UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROELETRÔNICA

JHON ALEXANDER GÓMEZ CAICEDO

CMOS Low-Power Threshold Voltage Monitor Circuits and Applications

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Microeletronics

Advisor: Prof. Dr. Hamilton Klimach
Coadvisor: Prof. Dr. Eric Fabris

Porto Alegre
May 2016

CIP – CATALOGING-IN-PUBLICATION

Gómez Caicedo, Jhon Alexander

CMOS Low-Power Threshold Voltage Monitor Circuits and Applications / Jhon Alexander Gómez Caicedo. – Porto Alegre: PGMICRO da UFRGS, 2016.

83 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Microeletrônica, Porto Alegre, BR–RS, 2016. Advisor: Hamilton Klimach; Coadvisor: Eric Fabris.

1. Threshold voltage. 2. CMOS analog design. 3. Threshold voltage monitor circuit. 4. High-PSRR. 5. Resistorless. 6. Ultra-low-power. 7. Voltage Reference. 8. Process Compensation. I. Klimach, Hamilton. II. Fabris, Eric. III. Título.

”We leave something of ourselves behind when we leave a place, we stay there, even though we go away. And there are things in us that we can find again only by going back there."

— PASCAL MERCIER

ACKNOWLEDGEMENTS

I would like to say thank to all the people that directly or indirectly collaborated in the development of this thesis, but first, I must say thank to professor Hamilton Klimach, that is the person whom me and this thesis owe him more. Thank for all the advices, suggestions, corrections, and especially, for all things that you taught me. Also, I thank to professors Eric Fabris and Sergio Bampi, for all the discussions that help me to mature and learn in the ICs field.

To the NSCAD team, for all our technical discussions, exchanges of knowledge, and to allow me to use their installations to work in my master. To my colleagues of the 110 lab, whom more than colleagues are friends.

A special thank, to my "novia" who supported me in turbulent moments, and reminded me, who I am, and why I am here. Also, to my dear friend Diogo who always was available to share a beer.

And finally to my family that with all their love, motivates me, strengthens me, and have conducted me to this moment.

ABSTRACT

A threshold voltage (V_{T0}) monitor is a circuit that ideally delivers the estimated V_{T0} value as a voltage at its output, for a given temperature range, without external biases, parametric setups, curve fitting or any subsequent calculation. It can be used in temperature sensors, voltage and current references, radiation dosimeters and other applications since the MOSFET V_{T0} dependence on the operation conditions is a very well modeled aspect. Also, it can be used for fabrication process monitoring and process variability compensation, since V_{T0} is a key parameter for the transistor behavior and modeling.

In this thesis, we present three novel circuit topologies, two of them being NMOS V_{T0} monitors and the last one being a PMOS V_{T0} monitor. The three structures are resistorless self-biased circuit topologies that present high power supply rejection, low line sensitivity, and allow the direct extraction of the threshold voltage for wide temperature and power supply voltage ranges, with small error. Its design methodology is based on the Unified Current Control Model (UICM), a MOSFET model that is continuous from weak to strong inversion and from triode to saturation regions. The circuits occupy small silicon area, consume just tens of nanoWatts, and can be implemented in any standard digital CMOS process, since they only use MOS transistors (does not need any resistor).

The V_{T0} monitors are used in different applications in order to prove their functionality, and behavior as part of a system. The applications vary from a reference voltage, that presents performance comparable with state-of-the-art works, to a configuration that allows to obtain a lower process variability, in the output of a self-biased circuit that generates a complementary to the absolute temperature (CTAT) voltage. In addition, exploiting the ability to operate as an specific current (I_{SQ}) generator, that the V_{T0} monitors presented here offer, we introduced a new self-biased circuit that produces a CTAT voltage and is less sensitive to process variations, and can be used in band-gap voltage references.

Keywords: Threshold voltage, CMOS analog design, Threshold voltage monitor circuit, High-PSRR, resistorless, ultra-low-power, Voltage Reference, Process Compensation.

Circuitos Monitores de Tensão de Limiar CMOS de baixa potência e Aplicações

RESUMO

Um monitor de tensão de limiar (V_{T0}) é um circuito que, idealmente, entrega o valor do V_{T0} como uma tensão na saída, para uma determinada faixa de temperatura, sem a necessidade de polarização externa, configurações paramétricas, ajuste de curvas ou qualquer cálculo subsequente. Estes circuitos podem ser usados em sensores de temperatura, referências de tensão e corrente, dosímetros de radiação e outras aplicações, uma vez que a dependência do V_{T0} nas condições de operação é um aspecto bem modelado. Além disso, estes circuitos podem ser utilizados para monitoramento de processos de fabricação e para compensação da variabilidade do processo, uma vez que o V_{T0} é um parâmetro chave para o comportamento do transistor e sua modelagem.

Nesta tese, são apresentadas três novas topologias de circuitos, duas são monitores de V_{T0} NMOS e a terceira é um monitor de V_{T0} PMOS. As três estruturas são topologias de circuito auto-polarizadas que não utilizam resistências, e apresentam alta rejeição a variações na alimentação, baixa sensibilidade de Linea, e permitem a extração direta da tensão de limiar para grandes intervalos de temperatura e de tensão de alimentação, com pequeno erro. Sua metodologia de projeto é baseada no modelo unificado controlado por corrente (UICM), um modelo MOSFET que é contínuo, desde o nível de inversão fraca a forte e para as regiões de operação de triodo e saturação. Os circuitos ocupam uma pequena área de silício, consomem apenas dezenas de nanowatts, e podem ser implementados em qualquer processo padrão CMOS digital, uma vez que só utilizam transistores MOS (não precisa de nenhum resistor).

Os monitores de V_{T0} são utilizados em diferentes aplicações, a fim de investigar a sua funcionalidade e comportamento como parte de um sistema. As aplicações variam de uma tensão de referência, que apresenta um desempenho comparável ao estado da arte, para uma configuração que permite obter uma menor variabilidade com processo na saída de um circuito auto-polarizado que gera um tensão CTAT. Além disso, explorando a capacidade de funcionar como um gerador de corrente específica (I_{SQ}) que os monitores de V_{T0} aqui apresentados oferecem, introduz-se um novo circuito auto-polarizado que gera um tensão CTAT, que é menos sensível a variações de processo, e pode ser usado em referências de tensão band-gap.

Palavras-chave: projeto analógico CMOS, Monitor de tensão de limiar, referência de tensão, ultra-baixo consumo, baixa alimentação.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BGR	Bandgap Voltage Reference
CMOS	Complementary Metal Oxide Semiconductor
CTAT	Complementary to Absolute Temperature
DUT	Device Under Test
LS	Line Sensitivity
MI	Moderate Inversion
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
PSR	Power Supply Rejection
PTAT	Proportional to Absolute Temperature
SI	Strong Inversion
TC	Temperature Coefficient
VR	Voltage Reference
WI	Weak Inversion

CONTENTS

1 INTRODUCTION

The semiconductor industry was a 336 billion dollar market in 2014 (World Semiconductor Trade Statistics, 2015), and it has registered a constant growth in the last decades, solely in the U.S it was the industry with the largest growth in the last 35 years (Semiconductor Industry Association, 2014). This turns the semiconductor industry in one of the most prominent industries in the world. In addition, with the increasing of the Internet of things (IoT) and the necessity of the global connectivity, the integrated circuits (ICs) are ever more present in all aspects of human life, from heart monitors to space navigation systems.

In the heart of this industry is found the transistor, the main element in ICs, that year after year decreases its minimum size thanks to the lithography advances, thus allowing to integrate more elements per area. This increase in the quantity of transistors per area, allows to implement in a smaller area the same circuits that were fabricated before, thereby reducing cost. Another possibility is to add more features to the original system, maintaining the area of previous versions; this results in a more advanced system at the same cost of previous one.

This dissertation discuss one of the most important parameters of the transistor, the threshold voltage (V_{T0}). This chapter will give the necessary context to understand the developed work. First, the V_{T0} is introduced, together with its definition and the method to extract it. After that, the importance of the V_{T0} in microelectronics will be highlighted, for this some applications will be exposed, and then the V_{T0} monitor will be explained. Next, the objectives of this work will be presented and finally the organization of the text will be detailed.

1.1 Threshold Voltage

Few months later of the first successful fabrication of a MOSFET the first model based on V_{T0} appeared. This model, proposed by IHANTOLA; MOLL (1964), inspired during decades the development of many MOSFET models for SPICE simulation, such as level = 1,2,3 and BSIM. There is no surprise, therefore, that all the models that followed it inherited the V_{T0} as a fundamental parameter (ORTIZ-CONDE et al., 2013).

Since the first MOSFET model that used the V_{T0} was developed, there have appeared many models with different approaches and different definitions of the V_{T0}. PAO; SAH (1966) proposed the first surface potential model, which was based on a double-integration. This model establishes the basis of contemporary compact models. It also inspired BREWS (1978) and BACCARANI; RUDAN; SPADINI (1978) whose independently formulated charge-sheet models. This diversity in the models makes difficult to obtain a unified definition of the V_{T0} and the same happens with the methods to extract it.

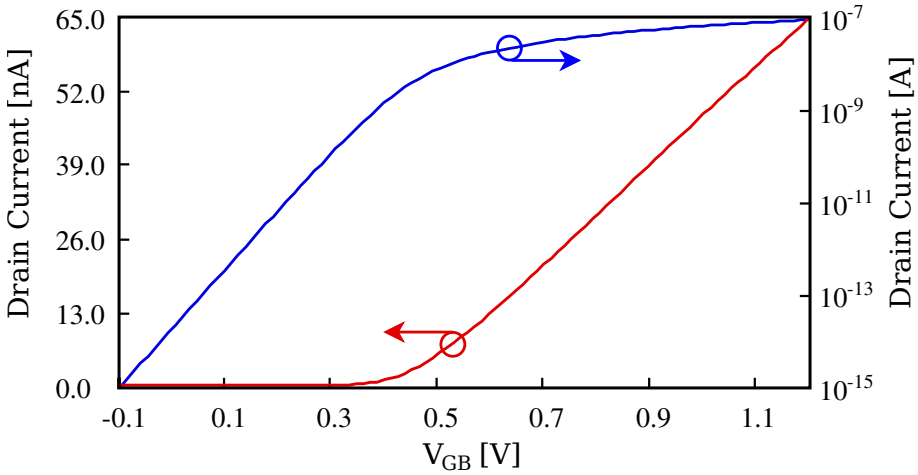
In this work the UICM model was used for its simplicity and physical sustentation. This model, which is detailed in appendix B, takes advantages of the charge-sheet model, and the incrementally linear relationship between the inversion charge density and the surface potential. Next, a definition of the V_{T0} using the UICM model is delivered as well as a method to obtain its theoretical value.

1.1.1 Definition

The threshold voltage can be understood as the voltage that must be applied in the gate of the transistor (V_G) in order to produce a behavioral change in the drain current (I_D) of the transistor. At the highest level of abstraction, this behavioral change refers to the voltage in which the current begins to flow. This definition even though is widely known and useful, especially for digital design, is not suitable for analog design, as will be explained.

When I_D is plotted as a function of V_G (Figure 1.1), it seems that in fact there is a voltage in which the current starts to flow; however, when the current is plotted in a logarithmic scale it is clear that the current actually exists for voltages below that point. As even small currents are important in analog design, this high-level definition of the V_{T0} results improper. The logarithmic scale also allows to see that I_D presents two well defined behaviors. First, I_D behaves exponentially with respect to V_G for small voltages (weak inversion) while, for larger voltages it behaves quadratically (strong inversion).

Figure 1.1: I_{DS} vs V_G in different scales.

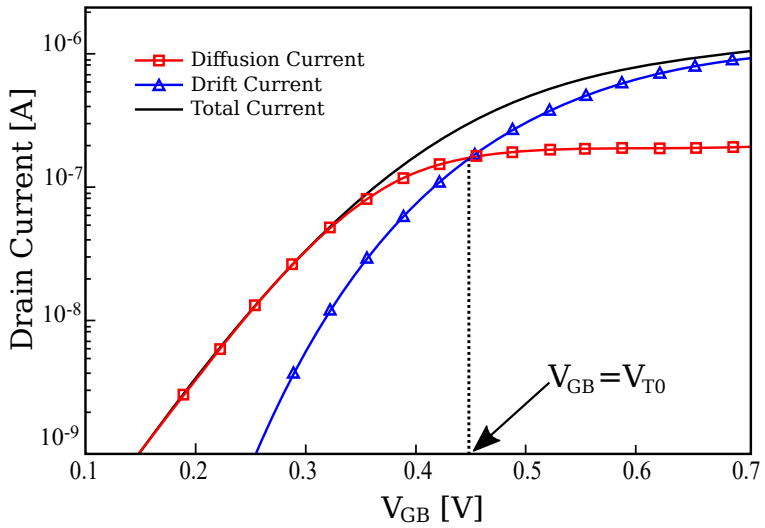


Source: The own author.

The change in the I_D behavior is because the physical process that generates the current for large and small values of V_G is different. As shown in Figure 1.2, the current in weak inversion is due to carrier diffusion, whereas the strong inversion current is mostly due to carrier drift (SIEBEL; SCHNEIDER; GALUP-MONTORO, 2012). This change in the behavior, which has a physical foundation, is used in order to create a more proper V_{T0} definition. Then, as shown in Figure 1.2, the V_{T0} is the voltage in which the drift and diffusion component of the current are equal.

Defined V_{T0} together with its physical meaning, it is possible to link the aforementioned concepts with the UICM model. This model defines I_D as a function of the forward

Figure 1.2: I_{DS} and its diffusion and drift components vs V_G for a MOSFET with V_{DS}=$\phi_t/2$.



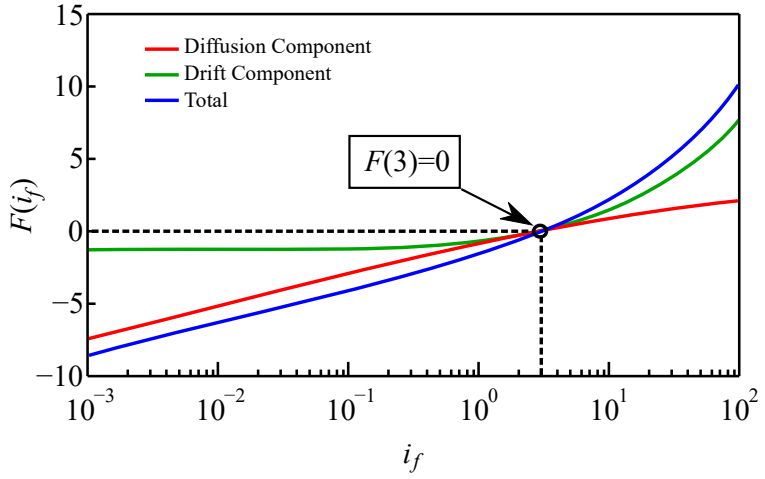
Source: SIEBEL; SCHNEIDER; GALUP-MONTORO (2012).

and reverse inversion levels ($F(i_f)$ and $F(i_r)$, respectively). The relationship between these inversion levels and the terminal voltages is given by

$$\frac{V_G - V_{T0}}{n\phi_t} - \frac{V_{S(D)}}{\phi_t} = F(i_{f(r)}) = \underbrace{\sqrt{1+i_{f(r)}} - 1}_{\text{Drift}} + \underbrace{\ln(\sqrt{1+i_{f(r)}} - 1) - 1}_{\text{Diffusion}} \tag{1.1}$$

where V_S and V_D are the source and drain voltages (all terminal voltages are referenced to the transistor bulk in the UICM model), ϕ_t is the thermal voltage and n is the subthreshold slope factor. From (1.1) it is possible to see that a transistor with the source grounded and with V_G equal to V_{T0} has $F(i_f)$ equal to zero. This, as expected, happens when the drift and diffusion components are equal, and as shown in Figure 1.3 occurs for an i_f equal to 3. Eq. (1.1) does not consider short-channel effects being valid only for long channel devices.

Figure 1.3: $F(i_f)$ and its drift and diffusion components vs i_f.



Source: The own author.

1.1.2 Extraction Method

Just as there are many V_{T0} definitions, exist countless number of methods to extract the V_{T0} value. ORTIZ-CONDE et al. (2013) revised the most used methods and summarized about 12 works. As the V_{T0} was defined physically and through UICM model, only methods using this concepts are shown here. In the constant current (CC) method the MOSFET operates in the saturation region and it must be biased with a specific value of I_D in order to obtain an accurate value of V_{T0}. The CC method is the most direct procedure but presents the disadvantage of requiring a previous estimation of the sheet normalization transistor current (I_{SQ}). On the other hand, in the g_{ch}/I_D and g_m/I_D methods, the MOS transistor operates in the linear region. These methods are similar and present a comparable accuracy; however, as the transconductance g_m is a more exploited parameter in comparison with the channel transconductance g_{ch}, it was preferred to use the g_m/I_D method in order to extract the theoretical value of V_{T0}.

The aim of the g_m/I_D extraction method is to obtain the V_{T0} value from a g_m/I_D vs V_G graph. For this, first it is necessary to get the value of the g_m/I_D relationship for the threshold condition ($i_f = 3$). To do this, it is necessary to give some definitions from the UICM model. First, the source and drain transconductances are be given by

$$g_{ms(d)} = \frac{2I_S}{\phi_t}(\sqrt{1 + i_{f(r)}} - 1) \tag{1.2}$$

where I_S is the specific current of the transistor. Then, using (B.1), (1.1), (1.2) and remembering that g_m is expressed mathematically as the derivative of I_D with respect to V_G, one obtains

$$\frac{g_m}{I_D} = \frac{dI_D}{I_D dV_G} = \frac{g_{ms} - g_{md}}{nI_D} = \frac{2}{n\phi_t(\sqrt{1 + i_f} + \sqrt{1 + i_r})} \tag{1.3}$$

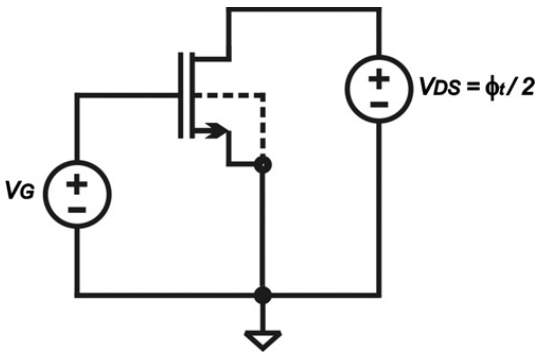
If it is assumed that the variation of n with V_G is small (n approximately constant for any i_f), it is possible to get the g_m/I_D value from (1.3) for a given i_f and i_r. Then, as i_f is set for the threshold condition, just rest to obtain the value of i_r, which can be obtained from (1.1) once V_D is set in $\phi_t/2$ in order to maintain the transistor in the linear region. Therefore, $i_r = 2.31$, and $g_m/I_D = 0.531 * (2/n\phi_t) = 0.531 * (g_m/I_D)_{\text{max}}$.

To apply the g_m/I_D extraction method the test-bench of Figure 1.4 is implemented. In this, V_{DS} is fixed at the desired bias condition ($\phi_t/2$), and V_G varies from 0 to 1.2 V. The current is saved to obtain the g_m/I_D curve. Finally with the maximum value of the g_m/I_D relationship the point that corresponds to the threshold voltage condition is found. In Figure 1.5 an example for a 130 nm NMOS transistor I/O type of the IBM process with W=2 µm and L=2 µm is given. The transistor presents a $(g_m/I_D)_{\text{max}} = 28.98$, and therefore for the point in which (g_m/I_D)=15.38 it is found $V_G = V_{T0} = 481.15$ mV.

1.2 Threshold Voltage Monitor

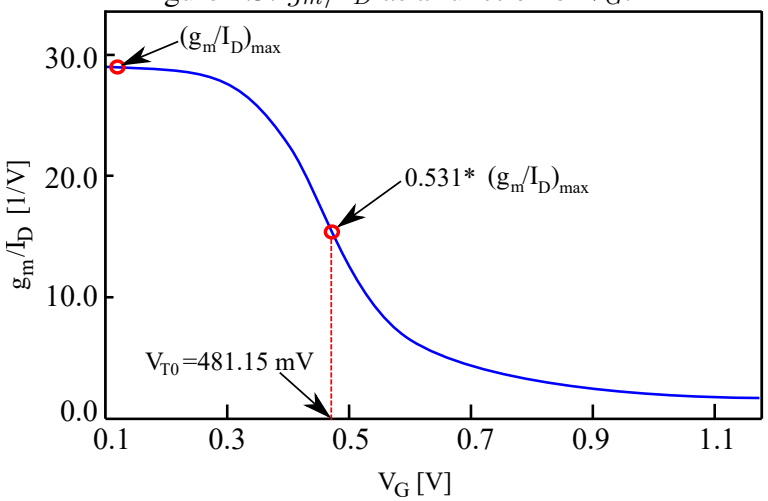
The threshold voltage is such an important parameter in almost all the MOSFET models that accurate values of it for various geometries and bias conditions have to be determined to characterize the transistor for optimizing circuit design (TSAY; LIU; WU, 1996). Threshold voltage is also used to monitor the success of the fabrication process and for assessing and predicting device performance variability, because it depends sensitively on the device geometries, the gate-oxide thickness and the channel doping, as well as other

Figure 1.4: Circuit configuration for measuring the g_m/I_D characteristic in the linear region.



Source: SIEBEL; SCHNEIDER; GALUP-MONTORO (2012).

Figure 1.5: g_m/I_D as a function of V_G.



Source: The own author.

operation reliability factors (ORTIZ-CONDE et al., 2013). Thus, the measurement of V_{T0} is a fundamental step in process characterization and circuit design.

Other uses for the V_{T0} measurement are the compensation of process parameters, the cancellation itself in I –V and V–I conversion circuits, the production of correct bias conditions (FIKOS; SISKOS, 2001) and, lately, in the implementation of MOSFET radiation dosimeters. Also taking advantage of its excellent linearity with respect to temperature, V_{T0} can be used to do on-chip absolute or relative temperature sensing and full CMOS voltage references (CILINGIROGLU; HOON, 2003).

This range of applications makes the V_{T0} measurement an area of large interest. Traditionally, the threshold voltages of MOSFETs are extracted either from graphical methods or from numerical methods (ORTIZ-CONDE et al., 2013; THOMAS; HOLMAN, 1999); unfortunately, these methods are highly time-consuming and/or computationally intensive and cannot be suitably used for real-time on-chip applications (DASGUPTA; SAHA; SAMANTA, 2012).

These problems can be overcome by using a threshold voltage monitor, which is a circuit that ideally delivers the estimated V_{T0} value as a voltage at its output, for a given temperature range, without external biases, parametric setups, curve fitting or any subsequent calculation. As was cited above, the V_{T0} measurement can be used in countless

applications. This makes the V_{T0} monitor a multipurpose circuit, and therefore, it must present high performance in order to maintain the total yield of a system. The most important aspects in V_{T0} monitors are:

- Large range of temperature operation
- High Power Supply Rejection (PSR)
- Low Line Sensitivity (LS)
- Small error
- Low power consumption
- Include a Start-up stage

1.3 Objectives

Based on the wide range of applications in which the V_{T0} monitors can be used, and in the fact that none of the available topologies presents the minimum performance to be used as multi-purpose circuits, it can be concluded that the study and design of these circuits is of great interest. This thesis aims the implementation of PMOS and NMOS V_{T0} monitors that present high performance, making possible its integration in different systems. In addition, this thesis intends to develop some circuits using the proposed V_{T0} monitors, this in order to prove the behavior of the monitors as part of a system.

1.4 Organization

The text is organized as follows. Chapter 2 makes a summarized bibliography review in chronological order of V_{T0} monitors, and in the end of this section, a comparison between state-of-the-art works is given. After that, Chapter 3 shows the circuit analysis, circuit design, and post-layout results of the proposed NMOS and PMOS V_{T0} monitors. It also introduces the I_{SQ} generation, a concept used in the PMOS V_{T0} monitor implementation. Chapter 4 presents some applications of the proposed V_{T0} monitors, and finally, the conclusions and considerations for future work are included in Chapter 5. Some additional information, as the UICM model, can be found in the Appendix at the end of this text.

2 THRESHOLD VOLTAGE MONITORS OVERVIEW

As was related in the last chapter, there are a lot of applications in which a V_{T0} monitor can be used. This versatility makes this circuit a structure of great interest, and has motivated through the years many authors to propose several topologies. First designs, in general, operate in strong inversion and use quadratic model. These designs were useful solutions at that time, but now they represent low-efficient circuits, since the strong inversion operation results in high power consumption. Recent works use the transistors in moderate and/or weak inversion, thus achieving low power consumption. Unfortunately, these circuits still present some problems associated to its integration, as it will be explained later on. In this chapter, we summarize chronologically the early designs in V_{T0} monitors, then we expose the most recent advances in the area, and finally we summarize and compare the most important works.

2.1 Early Designs

The first work related to the extraction of the V_{T0} value using a circuit was published by TSIVIDIS; ULMER (1979). In Figure 2.1 it is shown the proposed circuit. In this structure, the bias current I_R of the centre string is scaled, and mirrored by the PMOS current mirror (M_G-M_H) and the NMOS cascode current mirror (M_A-M_D). Since all the transistors that were used for this design operate in strong inversion, and in the saturation region, their drain current can be approximately given by:

$$I = \frac{1}{2} Sk(V_{GS} - V_{T0})^2 \tag{2.1}$$

where S is the aspect ratio of the transistor W/L (being W and L the width and length of the transistor respectively), and $k = \mu_n C'_{ox}$. Considering the remaining transistors and (2.1), we obtain an expression for the output voltage V_X,

$$V_X = m\left(V_{T0} + \sqrt{\frac{2I_E}{S_E k_E}}\right) - \left(V_{T0} + \sqrt{\frac{2I_F}{S_F k_F}}\right) \tag{2.2}$$

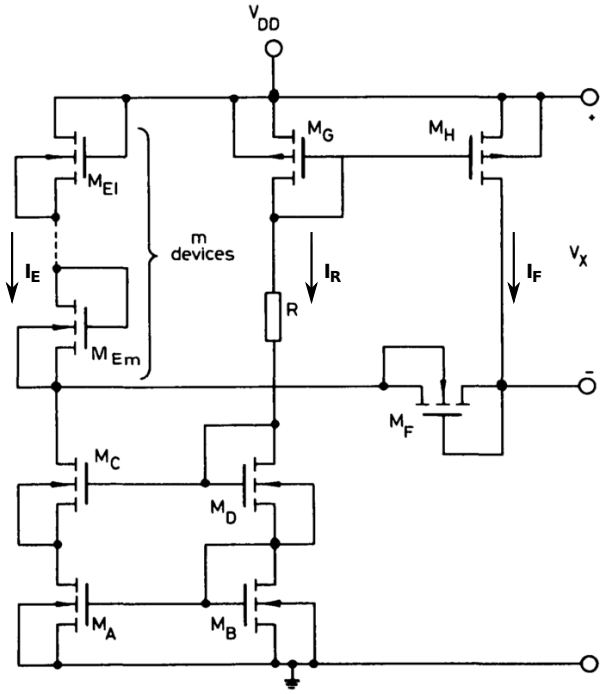
where I_E and k_E are the common values of I and k of the transistors M_{E1}-M_{Em}, and I_F and k_F are the values of I and k of the transistor M_F. Now, if the circuit is biased and sized in order to maintain the relation $I_F/I_E = m^2 S_F/S_E$ it is obtained that V_X is equal to:

$$V_X = (m-1)V_{T0} \tag{2.3}$$

It is clear that a voltage equal ($m = 2$) or proportional to V_{T0} can be obtained. In the same paper, Tsividis presents a second structure for which the same equations are valid.

