# First Principle Leakage Current Reduction Technique for CMOS Devices

Hippolyte Djonon Tsague
Council for Scientific and Industrial Research (CSIR)
Modelling and Digital Science (MDS)
Pretoria, South Africa
hdjonontsague@csir.co.za

Bhekisipho Twala
Faculty of Engineering University of Johannesburg (UJ)
Department of Electrical and Electronic Engineering Science
Johannesburg, South Africa

***Abstract*—This paper presents a comprehensive study of leakage reduction techniques applicable to CMOS based devices. In the process, mathematical equations that model the power-performance trade-offs in CMOS logic circuits are presented. From those equations, suitable techniques for leakage reduction as pertaining to CMOS devices are deduced. Throughout this research it became evident that designing CMOS devices with high-κ dielectrics is a viable method for reducing leakages in cryptographic devices. To support our claim, a 22nm NMOS device was built and simulated in Athena software from Silvaco. The electrical characteristics of the fabricated device were extracted using the Atlas component of the simulator. From this research, it became evident that high-κ dielectric metal gate are capable of providing a reliable resistance to DPA and other form of attacks on cryptographic platforms such as smart card.The fabricated device showed a marked improvement on the $I_{on}/I_{off}$ ratio, where the higher ratio means that the device is suitable for low power applications. Physical models used for simulation included $Si_3N_4$ and $HfO_2$ as gate dielectric with TiSix as metal gate. From the simulation result, it was shown that $HfO_2$ was the best dielectric material when TiSix is used as the metal gate.**

***Keywords—Differential power analysis; High-K dielectric gate; Smart card***

## I. INTRODUCTION

Smart Cards are a safe place to store valuable information such as cryptographic private keys making them an adequate medium capable of providing secure authentication and storage of secret data. Recent development in this field of study have redirected attention to a whole new class of attacks developed to retrieve the secret key stored in the card through the study and interpretation of the information leaked by processor power dissipation. Such attacks are known as Side Channel Attacks (SCA's) and are performed by taking advantage of the interdependence in Complementary Metal Oxide circuits between dynamic power dissipation on the processed data. Chief among those techniques is Differential Power Analysis (DPA) which has been widely shown to be a threat to the security of cryptographic devices [1]. For that reason a large number of side channel related countermeasures at various levels of abstraction have been proposed. In particular, at the transistor level, countless alternatives to first order DPA attacks have been developed and demonstrated. Researchers in [2] have successfully demonstrated that in sub-100 nm and related technologies, power leakages are as high as the dynamic power of the device and hence the leakage (static) supply current can be used as a new Side-Channel.

Off State leakages have been recognized by semiconductor manufacturers as a major bottleneck for future microcontroller integration. Off-state leakage is a static power current that leaks through transistors even when such devices are turned off. It is one of two principal sources of power dissipation in today's microcontrollers. The other source of leakage is of course dynamic power, which is caused by the repeated capacitance charge and discharge on the output of the hundreds of millions of gates in today's microcontrollers. Until recently, only dynamic power was identified as a significant source of power consumption, and Moore's law has helped to control it through shrinking processor technology. Dynamic power is proportional to the square of the supply voltage; therefore reducing the voltage significantly reduces the device's power consumption. Unfortunately smaller geometries aggravate current leakages problems; static power begins to dominate the power consumption equation in microcontroller design" [12].

CMOS technology was invented at Fairchild Semiconductor in 1963 by Wanlanss Frank. The original idea behind the technology was to design a low power alternative to Transistor-transistor Logic (TTL). The early adopters of the technology were watch designers who realized the importance of low battery power consumption over the device processing capabilities for electronic circuits. Nowadays, the CMOS technology has grown substantially to become the dominant technology integrated circuit design. "This is essentially because area occupation, operating speed, energy efficiency and manufacturing costs have benefited and continue to benefit from the geometric downsizing that comes with every new generation of semiconductor manufacturing processes. In addition, the simplicity and comparatively low power dissipation of CMOS circuits have allowed for integration densities not possible in similar techniques such as bipolar junction transistors (BJT)" [17]. Despite its many benefits, authors of [12] argue that the power that such devices consume has increased dramatically with increases in device speed and

**978-1-4673-9354-6/15/$31.00 ©2015 IEEE**

chip density; and so has the number of attacks on cryptographic devices.

## II. POWER FUNDAMENTALS

Five mathematical equations govern the power performance in the CMOS logic circuits according to [12]. In this paper, we present them in a way that addresses the basics of physics and logic circuitry design. The first mathematical equations related to CMOS power fundamentals are the basics of low power consumption [3] while the last two equations are more concerned with sub-threshold and gate-oxide leakage modeling in CMOS technologies.

