



Transistor Leakage Mechanisms and Power Reduction Techniques in CMOS VLSI Design

Akanksha Dixit

Asst Professor, Dept of Electronics and Communication, Gyan Ganga College of Technology, RGPV, Jabalpur, India

Abstract: A rapid growth in semiconductor technology and increasing demand for portable devices powered up by battery has led the manufacturers to scale down the feature size, resulting reduced threshold voltage and thereby enabling integration of extremely complex functionality on a single chip. In CMOS circuits, increased sub-threshold leakage current refers static power dissipation is the result of low threshold voltage. For the most recent CMOS technologies static power dissipation i.e. leakage power dissipation has become a challenging area for VLSI chip designers. According to ITRS (International technology road-map for semiconductors), leakage power is becoming a dominant part of total power consumption. To prolong the battery life of portable devices, leakage power reduction is the primary goal. The main objective of this paper is to present the analysis of leakage components, comprehensive study & analysis of leakage components and to present different proposed leakage power reduction techniques.

Keywords: CMOS, Leakage power, VLSI circuits, multimedia applications, Static power, Nano Scale, LSSR.

I. INTRODUCTION

The rapid growth in semiconductor technology through the use of deep-submicron processes has led the feature sizes to be shrinking; thereby integrating extremely complex functionality on a single chip. In the ever increasing market of mobile hand-held devices used all over the world today, the battery-powered electronic system forms the backbone. To maximize the battery life, the tremendous computational capacity of portable devices such as notebook computers, personal communication devices (mobile phones, pocket PCs, PDAs), hearing aids and implantable pacemakers has to be realized with very low power requirements. The power dissipation has become a very critical design metric due to device miniaturization and rapid growth towards wireless communication. The longer the battery lasts; the better is the device^[13]. The power dissipation has not diminished even with the scaling down of the supply voltage. The problem of heat removal and power dissipation is getting worse as the magnitude of power per unit area has kept growing. There is a little help from advanced cooling and packaging strategies the rapid increase in power consumption of present day chips. It is already reported that Component failure rate roughly doubles for every 10°C increase in operating temperature. Following Moore's law, with the on-chip devices doubling every two years, minimizing the power consumption has become currently an extremely challenging area of research. Leakage power of a CMOS transistor depends on gate length and oxide layer thickness. To decrease the dynamic power, the supply voltage is decreased which leads to the performance degradation. To increase the operating speed the device, the threshold voltage should

also be scaled down along with the supply voltage, which results in exponential increase in the sub-threshold leakage current, thereby increase in the static power dissipation. The main components of leakage current in a MOS transistor are shown in Figure 1.1

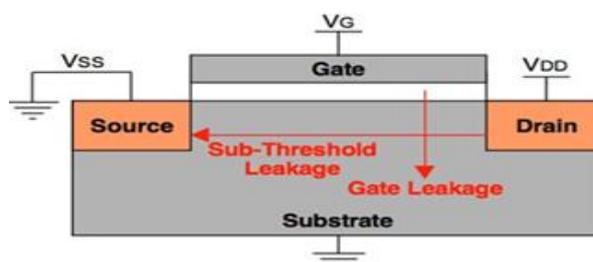


Figure 1.1. Static CMOS Leakage Power Sources [21]

II. TRANSISTOR LEAKAGE MECHANISMS

The exponential increase of the number of on chip active devices, the constant decrease of the threshold Voltage (V_{th}) and reduction of the gate oxide thickness (T_{ox}) result in a significant amount of static power in circuits designed in deep sub-micron technologies. Static power consumption is generated by leakage currents, that is the currents that flow through the devices when the total current should be 0A. Increase in leakage power is the a big concern that circuit designers need to address, particularly for circuits that have low-duty cycles, bursty operation and rely on batteries for long periods of time. There are many scenarios that generate leakage currents, and understanding such conditions is crucial to



understanding how to abate them. This chapter summarizes the main transistor leakage mechanisms for short channel CMOS devices. Although most of the discussion in this chapter is rendered for an n-type MOS transistor, an analogous analysis can be produced for a p-type MOS transistor.