The second circuit presents the advantage of delivering its output as a voltage referenced to the ground and not to V_{DD}. The author also comments about the possibility of the implementation of a PMOS V_{T0} monitor if a triple-well process is available. This circuit was for almost 10 years the state-of-the-art, although it presented at its output an error higher than 10%.

Figure 2.1: Schematic of the circuit proposed by TSIVIDIS; ULMER



Source: TSIVIDIS; ULMER (1979).

ALINI et al. (1992) proposed the circuit shown in Figure 2.2. The main idea is to bias with the same current (I_{do}) two transistors, M1 and M2, which present an aspect ratio relationship

$$S_2/S_1 = \alpha^2 \tag{2.4}$$

Using (2.1) and (2.4), the relationship between the overdrive voltages of M1 and M2 can be written as:

$$V_{ov1} = \alpha V_{ov2} \tag{2.5}$$

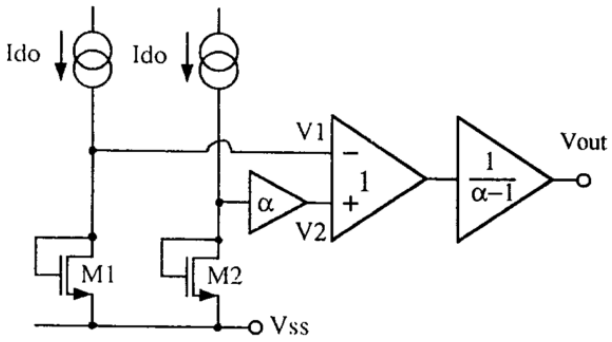
Then, if the V_{GS} of transistor M2 is amplified by α, it is obtained in the output of the subtractor:

$$V_2 - V_1 = \alpha(V_{ov2} + V_{T0}) - (V_{ov1} + V_{T0}) = (\alpha - 1)V_{T0} \tag{2.6}$$

Finally, after amplifying by a factor of 1/(α-1), the output of the circuit $V_{out} = V_{T0}$ is obtained. This implementation, even though presents an error relatively low (4%), has a big complexity due to the amplifiers and the subtractor, which in the case of Alini were implemented using BJT transistors. Another problem intrinsic to this structure is, that any mismatch between the bias current, or the M1-M2 transistors, results in an error in the V_{T0} value.

WANG (1992) found a similar solution to that presented by Alini, but instead of biasing two transistors with a current source, he generates the current by biasing the gate of

Figure 2.2: Schematic of the circuit proposed by ALINI et al.



Source: ALINI et al. (1992).

one of the transistors, and then, bias the other one by using a PMOS current mirror. In this way, he does not need a current source, but instead, he needs a bias voltage.

The circuit is shown in Figure 2.3, and works as follows. A voltage V_B bias the transistor M_0, which generates a current. This current is copied by a PMOS current mirror, into an array of nxm transistors (Figure 2.4). Assuming that all the transistors are equal, and using the quadratic model, we can obtain:

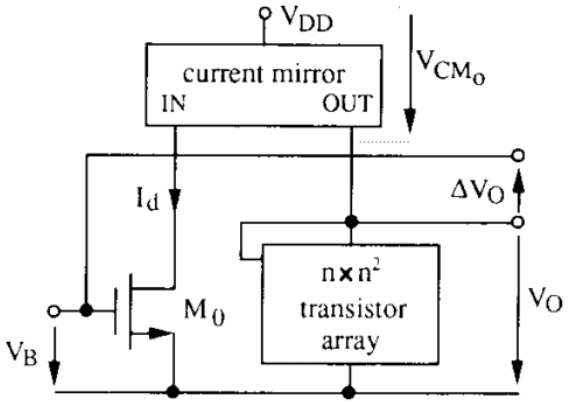
$$V_O = \frac{n}{\sqrt{m}} V_B + n\left(1 - \frac{1}{\sqrt{m}}\right) V_{T0} \tag{2.7}$$

If $m = n^2$ the expression is simplified to $V_O = V_B + (n-1)V_{T0}$. Now, taking the difference between V_B and the V_O, we obtain

$$\Delta V_O = V_B - V_O = (n-1)V_{T0} \tag{2.8}$$

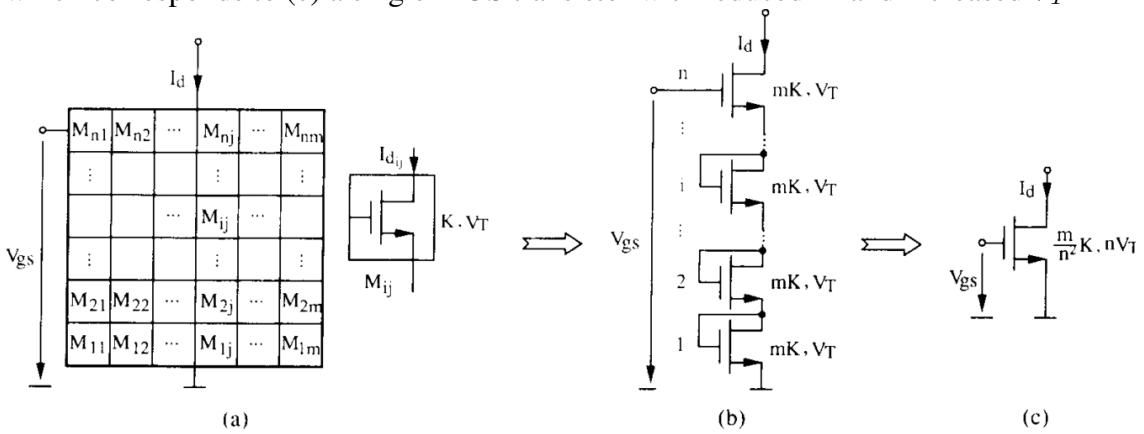
Then, an array of 2×4 transistors leads in the output, a voltage that is equal to the V_{T0} value. This implementation has the advantage of using transistors of the same size, which improves the matching of the transistors and also allows to implement common centroid structures. On the other hand, the structure presents the problem of using an additional voltage, as was above mentioned, and additionally, the circuit does not deliver the V_{T0} value referenced to ground but as a differential voltage.

Figure 2.3: Schematic of the circuit proposed by WANG



Source: WANG (1992).

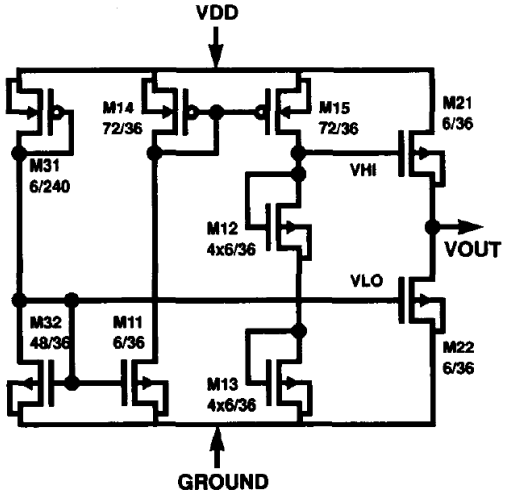
Figure 2.4: (a) A transistor array with a size of $n \times m$ and (b) its equivalent cascode circuit which corresponds to (c) a single MOS transistor with reduced K and increased V_T



Source: WANG (1992).

Figure 2.5 shows the V_{T0} monitor proposed by JOHNSON (1993). This topology presents two main differences with Wang's topology. First, a stage to bias transistor M11 (M_0 in Wang's circuit) was added, thus resolving the necessity for an additional voltage, and turning the circuit a self-bias structure. The second difference is the two-transistor differential amplifier added to the output. This stage takes the differential voltage (ΔV_O in Figure 2.3 or VHI-VLO in Figure 2.5) and delivers it as a ground-referenced voltage.

Figure 2.5: Schematic of the circuit proposed by JOHNSON



Source: JOHNSON (1993).

Equations (2.7) and (2.8) are valid also for this circuit, but for simplicity, the last one is repeated here, taking into account the nomenclature of Figure 2.5.

$$VHI - VLO = V_{T0} \tag{2.9}$$

Now, assume that transistor M21 and M22 are identical and that both are in saturation. Writing the V_{GS} voltages for these transistors, $V_{GS21} = VHI - VOUT$, and also $V_{GS22} = VLO$. Since M21 and M22 carry the same current, they must also have identical V_{GS}, thus:

$$VOUT = VHI - VLO = V_{T0} \tag{2.10}$$

This topology was known as an input-free V_{T0} monitor, because different from previous works, it does not need any voltage or current additionally to the supply voltage which is a big advantage. Nonetheless, the circuit cannot be implemented in any CMOS standard process, since it does not use a common voltage to bias the bulk of all the NMOS transistors.

Until now, all the structures are based on two transistors with a known aspect ratio relationship that are biased with the same current. These structures, in general, present the problem that any mismatch between the two transistors will directly affect the V_{T0} value delivered. YU; GEIGER (1994) proposed a solution to this problem by using only one transistor, and a switched capacitor implementation. The circuit is shown in Figure 2.6. In this circuit, the output of a current mirror I_{D1} (with S1 closed and S2 open) and I_{D2} (with S1 open and S2 closed) are applied to a test transistor, which operates in the saturation region. Using the quadratic model (2.1), we obtain:

$$I_{D1} = \frac{1}{2}Sk(V_{GS1} - V_{T0})^2 \tag{2.11}$$

$$I_{D2} = \frac{1}{2}Sk(V_{GS2} - V_{T0})^2 \tag{2.12}$$

Observe that k, S and V_{T0} are the same in (2.11) and (2.12) because only one test transistor is used. Then, using the relation $I_{D1} = nI_{D2}$, it is possible to obtain:

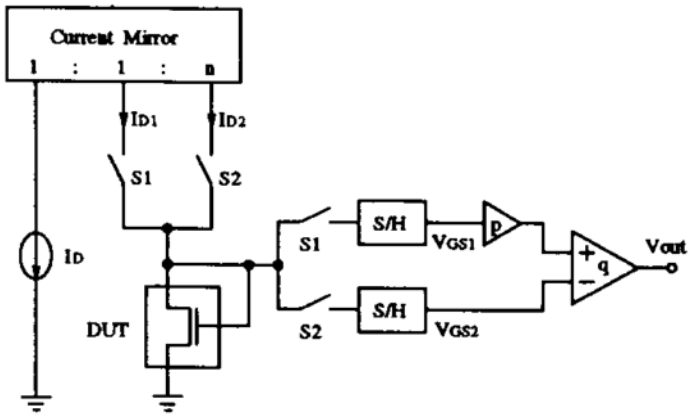
$$V_{T0} = \frac{1}{\sqrt{n} - 1}(\sqrt{n}V_{GS1} - V_{GS2}) \tag{2.13}$$

Assume that a complementary non-overlapping clock drives switches S1 and S2. When S1 is closed, V_{GS1} is sampled and multiplied by p, and when S2 is closed, V_{GS2} is sampled and subtracted from pV_{GS1}. The result is then multiplied by q. Hence, V_{out} is:

$$V_{out} = q(pV_{GS1} - V_{GS2}) \tag{2.14}$$

From (2.13), (2.14), and assuming $p = \sqrt{n}$ and $q = m/(\sqrt{n}-1)$, results in $V_{out} = mV_{T0}$. Therefore, this circuit can deliver a multiple of V_{T0} by choosing an integer m. Then, V_{T0} can be obtained by choosing $n = 4$ and $m = 1$, that results in $p = 2$ and $q = 1$.

Figure 2.6: Schematic of the circuit proposed by YU; GEIGER



Source: YU; GEIGER (1994).

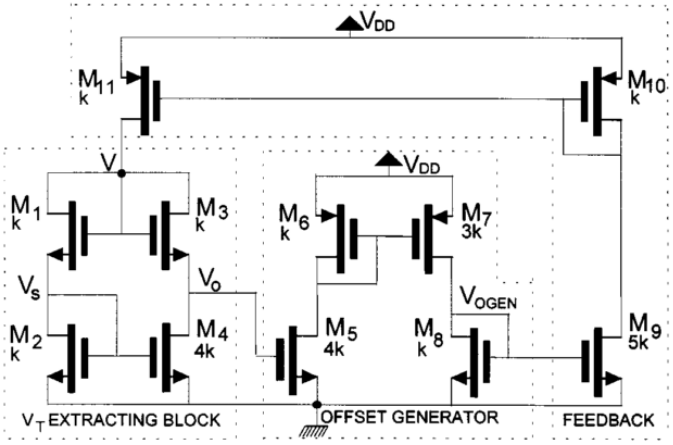
As was aforementioned, this implementation has a great advantage over the others, because its accuracy is not dependent on the matching of two transistors. Notwithstanding, to achieve this, the circuit uses many structures, as two sample and hold circuits and one differential amplifier, increasing the complexity, and the area when compared with other implementations.

After that, multiple solutions were developed, but it was, probably, Fikos's implementation the most remarkable structure. The circuit is shown in Figure 2.7 and its operation is explained as follows. Considering that all transistors of the circuit operate in saturation, with the ratios ($S_1 = S_2 = S_3 = S_4/4$), and assuming that the I_D of each transistor follows the simple quadratic law (2.1), then:

$$I_{D1} = I_{D2} \quad \therefore \quad V_S = V/2 \tag{2.15}$$

$$I_{D3} = I_{D4} \quad \therefore \quad V_O = V_{T0} \tag{2.16}$$

Figure 2.7: Schematic of the circuit proposed by FIKOS; SISKOS



Source: FIKOS; SISKOS (2001).

In order to compensate the mismatch problem and some second order effects, the authors added an offset block together with a feedback. The offset was implemented with a simple structure in cascade with the V_{T0} monitor, thus, this circuit achieves a low error, with a low complexity. Additionally, it was the first structure designed taking care of the power consumption, but even so, the operation in the strong inversion region makes its consumption higher than some tens of micro Watts.

2.2 Recent Development in V_{T0} Monitors

Other structures, besides those already presented, as CILINGIROGLU; HOON (2003); SENGUPTA (2004); WANG; TARR; WANG (2004); VLASSIS; PSYCHALINOS (2007) were developed before 2012. All of them operate using the same principle: bias two transistors in strong inversion with a given aspect ratio, and then compare its gate-to-source voltage. The big problem with this approach is the high power consumption, associated to the operation of the transistor in the strong inversion region.

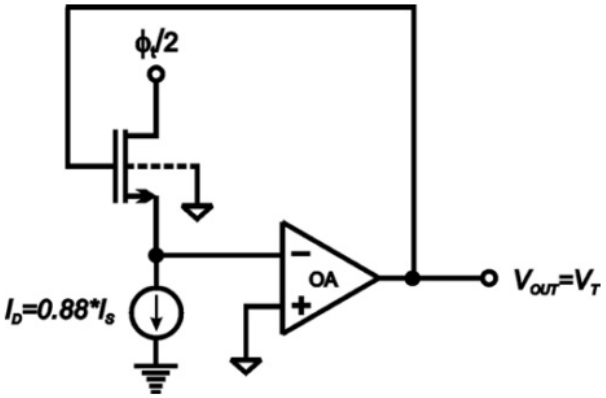
In the last 5 years new structures with low-power consumption have appeared. To accomplish that, these new circuits operate in the weak or/and moderate inversion, taking advantage that there is not speed requirement for V_{T0} monitors.

SIEBEL; SCHNEIDER; GALUP-MONTORO (2012) proposed the circuit shown in Figure 2.8. According to the authors "in this circuit, since the MOS transistor operates at low current levels and in the linear region, it is less affected by second order effects" (2012, p. 2).

As the circuit is based on the g_{ch}/I_D methodology, we present the mathematical explanation, before explain the circuit operation. Using (B.3) and (B.4) from the UICM model (see Appendix B), it is obtained that, if a transistor operates under an inversion level $i_f = 3$ and with $V_S = 0$, then its V_{GB} must be equal to V_{T0}. Now, if we also set $V_{DS} = \phi_t/2$, we obtain $i_r = 2.12$ and $I_D = 0.88 * I_S$. Then, looking to the circuit in Figure 2.8, we can recognize that, because the transistor is biased with a current equal to $0.88 \times I_S$, and presents a $V_S = 0$ and $V_D = \phi_t/2$, it must have $V_{GB} = V_{T0}$. This condition is achieved through the use of the OPAMP, that create a feedback.

This circuit was implemented using discrete elements, reason for which it is not possible to compare its behaviour with that of other structures. Also it is important to remember that, even when the circuit is simple, some problems need to be resolved in order to implement an integrated version, as for example find a circuit that generates a current exactly equal to $0.88 \times I_S$.

Figure 2.8: Schematic of the circuit proposed by SIEBEL; SCHNEIDER; GALUP-MONTORO



Source: SIEBEL; SCHNEIDER; GALUP-MONTORO (2012).

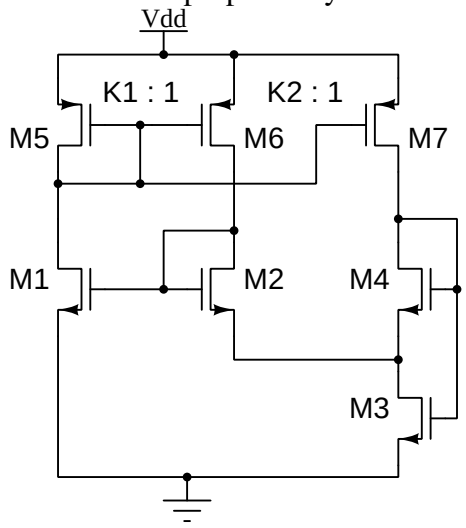
In 2014, MATTIA; KLIMACH; BAMPI proposed a new monitor, based on the well-known current source of Figure 2.9. It is important to highlight that, although the circuit is the same presented by SERRA-GRAELLS; HUERTAS (2003), its functionality is different.

This circuit, as the one proposed by Siebel, is designed using the UICM model, and is based on the fact that a transistor that operates under an inversion level $i_f = 3$ with $V_S = 0$ will have a $V_{GB} = V_{T0}$, as was above explained. Then, suppose that in the circuit shown in Figure 2.9, M1 operates under such condition ($i_f = 3$), being in the moderate inversion region. M2 is kept saturated too, and sharing the same gate voltage, but with a source voltage different from zero. Using (B.3) and (B.4) for M2, knowing that $V_{G1} = V_{G2} = V_{T0}$ and $I_{D1} = K1I_{D2}$, leads to (2.17).

$$V_{s2} = -\phi_t F(i_{f2}) = -\phi_t F\left(\frac{3S_1}{K1S_2}\right) \tag{2.17}$$

Thus, a voltage proportional to ϕ_t needs to be attached to the source terminal of M2, in order to maintain the equality of (2.17). This condition is achieved by using the self-cascode

Figure 2.9: Schematic of the circuit proposed by MATTIA; KLIMACH; BAMPI



Source: MATTIA; KLIMACH; BAMPI (2014a).

(SC) composed by transistors M3 and M4. In this configuration, the lower transistor (M3) works in triode, while the upper transistor (M4) operates in saturation. Consequently, the voltage in the output node of the SC is determined by:

$$V_{DS3} = \phi_t[F(i_{f3}) - F(i_{f4})] \tag{2.18}$$

As the circuit is designed to make $i_{f1} = 3$, and to keep transistor M1 in saturation, we can define, using (B.1), the drain current of transistor M1 as $I_{D1} = 3S_1 I_{SQ}$. This current is mirrored to M2 and to the SC pair M3-M4 by transistors M5-M7. Then, $I_{D2} = I_{D1}/K1$, $I_{D4} = I_{D1}/(K1K2)$ and $I_{D3} = (I_{D1}/K1)(1 + 1/K2)$. Therefore, we can express V_{DS3} as:

$$V_{DS3} = \phi_t \left[F\left(3\frac{S_1}{S_3}\frac{K2+1}{K1K2} + 3\frac{S_1}{S_4}\frac{1}{K1K2} \right) - F\left(3\frac{S_1}{S_4}\frac{1}{K1K2} \right) \right] \tag{2.19}$$

Finally, equaling (2.18) and (2.19) an expression to find the aspect ratio of transistors M1-M4 and the current gain $K1$ and $K2$ can be obtained. The aspect ratio of transistors M5-M7 is established in order to maintain all the transistors in the saturation region.

This structure achieves a low error, consuming only nano Watts, but it presents a poor supply regulation, resulting in poor PSR and LS.

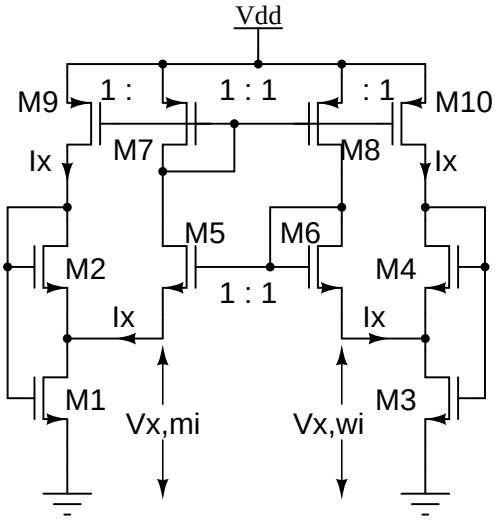
The next year, MATTIA; KLIMACH; BAMPI (2015) proposed another self-biased topology, now based on the equilibrium of two self-cascode cells, that can be made to operate at a much higher difference of inversion levels than the previous circuit. This makes its operating point more robust than the MATTIA; KLIMACH; BAMPI (2014a) circuit, and it also results into a lower supply sensitivity.

The circuit is shown in Figure 2.10. In this circuit, a voltage-following current mirror (transistors M5-M10) clamps two SC together. One of the SC operates in the moderate inversion region, and the other one in the weak inversion region. Using equations (B.1), (B.3) and (B.4) it is obtained:

$$V_{X,12} = \phi_t[F(i_{f1}) - F(i_{f2})] \tag{2.20}$$

$$V_{X,34} = \phi_t[F(i_{f3}) - F(i_{f4})] \tag{2.21}$$

Figure 2.10: Schematic of the circuit proposed by MATTIA; KLIMACH; BAMPI



Source: MATTIA; KLIMACH; BAMPI (2015).

Transistor M1 is biased in the $i_f = 3$ condition in order to obtain the V_{T0} value at its gate. Then, making use of the expression (2.20), and defining the inversion level i_{f2}, we can get $V_{X,12}$. After that, as the SC M3-M4 needs to be in an inversion level lower than M1-M2 SC, a relation between i_{f2} and i_{f3} can be established. Once the value of i_{f3} is found, using (2.21) and recalling that $V_{X,12} = V_{X,34}$ we can know the value of i_{f4}. Finally, the aspect ratio of all the transistors can be found, if a value for the current I_X is defined.

2.3 Comparison of recent works and State of the Art

To conclude the bibliographic study, we present in Table 2.1 a comparison between some relevant V_{T0} monitors in the last 15 years.

Older works (CILINGIROGLU; HOON, 2003; WANG; TARR; WANG, 2004; SENGUPTA, 2004; VLASSIS; PSYCHALINOS, 2007) were focused on the strong inversion quadratic MOSFET model for the drain current, being limited to a specific condition of operation and presenting high power consumption. These designs although were useful solutions, and present good performance in some aspects, as the best LS (VLASSIS; PSYCHALINOS, 2007), are now useless due to the low-efficiency that implies to work in strong-inversion for applications that do not need fast responses.

Recent works (MATTIA; KLIMACH; BAMPI, 2014a; MATTIA et al., 2015) use the transistors in moderate and/or weak inversion, thus maintaining low the power consumption. Additionally, these implementations deliver the V_{T0} value with low error. Unfortunately, they do not achieve good results regarding some other important performance characteristics, like Power Supply Rejection (PSR) or line sensitivity (LS), much needed features of a circuit that would be a part of a system.

Table 2.1: Comparison of recent V_{T0} monitor circuits

Characteristic	[1]	[2]	[3]	[4]	[5]	[6]	Units
Year	2015	2014	2007	2004	2004	2003	
PSRR @100 Hz	-38.9	-30	- - -	- - -	- - -	- - -	dB
Line Sensitivity	3600	46000	480	2562	8000	554	ppm/V
Temperature Range	-40 to 125	-40 to 125	0 to 100	0 to 100	20 to 80	-50 to 100	°C
Supply Range	0.6 to 1.8	0.6 to 1.2	1 to 3.6	1.9 to 2.1	2 to 2.5	3.5 to 6.5	V
Max. Error	1.64	1.3	- - -	11	4.3	4.9	%
Power Consumption	23	23	50000	290000	387500	- - -	nW
Model	UICM	UICM	quadratic	quadratic	quadratic	quadratic	

[1](MATTIA et al., 2015) [2](MATTIA; KLIMACH; BAMPI, 2014a)

[3](VLASSIS; PSYCHALINOS, 2007) [4](WANG; TARR; WANG, 2004)

[5](SENGUPTA, 2004) [6](CILINGIROGLU; HOON, 2003)

3 PROPOSED NMOS AND PMOS THRESHOLD VOLTAGE MONITORS

As was noticed in the last chapter, despite there are many different implementations of V_{T0} monitors, and although these have had a big advance in the last years, still there is not a topology that meets the necessary specifications to become part of a complete system.

