UNIVERSITÀ DELLA CALABRIA

UNIVERSITA' DELLA CALABRIA

Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica

Scuola di Dottorato

Archimede in Scienze, Comunicazione e Tecnologie

Indirizzo

Scienze e Tecnologie dei Sistemi Complessi

CICLO

XXVIII

TITOLO TESI

Circuit and architecture solutions for the low-voltage, low-power domain

Coordinatore: Prof. Pietro Pantano

Firma ______________________

Supervisore/Tutor: Prof. Felice Crupi

Firma ______________________

Dottorando/a: Dott. Domenico Albano

Firma ______________________

ABOUT THE AUTHOR

Domenico Albano was born in Soriano Calabro (VV), Italy, on October 15th, 1986.

He received the “diploma di geometra” with full grade (100/100) from Istituto Tecnico per Geometri, L. Einaudi, Serra San Bruno (VV), Italy.

He received the Bachelor of Science degree in Electronic Engineering cum laude from University of Calabria, Rende (CS), Italy, in 2009. The title of the thesis was: Optimization of VLSI systems. The advisor was Prof. Manlio Gaudioso.

In 2012 he received the Master of Science degree in Electronic Engineering cum laude from University of Calabria, Rende (CS), Italy. The title of the thesis was: Design of an ultra-low power temperature compensated voltage reference. The advisors were Prof. Felice Crupi and Prof. Giuseppe Iannaccone.

In 2012 he started to work toward the Ph.D. degree in Electronic Engineering at University of Calabria. During the Ph.D. he worked with different research groups and universities as University of Pisa, University of Pavia, University of Singapore and Holst Centre, Eindhoven, The Netherlands.

In April 2012 he was visiting student at the University of Pisa where he designed an ultra-low voltage, low-power voltage reference.

From October 2013 to November 2014 he was a visiting Ph.D. student at the Department of Electrical, Computer, and Biomedical Engineering of the University of Pavia where he worked on the design of a low-power successive approximation register analog to digital converter.

From April 2015 to October 2015 he was a visiting Ph.D. student at the Holst Centre (IMEC), Eindhoven, The Netherlands, where he worked on the design of a new silicon Physical Unclonable Function for security applications.

His main research interests include the design of ultra-low voltage, low-power voltage references, current references, ADCs and Physical Unclonable Functions.

LIST OF PUBLICATIONS

[1] F. Crupi, **D. Albano**, M. Alioto, J. Franco, L. Selmi, J. Mitard, G. Groeseneken, “Impact of High-Mobility Materials on the Performance of Near- and Sub-Threshold CMOS Logic Circuits,” *IEEE Transactions on Electron Devices*, vol. 60, n. 3, pp. 972-977, 2013.

[2] **D. Albano**, F. Crupi, F.Cucchi and G. Iannaccone, “A Picopower Temperature-Compensated Subthreshold CMOS voltage reference”, *International Journal of Circuit Theory and Applications*, 2013. DOI: 10.1002/cta.1925.

[3] M. Lanuzza, R.Taco and **D. Albano**, “Dynamic Gate-level Body Biasing for Subthreshold Digital Design,” in *IEEE 5th Latin American Symposium on Circuits and Systems* (LASCAS), Feb. 2014, pp. 25-28.

[4] **D. Albano**, F. Crupi, F.Cucchi and G. Iannaccone, “A sub kT/q voltage reference operating at 150 mV”, *IEEE Transactions on Very Large Scale Integrated Systems* (VLSI), pp. 1547 – 1551, Aug. 2015.

[5] **D. Albano**, M.Lanuzza, R.Taco, F. Crupi, “Gate-Level Body Biasing for Subthreshold Logic Circuits: Analytical Modeling and Design Guidelines”, *International Journal of Circuit Theory and Applications*, vol. 43, no 11, pp. 1523-1540, Nov. 2015.

[6] **D. Albano,** M. Grassi and P. Malcovati, “A Low Power 12-Bit ENOB SAR ADC for Silicon Drift X and Gamma Ray Detector Read-Out”, in *IEEE International Symposium on Circuits and Systems* (ISCAS), May 2015, pp. 297 – 300.

[7] R. Taco, M. Lanuzza, **D. Albano**, “Ultra-Low-Voltage Self-Body Biasing Scheme and Its Application to Basic Arithmetic Circuits,” *VLSI Design, Hindawi Publishing Corporation*, 2015.

INTRODUCTION TO THIS WORK

As a consequence of the growing request for long lifetime portable, implantable and wearable electronic devices, low-power, low-voltage operation represents a mandatory requirement of the modern electronics. Applications like energy-harvesting systems, Body Area Network (BAN) systems, Radio Frequency Identification (RFID) systems and implantable medical devices, are requiring substantial improvements in term of energy-efficient operation to the circuit designers' community. In these systems complex mixed-signal electronics is used to pick up, process and transmit the signals coming from the environment or the human body. Among the different circuital blocks, bias circuits and signal processing modules like Analog to Digital Converters (ADCs), represent compulsory components of all these applications. Additionally, applications like wireless BAN and RFID systems, such as Near-Field-Communication (NFC) systems, have to deal with privacy-related issues. As an example in the case of the BAN systems the private information and the actions required to the actuators must be protected from possible attacks since any involuntary action can cause a risk for the patient. At the same time in the case of the NFC systems, commonly used for money transfer, the private information must be protected from potential frauds. For such a reason, in the majority of the energy-constrained applications, security is becoming a compulsory requirement as well as the energy efficiency operation.

In this work different circuits capable to satisfy to the requirements of the energy-constrained applications are presented.

In chapter 1 a brief introduction to some of the energy-constrained applications is reported. The main requirements are highlighted as well as the implications at design level.

In chapter 2 the behaviour of the MOSFET at device level is analysed. The main design considerations and issues of the MOSFET operating in the low-power, low-voltage regime are highlighted.

In chapter 3 two ultra-low voltage, low-power subthreshold voltage references are presented. First a brief overview of the different solutions proposed so far is reported. The two proposed solutions are then introduced. The first solution proposes a voltage reference capable of operating with a supply voltage of 0.45 V and with a power consumption of 40 pW while the second solution consists in the first $k\mathrm{T}/q$ voltage reference capable of operating with a supply voltage of only 150 mV while consuming 26.1 pW. For both solutions the main design considerations are explained in detail. Measurements results are analysed and compared with the state-of-the-art solutions.

In chapter 4 the first subthreshold CMOS current reference operating at 0.5 V with a power consumption of only 40 nW is presented. Also here a brief overview on the main solutions proposed so far is reported. The operating principle of the proposed solution is analysed in detail. Measurements results are reported and compared with the other state-of-the-art solutions.

In chapter 5 a low-power Successive Approximation Register (SAR) Analog to Digital Converter (ADC) with an ENOB of 9 bits, a power consumption of only 0.27 mW and energy per conversion equal to 87 fJ/step, is reported. The architecture is explained in detail as well as the main design considerations. Measurement results are reported and compared with the other low-power SAR ADCs.

Finally in chapter 6 two innovative silicon Physical Unclonable Functions for data protection capable of operating at very low supply voltages are presented. In the first solution complementary current mirrors and an additional sense operational amplifier are employed to generate a random and robust secret key, while in the second solution a voltage divider consisting of two nMOS transistors is exploited to generate a secret key. For both solutions the main figures-of-merit in terms of uniqueness, randomness and reliability against malicious attacks are extracted. Simulation and measurement results are compared with the other state-of-the-art solutions.

INDEX

1. Low-Power, Low-Voltage Electronics

1.1. Introduction

Energy efficiency represents the main technology driver of the modern electronics. Indeed, the progresses in every single branch of the integrated electronics, from RF systems to data converters, wireless sensor networks, Internet of Things (IoT), digital and analog systems, are measured in terms of power consumption or energy efficiency [1]. These requirements are becoming so critical that both manufacturers and designers have been introduced different methodologies to address this problem at different levels, from variations in the semiconductor manufacturing process to the implementation of design techniques for low-power consumption. As an example in Figure 1.1 the different energy efficiency techniques exploited in microcontroller design during the last 15 years are reported [1]. Since the simple technology scaling is not in able to ensure good performances anymore, other techniques like ***aggressive voltage scaling***, ***Ultra-Low Power*** (ULP), ***Ultra Low-Leakage*** (ULL) technologies and ***compensation for variability on CMOS process*** have been exploited during the last years to achieve better results.

Nowadays the concepts of energy efficiency and minimum power consumption are becoming even more critical as a consequence of the explosion of applications like wearable, portable and handheld devices [2]-[5]. Indeed, in this kind of applications saving power is extremely important since the sizes of the devices are usually very small and only a limited space is available for the source of energy (i.e. battery). Because of this limited amount of energy, the complex electronics used to process and transmit the signals has to ensure the lowest possible amount of energy consumption both in idle and active mode. Additionally, in applications such as the implantable medical devices, saving energy is a mandatory requirement since battery replacement brings usually to an invasive procedure for the patient. With the main aim of understanding the different requirements of the low-power and low-energy electronics, in the following some of the main ultra-low power contexts are briefly introduced.

1.2. Body Area Networks

The extreme low-power, low-voltage operation is particular attractive in the Body Area Networks (BAN) scenario. BAN is an emerging technology that has the potential to revolutionize next-generation of healthcare, entertainment, and other personal applications [1]. This technology exploits wireless communications protocols that allow low-powered sensors to communicate with another one and transmit data to a local base station and to remote places like hospitals [3]. A particular kind of BAN is represented by the Body-Channel Communication (BCC) systems in which the human body is used as transmission medium. Both WBANs and WBCCs allow monitoring the person's vital signs, as shown in Figure 1.2.

Thus both WBANs and WBCCs use sensors on and/or in the human body, which are responsible for sensing the main biological signals such as temperature, heart rate, movements, electrocardiogram, blood pressure, oxygen saturation, etc. These data sets are collected ensuring reliability, security, and accuracy since any possible mistake can compromise the quality of life of the patient.

WBAN and WBCC systems can achieve high-speed communication with low energy consumption compared to other personal-area network (PAN) solutions such as ZigBee, Bluetooth, and UWB.

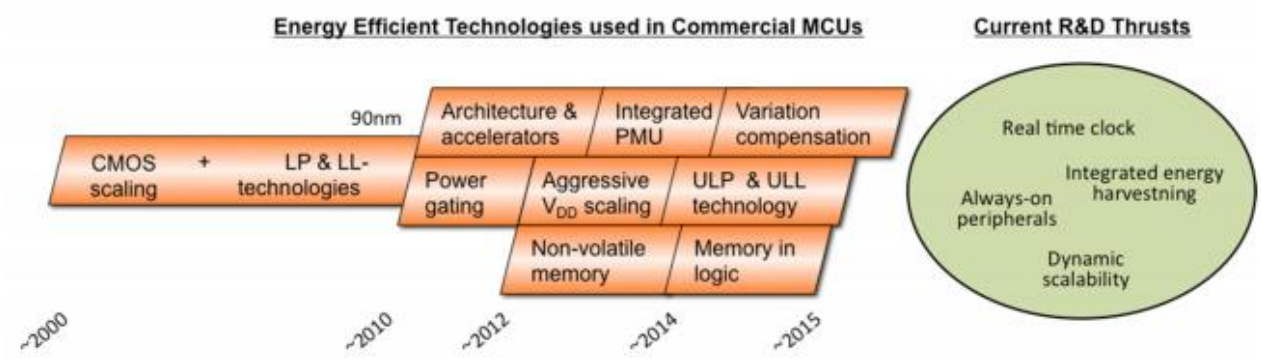


Figure 1.1. Energy efficiency techniques in microcontroller design.

From Figure 1.3 it is possible to note that BAN systems fall in a unique region of the power vs data rate trade-off that makes the concept of power consumption particular challenging. As explained before the key point of such systems is to continuously monitor a person's health status from a central station or hospital without impeding the person's mobility [3]. Depending on the application, some sensors will require a battery lifetime of months, years, or only a few hours [3].

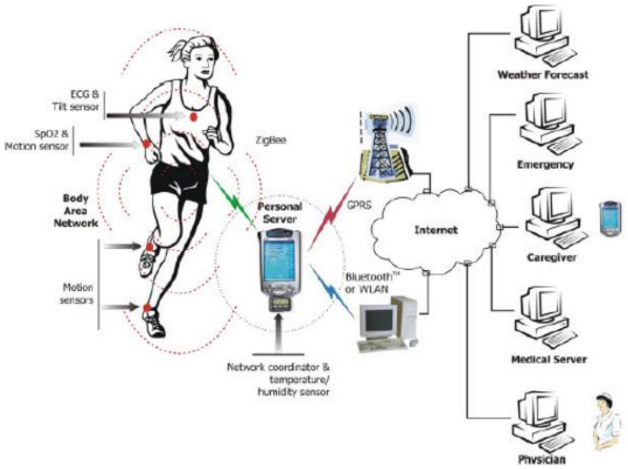


Figure 1.2. BAN system.

To be less invasive as possible, a BAN system is very small in size, which means that only a limited amount of energy for the proper operation of the different blocks is available. For such a reason a very power-efficient protocol (IEEE802.15.6) has been introduced for this particular application. However the power efficient protocol alone is not in able to achieve the previous targets in terms of lifetime. In fact the nodes (sensors) in a WBAN system are extremely energy constrained since they have to process the data in the fastest and most efficient way while consuming a very small amount of energy both in active and idle mode.

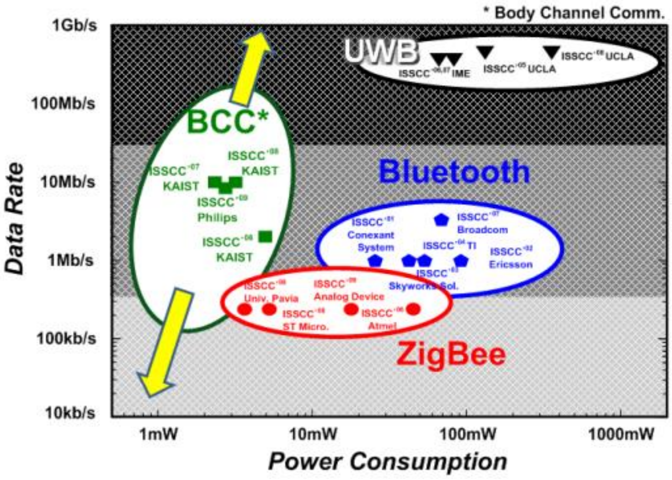


Figure 1.3. Power-data rate in WBANs.

From the design point of view this means that every single circuit involved in the signal processing of the vital data should be designed in order to consume the lowest amount of energy without compromising the performance in terms of resolution, speed, stability and reliability. The sensors which are commonly used in the previous low-energy contexts consist of analog, digital and mixed-signal circuits. For better understanding the electronics behind this application, in Figure 1.4 the architecture for a wireless biopotential acquisition is reported. The system consists of an analog front-end, a digital signal processor, an ADC, an RF transmitter and a harvesting and management unit control. In addition to the previous requirements, as reported before, in the context of the WBAN systems the identification is a critical step as well as the performance from the energy point of view. The private information and eventually the actions required to the sensors / actuators in WBANs systems should be protected from possible attacks since any involuntary action can cause a risk for the patient [4]. Despite the great efforts both in designing energy efficient electronics and reliable authentication protocols a lot of efforts are still necessary in WBANs systems design [4].

1.3. Radio Frequency Identification (RFID) Systems

The term Radio Frequency IDentification (RFID) system is used to indicate a short range radio-frequency technology in which a reader and an electronic tag can exchange data with the main goal of identification and tracking [5]. The very first implementation of an RFID system was reported in [6] in 1948 in the article entitled "*Communication by means of reflected power*".

In 1950 Watson-Watt, the responsible of a secret project in UK, developed the first active Identify Friend or Foe (IFF) system. A transmitter was putted on each British plane. When the plain received signals from radar stations on the ground, it began broadcasting a signal back that identified the aircraft as friendly. All the modern RFIDs systems work on this same basic concept. A signal is sent to a transponder, which wakes up and either reflects back a signal (passive system) or broadcasts a signal (active system) [7].

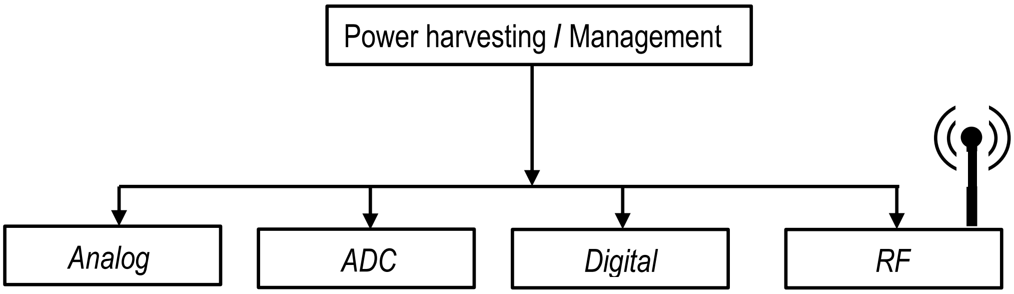


Figure 1.4. Architecture for wireless biopotential acquisition.

In the 1960 two companies *Sensormatic* and *Checkpoint* were founded for developing Electronic Article Surveillance (*EAS*). They introduced the first commercial prototype. In this case, the transponder was characterized by a single-bit output, representing its presence or absence. In 1990 IBM introduced the first high frequency (UHF) RFID system. The range covered by the system was extended to about 6 meters in the best condition, moreover a faster transfer data rate was guarantee compared to the 125 KHz and 13.56 MHz applications. This project was developed by IBM in collaboration with Wal-Mart, an American multinational retail corporation that operates a chain of discount department stores and warehouse stores. Despite that the UHF RFID developed by IBM was never commercialized [8]. Only in 1999 the RFID technologies start to become popular. Uniform Code Council, EAN International, Procter & Gamble and Gillette put up funding to establish the Auto-ID Center at the Massachusetts Institute of Technology to do some research about the feasibility of putting low-cost RFID tags on all products made to track them through the supply chain. Their idea was to put only a simple microchip that stored very little information instead to put a more expensive complex chip with more memory [7]-[8].

The RFIDs can be classified in different categories. The main classification takes into account the transponder power supply. Based on this the RFID can be: *active*, *semi-passive* and *passive*.

Active RFIDs incorporate a battery to transmit a signal to a reader antenna. They allow very low-level signals to be received by the tag. Usually they are designed in order to stay in idle mode for saving power for most of the time emitting a signal at a predefined interval or transmitting only when addressed by a reader. As a result of the built-in battery, an active RFID tag can be used at large distances from the reader. However they show limited life, due to the limited energy in the battery, and higher costs.

Semi-passive tags have an on-board power source, such as a battery, which is used to provide the power to the microchip’s circuitry. Despite the battery a semi-passive tag communicates by drawing power from the reader. Semi-passive tags show greater range than totally passive tags and have the ability to monitor sensor inputs even when they are not in the presence of an RF field.

Passive tags do not have an integrated power source thus they are generally powered by the reader antenna through an antenna located on the tag. The power transmission is performed through induction thanks to a specially designed antenna which generates a small voltage potential. Passive tags have in general a short-distance operating range. The power supply is typically generated using an energy harvester that converts the RF voltage coming from the antenna in a constant voltage used to supply the circuit.

RFID systems can operate at a number of designated frequencies, depending on the application requirements. Usually an RFID can be classified in the following categories: Low Frequency (125 kHz), High Frequency (13.56MHz), Ultra High Frequency (860-960 MHz) and Microwave (2.45 GHz).

A block diagram of a commercial RFID tag is reported in Figure 1.5 [9]. The main blocks are the envelope or peak detector, demodulator (consisting of an amplifier, a Schmitt trigger, an offset cancellation circuit, etc), the bias bandgap reference circuit (BGR) and the clock generator. All of these circuits are designed for low power operation in order to achieve an ultra-low power RFID tag chip.

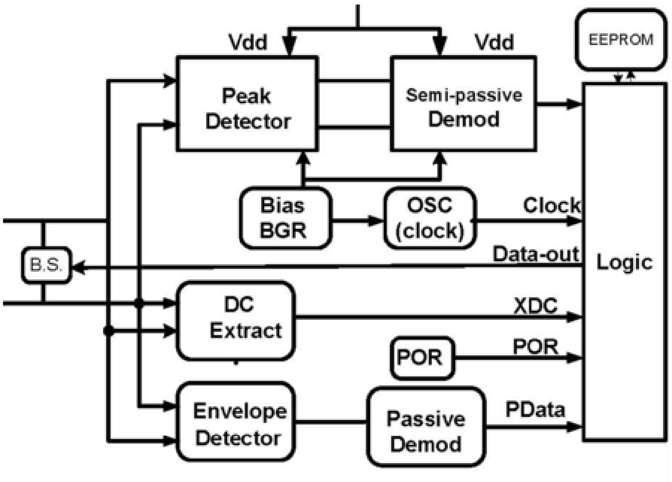
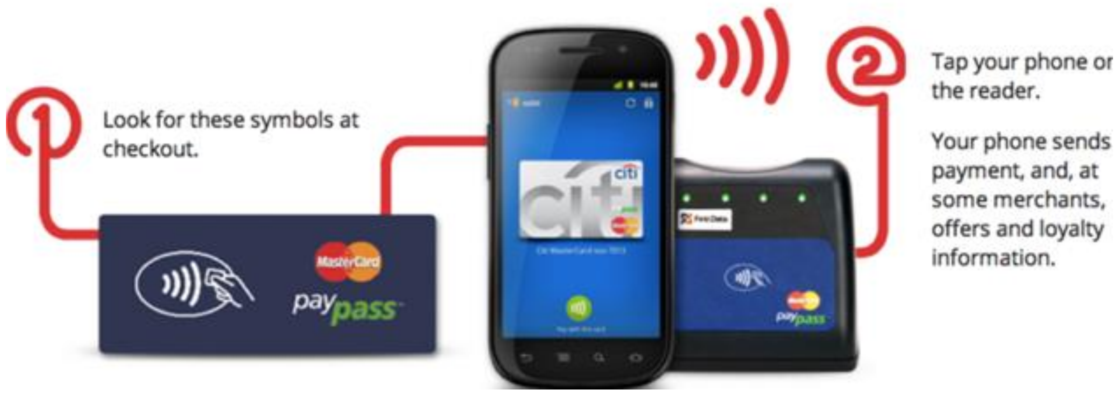


Figure 1.5. Semi-passive RFID Block Diagram [9].



Look for these symbols at checkout.

Tap your phone or the reader.

Your phone sends payment, and, at some merchants, offers and loyalty information.

Figure 1.6. *NFC mobile payment system*.

Another important issue regarding RFIDs is the level of security. RFIDs are increasingly being used in manufacturing, pharmaceutical and military sector for tagging, tracking and locating. The widespread item-level RFID tagging of products such as clothing and electronics raises public concerns regarding personal privacy.

Thus RFID chips carry on personal information which in some cases should be protected from attacks. As a consequence several works are trying to solve this problem by implementing reliable systems for the protection of the intellectual property at chip level. A promising solution consists in Physical Unclonable Function exploiting the variability in CMOS process [10]-[13].

A particular kind of RFID system in which security is the main issue is represented by the Near-Field-Communication (NFC). It operates at 13.56 MHz on ISO/IEC 18000-3 standard [14]. NFC is nowadays commonly used for mobile payments (see Figure 1.6) [15]-[17].

1.4. Energy Harvesting

Energy harvesting circuits convert the energy existing in a generic environment like linear motion, pressure, light, differences in temperature, into energy that can be used electrically.

A plenty of products that convert energy from Vibration (Piezo), PhotoVoltaic (Solar) and Thermal (TEC, TEG, Thermopiles, Thermocouples) sources are already on the market [18]. They provide high efficiency conversion to regulated voltages or to charge batteries and super capacitor storage elements. Usually these systems use boost capable of operating with a very low input signal, from 20mV [18] down to 10mV [19].

In this context of battery-free systems a small amount of energy is available, typically in the $\mu W/cm^2$ order of magnitude [20].

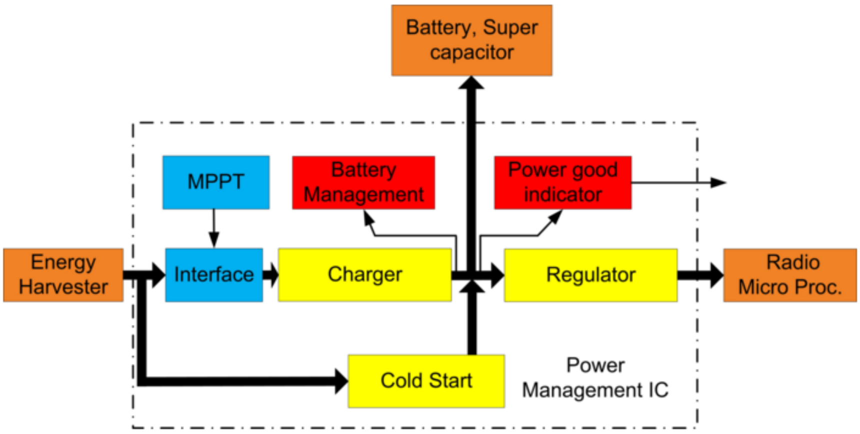


Figure 1.7. Energy-harvesting system.

As a consequence the complex IC which has to manage the harvesting process has to ensure ultra-low power operation. Moreover the electronics in these systems has to manage the problem of the intermittent and small amount of energy coming from the source. A typical energy-harvesting system is shown in Figure 1.7. It includes conversion, temporary storage block, a sophisticated power management circuit (MPPT), analog converters and ULP MCUs. From the design point of view the harvesting system poses several specifications like ultra-low leakage current, ultra-low voltage operation, very low power ADC with medium-high resolution.

1.5. Ultra-Low voltage VLSI design

Figure 1.7 reports the operating supply voltage as a function of the technology node considering the publications on digital circuits design [21]. Most of the published works show a typical operating voltage of about 150 -200 mV with a minimum operating voltage of only 100 mV. It is really interesting also to take a look to the evolution of the publications over the years considering the operating regime of the circuits. This can be depicted by inspecting Figure 1.8 where the publications are classified based on the operating regime [21]. It is possible to note the exponential increasing in the number of publications in which the subthreshold regime is exploited. As an example, considering the year 2013, the works in subthreshold regime are about 4 times more than the works considering the near-threshold regime.

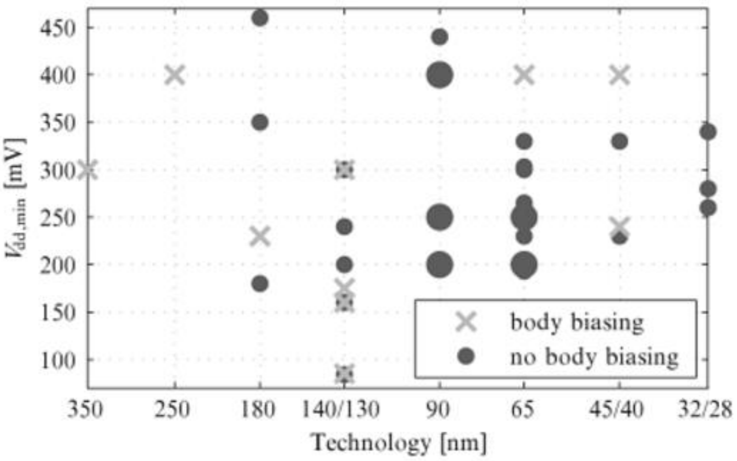


Figure 1.8. Minimum V_{DD} as a function of the technology node for digital systems [21].

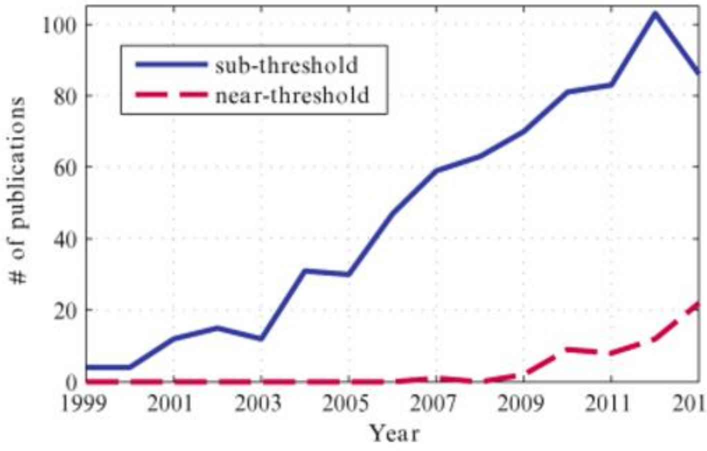


Figure 1.9. Publications on VLSI circuits over the years. The data are extracted from the IEEE Xplore database [21].

Again, the main reason of such a trend consists in the research of the maximum energy-efficiency or the minimum energy point (MEP). Combining the information reported in Figure 1.8 and Figure 1.9 the majority of the works in VLSI systems is performed in the direction of lowering the supply voltage in order to achieve subthreshold / near threshold conduction.

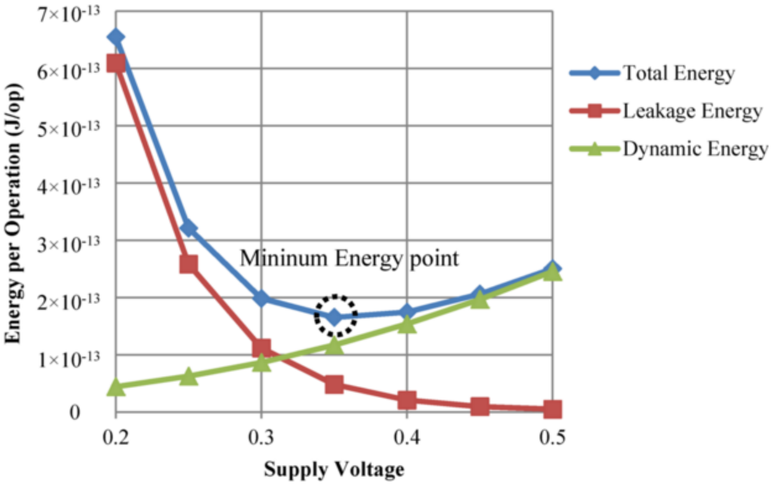


Figure 1.10. Dynamic, static and total energy in VLSI systems. The figure indicates the minimum energy point (MEP).

In fact, as well know, lowering the supply voltage has always a positive implication on the power consumed by the VLSI systems [22]-[24], as shown in Figure 1.10.
It is worth noting that subthreshold conduction ensures very low level of dynamic consumption while increasing the impact of the leakage current (subthreshold current). A lot of research is investigating what is the best region in terms of minimum energy point. Some claim the near-threshold conduction as the best [25]-[26], while in other systems the subthreshold seems the optimal operating regime for maximizing the energy efficiency of the systems [27]-[28].

1.6. Low power, low-voltage analog design

As shown before the supply voltage scaling is one of the most efficient ways to scale power consumption and to achieve energy efficient operation. This technique is particular efficient in the digital design context since the dynamic power consumption in a VLSI circuit scales by scaling the supply voltage. This concept is not true in the case of an analog circuit. Here the power consumption is defined by the energy necessary to ensure the proper operation of the circuit [29]. Specifically it is defined as the power consumed in analog signal processing circuits to maintain the signal energy above the fundamental thermal noise in order to achieve the required signal-to-noise ratio (SNR) [29]. The research of the minimum supply voltage and the minimum power consumption is still challenging from the design point of view of an analog system and a lot of research is still required in this field as reported in the technology trends report of the ISSCC [1]:

“The efficient control, storage, and distribution of energy are worldwide challenges, and are increasingly important areas of analog circuit research. While the manipulation and storage of information is efficiently performed digitally, the conversion and storage of energy is fundamentally performed with analog systems. Therefore, ***the key technologies for power management are predominantly analog****. ……….. There is also an explosion of technologies that allow energy to be collected from the environment via photovoltaic, piezoelectric, or thermoelectric transducers, with a trend toward the use of multiple sources at the same time.* ***A significant focus here is on analog circuits that are able to harvest sub-microwatt power levels from multiple energy sources at tens of millivolts, to provide autonomy for remote sensors, or to supplement conventional battery supplies in mobile devices. To achieve this, the attendant analog circuits have to consume extremely low power, so that some energy is left over to charge a battery or supercapacitor.*** *…….****Analog circuits also serve as bridges between the digital world and the analog real world****. Just like actual bridges,* ***analog circuits are often bottlenecks and their design is critical to overall performance, efficiency, and robustness****. Nevertheless, since digital circuits, such as microprocessors, drive the market, semiconductor technology has been optimized relentlessly over the past 40 years to reduce their size, cost, and power consumption.* ***Analog circuitry has proven increasingly difficult to implement using these modern IC technologies. For example, as the size of transistors has decreased, the range of analog voltages they can handle as well as their analog performance have decreased, while the variation observed in the analog parameters has increased.”***

Bibliography

[1] **ISSCC 2015 tech trends**. http://isscc.org/doc/2015/isscc2015_trends.pdf.

[2] K.A. A. Makinwa, A. Baschirotto, P. Harpe, Efficient Sensor Interfaces. **Advanced Amplifiers and Low Power RF Systems**, *Springer,* 2015.

[3] http://zeitgeistlab.ca/doc/Advanced_WBANs_for_an_Ageing_e-Health_Society.html

[4] E. Sazonov, M. Neuman, **Wearable Sensors**. 1st edition, *Elsevier*, 2014.

[5] J. Landt, "The History of RFID", in *IEEE Potentials,* vol. 24, no. 4, Oct.-November 2005, pp. 8-11.

[6] H. Stockman, "Communication be means of reflected power", in *Proceeding of IRE*, vol. 36, no. 10, October 1948, pp. 1196-1204.

[7] http://www.rfidjournal.com/articles/view?1338.

[8] L. Yang,A. R. Manos M. Tentzeris, **Design and Development of Radio Frequency Identification (RFID) and RFID-Enabled Sensors on Flexible Low Cost Substrates**. *Morgan & Claypool Publishers,* 2009.

[9] http://www.prototypexpress.com/rfidwhitepaper.htm.

[10] S. Devadas , E. Suh , S. Paral, R. Sowell, T. Ziola, V. Khandelwal, "Design and Implementation of PUF-Based Unclonable RFID ICs for Anti-Counterfeiting and Security Applications", in *IEEE International Conference on RFID The Venetian*, 2008, pp. 58-64.

[11] L. Bolotnyy and G. Robins, "Physically Unclonable Function -Based Security and Privacy", in RFID Systems, in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, 2007, pp. 211-220.

[12] http://www.rfidjournal.net/PDF_download/Verayo_031610_Webinar.pdf.

[13] H. Dirk, **RFID Security and Privacy**. *Springer*, 2008.

[14] https://en.wikipedia.org/wiki/Near_field_communication.

[15] http://www.mastercard.com/contactless/.

[16] http://www.maestrocard.com/at/privatkunden/innovation_kontaktlos.html.

[17] https://developer.visa.com/paywavemobile.

[18] http://www.linear.com/products/energy_harvesting.

[19] M. B. Machado, M. Sawan, M. Cherem Schneider; C Galup-Montoro, "10 mV – 1V step-up converter for energy harvesting applications", in *Proceeding of the 27th Symposium on Integrated Circuits and Systems Design* (SBCCI), 2014, pp. 1-5.

[20] http://www.mouser.com/pdfDocs/TI-ULP-meets-energy-harvesting-A-game-changing-combination-for-design engineers.pdf.

[21] N. Reynders**,** W. Dehaene, **Ultra-Low-Voltage Design of Energy-Efficient Digital Circuits**. *Springer* 2015.

[22] A. Wang, A. Chandrakasan, and S. Kosonocky, "Optimal supply and threshold scaling for subthreshold CMOS circuits," in *Proceeding of the IEEE Annual Symposium on VLSI*, Apr. 2002, pp. 5–9.

[23] A. Wang and A. Chandrakasan, "A 180-mV subthreshold FFT processor using a minimum energy design methodology," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.

[24] H. Soeleman, K. Roy, and B. C. Paul, "Robust subthreshold logic for ultralow power operation," *IEEE Transaction on Very Large Scale Integrated* (VLSI) *Systems*, vol. 9, no. 1, pp. 90–99, Feb. 2001.

[25] Ronald G. Dreslinski, Michael Wieckowski, David Blaauw, Dennis Sylvester, Trevor Mudge, "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits," in *Proceedings of the IEEE, Special Issue on Ultra-Low Power Circuit Technology*, vol. 98, no. 2, February 2010, pp. 253 – 266.

[26] Bo Zhai, Ronald G. Dreslinski, Trevor Mudge, David Blaauw, Dennis Sylvester, "Energy Efficent Near-threshold Chip Multi-processing," in *ACM/IEEE International Symposium on Low-Power Electronics and Design* (ISLPED), August 2007, pp.32-37.

[27] A. Pajkanovic, T.J. Kazmierski, B. Dokic, "Minimum energy point of sub-threshold operated pass-transistor circuits," in *IEEE Forum on Specification and Design Languages* (FDL), Sep. 2012, pp.202-2007.

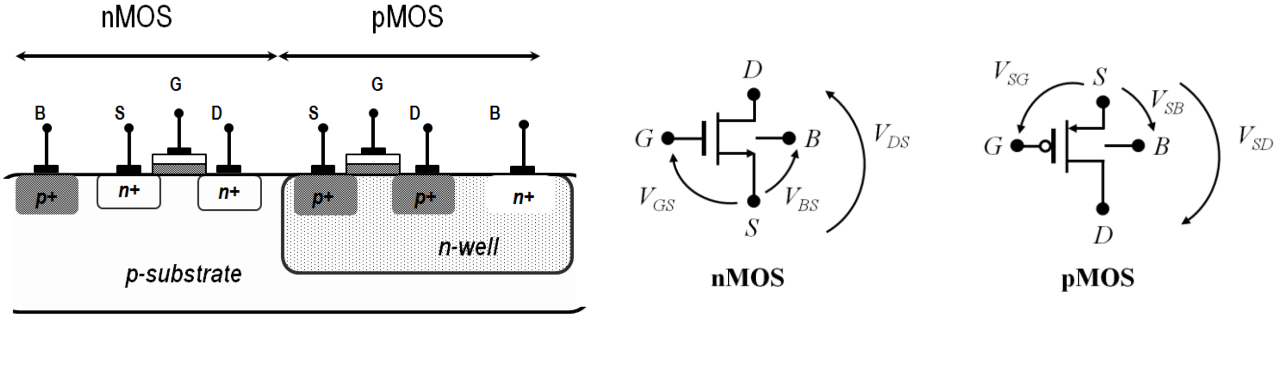
[28] A. Wang, B.H. Calhoun, A. Chandrakasan, **Sub-threshold Design for Ultra Low-Power Systems**. *Springer*, 2006.

[29] C. Enz and E. A. Vittoz, "CMOS low-power analog circuit design," in *Proceeding of IEEE International Symposium on Circuits and Systems* (ISCAS), chapter 1.2, Tutorials, pp. 79–132, 1996.

2. CMOS Design in weak inversion

2.1. Introduction

The Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET), is by far the most widespread component in the modern electronics [1]. The schematic and the cross section of the nMOS and pMOS transistor are reported in Figure 2.1(a) and 2.1(b) respectively. The four terminals (drain D, source S, body B and gate G) are used to change the operating regime of the transistor. The layer under the gate, also called channel, contains electrons in the case of the nMOSFET and holes in the cse of the pMOSFET. Thus an nMOS transistor has a p-type substrate while the pMOS transistor has an n-type substrate. The MOSFET is in able to implement the controlled switch which is ON (drop voltage equal to zero across drain and source) when the gate-source voltage is higher, in absolute value, than the threshold voltage and OFF (drain-source current equal to zero) when the gate-source voltage is lower than the threshold voltage. However the MOSFET operates in a more complex way than a simple switch. Since its introduction three regimes of operation have been identified for this component: the *cut-off regime*, the *triode regime* and the *saturation regime*. The operating regime in which the MOSFET works depends on the relationship between the gate-source (V_{GS}) voltage and the threshold voltage V_{TH} of the device.



(a) (b)

Figure 2.1. nMOS and pMOS cross section (a) and circuital symbols (b).

The operating regimes of the MOSFET as well as the models used to describe its behaviour have been revised during the years [2]-[7]. In particular the former cut-off regime has been revised discovering that the current flowing in this operating condition is not negligible, indeed the device shows an interesting exponential relationship which nowadays is widely exploited in all the low-power, low-voltage contexts [8]. This regime is nowaday knows as *weak inversion* or *subthreshold regime*.

In the following a brief overview on the different operating regimes of the MOSFET are reported with particular emphasis on the weak inversion regime of operation. The main considerations from the design point of view will be highlighted.

2.2. I-V MOSFET relationship

In this section the main relationships used to describe the behaviour of a MOSFET are reported. All the equations here reported refer to the nMOS transistor; nevertheless the equations for the pMOS can be simply derived by considering the absolute values of the voltages.

The most important parameter for defining the operating regime of the MOSFET is the threshold voltage V_{TH}. For a MOSFET the threshold voltage represents the gate-source voltage required to forms an inversion layer (channel) between the oxide and the substrate of the transistor. For an nMOSFET with long channel/width and uniform substrate doping concentration V_{TH} is defined as [4]:

$$V_{TH} = V_{TH,0} + \lambda_B \left(\sqrt{V_{SB} + \phi_S} - \sqrt{\phi_S} \right), \tag{2.1}$$

where in (2.1) ϕ_S is the surface potential, V_{SB} is the source-body voltage and V_{TH0} is the threshold voltage of a long channel device at zero substrate voltage. λ_B is the body coefficient equal to:

$$\lambda_B = \frac{t_{OX}}{\varepsilon_{OX}} \sqrt{2qN_A \varepsilon_{Si}}\ , \tag{2.2}$$

with t_{OX} and C_{OX} oxide thickness and oxide permittivity respectively, ε_{Si} silicon permittivity, q elementary charge and N_A doping concentration. Depending on the relationship between the gate-source voltage and the threshold voltage, the MOSFET can operate in different regimes of operation.

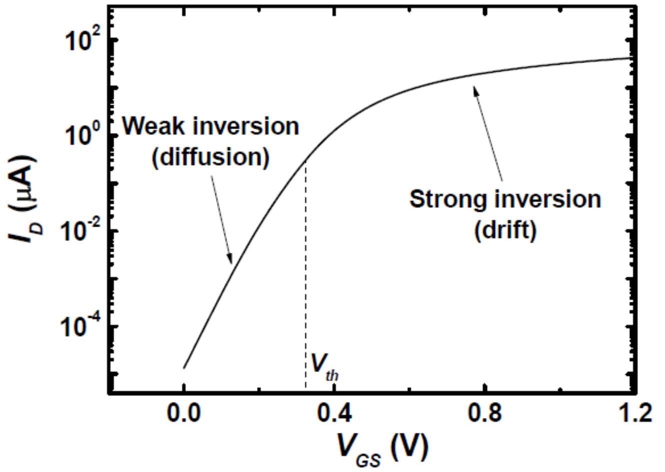


Figure 2.2. I_D versus V_{GS} for an nMOS in 180 nm technology.

Specifically, if

- ✓ $V_{GS}<V_{TH}$ the MOSFET is working in subthreshold regime or weak inversion regime;

- ✓ V_{GS}>V_{TH} the MOSFET is working in above-threshold regime. In this case if
  - V_{DS} < V_{GS}-V_{TH} then the MOSFET works in *triode regime*
  - V_{DS} > V_{GS}-V_{TH} the MOSFET is in saturation regime.

In Figure 2.2 the drain current (I_D) against V_{GS} for an nMOS in 0.18 μm is reported while Figure 2.3 shows the I_D-V_{DS} curve for an nMOS in the same technology.
There are different equations used to describe the I_D relationship in these different regimes of operation. Among them the models proposed in the BSIM [4]-[6] and in the EKV model [7] are considered nowadays as the industry standards. Despite the EKV model is more accurate in the description of the subthreshold operation, most of the industries release their design kits, commonly employed in SPICE simulators [8]-[9], using the BSIM model.

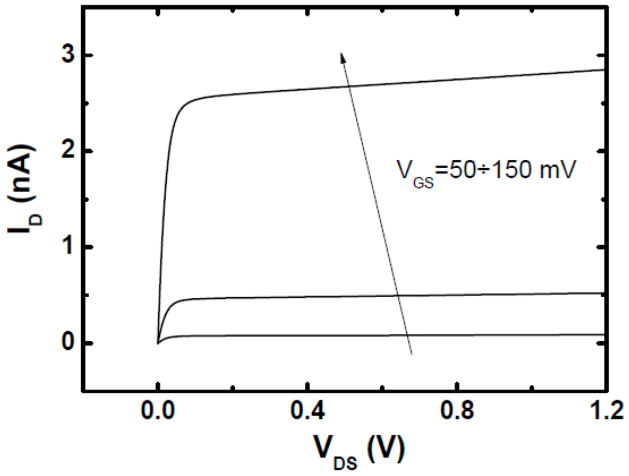


Figure 2.3. I_D versus V_{DS} for an nMOS in 180 nm technology.

In the equations (2.3)-(2.5) the most common expressions of the drain current for the subthreshold and above threshold conduction are reported. These equations represent the simplest version of the MOSFET model used in circuital simulations [10].

Regime of Operation	***Drain Current***	***Condition***	
Subthreshold	$I_{DS} = \mu_N C_{OX} \frac{W}{L} V_T^2 \exp\left(\frac{V_{GS}-V_{TH}}{nV_T}\right)\left(1-\exp\left(-\frac{V_{DS}}{V_T}\right)\right),$	$V_{GS} < V_{TH}$	(2.3)
Triode	$I_{DS} = \frac{1}{2}\mu_N C_{OX} \frac{W}{L}[2(V_{GS}-V_{TH})-V_{DS}]V_{DS},$	$V_{GS} > V_{TH}$ $V_{DS} < V_{GS} - V_{TH}$	(2.4)
Saturation	$I_{DS} = \frac{1}{2}\mu_N C_{OX} \frac{W}{L}[(V_{GS}-V_{TH})]^2(1+\lambda(V_{DS}-V_{DS,sat})).$	$V_{GS} > V_{TH}$ $V_{DS} > V_{GS} - V_{TH}$	(2.5)

Despite the simplicity with respect to the more complex models these equations are in able to provide the main information about the effect of the bias voltages on the MOS transistor.
In the equations (2.3)-(2.5) μ_N is the electron mobility, C_{OX} is the oxide capacitance per unit area transistor, W and L are the channel length and width respectively, V_T is the thermal voltage equal to kT/q (with k Boltzmann's constant, T absolute temperature and q elementary charge), n is the

subthreshold swing factor, V_{TH} is the threshold voltage, λ is the channel length modulation and $V_{DS,sat}$ the saturation drain-source voltage [11].

Since this work is focused on low power, low voltage design, in the following the parameters used in the description of the weak inversion regime are considered more in detail.

The expression (2.3) shows that the current flowing in the MOSFET depends exponentially from the V_{GS} voltage in subthreshold regime. This means that the gate terminal has a very good control of the channel and consequently of drain current. Indeed, the gate transconductance $g_m=\partial I_D/\partial V_{GS}$ in this regime is higher than in strong inversion regime as reported in the equations (2.6) and (2.7).

Regime of Operation	***Drain Current***	
$V_{GS} < V_{TH}$	$g_m = \frac{I_D}{nV_T}$	(2.6)
$V_{GS} > V_{TH}$	$g_m = \sqrt{2\mu_N C_{OX} \frac{W}{L} I_D}$	(2.7)

The parameter n, which is included in the expression of the drain current in subthreshold regime, represents a loss of coupling efficiency between the gate and channel caused by the body, which acts as a back gate. In weak inversion, n is related to the capacitive voltage division between the gate voltage and silicon surface potential resulting from the gate-oxide, depletion, and interface state capacitances [11]. It can be expressed as:

$$n \approx 1 + \frac{C_{DEP}}{C_{OX}}, \tag{2.8}$$

where C_{DEP} is the depletion capacitance per unit area. This parameter is usually expressed in the subthreshold swing (SS):

$$SS \approx \ln(10) n V_T, \tag{2.9}$$

which defines the increment in V_{GS} necessary to achieve an 10 × increment in the drain current. Graphically it represents the slope of the curve I_D-V_{GS} in weak inversion regime (linear region in Figure 2.2).

The expression of the subthreshold current reported in (2.3) shows that V_{DS} effects I_D only because of term (1-exp(-V_{DS}/V_T)). This contribution is negligible if V_{DS}>4V_T which is equal to about 104 mV at 25 °C. Thus, in such condition (saturation regime) the following equation can be used to describe the drain current in subthreshold regime:

$$I_{DS} \approx \mu_N C_{OX} \frac{W}{L} V_T^2 \exp\left(\frac{V_{GS} - V_{TH}}{nV_T}\right). \tag{2.10}$$

However V_{DS} influences the drain current also because of the Drain-Induced-Barrier-Lowering (DIBL) effect. From the design point of view this effect bring to a variation of the threshold voltage and consequently to a variation of the drain current. As reported in (2.1) also the body-source voltage can cause such a change. To model both effects, the following equation is commonly used [12]:

$$V_{TH} = V_{TH0} - \lambda_{DS} V_D - \lambda_B V_{BS}, \tag{2.11}$$

where V_{TH0} is the zero-biased threshold voltage extracted at $V_{DS}=V_{BS}=0$ while λ_D and λ_B represent the DIBL and body coefficient already defined in (2.2). The DIBL coefficient has an exponential relationship with the MOSFET channel length [13], thus long channel transistor can reduce this effect. In Figure 2.4 simulation results on λ_D obtained for an nMOS in 0.18 μm is reported. The SPICE simulations confirm the exponential relationship between λ_D and the channel length.

On the other hand the reduction of the body effect cannot be performed by simply adjusting the geometry of the transistor. The only solution to mitigate this effect consists in using a tripe well technology which allows performing the condition of $V_{BS}=0$ also for the stacked nMOS transistors. This results in a larger occupied area because of the additional masks [14].

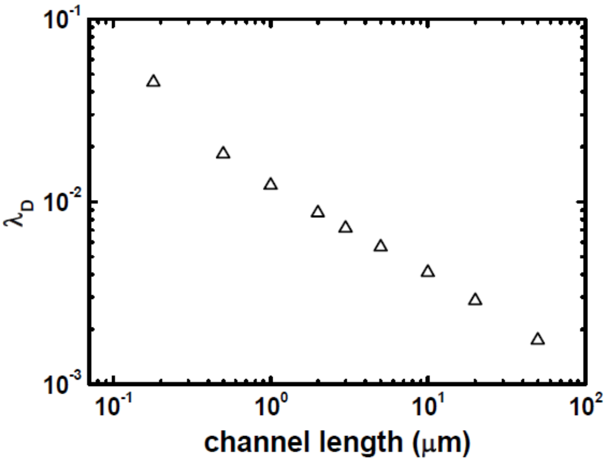


Figure 2.4. λ_D (mV/V) against V_{DS} for an nMOS in 180 nm technology.

2.2.1. Small signal behaviour

parameter	***sub-threshold***		***above-threshold***	
$g_m = \dfrac{\partial I_D}{\partial V_{GS}}$	$\dfrac{I_D}{nV_T}$	(2.12a)	$\sqrt{\dfrac{2\mu_N C_{OX} W I_D}{L}}$	(2.12b)
$g_{mb} = \dfrac{\partial I_D}{\partial V_{BS}}$	$\dfrac{\lambda_B I_D}{nV_T}$	(2.13a)	$\dfrac{g_m \lambda_B}{2\sqrt{\lvert V_{SB}\rvert + \phi_S}}$	(2.13b)
$r_d = \left[\dfrac{\partial I_D}{\partial V_{DS}}\right]^{-1}$	$\dfrac{nV_T}{\lambda_D I_D}$	(2.14a)	$\dfrac{1}{\lambda I_D}$	(2.14c)

The MOSFET equivalent circuit valid both in the case of the above threshold regime and the weak inversion regime is reported in Figure 2.5. Despite the same model the small signal parameters are different for the two operating regime. The values of these parameters in above- and sub-threshold regime are reported in the equations (2.12a)-(2.14c). It is possible to note that the transconductance efficiency g_m/I_D is higher in weak inversion than in strong inversion. Moreover the g_m/I_D ration in subthreshold regime is independent from the geometry of the transistor except for the dependence of n from transistor's area. The intrinsic DC gain offered by the MOSFET is significantly higher in weak inversion ($1/\lambda_D$) than in strong inversion regime ($\propto 1/\lambda\sqrt{\boldsymbol{I_D}}$).

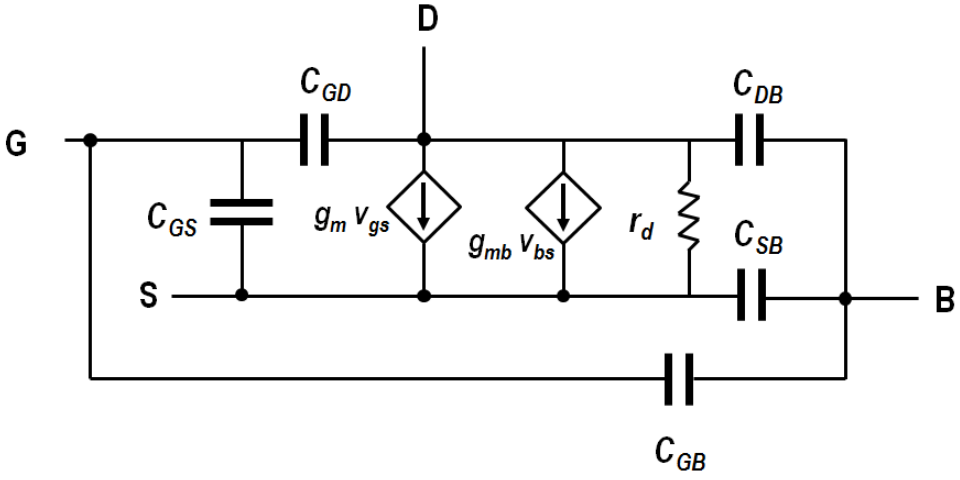


Figure 2.5. MOSFET small-signal equivalent circuit.

TABLE 2.I. GATE CAPACITANCE FOR DIFFERENT OPERATING REGIMES

Capacitance	***Sub-threshold***	***Triode***	***Saturation***
C_{GB}	WLC_{OX}	Negligible	Negligible
C_{GD}	WL_DC_{OX}	$\frac{1}{2}WLC_{OX}+WL_DC_{OX}$	WL_DC_{OX}
C_{GS}	WL_DC_{OX}	$\frac{1}{2}WLC_{OX}+WL_DC_{OX}$	$\frac{2}{3}WLC_{OX}$
$C_{GG}(total)$	$WLC_{OX}+2WL_DC_{OX}$	$2WL_DC_{OX}+WLC_{OX}$	$\frac{2}{3}WLC_{OX}+WL_DC_{OX}$

The small signal analysis allows also understanding the frequency behaviour of the MOSFET in the different operating regimes.

As reported in Table 2.I the values of the MOSFET capacitances are different in the different operating regimes [15]. The values reported in the table take into account also the parasitic capacitances due to the overlaps in the drain and source regions (L_D indicates the length of the overlap which is considered equal for the drain/source region) [15].

The frequency performances of the MOSFET can be investigated by considering the intrinsic bandwidth of the MOSFET, defined as the frequency where current gain from the gate input to the short circuit, drain output is equal to unity. It is defined as [11]:

$$f_{Ti}=\frac{g_m}{2\pi\left(C_{GS}+C_{GB}\right)}. \tag{2.15}$$

Considering the (2.15) and substituting the values of C_{GS} and C_{GB} in weak inversion and strong inversion, the following expressions are obtained [11]:

sub-threshold

$$f_{Ti}\approx\left(\frac{n}{n-1}\right)\frac{\mu_0V_T}{\pi L^2}IC, \tag{2.16a}$$

above-threshold

$$f_{Ti}\approx\left(\frac{n}{n-1/3}\right)\frac{\mu_0V_T}{\pi L^2}\sqrt{IC}, \tag{2.16b}$$

where in (2.16a) and (2.16b), the inversion coefficient (*IC*) is defined as [11]:

$$IC = \frac{I_D}{2\mu_0 n_0 C_{OX} V_T^2 \left(\frac{W}{L}\right)}. \tag{2.17}$$

In (2.16a), (2.16b) and (2.17) μ_0 and n_0 represent the mobility and the subthreshold swing factor at the room temperature. The value of *IC* is in able to define the operating regime of the MOSFET depending on the magnitude of the drain current I_D. Figure 2.6 reports the relationship between *IC* and the operating regime. Since for the purpose of this work the concept of the *IC* is not essential is will not deeply investigated, however the MOSFET modeling based on the *IC* is becoming very popular also in the circuital community thanks to the diffusion of the EKV model [7], [11]. According to the definition of the *IC* parameter, from (2.16a) - (2.16b), the intrinsic bandwidth of the transistor is strongly reduced by increasing the channel length in both regime of operation.

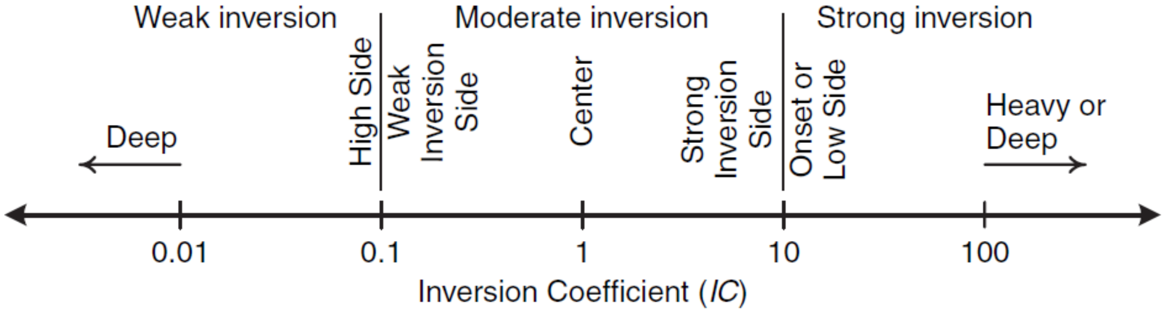


Figure 2.6. MOSFET operating regime as a function of the inversion coefficient [11].

