by analyzing these two parameters.

Threshold Voltage

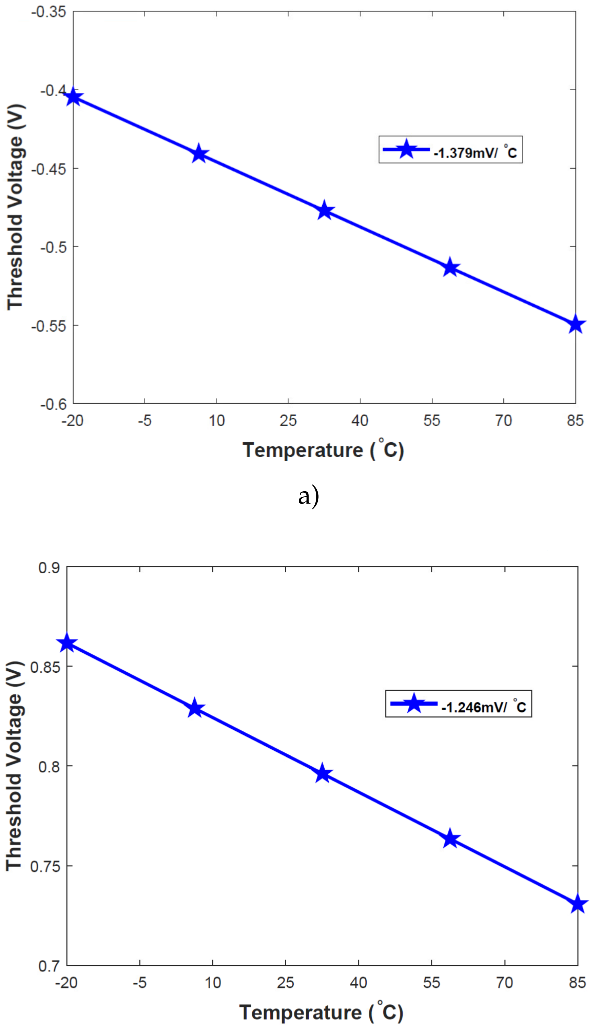
Threshold voltage of a CMOS device is described by the equation 4.4

$$V_{th} = V_{fb} + 2\phi_f + \gamma\sqrt{2\phi_f + V_{sb}}[2] \tag{4.4}$$

Where V_{fb}, is the flat band voltage, ϕ_f is the Fermi potential, γ is the body-effect coefficient, which is a process dependent parameter and V_{sb} is the source to body potential.

$$\gamma = \frac{\sqrt{2\epsilon_o\epsilon_{si}qN_b}}{C_{ox}}[2] \tag{4.5}$$

The ϵ_o is the dielectric permittivity of vacuum, ϵ_{si} is the dielectric permittivity of silicon, q is the electron charge density, N_b is the impurity concentration of bulk silicon and C_{ox} is the gate oxide capacitance. In equation 4.4 both V_{fb} and ϕ_b decrease with increasing temperature and this is why the threshold voltage V_{th} is complementary to absolute temperature [2]. Figure 4.6 presents the results of our simulation for the threshold voltage of all devices while sweeping temperature.

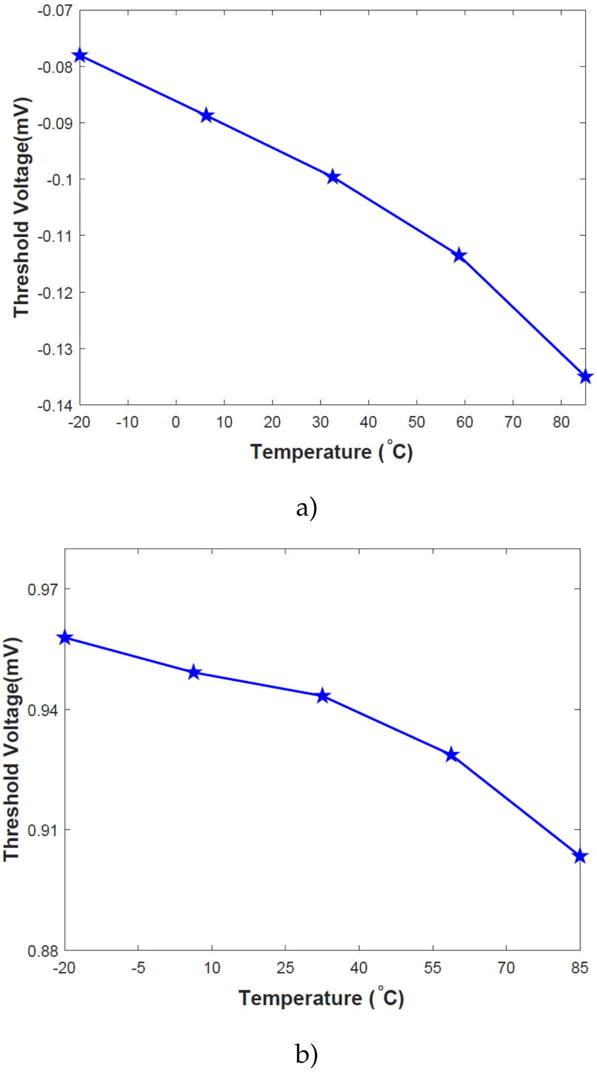


a)

b)

Figure 4.3: Temperature Response of Threshold Voltage (V_{th}) for a) $M1$ b)$M2$ and $M3$

As it can be seen from the plots, the threshold voltage of both the low threshold and the high threshold device show a linear CTAT response. The temperature coefficient of the threshold voltage (dV_{th}/dT) is depicted in figure 4.3. However, this is not the case when the body effect is present. When the body of both $M1$ and $M2$ are connected to the ground and not to the source terminal, the temperature response of the threshold voltage changes significantly (figure 4.5).



a)

b)

Figure 4.4: Temperature Response of Threshold Voltage (V_{th}) with Body Effect for a)$M1$ b)$M2$ and $M3$

The main reason for this observation is that when there is a source to bulk potential (V_{sb}), the width of the depletion region below the channel X_{dm} changes, and hence the threshold voltage can no longer be expressed by 4.4 [2]. Instead, V_{th} is now expressed as:

$$V_{th} = V_{fb} +_{si} + \frac{q(N_s - N_b)X_i}{C_{ox}} + \gamma\sqrt{_{si} + V_{sb} - V_o}[2] \qquad (4.6)$$

where N_b is the substrate doping concentration, N_s is the surface concentration, X_i is the width of the depletion region and $V_o = \frac{qX_i^2}{2\epsilon_o\epsilon_{si}}(N_s - N_b)$. ϵ_o is the vacuum permittivity and ϵ_{si} is the dielectric permittivity of silicon. Therefore, when there is a V_{sb} potential, both the value of V_{th} and its temperature coefficient change significantly.

Carrier Mobility

Carrier mobility (μ) is also sensitive to temperature [2]. More precisely, there are two scattering mechanisms that affect the electron and hole mobility, the lattice and impurity scattering [9]. Lattice scattering refers to the lattice vibrations, which reduce mobility as the temperature increases. On the other hand, impurity scattering occurs due to crystal defects and has the opposite temperature effects in mobility. However, impurity scattering is observed only in very low temperatures. So, for our measurement range of $-20^o - 85^o$ it is safe to take into account only the temperature effects from the lattice scattering. We then conclude that the mobility of both holes and electrons decreases with increasing temperature [9]. Carrier mobility can be modeled by the equation 4.7.

$$\mu_o(T) = \mu_o(T_o)(\frac{T}{T_o})^{-m}[2] \tag{4.7}$$

Where m is the slope of the logarithmic plot of the mobility μ_o versus temperature T. Here, μ_o is the mobility for low gate voltages. Because we do not use any high voltage, μ_o equals μ. The value of m varies from 1.4 to 1.6. The mobility yields a negative temperature coefficient, and for $m = 1.5$, its temperature coefficient can be expressed by equation 4.8 [2].

$$\frac{1}{\mu}\frac{d\mu}{dT} = -\frac{1.5}{T}[2] \tag{4.8}$$

The transconductance parameter (KP) follows the temperature response of the mobility (μ) where $KP = \mu C_{ox}^{'} = \mu \frac{e_{ox}}{t_{ox}}$ [3]. Therefore there is a reduction in KP with increasing temperature (equation 4.9).

$$KP(T) = KP(T_o)(\frac{T}{T_o})^{-m}[3] \tag{4.9}$$

In figure 4.5, simulations for KP and hence for μ show that, indeed, the mobility drops with increasing temperature.

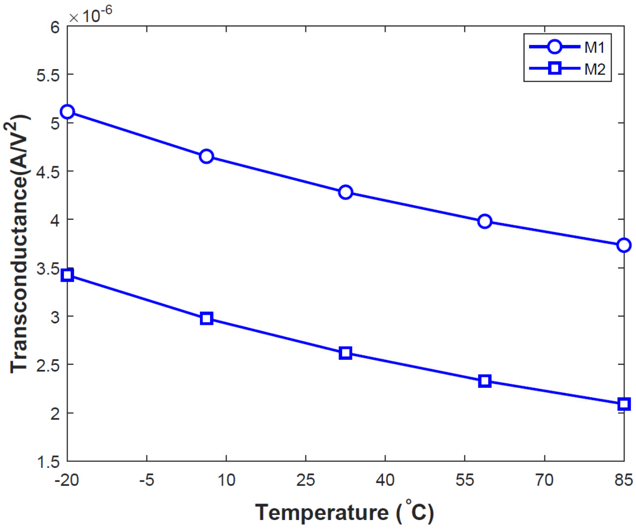


Figure 4.5: Temperature Response of the Transconductance Parameter KP

Thermal Voltage

As it is known, thermal voltage (V_T) exceeds a positive temperature coefficient and has a value of around $26mV$ at room temperature. Its temperature response is described by the equation 4.10 [3].

$$\frac{\partial V_T}{\partial T} = \frac{\partial}{\partial T}\left(\frac{kT}{q}\right) = \frac{k}{q} = 0.085mV/{}^{o}C[3] \tag{4.10}$$

Subthreshold Slope Factor

The subthreshold slope (S) is the slope of the $I_{ds} - V_{gs}$ curve of the subthreshold region. It is described as the gate to source voltage (V_{gs}), which is required in order to reduce the drain current by one decade [2]. It can be calculated by the equation 4.16.

$$S = 2.3\left[\frac{dV_{gb}}{d\ln I_{ds}}\right][2] \tag{4.11}$$

The factor 2.3 is the conversion from "log" base to "ln" equation 4.16 can be rewrite as :

$$S = 2.3\left[\frac{dV_{gb}}{d\phi_{ss}}\Bigg/\frac{d\ln I_{ds}}{d\phi_{ss}}\right][2] \tag{4.12}$$

By differentiating the two terms in 4.16 we get :

$$S = 2.3V_T\left[\left(1+\frac{C_d}{C_{ox}}\right)\Bigg/\left\{1-\frac{2V_T}{\gamma^2}\left(\frac{C_d}{C_{ox}}\right)^2\right\}\right][2] \tag{4.13}$$

For $\gamma >> C_d\sqrt{V_t/C_{ox}}$, then we get the following equation for the subthreshold slope.

$$S \cong 2.3V_T\left(1 + \frac{C_d}{C_{ox}}\right) = 2.3 \times V_T \times m[2] \tag{4.14}$$

The term $m = 1 + (C_d/C_{ox})$ is what we call the subthreshold slope factor usually referred to as η. The subthreshold slope factor implies the capacitive coupling between the gate and silicon surface [2]. It can also be expressed as :

$$m = 1 + (C_d/C_{ox}) = 1 + \frac{\gamma}{2\sqrt{2\phi_f + V_{sb}}}[2] \tag{4.15}$$

Equation 4.15 suggest that m is sensitive to temperature, process variations and supply voltage. The temperature response of m for $M1$ and $M2$ can be seen in the next figure.