In this chapter, first we present a high PSR and low LS self biased circuit topology, that allows the direct extraction of the threshold voltage for wide temperature and power supply voltage ranges. Its design methodology is based on the Unified Current Control Model (UICM) model (CUNHA; SCHNEIDER; GALUP-MONTORO, 1998). It occupies a small silicon area, consumes just tens of nano-Watts, and can be easily implemented in a standard digital CMOS process, since it only uses MOS transistors (does not need any resistor). This implementation works by finding an equilibrium point between two transistors that share the gate voltage, but present different inversion levels, and for this reason is known as unbalanced V_{T0} monitor.

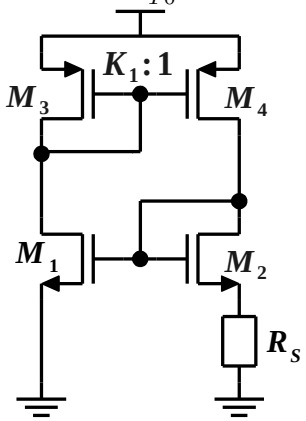
Then, an additional NMOS V_{T0} monitor is proposed, this is designed using the same MOSFET model than the previous one, and also uses the same technique to improve the PSR and LS, thereby sharing all the benefits that presents the unbalanced V_{T0} monitor. The difference is that this structure is based on the equilibrium of two Self-Cascode structures that are clamped together. Since, the structure is symmetrical, it is more robust against mismatch effects and attains a lower error. This structure is called balanced V_{T0} monitor.

Finally, a PMOS V_{T0} monitor is introduced. This circuit is implemented using a I_{SQ} current generator, that uses the PMOS version of the balanced V_{T0} monitor. The PMOS V_{T0} monitor presents the same advantages of the NMOS versions, and additionally overcomes one of the principal issues in PMOS V_{T0} monitors by delivering its output referenced to ground, and not to the supply voltage.

3.1 Unbalanced NMOS V_{T0} monitor

The unbalanced NMOS V_{T0} monitor circuit was introduced in MATTIA; KLIMACH; BAMPI (2014a). The circuit is based in the well known structure of Figure 3.1, which was originally presented as a PTAT voltage generator in VITTOZ; NEYROUD (1979). The operation of the circuit will be explained using the UICM model, later the circuit design procedure is exposed, and finally some simulation results are presented.

Figure 3.1: Unbalanced V_{T0} monitor circuit concept.



Source: VITTOZ; NEYROUD (1979).

3.1.1 V_{T0} Monitor concept

From (B.3) and (B.4)(appendix B), one can see that a saturated nMOSFET with grounded source, as is the case of transistor M_1 in Figure 3.1, presents a gate voltage equal to (3.1)

$$V_{G1} = V_{T0} + n\phi_t F(i_{f1}) \tag{3.1}$$

considering that transistor M_2 presents the same gate voltage of M_1, and again using (B.3) and (B.4) it is possible to obtain (3.2)

$$\frac{V_{G1} - V_{T0}}{n} - V_{S2} = \phi_t F(i_{f2}) \tag{3.2}$$

Now, equaling (3.1) and (3.2), (3.3) is obtained , which shows that the source voltage of transistor M_2 (V_{S2}) is proportional to the ϕ_t voltage. Therefore, if the output of the circuit is taken in the V_{S2} node, an ideal PTAT voltage can be obtained as long as the inversion level of transistor M_1 and M_2 remains constant with temperature.

$$V_{S2} = \phi_t[F(i_{f1}) - F(i_{f2})] \tag{3.3}$$

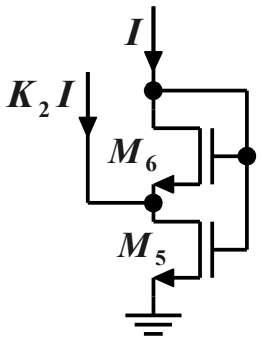
The topology proposed in VITTOZ; NEYROUD (1979) can be modified by changing the resistor R_S of Figure 3.1 for the lower transistor of a SC. This change in the topology turns the circuit into an MOSFET-only structure, which makes possible to implement it in any standard digital CMOS process.

A SC circuit is shown in Figure 3.2. This circuit is used to generate a PTAT voltage independent of process parameters (VITTOZ; FELLRATH, 1977).

In the SC the transistor M_5 has higher drain current than M_6 but smaller aspect ratio, leading to different inversion levels on each transistor. While M_6 must be in saturation, M_5 must be in triode. The difference between their gate-source voltages appear across the drain-source terminals of M_5. As done in ROSSI; GALUP-MONTORO; SCHNEIDER (2007), the use of (B.3) and (B.4) demonstrates that this voltage is proportional to the thermal voltage, as given by (3.4).

$$V_{DS5} = \phi_t[F(i_{f5}) - F(i_{f6})] \tag{3.4}$$

Figure 3.2: Self-Cascode structure.

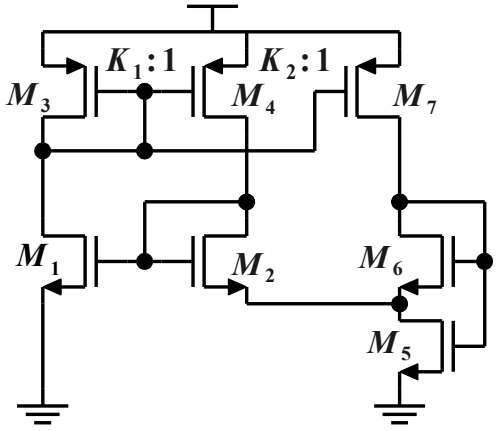


Source: The own author.

Usually both transistors operate in weak inversion, and the PTAT voltage is determined only by the ratio of current densities between the devices. But, as demonstrated in ROSSI; GALUP-MONTORO; SCHNEIDER (2007), (3.4) shows that as long as the inversion levels of both MOSFETs are kept constant over temperature, they generate an ideal PTAT voltage under any inversion level. This can be achieved by biasing both transistors with currents proportional to I_{SQ}.

Figure 3.3 shows the topology after the R_S replacing. This circuit reaches an equilibrium point when the voltage $V_{S2} = V_{DS5}$ and can operate as a PTAT voltage generator, a current source (CAMACHO-GALEANO; GALUP-MONTORO; SCHNEIDER, 2005) and/or a V_{T0} monitor as it will be explained.

Figure 3.3: Unbalanced V_{T0} monitor circuit.



Source: The own author.

From Chapter 1.1 we know that a saturated nMOSFET with grounded source and operating under a constant inversion level equal to 3 ($i_f = 3$) will have a gate voltage V_G equal to the threshold voltage V_{T0}, because under these conditions, the right side of (B.3) becomes zero. Suppose that in the circuit shown in Figure 3.3, M_1 operates under such condition, being in the moderate inversion region. M_2 is kept saturated too and sharing the same gate voltage, but with a source voltage different from zero. Using (3.3) and (B.1) for M_2, knowing that $V_{G1} = V_{G2} = V_{T0}$ and $I_{D1} = K_1 I_{D2}$, leads to:

$$V_{S2} = -\phi_t F(i_{f2}) = -\phi_t F\left(\frac{3S_1}{K_1 S_2}\right) \tag{3.5}$$

From (3.5) one can conclude that if the resulting M_1 current I_{D1}, that was chosen to keep M_1 operating with $i_{f1} = 3$, is also used to control the drain current of M_2 (through the M_3-M_4 current mirror), M_2 also operates under a constant inversion level, making $F(i_{f2})$ constant. Then, a non-zero equilibrium point can be reached in this circuit, $V_{G1} = V_{G2} = V_{T0}$ for any temperature and independently of process parameters, defined only by geometric ratios, if the SC composed by transistors M_5-M_6 is properly designed.

As the circuit is designed to make $i_{f1} = 3$, then it is necessary that $I_{D1} = 3S_1 I_{SQ}$, which means that I_{D1} is proportional to the I_{SQ}. This current is mirrored to M_2, and to the SC pair M_5-M_6, meaning that $I_{D2} = I_{D1}/K_1$, $I_{D6} = I_{D1}/(K_1K_2)$ and $I_{D5} = (I_{D1}/K_1)(1 + 1/K_2)$. Thus, equation (3.4) becomes:

$$V_{DS5} = \phi_t \left[F\left(3\frac{S_1}{S_5}\frac{K_2+1}{K_1K_2} + 3\frac{S_1}{S_6}\frac{1}{K_1K_2}\right) - F\left(3\frac{S_1}{S_6}\frac{1}{K_1K_2}\right)\right] \tag{3.6}$$

From (3.6) it is clear that the PTAT voltage generated by the SC cell depends only on geometrical factors, and not on fabrication process parameters.

Since $V_{G1} = V_{T0}$, the gate voltage of M_1 has a temperature dependence equal to that of the threshold voltage, that can be approximated by the linear equation (3.7).

$$V_{G1} = V_{T0(nom)} + K_T(T - T_{nom}) \tag{3.7}$$

where T is the absolute temperature, $V_{T0(nom)}$ is the threshold voltage at the nominal temperature T_{nom} and K_T is the thermal coefficient of the threshold voltage.

3.1.1.1 PSR and Line Sensitive improvement

The basic topology shown in Figure 3.3 exhibits a high sensitivity to changes in the supply voltage (V_{DD}), resulting in a poor LS and PSR, due to the limited output impedance of the current mirror, formed by the M_3, M_4 and M_7 transistors. A possible solution to increase its output impedance would be the use of cascode current mirrors, rather than single transistors, but the transistor stacking would increase the minimum operating supply voltage. A better option is to include an Operational Amplifier (OA)(BAKER, 2004) that add a high gain feedback path resulting an effective increase in the output impedance of the mirror without increasing the minimum supply voltage.

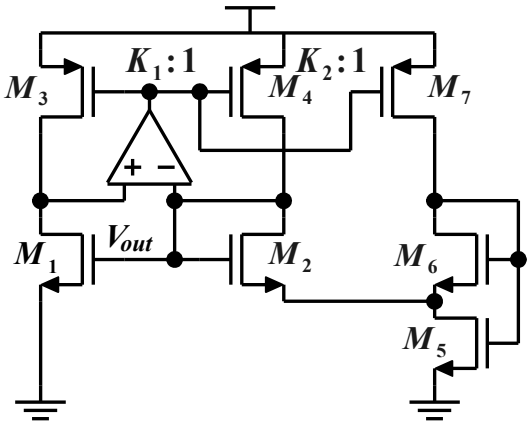
Figure 3.4 shows the V_{T0} monitor circuit with the OA connected. Note that M_3 is no longer diode-connected, so its drain can move to the same voltage of the drain of M_4. The OA compares the drain voltage of M_4 with the drain voltage of M_3 and force them to be approximately equal, adjusting the current mirror bias.

3.1.1.2 Start-up Circuit

As the V_{T0} monitor circuit is a self-biased structure it presents two DC stability operation points, one in the desired bias condition in which the SC has the same PTAT voltage as V_{S2}, and another when the current in all branches is zero. A Start-up circuit that can prevent the zero-current condition is necessary.

Figure 3.5 shows the topology with the start-up circuit. Note that the current through M_8 and M_9 is zero when the circuit is in normal operation because of the loaded C_c capacitance, which forces $V_{DS8} = 0$ and $V_{GB9} = 0$. This results in a zero extra current consumption, which is a desirable characteristic in nano-Watt circuits.

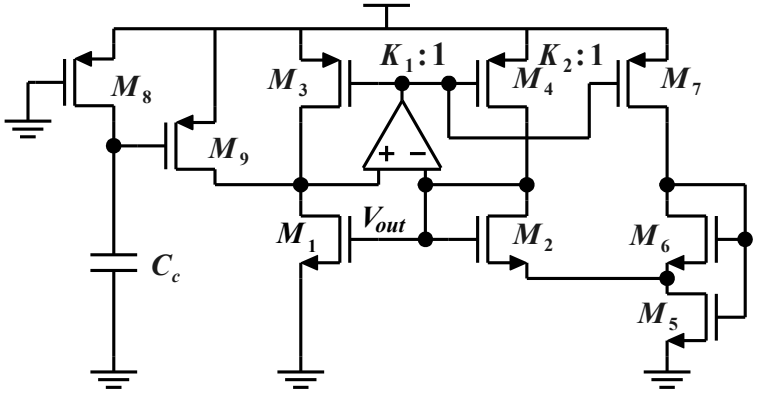
Figure 3.4: Unbalanced V_{T0} monitor with an OA connected to improve LS and PSR.



Source: The own author.

The start-up circuit works as follows: supposing the capacitor C_c discharged, when the V_{DD} voltage starts to increase, M_9 drives a current into the drain of M_1 initializing the circuit, and simultaneously, M_8 delivers a current into the capacitor C_c, charging it and eventually moving M_8 to deep triode, and M_9 to the cut-off state. The start-up circuit, was chosen due to its simplicity and zero steady state consumption, even so, it presents some disadvantages, which are discussed in Appendix C.

Figure 3.5: Schematic of the proposed unbalanced V_{T0} monitor.



Source: The own author.

3.1.2 Design Methodology

Before starting to talk about the design methodology that was used, it is important to emphasize that the design methodology aims to facilitate in the layout the use of common-centroid technique and dummy devices. It provides greater layout regularity, resulting in better device matching and consequent lowering circuit spread from local variability. For this reason, only integer values for the current mirror gains and for the width ratios of the transistors are considered.

As a starting point, the forward inversion level of M_1 is set as 3, and the inversion level of M2 (i_{f2}) is fixed by the current gain K_1 and by the ratio S_2/S_1 as (3.5) shows. It should be noted that a low i_{f2} reduces power consumption, but a higher V_{S2} value is needed to balance the circuit operation. Therefore, as the voltage V_{S2} has to be generated by a single SC, its value is limited to a maximum of 100 mV.

It is important also consider that, a small aspect ratio for M_1 and M_2 contributes to lower the power consumption, but the designer must remember that smaller transistor area has a severe drawback, namely the increase of the local mismatch from local variability (PELGROM; DUINMAIJER; WELBERS, 1989).

With these considerations a value of $K_1 = 6$ and $S_2/S_1 = 1$ was chosen in order to demonstrate a practical implementation, leading to a $V_{S2} \approx 58$ mV at 27 °C. Once the voltage V_{S2} and the current gain K_1 are defined, it is possible to proceed to sizing the self-cascode generator.

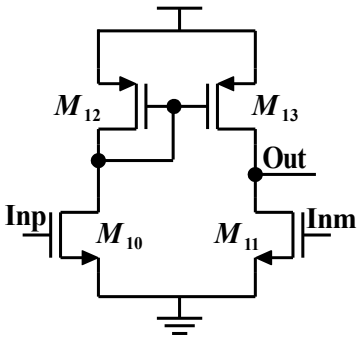
Since we have chosen $V_{S2} \approx 58$ mV for the equilibrium condition, considering the aforementioned trade-off between power consumption and area, and using (3.6), it is possible to use $S_5/S_6 = 4$ and $K_2 = 1$. Table 3.1 presents the sizing using the design above mentioned after a fine adjust through simulation.

Table 3.1: Core transistors sizing.

	M_1	M_2	M_5	M_6
i_f	3	0.5	0.833	0.126
W (μm)	1	1	1*2	1.1*6
L (μm)	30	30	50	50
Area (μm^2)	30	30	100	330

The OA was implemented using the low voltage pseudo-differential amplifier shown in Figure 3.6. In this, the PMOS transistors form a current mirror while the NMOS ones operate as a differential amplifier. The design of this structure requires a relationship between the additional area and power consumption that needs to be considered. In this design, we consider that an additional 10% of current consumption is adequate. Observe that the inversion level of the input transistors of the OA are set to 3, since their sources are grounded and their gate voltages are equal to V_{T0}. Once we know the current and the inversion level, it is easy to size the transistors. Table 3.2 presents the sizing of the auxiliary structures as the OA, start-up and current sources.

Figure 3.6: Pseudo-Differential Amplifier.



3.1.3 Results

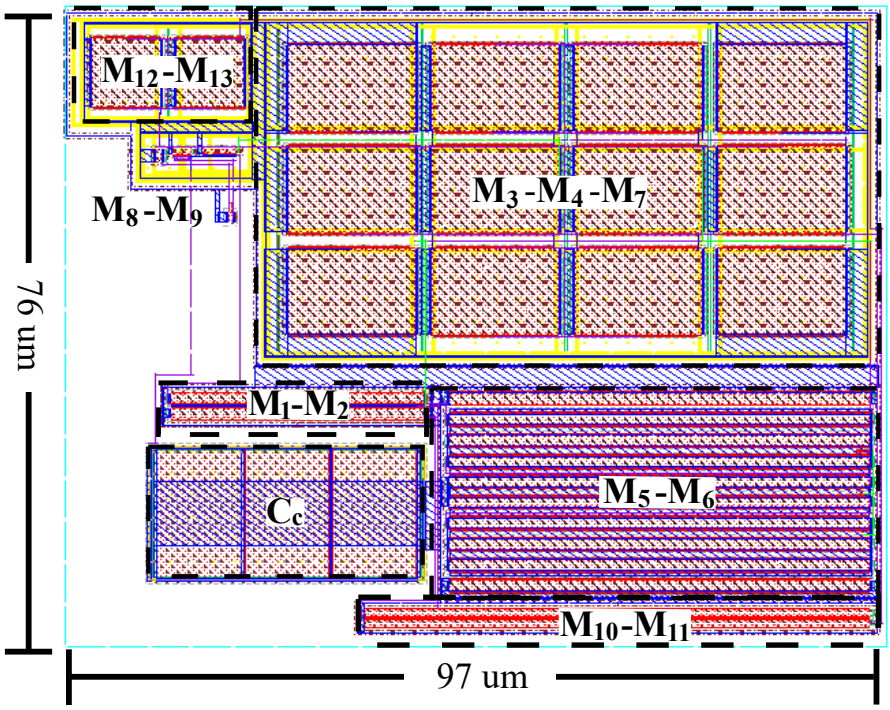
The results presented here are for Cadence Virtuoso post-layout simulations of our design implemented in IBM 130 nm process. The layout takes into consideration the good

Table 3.2: Sizing of the auxiliary structures.

	M_3	M_4	M_7	M_8	M_9	$M_{10,11}$	$M_{12,13}$
i_f	0.1	0.1	0.1	- - -	- - -	3	0.16
W (µm)	10*6	10	10	0.5	0.36	0.5	7.8
L (µm)	15	15	15	2	4	60	8
Area (µm^2)	900	150	150	1	1.4	30	62.4

layout matching practices such as common-centroid placement and dummy structures. The occupied silicon area is 0.0073 mm^2, as shown in Figure 3.7.

Figure 3.7: Unbalanced V_{T0} monitor layout.



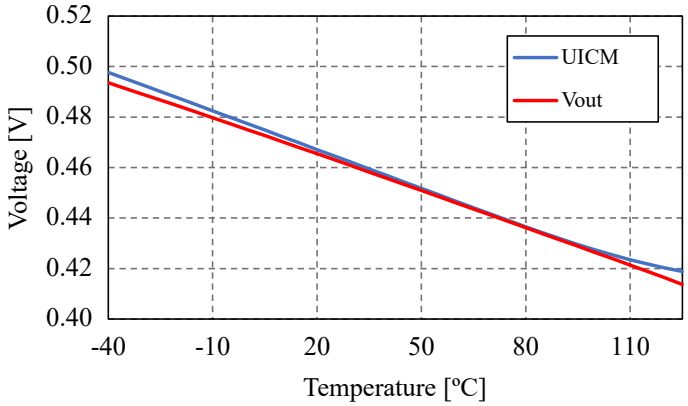
Source: The own author.

The MOSFETs used in this implementation are standard I/O type, that present higher threshold voltage and also allow a higher VDD voltage (VDD$_{max}$) than the core transistors in this CMOS process. Figure 3.8 presents the V_{T0} variation over temperature, estimated by the (g_m/I_D) method presented in Section 1.1.2 (labeled UICM), and simulated in the V_{T0} monitor circuit of Figure 3.5 (labeled Vout). As one can see the two lines are very close, to each other.

The error defined by (3.8) is presented in Figure 3.9. The circuit tracks the ideal threshold voltage with an error inferior to 1.3% within the -40 to 125 °C temperature range.

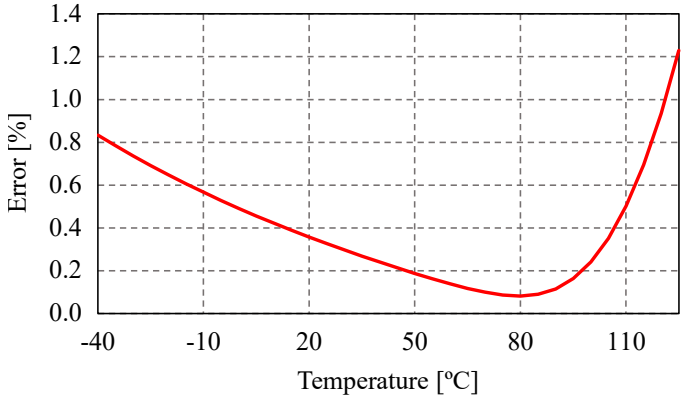
$$\epsilon(\%) = 100 \times \left(\frac{V_{out} - V_{UICM}}{V_{UICM}} \right) \tag{3.8}$$

Figure 3.8: Difference between the theoretical value of V_{T0} (UICM) and the simulated output of the V_{T0} monitor (Vout).



Source: The own author.

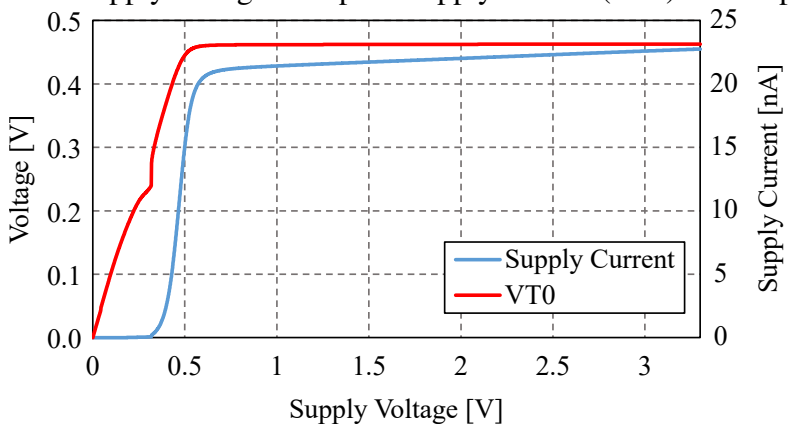
Figure 3.9: Percentual error in the V_{T0} extracted by the circuit.



Source: The own author.

As shown in Figure 3.10, the circuit starts operating at around 0.7 V. The line sensitivity of V_{G1} has been improved with respect to MATTIA; KLIMACH; BAMPI (2014a), being 550 uV/V from 0.7 V to 3.3 V, while the current consumption sensitivity is 640 pA/V. The whole circuit consumes 21.54 nA at 27 °C.

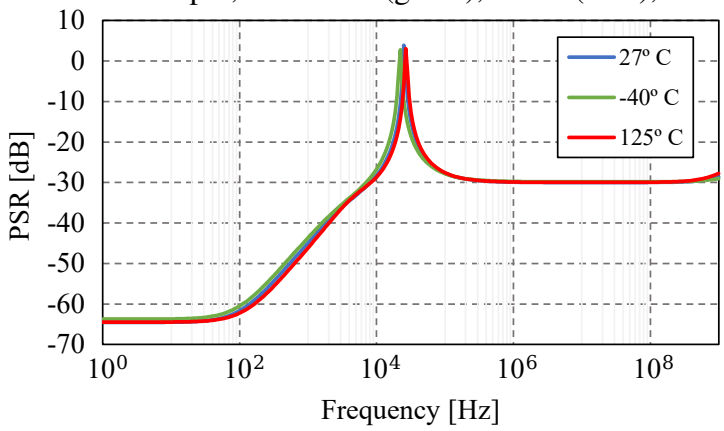
Figure 3.10: Supply voltage sweep vs. supply current (blue) and output (red).



Source: The own author.

The PSR simulated between 10 Hz and 1 GHz for 3 different temperatures (-40, 27 and 125 °C) is shown in Figure 3.21. PSR simulated at 100 Hz and V_{DD} = 1.2 V, is -63 dB for V_{G1}, which is almost the double of the result obtained in previous works. This result together with the LS, confirms that the approach used in order to improve the output impedance works properly.

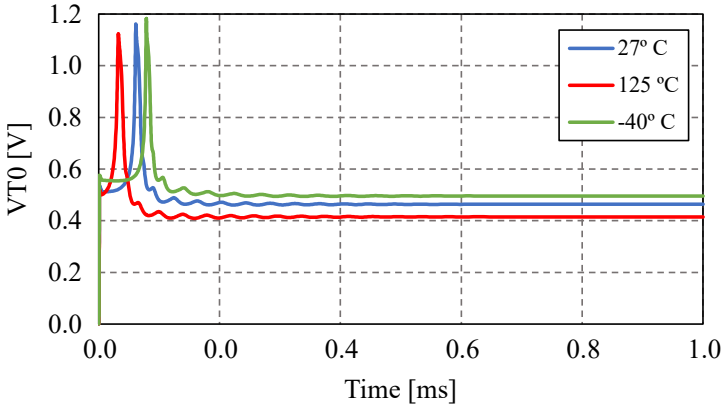
Figure 3.11: PSR at the output, for -40 °C (green), 27 °C (blue), and 125 °C (red).



Source: The own author.

Start-up behavior of the circuit was simulated for typical conditions and also for -40 and 125 °C in order to study the maximum settling time and the stability of the circuit. The simulations show that even in the worst case, that is -40 °C, the circuit presents a settling time less than 0.5 ms, which is acceptable for our proof of concept.

Figure 3.12: Start-up simulation, for -40 °C (green), 27 °C (blue), and 125 °C (red).

