IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 5, Issue 4, Ver. I (Jul - Aug. 2015), PP 31-36
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

LEAKAGE POWER REDUCTION TECHNIQUE IN CMOS CIRCUIT: A STATE-OF-THE-ART REVIEW

Himanshu Asija[1], Vikas Nehra[2], Pawan Kumar Dahiya[3]
[1](ECE Department, Govt. polytechnic, Sonipat, India)
[2,3](ECE Department, Deenbandhu Chhotu Ram University of Science & Technology, Murthal, India)

Abstract: *The demand for low power devices is increasing vastly due to the fast growth of battery operated applications such as smart phones and other handheld devices. It has become important to control the power dissipation throughout the design cycle beginning from the architectural level to final design at hardware level. Leakage current is the main factor which contributes to almost or more than 50% of total power dissipation. In many new high performance designs, the leakage component of power consumption is comparable to the switching component. More than 40% leakage in SRAM memory is due to leakage in transistors. This survey paper use the design of SRAM architecture to reduce the leakage current and hence the leakage power. The various leakage power reduction techniques have been evolved to tackle the problem and it is still in progress. In this paper mainly, there is study of various leakage power reduction techniques with SRAM architecture in fabrication Technology.*

Keywords: *CMOS, Dynamic Power, Leakage Control Transistor, SRAM, Sub-threshold Leakage Current, Threshold Voltage.*

I. Introduction

Static Random Access Memory (SRAM) has played a key role in high performance and low power VLSI applications. SRAM is most common choice for embedded-memory CMOS Integrated Circuits (ICs) [1]. System speed, power consumption and stability are the main concern for the modern processors. Growing demand for battery power handheld multimedia systems are becoming more and more popular day by day. But Complementary Metal oxide Semiconductors (CMOS) technology has continuously scaled down so it has been a major thrust to improve the performance and robustness of the memory used in these devices [2]. The power-sensitive portable devices also confronted with the need to reduce the dynamic and standby power consumption to meet the stringent battery-life requirement [3]. Power consumption of SRAM is important for increasing mobile and handheld applications where battery life is key design [4]. There are basically two types of power dissipation: 1. Static power 2. Dynamic Power. In active mode of operation, leakage is due to both dynamic and static components. In the standby mode of operation, the power dissipation is due to standby leakage current. The static power of a CMOS circuit is determined by the leakage current through each transistor. Dynamic power consists of switching power, consumed while charging and discharging the loads on a device, and internal power (also referred to as short circuit power), consumed internal to the device while it is changing state. The reduction in leakage current can be achieved at both circuit and process-level techniques. At the process-level leakage reduction can be achieved by controlling the dimensions length, oxide thickness, junction depth and doping profile in the transistors [5]. At the circuit level, threshold voltage and leakage current of transistors can be effectively controlled by controlling the voltages of different device terminals [drain, source, gate and the body (substrate)].There are various proposed techniques to reduce the leakage power. A number of leakage reduction techniques have been proposed in previous works like multiple threshold voltage (MT-CMOS) or variable threshold voltage technologies (VT-CMOS), Leakage reduction by using dynamic V_{TH}, Scaling supply voltage reduction, Leakage reduction by using Drowsy cache etc.

The main contribution of the paper is to present a state-of-the-art review of the techniques adopted by researchers to reduce the leakage power in CMOS circuits. The rest of the paper is organized as under:

Leakage mechanism in CMOS transistor is explained in Section II. Detailed survey of leakage reduction techniques is presented in Section III. Section IV explains six-transistor SRAM architecture. Leakage power reduction technique in SRAM cell is explained in Section V. Finally, the paper is concluded in the Section VI.

II. Dominant leakage mechanism in CMOS Transistor

There are four main source of leakage current in a CMOS transistor as shown in fig: 1 and are detailed as under:

a. Reverse biased junction leakage (I_{rev}): The junction leakage occurs from the source or drain to the substrate through the reverse biased diodes when a transistor is OFF. I_1 current shown in fig. 1 is the junction leakage current due to reverse-biased P-N junction. A reverse-biased P-N junction leakage has two main components: one in minority carrier diffusion/drift near the edge of the depletion region, the other is due the electron-hole pair generation [6]. The magnitude of the diode leakage current depends on the area of the drain diffusion and the leakage current density which is in turn determined by the doping concentration. If both the n and p regions are heavily doped, band-to-band tunneling (BTBT) dominates the p-n junction leakage [7]. Junction leakage has a very high dependency on the temperature.