The f_{Ti} increases in weak inversion at higher current and reach its maximum value near the moderate inversion regime. Due to the higher drain current, the frequency response of the MOSFET in weak inversion is reduced compared to the strong inversion regime.

2.2.2. Temperature effect

Unlike the above threshold regime in which the current is mostly due to a drift mechanism, the subthreshold conduction is dominated by the diffusion mechanism.

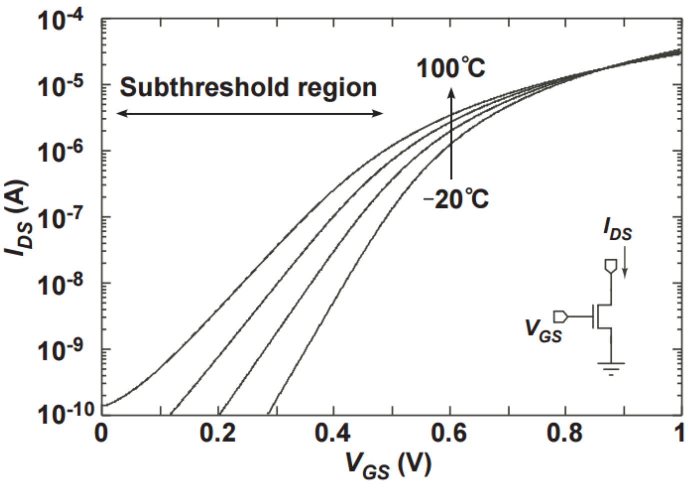


Figure 2.7. I_D current against V_{GS} for different temperatures in subthreshold and above threshold regime.

As a consequence the drain current depends on temperature more than in above threshold regime as shown in Figure 2.7. From the design point of view the main parameters which depend on temperature are the mobility, the thermal voltage, and the threshold voltage while a negligible influence of the temperature on the subthreshold swing factor is commonly assumed in the typical temperature range [11]. Equations (2.18)-(2.20) report the temperature dependence of these parameters [11].

Parameter	***Temperature Dependence***	
V_T	$V_T = 25.85mV \dfrac{T}{300K}$	(2.18)
μ_N	$\mu_N = \mu_N(T_0)\left(\dfrac{T}{T_0}\right)^{-m}$	(2.19)
V_{TH}	$V_{TH} = V_{TH}(T_0) - \kappa(T - T_0)$	(2.20)

In (2.19) $\mu_N(T_0)$ represents the electron mobility at the room temperature T_0, while m is the mobility exponent which defines the relationship between the electron mobility and the temperature. Typically m is equal to about 1.2 - 2 for the nMOS. In (2.20) κ is the coefficient which defines the drops of the threshold voltage at higher temperatures. From (2.20) the threshold voltage decreases linearly by increasing the temperature.

From the previous equations it is possible to note that the temperature-dependent parameters are the same both in the case of subthreshold and above-threshold conduction. Nevertheless, since the subthreshold current depends exponentially on threshold voltage in subthreshold regime, the current becomes more sensitive on temperature in weak inversion than in strong inversion.

2.2.3. Process stability

Working in subthreshold regime the designer has to take care, more than in above threshold design, of the problem related to the process variability. Due to the variations in the oxide thickness, doping concentration, Line Edge Roughness (LER) [16]-[17], the behaviour of the MOSFET can differ from the expected one. Because of this, the design should satisfy the expected requirements also in the worst case conditions of process variability. The latter concept is valid for any circuit independently from the operating regime. In subthreshold regime however this problem is emphasized since every single deviation from the nominal behaviour affects the threshold voltage which in turn affects exponentially the drain current. As a consequence a severe variation in the behaviour of the different components (delay, leakage current, stability, etc.) can be observed.

The variations of the process parameters can be classified into **inter-die** and **intra-die** variations. Inter-die or global variations refer to the lot-to-lot (L2L), wafer-to-wafer (W2W) and die-to-die (D2D) variations while the intra-die variations consider the variability at die level or within-die variations (WID) as shown in Figure 2.8 [17]. The intra-die variability can be classified in systematic and random. The systematic variations are usually strongly layout-dependent [17], while the random variations are completely unpredictable and result from the unpredictable process variability during fabrication. All the previous variations consider both the spatial as well as the temporal variations in the process which cause different dies and wafers to have different process parameters such as oxide thickness, dopant concentration, etc.

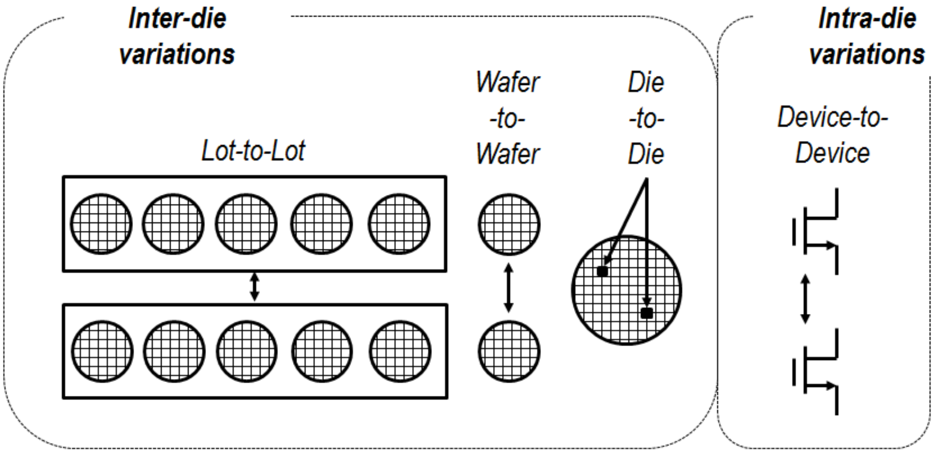


Figure 2.8. Variability in CMOS process.

These variations are caused by of the differences in the manufacturing equipment during the process. If the parameters vary rapidly over distances smaller than the dimension of a die they result in WID variations while if the variations are more gradual over the wafer they will cause D2D variations. Considering the process steps, even more variability is present from wafer to wafer (W2W variations) and between different manufacturing runs (L2L variations).

The estimation of the effects of process variations is performed through SPICE simulations using intra-die and inter-die mismatch models and process corners models provided by the foundry as part of the SPICE models in the process design kit (PDK).

The effect of the random variations is defined using the Pelgrom's law [18]. According to Pelgrom, the following equation describes the standard deviation on threshold voltage between the two matched transistors on the same wafer:

$$\sigma(V_{TH}) = \frac{A_{VT}}{\sqrt{W \times L}}, \tag{2.21}$$

where A_{VT} is a constant depending on the technology, while W and L are the effective channel width and length of the transistor respectively. In Figure 2.9 the standard deviation of the threshold voltage against the $1/(W \text{ X } L)^{0.5}$ factor is reported for 0.18 µm CMOS process.

On the other hand to model the L2L variability the corner analysis is performed. The foundry provides information about the maximum variability in the process such as maximum/minimum oxide thickness, maximum/minimum value of the doping concentration, maximum/minimum value of the threshold voltage. In this case the variations are classified considering separately the parameters in the nMOS and pMOS transistors, thus five corners are specified: TT (Typical nMOS, Typical pMOS), SS (Slow nMOS, Slow pMOS), FF (Fast nMOS, Fast pMOS), SF (Slow nMOS, Fast pMOS) and FS (Fast nMOS, Slow pMOS).

The problem of the process variability is particular important in structures in which a very good matching among different transistors is required such as input pair of operational amplifiers and current mirrors. Thus, to mitigate the effect of the mismatch, large area *W*/*L* transistors should be employed. Nevertheless it is worth noting that the equation (2.21) takes into account only for the random variations in CMOS process such as random dopant fluctuation but it does not take into account for systematic variations like temperature gradient or mechanical stress [19].

Usually the compensation for these effects is performed at layout level using interdigitated (Figure 2.10 (a)) or common centroid layout (Figure 2.10 (b)) [19]. Due to the layout rules required by process these solutions lead to a higher occupied area.

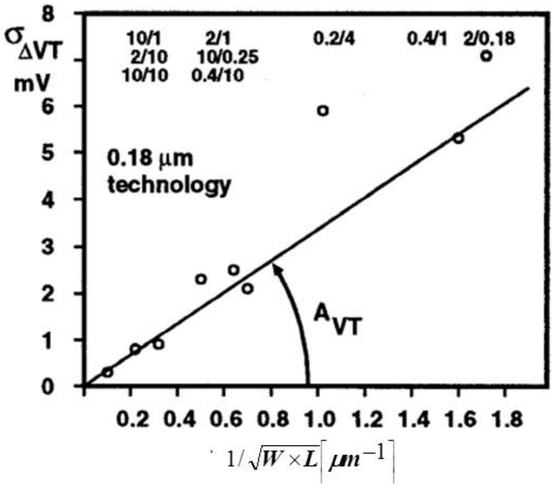


Figure 2.9. Standard deviation of V_{TH} as a function of transistor sizes in 0.18 μm CMOS.

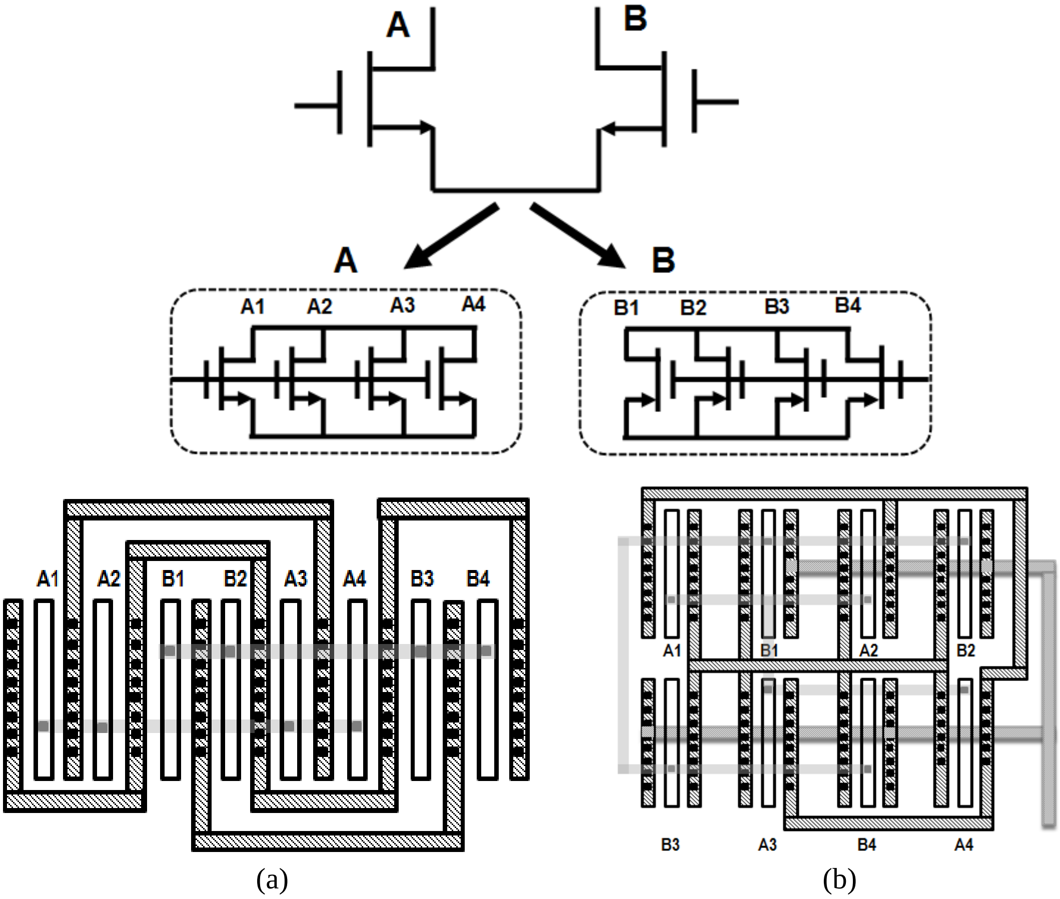


Figure 2.10. Interdigitated (a) and common centroid layout (b) for an nMOSFETs input pair.

Bibliography

[1] J. David Irwin. **The Industrial Electronics Handbook**, 2nd Edition, *CRC Press*, 2011.

[2] C.T. Sah, "Evolution of the MOS transistor from conception to VLSI", in *Proceeding of IEEE*, vol. 76, no 10, Oct. 1988, pp. 1280-1326.

[3] C.T. Sah, "A History of MOS Transistor Compact Modeling", in *Technical Proceeding of Workshop on Compact Modeling*, 1992, pp. 347-390.

[4] W. Liu, X. Jin, J. Chen, M. Jeng, Z. Liu, Y. Cheng, K. Chen, M. Chan, K. Hui, J. Huang, R. Tu, P. Ko, and C. Hu, " **BSIM3v3.2 MOSFET Model and Users' Manual**". http://www-device.eecs.berkeley.edu/~bsim3.

[5] W. Liu. **Mosfet Models for Spice Simulation: Including BSIM3v3 and BSIM4**, *John Wiley & Sons*, 2001.

[6] W. Liu, X. Jin, J. Chen, M. Jeng, Z. Liu, Y. Cheng, K. Chen, M. Chan, K. Hui, J. Huang, R. Tu, P. Ko and C. Hu. **BSIM4.6.4 MOSFET Model and Users' Manual**.

[7] C. C. Enz, F. Krummenacher, E.A. Vittoz, "An Analytical MOS Transistor Model Valid in All Regions of Operation and Dedicated to Low-Voltage and Low-Current Applications", *Analog Integrated Circuits and Signal Processing Journal on Low-Voltage and Low-Power Design*, 8. pp. 83–114, July 1995.

[8] L. W Nagel and D. O. Pederson, SPICE (Simulation Program with Integrated Circuit Emphasis), *Memorandum No. ERL-M382,* University of California, Berkeley, Apr. 1973.

[9] http://bwrcs.eecs.berkeley.edu/Classes/IcBook/SPICE/.

[10] http://people.rit.edu/lffeee/SPICE.pdf.

[11] D. Binkley, **Tradeoffs and Optimization in Analog CMOS Design**. *Wiley*, 2008.

[12] M. Alioto, "Understanding DC behaviour of subthreshold CMOS logic through closed-form analysis," *IEEE Transaction on Circuits and Systems I*, Regular Papers, vol. 57, no. 7, pp. 1597–1607, Jul. 2010.

[13] Z.H. Liu, C. Hu, J.H. Huang, T.Y. Chan, M.C. Jeng, P.K. Ko, and Y.C. Cheng, "Threshold Voltage Model For Deep-Submicrometer MOSFETs," *IEEE Transaction on Electron Devices*, vol. 40, pp. 86-95, Jan., 1993.

[14] S. H. Voldman, **Latchup**, chapter 5, *Wiley*, 2007.

[15] D. K. Schroder, **Semiconductor Material and Device Characterization**, 3rd Edition, *Wiley*, 2006.

[16] A. Asenov, S. Kaya, J.H. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," *IEEE Transaction on Electron Devices*, vol. 49, no 112, pp. 112-119, 2002.

[17] J. Kawa, **Design for Manufacturability and Yield for Nano-Scale CMOS**. *Springer*, 2007.

[18] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 1, pp. 1433-1439, Oct. 1989.

[19] A. Hastings, **The Art of Analog Layout. Englewood Cliffs**. *NJ: Prentice-Hall*, 2001.

3. Low-voltage, low-power subthreshold voltage references

3.1. Introduction

As a consequence of the explosion of applications that require low power consumption, subthreshold circuits have gained much interest in the design community. RFID, implantable medical devices, micro-sensor networks, microcontroller unit (MCU) and digital signal processing (DSP) of portable devices are typical examples of applications that benefit from low energy operation [1].

In these devices the requirement for small size and weight imposes the use of small batteries which provide a small amount of energy for the different building blocks. At the same time, as explained in Chapter 1, many portable applications impose also the requirement of long lifetime such as in the case of passive RFIDs and implantable medical devices. Therefore the energy and size requirements can be simultaneously guaranteed only by using circuits that operate at low supply voltage and with low power consumption.

In such a context, the design of the voltage references becomes particularly challenging. Voltage references are used in all analog, digital and mixed-signal systems to generate a constant output voltage irrespective of temperature, process and supply voltage variations. There are several approaches to design a voltage reference. The most common solution consists in the bandgap voltage reference (BGR) implemented in bipolar technology [2]. Nevertheless, in order to ensure a major compatibility with the rest of the system, several works have implemented the operating principle of the classical BGR in CMOS process by exploiting the parasitic vertical BJTs [3]-[7] or BiCMOS technology [8]. However, all these solutions exhibit power consumption and a minimum supply voltage that are both too large for the typical low-power applications.

Alternative approaches employ MOSFETs working in strong inversion regime [9]-[10], part in strong inversion and subthreshold regime [11]-[14] or all in subthreshold regime [15]-[17]. Some of them use MOSFETs with same threshold voltage [10]-[11], [13]-[16] while in other cases MOSFETs with two different threshold voltages are employed [9], [12], [17].

Among the different approaches, subthreshold design represents the most promising solution for extreme low-power, low-voltage applications. Subthreshold operation results in a very low minimum supply voltage and in nW power consumption [16]-[17].

However, despite the significant advantages in terms of minimum supply voltage and power consumption, subthreshold operation poses several design issues. Of the utmost importance is the high process sensitivity due to the exponential relationship between the drain current and the threshold voltage [18], which is the most important process-dependent parameter.

In the following sections of this chapter the main performance specifications for a voltage reference and a brief overview on the most significant solutions are reported. Finally two new very low-voltage, low-power solutions are introduced. The main design issues are addressed and the measurement results are reported and compared with the other solutions proposed so far.

3.2. Voltage reference performance specifications

From the operating point of view, a voltage reference is a three terminals device. It receives two inputs, the positive (V_{DD}+) and the negative (V_{DD}-) supply voltage, and generates a voltage as output (V_{OUT}). Since the main aim in the design of a voltage reference consists in generating a very stable output voltage, the main design specifications have the purpose to provide information about the stability of the generated voltage across different operating conditions like supply voltage V_{DD}, temperature, noise and process variations. Aside the above requirements, a voltage reference, especially if designed for the above mentioned applications, should satisfy the typical VLSI metrics such as the lowest possible power consumption and the minimum occupied area.

3.2.1. Line sensitivity

The line sensitivity (LS) of a voltage reference measures the stability of the output voltage against supply voltage variation. It is defined as

$$LS(\%) = 100 \times \frac{\Delta V_{REF}}{V_{REF} \times \Delta V_{DD}} \left[\frac{\%}{V} \right]. \tag{3.1}$$

In (3.1) ΔV_{REF} is the variation in Volts of the voltage reference (=$V_{REF,max}$-$V_{REF,min}$) over the supply voltage range ΔV_{DD}=$V_{DD,min}$-$V_{DD,max}$, with $V_{DD,min}$ and $V_{DD,max}$ minimum and maximum operating supply voltage respectively. V_{REF} is the nominal value of the reference voltage. Usually V_{REF} is defined as the voltage at the room temperature for the minimum supply voltage. Thus, for the maximum stability with respect to V_{DD} variations, the LS should be as low as possible.

3.2.2. Temperature coefficient

The temperature coefficient (TC) measures the stability of the reference voltage against temperature variations. It is defined as:

$$TC = 10^6 \times \frac{\Delta V_{REF}}{V_{REF} \times \Delta T} \left[\frac{ppm}{°C} \right]. \tag{3.2}$$

In (3.2) ΔV_{REF} is defined as the difference between the maximum and minimum voltage reference over the operating range of temperature ΔT=T_{MAX}-T_{MIN}. In analogy with the LS also here V_{REF} is defined as the nominal output voltage evaluated at the room temperature for the minimum operating voltage. For the best stability the TC has to be as low as possible.

3.2.3. Power-Supply-Rejection-Rate (PSRR)

The power supply rejection rate (*PSRR*) is a measure of the capability of a voltage reference to reject the noise coming from the supply voltage. It is evaluated as

$$PSRR = 20\log_{10}\left(\frac{v_{out}}{v_{dd}}\right)[dB], \tag{3.3}$$

where v_{out}/v_{dd} is the small signal ratio between the reference voltage and the supply voltage. The *PSRR* can be evaluated at different frequencies and considering the noise coming from V_{DD}+ (*PSRR* +) and V_{DD}- (*PSRR*-).

3.2.4. Process stability

One of the most important requirements for a voltage reference is the stability against process variations. Ideally, the same voltage reference should provide the same value if fabricated *N* times across different wafers or on the same wafer. To evaluate the process stability it is necessary to evaluate the *coefficient of dispersion* (σ/μ) of *N* voltage references produced in the same wafer and/or different wafers. It is defined as

$$\frac{\sigma}{\mu} = 100 \times \frac{\sqrt{\dfrac{\sum_{i=1}^{N-1}\left(V_{REF}(i) - \mu\right)^2}{N-1}}}{\dfrac{1}{N}\sum_{i=1}^{N-1} V_{REF}(i)}[\%], \tag{3.4}$$

where μ is the mean value while σ is the standard deviation of the *N* samples. It is clear that the minimum value of σ/μ is necessary for the maximum process stability.

3.3. Voltage references

The very first voltage reference circuit, **the bandgap voltage reference (BGR)**, was proposed by Widlar [2]. The basic concept behind the idea of the BGR is to perform compensation between a proportional-to-absolute temperature (PTAT) voltage and complementary-to-absolute temperature (CTAT) Voltage.

For the generation of the CTAT voltage the base-to-emitter (V_{BE}) voltage of a bipolar transistor or the forward voltage of a *pn* junction was exploited. Considering the case of a BJT, the collector current I_C can be expressed as:

$$I_C = I_S \exp\left(\frac{V_{BE}}{V_T}\right), \tag{3.5}$$

where I_S is the saturation current equal to $I_S = \mu k T n_i^2$ with μ minor carrier mobility, *k* Boltzmann’s constant, *T* absolute temperature in Kelvin and n_i the intrinsic minority carrier concentration. Introducing in (3.5) the temperature dependence of the mobility and carrier concentration, the following expression can be obtained [19]:

$$I_S = bT^{4+m} \exp\left(-\frac{E_g}{kT}\right). \tag{3.6}$$

In (3.6) b is a proportional factor, m is the temperature coefficient of the mobility and E_g the energy gap of the silicon equal to about 1.12 V. Considering that $V_{BE}=V_T ln(I_C/I_S)$, the TC of such a voltage is equal to:

$$TC(V_{BE})=\frac{\partial V_{BE}}{\partial T}=\frac{\partial V_T}{\partial T}\ln\left(\frac{I_C}{I_S}\right)-\frac{V_T}{I_S}\frac{\partial I_S}{\partial T}=\frac{V_{BE}-(4+m)V_T-E_g/q}{T}. \tag{3.7}$$

From (3.7) the TC of the V_{BE} voltage is a function of the magnitude of the same voltage. Thus, for a given value of V_{BE}, its TC can be evaluated. However since the TC is always lower than 0, V_{BE} is a CTAT voltage.

Regarding the generation of the PTAT voltage the difference between two V_{BE} voltages can be exploited [19]. In particular, if two different currents are injected into two BJTs with equal geometry, the difference between the two V_{BE} is a PTAT voltage. This can be demonstrated by simply considering the (3.5) and assuming in BJT1 a current $I_C=nI$ and in BJT2 a current $I_C=I$. In this case the difference between the two V_{BE} is equal to:

$$\Delta V_{BE}=V_{BE,1}-V_{BE,2}=V_T\ \ln(n). \tag{3.8}$$

Having a PTAT and a CTAT voltage, a temperature compensated voltage can be obtained. A possible circuit which can perform such compensation is reported in Figure 3.1. The output voltage V_{OUT} is equal to:

$$V_{OUT}=V_{BE2}+V_T\left(1+\frac{R_2}{R_3}\right)\ln(n). \tag{3.9}$$

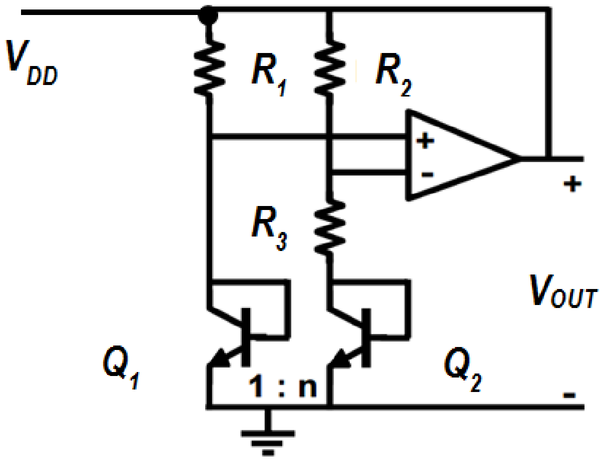


Figure 3.1. Bandgap voltage reference (BGR) [19].

Thus, for a zero TC voltage reference, the term R_2/R_3 should be properly chosen. Since at room temperature $\partial V_{BE}/\partial T$=-1.5 mV/K, and $\partial V_T/\partial T$=0.0087 mV/K, a possible solution to achieve temperature compensation consists in imposing

$$\left(1+\frac{R_2}{R_3}\right)\ln(n)=\frac{0.087}{1.5}=17.2, \tag{3.10}$$

which leads to a ratio R_2/R_3=4.

The classical implementation of the BGR is in able to generate an output voltage equal to about 1.25 V [2]-[19], quite close to the bandgap energy of the silicon. As a consequence the classical solution of the BGR does not allow obtaining a solution capable of operating at V_{DD} lower than 1.25 V. Another important problem of the classical BGR consists in the compatibility of such a component with the rest of the CMOS system.

As a consequence different solutions have been proposed to implement such a component also in CMOS process by exploiting the vertical parasitic BJT available in the standard CMOS process (see Figure 3.2) or by using a BiCMOS process.

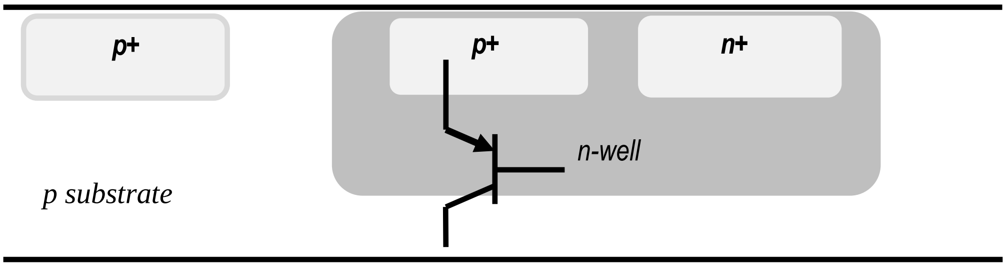


Figure 3.2. Parasitic BJT in the standard CMOS process.

A sub-1 V BRG voltage reference completely compatible with the standard CMOS process was proposed by Bamba [5]. The schematic of the voltage reference is reported in Figure 3.3.

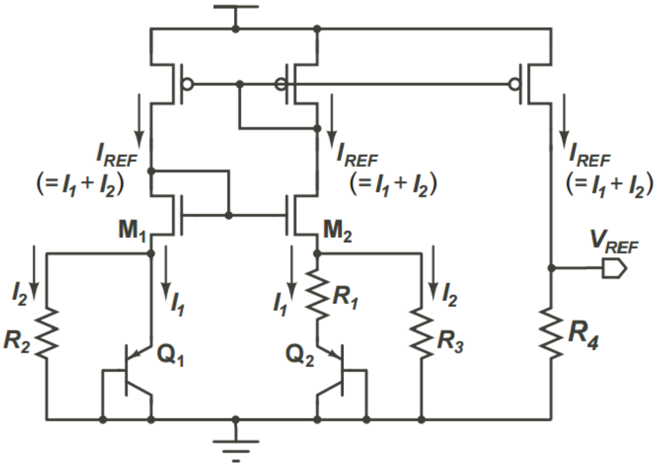


Figure 3.3. Sub-1V voltage reference proposed in [5].

In this solution the operating currents are equal to

$$
\begin{gathered}
I_1 = \frac{V_{BE2} - V_{BE1}}{R_1} = \frac{V_T \ln\left(K_2 / K_1\right)}{R_1}, \\
I_2 = \frac{V_{BE1}}{R_2},
\end{gathered}
\tag{3.11}
$$

where K_1 and K_2 indicate the geometry of the BJT Q1 and Q2 respectively. The current $I_{REF}=I_1+I_2$ is injected through the resistor R_4, generating the output voltage V_{REF}:

$$V_{REF} = \frac{R_4}{R_2} V_{BE1} + \frac{R_4}{R_1} V_T \ln\left(K_2 / K_1\right), \qquad (3.12)$$

thus imposing $\partial V_{REF}/\partial T=0$ the value of R_1, R_2, R_3, R_4, K_1 and K_2 which allows the temperature compensation can be obtained. The solution proposed in [5] operates at V_{DD}=0.84 V generating a reference voltage of about 0.5 V. The information on power consumption is not reported.

An alternative sub-1V BGR in BiCMOS technology is proposed in [8]. The schematic of the solution is reported in Figure 3.4.

In this solution the operational amplifier forces the voltages V_A and V_B to be equal. As a consequence the current flowing in the two nominally equal resistors R_1 and R_2 is equal and proportional to V_{BE}. Thus, the current flowing in M_1, M_2 and M_3 is equal to:

$$I_1 = \frac{V_{BE}}{R_1} + \frac{V_T \ln(n)}{R_0}. \qquad (3.13)$$

Since the output voltage is coincident with the voltage across R_3, it can be expressed as:

$$V_{OUT} = I_1 R_3 = \frac{R_3}{R_1}\left[V_T \frac{R_1}{R_0} \ln(n) + V_{BE}\right]. \qquad (3.14)$$

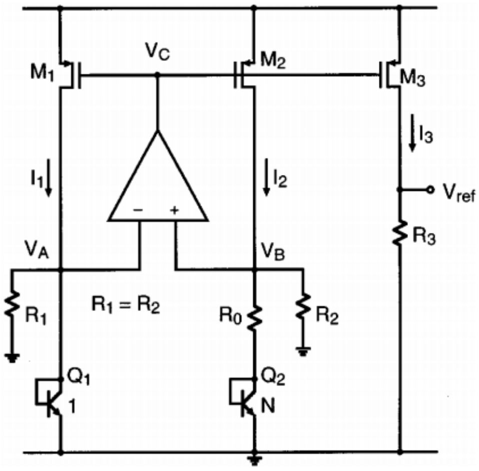


Figure 3.4. Voltage reference proposed in [8].

From (3.14) the temperature compensation of V_{OUT} is achieved by choosing the right value of n and R_1/R_0. Specifically from (3.14), to obtain TC=0, the following condition holds [8]:

$$\frac{R_1}{R_0} \ln(n) = 22. \qquad (3.15)$$

The circuit proposed in [8] starts to work from 1 V, generating a reference voltage of about 0.5 V. The power consumption is 92 μW.

3.3.1. Low-voltage, low-power voltage references

Despite the previous solutions are in able to ensure the sub-1V operation and the compatibility with the CMOS process, they result in high power consumption, widely over 1 μW. As a consequence several solutions have tried to obtain low voltage operation and low power consumption at the same time. An improvement in terms of power consumption can be obtained by using fully-CMOS solutions instead of Bi-CMOS solutions.

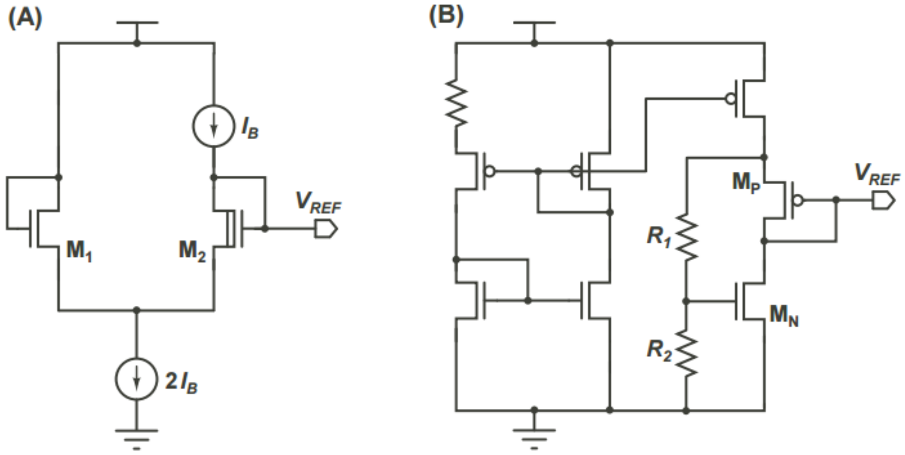


Figure 3.5. Voltage reference based on the difference between the gate-source voltages of two nMOS transistors [9] (a) and nMOS and pMOS transistors [10] (b).

In Figure 3.5 the solution proposed in [9] (a) and [10] (b) is reported. Both solutions exploit the difference between the gate-source voltages of two nMOS (a) or an nMOS and a pMOS (b) transistor to obtain a temperature compensated voltage reference. If the transistors work in strong inversion regime, the drain current can be expressed as:

$$I_{DS} = \frac{\beta}{2}\frac{W}{L}\left(V_{GS} - V_{TH}\right)^2 . \tag{3.16}$$

The expression (3.16) is equivalent to (2.5) with the condition of λ=0 and β=μC_{OX}. In the case of the circuit of Figure 3.5(a) two different threshold voltage devices are employed. The transistor M_1 is a regular nMOS transistor while M_2 is a high threshold voltage nMOS transistor. Assuming the same temperature coefficients for the two threshold voltages, V_{REF} becomes equal to:

$$V_{REF} = \left(V_{TH2}\left(T_0\right) - \kappa T\right) - \left(V_{TH1}\left(T_0\right) - \kappa T\right) + \sqrt{\frac{2I_B}{\beta}}\left(\frac{1}{\sqrt{K_1}} - \frac{1}{\sqrt{K_2}}\right). \tag{3.17}$$

Where in (3.17) K_i=$(W/L)_\mathrm{i}$. is the aspect ratio of the *i*-th transistor. By proper sizing M_1 and M_2 the V_{REF} can be compensated in temperature obtaining a value equal to:

$$V_{REF} \approx V_{TH2}\left(T_0\right) - V_{TH1}\left(T_0\right). \tag{3.18}$$

In the case of the solution proposed in [10], the difference between the gate-source voltage of a pMOS and an nMOS is exploited to obtain a stable reference voltage. In this case, according to the schematic of Figure 3.5(b), the voltage reference is equal to

$$V_{REF} = \left(1+\frac{R_1}{R_2}\right)V_{GSN} - V_{GSP}, \qquad (3.19)$$

thus adjusting the value of the two resistors a voltage equal to the difference between the two threshold voltages is obtained. However this method doesn't allow obtaining a compensation of the temperature dependence of the mobility.

Despite the total compatibility with the CMOS the solutions [9]-[10] are both too power hungry (>1μW) for the typical low-power applications.

A fully-CMOS solution based on transistors working in above threshold regime was proposed in [11]. Inspecting Figure 3.6 in this solution transistors M_1, M_2, M_3 and M_4 form a closed loop in which $V_{GS1}+V_{GS3}=V_{GS2}+V_{GS4}$. Thus I_B can be expressed as:

$$nV_T \ln\left(\frac{K_2}{K_1}\right) = \sqrt{\frac{2I_B}{\beta K_4}} - \sqrt{\frac{2I_B}{\beta K_3}}. \qquad (3.20)$$

From (3.20) the current I_B can be written as:

$$I_B = \frac{\beta K_4}{2} n^2 V_T^2 \ln^2\left(\frac{K_2}{K_1}\right)\left(\frac{\sqrt{K_3}}{\sqrt{K_3}-\sqrt{K_4}}\right)^2. \qquad (3.21)$$

Transistors $M_{5,}$ M_6, M_7 and M_8 receive the current I_B and generate the output voltage. Specifically these transistors are designed with the aim to observe the majority of the I_B current in M_7 and M_8 rather than through M_5 and M_6 in order to compensate for the temperature dependence of the mobility. If the latter assumption is satisfied, the output voltage can be expressed as:

$$V_{OUT} = V_{GS8} + V_{GS5} - V_{GS7} = V_{TH} + nV_T \ln\left(\frac{K_2}{K_1}\right)\frac{\sqrt{K_3K_4}}{\sqrt{K_3}-\sqrt{K_4}}\left(\frac{1}{\sqrt{K_8}}\left(1+\frac{\sqrt{K_6}}{\sqrt{K_5}}\right)-\frac{1}{\sqrt{K_7}}\right). \qquad (3.22)$$

Because V_{TH} in (3.22) has a negative TC and V_T has a positive TC, the output voltage V_{REF} can be compensated in temperature. As reported in [11], a measured TC of 12 ppm/°C and a power dissipation of 0.12 μW have been obtained. The minimum supply voltage is 1.5 V.

3.3.2. Subthreshold voltage references

To obtain a large reduction in terms of power consumption and minimum operating voltage, subthreshold voltage references have been proposed. In Figure 3.7 the voltage reference proposed in [12] is reported. Here M_1 and M_3 are two high threshold voltage transistors operating in subthreshold regime.

Since $V_{GS1}=V_{GS2}$ and $V_{GS3}=V_{GS4}$, the following equations hold:

$$V_{TH,H} + nV_T \ln\left(\frac{I_1}{K_1 I_0}\right) = V_{TH,L} + \sqrt{\frac{2I_2}{K_2\beta}}, \qquad (3.23)$$

$$V_{TH,H} + nV_T \ln\left(\frac{I_1}{K_3 I_0}\right) = V_{TH,L} + \sqrt{\frac{2I_2}{K_4\beta}}. \qquad (3.24)$$

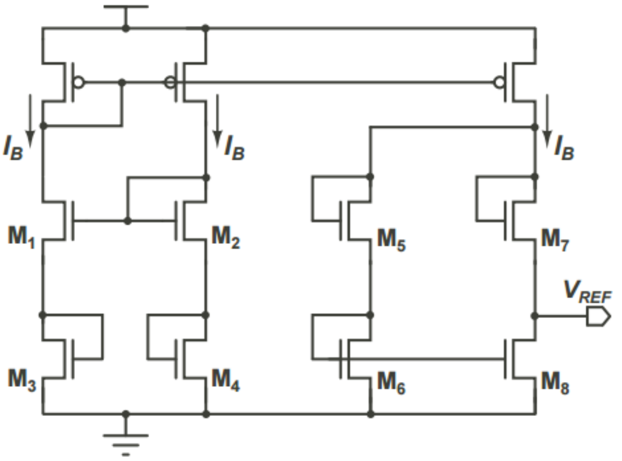


Figure 3.6. Voltage reference proposed in [11].

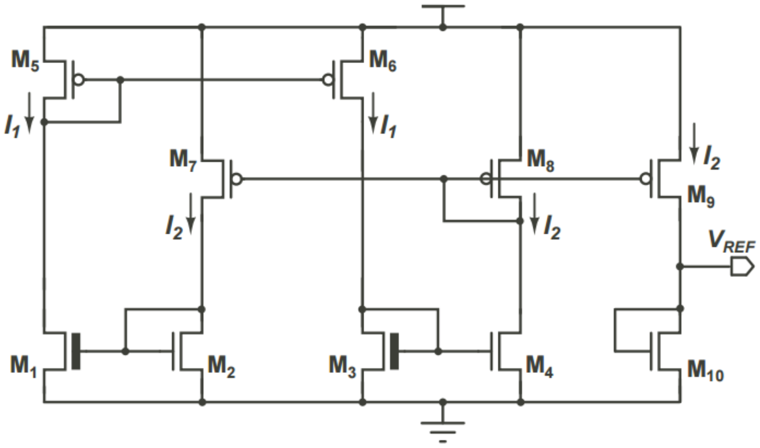


Figure 3.7. Subthreshold CMOS voltage reference proposed in [12].

In (3.23) and (3.24) $I_0=\mu_N C_{OX} V_T^2$. Using (3.23)-(3.24) the current I_2 can be expressed as:

$$I_2 = \frac{K_4}{2\left(\sqrt{K_4/K_2}\right)^2} n^2 V_T^2 \ln^2\left(K_3/K_4\right). \tag{3.25}$$

This current is mirrored into the diode connected transistor M_{10}, generating the reference voltage:

$$V_{REF} = V_{TH} + nV_T \ln\left(\frac{K_3}{K_1}\right)\frac{\sqrt{K_4/K_{10}}}{\sqrt{K_4/K_2}-1}. \tag{3.26}$$

Also in this solution the temperature behaviour can be compensated by cancelling the CTAT behaviour of V_{TH} with the PTAT term V_T. From experimental results the solution proposed in [12] is in able to achieve a TC of only 10 ppm/°C with a power consumption of 0.036 μW and a minimum operating voltage of 0.9 V.

3.3.3. Fully-subthreshold CMOS references

Voltage references with all transistors working in subthreshold regime are reported in [16]-[17].
The solution proposed in [16] uses an ultra-low power self-biased current reference to generate a bias voltage V_b. Since M$_2$ and M$_4$ work in subthreshold regime the drain current of these two transistors are equal to:

$$I_2 = K_2 I_0 \exp\left(\frac{V_{GS1} - V_{GS4} - V_{TH}}{nV_T}\right), \tag{3.27}$$

$$I_4 = K_4 I_0 \exp\left(\frac{V_{GS4} - V_{TH}}{nV_T}\right). \tag{3.28}$$

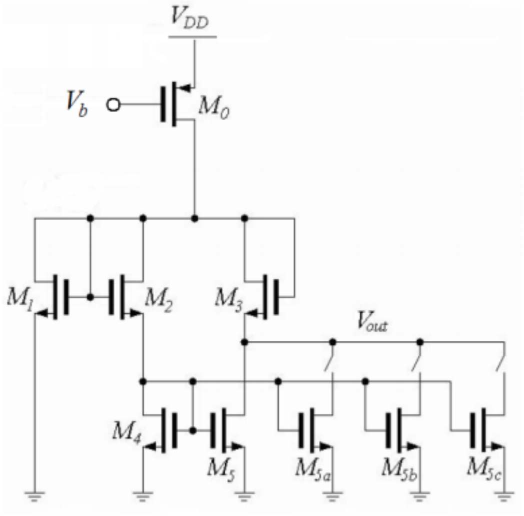


Figure 3.8. Fully-subthreshold voltage reference proposed in [16].

From Figure 3.8 I_2=I_4, thus V_{GS4} becomes equal to:

$$V_{GS4} = \frac{V_{GS1}}{2} + nV_T\left(\sqrt{\frac{(W/L)_2}{(W/L)_4}}\right), \tag{3.29}$$

since V_{OUT}=V_{GS4}+V_{GS2}-V_{GS3} the following condition holds:

$$V_{OUT} = \frac{V_{TH}}{2} + nV_T\left(\frac{(W/L)_3}{(W/L)_5}\sqrt{\frac{(W/L)_2}{(W/L)_4}}\right). \tag{3.30}$$

Choosing an opportune ratio for the term in curly brackets in (3.30), the temperature dependence of the threshold voltage can be compensated. From experimental results the solution proposed in [16] starts to work from 0.6 V consuming < 40 nW.
[17] proposed a state-of-the-art solution in terms of low-power, low-voltage operation. The schematic of the proposed solution is reported in Figure 3.9. The circuit consists of a start-up circuit, a current reference and a diode-connected nMOSFET. The transistor M$_2$ is a high-threshold voltage transistor while all the other transistors are regular threshold voltage. The current I_1 obtained in the current reference is equal to:

$$I_1 = \mu_N \frac{K_1 K_3}{K_2} V_T^2 \exp\left(-\frac{\Delta V_{TH}}{n V_T} \right), \tag{3.31}$$

where $\Delta V_{TH}=V_{TH1}+V_{TH3}-V_{TH2}$. This current is injected into the diode-connected nMOS transistor M_{10}, generating a voltage reference equal to:

$$V_{REF} = V_{TH10} - V_{TH1} + V_{TH3} - V_{TH2} + nV_T \ln\left(\frac{K_1 K_3}{K_2 K_{10}} \right), \tag{3.32}$$

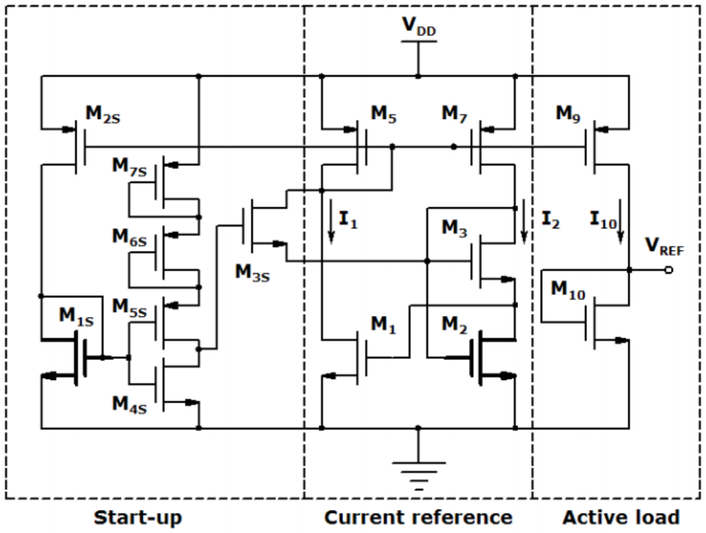


Figure 3.9. Fully-subthreshold voltage reference proposed in [17].

thus choosing the right value of K_1, K_2 , K_3 and K_{10}, a voltage reference roughly equal to the difference between the high threshold voltage transistor and the regular threshold voltage transistor is obtained. The solution reported in [17] is in able to work starting from only 0.45 V consuming only 2.6 nW.

3.4. Proposed Voltage Reference 1

3.4.1. Introduction

In this section a new low-voltage, low-power subthreshold voltage reference is proposed. The voltage reference consists of only two nMOS transistors with different threshold voltages. Measurements performed on 23 samples from a single batch show a mean reference voltage of 275.4 mV. The subthreshold conduction and the low number of transistors enable to achieve a mean power consumption of only 40 pW. The minimum supply voltage is 0.45 V, which coincides with the lowest value reported so far [17]. The mean TC in the temperature range from 0 to 120 °C is 105.4 ppm/°C while the mean line sensitivity is 0.46%/V in the supply voltage range 0.45-1.8 V. The occupied area is 0.018 mm^2. The PSRR without any filtering capacitor is -48 dB at 20 Hz and -29.2 dB at 10 kHz. Thanks to large area transistors and to a careful layout, the coefficient of variation of the reference voltage is only 0.62%. A new figure of merit, the voltage temperature parameter (*VTP*) which gives a direct measure of the overall percentage variation of the reference voltage on the typical 2D domain of supply voltage and temperature, is introduced. For the proposed circuit, the average *VTP* is 1.70% with a standard deviation of 0.21%. In order to investigate the effect of transistor area on process variability, a 4× replica of the proposed configuration has been fabricated and tested as well. Except for LS, the 4× replica doesn’t exhibit any appreciable improvement with respect to the basic voltage reference.

3.4.2. Operating principle

The schematic of the proposed configuration is reported in Figure 3.10. The circuit consists of only two nMOS transistors, a high-threshold-voltage transistor M$_\mathrm{H}$ ($V_{TH,H}$=600mV) and a low-threshold-voltage transistor M$_\mathrm{L}$ ($V_{TH,L}$=320mV).

The starting point of the analysis is the *I-V* relationship for an nMOS transistor working in subthreshold regime. Assuming the condition of $V_{DS}>4V_T$, inverting the expression of the subthreshold current reported in (2.10) the following equation provides the gate-source voltage of M$_\mathrm{H}$:

$$V_{GS} = V_{TH,H} + n_H V_T \ln\left(\frac{I_H}{I_{0,H} K_H V_T^2}\right). \tag{3.33}$$

$$I_{0H} = \mu_N C_{OX} V_T^2, K_H = \left(\frac{W}{L}\right)_H . \tag{3.34}$$

In (3.33) n_H and I_H are the subthreshold slope and the drain current flowing in M$_\mathrm{H}$, respectively. Considering also for M$_\mathrm{L}$ the assumption of $V_{DS}>4V_T$ the current injected by M$_\mathrm{L}$ into M$_\mathrm{H}$ is expressed as:

$$I_L = I_{0,L} K_L \exp\left(-\frac{V_{TH,L}}{n_L V_T}\right), \tag{3.35}$$

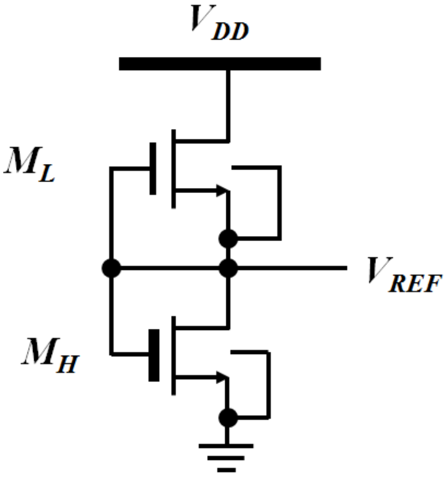


Figure 3.10. Schematic of the proposed voltage reference.

Since the reference voltage is coincident with the gate-source voltage of M$_H$ and I_H=I_L using (3.33) the following equation holds:

$$V_{REF} = V_{TH,H} + n_H V_T \ln\left(\frac{I_L}{I_{0,H} K_H V_T^2}\right). \tag{3.36}$$

Thus, replacing in (3.36) the (3.35), V_{REF} becomes equal to:

$$V_{REF} = V_{TH,H} - \frac{n_H}{n_L} V_{TH,L} + n_H V_T \ln\left(\frac{C_{ox,L}\left(\frac{W}{L}\right)_L}{C_{ox,H}\left(\frac{W}{L}\right)_H}\right). \tag{3.37}$$

By considering as a first approximation the subthreshold swing factors almost constant with temperature, the only temperature-dependent parameters in (3.37) are the two threshold voltages and V_T. Since the threshold voltage decreases with temperature and the thermal voltage increases with temperature, a compensation of these two terms can be performed by choosing a proper transistor size ratio. Therefore, expressing the temperature dependence of the threshold voltage to $V_{TH,H}$ and $V_{TH,L}$ and expressing the thermal voltage as kT/q the (3.37) can be rewritten as:

$$V_{REF} = V_{TH,H}(T_0) - |\kappa_H|(T - T_0) - \frac{n_H}{n_L}\left(V_{TH,L}(T_0) - |\kappa_L|(T - T_0)\right) + n_H V_T \ln\left(\frac{C_{ox,L}}{C_{ox,L}} K_R\right), \tag{3.38}$$

where K_R=K_L/K_H and κ is the temperature coefficient of the threshold voltage. The temperature variation of V_{REF} is then obtained by differentiating (3.38) with respect to temperature:

$$\frac{\partial}{\partial T} V_{REF} = -|\kappa_H| + \frac{n_H}{n_L}|\kappa_L| + n_H \frac{k}{q} \ln\left(\frac{C_{ox,L}}{C_{ox,H}} \left(\frac{W}{L} \right)_R \right). \qquad (3.39)$$

To obtain the condition of temperature compensated voltage reference the transistor size ratio which sets the (3.39) to 0 is considered:

$$\frac{\left(\frac{W}{L} \right)_L}{\left(\frac{W}{L} \right)_H} = \frac{C_{ox,H}}{C_{ox,L}} \exp\left[\frac{q}{n_H k} \left(|\kappa_H| - \frac{n_H}{n_L} |\kappa_L| \right) \right]. \qquad (3.40)$$

The temperature-compensated reference voltage is then evaluated by replacing the (3.40) into (3.38):

$$V_{REF} = V_{TH,H}(T_0) + |\kappa_H| T_0 - \frac{n_H}{n_L}\left(V_{TH,L}(T_0) + |\kappa_L| T_0 \right). \qquad (3.41)$$

A more simplified expression of the reference voltage is obtained by considering $n_H \approx n_L$ and $\kappa_H \approx \kappa_L$. Under these assumptions the reference voltage of the proposed configuration is expressed as

$$V_{REF} \approx \Delta V_{TH} = V_{TH,H}(T_0) - V_{TH,L}(T_0). \qquad (3.42)$$

Thus the reference voltage of the proposed configuration is approximated by the difference between the two threshold voltages at room temperature. Despite its simplicity, expression (3.42) gives a very accurate estimation of the reference voltage as confirmed by measurement results. The error between the measured mean value of V_{REF} and the value predicted from (3.42) by using the nominal values of $V_{TH,H}$ and $V_{TH,L}$ is only 1.67%.

3.4.3. Design considerations

3.4.3.1. Subthreshold conduction and threshold voltage constraint

In the proposed configuration the low-threshold voltage transistor has the gate, the source and the body terminals shorted. As a consequence, since the threshold voltage drops in temperature, the subthreshold conduction for this transistor is ensured if its threshold voltage is higher than 0 V at the maximum operating temperature. On the other hand, the reference voltage is coincident with the gate-source voltage of a diode-connected high-threshold-voltage transistor. Thus, to ensure subthreshod conduction for it, the output voltage has to be lower than its threshold voltage. Considering (3.41) the condition of $V_{TH,H}$>V_{REF} leads to

$$V_{TH,L}(T_0) > \frac{n_L}{n_H} |\kappa_H| T_{MAX} - |\kappa_L| T_0, \qquad (3.43)$$

where T_{MAX} represents the maximum operating temperature. At the same time, the condition of V_{REF}>4V_T has to be ensured in order to neglect, without significant loss of accuracy, the effect of V_{DS} on drain current. Thus from (3.34), imposing V_{REF}>4$V_{T,max}$ with the condition reported in (3.36), we obtain:

$$V_{TH,H}(T_0) > \frac{4kT_{\max}}{q} + |\kappa_H|(T_{\max} - T_0). \tag{3.44}$$

Thus if the conditions (3.43) and (3.44) are satisfied both transistors work in subthreshold regime.

3.4.3.2. Minimum supply voltage

In the proposed solution the supply voltage V_{DD} can be expressed as the sum of transistor drain-source voltages:

$$V_{DD} = V_{REF} + V_{DS,L}. \tag{3.45}$$

The current injected by M_L into M_H becomes almost independent from the drain voltage for $V_{DS,L}$>4$V_{T,max}$. This condition set the minimum supply voltage for the proper operation of the proposed solution:

$$V_{DD,MIN} > V_{REF} + 4V_{T,MAX}. \tag{3.46}$$

Considering the maximum operating temperature of 120 °C the expected minimum supply voltage is about 135.5 mV larger than the reference voltage. By using (3.46) is it possible to evaluate also the lowest value of $V_{DD,MIN}$ for the proposed solution. This value is equal to about 8$V_{T,MAX}$ (≈300 mV), since the condition of V_{REF}>4V_T has to be respected for subthreshold operation.

3.4.3.3. Supply voltage variations

V_{DS} can affect the drain current also because of drain-induced barrier lowering (*DIBL*) effect. As a result, the current injected by M_L into M_H depends on V_{DD} also if $V_{DS,L}$ is greater than 4V_T. Neglecting *DIBL* effect on the load transistor, the variation ΔV_{REF} on reference voltage caused by a variation ΔV_{DD} on supply voltage is evaluated by considering the *DIBL* effect on $V_{TH,L}$:

$$\Delta V_{TH,L} = \lambda_D \Delta V_{DS,L}. \tag{3.47}$$

In (3.47) λ_D is the DIBL coefficient of the low-threshold voltage transistor and $\Delta V_{DS,L}$ the variation of the drain-source voltage of M_L which is equal to $\Delta V_{DS,L}$=ΔV_{DD}-ΔV_{REF}. From (3.41) and (3.47) a variation $\Delta V_{TH,L}$ causes a variation on reference voltage ΔV_{REF} equal to

$$\Delta V_{REF} = \frac{n_H}{n_L} \Delta V_{TH,L}(T_0) = \frac{\frac{n_H}{n_L}\lambda_D}{1 + \frac{n_H}{n_L}\lambda_D} \Delta V_{DD}, \tag{3.48}$$

which leads to a line sensitivity, expressed as %/V, equal to

$$LS\left[\frac{\%}{V}\right]\approx 100\frac{\frac{n_H}{n_L}\lambda_D}{V_{REF}\left(1+\frac{n_H}{n_L}\lambda_D\right)}\approx 100\frac{\lambda_D}{V_{REF}\left(1+\lambda_D\right)}. \tag{3.49}$$

For LS minimization, the M_L transistor must be as long as possible since the *DIBL* increases exponentially as the channel length decreases [17].

3.4.3.4. Voltage temperature parameter

The two fundamental figures of merit of a voltage reference are LS and TC. There are two possible drawbacks in the use of these two parameters. The first drawback is that although LS (TC) depends on the temperature (supply voltage), LS (TC) is typically evaluated at a single temperature (supply voltage), thus giving only a very limited information on the overall variations of the reference voltage with supply voltage and temperature. The second drawback is that the reference voltage does not depend linearly on supply voltage and temperature, therefore the values of LS and TC measured in a given range does not allow evaluating the LS and TC in a smaller or larger range. As an example, the voltage reference typically exhibits a higher sensitivity to the supply voltage close to the minimum voltage and a lower sensitivity at higher voltages, therefore apparently better LS can be simply obtained by increasing the maximum supply voltage under consideration. In order to overcome these drawbacks, a new figure of merit the voltage temperature parameter (*VTP*) is introduced. It is defined as:

$$VTP[\%]=100\times\frac{V_{REF,MAX}-V_{REF,MIN}}{V_{REF,MEAN}}, \tag{3.50}$$

where $V_{REF,MAX}$, $V_{REF,MIN}$ and $V_{REF,MEAN}$ are the maximum, the minimum and the mean value of the reference voltage in a given 2D supply voltage-temperature domain. In order to compare *VTP*s obtained in the cases of different voltage references, a common range of operating temperature and V_{DD} can be assumed. Nevertheless it is obvious that the definition of this FOM can be performed in any V_{DD} and temperature range, thus making very easy the comparison of different solutions operating in different domains. A possible solution for the latter condition consists for example in choosing an equal range of V_{DD}s and temperatures variation or in an equal magnitude of the ΔV_{DD} and ΔT range.

