Sub-1 V, 4 nA CMOS Voltage References with Digitally-Trimmable Temperature Coefficient

Author:
Peter Luong

A Thesis
Presented to Lakehead University
in Partial Fulfillment of the Requirement for the Degree of
Master of Science
in
Electrical and Computer Engineering

Thunder Bay, Ontario, Canada

2014

Abstract

Voltage references are fundamental to mixed signal converters which are widely used in electronics. Hence there are significant advantages in having the voltage reference operate with less power while minimizing area consumption and maintaining performance. Past designs have suffered from issues related to process variations which adversely affect the temperature coefficient of the circuit output. To compensate for these process variations, a means to modify the temperature coefficient are proposed and experimentally verified with two circuit architectures.

Five test chip samples implement these architectures in a 0.35 μm CMOS process. Design methodologies for both architectures are presented. Design techniques include the use of a high-swing cascode to improve Line Sensitivity while minimizing additional power consumption, accounting for a well-matched layout, and the effect of leakage currents on the performance of the circuit.

Layout schematics, performance figures, test methodologies and results are presented. Each circuit dissipates less than 4 nW and operates down to 0.9 V or better with Line Sensitivity and Power Supply Rejection Ratio of less than 0.15 %/V and -58 dB respectively, while consuming an area of 0.053 mm^2 or less. The experimental average and median temperature coefficient was less than 26 ppm/°C and 22 ppm/°C respectively in the $-20\,^{\circ}$C to $80\,^{\circ}$C range, with the best performance being less than 8.1 ppm/°C. Areas of improvement and potential areas of future research are then identified to facilitate advancement of this work.

Acknowledgments

This work contains many thanks:

- Thanks Dr. Carlos Christoffersen for continuing to push the boundaries of what can be done and encouraging the free expression of new ideas.
- Thanks Dr. Conrado Rossi-Aicardi (of Universidad de la Republica, Uruguay) for significant assistance and advice on design, layout, and insisting on perfection in all endeavours.
- Thanks Dr. Carlos Dualibe (of Universite de Mons, Belgium) for input on technical accuracy and additional considerations that improved the completeness of the work.
- Thanks to the reviewers Dr. Krishnamoorthy Natarajan and Dr. Oleg Rubel for ensuring this thesis meets or exceeds a standard of quality for work produced from this university.
- Thanks Dr. Kam Leung and Tallie Macdonald for allowing the use of the thermometers, temperature-controlled freezers and incubators from the Lakehead Biology lab.
- Thanks CMC Microsystems for fabrication-related work and advice.
- Thanks National Research Council of Canada (NSERC) for the funding.
- Thanks Lakehead University and Thunder Bay for the environment that facilitates focus and perserverence.
- And most importantly, thanks family, friends, instructors, former co-workers, and peers for directly or indirectly instilling discipline, digilence, and the willingness to continously improve.

– *Peter Luong* (2014)

Contents

0.1 List of Symbols

Symbol	Definition
α_T	Change in output voltage with respect to temperature
C_{OX}	Gate oxide capacitance
E_G	Energy bandgap of semiconductor
ϵ_0	Permittivity of free space
γ	Body effect coefficient factor
g_m	Transconductance
I_C	Collector current (BJT)
I_D	Drain current
i_F, i_R	Forward / Reverse inversion coefficient
I_S	Saturation current (BJT)
I_{SQ}	Sheet normalization current
K_{VT}	Expected drop in threshold voltage at a reference temperature
λ	Channel length modulation coefficient
μ	Carrier mobility
n	Sub-threshold slope factor
N_A	Acceptor ion concentration
N_T	Number of unit transistors
ϕ_F	Fermi potential
$\Phi_{POLY/SI}$	Poly-Silicon work function

Symbol	**Definition**
S	Size, or Width/Length ratio, of a MOSFET
τ	Time constant
T, T_R	Temperature, Reference temperature (usually 300 K)
t_{OX}	Gate oxide thickness
$U_T = kT/q$	Thermal voltage, Boltzmann constant, temperature, and electron charge, respectively
V_{BE}	Base-emitter voltage
V_G	Gate-bulk voltage
V_{GS}	Gate-source voltage
V_D	Drain-bulk voltage
V_{DS}	Drain-source voltage
V_P	Pinch-off voltage
V_S	Source-bulk voltage
V_T	Threshold voltage
V_{T0}	Threshold voltage at 0 K

0.2 List of Figures

0.3 List of Tables

Chapter 1

Introduction

1.1 Chapter Overview

This chapter begins with addressing why this work is undertaken before defining what a voltage reference is and what it is used for. Historical details then follow, so that the reader is aware of the origins of the voltage reference and the subsequent developments that followed. The objectives and thesis overview are then stated to provide a rationale behind how the content is presented in this work.

1.2 Motivation

There is an on-going demand for electronic devices to be smaller, use less energy, and still work reliably. Devices such as A/D and D/A converters, medical implantable devices, sensors interfaces and instrumentation circuits rely on a fundamental building block called a voltage reference to achieve these qualities [1].

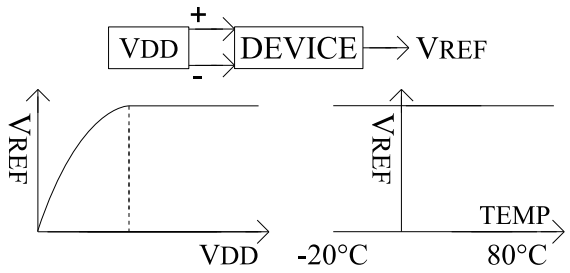


Figure 1.1: Qualities of an ideal voltage reference

1.3 Defining a Voltage Reference

Shown in Fig. 1.1, a voltage reference is a device which produces a constant output voltage. Once a certain minimum voltage is supplied to the device, the output ideally does not change even if there are voltage ripples in the power supply or if the power supply itself has increased in voltage. In addition, the voltage reference output ideally remains constant even with changes in temperature.

1.4 Applications of Voltage References

In A/D conversion, the quantization step is defined as a fraction of the output of a voltage reference as shown in Fig. 1.2.

Figure 1.2: Analog to digital conversion using a voltage reference

This is also the basis of digital temperature sensors if the input voltage has a known PTAT (proportional to absolute temperature) dependence [2]. Since the voltage reference is only meant to generate a

certain voltage without consideration for the current, they are generally unsuitable to be used directly as a voltage regulator. However, voltage references are still used in voltage regulators as the basis of the of regulated voltage [1] as depicted in Fig. 1.3.

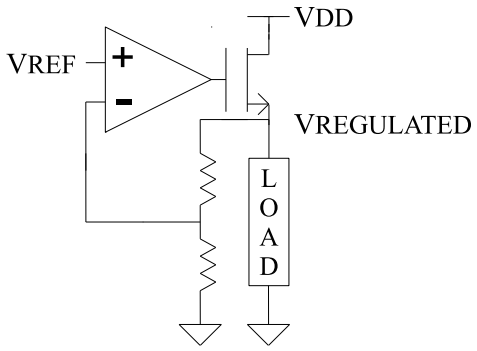


Figure 1.3: Use of the voltage reference output in a voltage regulator

1.5 Historical Background

1.5.1 Origins and Bandgap IC (1971)

Prior to 1971, Zener diodes were commonly used as voltage references [3]. Zener diodes had a relatively high breakdown voltage of 6.2V and exhibited a high level of noise. For those reasons, Zener diodes became unsuitable as technology advances demanded more reliable and power-efficient components. The solution to the problem with Zener diodes was the development of a bandgap IC as shown in Fig. 1.4 [3].

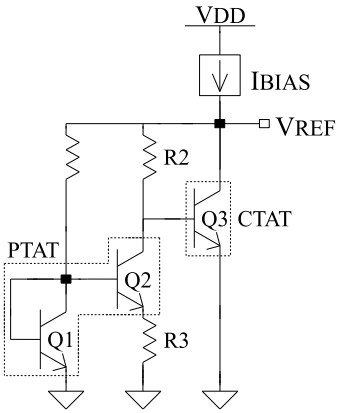


Figure 1.4: The first bandgap IC

The principle behind this circuit was that two components of the circuit with opposite temperature coefficients can be made to cancel out, resulting in an output that does not change with temperature. The two components are the base-emitter voltage, V_{BE}, and the difference between two base-emiiter voltages, ΔV_{BE}. V_{BE} is approximated by [4]

$$V_{BE} \cong E_{G,0K}\left(1 - \frac{T}{T_R}\right) + V_{BE,R}\ \left(\frac{T}{T_R}\right)\ , \tag{1.1}$$

where $E_{G,0K}$ is the energy band gap of the semiconductor extrapolated to zero Kelvin, T_R and $V_{BE,R}$ are the values of the temperature and base-emitter voltages respectively, at a reference point (usually 300 K). ΔV_{BE} is worked out by rearranging the equation for the collector current, I_C,

$$I_C = I_S\ \exp\left(\frac{V_{BE}}{U_T}\right)\ , \tag{1.2}$$

where I_S is the saturation current and $U_T = kT/q$ is the thermal voltage, where k is Boltzmann's constant, q is the electron charge and T is the absolute temperature. Solving for V_{BE}, then subtracting $V_{BE,Q2}$ from $V_{BE,Q1}$, and assuming I_S is identical in both transistors results in

$$\Delta V_{BE} = U_T\ \ln\left(\frac{I_{C1}}{I_{C2}}\right)\ . \tag{1.3}$$

Assuming $I_{E,Q2} \cong I_{C,Q2}$, the output voltage can be expressd as a sum of two different voltages from Eq. (1.1) and Eq. (1.3),

$$V_{REF} \cong [V_{BE,Q3}] + \left[\frac{\Delta V_{BE}}{R3}\right] R2\ . \tag{1.4}$$

Taking the derivative of Eq. (1.4) with respect to temperature and equating it to zero results in,

$$\frac{dV_{REF}}{dT} = \left[E_{G,0K}\left(-\frac{1}{T_R}\right) + V_{BE,R,Q3}\left(\frac{1}{T_R}\right)\right] + \left[\frac{k}{q}\ \ln\left(\frac{I_{C1}}{I_{C2}}\right)\right]\left(\frac{R2}{R3}\right)\ , \tag{1.5}$$

$$E_{G,0K} = V_{BE,R,Q3} + \left[U_{T,R}\ \ln\left(\frac{I_{C1}}{I_{C2}}\right)\right]\left(\frac{R2}{R3}\right)\ , \tag{1.6}$$

As terms in Eq. (1.6) are identical to those of Eq. (1.4) except at a reference temperature, the voltage reference output should not change with temperature if it is equivalent to $E_{G,0K}$. Assuming silicon is used, then the voltage reference output of this circuit is 1.17 V (the bandgap voltage of silicon extrapolated to zero Kelvin) [5].

1.5.2 Initial Drawbacks and Solutions

The bandgap IC had limitations for low power operation. A major limitation was primarily from the basis of the output on the band gap voltage of silicon, which placed a restriction on the minimum

voltage/current at which the circuit can operate. As well, to generate a lower biasing current, a circuit that generated a constant voltage drop over a large resistance was initially used [6], as depicted in Fig. 1.5.

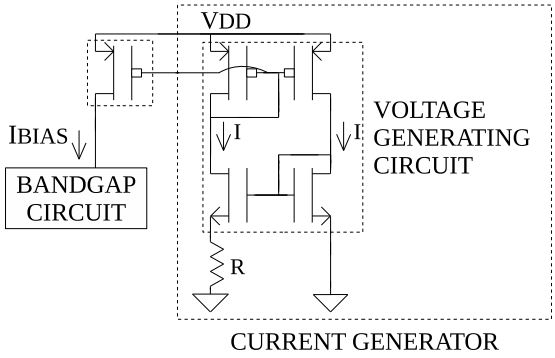


Figure 1.5: Using a resistor to generate a current

In addition, the design of typical bandgap voltage references [7,8] depended on the values of resistors. This reliance on resistors proved problematic however, as IC resistors may vary in value by ±20% or more due to random process variations, and take up a very large chip area [9].

Moreover, with the growing popularity of CMOS circuits due to simpler fabrication, lower costs, and lower power usage [10], more research had focused on low power MOSFET operation than on low power BJT operation. As a result, BJT models are not characterized as well as MOSFET models when the current is lowered to the order of nano-amperes or less. Thus, several MOSFET-based references have been proposed [11–15] . Also, several designs replaced large IC resistors with transistors biased to work at a fixed drain voltage in triode operation [13, 16–18]. as shown in Fig. 1.6.

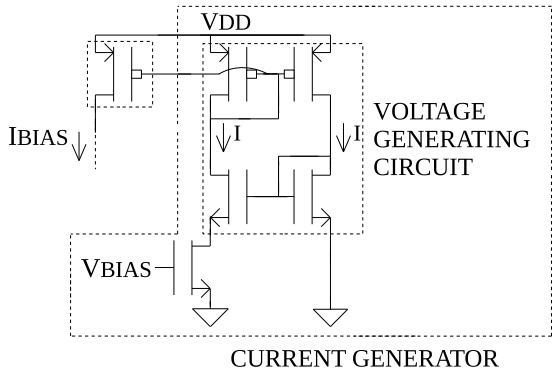


Figure 1.6: Using a triode MOSFET to generate a current

These transistors resulted in the decrease of both area consumption and susceptibility to random process variations (due to transistor matching techniques). To overcome the fundamental silicon bandgap voltage limitation, subsequent designs were based instead on different physical constants with a lower voltage, such as the threshold voltage at a given temperature [12, 13, 15].

1.6 Thesis Objectives and Overview

In this work, the objectives are to design, fabricate, and experimentally verify a voltage reference integrated circuit which meets the following criteria:

- low power
- changes minimally with temperature or supply voltage fluctuations
- small footprint

To realize these objectives, background information is presented first to provide a fundamental basis of microelectronics design using MOSFET transistors. This fundamental basis allows other recent publications to be understood and evaluated upon, which in turn helps identify the problems yet to be solved.

The engineering design work addresses these problems and begins with a planning phase, detailing the steps required to go from theoretical knowledge to physical microchips. The experimental results then demonstrate how well the microchips have performed under certain tests. An analysis of these results follow, proceeded by a summary of achievements. Finally, potential areas of future work are detailed, allowing for the advancement of this research.

Chapter 2

Background

2.1 Chapter Overview

This chapter covers all the fundamental knowledge required to understand and design MOSFET-based circuits. Standard models used in most works are reviewed with their deficiencies identified. To address these deficiencies, an alternative model, the ACM model, is detailed. The similarities between the ACM and standard models are then identified.

Current mirrors are fundamental to many voltage reference designs and cascoding helps improve circuit performance, and hence they are also explained as part of the prerequisite background knowledge. The general premise of voltage reference and ways of evaluating its performance are then detailed, putting into context the basis of the circuit and establishing a standard of good performance. Lastly, as many design decisions are based on the transistor layout design, these concepts are also covered.

2.2 MOSFET Transistors Standard Model

2.2.1 Inversion Levels

The terminals of the NMOS and PMOS transistors are represented in Fig. 2.1.

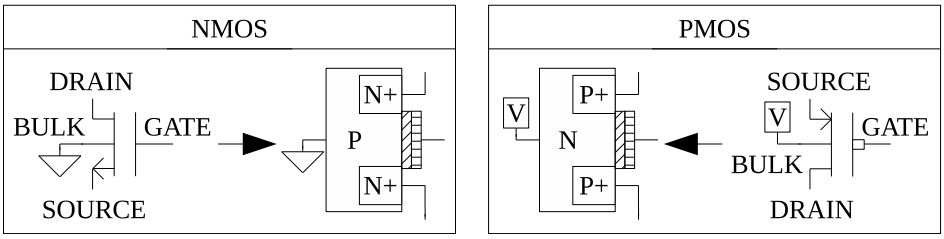


Figure 2.1: NMOS and PMOS, symbolic and physical diagrams

The inversion level of a MOSFET transistor depends on how well formed a channel of either electrons (for NMOS) or holes (for PMOS) are formed under its dielectric gate. These inversion levels are depicted in Fig. 2.2 and Fig. 2.3.

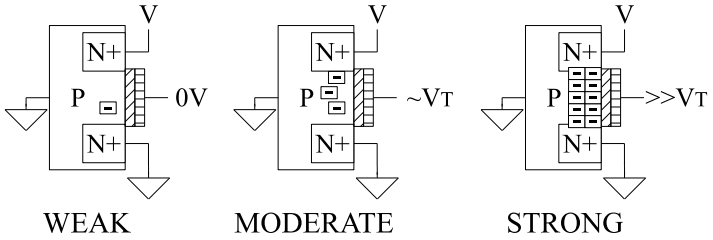


Figure 2.2: NMOS inversion levels

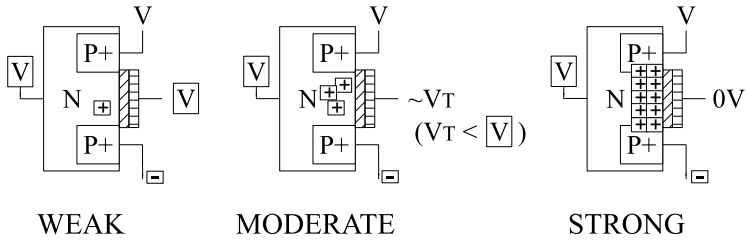


Figure 2.3: PMOS inversion levels

For an NMOS transistor in weak inversion, a given positive drain-source voltage attracts electrons to the drain at a lower rate than under strong inversion. Therefore the current is significantly lower

in weak inversion. Similarly, for a PMOS transistor in weak inversion, a negative drain-source voltage attracts holes to the drain at a lower rate than under strong inversion.

The inversion levels can be defined by the standard models as follows [9]:

- Weak: $V_{GS} - V_T \ < \ 0$
- Moderate: $V_{GS} - V_T \ \in \ [0 \ , \ 2nV_T]$
- Strong: $V_{GS} - V_T \ > \ 2nV_T$

where V_{GS} is the gate-source voltage, V_T is the threshold voltage, and n is the sub-threshold slope factor. [1]

2.2.2 Saturation and Triode

When the drain-source voltage (V_{DS}) is high enough, the drain current increases very little with further increases in V_{DS}, as shown in Fig. 2.4.

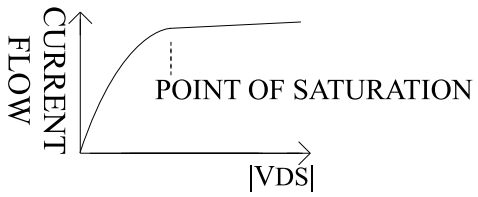


Figure 2.4: Transistor in saturation

At this level of V_{DS}, the MOS transistor is said to be operating in the saturation region. Depending on the transistor inversion level, the saturated currents are characterized as follows [21]:

$$I_{D,WEAK} = \left[S \ \left[\mu \, C_{OX} \ (n-1) \ {U_T}^2 \right] \ \exp\left(\frac{V_{GS} - V_T}{n \, U_T} \right) \right] \ \left(1 - \exp\left(-\frac{V_{DS}}{U_T} \right) \right) \ , \tag{2.1}$$

$$I_{D,STRONG} = S \ \left[\frac{1}{2} \, \mu \, C_{OX} \right] \ (V_{GS} - V_T)^2 \ (1 + \lambda V_{DS}) \ , \tag{2.2}$$

where S is the width/length ratio, μ is the carrier mobility, C_{OX} is the gate oxide capacitance, V_{DS} is the drain-source voltage, and λ is the channel length modulation coefficient. If slight increases in the current due to increased V_{DS} are neglected, then the expressions become:

$$I_{D,WEAK} = S \ \left[\mu \, C_{OX} \ (n-1) \ {U_T}^2 \right] \ \exp\left(\frac{V_{GS} - V_T}{n \, U_T} \right) \ , \tag{2.3}$$

[1] Detailed in App. B.

$$I_{D,STRONG} = S \left[\frac{1}{2}\,\mu\, C_{OX}\right] (V_{GS} - V_T)^2 \ . \tag{2.4}$$

Prior to the saturation of the transistor, triode operation is considered a point in which the drain-source voltage is low enough so that an increase of this voltage difference leads to a significant increase in current, operating in a similar manner to a resistor. In weak inversion, the current in triode operation is characterized by Eq. (2.1), where V_{DS} is approximately less than 4 U_T [22], while the strong inversion triode operation is characterized by

$$I_{D,STRONG,TRIODE} = S \left[\frac{1}{2}\,\mu\, C_{OX}\right] [2\ (V_{GS} - V_T) - V_{DS}]\ V_{DS} \ . \tag{2.5}$$

2.2.3 Motivation to use a different model

In the standard model, moderate inversion is not characterized, which leads to inherent inaccuracies as the gate voltage approaches the threshold voltage of the transistor. Simulators use empirical data to approximate moderate inversion and therefore may be subject to inaccuracies if conditions differ from the conditions that were present under the empirical testing. Hence, an alternative physics-based model that can characterize moderate inversion operation is pursued in the circuit design of this work. This alternative model is detailed in the next section.

2.3 ACM Model

This section follows [18, 23] in describing the Advanced Compact MOSFET (ACM) model [24, 25] which is a continuously differentiable model that is consistent through all operating regions of the MOS transistor. The drain current through a MOS transistor is characterized by the difference of forward and reverse currents as shown in Fig. 2.5.

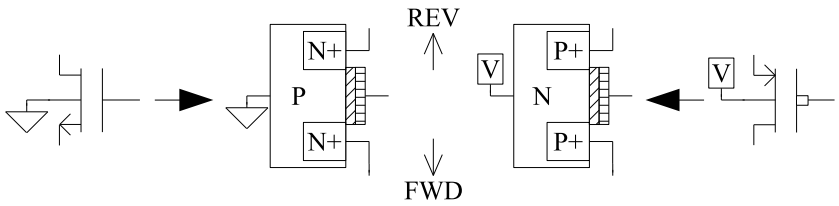


Figure 2.5: Forward and reverse currents as characterized by the ACM model

This drain current is quantified as follows:

$$I_D = S\, I_{SQ}\ (i_F - i_R)\ , \tag{2.6}$$

where i_F is the forward inversion coefficient, and i_R is the reverse inversion coefficient. I_{SQ} is the technology-dependent sheet normalization current defined as

$$I_{SQ} = \frac{1}{2}\, n\, \mu\, C_{ox}\, {U_T}^2\ . \tag{2.7}$$

I_{SQ} is not usually provided by technology data, and must be extracted using methods described in App. A. It is generally an accurate assumption for n to be 1.3 [25], but n can be calculated based on technology parameters as per App. B. The inversion coefficients are related to transistor terminal voltages by the following approximate equation:

$$\mathcal{F}(i_{F(R)}) \cong \frac{1}{U_T} \left| \frac{V_G - V_T}{n} - V_{S(D)} \right| \ , \tag{2.8}$$

where V_G , V_S and V_D are respectively the gate, source, and drain voltages referred to bulk and

$$\mathcal{F}(i_{F(R)}) = \sqrt{1 + i_{F(R)}} - 2 + \ln\left(\sqrt{1 + i_{F(R)}} - 1\right)\ , \tag{2.9}$$

If i_F is much larger than i_R , the transistor is considered to be operating in saturation and i_R can be neglected, resulting in

$$I_D \cong S\, I_{SQ}\, i_F\ . \tag{2.10}$$

In saturation, i_F determines the inversion level of the transistor. In this work, the inversion levels are defined as follows:

- Weak: $i_F \in (0,1)$
- Moderate: $i_F \in [1,100)$
- Strong: $i_F \geq 100$

The minimum drain-to-source voltage drop required for the saturated transistor to conduct its current is approximated by [24]:

$$|V_{DS,SAT}| \cong U_T\ \left(3 + \sqrt{1 + i_F}\right) \tag{2.11}$$

2.4 Similarities between ACM and Standard Model Equations

Since the ACM model is presently not widely used in MOSFET designs, proving that the ACM model can be used to derive equations similar to that of the standard models serve to justify the validity of the ACM model.

2.4.1 Weak Inversion

Approximating the ACM model for weak inversion requires assuming i_F to approach zero. Using Taylor approximations, and neglecting body effect (by assuming $V_S = 0$) results in the following expression,

$$\frac{1}{U_T}\left|\frac{V_G - V_T}{n}\right| \cong \mathcal{F}(i_F \to 0) \cong -1 + \ln\left(\frac{i_F}{2}\right) . \tag{2.12}$$

Rearranging Eq. (2.12) in terms of i_F and substituting into Eq. (2.10) yields

$$I_{D,WEAK,ACM} \cong S\ \left[\mu\, C_{OX}\, n\, {U_T}^2\right]\ \exp(1)\ \exp\left(\frac{V_G - V_T}{n\, U_T}\right) . \tag{2.13}$$

The ACM model equation varies slightly from that of the standard model, as per Eq. (2.3),

$$I_{D,WEAK,STANDARD} = S\ \left[\mu\, C_{OX}\ (n-1)\, {U_T}^2\right]\ \exp\left(\frac{V_{GS} - V_T}{n\, U_T}\right) ,$$

though it maintains the exponential increase in current with an increase in the gate voltage. A possible explanation for this difference may be a difference in how the threshold voltage was determined. Current flow exists at any gate voltage level and different models may arbitrarily define the threshold voltage depending on the current that flows for a particular gate voltage. To determine the difference in threshold voltage between the ACM and standard models, Eq. (2.13) is equated to Eq. (2.3), resulting in

$$n\, U_T\left[\ln\left(\frac{n}{n-1}\right) + 1\right] = V_{T,ACM} - V_{T,STANDARD} . \tag{2.14}$$

Another notable difference is that the threshold voltage as defined by the standard model is affected by the body effect, since a change in the source terminal affects its threshold voltage [9]. In comparison, the effect of the change in the source terminal is included in the ACM model threshold voltage [30].

2.4.2 Strong Inversion

Approximating the ACM model for strong inversion requires assuming i_F to approach infinity. Assuming $V_S = 0$ results in the following expression

$$\frac{1}{U_T}\left|\frac{V_G - V_T}{n}\right| \cong \mathcal{F}(i_F \to \infty) \cong \sqrt{i_F} . \tag{2.15}$$

Substituting this approximation into Eq. (2.10) yields

$$I_{D,STRONG,ACM} \cong S \left[\frac{1}{2}\,\mu\, C_{OX}\, n\right]\left[\frac{V_G - V_T}{n}\right]^2 . \tag{2.16}$$

Approximating n as one [2] results in

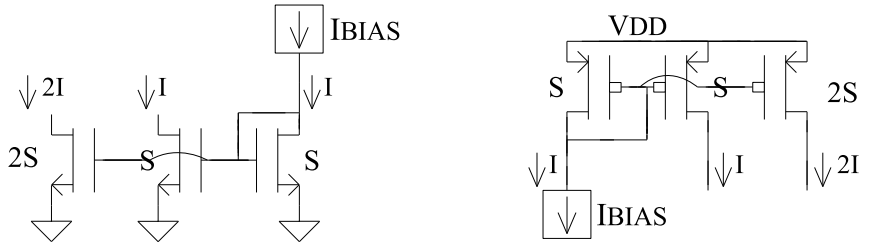
$$I_{D,STRONG,ACM} \cong S \left[\frac{1}{2}\,\mu\, C_{OX}\right](V_G - V_T)^2 , \tag{2.17}$$

which is nearly identical to the standard model equation from Eq. (2.4) (barring differences in threshold voltage as per Eq. (2.14)),

$$I_{D,STRONG,STANDARD} = S \left[\frac{1}{2}\,\mu\, C_{OX}\right](V_{GS} - V_T)^2 .$$

2.5 Current Mirrors

Current mirrors, used in most voltage reference designs, are a topology in which transistors can duplicate the current going through a diode-connected transistor (where the drain and gate share the same connection) [3] as shown in Fig. 2.6.



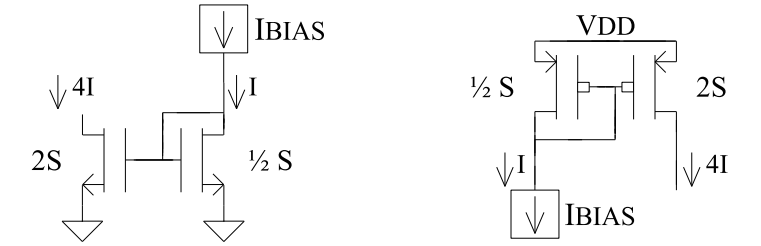
(a) NMOS current mirror (b) PMOS current mirror

Figure 2.6: NMOS and PMOS current mirrors

In a diode-connected transistor, the drain voltage is fed back to the gate. This allows the transistor to conduct any imposed external current in the same way as a forward-biased diode. Current flow in the diode-connected transistor creates a voltage potential at the drain and therefore to its gate as well. This voltage in turn can be connected to other transistor gates. The current produced is a multiple of the width/length ratio of the transistor with respect to the width/length ratio of the original diode-connected transistor, as depicted in Fig. 2.7.

[2] As strong inversion is by definition when $V_G >> V_T$, this assumption is valid as per Eq. (B.1) and Eq. (B.2) of App. B.

[3] This terminology is a relic from when BJT transistors were predominantly used; the collector of a transistor sharing a connection with its base resembles a diode.



(a) NMOS current mirror, 4x gain

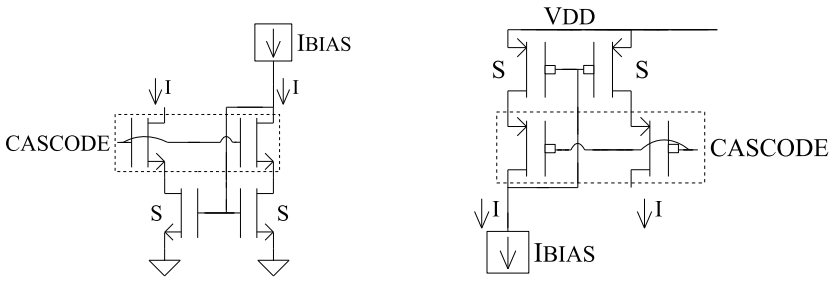
(b) PMOS current mirror, 4x gain

Figure 2.7: NMOS and PMOS current mirrors, 4x gain

The gate and source voltages are identical, hence, i_F is identical amongst the current mirror transistors as per Eq. (2.8). Therefore, changing the width/length ratio, S, results in a different saturated current, as per Eq. (2.10).