### *A. Investigation of Frequency and Voltage Relationships*

Equation (1) below depicts the supply voltage dependency of the operating frequency of the device as computed in [12]:

$$f \propto (V - V_{th})^{\alpha/V} \quad (1)$$

In this equation, $V$ represents the transistor's supply voltage while $V_{th}$ is the device's voltage. The exponent α is an experimentally derived constant with a value of 1.3 approximately [12]. Dynamic voltage scaling in CMOS devices is used to control switching power dissipation in battery operated systems. Also, power consumption minimization techniques rely on low voltage modes and lowered clock frequencies. In [12] authors have used the relation derived in (1) to compute an equation that depicts the relationship between frequency and supply voltage. The derivation begins with the selection of the device's working voltage and frequency defined as $V_{norm}$ and $f_{norm}$ respectively. The quantities selected are normalized entities depicting the relationship between the largest possible device's operating voltage $V_{max}$ and frequency $f_{max}$. This relationship is shown in (2) below:

$$V_{norm} = \beta_1 + \beta_2 \cdot f_{norm} = \frac{V_{th}}{V_{max}} + \left(1 - \frac{V_{th}}{V_{max}}\right) \cdot f_{norm} \quad (2)$$

From (1) it is evident that if $f = 0$, then (2) becomes:

$$V_{norm} = \beta_1 = \frac{V_{th}}{V_{max}}$$

The value of $V_{norm}$ can safely be approximated to 0.37. That approximation closely matches present day's industrial data [12]. It is also worth mentioning at this stage that $f_{max}$ is proportional to $V_{max}$ and that the frequency will drop to zero if V is equal to $V_{th}$, as clearly shown (1).

### *B. CMOS Power Dissipation*

Many origins of power consumption in CMOS devices exist in CMOS devices and that what the third equation attempts to clarify. The CMOS power consumption is the sum of dynamic and static power as shown in (3):

$$P = P_{dynamic} + P_{static} \quad (3)$$

The first term of (3) can be broken off into two distinct entities namely $P_{short}$ and $P_{switch}$. The first component $P_{short}$ is the power dissipated during gate voltage transient time while the second component $P_{switch}$ comes as a result of the many charging and discharging of capacitances in the device. The last term $P_{static}$ represents the power generated when the transistor is not in the process of switching. Equation (3) can be rewritten as:

$$P = (P_{short} + P_{switch}) + P_{static} = ACV^2 f + VI_{leak} \quad (4)$$

In (4) $A$ denotes the number of bits that are actively switching and $C$ is the combination of the device's load and internal capacitance. It is worth mentioning at this stage that for ease of simplicity, the power lost to spasmodic short circuit at the gate's output has been neglected.

From (4) it is evident that dropping the supply voltage leads to an important decrease in the device power consumption. Mathematically speaking, dividing the supply voltage by 2 or halving it will reduce the power consumption by four. The main drawback of that proposition is that it will reduce the processor's top operating frequency by more than half. A better approach suggested in [12] relies on the use parallel or pipelined techniques to compensate for the performance losses due to supply voltage reduction.

### *C. Computing Leakage Current*

Parallelism and pipelining techniques for power reduction were first proposed by [19]. Since then researchers have conducted studies aimed at optimizing the pipelining depth for dissipated power reduction in CMOS devices. Furthermore, researches have been conducted at a functional block level to compare the performances of pipelining and parallelism to find out which technique performs best when it comes to minimizing total switching power. In (3) it was shown that leakage current (source of static power consumption) is a combination of subthreshold and gate-oxide leakage i.e.:

$$I_{leak} = I_{sub} + I_{ox}$$

*Deriving Subthreshold Power Leakage:* Authors of [19] present an equation representing the direct relationship between a CMOS device threshold voltage, its subthreshold leakage current and the device supply voltage as follows:

