



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

Efficient Power Management Technique for Deep-Submicron Circuits

P.Sreenivasulu¹, Ch.Aruna² Dr. K.Srinivasa Rao³, Dr. A.Vinaya babu⁴

¹Research Scholar, ECE Department, JNTU Kakinada, A.P, INDIA.

² M.Tech (DECS) Student, Dr.S.G.I.T, Marakapur.

³Principal and Professor of ECE, T.R.R College of Engineering, Hyderabad, A.P, INDIA

⁴ principal of JNTU college of Eng., JNTUH, Kukatpally, Hyderabad, A.P, INDIA

Abstract—the state dependence of leakage can be exploited to obtain modest leakage savings in CMOS circuits. However, one can modify circuits considering state dependence and achieve larger savings. We identify a low leakage state and insert leakage control transistors only where needed. Leakage levels are on the order of 35% to 90% lower than those obtained by state dependence alone. Consequently, leakage control and reduction are very important, especially for low power applications. The reduction in leakage current has to be achieved using both process and circuit level techniques. Low supply voltage requires the device threshold to be reduced in order to maintain performance. Due to the exponential relationship between leakage current and threshold voltage in the weak inversion region, Leakage power can no longer be ignored. In this paper we present a technique to accurately estimate leakage power by accurately modeling the leakage current in transistor stacks.

Index Terms— Clock Frequency, process technology, Die Size.

I. INTRODUCTION

HIGH-PERFORMANCE VLSI design with low supply voltage (V_{DD}) becomes one of the most important issues in CMOS VLSIs, since main-stream will be scaled down to below 0.5 V in the coming years. The power and the delay dependence on the threshold voltage at 0.5 V are shown in Fig. 1. As seen from the figure, the threshold voltage has to be decreased to achieve high performance. Reducing, however, could cause a significant increase in the static leakage power component. In particular, when the threshold voltage is lower than 0.1 V, the leakage power becomes a dominant component in the total power consumption even in the active mode. In order to suppress the power consumption in low-voltage processors, it is necessary to reduce the leakage power component in the active mode.

Many embedded designs [11], instead of gating the cell, use circuit only techniques [7] and primarily rely on a dual-threshold voltage (dual- V_t) process technology [8, 9] to reduce leakage. Dual V_t makes sense till V_{DD} is 1V and V_t is around 0.25~ due to V_t spread [8] and also it requires extra process cost. By providing an alternative solution, our single V_t integrated circuit /architecture approach to reduce leakage for high-performance designs offers a key advantage over the dual- V_t approach. We adopted an IPS scheme which cuts off the virtual power links from the power supply and connects them intermittently during a standby penal. As a result, we realize a new data holding circuits without increase of area and off-leakage current.

Advanced logic CMOS technology, when scaled to the next generation, improves (1) transistor and interconnect performance, (2) transistor density, and (3) energy consumed per switching transition. Technology scaling with 30% reduction in minimum feature size per generation has three primary goals: (1) reduce gate delay by **30%**, (2) Double transistor density, and (3) reduce energy per transition by **30%** to **65%**, depending on the degree of Supply voltage reduction. These technology improvements, coupled with advances in circuits and micro architecture, are expected to sustain historical trends in clock frequency, die size, functional integration, and power dissipation of high-performance Microprocessors.

II. TRANSISTOR LEAKAGE MECHANISMS

The trends in transistor characteristics, clock frequency (at product introduction), transistor density, multilevel inter connections, power, and die size of high-performance microprocessors over the last few technology generations to evaluate how well the technology and design goals, as dictated by the 30% scaling theory, have been met. Possible Six short channel leakage mechanisms are as follows:

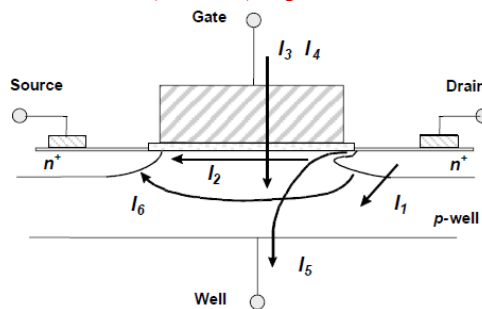


Fig. 1: Summary of leakage current mechanisms of deep submicron transistors

A. PN JUNCTION REVERSE BIAS CURRENT (I1)

For an MOS transistor, additional leakage can occur between the drain and well junction from gated diode device action (overlap of the gate to the drain-well *pn* junctions) or carrier generation in drain to well depletion regions with influence of the gate on these current components. *pn* junction reverse bias leakage (IREV) is a function of junction area and doping concentration. If both *n*- and *p*-regions are heavily doped (this is the case for advanced MOSFETs using heavily doped shallow junctions and halo doping), Band-To-Band Tunneling (BTBT) dominates the *pn* junction leakage [3].