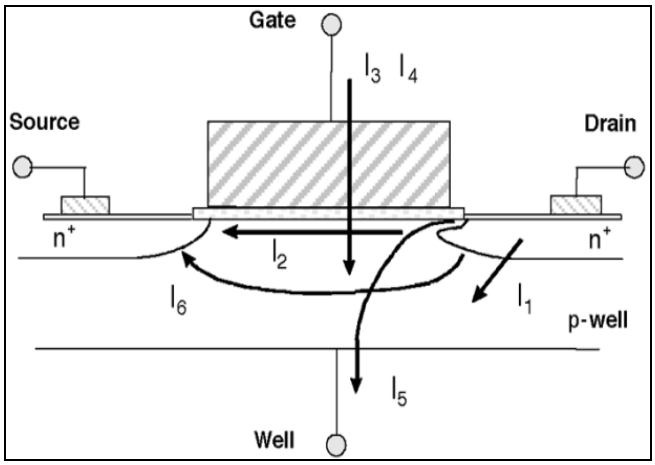

Fig. 1: Main source of leakage current in MOSFET [5].

b. Gate induced drain leakage (I_{GIDL}): The gate induced drain leakage (GIDL) is caused by high field effect in the drain junction of MOS transistors. I_5 current in fig. 1 [5] represents the leakage current due to gate induced drain leakage. For a NMOS transistor with grounded gate and drain potential at VDD, a significant band bending in the drain allow electron-hole pair generation through avalanche multiplication and the band to band tunneling. A deep depletion condition is created since the holes are rapidly swept out to the substrate. At the same time electrons are collected by the drain, resulting in GIDL current. This leakage mechanism is made work by high drain to body voltage and high drain to gate voltage. Thinner oxide and higher supply voltage increases GIDL current.

c. Gate direct tunneling leakage (I_G): With scaling of the channel length, maintaining good transistor aspect ratio by the comparable scaling of the gate oxide thickness, junction depth and depletion depth are important for ideal MOS transistor behaviour. With the technology scaling, maintaining good transistor aspect ratio has been a challenge. In other words, reduction of the vertical dimensions has been harder than that of horizontal dimensions with the silicon oxide gate thickness approaching scaling limits there is now a rapid increase in gate direct tunneling leakage current. FinFET and tri-gate MOS transistors that promise better aspect ratio are being explored.

d. Subthreshold leakage (I_{sub}): The sub threshold leakage is current flowing from drain to source when a transistor operated in weak inversion region. Unlike the strong inversion region in which the drift-current dominates, the sub threshold conduction is due to the diffusion current of the minority carriers in the channel for a MOS device. For instance, in case of an inverter with a low input voltage, the NMOS is turned off and the output voltage is high. In this case, although V_{GS} is 0V, there is still a current flowing in the channel of the OFF NMOS transistor due to the V_{DD} potential of the V_{DS}. The magnitude of the sub threshold current is a function of temperature, supply voltage, device size and the process parameter out of which the threshold voltage plays a dominant role [7]. For the current CMOS technologies, the sub threshold leakage current, I_{SUB} is much larger than the other leakage current components. This is mainly because V_T is lower in modern device.

e. Effect of Channel Length and V_{Th} Rolloff: Due to reduction in channel length the threshold voltage of MOSFET decreases. The reduction of threshold voltage with reduction of channel length is known as V_{TH} Rolloff [5]. As we know depletion region drain and source are surrounded by depletion regions. In long channel devices where drain and source are far apart, the depletion regions have not much effect on the

potential profile of the channel. But in case of short channel MOSFET where drain and source are at short distance. In this case depletion region near source drain have more effect on potential profile and field patterns. As due short channel depletion regions penetrate more and become the part of channel already depleted. So in this case gate voltage has to invert less bulk charge to turn a transistor on as a result there is more bending in the Si-SiO2 interface in short-channel devices, resulting in less threshold voltage in this case. This effect is more severe at high drain bias resulting in more decrease in threshold voltage, hence sub-threshold leakage current I_3 in fig 1 increases.