For the evaluation of the *VTP* on the proposed solution a 2D domain obtained by varying the supply voltage in the range [$V_{DD,MIN}\div V_{DD,MIN}$+1V] and the temperature range from 0 to 100 °C is considered.

3.4.3.5. Process variations

As reported in Chapter 2, process variations are particularly important in subthreshold design as a consequence of the exponential sensitivity of the drain current to the threshold voltage variations induced by the random dopant fluctuations and oxide thickness variation. The effect of the process variations on CMOS process is described by (2.2) [18]. Since the two transistors need different masks, the standard deviations of the threshold voltages are equal to [18]:

$$\sigma(V_{TH,H}) = \frac{A_H}{\sqrt{W_H L_H}}, \sigma(V_{TH,L}) = \frac{A_L}{\sqrt{W_L L_L}}. \tag{3.51}$$

As reported before in the proposed voltage reference V_{REF} is approximated by the difference between the two threshold voltages. For this reason, according to [18], to obtain good process stability large area (*W/L*) for both transistors is necessary. However, there are additional sources of variations which are not considered in (3.51), such as the effects of geometry and doping on a larger length scale of temperature and mechanical stress. Such effects are alleviated by using common centroid layout techniques [20].

3.4.3.6. Transistor sizing

The transistor sizing for the proposed configuration has been performed by taking into account three constraints: LS, TC and process stability. As reported in the previous sections:

- ✓ LS decreases as the channel length of M_L is increased;
- ✓ The effect of process variability decreases by increasing the area of both transistors;
- ✓ The reference voltage is temperature-compensated if the transistor size ratio is chosen according to (3.40).

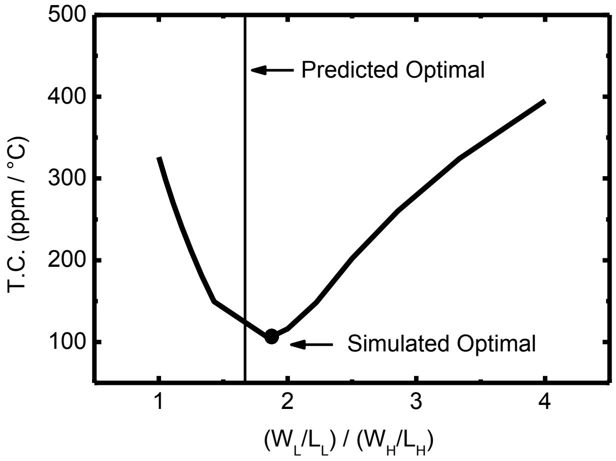


Figure 3.11. Simulated optimal TC as a function of transistor size ratio obtained by keeping L_L=*50μm*, W_L =*100μm*, L_H=*50μm* and by varying W_H. The simulated optimal transistor size ratio is equal to 1.85 instead of 1.67 predicted by (3.40).

Since the maximum *W* and *L* for the selected design kit are 100 µm and 50 µm, respectively, the following conditions can be imposed to satisfy the previous requirements:

- ✓ L_L=50 µm, in order to obtain the minimum value of LS;
- ✓ W_L=100 µm and L_H=50 µm to guarantee good process stability;
- ✓ W_H=54 µm in order to achieve the optimal transistor size ratio for temperature compensation.

In Figure 3.11 we report the simulated *TC* as a function of transistor size ratio. The simulation is performed by keeping L_L=L_H=50 µm and W_L =100 µm and by varying W_H. The simulated optimal value is 1.85 instead of a predicted optimal transistor size ratio of 1.67.

3.4.3.7. Process stability improvement

Since robustness to process variability improves as the transistor area increases [20], the stability can be improved by replicating the basic configuration an appropriate amount of times. The number of replicas is evaluated in order to find a proper trade-off among occupied area, power consumption and improvement in process variability. For a proper choice, first, the theoretical improvement provided by the *N* × replica is evaluated. Using (3.51) and (3.42) the standard deviation of V_{REF} can expressed as:

$$\sigma(V_{REF}) = \sqrt{\sigma^2\left(V_{TH,H}\right) + \sigma^2\left(V_{TH,L}\right) - 2\rho\sigma\left(V_{TH,H}\right)\sigma\left(V_{TH,L}\right)}, \tag{3.52}$$

where σ($V_{TH,H}$) and σ($V_{TH,L}$) are the standard deviations for V_{REF}, $V_{TH,H}$ and $V_{TH,L}$, respectively, while ρ represents the correlation factor between σ($V_{TH,H}$) and σ($V_{TH,L}$). Thus assuming ρ=0, an *N* times replicated version of the circuit decreases the standard deviation of the reference voltage by a factor equal to $\sqrt{N}$ which means that the most significant improvement in process stability without a high penalty in terms of area is obtained with a low number of replicas. In particular a number of replica equal to *N*=4 is chosen since the improvement in process stability given by a larger number of replicas is not commensurate to the penalty in occupation area and power consumption. To investigate experimentally the effect of the transistor area on process variability the basic and the 4 × replicated version, with a common centroid layout, have been fabricated and tested as well.

3.4.4. Experimental Results

The layout and the transistor sizing of the proposed voltage reference and its 4 × replica are reported in Figure 3.12 and Table 3.I, respectively

The active area is only 0.018 mm^2 for the basic voltage reference while an area of 0.056 mm^2 is occupied by its 4 × replica. The on-wafer electrical measurements have been performed on 23 samples containing the basic voltage reference and its 4 × replica by using a probe station SUMMIT 11861B Cascade with Temptronic thermal controller and a Keithley 4200-SCS parameter analyzer.

Figure 3.13 reports the measured output voltage against V_{DD} for the typical prototype of the basic solution. The circuit starts to work properly from V_{DD}=0.45 V, which is only 34 mV higher than the value predicted from (3.46) by considering the maximum operating temperature of 120 °C. The output voltage at room temperature for the minimum supply voltage is 275.4 mV. The mean LS over the supply voltage range of [0.45 V – 1.8 V] is 0.46 %/V corresponding to a variation of about 1.7 mV over a variation of 1.35 V in the supply voltage. It is worth noting that the measured value of LS is only 0.03 %/V higher than the value predicted by (3.49).

In Figure 3.14 the temperature dependence of the reference voltage is reported for different supply voltages.

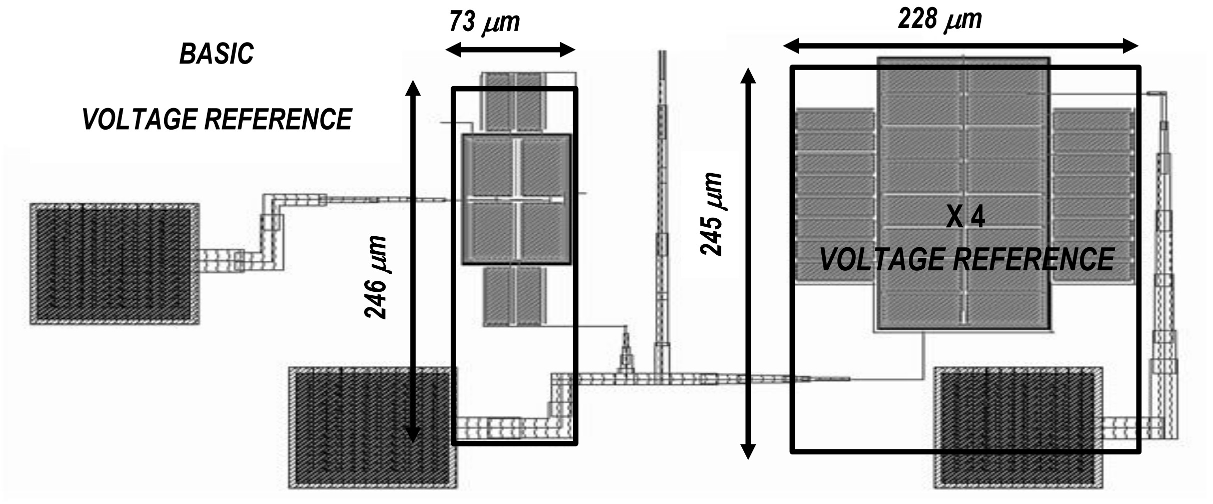


Figure 3.12. Layout of the basic voltage reference and its 4 × replica.

TABLE 3.I
TRANSISTOR SIZING FOR THE BASIC VOLTAGE REFERENCE AND ITS 4X REPLICA

Circuit	Transistor	*W/L*
Basic Voltage Reference	M_L	100µm / 50µm =(25 µm/50 µm) **x 4**
	M_H	54µm / 50µm =(13.5 µm/50 µm) **x 4**
4× replica	M_L	400µm / 50µm =(25 µm/50 µm) **x 16**
	M_H	216µm / 50µm =(13.5 µm/50 µm) **x 16**

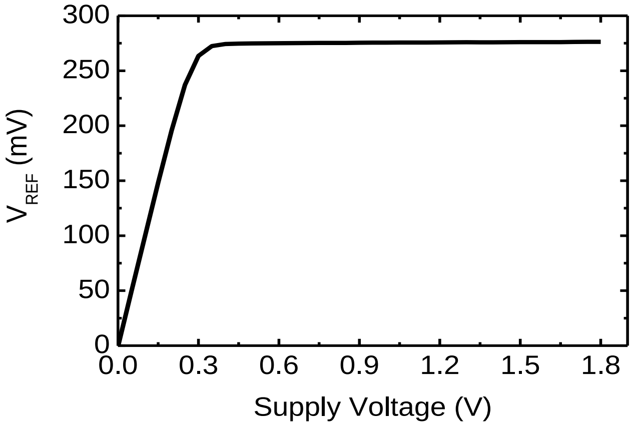


Figure 3.13. Output voltage against supply voltage variation (basic solution, typical chip).

Inspecting Figure 3.14 the temperature dependence is almost the same in the whole range of V_{DD}s. The reference voltage monotonously increases by decreasing the temperature and by increasing the supply voltage. The measured mean TC in the temperature range from 0 to 120 °C, averaged over the supply voltage range from 0.45 to 1.8 V, is 105.4 ppm/°C. Both supply voltage and temperature effect on V_{REF} can be depicted by inspecting Figure 3.15. From this figure becomes clear that the generate voltage is pretty stable in the whole 2D domain of operation.

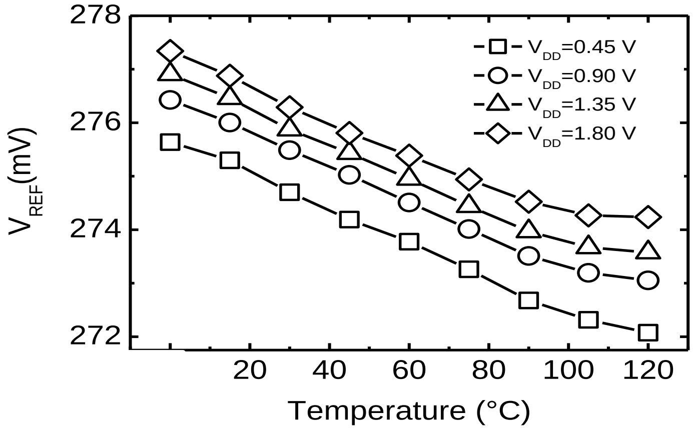


Figure 3.14. Temperature dependence of V_{REF} for different supply voltages (basic solution, typical chip).

The average *VTP* evaluated in the supply voltage range from 0.45 to 1.45 V and in the temperature range from 0 to 100 °C is 1.70 %.

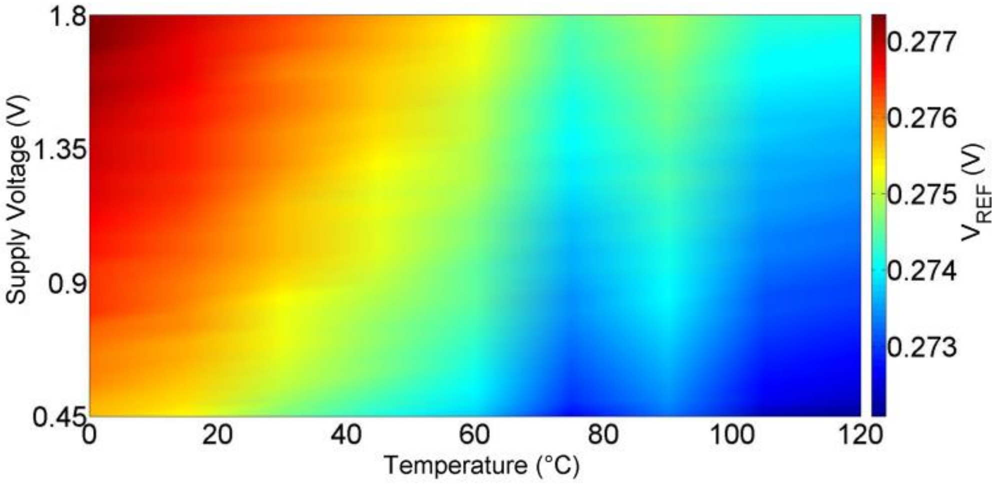


Figure 3.15. Temperature and Supply voltage dependence of V_{REF} (basic solution, typical chip).

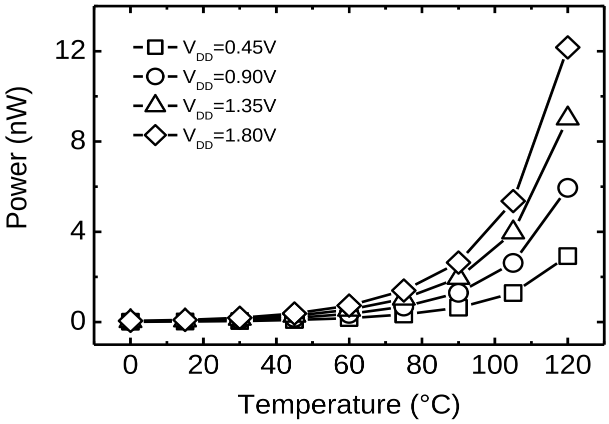


Figure 3.16. Power consumption against temperature for different supply voltages (basic solution, typical chip).

The power consumption at different temperatures and supply voltages is reported in Figure 3.16. At room temperature the basic solution consumes 36.7 pW for V_{DD}=0.45 V and 154 pW for V_{DD}=1.8

V. At the maximum operating temperature of 120°C the power consumption is 2.9 nW for V_{DD}=0.45 V and 12.2 nW for V_{DD}=1.8 V. Figure 3.17 reports the measured power supply rejection rate (*PSRR*) at room temperature for V_{DD} =0.45 V at different frequencies. The *PSRR* without any filtering capacitor is -48 dB at 20 Hz and -29.2 dB at 10 kHz. All previous data are consistently obtained also for the 4 × replica.

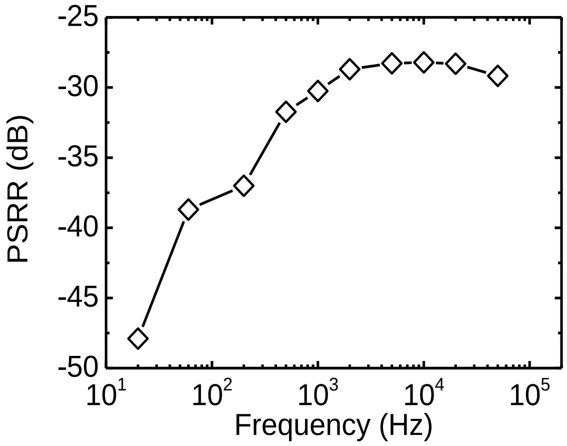


Figure 3.17. *PSRR* over frequency (basic solution, typical chip).

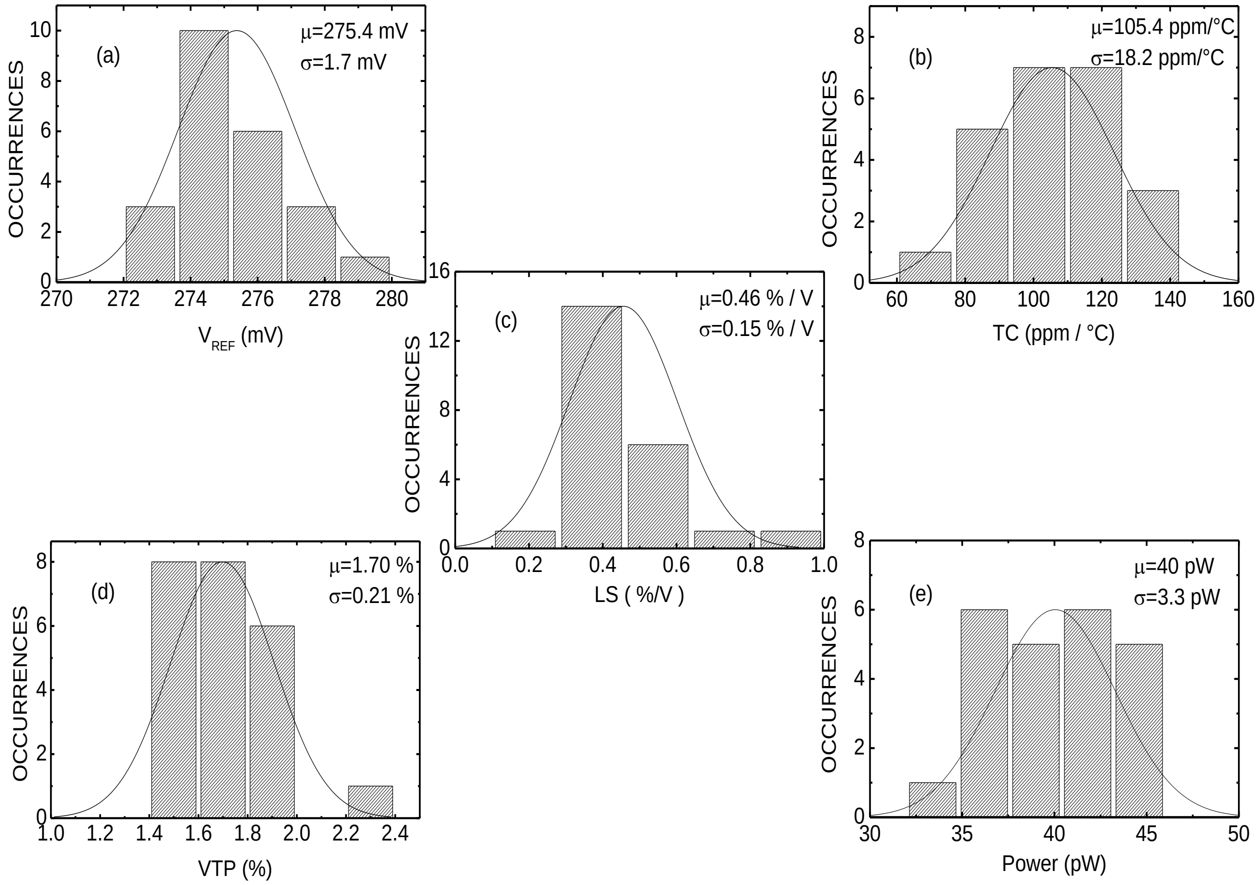


Figure 3.18. Statistical analysis over 23 samples of the basic voltage reference: (a) reference voltage at 25°C and V_{DD}=0.45V, (b) TC, (c) LS at 25°C, (d) *VTP*, (e) power consumption at 25°C and V_{DD}=0.45V (e).

The statistical analysis over 23 samples of the proposed circuits is summarized in Figure 3.18(a)-(e). The measured mean voltage reference at room temperature for the minimum supply voltage is

274.5 mV with a standard deviation of only 1.7 mV (Figure 3.18(a)) while the mean TC is 105.4 ppm/°C with a standard deviation of 18.2 ppm/°C (Figure 3.18(b)). The mean LS is 0.46 %/V with a standard deviation of 0.15 %/V (Figure 3.18(c)). The measured average *VTP* evaluated in the supply voltage range from 0.45 to 1.45 V and in the temperature range from 0 to 100 °C is 1.70 % with a standard deviation of 0.21% (Figure 3.18(d)). Across the 23 samples the mean power consumption at 25°C and for V_{DD}=0.45 V is 40 pW with a standard deviation of 3.3 pW (Figure 3.18(e)).

Table 3.II compares the mean value and the dispersion of the main figures of merit for the basic voltage reference and its 4 × replica. The only performance benefiting from the larger occupied area is the LS. The mean value of LS is 0.05 %/V lower in the X 4 version with respect to basic version and its standard deviation is 3 times smaller. The basic voltage reference is clearly more efficient in terms of power consumption, which is roughly four times smaller than in the case of the 4 × replica. All the other parameters are practically coincident. In particular, no improvement in the standard deviation of the reference voltage predicted by the $1/\sqrt{W \cdot L}$ law is observed for the 4 × version. Thus the basic voltage reference seems sufficiently large that a further enlargement does not provide a significant improvement against process variability.

Table 3.II
Comparison Between The Basic Voltage Reference And Its 4X Replica

	Basic Voltage Reference	**4 X Voltage Reference**
$V_{DD,MIN}$ [V]	0.45	0.45
Area [mm^2]	0.018	0.056
V_{REF} [mV]	σ=275.4	σ=275.0
	μ=1.7	μ=1.8
	σ/μ=0.62%	σ/μ=0.65%
TVC[%]	σ=1.70	σ=1.71
	μ=0.21	μ=0.22
	σ/μ=12.3%	σ/μ=12.9%
TC [ppm/°C]	σ=105.4	σ=115.7
	μ=18.2	μ=18.8
	σ/μ=17.3%	σ/μ=16.2%
LS [%/V]	σ=0.46	σ=0.39
	μ=0.15	μ=0.05
	σ/μ=32.6%	σ/μ=12.8%
Power [pW]	σ=40	σ=139.3
	μ=3.3	μ=6.9
	σ/μ=8.3%	σ/μ=4.9%
PSRR [dB]	-48 @ 20 Hz	-48.4 @ 20 Hz
	-29.2 @ 10 kHz	-23.7 @ 10 kHz

In Table 3.III the basic voltage reference is compared with the other low-power, low-voltage CMOS voltage references reported in literature. The minimum supply voltage is equal to [17], which represents the best result proposed so far. The line sensitivity is similar to [17] while the temperature coefficient is comparable to [16] and [17].

Except [16] the proposed solution has the lowest occupied area. Table 3.III reports also the comparison of the statistical analysis where data are available.

The coefficient of dispersion of the reference voltage in a single batch is lower with respect to the other CMOS voltage references [12]-[17] and even better than bipolar BGRs [5]-[6]. The observed robustness against process variation is ascribed to the large transistor areas and to a careful layout. A complete comparison of the dispersion of the other parameters can be performed only with [17] since it is the only work in literature which reports this information.

TABLE 3.III
COMPARISON WITH LOW-VOLTAGE LOW-POWER CMOS VOLTAGE REFERENCES

		This work *Ref. [21]*	*Ref. [17]*	*Ref. [16]*	*Ref. [15]*	*Ref. [12]*	*Ref. [13]*
Technology		0.18μm	0.18μm	0.18μm	0.18μm	0.35μm	0.35μm
Supply voltage (V)		0.45 to 1.8	0.45 to 2	0.6 to 2.3	0.85 to 2.5	0.9 to 4	1.1 to 4
Power @ room temperature		40pW@0.45V 0.18nW@1.8V	3.15nW@0.45V 14.4nW@1.8V	<40nW@0.7V –	3.3μW@0.85V Average	36nW @0.9V 220nW@4V	22nW@1.1V 88nW@4V
V_{REF} (mV)		275.4	263.5	~220	221	670	96.6vers-1 108.9vers-2
TC (ppm/°C) T range(°C)		105.4 [0:120]	142.1 [0:100]	127 [-20:100]	194 [-20:120]	10 [0:80]	11.4 – [-20:80], vers-1 9.2 – [-20:80], vers-2
Line Sensitivity (%/V)		0.46	0.44	~2.73	0.905	0.27	0.09 , vers-1 0.17 , vers-2
PSRR (dB) Low freq [≤100Hz] High freq [≥10MHz]		V_{DD}=0.45V -48 -29.2@10kHz	V_{DD}=0.45V -49.4 (-12.2 sim.)	- -41 -	- - -	V_{DD}=0.9V -47 -40	V_{DD}=3V <-60 <-40
σ/μ	***V_{REF}***	0.62%	3.9%	-	-	3.1%	4.1% vers.1 2.8% vers.2
	TC	17.3%	60.6%	-	-	-	-
	LS	32.6%	13.1%	-	-	-	-
	Power	8.3%	26.9%	-	-	-	-
Die area (mm^2)		0.018	0.043	0.004	0.0238	0.045	0.0189 , vers-1 0.0193 , vers-2

3.4.5. Conclusion

A temperature-compensated subthreshold CMOS voltage reference has been presented. The proposed solution, fabricated in UMC 0.18 μm triple-well CMOS technology, consists of only two nMOS transistors with different threshold voltages. A statistical experimental analysis of the proposed configuration has been performed over a set of 23 samples.

The mean reference value is 275.4 mV, which approximately corresponds to the difference between the two threshold voltages. The minimum supply voltage is 0.45 V, which coincides with the lowest value reported so far. The mean TC in the temperature range from 0 to 120 °C is 105.4 ppm/°C while the mean LS is 0.46 %/V in the supply voltage range from 0.45 to 1.8 V. The active area is 0.018 mm^2. Thanks to the large area transistors and to a careful layout, the coefficient of variation of the reference voltage is only 0.62 %, which is the lowest value among other CMOS voltage references reported in the literature. A new figure of merit the voltage temperature parameter (*VTP*), which gives a direct measure of the overall percentage variation of the reference voltage on the typical 2D domain of operating supply voltage and temperature, has been introduced. For the proposed circuit, the mean *VTP* in the supply voltage range from 0.45 to 1.45 V and in the temperature range from 0 to 100 °C is equal to 1.70 % with a standard deviation of 0.21 %.

In order to investigate the effect of transistor area on robustness to process variability, a 4 × replica of the proposed configuration has been fabricated and tested as well. The 4 × replica exhibits an appreciable improvement only on the mean value and the standard deviation of LS at the cost of the obvious penalty in the area occupation and power consumption.

3.5. Proposed Voltage Reference 2

3.5.1. Introduction

Reducing the minimum operating voltage is at the forefront of digital circuit research. Several works have reported the implementation of digital circuits operating with supply voltage lower than 200 mV [1],[22]-[24]. Indeed, as explained in Chapter 1, operation at the minimum supply voltage represents the most effective way to reduce power in a digital circuit. This concept cannot be extended tout-court to analog circuits where scaling the supply voltage does not ensure a reduction of power consumption [25]. However, it is worth noting that digital and analog circuits are usually used together in mixed-signal integrated circuits. Apparently, analog blocks represent a bottleneck for supply voltage scaling of such systems. In addition, in some emerging battery-free applications the minimum operating voltage is even a more important target than power consumption, as in the case of systems powered by energy harvesting devices (i.e. thermoelectric generators) or fuel cells [24]. In these applications, the energy harvested from machines or body heat can be considered an unlimited power supply [26]-[27], but only a limited voltage is available for the different building blocks.

In the context of power and supply voltage scaling, subthreshold operation has been intensely pursued by digital and analog designers. The fundamental limit to supply voltage scaling in circuits is represented by the voltage necessary to ensure a proper operation for all the employed transistors [28]. Since the bias required in the subthreshold regime is considerably less than in other operating conditions, it follows that this regime is the most promising for supply voltage scaling. In addition, the low current of subthreshold operation ensures a significant reduction in power consumption. However, as explained before, despite the apparent energy benefits, circuits biased in subthreshold pose several challenges related to speed limitations, temperature and process variability.

Ultra-low voltage, low-power voltage references have been obtained by biasing all MOSFETs in the subthreshold regime [17]-[29]. Both designs [17]-[29] use two MOSFET types with different threshold voltages in order to obtain a temperature-compensated voltage reference, and represent the state of the art in terms of minimum power consumption and supply voltage. The solution proposed in [29], the 2T (two transistors) voltage reference, exhibits a supply voltage as low as 0.5 V and a power consumption of only a few picowatts, while the solution reported in [17] shows a proper operation for a minimum supply voltage of only 0.45 V by consuming a power of 2.6 nW.

The configuration reported in [29] represents the best solution for the future low power applications since it allows a significant reduction in power consumption and occupied area. However, it exhibits degraded performance when the threshold voltages difference (equal to about V_{REF}) goes below 7.5V_T, where V_T is the thermal voltage (26 mV at 300 K). As a result the minimum operating supply voltage for the 2T voltage reference presented in [29] is predicted to be about 12-13 V_T, which corresponds to a voltage slightly less than 450 mV as in the case of [17]. Hence, it is crucial to understand if the solution of the 2T voltage reference is appropriate also for the future scenario of chips biased at V_{DD} lower than 450 mV. However it is apparent that for supply voltage scalability of this solution a new temperature compensation technique and new design considerations are simultaneously required.

In this section a reference voltage based on the 2T architecture capable to operate at the minimum supply voltage of only 150 mV while consuming 26.1 pW is presented. The implemented solution consists of two MOSFETs of the same type and exploits the dependence of the threshold voltage on the transistor size to generate a temperature-compensated reference voltage with a magnitude around the thermal voltage V_T. This makes the proposed circuit a valid solution to process the voltage generated by a thermoelectric generator (TEG) [26]-[27].

3.5.2. Operating principle and design considerations

The temperature compensation technique presented in this work is developed on 2T voltage reference proposed in [29]. In the proposed solution however, two MOSFETs of the same type are used, while MOSFETs with different threshold voltage are required in the reference presented in [29]. Figure 3.19 reports the schematic of the proposed circuit. In all references proposed so far the term (1-*exp*(-V_{DS}/V_T)) in the subthreshold current is neglected for all MOSFETs working in subthreshold regime [15]-[17], [29]. As already reported in Chapter 2, this approximation is true if $V_{DS}>4V_T$ which is equal to about 105 mV at room temperature.

However it is evident that this assumption is a strong limitation for supply voltage scaling. For this reason it is crucial to understand when this condition can be removed without any penalties in other design specifications.

By considering the contribution of the drain-source voltages on the currents I_1 and I_2 (current in M_1 and M_2, respectively), using equation (2.3), the following expressions are obtained:

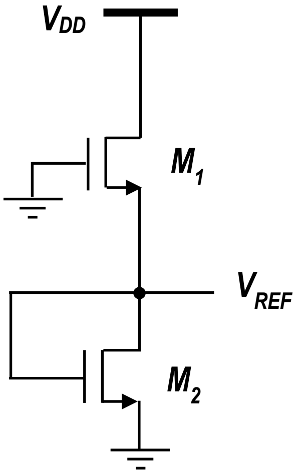


Figure 3.19. Schematic of the implemented voltage reference.

$$I_1 = \mu_{N,1} C_{ox,1}\left(\frac{W}{L}\right)_1 V_T^2 \exp\left(-\frac{V_{REF}+V_{TH,1}}{n_1 V_T}\right)\times\left(1-\exp\left(-\frac{V_{DD}-V_{REF}}{V_T}\right)\right), \tag{3.53}$$

$$I_2 = \mu_{N,2} C_{ox,2}\left(\frac{W}{L}\right)_2 V_T^2 \exp\left(\frac{V_{REF}-V_{TH,2}}{n_2 V_T}\right)\times\left(1-\exp\left(-\frac{V_{REF}}{V_T}\right)\right). \tag{3.54}$$

Since M_1 and M_2 are two nMOSFETs of the same type the approximation of $C_{ox,1}\approx C_{ox,2}=C_{ox}$ and $n_1\approx n_2=n$ is assumed.

In the 2T voltage reference transistor M_1 works as a current source. As pointed out in [17] when the reference voltage is coincident with the gate-source voltage of a diode-connected nMOS transistor, the stability of the generated reference voltage with respect to supply voltage variations mostly depends on the stability with respect to V_{DD} of the current injected into the load transistor. This condition provides the only constraint for supply voltage scaling in the 2T voltage reference, thus to ensure the best stability against V_{DD} variations, the drain-source voltage of M$_1$ has to be higher than $4V_T$, which means that V_{DD} has to be higher than V_{REF}+$4V_T$. On the other hand, the magnitude of the reference voltage, and consequently of the drain-source voltage of M$_2$, does not affect both supply voltage and temperature stability of the considered configuration. The latter observation suggests that in the 2T voltage reference it is possible to obtain a V_{REF} lower than $4V_T$ and consequently a minimum supply voltage lower than $8V_T$. However, since V_{REF} scales below $4V_T$ the temperature compensation technique proposed in [29] cannot be applied. As a result a new temperature compensation technique becomes necessary.

By assuming V_{DD}>V_{REF}+$4V_T$ and by equating I_1 and I_2, the following expression is obtained:

$$\ln\left\{\left(\frac{W}{L}\right)_R \frac{\mu_{N,1}}{\mu_{N,2}}\right\} = \frac{2V_{REF} + V_{TH,1} - V_{TH,2}}{nV_T} + \ln\left(1-\exp\left(-\frac{V_{REF}}{V_T}\right)\right), \tag{3.55}$$

where $(W/L)_R$=$(W/L)_1$/$(W/L)_2$. Although equation (3.55) has not explicit solution for V_{REF}, an approximate solution can be found by replacing the logarithmic term with its linear approximation

$$\ln\left(1-\exp\left(-\frac{V_{REF}}{V_T}\right)\right) \approx A + B\frac{V_{REF}}{V_T}, \tag{3.56}$$

where the two fitting parameters A and B depend only on the chosen interval of V_{REF}. Since in this work the main goal is to find the minimum achievable value of V_{REF}, the condition of $V_{REF} \in [V_T/2 - V_T]$ is imposed. In Figure 3.20 the goodness of the linear fitting is reported for the latter assumption on V_{REF} magnitude.

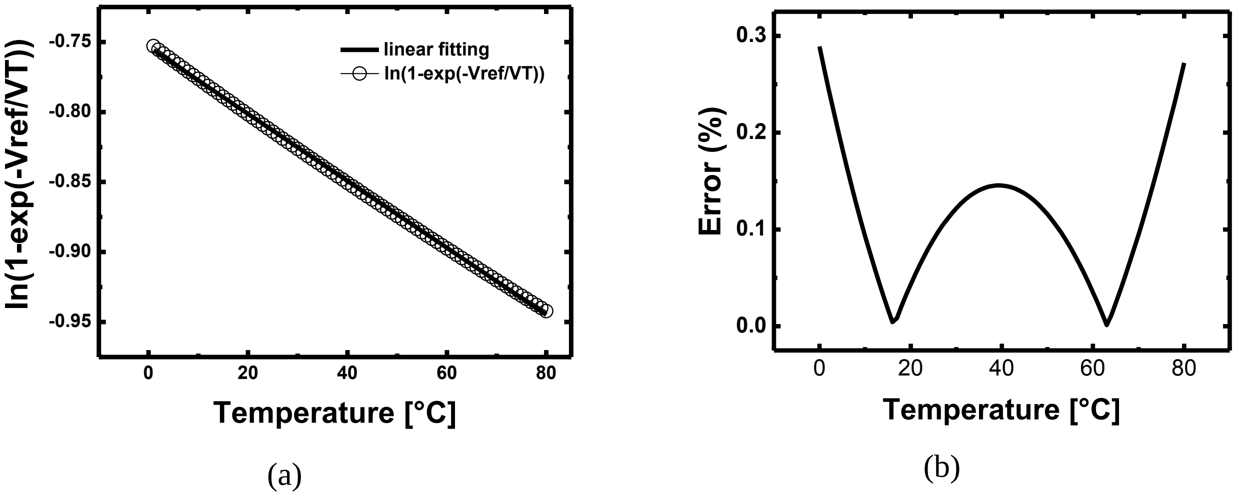


Figure 3.20. Linear approximation in (3.56) (a). The error due to the approximation is reported in (b).

The corresponding fitting parameters at room temperature are equal to A≈-1.35 and B≈0.92. However, it is worth noting that the proposed operating principle can be extended to a generic interval of V_{REF}, by evaluating the corresponding values of A and B. Introducing the threshold voltage dependence on temperature the reference voltage becomes equal to:

$$V_{REF} = \frac{nV_T \ln\left\{\left(\frac{W}{L}\right)_R \frac{\mu_{N,1}}{\mu_{N,2}}\right\} + V_{TH,2}(T_0) - V_{TH,1}(T_0)}{2+nB} + \frac{\left|k_{T,1}\right|\left(T-T_0\right) - nV_T A - \left|k_{T,2}\right|\left(T-T_0\right)}{2+nB}. \qquad (3.57)$$

Differentiating the (3.57) with respect to temperature and setting $\partial V_{REF}/\partial T=0$ the value of $(W/L)_R$ for temperature compensation is obtained:

$$\left(\frac{W}{L}\right)_R = \frac{\mu_{N,2}}{\mu_{N,1}} \exp\left\{\frac{q}{nk}\left[\left|k_{T,2}\right| - \left|k_{T,1}\right|\right] + A\right\}. \qquad (3.58)$$

Replacing (3.58) into (3.57), the following expression of the temperature-compensated reference voltage is obtained:

$$V_{REF} \approx \frac{V_{TH,2}(T_0) + \left|k_{T,2}\right|T_0 - V_{TH,1}(T_0) - \left|k_{T,1}\right|T_0}{2+nB}. \qquad (3.59)$$

The two threshold voltages at room temperature depend on drain-source voltage V_{DS} and body-source voltage V_{BS} according to equation [30]:

$$V_{TH}(T_0) = V_{TH0} - \lambda_D V_{DS} - \lambda_B V_{BS}, \qquad (3.60)$$

where V_{TH0} is the threshold voltage at V_{DS}=V_{BS}=0 V, while λ_D and λ_B are the *DIBL* and body coefficients, respectively. Replacing (3.60) into (3.59) we obtain the following expression:

$$V_{REF} \approx \frac{V_{TH0,2} + \left|k_{T,2}\right|T_0 - V_{TH0,1} - \left|k_{T,1}\right|T_0}{\left(2+nB\right)\left(1+\lambda_{D,1}+\lambda_{D,2}+\lambda_{B,1}\right)} \approx \frac{V_{TH0,2} + \left|k_{T,2}\right|T_0 - V_{TH0,1} - \left|k_{T,1}\right|T_0}{\left(2+nB\right)\left(1+2\lambda_D+\lambda_B\right)}. \qquad (3.61)$$

In analogy with [29], from (3.61) the reference voltage is proportional to the difference between the two threshold voltages at room temperature. As a consequence, to obtain a specific value of V_{REF}, a specific difference between the two threshold voltages has to be ensured. Since two nMOSFETs of the same threshold type are used in the proposed solution a significant difference between the two thresholds voltages can be obtained by using the dependence of the threshold voltage on transistor size. In Figure 3.21 the simulated threshold voltage as a function of channel length and width for an nMOSFET in 0.18 µm CMOS technology is reported. The simulation shows a substantial difference between the threshold voltage of a long and short channel nMOSFET. In order to satisfy the approximation reported in (3.56) we choose a combination for L_1 and L_2 that gives a value of $V_{TH,2}$-$V_{TH,1}$ in the range for V_{REF} (i.e. from V_T/2 to V_T in this case). From Figure 3.21, imposing L_1=25 µm and L_2=2 µm, the difference between the two threshold voltages is equal to about 39 mV, which means a rough value of V_{REF} equal to about 18 mV (by neglecting nB in (3.59)). After imposing the channel length of both transistors, W_1 and W_2 are chosen in order to compensate the temperature dependence of V_{REF}. By selecting W>30 µm for both transistors, V_{REF} depends only on the selected transistor length (from Figure 3.21 V_{TH} is almost independent of transistor width for W>30 µm).

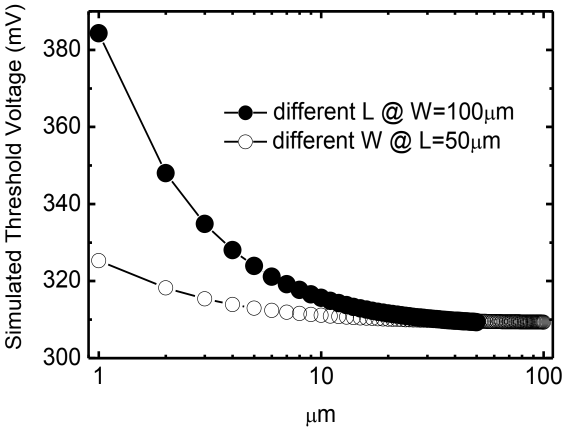


Figure 3.21. Simulated threshold voltage as a function of channel length and width.

From the approximated expression (3.58) the optimal transistor size ratio for temperature compensation is 0.09 very close to the optimal transistor size ratio of 0.1 obtained with SPICE simulations. In Figure 3.22 the simulated normalized temperature coefficient (TC) as a function of the transistor size ratio is reported. The simulation was performed by imposing L_1=25 µm and L_2=2 µm and W_2=50 µm. As a result, the optimal value of W_1 provided by circuit simulation is 63.9 µm, quite close to the optimal value of 56.2 µm provided by the simplified expression (3.58). Note that the optimal transistor size ratio evaluated in according to the temperature compensation technique reported in [29] would be 0.34, corresponding to W_1= 212.5 µm. These results confirm the necessity of a new temperature compensation technique for the 2T voltage reference with respect to the technique proposed in [29] when V_{REF} scales below $4V_T$ and confirm, at the same time, the validity of the proposed operating principle.

Since in the proposed scheme the reference voltage is obtained by using the threshold voltage dependence on transistor size, L_2 has to be notably lower than L_1. This choice does not affect significantly the robustness of the reference voltage against supply voltage variation since the line sensitivity of the implemented solution has a second order dependence on L_2. On the other hand, as reported before, large area MOSFETs can help to reduce the dispersion against intra-die process variations. Despite that since in the proposed design the difference in the threshold voltages of the two MOSFETs is simply obtained by using different geometries and not by using different process variables (e.g. oxide thickness, substrate doping) such as in previous works [17]-[29], a lower dispersion is expected. The power consumption of the proposed solution is also a function of transistor sizing mostly because in the 2T voltage reference the power consumption depends on the magnitude of the generated voltage reference.

Indeed, the current flowing in the circuit depends exponentially from V_{REF} according to the following equation:

$$I_{DD} \approx \mu_{N,1} C_{ox,1} \left(\frac{W}{L} \right)_1 V_T^2 \exp\left(-\frac{V_{REF} + V_{TH,1}}{n_1 V_T} \right), \tag{3.62}$$

consequently, the power consumption of the proposed circuit is given by

$$P \approx V_{DD} \times \mu_{N,1} C_{ox,1} \left(\frac{W}{L} \right)_1 V_T^2 \exp\left(-\frac{V_{REF} + V_{TH,1}}{n_1 V_T} \right). \tag{3.63}$$

From (3.63) the power consumption can be reduced by reducing V_{DD}, however since $V_{DD} \approx V_{REF} + 4V_T$, a reduction in V_{DD} causes an increment of the exponential term in (3.63) and consequently an increment in power consumption. For this reason the best choice to reduce V_{REF} (V_{DD}) without significant penalty in power consumption consists in the use of two high threshold voltage transistors for M_1 and M_2.

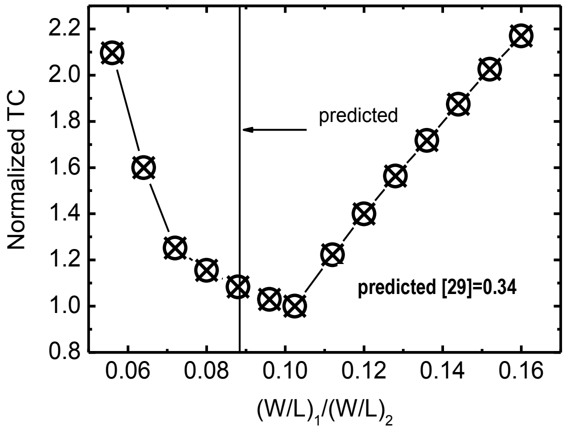
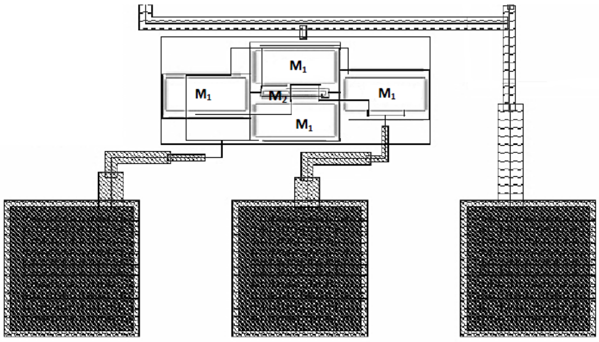


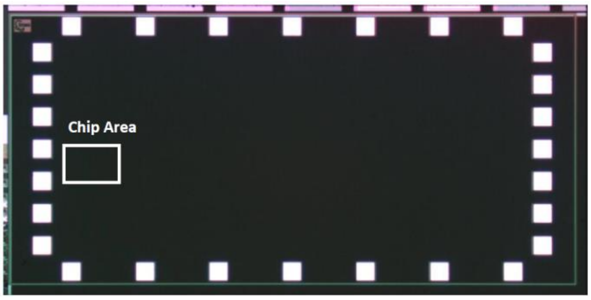
Figure 3.22. Simulated normalized TC as a function of the transistors size ratio. The simulated optimal transistor size ratio is equal to 0.10. According to (3.58) the predicted optimal transistor ratio is 0.09 while a value of 0.34 is evaluated in according to the temperature compensation technique proposed in [29].

3.5.3. Measurement results

The performances of the proposed sub-*kT*/*q* voltage reference have been tested over a set of 60 samples of two separated batches fabricated in UMC 0.18-µm, 1.8 V/3.3 V, CMOS process.



(a)



(b)

Figure 3.23. Layout (a) and chip photo (b) of the implemented voltage reference.

The layout and the chip photo of the proposed circuit are reported in Figure 3.23(a) and Fig 3.23 (b) respectively. In order to improve the process stability against systematic mismatch due to stress and temperature, transistor M_1 is divided in four parts placed in a common centroid configuration. This solution results in an active area of 1200 μm^2. The on-wafer electrical measurements have been performed by using a probe station SUMMIT 11861B with Temptronic thermal controller and a Keithley 4200-SCS parameter analyzer.

Figures 3.24-3.27 show the measured performance of a typical device. Figure 3.24 shows the reference voltage as a function of the supply voltage at the temperature of 25°C. It is worth noting that the circuit starts to work properly from only 150 mV which is by far the best result in this specification for a voltage reference.

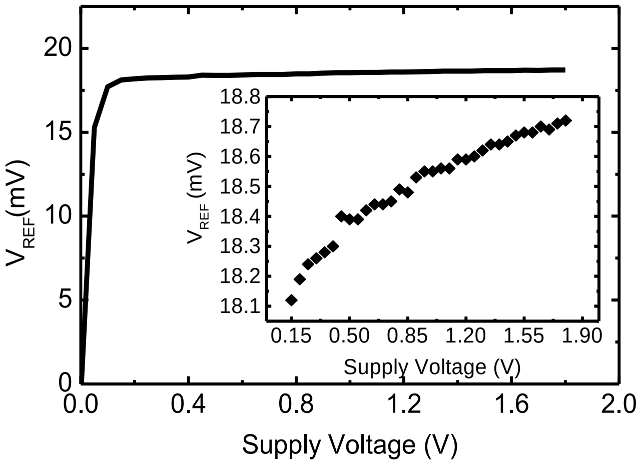


Figure 3.24. Measured V_{REF} as a function of supply voltage at 25 °C and its zoom in the operating range. The measured mean value of V_{REF} is 17.69 mV against a simulated value of 15.33 mV.

The temperature dependence of V_{REF} for different supply voltages is reported in Figure 3.25 while Figure 3.26 gives an overview of the measured V_{REF} values in the 2D domain of operating temperatures and supply voltages. The power consumption for different temperatures and supply voltages is reported in Figure 3.27.

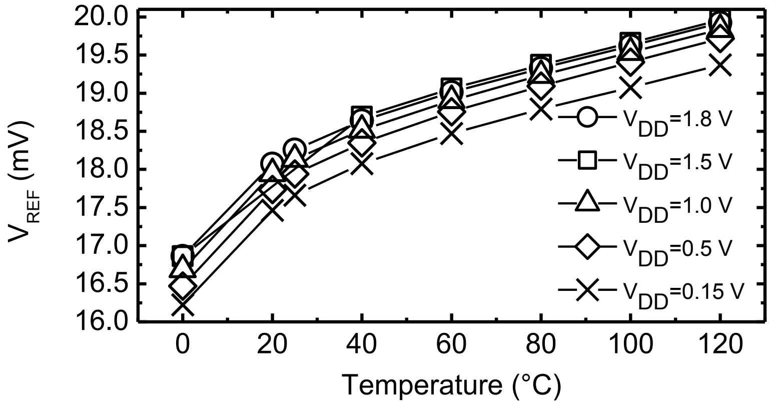


Figure 3.25. Measured V_{REF} versus temperature for different supply voltages.

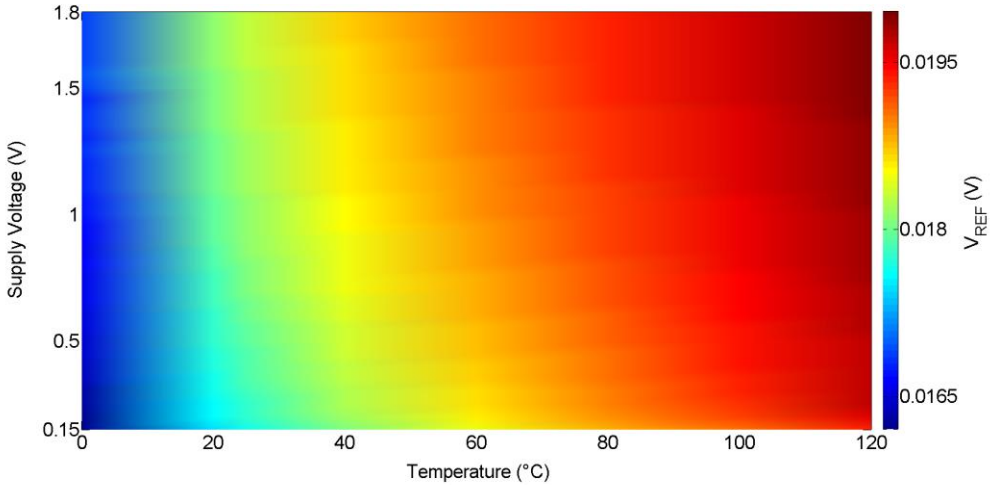


Figure 3.26. Measured V_{REF} versus temperature and supply voltage.

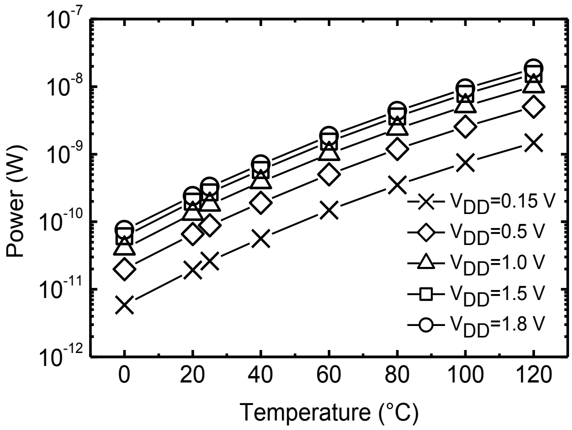


Figure 3.27. Measured power consumption as a function of temperature for different supply voltages.

The results of the statistical analysis performed over a set of 60 samples are summarized in Figure 3.28 and Table 3.IV, where the mean (μ) and the standard deviation (σ) of the reference voltage, the temperature variation, the supply voltage variation and the power consumption are reported in detail. In Figure 3.28(a) the measured value of V_{REF} at the temperature of 25°C for V_{DD}=0.15 V is reported. The mean value of V_{REF} is 17.69 mV, against a simulated value of 15.33 mV, with a standard deviation of only 0.29 mV. In Figure 3.28(b) the variation of the reference voltage in the temperature range from 0 to 120 °C is reported. The measured mean variation of V_{REF} against temperature is 26.74 µV/°C with a standard deviation of 5.57 µV/°C. By normalizing for V_{REF}, the TC is equal to 1462.4 ppm/°C with a standard deviation of 324 ppm/°C.

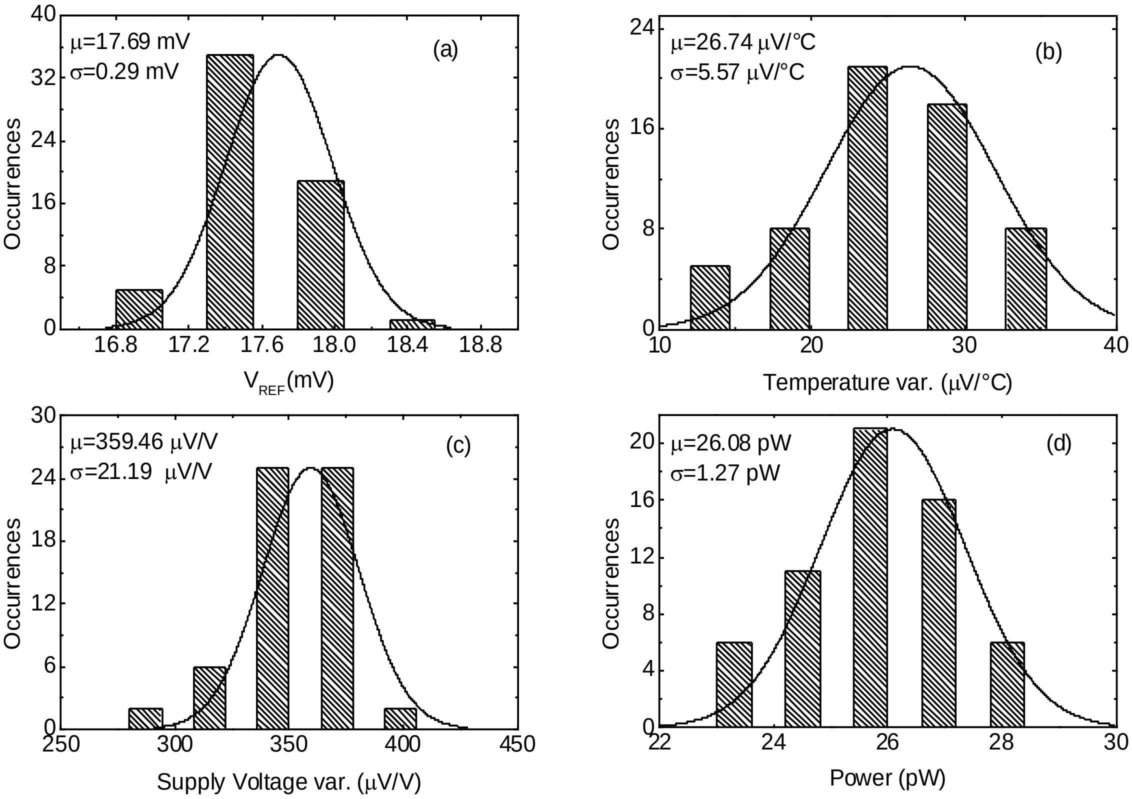


Figure 3.28. Measured distribution over 60 samples of: (a) reference voltage @ 25°C; (b) absolute temperature variation of V_{REF}; (c) supply voltage variation versus V_{DD}; (d) power consumption @ 25°C and V_{DD}=0.15 V.

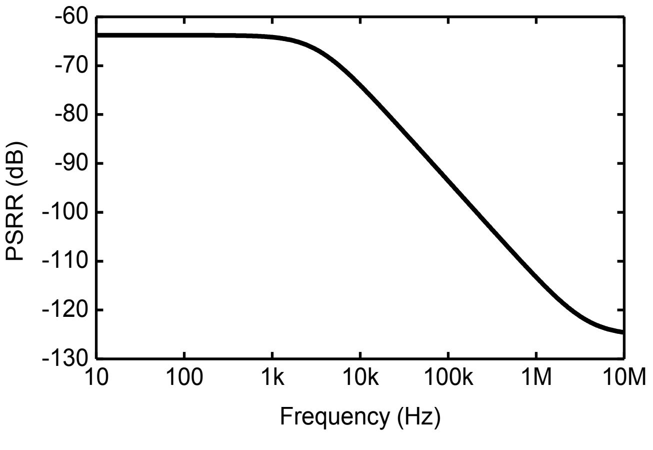


Figure 3.29. Simulated *PSRR* at V_{DD}=0.15 V.

In Figure 3.28(c) the distribution of the variation of V_{REF} against V_{DD} variations is reported. The mean variation of V_{REF} for V_{DD} ranging from 0.15 V to 1.8 V, at the temperature of 25°C, is equal to

359.46 µV/V with a standard deviation of 21.19 µV/V. By normalizing for V_{REF}, the mean LS is equal to 2.03%/V with a standard deviation of 0.11 %/V. The statistical analysis of power consumption at V_{DD}=0.15 V and T=25°C is reported in Figure 3.28(d). The mean power consumption is 26.08 pW with a standard deviation of 1.27 pW. The simulated *PSRR* without any filtering capacitor, for the minimum operating voltage of V_{DD}=0.15 V, is reported in Figure 3.29.

The *PSRR* is equal to about -64 dB at low frequencies and -124 dB at 10 MHz. As reported in [29], the 2T voltage reference acts as a low-pass filter, so the *PSRR* improves at higher frequencies. In Table 3.V the performances of the proposed solution are compared with the two best low power, low voltage solutions reported in literature [17],[29].

TABLE 3.IV
STATISTICAL ANALYSIS OVER 60 SAMPLES

	μ	σ
V_{REF} [mV]	17.69	0.29
TC [ppm/°C]	1462.4	324
Temp. variation [µV/°C]	26.74	5.57
LS [%/V]	2.03	0.11
Supply Voltage var. [µV/V]	359.46	21.19
Power [pW]	26.08	1.27

The proposed circuit starts to work properly from 150 mV, which improves by 350 mV the result obtained in [29] and by 300 mV the solution proposed in [17], which represents so far the best solution in terms of minimum operating V_{DD} for a voltage reference.