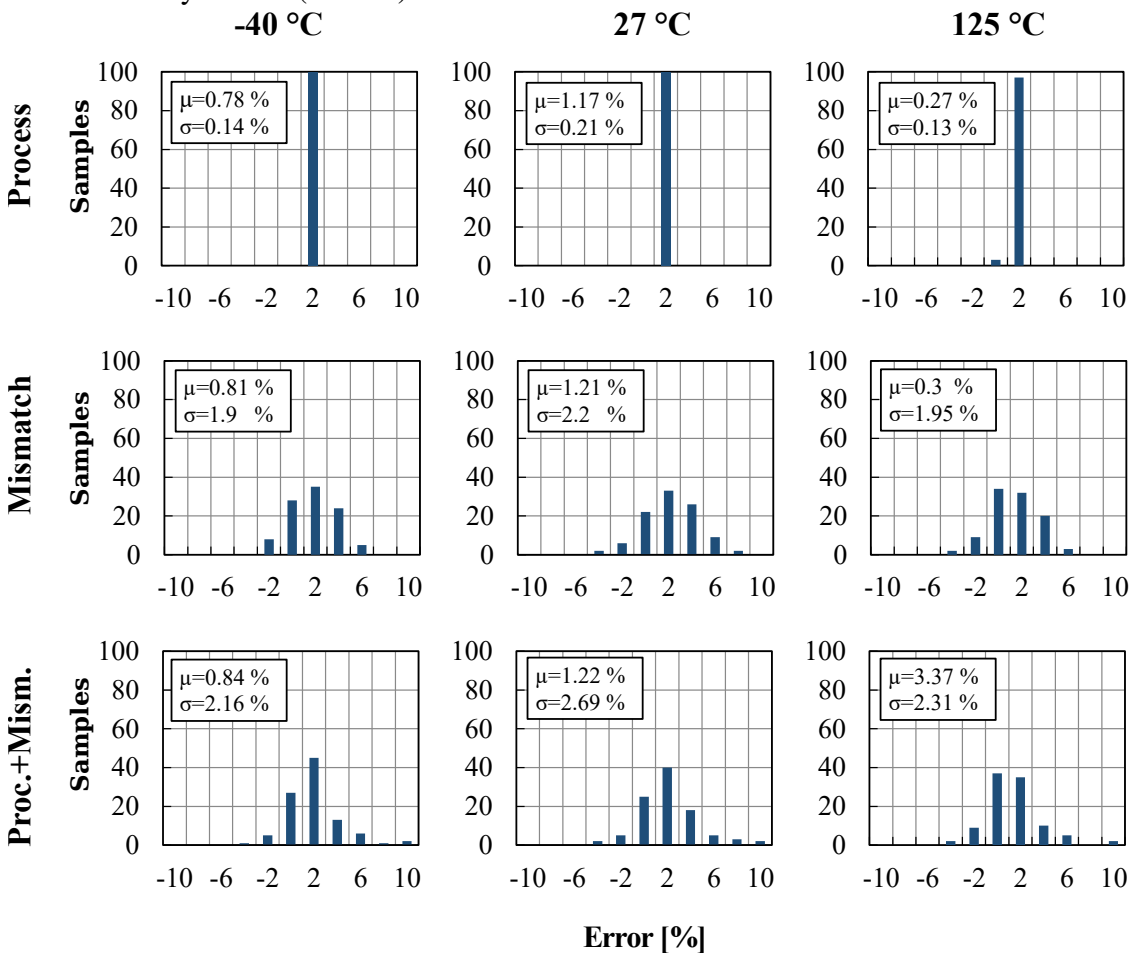


Source: The own author.

To analyze the impact of the fabrication variability effects on the output of the circuit, Monte Carlo (MC) simulations were run separately for local mismatch effects and average process variations, with 100 samples each. These simulations emulate the statistical variations that the transistor's parameters suffer due to fabrication variability. Also, the variability of transistor M_1 was excluded because this analysis aims to determine the error in the extracted V_{T0} due to the variability of the transistors that compose the monitor circuit, and not the variability of the transistor that is delivering the V_{T0} value. For average process MC the transistors have their parameters changed equally in each run - Figure 3.13 (top histograms). For local mismatch MC, the parameters of each transistor are varied individually in each run - Figure 3.13 (middle histograms). Both effects are taken

into account in a full variability analysis, shown in Figure 3.13 (bottom histograms). The results presented are for $V_{DD} = 1.2$ V under three different temperatures, -40, 27 and 125 °C.

Figure 3.13: Monte Carlo simulation results for Process (top), Mismatch (middle) and both variability effects (bottom).



Source: The own author.

As shown in the design methodology, the circuit's equilibrium point depends only on geometrical factors. It is thus less sensitive to average process variations, where V_{G1} tracks the threshold voltage value with a maximum error (mean and standard deviation) of $\epsilon(\pm 3\sigma) = 1.17 \pm 0.63\%$, comprising 99.7% of the samples. Local mismatch analysis, however, affect the current mirror and aspect ratio gains that define this equilibrium, resulting in a higher spread of $\epsilon(\pm 3\sigma) = 1.21 \pm 6.6\%$. A combined analysis yields a maximum error of $\epsilon(\pm 3\sigma) = 3.37 \pm 6.93\%$ for the whole operating temperature range. The mean and the sigma of each temperature and variability condition is shown in Figure 3.13. This circuit was fabricated using the Mosis's educational program, and we soon expect to measure it.

3.2 Balanced NMOS V_{T0} Monitor

The threshold voltage monitor circuit, shown in Figure 3.14, was introduced in (MATTIA et al., 2015) and is based on a self-biased current source topology proposed in

(CAMACHO-GALEANO; GALUP-MONTORO; SCHNEIDER, 2005). Its DC operating point is established by the equilibrium condition of the proportional-to-absolute-temperature (PTAT) voltages generated by two self-cascode (SC) cells that are clamped together. One of the SC cells operates in moderate inversion ($M_{1,2}$), while the other SC cell operates in weak inversion ($M_{3,4}$). Transistors M_5-M_{10} act as a voltage-following current mirror (GILBERT, 2004), making all the currents equal to I_X and forcing $V_{X1,2} = V_{X3,4}$.

3.2.1 Circuit Description

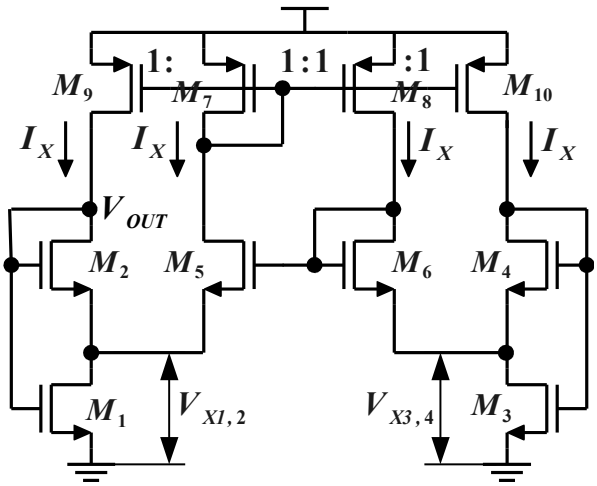
The voltage at the intermediate node of a self-cascode cell has been already shown to be a PTAT voltage, whenever both transistors operate at constant inversion levels (ROSSI; GALUP-MONTORO; SCHNEIDER, 2007), and the absolute value of the PTAT voltage (and its derivative) can be adjusted by the inversion levels of the transistors. Additionally, the upper transistors in a SC cell ($M_{2,4}$ in Figure 3.14) have to be in saturation, whereas the lower transistors $M_{1,3}$ are in triode. The use of (B.1), (B.3) and (B.4) demonstrates that

$$V_{X1,2} = \phi_t[F(i_{f1}) - F(i_{f2})] \tag{3.9}$$

$$V_{X3,4} = \phi_t[F(i_{f3}) - F(i_{f4})] \tag{3.10}$$

where $V_{X1,2}$ and $V_{X3,4}$ are ideally PTAT for any inversion level, as long as i_{f1-4} are kept constant over temperature. From (B.3) and (B.4), one can see that a NMOSFET with

Figure 3.14: V_{T0} monitor circuit basic topology.



Source: MATTIA; KLIMACH; BAMPI (2015)

grounded source and with a gate voltage V_G equal to the threshold voltage V_{T0} operates under a constant forward inversion level equal to 3 ($i_f = 3$). Suppose that M_1 operates under such condition, being in the moderate inversion region. The current I_X is defined based on the inversion levels of M_1 and M_2. Remembering that $i_{f2} = i_{r1}$, allows us to write

$$I_{D1} = S_1 I_{SQ}(3 - i_{f2}) = I_{D2} + I_{D5} = 2I_X \tag{3.11}$$

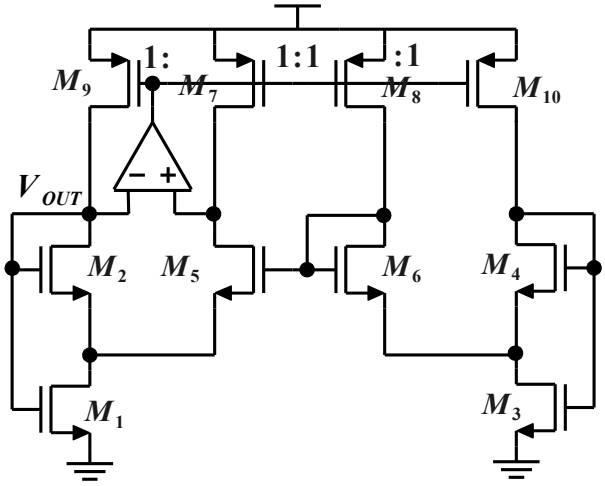
where i_{f2} defines the voltage $V_{X1,2}$ according to (3.9). The inversion levels i_{f3-4} can then be defined to make $V_{X1,2} = V_{X3,4}$, and the circuit operates in equilibrium.

3.2.1.1 PSR and Line Sensitive improvement

The basic topology shown in Figure 3.14 exhibits a high sensitivity to changes in the supply voltage (V_{DD}), resulting in a poor LS and PSR, due to the limited output impedance of the current mirror, formed by the M_7-M_{10} transistors. A possible solution to increase its output impedance would be the use of cascode current sources, rather than single transistors, but the transistor stacking would increase the minimum operating supply voltage. A better option is to include an Operational Amplifier (OA)(BAKER, 2004) that add a high gain feedback path resulting an effective increase in the output impedance of the mirror without increasing the minimum supply voltage.

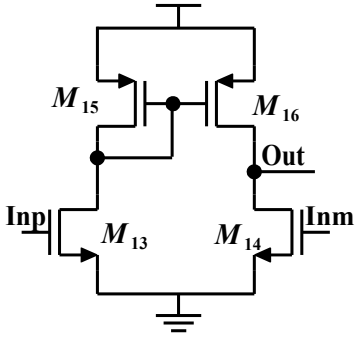
Figure 3.15 shows the V_{T0} monitor circuit with the OA connected. Note that M_7 is no longer diode-connected, so its drain can move to the same voltage of the drain of M_9. The OA compares the drain voltage of M_2 with the drain voltage of M_5 and force them to be approximately equal, adjusting the current mirror bias. Figure 3.16 shows the low voltage pseudo-differential amplifier that was used, where the PMOS transistors form a current mirror while the NMOS ones operate as a differential amplifier. When both inputs are equal, both branches of the mirror are in equilibrium. If the inputs are not equal, this imbalance causes the amplifier output to swing up or down providing the desired action.

Figure 3.15: V_{T0} monitor topology with improved LS and PSR.



Source: The own author

Figure 3.16: Pseudo-Differential Amplifier.



Source: The own author

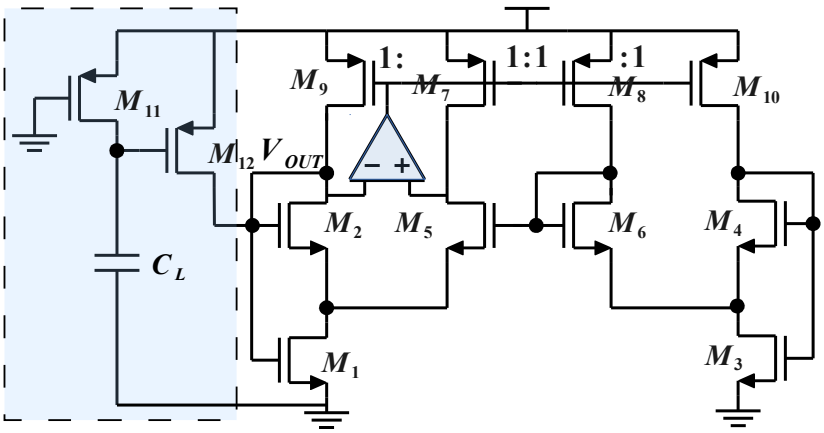
3.2.1.2 Start-up Circuit

As the V_{T0} monitor circuit is a self-biased structure it presents two DC stability operation points, one in the desired bias condition in which the SCs have the same PTAT voltage, and another when the current in all branches is zero. A Start-up circuit that can prevent the zero-current condition is necessary.

Figure 3.17 shows the topology with the start-up circuit. Note that the current through M_{11} and M_{12} is zero when the circuit is in normal operation because of the loaded C_L capacitance, which forces $V_{DS11} = 0$ and $V_{GB12} = 0$. This results in a zero extra current consumption, which is a desirable characteristic in nano-Watt circuits.

The start-up circuit works as follows: supposing the capacitor C_L discharged, when the V_{DD} voltage starts to increase, M_{12} drives a current into the SC $M_{1,2}$ initializing the circuit, and simultaneously, M_{11} delivers a current into the capacitor C_L, charging it, and eventually moving M_{11} to deep triode, and M_{12} to the cut-off state.

Figure 3.17: V_{T0} monitor with Start-Up circuit and improved LS and PSR



Source: The own author

3.2.2 Circuit Design

In the previous analysis it was supposed for simplicity the same current in all branches of the current mirror, and now in the design section it is convenient that these currents I_X are defined as a fraction of the specific current, according to the transistor current model being used herein. I_X is then normalized as follows:

$$I_X = I_{SQ}/A \tag{3.12}$$

where A is a design constant factor that can be used to determine the power consumption of the whole circuit ($I_{\text{TOTAL}} = 4I_X$). We also define another design constant $B = i_{f2}/i_{f3}$, that sets the ratio between the inversion levels of the transistors of each SC cell.

Since our objective is to set the gate-source voltage of M_1 to be equal to V_{T0}, we must chose $i_{f1} = 3$, and then $V_{X,12}$ is determined solely from i_{f2} according to (3.9). Then the forward inversion level of M_3 can be defined by the design constant B, and finally the ratio i_{f4}/i_{f3} can be adjusted to make $V_{X3,4}$ equal to $V_{X1,2}$, that is a condition presented in Section 3.2.1.

As a design example we choose $i_{f2} = 0.5$, $A = 10$ and $B = 5$, leading to $V_{X1,2} = V_{X3,4} = 58$ mV and $i_{f3} = 0.1$. Once $V_{X3,4}$ and i_{f3} are known, the value of $i_{f4} = 0.001$ is easily obtained, and the sizing of the transistors can be determined from (B.1) and (3.11).

A design technique that can be used to improve the circuit behavior is the implementation of the main transistors (M_1-M_4) through the parallel and series composition of unitary devices, that are assumed to have the same process characteristics (V_{T0}, I_{SQ}, n, and so on). Using the common-centroid layout strategy also helps to improve the circuit, since a regular layout and the use of dummy devices can minimize the mismatch of the threshold voltage. The sizes of M_1-M_4 becomes that of Table 3.3 for the design example previously showed and using a unitary transistor of W=2 μm and L=5 μm.

Table 3.3: Sizing of M_1-M_4.

A = 10, B = 5	M_1	M_2	M_3	M_4
i_f	3	0.5	0.1	0.001
W (μm)	2	2	5*2	20*2
L (μm)	5*5	2*5	5	5
Area (μm^2)	50	20	50	200

The OA design requires a relationship between the additional area and power consumption which needs to be considered. In this design we consider that an additional 5% of current consumption is adequate. Observe that the inversion level of the input transistors of the OA are set to 3, since their sources are grounded and their gate voltages are equal to V_{T0}. Once we know the current and the inversion level (i_f) it is easy to size the transistors. Table 3.4 presents the sizing of the auxiliary structures as the OA, start-up and current sources.

Table 3.4: Sizing of the auxiliary structures.

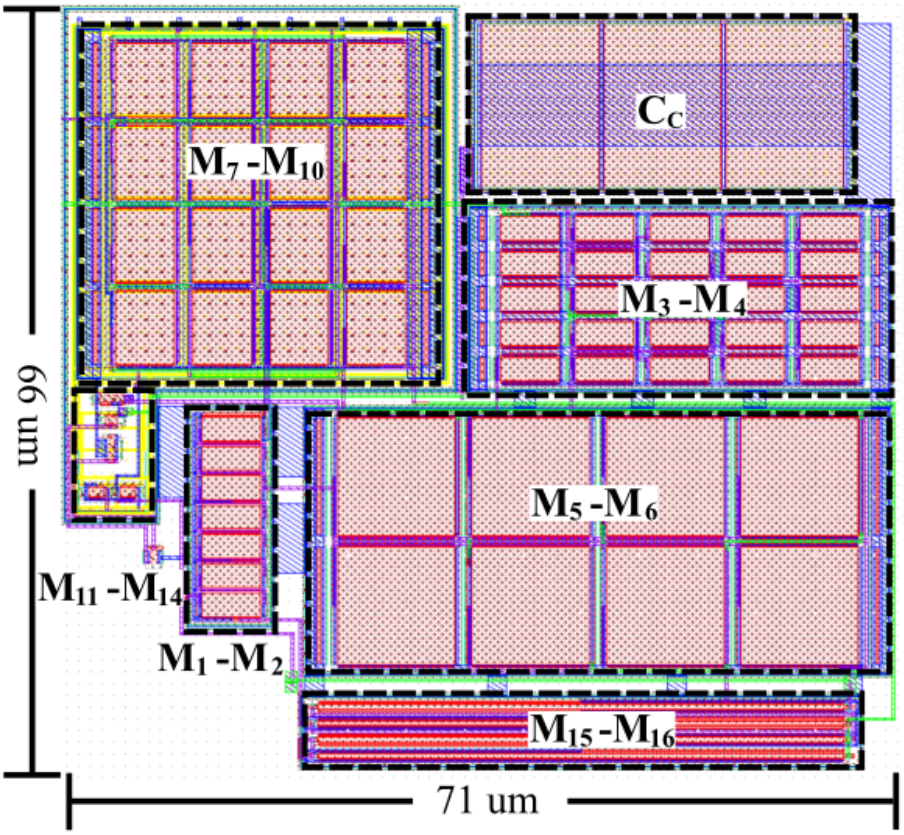
	$M_{5\text{-}6}$	$M_{7\text{-}10}$	M_{11}	M_{12}	$M_{13\text{-}14}$	$M_{15\text{-}16}$
i_f	0.1	0.08	- - -	- - -	3	0.13
W (μm)	10*4	6.25*4	0.5	0.5	0.4	0.76
L (μm)	10	5	1	2	45*2	1.3
Area (μm^2)	400	125	0.5	1	36	1

3.2.3 Simulations Results

The results presented here are for Cadence Virtuoso post-layout simulations of our design implemented in IBM 130 nm process. The layout takes into consideration the good layout matching practices such as common-centroid placement and dummy structures. The occupied silicon area is 0.0047 mm^2, as shown in Figure 3.18.

The MOSFETs used in this implementation are standard I/O type, that present higher threshold voltage and also allow a higher VDD voltage (VDD_{max}) than the core transistors in this CMOS process. Figure 3.19(a) presents the V_{T0} variation over temperature estimated by the (g_m/I_D) method introduced in Section 1.1.2 (labeled UICM), and simulated in the V_{T0} monitor circuit of Figure 3.17 (labeled Vout). As one can see the two

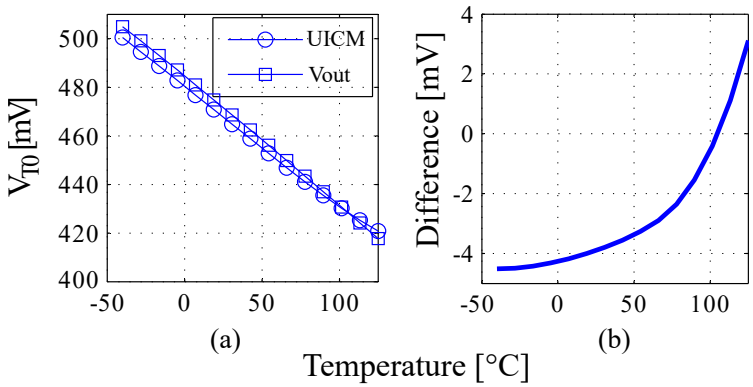
Figure 3.18: Complete Circuit Layout



Source: The own author

lines are very close. Figure 3.19(b) presents the difference between the V_{T0} from the circuit (Vout) and the one estimated analytically by the UICM model, resulting a maximum deviation around 5 mV (1%).

Figure 3.19: (a) V_{T0} value and (b) difference from the simulated monitor circuit (Vout) and from the g_m/I_D model (UICM) vs temperature



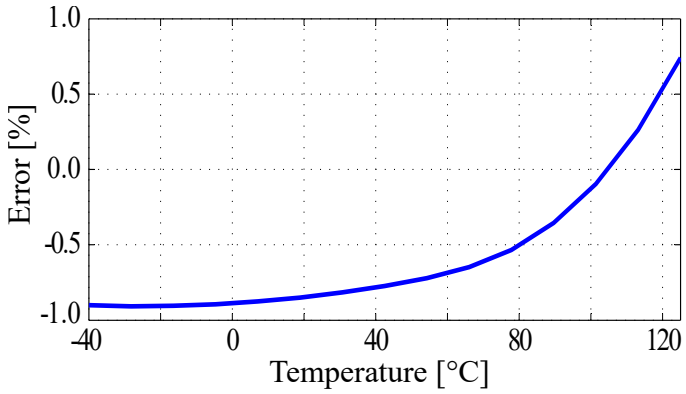
Source: The own author

The error calculation is defined by (3.13) and presented in Figure 3.20. The monitor circuit tracks the modeling threshold voltage with an absolute error lower than 1% for the

-40 to 125 °C temperature range.

$$\epsilon(\%) = 100 \times \left(\frac{V_{out} - V_{UICM}}{V_{UICM}} \right) \tag{3.13}$$

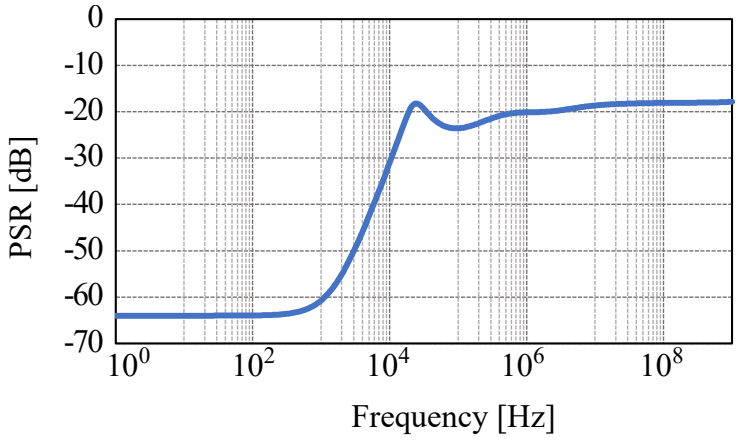
Figure 3.20: Percentual error in the V_{T0} monitored value over temperature



Source: The own author

Figure 3.21 shows the Power Supply Rejection (PSR) of the output, resulting -64 dB from 0 Hz to almost 1 kHz and for $V_{DD} = 1.2$ V, which is almost double of the result obtained in previous works.

Figure 3.21: PSR of the output over frequency



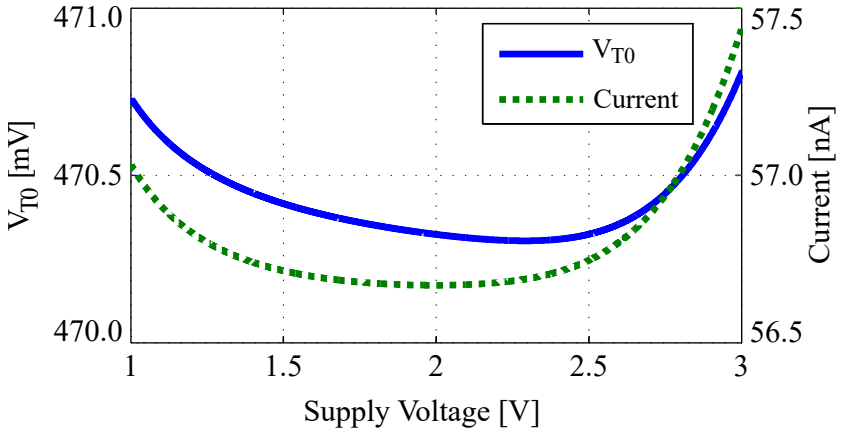
Source: The own author

The circuit consumes only 57 nW at room temperature for $V_{DD} = 1.2$ V, and reaches a maximum of 70 nW at 125 °C. Figure 3.22 presents the V_{T0} output value and the current consumption over the supply voltage. The maximum line sensitivity is around 250 ppm/V, while the current consumption sensitivity is 378 pA/V, both at 27 °C and for a supply voltage range from 1 V to 3 V.

The start-up behavior was simulated for the corner process cases and presented in Figure 3.23, resulting a settling time of less than 25 ms in the worst case at 125 °C, which is acceptable for our proof of concept.