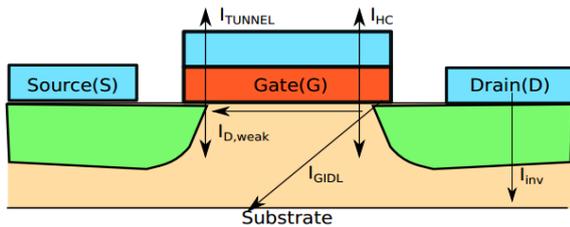


Figure 2.1 Traversal view of a MOSFET device showing all Leakage currents

Table 2.1: Leakage currents in a MOSFET device

I_{OFF}	Subthreshold drain current ($I_{D,WEAK}$) Reverse biased current (I_{INV}) Gate Induced drain leakage (I_{GIDL})
I_{GATE}	Gate tunneling (I_{TUNNEL}) Hot Carrier Injection (I_{HC})

Rather than a single component, the total leakage current is the addition of several parasitic currents. These currents can be classified in two main categories:

- i) drain leakage current that flows from the drain to the source or the body (I_{OFF}); and
- ii) leakage current that dribbles through the gates of the transistor (I_{GATE}).

I_{GATE} and I_{OFF} are the compound of currents generated by multiple physical effects under specific circumstances. Table 2.1 summarizes the main components that contribute for I_{GATE} and I_{OFF} . Figure 2.1 shows the leakage currents in a traversal cut of an n-type transistor. The rest of this chapter briefly describes the conditions that allows static power consumption in digital circuits. Subthreshold leakage current (I_{OFF})

The subthreshold leakage current of a transistor, I_{OFF} , is defined as the drain current when $|V_g| - |V_s| = 0$ and $V_d \geq 0$. I_{OFF} is dependent on the circuit topology and device physics such as V_{dd} , V_{th} , doping concentration, and gate oxide thickness T_{ox} . I_{OFF} is composed of several sub-components as shown in Fig. 1.2, and can be expressed as the sum of these components as shown in Eq. 2.1.

$$I_{OFF} = I_{INV} + I_{D,WEAK} + I_{GIDL} \quad 2.1$$

Reverse biased diode current (I_{INV})

I_{INV} is the current that flows through the reverse biased diode between the drain (n-region) and the p-region of the transistor, and it is dependent on the junction area between

the Source/Drain terminals and the body and exponentially dependent to the temperature [23, 3].

The leakage current for the inverse biased diode is given by Eq. 2.2, where U_T is the thermal Voltage, a parameter linearly dependent on the voltage. I_s is an intrinsic parameter for the device, usually known as the reverse saturation current and V_d is the voltage between the drain and the body of the transistor. Because of the exponential dependence on the voltage, any small perturbation will set the value of the reverse biased diode current (I_{INV}) near the value of the saturation current (I_s). Sometimes, electric data-sheets only provide the saturation current density J_{inv} , in which case one can compute the reverse biased diode current by multiplying the current density by the diffusion area A_d as described by equation 2.3.

$$I_{INV} = I_s(e^{(V_d/U_T)} - 1) \quad 2.2$$

$$I_{INV} = A_d \times J_{inv} \quad 2.3$$

Subthreshold drain current ($I_{D,WEAK}$)

Consider a transistor with $V_g < V_{th}$, $|V_d| \geq 0.1$ and $V_s = V_b = 0$. Under these conditions, the transistor is said to be in weak inversion. A transistor in weak inversion has a constant voltage across the channel and the magnitude of the longitudinal component of the electric field across the channel is 0. Hence, there is no drift current. Instead, the leakage current $I_{D,WEAK}$ is produced by the diffusion of majority carriers across the channel [36]. $I_{D,WEAK}$ can be modeled as described in Eq. 2.4, where U_t is the thermal voltage and I_0 is an initial DC offset in the drain current. It is important to note the exponential dependency of $I_{D,WEAK}$ on V_{gs} as well as a linear offset based on V_{DS} .

$$I_{D,WEAK} = \frac{W}{L} \times I_0 \times e^{(V_{gs} - V_{th})/(mU_T)} \times 1 - e^{-V_{ds}/mU_T} \quad 2.4$$

Gate-induced drain current (I_{GIDL})

Gate-induced drain leakage, I_{GIDL} , is generated when a large-enough gate-to-drain (V_{gd}) voltage is applied to produce a band-to-band electron tunneling near the interface between the gate oxide and the semiconductor of the drain.