III. Survey Of Leakage Power Reduction Technique

Laxmi Singh et. al. [8] present a new 5T SRAM cell for ultra low-voltage applications. They found 36% improvement as compared to conventional 6T SRAMs. Moving from 130nm to 65nm technology area increased from 71% to 82%. Leakage power was reduced upto 21%. The whole work was done in 65nm technology. S. Mutoh [9] present approach, named "sleepy keeper," which reduces leakage current while saving exact logic state. For the most recent CMOS feature sizes (e.g., 90nm and 65nm), leakage power dissipation has become an overriding concern for VLSI circuit designers. ITRS reports that leakage power dissipation may come to dominate total power consumption. Sleepy keeper uses traditional sleep transistors plus two additional transistors – driven by a gate's already calculated output to save state during sleep mode. for applications spending the vast majority of time in sleep or standby mode while also requiring low area, high performance and maintenance of exact logic state, the sleepy keeper approach provides a new weapon in a VLSI designer's. Michael Powell et. al. [10] explore an integrated architectural and circuit level approach to reducing leakage energy dissipation in instruction caches. Deep-submicron CMOS designs have resulted in large leakage energy dissipation in microprocessors. They present gated-Vdd, a circuit-level technique to gate the supply voltage and reduce leakage in unused SRAM cells. Their results indicate that gated-Vdd together with a novel resizable cache architecture reduces energy-delay by 62% with minimal impact on performance. The ever-increasing levels of on-chip integration in the recent decade have enabled phenomenal increases in computer system performance. Unfortunately, the performance improvement has been also accompanied by an increase in a chip's power and energy dissipation. Higher power and energy dissipation require more expensive packaging and cooling technology, increase cost, decrease product reliability in all segments of computing market, and significantly reduce battery life in portable systems. Chetna et. al. [11] present Dual V_{TH} 5T SRAM cell and compared to dual V_{TH} 6T SRAM cell using 180nm CMOS technology. The leakage current for 5T SRAM cell was found to be 210mu watt and for 6T SRAM cell was found to be 390mu watt. Write delay for "0" and "1" was calculated to be 3.0038e-8ns and 1.53e-8ns respectively for 5T and 3.05e-8ns and 1.51e-8ns respectively for 6T. Vikas Nehra et. al. [2] analyze 8T SRAM in 65nm CMOS technology. They observed the read and write stability using N curve and analyzed at different corners. It is observed that write margin is better at FF corner than TT corner while noise margin is worse at FF corner and also voltage plays an important role in stability measurement.

IV. Six-Transistors SRAM Architecture

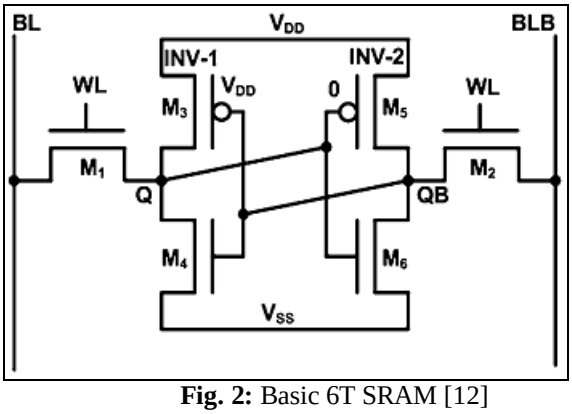

Fig. 2: Basic 6T SRAM [12]

SRAM is a type of random access memory. A random access memory is one in which the locations in the semiconductor memory can be written to or read from in any order, regardless of the last memory location that was accessed. The circuit for an SRAM memory cell is shown in Fig. 2. It consists of four transistors configured as two cross coupled inverters INV-1 and INV-2 connected in a positive feedback loop is a simple latch structure. In this the circuit has two stable states, and these are the logical "0" and "1" states. The internal nodes (Q and QB) of the bitcell contain complementary values. In addition to the four transistors in the basic

memory cell, and additional two pass transistors (M1 and M2) are required to control the access to the memory cell during the read and write operations also called access transistors [23]. They access transistors (M1 and M2) are controlled by wordline (WL) acts as switch between inverter pair and complementary pair of bitlines (BL and BLB) also called data-lines. This makes a total of six transistors and it is called as a 6T memory cell.