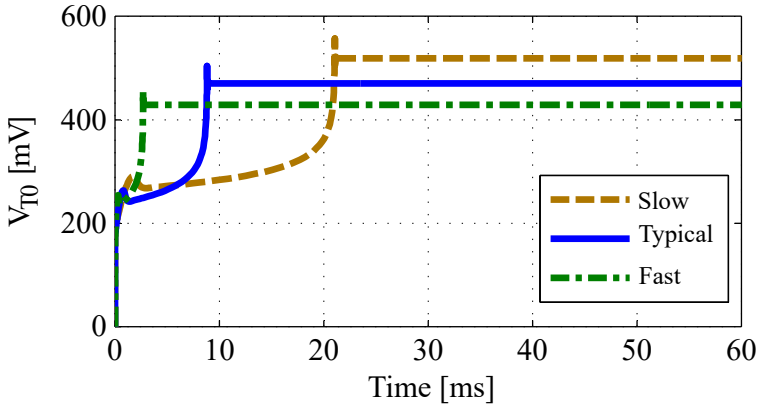
Table 3.5 presents a comparison of recently published threshold voltage monitors. One of the great advantages of our topology is the very low line sensitivity and high

Figure 3.22: V_{T0} monitored and Current consumption vs V_{DD} voltage



Source: The own author

Figure 3.23: Settling time for the corner process cases



Source: The own author

PSR, providing low error and comparable power consumption, working in a wide range of temperature and voltage. Except for the current consumption that depends on I_{SQ} and varies from one process to another, similar results can be obtained in other technologies. This circuit was fabricated using the Mosis's educational program, and we soon expect to measure it.

3.2.3.1 Fabrication Variability Effects

The impact of the fabrication variability effects on the output V_{T0} value was analyzed using Monte Carlo (MC) simulations, where local mismatch effects and average process variations were simulated separately with 400 runs each. Figure 3.24 shows the MC histograms for the error of the extracted V_{T0} from circuit simulations, with respect to the theoretical V_{T0} obtained from the analytical model, as this model does not take variations and mismatches into account. For average process MC all transistors have their parameters changed equally in each run - Figure 3.24 (top histograms). For local mismatch MC, the parameters of each transistor are varied individually in each run - Figure 3.24 (middle histograms). Both effects are also taken into account in a full variability analysis, shown in Figure 3.24 (bottom histograms). The results presented are for $V_{DD} = 1.2$ V and under three different temperatures: -40, +27 and +125 °C.

As shown in the design methodology, the circuit performance depends only on geo-

Table 3.5: Comparison of recent V_{T0} monitor circuits

Characteristic	This Work	[1]	[2]	[3]	[4]	[5]	[6]	Units
PSR @100 Hz	-63.9	-38.9	-30	- - -	- - -	- - -	- - -	dB
Line Sensitivity	252	3600	46000	480	2562	8000	554	ppm/V
Temperature Range	-40 to 125	-40 to 125	-40 to 125	0 to 100	0 to 100	20 to 80	-50 to 100	°C
Supply Range	0.97 to 3	0.6 to 1.8	0.6 to 1.2	1 to 3.6	1.9 to 2.1	2 to 2.5	3.5 to 6.5	V
Max. Error	1	1.64	1.3	- - -	11	4.3	4.9	%
Power	57	23	23	50000	290000	387500	- - -	nW
Model	UICM	UICM	UICM	quadratic	quadratic	quadratic	quadratic	

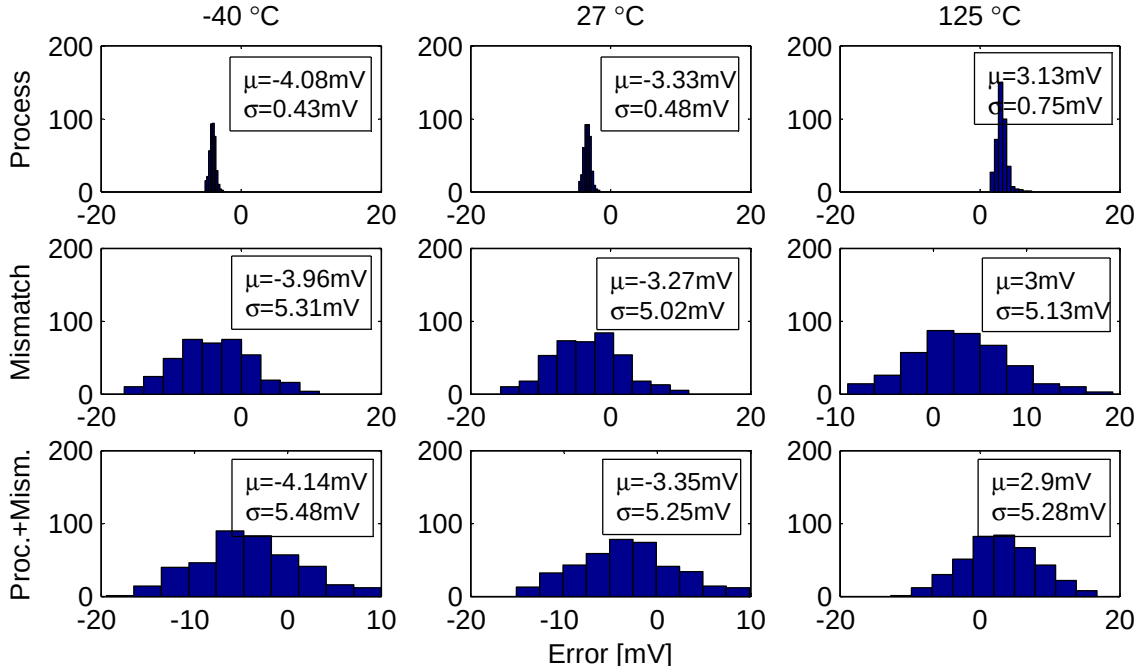
[1](MATTIA et al., 2015) [2](MATTIA; KLIMACH; BAMPI, 2014a)

[3](VLASSIS; PSYCHALINOS, 2007) [4](WANG; TARR; WANG, 2004)

[5](SENGUPTA, 2004) [6](CILINGIROGLU; HOON, 2003)

metrical factors, being less sensitive to process than to mismatch variations, which affect the currents balance and therefore the circuit behavior. One can verify the different sensitivities when comparing the maximum spread that results from MC results: $\epsilon(\pm 3\sigma) = -4.08 \pm 1.29$ mV for process, that is significantly less than $\epsilon(\pm 3\sigma) = -3.96 \pm 15.93$ mV for mismatch. Finally a maximum total error in the worst condition, combining process and mismatch analysis, yields $\epsilon(\pm 3\sigma) = -4.14 \pm 16.44$ mV that falls within ± 20 mV or 4.25% for the whole operating temperature range.

Figure 3.24: Monte Carlo simulations for Process (top), Mismatch (middle) and both variability effects (bottom)



Source: The own author

3.3 Proposed PMOS V_{T0} monitor

Although in CMOS processes the NMOS and PMOS transistors have the same importance, it is clear, when you look to the bibliographic study, that the topologies that extract the V_{T0} of PMOS transistors have been less developed in comparison with the topologies that extract V_{T0} of NMOS transistors.

In general, the authors were limited to design monitors that extract the V_{T0} value of NMOS transistors. Some mentioned that the concept that they used to obtain the NMOS V_{T0} monitor could be also used to obtain a PMOS version, but without any implementation to prove it.

Unfortunately, when the same NMOS technique is applied to implement a PMOS V_{T0} monitor, the output value results referred to the supply voltage, rather than to ground. This happens because normally the structures are just mirrored (the NMOS structures in the circuit are replaced with PMOS, and the PMOS structures are changed with NMOS structures), and consequently the V_{T0} value that in a NMOS monitor appears referred to ground, in the PMOS implementation appears referred to the supply voltage. This demands to know exactly the VDD value and also forces to make additional calculations, which are undesirable conditions when the circuit aims to act in a complex system.

In this section, a new PMOS V_{T0} monitor is presented. The monitor gives the threshold voltage of a PMOS transistor in module. The V_{T0} value is delivered by the circuit referenced to ground, and tracks the theoretical value with a low error, in a wide range of temperature and voltage.

3.3.1 Circuit Description

In Section 3.2 was introduced a NMOS V_{T0} monitor circuit. Then, although the monitor circuit was designed to obtain the threshold voltage of a transistor, it can also work as a current generator. Actually, in the operation point in which the circuit is designed to operate ($i_{f1} = 3$), it generates a current that is proportional to I_{SQ}, meaning that the circuit can be also used as an I_{SQ} current generator, as shown in CAMACHO-GALEANO; GALUP-MONTORO; SCHNEIDER (2005).

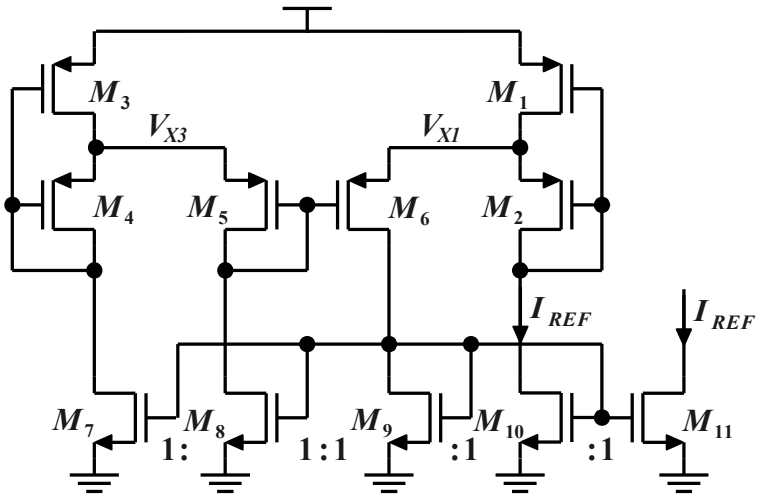
This characteristic is here exploited. The circuit of Figure 3.25 presents a PMOS version of the I_{SQ} current generator. Although the circuit analysis was already done in Section 3.2, it is again presented here for convenience, and also in order to prove that the same equations that were already presented are valid for the PMOS version of the circuit.

The DC operating point of the PMOS I_{SQ} current generator is established by the equilibrium condition of the complementary-to-absolute-temperature (CTAT) voltages generated by two PMOS self-cascode (SC) cells that are clamped together. One of the SC cells operates in moderate inversion ($M_{1,2}$), while the second SC cell operates in weak inversion ($M_{3,4}$).

Transistors M_5-M_{10} act as a voltage-following current mirror (GILBERT, 2004), making all the currents equal to I_{REF}, and forcing $V_{X1} = V_{X3}$.

The voltage at the intermediate node of a self-cascode cell, referred to its bulk, has been already shown to be a PTAT voltage, whenever both transistors have the same V_{T0}, and operate at constant inversion levels (ROSSI; GALUP-MONTORO; SCHNEIDER, 2007). Also, it has been shown that the absolute value of the PTAT voltage (and its derivative) can be adjusted by the inversion levels of the transistors. Additionally, the lower transistors in the SC cell ($M_{2,4}$ in Figure 3.25) have to be in saturation, whereas the

Figure 3.25: I_{SQ} current generator



Source: The own author

upper transistors $M_{1,3}$ are in triode. The use of (B.1), (B.3) and (B.4) demonstrates that

$$V_{BS2} = V_{DD} - V_{X1} = \phi_t[F(i_{f1}) - F(i_{f2})] \tag{3.14}$$

$$V_{BS4} = V_{DD} - V_{X3} = \phi_t[F(i_{f3}) - F(i_{f4})] \tag{3.15}$$

where V_{BS2} and V_{BS4} are ideally PTAT for any inversion level, as long as i_{f1-4} are kept constant over temperature. Consequently, if this condition is fulfilled, V_{X1} and V_{X3} will be CTAT voltages for any inversion level.

From (B.3) and (B.4), one can see that a PMOS with the source at the same voltage that its bulk, and with a gate voltage V_G equal to V_{T0}, operates under a constant forward inversion level equal to 3 ($i_f = 3$). Suppose that M_1 operates under such condition, being in the moderate inversion region. The current I_{REF} is defined based on the inversion levels of M_1 and M_2. Recalling that $i_{f2} = i_{r1}$, allows us to write

$$I_{D1} = S_1 I_{SQ}(3 - i_{f2}) = I_{D2} + I_{D5} = 2I_{REF} \tag{3.16}$$

where i_{f2} defines the voltage V_{BS2} according to (3.14). The inversion levels i_{f3-4} can then be defined to make $V_{BS2} = V_{BS4}$, and the circuit operates in equilibrium.

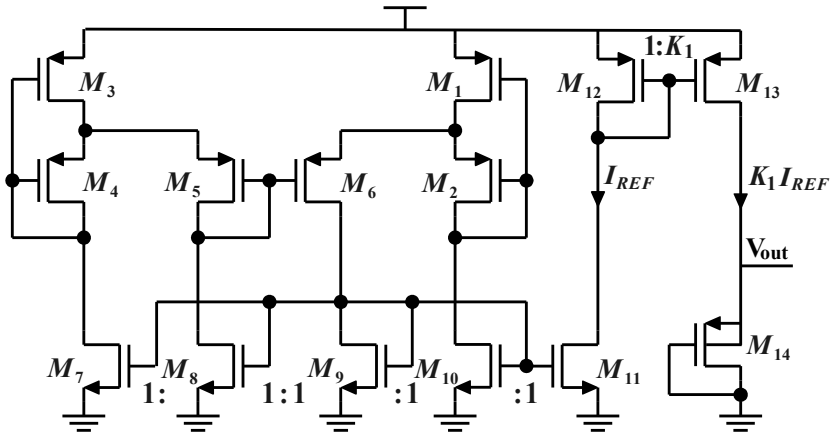
3.3.1.1 V_{T0} monitor

Once the I_{SQ} current generator operates properly, its output current is mirrored through the M_{12}-M_{13} PMOS current mirror into the M_{14} transistor, as Figure 3.26 shows. Then, considering that transistor M_{14} is always in saturation, and assuming that it operates under an inversion level $i_{f14} = 3$, its current is defined by

$$I_{D14} = K_1 I_{REF} = S_{14} I_{SQ} 3 \tag{3.17}$$

Finally, if M_{14} is in the $i_f = 3$ condition, and its source is at the same voltage of its bulk $V_{SB14} = 0$, it is obtained from (B.3) and (B.4) that necessarily $V_{BG14} = V_{T0}$. In Figure 3.26 you can noted that the gate of transistor M_{14} is grounded, therefore $V_G = 0$ and consequently $V_B = V_S = V_{T0}$.

Figure 3.26: PMOS V_{T0} monitor circuit.



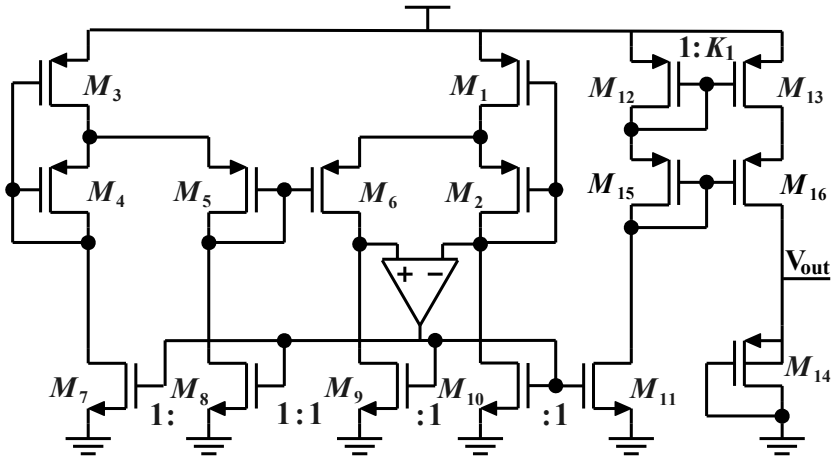
Source: The own author

3.3.1.2 PSR and Line Sensitive improvement

The basic topology shown in Figure 3.26 exhibits a high sensitivity to changes in the supply voltage (V_{DD}), resulting in a poor LS and PSR. One solution is to include an Operational Amplifier (OA)(BAKER, 2004) that add a high gain feedback path resulting an effective increase in the output impedance, as was done in Sections 3.1 and 3.2.

Figure 3.27 shows the V_{T0} monitor circuit with the OA connected. Note that M_9 is no longer diode-connected, so its drain can move to the same voltage of the drain of M_{10}. The OA compares the drain voltage of M_2 with the drain voltage of M_6 and force them to be approximately equal, adjusting the current mirror bias.

Figure 3.27: PMOS V_{T0} monitor circuit, with LS and PSR improvements.

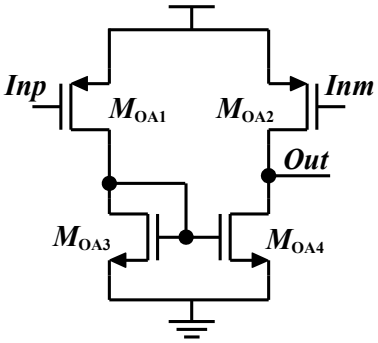


Source: The own author

Figure 3.28 shows the low voltage pseudo-differential amplifier that was used, where the NMOS transistors form a current mirror while the PMOS devices operate as a differential amplifier. When both inputs are equal, both branches of the mirror are in equilibrium. If the inputs are not equal, this imbalance causes the amplifier output to swing up or down providing the desired action. By doing this, the current sensitivity to changes in V_{DD} of the I_{SQ} generator is improved, and therefore, the line sensitive in the output is improved.

Finally, the circuit can be improved even more by changing the simple mirror M_{12}-M_{13} by a cascode mirror. For this, transistor M_{15} and M_{16} have been added.

Figure 3.28: PMOS pseudo-differential Amplifier.



Source: The own author

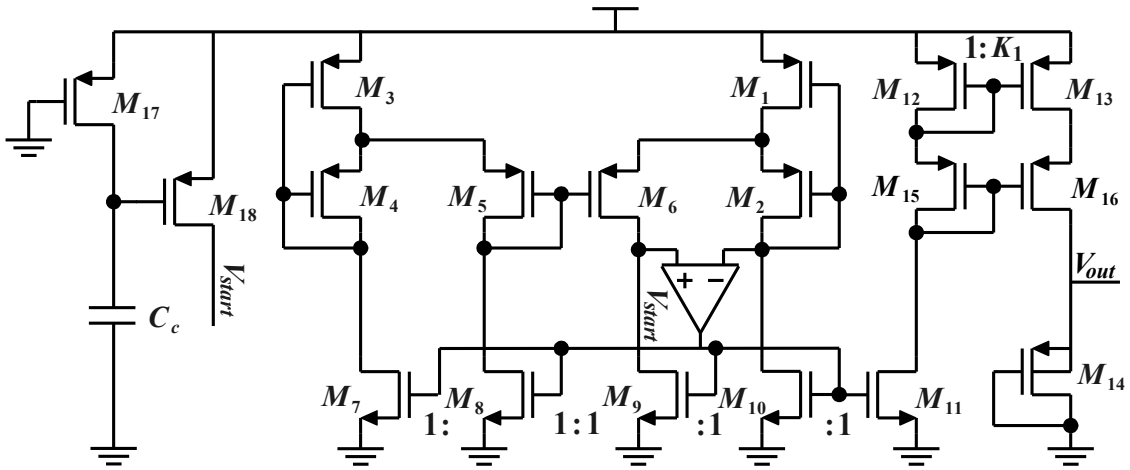
3.3.1.3 Start-up Circuit

As the V_{T0} monitor circuit is a self-biased structure, it presents two DC stability operation points, one in the desired bias condition in which the two SCs have the same intermediate voltage, and another when the current in all branches is zero. A Start-up circuit that can prevent the zero-current condition is necessary.

Figure 3.29 shows the topology with the start-up circuit. Note that the current through M_{17} and M_{18} is zero when the circuit is in normal operation because of the loaded C_C capacitance, which forces $V_{DS17} = 0$ and $V_{GB18} = 0$. This results in a zero extra current consumption, which is a desirable characteristic in nano-Watt circuits.

The start-up circuit works as follows: supposing the capacitor C_C discharged, when the V_{DD} voltage starts to increase, M_{18} drives a current into the M_9 transistor initializing the circuit, and simultaneously, M_{17} delivers a current into the capacitor C_C, charging it, and eventually moving M_{17} to deep triode, and M_{12} to the cut-off state.

Figure 3.29: Proposed PMOS V_{T0} monitor circuit.



Source: The own author

3.3.2 Circuit Design

In the previous analysis it was shown that the same currents I_{REF} pass through all the branches of the I_{SQ} current generator. Now in the design section, it is convenient that these currents I_{REF} are defined as a fraction of the specific current. I_{REF} is then normalized as follows:

$$I_{REF} = I_{SQ}/A \tag{3.18}$$

where A is a design constant factor that can be used to determine the power consumption of the whole circuit ($I_{\text{TOTAL}} = I_{REF}(5 + K_1)$). We also define another design constant $B = i_{f2}/i_{f3}$, that sets the ratio between the inversion levels of the transistors of each SC cell.

Now, to start the design we can chose $i_{f1} = 3$, and then V_{BS2} is determined solely from i_{f2}, according to (3.14). Then the forward inversion level of M_3 can be defined by the design constant B, and finally the ratio i_{f4}/i_{f3} can be adjusted to make V_{BS4} equal to V_{BS2}.

As a design example we choose $i_{f2} = 0.5$, $A = 10$ and $B = 5$, leading to $V_{BS2} = V_{BS4} = 58$ mV and $i_{f3} = 0.1$. Once V_{BS4} and i_{f3} are known, the value of $i_{f4} = 0.001$ is easily obtained, and the sizing of the transistors can be determined from (B.1) and (3.11).

A design technique that can be used to improve the circuit behavior is the implementation of the main transistors (M_1-M_4) through the parallel and series composition of unitary devices, that are assumed to have the same process characteristics (V_{T0}, I_{SQ}, n, and so on). This technique also facilitates the implementation of common-centroid layout structures, improving the circuit, since a regular layout and the use of dummy devices can minimize the mismatch of the threshold voltage. After some optimizations, and using unitary transistors of W=2 μm and L=5 μm, the sizes of M_1-M_4 and M_{14} for the design example previously described, becomes that of Table 3.6.

Table 3.6: Sizing of M_1-M_4 and M_{14}.

A = 10, B = 5	M_1	M_2	M_3	M_4	M_{14}
i_f	3	0.5	0.1	0.001	3
W (μm)	2	2	5*2	22*2	2
L (μm)	5*5	2*5	5	5	5*5
Area (μm^2)	50	20	50	220	50

The OA design requires a relationship between the additional area and power consumption which needs to be considered. In this design, we consider that an additional 20% of current consumption is adequate. Observe that the inversion level of the input transistors of the OA are set to 3, since their sources are in the supply voltage and their gate voltages are equal to V_{DD}-V_{T0}. Once we know the current and the inversion level (i_f) it is easy to size the transistors. Table 3.7 presents the sizing of the auxiliary structures as the OA, start-up and current sources.

3.3.3 Simulation Results

The results presented here are for Spectre simulations of our design implemented in IBM 130 nm process. As the design circuit was done taking care of use series and parallel

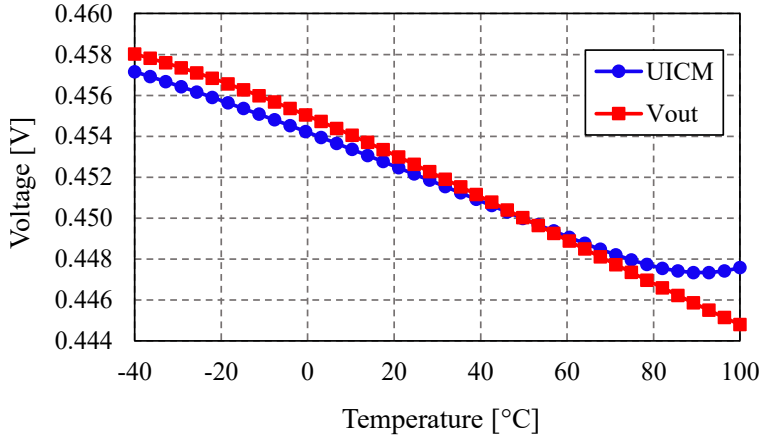
Table 3.7: Sizing of the auxiliary structures.

	$M_{5\text{-}6}$	$M_{7\text{-}11}$	$M_{12,15}$	$M_{13,16}$	M_{17}	M_{18}	$M_{OA1,2}$	$M_{OA3,4}$
i_f	0.0001	0.18	0.025	0.025	- - -	- - -	3	0.78
W (µm)	5*9	5	5*4	5*8	0.5	1	2	2
L (µm)	5	35	5	5	2	1	20	20
Area (µm^2)	225	175	100	200	1	1	40	40

transistors, the layout can be implemented taking into consideration good layout matching practices such as common-centroid placement and dummy structures. An estimate of the occupied silicon area is 0.012 mm^2.

The MOSFETs used in this implementation are standard I/O type, that present higher threshold voltage and also allow a higher VDD voltage (VDD$_{max}$) than the core transistors in this CMOS process. Figure 3.30 presents the V_{T0} variation over temperature, estimated by the (g_m/I_D) method presented in Section 1.1.2 (labeled UICM), and simulated in the V_{T0} monitor circuit of Figure 3.29 (labeled Vout). As one can see, the two lines are very close, indicating that the error is small in the entire temperature range.

Figure 3.30: Difference between the theoretical value of V_{T0} (UICM) and the simulated output of the V_{T0} monitor (Vout).

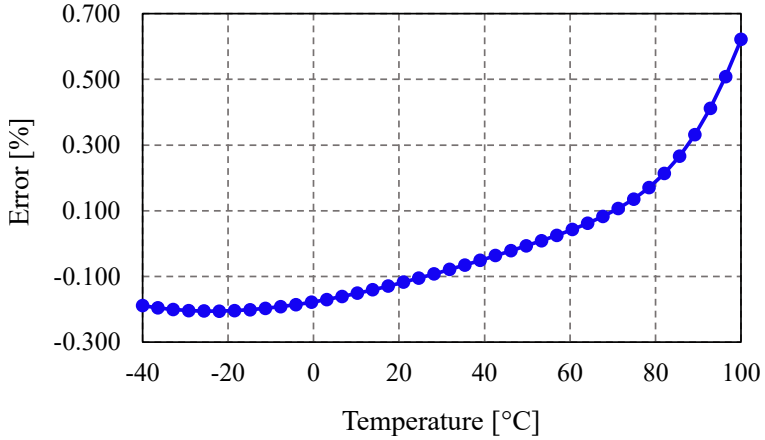


Source: The own author

Figure 3.31 shows the error in the output of the V_{T0} monitor. This error was found using the definition used in Sections 3.1 and 3.2 that is described by (3.8). As result, it is obtained that the circuit tracks the ideal threshold voltage with an error inferior to 1% under the -40 to 100 °C temperature range.