B. Sub threshold Leakage (I2)

In a short channel device, however, the source and drain depletion width in the vertical direction and the source-drain potential have a strong effect on the band bending over a significant portion of the device. Therefore, the threshold voltage and consequently the sub threshold current of short channel devices vary with the drain bias. This effect is referred to as Drain-Induced Barrier Lowering (DIBL).

$$I_{ds} = \mu_0 C_{ox} \frac{W}{L} (m - 1) (v_T)^2 \times e^{\frac{(V_g - V_{th})}{m v_T}} \times \left(1 - e^{-\frac{V_{ds}}{v_T}} \right)$$

$$m = 1 + \frac{C_{dm}}{C_{ox}} = 1 + \frac{\frac{\epsilon_{si}}{W_{dm}}}{\frac{\epsilon_{ox}}{t_{ox}}} = 1 + \frac{3 t_{ox}}{W_{dm}}$$

C. Tunneling Into and Through Gate Oxide (I3)

The high electric field coupled with low oxide thickness results in tunneling of electrons from substrate to gate and also from gate to substrate through the gate oxide, resulting in the gate Oxide tunneling current. The mechanism of tunneling between substrate and gate poly-silicon can be primarily divided into two parts, namely, (I) Fowler-Nordheim (FN) tunneling and (II) direct tunneling.

D. Injection of Hot Carriers from Substrate to Gate oxide (I4)

The injection from Si to SiO2 is more likely for electrons than holes as electrons have a lower effective mass than that of holes and the barrier height for holes (4.5 eV) is more than that for electrons (3.1 eV) [6]. a short channel transistor, due to high electric field near the Si/SiO2 interface electrons or holes can gain sufficient energy from the electric field to cross the interface potential barrier and enter into the oxide layer. This effect is known as hot carrier injection.

E. Gate Induced Drain Leakage (I5)

When the negative gate bias is large (i.e. gate at zero or negative and drain at VDD), the n+ drain region under the gate can be depleted and even inverted. This causes more field crowding and the peak field increase, resulting in a dramatic increase of high field effects such as avalanche multiplication and band-to-band tunneling. As a result of all these effects, minority carriers are emitted in the drain region underneath the gate. Since the substrate is at a lower potential for minority carriers, the minority carriers that have been accumulated or formed at the drain depletion region underneath the gate are swept laterally to the substrate, completing a path for the GIDL.

F. Punch through (I6)

As the channel length is reduced, if the doping is kept constant, the separation between the depletion region boundaries decreases. An increase in the reverse bias across the junctions (with increase in Vds) also pushes the junctions nearer to each other. When the combination of channel length and reverse bias leads to the merging of the depletion regions, punch through is said to have occurred. In sub-micron MOSFETs a Vth adjust implant is used to have a higher doping at the surface than that in the bulk. This causes a greater expansion of the depletion region below the surface (due to smaller doping there) as compared to the surface. Thus the punch through



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

occurs below the surface [8]. An increase in the drain voltage beyond the value required to establish the punch through, lowers the potential barrier for the majority carriers in the source. Thus, more of these carriers cross the energy barrier and enter into the substrate, and the drain collects some of them.

III. LEAKAGE REDUCTION TECHNIQUES

A. Transistor Scaling Trends

Over the last few technology generations, we have pursued the path of constant electric field scaling, where the maximum supply voltage is limited by gate oxide wear-out. Thus, the gate length, effective electrical gate oxide thickness, and supply voltage have scaled by approximately 30% per unit width. The worst-case sub threshold leakage current has remained approximately constant. Thus, reduction of threshold voltage from one technology to the next has been limited by our ability to (1) control short-channel effects through innovative channel and junction engineering, and (2) aggressively scale the control of all lateral and vertical dimensions, especially gate length. In spite of the stringent sub threshold leakage current limitation, we have been able to maintain approximately constant drive current per unit width of the transistor, at reduced supply voltages. All components of device capacitance, except the areal component of the drain-to-body junction capacitance, remain constant per unit width; the areal component of junction capacitance scales by 30% per unit width. As a result, the delay of a gate whose output load is dominated by device capacitances has reduced by 30% per technology generation. If the number of transistors per die remain constant, and no major transistor resizing is performed, the die area reduces by 50%, and the total capacitance on chip reduces by 30%. The transistor capacitance per unit area, however, increases by approximately 43%.