SRAM memory cell operation: The operation of the SRAM memory cell is relatively straightforward. When the cell is selected, the value to be written is stored in the cross-coupled flip-flops. The cells are arranged in form of matrix, with each cell individually addressable. Most SRAM memories select an entire row of cells at a time, and read out the contents of all the cells in the row along the column lines. While it is not necessary to have two bit lines, using the signal and its inverse, this in practice improves the noise margins and improves the data integrity. The two bit lines are passed to two input ports on a comparator to enable the advantages of the differential data mode to be accessed, and the small voltage swings that are present can be more accurately detected. Access to the SRAM memory cell is enabled by the Word Line. This controls the two access control transistors which control whether the cell should be connected to the bit lines. These two lines are used to transfer data for both read and write operations. The read and write operations in a standard 6T SRAM bitcell are as follows:

i. Read Operation: Due to Read operation data can be accessed from the cell. Conventially to read a bitcell, the bitlines (BL and BLB) are precharged to the supply voltage (V_{DD}) and then asserting the wordline (WL), enabling to turn on the pass-gate (PG) transistors. During read operation it is assume that internal data storage nodes Q and QB are at '0' and '1' respectively. Rise the WL from '0' to '1', result, one of the bitcell node stores the logic '0': that side of the bitcell is discharged through the pass-gate and pull-down transistors. If BLB goes to low, then the bitcell holds a logic '1' value. If BL goes to low, then the bitcell holds the logic '0' value. Depending upon whether the bitline BL or BLB is discharged, the bitcell is read as a logical '1' or '0'. A sense amplifier converts the differential signal exists on BL and BLB to a logic-level output. If the potential of the storage node goes over the trip point of the connected invertor, the stored data is flipped. For avoiding such data flipping conditions, the drivability of access transistor has to be weaker than that of Pull down (PD) transistor and this ratio called β-ratio [23]. In general, the bitcell ratio β can be varied from 1.25 to 2.5 depending on the target application and desired static noise margin (SNM).

ii. Write Operation: The write cycle starts when the BL pair is forced to the differential levels of "1" and "0" so this can be written to the corresponding storage nodes. If required to invert the storage data of the same cell, the BL pair is inverted from the differential levels compared to the previous one. When WL is made active, the pass-gate or access transistor connected to the BL is turned on and the potential of the corresponding storage node is lower and it is dependent on the ration of drivability of pass-gate and pull-up transistor. This is another SRAM design parameter ratio known as γ-ratio [12]. For a successful write operation into the cell, the critical level has to be lower than the trip point of inverter in the storage element.

iii. Hold Operations: When the world line WL is not active then SRAM cell is in data retention mode. An adequate amount of supply voltage (Vdd) is required to make the inverters on. Two cross-coupled inverter will make each other strong without any disturbance from BL through pass-gate transistors. As a result the SRAM data can hold the full potential difference of (Vdd-Vss). When the Vdd gets lower than a certain fixed value, it is known as SRAM data retention voltage (V_{hold}).

V. Leakage Reduction Techniques in SRAM Cell

The most efficient techniques used in recent memories are:

- Leakage current reduction (in active and standby mode) by utilizing multiple threshold voltage (MT-CMOS) or variable threshold voltage technologies (VT-CMOS): The replacement of faster Low-V_{TH} cells, which consume more leakage power, with slower High-V_{TH} cells, which consume less leakage power. Since the High-V_{TH} cells are slower, this swapping only occurs on timing paths that have positive slack and thus can be allowed to slow down (Fig. 3). As technologies have shrunk, leakage power consumption has grown exponentially, thus requiring more aggressive power reduction techniques to be used. Similarly, clock frequency increases have caused dynamic power consumption of the devices to outstrip the capacity of the power networks that supply them, and this becomes especially acute when high power consumption occurs in very small geometries, as this is a power density issue as well as a power consumption issue.

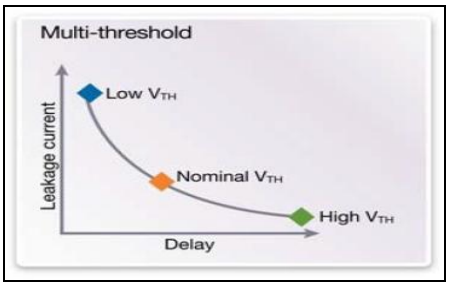

Fig. 3: Multi V_{TH} optimization [13]