The PSR has been simulated between 0.1 Hz and 1 GHz for 3 different stages of the monitor circuit for the purpose to prove the efficiency of the techniques used to improve the LS and the PSR. First, it was simulated the original circuit without any improvement. This results in a PSR of –29 dB at 100 Hz, which indicate that the circuit does not reject to much the variations in the supply voltage, as was expected. After that, it was simulated the circuit adding the OA. As the Figure 3.32 shows, the circuit improves its behavior in almost -20 dB for low frequencies, presenting -50 dB at 100 Hz of PSR. Finally, the PSR of the circuit including the OA, and the cascade current mirror was simulated, resulting

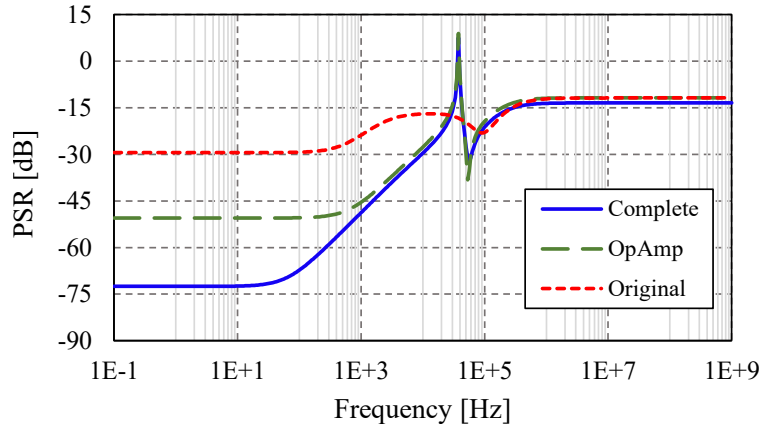
Figure 3.31: Percentual error in the V_{T0} monitored by the circuit over temperature.



Source: The own author

in a PSR at 100 Hz lower than -67 dB. This result together with the LS, confirms that the approach used in order to improve the output impedance works properly.

Figure 3.32: PSR at the output over frequency, for the original circuit (red), with OA connected (green), and final implementation (blue).



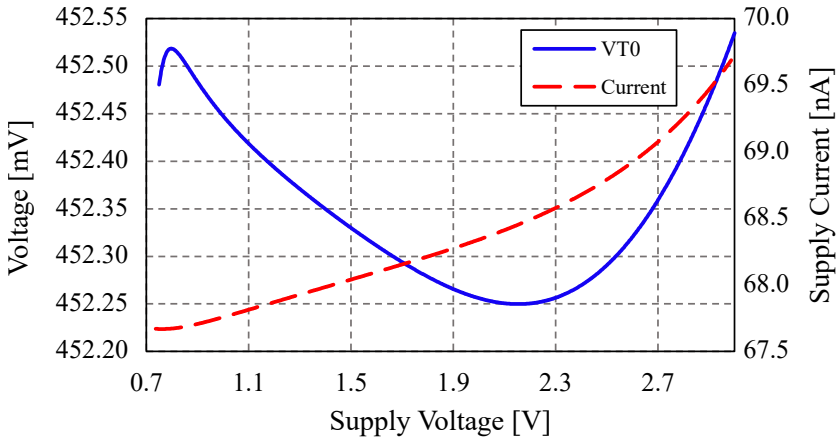
Source: The own author

As shows Figure 3.33 the circuit starts operating at around 0.75 V. The line sensitivity of the output is really good when compared with other topologies, being 120 µV/V from 0.7 V to 3 V, while the current consumption sensitivity is 897 pA/V. The whole circuit consumes 67.9 nA at 27 °C.

Start-up behavior of the circuit was simulated for typical conditions and also for -40 and 100 °C in order to study the maximum settling time and the stability of the circuit. The simulations show that even in the worst case, that is -40 °C, the circuit presents a settling time less than 50 ms, which is acceptable for our proof of concept.

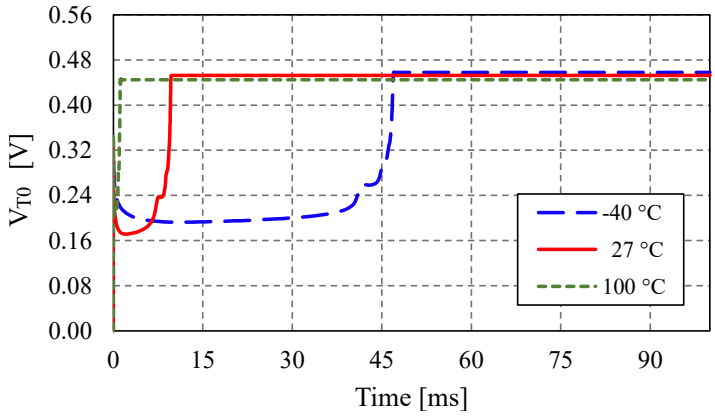
To analyze the fabrication variability of the error, Monte Carlo (MC) simulations were done separately for local mismatch effects and average process variations, with 200 runs each. For average process MC all the transistors have their parameters changed equally in each run - Figure 3.35 (top histograms). For local mismatch MC, the parameters of each transistor are varied individually in each run - Figure 3.35 (middle histograms). Both effects are taken into account in a full variability analysis, shown in Figure 3.35

Figure 3.33: V_{T0} monitored and Current consumption vs V_{DD} voltage.



Source: The own author

Figure 3.34: Start-up simulation, for -40 °C (blue), 27 °C (red), and 125 °C (green).

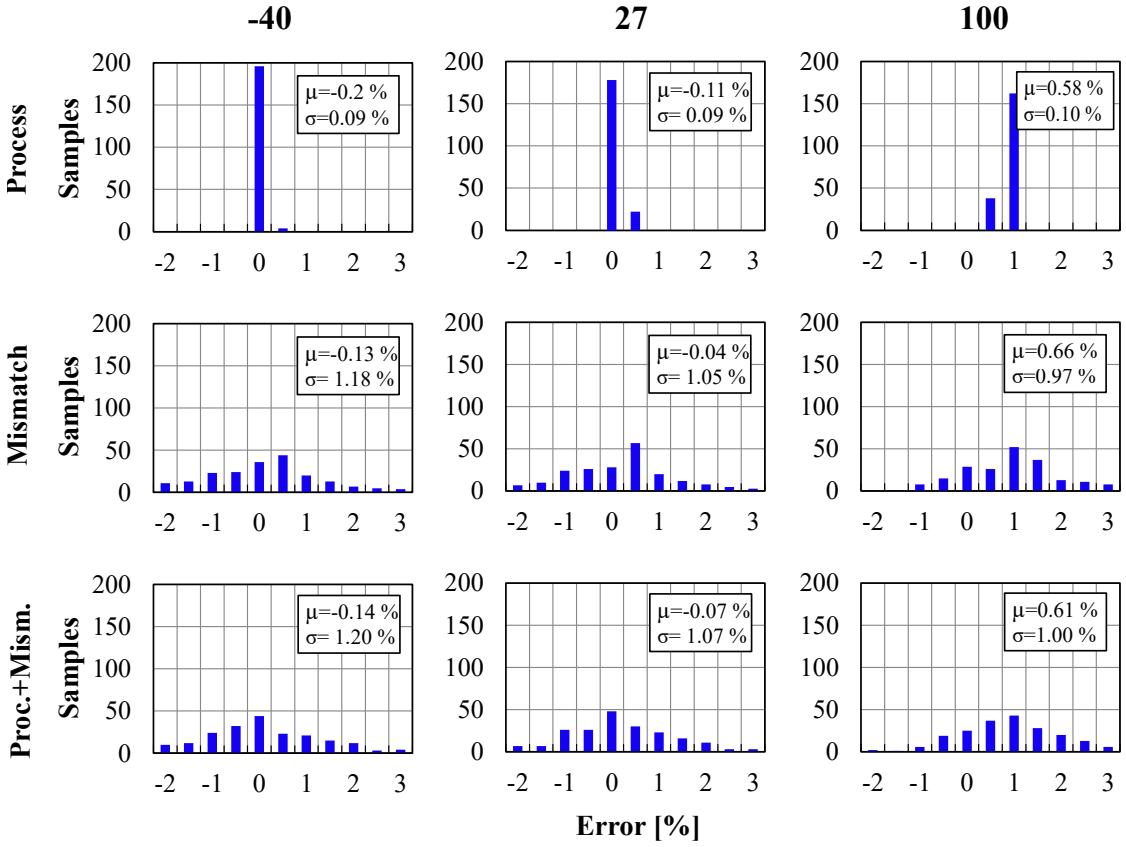


Source: The own author

(bottom histograms). The results presented are for $V_{DD} = 1.2$ V under three different temperatures, -40, 27 and 100 °C.

As shown in the design methodology, the circuit's equilibrium point depends only on geometrical factors. It is thus less sensitive to average process variations, where the output tracks the threshold voltage value with a maximum error (mean and standard deviation) of $\epsilon(\pm 3\sigma) = 0.58 \pm 0.1\%$, comprising 99.7% of the samples. Local mismatch analysis, however, affect the current mirrors and aspect ratio gains that define this equilibrium, resulting in a higher spread of $\epsilon(\pm 3\sigma) = 0.66 \pm 0.97\%$. A combined analysis yields a maximum error of $\epsilon(\pm 3\sigma) = 0.61 \pm 1\%$ for the whole operating temperature range. The mean and the sigma of each temperature and variability condition is shown in Figure 3.35.

Figure 3.35: Monte Carlo simulations for Process (top), Mismatch (middle) and both variability effects (bottom).



Source: The own author

4 APPLICATIONS

As was explained in the introduction, there are several applications in which the V_{T0} monitors can be used. This chapter aims to show, not only some possible applications of the proposed circuits, but also, that the characteristics presented for these, are enough to be part of a system. The first application exploited here consists in monitor and/or model the V_{T0} in a fabrication process. For this, a configuration with transistors of different lengths is biased with an I_{SQ} current. This implementation, allows obtaining the V_{T0} value in the entire range of lengths that a process offers, consuming only a few nano-Watts, and occupying little space. Then, is presented a voltage reference, working in the nano-Watt range, with high PSR and low LS. The circuit is CMOS-only being possible its implementation in any CMOS digital process. Finally, the compensation of process variations in circuits is investigated. For this, a V_{T0} monitor is used to cancel the dependence that a circuit presents with V_{T0}. Also, it is studied the possibility to replace the biasing circuit by a I_{SQ} current generator, taking advantage of the less process variability that the I_{SQ} presents.

4.1 Process Monitor and Modeling

Monitoring a CMOS process fabrication is perhaps the most obvious application of a V_{T0} monitor circuit. However not all the implementations can accomplish this task as some of them are not fully integrated or are dependent of extra calculations.

Different of that all the V_{T0} monitors that was presented here are fully integrated, CMOS-only, present nano-Watt consumption and occupy little space making possible to add many of them throughout the entire wafer and, in this way, it can be obtained a profile of the V_{T0}. This can help finding errors or variations in some important fabrication parameters as the oxide thickness or the dopants concentration.

The circuits presented herein were designed following a process independent methodology. This allows obtaining the sizing of the transistors without knowing the parameters of the process in which the circuit will be fabricated. Therefore, it is possible to implement the V_{T0} monitors in a process that has not been characterized yet and thereby model the V_{T0} value of the process. Off course, the V_{T0} value obtained is only valid for the UICM model forcing to work with this model to describe the entire process, or to somehow get an equivalence between the V_{T0} value of this model and the V_{T0} value described by other models.

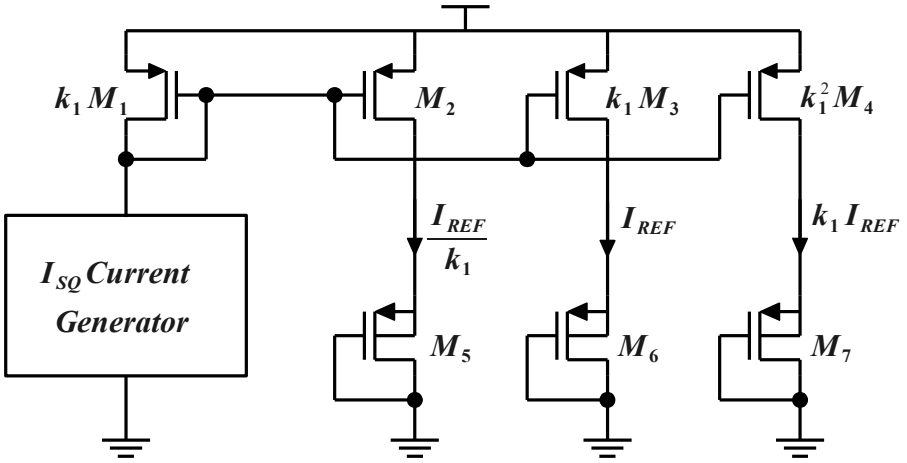
In general, the V_{T0} monitors deliver the V_{T0} value of a transistor with a given channel width and length. This could be a limitation when a new process is being characterized, because the V_{T0} changes with the value of both the channel length and width. Then, to properly model the V_{T0}, it is necessary to obtain the V_{T0} for transistors with lengths

that are between the entire range of possibilities that the process offers. Therefore, to accomplish that, many V_{T0} monitors must be designed and implemented (one by every length that is desired), turning it in a slowly and hardly task.

Then, it is desirable to obtain with only one design, and implementation, the V_{T0} of several transistors with different sizes. This can be done using the circuit of Figure 4.1. The concept is the same that was used to obtain the V_{T0} of a PMOS transistor in Chapter 3.3, and exploits the fact that a V_{T0} monitor delivers a voltage equal to the threshold voltage, and also a current that is proportional to the I_{SQ}.

Now, suppose that the I_{SQ} current generator of Figure 4.1 delivers a current I_{SQ}/A. This current is mirrored by the PMOS current mirror M_1-M_4, into the M_5-M_7 transistors, which leads $I_{REF} = I_{D1} = I_{D5}K_1 = I_{D6} = I_{D7}/K_1$. Once all the currents are known, using (B.1) and (B.2), it is possible to obtain the size of transistors M_5-M_7, in order to get an inversion level $i_f = 3$ in each one. Then, as these transistors are in the saturation region, their drains are grounded, and their sources are at the same voltage of their bulks, from (B.3) and (B.4), it is obtained that the voltage at the source of transistors M_5-M_7 should be equal to the V_{T0} value of each one.

Figure 4.1: Process monitor circuit.



Source: The own author

As a design example, we use the I_{SQ} current generator of Figure 3.25, with the same sizing of Chapter 3.3. This circuit generates a current equal to $I_{SQ}/10$. Then, defining $K_1 = 2$, the aspect ratio of transistors M_5-M_7 is set by (4.1), (4.2) and (4.3) respectively.

$$I_{D5} = \frac{I_{SQ}}{10 * 2} = S_5 I_{SQ} 3 \tag{4.1}$$

$$I_{D6} = \frac{I_{SQ}}{10} = S_6 I_{SQ} 3 \tag{4.2}$$

$$I_{D7} = \frac{2I_{SQ}}{10} = S_7 I_{SQ} 3 \tag{4.3}$$

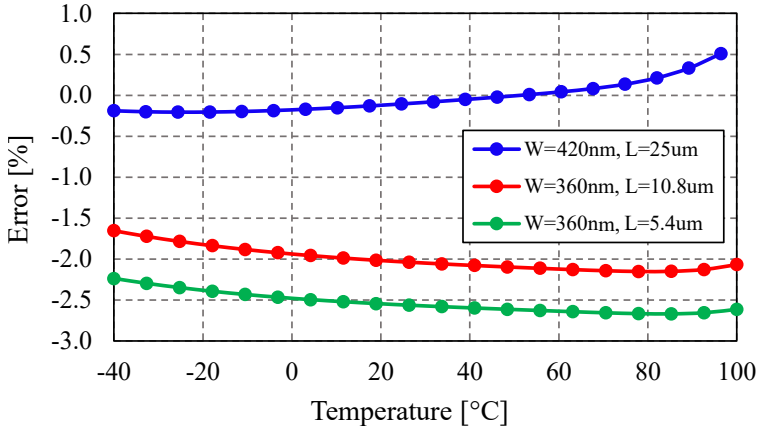
This design was implemented in IBM 130 nm process, using standard I/O MOSFETS. The final sizes of the transistors are presented in Table 4.1. The circuit was simulated using Spectre, and the results show that the error in the extracted $V_{T)}$ is larger for smaller channel lengths. This happens because the I_{SQ} current generator was designed to deliver an I_{SQ} current of a transistor with L=25 μm and therefore, the farther the lengths of the

used transistors are from this value, the greater the error in the output. Then, as Figure 4.2 shows, the lowest error is presented by transistor M_5 which has an L=25 µm and the largest error is presented by transistor M_7 which has an L=5.4 µm. Even so, the error is smaller than 3% in the -40 to 100 °C temperature range for all the cases.

Table 4.1: Sizing of M_1-M_4 and M_{14}.

A=10, K_1=2	M_1	M_2	M_3	M_4	M_5	M_6	M_7
i_f	0.025	0.025	0.025	0.025	3	3	3
W (µm)	10*2	10	10*2	10*4	0.42	0.36	0.36
L (µm)	5	5	5	5	25	10.8	5.4
Area (µm^2)	100	50	100	200	10.5	3.9	1.95

Figure 4.2: Error in the output of the process monitor for different channel lengths.



Source: The own author

4.2 Voltage References

Voltage references are key building blocks used in analog, mixed-signal, RF and digital systems. Most of this systems demands an accurate reference independently from temperature and power supply variations. Moreover the growing market of Internet of things and mobile systems incorporate ultra low power design as another important constraint. The basic concept of providing a temperature independent voltage is adding two voltages, one proportional to absolute temperature (PTAT) and the other complementary to absolute temperature (CTAT).

The CTAT voltage can be implemented in two ways. The first one is using a forward biased p-n junction, resulting the well known bandgap reference circuit (BGR), which presents low process variation sensitivity and a slightly non-linear behavior over temperature(TSIVIDIS, 1980), besides the need for higher supply voltages (p-n junction needs around 0.5 to 0.6 V for biasing). The non-linear behaviour can be minimized using a curvature compensation technique, but this has a direct impact in power and area. The other way is using the MOSFET threshold voltage (V_{T0}) as in recent voltage references

(MAGNELLI et al., 2011), (SEOK et al., 2012),(ZENG et al., 2013) where low power consumption and supply voltages were reached. The temperature range of these references varies widely but nano-watt and sub-1 V supply are usual to many of them.

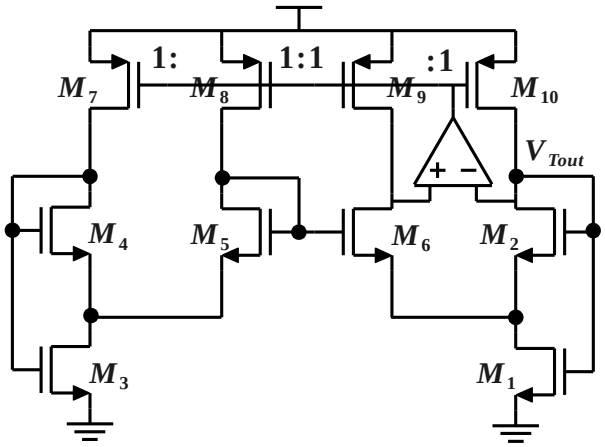
In this Section a voltage reference based on the sum of two almost linear temperature dependent terms is proposed. The threshold voltage monitor circuit presented in Chapter 3.2 provides the CTAT voltage, while the PTAT voltage is generated by two PMOS unbalanced differential pairs operating in weak inversion. The sum of these two linear terms results in a significant reduction of the thermal coefficient over a wide temperature range, while power consumption is kept low by the use of transistors operating in weak and moderate inversion.

The text is organized as follows: in section 4.2.1 is presented the proposed topology. A design methodology is developed in section 4.2.2, followed by simulation results in section 4.2.3.

4.2.1 Circuit Description

The balanced threshold voltage monitor circuit (GOMEZ C. et al., 2015), shown in Figure 4.3, is based on a self-biased current source topology proposed in CAMACHO-GALEANO; GALUP-MONTORO; SCHNEIDER (2005). Its DC operating point is established by the equilibrium condition of the PTAT voltages generated by two self-cascode (SC) cells that are clamped together. One of the SC cells operates in moderate inversion ($M_{1,2}$), while the other SC cell operates in weak inversion ($M_{3,4}$). Transistors M_5-M_{10} act as a voltage-following current mirror (GILBERT, 2004), making the currents equal in all the branches and forcing the voltage in the drains of transistors M_1 and M_3 to be equal. The Operational Amplifier (OA) adds a high gain feedback path that increase the effective output impedance of the mirror M_7-M_{10}, thereby improving the Line Sensitivity (LS) and the Power Supply Rejection (PSR).

Figure 4.3: V_{T0} monitor topology with improved LS and PSR.



Source: The own author

In a SC cell the upper transistors ($M_{2,4}$ in Figure 4.3) are usually in saturation, whereas the lower transistors $M_{1,3}$ are in triode. Using (B.1) and (B.3), and observing that $V_{D1(3)}$ can be calculated from the difference between $V_{GS1(3)}$ and $V_{GS2(4)}$, one can demonstrate that

$$V_{D1(3)} = \phi_t[F(i_{f1(3)}) - F(i_{f2(4)})] \tag{4.4}$$

where V_{D1} and V_{D3} are ideally PTAT for any inversion level, as long as i_{f1-4} are kept constant over temperature. This condition can be achieved by biasing the transistor with a current proportional to I_{SQ} (ROSSI; GALUP-MONTORO; SCHNEIDER, 2007).

From (B.3), one can see that a NMOS with grounded source and with a gate voltage V_G equal to the threshold voltage V_{T0} operates under a constant forward inversion level equal to 3 ($i_f = 3$). Suppose that M_1 operates under such condition, being then in the moderate inversion region. The current in each branch I_{DX} (I_{D7}-I_{D10}) is defined based on the inversion levels of M_1 and M_2. Recalling that $i_{f2} = i_{r1}$, it allows us to write

$$I_{D1} = S_1 I_{SQ}(3 - i_{f2}) = I_{D2} + I_{D6} = 2I_{DX} \tag{4.5}$$

The temperature dependence of the threshold voltage V_{T0} can be approximated (TSIVIDIS, 1987) by the linear equation

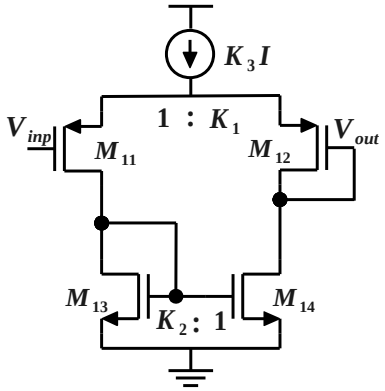
$$V_{T0}(T) = V_{T0(nom)} + \alpha T(T - T_{nom}) \tag{4.6}$$

where T is the absolute temperature, $V_{T0(nom)}$ is the threshold voltage at the nominal temperature T_{nom} and α is the negative thermal coefficient of the threshold voltage.

4.2.1.1 Unbalanced Differential Pair PTAT Generator

A common PTAT generator structure is the unbalanced differential pair (TSIVIDIS; ULMER, 1978), that operates in weak inversion. This circuit is shown in Figure 4.4, where K_1 and K_2 are the aspect ratio relationships between M_{11}-M_{12} and M_{13}-M_{14}, respectively.

Figure 4.4: Differential Pair PTAT generator schematic.



Source: The own author

In this circuit, transistors $M_{11,12}$ share the same source connection, and the PTAT voltage develops across the two gates. This circuit does not load the previous stage, since it is connected to a gate terminal. Using the UICM model, it is possible to extend the operation of this circuit to all inversion levels. Assuming that all transistors are in saturation and using (B.3), the PTAT voltage generated by the differential pair V_{DIFF} is given by:

$$V_{DIFF} = V_{G11} - V_{G12} = n\phi_t[F(i_{f11}) - F(i_{f12})] \tag{4.7}$$

This circuit generates a PTAT voltage independent of the inversion region, as long as the inversion levels i_{f11} and i_{f12} themselves are kept constant. This is achieved by biasing the differential pair with a current $I_{DIFF} = K_3 I_{SQ}$. Since the M_{11}-M_{12} pair

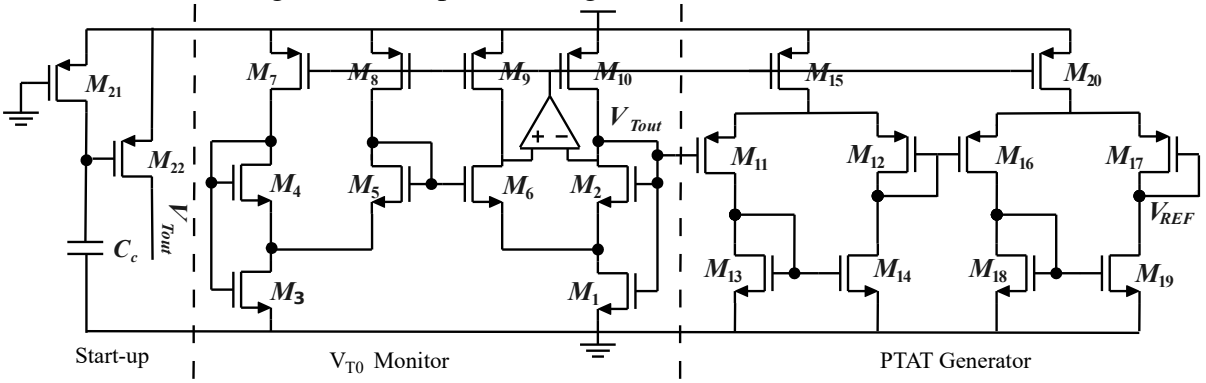
is biased by the M_{13}-M_{14} mirror, one can conclude that $i_{f11}/i_{f12} = K_1K_2$ and $I_{D12} = I_{DIFF}/(1+K_2)$, so (4.7) becomes

$$V_{DIFF} = n\phi_t \left[F\left(\frac{K_1K_2K_3}{(1+K_2)S_{12}}\right) - F\left(\frac{K_3}{(1+K_2)S_{12}}\right)\right] \tag{4.8}$$

4.2.1.2 Voltage Reference Circuit

The complete circuit of the proposed voltage reference can be seen in Figure 4.5. Transistors M_1-M_{10} form the V_{T0} monitor, whereas the unbalanced differential pairs are made of transistors M_{11}-M_{20}. Both unbalanced differential pairs present the same sizing in order to produce a total PTAT voltage that is twice that of a single cell.