B. Clock Frequency

Assuming that a technology generation spans 2-3 years, the data in Figure 1 clearly shows that the microprocessor frequency has been doubling every technology generation-not just by 43%. This could be attributed to several factors. Let us consider clock period/ average-static-gate-delay plotted on the right hand Y axis in Figure 1 (in arbitrary units). The average number of gate delays in a clock period is decreasing, indicating that the new micro architectures employ shorter pipelines for static gates, and utilize advanced circuit techniques to reduce the critical path delays even further. This could be one of the reasons why the Frequency is doubling every technology generation. One might suspect that this increase in frequency may be at the expense of over-design or over-sizing of transistors. Fig3.

C. Transistor Density

We define transistor density as the number of logic transistors packed per unit area. The transistor density is expected to double every technology generation, since the area reduces by 50% (scaling theory). Memory density has been scaling as expected, and therefore we focus our study on the logic transistor density. Figure 3 plots logic transistor density of some of the microprocessors, and their shrinks and compactions, across different technologies. The dotted line shows the expected 2X density trend. Notice that when a processor design is ported to the next process technology, it meets the density goal; however, a new processor micro architecture implemented on the same technology shows a drop in density. We suspect that this is due to the complexity of the new micro architectures, as well as limited resources available to accomplish a more complex design.

D. Interconnect Scaling

In order to meet the technology goals the interconnection system has to scale accordingly. In general, as the width and thickness of the interconnections are reduced, the resistance increases, and as the interconnections get closer, the capacitance increases. The size of a chip should reduce by 30%. which is true for shrinks and compactions of chips onto the next technology; however, to further exploit integration, new designs add more transistors on the chip. As a result, the average die size of a chip tends to increase over time. To account for increased parasitic (R & C), and increased integration and complexity, more interconnect layers are added. The thinner, tighter interconnect layers get used for local interconnections, and the new thicker and sparser layers get used for global interconnections and power distribution.

E. Power

Maximum power consumption of a chip depends on the technology as well as implementation. According to the scaling theory, a design when ported to the next generation technology would operate at 43% higher frequency. Since the total capacitance and the maximum supply voltage reduce by 30%, the total energy consumed in a clock cycle must reduce by 65%. Assuming that the average number of switching transitions per cycle has remained unchanged. Then, the power should reduce by 50%. If the electric field is below the maximum sustainable for gate oxide reliability, then the supply voltage may not reduce by 30%. In the extreme case, both the maximum supply voltage and threshold voltage will remain constant. The energy consumed per clock cycle

then reduces by 30%, and the power remains constant. Of course, there is then no room for adding more transistors to the chip without increasing the power budget from previous generation.

F. Die Size

We have not only taken advantage of increased transistor density, but have also increased the size of the chip (die) to further the level of integration. Figure 8 plots die size of lead microprocessors in mils (1 mil = 1/1,000 inch) over time, and shows that the die size tends to grow about 25% per technology generation. Loosely speaking, this satisfies Moore's Law.

IV. SIMULATION RESULTS

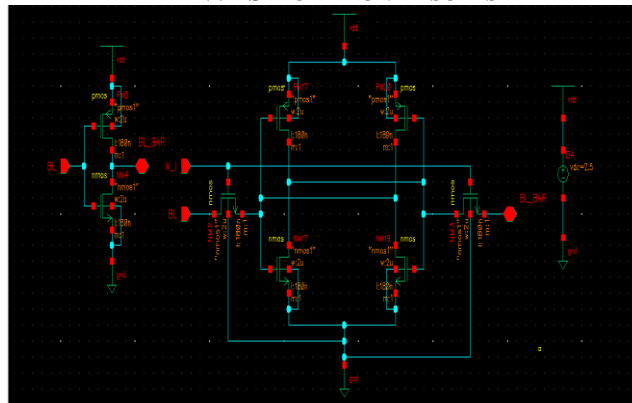


Fig. 2. Leakage feedback MTCMOS flip flop.

One PMOS pass gate switch supplies the normal supply voltage (VDD) (in the active mode), and the other supplies the low supply voltage (VDDLow) (in the stand-by mode) for the drowsy cache line. Each pass gate is a high-Vth device to prevent leakage current from the normal supply to the low supply through the two PMOS pass gate transistors.