2.6 Cascoding

Due to channel length modulation, the current flow through saturated transistors increase slightly as its $|V_{DS}|$ increases. This operation is undesirable since an ideal transistor has no increase in current once it is saturated, as ideal transistors have infinite output resistance. The output resistance of non-ideal transistors can be increased by using very long transistors [26], but this increases the required chip area. A solution to this problem is the use of cascoding, defined as a cascade of common-source and common-gate stages [26], as shown in Fig. 2.8.



(a) NMOS current mirror, cascoded

(b) PMOS current mirror, cascoded

Figure 2.8: NMOS and PMOS current mirrors, cascoded

The output resistance in a cascoded circuit can be designed to be orders of magnitude higher than

the output resistance of a non-cascoded one. [4]

2.7 CMOS Voltage References Introduction

2.7.1 Fundamental Premise

Common to many MOSFET-based voltage references is a current source being generated into a diode-connected NMOS transistor as shown in Fig. 2.9.

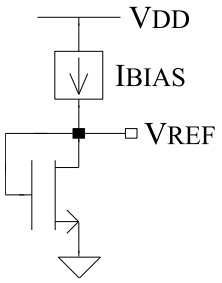


Figure 2.9: MOSFET-based voltage reference circuit common to all designs

Fulfilling the condition of $\frac{dV_{REF}}{dT} = 0$ (to remain constant with changes in temperature) requires either the current or transistor size ratio to be a certain magnitude.

2.7.2 Figures of Merit

The ideal performance of a voltage reference is shown in Fig. 2.10.

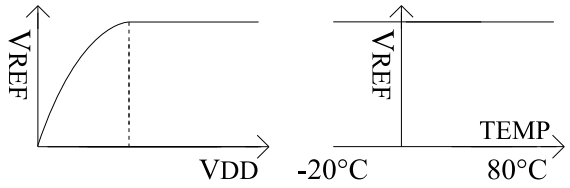
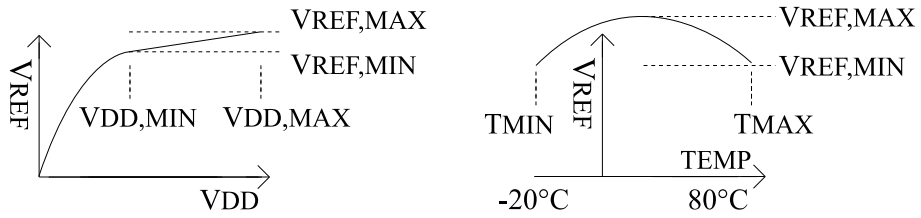


Figure 2.10: Ideal performance of a voltage reference

Actual performance differs however, as the voltage output tends to increase with the supply voltage and does not maintain the same voltage over the temperature range. These non-idealities are depicted in Fig. 2.11.

[4] Proven in App. C.



(a) Line sensitivity or PSRR evaluation (b) Temperature coefficient evaluation

Figure 2.11: Methods to evaluate voltage reference performance

Thus the performance of a voltage reference device must be evaluated by certain figures of merit. Common to most published works are the following [1]: [5]

- Line sensitivity (LS)
- Power supply rejection ratio ($PSRR$)
- Temperature Coefficient (TC)

2.7.3 Line Sensitivity

A low LS ensures that the output does not vary greatly with increases/decreases in the power supply. This low variation is important if the power supply may vary. Referencing Fig. 2.11a , LS is defined by [6]

$$LS = \frac{\left(\frac{V_{REF,MAX} - V_{REF,MIN}}{V_{DD,MAX} - V_{DD,MIN}}\right)}{V_{REF,MIN}} \times 100\% \,, \tag{2.18}$$

and is expressed in $\%/V$.

2.7.4 Power Supply Rejection Ratio

DC power supplies derived from AC sources may have ripples in their supply voltages. The $PSRR$ is the ability of the output to be unaffected by these higher-frequency ripples. The magnitude of the $PSRR$ is by definition highest when there are no frequency components (at 0 Hz) and gradually drops with increases in frequency as shown in Fig. 2.12 [1].

[5] The figures of merit amongst recently-published works are later compared in Table 5.5 of Page 93.

[6] Based on a combination of definitions from [1] and [27].

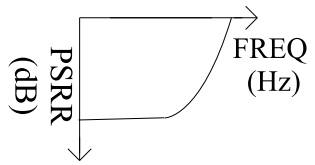


Figure 2.12: $PSRR$ of a typical voltage reference circuit

From Fig. 2.11a, the maximum $PSRR$ can be calculated by

$$PSRR = 20\log\left[\frac{\frac{V_{REF,MAX}-V_{REF,MIN}}{V_{REF,MIN}}}{\frac{V_{DD,MAX}-V_{DD,MIN}}{V_{DD,MIN}}}\right] , \tag{2.19}$$

and is expressed in Decibels (dB). If ripples exist, then the $PSRR$ of a particular frequency can be calculated by substituting Eq. 2.19 with the minimum and maximum values of the ripples that exist in the input and output.

2.7.5 Temperature Coefficient

The TC evaluates the variation in the output as temperature changes and is calculated based on Fig. 2.11b as follows:

$$TC = \frac{1}{V_{REF,NOM}}\left[\frac{V_{REF,MAX}-V_{REF,MIN}}{T_{MAX}-T_{MIN}}\right] , \tag{2.20}$$

where $V_{REF,NOM}$ is defined as

$$V_{REF,NOM} = \frac{1}{2}\left(V_{REF,MAX}+V_{REF,MIN}\right) . \tag{2.21}$$

The TC is expressed in parts-per-million-per-degree-of-temperature, ppm/K or $ppm/^{\circ}$C.

2.8 Layout

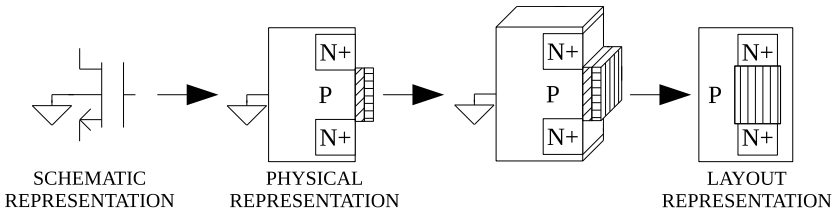


Figure 2.13: Schematic to layout

The layout is a way to represent the schematic design in a way which allows the circuit to be fabricated. During fabrication, random process variations occur which adversely affect the circuit performance. To counter these variations, the following factors [26] are considered:

- Diffusion
- Temperature gradients
- Ion implantation angle

These factors affect design decisions, and hence must be detailed prior to the start of the design.

2.8.1 Accounting for Diffusion

Due to the high temperatures involved in semiconductor fabrication (over 1000°C [9]), the N+/P+ impurities diffuse into the substrate, resulting in reduced transistor lengths as shown in Fig. 2.14.

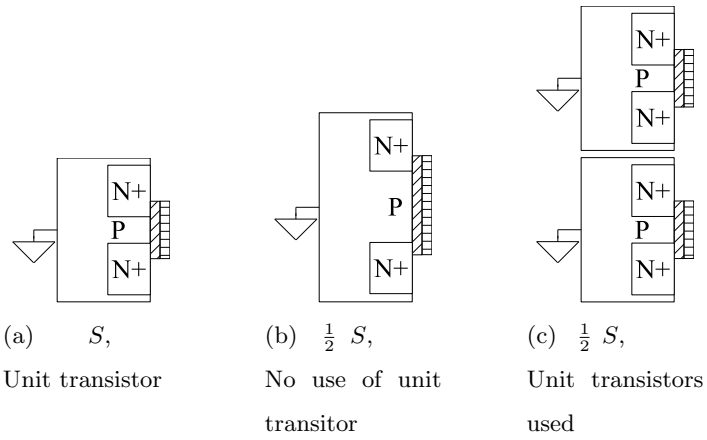


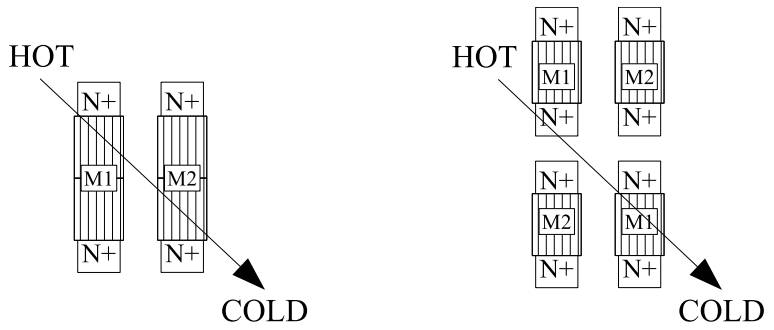
Figure 2.14: Use of unit transistors to counter the effect of diffusion

This diffusion reduces the effective transistor channel length, thereby reducing the effective threshold voltage since less energy would be required to form a channel. As the amount of diffusion is identical for all transistors, the effect of diffusion would account for a greater percentage of the total transistor length in a smaller transistor than a larger one. This difference in percentage would result in a mismatch of threshold voltage amongst two transistors of different lengths.

To avoid this mismatch, unit transistors are used, where all transistor dimensions are a multiple of a fundamental transistor width/length ratio. For instance, Fig. 2.14b and Fig. 2.14c have identical nominal transistor lengths, but the effective threshold voltage of Fig. 2.14c matches better in a circuit using the unit transistor of Fig. 2.14a.

2.8.2 Accounting for Temperature Gradients

During fabrication, there are sections of the die which increase in temperature faster than others. These gradients affect the diffusion of the impurities into the substrate and cause transistor mismatch as shown in Fig. 2.15a.



(a) Temperature gradients unaccounted for

(b) Temperature gradients accounted for

Figure 2.15: Effect of temperature gradients during fabrication

To counter the effect of temperature gradients, unit transistors from Sub-Sec. 2.8.1 can be placed for horizontal and vertical symmetry, as depicted in Fig. 2.15b. This symmetry ensures parts of the transistor resides in hot regions as well as the cold regions which balances the effects of temperature gradients.

2.8.3 Accounting for Ion Implantation Angle

Consider two dummy transistors and two design transistors as shown in Fig. 2.16 .

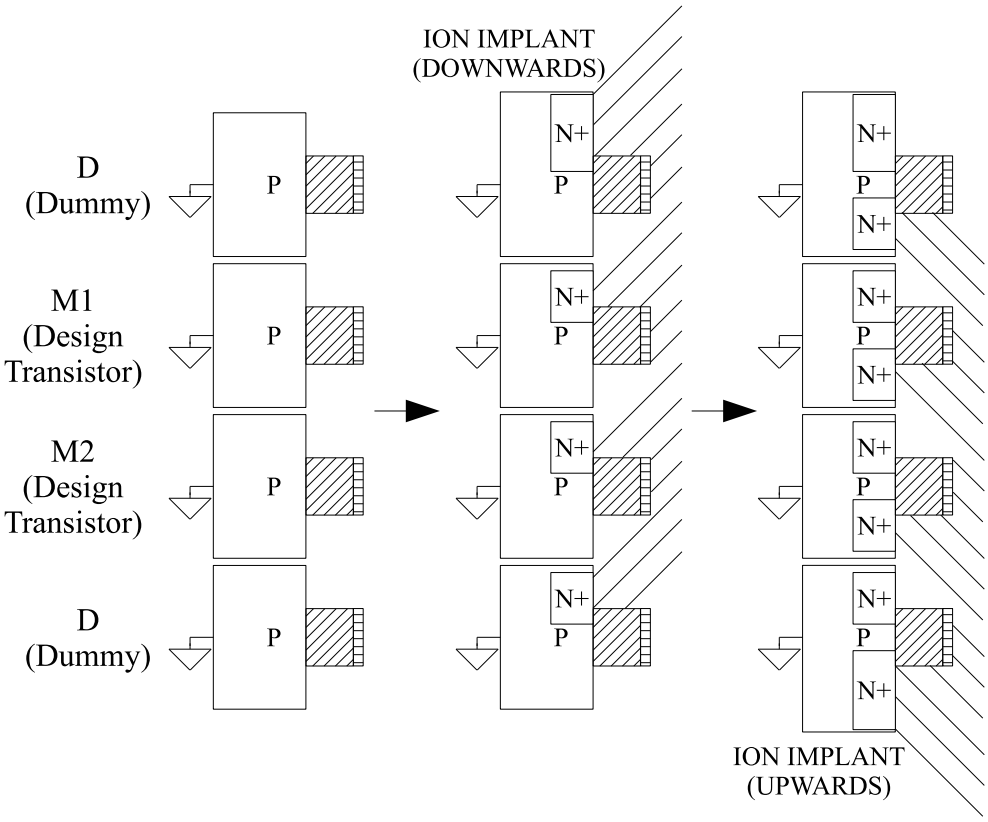


Figure 2.16: Dummy transistors to ensure uniform ion implantation, fabrication

The oxide has been exaggerated for clarity. The ion implantation during fabrication alternates at differing angles to ensure that both the drain and source are created. The dummy transistors are implanted excessively and thus have a larger drain or source terminal. However, the oxides of the dummy transistors prevent the same excess from occuring in the design transistors ensuring that their threshold voltages match. As this implantion can occur in all directions, the environments adjacent to the design transistors must be identical. This identical environment can be achieved by having dummy transistors surrounding the design transistors in all directions, as depicted in Fig. 2.17.

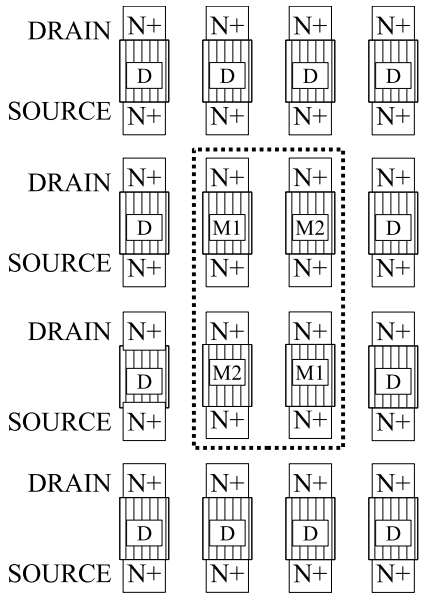


Figure 2.17: Dummy transistors to ensure uniform ion implantation, implementation

In addition, the current flowing through the transistors must always be in a single direction with metal traces routing the current to the opposite direction if required (i.e.: the source of M1 connecting to the drain of M2). To ensure no floating voltages exist, all the terminals of the dummy transistors are connected to the bulk.

Chapter 3

Literature Review

3.1 Chapter Overview

This chapter reviews recent work that have the best overall performance. Once this recent work is reviewed, current issues pertaining to this literature are identified and solutions are proposed.

3.2 De Vita et al. (2007)

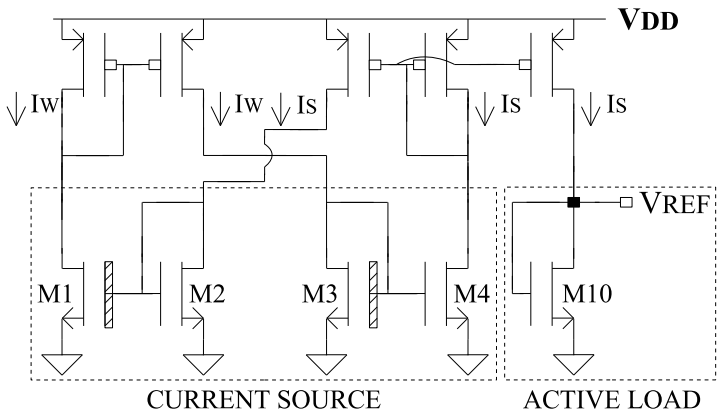


Figure 3.1: Voltage Reference circuit of de Vita et al. [12]

In the circuit of Fig. 3.1 [12], the active load transistor is operated by the current generated from a current source. The current source uses two sets of interdependent current mirrors. The branches of the current source circuit alternates with weak inversion currents ($I_{D,W}$) and strong inversion currents ($I_{D,S}$). This is accomplished by M1 and M3 having a thicker oxide, resulting in a higher threshold voltage which ensures sub-threshold operation. As per Eq. (2.1) and Eq. (2.2) these currents are defined as

$$I_{D,S} = S\ \frac{1}{2}\ \mu\ C_{OX}\ \ (V_{GS} - V_T)^2\ \ (1 + \lambda V_{DS})\ , \tag{3.1}$$

$$I_{D,W} = \left[S\ I_0 \exp\left(\frac{V_{GS} - V_T}{n\ U_T}\right) \right]\ \left(1 - \exp\left(-\frac{V_{DS}}{U_T}\right)\right)\ , \tag{3.2}$$

where

$$I_0 = \mu\ C_{OX}\ (n-1)\ {U_T}^2\ . \tag{3.3}$$

The justification for the alternating Weak and Strong inversion currents are so certain terms can cancel out, leading to an expression for the strong inversion current that flows through the load transistor M10. This expression is developed as follows.

Ignoring effects of channel length modulation (λ) and assuming $V_{DS} > 4\ U_T$, the gate-source voltages of a transistor is rearranged from Eq. (3.1) and Eq. (3.2) as

$$V_{GS,STRONG} = V_T + \sqrt{\frac{I_{D,S}}{S\ \frac{1}{2}\ \mu\ C_{OX}}}\ , \tag{3.4}$$

$$V_{GS,WEAK} = V_T + n\ U_T\ \ln\left(\frac{I_{D,W}}{S\ I_0}\right)\ . \tag{3.5}$$

From Fig. 3.1, the gate voltages of M1/M2 and M3/M4 are equivalent, resulting in

$$V_{TH,M1} + \ n\,U_T\ \ln\left[\frac{I_{D,M1}}{S_{M1}\,I_0}\right] = V_{T,M2} + \sqrt{\frac{I_{D,M2}}{S_{M2}\,\frac{1}{2}\,\mu\,C_{OX}}}\,, \tag{3.6}$$

$$V_{TH,M3} + \ n\,U_T\ \ln\left[\frac{I_{D,M3}}{S_{M3}\,I_0}\right] = V_{T,M4} + \sqrt{\frac{I_{D,M4}}{S_{M4}\,\frac{1}{2}\,\mu\,C_{OX}}}\,, \tag{3.7}$$

where V_{TH} is the threshold voltage of a thick-oxide transistor. Subtracting Eq. (3.6) from Eq. (3.7) and assuming that the weak inversion currents/threshold voltages (through M1 and M3) are equal and the strong inversion currents/threshold voltages (through M2 and M4) are also equal, the design equation for the strong-inversion current generated into the active load is as follows:

$$I_{D,ACTIVE} = \frac{1}{2}\,\mu\,C_{OX}\ \left(\frac{S_{M2}\,S_{M4}}{S_{M4} - 2\,\sqrt{S_{M4}}\,\sqrt{S_{M2}} + S_{M2}}\right)\left[n\,U_T\ \ln\left(\frac{S_{M3}}{S_{M1}}\right)\right]^2\,. \tag{3.8}$$

The voltage output is found by substituting Eq. (3.8) into Eq. (3.4) resulting in

$$V_{REF} = V_{T,M10} + \frac{1}{\sqrt{S_{M10}}}\sqrt{\frac{S_{M2}\,S_{M4}}{S_{M4} - 2\,\sqrt{S_{M4}}\,\sqrt{S_{M2}} + S_{M2}}}\left[n\,U_T\ \ln\left(\frac{S_{M3}}{S_{M1}}\right)\right]\,. \tag{3.9}$$

Evaluating the temperature dependence of Eq. (3.9) requires defining the threshold voltage in terms of temperature,

$$V_T = V_{T_{NOM}} - \left|\frac{dV_T}{dT}\right|\ (T - T_{NOM})\,, \tag{3.10}$$

where T_{NOM} is a reference temperature in which a threshold voltage is known (usually 300 K). Substituting Eq. (3.10) into Eq. (3.9) and taking its derivative with respect to temperature results in

$$\frac{dV_{REF}}{dT} = -\left|\frac{dV_{T,M10}}{dT}\right| + \frac{1}{\sqrt{S_{M10}}}\sqrt{\frac{S_{M2}\,S_{M4}}{S_{M4} - 2\,\sqrt{S_{M4}}\,\sqrt{S_{M2}} + S_{M2}}}\left[n\,\frac{k}{q}\ \ln\left(\frac{S_{M3}}{S_{M1}}\right)\right]\,. \tag{3.11}$$

Equating Eq. (3.11) to zero and rearranging obtains the Width/Length ratio of M10 neccessary for zero temperature coefficient,

$$S_{M10} = \left(\frac{1}{|dV_{T,M10}/dT|}\right)^2\ \frac{S_{M2}\,S_{M4}}{S_{M4} - 2\,\sqrt{S_{M4}}\,\sqrt{S_{M2}} + S_{M2}}\left[n\,\frac{k}{q}\ \ln\left(\frac{S_{M3}}{S_{M1}}\right)\right]^2\,. \tag{3.12}$$

Advantages and Drawbacks

The advantages of the voltage reference of De Vita et al. is its low power consumption (minimum 36 nW) while exhibiting a best-case temperature coefficient of 10 ppm/K comparable to or exceeding that of the best circuits at the time.

A drawback is that only strong and weak inversion operations have been accounted for. It had previously been shown that operating in Moderate and Weak inversion resulted in significantly less power while having a comparable die area [17]. As well, the zero temperature coefficient condition of Eq. (3.12) involves a change in threshold voltage with respect to temperature. This change is dependent on process variations and had not been accounted for in the design. As a consequence of a lack of cascoding in its topology, the line sensitivity is also poor at 0.27 $\%/V$. Another disadvantage is that this design requires thin and thick oxide NMOS transistors, resulting in extra fabrication steps and hence is more time-consuming and costly to produce.

3.3 Ueno et al. (2009)

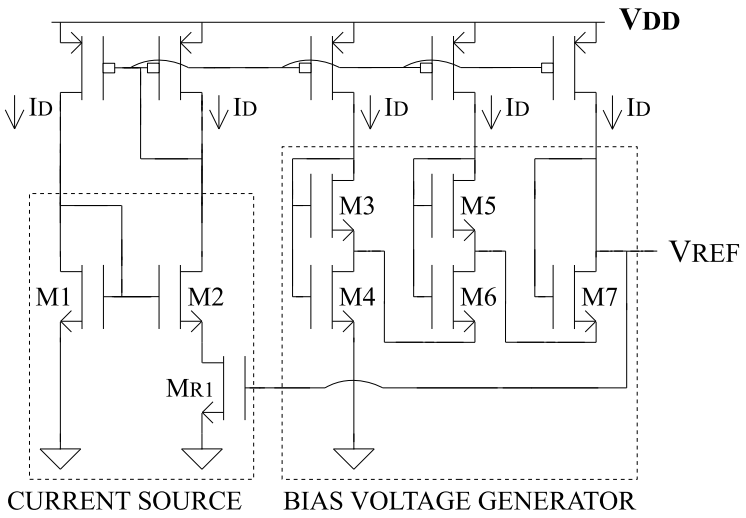


Figure 3.2: Voltage Reference circuit of Ueno et al. [13]

The voltage reference of Ueno et al. [13] is based on the principle of having the voltage output being equivalent to the threshold voltage extrapolated to zero Kelvin. With this equivalence, the temperature-dependent terms cancel out. Referencing Fig. 3.2, a summary of the design is as follows.

At the current source, the current being generated to the rest of the circuit depends on the strong-inversion triode-region MR1 transistor. The current is found by dividing the drain-source voltage of MR1, $V_{DS,MR1}$ by its equivalent resistance R_{MR1},

$$I_D = \frac{V_{DS,MR1}}{R_{MR1}} \ . \tag{3.13}$$

To find the $V_{DS,MR1}$, the weak-inversion, saturated, M1 and M2 transistors first must be related by

$$V_{GS,M1} = V_{GS,M2} + V_{DS,MR1} \ . \tag{3.14}$$

To find V_{GS}, the equation describing current flow through the transistors must be rearranged. The current of the M1 and M2 transistors is as per the well-known equation for weak inversion, Eq. (3.5). Assuming the V_{DS} is larger than $4\,U_T$, then the exponent term is considered negligible, [1] resulting in

$$I_D \cong S\ I_0 \exp\left(\frac{V_{GS} - V_T}{n\,U_T}\right) \ . \tag{3.15}$$

[1] This is verifiable from Eq. (2.1). Assuming $V_{DS} \geq 4\,U_T$, then the exponent term amounts to 0.0185 or less.

Rearranging Eq. (3.15) for V_{GS} yields

$$V_{GS} = n\, U_T\ \ln\left[\frac{I_D}{S\, I_0}\right] + V_T\ . \tag{3.16}$$

substituting Eq. (3.16) into Eq. (3.14),

$$V_{DS,R1} = n\, U_T\ \ln\left[\frac{S_{M2}}{S_{M1}}\right]\ . \tag{3.17}$$

To characterize the current from Eq. (3.13), the equivalent resistance of MR1 must then be found as follows:

$$R_{R1} = \frac{V_{DS,R1}}{I_{R1}}\ , \tag{3.18}$$

where I_{R1} is the strong-inversion triode-region current characterized by

$$I_{R1} = \frac{1}{2}\, S_{MR1}\, \mu\, C_{OX}\ \left[2\,(V_{GS,MR1} - V_T) - V_{DS,R1}\right]\ V_{DS,R1}\ . \tag{3.19}$$

Assuming that the transistor is in the deep triode region, with $(V_{GS,R1} - V_T) >> V_{DS,R1}$, the substitution of Eq. (3.19) into Eq. (3.18) yields

$$R_{R1} \cong \frac{1}{(V_{GS,MR1} - V_T)\, S_{MR1}\, \mu\, C_{OX}}\ . \tag{3.20}$$

With the gate of MR1 connected to V_{REF}, as per Fig. 3.2, substituting Eq. (3.17) and Eq. (3.20) into Eq. (3.13) allows the current to be defined with respect to all the transistors in the current source,

$$I_D = \left(n\, U_T\ \ln\left[\frac{S_{M2}}{S_{M1}}\right]\right)(V_{REF} - V_T)\, S_{MR1}\, \mu\, C_{OX}\ . \tag{3.21}$$

At the bias voltage generator, M3 to M7 are operating in weak inversion. These transistors are source-coupled translinear cells [28].

The analysis begins at the output, where

$$V_{REF} = V_{GS,M7} - V_{GS,M5} + V_{GS,M6} - V_{GS,M3} + V_{GS,M4}\ . \tag{3.22}$$

As per Eq. (3.16) , Eq. (3.21) and defining the threshold voltage with respect to temperature as follows

$$V_T = V_{T0} - \left|\frac{dV_T}{dT}\right|\ T\ , \tag{3.23}$$

where V_{T0} is the threshold voltage at zero Kelvin, Eq. (3.22) then becomes

$$V_{REF} = n\, U_T\ \ln\left[\frac{S_{M5}\, S_{M3}\, S_{MR1}}{S_{M7}\, S_{M6}\, S_{M4}}\ \frac{6}{U_T}\ \left(\frac{n}{n-1}\right)\ \left(V_{REF} - \left[V_{T0} - \left|\frac{dV_T}{dT}\right| T\right]\right)\ \ln\left(\frac{S_{M2}}{S_{M1}}\right)\right] + V_{T0} - \left|\frac{dV_T}{dT}\right| T\ . \tag{3.24}$$

Provided that $V_{REF} = V_{T0}$, and that the second-order derivative of dV_T/dT is close to zero, then taking the derivative of Eq. (3.25) with respect to temperature leads to the following approximation,

$$\frac{dV_{REF}}{dT} \cong n\,\frac{k}{q}\,\ln\left[\frac{S_{M5}\,S_{M3}\,S_{MR1}}{S_{M7}\,S_{M6}\,S_{M4}}\,\frac{6}{k/q}\,\left(\frac{n}{n-1}\right)\,\left|\frac{dV_T}{dT}\right|\,\ln\left(\frac{S_{M2}}{S_{M1}}\right)\right] - \left|\frac{dV_T}{dT}\right| . \tag{3.25}$$

In equating Eq. (3.25) to zero, and sizing the transistors accordingly, an output that has minimal change in temperature should result.

Advantages and Drawbacks

This design had been experimentally shown to have a very low change in the output with changes in the voltage supply or in other words, good line sensitivity. Accomplishing this line sensitivity involved an OP-AMP that maintained the drain voltages of the NMOS mirror which resulted in a higher power consumption at 1.4 V / 214 nA. The temperature coefficient was also very low, but process variations caused the performance to be inconsistent amongst every chip with a temperature coefficient averaging at 15 ppm/K.

As with the design by De Vita et al. [12], this design strictly used the Strong and Weak inversion levels, and did not account for the change in threshold voltage with respect to temperature due to process variations.