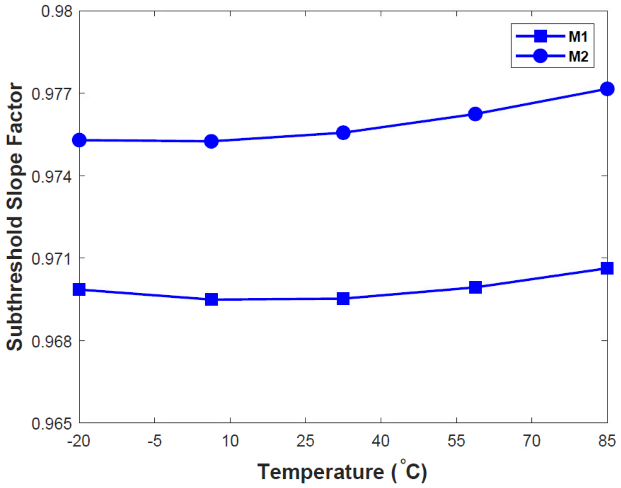


Figure 4.6: Temperature Response obtained by Simulation, of the Subthreshold Slope Factor m

[2]

Drain Current

As figure 4.8 shows, the drain current can have either a negative, a positive or a zero temperature coefficient depending on the region of operation [2].

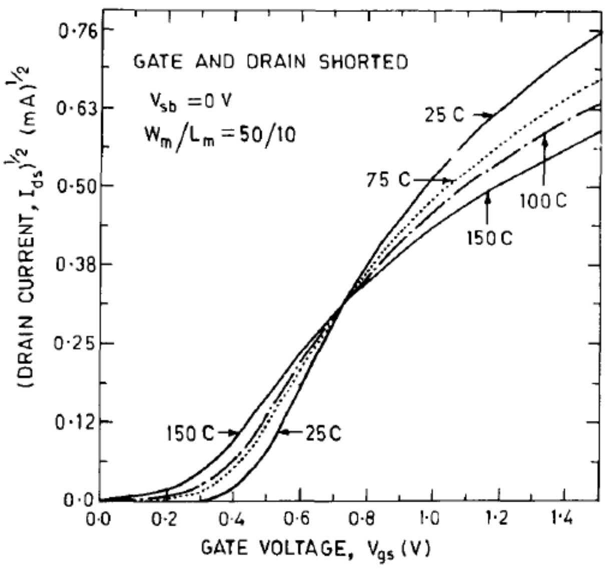


Figure 4.7: Temperature Response of the Subthreshold Slope Factor m for $M1$ and $M2$

[2]

That is due to the thermal effects of the threshold voltage V_{th} and the carrier mobility μ. Both V_{th} and μ are reduced with increasing temperature [3]. The temperature response of both μ and V_{ht} are simulated in presented in figure 4.3 and 4.5, respectively. In the saturation region the drain current can be expressed by the following equation.

$$I_{ds} = \frac{1}{2}\mu C_{OX}\frac{W}{L}\left(V_{gs} - V_{th}\right)^2 [2] \tag{4.16}$$

We can see from equation 4.16 that a decrease in mobility makes the drain current go down, while a decrease in threshold voltage will cause the drain current to go up [3]. For lower V_{gs} values, V_{th} dominates, and hence, the drain current presents a positive TC in weak inversion. For higher V_{gs} values, mobility dominates and hence the drain current has a negative TC in strong inversion [3].

Optimize Sizing for Low Temperature Coefficient

As mentioned earlier, the output voltage of the voltage reference circuit can be derived from equation 4.3 (repeated here for convenience).

$$V_{REF} = N\left[\left(\frac{m_1|V_{th2}| - m_2|V_{th1}|}{m_1 + m_2}\right) + \left(\frac{m_1 m_2 V_T}{m_1 + m_2}\right) ln\left(\frac{\mu_1 C_{OX1}\frac{W_1}{L_1}(m_1 - 1)}{\mu_2 C_{OX2}\frac{W_2}{L_2}(m_2 - 1)}\right)\right] [19] \tag{4.3}$$

Because of V_{th} the first term in 4.3 is complementary to temperature. The second term is proportional to temperature but its temperature coefficient can be changed by sizing $M1$ and $M2$. In order to achieve a zero temperature coefficient we can set $dV_{REF}/dT = 0$. In this way we can get the optimal transistor size for the lower temperature coefficient (equation 4.17).

$$\left(\frac{W_1/L_1}{W_2/L_2}\right)_{optimal} = \frac{\mu_2 C_{OX2}(m_2-1)}{\mu_1 C_{OX1}(m_1-1)} \times e^{\frac{q}{k}\left(\frac{1}{m_1}\frac{dV_{th1}}{dT}-\frac{1}{m_2}\frac{d|V_{th2}|}{dT}\right)}[19] \qquad (4.17)$$

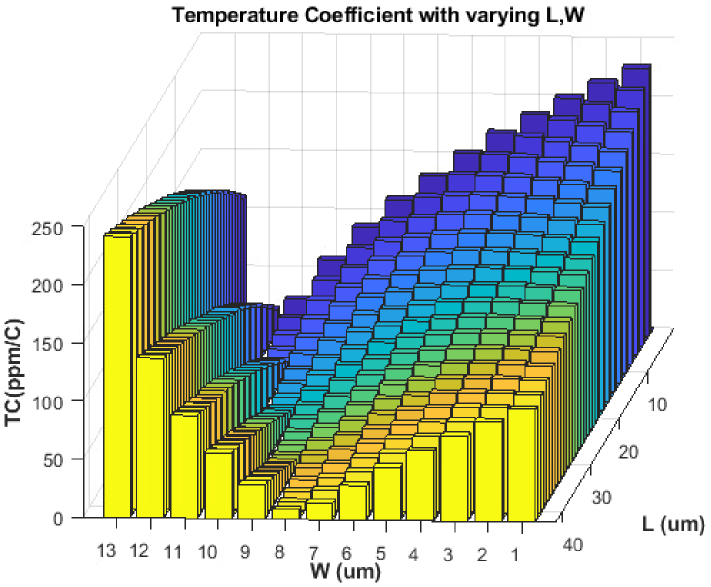


Figure 4.8: Simulation Results for Temperature Coefficient while sweeping the Length and Width of $M2$ and $M3$

As the three dimensional bar plot in figure 4.8 shows, a near zero ($7ppm/^oC$) TC can be achieved for $W1 = 8u$ and $L1 = 42um$. Here, we keep the width of $M1$ as low as possible and the length at $40m$. At the same time, we sweep $W1$ and $L1$ for a temperature range from -20^oC to $+85^oC$. In this way, a very low TC ($16ppm/^oC$) can be achieved with a shorter length of around $15um$ and a width of $11um$. However, this achievement comes with a cost in power consumption. The calculator tool of Cadence was used to find the TC in $ppm/^oC$ and the formula used was the one mentioned in section 2.2.1 (2.1).

Line Regulation

When $V_{sb} = 0V$, both the threshold voltage and the mobility remain almost stable with increasing supply voltage. Figure 4.9 presents the V_{th} response with and without the body effect. The body effect makes the threshold voltage change significantly, i.e., increasing, as we increase the supply voltage. This is not the case for the devices without body effect, in which the threshold voltage remains the same as the supply voltage is increasing.

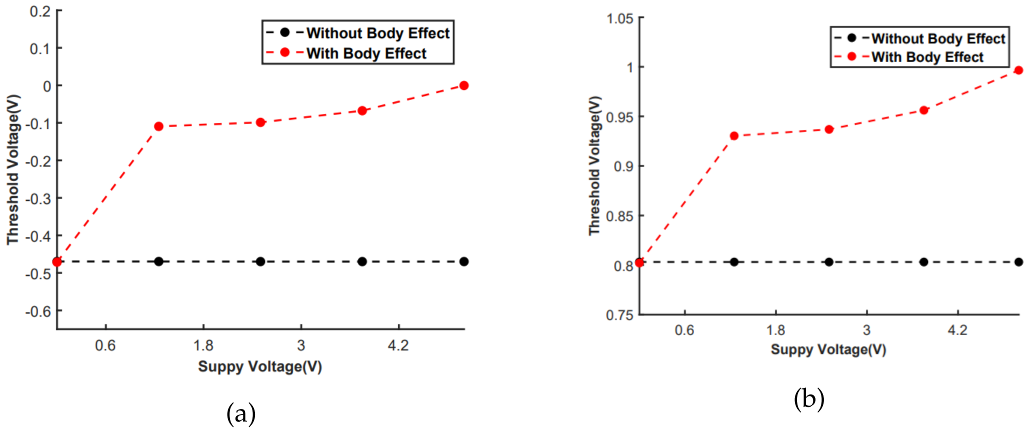


Figure 4.9: Threshold Voltage Response with Increasing Supply Voltage for a) $M1$ and b)$M2$

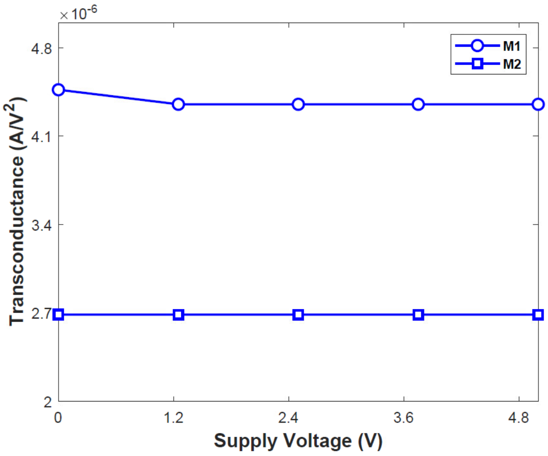


Figure 4.10: Transconductance (KP) with Varying Supply Voltage

As mentioned earlier, KP follows the temperature response of mobility μ. Mobility variations with increasing supply voltage are negligible when there is no body effect. Both W and L are stable from $0V$ to $5V$ of supply voltage. Therefore, we assume that the output of the voltage reference circuit would be stable from $1.2V$ to $5V$ ($1.2V$ is the minimum supply voltage). However, this was not the case, and we expected the output of the voltage reference to slightly vary in that supply range. The reason for this is the subthreshold slope factor (m), which we defined earlier as $1 + \frac{C_d}{C_{ox}}$ and the term $1 - e^{\frac{V_{ds}}{V_T}}$ of equation 4.1, that describes the subthreshold drain to source current. Both these terms are functions of V_{ds} and therefore we expect a slight increase in the output of the reference circuit with increasing supply voltage. The subthreshold slope factor m for both $M1$ and $M2$ is depicted in the following figure.