$$I_{sub} = K_1 W e^{-V_{th}/nV_0\left(1- e^{-V/V_\theta}\right)} \quad (6)$$

In (6) $K_1$ and $n$ are normally derived experimentally, W represents the device's gate width, and $V_\theta$ is its thermal voltage. The quantity $V_\theta$ can safely be approximated to 25 mV at room temperature (20 degrees Celsius). If $I_{sub}$ rises enough to generate, $V_\theta$ will rise as well and in that process cause an increase in $I_{sub}$ and this may result in thermal runaway. From (6) it becomes clear that two ways exist for reducing $I_{sub}$ which are (1) turning off the supply voltage and (2) stepping-up the threshold voltage. In [12] it is argued that "since this quantity shows up as a negative exponent, increasing that value could have a dramatic effect in even small increments. On the other hand, it is evident from (1) that increasing $V_{th}$ automatically creates a reduction in speed. The obvious problem with the first approach is loss of state; as for the second option, its major inconvenience relates to the loss of performance". The device's gate width W, its gate length $L_g$, the device's oxide thickness $T_{ox}$, and doping concentration $N_{pocket}$ are other major contributors to subthreshold leakage in CMOS based technologies. Processor designers often optimize one or a few of those leakage components as a convenient technique to reduce subthreshold leakage as will be seen in subsequent paragraphs.

*Subthreshold Power Leakage:* Gate leakage mechanisms, such as tunneling across thin gate oxide leading to gate oxide leakage current become significant at the 90nm node and smaller. Gate oxide leakage is not as well understood as subthreshold leakage. For the purpose of this research, a simplification of equations from the authors of [5] is sufficient to illustrate the point:

$$I_{ox} = K_2 W \left(\frac{V}{T_{ox}}\right)^2 e^{-\alpha T_{ox}/V} \quad (7)$$

Where $K_2$ and $\alpha$ are derived experimentally. In (5) we draw our attention to the oxide thickness, $T_{ox}$ component of the equation. "Increasing $T_{ox}$ will reduce the gate leakage. However, it also negatively affect the transistor's efficiency since $T_{ox}$ must decrease proportionately with process scaling to avoid short channel effects. Therefore, increasing $T_{ox}$ is not a viable option" [12]. A better approach to this problem may lie in the development of high-κ dielectric gate insulators. This approach is currently under heavy investigation by the research community.

## III. Reducing Static Power Consumption

Many researchers as well research groups have developed power models for reducing static power consumption for embedded devices. Power gating [4, 16] is slowly becoming a very popular design technique for decreasing leakage currents. Although effective in reducing static power consumption in many instances, its major drawback lies in its tendency to introduce delays by adding extra circuitry and wires and also uses extra area and power.

Another approach to static power reduction is based on the utilization of multiple threshold voltage techniques. "Present day's processes typically offer two threshold voltages. Microprocessor designers assign a low threshold voltage to some of the few identified performance-critical transistors and a high threshold voltage to the majority of less time critical transistors. This approach has the tendency to incur a high subthreshold leakage current for the performance-critical transistors, but can significantly reduce the overall leakage" [12].

Other techniques for reducing subthreshold leakage are closely related to gate tunneling current, however, their effects are still under investigations. Gate-oxide leakage has a negligible dependence on temperature [12]. Therefore, as it subsides with drops in temperature, gate-oxide related current leakage become important.

## IV. Fabrication Method

For this paper, a 20nm NMOS MOFSET was fabricated using the SILVACO Athena module and the device's electrical characteristic and performance were simulated and evaluated using the Atlas module from SILVACO. The specifications of the sample used in this experiment was p-type (boron doped) silicon substrate with doping concentration of $1.5e^{15}$atoms $cm^{-3}$ and <100> orientation. The next step consisted in developing the P-well by growing a 800 Å oxide screen on top of bulk silicon. This technique makes use of dry oxygen at a very high temperature (i.e. approximately 800°C) followed by Boron as dopant with a concentration of $3.75e^{13}$atoms $cm^{-3}$. In the third step, the deposited oxide layer is etched and there after annealed to ensure that all boron atoms are spread uniformly. This done at a temperature of 900°C using nitrogen followed by a futher rise in temperature to 950°C using dry oxygen. The next step was to isolate the neighbouring transistor by creating a shallow trench isolator with a thicknesses of 130 Å. After that step, the wafer was oxidized with dry oxygen for approximately 25 minutes at temperature of 1000°C. Two important processes were involved in the development of the STI namely, Low Pressure Chemical Deposition (LPCVD) and reactive ion etching (RIE). The LPCVD process starts with the deposition of a 1000 Å nitride layer on top of the STI oxide layer, followed by a photo-resistor deposition on the wafer. The RIE process consisted in etching the unnecessary part on the top of the STI area. Both chemical and mechanical polishing was implemented to strip away any extra oxide on the wafer. STI was further annealed for approximately 15 minutes at a temperature of 850°C. As a final STI step in the process, an oxide layer was carefully deposited and etched to eliminate possible defects that may have occurred on the surface.