3.4 Seok et al. (2012)

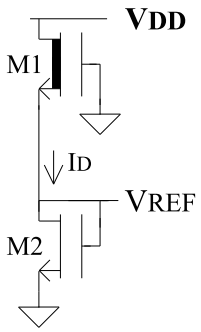


Figure 3.3: Voltage Reference circuit of Seok et al. [15]

The principle behind the circuit of Fig. 3.3 is that transistors are sized precisely to result in a condition that has minimal change with temperature. It requires two transistors in which the threshold voltages are very different from one another. This work uses a standard diode-connected NMOS transistor with current fed from a near-zero-threshold native transistor. The weak-inversion current flowing through the two transistors are as follows

$$I_D = S_{M1}\,\mu_{_{M1}}\,C_{OX,M1}\,(n_{_{M1}}-1)\,{U_T}^2\exp\left(\frac{[0-V_{REF}]-V_{T,M1}}{n_{_{M1}}\,U_T}\right)\quad, \tag{3.26}$$

$$I_D = S_{M2}\,\mu_{_{M2}}\,C_{OX,M2}\,(n_{_{M2}}-1)\,{U_T}^2\exp\left(\frac{[V_{REF}]-V_{T,M2}}{n_{_{M2}}\,U_T}\right)\quad. \tag{3.27}$$

Relating Eq. (3.26) and Eq. (3.27) and isolating the V_{REF} term results in

$$V_{REF} = \left(\frac{n_{_{M2}}\,n_{_{M1}}}{n_{_{M1}}+n_{_{M2}}}\right)\left(U_T\,\ln\left[\frac{S_{M1}\,[\mu_{_{M1}}\,C_{OX,M1}\,(n_{_{M1}}-1)]}{S_{M2}\,[\mu_{_{M2}}\,C_{OX,M2}\,(n_{_{M2}}-1)]}\right]+\frac{V_{T,M2}}{n_{_{M2}}}-\frac{V_{T,M1}}{n_{_{M1}}}\right)\,. \tag{3.28}$$

The conditions for minimal TC exists when $dV_{REF}/dT = 0$. Applying this condition to Eq. (3.29) and expressing in terms of the transistor size ratio yields

$$\frac{S_{M1}}{S_{M2}} = \frac{\mu_{_{M2}}\,C_{OX,M2}\,(n_{_{M2}}-1)}{\mu_{_{M1}}\,C_{OX,M1}\,(n_{_{M1}}-1)}\exp\left[\frac{q}{k}\left(\frac{|dV_{T,M2}/dT|}{n_{_{M2}}}-\frac{|dV_{T,M1}/dT|}{n_{_{M1}}}\right)\right]\,. \tag{3.29}$$

The work found the optimal size ratios of Eq. (3.29) based on simulation results.

To account for random process variations, trimming was used, which allowed for the width/length ratio of M1 and M2 to be varied via digital signals that turn trimming transistors on or off. This is depicted in Fig. 3.4

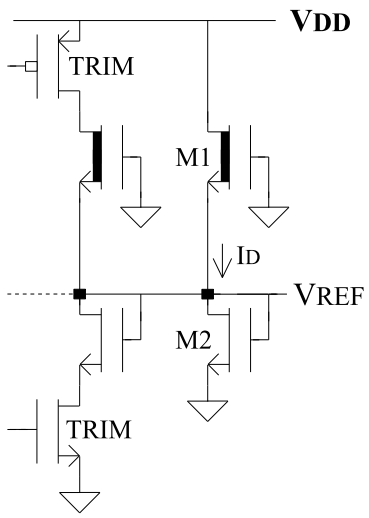


Figure 3.4: Digitally Trimmable Voltage Reference circuit of [15]

Advantages and Drawbacks

Since the design only requires two transistors, it uses two orders of magnitude less power (as low as 2.2 pW for the non-trimmable architecture and 10.85 pW for the trimmable one) than any other voltage reference circuit. In addition, the circuit requires much less die area, even when including the trimming transistors. However, in terms of its temperature coefficient, there is a big variation in performance even after the trimming was applied (5.3 ppm/K to 47.4 ppm/K). As well, its use of native transistors may not available in some CMOS processes.

3.5 Current Issues

Despite the developments described in previous work, there are still unresolved issues related to the standard transistor models and process variations.

3.5.1 Inaccurate Standard Equations

As detailed in Ch. 2, very weak inversion and very strong inversion are accurately characterized with standard equations. However, the means of predicting what happens in moderate inversion cannot be done with these equations [9]. For this reason, most designs that consider multiple inversion levels [12,13] ignore the moderate inversion operation. It was found that a combination of Weak and Moderate inversion levels resulted in the optimal power and chip usage [17]. Therefore there is a significant disadvantage to ignoring the moderate inversion operation.

3.5.2 Adverse Effects of Process Variation

Another issue with current designs is that they are still subject to random process variations. The range of performance that could be yielded in one production run of [13] is shown in Fig. 3.5.

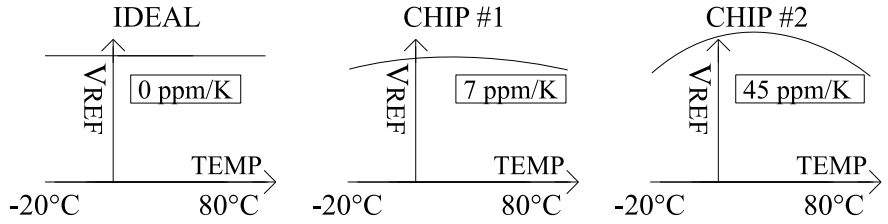


Figure 3.5: Range of circuit performance compared to the ideal case for [13]

Even though Seok et al. [15] attempted to account for these variations by implementing a means to digitally trim the transistor sizes, there was no published design consideration behind the trimming transistors on how it may affect the performance of the voltage reference.

Additionally foundries typically provide one value for the change in threshold voltage with respect to temperature and hence all designs to date assumed that this parameter was the same amongst all transistors. However, process variations are suspected to affect this dV_T/dT parameter which may explain the variation in the experimental results of other designs.

3.6 Solving Current Issues

To address the inability of standard equations to account for the moderate inversion level, an alternative MOSFET model (introduced in Sec. 2.3), the ACM model, is used. This model can accurately characterize all levels of inversion [24].

To address the effect of process variations, the voltage reference designed in this work includes the digital trimming of the circuit as proposed in [18], where it was experimentally verified that the trimming of transistor currents can result in an output voltage that changed minimally with temperature. The trimming transistors will be specifically designed with respect to its effect on the temperature performance of the voltage reference as well as accounting for possible variations in dV_T/dT. The design considerations involved in the trimming transistors will theoretically allow the circuit to work as well as it is designed for regardless of random process variations.

Chapter 4

Circuit Design

4.1 Chapter Overview

The chapter begins by detailing out the step-by-step tasks required to go from theory to the fabrication of the voltage reference microchip. The fundamental topology is then described. Based on the fundamental topology, the circuit design is then detailed one block at a time. The details include design concepts for the current source and the voltage reference itself. The layout design is then discussed. Lastly, the final design is detailed in the form of a parameter summary, schematics, layout and micrograph.

4.2 Engineering Work Flow

Prior to any work of engineering, it is useful to maintain overarching objectives as depicted in the work flow diagram of Fig. 4.1.

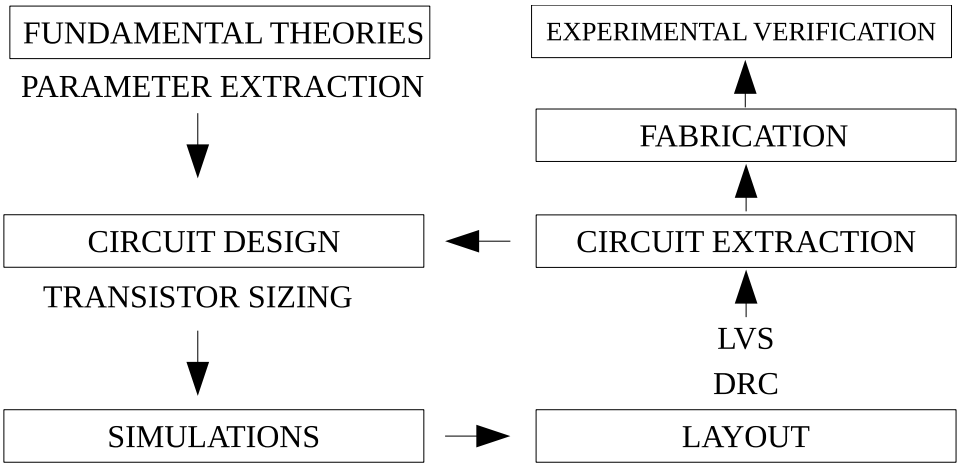


Figure 4.1: Engineering work flow diagram

Following these overarching objectives maintain the organization required to go from theoretical knowledge to a physical microchip. The steps are detailed as follows:

- Fundamental theories and parameter extraction (App. A) are investigated.
- The design involves using fundamental theories and parameters to size transistors.
- Simulations are then done to ensure that the circuit works prior to the start of the layout design.
- The layout is done in accordance to principles that ensure transistor threshold voltages match. During this process, the DRC, or Design Rule Checks, must be passed to ensure no unintended operation such as transistor latch-up [30] .
- The LVS, or Layout versus Schematic check, must also be passed to ensure that the layout and circuit schematics are identical.
- An equivalent circuit that includes parasitic components (capacitances, diodes, etc) is then extracted from the layout.
- Simulations are performed once once again with this extracted circuit.

- If the simulation of the extracted circuit differs significantly from the original simulations, then a redesign of the circuit is required, otherwise the layout can then be sent to an external organization for fabrication.

- Upon completion of fabrication, the microchips are experimentally verified and results are compared with the original simulations as well as other work.

This work involves all aspects of the work flow depicted in Fig. 4.1, except for the microchip fabrication itself.

4.3 Fundamental Topology for Achieving a Constant Voltage

The fundamental topology for the voltage reference circuit [18] is described in this section. It follows the same premise as all the other recent voltage reference designs (as per the literature review) in that a diode-connected transistor is evaluated for a condition that produces zero temperature coefficient. A constant inversion level (i_F is constant) current that biases a transistor, M1, is shown in Fig. 4.2.

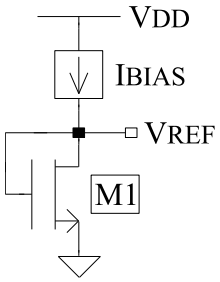


Figure 4.2: Basic voltage reference circuit

Being diode-connected, M1 operates in saturation and Eq. (2.6) can be approximated as:

$$I_{D,M1} = S_{M1}\, I_{SQ}\, i_F\ . \tag{4.1}$$

As $V_G = V_{REF}$ and $V_S = 0$, from Eq. (2.8),

$$\mathcal{F}(i_F) = \frac{V_{REF} - V_T}{n\, U_T}. \tag{4.2}$$

To relate Eq. (4.2) to temperature, the thermal voltage (U_T) and threshold voltage (V_T) needs to be re-defined. U_T can be written as

$$U_T = U_{TR}\, \frac{T}{T_R}\ , \tag{4.3}$$

where U_{TR} is the thermal voltage at an arbitrary reference temperature, T_R (usually 300 K). The threshold voltage dependence with temperature is well approximated in [31] as

$$V_T = V_{T0} - K_{VT} \frac{T}{T_R} , \tag{4.4}$$

where V_{T0} is the extrapolation of the threshold voltage at 0 K. K_{VT} is the expected drop in threshold voltage at a reference temperature and is defined as

$$K_{VT} = T_R \left| \frac{\partial V_T}{\partial T} \right| . \tag{4.5}$$

By substituting Eq. (4.3) and Eq. (4.4) into Eq. (4.2), the expression becomes:

$$V_{REF} = [\mathcal{F}(i_F)\, n\, U_{TR} - K_{VT}] \frac{T}{T_R} + V_{T0} . \tag{4.6}$$

The condition for a temperature-independent V_{REF} is obtained when dV_{REF}/dT is set to zero or equivalently, when V_{REF} is equated to V_{T0}. Thus,

$$\mathcal{F}(i_F) = \frac{K_{VT}}{n\, U_{TR}} . \tag{4.7}$$

4.4 Current Source Design

As the voltage reference circuit of Fig. 4.2 requires a constant inversion current source to function, its design will be detailed first. Consider the current source of Fig. 4.3.

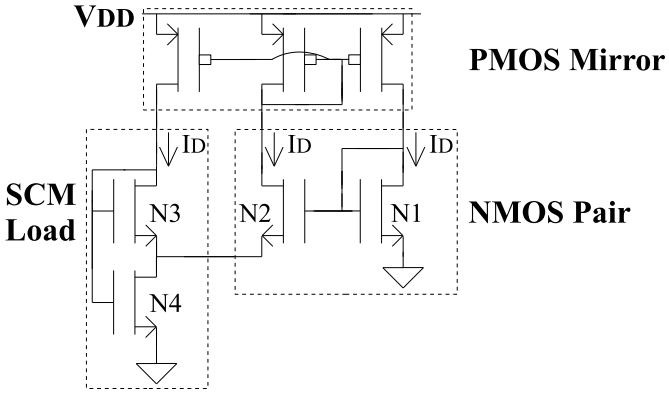


Figure 4.3: Current Source

A PMOS current mirror feeds current into an NMOS pair (N1 and N2) [26]. This creates a PTAT voltage potential at the source of N2. The PTAT potential is placed into the drain of the N4 transistor

in the SCM load.[1] This potential keeps N4 in the triode region of operation. The net result is a constant inversion level current generated in all branches of the current source. For a stable output current, it is essential that all NMOS transistors are well matched and the inversion levels in each branch are far apart [16].

4.4.1 Design Equations and Layout Considerations

As per [18, 19], it can be shown that the circuit in Fig. 4.3 must satisfy the following,

$$I_{D,N1} = I_{D,N2} = I_{D,N3} = \frac{1}{2}\, I_{D,N4} \ . \tag{4.8}$$

Assuming I_{SQ} of every transistor is equivalent,[2] then the following approximations can be made.

$$S_{N1}\, i_{F,N1} \cong S_{N2}\, i_{F,N2} \cong S_{N3}\, i_{F,N3} \cong \frac{1}{2}\, S_{N4}\ (i_{F,N4} - i_{R,N4}) \ . \tag{4.9}$$

Due to the source of N2 being shared amongst the NMOS pair and SCM load, the following expressions apply

$$\mathcal{F}(i_{F,N2}) = \mathcal{F}(i_{F,N1}) - \frac{V_{S,N2}}{U_T} \ , \tag{4.10}$$

$$\mathcal{F}(i_{F,N3}) = \mathcal{F}(i_{F,N4}) - \frac{V_{S,N2}}{U_T} \ . \tag{4.11}$$

From Eq. (4.10) and Eq. (4.11), the following is inferred,

$$\mathcal{F}(i_{F,N1}) - \mathcal{F}(i_{F,N2}) = \mathcal{F}(i_{F,N4}) - \mathcal{F}(i_{F,N3}) \ , \tag{4.12}$$

where

$$i_{F,N3} = i_{R,N4} \ . \tag{4.13}$$

Eq. (4.9) and Eq. (4.12) completely determine the inversion levels of all NMOS transistors as a function of transistor width/length ratios. Thus the circuit produces a current that has a constant inversion level, independent of technology parameters, voltage supply, and temperature [18, 19].

[1] N3 and N4 are connected in a Self-Cascode-MOSFET (SCM) configuration that results in area savings [17].

[2] From Eq. (2.7), I_{SQ} depends on n, which, from Eq. (B.1) and Eq. (B.2) of App. B , is in turn dependent on the gate voltage. I_{SQ} also depends on μ, which has dependence on the gate voltage as well [32]. The N1/N2 gate voltages are not equal to N3/N4, but the resulting I_{SQ} is considered to be similar enough due to the other terms (except n and μ) being equivalent. Hence the assumption in which all I_{SQ}'s are approximately equal can be made.

Sizing N1 and N2

The design steps begin with setting a target I_D. With this I_D, $i_{F,N1}$ and $i_{F,N2}$ are set to weak inversion for a stable output [16]. For low sensitivity to process variations, $i_{F,N1}$ and $i_{F,N2}$ must be differentiated by a large margin.[3] One possibility is for $i_{F,N1}$ to be ten times larger than $i_{F,N2}$. S_{N1} and S_{N2} can then be determined from Eq. (4.9).

Sizing N3 and N4

Again, for low sensitivity to process variations, $i_{F,N4}$ must also be as different as possible from both $i_{F,N1}$ and $i_{F,N2}$. A possibility is to have N4 operate in moderate/strong inversion, with $i_{F,N4}$ being around a hundred times larger than $i_{F,N1}$. Once $i_{F,N4}$ has been decided upon, $i_{R,N4}$ can be found using Eq. (4.9). Finally, S_{N3} and S_{N4} can be determined from Eq. (4.9) and Eq. (4.13).

Layout Considerations

As N1 and N2 operate in weak inversion, they require a much higher width/length ratio than N3 and N4, assuming they operate in moderate/strong inversion (Eq. (4.9)). For that reason N1 and N2 are usually implemented using parallel connections of a unit transistor, N3 and N4 are implemented with series connections of the same unit transistor, while a series/parallel combination allows for fractional equivalents, as depicted in Fig. 4.4.

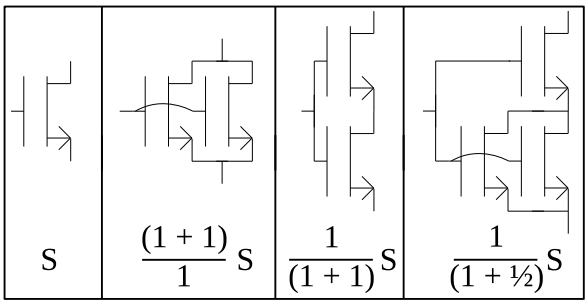


Figure 4.4: Transistor sizes using unit transistors

For a well-matching layout design (Sec. 2.8), an $S_{N2} : S_{N1}$ ratio that allows for a common-centroid layout is considered. For instance an 8:1 ratio allows for transistors to be laid out in a 3x3 manner,

[3] Assuming the process variations to be the change in the width/length ratio which affects the inversion coefficients as per Eq. (2.6), having a larger difference between $i_{F,N1}/i_{F,N2}$ as well as $i_{F,N3}/i_{F,N4}$ allows Eq. (4.12) to be approximately true.

where the N1 transistor can be in the center while the N2 transistors surround it in all directions, as shown in Fig. 4.5.

N2 N2 N2

N2 N1 N2

N2 N2 N2

Figure 4.5: Common centroid configuration

Usage

The resulting output bias current is [18]:

$$I_{BIAS} = G\,(S_{N2}\,I_{SQ}\,i_{F_{N2}})\ , \tag{4.14}$$

where G is the gain of the output branch of the PMOS current mirror. This gain is achieved with the appropriate multiple in either the PMOS or NMOS transistor width/length ratio. The output bias current is as shown in Fig. 4.6.

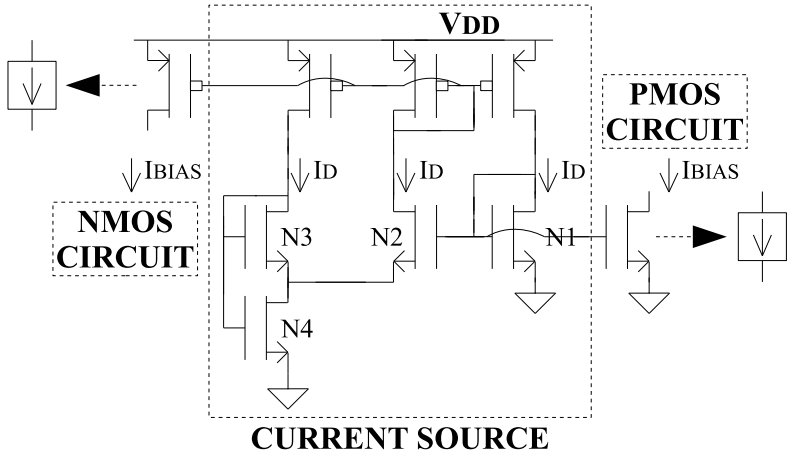


Figure 4.6: Output bias current

A PMOS transistor provides the bias current, I_{BIAS} for an NMOS circuit, while an NMOS transistor provides I_{BIAS} for a PMOS circuit. Transistor ratios along with the gain determine the inversion level (i_F) produced by this current source.

4.5 Cascode Design

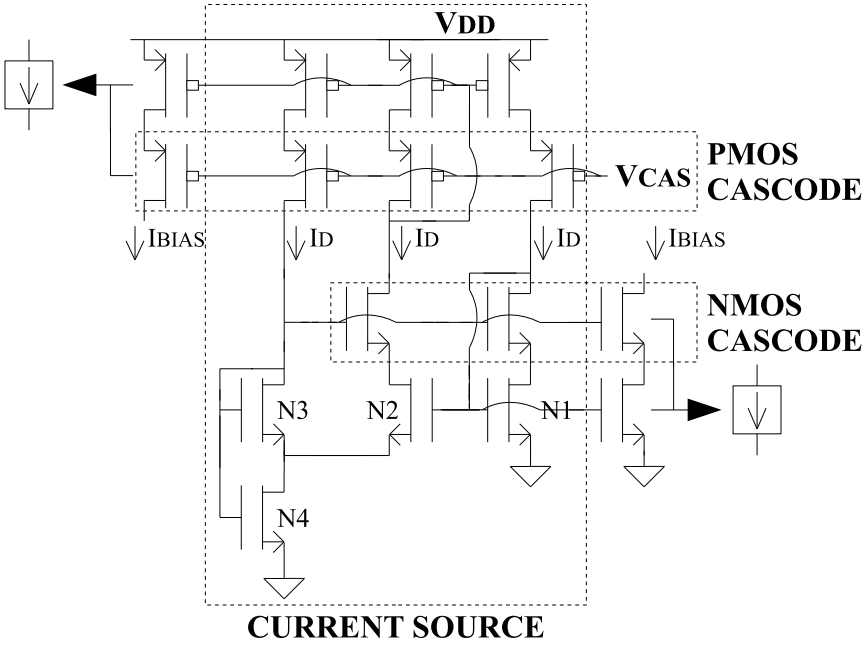


Figure 4.7: Output bias current with cascodes

Fig. 4.7 is the same as Fig. 4.6 except that a cascode configuration is used. Both the PMOS mirror and NMOS pair are cascoded for improved LS and $PSRR$. The PMOS cascode relies on a biasing circuit [33] that fixes its gate-to-bulk voltage, while the NMOS cascode has a fixed gate-to-bulk voltage available from the N3/N4 branch.[4] The cascode transistors are sized to operate in weak inversion, which minimizes the required $|V_{DS}|$ to operate in saturation as per Eq. (2.11).

[4]Note that the PMOS bulk (usually V_{DD}) is not the same as the NMOS bulk (usually Ground).

4.5.1 NMOS Cascode

The NMOS cascode portion of the current source is depicted in Fig. 4.8.

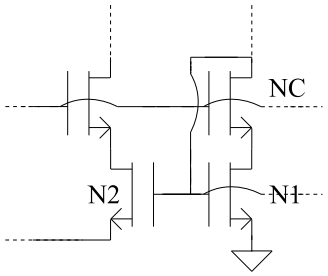


Figure 4.8: NMOS cascode

The design consideration of this portion follows the following expression,

$$V_{G,N1} \geq V_{DS,SAT,NC} + V_{DS,SAT,N1} \ , \tag{4.15}$$

where $V_{DS,SAT}$ is the minimum saturation voltage as defined by Eq. (2.11). Hence the sizing of the NMOS pair and cascodes should be for weak inversion where the i_F is close to zero. Satisfying Eq. (4.15) allows for cascoding without requiring any additional voltage to operate.

4.5.2 PMOS Cascode

The PMOS cascode portion of the current source is depicted in Fig. 4.9.

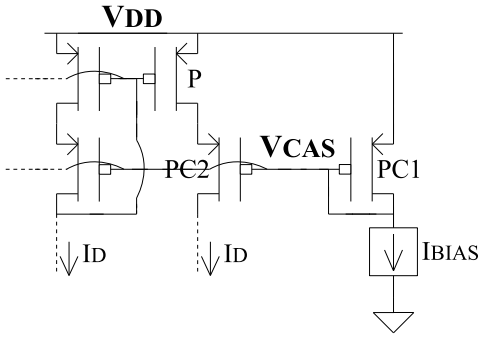


Figure 4.9: PMOS cascode

The design of the PMOS cascode circuit follows the method proposed in [33]. Like its NMOS counterpart, the PMOS cascode is also restricted to the following,

$$|V_{G,P}| \geq |V_{DS,SAT,P}| + |V_{DS,SAT,PC2}| \ . \tag{4.16}$$

Since the gate voltages of PC1 and PC2 are the same, then as per Eq. (2.8),

$$\mathcal{F}(i_{F,PC2}) = \mathcal{F}(i_{F,PC1}) - \frac{|V_{S,PC2}|}{U_T} \ . \tag{4.17}$$

The source voltage of PC2 is the drain-source voltage of P with an added safety margin voltage, V_{MARGIN} (accounting for process variations),

$$\frac{|V_{S,PC2}|}{U_T} = \frac{|V_{DS,SAT,P}| + V_{MARGIN}}{U_T} \ . \tag{4.18}$$

Defining $|V_{DS,SAT}|$ requires first defining $|V_{DS}|$ as follows,

$$\frac{|V_{DS,P}|}{U_T} = \mathcal{F}(i_{F,P}) - \mathcal{F}(i_{R,P}) = \frac{-|V_{S,P}| + |V_{D,P}|}{U_T} \ ,^{5} \tag{4.19}$$

$$\frac{|V_{DS,P}|}{U_T} = \sqrt{1+i_{F,P}} - \sqrt{1+i_{R,P}} + \ln\left[\frac{\sqrt{1+i_{F,P}}-1}{\sqrt{1+i_{R,P}}-1}\right] \ . \tag{4.20}$$

For a saturated transistor, a good approximation [33] of the term within the natural logarithm in Eq. (4.20) is

$$\left[\frac{\sqrt{1+i_{F,P}}-1}{\sqrt{1+i_{R,P}}-1}\right] \cong 100 \ . \tag{4.21}$$

Eq. (4.21) implies that $i_F >> i_R$ which is the expected case for a transistor in saturation, hence Eq. (4.20) can be further simplified as

$$\frac{|V_{DS,SAT,P}|}{U_T} \cong \sqrt{1+i_{F,P}} - 1 + \ln\left[100\right] \ . \tag{4.22}$$

Substituting Eq. (4.22) back into the original expression of Eq. (4.17) leads to

$$\mathcal{F}(i_{F,PC2}) \cong \mathcal{F}(i_{F,PC1}) - \left[\left(\sqrt{1+i_{F,P}} - 1 + \ln\left[100\right]\right) + \frac{V_{MARGIN}}{U_T}\right] \ . \tag{4.23}$$

Defining Eq. (4.23) in terms of the definition of the inversion coefficient function (Eq. (2.9)) and rearranging results in

$$\sqrt{1+i_{F,PC1}} - \sqrt{1+i_{F,PC2}} + \ln\left[\frac{\sqrt{1+i_{F,PC1}}-1}{\sqrt{1+i_{F,PC2}}-1}\right] \cong \left[\left(\sqrt{1+i_{F,P}} - 1 + \ln\left[100\right]\right) + \frac{V_{MARGIN}}{U_T}\right] \ . \tag{4.24}$$

Using Eq. (4.24) helps determine the inversion coeffient, i_F. By deciding on the current through each branch, the width/length ratio of transistors P, PC1, and PC2 can also be determined.

[5] The bulk of a PMOS is usually the highest potential in a circuit (i.e.: V_{DD}). This means the voltage potential of a PMOS terminal with respect to the bulk is a negative value. As only the voltage difference between the terminal and the bulk is of significance, absolute values are used.

A design tradeoff is that weak inversion P and PC2 results in a strong inversion PC1. It is desirable for P and PC2 to be in weak inversion since it results in a lower voltage requirement as per Eq. (4.16). A strong inversion PC1 (larger $i_{F,PC1}$) however, means a lower width/length ratio resulting in an increased transistor length. This increased length requires more die area and hence, the inversion level of PC1 is limited to the available chip space.

Another consideration is the current flowing through the biasing circuit. The gain of I_{BIAS} from Fig. 4.7 can be adjusted. The area consumption can be reduced at a cost of increased current and vice-versa.

4.6 Voltage Reference Design

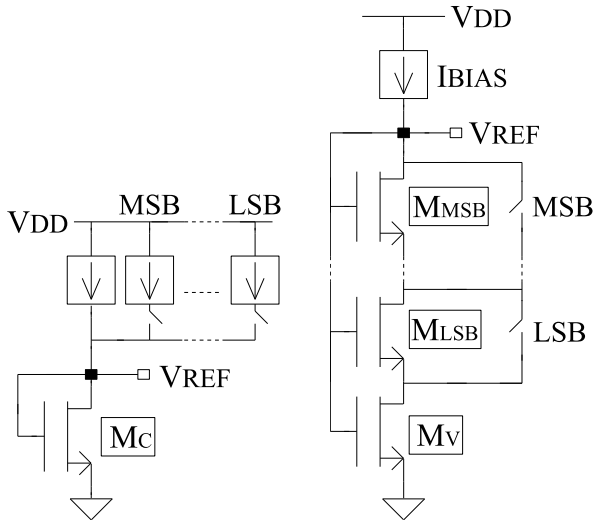
As per Eq. (4.7),

$$\mathcal{F}(i_F) = \frac{K_{VT}}{n\, U_{TR}} \ ,$$

to account for the possible change in K_{VT} due to process variations, i_F can be made to be trimmable. With reference to Fig. 4.2, this trim is accomplished from Eq. (4.1),

$$I_{D,M1} = S_{M1}\, I_{SQ}\, i_F \ ,$$

using two possible means: (i) adjust $I_{D,M1}$ or (ii) adjust S_{M1}, either of which allows for a desirable TC to be obtained. The trimming architectures are illustrated in Fig. 4.10.



(a) Constant load (b) Variable load

Figure 4.10: Proposed voltage reference architectures

Both of these architectures are explored in this work.

4.6.1 Range of i_F

The TC (temperature coefficient) of the voltage reference is re-defined as follows:

$$TC = \frac{\partial V_{REF}}{\partial T} \frac{1}{V_{REF}} = \alpha_T \frac{1}{V_{REF}} , \tag{4.25}$$

where V_{REF} is the average output voltage throughout its temperature range. From Eq. (4.6),

$$\alpha_T = \frac{\partial V_{REF}}{\partial T} = ([\mathcal{F}(i_F)]\ n\, U_{TR} - K_{VT}) \frac{1}{T_R} . \tag{4.26}$$

Neglecting the temperature dependence of n, the required range of i_F is determined by the dispersion of K_{VT} due to process variations. Since data on the dispersion of K_{VT} was not available among the technology parameters, information based on measurements from an existing device fabricated with the same technology with an additional safety margin was used to determine $i_{F,MIN}$ and $i_{F,MAX}$.