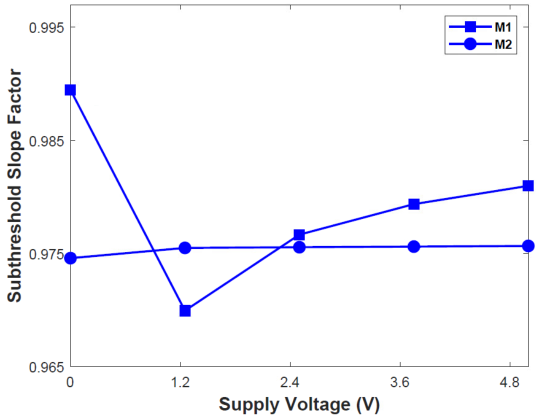


Figure 4.11: Transconductance (KP) with Varying Supply Voltage

Process Variations

The design suffers from process variations since almost all terms in 4.1, such as threshold voltage V_{th}, oxide thickness C_{OX} and even W/L, m are all sensitive to process variations [19, 2]. Figure 4.9 demonstrates how process variations affect the output of the voltage reference both in terms of TC and output voltage level. The plot shows the response of the reference circuit when simulated through corners.

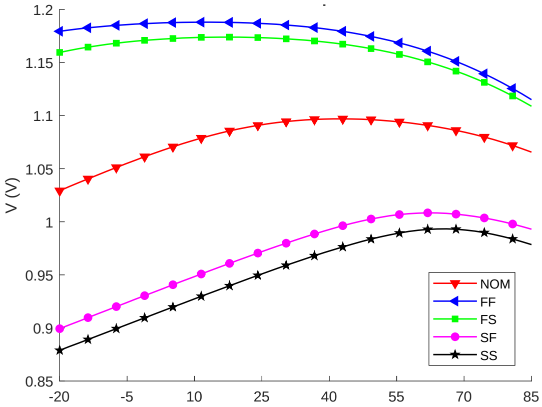


Figure 4.12: Corners Simulation of the Output Voltage Reference

It is shown that there is an offset of around $150mV$ from the nominal to the other corners. We can also observe that the TC is changing significantly. The nominal (NOM) corner along with the slow-fast (SF) and slow-slow (SS) corners present a PTAT response, while on the other hand the fast-slow (FS) and fast-fast (FF) corners both have a CTAT response.

To address these variations and deficiencies of the voltage reference, we propose a trimming technique that is capable of adjusting both the output voltage level and the temperature response of the voltage reference. The trimming circuit we propose can change the temperature response of the reference from PTAT to CTAT or vise versa by introducing a leakage current into the reference itself. That gives a great amount of freedom since the temperature response can be really unpredictable. By using the same trimming topology it is possible to bias the body of $M0$ in order to adjust the output voltage level, by either increasing or decreasing the Threshold voltage of $M1$. This is really important since the output could vary from $0.94V to 1.18V$ at room temperature as shown in figure 4.5. The trimming circuit does not use any resistor or blow-up fuses, only CMOS devices and diodes. The proposed circuit is carefully described in the next sections.

4.2 Leakage Current for Temperature Compensation

As it was presented in figure 4.12 process variations can significantly affect the temperature response of the proposed circuit, even if we size our devices to achieve the lowest TC. For that reason we propose a trimming technique that uses a leakage current in order to correct the temperature response of a voltage reference post-silicon. Most trimming techniques use blown fuses or digital signals to adjust a physical parameter of a device (e.g., W/L ration of a CMOS device) [18]. The proposed trimming circuit is able to change the temperature response of the output from PTAT to CTAT or vise versa without targeting a specific voltage level. This is done by either injecting into, or drawing a current from, the output voltage reference. The block diagram of the proposed trimming circuit is presented in the next figure.

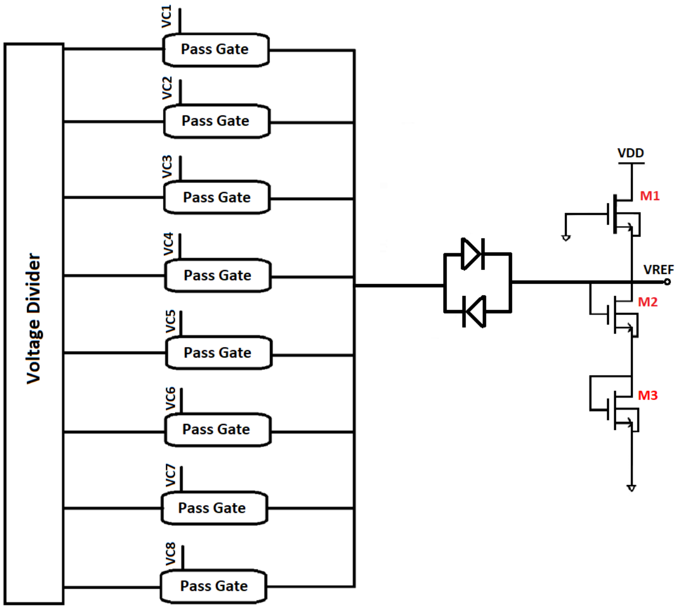


Figure 4.13: Block Diagram of the Temperature Domain Trimming

The voltage divider which is connected to the same supply voltage as the reference circuit, provides eight input voltages in the pass gates. These pass gates are transmission gates which are controlled off-chip through signals $VC1$ to $VC8$. When a switch is enabled i.e., when a control signal goes high, the voltage potential V_{trim} appears in the left side of the diode configuration see figure 4.1 .

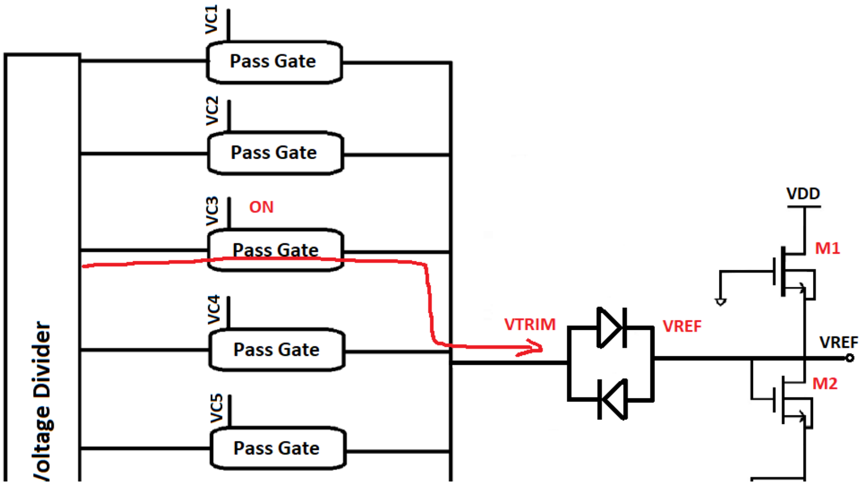
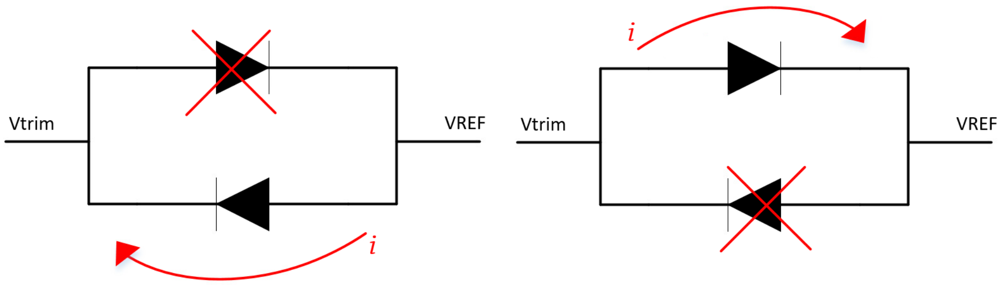


Figure 4.14: Block Diagram of the Temperature Domain Trimming

When $V_{TRIM} > V_{REF}$, the bottom diode is OFF, and the top diode is ON and starts leaking current to the voltage reference as seen in Fig 4.15. On the other hand, when $V_{TRIM} < V_{REF}$ the top diode is OFF, the bottom diode is ON and current is leaking from the voltage reference (figure 4.15).



(a) Current Flow when $V_{REF} > V_{TRIM}$ (b) Current Flow when $V_{REF} < V_{TRIM}$

Figure 4.15: Diode Configuration that Enables Current Injection for Temperature Compensation

When current is injected into the drain of $M2$ then the temperature response of the voltage reference output becomes proportional to temperature. On the other hand, when current is drawn from that point, the output becomes complementary to temperature. The reason for that is the change in the subthreshold slope factor (m) of transistor $M1$. In the next to figures, it is presented the temperature response of m while one switch is enabled at a time.

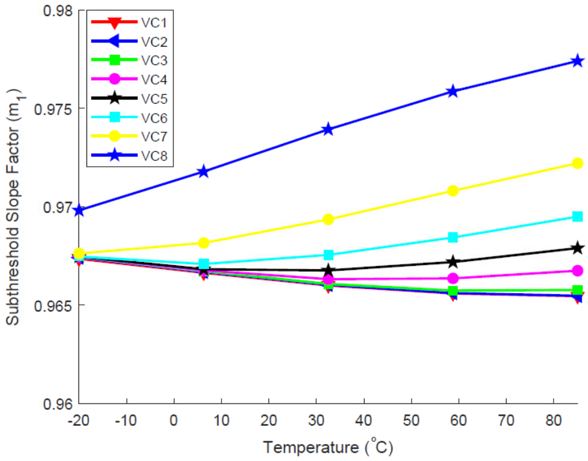


Figure 4.16: Block Diagram of the Temperature Domain Trimming

Initially the aim was to use a MOS capacitor instead of the two diode configuration to achieve the temperature compensation (see figure 4.18). The idea was to use one of the leakage current mechanisms that occur in MOSFET devices, direct tunneling[26].

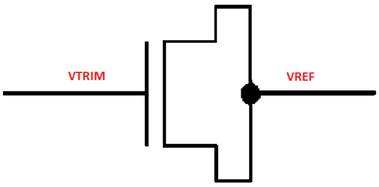


Figure 4.17: MOS Cap for Temperature compensation

In [26] they present the mechanisms of leakage current in a NMOS transistor. These leakage mechanisms can be seen in 4.18. These leakage mechanisms occur due to device scaling and have a negative effect in the device operation especially in digital circuits [26].

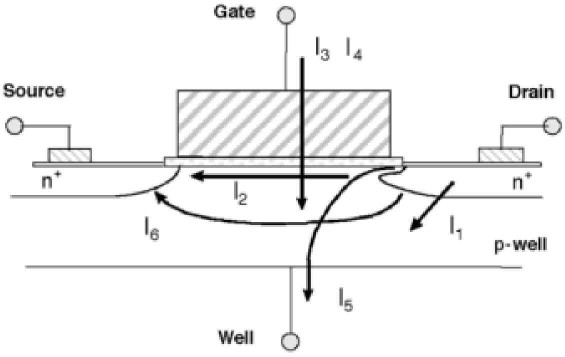


Figure 4.18: Leakage Current Mechanisms in Submicrometer Transistors [26]

Here I_3 and I_4 are leakage currents from the gate to the body of the transistor. These were the currents that wanted to be used to get the leakage current. I_4 usually occurs in small channel devices because of high electric fields. Electrons or holes gain sufficient energy and enter into the oxide layer. This phenomenon is know as the hot-carrier injection. I_3 occurs due to electrons or holes tunneling from the gate oxide into and through the gate oxide. These effects appear in devices with thin gate oxides. There are two known mechanisms responsible for that, the Fowler-Nordbeim (NF) tunneling and the direct tunneling. The FN tunneling is most important at high voltage and moderate oxide thickness and is used to program EEPROM memories. Direct tunneling is most important at lower voltage with thin oxides and is the dominant leakage component.

$$I_{Gate} = WA \left(\frac{V_{DD}}{t_{ox}} \right)^2 e^{-B \frac{t_{ox}}{V_{DD}}} [26] \tag{4.18}$$

In this work both subthreshold current and gate leakage current were investigated and tried to either source or sink leakage current from the output of the 3T voltage reference circuit in order to achieve the desired temperature response.