The power consumption at room temperature for the minimum operating supply voltage is about one order of magnitude larger than in [29] and about two orders of magnitude smaller than the solution proposed in [17]. The mean values of TC and LS are significantly higher than the other solutions [19]-[20].

This is due because both parameters are obtained by normalizing for V_{REF}, which in the proposed design is significantly lower than the other solutions. Therefore, for a fair comparison, in Table 3.VI the principal performance indicators are compared in terms of absolute variations.

From this comparison, the proposed solution exhibits better temperature stability with respect to the other solutions and a supply voltage stability better than the one reported in [17], but worse than the one reported in [29]. In Table 3.VI the comparison is performed also in terms of dispersion of the main figures of merit. The relative standard deviation σ/μ of the reference voltage is 1.6% compared to 0.72 % of [29] and 3.9% of [17]. Nevertheless, the absolute variation of V_{REF} is better compared to the other solutions. The standard deviation of V_{REF} is only 0.29 mV, compared to about 1.3 mV of [17] and 10 mV of [29] (Table 3.VI). The stability of the generated reference voltage against intra- and inter-die variations has been investigated by means of Monte Carlo simulations.

The simulated relative standard deviation of V_{REF} over 1000 samples is about 0.9 %. The proposed solution overcomes the other low-power, low-voltage voltage references in terms of dispersion of TC, LS and power consumption. The relative standard deviation of TC is 22 % which is about 3 times smaller than the compared solutions, while the relative standard deviation of LS and power consumption is respectively about 2 and 5 times smaller than in [17] ([29] does not reports dispersion data on these two parameters). The lower dispersion of the main figures of merit observed in the proposed reference voltage is ascribed to the use of the same threshold-type of MOSFETs, whereas the voltage reference design presented in [17]-[29] are based on two different threshold-types of MOSFET.

TABLE 3.V
COMPARISON WITH LOW-VOLTAGE LOW-POWER CMOS VOLTAGE REFERENCES

		This work	***Ref.* [29]** ***(no trimming)***	***Ref.* [17]**
Technology		0.18μm	0.13μm	0.18μm
Supply voltage (V)		0.15 to 1.8	0.5 to 3.0	0.45 to 2
Power ***@ room temperature***		26.1 pW@0.15V 342.3 pW@1.8V	2.2pW@0.5V -	3.15nW@0.45V 14.4nW@1.8V
V_{REF} (mV)		17.69	176.1	263.5
TC (ppm/°C) ***T range(°C)***		1462.4 (average) [0:120]	62 (average) [-20:80]	142.1 [0:100]
LS (%/V)		2.03	0.033	0.44
PSRR (dB) ***Low freq [≤100Hz]*** ***High freq [≥10MHz]***		V_{DD}=0.15V -64 (sim.) -124(sim.)	- -53 -62	V_{DD}=0.45V -49.4 (-12.2 sim.)
Process ***Sens.σ/μ***	***V_{REF}***	1.6%	0.72%	3.9%
	TC	22%	66%	60.6%
	LS	5.4%	-	13.1%
	Power	5%	-	26.9%
Die area (mm^2)		1200 μm^2	1350 μm^2	0.0430

TABLE 3.VI
ABSOLUTE VARIATION OF THE PRINCIPAL FIGURES OF MERIT FOR THE COMPARED VOLTAGE REFERENCES

		This work	***Ref.* [29]** ***(no trimming)***	***Ref.* [17]**
Temperature var. (μV/°C)		26.74	35	37.6
Supply voltage var. (μV/V)		359.5	57.72	1185.2
Process ***variations***	***σ/μ*** ***V_{REF} (mV)***	0.29	~1.3	10

3.5.4. Conclusion

A new sub-*kT*/*q* voltage reference capable of operating with a minimum supply voltage of only 150 mV was presented. Although the proposed voltage reference is based on the same 2T architecture proposed in [29], it presents two fundamental features of novelty. The first one is that it is based on different equations since it works at operating voltages significantly lower than the solution reported in [29]. The second one is that the proposed design does not require two different threshold-types of MOSFET, since it exploits the dependence of the threshold voltage on transistor size. Measurements performed over a set of 60 samples from two separated batches show a mean reference voltage of 17.69 mV with a standard deviation of only 0.29 mV. The temperature variation in the range from 0 to 120°C is 26.74 µV/°C which is the best among the low-power, low-voltage solutions proposed so far. The line sensitivity of the reference voltage for supply voltage ranging from 0.15 to 1.8 V is 359.46 µV/V. The power consumption at room temperature for a supply voltage of 0.15 V is 26.08pW. The occupied area is 1200 µm^2. In addition, the dispersion of the temperature coefficient, supply voltage sensitivity, and power consumption are smaller than the state-of-the-art solutions. The extremely low-voltage operation and the decapicowatt power consumption make the proposed solution very attractive for battery-free, energy-harvesting applications.

Bibliography

[1] A. Wang, C. B. Highsmith, A. P. Chandrakasan. **Sub-threshold design for Ultra low-power systems**. *Springer*, 2006.

[2] R. J. Widlar, “New developments in IC voltage regulators,” in *IEEE Journal of Solid-State Circuits*, vol. 6, no. 1, pp. 2-7, Feb. 1971.

[3] M. D. Ker, J. S. Chen,”New Curvature-Compensation Technique for CMOS Bandgap Reference with Sub-1-V Operation,” *IEEE Transaction on Circuits and Systems II, Exp. Briefs*, vol. 53, no. 8, Aug. 2006.

[4] B. S. Song and P. R. Gray, “A precision curvature-compensated CMOS bandgap reference,” *IEEE Journal of Solid State Circuits*, vol. DC-18, no. 6, pp.634–643, Dec. 1983.

[5] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, “A CMOS bandgap reference circuit with sub-1-V operation,” *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 670-674, May. 1999.

[6] T. Hirose, K. Ueno, N. Kuroki and M. Numa, “A CMOS bandgap and sub-bandgap voltage reference circuits for nanowatt power LSIs,” in *Proceeding of IEEE Asian Solid-State Circuits Conf.* (A-SSCC), Nov. 2010, pp. 1-4.

[7] K. N. Leung, P. K. T. Mok, “A CMOS voltage reference based on weighted ΔVGS for CMOS low-dropout linear regulators,” *IEEE Journal of Solid-State Circuits*, vol. 38, no. 1, pp. 146-150, Jan. 2003.

[8] P. Malcovati, F. Maloberti, “Curvature-Compensated BiCMOS Bandgap with 1-V Supply Voltage,” *IEEE Journal of Solid-State Curcuits*, vol.36, no.7, pp. 1076 – 1081, Jul 2001.

[9] B. S. Song and P. R. Gray, “Threshold-voltage temperature drift in ion-implanted MOS transistors,” *IEEE Journal of Solid-State Circuits*, vol. SC-17, no. 2, pp. 291-298, Apr. 1982.

[10] K. N. Leung, P. K. T. Mok, “A CMOS voltage reference based on weighted ΔVGS for CMOS low-dropout linear regulators,” *IEEE Journal of Solid-State Circuits*, vol. 38, no. 1, pp. 146-150, Jan. 2003.

[11] G. De Vita, G. Iannaccone, and P. Andreani, “A 300 nW, 12 ppm/°C voltage reference in a digital 0.35μm CMOS process,” in *Symposium of VLSI Circuits Dig. Tech. Papers*, Honolulu, HI, 2006, pp. 81–82.

[12] G. De Vita, G. Iannaccone, “A Sub-1-V, 10 ppm/°C, Nanopower Voltage Reference Generator”, *IEEE Journal of Solid-State Circuits*, vol. 42, no. 7, pp. 1536-1542, Jul. 2007.

[13] W. Yan, W. Li and R. Liu, “Nanopower CMOS sub-bandgap reference with 11 ppm/°C temperature coefficient”, *Electron Letters*, vol. 45, no. 12, pp. 627-629, Jun. 2009.

[14] K. Ueno, T. Hirose, T. Asai, Y. Amemiya, “A 300 nW, 15 ppm/°C, 20 ppm/V CMOS Voltage Reference Circuit Consisting of Subthreshold MOSFETs,” *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 2047-2054, Jul. 2009.

[15] P.H. Huang, H. Lin and Y.T. Lin,” A simple subthreshold CMOS voltage reference circuit with channel length modulation compensation”, *IEEE Transaction on Circuits and Systems II, Exp. Briefs*, vol.53, no.9, pp. 882-885, Sep. 2006.

[16] H. Wang and Q. Ye, “A CMOS voltage reference without resistors for ultra-low power applications,” in *Proceeding of the 7th Int. Conf. ASIC 2007* (ASICON’07), Oct. 2007, pp. 526–529.

[17] L. Magnelli, F. Crupi, P. Corsonello, C. Pace, and G. Iannaccone, “A 2.6 nW, 0.45 V Temperature Compensated Subthreshold CMOS Voltage Reference”, *IEEE Journal of Solid-State Circuits*, vol.46, no. 2, pp. 465-474, Feb. 2011.

[18] M. Pelgrom, A. Duinmaijer, and A. Welbers, “Matching properties of MOS transistors,” *IEEE Journal of Solid-State Circuits*, vol. 24, no. 1, pp. 1433-1439, Oct. 1989.

[19] B. Razavi, **Design of Analog CMOS Integrated Circuit**, *McGraw-Hill Higher Education*. 2003.

[20] A. Hastings, **The Art of Analog Layout**. Englewood Cliffs, *NJ: Prentice-Hall*, 2001.

[21] D.Albano, F.Crupi, F. Cucchi, G. Iannaccone, "A picopower temperature-compensated, subthreshold CMOS voltage reference", *International Journal of Circuit Theory and Applications*, vol. 42, no 12, pp 1306-1318, Dec. 2014.

[22] M.-E. Hwang and K. Roy, "A 135 mV 0.13 µW process tolerant 6T subthreshold DTMOS SRAM in 90 nm technology," in *Proc. IEEE Custom Integrated Circuits Conf.* (CICC), 2008, pp. 419–422.

[23] J. Burr, "Cryogenic ultra low-power CMOS," in *Proc. IEEE Symposium on Low Power Electronics*, Oct. 1995, pp. 82–83.

[24] N. Lotze and Y. Manoli, A 62 mV 0.13 µm CMOS Standard-Cell-Based Design Technique Using Schmitt-Trigger Logic," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 47-70, Jan. 2012.

[25] C. C. Enz and E. A. Vittoz, "CMOS low-power analog circuit design," in *Proceeding of IEEE International Symposium on Circuits and Systems* (ISCAS'96), chapter 1.2, Tutorials, pp. 79–132.

[26] . J. Carlson, K. Strunz and B. P. Otis,"A 20 mV Input Boost Converter With Efficient Digital Control for Thermoelectric Energy Harvesting," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 741-750, Apr. 2010.

[27] Y. K. Ramadass, A. P. Chandrakasan "A Batteryless Thermoelectric Energy-Harvesting Interface Circuit with 35mV Startup Voltage," in *IEEE International Solid-State Conference* (ISSCC) Dig. Tech. Papers, pp. 486-487, Feb. 2010.

[28] E. Seevinck, E. A. Vittoz, M. D. Plessis, T.-H. Joubert, W. Beetge ",CMOS Translinear Circuits for Minimum Supply Voltage,"*IEEE Transaction on Circuits and Systes II*: *Analog and Digital Signal Processing*, vol. 47, no. 12, pp. 1560-1564, Dec. 2000.

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

[30] M. Alioto. "Understanding DC Behavior of Sub-Threshold CMOS Logic through Closed-Form Analysis," *IEEE Transaction on Circuits and Systems I*, Regular Papers, vol. 57, no. 7, pp. 1597–1607, Jul. 2010.

4. Design of a low-power, low-voltage subthreshold current reference

4.1 Introduction

As reported in the previous chapter, bias circuits like current references and voltage references are essential components in analog to digital and digital to analog converters, amplifiers, PLLs, filters and sensors. As more of these systems are used remotely, the demand for low power, low-voltage operation increases.

Several solutions for the generation of a reference current have been proposed [1]-[15]. The classical current reference consists of two complementary current mirrors and a resistor [1]. In such solution all the transistors work in strong-inversion regime. The temperature compensation is achieved by using a very large value of the resistor; as a consequence this solution shows a large occupied area and exhibits a large process variation due to the sensitivity of the resistor.

Other solutions without resistors are based on the principle of the sum between a PTAT and CTAT current [2]-[3]. Despite the good performances they exhibit high power consumption and a minim supply voltage above 1 V.

Other solutions like [4]-[6] exploit the concept of the root square circuit to obtain a stable current.

In [7]-[8] the zero TC (ZTC) bias point of a MOSFET is exploited. Both [7]-[8] show minimum supply voltage and power consumption widely over 1V and 1 μW respectively.

[9] proposed a current reference with MOSFETs working in subthreshold and strong inversion regime. The power consumption is only 55 nW. However the minimum supply voltage of [9] is 1.5 V which is too high for the low- voltage contexts. It is obvious that despite the good performances in terms of temperature and supply voltage stability, all the previous solutions do not suit the requirements of the low-power, low-energy applications. A significant improvement in power consumption is offered by the solution reported in [10]. Here a fully-subthreshold CMOS current reference is reported. The solution shows a power consumption of only few nW, however, despite the power saving, the minimum supply voltage is still above 1 V. Additionally, the solution proposed in [10] shows large temperature sensitivity in comparison with the all-above threshold solutions and a variation of the reference current of about 10 %.

An ultra-low power solution has been proposed in [11] where a subthreshold current reference consuming only 23 pW is reported. Despite that also this solution shows a minimum operating voltage above 1 V.

A sub-1 V current reference is presented in [12]. The solution is in able to work properly starting from 0.8 V while consuming 290 nW. Despite the good performance also in terms of process stability the solution reported in [12] exhibits large area occupation. Moreover it uses resistors and BJTs, thus making really difficult the implementation in the battery-free contexts.

In this chapter, the first nanowatt subthreshold current reference working down to 0.5 V is presented. The proposed solution was implemented in 0.18 μm CMOS technology. A statistical analysis over 17 samples shows a mean output current of 68.27 nA at the minimum operating voltage of V_{DD}=0.5 V. The mean power consumption at room temperature and for V_{DD}=0.5 V is 40 nW. The mean line sensitivity is 0.03 %/V while the mean temperature coefficient is 2202 ppm/°C. The coefficient of variation σ/μ is equal to 1.1 % for the output current, 66.6 % for the line sensitivity and 9.8 % for the temperature coefficient. The occupied area of the proposed solution is only 0.016 mm^2.

In the following section a brief overview on some of the solutions proposed so far is reported, then the proposed scheme is introduced. The main design considerations and the measurement results are reported. Finally the proposed scheme is compared with other low-power, low-voltage solutions.

4.2. CMOS current references

The classical CMOS current reference is shown in Figure 4.1 [1]. It consists of a pMOS and nMOS current mirror and a resistor *R*. The solution is also know as “self-biased beta multiplier” current reference. In this configuration a positive feedback with gain less than 1 is used to reduce the sensitivity of the reference current with respect to supply voltage changes. Since $\beta_2=N\beta_1$,considering all the MOSFETs working in above threshold regime, the following expressions can be found:

$$V_{GS1} = V_{GS2} + I_{REF} R; \tag{4.1}$$

$$V_{GS1} = V_{TH1} + \sqrt{\frac{2I_{REF}}{\beta_1}}; \tag{4.2}$$

$$V_{GS2} = V_{TH2} + \sqrt{\frac{2I_{REF}}{N\beta_1}}. \tag{4.3}$$

Assuming the same value for the threshold voltage of M$_1$ and M$_2$, from (4.1)-(4.3) I_{REF} is equal to:

$$I_{REF} = \frac{2}{R^2\beta_1}\left[1-\frac{1}{N}\right]^2. \tag{4.4}$$

Equation (4.4) shows that the reference current does not depend on supply voltage. The temperature coefficient of the reference current is evaluated as:

$$TC = \frac{1}{I_{REF}} I_{REF} = -\frac{1}{\mu_P}\frac{\partial}{\partial T}\mu_P - \frac{2}{R}\frac{\partial}{\partial T}R. \tag{4.5}$$

Since both terms in (4.5) are process parameters the TC cannot be forced to zero. This solution requires very high value of resistance to achieve nA current which in turn results in a very high area occupancy and high sensitivity to process variations.

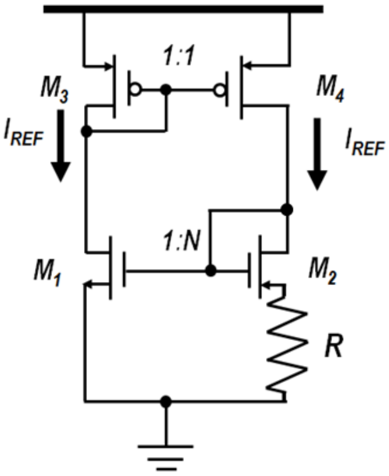


Figure 4.1. Self-biased beta multiplier CMOS current reference [1].

To overcome these problems resistorless solutions have been proposed.

In [13] a solution based on the self-biased beta multiplier was proposed. Here the resistor is replaced with an nMOSFET (M$_3$) biased by a diode connected nMOS (M$_4$) working in in strong inversion regime (see Figure 4.2) while transistors M$_1$ and M$_2$ work in weak inversion regime. According to the previous constraints on the operating regime of M$_1$-M$_4$, the currents I_3 and I_4 can be expressed as:

$$I_3 \approx \beta K_3 (V_{GS3} - V_{TH3}) V_{DS3}. \tag{4.6}$$

$$I_4 = \beta K_4 (V_{GS4} - V_{TH4})^2. \tag{4.7}$$

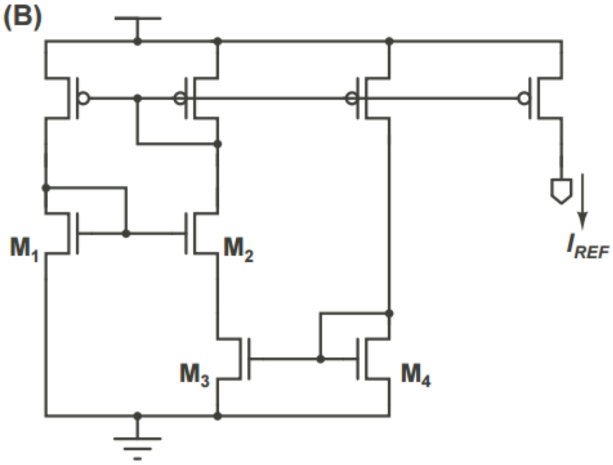


Figure 4.2. Resistorless CMOS self-biased beta multiplier current reference propose in [13].

Since V_{GS3}=V_{GS4}, combining (4.6) and (4.7) the reference current is :

$$I_{REF} = k_3 \beta \sqrt{\frac{2I_{REF}}{\beta K_4}} V_{DS3} = 2\frac{K_3^2}{K_4} \beta n^2 V_T^2 \ln^2\left(\frac{K_2}{K_1}\right). \tag{4.8}$$

The TC of (4.8) is equal to

$$TC = \frac{1}{I_{REF}} I_{REF} = \frac{1}{\mu_P}\frac{\partial}{\partial T}\mu_P + \frac{1}{V_T^2}\frac{\partial}{\partial T}\frac{1}{V_T^2} = \frac{2-m_N}{T}. \tag{4.9}$$

In the CMOS process m_N, the mobility coefficient for the nMOSFET, is about 1.5-2 thus the TC is never zero except in the case of $m_N=2$. From experimental results the TC of [13] is equal to about 1100 ppm/°C while the minimum supply voltage is equal to 1.2 V.

A fully-CMOS solution was proposed also in [4] (Figure 4.3). Transistors M_2–M_{11} operate in the subthreshold region, while M_1 and M_{12} operate in the strong inversion region. The gate-source voltages of M_1–M_{12} form a closed loop where

$$V_{GS1} = \sum_{j=12,10,8,6,4,2} V_{GSj} - \sum_{i=11,9,7,5,3} V_{GSi}. \tag{4.10}$$

Substituting the expression of V_{GS} in strong inversion for M_1 and M_{12} and the V_{GS} expression in weak inversion for M_2–M_{11}, the following expression is obtained:

$$I_{REF} \approx \frac{\beta}{2} n^2 V_T^2 \ln^2\left(120\frac{\prod\limits_{11,9,7,5,3} K_i}{\prod\limits_{10,8,6,4,2} K_j}\right)\left(\frac{K_1 K_{12}}{K_{12}-K_1}\right). \tag{4.11}$$

Thus the TC of the reference current is equal to:

$$TC = \frac{1}{I_{REF}} I_{REF} = \frac{1}{\mu_P}\frac{\partial}{\partial T}\mu_P + \frac{1}{V_T^2}\frac{\partial}{\partial T}\frac{1}{V_T^2} = \frac{2-m_N}{T}. \tag{4.12}$$

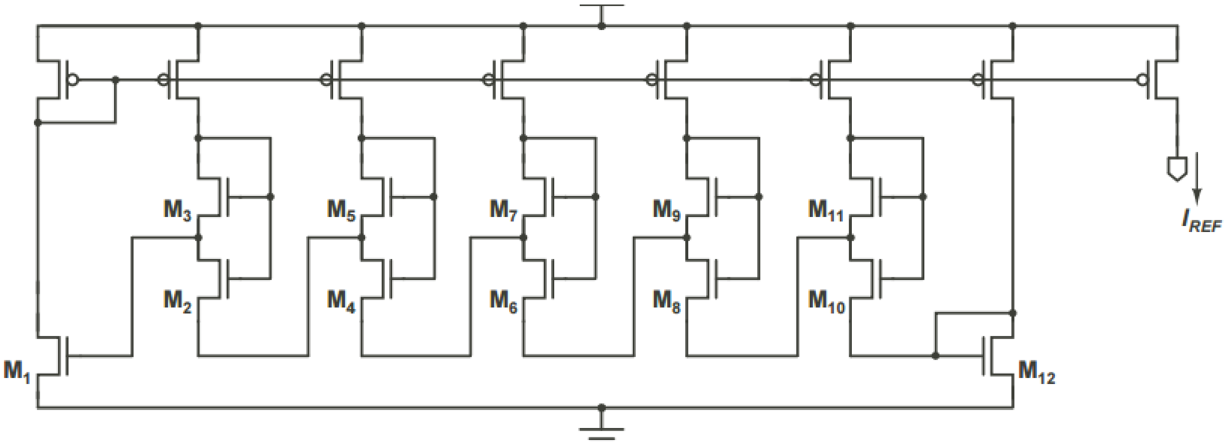


Figure 4.3. Current reference proposed in [4].

Experimental results reported in [4] show a TC of 375 ppm/°C, a power consumption of 10 µW and a minimum supply voltage of 3.5 V. The high power consumption and the large minimum supply voltage make this solution not suitable for the typical low-power contexts.

In [8] a current reference based on the zero TC bias point (ZTC) of the MOSFET is proposed. A voltage reference generates the gate-source voltage for an nMOSFET biased in strong inversion regime in order to achieve the temperature compensation of the drain current. Figure 4.4 (a). reports the measured transconductance characteristic of the MOSFET used as stable current source (Figure 4.4(b)).

According to Figure 4.4(b), assuming the above threshold conduction for M$_4$, the expression of the drain current I_{REF} becomes equal to:

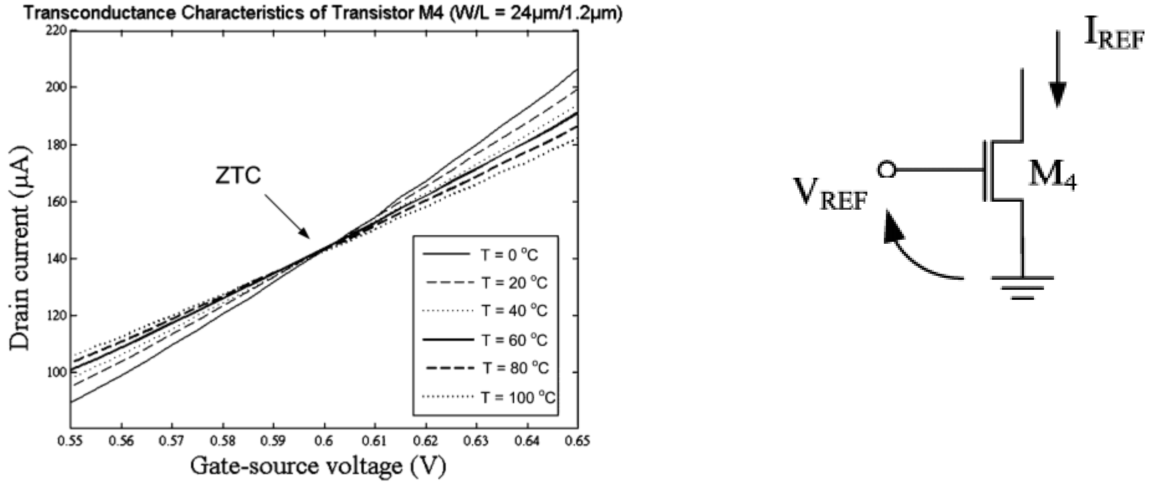
$$I_{REF} = \frac{\beta}{2} K_4 (V_{REF} - V_{TH})^2. \tag{4.13}$$

Introducing all the temperature effects, the (4.13) can be rewritten as:

$$I_{REF} = \frac{C_{OX}}{2} \frac{W_4}{L_4} \mu_N (T_0) \left[\frac{T}{T_0}\right]^{-mN} [V_{REF} - V_{TH}(T_0) + \kappa_4 (T - T_0)]^2. \tag{4.14}$$

If $V_{REF} = V_{TH}(T_0) - \kappa_4 T_0$ the current reference becomes equal to:

$$I_{REF} = \frac{C_{OX}}{2} \frac{W_4}{L_4} \mu_N (T_0) \left[\frac{T}{T_0}\right]^{-mN} [\kappa_4 T]^2. \tag{4.15}$$



(a) (b)

Figure 4.4. Zero TC voltage (ZTC) (a) for an nMOSFET used as current reference source (b) in [8].

Thus the TC of the current has an expression similar to the one already defined in (4.12). The authors claim that the value of m_N is nearby 2, thus a stable reference current can be obtained. The expression of such a current is equal to [5]:

$$I_{REF} = \frac{C_{OX}}{2} \frac{W_4}{L_4} \mu_N (T_0) [\kappa_4 T_0]^2. \tag{4.16}$$

Experimental results show a current reference equal to 144 μA with a TC of 185 ppm/°C. The nominal supply voltage is 1 V and the power consumption widely above 1 μW.

The same operating principle is exploited also in [7]. The solution consists of a threshold voltage monitoring circuit which generates a DC voltage equal to the threshold voltage at the room temperature (V_{TH0}), and the reference current generator. The schematic of the proposed solution is reported in Figure 4.5. In this solution the reference current can be expressed as:

$$I_{REF} = \frac{C_{OX}}{2}\frac{W}{L}\mu_N\left(V_{TH0}\left(1+\frac{R_1}{R_2}\right)-V_{TH}\right)^{\alpha}. \tag{4.17}$$

Thus, the TC of the reference current is equal to

$$TC = -\frac{m_N}{T} + \frac{\alpha\kappa}{V_{TH0}\left(1+\frac{R_1}{R_2}\right)-V_{TH}}. \tag{4.18}$$

The value of R_1/R_2 which ensures the condition of TC =0 is equal to

$$\frac{R_1}{R_2} = \left(\frac{\alpha}{m_N}-1\right)\frac{\kappa T}{V_{TH0}}. \tag{4.19}$$

The measured TC is equal to only 46 ppm/°C, the operating voltage ranges from 1.4 V to 3 V, the reference current is 18.4 μA. The authors do not report the total power consumption, hower since it is much higher than 1 μA also this solution is not compatible with the low-energy applications.

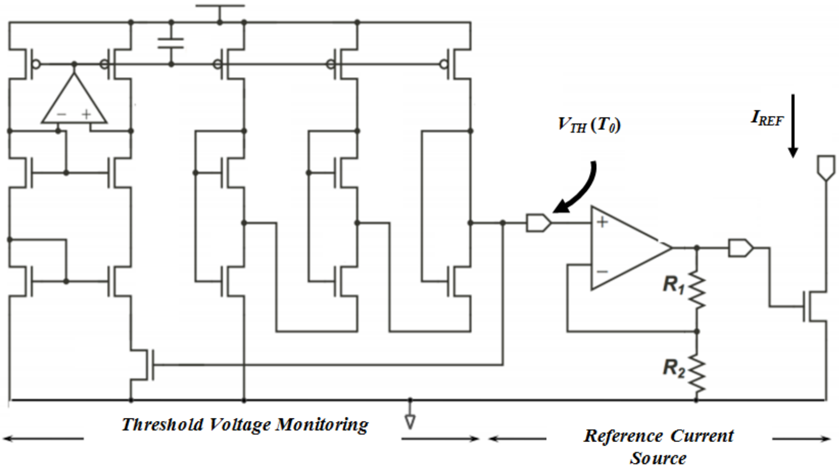


Figure 4.5. CMOS current reference proposed in [7]. The solution exploits the ZTC point of an nMOS to generate a reference current.

A low power, low voltage subthreshold current reference is proposed in [9]. The schematic of the proposed solution is reported in Figure 4.6. The transistors M_1, M_2, M_3, M_5 form a loop in which the following relationship holds:

$$V_{GS1} + V_{DS3} = V_{GS2} + V_{DS5}. \tag{4.20}$$

M_3 and M_5 work in subthreshold regime. Since the current flowing in both transistors is equal, then

$$V_{DS5}-V_{DS3}=nV_T\ln\left(\frac{K_2}{K_1}\right). \tag{4.21}$$

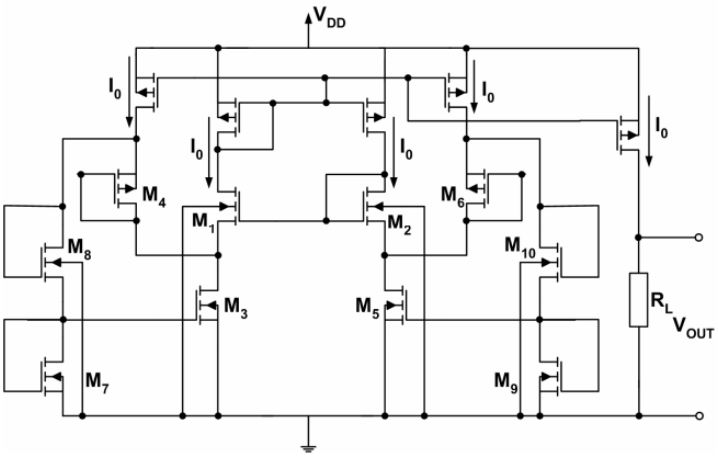


Figure 4.6. Subthreshold current reference proposed in [9].

M$_7$ and M$_8$ work in strong inversion regime, thus considering that V_{GS7}=V_{GS3}, the expression of the source of M$_1$ (M$_4$) is equal to

$$V_{S1}=V_{S4}=V_{THN}\left(1-\sqrt{\frac{K_7}{K_8}}\right)+\left(1+\sqrt{\frac{K_7}{K_8}}\right)V_{GS3}. \tag{4.22}$$

If the current flowing into M$_7$ and M$_8$ is negligible with respect to the current flowing in M$_4$, then the current flowing in M$_4$ is I_0 and the current in M$_3$ is $2I_0$.
Under this assumption, using (4.22), the expressions of V_{DS3} and V_{DS5} can be derived:

$$V_{DS3}=V_{THN}\left(1-\sqrt{\frac{K_7}{K_8}}\right)+\left(1+\sqrt{\frac{K_7}{K_8}}\right)\left(V_{THN}+\sqrt{\frac{4I_0}{K_3}}\right)-|V_{THP}|-\sqrt{\frac{2I_0}{K_3}}. \tag{4.23}$$

$$V_{DS5}=V_{THN}\left(1-\sqrt{\frac{K_9}{K_{10}}}\right)+\left(1+\sqrt{\frac{K_9}{K_{10}}}\right)\left(V_{THN}+\sqrt{\frac{4I_0}{K_5}}\right)-|V_{THP}|-\sqrt{\frac{2I_0}{K_6}}. \tag{4.24}$$

Combining (4.23) and (4.24) with (4.21), with the condition K_7/K_8=K_9/K_{10}, the expression of I_0 can be derived:

$$I_0=\frac{K_5}{2}\frac{n^2V_T^2\ln\left(\frac{K_2}{K_1}\right)}{\left[\sqrt{2}s\left(1-\sqrt{p}\right)+\sqrt{\frac{\mu_N}{\mu_P}}\left(\sqrt{l}-\sqrt{t}\right)\right]^2}. \tag{4.25}$$

with p=K_5/K_3, s=$1+\sqrt{K_9/K_{10}}$, t=K_5/K_6 and l=K_5/K_4.

From experimental results the temperature coefficient is 46 ppm/°C at V_{DD}=3 V. The power consumption at the minimum operating V_{DD} of 1.3 V is equal to 46.8 nW.

A very low-power current reference is proposed in [10]. In this solution a self-cascode MOSFET (SCM) is exploited to obtain a stable current. The schematic of the solution proposed in [10] is reported in Figure 4.7.

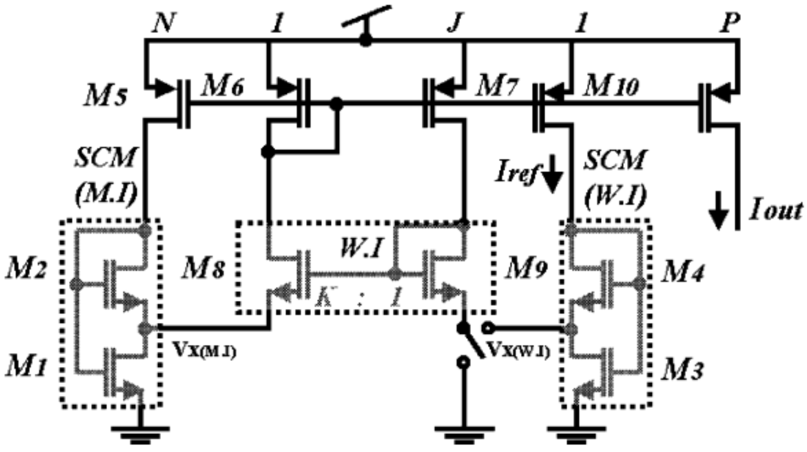


Figure 4.7. Fully-subthreshold current reference proposed in [10].

If M_8 and M_9 work in subthreshold regime a PTAT voltage shift is observed across the two devices. If the switch V_X(W,J) is connected to ground, the PTAT voltage is given by :

$$V_{GS9} = nV_T \ln(JK), \tag{4.26}$$

with J=(W_7/L_7)/(W_6/L_6) and K=(W_8/L_8)/(W_9/L_9). This topology is stable for JK>1 and it is very accurate for JK>10. If the switch is connected to M_3-M_4, then V_{S9} is given by

$$V_{S9} = V_T \ln\left(1 + (1+J)\frac{K_4}{K_3}\right). \tag{4.27}$$

This results in improved symmetry and matching of the structure. Both PTAT voltage references expressed by (4.26) and (4.27) are relatively immune to supply voltage as well as to technological parameters. In weak inversion, the SCM M_3-M_4 generates a sub-100-mV PTAT reference independent of current level and technology equal to:

$$I_{REF} = \frac{\mu_N}{2} C_{OX} n V_T^2. \tag{4.28}$$

The current (4.28) shows a severe dependence from the temperature. The results reported in [10] show that this solution is in able to work starting from 1.1 V while consuming only 2 nW. The reference current is equal to 400 pA with a TC of 2500 ppm/°C.

Finally in [11] a utra-low power current reference based on the ZTC point is proposed (Figure 4.8). In this solution all the transistors work in subthreshold regime, thus allowing a severe reduction of the power consumption. Specifically the circuit proposed in [11] generates a gate-source voltage for the MOSFET working as current source which scales linearly with temperature in order to obtain a first-order compensated current I_{REF}. Specifically the V_{GS} for the stacked transistors equal to

$$V_{GS} = V_{GS0} - k_{VGS} T, \tag{4.29}$$

Where V_{GS0} is a constant voltage and k_{VGS} is the temperature coefficient of V_{GS}. This allows obtaining a current equal to:

$$I_{REF} = \mu_N(T_0)\left(\frac{T}{T_0}\right)^{-m_N} C_{OX}\frac{W}{L}nV_T^2\exp\left(q\frac{V_{GS0}-V_{TH0}}{nk}\right)\exp\left(q\frac{k_{GS0}-\kappa_T}{nk}\right). \tag{4.30}$$

The expression (4.30) shows that the first-order dependence of the temperature is cancelled while the second-order dependence remains [11]. Measurement results show a current of 20 pA, a power consumption of 23 pW and a minimum supply voltage of 1.2 V.

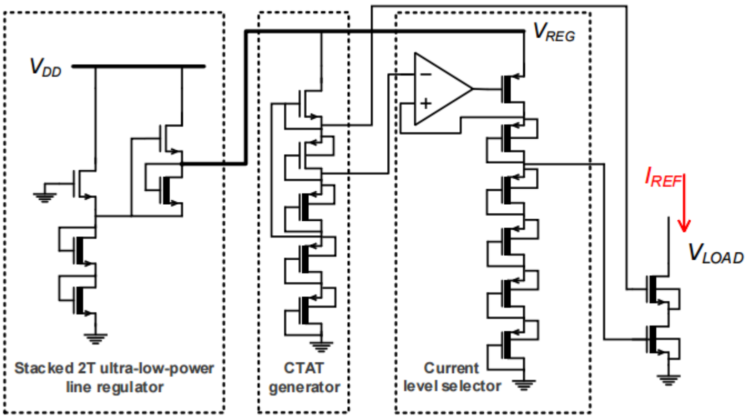


Figure 4.8. Fully-subthreshold current reference proposed in [11].

4.3. Proposed current reference

4.3.1. Introduction

In this section a new low-voltage subthreshold current reference working at 0.5 V is presented. The proposed solution was implemented in 0.18 μm CMOS technology and exploits the ZTC bias point of a MOSFET to generate a stable current. A statistical analysis over 17 samples shows a mean output current of 68.27 nA at the minimum operating voltage of V_{DD}=0.5 V. The mean power consumption at room temperature and for V_{DD}=0.5 V is 40 nW. The mean line sensitivity is 0.03 %/V while the mean temperature coefficient is 2202 ppm/°C. The coefficient of variation σ/μ is equal to 1.1 % for the output current, 66.6 % for the line sensitivity and 9.8 % for the TC. The occupied area of the proposed solution is only 0.016 mm^2.

4.3.2. Schematic and main design considerations

The schematic of the proposed current reference is reported in Figure 4.9. M_{P1} and M_{P3} are high threshold voltage transistors, M_{P2} and M_{P4} are regular threshold voltage transistos and M_{N2} and M_{N3} are regular nMOSFET transistors.

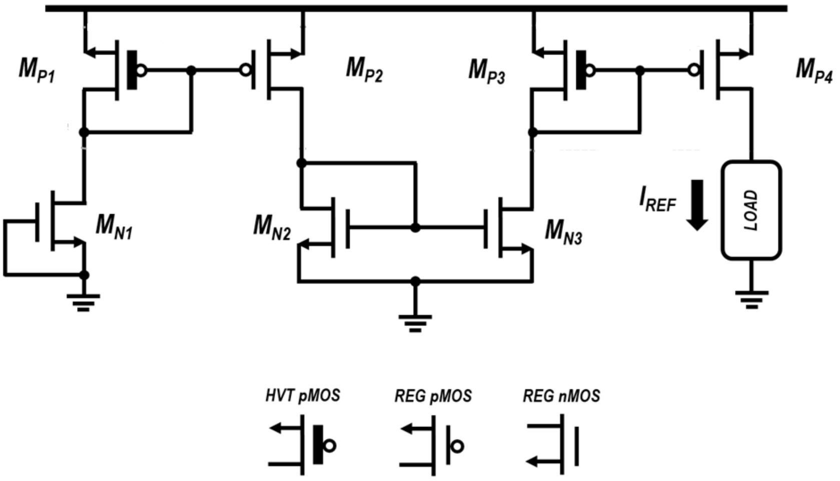


Figure 4.9. Proposed current reference.

In the proposed solution M_{P1} and M_{P2} have the same gate to source voltage. Since M_{P1} works in subthreshold and M_{P2} in above threshold regime, the following equation holds:

$$\left|V_{TH,P1}\right| + nV_T \ln\left(\frac{I_1}{\mu_P C_{OX1}\left(\frac{W}{L}\right)_{P1} V_T^2}\right) = \left|V_{TH,P2}\right| + \left(\frac{2I_2}{\mu_P C_{OX2}\left(\frac{W}{L}\right)_{P2}}\right)^{\frac{1}{\alpha}}, \qquad (4.31)$$

where α is the exponent defining the I_D-V_{GS} relationship in above threshold regime in the scaled technologies [17]. Since the current in M_{P1} is equal to the current in M_{N1}, the (4.31) becomes equal to:

$$\left|V_{TH,P1}\right| + nV_T \ln\left(\frac{\mu_N C_{OX}\left(\frac{W}{L}\right)_{N1} V_T^2 \exp\left(-\frac{V_{THN}}{nV_T}\right)}{\mu_P C_{OX1}\left(\frac{W}{L}\right)_{P1} V_T^2} \right) = \left|V_{TH,P2}\right| + \left(\frac{2I_2}{\mu_P C_{OX2}\left(\frac{W}{L}\right)_{P2}} \right)^{\frac{1}{\alpha}} . \tag{4.32}$$

Solving the (4.32) with respect to *I₂* the following expression is obtained:

$$I_2 = \frac{\mu_P C_{OX2}}{2}\left(\frac{W}{L}\right)_2 \left\{ \Delta V_{TH} + nV_T \ln\left[\left(\frac{\mu_N}{\mu_P}\right)\left(\frac{W}{L}\right)_R\right] \right\}^{\alpha} , \tag{4.33}$$

where:

$$\Delta V_{TH} = \left|V_{TH,P1}\right| - \left|V_{TH,P2}\right| - V_{TH,N1} , \tag{4.34}$$

$$\left(\frac{W}{L}\right)_R = \left(\frac{W}{L}\right)_{N1} / \left(\frac{W}{L}\right)_{P1} . \tag{4.35}$$

Introducing the temperature dependence of the threshold voltage, the (4.34) can be rewritten as:

$$\Delta V_{TH} = \left|V_{TH,P1}\right| - \left|V_{TH,P2}\right| - \left|V_{TH,N1}\right| = \Delta V_{TH}(T_0) - \Delta\kappa T , \tag{4.36}$$

with

$$\begin{aligned} &\Delta V_{TH}(T_0) = \left|V_{TH,P1}(T_0)\right| - \left|V_{TH,P2}(T_0)\right| - V_{TH,N1}(T_0) - \left(\left|\kappa_{P1}\right| + \left|\kappa_{P2}\right| + \left|\kappa_{N1}\right|\right)T_0 , \\ &\Delta\kappa = -\left|\kappa_{P1}\right| + \left|\kappa_{P2}\right| + \left|\kappa_{N1}\right| . \end{aligned} \tag{4.37}$$

In (4.37) $V_{TH}(T_0)$ is the threshold voltage at room temperature and κ the temperature coefficient for the threshold voltage. Considering the temperature dependence of the mobility, the logarithmic term in the (4.33) can be rewritten as:

$$nV_T \ln\left[\left(\frac{\mu_N}{\mu_P}\right)\left(\frac{W}{L}\right)_R\right] = nV_T \ln\left(\frac{W}{L}\right)_R + nV_T \ln\left[\frac{\mu_N(T_0)}{\mu_P(T_0)}\right] + nV_T\left(m_P - m_N\right)\ln[T] . \tag{4.38}$$

The (4.38) shows a linear dependence from temperature except for the term *Tln*(*T*). However considering the operating range of [0:80] °C, *Tln*(*T*) can be approximated with its linear regression as depicted from Figure 4.10. Thus

$$nV_T\left(m_P - m_N\right)\ln[T] \approx \frac{nk}{q}\left(m_P - m_N\right)\left(A + BT\right) = A_X + B_X T . \tag{4.39}$$

where $A_X = Ank(m_P - m_N)/q$ and $B_X = Bnk(m_P - m_N)/q$. The values of *A* and *B* can be easily extracted after evaluating the linear approximation of *Tln*(*T*) around the centre ($T = T_M$) of the considered temperature range. For the temperature range of [0:80] °C T_M is equal to 313.15 K, thus *A*=-313.15 and *B*=6.747. Using (4.39), the (4.38) can be rewritten as:

$$nV_T \ln\left[\left(\frac{\mu_N}{\mu_P}\right)\left(\frac{W}{L}\right)_R\right] = nV_T \ln\left(\frac{W}{L}\right)_R + nV_T \ln\left[\frac{\mu_N(T_0)}{\mu_P(T_0)}\right] + A_X + B_X T. \tag{4.40}$$

Replacing (4.40) in (4.33), the expression of the current I_2 becomes equal to:

$$I_2 \approx \frac{\mu_P C_{OX1}}{2}\left(\frac{W}{L}\right)_2 \left\{\Delta V_{TH}(T_0) - \Delta\kappa T + nV_T \ln\left(\frac{W}{L}\right)_R + nV_T \ln\left[\frac{\mu_N(T_0)}{\mu_P(T_0)}\right] + A_X + B_X T\right\}^{\alpha}. \tag{4.41}$$

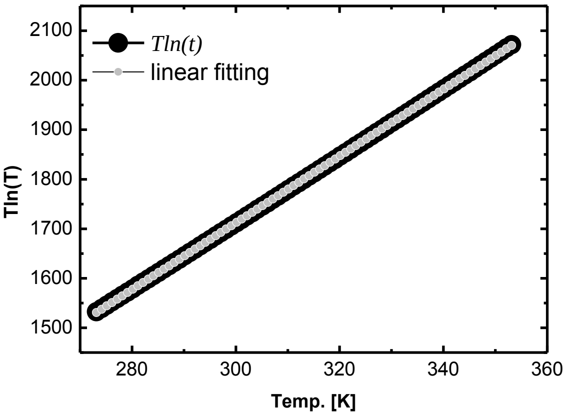


Figure 4.10. *Tln(T)* in the temperature range of [0:80] °C. The figure shows the goodness of the linear approximation.

Since the term in square brackets in (4.41) consists of a constant term and in a temperature dependent term (linearly dependent), for the sake of simplicity the (4.41) can be rewritten as:

$$I_2 = \frac{\mu_P C_{OX2}}{2}\left(\frac{W}{L}\right)_{P2} (V + CT)^{\alpha}, \tag{4.42}$$

where in (4.42) V and C are equal to:

$$V = \Delta V_{TH}(T_0) + A_X,$$

$$C = \left[\frac{nk}{q}\ln\left(\frac{W}{L}\right)_R + \frac{nk}{q}\ln\left(\frac{\mu_N(T_0)}{\mu_P(T_0)}\right) + B_X - \Delta\kappa\right]. \tag{4.43}$$

The current I_2 is mirrored through the current mirror M_{N2}-M_{N3} into M_{P3} which is a high threshold voltage transistor. As a consequence, considering the subthreshold conduction for it, the following equation holds:

$$V_{SG3} = |V_{THP3}| + nV_T \ln\left(R\frac{\frac{\mu_P C_{OX2}}{2}\left(\frac{W}{L}\right)_2 (V + CT)^{\alpha}}{\mu_P C_{OX3}\left(\frac{W}{L}\right)_3 V_T^2}\right). \tag{4.44}$$

In (4.44) *R* defines the aspect ratio of the current mirror M$_{N2}$-M$_{N3}$. After some manipulations the (4.44) can be rewritten in a more convinient form:

$$V_{SG3} = |V_{THP3}| + nV_T \ln\left(\frac{R}{2} \frac{C_{OX2}\left(\frac{W}{L}\right)_{P2}}{C_{OX3}\left(\frac{W}{L}\right)_{P3}} \right) + \alpha n V_T \ln(V + CT) - 2nV_T \ln(V_T). \tag{4.45}$$

As reported before, in the temperature range of interest the two logarithmic terms can be replaced with their linear approximations. Specifically:

$$\begin{aligned} &\alpha n V_T \ln(V + CT) \approx D + ET \\ &2nV_T \ln(V_T) \approx F + GT \end{aligned} . \tag{4.46}$$

As a consequence V_{SG3} can be rewritten in a more convient form:

$$V_{SG3} = V_3 - K_{3T}T, \tag{4.47}$$

where

$$\begin{aligned} &V_3 = |V_{THP3}(T_0)| + D - F, \\ &K_{3T} = |\kappa_{P3}| + \frac{nk}{q}\left(\frac{R}{2} \frac{C_{OX2}\left(\frac{W}{L}\right)_{P2}}{C_{OX3}\left(\frac{W}{L}\right)_{P3}} \right) + E - G. \end{aligned} \tag{4.48}$$

The source-gate voltage of M$_{P4}$ is coincident with V_{SG3}. M$_{P4}$ is regular threshold voltage transistor working in above threshold regime. Thus its ZTC voltage is equal to [8] :

$$V_{SG}(ZTC) = |V_{THP4}(T_0)| - |\kappa_{P4}| T_0. \tag{4.49}$$

If κ_{3T} is different from κ_{P4}, and $V_3 = |V_{THP4}(T_0)| - |\kappa_{P4}| T_0$, the current generated by M$_{P4}$ becomes equal to

$$I_{REF} = \frac{\mu_P(T_0) C_{OX4}}{2T_0^{-m_P}} \left(\frac{W}{L}\right)_4 (|\kappa_{P4}| - K_{3T})^{\alpha} [T]^{\alpha - mP}. \tag{4.50}$$

Thus the TC of the (4.50) is equal to the typical TC observed in the solutions in which the ZTC point is exploited:

$$TC = \frac{\alpha - m_P}{T}. \tag{4.51}$$

The (4.51) shows that also in the proposed approach the TC depends on how close m_P and α are.
The previous analysis shows the possibility of genereting the ZTC bias for the output transistor M$_{P4}$ exploited as current generator. At the same time the proposed analysis shows that the sizing of the

transistors M_{P1}-M_{P2}-M_{N1} is responsible of the fullfillment of the condition of $V_3 = |V_{THP4}(T_0)| - |\kappa_{P4}|T_0$ while the sizing of M_{P3}-M_{P4} and M_{N2}-M_{N3} allows the fulfillment of the condition $\kappa_{3T} \neq \kappa_{P4}$.

4.3.3. Experimental results

The experimental measurements have been performed at wafer-level using a Cascade SUMMIT 11861B prober equipped with a Temptronic chuck temperature controller and a Keithley 4200-SCS semiconductor parameter analyser. The circuit was implemented in UMC 0.18 μm CMOS technology. The sizes of the transistors are reported in Table 4.I. 17 samples have been successfully tested. Figure 4.11 shows the measured output current as a function of the supply voltage for a typical chip. It is worth noting that the minimum operating supply voltage is 0.5 V. This result represents so far the state-of-the-art for the minimum operating voltage of a current reference. In the inset of the same Figure 4.11 it is clear that the output current is very stable against supply voltage variations. The typical prototype of the proposed solution shows a line sensitivity of only 0.03 %/V.

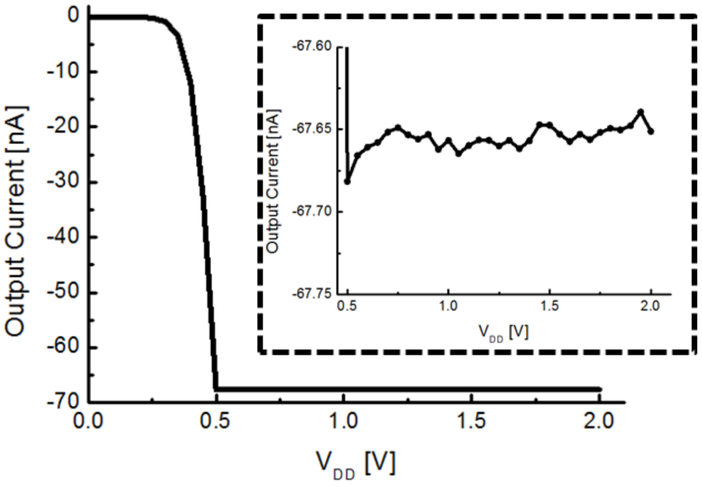


Figure 4.11. Output current as a function of supply voltage.

In Figure 4.12 the simulated current reference against supply voltage is reported. The figure shows that according to SPICE simulations the minimum supply voltage is 0.8 V with a LS of about 2 %/V. Thus better results are observed from experimental measurements for these two specifications.

TABLE 4.I. TRANSISTOR SIZES FOR THE PROPOSED CURRENT REFERENCE

Transistor	(*W* / *L*)
M_{N1}	(25μm / 50μm) **×4**
M_{P1}	(500 nm / 3μm)
M_{P2}	(1.5 μm / 50μm) **×2**
M_{N2}	(5.25 μm / 4μm) **×4**
M_{N3}	(6 μm / 25μm) **×4**
M_{P3}	(3 μm / 500nm) **×2**
M_{P4}	(500 nm / 7μm) **×2**

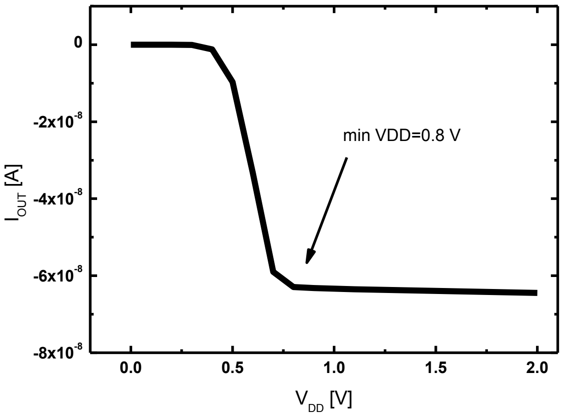


Figure 4.12. Simulated output current as a function of supply voltage.

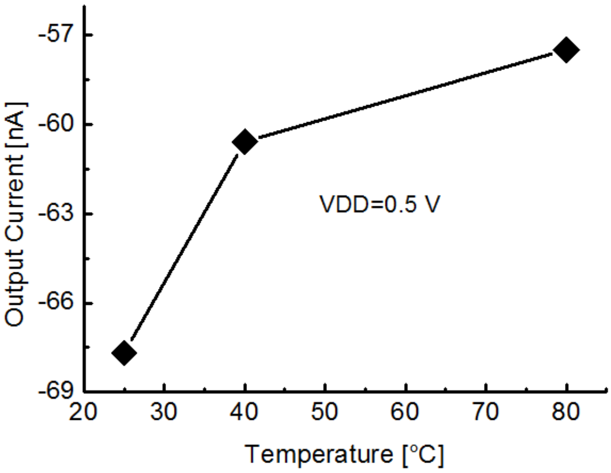


Figure 4.13. Output current against temperature for $V_{DD}=0.5\ V$.

In Figure 4.13 the temperature dependence of the output current for the minimum operating voltage of V_{DD}=0.5 V is reported. From this figure it is possible to note that the current increases at higher temperatures. The measured value of the TC, averaged over the V_{DD} ranging from 0.5 V to 2 V, is equal to 2202 ppm/°C.

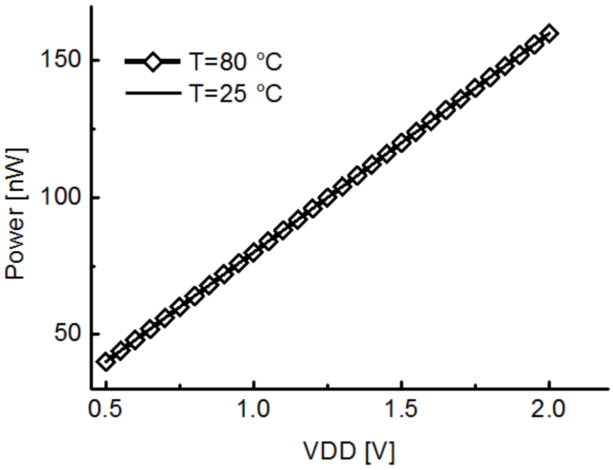


Figure 4.14. Power consumption as a function of V_{DD} at T=25°C and T=80°C.

This result is quite distant of the simulated value of TC, even in the worst case corner. Indeed, the simulated nominal value of TC is equal about 507 ppm / °C at the minimum V_{DD} of 0.8 V while a value of 1500 ppm/°C was observed in the worst case (corner SS).

The power consumption of the typical prototype is equal to 40 nW at the room temperature for the minimum operating voltage of V_{DD}=0.5 V and becomes equal to 160 nW for V_{DD}=2 V.

TABLE 4.II.
SIMULATED VALUES OF THE REFERENCE CURRENT
OVER DIFFERENT PROCESS CORNERS

Process Corner	Current
TT	27 nA
SS	22.6 nA
FF	35 nA
SF	21 nA
FS	37 nA

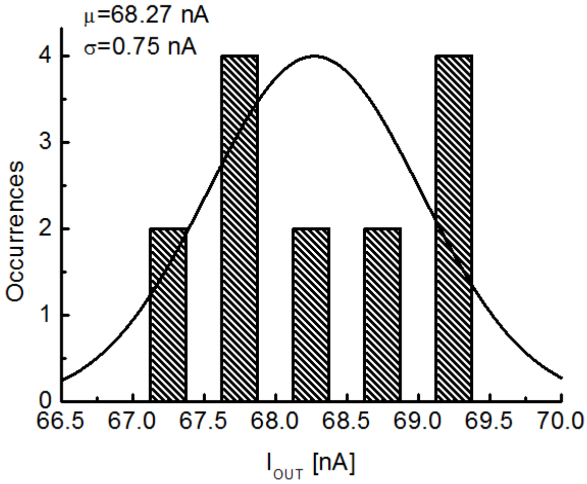


Figure 4.15(a). Statistical analysis over 17 samples: dispersion of the reference value.