The effect of the variation of K_{VT} on Eq. (4.26) is illustrated by Fig. 4.11,

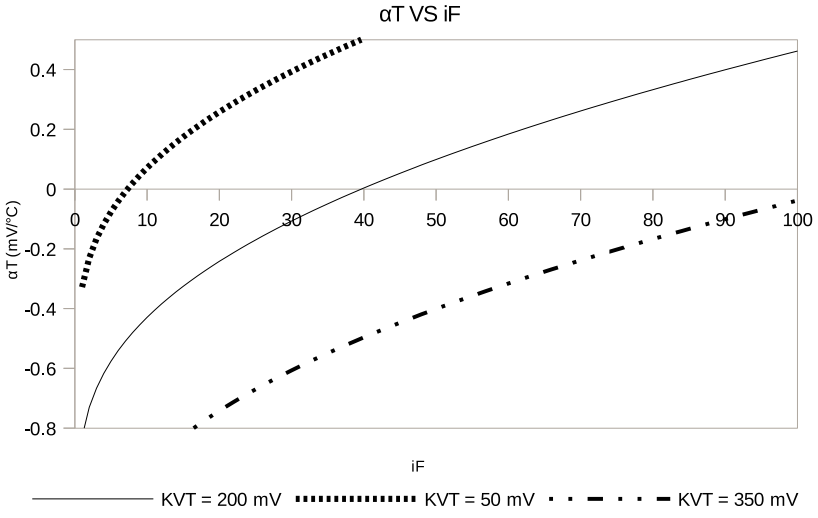


Figure 4.11: α_T, or $\frac{dV_{REF}}{dT}$, with respect to i_F

where $\mathcal{F}(i_F)$ is calculated as per Eq. (2.9) with $i_F \in [1, 100]$, $n = 1.3$, $U_{TR} = 25.4$ mV, and $T_R = 300$ K. Setting α_T to zero in Eq. (4.26) corresponds to the ideal TC of 0 ppm/°C, resulting in an identical expression to Eq. (4.7),

$$\mathcal{F}(i_F) = \frac{K_{VT}}{n\, U_{TR}} ,$$

from which i_F can be determined with each extreme of the K_{VT} range.

4.6.2 Resolution of i_F

Determining the appropriate resolution of i_F requires working out the Δi_F which maintains the α_T corresponding to the desired TC, as per Eq. (4.25). Insight on how this can be accomplished can be found in Fig. 4.11.

Too large of a Δi_F could cause an overshoot of the desired value of α_T, resulting in worse performance. To prevent this overshoot, the slope of α_T with respect to i_F is to be determined. With this slope, and the known upper and lower bounds of α_T (based on a chosen TC), an acceptable Δi_F can then be found. This premise is illustrated in Fig. 4.12. [6]

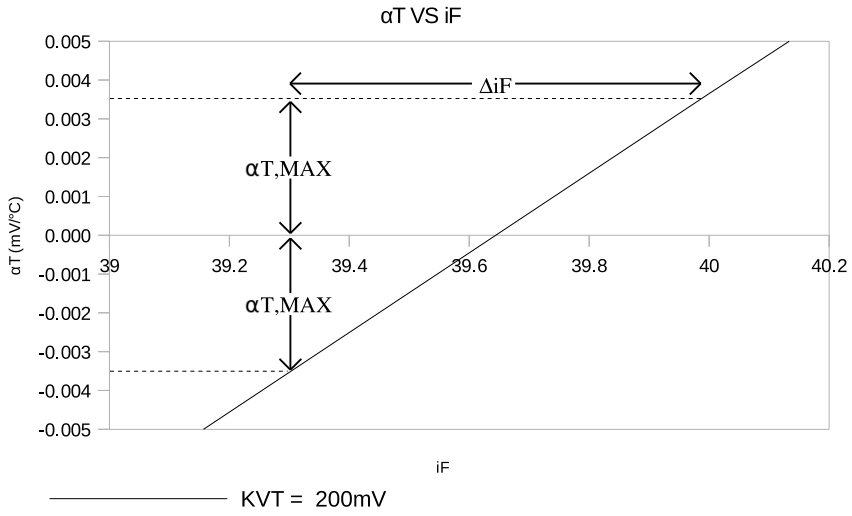


Figure 4.12: α_T, or $\frac{dV_{REF}}{dT}$, with respect to i_F, close up

To account for worst case, the maximum slope is calculated based on Eq. (4.26) as follows

$$\left.\frac{d\alpha_T}{di_F}\right|_{MAX} = \left(\left[\left.\frac{d\mathcal{F}(i_F)}{di_F}\right|_{MAX} \right] \; n\, U_{TR} \right) \; \frac{1}{T_R} \,, \tag{4.27}$$

where

$$\frac{d\mathcal{F}(i_F)}{di_F} = \frac{d\mathcal{F}(i_F)}{d\sqrt{1+i_F}} \; \frac{d\sqrt{1+i_F}}{di_F} = \frac{1}{2} \; \frac{1}{\sqrt{1+i_F}-1} \,, \tag{4.28}$$

$$\therefore \left.\frac{d\mathcal{F}(i_F)}{di_F}\right|_{MAX} = \frac{1}{2} \; \frac{1}{\sqrt{1+i_{F,MIN}}-1} \,. \tag{4.29}$$

[6] This figure uses the following example values: (1) $V_{REF} = 700$ mV, (2) $TC = 5$ ppm/°C. Using Eq. (2.20), the corresponding α_T (from Eq. (4.25)) = 0.0035 mV/°C.

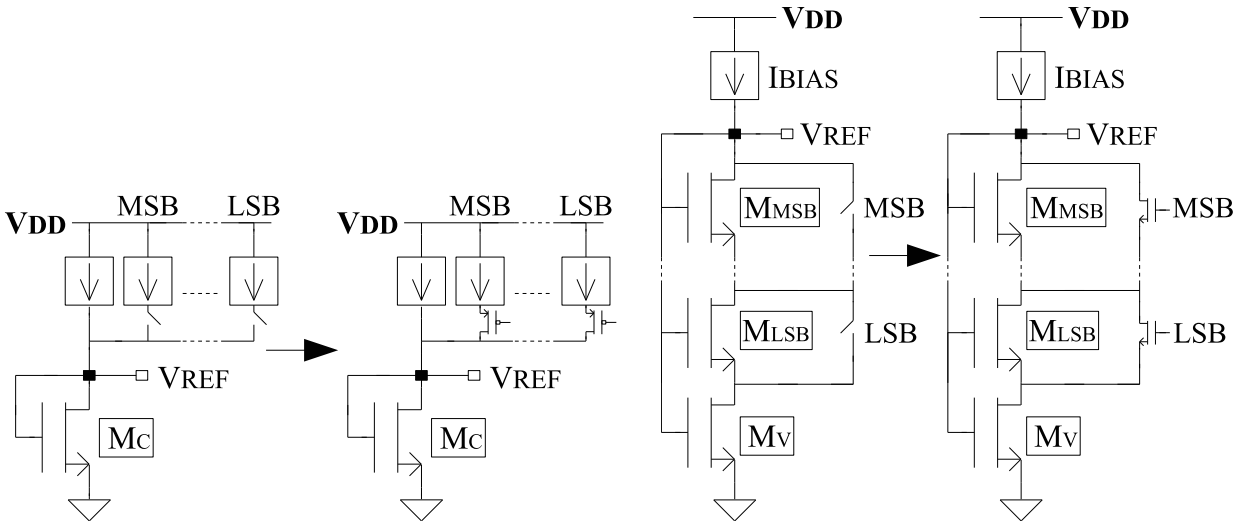
From Fig. 4.12, it is shown that Δi_F must be bound by the following to ensure the desired performance:

$$\Delta i_F \leq \frac{2\ \alpha_{T,MAX}}{\left.\frac{d\alpha_T}{di_F}\right|_{MAX}} = \frac{4\ \alpha_{T,MAX}\ (\sqrt{1+i_{F,MIN}}-1)\ T_R}{n\ U_{TR}}\ . \tag{4.30}$$

Hence, a single LSB of change in either of the voltage reference architectures should only change Δi_F as expressed in Eq. (4.30).

4.6.3 Number of Bits

The switches depicted in Fig. 4.10 are implemented with transistors. These trimming transistors are arranged in a bit-wise manner, with the least-significant-bit (LSB) producing the smallest change of i_F (Δi_F) and each successive bit increasing the change by powers of two. [7] This is shown in Fig. 4.13.



(a) Constant load (b) Variable load

Figure 4.13: Trimming transistors

With the range and resolution of i_F established, the number of bits required can now be determined as follows:

$$\frac{i_{F_{MAX}} - i_{F_{MIN}}}{\Delta i_F} \leq 2^N - 1\ , \tag{4.31}$$

where N is the number of bits as well as the number of trimming transistors required.

[7] The trimming transistors do not produce the change in i_F but the LSB to MSB transistors themselves. Hence, the trimming transistors can be powered by either V_{DD} or ground.

4.6.4 Use of Unit Transistors

For better matching in the fabrication layout, the same NMOS and PMOS unit transistor (as detailed in Sub-Sec. 2.8.1) is used throughout the loads of Fig. 4.13, as was done with the current source (Sub-Sec. 4.4.1). Exceptions to the use of unit transistors include parts of the design where process variations in dimensions and threshold voltages are non-critical, such as the NMOS trimming transistors and NMOS cascode.

4.6.5 Biasing Strategy

Minimum current consumption is limited by the errors introduced by leakage currents (I_{LEAK}) and off-state currents (I_{OFF}). I_{LEAK} occurs when the source or drain junctions have a non-zero voltage (with respect to the bulk), resulting in reverse-biased junctions that leaks some current away from the transistor. I_{OFF} occurs when the potential at the gate is the same as the bulk, resulting in a weak-inversion current as per Eq. (2.8) and Eq. (2.10). I_{OFF} must be accounted for in the trimming transistors since it means additional current flows when the transistors are "off". These off-state and leakage currents are illustrated in Fig. 4.14.

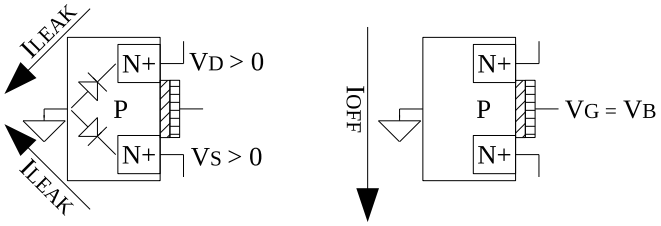


Figure 4.14: I_{LEAK} compared with I_{OFF}

To maintain the desired TC, the sum of the off-state and leakage must account for up to the minimum change of inversion level. To do so, the following condition must be fulfilled,

$$\frac{\sum I_{LEAK} + \sum I_{OFF}}{I_{MIN}} \leq \frac{\Delta i_F}{i_{F_{MIN}}} \, . \tag{4.32}$$

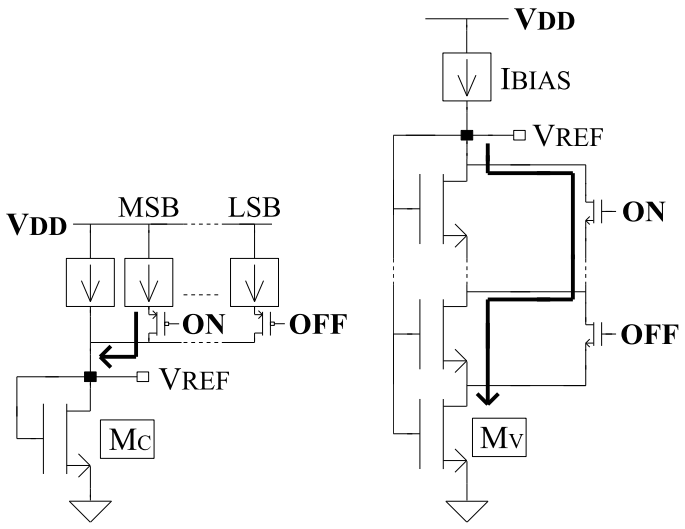
As unit transistors are used (Sub-Sec. 4.6.4), $\sum I_{LEAK}$, in turn, depends on N_T, the number of unit transistors in MV along with the all the LSB and MSB transistors of Fig. 4.13b,

$$N_T = \frac{i_{F_{MAX}}}{\Delta i_F} \, . \tag{4.33}$$

In this work, the constant load was designed to operate at currents close to that of the variable load. Hence, the number of unit transistors that MC consists of (from Fig. 4.13a) correlates closely to the I_{MIN} (for Fig. 4.13b) as determined from Eq. (4.32).

4.6.6 Trimming Transistor Design

The design of the trimming transistors account for the off-current when the transistor has been turned off and drain-source voltage drop [8] when the transistor has been turned on. Possible paths in which the current can travel are as depicted in Fig. 4.15 .



(a) Constant load (b) Variable load

Figure 4.15: Current path as a result of trimming transistors

[8] Note this drain-source voltage drop is far more significant in the variable load architecture of Fig. 4.15b than the constant load architecture of Fig. 4.15a.

Trimming transistor off

The off-current is as per the following,

$$I_{OFF} = S_{TRIM}\ I_{SQ}\ (i_F - i_R)|_{V_G=V_{BULK}}\ . \tag{4.34}$$

I_{OFF} is variable and depends on S_{TRIM}, the width/length ratio of the trimming transistor. As the majority of the current would be flowing in the load transistors when the trimming transistor is off, the drain-source voltage drop of the trimming transistors is assumed to be the same as that of the load transistors. Assuming a condition with zero TC, all the drain-source voltages of the load transistors do not change and hence $(i_F - i_R)$ of the trimming transistor remains constant.

Trimming transistor on

The current that flows when the transistors are on is the same current that is provided by the current source and hence assumed to be constant,

$$I_{ON} = I_{BIAS} = S_{TRIM}\ I_{SQ}\ (i_F - i_R)|_{V_G=ON}\ . \tag{4.35}$$

With a constant current, changes in S_{TRIM} would directly affect the value of $(i_F - i_R)$, which affects the V_{DS} drop since

$$(\mathcal{F}(i_F) - \mathcal{F}(i_R))\ \propto\ V_{DS}\ . \tag{4.36}$$

The worst-case drop in drain-source voltage is when all the trimming transistors are on. Hence the design consideration is that total drain-source voltage drop must account for less than the minimum acceptable change of the forward inversion coefficient (as per Sub-Sec. 4.6.2), characterized by

$$\frac{\sum V_{DS}}{V_{REF,NOM}} \leq \frac{\Delta i_F}{i_{F,NOM}}\ , \tag{4.37}$$

where $V_{REF,NOM}$ and $i_{F,NOM}$ is the nominal output and forward inversion coefficient, respectively, which assumes no dispersion in K_{VT} (Sub-Sec. 4.6.1).

Trimming transistor design tradeoff

The requirements in Eq. (4.34) and Eq. (4.35) conflict with each other. I_{OFF} is variable, and is minimized by reducing S_{TRIM}. The V_{DS} drop in this state of operation is identical to that across the load transistors. Conversely, I_{ON} is fixed, and V_{DS} is minimized by increasing S_{TRIM}. Hence S_{TRIM} must be sized as such that both the resulting I_{OFF} and V_{DS} drops account for less than an acceptable value of Δi_F.

4.7 Design Summary and Circuit Schematics

Table 4.1: Design values and choices

i_F Range	37–59
$TC = \alpha_T / V_{REF}$	5 ppm/°C
N	6 bits
I_D (Fig. 4.3)	0.6 nA
$I_{D,MC}$ (Fig. 4.10a)	1–1.6 nA
$I_{D,MV}$ (Fig. 4.10b)	1.2 nA
$S_{U,NMOS}$	1 μm / 24.9 μm
$S_{U,PMOS}$	2 μm / 1 μm

Table 4.1 outlines the design values and choices made for this work. S_U is the width/length ratio of the unit transistor. A top-level diagram of the circuit is shown in Fig. 4.16.

Figure 4.16: General diagram

The current is copied from the current source to the variable and constant load architectures.

4.7.1 Current Source Schematic

The complete current source is shown in Fig. 4.17. A starter circuit (see App. E) is included to ensure that a current path exists as V_{DD} ramps up from 0 V. This path ensures that the PMOS mirror as well as the PMOS cascode conducts enough current to be operational.

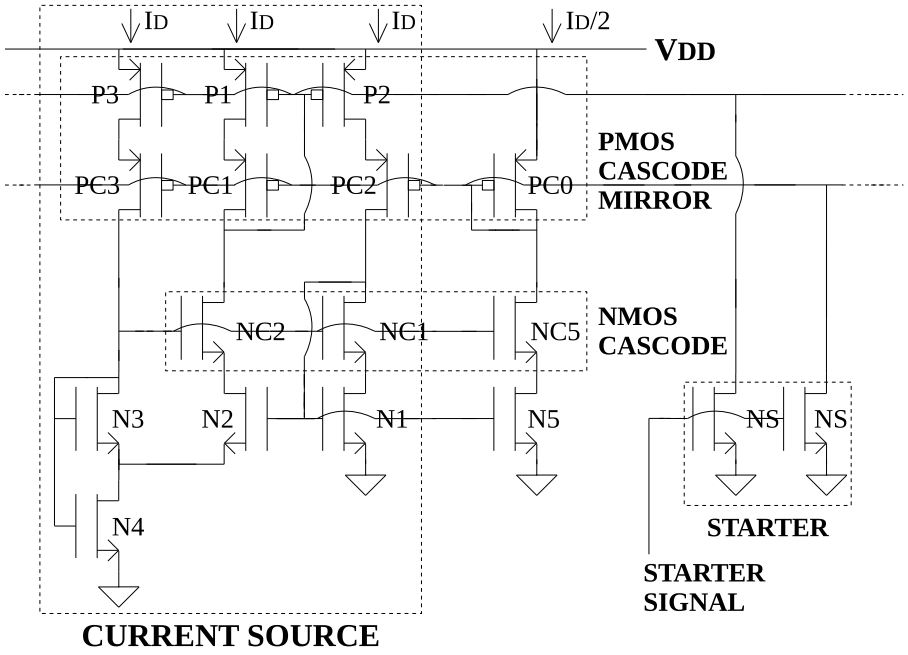


Figure 4.17: Current source schematic

Table 4.2: Transistor width/length ratios of Fig. 4.17

P1, P2, P3	4(P) - each
PC1, PC2, PC3, PC0	6(P), 6(P), 6(P), 60(S),
NC1, NC2	2 μm/2 μm - each
NC5	2 μm/2 μm - 2(S)
N1, N2, N3, N4, N5	2(P), 16(P), 20(S), 22(S), 1
NS	1

Table 4.2 shows the transistor width/length ratios for the current source. All values indicate a multiple of the PMOS or NMOS unit transistor, unless otherwise stated. (S) and (P) indicate that the unit transistors are connected in series or parallel, respectively.

4.7.2 Variable Load Schematic

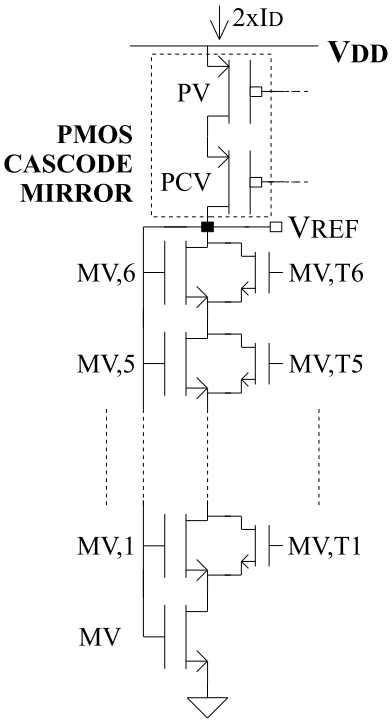


Figure 4.18: Variable load schematic

Table 4.3: Transistor width/length ratios of Fig. 4.18

PV, PCV	8(P), 12(P)
MV,6 to MV,1, MV	32(S), 16(S), 8(S), 4(S), 2(S), 1, 106(S)
MV,T6 to MC,T1	1 μm/0.45 μm - each

4.7.3 Constant Load Schematic

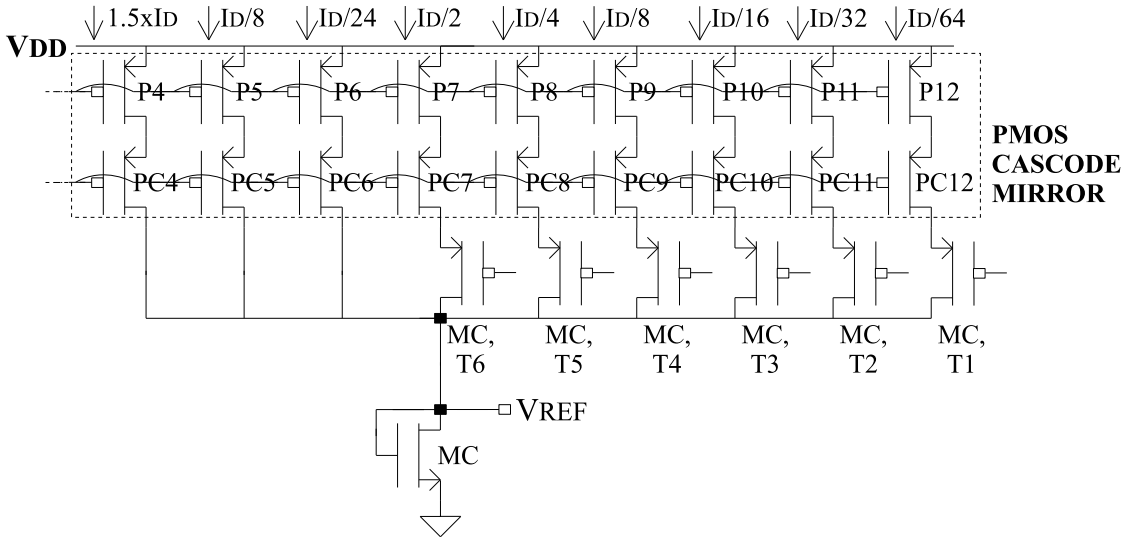


Figure 4.19: Constant load schematic

Table 4.4: Transistor width/length ratios of Fig. 4.19

P4, P5, P6	6(P), 2(S), 6(S)
PC4, PC5, PC6	9(P), 1, 1
P7, P8, P9, P10, P11, P12	2(P), 1, 2(S), 4(S), 8(S), 16(S)
PC7, PC8, PC9, PC10, PC11, PC12	3(P), 2(P), 1, 1, 1, 1
MC,T6 to MC,T1	1 - each
MC	130(S)

4.7.4 Current Output Schematic

The schematic of Fig. 4.20 allows a multiple of the current to be outputed either by sourcing ($I_{OUT,PLAIN}$) or sinking ($I_{OUT,BUFFERED}$). Their only purpose is for experimental verification.

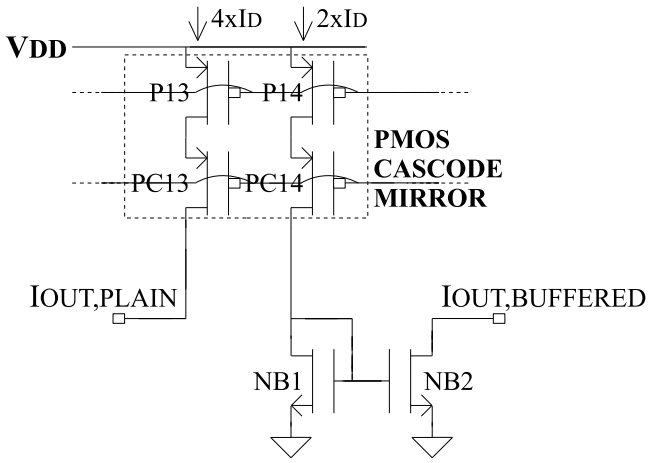


Figure 4.20: Current output schematic

Table 4.5: Transistor width/length ratios of Fig. 4.20

P13, P14	16(P), 8(P)
PC13, PC14	24(P), 12(P)
NB1, NB2	1, 214(P)

4.8 Layout Details

4.8.1 General Layout

The general layout is shown in Fig. 4.21.

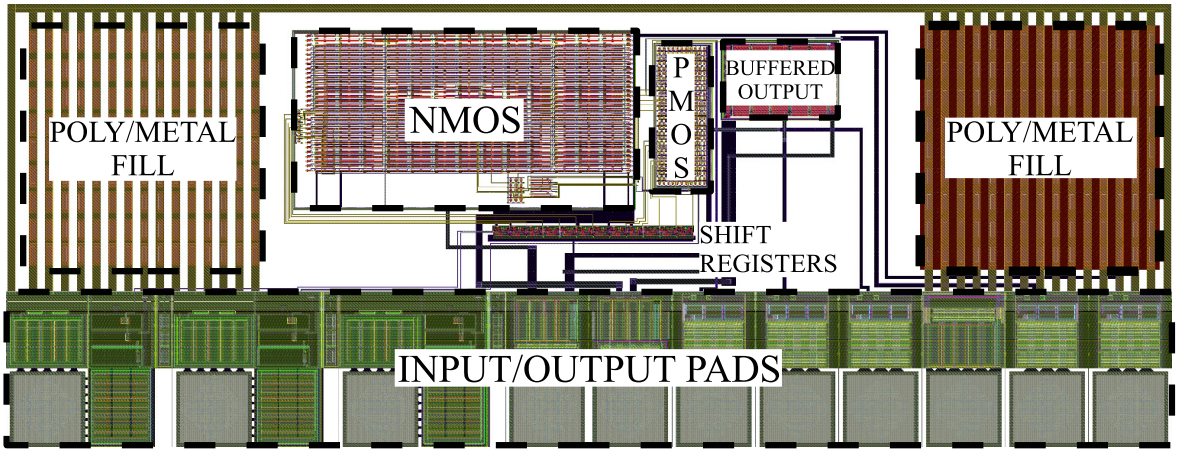


Figure 4.21: Overview of layout

The NMOS, PMOS and shift register sections are required for the circuit to function. The shift register allows the trimming transistors to be set from a single, clocked input and was a pre-made design provided by the foundry. [9] The pads serve as an interface to the pins of the chip and the Poly/Metal fill fulfills the DRC (design rule check) that requires the die to have certain percentages of poly and metal. The buffered output is used to aid the testing of the chip by multiplying the current several hundred times via an NMOS mirror.

[9] One aspect of using this pre-made shift register was that connecting the V_{DD} and GND terminals resulted in a failure of the LVS (Layout vs. Schematic) check. Hence, these connections were among the last to be made, after the LVS check had passed for all other parts of the circuit.

4.8.2 Unit Transistors

Fig. 4.22 and 4.23 show the unit transistors used throughout most of the circuit.

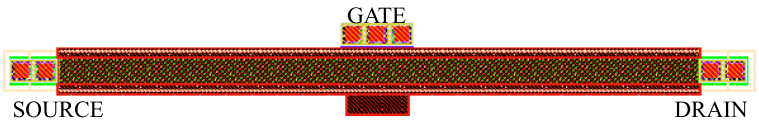


Figure 4.22: Unit Transistor, W=1 μm, L=24.9 μm, NMOS

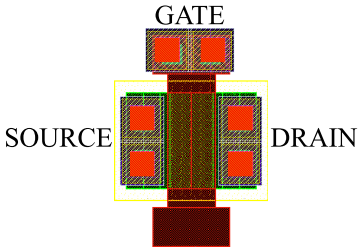


Figure 4.23: Unit Transistor, W=2 μm, L=1 μm, PMOS

To ensure identical environments, the transistors were designed to be symmetrical in both the x and y directions.

4.8.3 NMOS Section

The NMOS portion of the layout is depicted in Fig. 4.24.

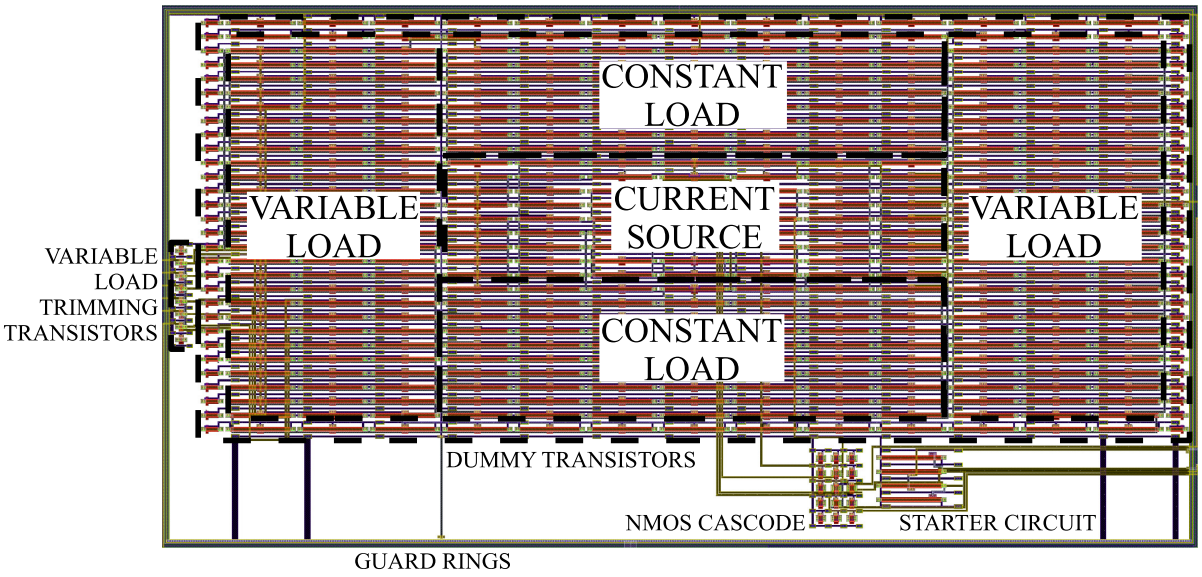


Figure 4.24: Layout: NMOS Overview

Both architectures as well as the current source are incorporated and are surrounded by dummy transistors to ensure uniform doping (Sub-Sec. 2.8.3). The dummy transistors to the far left and far right of the variable load transistors had been shortened to a length of 5 μm (compared to 24.9 μm in the unit transistor) due to concerns related to space constraints in the available die area. Guard rings surround all the transistors to prevent latch-up.