- Leakage reduction by using dynamic V_{TH}: In this reduction method threshold voltage is changed dynamically to adjust the operating state of the circuit. A high threshold voltage in standby mode gives low leakage current whereas a low threshold voltage allows for higher current drives in the active mode of operation. Dynamic threshold CMOS is achieved by joining the gate and body together [14]. The p-n junction diode between source and body should be reverse biased. DTMOS can be developed in bulk technologies by using triple wells [15]. This technique is only suitable for ultralow voltage less than equal to 0.6V circuits in bulk CMOS.
- Scaling supply voltage reduction: Supply voltage scaling was originally developed for switching power reduction. It is an effective method for switching power reduction because of the quadratic dependence of the switching power on the supply voltage. Supply voltage scaling also helps reduce leakage power, since the subthreshold leakage due to DIBL decreases as the supply voltage is scaled down [51]. For a 1.2-V 0.13µ m technology, it is observed that the supply voltage scaling has significant impacts on subthreshold leakage and gate leakage (reductions in the orders of V^3 and, respectively) [13]. To achieve low-power benefits without compromising performance, two ways of lowering supply voltage can be employed: static supply scaling and dynamic supply scaling.

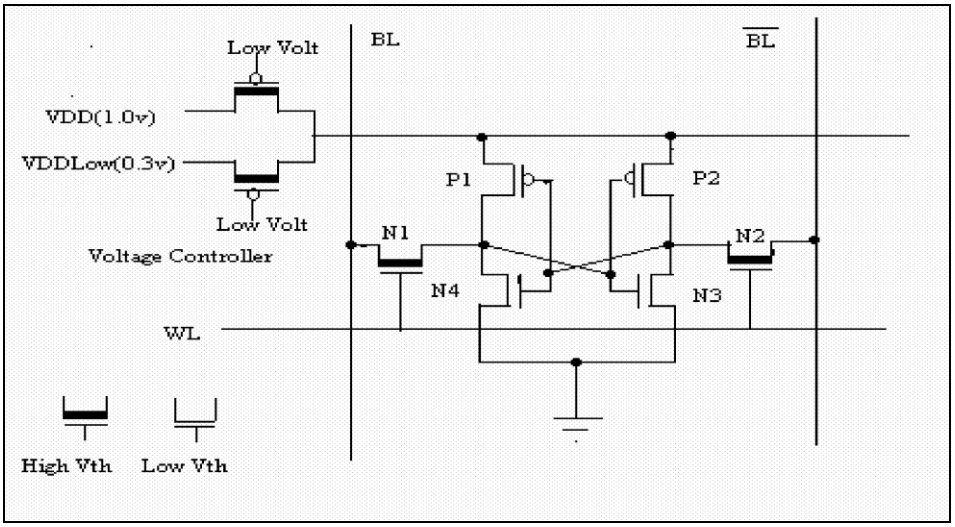

Fig. 4: Schematic of drowsy memory circuit [17].

- Leakage reduction by using Drowsy cache: By putting the cache into low power drowsy mode there is a significant leakage reduction. In drowsy mode, the information is preserved. The technique for implementing a drowsy cache is to switch between to different supply voltage in each cache line. Due to SCE in deep-submicrometer devices, subthreshold leakage current reduces significantly with voltage scaling. So the combine effect of reduced leakage and supply voltage gives large reduction in the leakage power. The schematic of a SRAM cell connected to voltage controller is shown in fig.4. Here PMOS pass-gate switch supplies to voltages one is (V_{DD}) and other is low supply voltage (V_{DDLow}). The PMOS pass-gate used are of high V_{TH}. High V_{TH} devices are used as pass transistors that connect the internal inverters of the memory cell to the read/write lines (N1 and N2). This reduces the leakage through the pass transistors, since the read/write lines are maintained in high power mode [17].

VI. Conclusion and Future Directions

The paper is mainly aimed to give a review of the various steps taken towards the reduction of the leakage power for VLSI designs. Memory hardware is very important and mandatory part of all real-time processing hardware in the present world of emerging electronic applications. So, a power efficient design is always an expectation of fabrication technology from the hardware designers. Growing complexity of mobile applications and other latest wireless applications are the features which make the battery power problem more challenging further, the existing techniques can be improved in future works in accordance with the day-by-day advancing fabrication methodologies to obtain more improved performance of the memories and other operational circuits. It may be concluded that the leakage power reduction techniques play an important role in reducing the power requirements. Researchers have contributed significantly towards this aim. However, more synergetic approaches are required to meet this aim.

References

[1]. E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," IEEE J. Solid-State Circuits, vol.SC-22, no.5; pp.748-754, Oct.1987.