However, using the gate leakage proved impossible as current did not flow either when $V_{TRIM} > V_{REF}$, or $V_{TRIM} < V_{REF}$. Simulations made in 65nm showed that using the MOS capacitor connection could definitely work. The reason why the mos capacitor configuration did not work for this process was the thickness of the gate oxide. As [37] and [26] clearly report, direct tunneling under the gate with supply voltage being in nominal range of the device ratings, occurs for oxide thickness below $3nm$. Figure 4.19 displays the current density as CMOS technology and therefore oxide thickness scales down.Thus it proved impossible to use the MOS capacitor configuration to introduce the leakage current.

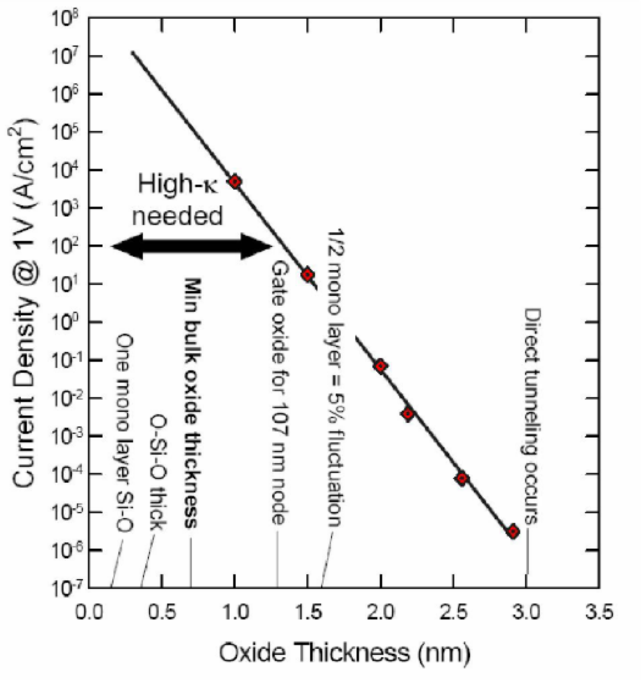


Figure 4.19: Direct tunneling current in thin SiO_2 [37]

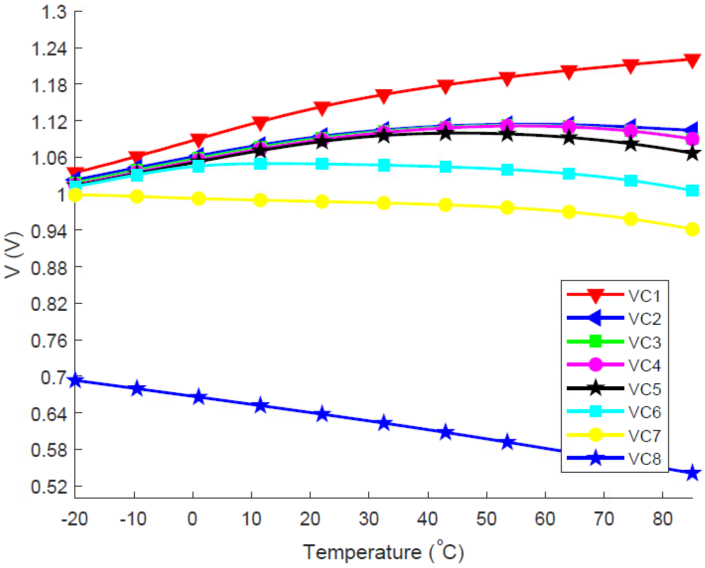


Figure 4.20: Simulation Results of the Effect of the Proposed Trimming Circuit on the Output Voltage

4.3 Body Biasing for Voltage level Regulation

The same trimming topology was used to trim the output of the voltage reference in the voltage domain. The output of the proposed design relies on the difference in threshold voltage of $M1$ and $M2$. As we stated in section 2.3.1, V_{th} is highly dependent on process variations which will cause the voltage level of the reference to deviate quite a lot from the intended value. To adjust this issue, we use the proposed trimming circuit to bias the body of transistor $M1$. This might not be possible in all cases since not all processes offer deep n-well devices. The difference here is that we did not use the two diodes to do that. The block diagram of the topology is presented in the next figure.

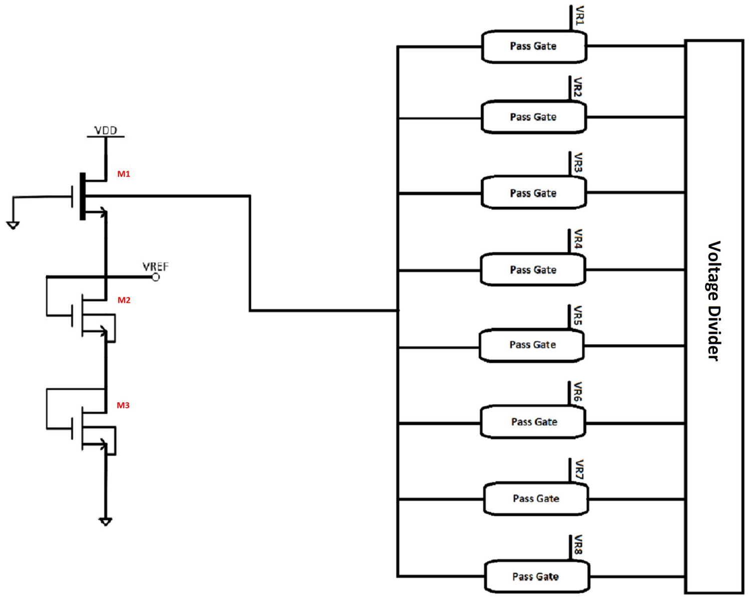


Figure 4.21: Block Diagram of the Proposed Trimming Scheme used to Bias the Body of $M1$

This way, by enabling one or more of the pass gates, we can bias the body of $M1$. As it was presented in equations 4.1, 4.7 the source to body potential plays a crucial role defining the threshold voltage V_{th}. When V_{sb} in a device increases the threshold voltage goes up, and when V_{sb} goes down the threshold voltage follows. On the other hand the subthreshold slope factor m has the exactly opposite response. It is not clear, which of the two parameters dominates or if there is another mechanism that defines if the output voltage will either increase or decrease with different values of V_{sb}. Simulations show that when $M1$ is biased with a higher voltage, hence V_{sb} becomes smaller, then the output voltage increases. On the other hand voltage reference drops when V_{sb} increases. Next figure presents the simulation results of the output voltage when the body of $M1$ is biased by different voltages. $VR1$ to $VR8$ are the signals that enable the pass gates. When $VR1$ is enabled, the body of $M1$ is biased by the higher voltage of the voltage divider. $VR8$ will bias the body of the device with the lowest voltage given by the voltage divider circuit.

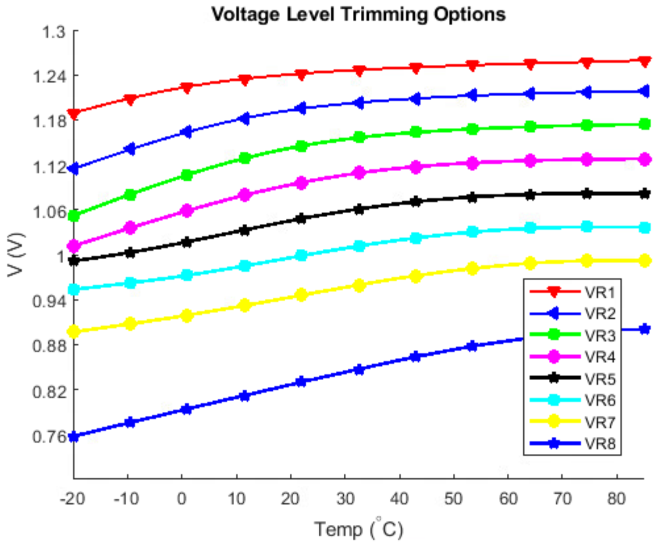


Figure 4.22: Simulation Results of the Output Voltage With different Body Biasing for

4.4 Voltage Divider Ladder

The voltage divider which is part of the proposed trimming topology is a CMOS only configuration which consists of 8 PMOS diode connected devices. A simplified version with only 3 NMOS devices is presented in the next figure.

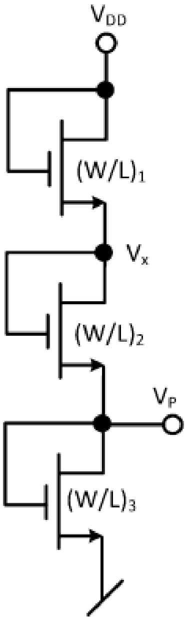


Figure 4.23: CMOS voltage divider
[24]

Device	P1-P8
Widht (μm)	1
Length (μm)	6

Table 4.2: Transistor size of the voltage divider

The working principle is the same for both types of devices. The current that flows through all transistors is the same, and, hence, the point V_p can be defined as:

$$V_p = \frac{\sqrt{\frac{(W/L)_1}{(W/L)_2}}(V_{DD} - V_{th})}{1 + \sqrt{\frac{(W/L)_1}{(W/L)_2}} + \sqrt{\frac{(W/L)_1}{(W/L)_3}}}[24] \tag{4.19}$$

If we make the devices identical then $V_P = (V_{DD} - V_{th})/3$. That means that the drain voltage at any of these diode connected devices will be a portion of $(V_{DD} - V_{th})$ depending on how many transistors we use. This way, in our configuration we can create eight voltages from $(V_{DD} - V_{th})$ to $(V_{DD} - V_{th})/8$. The layout of the voltage divider is presented in the next figure.

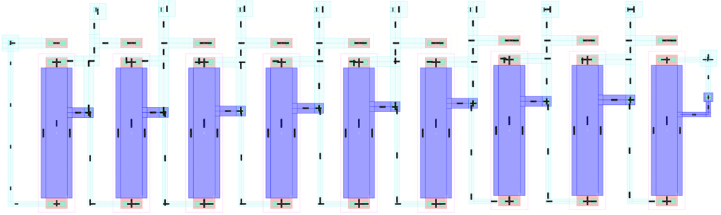


Figure 4.24: Layout of the voltage divider

4.5 Pass Gate

The pass gate or transmission gate, takes as input, each of the output of the voltage divider. Then by the control signals V_C and V_R passes that voltage to its output. There are eight pass gates, one for each output voltage of the divider. The outputs of the pass gates are shorted together and are connected either to the two diode configuration (figure 4.13) or into the body of $M1$ (figure 4.22). The schematic and layout of the pass gate are shown below.