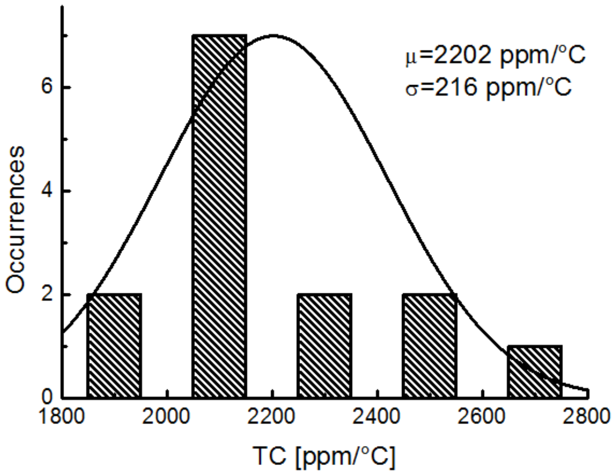


Figure 4.15(b). Statistical Analysis over 17 samples: TC dispersion over the different samples.

As shown in Figure 4.14 the power consumption doesn't show any appreciable variation in temperature. At the same time the performed measurements do not show significant variations in the reference current by changing the load value.

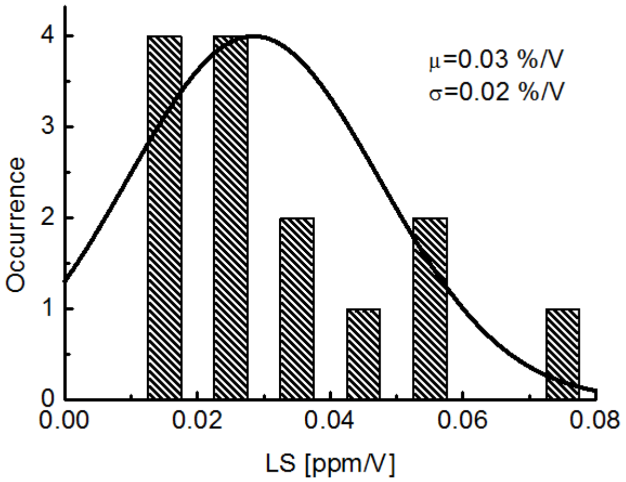


Figure 4.15(c). Statistical Analysis over 17 samples: LS dispersion over the different samples.

To investigate the effect of the process variability on the proposed solution a statistical analysis over the 17 samples have been performed. The results of such analysis are reported in Figure 4.15(a)-(c). In particular Figure 4.15(a) reports the statistical analysis on the reference value extracted at room temperature for the operating condition of V_{DD}=0.5V. The mean value of the output current is equal to 68.27 nA with a standard deviation of 0.75 nA. As a consequence the coefficient of dispersion σ/μ is only 1.1%, thus demonstrating a good stability against process variations. In Table 4.II the simulated reference current over the different process corners for the simulated minimum operating voltage of 0.75 V are reported. The corner analysis shows a σ/μ equal to 22.6 %. Moreover from the same Table 4.II it is clear the difference between the simulated and measured value of the reference current thus confirming the lack of accuracy of the SPICE models in subthreshold design.

TABLE 4.III. COMPARISON WITH THE OTHER LOW-POWER, LOW-VOLTAGE CURRENT REFERENCES

	This work	[1]	[2]	[3]	[4]	[5]	[11]
Technology	0.18 μm	0.35 μm	0.35 μm	0.35 μm	1.5 μm	2 μm	0.18 μm
I_{REF} (nA)	68.3	94.9	96	9.1	0.4	1-100	0.02
Temp. Range [°C]	[25:80]	[-20:100]	[0:80]	[0:80]	[-20:70]	[-40:80]	[-40:80]
Min V_{DD}	**0.5**	1.8	1.8	1.3	1.1	1.2	1.2
TC [ppm/°C]	2202	523	520	44	2500	1100	780
Power [nW]	40	598	1000	54.6	2	70	0.023
LS [ppm/V]	**337**	1780	2000	569	60000	100000	5800
Die Area [mm^2]	0.016	0.055	0.015	0.035	0.046	0.06	0.038

In Table 4.III the comparison with the other low-power, low-voltage current references is reported. The proposed solution shows a minimum V_{DD} of only 0.5 V which is 300 mV lower than the best solution proposed so far in this specification [6]. In terms of power consumption only [4] and [11] report better results. However it is worth noting that in these solutions the reference current is notably lower than the proposed one. On the contrary the TC is very high, only [4] shows a worst result in this specification. Despite that in the context of very low- voltage, low-power application the proposed solution is clearly the most promising. This can be easily observed in Figure 4.16 where the proposed solution is compared with the most significant publications on current references. Inspecting this figure is clear that the presented solution works in a new area of the V_{DD}-

power consumption plane. This area is particularly suitable for the energy harvesting applications or in healthcare devices which usually work in very strict conditions in terms of energy but in a limited range of operating temperatures like in the case of human body.

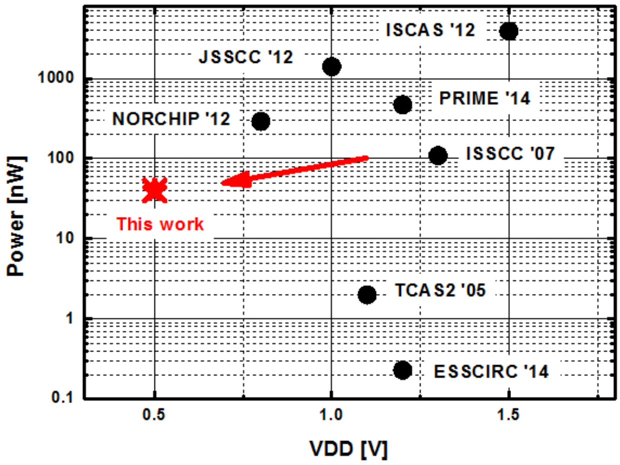


Figure 4.16. Current references in the supply voltage-power consumption plane.

4.4. Conclusion

In this chapter a new extremely low-voltage, low-power subthreshold current reference have been presented. The proposed solution implemented at silicon level using UMC 0.18 μm CMOS technology shows a minimum supply voltage of only 0.5 V while consuming 40 nW. The reference current shows a very good stability against intra-die variationsand supply voltage variations. Due to the critical matching with the ZTC bias voltage, the temperature compensation is very critical. As a consequence a trimmed solution appears necessary to achieve very low values of TC. Despite that, the minimum operating voltage and the low power consumption make the proposed scheme a very interesting solution in the low-power, low voltage applications.

Bibliography

[1] H. W. Li, R. J. Baker, D. C. Thelen. **Analog Circuits and Devices**. Chapter 3. *CRC Press LLC*. 2003.

[2] W.T. Harrison, J.A. Connelly and R. Stair, "An improved current-mode cmos voltage reference," in *IEEE Symposium on Mixed-Signal Design* (SSMSD), Feb. 2001, pp. 23-27.

[3] C. Yoo and J. Park, "CMOS current reference with supply and temperature compensation," *Electronic Letters*, Dec. 2007, vol. 46 no 25.

[4] W.M. Sansen, F. O. Eynde, M. Steyaert, "A CMOS temperaturecompensated current reference," *IEEE Journal Solid-State Circuits*, vol. 23, no. 3, pp. 821-824, Jun. 1988;

[5] K. Ueno, T. Hirose, T. Asai, Y. Amemiya, "A 300 nW, 15 ppm/°C, 20 ppm/V CMOS Voltage Reference Circuit Consisting of Subthreshold MOSFETs," *IEEE Journal Solid State Circuits*, vol. 44, no. 7, pp. 2047-2054, Jul. 2009.

[6] C.-H. Lee, H.-J. Park, "All-CMOS temperature-independent current reference," *Electronics Letters*, vol. 32, pp. 1280-1281, Jul. 1996.

[7] K. Ueno, T. Hirose, T. Asai, Y. Amemiya, "A 46-ppm/°C temperature and process compensated current reference with on-chip threshold voltage monitoring circuit," in *Proceeding of the IEEE Asian Solid-State Circuits Conference* (A-SSCC), pp. 161-164, 2008.

[8] A. Bendali, Y. Audet, "A 1-V CMOS Current Reference with Temperature and Process Compensation," *IEEE Transaction on Circuits and Systems I: Regular paper*, vol.54, no 7, Jul 2007.

[9] G. De Vita and G. Iannaccone, "A 109 nW, 44 ppm/°C CMOS current reference with low sensitivity to process variations," in *Proceedings of the IEEE International Symposium on Circuits and Systems* (ISCAS), May 2007, pp. 3804-3807.

[10] E.M. Camacho-Galeano, C. Galup-Montoro, M.C. Schneider , "A 2-nW 1.1 V self-biased current reference in CMOS technology", *IEEE Transaction on Circuits Syst. II, Exp. Briefs*, vol. 52, no. 2, pp. 61–65, Feb. 2005.

[11] M. Choi, I. Lee, T.-K. Jang, D. Blaauw, D. Sylvester, "A 23pW, 780ppm/°C Resistor-less Current Reference Using Subthreshold MOSFETs," in *IEEE European Solid State Circuits Conference* (ESSCIRC), pp. 119-122, Oct. 2014.

[12] F. Cucchi, S. Di Pascoli, G. Iannaccone, "Variability Aware Design of 55 nA current reference with 1.4% standard deviation and 290 nW power consumption" *in Proceedings of NORCHIP Conference*, Nov. 2012.

[13] H. J. Oguey and D. Aebischer, "CMOS current reference without resistance," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 7, pp. 1132-1135, Jul. 1997.

[14] J. Lee; S.-H. Cho, "A 1.4-µW 24.9-ppm/°C Current Reference with Process-Insensitive Temperature Compensation in 0.18-µm CMOS," *IEEE Journal of Solid-State Circuits*, vol.47, no.10, pp.2527-2533, Oct. 2012.

[15] S.S. Chouhan, K. Halonen, "A modified CMOS nano-power resistorless current reference circuit," in *IEEE Conference on PhD Research in Microelectronics and Electronics* (PRIME), pp. 1-4, June 2014.

[16] C. Quemada, T.L. Cochran, H. D. Sam, "A compact resistorless 1.5-V CMOS current reference with 16.5-ppm/°C temperature coefficient, " in I*EEE International Symposium on Circuits and Systems* (ISCAS), pp. 3146 – 3149, 2012.

[17] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFETs: Part I-Effects of substrate impurity concentration," in *IEEE Transaction on Electron Devices*, vol. 41, no. 12, pp. 2357–2362, Dec. 1994.

5. Design of a low Power SAR Analog to Digital Converter

5.1. Introduction

The Analog to Digital converter (ADC) represents the fundamental block in the connection between the real world, purely analog, and the world of the computers, purely digital. Specifically the A/D converter transforms a continuous-time, continuous-value signal into a discrete-time, discrete-value signal. The conversion is partially reversible under opportune conditions.
The block diagram of the AD conversion is reported in Figure 5.1 [1].

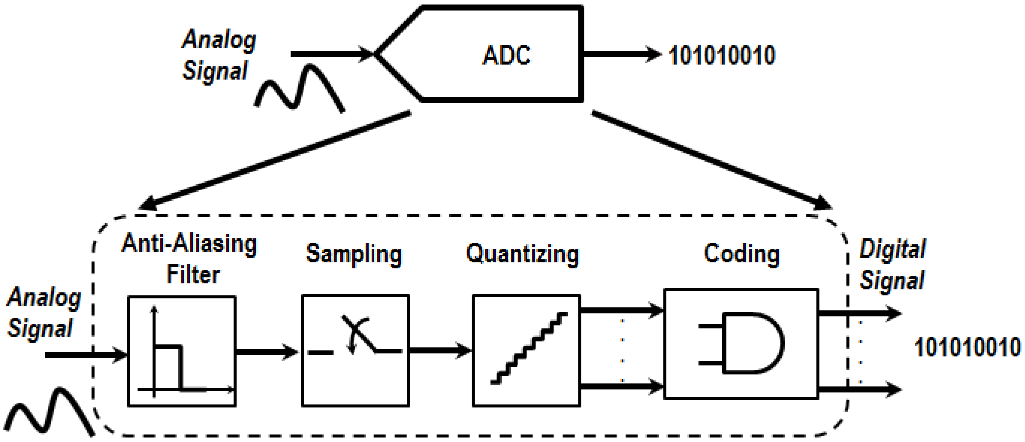


Figure 5.1. A/D conversion chain.

The conversion from an infinite number of values into a finite number of values is usually performed by sampling the signal with a fixed time step T_S (or with a sampling frequency of $f_S=1/T_S$). As result of the sampling, the continuous time signal $v(t)$ is converted in a discrete time signal expressed as:

$$w(nT_S)=\sum_{n=-\infty}^{+\infty}v(t)\delta(t-nT_S). \tag{5.1}$$

The signal reported in the (5.1) is a discrete-time signal that at each sampling instant nT_S produces a Dirac function with weight equal to the sampled value. The main problem in the sampling process consists in finding a value of T_S which allows obtaining the signal $v(t)$ starting from $w(nT_S)$ without losing any information. The value of T_S that allows the reconstruction of the input signal can be found by considering the frequency domain. The Fourier transform of the signal $w(nT_S)$ is equal to:

$$F(\omega)=\frac{1}{T_S}\sum_{n=-\infty}^{+\infty}F(\omega-n\omega_S), \qquad (5.2)$$

thus, from (5.2), the spectrum of the sampled signal consists in an infinite number of replicas of the spectrum of the original signal. Specifically the replicas are located at the frequencies $2n\pi/T_S$ and scaled in amplitude by a factor $1/T_S$. In order to reconstruct the signal without any loss of information, the replicas have to be separated each other without any overlap. If ω_B represents the bandwidth of the sampled signal, the reconstruction if performed without any overlap if the following condition is satisfied:

$$\omega_S \geq 2\omega_B\,. \qquad (5.3)$$

The (5.3) is also known as Shannon's law. To better understand the condition expressed by the Shannon's law, Figure 5.2 shows the spectrum of the sampled signal in the condition of $\omega_S=2\omega_B$ (Figure 5.2(a)) and $\omega_S< 2\omega_B$ (Figure 5.2(b)). In the latter condition the replicas are overlapped each other generating the *aliasing effect*. As a consequence the original signal cannot be reconstructed.

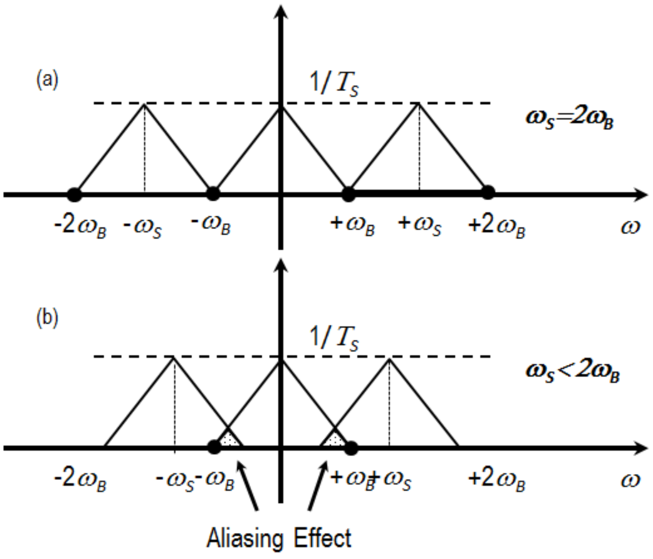


Figure 5.2. Spectrum of sampled signal in the case of $\omega_S=2\omega_B$ (a) and $\omega_S<2\omega_B$ (b).

Based on (5.3), the sampling frequency must be at least twice the maximum frequency of the spectrum of the sampled signal. For this reason, before sampling, the signal is filtered by an Anti-Aliasing-Filter (AAF). The AAF can be either a low-pass or a band-pass filter. The idea is to remove all the spurious components at frequencies higher or equal to the half of the sampling frequency in order to satisfy the Shannon's law, thus ensuring the reconstruction of the analog signal. It is worth noting that the assumption of an ideal sampling frequency is not typically true in a real system. In fact the sampling time can differ because of the *jitter* in the clock [2]. This can reduce drastically the performance of the ADC. The Signal-to-Noise Ratio (*SNR*) due to the jitter in the sampling clock is equal to [2]:

$$SNR_{jitter}=-20\log(2\pi f_{sign}t), \qquad (5.4)$$

where f_{sign} is the signal frequency and t is the jitter time. Thus, with a fixed amount of clock jitter, the *SNR* degrades as the input frequency increases. Another factor that can reduce the ADC resolution is the thermal noise (kT/C) [1]. The problem in this case is related to the switch used to perform the sampling phase. Figure 5.3 describes the equivalent circuit in the sampling phase. The input voltage charges the load capacitor through the switch which offers a resistance R_S during its sampling phase.

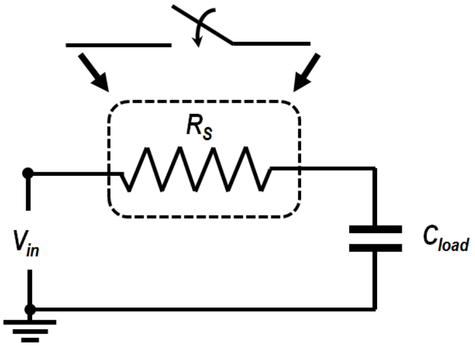


Figure 5.3. Charge model during the sampling phase.

In order to perform correctly the sampling, the time constant $\tau_S=R_SC_{load}$ should be negligible compared to the variation of the input, furthermore the bandwidth of the signal to be sampled must be much lower than $1/\tau_S$. The power spectral noise density (PSD) of the thermal noise of R_S is $v^2_{n,R_S} = 4kTR_S$. Since the entire system acts as a low-pass filter, the following equation describes the power spectral noise of the sampling capacitor [1]:

$$v_n^2(sampling) = \frac{4kTR_S}{1+\left(\omega R_S C_{load}\right)^2} . \tag{5.5}$$

When the switch is open, C_{load} holds both the sampled input voltage as well as the noise. Since the cut-off frequency of the R_SC_{load} filter is much higher than the Nyquist frequency ($2\omega_B$), the total power noise stored on C_{load} has a white spectrum given by all the folded bands into the base-band [1]. The total power noise in the load capacitor is then equal to:

$$P_n = \int_0^\infty v_n^2(sampling)df = \int_0^\infty \frac{4kTR_S}{1+\left(\omega R_S C_{load}\right)^2} df = \frac{kT}{C_{load}} . \tag{5.6}$$

As shown in the (5.6) the power noise does not depend on the value of the resistance offered by the switch. In fact if R_S increases the corner frequency decreases as well, compensating the variation.
After the sampling phase, the signal is quantized. In the quantization process the generic interval [-V/2;+V/2] is divided in l equal levels (uniform quantization) with amplitude equal to V/l. The uniform interval $\Delta=V/l$ is called quantization step. Using $l=2^n$ intervals, n bits can be used to identify all the quantization intervals. Since all the values that fall in the same quantization interval are approximated with the same binary code, the quantization introduces intrinsically an error which is not reversible. If Q define the quantization function, then the quantization error is evaluated as $e=Q(x)-x$, where x is the sampled value. The quantization error can be classified in *overload error* if $x< -V/2$ or $x> +V/2$ or in *granular error* if $-+V/2\leq x\leq +V/2$.

The signal to quantization noise ratio *SNRQ* expressed in decibels is equal to:

$$SNRQ = 10\log\left(\frac{P_{signal}}{P_{noise}}\right). \tag{5.7}$$

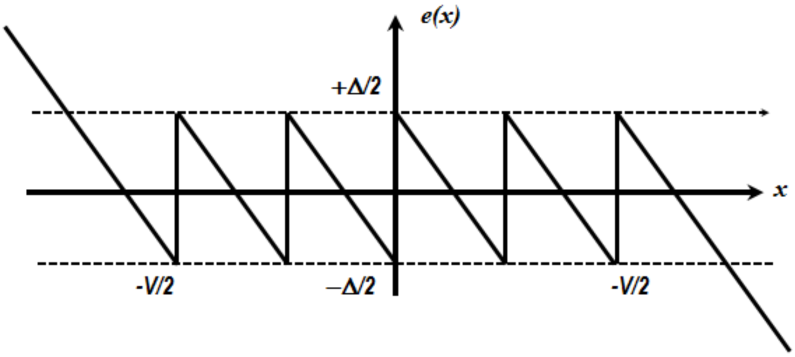


Figure 5.4. Error function in the quantization process.

For a uniform quantizer, the error has the distribution reported in Figure 5.4, thus the probability distribution of the error *e* is equal to:

$$p(e) = \begin{cases} \frac{1}{\Delta} if & e \in \left[-\frac{\Delta}{2};+\frac{\Delta}{2}\right] \\ 0 & otherwise \end{cases}. \tag{5.8}$$

The power of *e* is given by

$$P = \int_{-\infty}^{+\infty} e^2 p(e) de = \int_{-\Delta/2}^{+\Delta/2} \frac{e^2}{\Delta} de = \frac{\Delta^2}{12}. \tag{5.9}$$

Considering a sine wave as input signal *v*(*t*)=(*A/2*)*sin*(*2πft*), the corresponding power is equal to

$$P_{\sin e} = \frac{1}{T}\int_0^T \frac{A^2}{4}\sin^2(2\pi f t)dt = \frac{(\Delta \cdot 2^n)^2}{8}. \tag{5.10}$$

In (5.10) *n* is the number of bits used in the quantization process. Thus combining (5.9) and (5.10) with the (5.7), the well-known expression of the *SNRQ* is obtained [1]:

$$SNRQ = 6.02n + 20\log\left(\frac{A}{V}\right) + 1.76dB. \tag{5.11}$$

From (5.11), the *SNRQ* ratio is improved if the signal uses all the available range of the quantizer (*A*≈*V*) and an improvement of 6.02 dB is obtained for each bit added to the AD conversion.

The last operation performed by the ADC is the coding. In this phase a binary code is associated to the signal depending on the quantization interval in which the sampled signal value falls. Different codes can be used to generate the digital word [1]:

- ✓ ***Thermometric***. It uses a set of (2^n−1) binary levels to represent *n* bits.
- ✓ ***Unipolar Straight Binary*** (USB). It uses the code 00…0 to represent the last level of quantization and increases by one for each quantization level until the last code (111…1).

- ✓ ***Complementary Straight Binary*** (CSB). It is the opposite of USB coding. It represent the full scale code (maximum input level) with 0...000 code and the first quantization level (minimum input value) with 1...111 code.
- ✓ ***Binary Two's Complement*** (BTC). This coding scheme allows performing subtractions. The most significant bit (MSB) indicates the sign (0 for positive inputs and 1 for negative inputs) while the others bits indicate the quantization level.

5.2. ADC specifications

The ADCs can be classified into two main categories: *Nyquist rate* ADC or *Oversampling* ADC. In the case of the Nyquist rate converters the maximum achievable bandwidth of the signal is half the sampling frequency in order to respect the Nyquist condition, while in the case of the oversampling ADCs the frequency of the input signal is much lower than the sampling frequency. Both ADCs categories are evaluated using the same specifications which can be applied to all the architectures. The first specification which defines the performances of an ADC is the **resolution**. It is defined as the number of bits that the ADC uses to represent the input signal. Generally the Nyquist rate ADCs can reach medium resolution (up to14 bits), while the oversampled ADCs can achieve very high resolutions (18 up to 30 bits). The other specifications which are commonly used to describe the ADC can be classified in **static** and **dynamic** performances [1]. In the following these performances are explained in detail.

5.2.1. Static performances

The main static performances of an ADC are the *analog resolution*, *analog input signal*, *offset*, *gain error*, *differential non linearity* (*DNL*), *integral non linearity* (*INL*) and *temperature drift* [1]. Specifically:

- ✓ *analog resolution* defines the smallest variation in the analog input which generates a variation of 1 *LSB* in the output code. It is defined as:

$$LSB = \frac{V_{MAX} - V_{MIN}}{2^n}, \tag{5.12}$$

where V_{MAX}-V_{MIN} defines the input range while n is the number of bits.

- ✓ *analog input signal* defines the peak-to-peak variation in the input signal manageable from the ADC.
- ✓ *offset* is defined as the difference between the ideal and real input signal values to get a null output signal. It is typically expressed in Volts, *LSB* or percentage of the *LSB*. For an ideal data converter, the first transition occurs at 0.5 *LSB* above zero; however in a real ADC this transition can occurs before or after this value. This error affects systematically all the quantization steps by the same quantity. The graphical interpretation of the offset error is reported in Figure 5.5(a).
- ✓ *gain error* measure the error in the slope of the straight line interpolating the transfer curve. Ideally the slope of this curve is equal to 1. Figure 5.5(b) describes such an error.

✓ *differential non linearity* (*DNL*) provides a measure of how far a code is with respect to its neighbors code [4]. For an ideal ADC, in which the differential nonlinearity coincides with *DNL*=0 *LSB*, each analog step has a width Δ_i=1*LSB*. As a consequence all the codes are equally spaced with a distance equal to 1 *LSB*. Indicating with x_k the transition point between two successive codes in the real ADC, the width of the k^{th} step is equal to $\Delta_k=(x_{k+1}-x_k)$. Thus the *DNL* of the k^{th} step is equal to:

$$DNL(k) = \frac{\Delta_k - \Delta_i}{\Delta_i}. \tag{5.13}$$

DNL is usually expressed in *LSB*. A *DNL* error lower than ± 1 *LSB* means no missing code while a *DNL* equal to ± 1 *LSB* does not necessarily guaranteed to have no missing codes. In Figure 5.5(c) the *DNL* error is reported.

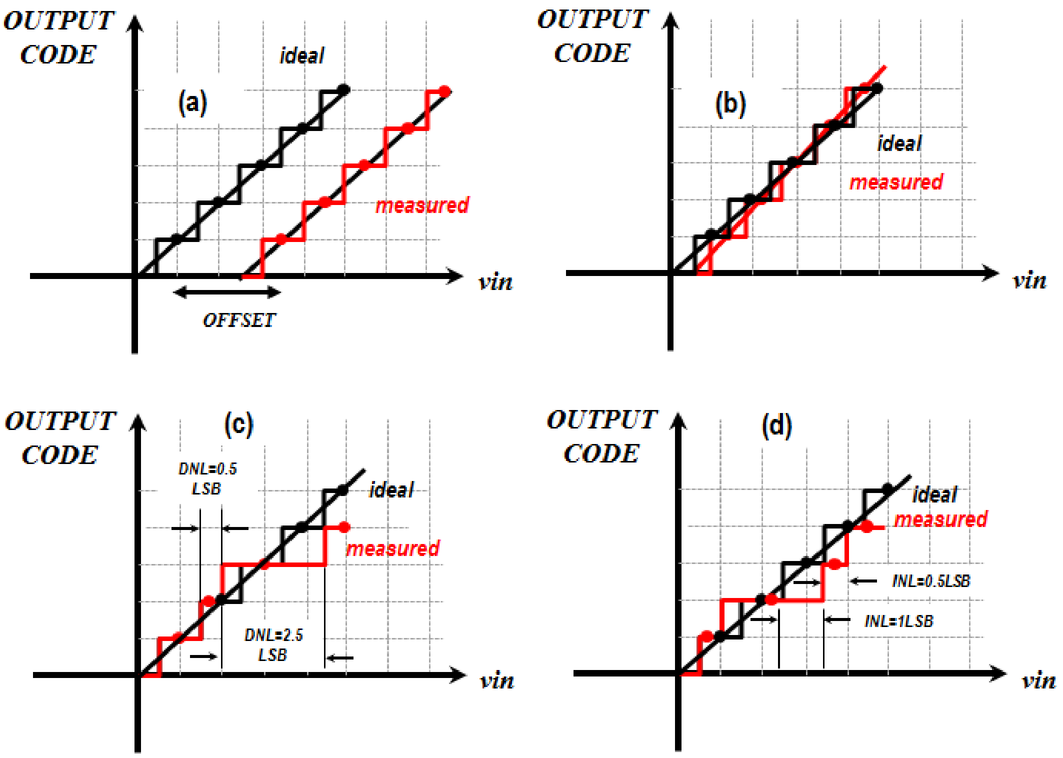


Figure 5.5. Main DC performance for an ADC: offset (a), gain error (b), *DNL* (c) and *INL* (d).

✓ *Integral Non-Linearity* (*INL*) is defined as the integral of the *DNL* error. *INL* error is described as the deviation, in *LSB* or percent of full-scale range (FSR), of a measured transfer function from a straight line [3]. The *INL*-error magnitude depends on the method chosen for the straight line evaluation. Two definitions are commonly used: the best straight-line *INL* and the end-point *INL*. The best straight-line determines, in the form of a straight line, the closest approximation to the ADC’s actual transfer function while the end-point *INL* determines the straight line through end points of the converter’s transfer function. For an *n*-bit ADC the end-point line is defined by its zero (all zeros) and its full-scale (all ones) outputs [5]. Between the two approaches, best straight-line is usually preferred. The name integral nonlinearity derives from the fact that the sum of the values of the *DNL* from step 0 to step *k*, determines the value of the *INL* for step *k*. In particular, considering the evaluation of the *INL* which takes into account the ideal slope equal to 1, the following definition holds [2]:

$$INL(k)=\sum_{j=0}^{k} DNL(k)\,. \tag{5.14}$$

The graphical interpretation of the *INL* error is reported in Figure 5.5(d).

- ✓ *temperature drift.* The temperature drift affects the performance of an ADC converter based on resolution. As an example for a 12-bit converter to maintain accuracy over the extended temperature range (-40°C to +85°C), the drift in the reference voltages must be at maximum 4 ppm/°C [1]. Thus it is usual for an ADC to define its resolution at different temperatures.

5.2.2. Dynamic performances

Frequently an ADC performs well when operates in DC. However if the *DNL* and *INL* meet the requirements this does not mean that the same performance can be achieved when an AC signal is applied. For this reason an ADC is evaluated also by considering its dynamic behaviour.

The main AC performances for an ADC are the *signal-to-noise ratio* (*SNR*), *signal-to-noise and distortion ratio* (*SINAD*), *total harmonic distortion* (*THD*) and *spurious-free dynamic range* (*SFDR*) [6]. In particular:

- ✓ *signal-to-noise ratio SNR* reveals where the noise floor of the converter is. It is defined as the ratio between the power of the signal (normally sinusoidal) and the power of quantization noise and circuit noise.
- ✓ *signal-to-noise and distortion ratio* (*SINAD*) is defined as is the ratio between the root-sum-square (*rms*) signal amplitude and the mean value of the root-sum-square (rss) of all other spectral components, including harmonics, but excluding the DC components. *SINAD* is often plotted for various input amplitudes and frequencies [6].
- ✓ *Effective Number of Bits* (*ENOB*) represents the effective resolution of the ADC considering the *SINAD* parameter. For a sinusoidal input signal the *ENOB* is given by

$$ENOB=\frac{SINAD(dB)-1.76}{6.02}; \tag{5.15}$$

- ✓ *total harmonic distortion* (*THD*) is the ratio of the *rms* value of the fundamental signal and the mean value of the root-sum-square of its harmonics plus all noise components excluding the DC component. The bandwidth over which the noise is measured must be specified.
- ✓ *spurious-free dynamic range* (*SFDR*) is the difference between the amplitude of the signal and the highest spurious spectral component in the first Nyquist bandwidth.

Usually the Figure-of-Merit (*FOM*) used to perform the comparison between different ADCs is defined as

$$FOM=\frac{power}{2BW\,2^{ENOB}}, \tag{5.16}$$

where in (5.16) *power* is the power consumption of the ADC while *ENOB* and *BW* are the effective number of bits and the bandwidth of the input signal, respectively.

5.3. ADC architectures

In this section a brief overview on the main architectures used in the AD conversion are reported [7]-[9]. For each solution the main performances, advantages and drawbacks are reported.

5.3.1. Flash ADC

The flash ADC represents the simplest and the fastest architecture for an ADC. Here the analog signal is compared with 2^n reference voltages generated in a resistive path consisting of 2^n well matched resistances. Every single voltage generated in the resistive path is compared with the input signal trough a dedicated comparator, thus 2^n-1 comparators are used to obtain an *n* bit conversion.

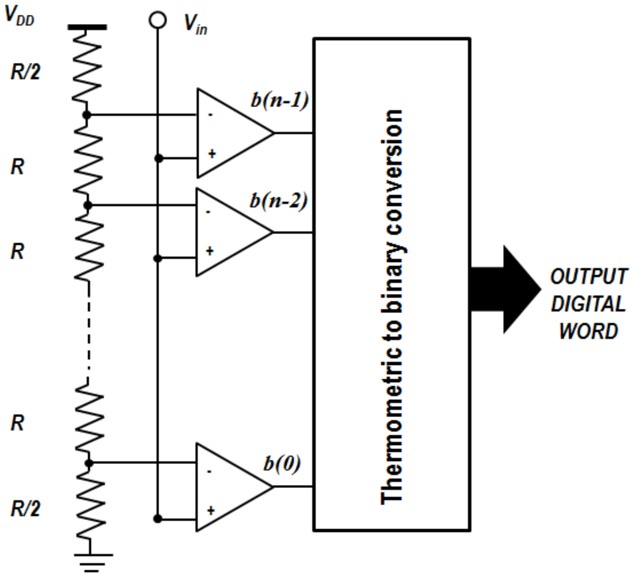


Figure 5.6. Flash ADC architecture.

The comparators generate a thermometric code which is converted in a binary code by an opportune logic network. The architecture of a Flash ADC is reported in Figure 5.6. Flash ADCs can achieve very high speed (gigasample per second) since only one clock period is sufficient for the comparison and the generation of the digital output code. Despite that this architecture suffers from many drawbacks. One of the main drawbacks consists in the high number of components which increases exponentially at higher *n*. Moreover this solution is very expensive in terms of power consumption and the matching between the different resistances effects drastically the resolution of the ADC. Because of this the Flash ADC is used in contexts in which the high speed conversion is the first requirement. Typically Flash ADCs have a resolution around 6 bits.

5.3.2. Pipeline ADC

In the pipeline ADC the input signal is converter in a binary output through successive steps. Specifically a pipelined ADC employs a parallel structure in which each stage works on few bits concurrently. While the first stage is converting the input signal, the second stage is converting the signal already elaborated by the first stage in the previous clock time. This process is repeated for all the stages. Figure 5.7 reassumes the architecture of a pipeline ADC. Each single stage generates *n* bits which together generates the output word. Every stage in the chain converts the residual coming from the previous stage. The residual of each stage is obtained by subtracting the input with the signal obtained by a D/A converter which receives as input the signal converted in the same stage. Thus the residual is simply the quantization error of the conversion stage. Apart an initial latency of *k* clock periods, the pipeline is in able to produce a new conversion every clock period. Pipelined ADC is used for sampling rates from a few megasamples per second (Msps) up to 100Msps. The resolution ranges from 8 bits at the faster sample rates up to 16 bits at the lower rates.

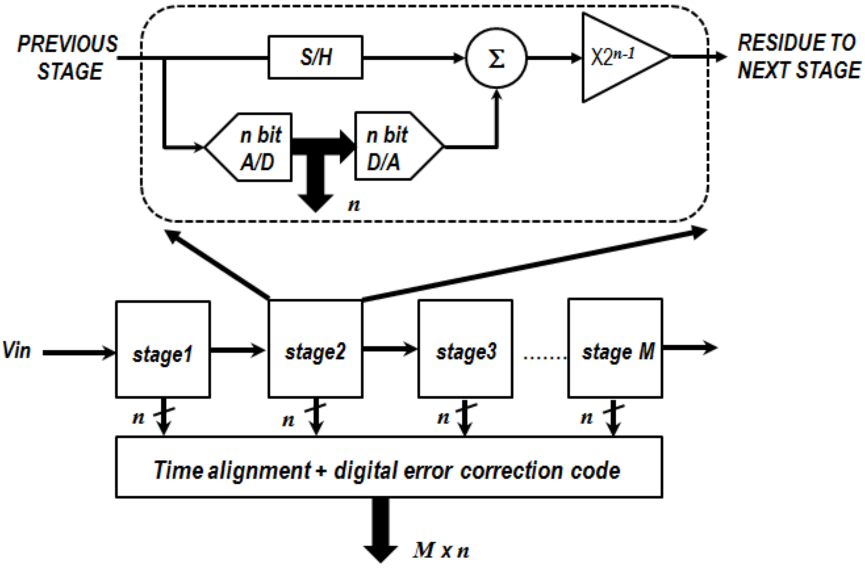


Figure 5.7. Pipeline ADC architecture.

5.3.3. Sigma-Delta (ΣΔ) ADC

The ΣΔ ADC falls in the category of the oversampling ADCs.

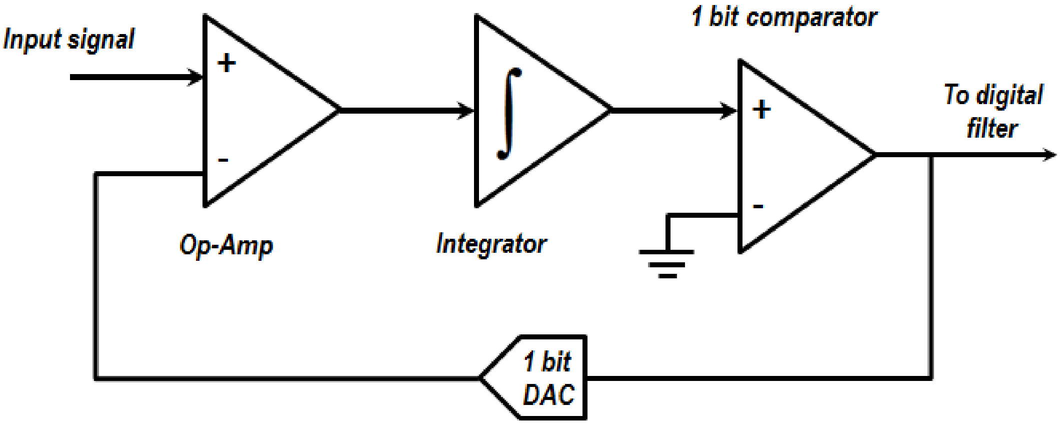


Figure 5.8. First order sigma-delta ADC.

The key advantage of oversampling is that the signal occupies a small portion of the Nyquist band, thus if a digital filter is used after the A/D conversion, the noise in the signal bandwidth can be notably reduced. The oversampling reduces the floor noise because it becomes distributed over a wider range of frequencies. A ΣΔ ADC converter exploits this effect by following the 1-bit ADC with a digital filter (Figure 5.8). This action enables ΣΔ converters to achieve wide dynamic range from a low-resolution ADC. Indeed since the *SNR* for a 1-bit ADC is 7.78dB (6.02 + 1.76), an oversampling factor equal to 4 increases the *SNR* by 6 dB which is equivalent to gaining one bit. It is obvious that such a technique cannot be used to achieve very high resolution because of the limits in the oversampling. Thus other techniques like *noise shaping* are exploited to obtain higher resolutions. The main point of the noise shaping consists in changing the power noise distribution over frequencies. Here most of the components are pushed at higher frequencies compared to the signal. The noise components are then removed through a digital filter and finally the output data rate can be reduced (*decimation*) back to the original sampling rate. This technique allows improving the *SNR* more than 6 dB by doubling the sample rate. Figure 5.9 shows the oversampling, digital filtering and the decimation processes [8].

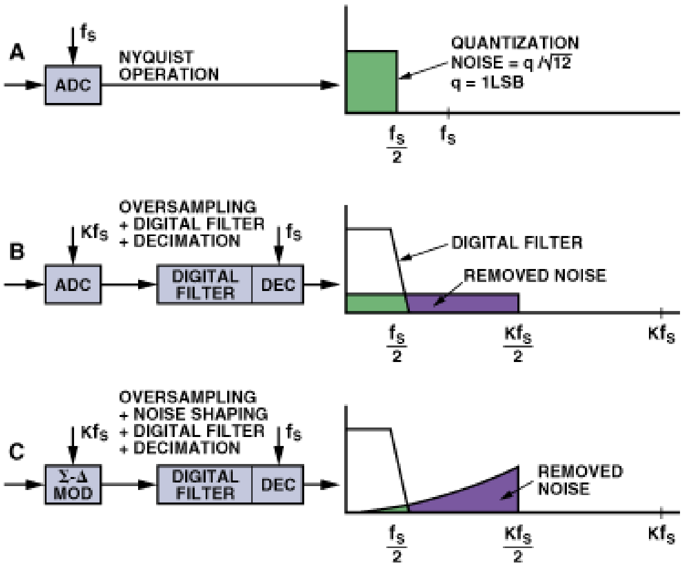


Figure 5.9. Oversampling, digital filtering, noise shaping and decimation in Σ-Δ ADC [8].

5.3.4. Successive Approximation Register (SAR) ADC

The basic architecture of a SAR ADC is reported in Figure 5.10. The input voltage (V_{IN}) is held on a sample and hold circuit. The *n*-bit register is first set to midscale (100... .00 where the *MSB* is set to 1).This forces the DAC output (V_{DAC}) to be $V_{REF}/2$, where V_{REF} is the reference voltage provided to the ADC. The comparator determines if V_{in} is less than, or greater than, V_{DAC}. If V_{in} is greater than V_{DAC}, the comparator output is equal to logic 1 (V_{DD}), and the MSB of the *n*-bit register remains at 1. On the other hand, if V_{IN} is less than V_{DAC}, the comparator output is equal to logic 0 (GND) and the MSB of the register is cleared to logic 0. The SAR control logic then moves to the next bit down, forces that bit high, and starts with another comparison. This procedure continues until the LSB, thus *n+1* clock cycles are necessary to achieve an *n* bit output. After that the SAR register generates an End-Of-Conversion (EOC) signal which indicates the end of the conversion for the sampled value.

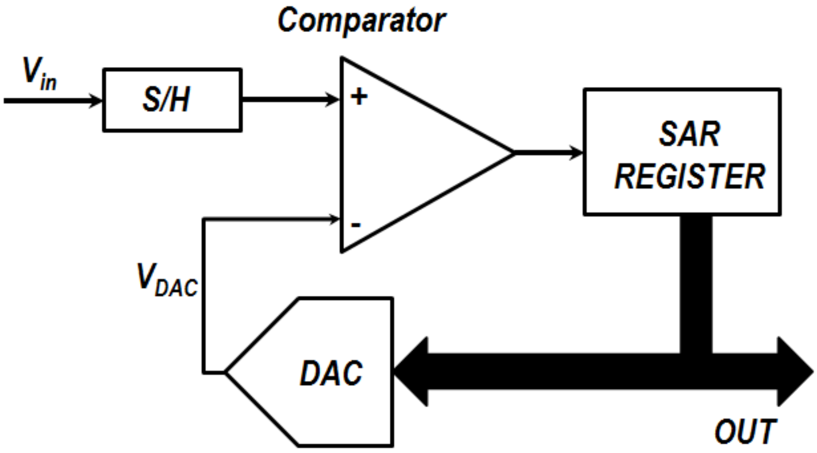


Figure 5.10. SAR ADC architecture.

The performances of the SAR ADC are generally limited by the comparator and the matching between the elements in the DAC. As a consequence of these effects the resolution offered by a SAR ADC is typically lower than 12 bits while the maximum speed is typically equal to 10 MS/s-100 MS/s.

5.3.5. ADCs comparison

Figure 5.11 reports the comparison among the different ADC architectures in terms of resolution and sampling rate. As reported before flash ADC is extremely power-hungry and nowadays used only in some applications or as part of other converters. Indeed for flash converters, every bit increase in resolution almost doubles the size of the ADC core as well as the power. In contrast, a SAR, pipelined, or sigma-delta ADC die size will increase linearly with resolution.

Inspecting Figure 5.11, sigma-delta converters allow obtaining very high resolutions on low frequency signals while pipeline ADCs show medium resolution but in high-speed applications. On the contrary SAR ADCs achieve medium resolution in a medium range of sampling rates.

TABLE 5.I. COMPARISON BETWEEN THE DIFFERENT ADC ARCHITECTURES

ADC	Conversion frequency	resolution	Advantages/drawbacks
ΣΔ	< 4Ksps	< 31 bit	High Resolution Moderate-cost
	< 4Msps	< 24 bit	
	< 10Msps	< 16 bit	
SAR	< 4Msps	< 16 bit	Low-cost Low-power
	< 1.25Msps	< 18 bit	
Pipeline	< 200 Msps	< 16	Fast, expensive, higher power requirements
	< 250 Msps	< 14	
	< 550 Msps	< 12	

A detailed comparison on the performances of the different ADCs is reported in Table 5.I [9]. In Figure 5.12 the energy-per-step-conversion against the Nyquist sampling rate for the different ADCs is reported. The graph is extracted from the 2015 ISSCC tech trends report [10].

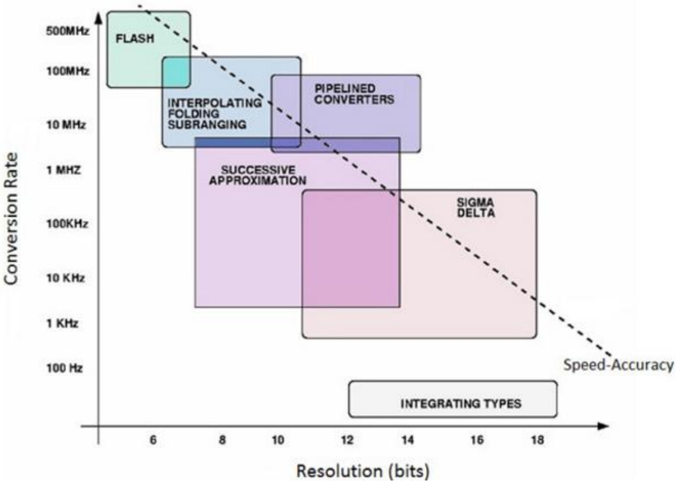


Figure 5.11. Conversion rate and resolution for the different ADC architectures.

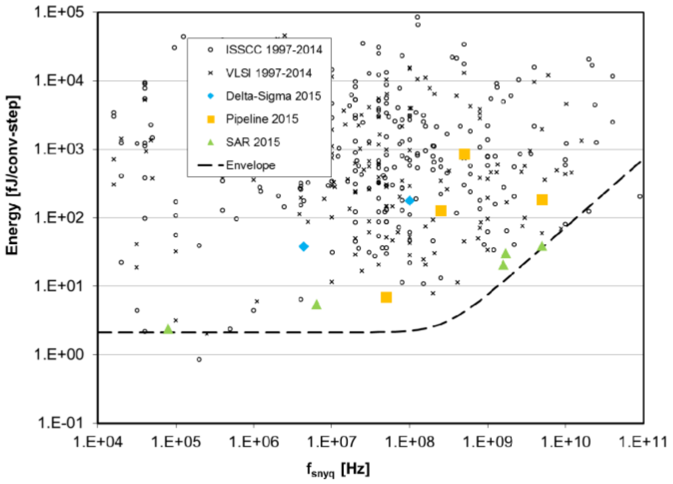


Figure 5.12. Energy vs Nyquist sampling frequency in ADCs [10]. The figure refers to the works published at the international solid-state conference (ISSCC) and international conference on very large scale integration (VLSI) systems.

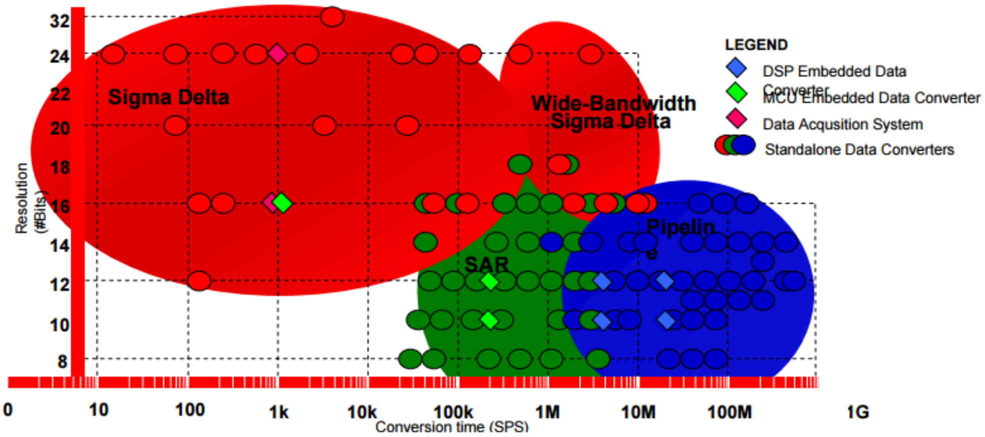


Figure 5.13. Applications of the different ADC architectures [5].

Also here the SAR converter appears as the best solution in terms of power-performances trade-off. Finally as last comparison, Figure 5.13 reports the application contexts of the different architectures. The ΣΔ and the SAR converter are the most popular choices in embedded systems while pipeline is commonly used in standalone applications due to the higher power consumption. The choice between SAR and ΣΔ mainly depends on the requirements in terms of resolution, power consumption and speed of the input signal, however if power is the main concern of the design, SAR becomes the only possible solution to meet the requirement of moderate speed, moderate resolution and very low-power consumption.

5.4. Proposed SAR ADC

5.4.1. Introduction

As reported in the previous section successive approximation analog-to-digital converter is becoming a popular solution in the field of low power applications such as implantable medical devices, portable instruments and generally in battery-powered applications or self-supplied devices [11]-[13]. Thanks to its architecture, such a component is in able to achieve low-to-medium resolution (8-12 bit) and medium speed (10 MS/s-100 MS/s) by consuming a limited amount of energy. In this section a high-resolution, low power SAR ADC is presented. The proposed scheme, implemented in 0.35 μm CMOS technology consists in an input buffer, a high resolution comparator, a logic control circuit and a capacitive DAC. Simulations carried out in worst case matching between the capacitance elements of the DAC and by considering intra- and inter-die variations have shown an effective number of bits equal to 11.97 and a SNDR of 73.8 dB. The simulated DNL performance is –0.1 ÷ +0.1 LSB while the INL is equal to –0.14 ÷ + 0.14 LSB. The mean power consumption is 0.27 mW while the ADC FoM is only 87 fJ/step. Silicon measurements have been performed as well. Without calibration bank the designed ADC shows a DNL of ±4 LSB and an ENOB equal to 9 bits while confirming the simulated values of power consumption and FoM.

5.4.2. Architecture and main design considerations

The complete block diagram of the proposed SAR ADC converter is reported in Figure 5.14.

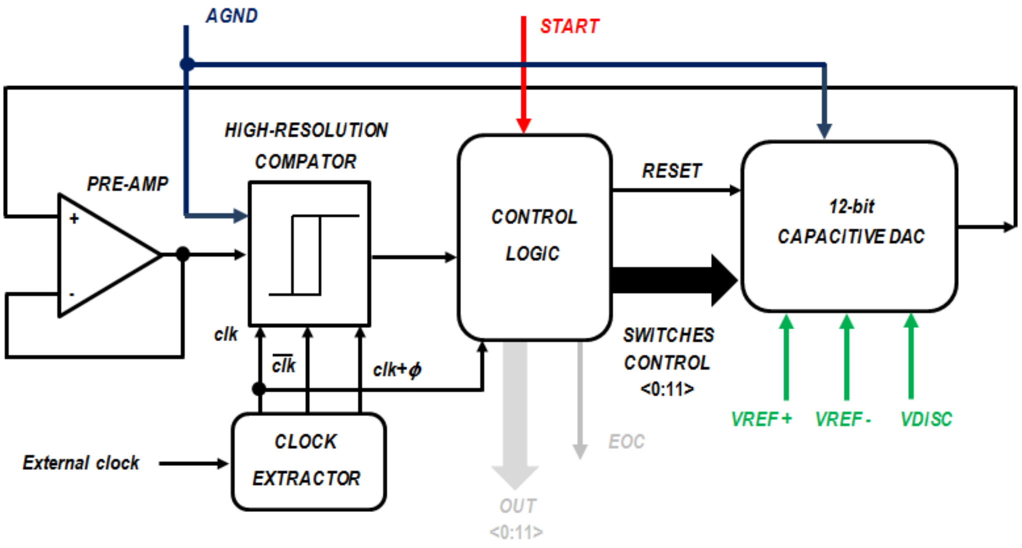


Figure 5.14. Schematic of the proposed SAR ADC.

The proposed solution consists of a pre amplifier, a high resolution comparator, a successive approximation register, a capacitive DAC and a non-overlapped clock generator. The specifications required to the ADC are summarized in Table 5.II. Considering these specifications, in the following the design of the different components of the ADC are discussed in detail. All the different components in the system were designed in order to achieve the maximum required performance in terms of resolution and speed.

TABLE 5.II. MAIN SPECIFICATIONS FOR THE DESIGNED ADC

VDD	3.3 V
Vin	0.8 V - 2.5 V
Full scale voltage	2.5 V - 0.8 V
ENOB	8 - 12
Power	≤ 1mW
Signal frequency	≤ 1KHz
Area	< 1 mm^2

5.4.2.1. Comparator

Since the input signal ranges from 0.8 V to 2.5 V, considering the goal of an ENOB equal to 12 bits, the LSB of the ADC is equal to 415 μm. This specification is fundamental for the design of the comparator due to the fact that it should be in able to discriminate signals which differs each other from this quantity or even less. The high resolution comparator employed in the proposed ADC is reported in [14]. The schematic of the comparator is shown in Figure 5.15. In contrast with the solution proposed in [14] in which two non-overlapped clocks are used, here 3 clock signals are employed to boost the performances of the comparator. Specifically during the phase *clk* nodes A, B are forced to V_{DD} while nodes C and D are forced to *GND*. During the phase *clk*+ϕ, the comparator starts to compare the signals in1 and in2. The nodes A and B start to change according to the unbalancing in the input pair. These signals are then sent to the SR latch which generates the output.

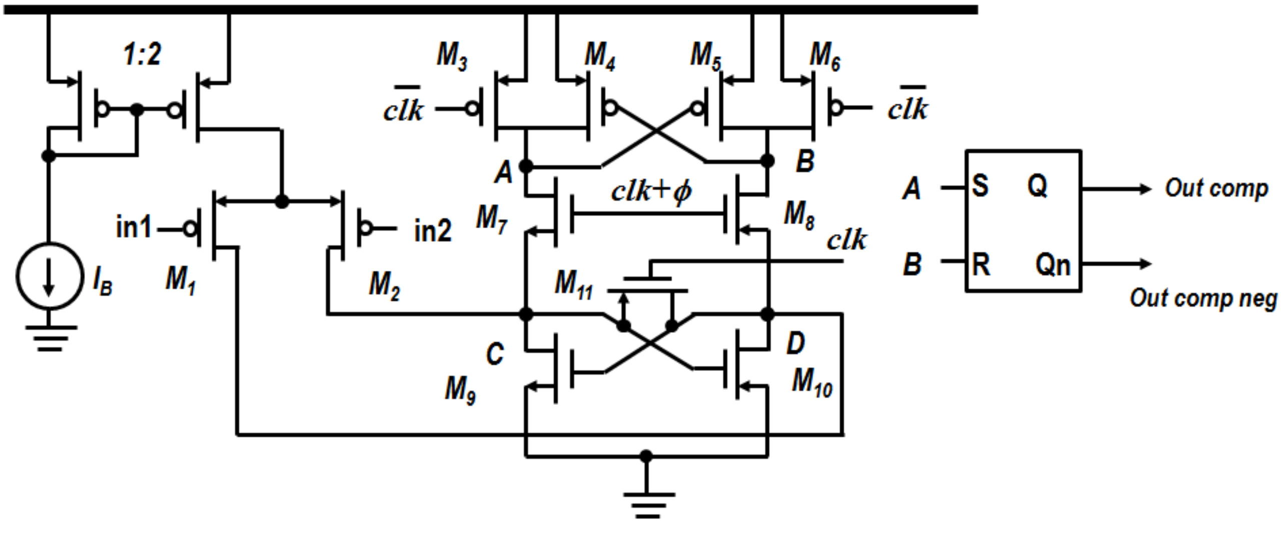


Figure 5.15. Schematic of the high resolution comparator [14].

The design of the comparator was performed in order to obtain the desired specifications in terms of speed and resolution. It is worth noting that the offset of the comparator leads to a non-linearity in the ADC output thus it is extremely important to alleviate this error. Regarding the speed of the comparator, to ensure a good functionality of the entire system and in particular a correct interpretation of the comparison result by the successive approximation register, the output of the comparator should reach the steady state before $T_{clk}/2$ where T_{clk} is the period of the clock signal. This condition has to be ensured for all the process corners and temperatures.

In the comparator of Figure 5.15 the regeneration time constant is equal to [14]:

$$\tau = \frac{C_{C(D)}}{\left(g_{m9(10)} - g_{d11}\right)}, \tag{5.17}$$

where is $C_{C(\mathrm{D})}$ is the capacitance at node C(D), $g_{m9(10)}$ is the gate transconductance of transistor M_9 (M_{10}) and g_{d11} the conductance of M_{11}. Thus to ensure an high-speed regeneration $g_{m9(10)}$ should be maximized. This is not trivial since the value of $g_{m9(10)}$ has to be maximized without increasing the value of $C_{C(\mathrm{D})}$. In the design of the comparator is also extremely important to take care of the problem related to the noise. Different kinds of noise can affect the resolution and performances of the comparator; however the thermal noise is one of the most critical factors limiting the accuracy of the comparator. Since in the dynamic comparator the transistors operate in different regimes at different times, the noise analysis is quite challenging. In [15] the authors show that in a dynamic latch comparator the thermal noise has the typical kT/C behaviour [16]:

$$v_{nc}^2 = \kappa_F \frac{kT\gamma}{C_{LC}}. \tag{5.18}$$

In (5.18) γ is the thermal noise factor (equal to 2/3 for long channel devices and >2/3 in short channel devices), κ_F is a constant which depends on the comparator's architecture and C_{LC} is the load capacitance at the bandwidth-limiting node of the comparator. Thus in order to limit the input referred thermal noise, long channel devices have to be employed. Taking into account all the previous consideration the comparator was designed in order to work properly at the maximum operating frequency of f_{clk}=10 MHz. In the proposed solution, in each phase of the conversion, the comparator discriminates the signal generated by the DAC with respect to the analog ground $AGND=V_{DD}/2$=1.65 V

TABLE 5.III. COMPARATOR'S PARAMETERS

Transistor	***W/L***
M_1=M_2	250µm / 1µm =(10 µm/ 50 µm) **x 25**
M_3=M_6	5µm / 1µm =(1 µm/ 1 µm) **x 5**
M_4=M_5	10 µm / 1µm =(2.5 µm/ 1 µm) **x 4**
M_7=M_8	1µm / 1µm =(0.5 µm/ 1 µm) **x 2**
M_9=M_{10}	20µm / 1µm =(2 µm/ 1 µm) **x 10**
M_{11}	8 µm / 1µm
Bias Current	**I_B=6 uA**

Because of this the comparator works with differential input voltages always around *AGND*, thus allowing avoiding the common mode issues due to the different bias conditions. As results the maximum performance in terms of resolution and linearity are obtained. The transistors size and the value of the bias current were chosen to obtain a settling time lower than 100nsec ($T_{clk}/2$) also in the worst case condition. From noise simulations the thermal noise was estimated to be 150 μV. As a consequence a nominal resolution of *LSB*/2 was imposed during the design. The transistor sizes and the bias current of the comparator that ensure all the previous specifications are summarized in Table 5.III. It is worth noting the large input pair necessary to reach the previous requirements in terms of resolution and settling time in all the process corners. At the same time the large input pair ensures a good robustness against process variation and allows minimizing the effect of the offset of the comparator.