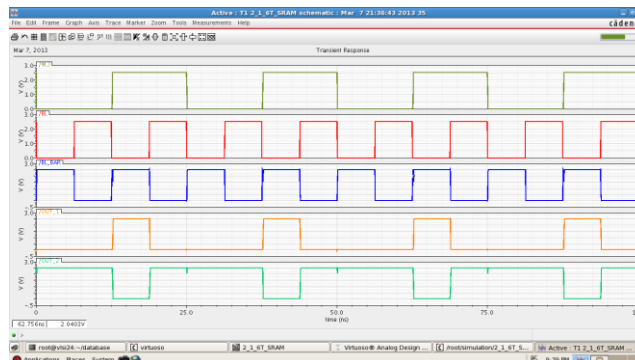


Fig. 3. Waveforms of Leakage feedback MTCMOS flip flop.

Significant leakage reduction can also be achieved by putting the cache into a low power drowsy mode [33]. In the drowsy mode, the information in the cache line is preserved. However the line has to be reinstated to a high power mode before its contents can be accessed. One technique for implementing a drowsy cache is to switch between two different supply voltages in each cache line

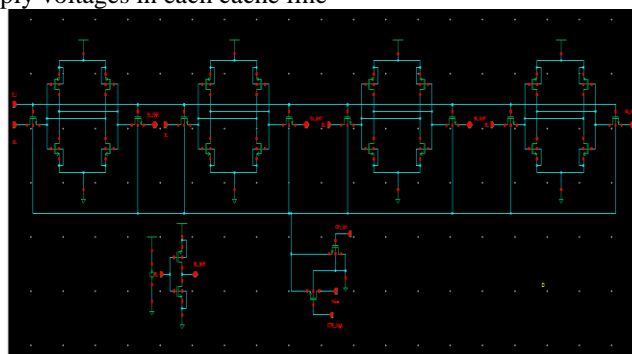


Fig. 4. Schematic of a dynamic Vth SRAM set



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

The MT-CMOS circuit technology utilizes MOSFETs with both high-V_{th} and low-V_{th} in one LSI chip. Figure 1 shows the basic MT-CMOS circuit scheme.

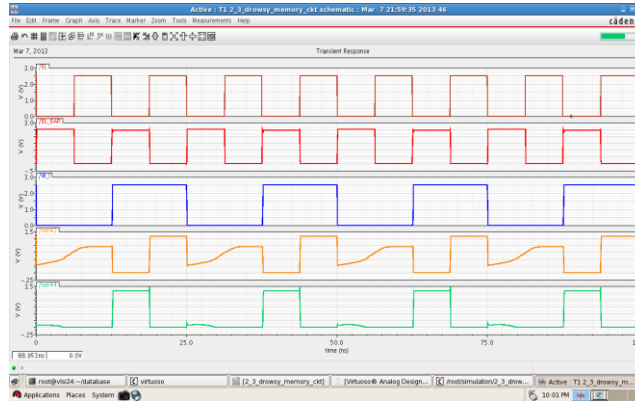


Fig. 5. Waveforms of dynamic V_{th} SRAM set.

To achieve low-power benefits without compromising performance, two ways of lowering supply voltage can be employed: static and dynamic supply scaling. In static supply scaling, multiple supply voltages are used as shown in Fig. 17.

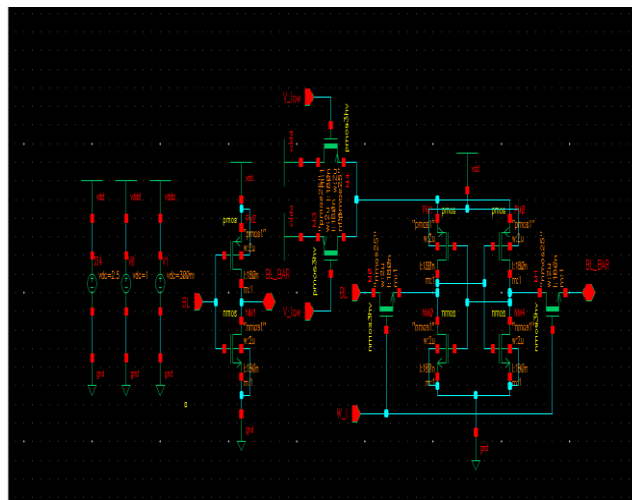


Fig. 6. Schematic of drowsy memory circuit

The error signal, which is the difference between the reference clock frequency and the oscillator frequency, is fed into the feedback controller.