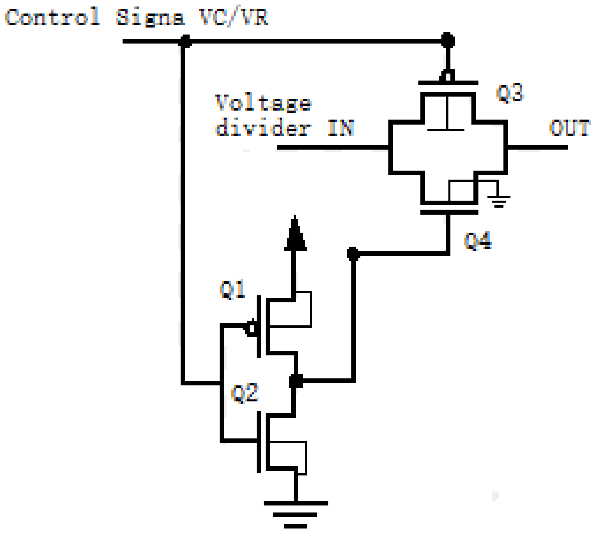


Figure 4.25: Buffer

Device	Q1	Q2	Q3	Q4
Widht (μm)	2	0.72	20	20
Length (μm)	0.18	0.18	0.5	0.6

Table 4.3: Transistor size of the transmission gate

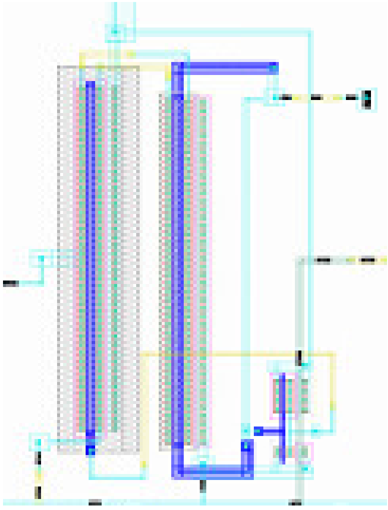


Figure 4.26: Layout of The Pass Gate/Transmission Gate

4.6 Buffer Output

As the voltage reference is shown to consume very little current, the circuit becomes sensitive to any kind of distortion. For example, when doing testing, even an oscilloscope probe can load the output of the voltage reference, resulting in wrong results. For that reason it is essential to use an opamp connected in a buffer configuration to drive the output voltage of our circuit (see figure 4.24).

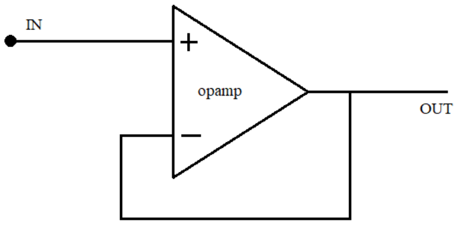


Figure 4.27: Buffer

In the input we connect the output of the voltage reference and the output is connected to one of the pads of the the chip. The buffer topology used was provided by supervisor Philipp Dominik Häfliger. The schematic and the layout of the buffer are illustrated in figures 4.28 and 4.29 respectively.

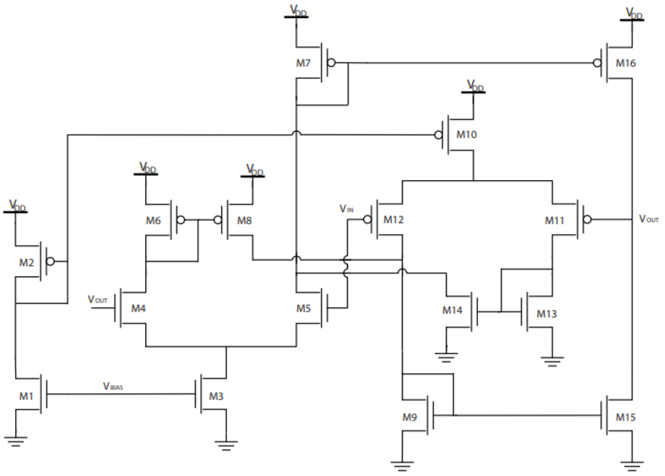


Figure 4.28: Schematic of the Output Buffer

Device	M1-M14(NMOs)	M1-M14(PMOS)	M15	M16
Widht (μm)	1.5	3	20	40

Table 4.4: Transistor Sizes of the Output Buffer $L = 0.5\mu m$

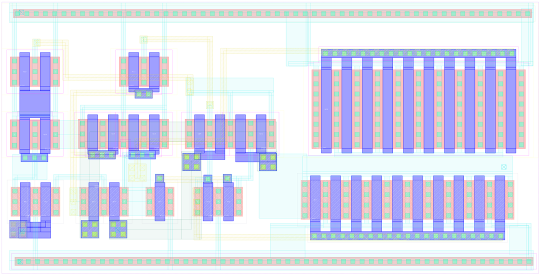
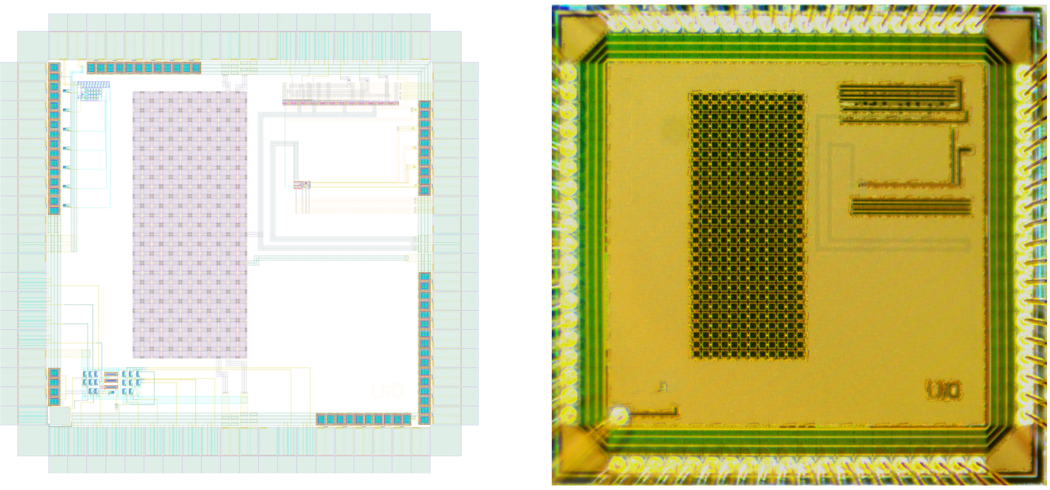


Figure 4.29: Layout of the Output Buffer

Chapter 5

Simulation and Testing Results

This chapter presents the simulation and measurement results of the presented design. Since, only the entire circuit (trimming included) was taped out, there are only simulation results for the 3T topology. The circuit was implemented in a 180nm BCD 2 process by TSMC. The entire layout together with an image of the actual fabricated chip can be seen in figure 5.1.



(a)

(b)

Figure 5.1: a) Layout of the Entire Chip. b) Microscope Image of the Entire Chip

5.1 3 Transistor Core Cell (3T) Simulation Results

5.1.1 Line Regulation

The 3T topology has an output voltage of 1.1V with the lowest operating supply 1.2V at room temperature. It achieves a really low LS of $0.6\%/V$ from 1.2 to $5V$. The output voltage with respect to different supplies is illustrated in the next figure.

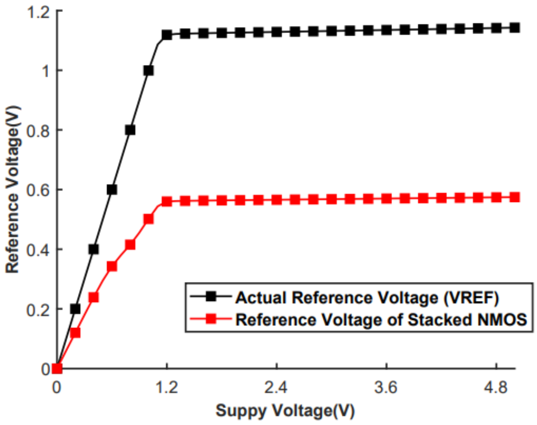


Figure 5.2: Voltage reference Output and stacked NMOS Drain Voltage

The red line in the above figure represents the drain voltage of transistor $M3$. This voltage is $0.55V$ at room temperature and has an LS of $0.65\%/V$. The output voltage of this stacked diode connected NMOS can be used for applications which require lower operating voltages.

However this is not the case when body effect is present. As it is presented in figure 5.3 both V_{REF} and the drain voltage of $M3$ differ quite a lot. We see here that the voltage level drops and LS increases. The design presents an LS of $5.6\%/V$ from $0.9V$ to $5V$.

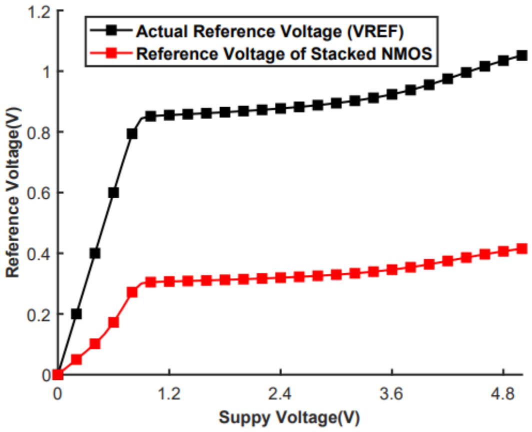


Figure 5.3: Voltage reference Output and stacked NMOS Drain Voltage With Body Effect

5.1.2 Temperature Coefficient

In the next figure, the temperature response of the output voltage is shown. With device sizing for optimal TC, the design shows a near zero temperature coefficient of $7ppm/^oC$ at a temperature range from -20^oC to 85^oC. The sizing of the devices for this measurement are shown in table 5.1. It is important to notice here, that this TC can be achieved only at a power supply of $1.2V$. If the supply varies, this TC is set to change.

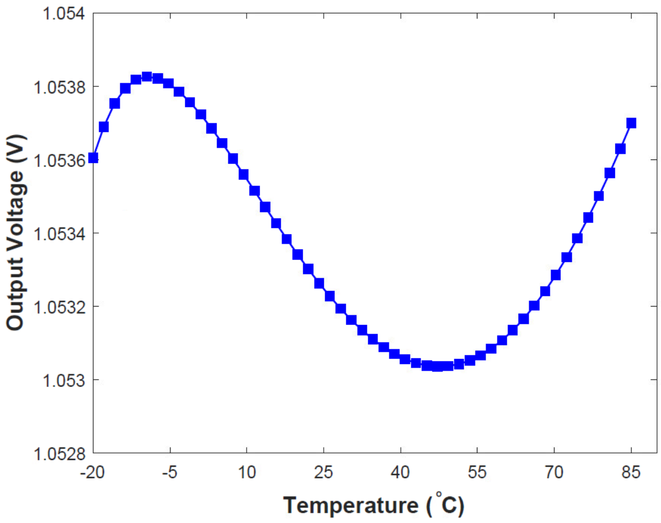


Figure 5.4: Temperature Response of the Output Voltage

Device	M1	M2	M3
Width (μm)	2	7	7
Length (μm)	40	39	39

Table 5.1: Transistor Sizing for best TC

The next figure illustrates how the body effect can reduce the output voltage and alter the temperature response of the voltage reference for the same temperature range..