5.4.2.2. Pre-Amplifier

Despite the comparator is in able to accomplish to its specifications, during simulations the kickback noise has degraded the performance of such component and of the entire ADC chain. The kickback noise is a consequence of the positive feedback mechanism for regeneration of the output voltage in a latched comparator. Here the large variations on the regeneration nodes are coupled to the inputs of the comparator through the parasitic capacitances (C_{PAR}) of the input pair (see Figure 5.16).

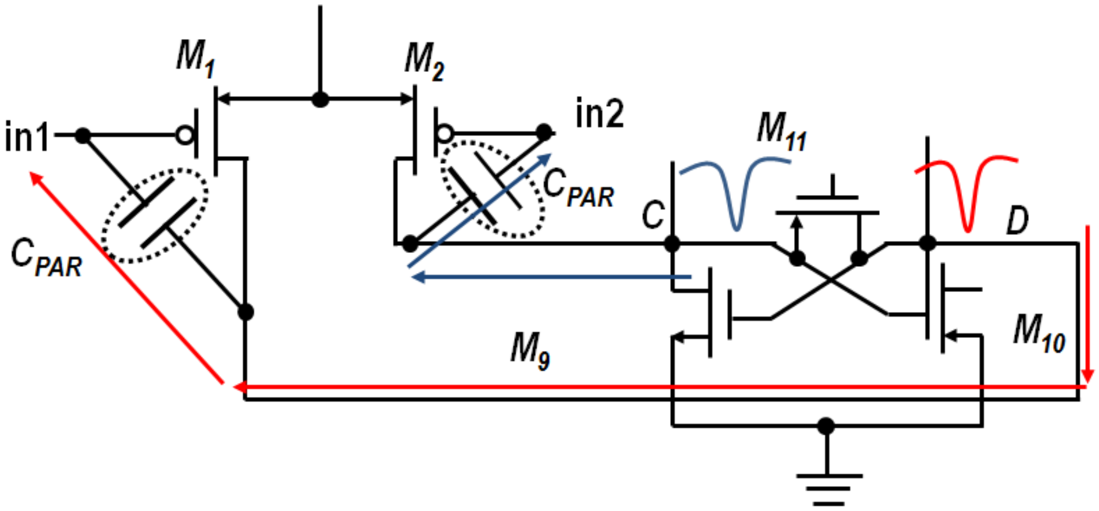
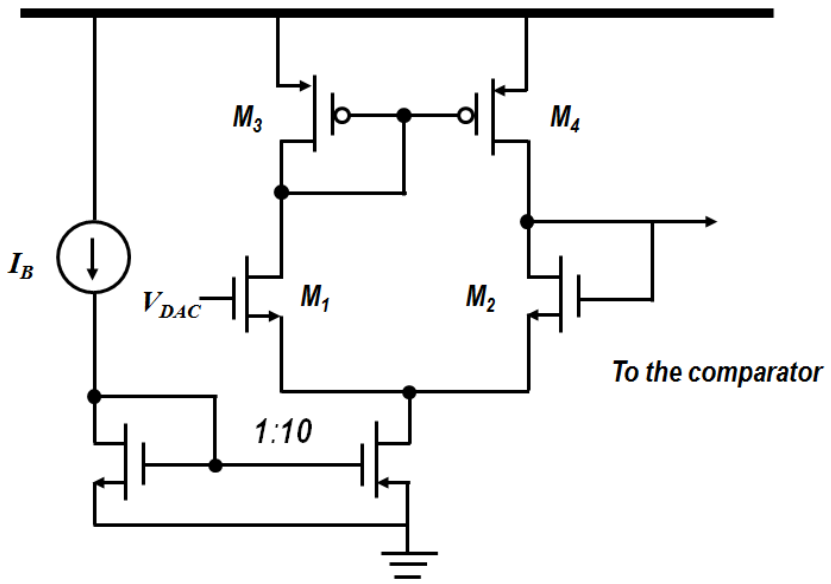


Figure 5.16. Kickback noise in the comparator of Figure 5.15.

As depicted in Figure 5.16 drain-gate capacitance is the main responsible for such effect, thus the large input pair of the proposed solution, which is necessary to accomplish the requirements of settling time and resolution, emphasizes this parasitic effect.

To mitigate the effect of the kickback noise, a buffer working as closed-loop voltage follower has been employed to generate the signal which has to be compared with *AGND*. This allows decoupling the input of the comparator from the regenerative nodes. The pre-amp stage amplifies the input signal to improve the comparator sensitivity (i.e., increases the minimum input signal with which the comparator can make a decision) and isolates the input of the comparator from switching noise coming from the positive feedback stage [17]. The schematic of the pre-amplifier and the main design parameters for this component are reported in Figure 5.17.



Transistor	*W/L*
M_1=M_2	200µm / 1µm
M_3=M_4	4µm / 1µm
I_B=3 uA	

Figure 5.17. Schematic of the pre-amp for kickback noise suppression.

5.4.2.3. Clock Generator

The phases necessary for the comparator are generated by a non-overlapped clock generator. The circuit receives an external clock and generated the signals *clk*, $\overline{clk}$ and the non-overlapped phase $clk+\phi$. The schematic of the non-overlapped generator is reported in Figure 5.18.

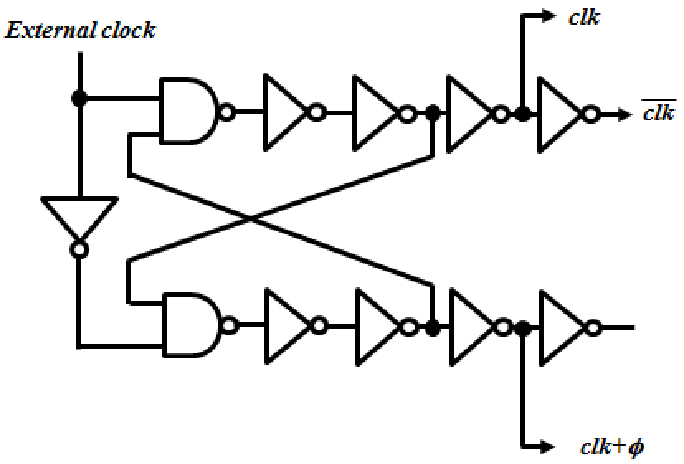


Figure 5.18. Schematic of the non-overlapped clock generator.

5.4.2.4. Capacitive DAC

For the digital to analog conversion a binary-weighted split capacitive array was used. The schematic of the DAC is reported in Figure 5.19.

The DAC consists of two equal arrays of binary-weighted capacitors with a value for the unit capacitor equal to C_U. Unlike the conventional solution, in the adopted scheme also the attenuator bridging capacitor is coincident with the unit capacitor C_U. Furthermore the additional unit capacitor in the MSB is removed. This allows mitigating the problem of matching the ratio between the capacitors in the array and the attenuator bridging capacitor. Moreover it allows saving switching energy and area. As pointed out in [18], this solution results in an error equally distributed through the quantization levels, causing a 1-*LSB* gain error which can be easily compensated by the preceding analog chain or in digital domain. For area and power saving purposes the unit capacitor C_U has to be chosen as small as possible. However different elements should be considered in the selection of the minimum value of C_U. The first element to consider is the *kT*/*C* noise.

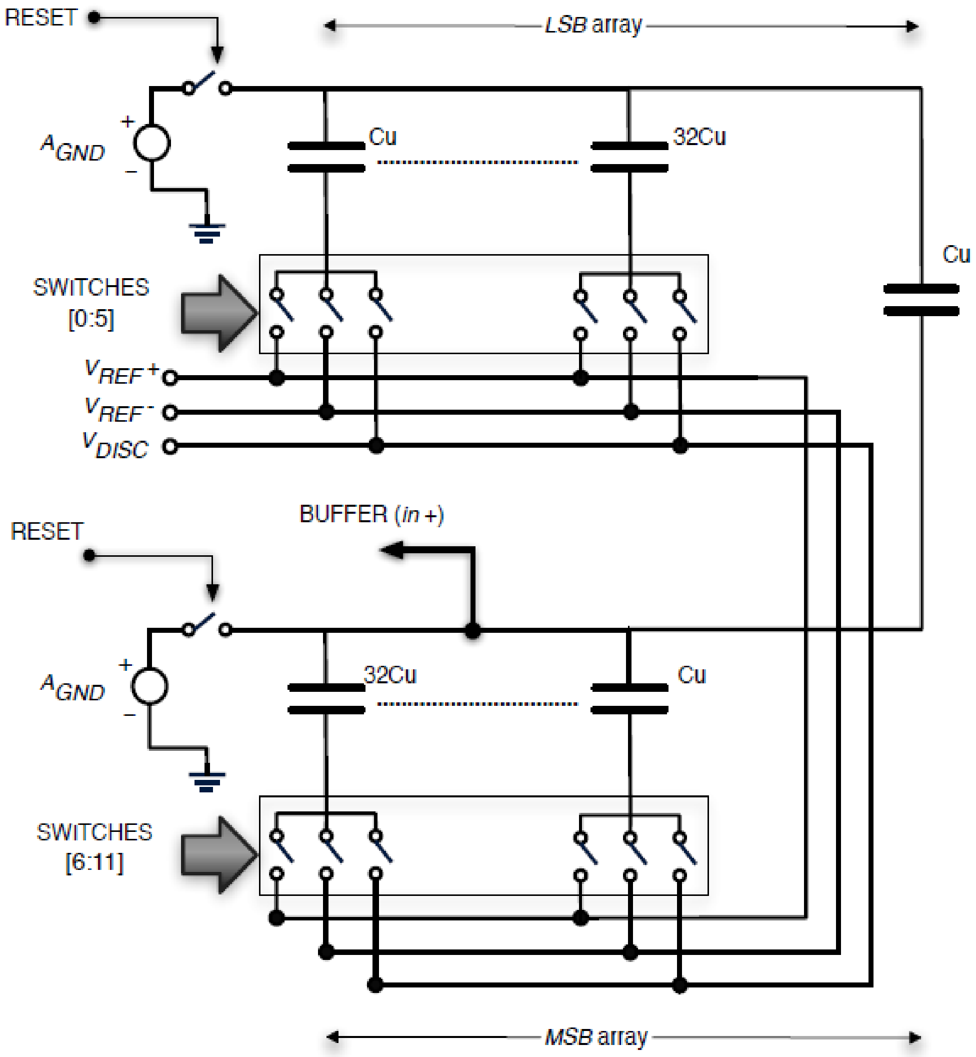


Figure 5.19. Binary-weighted split-capacitive DAC.

The value of C_U should be chosen in order to ensure that the noise introduced by the capacitor is less than the quantization noise:

$$\frac{kT}{C_{TOT}} < \frac{\Delta^2}{12} = \frac{1}{12}\left(\frac{V_{REF+} - V_{REF-}}{2^n}\right)^2 . \qquad (5.19)$$

From (5.19) the total capacitance of the DAC has to satisfy the following relationship:

$$C_{TOT} > \frac{12kT}{\left(\frac{V_{REF+} - V_{REF-}}{2^n}\right)^2}. \tag{5.20}$$

From (5.20) the total capacitance C_{TOT} has to be higher than 286.63 fF. Since the total capacitance of the proposed architecture is:

$$C_{TOT} = \left[2^{n/2} + \left(2^{n/2} - 1\right)\right]C_U = 127C_U, \tag{5.21}$$

the corresponding value of C_U is equal to 2.26 fF. This value of capacitance doesn't represent the real limit to the value of C_U. The mismatch between two unit capacitors in fact represents a worst condition to the value of C_U since it brings to a *DNL* and *INL* error. The worst case standard deviation of *DNL* and *INL* occurs when the DAC passes from the *LSB* bank to the *MSB* bank. According to [19] the following equations can be used to describe the standard deviation on *INL* and *DNL* respectively:

$$\sigma_{INL,MAX} = \sqrt{2^n - 1}\,\frac{\sigma_U}{C_U}\,LSB, \tag{5.22}$$

$$\sigma_{DNL,MAX} = \sqrt{2^{n-1}}\,\frac{\sigma_U}{C_U}\,LSB. \tag{5.23}$$

Since the value is related to the process, the foundry usually defines the standard deviation on the capacitor σ_U. From (5.22) and (5.23) the worst case is represented by the $\sigma_{DNL,MAX}$. For a capacitor the standard deviation is defined as:

$$\sigma\left(\frac{\Delta C}{C}\right) = \frac{k_C}{\sqrt{A}}. \tag{5.24}$$

In (5.24) k_C is defined as the matching coefficient for the technology while *A* is the area. *C* can be expressed as $C=\rho A$ with ρ capacitor density parameter expressed in Farad/μm^2. The standard deviation of a single capacitor from its nominal value is $\sqrt{2}$ times smaller than the matching between two capacitors, thus

$$\frac{\sigma_U}{C_U} = \frac{1}{\sqrt{2}}\sigma\left(\frac{\Delta C}{C}\right). \tag{5.25}$$

Imposing the condition of $6\sigma_{DNL,MAX}<0.5$ *LSB* for high yield, the following equation for C_U is obtained:

$$C_U \geq 36\left(2^n - 1\right)k_C^2\rho. \tag{5.26}$$

Substituting in (5.26) the value of density parameter and the matching factor of the chosen technology (k_C=0.45 %/μm, ρ=0.96 F/μm^2) the resulting C_U is equal to 11.5 fF.

Despite that, the value of C_U has to be chosen also by taking into account the parasitic capacitances which can modify the binary ratio of the capacitances in the DAC, causing a degradation of the conversion linearity. For this reason, in the following, the effect of the parasitic capacitance on the output voltage V_{OUT} is considered.

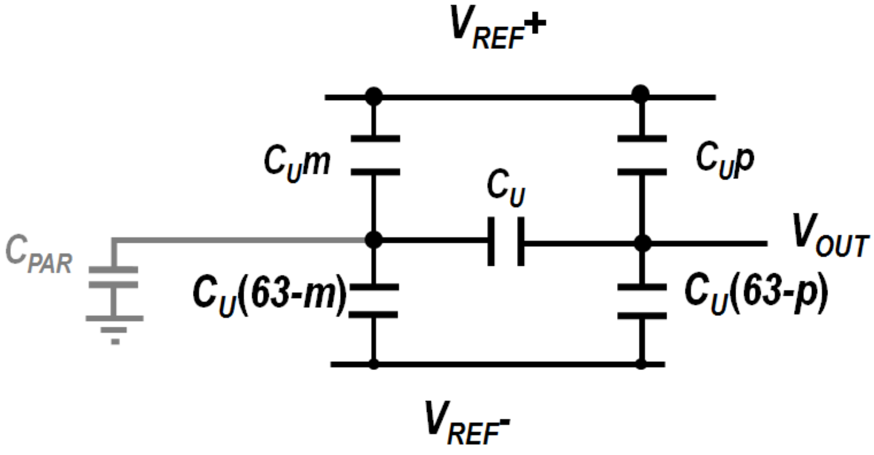


Figure 5.20. Equivalent circuit of the binary-weighted split-capacitive DAC of Figure 5.19.

To this aim, the equivalent structure of the DAC is reported in Figure 5.20. This scheme can be used to evaluate V_{OUT} as a function of the digital input. It is worth noting that only the parasitic capacitance of the *LSB* array is considered since the parasitic capacitance of the *MSB* array does not affect the linearity of the converter. From Figure 5.20, the output voltage V_{OUT} can be expressed as:

$$V_{OUT} = \frac{\left(64 + \frac{C_P}{C_u}\right)p + m}{\left[64\left(64 + \frac{C_P}{C_u}\right) - 1\right]}\left(V_{REF}+ - V_{REF}-\right), \tag{5.27}$$

where C_P represents the parasitic capacitance between the top plate of the *LSB* array and ground, while V_{REF}+ and V_{REF}- represent the positive and negative input signals for the DAC. From (5.27) it is clear the non-linearity effect due to the parasitic capacitance C_P and its effect on the output voltage V_{OUT}. The previous expression can be used to obtain an opportune value of C_U. This can be done by imposing the condition of

$$\varepsilon = \left|V_{OUT}(C_P) - V_{OUT}(C_P = 0)\right| < LSB, \tag{5.28}$$

where V_{OUT} (C_P) represents the value of V_{OUT} considering the parasitic capacitance as in (5.17), while V_{OUT} ($C_P=0$) represents the value of V_{OUT} imposing the condition of $C_P=0$. After performed parasitic extraction a reasonable C_P/C_U ratio, which ensures the condition of $\varepsilon<LSB$ consists of choosing $C_U>10C_P$. Considering the trade-off in area overhead, power consumption, and the condition reported in (5.28), a value of C_U=100 fF was used for the capacitive DAC, resulting in a total capacitance equal to 12.6 pF.

Regarding the operating condition of the DAC, during the reset phase (RESET=1) the top plate of both *LSB* and *MSB* arrays are connected to A_{GND} while the bottom plates are pre-charged, before each conversion, to $V_{DISC}=V_{IN}$.

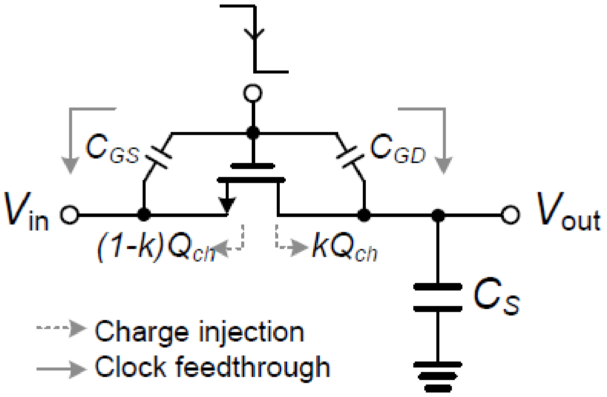


Figure 5.21. Charge injection and clock feedthrough effect in switch.

This allows an input signal for the comparator always centred on A_{GND} for any value of V_{IN}. As reported in Figure 5.19, each capacitor in the array is connected to the V_{REF}+, V_{REF}- and V_{DISC} signals through a dedicated pass transistor gate. The pass gate is dimensioned in order to mitigate the effect of charge injection and clock feedthrough (Figure 5.21) [20]. The control signals for the pass transistors, as well as the generation of the EOC signal and the output bits OUT[11:0] are generated by a dedicated control logic.

5.4.2.5. Successive approximation register

The successive approximation register receives the signal coming from the comparator and generates the right sequence of control signals for the switches in the DAC in order to implement the successive approximation algorithm. The description of this algorithm for a 3-bit SAR is reported in Figure 5.22.

START		DECISION b_2		DECISION b_1		DECISION b_0
100	>	110	>	111	>	111
					<	110
			<	101	>	101
					<	100
	<	010	>	011	>	011
					<	010
			<	001	>	001
					<	000

Figure 5.22. Successive approximation algorithm for a 3 bit SAR.

The SAR is also responsible for the generation of the *EOC* signal which provides information about the ending of the conversion of the sampled data. As explained before the *EOC* signal allows charging the DAC at V_{IN} before starting with the new conversion. The implementation of such a block was performed in VHDL. The code was synthesized using the standard cells provided by the selected technology (AMS 0.35 µm).

5.4.3. Layout

In this section the layout of the different blocks are reported. The technology used for the final prototype is the AMS 0.35 μm CMOS technology. It is worth noting that all the components are implemented in order to compensate the effects of the systematic mismatches. Careful layout techniques as interdigitate fingers, common centroid and dummy transistors were exploited in each block. In Figure 5.23 the layout of the comparator is reported. Most of the area occupied by this component is due to the matched input pair consisting in interdigitated fingers and dummy transistors.

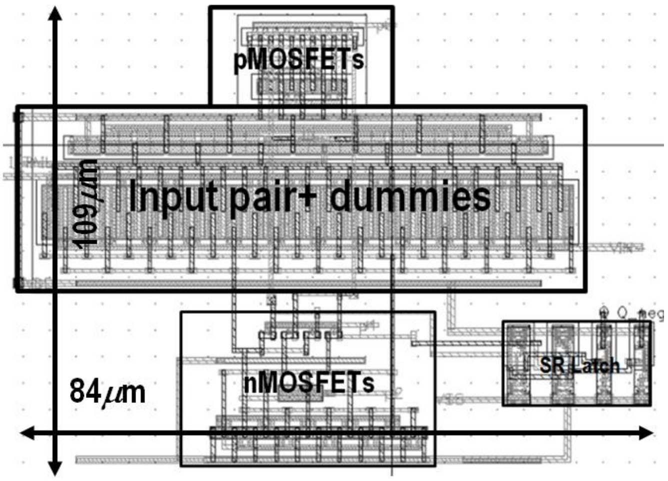


Figure 5.23. Layout of the comparator.

The layout of the pre-amplifier is reported in Figure 5.24. Regarding the DAC, to reduce the mismatch between the capacitance, a common centroid solution was implemented. To this aim the higher capacitances ($2C_U$, $4C_U$,….$32C_U$) are generated as the sum of the unit capacitor C_U. Figure 5.25 reports the arrangement of the capacitances in the DAC.

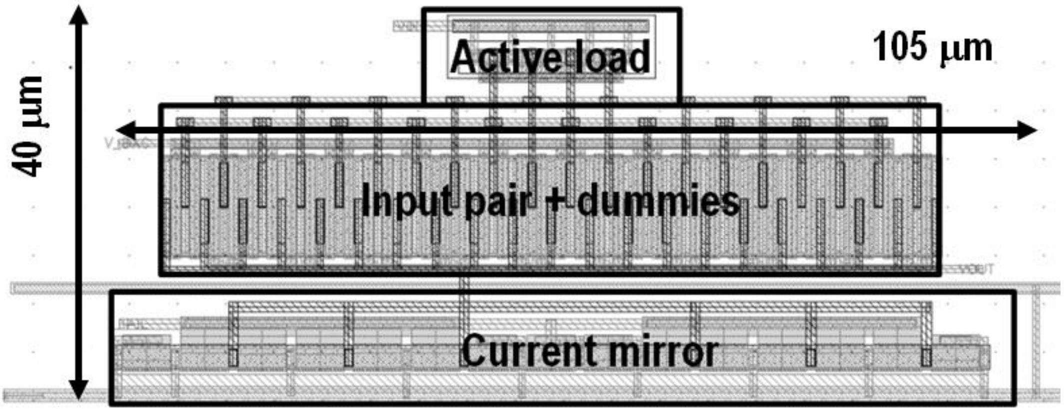


Figure 5.24. Layout of the pre-amplifier.

←MSB array→ **←LSB array→**

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	32	32	32	32	32	32	32	32	32	0	32	32	32	32	32	32	32	32	32	0
0	32	32	16	16	32	16	16	32	32	0	32	32	16	16	32	16	16	32	32	0
0	0	32	8	8	0	8	8	32	0	0	0	32	8	8	0	8	8	32	0	0
0	0	16	16	4	0	4	16	16	0	0	0	16	16	4	0	4	16	16	0	0
0	0	0	0	2	1	2	0	0	0	1	0	0	0	2	1	2	0	0	0	0
0	0	16	16	4	0	4	16	16	0	0	0	16	16	4	0	4	16	16	0	0
0	0	32	8	8	0	8	8	32	0	0	0	32	8	8	0	8	8	32	0	0
0	32	32	16	16	32	16	16	32	32	0	32	32	16	16	32	16	16	32	32	0
0	32	32	32	32	32	32	32	32	32	0	32	32	32	32	32	32	32	32	32	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 5.25. Common centroid configuration for the DAC.

In Figure 5.25 the number 0 indicates the dummy cell. The capacitance with weight 32 is obtained by putting together 32 unit capacitors while the capacitance with weight 16 is obtained by putting together 16 unit capacitors. Exploiting the same principle the capacitors with weight 8, 4 and 2 are obtained. The cell 1 indicates the two unit capacitors in the *MSB* and *LSB* array and the bridge capacitor. In Figure 5.26 the layout of the unitary cell (unit capacitor) is reported. The single cell can be connected to the other cells and to the bias voltages through two groups of connection lines, one in the upper part and one in the lower part of the cell. Since the connections to V_{REF}+, V_{REF}-, V_{DISC}, *AGND* and to the control signals coming from the SAR are obtained using horizontal lines, to avoid shorted connections, the capacitors with weight 32, 4, 1 are connected together using the upper lines while the capacitances with weight 16, 8, 2 are connected together using the connections in the lower part of the cell. In the first case the connections in the bottom of the cell are floating, vice-versa in the second case. This solution allows every cell to have the same amount of parasitic, thus reducing the unbalancing between the capacitances ratio. The complete layout of the DAC is reported in Figure 5.27 while in Figure 5.28 the layout of the SAR logic implemented with standard cells is reported.

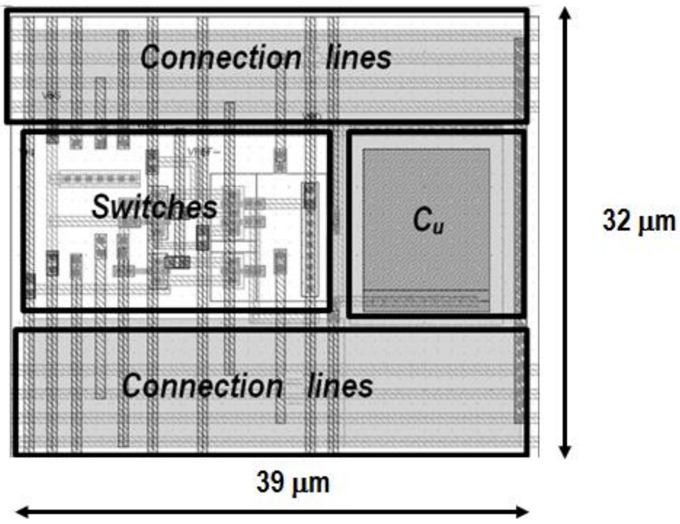


Figure 5.26. Layout of the unitary capacitance.

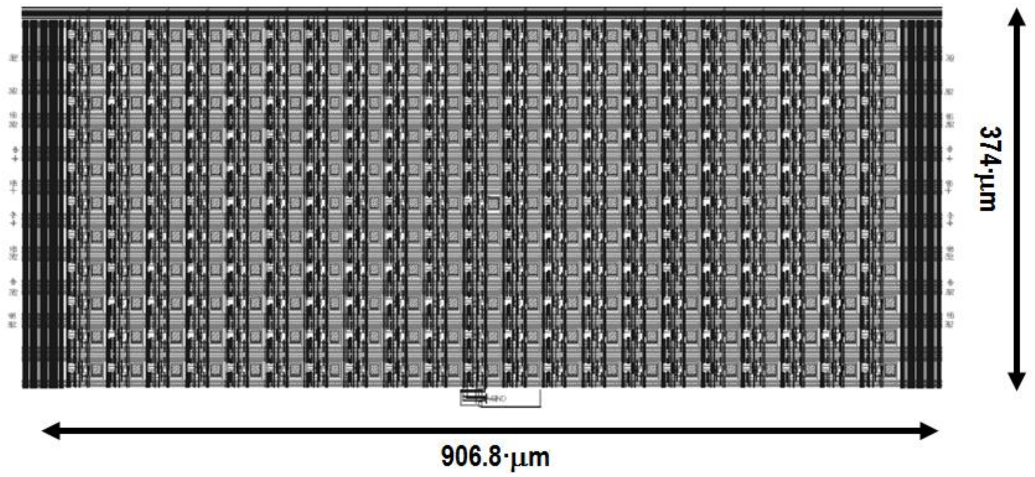


Figure 5.27. Layout of the DAC.

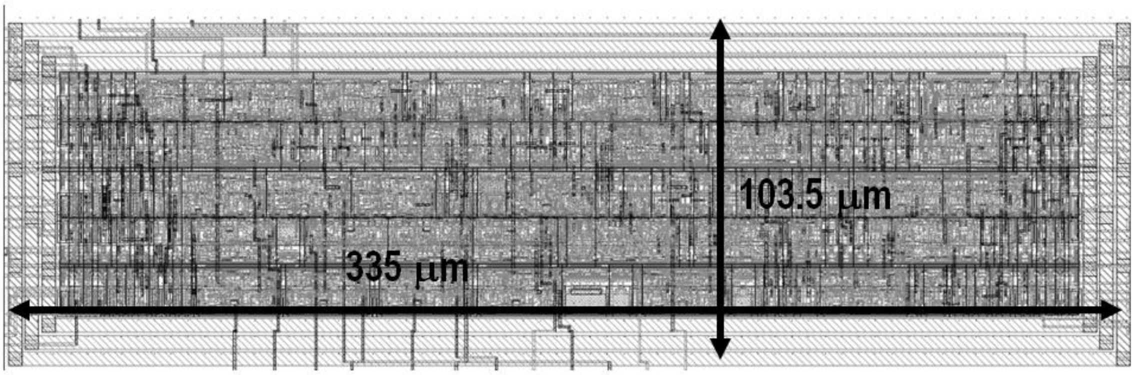


Figure 5.28. Layout of the SAR logic.

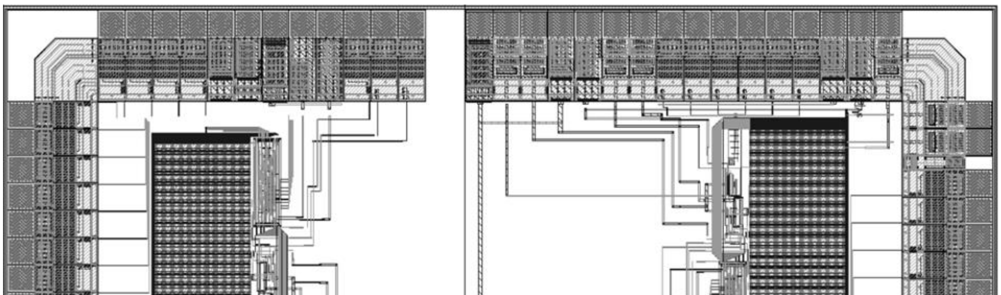


Figure 5.29. Layout of the entire chip.

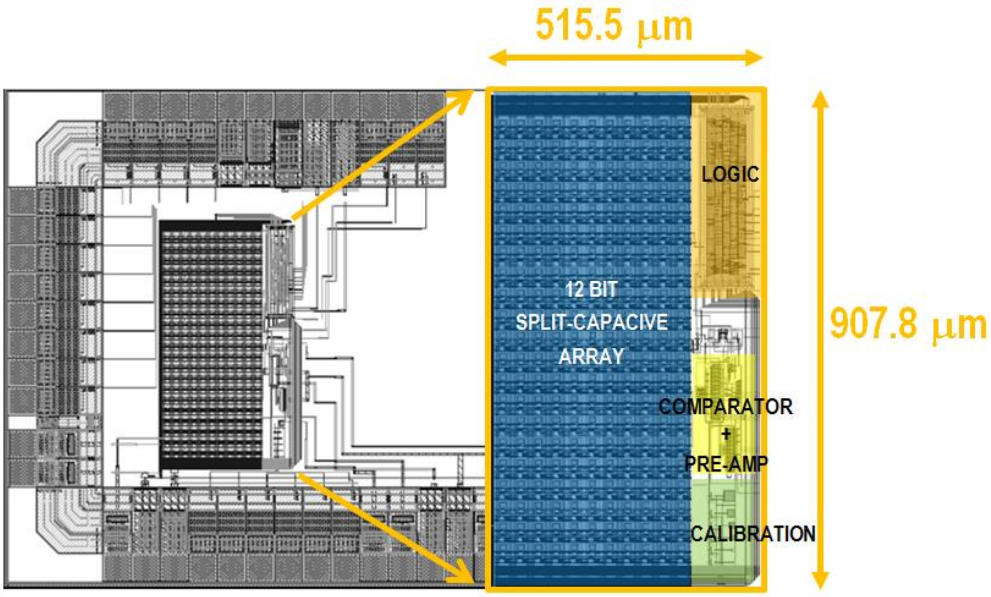


Figure 5.30. Layout of the ADC. The single elements in the structure are highlighted.

The layout of the complete system is shown in Figure 5.29. Two versions of the same prototype were fabricated. The two solutions are the same except for the extra pads in one replica. Here the output of the comparator, the input and the output of the pre-amp are connected to the padring with the aim of monitoring the functionality of the different components. Finally in Figure 5.30 the layout of the single ADC is reported. The figure puts in evidence the different blocks in the system. An additional calibration bank was inserted to achieve higher resolution.

5.4.4. Simulation and experimental results

The proposed SAR ADC was implemented at transistor level using AMS 0.35 μm CMOS technology. The simulated *DNL* and *INL* are reported in Figure 5.31(a) and Figure 5.31(b), respectively.

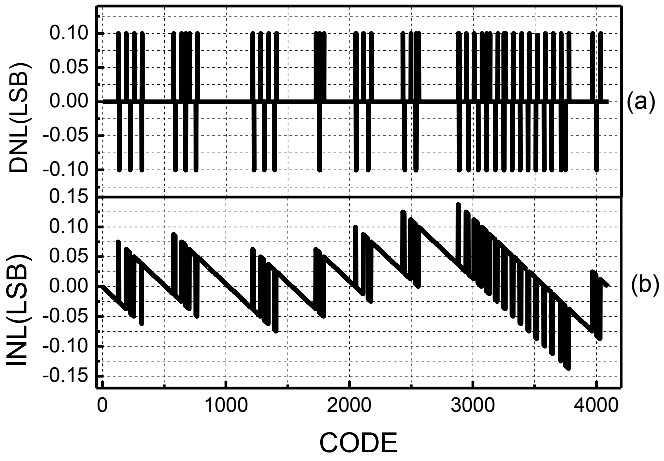


Figure 5.31. Simulated *DNL*(a) and *INL*(b).

For the extraction of the simulated values of *INL* and *DNL* a ramp signal was applied. The slope of the input signal was chosen in order to obtain 10 codes for each quantization level. It is worth noting that the simulation results are obtained in the worst case matching between the unit capacitors and by considering both intra-die and inter-die process variations. In such conditions the simulated *DNL* ranges from -0.1 *LSB* to +0.1 *LSB*, while the simulated *INL* ranges from -0.14 *LSB* to +0.14 *LSB*. In the same running conditions the simulated output spectrum was evaluated by using 4096 samples by applying an input sine wave signal with a frequency equal to 1.03 kHz (Figure 5.32).

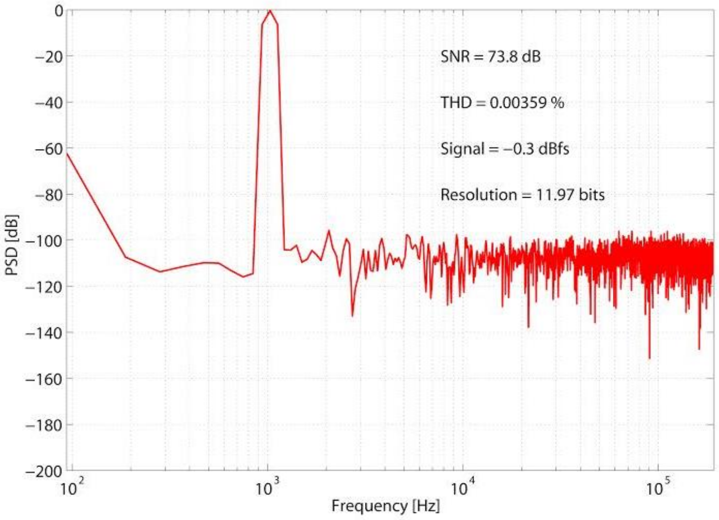


Figure 5.32. Simulated FFT. The input frequency is 1 KHz while the clock frequency is 10 MHz.

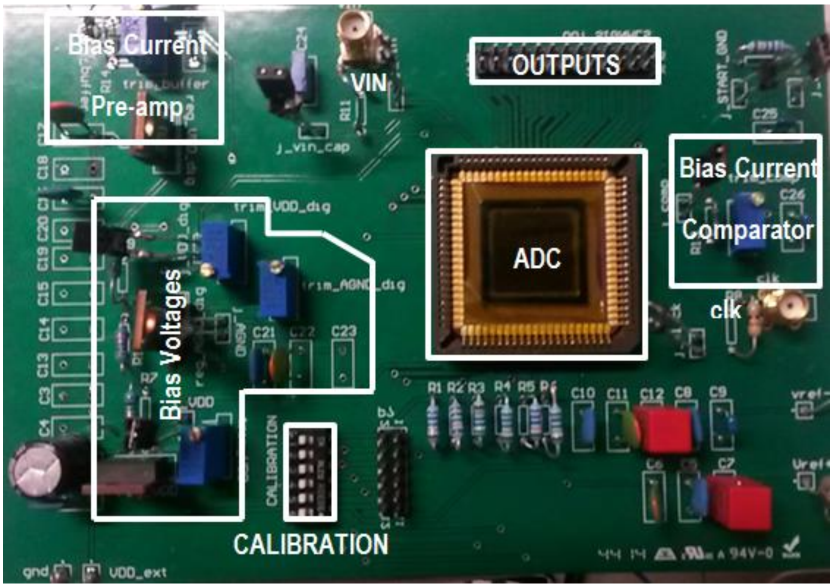


Figure 5.33. PCB used for the testing of the ADC.

The resulting *SNDR* is 73.8 dB and thus the *ENOB* is 11.97 bits while the mean power consumption is equal to 0.27 mW. The Figure of Merit for the ADC defined in (5.16) is equal to 87 fJ/step. The silicon prototype of the proposed ADC was tested as well. The PCB used for the testing was implemented using Altium®. A picture of the PCB is reported in Figure 5.33. Also for the testing of the silicon prototype the AC and DC specifications were extracted. For the DC test the control of the different instruments (waveform generator, digital multimeter, universal source) and the acquisition of the output bitstream were performed using Labwiew®. As in the case of the simulations, for the *DNL* and *INL* extraction a ramp signal was applied to the ADC by using a ramp signal generator.

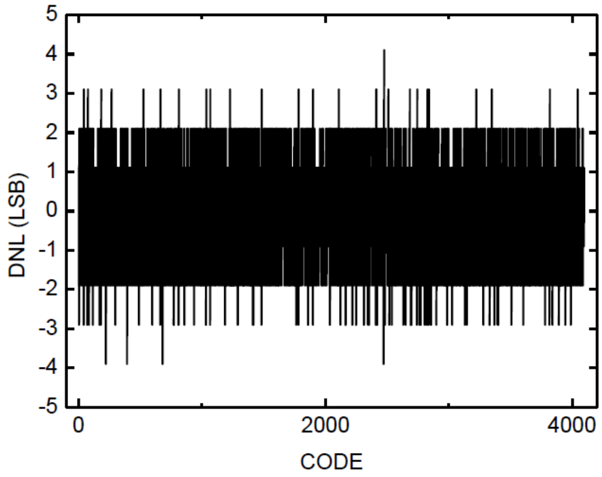


Figure 5.34. Measured *DNL*.

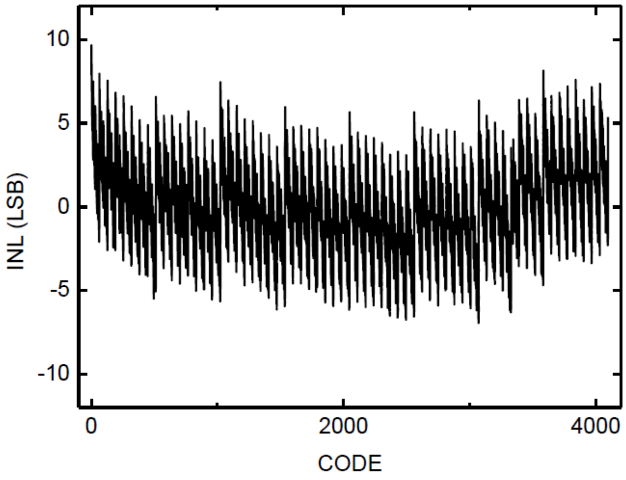


Figure 5.35. Measured *INL*.

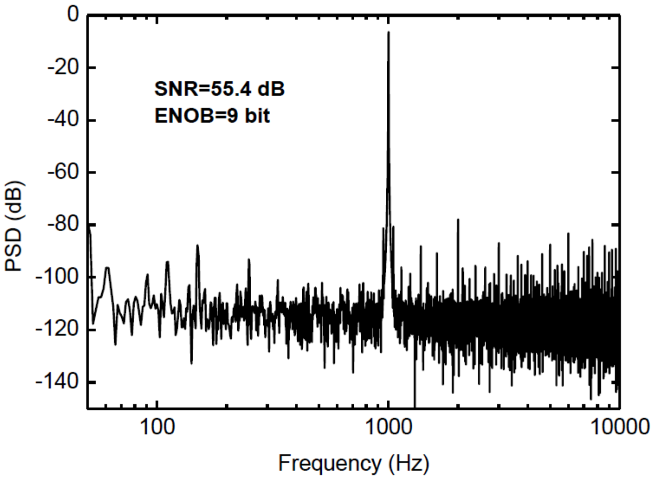


Figure 5.36. Measured *FFT*.

TABLE 5.IV. COMPARISON WITH THE OTHER LOW-POWER SAR ADCS

	[21]	[22]	[23]	[12]	[24]	**This work**
***Technology* (*nm*)**	90	130	90	130	350	**350**
***Resolution* (bit)**	9	10	10	9	12	**12**
***SNDR* (dB)**	53.3	52.8	56.6	-	-	**55.4**
***ENOB* (bit)**	8.6	8.5	9.1	9.1	10.2	**9**
***DNL* (LSB)**	+0.7 / -0.45	+0.88 / -1	+0.79 / -0.27	+0.5 / -0.5	+0.8 / -0.8	**+4/-4**
***INL* (LSB)**	+0.56 / -0.65	+2.2 / -2.09	+0.86 / -0.78	+0.5 / -0.5	+1.4 / -1.4	**+10/-5**
Sample Rate	40 MS/s	50 MS/s	100 MS/s	1 kS/sec	1 kS/sec	**0.77 MS/sec**
***Power* (mW)**	0.82	0.92	3	53 nW	895 nW	**0.27**
E_{CONV} (fJ)	54	52	55	94.5	195	**87**

Also here the slope was chosen in order to obtain 10 conversions for each quantization level, which results in a total number of conversions equal to 40960. The measured *DNL* and *INL* are reported in Figure 5.34 and Figure 5.35 respectively. From Figure 5.34 the measured *DNL* is equal to +4/- 4 *LSB* thus a DC resolution of 8 bits was achieved. The *INL* waveform on the other hand shows that the ADC exhibits the maximum non linearity in the beginning of the characteristic.

By analyzing the figure of the *DNL* it becomes clear that the missing codes are concentrated in the points in which the conversion moves from the *LSB* to the *MSB* bank, meaning that the parasitic capacitance on the *LSB* bank is much higher than the simulated one. The AC test was performed by applying a sine wave with a frequency of 1KHz. The clock frequency was settled to 5MHz. The measured output spectrum is reported in Figure 5.36. Despite the input filter (notch filter around f_{in}) spurious components at frequencies higher and lower than the fundamental are introduced. As a consequence the *SNR* is equal to 55.4 dB, thus the measured *ENOB* is equal to 9 bit.

In Table 5.IV the designed SAR ADC is compared with other low power SAR ADCs proposed in literature. The solutions chosen to perform the comparison have a resolution comparable to the one obtained in the proposed ADC. They are implemented in different technology nodes and exhibit different performances in terms of power consumption and speed. Ultra-low power solutions presented in [12] and [24] are in able to work properly by consuming nW, however they have very limited performance in terms of sampling rate (1 KS/sec) and energy per operation. On the other hand the very high speed solutions [21]-[23] show a power consumption nearby or higher 1 mW with a very low value of energy per conversion. Thus from Table 5.IV the proposed solution represents a very good option in the context of low-energy applications in which moderate performances in terms of speed, power and energy per conversion are required.

5.4.5. Conclusion

In this chapter a Successive Approximation Register ADC suitable for low-power applications was developed. The proposed solution consists of an input buffer, a high-resolution comparator, a control logic, and a low energy binary-weighted split-capacitive DAC. The proposed ADC was validated trough simulations performed at transistor level by using AMS 0.35 µm CMOS technology, considering the worst case matching between the capacitance units in the DAC and both inter-die and intra-die process variations. The simulated ENOB is 12 bit. The simulated DNL output range is equal to +0.1 LSB / –0.1 LSB while the INL output range is equal to +0.14 LSB / –0.14 LSB. The measured DNL and INL are +4/-4 LSB and +10/-4 LSB respectively while the measured ENOB is 9. The degraded performances are ascribed to the parasitic capacitance in the LSB bank, which was observed to be much higher than the simulated one. The measured mean power consumption is equal to 0.27 mW while the energy per conversion is 87 fJ/step. These performances and the achieved area occupancy of 0.47 mm^2 are very satisfactory for the required SAR ADC specifications.

Bibliography

[2] F. Maloberti. **Data Converters**, *Springer*, 2007.

[3] T. Neu, “Clock jitter analyzed in the time domain, Part 1”, *Texas Instruments Incorporated*, pp. 5-9. Online: http://www.ti.com/lit/an/slyt379/slyt379.pdf

[4] B. Brannon, **Sampled Systems and the Effects of Clock Phase Noise and Jitter**, AN-756 APPLICATION NOTE, *Analog Device*, Inc. Norwood, MA.

[5] Maxim Integrated. TUTORIAL 641. **ADC and DAC Glossary**. Available at https://www.maximintegrated.com/en/app-notes/index.mvp/id/641.

[6] *Maxim Integrated*. TUTORIAL 283. **INL/DNL Measurements for High-Speed Analog-to-Digital Converters (ADCs)**. https://www.maximintegrated.com/en/app-notes/index.mvp/id/283.

[7] W. Kester. MT-003 TUTORIAL. **Understand SINAD, ENOB, SNR, THD, THD + N, and SFDR so You Don't Get Lost in the Noise Floor**. *Analog Devices*. Online: http://www.analog.com/media/en/training-seminars/tutorials/MT-003.pdf.

[8] P. Malcovati. **ADC Conversion**.

[9] W. Kester. **Which ADC Architecture Is Right for Your Application?** Analog Dialogue 39-06, pp. 1-8. June (2005).

[10] **Choose the right A/D converter for your application**. *Texas Instrument*. Online: http://www.ti.com/europe/downloads/Choose%20the%20right%20data%20converter%20for%20your%20application.pdf.

[11] **ISSCC 2015 tech trends.** Online: http://isscc.org/doc/2015/isscc2015_trends.pdf.

[12] H. Tang, Z.C. Sun, K.W.R. Chew, and L. Siek, “A 5.8 nW 9.1-ENOB 1-kS/s Local Asynchronous Successive Approximation Register ADC for Implantable Medical Device,” in *IEEE Transactions on Very Large Scale Integration* (VLSI) *Systems*, vol. 22, no. 10, pp. 2220-2224, Oct. 2014.

[13] D. Zhang, A. Bhide, and A. Alvandpour, “A 53-nW 9.1-ENOB 1-kS/s SAR ADC in 0.13- µm CMOS for Medical Implant Devices,” *IEEE Journal of Solid-State Circuits*, vol. 47, no. 7, pp. 1585-1593, Nov. 2012.

[14] Chao Yuan and Yvonne Lam, “A Novel Low-voltage Low-power SAR ADC for biomedical applications,” in *IEEE International New Circuits and Systems Conference* (NEWCAS); pp. 101-104, 2011.

[15] G.M. Yin, F. Op't Eynde; W. Sansen, “A high-speed CMOS comparator with 8-b resolution,” *IEEE Journal of Solid State Circuits*, vol.25, no 22, pp. 208-211, Aug. 1992.

[16] P. Nuzzo, F. De Bernardinis, P. Terreni, G. Van der Plas, “Noise Analysis of Regenerative Comparators for Reconfigurable ADC Architectures,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol 55, no 6, pp 1441-1454, 2008.

[17] T. Sundström, B. Murmann, and C. Svensson, “ Power Dissipation Bounds for High-Speed Nyquist Analog-to-Digital Converters,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol 56, no 3, pp 509-518, Mar. 2009.

[18] R.J. Backer, **CMOS Circuit Design, Layout and Simulation**. 3rd edition, Wiley 2010.

[19] A. Agnes, E. Bonizzoni, P. Malcovati and F. Maloberti, "A 9.4-ENOB 1V 3.8µW 100kS/s SAR ADC with Time-Domain Comparator", in *IEEE International Solid-State Circuits Conference* (ISSCC), Feb. 2008, pp. 246-347.

[20] T. Wakimoto, H. Li, and K. Murase, “ Statistical analysis on the effect of capacitance mismatch in a high-resolution successive-approximation ADC,” in *IEEJ Transaction on Electrical and Electronic Engineering*, vol 6, no s1, pp. 89-93, 2011.

[21] B. Razavi. **Design of Analog CMOS Integrated Circuit**. McGraw-Hill Higher Education. 2003.

[22] V. Giannini, P. Nuzzo, V. Chironi, A. Baschirotto, G. Van der Plas, and J. Craninckx, “An 820 µW 9b 40 MS/s noise-tolerant dynamic-SAR ADC in 90 nm digital CMOS,” in *IEEE International Solid-State Circuits Conference* (ISSCC), Feb. 2008,pp. 238-610.

[23] C. C. Liu, S.-J. Chang, G.-Y. Huang, and Y.-Z. Lin, “A 0.92 mW 10-bit 50-MS/s SAR ADC in 0.13 µm CMOS process,” in *Symposium on VLSI Circuits*, Jun. 2009, pp. 236-237.

[24] Y. Zhu, C. H. Chan, U-F. Chio, S.W. Sin, U. Seng-Pan, R. P. Martins, and F. Maloberti, “A 10-bit 100-MS/s reference-free SAR ADC in 90 nm CMOS,” *IEEE Journal of Solid-State Circuits*, vol. 45, no. 6, pp. 1111-1121, Jun. 2010.

[25] Xiaodan Zou, Xiaoyuan Xu, Libin Yao, and Yong Lian “A 1-V 450-nW fully integrated programmable biomedical sensor interface chip. *IEEE Journal of Solid-State Circuits*,” vol. 44, no 4, pp. 1067 –1077, April 2009.

6. A new class of low-voltage Analog Physical Unclonable Functions

6.1. Introduction

Hardware security is becoming a crucial issue in a wide range of systems and applications, involving many processes ranging from device authentication to encryption and protection of intellectual property and sensitive data [1]-[22], [31]. All these applications largely benefit from the availability of circuits for ubiquitous hardware security that can internally generate a unique and reliable signature for each single device. At the same time many applications require energy-constrained and low-cost devices as in the case of handheld and mobile devices, sensor nodes and RFID systems. Hence, circuits for ubiquitous hardware security must operate at very low energy consumption using very little silicon area, maintaining at the same time requirements on the robustness against environmental variations and malicious attacks. In such a context silicon-based Phisycal Unclonable Functions (PUFs) are becoming a very attractive solution for security since they allow obtaining a unique signature down to chip level at the expense of a resonable energy and area consuption.

6.2. Overview on Physical Unclonable Functions

Starting from 2000, PUFs have emerged as a very promising solution to address the above open issues related to ubiquitous hardware security [1]-[22], [31]. A PUF is a function that generates a map of *Challenge-to-Response* pairs (**CPRs**) that is easy to evaluate but impossible to reproduce. Specifically, the PUF is interrogated with a bitstream called *challenge* while the bitstream obtained as reply is called *response*. The length of the challenge and of the response can be different. The unclonability of a PUF stems from the fact that the response generated by this device is a complex function of several physical quantities that even the manufactures cannot control. As a consequence the response generated by the PUF depends from its unique physical properties. Since its introduction in 2001 [1], different operating principles have been explored to implement such a component.

A classification of the PUFs is reported in Figure 6.1. The two main groups in which a PUF can be classified are *extrinsic* and *intrinsic* PUFs. In the case of an extrinsic PUFs the randomness is extrinsically introduced as in the case of the *coating* and *optical* PUFs [1]-[2], while in the case of the intrinsic PUFs the randomness is intrinsically introduced inside the components, like in the case of the *silicon* PUFs [3]-[22], [31]. Another possible classification of the PUFs take into account the number of challenge-response pairs that a PUF can handle according to its architecture. According to this, a PUF can defined ***strong*** or ***weak***.

A strong PUF generates a huge number of CRPs which cannot measured in a reasonable time frame, as a consequence in a strong PUF the prediction of the response to a random challenge is almost impossible, even with the prior knowledge of a limited number of CRPs. This implies that such a PUF should not be susceptible to modeling attacks.

On the contrary, in the case of the weak PUFs, a limited set of CRPs are available. Generally a weak PUF offers a safer mechanism for the generation of the secret keys compared to storing them in non volatile memory, nevertheless the secret keys can be stolen using side channel attacks [20]. Typical examples of weak PUFs are the memory-based PUFs.

6.2.1. PUF specifications

The main aim of a PUF consists in generating a reliable and unpredictable signature. Since each PUF should be completely different from another one and reliable at the same time, the main purpose of the different FOMs consist in evaluating such requirements. The different FOMs used for the evaluation of a PUF can be classified into two main groups: *security* FOMs and *VLSI* FOMs. The main security FOMs are *uniqueness*, *randomness*, and *reliability* while the main VLSI FOMs are power consumption, occupied area, minimum operating voltage. In the following sub-sections a detailed description of the security metrics is reported.

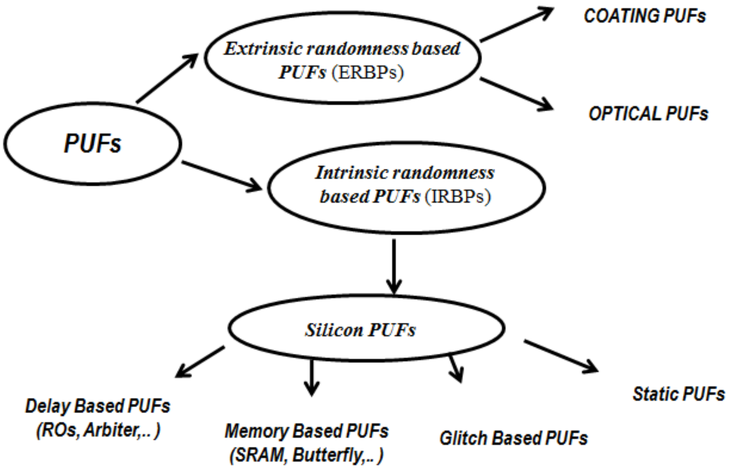

COATING PUFs
Extrinsic randomness based PUFs (ERBPs)
OPTICAL PUFs
PUFs
Intrinsic randomness based PUFs (IRBPs)
Silicon PUFs
Static PUFs
Delay Based PUFs (ROs, Arbiter,..)
Memory Based PUFs (SRAM, Butterfly,..)
Glitch Based PUFs

Figure 6.1. PUFs classification.

6.2.1.1. Uniqueness

Uniqueness is a measures of the inter-die randomness. It measures how different a PUF on a die is from another one built using the same design file in the same technology process. In particular, it provides a measure of how uncorrelated the response bits are across different dies. Ideally the response bits should differ with a probability of 0.5 in order to ensure the maximum level of uniqueness. The uniqueness can be improved by minimizing any systematic bias in the design (i.e. parasitic connections). A common way to measure the uniqueness of a PUF consists in the evaluation of the Hamming Distance (HD) between the words generated by different PUFs to a same challenge. In the ideal case, the fractional HD between the responses generated by two PUFs should be equal to 50 %.

6.2.1.2. Randomness

Randomness or intra-die randomness is a measure of the unpredictability of the response generated by a PUF. Ideally the response of a PUF should be unpredictable to a new challenge despite the prior knowledge of a large number of challenge-response pairs (CRPs). Moreover for a completely random response the numbers of 0's and 1's in the binary response have to be equally probable, thus ensuring that the response is totally unbiased. Several tests can be used to evaluate the randomness of the generated responses. Among them NIST test [25] represents an efficient method for evaluating the randomness in a binary sequence. It is worth noting that despite a failure in finding a lack in randomness, from the statistical point of view no set of finite tests can guarantee an absolute answer in terms of randomness. As a consequence a test can only provide information about the level of randomness for a given set of CRPs which becomes more reliable by increasing the number of considered CPRs. Figure 6.2 reports the concept of the intra-die and inter-die HD distance for a PUF. A PUF (user) can be identified only if a large separation between intra- and inter- PUF distance is guaranteed.

6.2.1.3. Reliability

Reliability measures the repeatability in the generation of the PUF response considering internal variations, environmental variations, noise and aging. Since the PUF response is used to identify an user, it is extremely important that the PUF is in able to provide the same answer for a give challenge independently of the external conditions. The consequence of the lack of reliability of the PUF is the bit flipping phenomenon whose description is reported in Figure 6.3. As an example, due to variations in the operating temperature or in the supply voltage, one or more bits in the PUF response can switch from the high logical state (V_{DD}) to the low logical state (0) or vice-versa. The reduction of this phenomenon is one of the main challenges during the design phase of a PUF. For this reason, to improve reliability of the PUF response, an error correction codes (ECC) circuit is commonly used when high level of reliability is required. However the additional ECC circuit introduces large penalties in terms of power consumption and silicon area.