[2]. Vikas Nehra, Rajesh Singh, Neeraj Kumar Shukla, Shilpi Birla, Mahima Kumar, Ankit Goel, "Simulation & Analysis of 8T SRAM Cell's Stability as Deep Sub-Micron CMOS Technology for Multimedia Applications," Canadian Journal on Electrical and Electronics Engineering Vol. 3, No.1, January 2012.

[3]. Yih Wang, Hong Jo Ahn, Uddalak Bhattacharya, Zhanping Chen, Tom Coan, Fatih Hamzaoglu, Walid M.Hafez, Chia-Hong Jan, Pramod Kolar, Sarvesh H. Kulkarni, Jie-Feng Lin, Yong-Gee Ng, lan Post, Liqiong Wei, Ying Zhang, Kevin Zhang, and Mark Bohr, "A 1.1 Ghz 12u A/Mb-Leakage SRAM Design in 65 nm Ultra-Low-Power CMOS Technology With Integrated Leakage Reduction for Mobile Applications," IEEE Journal of Solid-State Circuits, Vol. 43, No. 1, January 2008.

[4]. Faith Hamzaoglu, Kevin Zhang, Yith Wang, Hong Jo Ahn, Uddalak Bhattacharya, Zhanping Chen, Yong-Geo Ng, Andrei Pavlov, Ken Smits and Mark Bohr "A 3.8 153 Mb SRAM Design With Dynamic Stability Enhancement and Leakage Reduction in 45 nm High-k Metal Gate CMOS Technology," IEEE Journal of Solid-State Circuits, Vol.44, No. 1, January 2009.

[5]. Kaushik Roy, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits Proceedings of the IEEE," Vol. 91, NO. 2, Feb. 2003.

[6]. R. Pierret, Semiconductor Device Fundamentals. Reading, MA: Addison-Wesley, 1996, ch. 6, pp. 235–300.

[7]. Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices, New York: Cambridge Univ. Press, 1998, ch. 2, pp. 94–95.

[8]. Laxmi Singh and Ajay Somkuwar, "Design a 5T SRAM by using self controllable voltage level leakage reduction technique with CMOS".

[9]. S. Mutoh, "1-V Power Supply High-speed Digital Circuit Technology with Multithreshold-Voltage CMOS," IEEE Journal of Solis-State Circuits, Vol. 30, No. 8, pp. 847-854, August 1995.

[10]. Michael Powell Se-Hyun Yang, Babak Falsafi, Kaushik Roy, and T. N. Vijay Kumar, "Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep-Submicron Cache Memories," in proceeding ACM, ISLPED, koria 2000.

[11]. Chetna, Abhijeet, "Design of Low Power 5TDual V_{TH} SRAM-Cell," IOSR Journal of Engineering, May. 2012, Vol. 2(5) pp. 1128-1132.

[12]. Jawar Singh, Saraju P.Mohanty, Dhiraj K. Pradhan, Robust SRAM Designs and Analysis, Springer Publication, 2013.

[13]. S. Tyagi, M. Alavi, R. Bigwood, T. Bramblett, "A 130 nm generation logic technology featuring 70 nm transistors, dual V_{TH} transistors and 6 layers of Cu interconnects," Dig. Tech. Papers Int. Electron Devices Meeting, 2000, pp.267-270.

[14]. F. Assaderaghi, D. Sinitsky, S. Parke, J. Bokor, P. K. Ko, and C.Hu, "A dynamic threshold voltage MOSFET (DTMOS) for ultra-low voltage operation," Dig. Tech. Papers IEEE Int. Electron Devices Meeting, pp. 809–812, 1994.

[15]. C.Wann, F. Assaderaghi, R. Dennard, C. Hu, G. Shahidi, and Y.Taur, "Channel profile optimization and device design for low-power high-performance dynamic-threshold MOSFET," in Dig. Tech. Papers IEEE Int. Electron Devices Meeting, 1996, pp. 113–116.

[16]. A. J. Bhavnagarwala, B. L. Austin, K. A. Bowman, and J. D. Meindl, "A minimum total power methodology for projecting limit on CMOS GSI," IEEE Trans. VLSI Syst., vol. 8, pp. 235–251, June2000.

[17]. K. Flautner, N. S. Kim, S. Martin, D. Blaauw, and T. Mudge, "Drowsy caches: Simple techniques for reducing leakage power," in Proc. 29th Annual Int. Symp. Computer Architecture, 2002, pp.148-157.