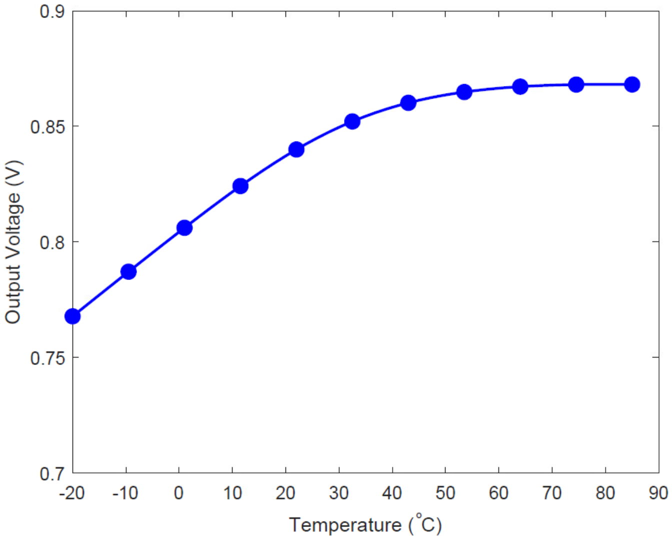


Figure 5.5: Output of the Voltage Reference Circuit With Varying Temperature When there is body effect for $M1$ and $M2$

Temperature coefficient for determined transistor sizes is only measured for a specific supply voltage. Therefore, it is important to demonstrate demonstrate how TC of specific transistor sizes can be changed with varying supply voltages. The next figure illustrates the TC both with and without the body effect.

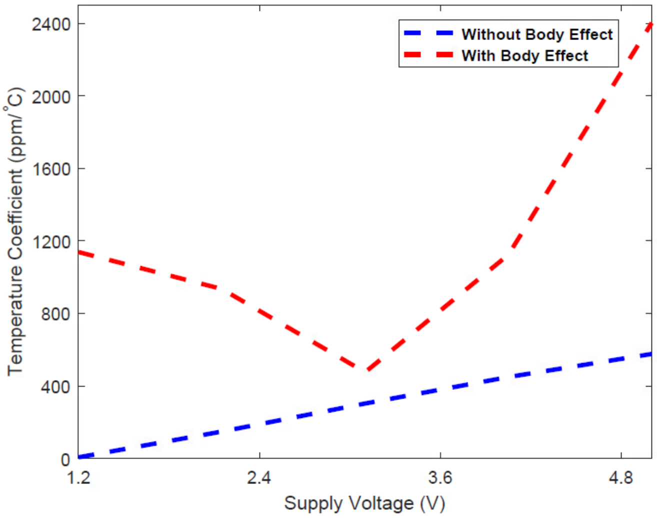


Figure 5.6: TC for different supply voltages

5.1.3 Power Consumption

The power consumption of the design over a specific temperature range and for supply voltages from $1.2V$ to $5V$ is presented in figure 5.5. The design has a really small power consumption of only $33pWatts$ at room temperature when $VDD = 1.2V$. However, the power consumption shown here matches transistors that are not sized for optimal TC (see table 4.1).

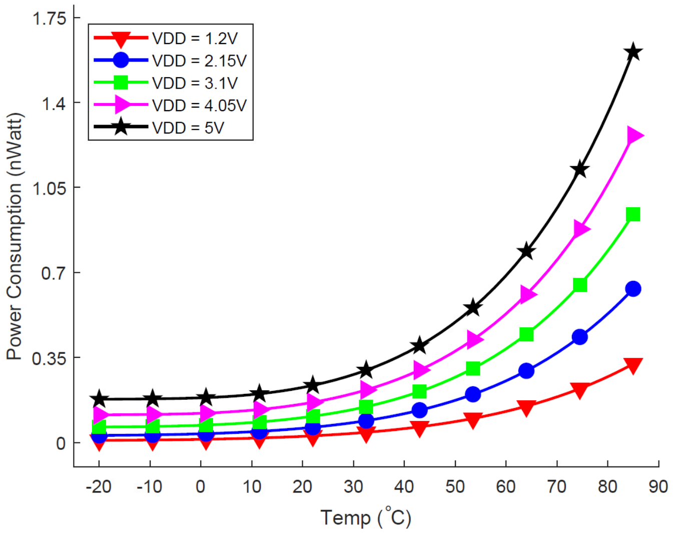


Figure 5.7: Power Consumption

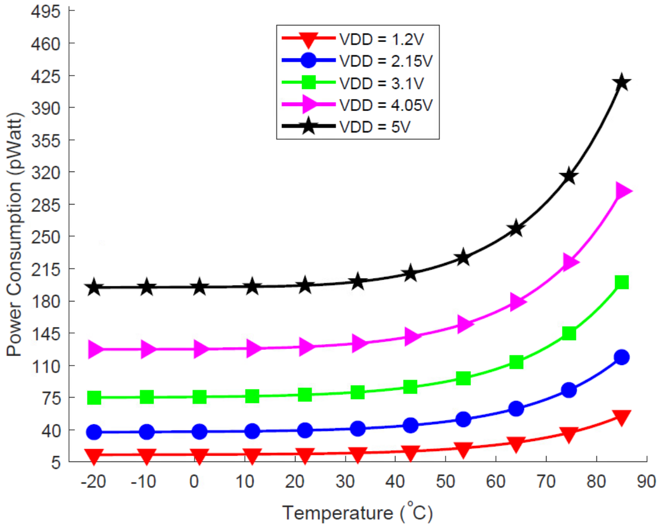


Figure 5.8: Power Consumption With body effect

As it is shown in figure 5.8 the body effect has a crucial role when it comes to power consumption. The design is now conducting even less current as VDD increases. In particular, it shows a power consumption of only $200pW$ at room temperature when $VDD = 5V$. This value only goes up to around $430pW$ when temperature increases to 80^oC.

5.1.4 PSRR & Output Noise

Here we present measurements done for PSRR and the output noise of the design when different capacitors connected at the output.

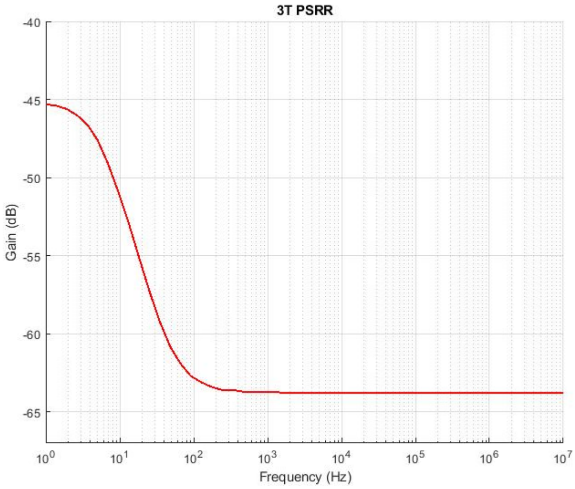


Figure 5.9: 3T PSRR

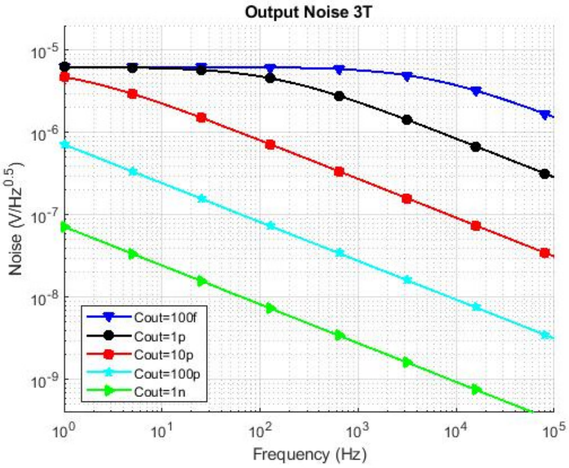


Figure 5.10: Output Noise with Different Capacitor Values

5.1.5 Process Variations

To predict the effects of process variations in the 3T topology, Monte Carlo simulation was done. Both the Monte Carlo and mismatch library were used for this simulation. The results are illustrated in the figure below. As it was expected, the design is quite sensitive to mismatches. The output voltage at room temperature with a supply of 1.2V deviates from the mean value (1.05V).

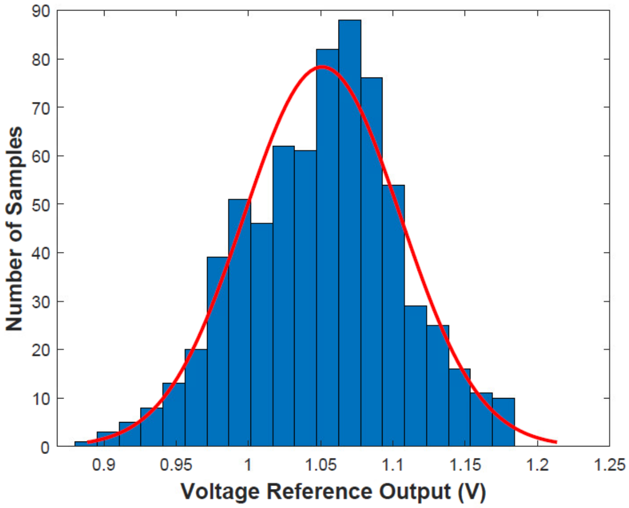


Figure 5.11: Monte Carlo Analysis

5.2 Post Layout Simulation & Testing Results For The Tapped-Out Design (Trimming Included)

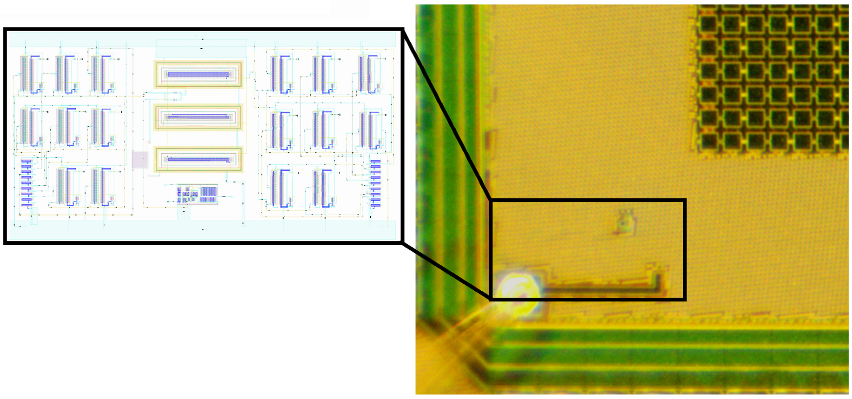


Figure 5.12: The Entire Design Placed in the Bottom Left Corner of the Chip

5.2.1 Line Sensitivity

In the post layout simulations done for the tapped out circuit (trimming topology included), the voltage reference output exceeds an LS of $5.6\%/V$. The output voltage is presented in the figure below. Since the trimming topology enables the biasing of the body of $M1$ in order to trim in the voltage domain, we target this output voltage by enabling the pass gates $VR1$ and $VR2$.