Figure 4.5: Proposed voltage reference schematic.



Source: The own author

To bias the PTAT structures the current generated in the V_{T0} monitor is mirrored through PMOS transistors M_{15} and M_{20}. This current is proportional to I_{SQ} maintaining constant i_{f11}, i_{f12}, i_{f16} and i_{f17}. Also by the effect of the feedback added with the OA, the bias current presents low sensitivity to power supply variations improving the PSRR and LS at the output of the voltage reference.

The reference voltage V_{REF}, at the gate of M_{17}, is the sum of a CTAT term given by (4.6) and twice the PTAT term given by (4.7), hence:

$$V_{REF} = V_{T0} + 2n\phi_t[F(i_{f11}) - F(i_{if12})] \tag{4.9}$$

The start-up circuit works as follows: supposing the capacitor C_C discharged, when the V_{DD} voltage starts to increase, M_{22} drives a current into the SC $M_{1,2}$ initializing the circuit, and simultaneously, M_{21} delivers a current into the capacitor C_C, charging it, and eventually moving M_{21} to deep triode, and M_{22} to the cut-off state.

4.2.2 Circuit Design

As the current in all the branches of the monitor are equal and proportional to the I_{SQ} may be convenient to define that current as $I_X = I_{SQ}/A$, where A is a design constant factor. We also define another design constant $B = i_{f2}/i_{f3}$, that sets the ratio between the inversion levels of the transistors of each SC cell.

Since our objective is to set the gate-source voltage of M_1 to be equal to V_{T0}, we must chose $i_{f1} = 3$, and then V_{D1} is determined solely from i_{f2}, according to (4.4). Then the forward inversion level of M_3 can be defined by the design constant B, and finally the

ratio i_{f4}/i_{f3} can be adjusted to make V_{D3} equal to V_{D1}, that is a condition presented in section 4.2.1.

As a design example we choose i_{f2} = 0.5, A = 10 and B = 5, leading to V_{D1} = 58 mV (for 300 K) and i_{f3} = 0.1. Once V_{D3} = V_{D1} and i_{f3} are known the value of i_{f4} = 0.001 is easily obtained, and the sizing of the transistors can be determined from (B.1) and (4.5).

Once the threshold voltage monitor is designed, differentiating (4.9) with respect to temperature (considering n=1.2), and equating it to zero provides the necessary inversion level of M_{11}-M_{12} and M_{16}-M_{17}, together with the aspect ratios, and current gains K_1 and K_2 to make the output V_{REF} temperature independent.

Since M_{11}-M_{20} are all operating in weak inversion, the value of the current mirror gain K_3 does not affect heavily the PTAT voltage generated. It is designed to guarantee that all transistors are kept in saturation. The same is true for aspect ratios of the NMOS current mirrors M_{13}-M_{14} and M_{18}-M_{19}. Table 4.2 presents the size of the main transistors of the design example after some optimization steps.

Table 4.2: Sizing of the core transistors of the voltage reference.

	M_1	M_2	M_3	M_4	M_{11}	M_{12}	M_{13}	M_{14}
i_f	3	0.5	0.1	0.001	0.25	0.025	0.06	0.06
W (μm)	2	2	5*2	20*2	2	2*3	2*4	2
L (μm)	5*5	2*5	5	5	17.1	20	20	20
Area (μm^2)	50	20	50	200	34.2	120	160	40

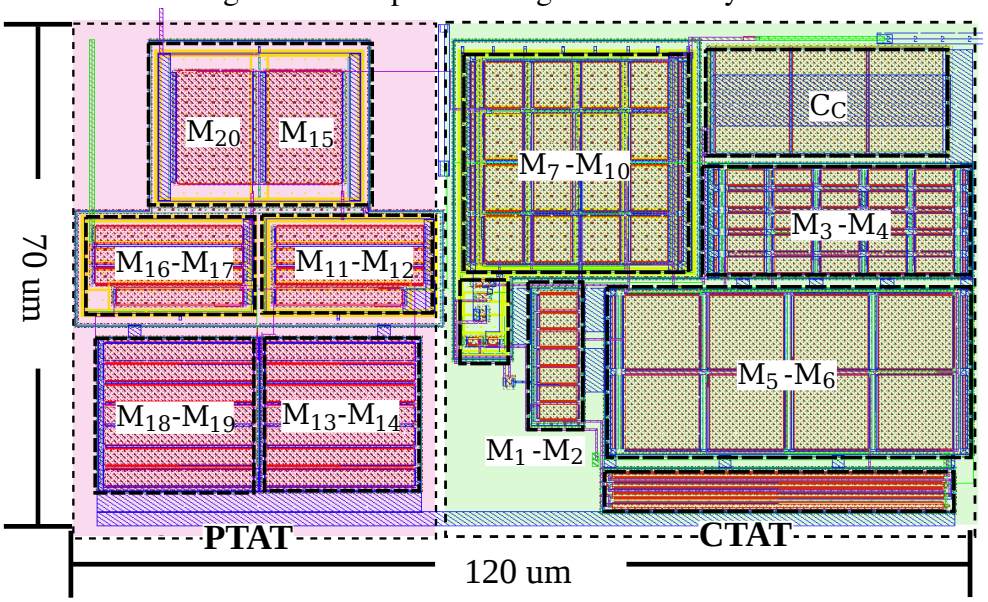
4.2.3 Simulations

Results presented here are for Cadence Virtuoso post-layout simulations of our design implemented in IBM 130 nm process. Layout takes care of good matching practices such as common-centroid placement and dummy structures. The occupied area is 0.0084 mm^2, as shown in Figure 4.6. MOSFETs used in this implementation are I/O type that present higher threshold voltage and also allow a higher supply voltage than the standard transistors in this CMOS process.

The temperature behavior of the main voltages is shown in Figure 4.7, where one can see that the PTAT voltages of the two unbalanced differential pair are equal. The reference voltage (Figure 4.8(a)) presents a low TC around 1.5 ppm/V with a maximum variation of 160 μV in this temperature range (-40 to 125 °C). Figure 4.8(b) presents the total current of the circuit as a function of the temperature. The current increases with the temperature to a maximum of 95 nA at 125 °C.

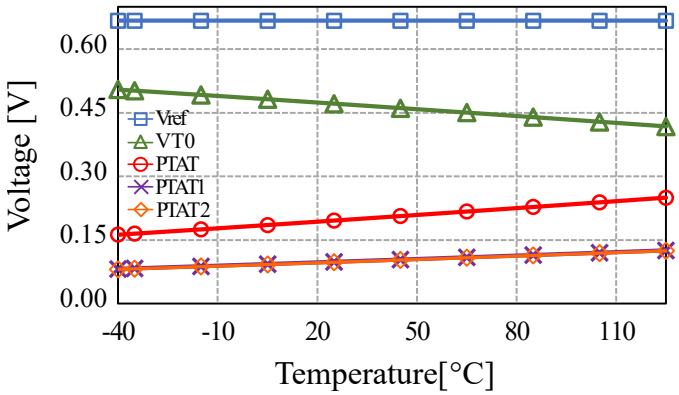
Figure 4.9 presents the PSR of the output, resulting -70 db from 0 Hz to almost 100 Hz for V_{DD}=1.2 V. Figure 4.10(a) shows that the output of proposed topology varies only 1.15 mV when variations between 1 to 3 V are done in V_{DD}, resulting a LS of 576 μV/V, meaning that the feedback path added in the V_{T0} monitor by the OA works fine. Another advantage of this implementation is that the TC presents a low supply sensitivity as shown in Figure 4.10(b).

Figure 4.6: Proposed voltage reference layout.



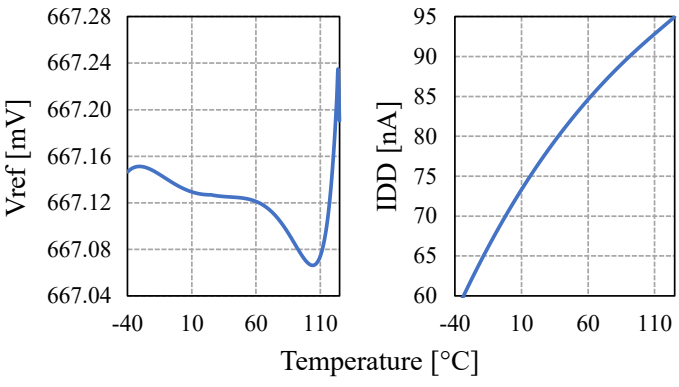
Source: The own author

Figure 4.7: CTAT, PTAT and V_{REF} voltages.



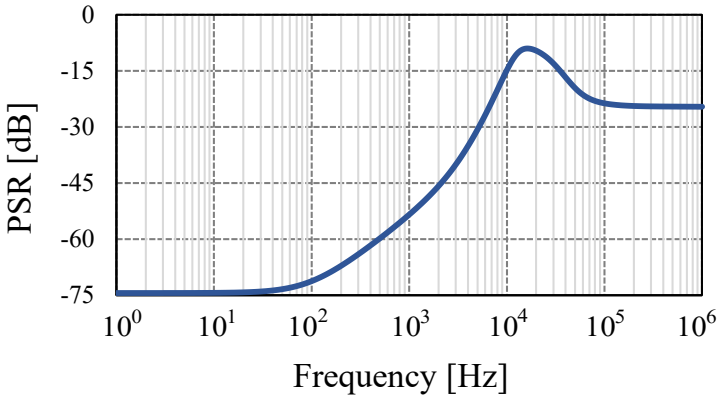
Source: The own author

Figure 4.8: V_{REF} voltage, and total current vs temperature.



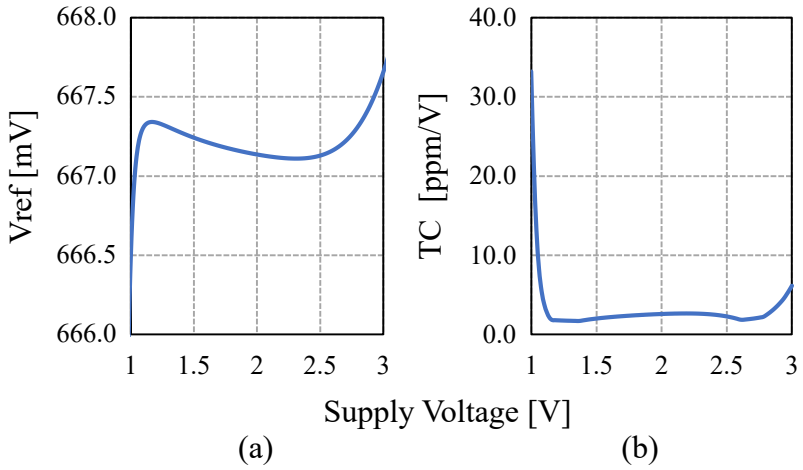
Source: The own author

Figure 4.9: PSR in the output over frequency.



Source: The own author

Figure 4.10: LS of V_{REF} and TC.



Source: The own author

Table 4.3 presents a comparison of recently published resistorless voltage references. One of the great advantages of our topology is the very low line sensitivity and high PSRR, providing a low TC, working in a wide range of temperature and voltage. The power consumption and area of this circuit is comparable with other state of the art circuits.

The fabrication variability of the circuit is analysed through Monte Carlo (MC) simulations (200 runs - Figure 4.11). The TC histograms (top) are almost equally affected by both average process variations and mismatch effects, but even including both effects 91% of the samples present a TC lower than 30 ppm/°C. The reference voltage histograms (bottom) present a maximum variation 3σ= 78 mV for combining effects, with σ/μ= 3.9 %, being the process variations 3σ= 69 mV the main contribution. High process dependence is usual in voltage references based on threshold voltage, but it can be overcome by trimming. This circuit was fabricated using the Mosis's educational program, and we soon expect to measure it.

Table 4.3: Comparison of recent CMOS Voltage References

	[1]$^{+}$	[2]$^{+}$	[3]$^{+}$	[4]*	[5]*	**This Work***	Unit
Process	0.18	0.18	0.18	0.18	0.18	**0.13**	µm
CTAT	V_{T0}	V_{T0}	V_{EB}	V_{T0}	V_{EB}	$\mathbf{V_{T0}}$	–
Temperature Range	0 to 125	-20 to 80	-40 to 120	-20 to 80	0 to 125	**-40 to 125**	°C
TC§	39	54.1	114	19.4	7	**1.5**	ppm/°C
Power	3.15	0.011	52.5	180	5.7	**77**	nW
V_{REF}	263.5	328	548	633	479	**667**	mV
V_{DD}	0.45 2	0.5 3.6	0.7 1.8	0.85 2.5	0.85 1.8	**1** **3**	V
LS	4400	440	-	24	2112	**576**	ppm/V
PSRR	-45	-49	-56	-76	-48	**-71.21**	dB
Area	0.043	0.0014	0.0246	–	0.0014	**0.0084**	mm^2

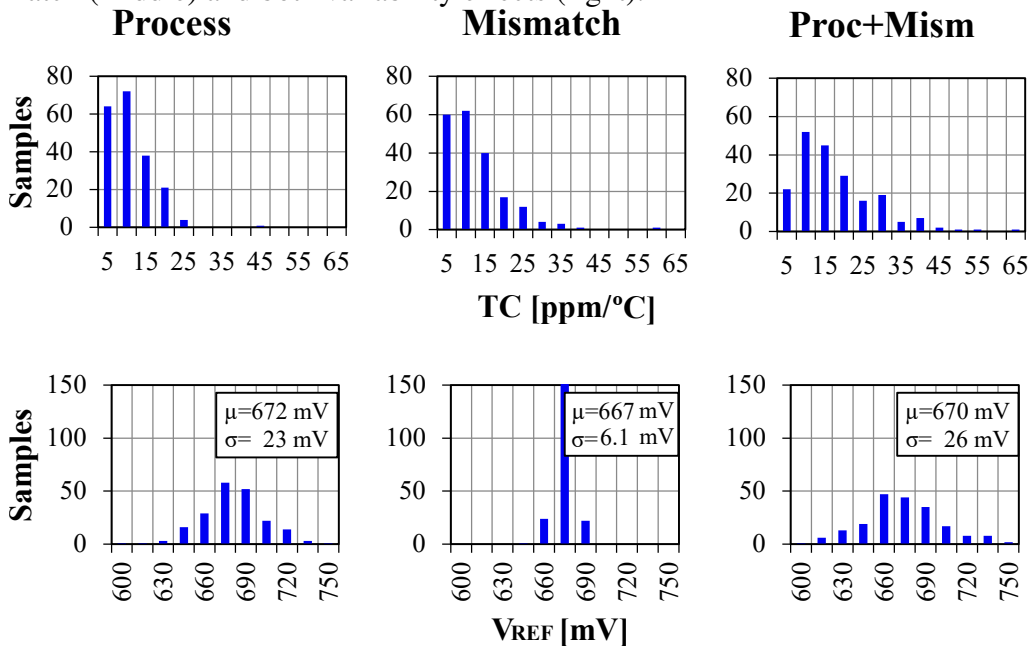
($^{+}$) Measurement (*) Simulation ($\S$)Reported Best Cases

[1](MAGNELLI et al., 2011) [2](SEOK et al., 2012)

[3](OSAKI et al., 2013) [4](ZENG et al., 2013)

[5](MATTIA; KLIMACH; BAMPI, 2014b)

Figure 4.11: Monte Carlo simulations of TC (top) and Vref (bottom) for Process (left), Mismatch (middle) and both variability effects (right).



Source: The own author

4.3 Process Variability Compensation Circuit

In integrated circuits, the output quality depends on the correspondence of internal parameters. For example, in an operational amplifier, the input offset is affected if the input transistors parameters are slightly different. The same phenomena can be seen in digital circuits, where the delays depend directly of the V_{T0} value of the transistors. Thus, any change in the expected value of V_{T0}, can increase or decrease the delay, which could produce setup or hold time issues.

Then, process variability is a big problem to consider in circuits that are implemented using CMOS processes. It is possible to divide the process variability in three different effects. The first one is the average process variability which affects the entire wafer, and normally happens due to small differences in the fabrication process, making that some parameters, as the oxide thickness, change its value from wafer to wafer. The second one is the global intra-wafer effect, that is due to gradients, and makes that device parameters vary across the entire wafer. Finally, there is the local effect which occurs randomly, and makes that the parameters vary from one device to another one (mismatch) into the same wafer (FRANCA; TSIVIDIS, 1994).

The mismatch variations can be mitigated by the designer, increasing the area of the devices and the inversion level (in the transistors case). The intra-wafer variations can be attenuated at technological level, but especially, can and must be compensated by the designer with suitable layout techniques, as common centroid structures, dummies, etc. On the other hand, the average process variation does not offers many options for the designer to mitigate its effects, and is normally controlled at the technological level by adding extra steps to the manufacturing process to gain a finer control over it, and/or to perform corrections over some devices or entire sections of the chip (ANDRADE; CALOMARDE; RUBIO, 2010). Traditionally, the option that has the designer to reduce the global process variations effects is to calibrate each circuit. However this technique slows the time to market increasing the cost of a chip, making it desirable to avoid it whenever possible.

One of the parameters that most influences the performance of analog and digital circuits, and that more varies with process, is the V_{T0} of the transistor. Then, one possibility to diminish the effect of the process variation on the behavior of a circuit is to compensate the variability in the V_{T0}. Through the years, some works has been oriented in this way. In Pineda De Gyvez; LEONE (2008), for example, was patented a feedback system that modifies the bias in the transistor's bulk, and consequently changes the effective V_{T0}, depending on the output value of a V_{T0} monitor. Other authors as PING-CHUA et al. (2010), and PAPPAS; SISKOS; DIMITRIADIS (2010) also use the idea to compensate the circuit to reduce process variation, temperature variation or even aging effects. All of them have in common the use of a V_{T0} monitor to sense the V_{T0} value.

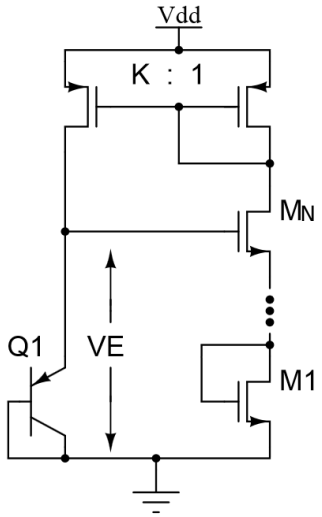
In this section the usability of the V_{T0} monitors presented herein is studied, in order to compensate the process variability in a circuit. For this, it is used the circuit presented in MATTIA; KLIMACH; BAMPI (2014c), which presents a direct dependency on the V_{T0} value, and therefore of its variability, as will be shown below.

4.3.1 Resistorless BJT bias

The basic concept of the topology to be compensated is shown in Figure 4.12. In this circuit, the BJT junction voltage is counterbalanced by the gate-source voltage of N stacked nMOS transistors. The resulting V_{GS} defines the BJT bias current, through a

feedback path that uses a current mirror with gain K. By defining N and K, a non-zero equilibrium DC point can be reached, which reflects the current-voltage behavior of both the BJT and the MOSFETs.

Figure 4.12: BJT bias concept.



Source: MATTIA; KLIMACH; BAMPI (2014c)

The emitter current I_E of the bipolar transistor is given by:

$$I_E = I_{SE} e^{\left(\frac{V_E}{m\phi_t}\right)} \tag{4.10}$$

where V_E is the emitter-base voltage, m represents the slope factor and I_{SE} is the reverse saturation current for the p-n junction. Assuming that all MOSFETs are operating in subthreshold regime and saturated (V_{DS} > 100 mV), the drain current I_D, according to the UICM MOSFET model CUNHA; SCHNEIDER; GALUP-MONTORO (1998), is given by:

$$I_D = 2I_{SQ}\frac{W}{L} e^{\left(\frac{V_G - V_{T0}}{n\phi_t} - \frac{V_S}{\phi_t}\right)} \tag{4.11}$$

where I_{SQ} is the specific current per square, which is a process dependent parameter defined by carrier mobility, oxide capacitance and temperature, V_{T0} is the threshold voltage for zero bulk-source voltage, and V_G and V_S are the gate and source voltages referred to the substrate, respectively. Assuming that all MOSFETs have the same width W and length L, and present the same slope factor n (since the bulk-source voltage is zero), leads to $V_{GS} = V_E/N$. Substituting (4.10) and (4.11) into the equality $KI_D = I_E$, and solving for the junction voltage V_E, leads to:

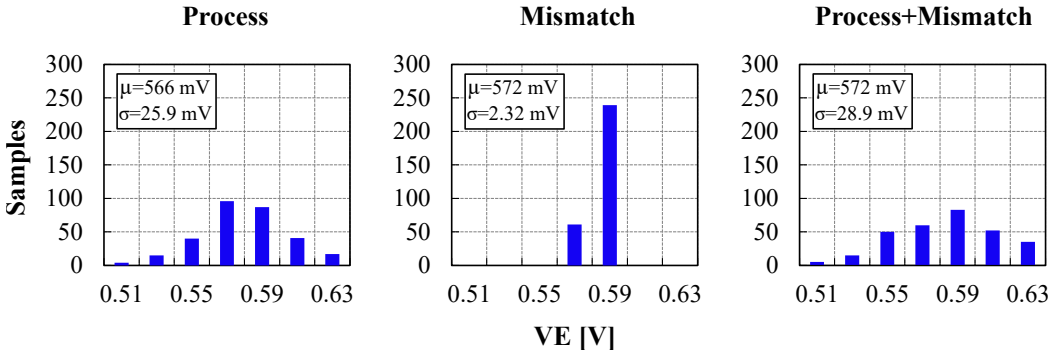
$$V_E = \frac{\phi_t}{\left(\frac{1}{m} - \frac{1}{nN}\right)} \left[ln\left(2K\frac{W}{L}\frac{I_{SQ}}{I_{SE}}\right) - \frac{V_{T0}}{n\phi_t} \right] \tag{4.12}$$

4.3.2 Resistorless BJT bias compensated

The circuit introduced in the last section presents ultra-low power consumption, is simple, and can be implemented in any CMOS standard process. Unfortunately, as (4.12) exposed, the circuit also depends directly on the V_{T0} value, and therefore presents a high variability in the output voltage V_E.

The impact of the fabrication variability effects on the output V_E value was analyzed using Monte Carlo (MC) simulations, where local mismatch effects and average process variations were simulated separately with 300 runs each. For average process MC all transistors have their parameters changed equally in each run - Figure 4.13 (left histograms). For local mismatch MC, the parameters of each transistor are varied individually in each run - Figure 4.13 (middle histograms). Both effects are also taken into account in a full variability analysis, shown in Figure 4.13 (right histograms).

Figure 4.13: Histograms of the Monte Carlo simulations.



Source: The own author.

As was shown in the circuit analysis, the circuit performance is strongly dependent on process parameters as V_{T0}, I_{SQ} and I_{SE}, and not to geometrical factors, being less sensitive to mismatch than to process variations. One can verify the different sensitivities when comparing the maximum spread that results from MC results: $\epsilon(\pm 3\sigma) = 572 \pm 6.96$ mV for mismatch, that is significantly less than $\epsilon(\pm 3\sigma) = 566 \pm 77.7$ mV for process. Finally a maximum total error in the worst condition, combining process and mismatch analysis, yields $\epsilon(\pm 3\sigma) = 572 \pm 86.7$ mV that is not much different from the process result.

Now, an important point to note is that the principal source of variation is the V_{T0}, because although the circuit also depends on I_{SQ} and I_{SE}, these parameters vary less, and furthermore are into a napierian logarithm, which attenuates the effect of the variations.

Then, making use of the circuits proposed herein, it is possible to compensate the circuit by compensating the V_{T0} value, making V_E independent of V_{T0}, and of its variability. As a proof of concept, we add mathematically to V_E, the V_{T0} value delivered by the unbalanced (Section 3.1) and balanced (Section 3.2) V_{T0} monitor circuit, and compare the resulting variability with the original value.

Monte Carlo simulations were done in order to analyze the impact that would have the fabrication variability in each configuration. Table 4.4 summarize the results found for the 300 MC simulations, taking into account average process variations. On the other hand, Table 4.4 compile the results for 300 runs of a full variability analysis, including the effect of process and mismatch variations. In both cases, were analyzed the original circuit (V_E), the original circuit adding the V_{T0} value delivered by the unbalanced V_{T0} monitor circuit (V_E+Unbalanced), and the original circuit adding the V_{T0} value delivered by the balanced V_{T0} monitor circuit (V_E+Balanced).

With the final purpose to do an equitable comparison, it is proposed to not compare neither the value of the average (μ) nor the deviation (σ), but rather, the ratio between these two values (σ/μ). The column "normalize" in Table 4.4, shows the ratio between the $(\sigma/\mu)_{initial}$ of the initial circuit and the (σ/μ) of each configuration, and allows to

Table 4.4: Results of Monte Carlo simulations considering average process.

	Average μ [mV]	Deviation σ [mV]	σ/μ	Normalized
V_E	569	25.97	0.046	1
V_E+Unbalanced	1035	7.81	0.008	6.04
V_E+Balanced	1052	8.24	0.008	5.827

Table 4.5: Results of Monte Carlo simulations considering average process and mismatch variations.

	Average μ [mV]	Deviation σ [mV]	σ/μ	Normalized
V_E	574	28.99	0.051	1
V_E+Unbalanced	1030	29.04	0.028	1.8
V_E+Balanced	1050	9.80	0.009	5.4

see, that the configurations using the monitors, improve the performance of the circuit in almost 6 times.

Finally, when process and mismatch variations are considered together, the configuration that used the unbalanced V_{T0} monitor gets worse comparing to its performance when only process variations were considered, but even so, its behavior is 1.8 times better than the original circuit. Furthermore, the configuration that used the balanced V_{T0} monitor is not very affected by the addition of the mismatch variations, and remains with an improvement higher than 5 times with respect to the original circuit.

4.3.3 Resistorless BJT bias, using I_{SQ} current

As it was demonstrated above, the V_E value gets more insensitive to process variations, if it is canceled its V_{T0} dependency. This can be done, adding directly the V_{T0} from a V_{T0} monitor circuit, as it was done before. However, it is important to remember, that to implement this idea, a circuitry must be inserted in order to do the addition between V_E, and the V_{T0} delivered by a monitor. After that, another configuration will be needed to divide the resulting value, because a value higher than 0.7 V for CTAT voltage is not useful nowadays.