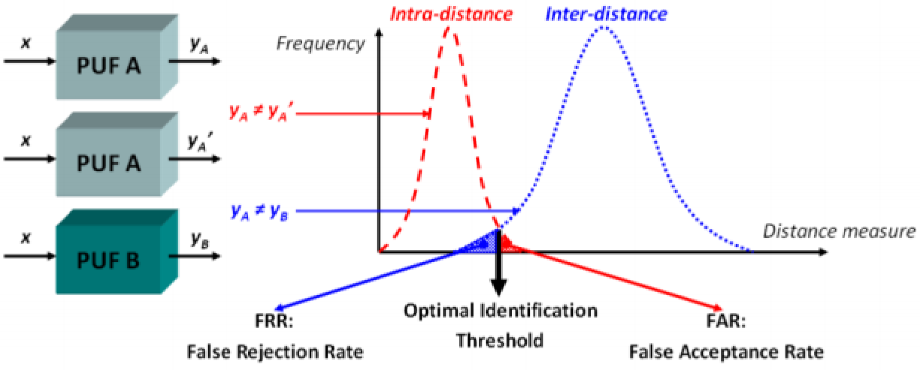


Figure 6.2. PUF inter-die and intra-die distance. To ensure the maximum level of reliability the intra-PUF HD has be as low as possible (ideally 0), while for the maximum level of randomness the inter-PUF HD has be equal to 50 %.

6.2.2. Main PUF solutions

The very first PUF was introduced by Pappu under the name of *Physical one-way function* [1]. In this solution the generation of the random bitstream was obtained by exploiting an optical mechanism. The challenge consists of a specific point and angle of incidence of the applied laser beam while the corresponding response is the result of an image transformation (Gabor transformation in the case of Pappu's) applied to the raw speckle pattern. The concept behind the Pappu's PUF is shown in Figure 6.4. There are several advantages in the optical version of the PUF [17]: 1) Low costs, since a non-integrated optical PUF consists of an inexpensive plastic platelet with randomly distributed light scatterers inside and no microelectronic or silicon circuitry are required. 2) High output complexity, since each PUF response consists of thousands of bits resulting from a very complex optical interference process. 3) High security against modeling attacks.

Despite that, due to the issues in the integration and in the miniaturization of an optical PUF, the Pappu et Al.'s idea didn't have a great diffusion.

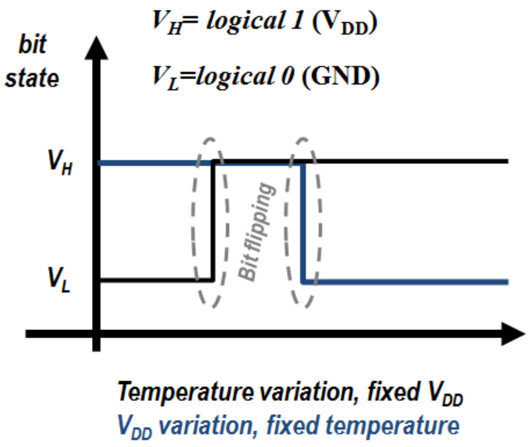

V_H= logical 1 (V_{DD})
V_L=logical 0 (GND)
bit state
V_H
V_L
Bit flipping
Temperature variation, fixed V_{DD}
V_{DD} variation, fixed temperature

Figure 6.3. Bit flipping due to temperature and V_{DD} variations.

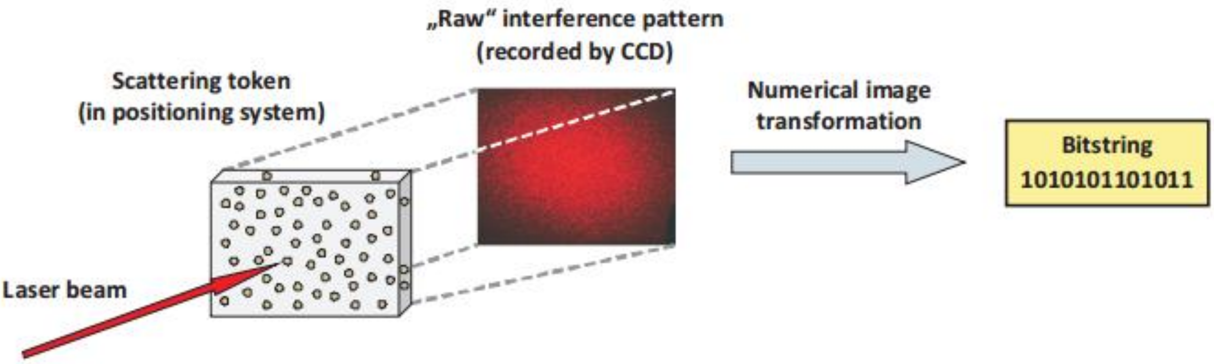


Figure 6.4. Pappu's optical PUF.

6.2.2.1. Silicon PUFs

Silicon PUFs are gaining a lot of attention as a potential high-efficiency and low-cost solution for security applications [3]-[15]. The silicon-based PUFs rely on the unpredictable variations of physical quantities during the fabrication, like oxide thickness and doping concentration. Among the different solutions for silicon-PUFs, delay-based PUFs [3]-[9] and memory-based PUFs [10]-[12] are widely considered as the most promising solutions for security at chip level. Other silicon-based solutions exploit other sources of randomness in the CMOS process/circuits like mask [14], maximum gain point [18], dynamic thresholding [19] and the oxide breakdown phenomenon [22]. Despite the good VLSI (power, speed, area) and security performance, the delay-based and memory- based PUFs are still considered the most efficient solutions for the implementation of silicon PUFs. For such a reason in the following a description of these two solutions is reported.

6.2.2.1.1. Delay-based PUFs

In the delay-based PUFs each random bit is obtained by exploiting the random difference in two nominally identical paths. The difference in the signal delay propagation across the two paths is processed to generate the random bit in the response. Two main architectures of delay-based PUFs have been proposed: the arbiter PUFs [3]-[5] and the ring oscillator (RO) PUFs [6]-[9].

6.2.2.1.1.1. Arbiter PUFs

In the arbiter PUFs, whose schematic is shown in Figure 6.5, an input signal is propagated trough two different paths according to the bits in the challenge. Nominally these two paths are identical, however due to the random and unpredictable mismatch between the components, one path is randomly faster or slower than the other one. The secret bit is generated by recognizing which one of the two paths is the fastest.

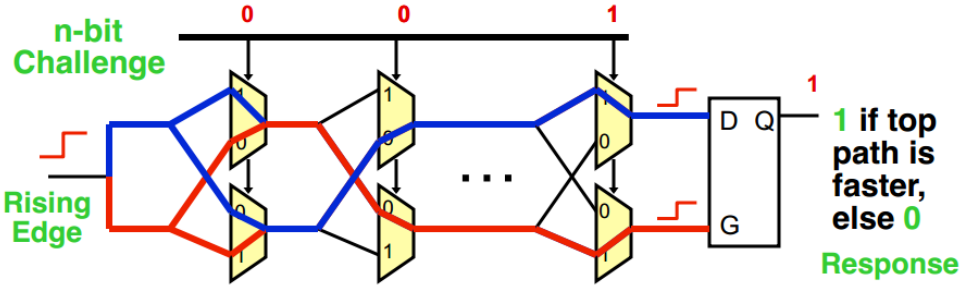


Figure 6.5. Arbiter PUF.

Despite the good performance in terms of uniqueness, the arbiter-based PUF is extremely fragile with respect to machine learning attack due to its linear behaviour [20]. An improved version of the arbiter PUF consists in the *XOR arbiter PUF* [4]. The schematic of this solution is reported in Figure 6.6. The XOR gate is used to process the signals generated by different PUFs, increasing the machine learning complexity as *O*(*nk*) for *k* XOR gates over *n*-stage PUFs. However it is obvious that this solution results in a higher silicon area. The size of circuit in fact grows as *O*(*nk*).

Other Arbiter-based architectures were proposed to overcome the weakness against machine learning. Among them the *lightweight arbiter PUF* [5] (Figure 6.7 (a)) and *feed-forward arbiter PUF* (Figure 6.7 (b)) are here reported.

In the *lightweight arbiter* PUF the input network includes XOR gates to create different combinations of challenge bits to each of the PUFs. Usually also an output logic network is included in such a PUF also consisting of XOR gates. In this way the responses from different PUFs are combined increasing the level of security. As result the lightweight PUF is more resistant to reverse engineering and emulation attacks than the basic arbiter PUF [20].

In the *feed-forward arbiter* PUF some of the challenge bits for the signal propagation are generated by intermediate signals along the PUF structure. As a consequence the non-linearity of the CRPs increases making the machine learning on the PUF infeasible [20], [23].

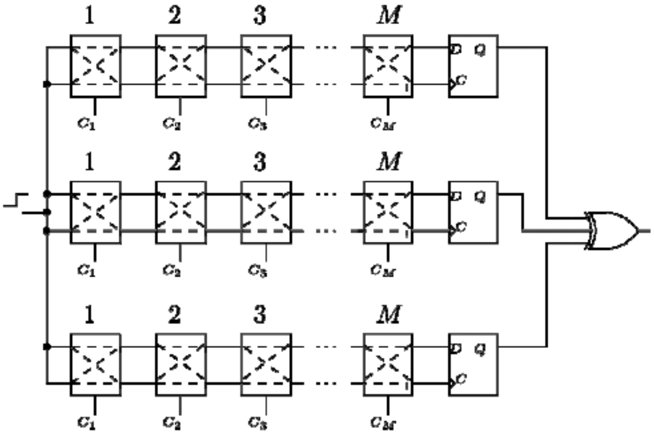


Figure 6.6. XOR arbiter PUF.

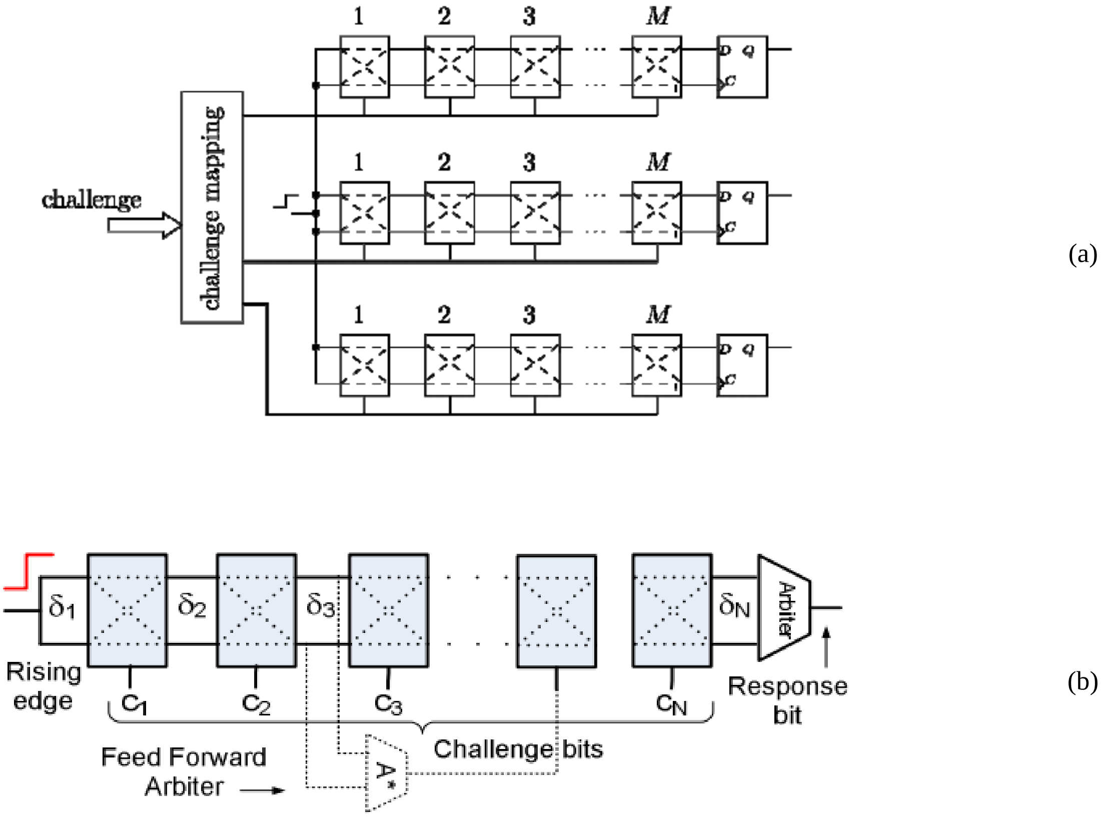


Figure 6.7. Lighweight (a) and Feed-Forward (b) arbiter PUF.

6.2.2.1.1.2. Ring oscillator PUFs

In the case of the Ring Oscillator (RO) PUFs, the secret bit is generated by comparing the oscillating frequency of two nominally identical ROs. Due to manufacturing variations, each ring oscillator generates a signal which is different in frequency from another one. Using the challenge bits, two of the *N* oscillators are selected and their frequencies are compared in order to generate the bit in the secret response. The basic architecture of a RO PUF is reported in Figure 6.8.

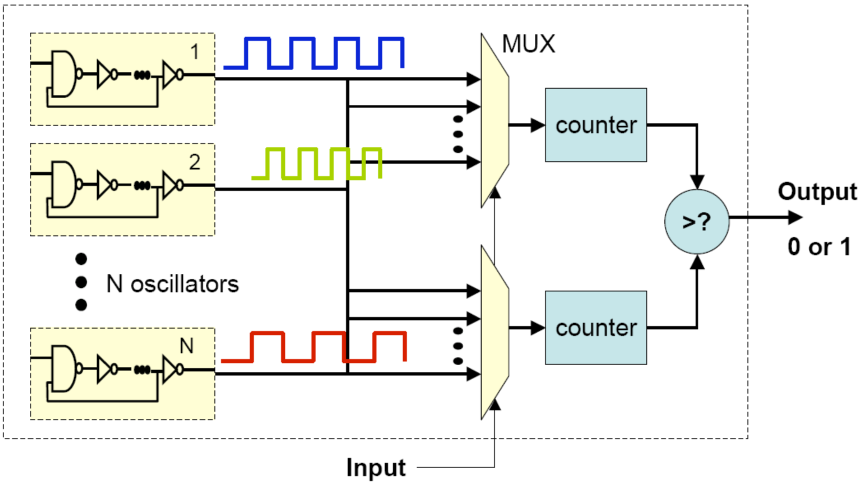


Figure 6.8. Ring Oscillator (RO) PUF.

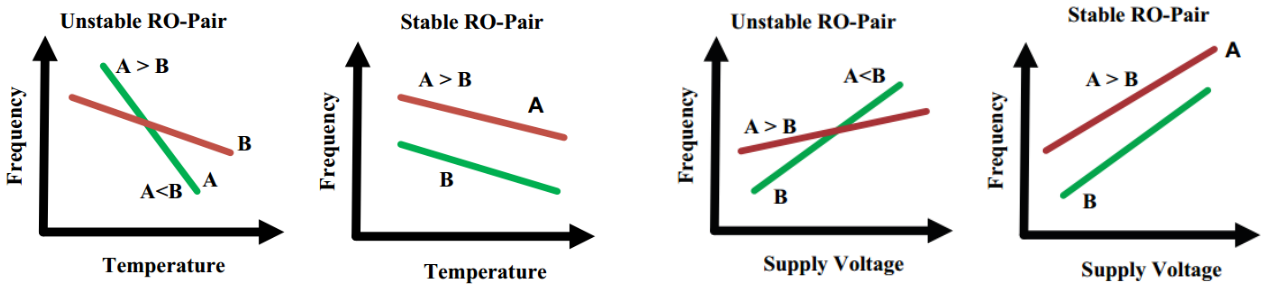


Figure 6.9. Bit-flipping in RO PUFs.

Despite the good performance in terms of randomness in the nominal operating conditions, RO PUFs suffer from reliability problems. Indeed, the relationship between the frequencies of the two oscillators can change by varying the operating supply voltage or the operating temperature as shown in Figure 6.9. As a consequence a large number of bit-flipping events are observed in such a solution [15]. To overcome these problems solutions with multiple supply voltages [8]-[9] have been proposed.

6.2.2.1.2. Memory- Based PUFs: Arbiter

In the memory-based PUFs, like Butterfly (Figure 6.10) [10] and SRAM PUFs (Figure 6.11) [11]-[13], a bi-stable structure consisting of two cross-coupled inverters is exploited to generate the output bit. When powered-up the memory PUF takes a random initial state depending upon the intrinsic mismatch between the two inverters [13]. Thanks to the positive feedback in the bi-stable structure, the unbalancing between the two inverters is amplified. As result the data stored in the memory cell becomes equal to logical 0 (0 V) or logical 1 (V_{DD}) in a random way.

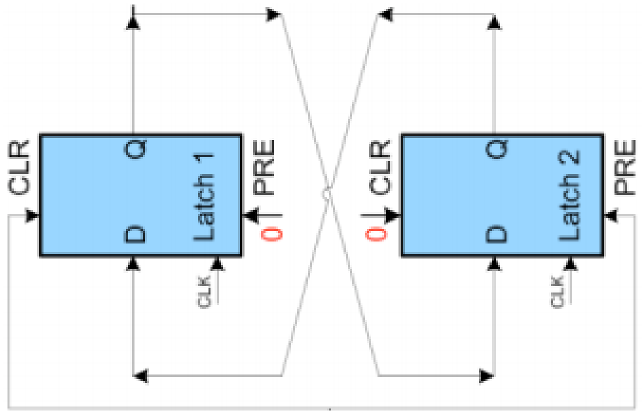
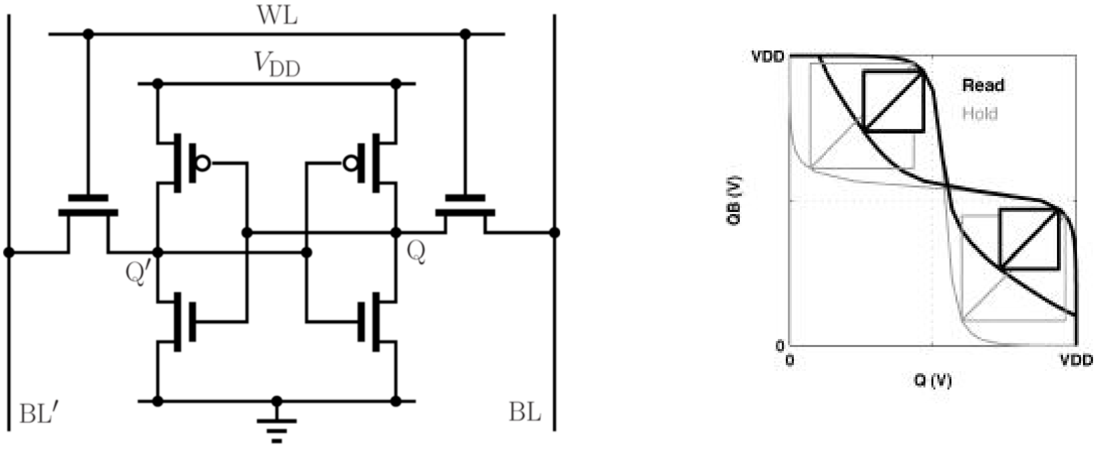


Figure 6.10. Butterfly PUF.



(a)

(b)

Figure 6.11. SRAM cell (a) and voltage transfer characteristic (b).

6.2.2.1.3. Considerations on delay-based and memory-based PUFs

The success of the delay-based and memory-based PUFs consists in the high level of technology migration offered by these two solutions. Indeed they can be implemented in any technology without any effort from the design point of view. The memory-based PUFs also have the advantage of reusing resources (memories), which already exist in most of the systems. Despite that, several works have shown several drawbacks both in the delay-based and memory-based PUFs [8]-[9], [13]-[15], [20],[23]. Considering the SRAM PUFs, in the ideal case the eye diagram of the SRAM cell should scale symmetrically by varying V_{DD} as reported in Figure 6.12. In reality, due to the process variations, the eye diagram shows an unbalance between the two logical states by varying the operating temperature, supply voltage or as a consequence of the external noise. A real eye diagram, which takes into account the effects of the variations in the operating conditions, is reported in Figure 6.13. As a consequence of the non-ideal behaviour, a logic state becomes dominant with respect to another one. This results in several bit-flipping events which dramatically reduce the reliability of the response. [15]-[20]. The reproducibility of the response generated by the SRAM PUFs is also severely affected by the ramp-up speed of the bias voltage and by the temperature variations [13]. SRAM-based PUFs also suffer from aging since the memory cells used to generate the secret keys are also employed as data memory for the rest of their lifetime.

Aside the above drawbacks, being not steady, both memory-based and delay-based PUFs show a high sensitivity to noise which easily leads to a high percentage of bit-flipping [15]. In the case of the RO PUFs, the oscillation frequencies must be not too high in order to allow the counters, which generate the random output bit, to count oscillations [9]. Thus, a minimal number of inverters in every ring oscillator are necessary to ensure a suitable oscillating frequency (typically more than 10-20 [9]).

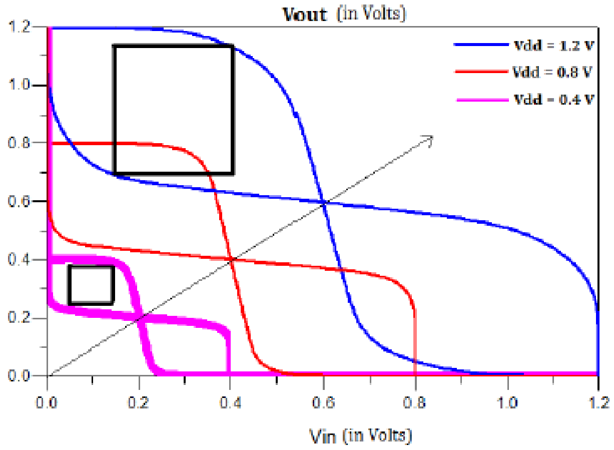


Figure 6.12. Ideal SRAM VTC at different V_{DD}s.

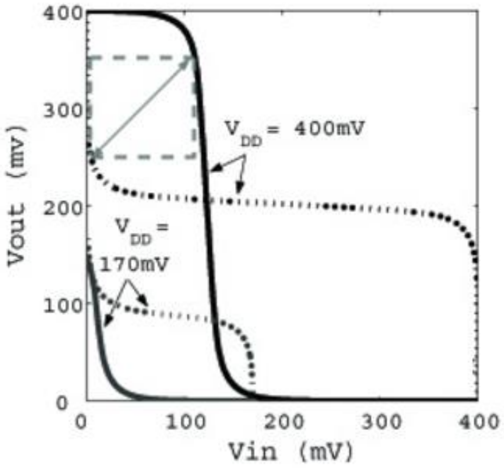


Figure 6.13. Deterioration of the SRAM VTC at low-V_{DD}.

However, with a higher number of RO stages the oscillating frequencies of the compared ROs become close, thus making the comparison very sensible to supply (V_{DD}) and temperature variations [8]. In addition, RO PUFs exhibit large power consumption [8].

6.3. Proposed Solution 1: Complementary current mirrors based PUF

6.3.1. Introduction

In this section a new class of physical unclonable functions for secret keys generation at silicon level is presented. The proposed solutions exploit the variability of complementary current mirrors to generate a voltage which is completely random and robust against supply voltage, temperature and noise variations. An extensive set of MonteCarlo simulations were performed using 65 nm and 180 nm CMOS technology and, where available, simulations results are compared with experimental results. Both experimental and simulation results show a very high level of uniqueness, randomness and reliability. For all the reported solutions the *intra*-PUF and *inter*-PUF fractional hamming distance are close to the ideal values of 0% and 50 % respectively while maintaining a very low percentage of unstable bits. Using NIST test it was possible to observe that the proposed schemes are capable of generating a bitstream which is completely random and unpredictable. Moreover all the proposed solutions can work also in the context of ultra-low voltage scenario of V_{DD}=0.3 V. Finally, being static, a very high level of robustness against supply voltage noise was observed.

6.3.2. Architecture of the proposed PUF

Figure 6.14 shows the architecture of the proposed PUF. The generation of the secret key is obtained through a combinational function (*COMB*) which receives the challenge and the random bitstream generated by the PUF key generator. The function implemented by *COMB* is purely deterministic while the PUF key generator is completely random.
In order to obtain a random response, the function *COMB* has to respect some properties. In particular:

- ✓ it has to show a good statistical response. The number of logical 0’s and logical 1’s has to be the same in the ideal case, thus ensuring a completely unbiased response.
- ✓ it should be robust against any possible attack. As a consequence the response for a given challenge should provide minimal (zero in the ideal case) information about the secret word *k*.

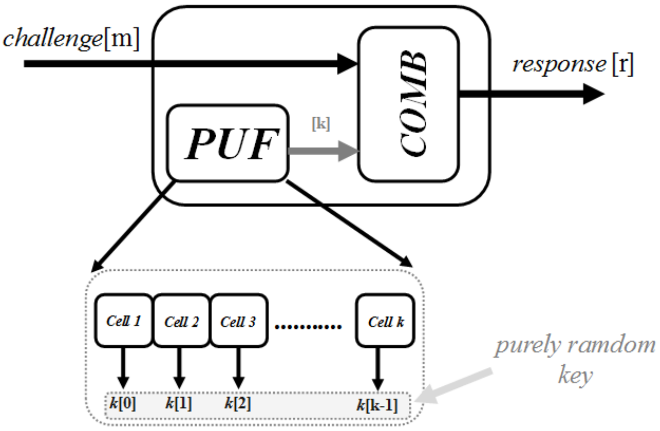


Figure 6.14. Architecture of the proposed PUF.

A possible implementation of such a function is reported in the Figure 6.15. Using this function, for a given word *m*, the complexity is estimated to be *m* XOR gates+$log_2(m)$.
From the previous considerations, in the proposed scheme the randomness of the response is generated only in the cells of the PUF key generator.
The above described novel class of PUFs has several advantages over existing classes.
First, the unpredictability of the PUF relies only on the variations within the array of *N* cells inside the random generator. As a consequence the design and verification phases are limited to the single cell generating the unpredictable bits, instead of being widely distributed as occurs in delay-based PUFs. Thanks to this, the designer can focus on a very small structure to ensure that the desired statistical features and adequate robustness against voltage/temperature variations are achieved. As a consequence the design and the verification efforts are much lower than existing PUFs, thus ensuring easier technology portability.
Secondly, the proposed class of PUFs is purely static, hence interconnects do not impact at all the response of the PUF to each challenge. Indeed the response purely relies on transistors variability, which is well modelled in today's design kits.

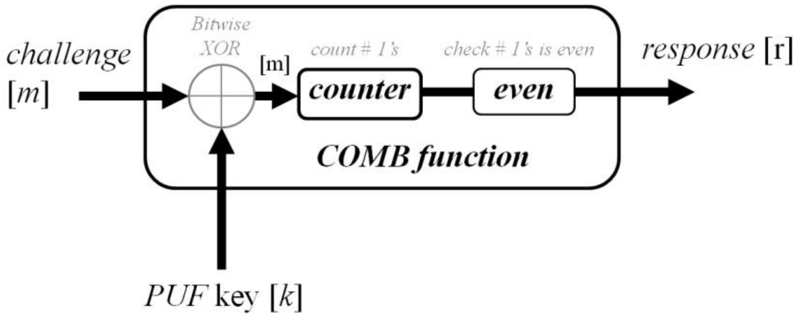


Figure 6.15. Possible implementation of the *COMB* function.

As a consequence the designer can develop a deep understanding of the statistical features and robustness against voltage/temperature variations at design time, rather than having to wait for silicon measurements. Instead, existing PUFs, like ROs and SRAM PUFs, rely on transient phenomena that are influenced by interconnects, which are poorly characterized in terms of variations.
Third, it is immune to modelling attacks. This because the proposed PUF class relies on bits that are basically uncorrelated and their individual value cannot be predicted from the output.
Fourth, the silicon area of the proposed PUF class is very small, as each of the *N* bitcells contains very few and minimum-sized transistors.
Fifth, energy is small since the proposed PUF class is very compact and is suitable for ultra-low voltage operation. Indeed, the purely digital block has consistent output over a very wide range of voltages, and the bitcells are designed to have the same output at the same voltage range. In contrast, as shown before, existing PUF classes are not suitable for ultra-low voltage operation, since their output strongly depends on voltage and temperature.
Sixth, aging is not an issue since the bitcells are read only occasionally (e.g., at the chip boot) and then stored in a register for run-time utilization.
In the following section the architecture for the generation of the secret bit (basic bit-cell key generator) is reported.

6.3.3. Basic solution: complementary current mirrors (CCMs)

Figure 6.16 shows the simplest version of the proposed bitcell key PUF exploited for the generation of the single random bit. The solution consists in a couple of complementary current mirrors and a simple inverter (*INV*) which generates the output bit *OUT*. The aspect ratio of both current mirrors is equal to 1 ($(W/L)_{N1}=(W/L)_{N2}$, $(W/L)_{P1}=(W/L)_{P2}$). Since obtaining a process-dependent voltage is the first aim of the proposed circuit, all the transistors are biased in subthreshold regime.

Considering the DC behaviour of the proposed architecture, the two diode-connected transistors M_{N1}-M_{P1} act as a voltage divider for the supply voltage V_{DD}.

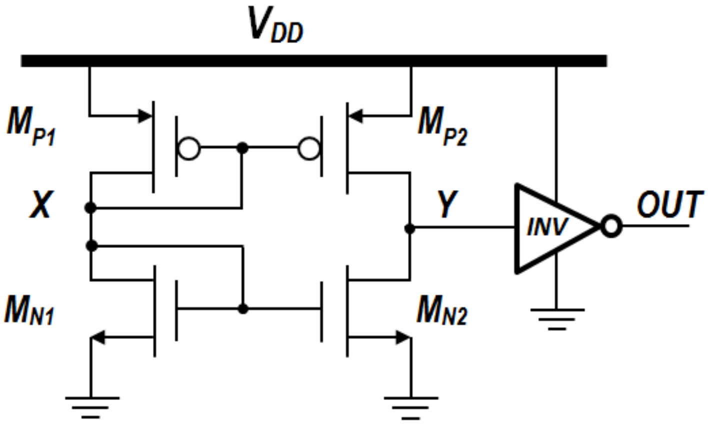


Figure 6.16. Proposed PUF key cell based on Complementary Current Mirrors (CCMs).

By assuming the subthreshold conduction for both transistors, by equating the currents in M_{P1} and M_{N1}, the following equation holds:

$$\beta_{N1}\exp\left(\frac{V_X - V_{TH,N1}}{nV_T}\right) = \beta_{P1}\exp\left(\frac{V_{DD} - V_X - |V_{TH,P1}|}{nV_T}\right). \tag{6.1}$$

Solving the (6.1) with respect to V_X (voltage on node *X*) the following expression is obtained:

$$V_X = \frac{V_{DD}}{2} + \frac{V_{TH,N1} - |V_{TH,P1}|}{2} + \frac{nV_T}{2}\ln\left(\frac{\beta_{N1}}{\beta_{P1}}\right). \tag{6.2}$$

In (6.1) and (6.2) the current factors β_N and β_P are defined as:

$$\beta_{N1} = \mu_N C_{OX,N}\left(\frac{W}{L}\right)_{N1} V_T^2 , \tag{6.3}$$

$$\beta_{P1} = \mu_P C_{OX,P}\left(\frac{W}{L}\right)_{P1} V_T^2 . \tag{6.4}$$

Without any loss of generality, an equal value for the subthreshold slope factor of the nMOS and pMOS transistor is assumed. This approximation does not influence the validity of the proposed analysis since the value of *n* is pretty close for the pMOS and nMOS transistors [24]. Additionally it is worth noting that in (6.1) it is assumed that the drain current of the nMOS (pMOS) transistor is independent from V_{DS} (V_{SD}), which is satisfied if V_{DS} (V_{SD}) is greater than $4V_T$.

From (6.2), by balancing the nMOS and pMOS strength ($\beta_{N1}= \beta_{P1}$), V_X becomes equal to $V_{DD}/2$. Additionally from (6.2) it is possible to note that V_X linearly depends from the difference of the absolute values of the two threshold voltages. Since the standard deviation of V_{TH} is proportional to $1/(WL)^{0.5}$ [25], small area transistors can be employed to emphasize the effects of the process variability [25].

Nominally the voltage on node Y tracks the voltage on node X. However, due to the mismatch between M_{N1}-M_{N2} and M_{P1}-M_{P2}, the voltage V_Y largely differs from V_X. Using the small-signal analysis, the relationship between V_X and V_Y can be expressed as:

$$V_Y = -\frac{g_{m,N2} + g_{m,P2}}{g_{d,N2} + g_{d,P2}} V_Y \approx -g_m r_{ds} V_X \, , \tag{6.5}$$

where g_m and $g_{ds}=1/r_{ds}$ represent the gate transconductance and the drain conductance for the nMOS (pedix N) and pMOS (pedix P) transistor, respectively. Note that these two parameters are assumed equal for the two transistors (due to balanced strengths). From (6.5) the process-induced variations are amplified by a factor equal to $g_m r_{ds}$.

6.3.4. Complementary stacked current mirrors (*k*-CCMs)

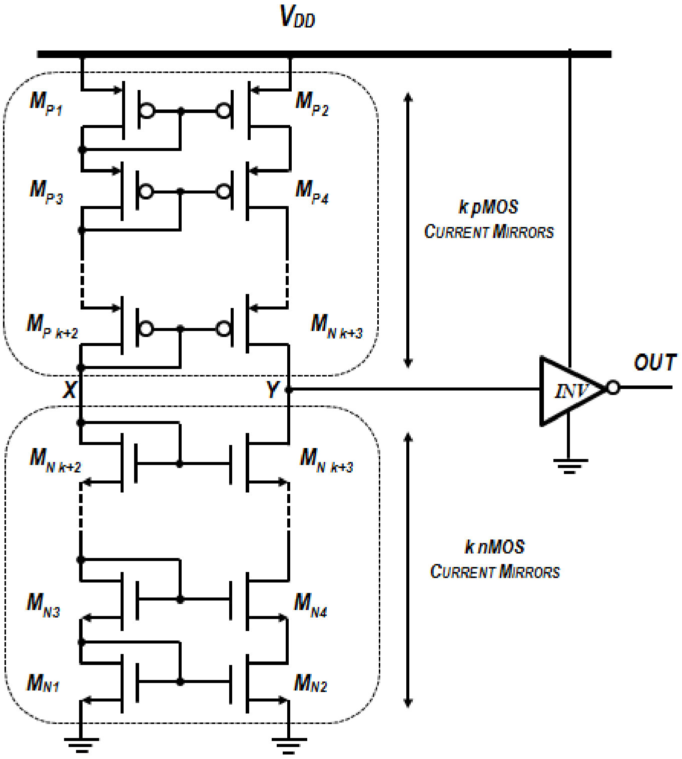


Figure 6.17. Schematic of the *k*-stacked CCMs PUF (*k*-CCMs).

Figure 6.17 shows the generalized schematic of the proposed random bit generator. As depicted from Figure 6.17 the random bit is simply generated by using *k* stacked 1:1 complementary current mirrors (*k*-CCMs). Assuming for all the transistors the subthreshold operation, the generalized expressions for V_X and V_Y are equal to:

$$V_X\left(k>1\right)\approx\frac{V_{DD}}{2}+\frac{nV_T}{2}\ln\left(\frac{\beta_{Pk+2}}{\beta_{Nk+2}}\right)+\frac{1}{2}\sum_{i=}^{k+2}\left[V_{TH,Ni}-\left|V_{TH,Pi}\right|+nV_T\ln\left(\frac{\beta_{Nk+2}}{\beta_{Ni}}\right)-nV_T\ln\left(\frac{\beta_{Pk+2}}{\beta_{Pi}}\right)\right] \qquad (6.6)$$

$$V_Y(k)\approx-g_m{}^{k-1}r_{ds}{}^{k}V_X \qquad \text{for } k>1.$$

Equation (6.6) shows that both V_X and V_Y are equal to $V_{DD}/2$ if a properly balancing between the strength of the pMOS and nMOS transistors is performed, independently from the stacking factor k. From (6.6) a higher number of stacked transistors (i.e. higher k) results in a higher process-dependence thanks to the body effect on the stacked transistors and the deeper subthreshold conduction. Moreover from (6.6) a higher k factor results in a higher small-signal gain, which can reduce significantly the number of unstable bits. On the contrary, increasing k, larger devices are required to balance the strength between the nMOS and pMOS transistors. According to the Pelgrom's law this can results in a lower level of randomness [25]. As a consequence, it is not trivial to understand which is the optimum number of stacked devices to implement a bitcell key PUF.

6.3.5. Sense op-amp PUFs

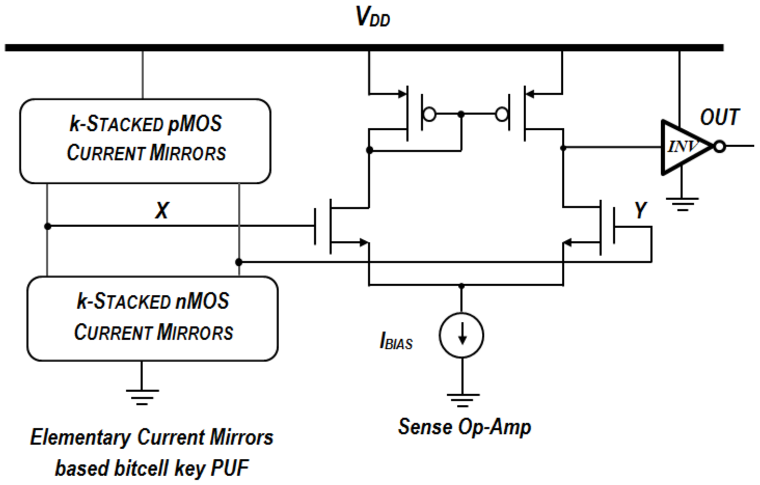


Figure 6.18. Elementary bitcell key PUF with sense op-amp (*k*-CCM_OA).

In the proposed solutions, the number of unstable bits depends from the gain of the output stage. A high gain ensures that also a small deviation with respect to the nominal voltage ($V_{DD}/2$) can result in an output voltage equal to V_{DD} or 0, thus reducing significantly the number of output states which fall in the indefinite region of the output inverter. Different gain-boost techniques (i.e. different output stage amplifiers) can be used to this purpose. Among them, the solution of a sense operational amplifier (Op-Amp) is here explored.

Despite of its simplicity, this solution allows obtaining a significant improvement in terms of gain without compromising the randomness of the core (CCMs), as in the case of more complex solutions (i.e. folder cascode), moreover it allows maintaining a small area for the single cell. The general PUF bitcell generator with a sense Op-Amp (*k*-CCMs_OA) is reported in Figure 6.18. The input pair of the sense Op-Amp receives the voltages from nodes X and Y, which are both equal to $V_{DD}/2$ at the nominal conditions. However, as shown in the previous sub-section, due to the process variations V_X and V_Y can differ in an unpredictable way. By properly sizing the input pair and the active load, the sense Op-Amp allows introducing other sources of variability like the offset and the small signal gain. Additionally, since the input voltages of the sense op-amp can differ substantially each other, the effect of the load mismatch in the common mode response is introduced [26]. Considering the nominal case the sense op-amp boosts the small signal gain of the complementary current mirrors by a factor $gm_N \; rds_P \; // \; rds_N$.

6.3.6. Simulated and measured results

In this section, the validity of the proposed schemes is investigated through simulation results obtained using UMC 65 nm CMOS technology. The main security FOMs were extracted with the main aim to obtain information about the randomness, uniqueness and stability of the response. To obtain information about the goodness of the simulation results a comparison with experimental results is reported where available. Finally, the portability of the proposed approach in the different technology nodes is shown by comparing the simulation results obtained in 65 nm CMOS technology with the results obtained using 180 nm CMOS technology.

6.3.6.1. Impact of the stacking factor *k*

As explained before, it is not trivial to understand the optimum value of the stacking factor *k* for a PUF application. For this reason the effect of the *k* factor on the main security metrics was investigated. For a fair comparison each solution is dimensioned in order to obtain $V_{DD}/2$ on node *X* and *Y* at the nominal conditions of V_{DD}=1 V and T=27°C. In addition, to emphasize the effects of the process variability, the length of the transistors (both nMOS and pMOS) is settled to the minimum value allowed by the chosen design kit (60 nm). The minimum allowed width (80 nm) is employed for all the nMOS transistors. Thus, the only parameter used to balance the strength between β_N and β_P is the width of the pMOS transistors. To enhance the effects of the variations, in the case of the sense op-amp solutions, all the transistors in the amplifier are settled to the minimum size allowed by the chosen design kit. This results in an equivalent DC gain of 45 dB using a bias current of 238 nA.

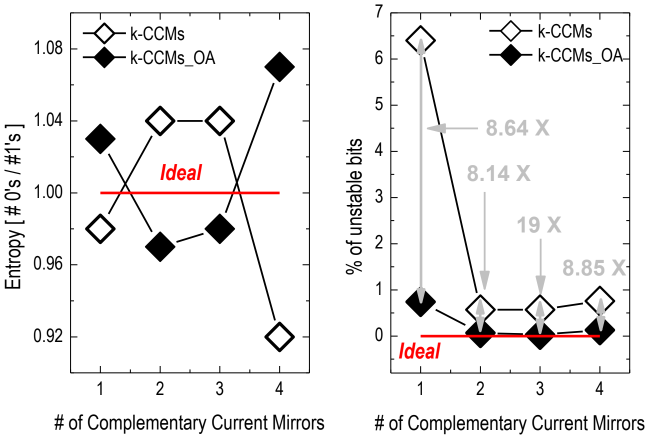


Figure 6.19. Logical 0's / logical 1's ratio in the response and percentage of unstable bits as a function of *k* in the *k*-CCMs and *k*-CCMs_OA solutions.

In Figure 6.19 the effect of the stacking factor *k* on the bias response (0's /1's ratio) and on the percentage of unstable bits is reported both in the case of the *k*-CCMs and *k*-CCMs_OA solution. The results refer to MonteCarlo simulations performed over 10000 runs considering the only intra-die variations. Inspecting Figure 6.19, the balancing between the number of logical 0's and logical 1's is pretty close to the ideal value of 1 in all the simulated cases. The deviation from the ideal value is about 3.5 % in all the cases except for the 1-CCMs and 4-CCMs_OA schemes, where the deviation is about 7 %. The same Figure 6.19 shows that the sense op-amp does not introduce appreciable effects in terms of randomness (deviation from 1 in absolute value). The proposed schemes show good performance also in terms of unstable states. The percentage of "noisy" bits is always below 1% except for the 1-CCMs, where a value of 6.4% was observed. Moreover, as expected, simulation results show a significant benefit in the number of unstable bits in the case of the solutions with sense op-amp. Comparing the results obtained in the *k*-CCMs and the *k*-CCMs_OA, using the sense op-amp, the percentage of unstable bits is reduced by a factor equal to 8.64 × in the case of *k*=1, 8.14 × for *k*=2, 19 × for *k*=3 and 8.85 × for *k*=4. In absolute terms, the best result on the number of unstable bits at the nominal operating conditions are obtained in the 2-CCMs scheme in the case of the solutions without sense op-amp (0.56%) and in the 3-CCMs scheme in the case of the solutions with sense op-amp (0.07%).

6.3.6.2. Impact of the supply voltage and temperature variations

To be robust against possible side attacks [20], the PUF response has to be stable against temperature and supply voltage variations. Table 6.I reports the variation on the entropy (0's /1's ratio) and on the percentage of unstable bits when the operating temperature rises from 27°C to 85°C.

Considering the entropy, the worst case variation is observed in the case of 1-CCMs and 4-CCMs, with a variation of about + 5% in both cases. Better results are observed in the case of the solutions with sense op-amp, where no variation at all are observed except for the 3-CCMs_OA solution, where a variation of -1.02 % is observed.

As in the case of the entropy, also in the case of the percentage of unstable bits, the solutions with sense op-amp show a substantial improvement in the robustness. The only solution which shows a significant variation is the 1-CCMs_OA with an increment on the percentage of unstable bits equal to 2.7 %. Despite that it is worth noting that this variation is negligible since it corresponds to a variation from 0.74% @ 27°C to 0.76 % @ 27°C, representing only two extra unstable cells over 10^4 simulated cells.

TABLE 6.I. TEMPERATURE EFFECT ON ENTROPY AND ON THE PERCENTAGE OF UNSTABLE BITS

Solution	**Entropy**	**% of unstable bits**
	27°C → 85°C	**27°C → 85°C**
1-CCMs	+ 5.10 %	+ 5.47 %
2-CCMs	-0.97 %	- 7.02 %
3-CCMs	- 1.96 %	- 78.95 %
4-CCMs	+ 5.15 %	- 97.4%
1-CCMs_OA	+ 0 %	+ 2.70 %
2-CCMs_OA	+ 0 %	+ 0 %
3-CCMs_OA	-1.02 %	+ 0 %
4-CCMs_OA	+ 0 %	+ 0 %

The variation in the percentage of unstable bits is higher in the case of the *k*-CCMs solutions. Only in the case of the 1-CCMs solution the percentage of unstable bits increases at higher temperatures (+5.47 %), while in all the other cases the percentage of bits which fall around the undefined region of the characteristic of the output stage becomes lower at higher temperatures. In particular the percentage of unstable bits becomes -7.02 %, -78.95% and - 97.4 % lower in the case of the 2-CCMs, 3-CCMs and 4-CCMs, respectively, when the operating temperature rises from 27°C to 85°C. Nevertheless, despite of the large percentage variations, also in this case the absolute variations in the number of unstable bits are very low. Indeed, the percentage of unstable bits goes from 0.56 % to 0.53 % in the case of 2-CCMs, from 0.57 % to 0.12 % for the *3*-CCMs and from 0.76 % to 0.02 % in the case of the 4-CCMs, confirming a very stable behaviour of the proposed architecture also in the case of significant temperature variations. It is important to underline that since every variation in the number of unstable bits represent a bit flipping, the total number of unstable bits coincides with the highest value in the operating range of temperature.

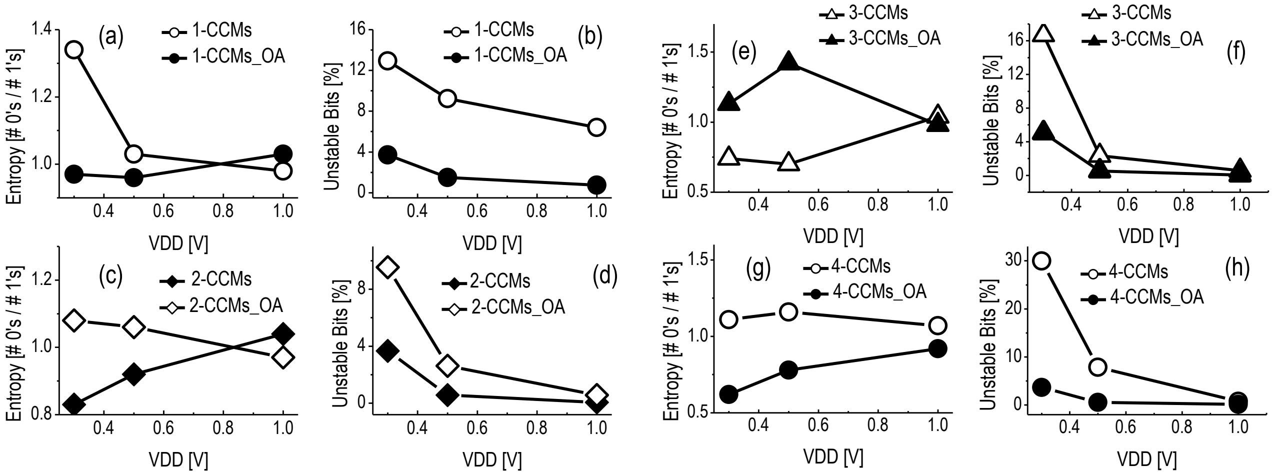


Figure 6.20. Entropy (# logical 0's/ # logical 1's) and percentage of unstable bits for different V_{DD}s.

The effect of the bias variation on the entropy and on the percentage of unstable bits for all the proposed solutions are reported in Figure 6.20 (a)-(h). In particular, the simulations show the variation in the entropy of the PUF response and in the percentage of unstable bits when V_{DD} scales from the nominal value of 1 V (at which the circuits are optimized) to 0.5 V and 0.3 V. As result this analysis considers an extreme variation in the supply voltage, which corresponds to a completely different application scenario with respect to the nominal one. This analysis allows obtaining information about the bias in the PUF response when V_{DD} scale significantly, providing additional information on the sensitivity to side attacks through V_{DD} variations.

Since the output voltage of the single bitcell depends on the balancing between the strength of the nMOS and pMOS transistors, a significant variation of V_{DD} can cause a severe variation in the balancing, which can result in a higher probability of obtaining a logical 0 or 1 on n the response. Inspecting Figure 6.20, the 2-CCMs (Figure 6.20 (c)), 3-CCMs_OA (Figure 6.20 (e)) and 4-CCMs (Figure 6.20 (f)) solutions show a very stable entropy, even for large V_{DD} variations, while in all the other solutions the entropy degrades when V_{DD} scales at very low values (i.e. V_{DD}=0.3 V). The entropy degradation is more significant in the case of the 1-CCMs, 2-CCMs_OA and 4-CCMs_OA for V_{DD}=0.3V.

The same Figure 6.20 shows the variation of the percentage of unstable bits when V_{DD} scales from 1 V to 0.3 V. In all the simulated structures, the percentage of unstable bits becomes higher at lower V_{DDs}. The 2-CCMs and the 2-CCMs_OA schemes show the lowest sensitivity. At the same time, simulation results show large benefits in the number of unstable bits in the solutions with sense op-amp, even in the case of very low values of V_{DD}s.

6.3.6.3. Inter-PUF and intra- PUF HD

Ideally, to obtain a unique and reliable signature, a PUF has to generate a map of CRPs which exhibits a large HD (ideally 50 %) when compared to the CRPs obtained from other PUFs (inter-PUF *HD*) and an HD equal to zero when compared to the CPRs generated by the same PUF (intra-PUF HD even in the case of different $V_{DD}s$ and temperatures.

TABLE 6.II. SIMULATED INTER-PUF HD

	Inter-PUF HD			*Intra-PUF HD*
	V_{DD}=1 V	V_{DD}=0.5 V	V_{DD}=0.3 V	*@ diff $V_{DD}s$ and Temp.*
1-CCMs	128.00	127.94	128.06	0.30
2-CCMs	127.79	127.75	127.34	0.22
3-CCMs	127.49	123.93	125.21	0.36
4-CCMs	127.74	126.04	120.61	0.39
1-CCMs_OA	127.23	127.83	127.77	0.21
2-CCMs_OA	127.49	128.05	127.79	0.29
3-CCMs_OA	127.49	123.96	127.64	0.28
4-CCMs_OA	127.49	127.52	127.52	0.33

To obtain information about the uniqueness and the reliability of the proposed schemes, the intra- and inter- PUF HD is evaluated considering a 256-bit word as key. Thus, for the chosen word length, the ideal inter-PUF HD is equal to 128, while the ideal intra-PUF HD is equal to 0. In Table 6.II the simulated inter-PUF and intra-PUF HD for all the proposed schemes are reported. The inter-PUF HD was evaluated at V_{DD}=1V, V_{DD}=0.5V and V_{DD}=0.3V, respectively. The values reported in Table 6.II are evaluated by averaging the HD over 10^5 comparisons. Inspecting Table 6.II the inter-PUF HD is pretty close to the ideal value of 128 in all the solutions, both in the case of the nominal operating condition (V_{DD}=1V) and in the case of lower $V_{DD}s$. Only the 3-CCMs and 4-CCMs solutions show a significant degradation of the inter-PUF HD at lower $V_{DD}s$. This confirms the good performance of the proposed architecture in terms of uniqueness; moreover it confirms the high level of robustness against V_{DD} variations. The intra-PUF HD evaluated in the same condition (by averaging the HD of 10^5 comparisons using 256-bit words) compares the CRPs generated by the same PUF at different $V_{DD}s$ (1V, 0.5 V, 0.3V) and temperatures (from 0 °C up to 85 °C). Also in this case all the solutions show a value that is pretty close to the ideal value of 0, thus confirming a high level of robustness even in the case of severe temperature variations and supply voltage variations.

To obtain information about the randomness of the different solutions, the bias of the response is evaluated. For a completely unbiased response the bias has to be equal to 128, which correspond to an equal number of logical 1's and 0's in the response. The results of such analysis are reported in Table 6.III. For each solution the bias is evaluated at the nominal supply voltage of V_{DD}=1 V and at V_{DD}=0.5 V and V_{DD}=0.3 V. Also here the results are obtained by averaging, for each solution, the bias of 10^5 keys. At the nominal operating supply voltage (V_{DD}=1 V) the best solution in terms of randomness is the 1-CCMs, anyway a good level of randomness is observed in all the proposed solutions. The worst case deviation from the ideal value is observed in the case of the 4-CCMs solution, where the variation from the ideal bias is only 3.9 %.

When V_{DD} scales from V_{DD}=1 V to V_{DD}=0.5 V, a severe degradation of the bias is observed in the case of the 3-*CCM*s and 3-*CCMs*_OA solutions. In these solutions the deviation from the ideal bias becomes equal to 19.39 % and 18.69 %, respectively, while in the other solutions a deviation below the 10% is observed.

When V_{DD} scales from V_{DD}=1 V to V_{DD}=0.3 V the worst absolute deviation from the ideal bias is observed in the 3-CCMs (17.06%) and 4-*CCMs* (19.84%) solution, while in all the other solutions the deviation from the ideal value is below 11 %. From this analysis the best solutions in terms of randomness, also in the case of extreme V_{DD} variations, are the 1-CCMs and 1-CCMs_OA with a deviation from the ideal bias of only 1.20% (0.87%) @ V_{DD}=1V, 1.35% (1.91%) @ V_{DD}=0.5V and 1.52 % (1.46%) @ V_{DD}=0.3V for the 1-CCMs (1-CCMs_OA). Finally the randomness was analysed using NIST test [28]-[29]. For a better understanding of the meaning of the different suites in the test, the reader is referred to [29]. For the NIST tests, a value of the threshold *p* equal to 0.001 was used. This value indicates that there is 99.9% confidence that the data is random. The results of this analysis performed over the different schemes are reported in Table 6.IV. Except for the 3-CCMs OA and the 4-CCMs OA, which fail both in one test suite (block frequency and frequency respectively), all other solutions pass all the suites, demonstrating a very high level of randomness. It is worth noting that the selected value of the threshold allows considering as "passed" only the solutions which have the 100 % of confidence of generating a completely random bitstream.

To proof the validity of the proposed bitcell key PUF, silicon prototypes of the 2-CCMs and 2-CCMs_OA have been recently fabricated and tested [15]. The performance of these solutions against the state-of-the-art solutions are reported in Table 6.V.

TABLE 6.III. SIMULATED INTRA-PUF HD

	V_{DD}=1 V	V_{DD}=0.5 V	V_{DD}=0.3 V
1-CCMs	129.53	126.27	126.06
2-CCMs	125	133.33	139.71
3-CCMs	125.37	150.19	147.21
4-CCMs	132.96	143.96	158.36
1-CCMs_OA	126.89	130.44	129.87
2-CCMs_OA	125.03	124.05	123.43
3-CCMs_OA	129.56	105.64	120.06
4-CCMs_OA	123.42	118.78	121.58

TABLE 6.IV. NIST TEST RESULTS

	w/o sense Op-Amp				*sense Op-Amp*			
	k=1	*k=2*	*k=3*	*k=4*	*k=1*	*k=2*	*k=3*	*k=4*
Approximated Entropy	0.205	0.429	0.044	0.299	0.291	0.820	0.344	0.035
Block Frequency	0.764	0.593	0.722	0.423	0.001	0.815	0.144	0.045
Cumulative Sums	0.241	0.288	0.058	0.061	0.024	0.720	0.137	0.002
FFT	0.305	0.766	0.487	0.716	0.128	0.064	0.233	0.643
Frequency	0.182	0.182	0.036	0.030	0.016	0.639	0.083	0.001
Linear Complexity	0.095	0.353	0.849	0.934	0.744	0.306	0.217	0.744
Longest Run	0.210	0.285	0.798	0.031	0.230	0.799	0.223	0.042
Non Overlapping Template	0.241	0.446	0.597	0.509	0.571	0.291	0.090	0.851
Rank	0.334	0.334	0.334	0.793	0.334	0.334	0.334	0.334
Runs	0.561	0.384	0.266	0.531	0.468	0.888	0.899	0.362
Serial	1.000	0.340	0.784	0.820	0.920	0.474	0.243	0.085

***Failed*=** ▨ ***Passed*=** □

TABLE 6.V. COMPARISON BETWEEN 2-CCMS AND THE OTHER STATE-OF-THE-ART SOLUTIONS

	[19]	[15]	[18]	[14]	[21]	[22]
Technology	40 nm	65 nm	22 nm	90 nm	350 nm	65 nm
Architecture	Digital	Analog	Digital	Analog	Analog	Digital
CRPs	**5.5 E+28**	3040	1 (chip ID)	1 E+ 25	-	1 (chip ID)
BER in tipycal case	**0**	<2%	-	0.009 %	1.3 %	0
Temperature (°C)	-25: 125	25 : 85	25 : 50	25 : 125	-25 : 125	0 : 85
V_{DD} (V)	0.7 – 1.2	0.6 – 1	0.7 – 0.9	**0.6 ±10 %**	1.1 – 5	1.1±10 %
BER (worst-case)	**< 1 E-8**	<6%	0.97 %	0.1 %	5 %	0
Bit rate (Mb/sec)	1.6	-	**2000**	0.006	1.5 (max)	625
Mean Inter HD	**50.01 %**	50.13 %	51 %	-	-	63 %
Mean Intra HD	**0.01**	0.86	6.87	-	-	-
Core area (µm^2)	**845**	65700	1193	35000	23496	1242
Power (µW)	28.4	>50	**25**	38	250 (typ.)	212.5
Efficiency (pJ/bit)	17.75	**0.015**	0.19	6080	8330	0.34

In bolt the best performance are highlighted. The proposed PUF [15] shows the best performance in terms of efficiency while maintaining a very high level of Bit-Error-Rate (BER) and a value of the intra-HD and inter-HD close to the ideal ones. Moreover the experimental results have confirmed the superiority of the proposed solution with respect to the conventional delay-based PUFs and memory-based SRAM PUFs [15] in terms of stability against temperature and V_{DD} variations. In the next section, a comparison between simulated and measured results is reported where data are available.