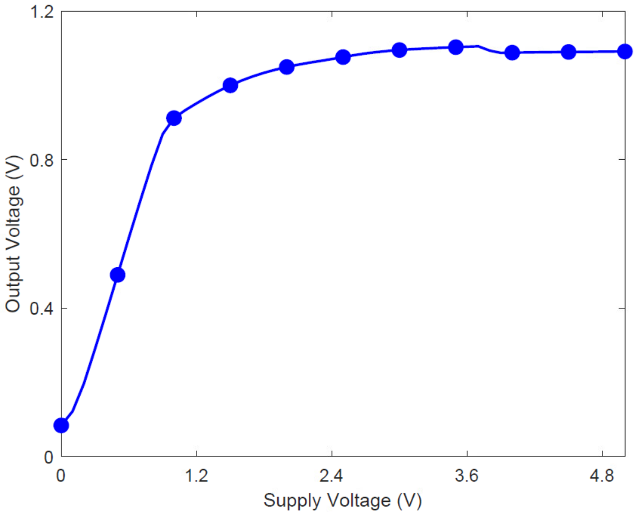


Figure 5.13: Line sensitivity post layout simulations

Power Consumption

Because of the implementation of the trimming configuration, the power consumption of the design rises scientifically. The plot here demonstrates the power consumption for only two supply voltages, $1.2V$ and $2.15V$. The reason for that is that for supply voltages higher than $2.5V$ the design starts consuming tens of $\mu watts$ of power. Even when $VVD = 2.15$ the amount of power the design consumes is still high, in comparison with when the trimming topology was not implemented. However for $VDD = 1.2V$ the circuit consumes $270pW$ at room temperature, which value is still in the ultra-low temperature range.

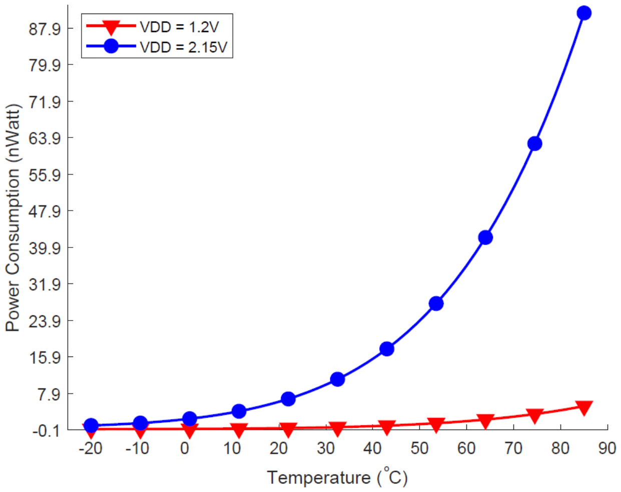


Figure 5.14: Power consumption post layout

5.2.2 Temperature Coefficient

The proposed trimming circuit has many option for trimming the circuit both in the in the temperature and in the voltage domain by enabling any of the 16 switches (pass gates). Simulation results show that the best TC of $85ppm/^o$ can be achieved when $VR_{2,3}$ and $VC_{6.7}$ are on.

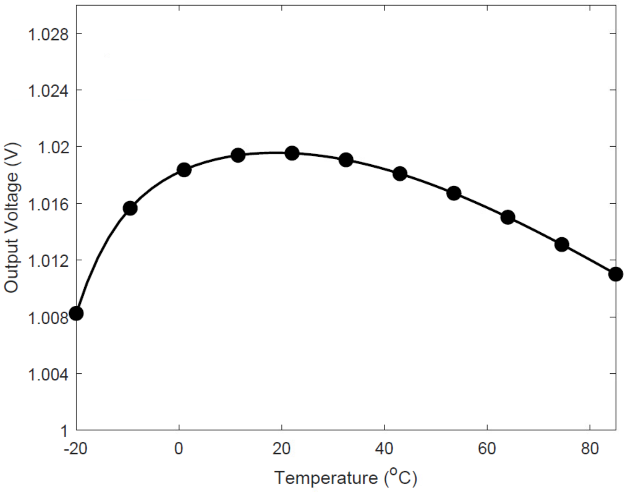
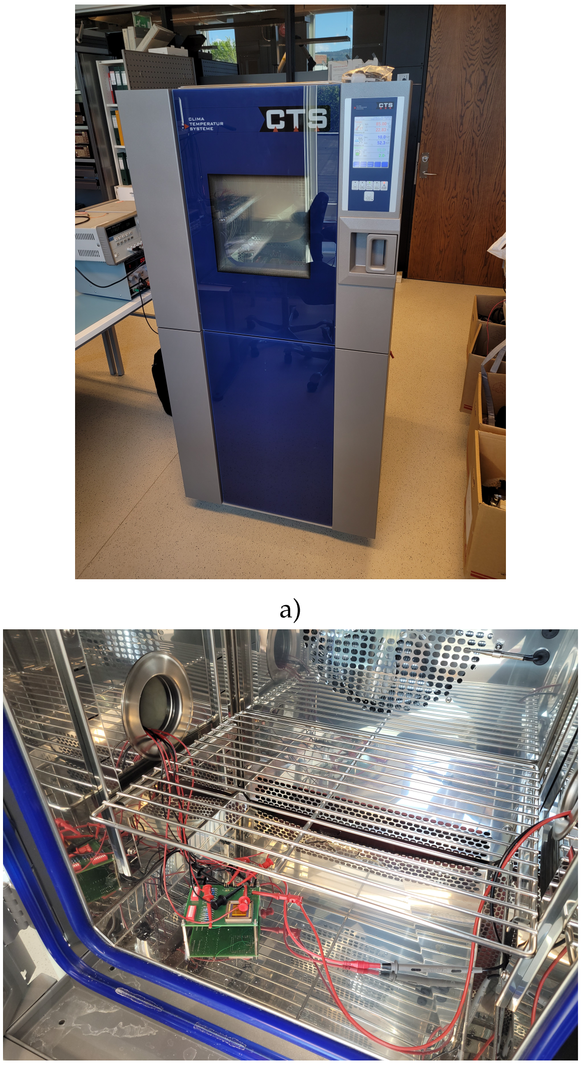


Figure 5.15: Post Layout Simulation Results for the Output Voltage with lower TC

5.3 Testing Results

Testing was done using the equipment in the lab of the University of Oslo. A temperature chamber was used to test the temperature response of the design. Due to limited time not excessive testing was done in order to see the full operation of the proposed circuit.



a)

b)

Figure 5.16: Temperature Chamber

Figure 5.17 shows how the temperature response of the output was trimmed. The TC without trimming is 495.4ppm^oC and post trimming it drops to 265$ppm/^oC$.

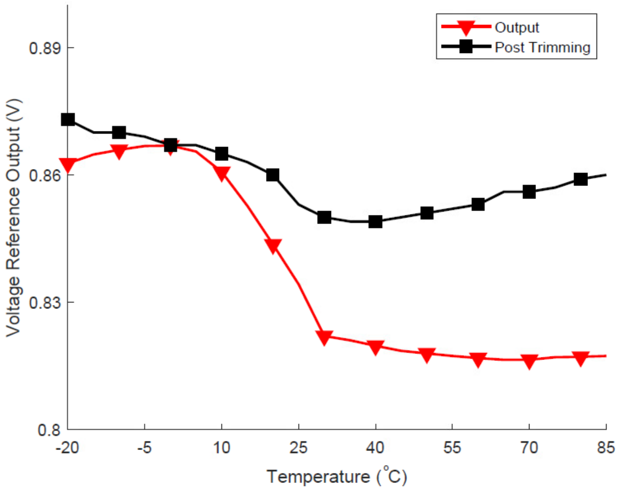


Figure 5.17: TC Before and After Trimming for one Sample

LS improved from 53%/V to 13%/V post trimming for supply voltages from 1.2V to 2.2V (figure (5.15). The targeted voltage in this case was 1.1V

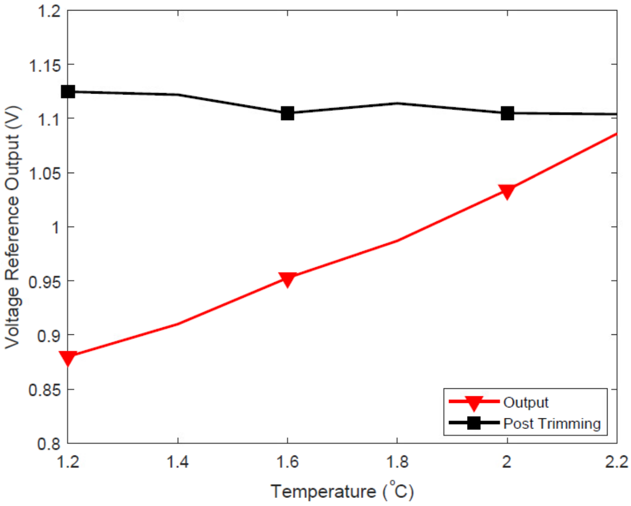


Figure 5.18: LS Before and After Trimming for one Sample

It is important to note here that not every switch combination of the trimming circuit was tested, therefore both LS and TC could be further improved.

Chapter 6

Conclusion

In this thesis′ we explored the recent developments of ultra-low power always/on voltage reference circuits. Special attention was given to the designs operating in the picowatt power consumption range. The main focus of such circuits lies on the Internet of Things (IoT) and other low power applications, such as energy harvesting systems and energy-autonomous platforms. Because of the extremely constrained power budget of these applications, the voltage reference circuit design needs to be adjusted to this constraint. Prior work showed that these ultra-low power references are often sensitive to process variations, which reduce the design yield and make it unreproducible. In addition, the reported output voltage of many of these devices is low, which reduces the dynamic performance of the analog circuits.

Therefore, we explored and implemented a 3-transistor (3T) configuration, which consists of two diode-connected, high-threshold voltage, NMOS devices, which are biased by a depletion (near-$0V$ threshold voltage) NMOS in the weak inversion region. We then propose a trimming technique, which in comparison to prior work does not use any blow fuses or signals to alter the physical parameters of the device (e.g the W/L ratio of a CMOS device). Instead, the proposed trimming circuit introduces a leakage current on the output of the voltage reference in order to compensate for the temperature response of the design. The trimming technique was also used to trim the reference in the voltage domain by biasing the body of the current source transistor. Both the 3T topology and the trimming circuits are described in detail, and design considerations are discussed. The final design was fabricated in a BCD GEN2 $180nm$ CMOS process.

The simulation results show that the 3T configuration can operate at 1.2V minimum supply voltages, consuming merely $33.1pWatts$ (best) at room temperature. The proposed circuit achieves a line sensitivity of $0.6\%/V$ measured from 1.2 to $5V$, and a power supply rejection of $-44.9dB$. The silicon area that the 3T design occupies is $0.005605mm^2$. A detailed designed methodology for the implementation of the trimming circuit is described. The post layout simulation results show that, after the trimming implementation, the power consumption of the design is

$270pW$ at room temperature with a supply of $1.2V$. The TC of the voltage reference after trimming has been found to be $87ppm^o/C$ (best) and the output voltage has an LS of 5.6% from 1.2 to $5V$. The area that the trimming topology consumes is $0.02304mm^2$. The trimming topology is set to make the reference better protected against process variations with an acceptable drop in performance. The design also has an output of $1.02V$ at room temperature, which is higher to that of most of the prior works. The given performances make the present proposed circuit suitable for IOT and other ultra low-power applications, which require a higher operating supply.