With the increase in the leakage current more and more, as will be seen that it becomes proportional to the total power dissipation as given by following equation.

$$P_{leak} = I_{leak} \times V_{dd} \quad 2.5$$

In today's world there is a huge demand for minimizing the dynamic power dissipation and scaling down the supply voltage by pushing the circuits design towards ever-shortening channel lengths in CMOS technology [19]. To maintain the circuit speed, the transistor threshold voltages must also be scaled down [1]. This can be easily seen from the first-order propagation delay equation of a transistor given by following equation.



$$\tau = \frac{C \cdot V_{dd}}{(V_{dd} - V_t)^x}$$

Where C is the load capacitance, V_t is the threshold voltage; x (which is greater than 1 but less than 2) models the short channel effect. The sub threshold leakage current exponentially increases as V_t is reduced. It has been shown that as the technology scales down below 100nm which is the shrinking of feature size of transistor, the channel length decreases, thereby increasing the amount of leakage power in the total power dissipated. As we can see as the technology is moving towards lower nanotechnology the sub-threshold leakage increases thereby affecting the battery life. Thus there were various technique developed to deal with this problems. The various techniques will be discussed in the next section along with their drawbacks.

III. STATIC POWER REDUCTION TECHNIQUES

Static power reduction techniques can be classified in three main categories depending on the granularity at which can be applied:

1. Device Engineering. It refers to techniques that are implemented on the underlying transistors that conform circuits.
2. Circuit Engineering. It refers to techniques that are applied to gates, which are clusters of transistors that perform a small computation like NAND, NOR.
3. System Engineering. It refers to techniques that can be applied to macroblocks that are part of a big datapath or micro-chip.

Device Engineering

Researchers have developed engineering techniques to reduce leakage on CMOS devices. Unfortunately, these techniques usually have trade-offs in performance and area. In order to control static power, device engineers can modify certain dimensions of the device (e.g., L_{eff}, T_{ox}, substrate depth, (Source-Drain overlap), the nominal values of V_{dd} and V_{th}, the materials used (choice of gate dielectric, semiconductor), the FET-type (depleted devices, bulk devices, multiple gate devices), the doping profile and doping halo[30]. The advantages of tuning transistors are vast since leakage reduction is usually significant and devices are agnostic to circuit paradigms and logic families. The main disadvantage of device engineering is that oftentimes the circuit designer has no control over the selection of devices and most likely lacks the expertise, time, and budget to modify the underlying transistors that conform circuits.

Circuit Engineering

There exists a collection of techniques that assist the circuit designer to reduce leakage power at the gate level. Probably the most known technique is forced transistor

2.6

stacking, which exploits the dependence of the subthreshold current I_{d,weak} on V_{th} and V_{gs} as described by Eq. 2.4. There are two flavors of transistor stacking, the first flavor consists on exploiting the natural transistor stacks like the ones found in NAND and NOR gates by setting proper input vectors to increase the number of off transistors in a series of transistors.

The second flavor is to forcefully add extra transistors in series to the gates to reduce the leakage power. Forced stacking generally comes at expense of an increased gate delay in active operation. Figure 3.1[A] shows an example of natural transistor stacking on an NAND gate. The main problem of low-level circuit engineering is that it usually requires a lot of manual engineering of the gates and circuit designers generally need to work really hard to achieve a significant reduction on the leakage power.

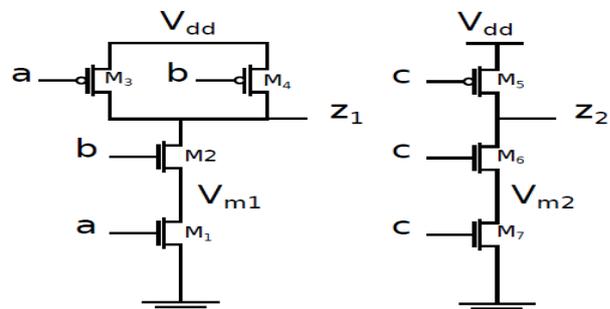


Figure 3.1: [A] Natural stacks found in a NAND gate. [B] Forced stacking on the pull-down network of an inverter