4.8.4 PMOS Section

The PMOS portion of the layout is depicted in Fig. 4.25.

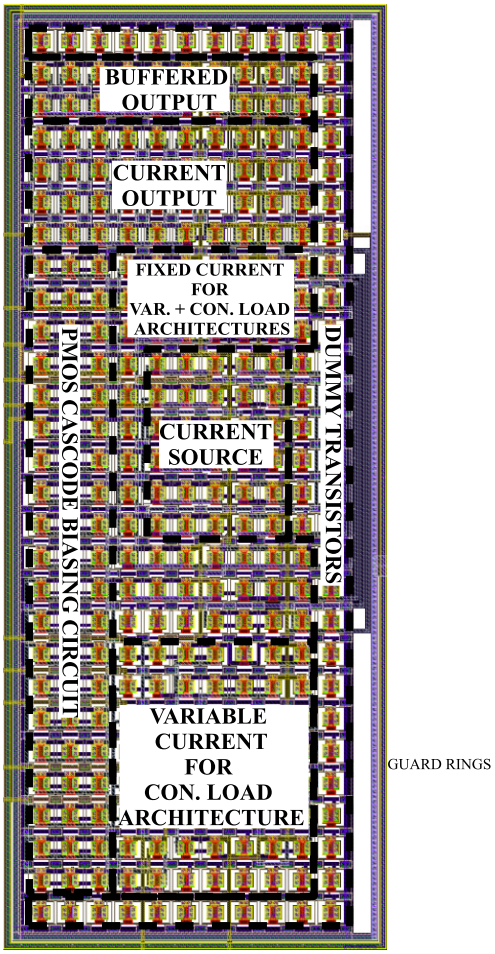


Figure 4.25: Layout: PMOS Overview

Due to space-related concerns, dummy transistors did not surround the left portion of the circuit. This portion is dominated by cascode transistors, of which precise matching is assumed to be less critical.

4.8.5 Current Source transistors in layout

The current source is fundamental to the operation of both of the voltage reference architectures, and hence, transistor matching in this circuit is more critical. The emphasis on matching is shown in Fig. 4.26 and Fig. 4.27.

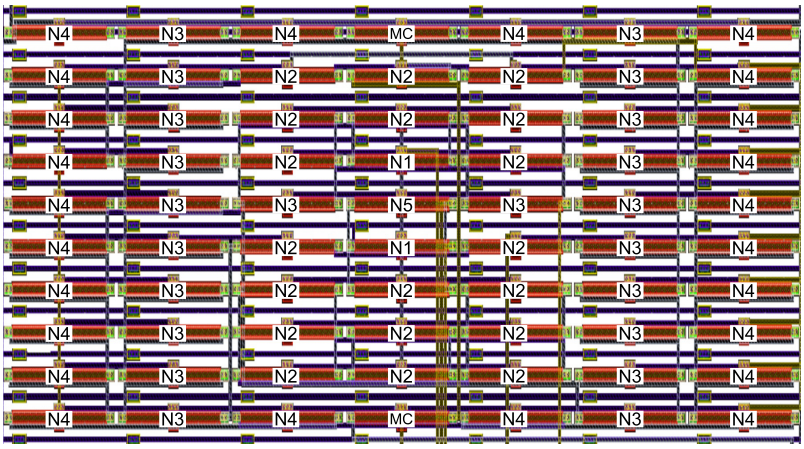


Figure 4.26: Layout: Current Source, NMOS Section

Figure 4.27: Layout: Current Source, PMOS Section

The transistors are based on the current source schematic from Fig. 4.17. Requirements regarding unit transistors, symmetry and identical environments have all been met. (Sec. 2.8).

4.9 Micrographs and dimensions

A micrograph of the die including the I/O pads is shown in Fig. 4.28 and is compared with the layout created by software shown in Fig. 4.29. Five chips were fabricated, with an area usage (excluding I/O pads) of approximately 0.048 mm^2 for the constant load architecture (Fig. 4.19) and 0.053 mm^2 for the variable load architecture (Fig. 4.18).

The dimensions of the parts are as follows:

- PMOS: 74.55 um x 184.30 um $\cong$ 0.014 mm^2
- NMOS: 433.00 um x 226.45 um $\cong$ 0.098 mm^2
- Shift Registers: 126.45 um x 15.20 um $\cong$ 0.0019 mm^2

Additionally the area consumption of the I/O pads are 1470.30 um x 200.00 um $\cong$ 0.29 mm^2.

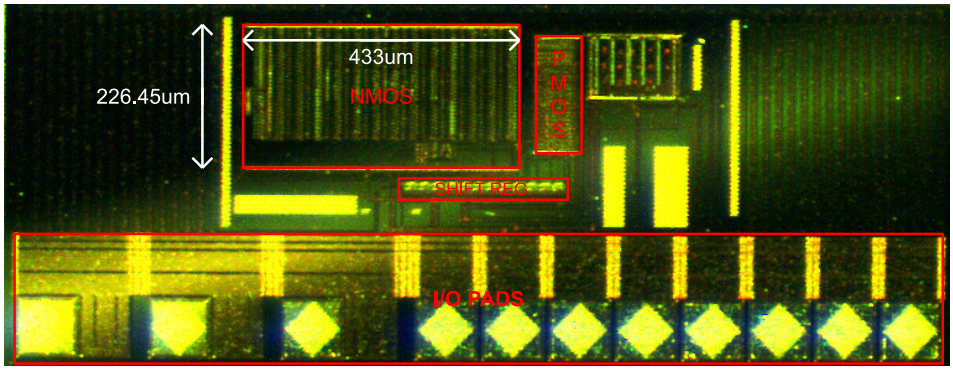


Figure 4.28: Micrograph of fabricated chip, entire die

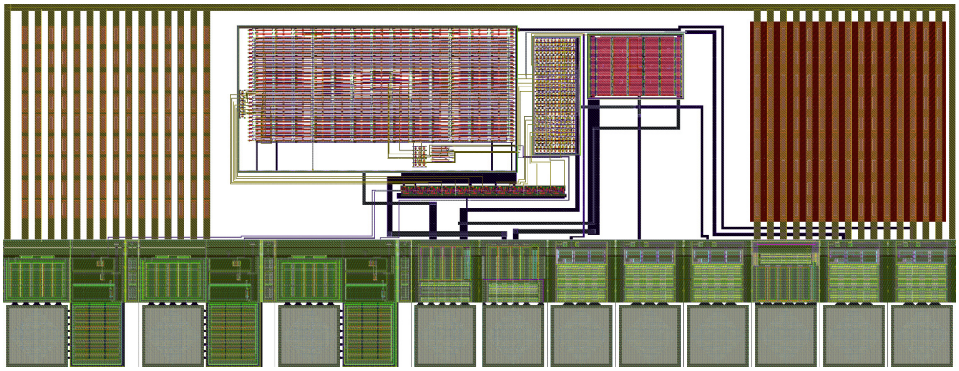


Figure 4.29: Layout of die, pre-fabrication

Chapter 5

Results and Analysis

5.1 Chapter Overview

In this chapter the experimental methodology is presented first to aid in the reproduction of the results of this work. Also presented are issues encountered during the experiments, which identifies limitations and areas of potential improvements.

The simulations and experimental results pertaining to current output, changes in supply voltage and changes in temperature are then detailed, followed by statistical analysis of all results. The results are analyzed qualitatively, with explanations for any anomaly that had occurred along with any new information that can be inferred from the results.

Finally, the results are compared with other recent publications to put into proper context of whether or not the circuit had performed well.

5.2 Experimental Verification Methodology

This section details the methods used to experimentally verify the operation of the voltage reference microchip.

5.2.1 Scope of testing

Both 6-bit architectures were tested to confirm operation. The trimming transistor bits were activated via an on-chip shift register, as shown in Fig. 5.1.

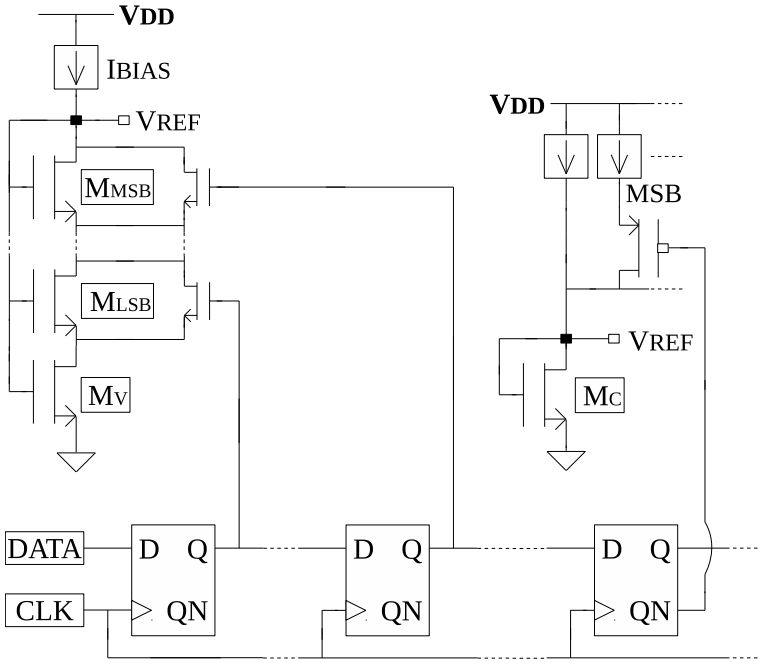


Figure 5.1: Shift register connections

It was expected that all bits being turned on would result in an increasing-with-temperature output, while all bits being turned off will result in a decreasing-with-temperature one. Therefore at an intermediate point, there exists a combination of bits that results in the minimum TC.

5.2.2 Electrostatic Discharge Protection

The primary consideration prior to any experimental verification is ESD protection since exposed pins may be subjected to 3 kV as per the Human Body Model [34]. Proper grounding of specialized mats and wrist straps that dissipate any build-up of charges as well as the avoidance of working on carpeted areas help reduce the risk of ESD.

5.2.3 Voltage Output Measurement

The current through the voltage reference is in the order of nA's, hence, conventional measuring devices cannot directly measure the output. The issue with low currents can be resolved with the use of a CMOS-based operational amplifier (OP-AMP) connected as a voltage follower as shown in Fig. 5.2.

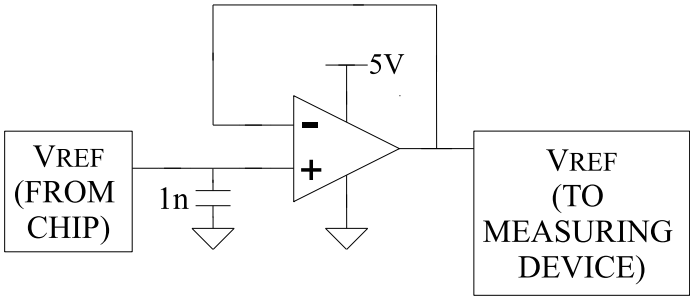


Figure 5.2: Op-amp buffer

The output remains steady if both terminals are approximately the same. In this work, the OP-AMP used was an ALD1722. This particular OP-AMP has an input bias current of 0.01 pA and a nominal voltage offset of 25 μV with a drift of 4 μV/°C. However, the test conditions differed from those in the OP-AMP datasheet, and hence, the voltage offset between the positive terminal and output was determined at all measured temperatures prior to measurement of the voltage reference.

5.2.4 Current Output Measurement

The same OP-AMP buffers from Sub-Sec. 5.2.3 can be used to measure the current of the voltage reference chip. As the current could not be measured directly with conventional equipment, the voltage drop over a large resistance was output to the OP-AMP buffer and then measured. The voltage and resistance was then be used to work out the current. Two approaches can be taken to achieve this voltage drop and are shown in Fig. 5.3 and Fig. 5.4.

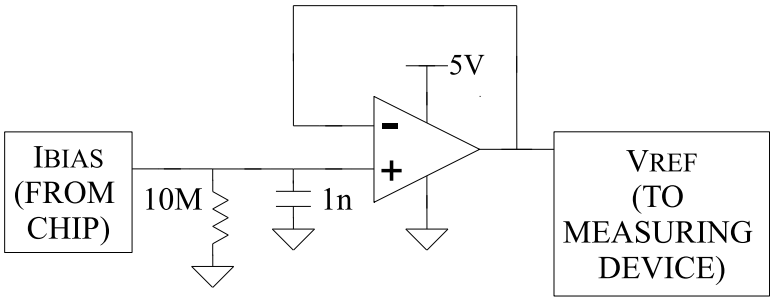


Figure 5.3: Current measurement, sourcing

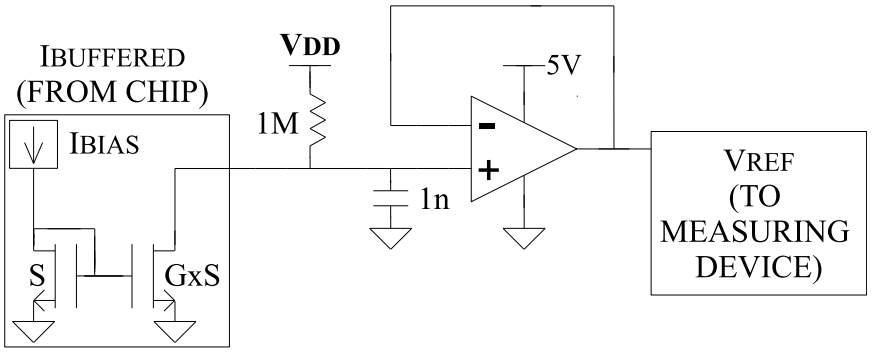


Figure 5.4: Current measurement, sinking

5.2.5 Bit Setting

The clock signal was set manually by connecting to the supply voltage and then to ground to dissipate excess charge. A Schmidt trigger was used to ensure no signals turn on and off momentarily (i.e.: the mechanical bounce of a contact between two metal connections) upon being set, as shown in Fig. 5.5.

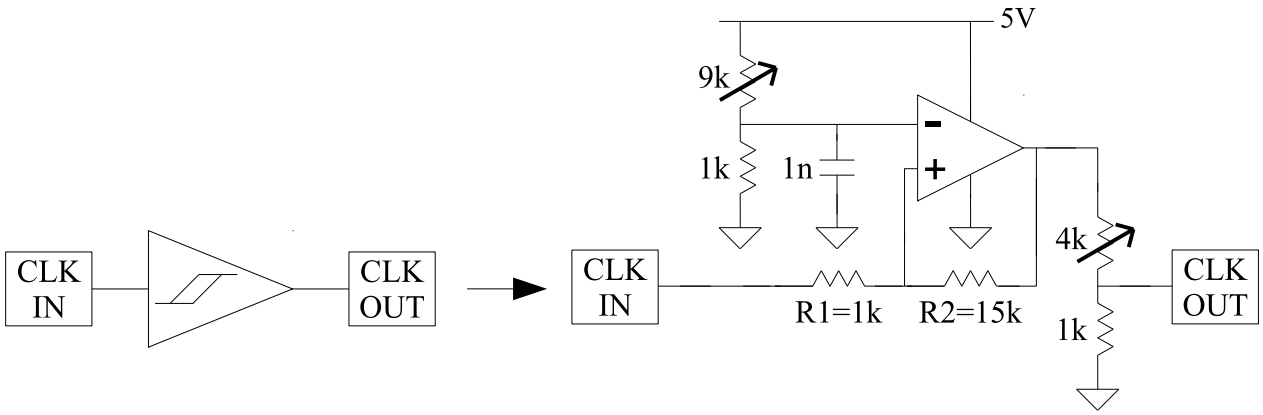


Figure 5.5: Schmidt trigger

To verify that the bits had been set correctly, a single bit was transferred through the shift register of every chip. Doing so confirmed the bit operation as each activated bit resulted in an increased output that corresponded to its bit value as depicted in Table 5.1.

Table 5.1: Effect of activating one bit at a time

Bit Active	NONE	LSB	LSB+1	LSB+2	...	MSB
Output	V_{MIN}	V_{MIN} + 1 mV	V_{MIN} + 2 mV	V_{MIN} + 4 mV	...	V_{MIN} + x mV

5.2.6 Temperature Change

The total temperature range tested was $-20\,^{\circ}\text{C}$ to $80\,^{\circ}\text{C}$. Due to limitations in available equipment, only the following ranges were measured:

- $-20\,^{\circ}\text{C}$
- $10\,^{\circ}\text{C}$ to $20\,^{\circ}\text{C}$
- $30\,^{\circ}\text{C}$ to $80\,^{\circ}\text{C}$

The lower ranges from $-20\,^{\circ}\mathrm{C}$ to $20\,^{\circ}\mathrm{C}$ were measured using temperature-controlled laboratory refrigerators, while the higher ranges from $30\,^{\circ}\mathrm{C}$ to $80\,^{\circ}\mathrm{C}$ were accomplished with an incubator oven.

5.2.7 Issues Encountered, Noise

It was observed that the voltage reference output signal may continuously vary by up to +/- 0.2 mV. Factors that contribute to this variation may be noise due a combination of the following:

- Wire length
- Consistency of temperature

The wires that connect to the voltage reference microchip were affected by electromagnetic interference. An excess of physical vibrations also caused interference to be induced. Minimizing wire length or movement (ideally a PCB board) reduces this interference.

It was also observed that the consistency of temperature affected the output. For instance if there is a great distance between the heating/cooling elements and the microchip, the output will fluctuate by a large margin. Hence, for testing purposes, the voltage reference chip should be in an enclosed environment and close to the heating/cooling elements.

5.2.8 Issues Encountered, Unintended Bit Shifts

Another issue that was frequently observed was that the bit settings changed upon a sudden large change in temperature. This occurence was likely due to how the shift register maintained the voltage[1] to the trimming transistor gates and can be an avenue of future research.

[1] 0 V (logic low) or V_{DD} (logic high).

5.3 Simulations, Experimental Results, and Analysis

In this section, the general order of presentation is as follows: (i) simulations, (ii) experimental results, (iii) analysis. A summary of the experimental results can be found in Sub-Sect. 5.3.6 followed by a statistical analysis presented in Sub-Sect. 5.3.7.

5.3.1 Current Output

Shown in Fig. 5.6 are the simulated and measured current outputs.

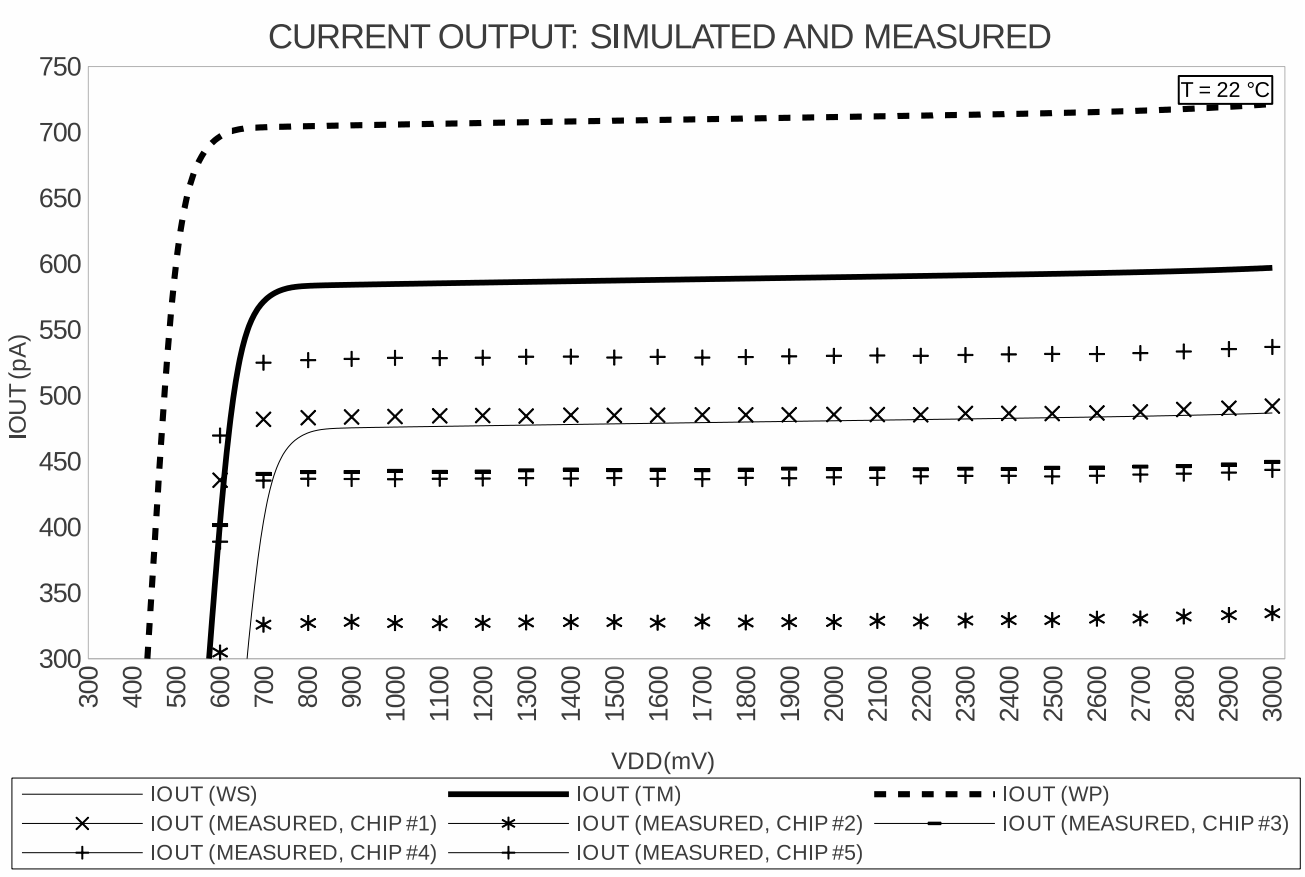


Figure 5.6: I_{OUT} simulated and measured

All simulations are evaluated under the process corners, Worst Power (WP) and Worst Speed (WS) to determine the maximum expected dispersion in the output. The unexpected result of three of the chips (Chip #2, 3 and 4) outputing less than that of the simulated current is investigated further in the proceeding paragraphs.

From simulations, the current flow for the extreme corners corresponded to $I_{BIAS}\,\epsilon$ (475 pA , 705 pA)

with a typical (most statistically likely) value of 584 pA. However, some of the measured values fell below the lowest extreme range, measuring from 334 pA to 450 pA. The result of less current results in a lower inversion level coefficient, i_F, as per Eq. (2.10),

$$I_D = S\, I_{SQ}\, i_F\ ,$$

assuming I_{SQ} is the typical value.

A possible explanation behind the low output current may lie in the experimental setup, since there may have been losses from the resistor being used in the measurement (Sub-Sec. 5.2.4). However, a specialized meter capable of measuring the current directly was subsequently used and identical anomalous results were obtained.

Alternatively, the low current may have been caused by transistor mismatch in the PMOS section since dummy transistors did not surround the side dominated by cascode transistors, in which matching was thought to not have been critical (Sub-Sec. 4.8.4).

Regardless of the reason behind the low current output, it is worth noting that the voltage outputs (Fig. 5.7 to 5.11) fell well within the extremes. These voltage outputs resembled the output of a circuit running at currents close to the typical value.

5.3.2 Evaluation of LS and $PSRR$

Fig. 5.7 to 5.11 show the output of the voltage reference as the supply voltage is varied up to 3 V. Turning all the bits off or on resulted in the minimum or maximum outputs, all of which were well within the worst-corner limits established by simulation. The average LS and $PSRR$ (at 0 Hz) were 0.14 %/V and -57.76 dB respectively for the variable load architecture, and 0.13 %/V and -58.44 dB respectively for the constant load architecture. These figures of merit were calculated from Eq. (2.18) and (2.19) with V_{DD} ranging from 0.9 V to 3.0 V.

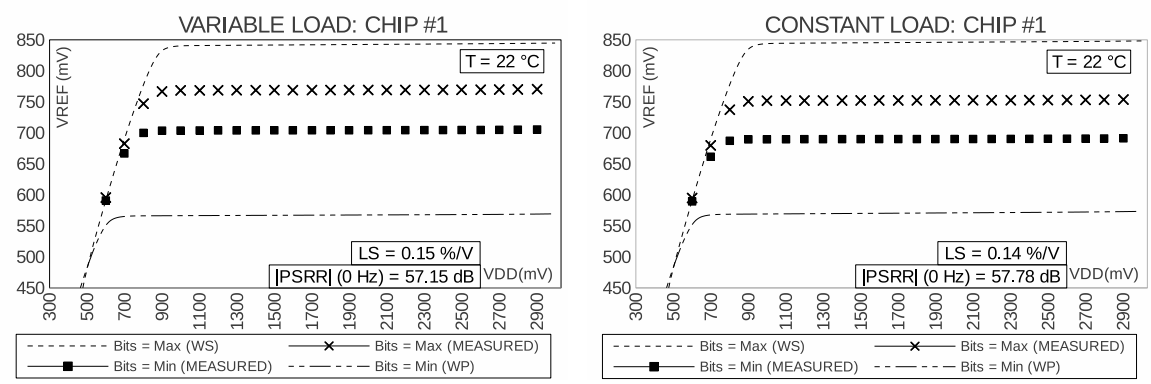


Figure 5.7: V_{REF} output vs supply voltage, Chip #1

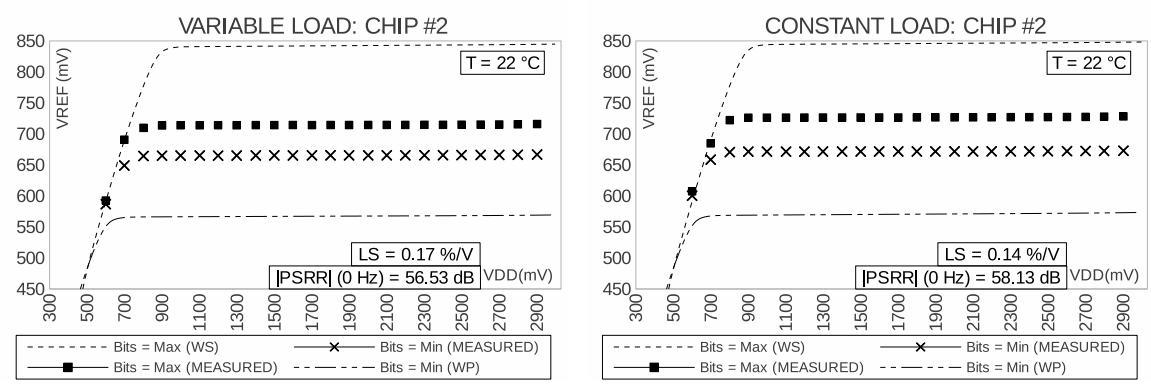


Figure 5.8: V_{REF} output vs supply voltage, Chip #2

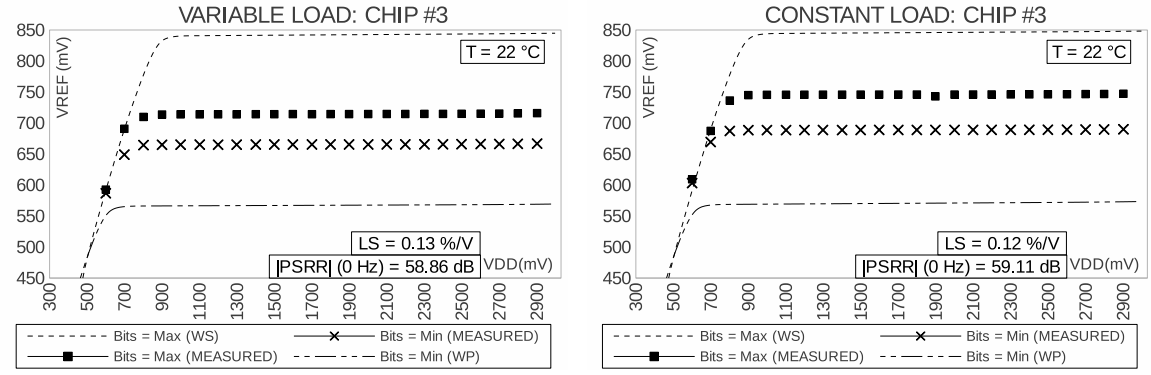


Figure 5.9: V_{REF} output vs supply voltage, Chip #3

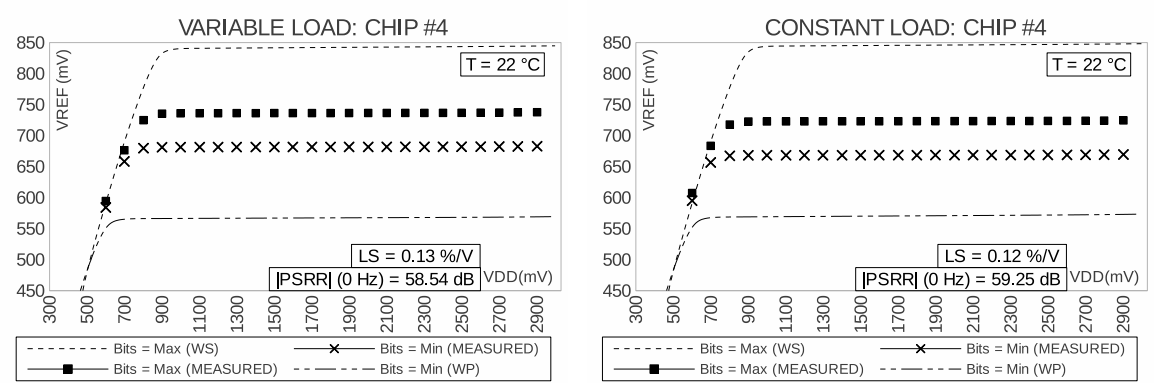


Figure 5.10: V_{REF} output vs supply voltage, Chip #4

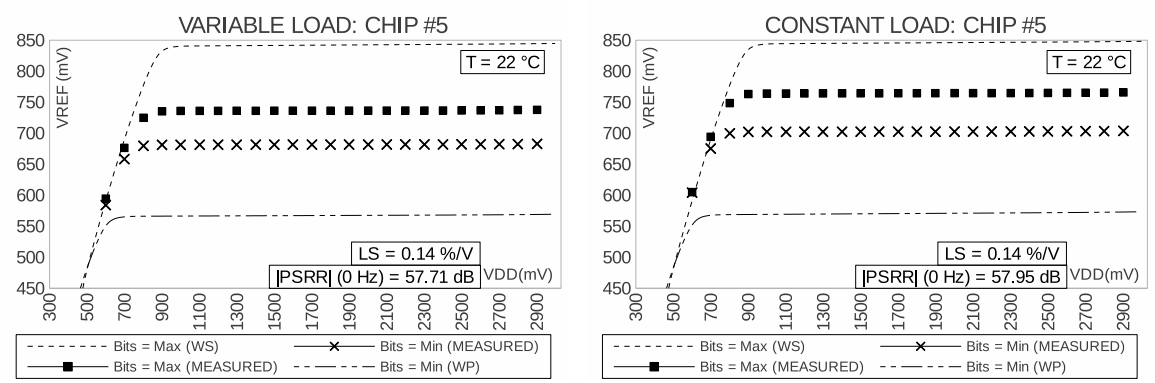


Figure 5.11: V_{REF} output vs supply voltage, Chip #5

The low level of change in the output with respect to changes in the supply voltage (Fig. 5.7 to 5.11) was the result of the cascode circuits used. Since the cascode design required little to no additional voltage (Sec. 4.5), sub-1 V operation was achieved on all the samples.