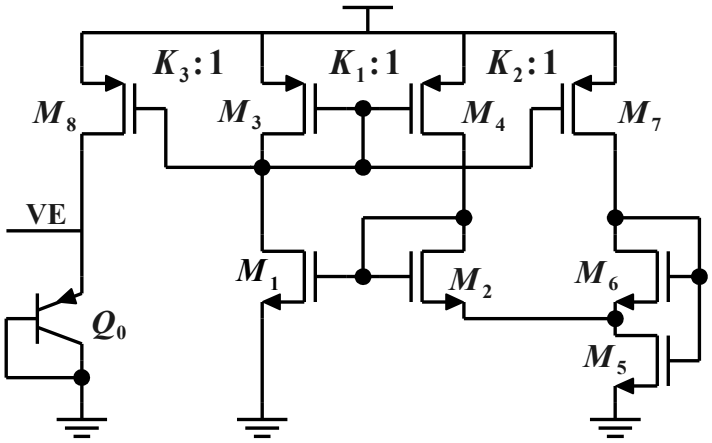
Hence, as the complexity of the circuit will be increased, the number of devices also will increase, and so the response of the circuit to mismatch and process variations will be worst in comparison with that shown in the last section. An alternative to this is biasing the BJT using a I_{SQ} current generator, as shown in Figure 4.14. In this way, it is possible to obtain a V_E independent of the V_{T0} value and of its process variability, as it will be demonstrated below. Suppose that the I_{SQ} current generator delivers a current I_{SQ}/A into the BJT, then using (4.10), and solving for V_E, it is obtained

$$V_E = m\phi_t \left[ln \left(\frac{K_3 I_{SQ}}{A I_{SE}} \right) \right] \tag{4.13}$$

Then, changing the current gain K_3, and the A factor of the I_{SQ} current generator,

different V_E values, and derivatives of $\partial V_E/\partial T$ can be generated.

Figure 4.14: BJT bias with the I_{SQ} current generator.



Source: The own author.

The circuit was implemented using the circuit of Section 3.1, at schematic level, and K_3 was set, in order to obtain the same V_E value that the original circuit delivers at 27 °C. Monte Carlo simulations were done in order to analyze the impact that would have the fabrication variability in the proposed circuit. Table 4.6 summarize the results found for the 300 MC simulations, taking into account average process variations. On the other hand, Table 4.7 compile the results for 300 runs of a full variability analysis, including the effect of process and mismatch variations.

Table 4.6: Results of Monte Carlo simulations considering average process.

	Average μ [mV]	Deviation σ [mV]	σ/μ	Normalized
V_E	569	25.97	0.046	1
I_{SQ} biasing	562	5.13	0.009	5

Table 4.7: Results of Monte Carlo simulations considering average process and mismatch variations.

	Average μ [mV]	Deviation σ [mV]	σ/μ	Normalized
V_E	569	25.97	0.046	1
I_{SQ} biasing	561	8.79	0.016	3.2

As was done before, the effectivity of the technique is evaluated, by looking to the σ/μ ratio. Table 4.6 shows that the proposed circuit improves the performance by 5 times in comparison with the original circuit when average process variations are considered. When process and mismatch variations are considered together (Table 4.7), the proposed circuit gets worse, comparing to its performance when only process variations are considered, but even so, its behavior is 3.2 times better than the original circuit.

It is important to highlight that the solution here presented, is a complete solution, which can be implemented in any CMOS standard process, and can be designed to consume only nano-Watts. In addition, the I_{SQ} current generator can be implemented using the circuit of Section 3.2, which presents better performance against mismatch variations, making possible to obtain even better results.

5 CONCLUSION

The threshold voltage is a fundamental parameter in almost all the MOSFET models, and also, is an important indicator in the monitoring of the success of a fabrication process. A threshold voltage monitor is a circuit that ideally delivers the estimated threshold voltage value, as a voltage at its output for a given temperature range, without external biases, parametric setups, curve fitting or any subsequent calculation. It can be used in temperature sensors, voltage and current references, radiation dosimeters and other applications.

In this thesis, after a complete revision of the progress made through the years, and the study of state of the art in threshold voltage monitors, three new topologies had been developed. Two of them monitor/sense the threshold voltage of NMOS transistors. These topologies overcome the principal nowadays monitors issue, which is its dependence with supply voltage, by adding a feedback path that increases the effective output impedance, which improves the PSR and LS. The third monitor that was presented senses the threshold voltage of PMOS transistors, for this, it uses an I_{SQ} current generator. The PMOS threshold voltage monitor is an important contribution, because is the first fully integrated PMOS topology that operates in the nanoWatt range, delivering the threshold voltage value without the needs of any additional calculation, since its output is delivered referred to the ground, and not to the supply voltage as done in previous implementations. This circuit has a similar performance to that presented by its NMOS counterparts, as it also used techniques to improve its behaviour against supply voltage variations.

In general, the threshold voltage monitors herein presented, and designed, are resistorless self-biased, and consume only tens of nW. The NMOS threshold voltage monitors works over the -40 to +125 °C temperature range, while the PMOS operates under the -40 to +100 °C temperature range. All of them works in 1 to 3 V supply voltage range with errors lower than 1%, and presents PSR and line sensitive lowers than -60 dB (at 100 Hz) and 450 ppm/V respectively. Monte Carlo simulations for an IBM 130 nm process support the design robustness to process and mismatch variations. A maximum error of 9% was found for the worst case over the entire operating temperature range, including fabrication variability effects, for all the circuits. The results shown for the NMOS threshold voltage circuits, are for Post-Layout simulations, meaning that the effect of parasitic elements that could appear due to the fabrication process were considered. We would like to emphasize that for the balanced threshold voltage monitor, and the PMOS threshold voltage monitor, the maximum error in the worst case was less than 4.25% and 3.6% respectively.

In addition, there have been presented some applications, in which are used our monitors. First, using the threshold voltage monitor as an I_{SQ} current generator, and taking advantage of its process-independent design methodology, it was proposed a structure that allows modeling the threshold voltage, in the entire range of lengths that a process offers. The simulation results shown lower errors than 3%, making extraction of transistors with lengths between 5 μm to 25 μm. This range can be extended by aggregating more branches to the circuit, but it is important to consider that between more far of the length use it by the I_{SQ} current generator, higher will be the error. This problem can be overcame, just by using two or more I_{SQ} current generators that extract the I_{SQ} from transistor with different lengths.

After that, a resistorless voltage reference consuming only tens of nano-Watts and with typical TC of 1.5 ppm/°C was presented. The circuit is composed by a threshold voltage monitor that implements the CTAT voltage, and two unbalance differential pairs implementing the PTAT voltage. The circuit performance is comparable with other voltage references in the literature. It works over the -40 to +125 °C temperature range and with a supply voltage range from 1 to 3 V, presenting -71 dB of PSR and 576 ppm/V of LS. Post-layout and Monte Carlo simulations for a 130 nm CMOS process show the design robustness to process and mismatch variations, resulting that σ/μ= 3.9 $\%$ for V_{REF} and that 91% of the samples presents a TC lower than 30 ppm/°C.

Finally, it was studied the variability of a self-biased circuit that generates a CTAT voltage that can be used to BGRs implementation, and was proved through simulations, that its variability can be improved by a factor of five, by adding the threshold voltage value delivered by our monitors. Also, it was proposed a new self-biased CTAT circuit. It uses an I_{SQ} current generator to bias a BJT. Three hundred runs of Monte Carlos simulations was done at schematic level, in order to test the circuit performance, against average process and mismatch variations. As a result, the circuit improves the performance by a factor of three, and can be even better, if it is used the balanced threshold voltage monitor to implement the I_{SQ} current generator, since this circuit is less sensitive to mismatch variations, than the unbalance threshold voltage monitor that was used.

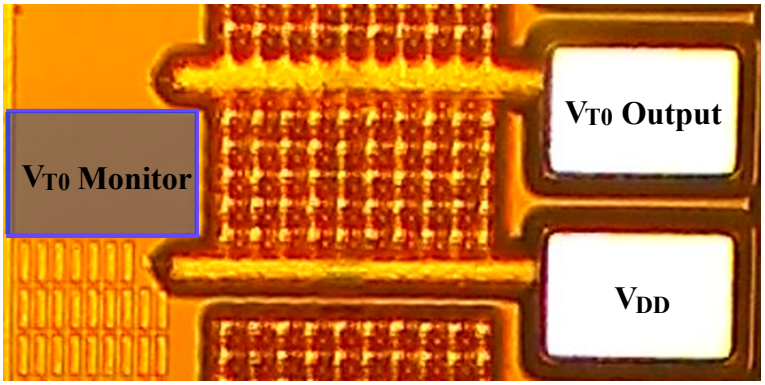
5.1 Future Works

The proposed monitors achieve excellent results, but it is important to remember that their performance have only been tested through simulations. Then, although it is expected that the DC circuits behavior does not change due to parasitic elements that can be appear in the fabrication process, its variability itself may change, as the models that the foundry deliver for process variability are not 100% reliable.

Therefore, as a next step, it is necessary to obtain the experimental data of the circuits that were sent to fabrication using the IBM 130 nm (Figure 5.1) and Silterra 130 nm process. After that, we aim to run at least one more fabrication to address batch-to-batch variations in the V_{T0} monitors. Additionally, even though we showed some applications in which the V_{T0} monitors were used, several other circuits can be implemented using them. In fact, just in the field of process compensation, there are a lot of possibilities that can be explored, then this topic can be the research focus of another dissertation.

Other topologies were designed during the M.Sc. but were left out of this thesis, some of them need to be measured, and could be improved as well.

Figure 5.1: Micro-photography of the balanced V_{T0} monitor circuit fabricated in IBM 130 nm process.



Source: The own author

REFERENCES

ALINI, R. et al. ACCURATE MOS THRESHOLD VOLTAGE DETECTOR FOR BIAS CIRCUITRY. In: IEEE INT. SYMP. CIRCUITS SYST., 1992. **Anais...** [S.l.: s.n.], 1992.

ANDRADE, D.; CALOMARDE, A.; RUBIO, A. A comprehensive compensation technique for process variations and environmental fluctuations in digital integrated circuits. **Midwest Symp. Circuits Syst.**, [S.l.], p.141–144, 2010.

BACCARANI, G.; RUDAN, M.; SPADINI, G. Analytical IGFET model including drift and diffusion currents. **Solid-State Electron Devices, IEE J.**, [S.l.], v.2, n.2, p.62–68, 1978.

BAKER, R. J. **CMOS Circuit Design, Layout, and Simulation, Second Edition**. [S.l.]: Wiley-IEEE Press, 2004.

BREWS, J. A charge-sheet model of the MOSFET. **Solid State Electron.**, [S.l.], v.21, n.2, p.345–355, feb 1978.

CAMACHO-GALEANO, E. M.; GALUP-MONTORO, C.; SCHNEIDER, M. C. A 2-nW 1.1-V self-biased current reference in CMOS technology. **Circuits and Systems II: Express Briefs, IEEE Transactions on**, [S.l.], v.52, n.2, p.61–65, Feb 2005.

CILINGIROGLU, U.; HOON, S. K. An optimally self-biased threshold-voltage extractor [MOSFET circuit parametric testing]. **Instrumentation and Measurement, IEEE Transactions on**, [S.l.], v.52, n.5, p.1528–1532, Oct 2003.

CUNHA, A.; SCHNEIDER, M. C.; GALUP-MONTORO, C. An MOS transistor model for analog circuit design. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.33, n.10, p.1510–1519, 1998.

DASGUPTA, R.; SAHA, D.; SAMANTA, J. Implementation of a New Offset Generator Block for the Low-Voltage , Low-Power Self Biased Threshold. **16th Int. Symp. VDAT 2012**, [S.l.], p.156–165, 2012.

FIKOS, G.; SISKOS, S. Low-voltage low-power accurate CMOS VT extractor. **Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on**, [S.l.], v.48, n.6, p.626–628, Jun 2001.

FRANCA, J.; TSIVIDIS, Y. **Design of analog-digital VLSI circuits for telecommunications and signal processing**. [S.l.: s.n.], 1994.

GILBERT, B. Current Mode, Voltage Mode, or Free Mode? A Few Sage Suggestions. **Analog Integr. Circuits Signal Process.**, Hingham, MA, USA, v.38, n.2-3, p.83–101, Feb. 2004.

GOMEZ C., J. A. et al. High PSRR Nano-Watt MOS-Only Threshold Voltage Monitor Circuit. In: SYMP. INTEGR. CIRCUITS SYST. DES. - SBCCI '15, 28., 2015. **Proceedings...** ACM, 2015.

IHANTOLA, H. K. J.; MOLL, J. L. Design theory of a surface field effect transistor. **Solid State Electron.**, [S.l.], v.7, n.6, p.423–430, 1964.

JOHNSON, M. An input-free VT extractor circuit using a two-transistor differential amplifier. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.28, n.6, p.704–705, Jun 1993.

MAGNELLI, L. et al. A 2.6 nW, 0.45 V Temperature-Compensated Subthreshold CMOS Voltage Reference. **Solid-State Circuits, IEEE J.**, [S.l.], v.46, p.465–474, 2011.

MATTIA, O. E. et al. 0.7 V Supply Self-Biased NanoWatt MOS-Only Threshold Voltage Monitor. **IEEE Int. Symp. Circuits Syst.**, [S.l.], n.2, p.2–5, 2015.

MATTIA, O. E.; KLIMACH, H.; BAMPI, S. Sub-1 V Supply Nano-Watt MOSFET-Only Threshold Voltage Extractor Circuit. In: SYMPOSIUM ON INTEGRATED CIRCUITS AND SYSTEMS DESIGN, 27., 2014, New York, NY, USA. **Proceedings...** ACM, 2014. p.23:1–23:6. (SBCCI '14).

MATTIA, O. E.; KLIMACH, H.; BAMPI, S. 0.9 V, 5 nW, 9 ppm/°C Resistorless Sub-Bandgap Voltage Reference in 0.18 μm CMOS. **Circuits Syst. (LASCAS), 2014 IEEE Fifth ...**, [S.l.], n.1, p.0–3, 2014.

MATTIA, O. E.; KLIMACH, H.; BAMPI, S. Resistorless BJT bias and curvature compensation circuit at 3.4 nW for CMOS bandgap voltage references. **Electron. Lett.**, [S.l.], v.2, n.50, p.3–4, 2014.

MATTIA, O. E.; KLIMACH, H.; BAMPI, S. 0.7 V Supply Self-Biased NanoWatt MOS-Only Threshold Voltage Monitor. **Proceedings of the 27th Symposium on ...**, [S.l.], n.2, p.2–5, 2015.

ORTIZ-CONDE, A. et al. Revisiting MOSFET threshold voltage extraction methods. **Microelectron. Reliab.**, [S.l.], v.53, n.1, p.90–104, 2013.

OSAKI, Y. et al. 1.2-V Supply, 100-nW, 1.09-V Bandgap and 0.7-V Supply, 52.5-nW, 0.55-V Subbandgap Reference Circuits for Nanowatt CMOS LSIs. **Solid-State Circuits, IEEE J.**, [S.l.], v.48, n.6, p.1530–1538, 2013.

PAO, H.; SAH, C. Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors. **Solid. State. Electron.**, [S.l.], v.9, n.10, p.927–937, oct 1966.

PAPPAS, I.; SISKOS, S.; DIMITRIADIS, C. A. A new threshold-voltage compensation technique of poly-Si TFTs for AMOLED display pixel circuits. **J. Soc. Inf. Disp.**, [S.l.], v.18, n.10, p.721–731, 2010.

PELGROM, M.; DUINMAIJER, A. C. J.; WELBERS, A. Matching properties of MOS transistors. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.24, n.5, p.1433–1439, Oct 1989.

Pineda De Gyvez, J. D. J.; LEONE, M. **CIRCUIT AND METHOD FOR CONTROLLING THE THRESHOLD VOLTAGE OF TRANSISTORS**. United States Patent US007332953B2, 2008.

PING-CHUA, W. et al. **METHOD AND APPARATUS FOR EXTENDING THE LIFETIME OFA SEMICONDUCTOR CHIP**. United States Patent US 7,821,330 B2, 2010.

ROSSI, C.; GALUP-MONTORO, C.; SCHNEIDER, M. C. PTAT Voltage Generator based on an MOS Voltage Divider. In: NSTI NANOTECH, 2007. **Anais...** [S.l.: s.n.], 2007. v.3, p.625–628.

Semiconductor Industry Association. **The U.S. Semiconductor Industry**: a key contributor to U.S. economic growth. Accessed: 2016-01-30, Available: <http://www.semiconductors.org/main/resources/>.

SENGUPTA, S. AN INPUT-FREE NMOS V_T EXTRACTOR CIRCUIT IN PRESENCE OF BODY EFFECTS. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, 2004. **Anais...** [S.l.: s.n.], 2004.

SEOK, M. et al. A Portable 2-Transistor Picowatt Temperature-Compensated Voltage Reference Operating at 0.5 V. **Solid-State Circuits, IEEE J.**, [S.l.], v.47, n.10, p.2534–2545, 2012.

SERRA-GRAELLS, F.; HUERTAS, J. L. Sub-1-V CMOS proportional-to-absolute temperature references. **IEEE J. Solid-State Circuits**, [S.l.], v.38, n.1, p.84–88, 2003.

SIEBEL, O. F.; SCHNEIDER, M. C.; GALUP-MONTORO, C. MOSFET threshold voltage: definition, extraction, and some applications. **Microelectronics Journal**, [S.l.], v.43, n.5, p.329 – 336, 2012.

THOMAS, F.; HOLMAN, W. MOSFET threshold voltage extractor circuits based on square-law behavior. In: MIDWEST SYMP. CIRCUITS SYST., 1999. **Anais...** [S.l.: s.n.], 1999. v.2.

TSAY, J.-H.; LIU, S.-I.; WU, Y.-P. Measuring the Threshold Voltage of a MOSFET with an Active Attenuator. In: IEEE INT. SYMP. CIRCUITS SYST., 1996. **Anais...** [S.l.: s.n.], 1996. p.397–400.

TSIVIDIS, Y. **Operation and Modeling of the MOS Transistor**. [S.l.]: Oxford University Press, USA, 1987.

TSIVIDIS, Y. P. Accurate Analysis of Temperature Effects in IC-VBE Characteristics with Application to Bandgap Reference Sources. **IEEE J. Solid-State Circuits**, [S.l.], v.SC-15, n.6, p.1076–1084, 1980.

TSIVIDIS, Y.; ULMER, R. Threshold voltage generation and supply-independent biasing in c.m.o.s. integrated circuits. **IEEE J. Electron. Circuits Syst.**, [S.l.], v.3, n.1, p.1–4, 1979.

TSIVIDIS, Y.; ULMER, R. W. A CMOS voltage reference. **Solid-State Circuits, IEEE J.**, [S.l.], v.13, n.6, p.774–778, 1978.

VITTOZ, E. A.; NEYROUD, O. A low-voltage CMOS bandgap reference. **Solid-State Circuits, IEEE J.**, [S.l.], v.14, n.3, p.573–579, 1979.

VITTOZ, E.; FELLRATH, J. CMOS analog integrated circuits based on weak inversion operations. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.12, n.3, p.224–231, 1977.

VLASSIS, S.; PSYCHALINOS, C. Low-voltage CMOS VT extractor. **Electronics Letters**, [S.l.], v.43, n.17, p.921–923, August 2007.

WANG, Y.; TARR, G.; WANG, Y. Input-Free cascode V_{thn} and V_{thp} Extractor Circuits. **IEEE International Conference on Electronics, Circuits, and Systems**, [S.l.], n.3, p.282–285, 2004.

WANG, Z. Automatic VT extractors based on an $n \times n^2$ MOS transistor array and their application. **IEEE Journal of Solid-State Circuits**, [S.l.], v.27, n.9, p.1277–1285, 1992.

World Semiconductor Trade Statistics. Accessed: 2016-01-30, Available: <https://www.wsts.org/>.

YU, C.-G.; GEIGER, R. L. An accurate and matching-free threshold voltage extraction scheme for MOS transistors. In: IEEE INT. SYMP. CIRCUITS SYST., 1994., 1994. **Anais...** [S.l.: s.n.], 1994. v.44, p.115–118.

ZENG, Y. et al. An ultra-low-power CMOS voltage reference generator based on body bias technique. **Microelectronics J.**, [S.l.], v.44, n.12, p.1145–1153, 2013.

APPENDIX A LIST OF PUBLICATIONS

1. GOMEZ C., J. A., KLIMACH, H., FABRIS, ERIC. and MATTIA, O. E. *"High PSRR Nano-Watt MOS-Only Threshold Voltage Monitor Circuit"*. Proceedings of the 28th Symposium on Integrated Circuits and Systems Design **(SBCCI'15)**. Salvador, Brazil. September 2015.

2. GOMEZ C., J. A., KLIMACH, H., FABRIS, ERIC. and BAMPI, SERGIO. *"1.5 ppm/°C Nano-Watt Resistorless MOS-Only Voltage Reference"*. Proceedings of the VII IEEE Latin American Symposium on Circuits and Systems **(LASCAS'16)**. Florianopolis, Brazil. February-March 2016.

3. CAMPOS DE OLIVEIRA, ARTHUR, GOMEZ C., J. A., KLIMACH, H. and BAMPI, SERGIO. *"0.3 V Supply 17 ppm/°C 3-Transistor Picowatt Voltage Reference"*. Proceedings of the VII IEEE Latin American Symposium on Circuits and Systems **(LASCAS'16)**. Florianopolis, Brazil. February-March 2016.

4. GOMEZ C., J. A., MATTIA, O. E., KLIMACH, H. and BAMPI, S. *"0.75 V Supply NanoWatt Resistorless Sub-Bandgap Curvature-Compensated CMOS Voltage Reference"*. Analog Integr. Circuits Signal Process (AICSP), Springer, Apr. 2016

APPENDIX B ACM MOSFET MODEL

In the ACM model, the drain current I_D of a long-channel MOSFET is expressed as

$$I_D = I_F - I_R = SI_{SQ}(i_f - i_r) \tag{B.1}$$

where I_F and I_R are the forward and reverse currents, $S = W/L$ is the aspect ratio, W being the width and L the length of the transistor. i_f and i_r are the forward and reverse inversion coefficients, related to the source and drain inversion charge densities, while I_{SQ} is the sheet normalization transistor current

$$I_{SQ} = \frac{1}{2}n\mu C'_{ox}\phi_t^2 \tag{B.2}$$

where n is the subthreshold slope factor, μ is the channel effective mobility (both slightly dependent on the gate voltage V_G), C'_{ox} is the gate capacitance per unit area, and ϕ_t is the thermal voltage. The relationship between inversion levels i_f and i_r and terminal voltages is given by

$$\frac{V_P - V_{S(D)}}{\phi_t} = F(i_{f(r)}) = \sqrt{1 + i_{f(r)}} - 2 + \ln(\sqrt{1 + i_{f(r)}} - 1) \tag{B.3}$$

where V_S and V_D are the source and drain voltages (all terminal voltages are referenced to the transistor bulk), and V_P is the pinch-off voltage, approximated by

$$V_P \simeq \frac{V_G - V_{T0}}{n} \tag{B.4}$$

where V_G is the gate voltage, and V_{T0} is the threshold voltage for zero bulk bias. Over time, many operational and device physics dependent (inversion charge, for instance) definitions for V_{T0} were used. In the ACM MOSFET model, the threshold voltage has a universal physical meaning, defined as the condition where the drift and diffusion components of the drain current have equal magnitude.

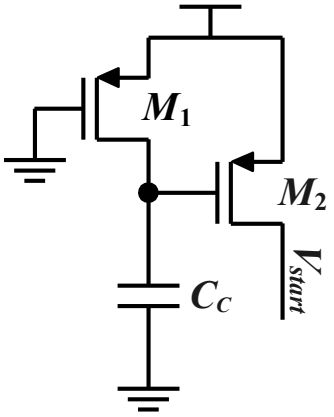
The first term (the square root one) in the right side of (B.3) is related to the drift component of the drain current, being predominant under strong inversion. The last term (the logarithmic one) is related to the diffusion component, being predominant under weak inversion operation. In the forward saturation condition, $I_F \gg I_R$, and consequently, $I_D \simeq I_F = SI_{SQ}i_f$. In this thesis the V_{T0} value is then rigorously defined based on (B.3).

APPENDIX C START-UP ANALYSIS

All the circuits that were herein presented are self-biased structures that present two equilibrium points, which results in two different DC steady state conditions. The first equilibrium point occurs when all the currents are zero, and the second one is the desirable equilibrium point. Therefore, the possibility that the circuit operates in an undesirable condition, forces the inclusion of a start-up stage, that drives the structure into the desirable operation point.

For this job was used the circuit shown in Figure C.1, which as was already mentioned, was chosen due to its simplicity and zero DC consumption. Although the normal operation of this structure was explained several times, it presents some disadvantages which will be discussed below.

Figure C.1: Start-up schematic.



Source: The own author

First, in the steady state of the circuit, it is considered that the V_{DS} of transistor M_1 and the V_{GS} of transistor M_2 are zero, because the voltage in the capacitor C_C is equal to the supply voltage V_{DD}, making possible to guarantee that the current consumption of the start-up in the steady state is zero. Unfortunately, although this consideration is correct, DC variations may occur in the supply voltage after the circuit reaches its steady state condition, which will force that the start-up stage operates once again, delivering an undesirable current into the circuit, and affecting its normal behaviour.

Other issue is the fact that the start-up structure delivers a current impulse, independently of the initialization state of the circuit. This can generate that even after the start-up stage injects the current, the circuit remains in its zero state, or can also generate that the start-up stage continue injecting current after the circuit reaches its operation point, which are both undesirable cases. This problem can be overcome by using a start-up topology

that uses a feedback. The start-up stage will deliver a current and simultaneously will sense a node in the circuit, and then, when the node reaches its desirable value, the start-up stage will turn off. The problem with this kind of start-up is that usually represent complex structures, increasing the area and the design effort. Additionally, this structures can continue consuming a small current in steady state, which in our case (nano-watts consumption) can be a large percentage of the total current, which means an increasing in the power consumption.