Fig. 7. Waveforms of drowsy memory circuit

A separate voltage controller is needed for each cache line. By scaling the voltage of the cells to approximately 1.5 times of V_{th} the state of the memory cell can be maintained. For a typical 70nm process, the drowsy voltage is about 0.3V.

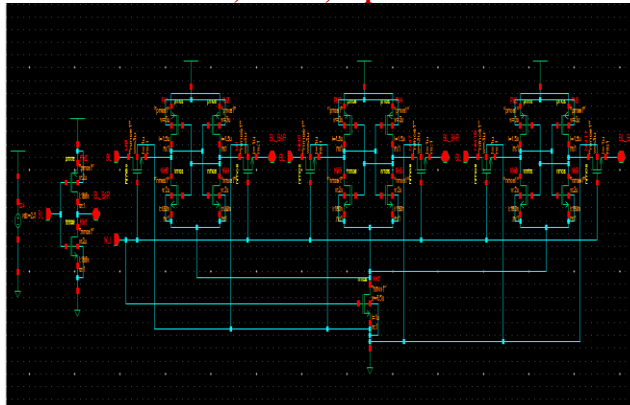


Fig. 8. Anatomy of Data Retention Gated-Ground Cache (DRGCache)

In order to suppress the leakage power further, combining the -hopping scheme and the dual- scheme could be useful. Fig. 10 shows the schematic of this scheme. In this scheme, -hopping is used only in the critical paths. On the other hand, of the noncritical gates is set to a considerably higher value (), which is not changed for the entire time.

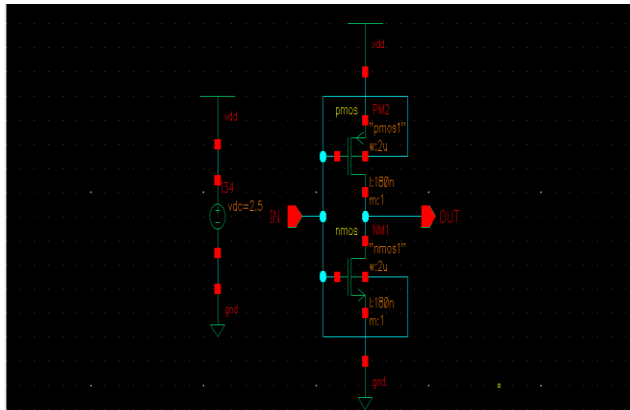


Fig. 9. Schematic of DTMOS inverter

The highest supply voltage delivers the highest performance at the fastest designed frequency of operation. When performance demand is low, supply voltage and clock frequency is lowered, delivering reduced performance but with substantial power reduction.

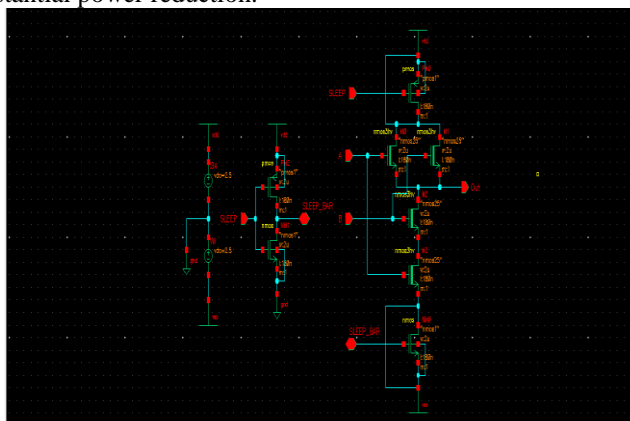


Fig. 10. Schematic of MTCMOS circuit

These two leakage paths make up a high percentage of the total leakage [35]. Fig. 22 shows the schematic of a DTSRAM cache line.