System Engineering

Finally, there exist some techniques that can be applied system-wide to reduce leakage and usually render a lot of static power savings. Some of these techniques are design specific, for example the choice of an SRAM cell or a register cell or the specific implementation of a datapath unit. However there are two general techniques that can be applied to (almost) any circuit: Clock Gating and Power Gating. Asynchronous circuits are data driven in nature, implementing an equivalent fine-grain clock gating design. Hence, power gating is the unrivaled systematic technique to reduce static power in an asynchronous pipeline. This thesis explores all the power gating techniques as well as the best way to implement power gating in asynchronous pipelines and how to exploit the unique capabilities of asynchronous circuits in the context of power gating.

IV. RELATED WORK DONE

Many techniques have been come into existence to overcome the leakage power problem in the nano-scale technology, but those techniques have tradeoff between area, delay and also active power. Some of those techniques are as described in this section.



4.1. Dual Vt and MTCMOS

This was the earliest suggested technique to reduce the leakage power. As stated in [1,7] Dual VT technique is a variation in MTCMOS, in which the gates in the critical path use low-threshold transistors and high-threshold transistors for gates in non-critical path. According to the authors both the methods requires additional mask layers for each value of Vt in fabrication, which is a complicated task depositing two different oxides thickness, hence making the fabrication process complex. Moreover the techniques also suffer from turning-on latency i.e., the idle of circuit cannot be used immediately after reactivated since sometime is needed to return to normal operating condition. The latency is typically a few cycles for former method, and for Dual technology, is much higher. When the circuit is active, these techniques are not effective in controlling the leakage power.

4.2. Sleep Mode Approach

This method was developed to overcome the disadvantages of the dual Vt and MTCMOS technique. According to [8] it is one of the most commonly known traditional approaches for sub threshold leakage power reduction is the sleep transistors (sleep transistors) are inserted in between the power supply and ground transistors (sleep transistors) are inserted in between the power supply and ground As explained in [1] in this technique an additional "sleep" PMOS transistor is placed between VDD and the pull-up network of a circuit and an additional "sleep" NMOS transistor is placed between the pull-down network of the circuits and GND [6]. These sleep transistors turn off the circuit by cutting off the power rails.

The sleep transistors are turned on when the circuit is active and provide very low resistance in the conduction path so that circuit's performance will not get affected due to these additional transistors. During the standby mode the leakage power is reduced in the circuit by making transistors turned off which introduces large resistance in the conduction path. Thus leakage power can be reduced effectively by switching off the power source. These types of techniques are also called gated-VDD and gated- GND.

4.3 Stack Approach

The sleep technique though proved to be better than dual Vt and MTCMOS technique but however could not give a satisfying result in reducing the leakage power. This led to the authors in [8] to design a new better circuit and in this race they suggested a new technique called the stack technique which forces a stack effect by breaking down an existing transistor into two half size transistors. The authors suggested the circuit as shown in Figure 2. It is shown that induced reverse bias between the two transistors results when the two transistors are turned off together resulting in sub-threshold leakage current reduction. But the disadvantage is increase delay significantly between

divided transistors which could limit the usefulness of the approach.

4.4. Sleepy Keeper Approach

The basic problem with traditional CMOS is that the transistors are used only in their most efficient, and naturally inverting, i.e., PMOS transistors are connect to VDD and the NMOS transistors are connect to GND. It is a

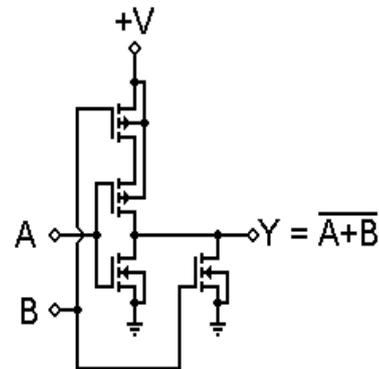


Figure 4.1 Stack mode approach using 2 input NOR gate [22]

well known fact that the PMOS transistors are not efficient at passing GND and that the NMOS transistors are not efficient at passing VDD. However, to maintain a value of '1' in sleep mode, given that the '1' value has already been calculated, the sleepy keeper approach uses this output value of '1' and an NMOS transistor connected to VDD to maintain output value equal to '1' when in sleep mode. To tackle this problem the authors suggested a new design using sleepy keeper approach in which an additional single. NMOS transistor placed in parallel to the pull-up sleep transistor connects VDD to the pull-up network. When in sleep mode, this NMOS transistor is the only source of VDD to the pull-up network since the sleep transistor is off. An additional single PMOS transistor placed in parallel to the pull-down sleep transistor is the only source of GND to the pull-down network.