It is important to mention at this stage that the deposition of

high-κ dielectric process with gate oxide thickness is selected so that they have the same equivalent oxide thickness as $SiO_2$ . Furthermore, the length of the high-κ material was scaled so as to get the equivalent 22nm gate length of the transistor. The next step consisted of the deposition of Titanium Silicide (TiSix) on top of the high-κ dielectrics ($Si_3N_4, HfO_2$) followed by halo implantation of indium dose to obtain the optimum value of the NMOS device [6, 7]. The next step involved the formation of the sidewall spacer that would serve as source and drain electrodes for the device. In this case, implantation with arsenic is followed by a dose of phosphor to ensure an uninterrupted flow of current in the fabricated NMOS device [8].

The next step consisted in the deposition of a layer (0.5 µm) of Boron Phosphor Silicate Glass (BPSG) to act as the pre-metal dielectric [9]. Once more, annealing was done at 950°C on the wafer to strengthen the structure. The next step consisted of compensation implantation with a layer of phosphorous. The last step involved the deposition of aluminium layer which served as metal electrode (source and drain). Once the model structure design was completed we then proceeded with device simulation using Atlas.

## V. RESULTS AND DISCUSSIONS

The complete NMOS structure is shown in Fig. 1 below. The fabrication process is the same for all high-κ devices fabricated for this research except that the dielectric materials were varied.

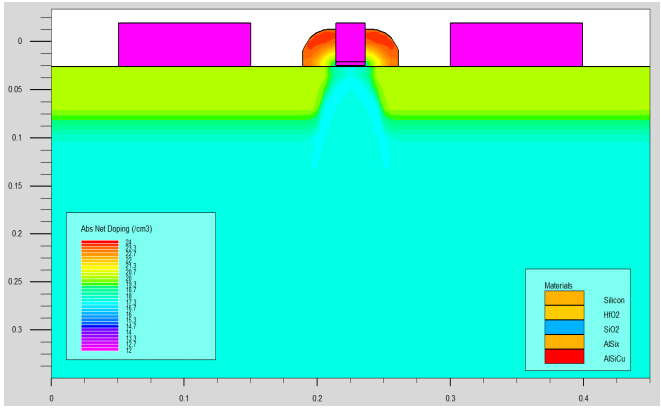


Fig. 1. Cross Section of the fabricated 22 nm NMOS device

Fig.2 on the other hand shows the doping profile of one of the designed structure with gate length of 22nm NMOS.

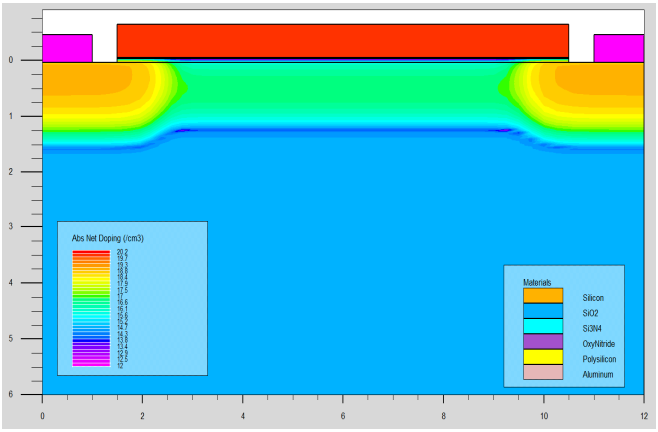


Fig. 2. Doping profile of the fabircated device

Results of electrical characteristic simulation are obtained in Fig. 3 below.The plots are also known as "$V_t$ Curves", because devices designers use them extensively to extract the threshold voltage ($V_t$), which defines an approximation of when a transistor is "on" and allows current to flow across the channel.

For this research, the Figure presents the drain ($I_d$) vs gate voltage ($V_{GS}$) curves for $Si_3N_4$ ($k$~$29$), $HfO_2$ ($k$~$21$) and aconventional device made of $SiO_2$ . Typically, the fabricated device's drain voltage $V_{DS}$was fixed when ($I_d$) vs ($V_{GS}$) was plotted. The threshold voltage ($V_{th}$), state on current ($I_{on}$) and state off current ($I_{off}$) can be extracted from the ($I_d$) vs ($V_{GS}$) curve.