5.3.3 Evaluation of TC

The TC was evaluated from $-20\,^{\circ}\mathrm{C}$ to $80\,^{\circ}\mathrm{C}$ as per Eq. (2.20). The supply voltage was kept fixed at 1.5 V, well within the minimum and maximum operational range of the circuit. The output voltage as a consequence of all the bits being activated had an average positive slope with respect to temperature, while all the bits being de-activated resulted in an average negative slope. Based on where these slopes intersected, the bit combination required for a minimum TC was determined. Simulations resulted in a TC of less than 18 ppm/°C as shown in Fig. 5.12.

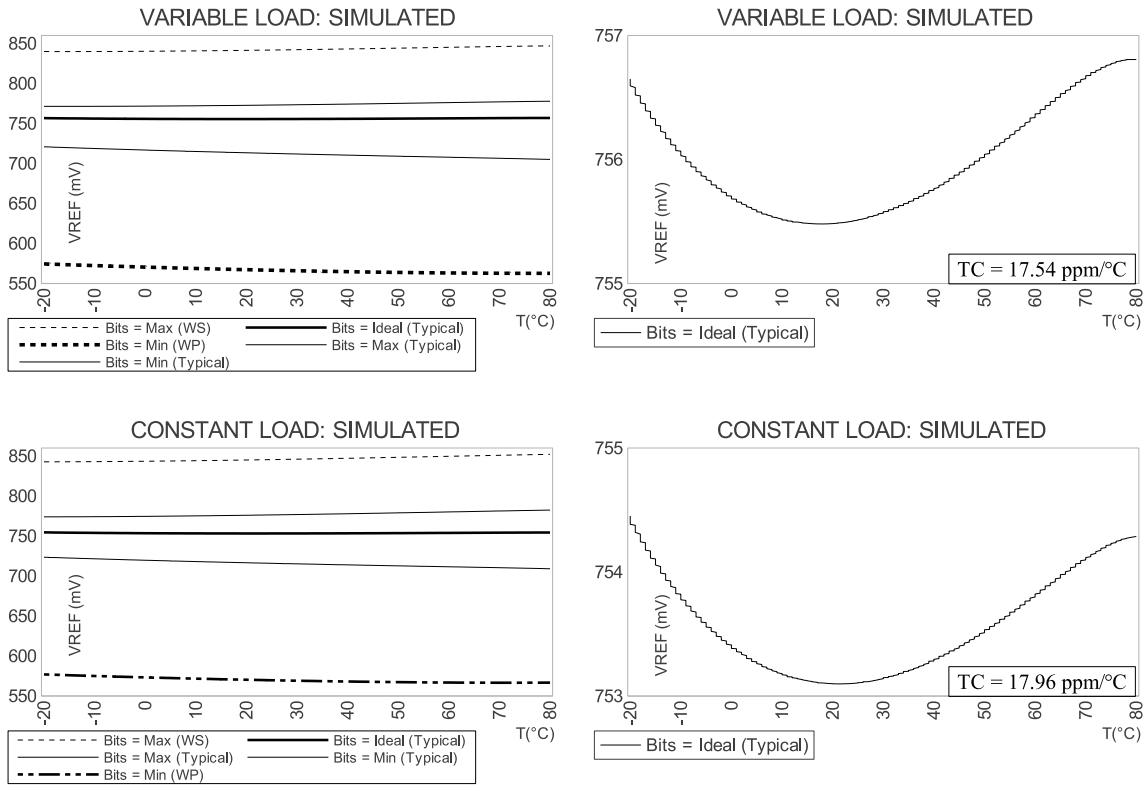


Figure 5.12: V_{REF} measured as temperature is increased, Simulations

5.3.4 Evaluation of TC: Testing of individual bits

To ensure that the bits of the experimental chips can be set to a desired voltage, a single value was cycled through every bit with its output measured at room temperature, as shown in Table 5.2, where Δ represents the difference in voltage compared to when no bits were set.

Table 5.2: Effect of activating one bit at a time, T=22°C

Chip	Architecture	Bit Active	NONE	1	2	3	4	5	6
1	Variable Load	Output (mV)	698.10	699.24	700.50	702.61	707.03	715.62	731.00
		Δ (mV)		1.14	2.40	4.51	8.93	17.52	32.90
1	Constant Load	Output (mV)	711.34	712.33	713.10	714.91	**727.76**	**727.82**	744.46
		Δ (mV)		0.99	1.76	3.57	16.42	16.48	33.12
2	Variable Load	Output (mV)	663.63	664.36	665.49	667.43	671.42	679.02	693.76
		Δ (mV)		0.73	1.86	3.80	7.79	15.39	30.13
2	Constant Load	Output (mV)	672.93	673.67	674.59	676.40	679.19	687.77	700.57
		Δ (mV)		0.74	1.66	3.47	6.26	14.84	27.64
3	Variable Load	Output (mV)	681.06	682.26	683.31	685.32	689.48	697.69	713.26
		Δ (mV)		1.2	2.25	4.26	8.42	16.63	32.2
3	Constant Load	Output (mV)	680.16	681.05	682.27	684.92	691.63	696.22	711.78
		Δ (mV)		0.89	2.11	4.76	11.47	16.06	31.62
4	Variable Load	Output (mV)	660.54	661.75	662.73	664.63	668.58	676.43	691.12
		Δ (mV)		1.21	2.19	4.09	8.04	15.89	30.58
4	Constant Load	Output (mV)	689.23	690.19	691.22	692.41	696.82	**710.97**	**713.2**
		Δ (mV)		0.96	1.99	3.18	7.59	21.74	23.97
5	Variable Load	Output (mV)	691.84	693.05	694.23	696.56	701.13	709.97	726.66
		Δ (mV)		1.21	2.39	4.72	9.29	18.13	34.82
5	Constant Load	Output (mV)	700.22	701.59	702.46	704.84	**705.23**	**722.39**	**725.49**
		Δ (mV)		1.37	2.24	4.62	5.01	22.17	25.27

The general trends show that the active bits increase the output by an amount proportionate to the weight of the bit. Exceptions are bolded and will be investigated in the proceeding paragraphs.

All the bits being de-activated resulted in the lowest output voltage, while all the bits being activated resulted in the highest. This was the expected result since the activation of bits represented a higher i_F, which according to Eq. (2.8), increased the output gate voltage.

It was expected that the increase of the voltage would correspond directly with the weight of the bits. As per Table 5.2, this expectation had been consistently met with the variable load architecture. However, with the constant load architecture, there were several instances in which the bit activation did not lead to the expected increase in voltage. This result may have compromised the resolution of i_F, undermining the ability of the constant load to switch to the correct voltage for minimum TC. Due to this problem, the constant load architecture had a higher measured average TC (as per Table 5.4).

A probable cause of this result may be due to an incorrect current being switched into the load. Since one side of the PMOS structure had no dummy transistors (Fig. 4.25), a possible culprit for this problem may be due to transistor mismatch in the PMOS cascode.

5.3.5 Evaluation of TC: Experimental Results

Measurements of the TC are depicted from Fig. 5.13 to 5.17. The lowest TC obtained was 8.05 ppm/°C (Fig. 5.16).

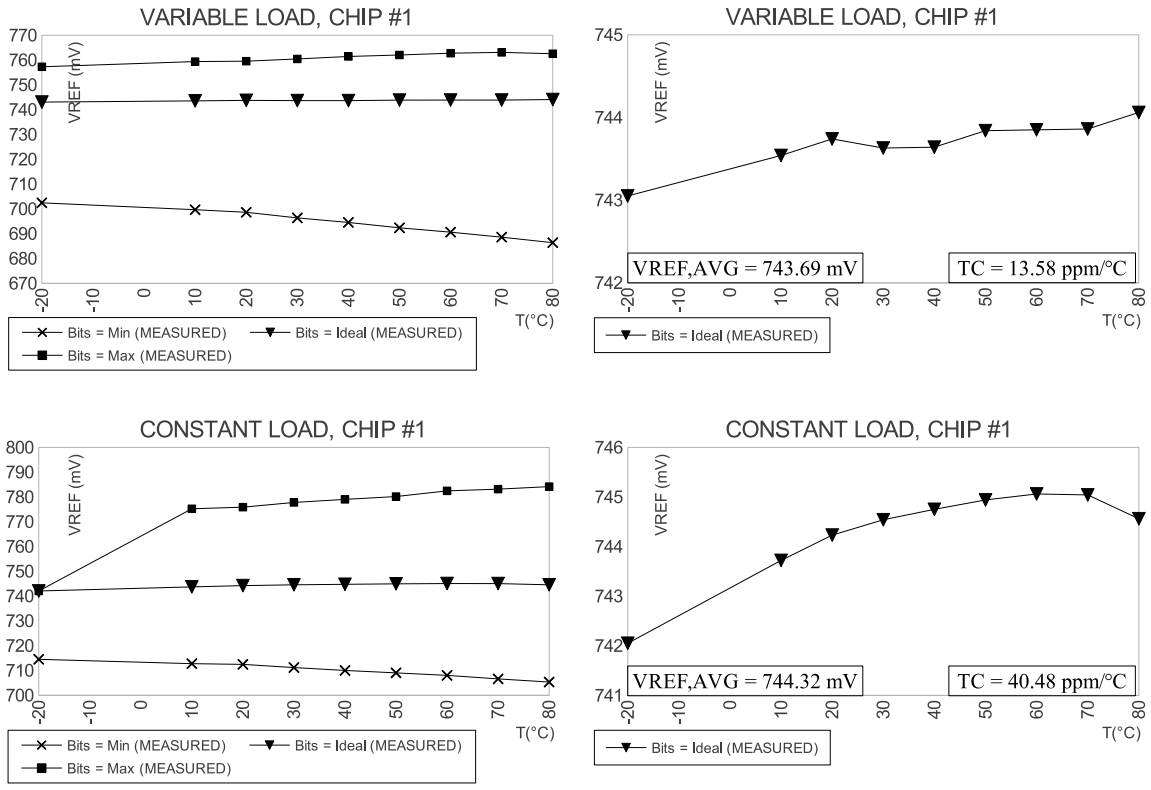


Figure 5.13: V_{REF} measured as temperature is increased, Measurements, Chip #1

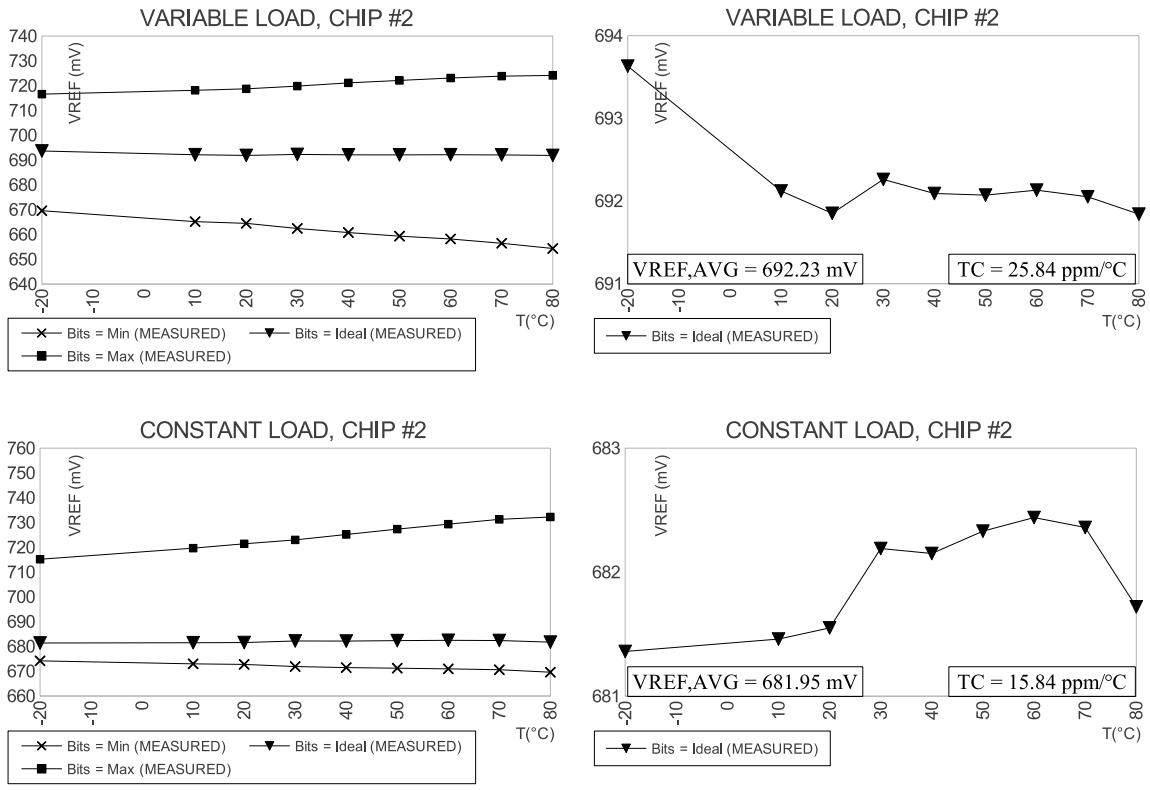


Figure 5.14: V_{REF} measured as temperature is increased, Measurements, Chip #2

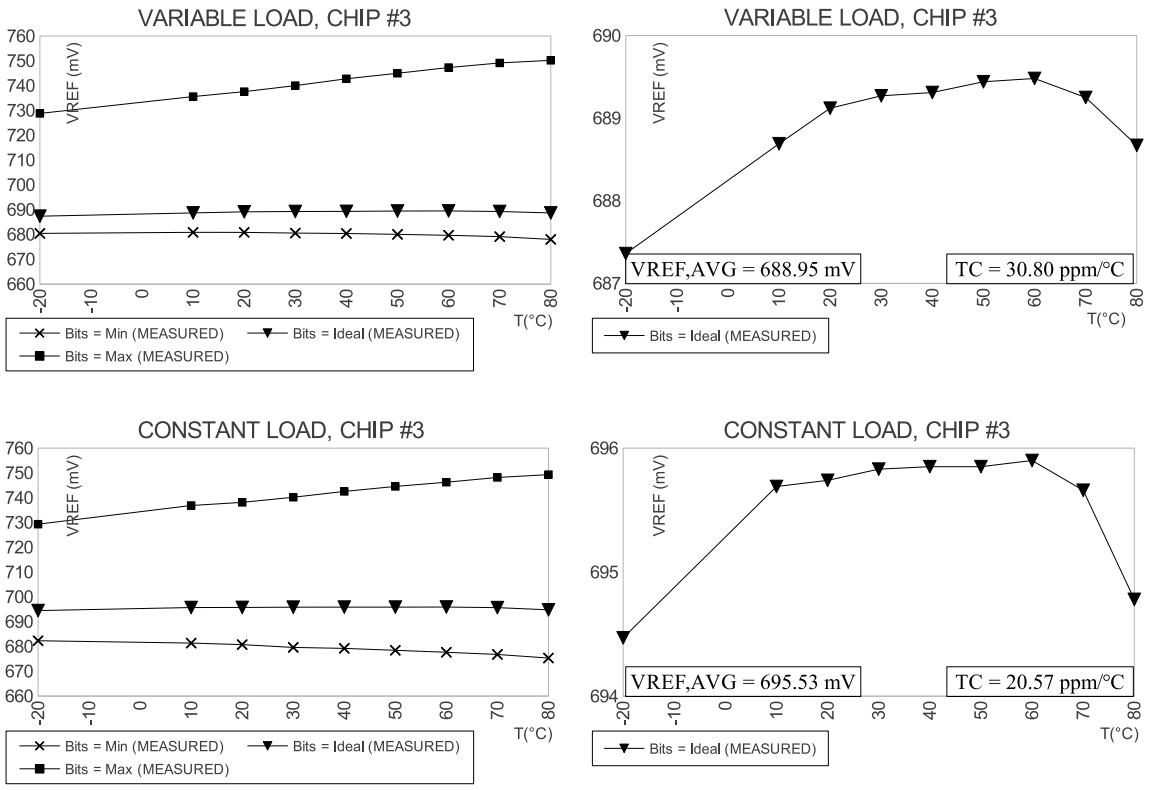


Figure 5.15: V_{REF} measured as temperature is increased, Measurements, Chip #3

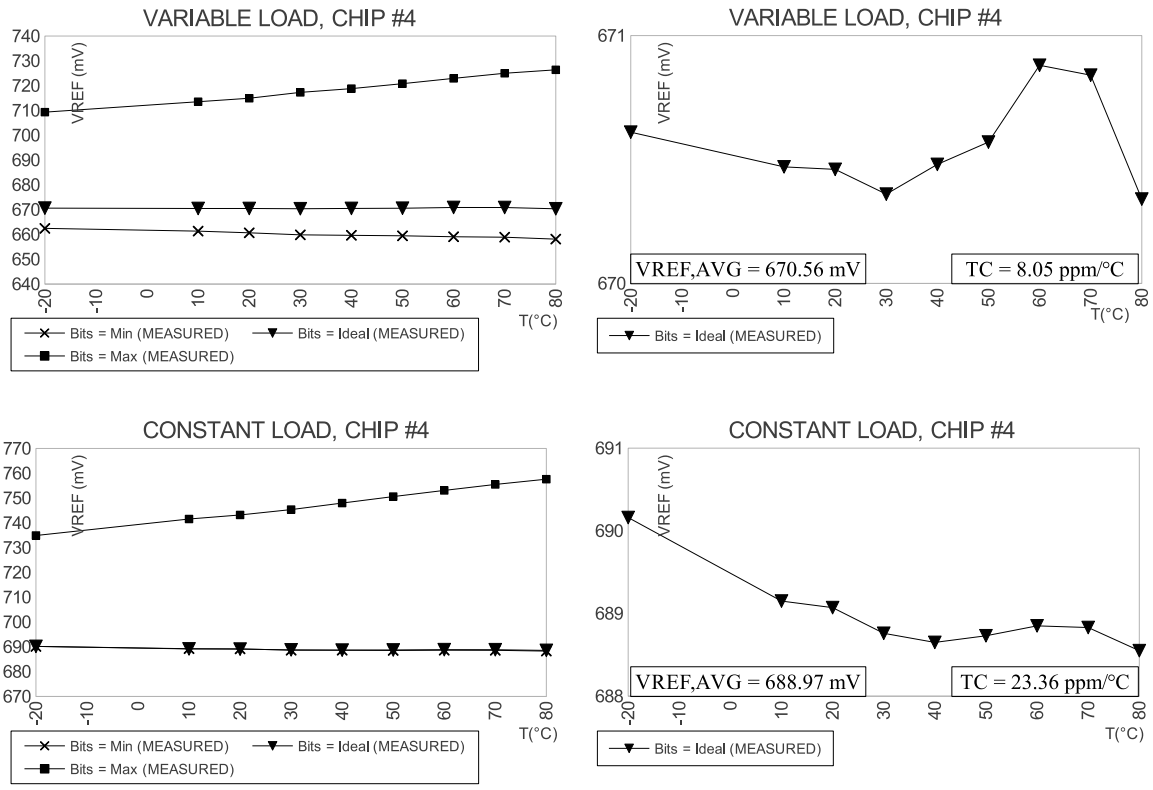
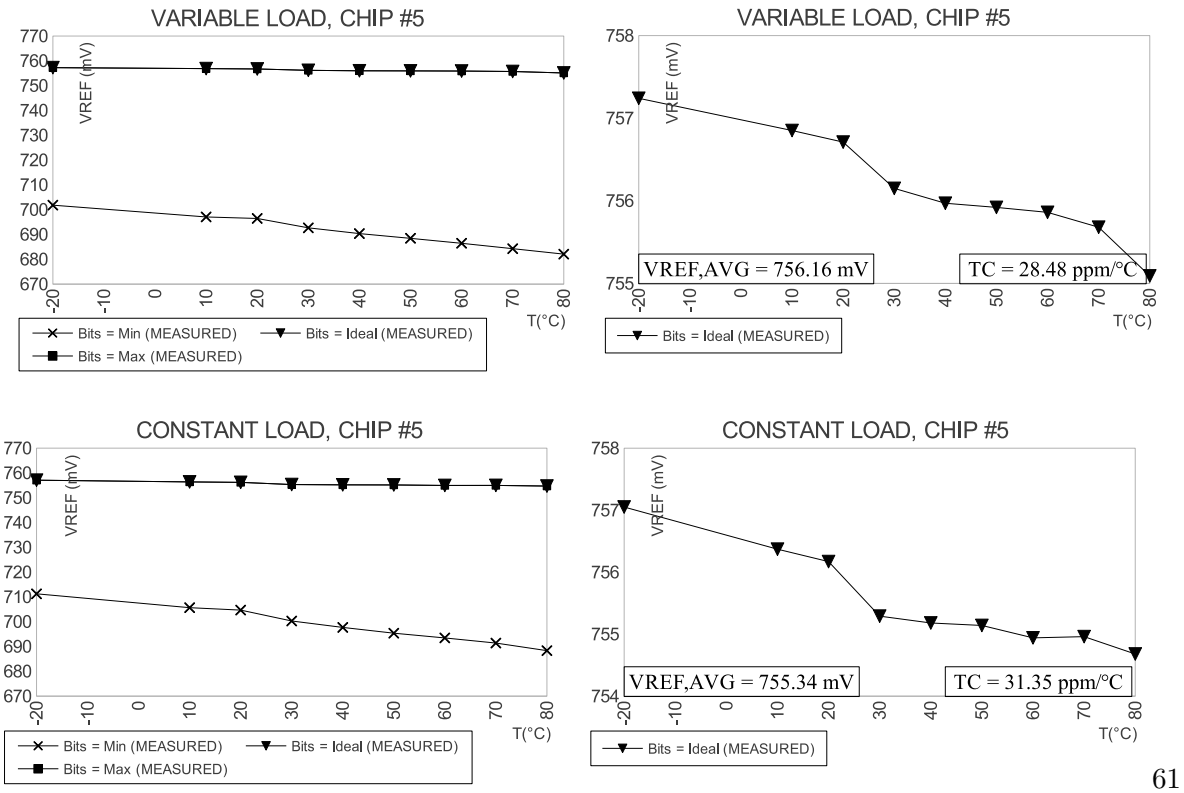


Figure 5.16: V_{REF} measured as temperature is increased, Measurements, Chip #4



61

Figure 5.17: V_{REF} measured as temperature is increased, Measurements, Chip #5

Analysis: Evaluation of TC

As per Table 4.1, the circuit was designed for a TC of 5 ppm/°C. However, the measured and simulated performance was beyond the designed value and varied by a large margin as depicted in Table 5.4. A possible explanation for this discrepancy are second order effects. These effects include the variation n (slope factor) with temperature, the approximation of Eq. (2.8), and the variation of the threshold voltage with respect to temperature not being entirely linear as assumed in Eq. (4.4). Second order effects are not accounted for in the simplified ACM model [2] used for the design and to account for them would go beyond the scope of this work.

Another possible explanation are the limitations in the experimental setup. As detailed in Sub-Sec. 5.2.7, observed noise associated with the measurement technique may have introduced errors to the recorded results. The design objective of 5 ppm/°C implied approximately 0.35 mV of difference between -20°C to 80°C. [3] However, noise sources may affect the readings by up to ±0.2 mV, representing a large fraction of the precision required to determine the TC. There may also be noise intrinsic to the chip, which may be an area of further research.

Analysis: Average V_{REF}

It was expected that the average V_{REF} within a single chip would be identical amongst both architectures. The V_{REF} is ideally equivalent to the threshold voltage extrapolated to zero Kelvin. With sufficient matching in the layout, the transistors should have identical threshold voltages. However, amongst three of the chips, as depicted from Fig. 5.14 to Fig. 5.16, the average voltage values differed by approximately 10 mV, implying a difference in threshold voltage amongst the architectures. The output voltage depended largely on the NMOS transistors, and the NMOS layout was done in accordance to known measures to prevent a mismatch in the threshold voltage (Sec. 2.8). Despite these measures, the likeliest cause of this discrepancy may be due to random process variations in the fabrication process.

Analysis: Design

Fig. 5.17 had demonstrated that despite all the bits being set on, the output decreased with temperature. This result implied that the estimated range of i_F was insufficient (Sec. 4.6.1), which in turn meant that the a spread in the temperature coefficient of the threshold voltage, $\frac{\partial V_T}{\partial T}$, was larger than anticipated.

[2] The ACM model used in this work is in fact a simplication of a more complex and complete model presented in [30].

[3] Assuming $V_{REF,MIN}$ = 700 mV, $V_{REF,MAX}$ = 700.35 mV, and calculating with Eq. (2.20).

This implication is as per Eq. (4.7),

$$\mathcal{F}(i_F) = \frac{K_{VT}}{n\, U_{TR}} \,,$$

where a larger $\frac{\partial V_T}{\partial T}$ requires a larger i_F, since according to Eq. (4.5),

$$K_{VT} = T_R \left| \frac{\partial V_T}{\partial T} \right| \,.$$

As well, this spread was larger than what could be calculated using App. D.

The datasheets from the foundry provided only a single value for $\frac{\partial V_T}{\partial T}$. Since this work had shown that there is a considerable spread in $\frac{\partial V_T}{\partial T}$ with process variations, more testing from the foundry is required to accurately characterize $\frac{\partial V_T}{\partial T}$.

5.3.6 Summary of Results

A summary of results for every chip and architecture is shown in Table 5.3,

Table 5.3: Summary of results

Chip	Architecture	LS (%/V)	$PSRR$ (dB)	TC (ppm/°C)	$V_{REF,TC}$ (mV)
1	Variable Load	0.15	57.15	13.58	743.69
1	Constant Load	0.14	-57.78	40.48	744.32
2	Variable Load	0.17	-56.54	25.84	692.23
2	Constant Load	0.14	-58.13	15.84	681.95
3	Variable Load	0.13	-58.86	30.80	688.95
3	Constant Load	0.12	-59.11	20.57	695.53
4	Variable Load	0.13	-58.54	8.05	670.56
4	Constant Load	0.12	-59.25	23.36	688.97
5	Variable Load	0.14	-57.71	28.43	756.16
5	Constant Load	0.14	-57.95	31.35	755.34

where LS (Eq. (2.18)) and $PSRR$ (Eq. (2.19)) was evaluated at an ambient temperature of 22 °C with the power supply being varied from 0 V to 3 V. The TC (Eq. (2.20)) was based on the voltage output from −20 °C to 80 °C, and $V_{REF,TC}$ is the average value of V_{REF} over the same tested temperature range.

5.3.7 Statistics of Results

The overall statistical performance of the circuit is shown in Table 5.4,

Table 5.4: Statistics of results, overall

Output	Architecture	μ	σ	σ/μ	MIN	MAX
LS	Variable Load	0.1445 %/V	0.018 %/V	12.79%	0.1211 %/V	0.1655 %/V
LS	Constant Load	0.1314 %/V	0.011 %/V	8.32%	0.1212 %/V	0.1434 %/V
$PSRR$	Variable Load	-57.77 dB	1.11 dB	1.92%	-56.54 dB	-58.86 dB
$PSRR$	Constant Load	-58.57 dB	0.72 dB	1.23%	-57.78 dB	-59.25 dB
TC	Variable Load	19.57 ppm/°C	10.55 ppm/°C	53.91%	8.05 ppm/°C	30.80 ppm/°C
TC	Constant Load	25.06 ppm/°C	10.74 ppm/°C	42.84%	15.84 ppm/°C	40.48 ppm/°C
$V_{REF,TC}$	Variable Load	698.86 mV	31.37 mV	4.49%	670.56 mV	743.69 mV
$V_{REF,TC}$	Constant Load	702.69 mV	28.3 mV	4.03%	681.95 mV	744.32 mV

where μ is the average value, σ is the standard deviation from the average value, and σ/μ is the coefficient of variation. It is important to note that one of the samples (Chip #5) had been omitted from this statistical analysis. This omission was due to its output with respect to temperature decreasing with temperature despite having all the bits set on (Fig. 5.17), which is not considered representative of the circuit design since it implies a larger range of inversion levels than the design was based on.

As only four chip samples are considered, μ and σ are insignificant [4] and hence further statistical analysis is required for the significance of these results to be known. Using methods described in [29], it was found that the population median falling within the MIN and MAX of any 4 samples had an 87.5% probability of being true for an infinitely large (theoretical) sample size. [5]

[4] These statistical values are only included since they are the only ones mentioned in most publications, and hence are used for purposes of comparison.

[5] A way to verify this figure is shown in App. H.