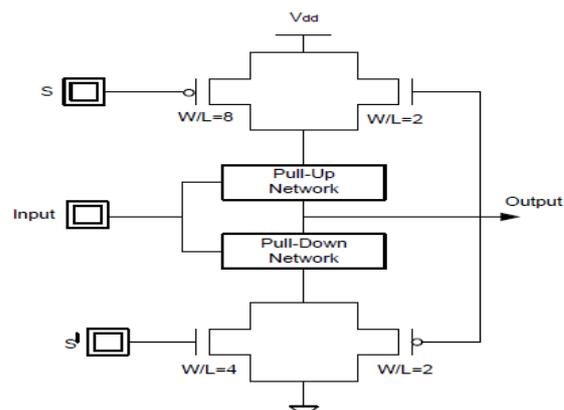


Figure 4.2. Sleepy Keeper circuit [11,23]



4.5. LECTOR Technique

This is one of the low power retention techniques as stated in [1]. The authors suggested a CMOS circuit in which two extra Leakage Control Transistors (a P-type and an N-type) is inserted within the gate, in which the gate terminal of each Leakage Control Transistor is controlled by the source of the other. The circuit can be viewed in Figure 4.3.

The basic idea behind their approach was for reduction of leakage power is the effective stacking of transistors in the path from supply voltage to ground. The authors in [8, 10, 16], and [17] made an observation that “a state with more than one transistor OFF in a path from supply voltage to ground is far less leaky than a state with only one transistor OFF in any supply to ground path.” In their method they introduced two leakage control transistors (LCTs) in each CMOS gate such that one of the LCTs is near its cutoff region of operation.

4.6 LSSR (Lector Stack State Retention Technique)

In order to achieve low power they formed this new circuit design by combining two previously done approaches namely LECTOR approach and Forced stack approach. Since it combines the two above mentioned techniques it has the features of both the approaches and thus is much beneficial than the previous works done.

The authors in [1] have proposed the circuit by introducing two gated leakage transistors between pull up and pull down networks with high threshold voltage, and then stack effect is added to pull up and pull down networks by dividing each transistor in to half size transistors.

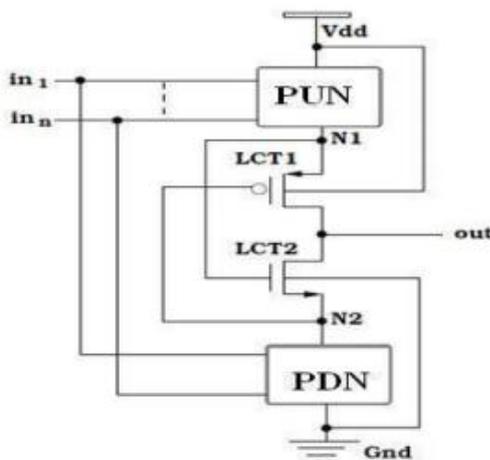


Figure 4.3. LECTOR circuit design [1,8]

this new technique LSSR can be proved to be much better than the earlier works done which can achieve better leakage reduction by maintaining exact logic state(state retention) than the other techniques.

V. CONCLUSION

The main of this paper was to give a review of the various steps taken towards the reduction of the leakage power for VLSI designs. A major thrust towards the low power design of CMOS is actually due to recent technological advances in wireless communication because the usable time of a mobile device is heavily restricted by its battery life. With the growing complexity of mobile devices, such as with a digital camera, multimedia services, Video Conferencing, global positioning system (GPS) etc are the features which make the battery power problem more challenging. To solve the problem faced various works have been implemented and still technicians are working on this field. However through this paper we get to know the advantages and disadvantages of various works done. Finally it is concluded that the optimized layout will also play an important role in reducing the leakages.