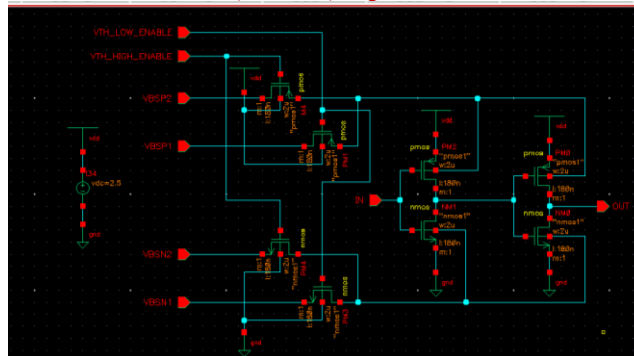


Fig. 11. Schematic diagram of Vth-hopping

The NMOS substrate can be switched to 0V for high performance. When the cache line is not in use, the substrate can be switched to a negative voltage (Vbs) to reduce the leakage. Since the transition energy required for a single substrate bias transition is much more than the leakage energy saved during one clock cycle, Vth transition cannot be made every clock cycle [35]. Moreover, the performance loss due to negative body bias (i.e. high Vth) is considerable. To overcome these difficulties, properties of temporal and spatial locality of cache access can be used. In [35], a time based scheme is described, which instead of turning a cache line to high Vth state right after its access, leaves the cache line in low Vth for a certain time period (30 μ s ~100 μ s).

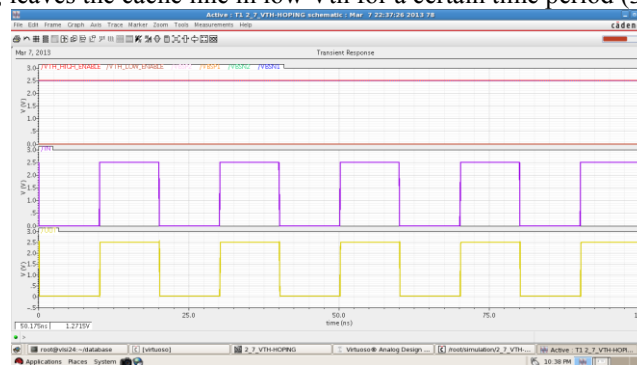


Fig. 12. Waveforms of Vth-hopping

V. CONCLUSION

A threshold voltage hopping (-hopping) scheme is proposed where the threshold voltage, V_{th} , is dynamically controlled through software depending on the workload of a processor. The -hopping scheme can achieve 82% power saving compared with the fixed low- circuits in a 0.5-V supply voltage regime for multimedia applications. -hopping is effective in the low designs where V_{th} is low and the active leakage component is dominant in total power consumption. The reduction in leakage current has to be achieved using both process and circuit level techniques. Low supply voltage requires the device threshold to be reduced in order to maintain performance. Due to the exponential relationship between leakage current and threshold voltage in the weak inversion region, Leakage power can no longer be ignored. In this paper we present a technique to accurately estimate leakage power by accurately modeling the leakage current in transistor stacks.

REFERENCES

- [1] V. De and S. Borkar, "Technology and Design Challenges for Low Power and High Performance," in Proceedings of International Symposium on Low Power Electronics and Design, pp. 163–168, August 1999.
- [2] 2001 International Technology Roadmap for Semiconductors, <http://public.itrs.net/>
- [3] Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices, New York, USA: Cambridge University Press, 1998, ch. 2, pp. 94-95.
- [4] Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices, New York, USA: Cambridge University Press, 1998, ch. 3, pp. 120-128.
- [5] A. Keshavarzi, K. Roy, and C. F. Hawkins, "Intrinsic Leakage In Low Power Deep Submicron CMOS Ics," in Proceedings of International Test Conference, pp. 146–155, 1997.



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

- [6] Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices, New York, USA: Cambridge University Press, 1998, ch.2, pp. 97-99.
- [7] K. Roy and S. C. Prasad, Low-Power CMOS VLSI Circuit Design, New York, USA: Wiley Interscience Publications, 2000, ch. 2, pp. 28-29.
- [8] K. Roy and S. C. Prasad, Low-Power CMOS VLSI Circuit Design, New York, USA: Wiley Interscience Publications, 2000, ch. 2, pp. 27-28.
- [9] S. Thompson, et. al., "MOS Scaling: Transistor Challenges for the 21st Century," Intel Technology Journal Q3'98.
- [10] Z. Chen, et. al., "Estimation of Standby Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks," in Proceedings of International Symposium on Low Power Electronics and Design, pp. 239-244, 1998.
- [11] M. C. Johnson, D. Somasekhar, and K. Roy, "Leakage Control with Efficient Use of Transistor Stacks in Single Threshold CMOS", Design Automation Conf., pp.442-445, 1999.
- [12] N. Sirisantana, L. Wei, and K. Roy, "High-Performance Low- Power CMOS Circuits Using Multiple Channel Length And Multiple Oxide Thickness," in Proceedings of International Conference on Computer Design, pp. 227 –232, 2000.