5.4 Comparison with Other Designs

Table 5.5: Comparison of experimental results with other designs

	Variable Load	Constant Load	[11]	[12]	[13]	[14]	[15]	[15]
Year	2014	2014	2002	2007	2009	2011	2012	2012
CMOS Tech. (μm)	0.35	0.35	0.5	0.35	0.35	0.18	0.13	0.13
Voltage (V)	0.9–3	0.9–3	3.7	0.9–4	1.4–3	0.45–2	0.5–3.0	0.5–3.0
Current (nA)	3.3	4.3	378×10^3	40	214	5.8	0.0044	0.0217
V_{REF} (mV)	698.9	702.7	1121.9	670	745	257.5	176	176
LS (%/V)	0.14	0.13	n/a	0.27	0.002	0.440	0.033	0.036
PSRR (dB)	-58.8	-58.6	-45	-47/-40	-45	-45	-53/-62	-51/-64
/f (Hz)	(DC)	(DC)	10	100/10M	100	<100	100/10M	100/10M
TC (ppm/°C)			145					
(avg)	19.6	25.1	n/a	n/a	15	165	49	29
(best)	8.1	15.9	n/a	10	7	39	16.9	5.3
(worst)	30.8	40.5	n/a	n/a	45	357	231	47.4
(# of samples)	(4)	(4)	(4)	(20)	(17)	(40)	(49)	(30)
	(post-trim)	(post-trim)						(post-trim)
Area (mm^2)	0.053	0.048	0.4	0.045	0.055	0.043	0.00135	0.0093

Table 5.5 compares the mean experimental results of the two proposed designs with other published voltage references. The performance of the proposed circuits compares well to other designs except for the very low consumption in [15]. However, the circuit in [15] is based on two transistor types of different threshold voltages, which are not always available to designers. Hence the design presented in this work is a generic alternative, requiring no special technology options and hence usable with any process.

Chapter 6

Conclusions

6.1 Achievements of this work

Two MOS-only voltage reference circuits with their respective trimming methods have been designed, fabricated and experimentally verified. A design methodology that accounted for the layout, performance improvement using cascodes, and the effect of TC was developed. In addition, a complete layout was created with techniques used to ensure well-matched transistors.

Experimental methods were developed to allow the measurement of nano-watt-power device outputs. Methods were also developed to verify bit operations and bit settings for the optimal TC. The objectives of low power, low area, and high performance were met:

1. Both circuits operated down to a 0.9 V supply, each consuming less than 5 nA.
2. Each circuit occupied an area of 0.053 mm^2 or less.
3. An LS and $PSRR$ below 0.15 %V and -58 dB was achieved.
4. The circuit had an average TC of less than 26 ppm/°C, with a median TC of less than 22 ppm/°C, both of which represent approximately 2 mV of variation from $-20\,°C$ to $80\,°C$. The best-performing circuit had a TC under 8.1 ppm/°C, representing less than 0.6 mV of variation over the same temperature range.

6.2 Potential Improvements

6.2.1 Noise Considerations

It remains inconclusive whether the design objective of 5 ppm/°C (as per Sec. 4.7) could have been met or exceeded, since noise consistently affected the output signals despite having filtering capacitors on every output. Noise can be reduced by testing the chip in an environment that has minimal fluctuations in physical vibrations and temperature. These conditions require improved facilities and equipment. Noise reduction can also achieved by avoiding the use of wires to connect test components, such as the use of a PCB.

6.2.2 Design Considerations

The PMOS sections could have improved matching by ensuring dummy transistors surround the entire perimeter as well as ensuring better symmetry and distribution of unit transistors in the PMOS layout. As well, the width/length dimensions of the PMOS unit transistor could have been increased. With this increase, differences in the transistor dimensions due to process variations would account for a lower percentage of the overall width/length. Doing so can rule out transistor mismatch as the cause of lower currents (Sub-Sec. 5.3.1), bit malfunctions (Sub-Sec. 5.3.4), and differences in the 0 K threshold voltage (Sub-Sec. 5.3.5).

6.2.3 Testing Considerations

The following considerations may improve the comprehensiveness or ease of the testing methods.

- The $PSRR$ at other frequencies can be characterized with the use of a waveform generator and oscilloscope.
- The bit and clock switching was done by hand in this work, aided by a Schmidt Trigger (Sub-Sec. 5.2.5). The automation of the bit and clock switching by having it programmed with a microcontroller and using appropriate clock drivers could improve the efficiency and reliability of such switching.

6.3 Future Research

- The temperature coefficient of threshold voltage was shown to span a wider range than expected, though only a single value in the datasheet was provided for this parameter. Therefore the foundry may need to do further testing to better characterize the TC of $\frac{\partial V_T}{\partial T}$. Alternatively, designing for a higher range of i_F can also compensate for this wider range.

- Sudden temperature changes affected the ability of the shift register to hold the bits (Sub-Sec. 5.2.8), hence, there should be design considerations in the creation of a shift register to address this problem.

- Upon a loss of power, the bits have to be set again. Thus, a means to store the "correct" bit setting into the chip and automatically load upon startup is desirable. This function may be accomplished by incorporating EEPROM cells into the chip.

- The voltage reference itself is not very useful for practical applications. Therefore the implementation of the circuit into microchip temperature sensors, mixed signal converters or voltage regulators (Sec. 1.4) may be another avenue of research.

- During the design phase, it was found that various degrees of weak inversion levels in the NMOS pair of the current source lead to a different simulated TC for the voltage reference output. Various simulations found that there existed an optimal TC for a certain ratio of the NMOS pair. An example of this optimal ratio is shown in Fig. 6.1.

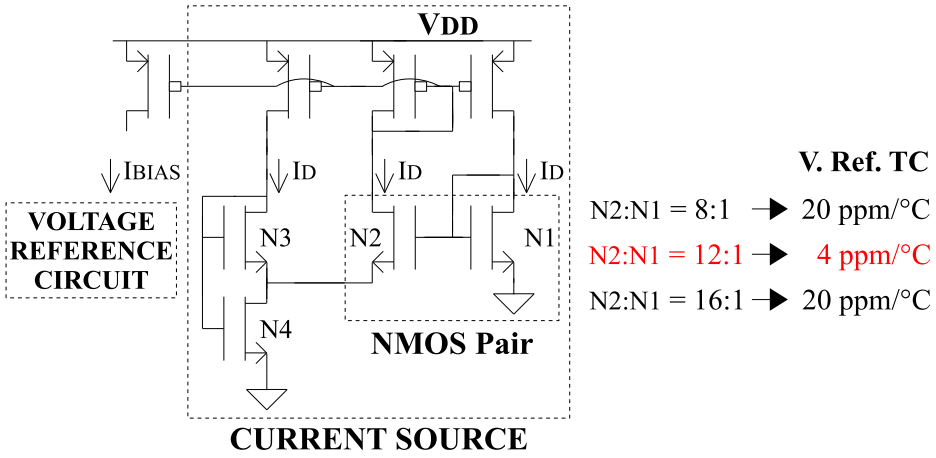


Figure 6.1: Optimal NMOS pair, based on simulations

A similar "trial-and-error" method of finding the optimal TC based on simulation results was also how the voltage reference of [15] was designed. The results from this effort were not used since it was not certain at the time whether or not simulation results based on a large temperature range could be relied upon. However, experimental results have thus far shown that the simulations were reasonably accurate. Hence, improved performance may be achieved by investigating why a certain weak-inversion NMOS pair in the current source would result in the optimal TC of the voltage reference output.

6.4 Future Research: Special Consideration

As the work of Seok et al. [15] had superior power and area consumption while exhibiting comparable performance, a special consideration is made for possible future research based on the ideas of this work. The drawback of the voltage reference presented in [15] was the use of native transistors, which may not always be available, nor at a comparable cost to older technologies with no special process options. A possible solution to this drawback is to replace the native transistor with a PMOS transistor as shown in Fig. 6.2.

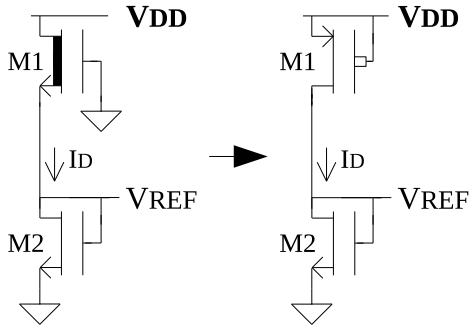


Figure 6.2: Voltage Reference of [15], possible alternative topology

The ACM model and trimming design considerations can then be applied to this alternative topology.

Appendices

Appendix A

Parameter Extraction

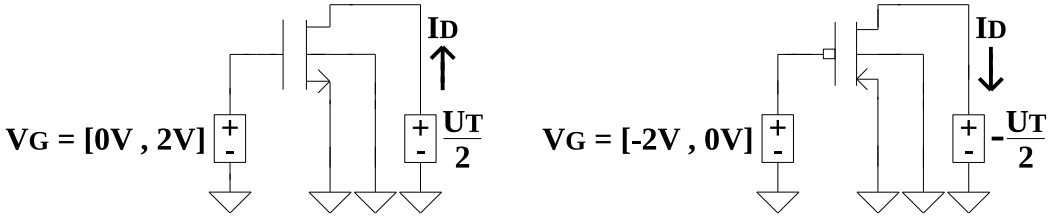
A.1 Motivation

The sheet normalization current, I_{SQ}, is not provided by transistor technology datasheets, and the threshold voltage, V_T, is also defined differently according to the ACM model (as per Sec. 2.4). Hence, an extraction method [30] is required to determine I_{SQ} and V_T . To fully account for process variations, this extraction method must be used for all the extreme process corners of the transistor. [1]

The methodology will first be stated, followed by a justification of methods used.

A.2 Methodology

The circuits used for parameter extraction are shown in Fig. A.1.



(a) NMOS extraction circuit (b) PMOS extraction circuit

Figure A.1: Circuits used to extract I_{SQ} and V_{TH}

[1] Only by simulations will it be possible to account for the extreme process corners. Otherwise, it is also possible to extract these parameters experimentally.

A multiple of the thermal voltage, $|U_T/2|$, is set as the drain voltage of the transistor while the gate voltage is varied from 0 to $|2\ V|$. The transconductance-to-current ratio, g_m/I_D, is required and can be derived from the output current, I_D, as follows,

$$\frac{g_m}{I_D} = \frac{1}{I_D}\,\frac{dI_D}{dV_G} = \frac{d\ln(I_D)}{dV_G}\ . \tag{A.1}$$

From the output in Eq. (A.1), 53% of the peak g_m/I_D is where the gate voltage is equivalent to the threshold voltage,

$$V_T = V_G\ @\ \left(0.53\ \left|\frac{g_m}{I_D}\right|_{MAX}\right)\ . \tag{A.2}$$

Finding the current output for the V_G from Eq. (A.2) and dividing by 0.885 as well as the transistor width/length ratio results in I_{SQ},

$$I_{SQ} = \frac{I_D\ (@V_G = V_T)}{0.885\ S}\ . \tag{A.3}$$

A.3 I_{SQ}, Justification of extraction methodology

Eq. (A.3) relies on $V_G = V_T$. The justification of the I_{SQ} extraction requires working out the inversion coefficients when this condition is true. The relationship between $i_{F(R)}$ and the transistor terminals are as follows,

$$\frac{V_G - V_T}{n\ U_T} - \frac{V_{S(D)}}{U_T} \cong \mathcal{F}(i_{F(R)}) = \sqrt{1 + i_{F(R)}} - 2 + \ln\left(\sqrt{1 + i_{F(R)}} - 1\right)\ . \tag{A.4}$$

With respect to Fig. A.1, if in Eq. (A.4) the source is 0 V and i_F is 3, then the following is true,

$$V_G = V_T\ (@i_F = 3)\ . \tag{A.5}$$

Again from Fig. A.1, substituting $|V_D| = U_T/2$ as well as the condition of Eq. (A.5) allows i_R to be found,

$$\frac{1}{2} = \sqrt{1 + i_R} - 2 + \ln\left(\sqrt{1 + i_R} - 1\right)\ . \tag{A.6}$$

From numerical analysis, Eq. (A.6) is true when $i_R = 2.115$. With the value of i_F and i_R determined, substituting into the equation for current results in

$$I_D = S\ I_{SQ}\ \left([i_F = 3] - [i_R = 2.115]\right) = S\ I_{SQ}\ (0.885)\ . \tag{A.7}$$

Rearranging Eq. (A.7) yields Eq. (A.3).

A.4 V_T, 53% Justification of extraction methodology

Eq. (A.2) requires 53% of the maximum g_m/I_D. Justifying this value first requires the derivation of transconductance,

$$g_m = \frac{dI_D}{dV_G} = S\ I_{SQ}\ \left(\frac{di_F}{dV_G} - \frac{di_R}{dV_G}\right)\ . \tag{A.8}$$

To determine $di_{F(R)}/dV_G$, its inverse must first be found by isolating V_G from Eq. (A.4) and then taking its derivative with respect to $i_{F(R)}$ as follows

$$V_G = \left[\mathcal{F}(i_{F(R)})\ U_T + V_S\right]\ n + V_T\ , \tag{A.9}$$

$$\frac{dV_G}{di_{F(R)}} = \left[\frac{d\mathcal{F}(i_{F(R)})}{di_{F(R)}}\ U_T\right]\ n\ , \tag{A.10}$$

where

$$\frac{d\mathcal{F}(i_{F(R)})}{di_{F(R)}} = \frac{d\mathcal{F}(i_{F(R)})}{d\sqrt{1+i_{F(R)}}}\ \frac{d\sqrt{1+i_{F(R)}}}{di_{F(R)}} = \frac{1}{2}\ \frac{1}{\sqrt{1+i_{F(R)}}-1}\ . \tag{A.11}$$

Substituting Eq. (A.11) into Eq. (A.10) and inverting the expression yields

$$\frac{di_{F(R)}}{dV_G} = \left[2\ \left(\sqrt{1+i_{F(R)}}-1\right)\ \frac{1}{U_T}\right]\ \frac{1}{n}\ . \tag{A.12}$$

Eq. (A.12) can be used for Eq. (A.8) which in turn gets simplified as

$$g_m = S\ I_{SQ}\ \frac{1}{n\ U_T}\ 2\ \left(\sqrt{1+i_F}-\sqrt{1+i_R}\right)\ . \tag{A.13}$$

Dividing g_m by I_D to obtain the transconductance-to-current ratio results in

$$\frac{g_m}{I_D} = \frac{1}{n\ U_T}\ \left(\frac{2}{\sqrt{1+i_F}+\sqrt{1+i_R}}\right)\ . \tag{A.14}$$

In Eq. (A.14) the maximum g_m/I_D occurs when $i_{F(R)} \to 0$, hence,

$$\left[\frac{g_m}{I_D}\right]_{MAX} = \frac{1}{n\ U_T}\ . \tag{A.15}$$

Substituting the $i_{F(R)}$ neccessary to satisfy the conditions for $V_G = V_T$ (from Sect. A.3) results in value that is 53% of the maximum g_m/I_D,

$$\left[\frac{g_m}{I_D}\right]_{V_G=V_T,V_D=0.5\ U_T} = \left[\frac{g_m}{I_D}\right]_{MAX}\ \left(\frac{2}{\sqrt{1+3}+\sqrt{1+2.115}}\right) \cong \left[\frac{g_m}{I_D}\right]_{MAX}\ (0.53)\ . \tag{A.16}$$

A.5 Sample Extracted Parameters

Sample extracted parameters for an undisclosed 0.35μm technology (due to legal restrictions) are shown in Table A.1,

Table A.1: 0.35μm technology, NMOS extracted parameters

	$I_{SQ,N}$			$V_{T,N}$			$I_{SQ,P}$			$\|V_{T,P}\|$		
T (°C)	TM	WP	WS	TM	WP	WS	TM	WP	WS	TM	WP	WS
-23.15	69.93	79.97	69.95	559.90	417.98	622.54	15.84	17.89	13.85	775.04	652.69	856.30
-13.15	73.69	84.10	97.21	552.40	410.43	514.86	17.27	19.61	15.05	760.80	638.51	842.00
-3.15	77.09	88.08	70.41	544.78	402.94	607.27	18.71	21.29	16.27	746.64	624.39	827.84
6.85	80.38	92.05	73.50	537.21	395.50	599.71	20.12	22.96	17.43	732.47	610.26	813.62
16.85	83.71	95.84	76.47	529.74	388.06	592.19	21.55	24.65	18.66	718.37	596.24	799.49
26.85	86.76	99.36	79.15	522.24	380.66	584.63	22.92	26.32	19.84	704.25	582.20	785.43
36.85	89.96	102.95	82.02	514.79	373.25	577.17	24.40	28.01	21.07	690.21	568.20	771.38
46.85	92.89	106.45	84.58	507.35	365.98	569.64	25.84	29.69	22.28	676.25	554.22	757.33
56.85	95.67	110.00	87.15	499.90	358.85	562.23	27.26	31.38	23.46	662.27	540.31	743.32
66.85	98.44	113.67	89.67	492.54	351.90	554.84	28.68	33.10	24.66	648.31	526.43	729.38
76.85	101.14	117.48	92.21	485.25	345.09	547.51	30.17	34.80	25.86	634.48	512.62	715.51
79.85	102.00	118.71	92.99	483.07	343.09	545.37	30.56	35.30	26.25	630.32	508.47	711.40

where $I_{SQ,N}$ and $I_{SQ,P}$ are the sheet specific currents for the NMOS and PMOS respectively, and are expressed in nA. $V_{T,N}$ and $|V_{T,P}|$ are the threshold voltages for NMOS and PMOS respectively, and are expressed in mV. TM (Typical Mean) represents the state of transistors which are most statistically likely to occur, while WP (worst power, fast-NMOS-fast-PMOS) and WS (worst speed, slow-NMOS-slow-PMOS) accounts for the worst-case process variations.

Appendix B

Subthreshold Slope Factor Calculation

B.1 Motivation

Essential design equations utilize the subthreshold slope factor, n. These equations include the conditions for zero temperature coefficient (Eq. (4.7)) and the trimming resolution (Eq. (4.30)).

B.2 Definition

In subthreshold operation, the current has an exponential dependence on the gate voltage as per Eq. (3.2). This dependence means that an increase in gate voltage causes an exponential increase in current, resulting in a linear characteristic as shown in Fig. B.1 [25].

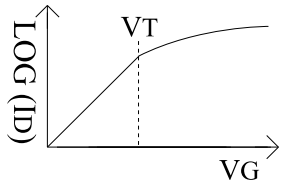


Figure B.1: Linear characteristic when $V_G < V_T$

The resulting slope from this subthreshold region of operation corresponds to n. For instance, a slope of 77 mV/dec corresponds to $n \cong 1.3$ at 25°C [25]. The calculation of n involves the following [30],

$$n \cong 1 + \frac{\gamma}{2\sqrt{2\,\phi_F + V_P}}\,, \tag{B.1}$$

where γ is the body effect coefficient factor, V_P is the pinch-off voltage, and ϕ_F is the Fermi potential.

γ can be found as a parameter from the transistor datasheets, while V_P is defined by [30]

$$V_P = \left[|V_G| - |V_T| + \left(\sqrt{2\ \phi_F + \frac{\gamma}{2}} \right)^2 - \frac{\gamma}{2} \right]^2 - 2\ \phi_F\ . \tag{B.2}$$

Φ_F is defined as

$$\phi_F|_{T=300K}\ = U_T\ \ln \left[\frac{N_A}{n_i|_{T=300K}} \right]\ , \tag{B.3}$$

where N_A is the acceptor ion concentration and n_i is the intrinsic carrier concentration. n_i is commonly cited as $1.45 \cdot 10^{16}\ m^{-3}$ [9] but experimental evidence supports a value closer to $1.0 \cdot 10^{16}\ m^{-3}$ [20].

N_A is derived using γ from the following [9],

$$\gamma = \frac{\sqrt{2\ \epsilon_0\ \epsilon_{Si}\ q\ N_A}}{C_{OX}}\ , \tag{B.4}$$

where ϵ_0 is the permittivity of free space ($10^{-9}/36\pi\ F/m$), ϵ_{Si} is relative permittivity of silicon (11.7), q is the electron charge ($1.602 \cdot 10^{-19}\ C$), and C_{OX} is the oxide capacitance.

C_{OX} can be found by [9]

$$C_{OX} = \frac{\epsilon_0\ \epsilon_{OX}}{t_{OX}}\ , \tag{B.5}$$

where t_{OX} is the oxide thickness (usually provided by technology datasheets), and ϵ_{OX} is the relative permittivity of the oxide (3.9 for silicon oxide).

Appendix C

Proof of Cascode increasing Output Resistance

Proving the cascode transistors increase the small signal output resistance requires small signal analysis.

C.1 Small Signal Analysis Model

The small signal analysis model represents the circuit as a result of miniscule changes in current and voltage and is depicted in Fig. C.1 [30],

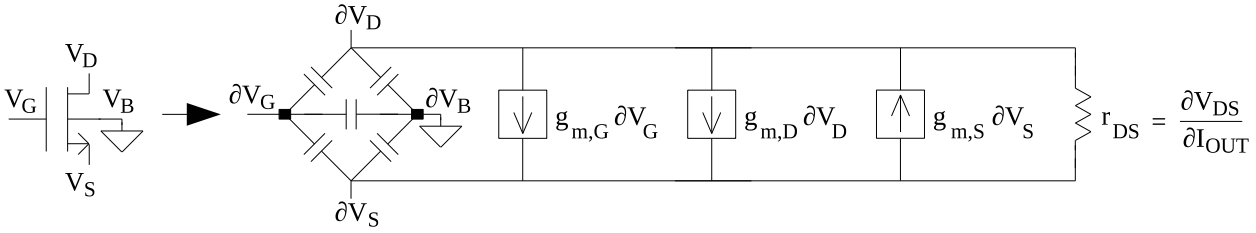


Figure C.1: Small signal model, complete

where current flow is shown as a consequence of small changes in terminal voltages and the drain to the source is characterized as a resistance, r_{DS}. The transconductance, g_m, represents small changes in the output current caused by a small change in a particular terminal voltage and is expressed as

$$g_{m,TERMINAL} = \frac{\partial I_{OUT}}{\partial V_{TERMINAL}} \, . \tag{C.1}$$

Assuming that the transistor is in saturation, ∂V_S causes a much larger ∂I_{OUT} than ∂V_D, therefore

as per Eq. (C.1), $g_{m,S} >> g_{m,D}$ [30]. Neglecting the capacitances and further assuming a purely DC gate voltage (hence no small-signal fluctuations), Fig. C.1 can be simplified as shown in Fig. C.2.

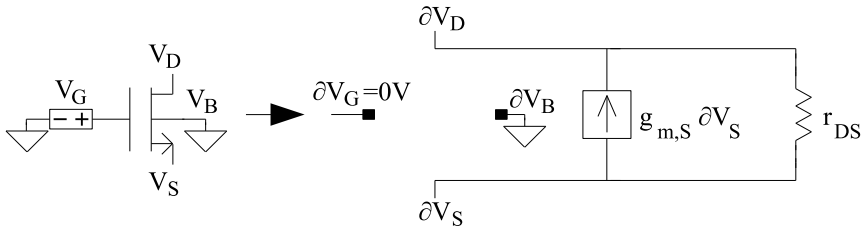


Figure C.2: Small signal model, simplified

C.2 Output Resistance

The output resistance, R_{OUT}, is defined as resistance due to a theoretical fluctuation of voltage (∂V_X) and current (∂I_X) at a particular point in a circuit [26]. This resistance is expressed as

$$R_{OUT} = \frac{\partial V_X}{\partial I_X} \ . \tag{C.2}$$

A non-cascoded transistor connected the the gate of a current mirror is shown in Fig. C.3.

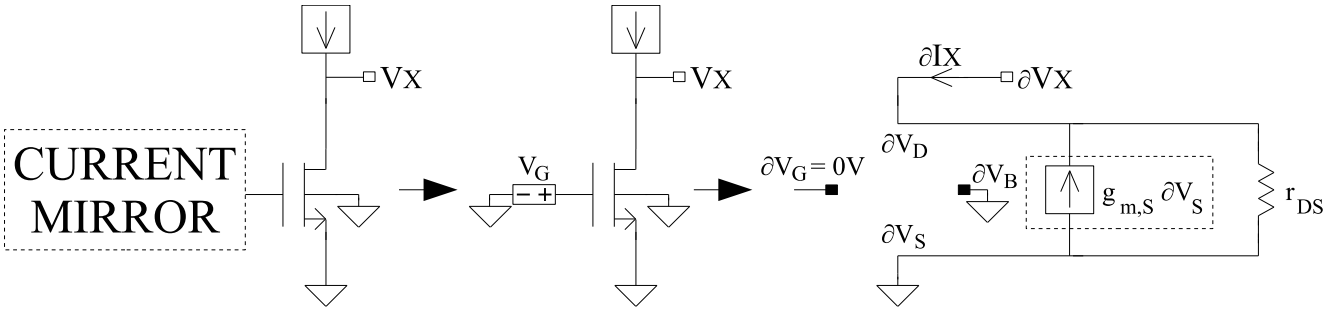


Figure C.3: Current mirror transistor, small signal equivalent

As $\partial V_S = 0$, its corresponding current can be neglected and hence, for a non-cascoded configuration,

$$R_{OUT,NON-CASCODED} = r_{_{DS}} \ . \tag{C.3}$$

For purposes of comparison, a cascoded transistor configuration is depicted in Fig. C.4 .

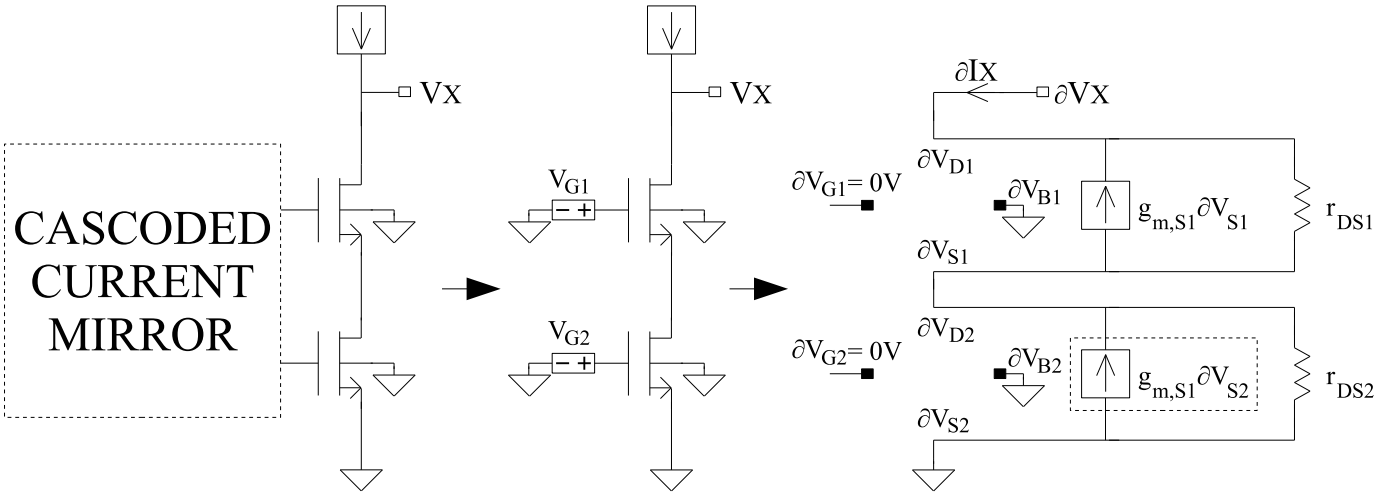


Figure C.4: Cascoded current mirror transistor, small signal equivalent

To determine $\partial V_X/\partial I_X$, ∂V_X must first be related to the small-signal voltage drops of the cascoded configuration,

$$\partial V_X = \partial V_{r_{DS,1}} + \partial V_{r_{DS,2}} \ , \tag{C.4}$$

where the small-signal drop across $r_{DS,1}$ is as follows,

$$\partial V_{r_{DS,1}} = \left[\partial I_X + g_{m,S1}\partial V_{S1}\right]\ r_{DS,1} \ , \tag{C.5}$$

and $\partial V_{r_{DS,2}}$ is characterized by

$$\partial V_{r_{DS,2}} = \partial V_{S1} = \left[\partial I_X + g_{m,S1}\partial V_{S1}\right]\ r_{DS,2} \ . \tag{C.6}$$

To aid in simplification of the expressions, ∂V_{S1} in Eq. (C.6) is isolated in terms of the cascode configuration as follows,

$$\partial V_{S1} = \partial I_X \left[\frac{r_{DS,2}}{1 - r_{DS,2}\ g_{m,S1}}\right] \ . \tag{C.7}$$

Substitution of Eq. (C.5, C.6, C.7) into Eq. (C.4) results in

$$\partial V_X = \left[\partial I_X + g_{m,S1}\left(\partial I_X \left[\frac{r_{DS,2}}{1 - r_{DS,2}\ g_{m,S1}}\right]\right)\right] + \left(\partial I_X \left[\frac{r_{DS,2}}{1 - r_{DS,2}\ g_{m,S1}}\right]\right) \ . \tag{C.8}$$

Hence,

$$R_{OUT,CASCODED} = \frac{\partial V_X}{\partial I_X} = r_{DS,2} + \left[\frac{r_{DS,2}}{1 - r_{DS,2}\ g_{m,S1}}\right]\left(g_{m,S1}\ r_{DS,1} + 1\right) \ . \tag{C.9}$$

which is a larger output resistance than without the cascode configuration as per Eq. (C.3).

Appendix D

Threshold Voltage Temperature Coefficient Calculation

D.1 Motivation

As per Sect. 4.6.1, the data on the dispersion of K_{VT} was not known and empirical data along with a safety margin was used to estimate it. To verify this estimation, this Appendix shows how such dispersion can be calculated. As per Eq. (4.5), K_{VT} is the expected drop in threshold voltage from 0 K,

$$K_{VT} = T_R \left| \frac{\partial V_T}{\partial T} \right| . \tag{D.1}$$

The $\partial V_T/\partial T$ is given as a single value in technology datasheets. However, $\partial V_T/\partial T$ may change due to process variations. This change can be derived from equations that define the threshold voltage.