6.1 Future Work

Due to limited time, only few measurements were done to test the performance of the tapped out design. More testing needs to be done in order to ensure the proper operation of the voltage reference and the proposed trimming circuit. Moreover, the simulation results of the proposed trimming topology showed that the design draws an excessive amount of current for supply voltages higher than $2V$, which occurs possibly due to the trimming circuit. Designing the proposed trimming in a different way (e.g., using fewer pass gates and fewer devices in the voltage divider), could make this design suitable for operating in a wider range of supplies without exceeding the picowatt range and at the same time consuming a smaller silicon area.

Bibliography

[1] Ziyad Al Tarawneh. 'The effects of process variations on performance and robustness of bulk CMOS and SOI implementations of C-elements'. PhD thesis. Newcastle University, 2011.

[2] Narain Arora. *Mosfet modeling for VLSI simulation: theory and practice*. World Scientific, 2007.

[3] R Jacob Baker. *CMOS: circuit design, layout, and simulation*. John Wiley & Sons, 2019.

[4] STT Bandung, STT Bina Tunggal and STT Dr Khez Muttaqien. 'International technology roadmap for semiconductors'. In: (2013).

[5] Saurav Bandyopadhyay et al. 'A 1.1 nW energy-harvesting system with 544 pW quiescent power for next-generation implants'. In: *IEEE journal of solid-state circuits* 49.12 (2014), pp. 2812–2824.

[6] Duane Boning and Sani Nassif. 'Models of process variations in device and interconnect'. In: *Design of high performance microprocessor circuits* (2000), p. 6.

[7] A Paul Brokaw. 'A simple three-terminal IC bandgap reference'. In: *IEEE Journal of Solid-State Circuits* 9.6 (1974), pp. 388–393.

[8] Arne E Buck et al. 'A CMOS bandgap reference without resistors'. In: *IEEE Journal of Solid-State Circuits* 37.1 (2002), pp. 81–83.

[9] William D Callister, David G Rethwisch et al. *Materials science and engineering: an introduction*. Vol. 9. Wiley New York, 2018.

[10] Maurizio Capra et al. 'Edge computing: A survey on the hardware requirements in the internet of things world'. In: *Future Internet* 11.4 (2019), p. 100.

[11] T Chan Carusone, D Johns and K Martin. 'Analog integrated circuit design'. In: *New York: Wiley. Hannane Gholamnataj was born in Babolsar, Iran, on September* 16 (2012), p. 1984.

[12] R Dobkin. 'Monolithic temperature stabilized voltage reference with 0.5 ppm/° drift'. In: *1976 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*. Vol. 19. IEEE. 1976, pp. 108–109.

[13] Chrispin Alfred Gray. 'Energy consumption of Internet of Things applications and services'. PhD thesis. 2018.

[14] Vinayak Hande and Maryam Shojaei Baghini. 'Survey of Bandgap and Non-bandgap based Voltage Reference Techniques'. In: *Scientia Iranica* 23.6 (2016), pp. 2845–2861.

[15] Linden T Harrison. *Current Sources and Voltage References: A Design Reference for Electronics Engineers*. Elsevier, 2005.

[16] David Hilbiber. 'A new semiconductor voltage standard'. In: *1964 IEEE international solid-state circuits conference. Digest of technical papers*. Vol. 7. IEEE. 1964, pp. 32–33.

[17] Hector Hung and Vladislav Adzic. 'Monte carlo simulation of device variations and mismatch in analog integrated circuits'. In: *Proc. NCUR 2006* (2006), pp. 1–8.

[18] Chi-Wah Kok and Wing-Shan Tam. *CMOS voltage references: an analytical and practical perspective*. John Wiley & Sons, 2012.

[19] Inhee Lee, Dennis Sylvester and David Blaauw. 'A subthreshold voltage reference with scalable output voltage for low-power IoT systems'. In: *IEEE Journal of Solid-State Circuits* 52.5 (2017), pp. 1443–1449.

[20] Luca Magnelli et al. 'A 2.6 nW, 0.45 V temperature-compensated subthreshold CMOS voltage reference'. In: *IEEE Journal of Solid-State Circuits* 46.2 (2010), pp. 465–474.

[21] Eiju Matsumoto. 'Edward Weston made his mark on history of instrumentation'. In: *IEEE instrumentation & measurement magazine* 6.2 (2003), pp. 46–50.

[22] Harry Neuteboom, Ben MJ Kup and Mark Janssens. 'A DSP-based hearing instrument IC'. In: *IEEE Journal of Solid-State Circuits* 32.11 (1997), pp. 1790–1806.

[23] Arthur Campos de Oliveira. 'Temperature compensated subthreshold CMOS voltage references for ultra low power applications'. In: (2017).

[24] Ajishek Raj, Data Ram Bhaskar and Pragati Kumar. 'Two quadrant analog voltage divider and square-root circuits using OTA and MOSFETs'. In: *Circuits, Systems, and Signal Processing* 39.12 (2020), pp. 6358–6385.

[25] Behzad Razavi. *Design of analog CMOS integrated circuits*., 2005.

[26] Kaushik Roy, Saibal Mukhopadhyay and Hamid Mahmoodi-Meimand. 'Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits'. In: *Proceedings of the IEEE* 91.2 (2003), pp. 305–327.

[27] Mingoo Seok et al. 'A portable 2-transistor picowatt temperature-compensated voltage reference operating at 0.5 V'. In: *IEEE Journal of Solid-State Circuits* 47.10 (2012), pp. 2534–2545.

[28] Ashish Srivastava, Dennis Sylvester and David Blaauw. *Statistical analysis and optimization for VLSI: Timing and power*. Vol. 59. Springer, 2005.

[29] Yuan Taur and Tak H Ning. *Fundamentals of modern VLSI devices*. Cambridge university press, 2021.

[30] Y.P. Tsividis and R.W. Ulmer. 'A CMOS voltage reference'. In: *IEEE Journal of Solid-State Circuits* 13.6 (1978), pp. 774–778. DOI: 10.1109/JSSC.1978.1052049.

[31] Giuseppe Vita and Giuseppe Iannaccone. 'A Sub-1 V, 10 ppm/? C, Nanopower Voltage Reference Generator'. In: *2006 Proceedings of the 32nd European Solid-State Circuits Conference*. 2007.

[32] E. Vittoz and J. Fellrath. 'CMOS analog integrated circuits based on weak inversion operations'. In: *IEEE Journal of Solid-State Circuits* 12.3 (1977), pp. 224–231. DOI: 10.1109/JSSC.1977.1050882.

[33] E.A. Vittoz and O. Neyroud. 'A low-voltage CMOS bandgap reference'. In: *IEEE Journal of Solid-State Circuits* 14.3 (1979), pp. 573–579. DOI: 10.1109/JSSC.1979.1051218.

[34] Hui Wang and Patrick P Mercier. 'A 420 fW self-regulated 3T voltage reference generator achieving 0.47%/V line regulation from 0.4-to-1.2 V'. In: *ESSCIRC 2017-43rd IEEE European Solid State Circuits Conference*. IEEE. 2017, pp. 15–18.

[35] Neil HE Weste and David Harris. *CMOS VLSI design: a circuits and systems perspective*. Pearson Education India, 2015.

[36] Robert J Widlar. 'New developments in IC voltage regulators'. In: *IEEE Journal of Solid-State Circuits* 6.1 (1971), pp. 2–7.

[37] Hei Wong. 'The current conduction issues in high-k gate dielectrics'. In: *2007 IEEE Conference on Electron Devices and Solid-State Circuits*. IEEE. 2007, pp. 31–36.

[38] T Ytterdal. 'CMOS bandgap voltage reference circuit for supply voltages down to 0.6 V'. In: *Electronics letters* 39.20 (2003), pp. 1427–1428.

Appendix A

A.1 PCB

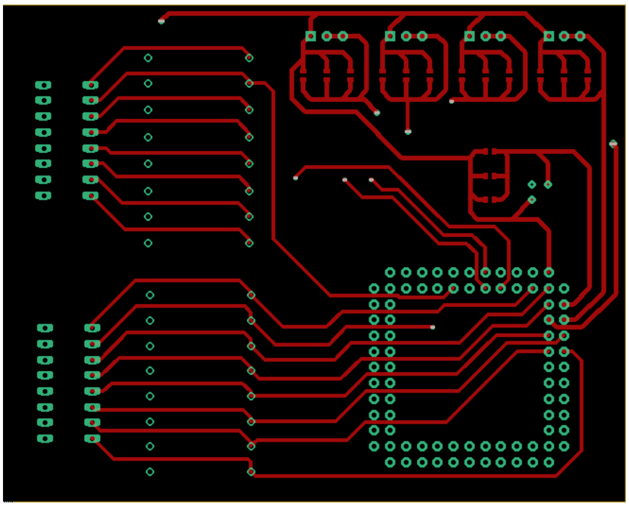


Figure A.1: PCB Design.Top Electric Layer

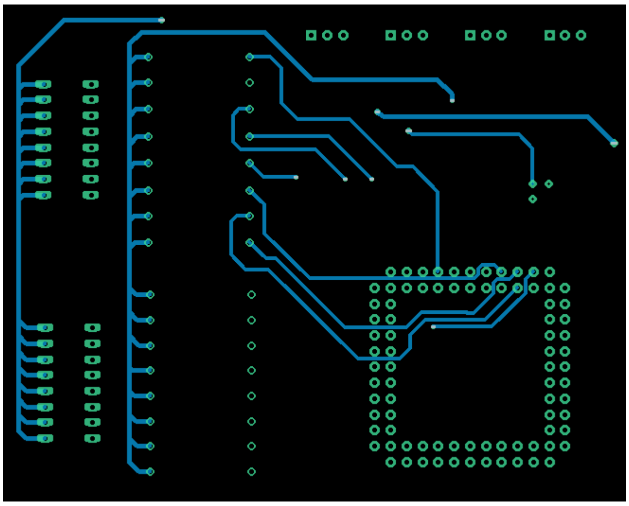


Figure A.2: PCB Design.Bottom Electric Layer

A.2 ASIC

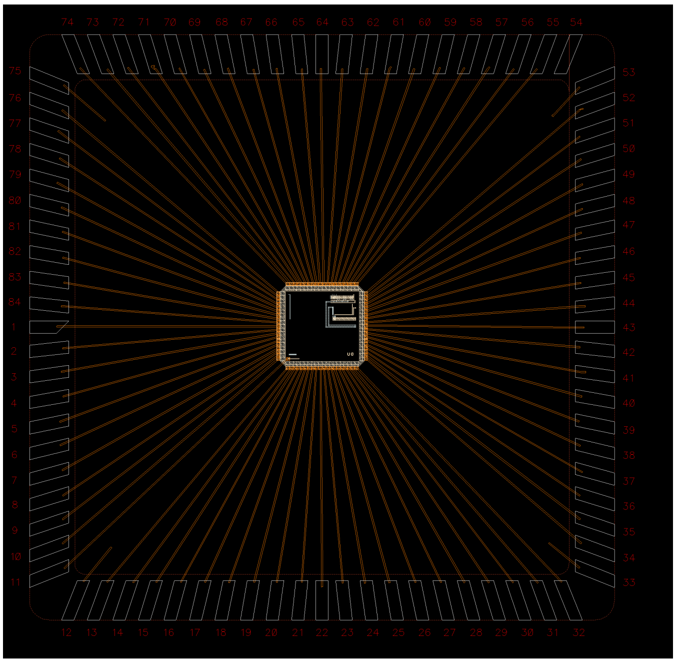


Figure A.3: Boning Diagram of the ASIC to THE JLCC84 Package