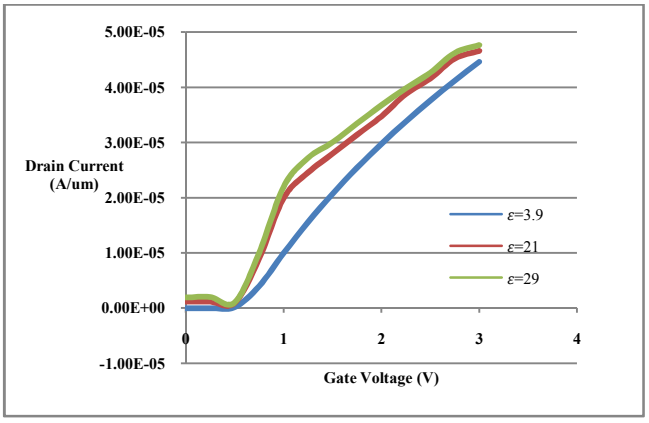


Fig. 3. Drain cureent Vs gate voltage curves (threshold curves)

A good doping concentration is one of the critical factorsthat ensure that the transistor works well and emits fewer leakage currentsso as to enhance gate control [10]. There are four factors that influence the threshold voltage; these factors include (1) threshold voltage adjustment implant, (2) halo implant, (3) channel implant and (4) compensation implant. However for the purpose of this research, where the main requirement was to investigate the leakage current

emissions of the device while varying the gate material used, the threshold voltage adjustment implant technique was utilised. To get a threshold voltage of 0.302651 as stipulated by the ITRS the best doping concentration with boron was set to $8.5763x10^{13}cm^{-2}$ for $HfO_2$ and $9.73654x10^{13}cm^{-2}$ for $Si_3N_4$. This doping concentration variation is to account for the physically thicker and stronger dielectric materials utilised. As expected, the drain current for both $Si_3N_4$ and $HfO_2$ dielectric are decreased compared to the drain current of the device made of $SiO_2$ as can be seen from Fig. 3 above.

Drain leakage current ($I_{off}$) or sub-threshold leakage current occurs when the gate voltage ($V_{GS}$) is lower than the threshold voltage ($V_{th}$). In ideal case, when the transistor is turned off, $V_{GS}$= 0 volt and $V_{DS}$= $V_{DD}$(voltage supply), there is no current flow through the channel ($I_{off}$= 0). Again from Fig. 3 it is clear that the leakage current through $HfO_2$dielectric is lowest compared to both $Si_3N_4$and $SiO_2$ dielectrics. This result suggests that $HfO_2$dielectric material is more compatible with silicon and appears to be the most stable oxide with the highest heat of formation. This observation matches the findings of [12, 13].

Table 1 show the simulated results for $Si_3N_4$ and for $HfO_2$dielectric materials with TiSix as metal gate for 22nm NMOS. Clearly, the $I_{on}$values from the simulation are larger compared to prediction value.

TABLE I. SIMULATED RESULTS OF THE UTILISED DIELECTRICS

| Parameter | $Si_3N_4$ | $HfO_2$ | ITRS 2011 prediction |
|---|---|---|---|
| $V_{th}(V)$ | 0.30226 | 0.30226 | 0.302 |
| $I_{on}$ | $3.03345^{-4}$ | $2.63345^{-4}$ | $1.03^{-7}$ |
| $I_{off}$ | $2.83345^{-15}$ | $1.03345^{-15}$ | $1.495^{-6}$ |
| $\frac{I_{on}}{I_{off}}$ | $1.071^{11}$ | $2.541^{11}$ | $6.6071^{-2}$ |

From the Table it can be seen that the simulation results for $I_{off}$are slightly lower than prediction value from the ITRS-2011. This is an indication that the above selected high-κ dielectric materials are suitable combination with metal gate as well as compatible with silicon for building low leakage transistors. Again as stated earlier $HfO_2$ appears to be the better option.

## VI. CONCLUSION

NMOS structure with 22nm gate length were successfully designed and simulated to study the electrical performance of the device with a major emphasis on the device leakage current emissions. Two high-κ dielectric materials namely $HfO_2$ and $Si_3N_4$were investigated in this study. Throughout the study, it became evident that materials with high-κ dielectric are possible replacement to silicon oxide ($SiO_2$ ). Replacing $SiO_2$ with materials such as $HfO_2$(Hafniumoxide) or $Si_3N_4$ (Silicon nitride) could be some of the best mechanisms to limit leakage current emissions in CMOS devices and help in providing better resistance to attacks such as simple or differential power analysis or even electromagnetic power analysis that is gaining momentum in cryptographic devices such as smart cards.