D.2 Definition

To determine $\partial V_T/\partial T$, the Poly-Silicon work function, $\Phi_{POLY/SI}$, must first be determined from the definition of the threshold voltage [35]

$$V_T(T) = \Phi_{POLY/SI}(T) + 2\,\phi_F(T) + \gamma\,\sqrt{2\,\phi_F(T) + V_S}\,. \tag{D.2}$$

V_T is extracted from App. A, while ϕ_F is determined from Eq. (B.3). Using $\Phi_{POLY/SI}$ and all the known values from Eq. (D.2), $\partial V_T/\partial T$ can be found by [35]

$$\frac{dV_T(T)}{dT} = \frac{\Phi_{POLY/SI}(T)}{T} + \frac{3k}{q} + \frac{E_{G,0K}}{qT} + \left(2 + \frac{\gamma}{C_{OX}}\sqrt{\frac{1}{2\,\Phi_F(T) + V_S}}\right)\left(\frac{\Phi_F(T)}{T} - \frac{3k}{2q} - \frac{E_{G,0K}}{2qT}\right). \tag{D.3}$$

The dispersion of K_{VT} can be determined by evaluating Eq. (D.3) at all process corners.

Appendix E

Automatic Starter Circuit

E.1 Motivation

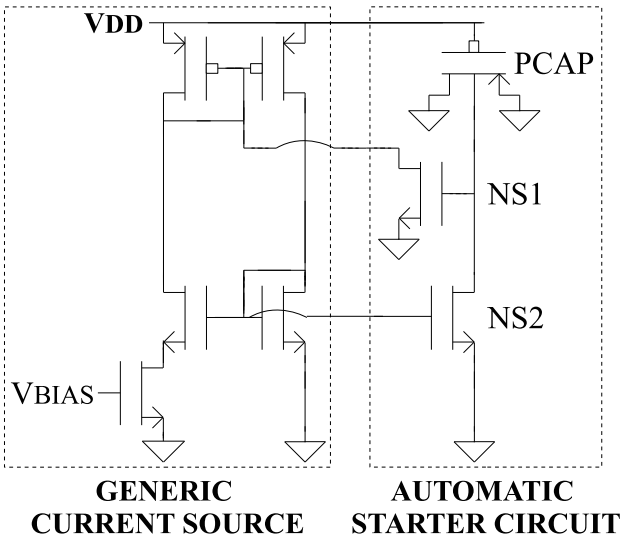


Figure E.1: Starter circuit

In Fig. E.1, as V_{DD} ramps up from 0 V, the PMOS current mirror will initially conduct more current. To conduct the full current requires the NMOS pair to be turned on with a high enough gate voltage. If the NMOS pair does not turn on sufficiently, then the current flow may cease. To prevent this state of operation, a starter circuit is neccessary. The starter circuit depicted in Fig. 4.17 requires a starting signal to be manually applied and then removed, which is undesirable for use in a practical application. Hence, an alternative starter circuit is required to have a circuit that can work

automatically.

E.2 Explanation

As V_{DD} increases, a PMOS connected as a capacitor, PCAP, creates a positive potential at the gate of NS1, allowing NS1 to divert current from the diode-connected PMOS to ground. This path ensures that the sufficient current is generated by the PMOS mirror into the NMOS mirror. Once the NMOS pair conducts enough current, NS2 would be turned on sufficiently to dissipate the positive potential at NS1, allowing the path from the diode-connected PMOS to close.

E.3 Another Alternative

A much simpler alternative starter circuit is shown in Fig. E.2.

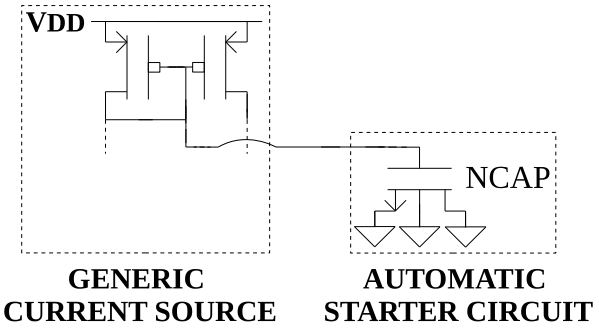


Figure E.2: Starter circuit, another alternative

The general premise is that when the circuit is turned on, NCAP charges through the diode-connected PMOS, allowing the rest of the circuit to start.

Appendix F

Chip Connection Diagram

Shown in Fig. F.1 is how the voltage reference microchip was connected. This diagram can be used to aid future testing of the microchips produced in this work.

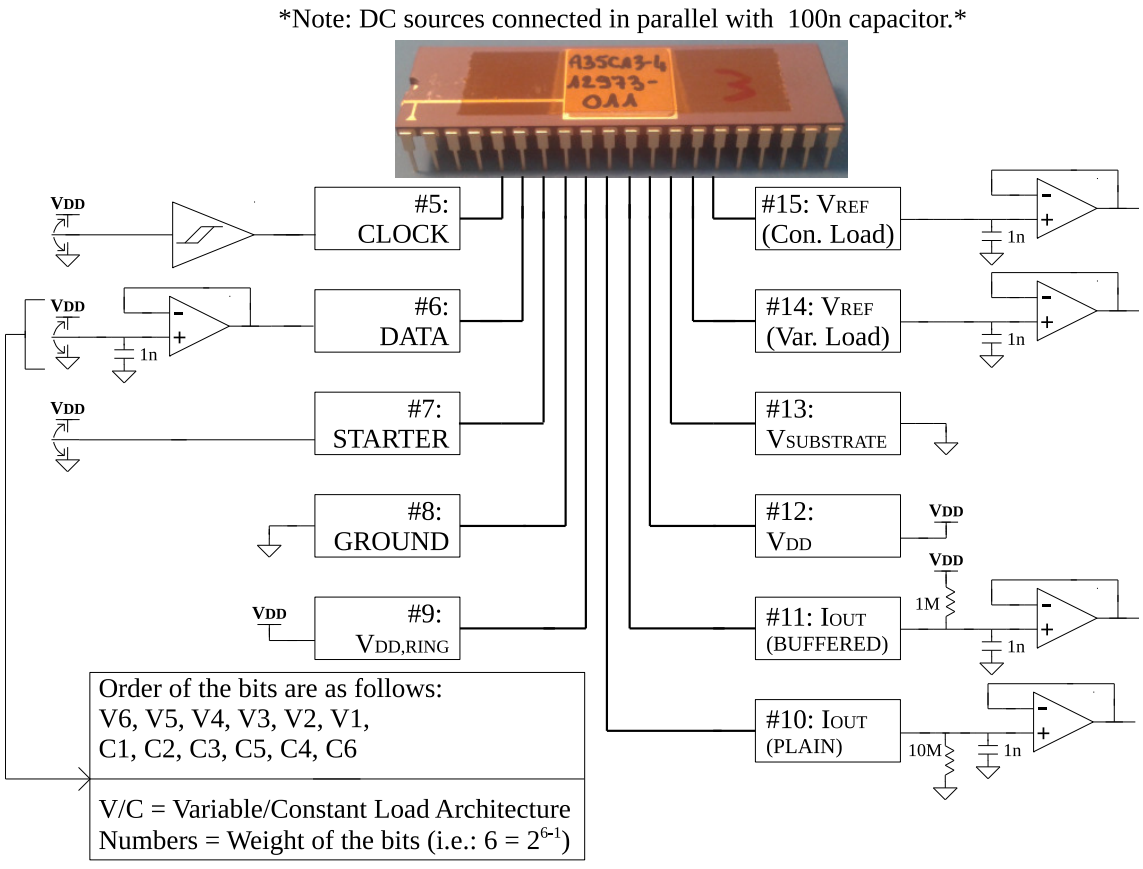


Figure F.1: Chip connection diagram

Appendix G

Measurement Raw Data

G.1 DC Measurements

DC measurements were made as the supply voltage was ramped up from 0.6 V up to 3.0 V at an ambient temperature of 22°C.

Table G.1: Current Measurements, T=22°C, All Chips

Chip # →	1	2	3	4	5
V_{DD} (mV) ↓	I_{OUT} (pA) ↓	I_{OUT} (pA) ↓	I_{OUT} (pA) ↓	I_{OUT} (pA) ↓	I_{OUT} (pA) ↓
600	435.58	304.66	401.51	388.87	469.55
700	481.85	325.81	440.33	435.31	524.92
800	483.63	327.97	441.77	436.46	527.76
900	484.01	327.04	442.62	436.29	528.59
1000	484.49	326.91	441.89	436.69	528.34
1100	484.76	327.16	442.15	436.8d4	528.69
1200	484.34	327.52	443.08	437.09	529.42
1300	485.04	327.79	443.68	436.81	529.52
1400	484.84	327.82	443.20	437.19	528.81
1500	485.04	327.32	443.53	436.59	529.32
1600	485.27	328.17	443.18	436.34	528.79
1700	485.27	327.47	443.58	437.34	529.17
1800	485.34	327.74	444.33	437.02	529.69
1900	485.52	327.84	444.03	437.72	530.10
2000	485.47	328.72	444.46	437.29	530.37
2100	485.34	328.25	443.93	438.45	530.12
2200	486.35	328.98	444.28	438.83	530.73
2300	486.40	329.28	444.11	438.88	531.20
2400	486.25	329.30	444.99	438.42	531.53
2500	486.67	330.36	445.19	438.95	531.45
2600	487.48	330.51	445.89	439.86	532.16
2700	489.34	332.02	446.24	440.56	533.42
2800	483.05	327.09	441.72	436.69	526.80
2900	490.32	333.10	447.50	441.37	535.20
3000	491.93	334.53	449.46	443.40	536.86

Table G.2: V_{OUT} as V_{DD} increases, T=22°C, Chip #1

Architecture →	Variable	Variable	Constant	Constant
Bit Config. →	Min	Max	Min	Max
V_{DD} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓
600	590.95	595.79	597.14	602.64
700	666.98	682.82	669.24	687.46
800	700.10	747.28	694.95	745.04
900	703.67	766.81	697.20	758.50
1000	703.85	768.59	697.35	759.57
1100	703.98	768.64	697.42	759.69
1200	704.05	768.78	697.53	759.79
1300	704.09	768.91	697.53	759.80
1400	704.18	768.93	697.62	759.91
1500	704.24	769.03	697.68	759.99
1600	704.24	768.99	697.69	760.02
1700	704.31	769.11	697.73	760.01
1800	704.35	769.15	697.77	760.14
1900	704.34	769.22	697.75	760.23
2000	704.40	769.27	697.81	760.23
2100	704.44	769.25	697.85	760.28
2200	704.51	769.33	697.90	760.32
2300	704.59	769.38	698.00	760.30
2400	704.66	769.47	698.09	760.47
2500	704.72	769.59	698.19	760.52
2600	704.84	769.69	698.30	760.72
2700	705.00	769.91	698.44	760.90
2800	705.25	770.13	698.63	761.09
2900	705.52	770.59	698.98	761.46
3000	705.95	771.01	699.30	761.93

Table G.3: V_{OUT} as V_{DD} increases, T=22°C, Chip #2

Architecture →	Variable	Variable	Constant	Constant
Bit Config. →	Min	Max	Min	Max
V_{DD} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓
600	586.37	592.20	592.39	599.59
700	648.73	690.42	650.68	676.96
800	664.04	709.52	662.75	714.32
900	664.55	713.34	663.58	718.27
1000	664.75	713.65	663.56	718.32
1100	664.89	713.63	663.53	718.40
1200	664.64	713.73	663.57	718.48
1300	664.79	713.68	663.63	718.59
1400	664.91	713.88	663.73	718.58
1500	665.01	713.92	663.76	718.63
1600	664.97	714.03	663.77	718.61
1700	665.07	713.90	663.80	718.61
1800	665.13	714.05	663.85	718.75
1900	665.07	714.07	663.84	718.79
2000	665.14	714.15	663.91	718.84
2100	665.22	714.25	663.94	718.87
2200	665.17	714.21	664.00	718.96
2300	665.24	714.34	664.03	719.12
2400	665.35	714.39	664.04	719.12
2500	665.48	714.58	664.16	719.25
2600	665.74	714.67	664.28	719.37
2700	665.80	714.87	664.56	719.54
2800	666.05	715.21	664.70	719.98
2900	666.40	715.61	665.07	720.33
3000	666.86	716.21	665.50	720.90

Table G.4: V_{OUT} as V_{DD} increases, T=22°C, Chip #3

Architecture →	Variable	Variable	Constant	Constant
Bit Config. →	Min	Max	Min	Max
V_{DD} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓
600	587.63	593.83	595.2	601.41
700	654.08	672.11	661.43	679.38
800	671.59	721.54	679.21	728.37
900	672.82	730.12	680.55	737.15
1000	672.88	730.51	680.65	737.63
1100	672.98	730.56	680.65	737.64
1200	673.00	730.60	680.72	737.77
1300	673.04	730.61	680.76	737.83
1400	673.04	730.57	680.74	737.86
1500	673.12	730.72	680.81	737.90
1600	673.17	730.76	680.91	737.89
1700	673.18	730.74	680.84	737.97
1800	673.21	730.77	680.88	737.97
1900	673.11	730.85	680.92	735.05
2000	673.26	730.88	680.95	738.00
2100	673.33	730.92	680.95	738.06
2200	673.33	730.94	681.05	738.15
2300	673.39	730.95	681.03	738.23
2400	673.41	731.07	681.15	738.30
2500	673.58	731.19	681.21	738.42
2600	673.62	731.31	681.30	738.48
2700	673.72	731.51	681.49	738.71
2800	674.07	731.70	681.69	738.91
2900	674.25	732.09	681.94	739.29
3000	674.61	732.59	682.31	739.69

Table G.5: V_{OUT} as V_{DD} increases, T=22°C, Chip #4

Architecture →	Variable	Variable	Constant	Constant
Bit Config. →	Min	Max	Min	Max
V_{DD} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓
600	583.76	594.23	587.07	599.61
700	658.01	676.11	648.82	675.66
800	679.41	724.59	659.53	709.95
900	680.96	735.08	660.33	714.84
1000	681.08	735.76	660.45	715.05
1100	681.18	735.82	660.41	715.08
1200	681.26	735.79	660.44	715.13
1300	681.22	735.94	660.45	715.23
1400	681.28	735.89	660.49	715.25
1500	681.28	736.01	660.51	715.28
1600	681.36	736.01	660.58	715.29
1700	681.37	736.07	660.59	715.3
1800	681.42	736.08	660.55	715.42
1900	681.45	736.15	660.71	715.41
2000	681.49	736.18	660.63	715.50
2100	681.55	736.19	660.73	715.51
2200	681.65	736.30	660.72	715.67
2300	681.61	736.32	660.72	715.64
2400	681.65	736.40	660.79	715.68
2500	681.77	736.46	660.93	715.79
2600	681.88	736.61	660.99	715.86
2700	682.00	736.78	661.13	716.09
2800	682.20	737.15	661.29	716.34
2900	682.59	737.41	661.66	716.72
3000	682.84	737.87	662.01	717.00

Table G.6: V_{OUT} as V_{DD} increases, T=22°C, Chip #5

Architecture →	Variable	Variable	Constant	Constant
Bit Config. →	Min	Max	Min	Max
V_{DD} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓
600	589.78	589.48	596.27	597.10
700	663.82	678.93	667.43	686.35
800	692.29	733.22	691.99	740.65
900	694.79	747.06	694.25	755.28
1000	694.87	747.79	694.38	756.10
1100	695.02	747.89	694.39	756.21
1200	695.05	747.87	694.60	756.29
1300	695.04	747.98	694.52	756.33
1400	695.11	748.02	694.55	756.42
1500	695.14	748.10	694.64	756.51
1600	695.23	748.13	694.65	756.45
1700	695.23	748.16	694.63	756.62
1800	695.33	748.09	694.71	756.61
1900	695.30	748.21	694.69	756.70
2000	695.35	748.34	694.77	756.76
2100	695.36	748.36	694.83	756.75
2200	695.42	748.38	694.89	756.80
2300	695.45	748.42	694.88	756.86
2400	695.55	748.54	694.99	756.90
2500	695.63	748.65	695.08	757.07
2600	695.68	748.78	695.25	757.25
2700	695.85	749.03	695.40	757.41
2800	696.18	749.15	695.53	757.59
2900	696.49	749.61	695.94	757.92
3000	696.90	750.13	696.30	758.48

G.2 Temperature-related Measurements

This section includes the raw data of measurements made over a temperature range from -20°C to 80°C with a supply voltage kept at 1.5 V. Details of the bit configuration are also included, where 6...1 represents the MSB...LSB.

Table G.7: V_{OUT} as temperature is changed, V_{DD}=1.5 V, Chip #1

Architecture →	Variable	Variable	Variable	Constant	Constant	Constant
Bit Config. →	Min	Max	Ideal (6,4,3,1)	Min	Max	Ideal (6)
T (°C) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{DD} (mV) ↓	V_{DD} (mV) ↓
-20	702.45	757.28	743.05	714.51	742.05	742.05
-10	n/a	n/a	n/a	n/a	n/a	n/a
0	n/a	n/a	n/a	n/a	n/a	n/a
10	699.66	759.36	743.54	712.75	775.2	743.72
20	698.62	759.53	743.74	712.47	775.86	744.23
30	696.36	760.42	743.63	711.17	777.77	744.54
40	694.54	761.43	743.64	710.01	779.00	744.75
50	692.38	762.02	743.84	709.04	780.15	744.94
60	690.62	762.75	743.85	708.03	782.44	745.06
70	688.60	763.08	743.86	706.60	783.12	745.04
80	686.38	762.52	744.06	705.33	784.14	744.56

Table G.8: V_{OUT} as temperature is changed, V_{DD}=1.5 V, Chip #2

Architecture →	Variable	Variable	Variable	Constant	Constant	Constant
Bit Config. →	Min	Max	Ideal (5,4,3,2)	Min	Max	Ideal (4,3)
T (°C) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{DD} (mV) ↓	V_{DD} (mV) ↓
-20	669.56	716.54	693.63	674.16	715.15	681.36
-10	n/a	n/a	n/a	n/a	n/a	n/a
0	n/a	n/a	n/a	n/a	n/a	n/a
10	665.16	718.10	692.12	672.96	719.63	681.46
20	664.48	718.69	691.85	672.73	721.39	681.55
30	662.42	719.79	692.26	671.93	722.95	682.19
40	660.77	721.10	692.09	671.67	725.14	682.15
50	659.31	722.08	692.07	671.46	727.31	682.33
60	658.18	723.06	692.13	671.18	729.30	682.44
70	656.44	723.80	692.05	670.41	731.25	682.36
80	654.33	724.10	691.84	669.55	732.20	681.72

Table G.9: V_{OUT} as temperature is changed, V_{DD}=1.5 V, Chip #3

Architecture →	Variable	Variable	Variable	Constant	Constant	Constant
Bit Config. →	Min	Max	Ideal (4)	Min	Max	Ideal (5)
T (°C) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{DD} (mV) ↓	V_{DD} (mV) ↓
-20	680.43	728.79	687.36	682.30	729.32	694.47
-10	n/a	n/a	n/a	n/a	n/a	n/a
0	n/a	n/a	n/a	n/a	n/a	n/a
10	680.86	735.57	688.69	681.36	736.85	695.69
20	680.85	737.56	689.12	680.73	738.11	695.74
30	680.55	739.99	689.27	679.61	740.17	695.83
40	680.39	742.74	689.31	679.22	742.51	695.85
50	680.04	744.96	689.44	678.44	744.54	695.85
60	679.66	747.28	689.48	677.64	746.24	695.90
70	679.12	749.12	689.25	676.77	748.16	695.66
80	678.01	750.18	688.67	675.32	749.29	694.78

Table G.10: V_{OUT} as temperature is changed, V_{DD}=1.5 V, Chip #4

Architecture →	Variable	Variable	Variable	Constant	Constant	Constant
Bit Config. →	Min	Max	Ideal (4,2)	Min	Max	Ideal (0)
T (°C) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{DD} (mV) ↓	V_{DD} (mV) ↓
-20	662.45	709.30	670.61	690.16	734.88	690.16
-10	n/a	n/a	n/a	n/a	n/a	n/a
0	n/a	n/a	n/a	n/a	n/a	n/a
10	661.34	713.50	670.47	690.24	741.53	689.15
20	660.65	714.90	670.46	689.16	743.15	689.07
30	659.84	717.26	670.36	688.63	745.32	688.76
40	659.64	718.77	670.48	688.63	747.96	688.65
50	659.45	720.76	670.57	688.54	750.56	688.73
60	659.07	722.90	670.88	688.63	753.05	688.85
70	658.89	724.96	670.84	688.59	755.47	688.83
80	658.12	726.35	670.34	688.28	757.58	688.55

Table G.11: V_{OUT} as temperature is changed, V_{DD}=1.5 V, Chip #5

Architecture →	Variable	Variable	Variable	Constant	Constant	Constant
Bit Config. →	Min	Max	Ideal (6,5,4,3,2,1)	Min	Max	Ideal (6,5,4,3,2,1)
T (°C) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{OUT} (mV) ↓	V_{DD} (mV) ↓	V_{DD} (mV) ↓
-20	701.81	757.24	757.24	711.24	757.05	757.05
-10	n/a	n/a	n/a	n/a	n/a	n/a
0	n/a	n/a	n/a	n/a	n/a	n/a
10	697.06	756.85	756.85	705.66	756.37	756.37
20	696.45	756.71	756.71	704.68	756.17	756.17
30	692.67	756.15	756.15	700.28	755.29	755.29
40	690.36	755.97	755.97	697.69	755.18	755.18
50	688.44	755.92	755.92	695.37	755.14	755.14
60	686.44	755.86	755.86	693.46	754.94	754.94
70	684.27	755.68	755.68	691.39	754.96	754.96
80	682.07	755.09	755.09	688.33	754.68	754.68

Appendix H

Statistical Analysis Verification

Verification of the statistical analysis was done by software and is shown in Fig. H.1.

```
#!/usr/bin/python
import numpy as np
# Generate 10000 random numbers
set = np.random.rand(10001,1)[:,0]
# Calculate median
med = np.median(set)
# repeat the results 100 times
for k in range(100):
    # repeat experiment many times
    success = 0
    times = 10000
    for j in range(times):
        # Draw n samples
        n = 4
        top = 0.
        bottom = 1.
        for i in range(n):
            x = set[int(np.random.rand() * 10001)]
            if x > top:
                top = x
            if x < bottom:
                bottom = x
        # Check success
        if (med > bottom) and (med < top):
            success += 1
# Print result
result = float(success)/times*100
print result, '%'
# NOTE: result approaches the calculated value
# as the number of times the experiment is
# repeated approaches infinity
```

Figure H.1: Python code used to verify statistical analysis

A set of 10,000 random numbers ranging from 0 to 1 were generated and the population median was found. 4 samples were randomly drawn from the set. The probability of the population median being within the minimum and maximum values of the 4 samples was then evaluated.

Bibliography

[1] C.W. Kok and W.S. Tam, *CMOS Voltage References: An Analytical and Practical Perspective*, Singapore, Wiley, 2013.

[2] A. Bakker, "CMOS smart temperature sensors - an overview," *IEEE Sensors, Proc.*, pp. 1423–1427 Vol. 2, 2002.

[3] R.J. Widlar, "New developments in IC voltage regulators," *IEEE J. of Solid-State Circuits,* Vol. 6, No. 1, pp. 2–7, Feb. 1971.

[4] J.S. Brugler, "Silicon transistor biasing for linear collector current temperature dependence," *IEEE J. of Solid-State Circuits,* Vol. SC-2, pp. 57–58, Jun. 1967.

[5] C. Kittel, *Introduction to Solid State Physics*, New Jersey, Wiley, 2005.

[6] E.A. Vittoz, and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operations," *IEEE J. of Solid-State Circuits,* Vol. 12, No. 3, pp. 224–231, Jun. 1977.

[7] E.A. Vittoz, and O. Neyroud, "A low-voltage CMOS bandgap reference," *IEEE J. of Solid-State Circuits,* Vol. 14, No. 3, pp. 573–577, Jun. 1979.

[8] A. Bansal, M.K. Raja, and J. Minkyu, "A 60nW voltage reference circuit generating 1.0V using BJTs and subthreshold MOSFET," *2012 IEEE Int'l Symposium on Radio-Frequency Integration Technology (RFIT), Proc.,* pp. 113–115, Nov. 2012.

[9] P.R. Gray, P.J. Hurst, S.H. Lewis and R.G. Meyer, *Analysis and Design of Analog Integrated Circuits,* New York, Wiley, 2001.

[10] A. Ortiz-Conde, F.J. Garcia Sanchez, and J.J. Liou, "On the extraction of threshold voltage, effective channel length and series resistance of MOSFETs," *J. of Telecom. and IT,* pp. 43–58, 2000.

[11] A. E. Buck, C. L. McDonald, S. H. Lewis, and T. R. Viswanathan, "A CMOS bandgap reference without resistors," *IEEE J. of Solid-State Circuits,* Vol. 37, No. 1, pp. 81–83, Jan. 2002.

[12] G. De Vita and G. Iannaccone, "A sub-1 V, 10 ppm/°C, nanopower voltage reference generator," *IEEE J. of Solid-State Circuits,* Vol. 42, No. 7, pp. 1536–1542, Jul. 2007.

[13] K. Ueno, T. Hirose, T. Asai and Y. Amemiya, "A 300 nW, 15 ppm/°C, 20 ppm/V CMOS voltage reference circuit consisting of subthreshold MOSFETs," *IEEE J. of Solid-State Circuits,* Vol. 44, No. 7, pp. 2047–2054, Jul. 2009.

[14] L. Magnelli, F. Crupi, P. Corsonello, C. Pace, and G. Iannaccone, "A 2.6 nW, 0.45 V temperature-compensated subthreshold CMOS voltage reference," *IEEE J. of Solid-State Circuits,* Vol. 46, No. 2, pp. 465–474, Feb. 2011.

[15] M. Seok, G. Kim, D. Sylvester, and D. Blaauw, "A portable 2-transistor picowatt temperature-compensated voltage reference operating at 0.5 V," *IEEE J. of Solid-State Circuits,* Vol. 47, No. 10, pp. 2534–2545, Oct. 2012.

[16] H.J. Oguey and D. Aebischer, "CMOS current reference without resistance," *IEEE J. of Solid-State Circuits,* Vol. 32, No. 7, pp. 1132–1135, Jul. 1997.

[17] E.M. Camacho-Galeano, C. Galup-Montoro, and M.C. Schneider, "A 2-nW 1.1-V self-biased current reference in CMOS technology," *IEEE Trans. on Circuits and Systems II: Express Briefs,* Vol. 52, No. 2, pp. 61–65, Feb. 2005.

[18] C. Rossi-Aicardi, J. Oreggioni, F. Silveira and C. Dualibe, "A MOSFET-only voltage source with arbitrary sign adjustable temperature coefficient," *2011 IEEE NEWCAS Conf. Proc.,* pp. 366–369, Jun. 2011.

[19] C. Rossi-Aicardi, "Techniques for ultra low power integrated temperature sensors," *Ph.D. Thesis,* Universidad de la Republica: Uruguay, Feb. 2013.

[20] A.B. Sproul, M.A. Green, and J. Zhao, "Improved value for the silicon intrinsic carrier concentration at 300 K," *J. Appl. Phys.,* Vol. 70, No. 2, pp. 846–854, Jul. 1991.

[21] C. Fonstad, "6.012 Microelectronic Devices and Circuits," *MIT OpenCourseWare: Massachusetts Institute of Technology,* lec. 12, slide 28, 2009.

[22] E. A. Vittoz, and J. Fellrath, "Low-power design: ways to approach the limits," *1994 IEEE Int'l Solid-State Circuits Conference, Proc.,* pp. 14–18, Feb. 194.

[23] C. Rossi, C. Galup-Montoro, and M.C. Schneider, "PTAT voltage generator based on an MOS voltage divider," *Technical Proceedings of the 2007 NSTI Nanotechnology Conference and Trade Show,* Santa Clara, CA, USA, pp. 625–628, May 2007.

[24] A.I.A. Cunha, M.C. Schneider, and C. Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. of Solid-State Circuits,* Vol. 33, No. 10, pp. 1510–1519, Oct. 1998.

[25] C. Galup-Montoro and M.C. Schneider, *MOSFET modeling for circuit analysis and design,*, World Scientific Pub. Co., 2007.

[26] B. Rasavi, *Design of Analog CMOS Integrated Circuits,* New York, McGraw-Hill, 2001.

[27] D. Megaw, "Voltage Reference Selection Basics," *POWER|designer,* Texas Instruments, Lit. No. SNVA602, 2011.

[28] B. Gilbert, "Translinear circuits: a proposed classification," *Electronics Letters,* Vol. 11, No. 1, pp. 14–16, Jan. 1975.

[29] H. Schmid and A. Huber, "Measuring a small number of samples and the 3σ fallacy," *IEEE Solid-State Circuits Magazine,* Vol. 6, No. 2, pp. 52–58, Spring 2014.

[30] M.C. Schneider and C. Galup-Montoro, *CMOS Analog Design Using All-Region MOSFET Modeling,* New York, Cambridge University Press, 2010.

[31] I.M. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications,* Vol. 48, No. 7, pp. 876–884, Jul. 2001.

[32] K. Chen, H.C. Wann, J. Dunster, P.K. Ko, C. Hu, and M. Yoshida, "MOSFET carrier mobility model based on gate oxide thickness, threshold and gate voltages," *Solid-State Electronics,* Vol. 39, Issue 10, pp. 1515–1518, Oct. 1996.

[33] P. Aguirre and F. Silveira, "Bias circuit design for low-voltage cascode transistors," *19th annual symposium on Integrated circuits and systems design, Proc.,* ACM, 2006.

[34] C. Putnam, R. Gauthier, M. Muhammad, K. Chatty, and M. Woo, "Human body model ESD protection concepts in SOI and bulk CMOS at the 130 nm node," *2003 IEEE Int'l SOI Conference, Proc.,* pp. 23–25, 2003.

[35] V. d'Alessandro, and P. Spirito, "Achieving accuracy in modeling the temperature coefficient of threshold voltage in MOS transistors with uniform and horizontally nonuniform channel doping," *Solid-State Electronics,* Vol. 49, issue 7,Jul. 2005.