6.3.6.4. Comparison between measured and simulated results

The comparison between the simulated and measured results has the main aim to demonstrate the validity of the simulation results reported in the previous sections. An exhaustive comparison between simulated and measured results is reported in Table 6.VI and Figure 6.21 (a)-(b).

Specifically Table 6.VI reports the comparison between the simulated and measured results obtained at the nominal operating conditions of V_{DD}=1V and T=25°C. A good agreement is observed in the case of the bias (entropy), inter-PUF HD and intra- PUF HD, while the measured percentage of unstable bits is about 2.85 × and 19.5 × times higher than the simulated one in the case of the 2-CCMs and 2-CCMs_OA, respectively. Figure 6.21 shows the comparison between simulated and measured percentage of unstable bits in the case of the 2-CCMs (Figure 6.21 (a)) and 2-CCMs_OA (Figure 6.21 (b)) when V_{DD} scales.

TABLE 6.VI. SIMULATED AND MEASURED PERFORMANCE FOR THE 2-CCMS AND 2-CCMS_OA SCHEME

	Entropy [0's / 1's]		**% of unstable bits**		**Inter-PUF HD**		**Intra-PUF HD**	
	Sim.	*Meas.*	*Sim.*	*Meas.*	*Sim.*	*Meas.*	*Sim.*	*Meas.*
2-CCMs	1.04	0.99	0.56	1.60	127.79	128.36	0.22	0.34
2-CCMs_OA	0.97	1	0.07	1.56	127.49	128.38	0.29	0.34

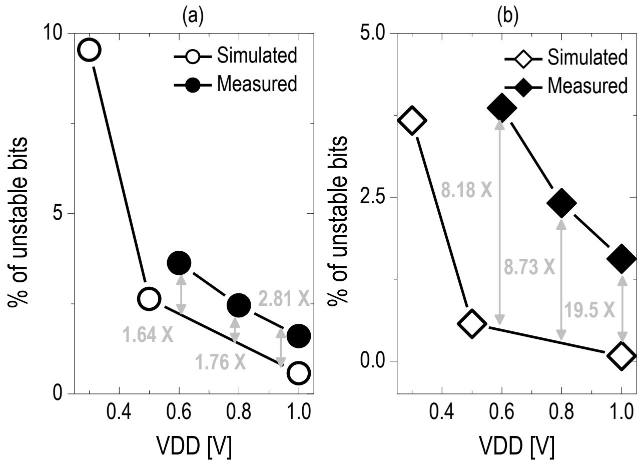


Figure 6.21. Simulated and measured percentage of unstable bits in the case of the 2-CCMs (a) and 2-CCMs_OA (b).

In both cases, the simulated percentage of unstable bits is lower than the measured one. In particular, in the case of the 2-CCMs the simulated percentage of unstable bits is about 2.81 ×, 1.76 × and 1.64 × times lower than the measured values at V_{DD}=1 V, V_{DD}=0.5V and V_{DD}=0.3V, respectively. In the case of the 2-CCMs_OA, the simulated percentage of unstable bits is 19.5 × times lower than the measured value at V_{DD}=1V, while the simulated value is 8.73 × and 8.18 × times lower than the measured value at V_{DD}=0.5 V and V_{DD}=0.3 V, respectively. The latter comparison confirms the optimistic estimation in the percentage of unstable states in the case of the simulations. In addition, measurement results show a negligible improvement in the number of unstable bits by using the sense op-amp, while from simulation results a large benefit in the unstable bits is observed for the solutions with sense op-amp. Both simulation and measurement results show a similar trends in the variation of the unstable states when V_{DD} scales at lower values.

6.3.6.5. Stability against supply noise

One of the main advantages of the proposed scheme consists in its static behaviour. Differently of the delay-based and memory-based PUFs, in the proposed approach the single random bit is statically generated, thus no transient effect is involved in the generation of the secret key [15]. As a consequence, the proposed approach is in able to guarantee a high level of robustness against any kind of noise (i.e. supply voltage noise) compared to the memory-based and delay-based PUFs, where a significant percentage of the bit-flipping due to the supply noise is observed [13].

Generally speaking, the bit-flipping happens when V_{OUT} changes its logical state as a consequence of a variation in V_{DD} (i.e. noise on supply voltage). In this case in fact a bit located in the voltage range defining the logical state 1 (0) can fall in the indefinite region of the output stage or even in the range of voltages defining the logical 0 (1). It is quite obvious that the V_{DD} noise also affects the output stage which defines the voltages range corresponding to the logical state 1 and logical state 0. A severe variation of this range can cause a high number of bit-flipping independently from the stability of the solution employed as PUF. To proof the stability of the proposed static PUF generator, a Gaussian White Noise (GWN) was applied on the supply voltage of each solution. Without loss of generality the nominal condition of V_{DD}=1V and T=27°C were considered. Two different amplitudes of the GWN were applied, 5% and 10% of V_{DD}, corresponding respectively to an optimistic and pessimistic case of the supply voltage noise. The simulation considers also the effect of the noise on the output stage (i.e. buffer inverter). For better understanding the possible effects of the supply voltage noise on the output bits, in Figure 6.22 the main transient behaviours of the output bit in the case of a GWN are reported. In the ideal case, represented by the curve (a) of Figure 6.22, the GWN does not change the logical state of the output voltage, thus the output bit is insensitive to the V_{DD} noise. This behaviour can be observed also in the case of a native unstable bit. On the other hand, the V_{DD} noise can change the logical state of the output bit as in the case of the curves (b) (from logical 1 to logical 0), (c) (from logical 0 to logical 1) and (d) (from logical 0 to unstable state) of Figure 6.22. In all these cases a bit flip is observed, thus the output bit is sensible to the V_{DD} noise.

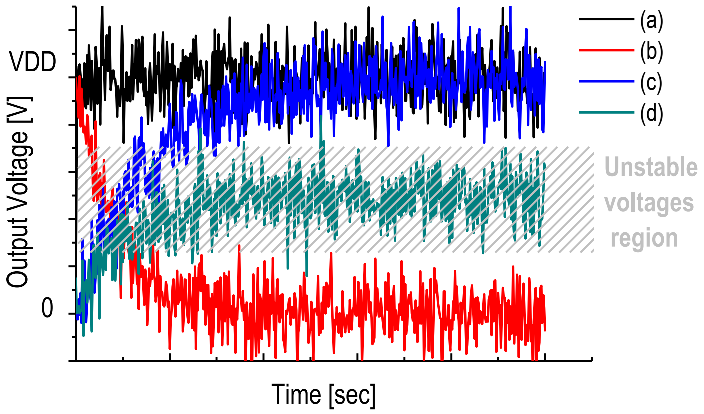


Figure 6.22. Main transient behaviours of the output bit in the case of a GWN on supply voltage.

TABLE 6.VII. EFFECT OF THE GWN ON THE PERCENTAGE OF UNSTABLE BITS

	nominal	GWN=5 %	GWN=10 %
1-CCMs	6.40 %	6.91 %	7.91 %
2-CCMs	0.56 %	0.56 %	0.57 %
3-CCMs	0.27 %	0.27 %	0.27 %
4-CCMs	0.80 %	0.91 %	1.01 %
1-CCMs_OA	0.80 %	0.83 %	0.83 %
2-CCMs_OA	0.07 %	0.08 %	0.11 %
3-CCMs_OA	0.04 %	0.06 %	0.08 %
4-CCMs_OA	0.11 %	0.14 %	0.32 %

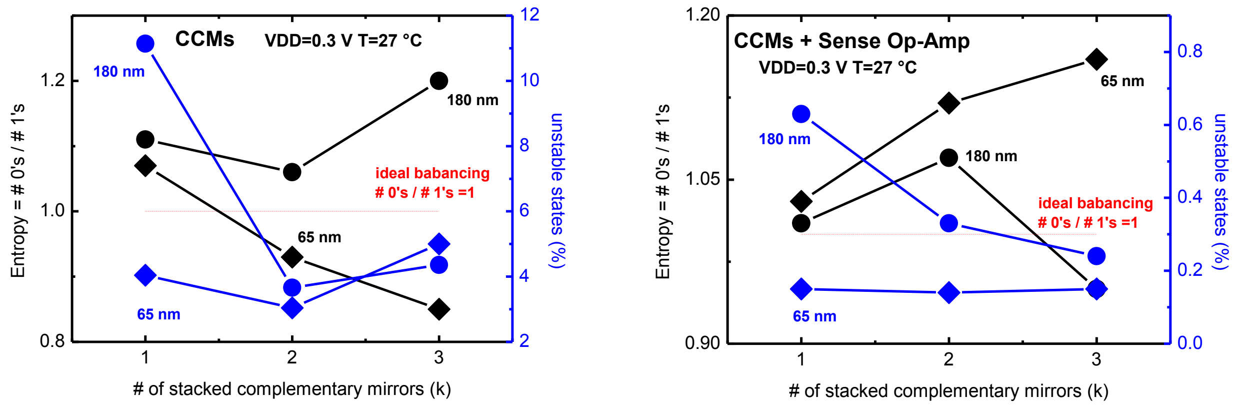
In Table 6.VII the effect of the GWN on the percentage of unstable bits is reported both in the case of the *k*-CCMs and of the *k*-CCMs_OA. As previously explained, in the unstable bits are also included the bits passing from stable to unstable states and vice-versa.

As expected, due to the higher *PSRR*, all the solutions with the sense op-amp show better results with respect to the mirrors-only counterparts both in the case of a GWN=5% and GWN=10%. However, except for the 1-CCMs solution, all the schemes exhibit a percentage of unstable bits lower than about 1 % even in the case of high level of noise on V_{DD} (i.e. GWN=10%), thus confirming the very low level of native unstable bits in the proposed schemes and the high stability against supply voltage noise. Additionally, this analysis confirms that the lowest percentage of unstable bits is observed in the 3-CCMs_OA solution. At the same time the 1-CCMs represents the worst solution both in terms of number of unstable bits and in the noise sensitivity.

6.3.6.6. Technology impact

In this section the performance of the proposed solutions are explored considering the impact of the technology node. To this aim simulation results obtained by using 65 nm CMOS technology are compared with the results obtained by using 180 nm CMOS technology. The comparison is performed considering the operating conditions of V_{DD}=0.3 V and T=27°C. All the solutions are optimized for working at V_{DD}=0.3 V which can be easily obtained by dimensioning the transistors in order to obtain 0.15 V on node X and Y in the nominal process corner. The comparison is performed considering *k*=1, 2, 3 and by considering both the entropy and the percentage of unstable bits. For a fair comparison all the transistors in both technologies are selected in order to ensure the subthreshold conduction. Finally, the sense op-amp is biased to obtain in the nominal operating conditions the same DC gain in both technologies.

Figure 6.23(a) reassumes the impact of the number of complementary current mirrors in the case of the *k*-CCMs solutions in the two different CMOS technologies. When optimized to work at V_{DD}=0.3 V, the balancing between the number of logical 0's and 1's is 1.07 in the case of the 1-CCMs, 0.93 for 2-CCMs and 0.85 for 3-CCMs when implemented by using 65 nm CMOS technology. Using 180 nm CMOS technology, the balancing becomes 1.11 in the case of 1-CCMs scheme, 1.06 for 2-CCMs and 1.11 for 3-CCMs.



(a) (b)

Figure 6.23. Effect of the # of stacked complementary mirrors in 65 nm and 180nm CMOS technology on randomness and unstable states. The figure 6.23(a) reports the comparison in the case of the *k*-CCMs solutions while figure 6.23(b) reports the same comparison in the case of the *k*-CCMs_OA solutions.

Considering the results reported in Figure 6.23(a) the 2-CCMs solution shows the balancing in both technologies. Regarding the percentage of unstable bits, due to the higher gain, the 65 nm CMOS technology shows the best results in the case of stacking factors equal to 1 and 2.

The percentage of noisy bits in the case of 1-CCMs is equal to 4.04 % which is about 2.56 × times lower than the implementation in 180 nm (11.14 %), while in the case of the 2-CCMs solution the percentage of unstable bits is 3.04 % about 1.2 × times lower than the equivalent 180 nm implementation (3.04%).

Considering the 3-CCMs scheme, the percentage of unstable bits is 5% when implemented in 65 nm which is 1.14 × times higher with respect to the case of 180 nm technology (4.36%).

Thus, also regarding the unstable states, the 2-CCMs solution shows the best results in both technologies.

In Figure 6.23(b) the same comparison is performed by considering the effect of the additional sense op-amp. It is worth noting that the number of unstable states is significantly lower in all the simulated cases (<1 %) in comparison to the case of the solutions without op-amp. When implemented in 65 nm the percentage of unstable states is equal to 0.15% for the 1-CCMs_OA solution, 0.14% for the 2-CCMs_OA and 0.15% for 3-CCMs_OA while a percentage of 0.63 %, 0.33 % and 0.24% are observed for the same solutions implemented in 180 nm.

Comparing the randomness of the output bits generated by the different solutions, simulation results on 65 nm show a ratio between the number of logical 0's and 1's equal to 1.03 versus 1.01 obtained in 180 nm in the case of the CCMs_OA scheme. In the case of the 2-CCMs_OA solution this ratio becomes equal to 1.07 if implemented in 180 nm while a ratio of 1.12 is observed in the 65 nm implementation. Finally, in the case of the 3-CCMs_OA the balancing is equal to 0.95 in the case of the 180 nm against a value of 1.16 observed in the case of the 65 nm implementation.

6.3.7. Conclusion

A new class of silicon Physical Unclonable Functions consisting in complementary current mirrors was presented. In order to understand the sources of randomness and to provide the main design considerations, an analytical analysis was presented. The design considerations for enhancing the randomness while maintaining stability against temperature, supply voltage and noise variations have been discussed. To understand the performances of the different solutions in terms of uniqueness, randomness and reliability, also in the case of very low-voltage applications and for a wide range of operating temperatures, for all the proposed solutions an extensive set of MonteCarlo simulations were performed.

Where available, simulations results were compared with the experimental results. Simulations results track well experimental results in estimating bias response, intra- and inter- PUF Hamming Distance while simulations results underestimate the percentage of unstable bits especially in the case of the solutions which use the sense op-amp amplifier. Despite that, also in the case of the experimental results, the percentage of unstable bits is lower than 2 % at the nominal operating conditions of V_{DD}=1 V and T=25°C, demonstrating a high level of reliability of the generated bitstream. For all the proposed solutions a very high level of robustness against supply voltage and temperature variations is observed. Inter- and intra- Hamming distances are very close to the ideal values of 50 % and 0% respectively. Considering the NIST test results, all the solutions show additionally a very high level of randomness. Finally the validity of the proposed approach was demonstrated through different technologies.

6.4. A 2T voltage divider PUF

6.4.1. Introduction

In this section, a new extremely compact circuit solution for silicon-based static PUFs is presented. The proposed solution exploits the variability of a simple voltage divider implemented by two identical series-connected nMOSFETs in UMC 65nm CMOS process to generate a random and stable nanokey. Both the transistors are biased in subthreshold regime to enhance the output voltage dispersion and consequently the variability in the PUF response. The key generation is obtained by comparing the analog outputs of a pair of voltage dividers. Monte Carlo simulations on 10,000 samples have been performed to deduce the design guidelines for transistors' sizing aimed at ensuring a high robustness of the PUF response against noise, supply voltage and temperature variations. When compared to some state-of-the-art PUF designs, the proposed circuit solution proves to be a very promising and competitive candidate for implementing analog and static PUFs featuring small area occupancy, low-power features, and high reliability.

6.4.2. Proposed circuit solution

As shown in Figure 6.24, the core of the proposed PUF solution consists of a 2T voltage divider implemented by the series of two identical nMOSFETs operating in the deep subthreshold region (V_{GS}=0). Thus, considering $V_{DS}>4V_T$, the current in M$_1$ and M$_2$ becomes as follows:

$$I = \beta_0 \exp\left(-\frac{V_{TH}}{nV_T}\right), \qquad (6.7)$$

where $\beta_0=\mu_N C_{OX}(W/L)V_T^2$. Considering the relationship between V_{TH}, V_{DS} and V_{BS} already defined in (2.11), the two threshold voltages become equal to:

$$V_{TH1} = V_{TH0,1} - \lambda_D\left(V_{DD} - V_{OUT}\right), \qquad (6.8)$$

$$V_{TH2} = V_{TH0,2} - \lambda_D V_{OUT}\,. \qquad (6.9)$$

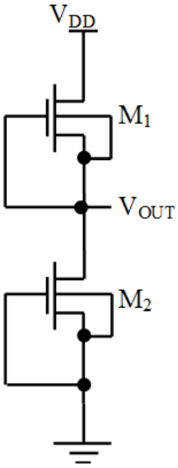


Figure 6.24. Schematic of the proposed basic circuit based on a 2T subthreshold nMOS voltage divider.

Setting the current through M$_1$ and M$_2$ equal and combining (6.7), (6.8) and (6.9), an analytical expression for V_{OUT} can be easily obtained:

$$V_{OUT} = \frac{V_{DD}}{2} + \frac{V_{TH0,2} - V_{TH0,1}}{2\lambda_D} + \frac{nV_T \ln\left(\frac{\beta_{0,1}}{\beta_{0,2}}\right)}{2\lambda_D}, \tag{6.10}$$

where λ_D and n are assumed to be equal in the two transistors. The expression in (6.10) shows that the output voltage of the 2T voltage divider is strongly dependent on the mismatch between M_1 and M_2, caused by the random process variations, in terms of the difference of the threshold voltages, related to the second term in (6.10), and of the ratio $\beta_{0,1}/\ \beta_{0,2}$, related to the third term in (6.10). It is worth pointing out that, in absence of mismatch between M_1 and M_2, the second and third terms in (6.10) become zero and, accordingly, V_{OUT} is the half of the supply voltage (V_{DD}). By considering only threshold voltage mismatch and by assuming that the variations on the V_{TH0} of the two transistors are independent, the standard deviation of V_{OUT} is related to the standard deviation of V_{TH0} by the following relationship:

$$\sigma_{V_{OUT}} = \frac{\sigma_{V_{TH0}}}{\sqrt{2}\lambda_D}. \tag{6.11}$$

Since λ_D values are significantly lower than one, especially for relatively long MOSFETs, the threshold voltage variability is emphasized on the output node of the proposed circuit. In the same way, the β_0 mismatch in (6.10) gives an additional contribution to the output voltage variability and also this contribution is inversely proportional to λ_D. The amplification effect of λ_D on the random transistor mismatch represents the key property of the proposed circuit.

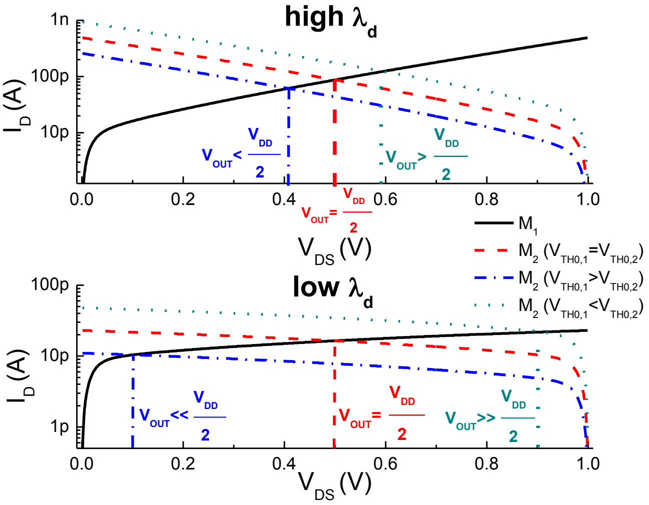


Figure 6.25. I_D-V_{DS} curves for the two transistors belonging to the 2T voltage divider in case of a (top) high and (bottom) low value of λ_D. In the two graphs, the operating point of the divider is highlighted in case of absence of mismatch between M_1 and M_2 (i.e., $V_{TH0,1}$=$V_{TH0,2}$) and in case of threshold voltage mismatch.

The effect of λ_D on the operating point of the circuit is well illustrated in the sketches of Figure 6.25, where the I_D-V_{DS} curves of M_1 and M_2 are reported for different values of λ_D and different V_{TH0} mismatches. According to (6.10), in the case of absence of mismatch between M_1 and M_2, V_{OUT} is equal to $V_{DD}/2$, while, in presence of mismatch, the operating point of the circuit deviates from $V_{DD}/2$. In particular, for a given mismatch condition, a lower value of λ_D can lead to a significant deviation of V_{OUT} from $V_{DD}/2$, thus resulting in a strong amplification of random mismatch.

The amplification effect of λ_D on the random mismatch is also quantified in Figure 6.26, where the statistical distributions of V_{TH0} and V_{OUT} are reported for different values of λ_D. It can be noted that, starting from a V_{TH0} standard deviation of a few millivolts, the amplification effect due to low values of λ_D can lead to a very high standard deviation of V_{OUT}, in agreement with (6.11). Moreover, Figure 6.26 demonstrates that the major contribution to the variability of V_{OUT} can be ascribed to the second term in (6.10), related to the V_{TH0} mismatch between M_1 and M_2. In fact, for the two λ_D values here considered, the statistical fluctuations of V_{TH0} determine more than 85% of the V_{OUT} standard deviation in both the two cases.

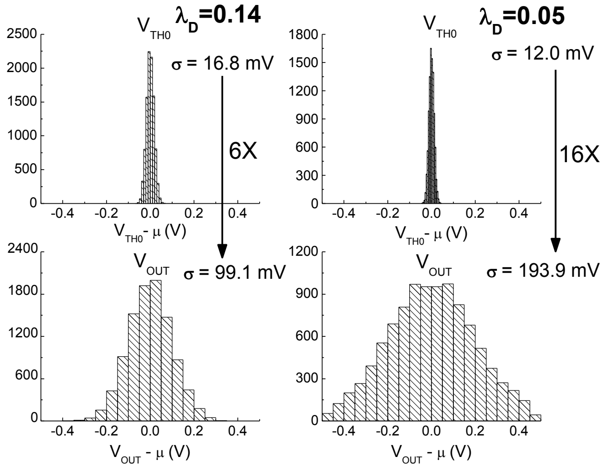


Figure 6.26. V_{TH0} (top) and V_{OUT} (bottom) statistical distributions at V_{DD}=1V in case of a (left) high and (right) low value of λ_D. The graphs refer to the use of low threshold voltage (*LVT*) transistors of the 65nm CMOS technology with different channel length values, *L*=60nm and *L* =120nm, respectively, and the same channel width *W*=500nm.

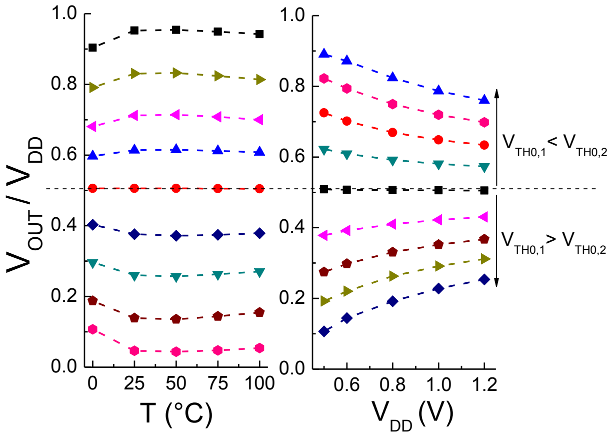


Figure 6.27. V_{OUT}/V_{DD} as a function of (left) temperature (for V_{DD}=1V) and (right) supply voltage (for T=25°C) under transistor mismatch variations (the graphs refer to the use of low threshold voltage (LVT) transistors of the 65nm CMOS technology adopted in this work with *L*=120nm and *W*=500nm).

As it will be clarified in the next section, in order to generate a PUF response robust against temperature and bias variations, the output voltage of the proposed circuit should be insensitive to temperature and should scale proportionally with V_{DD}. Figure 6.27 reports the trend of V_{OUT} normalized with respect to the supply voltage (V_{OUT}/V_{DD}) as a function of temperature (within the range 0-100°C for V_{DD}=1V) and supply voltage (within the range 0.5-1.2V for T=25°C) for different transistor mismatch cases. According to (6.10), we can observe that, in the case of a small mismatch between M_1 and M_2, V_{OUT}/V_{DD} is quite insensitive to *T* and V_{DD}. On the contrary, due to the presence of the second and third terms in (6.10), a greater transistor mismatch leads to a

V_{OUT}/V_{DD} more sensitive to T and V_{DD} variations, especially in the range of lower T and V_{DD}, as shown in Figure 6.27.

Moreover, it is worth noting that the behaviour of the proposed circuit does not suffer of aging [30], since both nMOSFETs are biased in the subthreshold region with V_{GS}=0.

6.4.3. Simulation setup and FOMs

As above claimed, the proposed PUF is based on the amplification of random mismatches between the devices belonging to a subthreshold 2T voltage divider. This effect can be exploited in single-ended or in differential sensing mode to generate random bits. In single-ended mode, the random bit can be generated by comparing the output of a generic 2T voltage divider with a reference voltage at $V_{DD}/2$. In differential mode, the random bit can be generated as a function of the difference between the analog outputs of two nominally identical voltage dividers. In this work, the differential-mode configuration is exploited, whose output is given by:

$$\Delta V_{OUT} = V_{OUT\_1} - V_{OUT\_2} = \frac{\left(V_{TH0,2\_1} + V_{TH0,1\_2}\right) - \left(V_{TH0,1\_1} + V_{TH0,2\_2}\right)}{2\lambda_D} + \frac{nV_T \ln\left(\frac{\beta_{0,1\_1}\beta_{0,2\_2}}{\beta_{0,2\_1}\beta_{0,1\_2}}\right)}{2\lambda_D}, \tag{6.12}$$

where V_{TH0,i_j} and β_{0,i_j} are parameters of the *i-th* transistor belonging to the *j-th* voltage divider.

The differential output is a random variable with a zero mean and a standard deviation ($\sigma\Delta V_{OUT}$), which, in the same hypothesis of (6.11), is given by:

$$\sigma_{\Delta V_{OUT}} = \frac{\sigma_{V_{TH0}}}{\lambda_D}. \tag{6.13}$$

Therefore, an output comparator can be used to find the polarity of ΔV_{OUT} and, consequently, to generate a random response bit for the PUF secret key [31].

An extensive set of Monte Carlo (MC) simulations on 10,000 samples has been performed to investigate the effect of transistor sizing in the proposed circuit on the variability and the robustness of the PUF response. Voltage dividers have been implemented using low voltage threshold (LVT) nMOS transistors of the low-leakage (LL) 65nm UMC process technology. It is worth pointing out that, in the first part of the reported analysis, the simulation setup did not include the output comparator with the aim of decoupling the performance evaluation of the voltage divider pair from the specific design of the output stage. Nevertheless, in order to quantify the number of unstable (or "noisy") bits, i.e. the bits that cannot be used in the PUF secret key because they fall around the undefined region of the characteristic of the output stage and, consequently, are the most sensitive to noise and then can flip even at nominal conditions [15], [31], in the proposed analysis a parameter called "decision margin" ($M_{\Delta V_{OUT}}$) has been introduced. Therefore, at a given environmental condition, in the performed simulations, the "noisy" bits have been found in correspondence of absolute values of ΔV_{OUT} lower than the considered decision margin.

According to this approach, the following FOMs have been considered to evaluate the robustness of the PUF response:

- the percentage of nominal unstable bits (ub_N), related to the "noisy" bits at nominal condition (T=25°C and V_{DD}=1V);
- the percentage of temperature bit-flipping (bf_T), related to the bits which change the logical state and/or become "noisy" at operating temperatures different from the nominal one (within the range 0-100°C for V_{DD} fixed to 1V);

- the percentage of supply voltage bit-flipping (bf_V), related to the bits which change the logical state and/or become “noisy” at operating V_{DD}s different from the nominal one (by considering a ±20% V_{DD} variation at T=25°C);

- an overall *bit-robustness FOM* (FOM_{rob}), calculated as the quadratic sum of the previously defined three FOMs [31]:

$$FOM_{rob} = \sqrt{ub_N^{\,2} + bf_T^{\,2} + bf_V^{\,2}}\ . \tag{6.14}$$

Obviously, the evaluation of the previously defined FOMs depends on the chosen value for the decision margin. This is clarified in Figure 6.28, which reports an example of ub_N calculation as a function of $M_{\Delta V_{OUT}}$. As easily understandable, a higher $M_{\Delta V_{OUT}}$ leads to lower robustness.

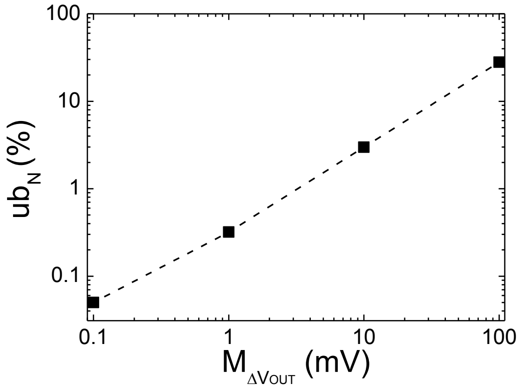


Figure 6.28. Nominal unstable bits (ub_N) vs decision margin ($M_\Delta V_{OUT}$).

In the following, a $M_{\Delta V_{OUT}}$ equal to 10mV is considered which could be a conservative value to take into account the effect of noise on the PUF performance in presence of a non-ideal output voltage comparator. Moreover, it is worth emphasizing that this choice did not affect the main results of the performed comparative analysis, reported in the following section, aimed at providing the guidelines for the sizing of the devices belonging to the 2T voltage divider pair.

6.4.4. Simulation results

6.4.4.1. Analysis and optimization of the voltage divider pair

This section reports the simulation results in terms of variability and robustness of the PUF response by varying the transistor sizing of the 2T voltage divider pair. Figure 6.29 illustrates the standard deviation of ΔV_{OUT} and u_{bN} trends as a function of the transistor channel width (W) and length (L) at nominal condition. As expected, a higher variability of ΔV_{OUT} corresponds to a smaller value of nominal unstable bits. In particular, for a given L, lower W values lead to higher ΔV_{OUT} standard deviations and hence lower u_{bN}. On the contrary, the effect of L on the ΔV_{OUT} variability is more complex due to the fact that the DIBL coefficient λ_D typically increases as L decreases [32]. Therefore, for a given W, starting from the minimum allowed length (60nm for the chosen design kit) and increasing L, λ_D correspondingly decreases and leads to an increase of ΔV_{OUT} variability, according to (6.12), and, consequently, to a reduction of u_{bN}, despite the reduction of random mismatch variations due to the larger size.

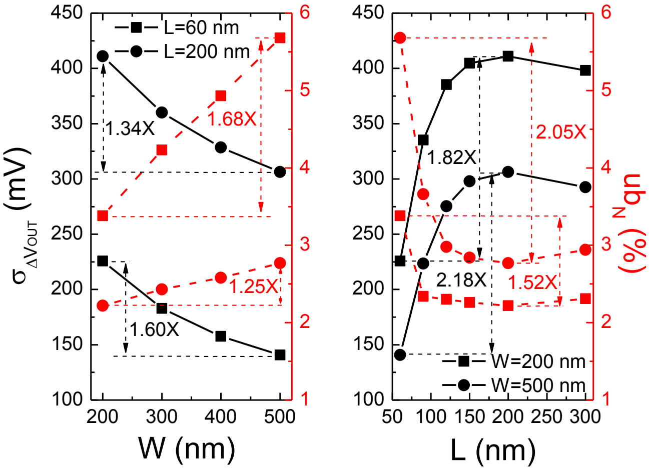


Figure 6.29. ΔV_{OUT} variability and ub_N as a function of transistor channel (right) width (W) and (left) length (L) at nominal condition (T=25°C and V_{DD}=1V).

This effect is predominant up to a certain value of L (in Figure 6.29, up to 200nm), after which the decrease of random process variations at larger sizes becomes the dominant effect and, accordingly, a further increase of L results in a decrease of ΔV_{OUT} variability. Figure 6.29 shows that the ΔV_{OUT} of the voltage divider pair can reach very high variability ($\sigma\Delta V_{OUT}$ ranges from 150mV up to about 400mV), corresponding to u_{bN} down to about 2%.

Figure 6.30 reports the simulation results in terms of b_{fV}, calculated at T=25°C, as a function of transistor W and L.

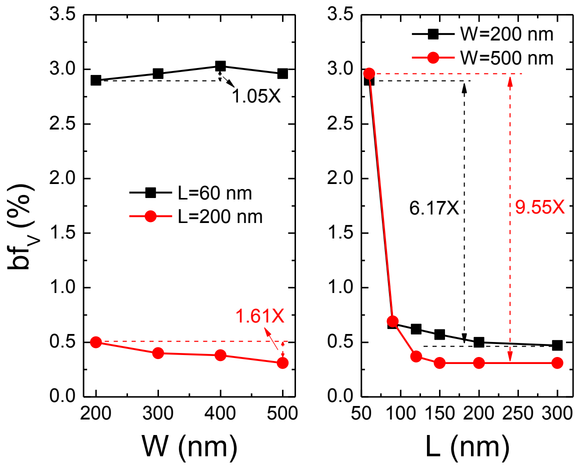


Figure 6.30. Supply voltage bit-flipping (bf_V) as a function of transistor channel (right) width (W) and (left) length (L) at T=25°C.

The evaluation of bf_V across the considered V_{DD} range (±20% V_{DD} variation with respect to the nominal condition of 1V) shows a slight dependence on the W values for a given L, while a strong dependence of bf_V can be observed as a function of L owing to the significant DIBL effect at lower L values. As matter of fact, starting from the minimum allowed length of 60nm, for a given W, an increase of L up to 300nm can lead to a bf_V reduction of 9.55×.

Figure 6.31 reports the simulation results in terms of bf_T, calculated at V_{DD}=1V, as a function of transistor W and L.

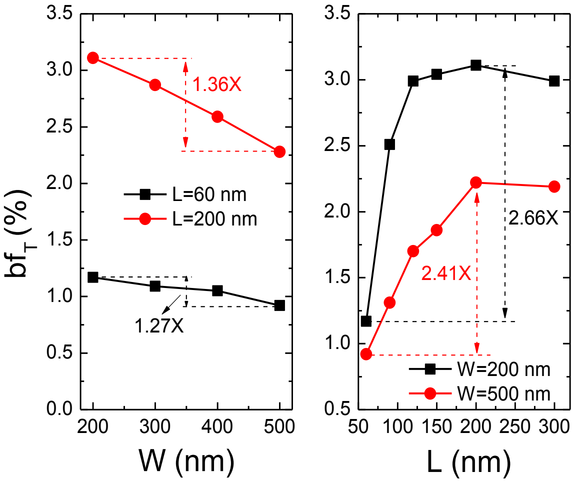


Figure 6.31. Temperature bit-flipping (bf_T) as a function of transistor channel (right) width (W) and (left) length (L) at V_{DD}=1V.

From the comparison between data of Figure 6.29 and Figure 6.31, bf_T follows the same trend of ΔV_{OUT} variability as a function of transistor size. This is very likely due to the effect of the second term in (6.12), which is particularly sensitive to temperature variations. In particular, the effect of this term on the temperature stability of ΔV_{OUT} is amplified by random mismatch variations at smaller transistor size and by low values of λ_D at larger L values.

Above results provide the following design guidelines for transistor sizing of the 2T voltage divider pair:

- low ub_N values require a high ΔV_{OUT} variability and, hence, a low W and a L greater than the allowed minimum value;
- low bf_V values require high L values;
- low bf_T values require a low ΔV_{OUT} variability and, hence, a high W and a low L.

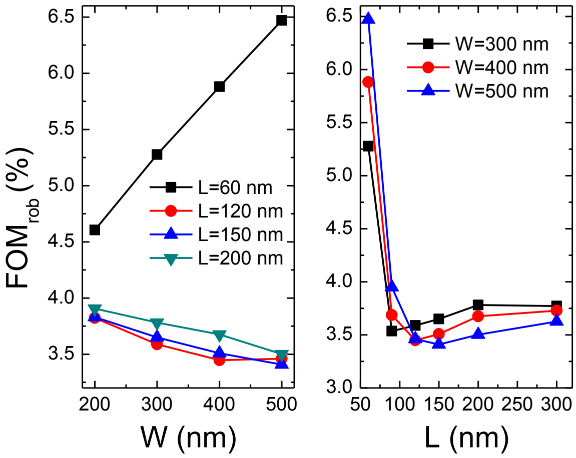


Figure 6.32. Bit-robustness FOM (FOM_{rob}) as a function of transistor channel width (left) and length (right).

Figure 6.32 reports the FOM_{rob} as a function of transistor size. In particular, Figure 6.32 shows that a proper transistor sizing of the voltage divider pair can be achieved by choosing high W and an L greater than the allowed minimum value, specifically in the range from 90 to 150nm. By taking into account the results of the performed analysis and also considering the silicon area occupation, a good choice for the transistor sizing of the proposed circuit can be obtained with a W equal to 400nm and an L equal to 120nm.

According to this choice, the summary results are shown in Table 6.VIII, which also reports simulation results corresponding to a less conservative value of 1mV for $M_{\Delta V_{OUT}}$.

A further analysis on robustness performance of the proposed circuit is reported in Figure 6.33 in terms of *intra*-PUF HD, which is ideally equal to 0 [15], considering a 256-bit word as key and a $M_{\Delta V_{OUT}}$ of 10mV.

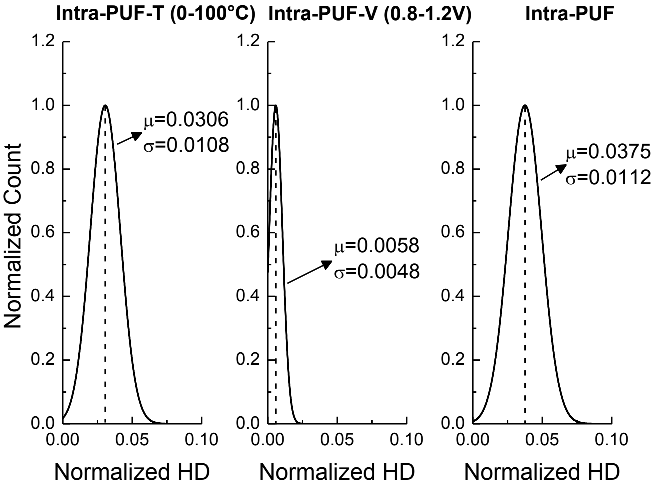


Figure 6.33. Normalized *intra*-PUF Hamming Distance (HD) under (left) temperature variations, (centre) V_{DD} variations and (right) both temperature and V_{DD} variations, considering a 256-bit word for the PUF key.

TABLE 6.VIII. SUMMARY RESULTS

Process	*L* (nm)	*W* (nm)	V_{TH0} (mV)	$\sigma_{\Delta VOUT}$ (mV)	ub_N (%)	bf_V (%)	bf_T (%)	FOM_{rob} (%)	*Norm. Mean Intra-PUF HD*
Decision margin ($M_{\Delta V_{OUT}}$) = 10mV									
65nm LVT	120	400	~465	299.6	2.75	0.35	2.05	3.45	0.0375
Decision margin ($M_{\Delta V_{OUT}}$) = 1mV									
65nm LVT	120	400	~465	299.6	0.32	0.14	1.84	1.87	0.0225

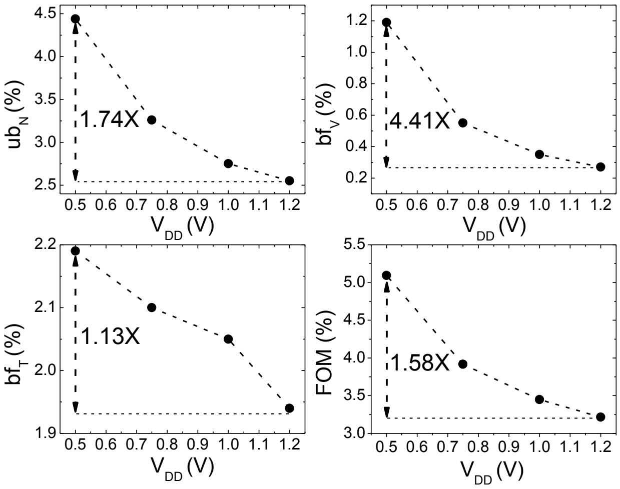


Figure 6.34. Robustness FOMs vs V_{DD} for the proposed PUF.

Finally, Figure 6.34 evaluates the robustness for different V_{DD}s, by considering a $M_{\Delta V_{OUT}}$ of 10mV. As expected, the V_{DD} scaling affects the stability of the PUF response, especially in terms of ub_N and bf_V. Overall, by reducing the V_{DD} from 1.2V down to 0.5V, the proposed PUF shows a FOM_{rob} degradation of 1.58×.

6.4.4.2. Impact of the output stage

After the first analysis, aimed at investigating the design of the voltage divider pair, an output voltage comparator, consisting of a simple sense amplifier and a buffer, is introduced. In this analysis, the general architecture shown in Figure 6.35 is considered. It is worth emphasizing that in the solution of Figure 6.35, the elementary bitcell (BC) is composed only by a voltage divider pair with the aim of saving area, while the output comparators are shared by a set of bitcells, similarly to the scheme proposed in [31].

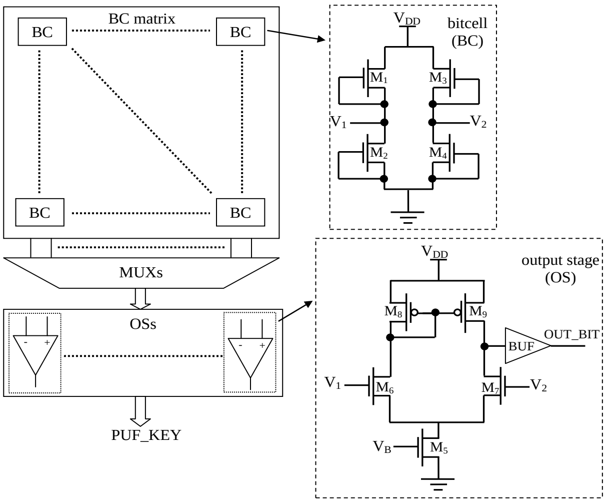


Figure 6.35. PUF architecture. The figure shows the schematics of the elementary bitcell (BC) and the shared output stage (OS).

TABLE 6.IX. COMPARATIVE RESULTS

	[15] INV_PUF (measured)	[15] SA_PUF (measured)	[31] (measured)	**This work** (simulated)
Technology	65nm	65nm	65nm	65nm
Nominal conditions	V_{DD} = 1V T = 25°C	V_{DD} = 1V T = 25°C	V_{DD} = 1V T = 20°C	V_{DD} = 1V T = 25°C
$\boldsymbol{ub_N}$ **[%]**	1.73	1.56	2.0	1.07
$\boldsymbol{T}$ **range [°C]**	25-85	25-85	0-80	0-100
$\boldsymbol{bf_T}$ **[%]**	2.83	2.95	3.5	1.7
$\boldsymbol{bf_T}$ **per 10°C [%]**	0.47	0.49	0.44	0.17
$\boldsymbol{V_{DD}}$ **range [V]**	0.7-1	0.7-1	0.6-1.2	0.5-1.2
$\boldsymbol{bf_V^*}$ **[%]**	2.08	2.27	0.78	0.86
$\boldsymbol{bf_V^*}$ **per 0.1V [%]**	0.69	0.76	0.13	0.12
$\boldsymbol{FOM_{rob}}$ **[%]**	3.92	4.04	4.12	2.19
$\boldsymbol{\sigma_{\Delta VOUT}}$ **[mV]**	N/A	N/A	31.0	299.6

* bf_V has been calculated by considering a V_{DD} variation in the whole considered range

Aside from allowing the implementation of the sensing differential-mode configuration, this output stage (OS) typically leads to increase the voltage gain of the circuit, thus amplifying the difference between the two outputs of a voltage divider pair and, consequently, reducing the probability of generating unstable bits, and introduces additional random mismatch related to the offset of the sense amplifier [15]. Here, the evaluation of unstable (or "noisy") bits at nominal condition has been carried out by applying a sinusoidal noise to the supply voltage. In particular, by considering a 100MHz sinusoidal noise with an amplitude of 100mV (200mV peak-to-peak), the ub_N is equal to 1.07%, which is an intermediate value between the two results previously obtained by considering a decision margin of 1mV and 10mV (see Table 6.VIII). In addition, simulation results with the OS in terms of bf_V and bf_T are 0.23% and 1.7%, respectively, thus leading to an overall FOM_{rob} of 2.02%, which is again an intermediate value between the two results obtained for a $M_{\Delta V_{OUT}}$ of 1mV and 10mV (see Table 6.VIII). As shown in Table 6.IX, when compared to more recent robust PUF topologies based on static and mono-stable analog circuits [15],[31], the proposed PUF shows superior performance in terms of robustness against noise, bias and temperature variations. These advantages are mainly ascribed to the extremely high variability offered by the proposed solution. In particular the ΔV_{OUT} dispersion is around 300mV, which is about one order of magnitude higher compared to the variability of the solution reported in [31]. It is worth noting that the above claimed advantages have been obtained with a power consumption of only 35.6 pW at 1V and an active silicon area less of 1 μm^2 per bitcell.

6.4.5. Conclusion

In this section, an extremely compact circuit solution for static and mono-stable silicon-based PUFs, based on pairs of 2T voltage dividers consisting of a series of two identical nMOSFETs, is presented. The proposed PUF mainly exploits the statistical mismatch variations of threshold voltages of LVT transistors in a commercial 65nm CMOS technology to produce per each pair a random differential output voltage with an extremely high variability. The polarity of this differential voltage determines the generation of a random bit in the PUF response. An extensive analysis by means of MC simulations has been performed to investigate the effect of transistor sizing on the variability and the stability of the PUF response. Simulation results prove how a proper transistor sizing allows obtaining high robustness against noise, supply voltage and temperature variations.

From comparison with some state-of-the-art PUF designs, the proposed solution proves to be a promising and competitive candidate for analog and static PUFs.

Bibliography

[1] R. Pappu, “Physical one-ways functions,” Ph.D. dissertation, Massachusetts Institute of Technology, *Cambridge,* MA, 2001.

[2] R. Pappu,

[3] *International Conference on Computer-Aided Design*, pp. 670-673, 2008.

[4] G. Qu and C.-E. Yin, “Temperature-aware cooperative ring oscillator PUF,” in *IEEE International Workshop on Hardware-Oriented Security and Trust* (HOST), 2009, pp.36-42.

[5] J. Murphy, A. Dziech, A.Czyżewski, “Asynchronous Physical Unclonable Functions ASYNC PUF,” in *Multimedia Communications, Services and Security*; Krakow 2012.

[6] V. Vivekraja and L. Nazhandali, “Feedback based supply voltage control for temperature variation tolerant PUFs,” in *International Conference on VLSI Design,* Jan. 2011, pp. 214 –219.

[7] S. Mansouri and E. Dubrova, “Ring Oscillator Physical Unclonable Function with Multi Level Supply Voltages”, in *IEEE international Conference on Circuit Design* (ICCD), 2012, pp. 520-521.

[8] J. GuajardoB. Recht, J. Taylor, and N. Gershenfeld, “Physical one-way functions,” *Science*, vol. 297, pp. 2026–2030, 2002.

[9] B. Gassend, D. Clarke, M. van Dijk, S. Devadas “Silicon physical random functions”, in *ACM Conference on Computer and Communications Security,* pp. 148–160, 2002.

[10] E.E. Suh and S. Devadas. “Physical unclonable functions for device authentication and secret key generation,” in *IEEE Design Automation Converence* (DAC), pp. 9–14, 2007.

[11] M. Majzoobi, F. Koushanfar, M. Potkonjak, “Lightweight secure PUFs”, in *Proceedings of the IEEE/ACM,* S.S. Kuma, G.-J. Schrijen, P. Tuyls, “FPGA intrinsic PUFs and their use for IP protection,” in *Workshop on Cryptographic Hardware and Embedded Systems*, 2007, pp. 63–80.

[12] A.R. Krishna, S Narasimhan, X Wang, S Bhunia, “MECCA: a robust low-overhead PUF using embedded memory array,” in *Cryptographic Hardware and Embedded Systems* (CHES), 2011, pp. 407-420.

[13] S. Eiroa et Al., “Reducing bit flipping problems in SRAM physical unclonable functions for chip identification” in *IEEE International Conference on Circuits and Systems* (ICECS), 2012, pp. 392-395.

[14] M. Cortez, A. Dargar, S. Hamdiuri, G.-J. Schrijen, “Modeling SRAM Start-Up Behaviour for Physical Unclonable Functions,” in *IEEE International Symphosium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems* (DFT), Oct. 2012, pp. 1-6.

[15] S. Stanzione, D. Puntin, and G. Iannaccone, “CMOS Silicon Physical Unclonable Functions Based on Intrinsic Process Variability,” in *IEEE Journal of Solid-State Circuits* (JSSC), vol.46, no 6, Jun. 2006, pp. 1456-1463.

[16] A. Alvarez, W. Zhao, M. Alioto, “15-fJ/bit Static Physically Unclonable Functions for Secure Chip Identification with <2% Native Bit Instability and 140X Intra/Inter PUF Hamming Distance Separation in 65nm,” in *IEEE International Solid-State Circuits Conference* (ISSCC), 2015.

[17] A. Rukhin et. Al, “A Statistical Test Suite for Random and Pseudorandom Generators for Cryptographic Applications,” NIST,800-22 (Rev.1a), 2010.

[18] U. Ruhrmair, C. Hilgers, S. Urban, A. Weiersh, E. Dinter, B. Forster, C. Jirauschek, “Optical PUFs Reloaded,” *IACR Cryptology ePrint Archive,* 2013.

[19] S. K. Mathew, S. K. Satpathy, M. A. Anders, H. Kaul, S. K. Hsu, A. Agarwal, G. K. Chen, R. J. Parker, R. K. Krishnamurthy, and V. De, “A 0.19pJ/b PVT-Variation-Tolerant Hybrid Physically Unclonable Function Circuit for 100% Stable Secure Key Generation in 22nm CMOS,” in *IEEE International Solid-State Circuits Conference* (ISSCC), 2014, pp. 278–280.

[20] Kaiyuan Yang, Qing Dong, David Blaauw, Dennis Sylvester, "A Physically Unclonable Function with BER <10-8 for Robust Chip Authentication Using Oscillator Collapse in 40nm CMOS," in *IEEE International Solid-State Circuits Conference* (ISSCC), 2015.

[21] J. Delvaux, I. Verbauwhede, "Fault Injection Modeling Attacks on 65 nm Arbiter and RO Sum PUFs via Environmental Changes," *IEEE Transaction on Circuits and Systems I, Regular Paper*, vol. 61, no 6, pp. 1701-1713, Jun. 2014.

[22] K. Lofstrom, et *Al*., "IC Identification Circuit Using Device Mismatch", in *IEEE International Solid-State Circuits Conference* (ISSCC), pp. 372-373, 2000.

[23] N. Liu, *et al*., "OxID: On-chip One-Time Random ID Generation Using Oxide Breakdown," *IEEE Symp. VLSI Circuits*, pp. 231-232, 2010.

[24] Ulrich Ruhrmair, Jan Solter, "PUF Modeling Attacks: An Introduction and Overview", in *Design, Automation and Test in Europe Conference and Exhibition* (DATE), 2014, pp. 1-6.

[25] M. Alioto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: A Tutorial," *IEEE Trans. on Circuits and Systems* – part I (invited), vol. 59, no. 1, pp. 3-29, Jan. 2012.

[26] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors", *IEEE Journal of Solid-State Circuits* (JSSC), vol. 24, pp. 1433 - 1439, 1989.

[27] B. Razavi, **Design of Analog CMOS Integrated Circui**t. *McGraw-Hill Higher Education*, 2003.

[28] M.J.M. Pelgrom, A.C.J. Duinmaiger, and A. P. G. Welbers ,"Matching properties of MOS transistors for precision analog design," *IEEE Journal of Solid-State Circuits* (JSSC), vol. 24, no 5, pp. 1433-1439, Oct. 1989.

[29] A.Rukhin, J.Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, S.Vo, **A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications**, Special Publication 800-22 Revision 1a, 2001. http://csrc.nist.gov/publications/nistpubs/800-22-rev1a/SP800-22rev1a.pdf.

[30] http://csrc.nist.gov/groups/ST/toolkit/rng/stats_tests.html.

[31] P.Magnone, F. Crupi, N. Wils, R. Jain, H. Tuinhout, P. Andricciola, G. Giusi, C. Fiegna, "Impact of Hot Carriers on nMOSFET Variability in 45 nm and 65 nm CMOS Technologies", *IEEE Transactions on Electron Devices*, vol. 58, n. 8, pp. 2347-2353, 2011.

[32] J. Li and M. Seok, "A 3.07um^2/bitcell Physically Unclonable Function with 3.5% and 1% Bit-Instability across 0 to 80 ^{o}C and 0.6 to 1.2V in a 65nm CMOS," in *Proc. Symposium on VLSI Circuits*, 2015, pp. C250-C251.

[33] J. H. Huang, Z. H. Liu, M. C. Jeng, P. K. Po, and C. Hu, "A Physical Model for MOSFET Output Resistance," in *Proceedings of the IEEE International Electron Devices Meeting* (IEDM), 1992, pp. 569-572.

CONLUSIONS AND FUTURE DEVELOPEMENTS

In this work different circuits capable of operating at ultra-low supply voltages and with low power consumption have been presented.

Two ultra-low voltage, low power voltage references have been implemented in 0.18 μm CMOS technology. The first voltage reference operating at the minimum supply voltage of only 150 mV has been presented. Measurement results showed that the two proposed solutions are in able to guarantee good performance in terms of stability against supply voltage, temperature and process variations.

An ultra-low voltage, low power current reference, implemented in 0.18 μm CMOS technology, has been presented. The proposed solution represents the first current reference capable of operating with a supply voltage of only 0.5 V. Measurement results showed excellent performance in terms of stability against supply voltage and process variations. Regarding the temperature behaviour, considering the lack of accuracy of the models used in the design phase and considering the sensitivity of the proposed solution with respect to the matching with the zero TC bias point, a trimmed solution is necessary to achieve better temperature stability.

A low power, low energy-per-conversion SAR ADC, implemented in 0.35 μm CMOS technology, has been presented. The occupied area and the measured DC and AC performances resulted very satisfactory for the project requirements. The performance in terms of speed, supply voltage operation, area occupancy and resolution can be easily boosted with the implementation of the proposed scheme in a scaled technology.

Finally two innovative PUF solutions for data protection capable of operating at very low supply voltages and with low power consumption have been presented. An extensive set of simulation results obtained in different process conditions and on different technologies showed excellent performance in terms of uniqueness, randomness and reliability. In the case of the complementary current mirrors solution, the goodness of the proposed scheme has been confirmed by comparing simulation results with measurement results where available. For the experimental validation of the 2T voltage divider solution the silicon implementation is suggested.

ACKNOWLEDGEMENTS (in Italian)

A conclusione di questa esperienza di dottorato non posso che ritenermi fortunato in merito alle persone con cui ho avuto il piacere e il privilegio di lavorare. Mi limiterò a citare esplicitamente chiunque abbia dato un contributo tecnico a questo lavoro di tesi o abbia contribuito alla mia crescita professionale durante questi tre anni.

La prima persona ad aver dato un contributo essenziale all'elaborato di tesi è il mio tutor, il Prof. Felice Crupi o per megio dire Felice. Senza di lui non avrei sicuramente scelto questo percorso. Conoscendomi ritengo sia uno dei pochi (se non l'unico) con cui avrei potuto svolgere il percorso del dottorato di ricerca. Spero di aver meritato la fiducia che ha riposto in me.

La seconda persona che vorrei ringraziare è il Prof. Piero Malcovati, il Piero, che mi ha dato la possibilità di lavorare tra le presone più competenti al mondo in merito alla progettazione circuitale. Un grazie infinito va anche a Marco Grassi che mi ha seguito dal punto di vista tecnico durante la mia intera esperienza a Pavia. Ringrazio inoltre tutti i ragazzi dei laboratori SMS e IMS dell'Università di Pavia. Si sono dimostrati sempre disponibili e cordiali, aiutandomi sia durante la fase progettuale sia durante la fase di testing.

Ringrazio il Prof. Marco Lanuzza con il quale ho avuto il piacere di realizzare diversi lavori scientifici.

Tra le persone che sento di dover ringraziare c'è anche il Prof. Giuseppe Iannaccone, il quale ha consentito più volte di realizzare i circuiti da me progettati e il Prof. Massimo Alioto con il quale ho avuto il piacere di collaborare su più attività di ricerca.

Anche se l'attività svolta insieme a loro non è riportata in questo lavoro di tesi, mi sento in obbligo di ringraziare Stefano Stazione, il mio tutor insieme a Samira Zaliasl durante la mia esperienza all'High Tech Campus di Eindhoven. Entrambi hanno indubbiamente contribuito alla mia crescita professionale.

Durante questi anni di dottorato nelle mie esperienze dentro e fuori l'Unical ho avuto il piacere di conoscere molte persone con cui ho trascorso momenti che talvolta ricordo con grande nostalgia. Non è retorica, farne un'elenco è quasi impossibile. A loro va il mio più sentito grazie, mi hanno sempre fatto sentire a casa, credo sinceramente che senza di loro sarebbe stato molto più difficile completare questo percorso.