## REFERENCES


[1] P.C Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in Proc. Of CRYPTO '99, pp. 388-397, 1999

[2] M. Alioto, L. Giancane, G. Scotti, A. Trifiletti, "Leakage PowerAnalysis Attacks: a Novel Class of Attacks to NanometerCryptographic Circuits," IEEE Trans. on Circuits and Systems – partI, vol. 57, no. 2, pp. 355-367, Feb. 2010

[3] H. Djonon Tsague and Bheki Twala, "Simulation and Parameter Optimization of Polysilicon Gate Biaxial Strained Silicon MOSFETs," The Fifth International Conference on Digital Information Processing and Communications (ICDIPC2015), Switzerland, Geneva 2015 in press.

[4] Y. Shin, J. Seomun, K.M. Choi and T. Sakurai, "Power gating: circuits, design methodologies, and best practice for standard-cell VLSI designs," ACM Trans. Des. Autom. Electron Syst. 15(4), 28:1–28:37 (2010).

[5] Integrated Circuit Characterization and Analysis Program (IC-CAP) modeling software, available from http://www.home.agilent.com, [last accessed: Sept. 2015]

[6] H.A Elgomati, B.Y Majlis, I. Ahmad, F. Salahuddin, F.A. Hamid, A. Zaharim and P.R Apte, "Investigationof the Effect for 32nm PMOS Transistor and Optimizing Using TaguchiMethod," Asian Jounal of Applied Science, 2011

[7] S.Yuan Chen et al., "Using simulation to Characterize Highperformance 65nm Node Planar," International Symposium on NanoScience and Technology, Taiwan, 20-21 Nov 2004.

[8] G. Darbandy, A. Jasmin, J. Seldmeir, U. Monga, I. Garduño, A. Cerdeira, and B. Iñiguez, "Temperature Dependent Compact Modeling for Gate Tunneling Leakage Current in Double Gate MOSFETs," Solid-State Electronics, 2013

[9] H. Wong and H. Iwai, "On the Scaling Issues and High-k Replacement of Ultrathin Gate Dielectric for Nanoscale MOS Transistor," Microelectronic Engineering, pp.1867-1904, 2013.

[10] G. He and Z. Sun, "High-k Dielectrics for CMOS technologies," John Willey and son's Inc. Aug. 2012

[11] M.G Priya, K.Baskaran and D. Krishnaveni, "Leakage Power Reduction Techniques in Deep Submicron Technologies for VLSI Applications," International Conference on Communication Technology and System Design. Vol. 30, pp 1163-1170, 2012

[12] N.S. Kim et al., "Leakage Current: Moore's Law Meets Static Power,"

[13] S.Y. Chen et al., "Using simulation to Characterize High performance 65nm Node Planar," International Symposium on Nano Science and Technology, Taiwan, 20-21 Nov 2014.

[14] F.salahuddin, I.Ahmad, and A.Zaharim, "Impact of Different Dose and Angle in Halo structure for NMOS Device," ICMST 2010, pp. 81-85, 26-28 Nov 2010.

[15] K. Ashok "Optimization of Device Performance Using Semiconductor TCAD Tools", Silvaco International Product Description, Silvaco International http://www.silvaco.com/products/descriptions.

[16] J Greer, A. Korkin and J.Lebanowsky, "Nano and Giga Challenges in Microelectronics," Molecular and Nano Electronics: Analysis, Design and Simulation, 1st Edition, 24 Oct 2003

[17] S. Chattererjee, Y. Kuo, J. Lu, J. Tewg and P. Majhi, "Electrical reliability aspects of HfO2 high-K gate dielectric with TaN metal gate electrodes under constant voltage stress", Microelectronics Reliability,46, 69-76.

[18] N. Kim, T. Austin, D. Blaauw, T. Mudge, K. Flautner, J. Hu, M. J. Irwin, M. Kandemir and V. Narayanan, "Leakage Current: Moore's

Law Meets Static Power," IEEE Computer, Vol. 36, no. 12, pp68-75, 2003
[19] Ganymede, Complementary Metal Oxide Semiconductor (CMOS), [online] http://ganymede.meccahosting.com/~a0006505/Dictionary/CMOS_Dictionary.htm, [accessed 20 October 2015]
[20] E. N. Shauly, "CMOS Leakage and Power Reduction in Transistors and Circuits: Process and Layout Considerations," Journal of Low Power Electronics and Applications, open access January 2012
[21] A. Chandrakasan, et al., "Low-power CMOS digital design," IEEE JSSC, 27(4), pp. 473~484, Apr. 1992.