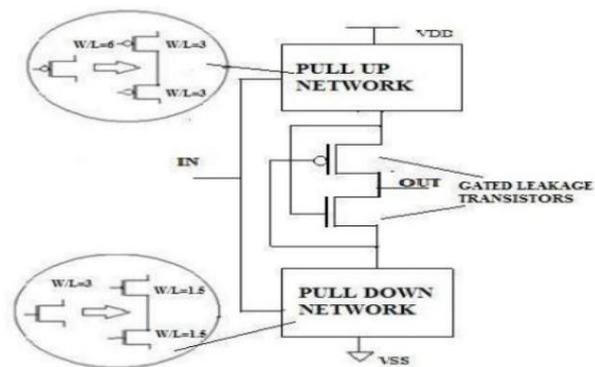


Figure 4.4. LSSR circuit design

REFERENCES

- [1] Praveen Kumar, Pradeep SR, Pratibha SR, “LSSR: LECTOR Stacked State Retention Technique a novel leakage reduction and state retention technique in low power VLSI design,” IJERT, vol. 2, pp. 1-4, October 2013.
- [2] Velicheti Swetha, S.Rjeshwari, ”Design and Power Optimization of MT-CMOS circuits using power gating techniques,” IJAREEIE, vol. 2, August 2013.
- [3] Vinay Kumar Madasu, B Kedharnath, “Leakage power reducing by using sleep method,” IJECS, vol.2, pp. 2842-2847, September 2013.
- [4] Hina malviya, Sudha Nayar, “A new approach for Leakage Power Reduction Techniques in Deep Submicron Technologies in cmos circuit for vlsi applications .” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013.
- [5] B.Dilip, P.Surya Prasad and R.S.G. Bhavani, “Leakage power reduction in CMOS circuits using leakage control transistor technique in nanosacle technology,” IJESS, vol. 2, 2012.
- [6] Dhananjay E. Upasani, Sandip B. Shrote, “Standby leakage reduction in nanoscale CMOS VLSI circuits,” International Journal on computer applications, vol. 7, September 2010.
- [7] Jun Seomun and youngsoo Shin, “Design and optimization of power –gated circuits with autonomous data retention,”IEEE transactions on VLSI system, vol. 19, no. 2, February 2011.
- [8] Narender Hanchateand and Nagarajan Ranganathan, “LECTOR: A technique for leakage reduction in CMOS circuits,”IEEE



- transactions on VLSI systems, vol. 2, no. 2, pp. 196-200, February 2004.
- [9] K.Flautner, S.Reinhardt, T.Mudge, "Automatic performance setting for dynamic voltage scaling,"7th International Conference on Mobile Computing and Networking, Rome, Italy, 2001. .
- [10] S.H Kim and V.J. Mooney, "Sleepy Keeper: A new approach to low leakage power VLSI design," IFIP, pp. 367-372, 2006.
- [11] Kyung Ki Kim, Yong-Binki, "Optimal body biasing for minimum leakage power in standby mode,"IEEE International Symposium on Circuits and Systems, pp. 27-30, May 2007..
- [12] M. Johnson, D. Somasekhar, L.-Y.Chiou, and K. Roy, "Leakage control with efficient use of transistor stacks in single threshold CMOS," IEEE Trans. VLSI Systems., vol. 10, no. 1, pp. 1-5, Feb. 2002
- [13] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multi-threshold voltage CMOS," IEEE J. Solid-State Circuits, vol. 30, pp. 847-854, Aug. 1995.
- [14] Steven Keeping, "Design techniques for extending Li-ion battery life", November 19, 2013.
- [15] Ken Bigelow, "Inside Computer Logic Gates".
- [16] Se Hun Kim and Vincent J. Mooney III , "Sleepy Keeper : a New Approach to Low-Leakage Power VLSI Design", in VLSI SOC conference 2006, PP. 367-372.
- [17] N. Hanchate and N. Ranganathan," Lector: A technique for leakage reduction in CMOS circuits", IEEE Transactions on Very Large Scale Integration (VLSI) Systems , vol. 12, no. 2, pp. 196-205, February 2004.