Design and Development of Minimum-Leakage Standard Cell Library using MTCMOS Technique for Full Custom Power Aware Designs

Dr. T.Esther Rani

Professor of ECE, CVR College of Engineering
Hyderabad, India

ABSTRACT— The main objective of this work is to design a Low Power Standard Cell Library that contains a collection of components standardized at the logic or functional level. The technology is adopted for a channel length of 90nm and the supply voltage of 1V, for efficiency. The library is successfully designed with standard cells consisting of all logic gates with low leakage power incorporating the MTCMOS technique using the Cadence Tools, satisfying the design rules. The power dissipation is calculated for different cells of the library designed for minimum area, high speed and minimum leakage with different loads. The cells are designed to handle different loads by varying the size of the sleep transistor. After post layout simulation it is observed that there is a reduction in power dissipation up to 50% using the MTCMOS technique.

Keywords— Low Power, Standard Cell Library, MTCMOS technique, sleep transistor, minimum leakage.

I. Introduction

The major concerns of the VLSI designer were area, performance, cost, reliability and the power dissipation was generally of only secondary importance. In recent years this has instigated to change for several factors have contributed to this trend. The primary influence has been the remarkable success and growth of the class of personal computing devices and wireless communications systems, which demand high-speed computation and complex functionality with low power consumption. Also the continuing decrease in feature size and corresponding increase in chip density and operating frequency made power consumption a major concern in VLSI design. Excessive power dissipation in integrated circuits discourages their use in portable systems. It also causes overheating, which degrades the performance and reduces chip lifetime. To control the on-chip temperature, the chips need specialized and costly packaging and cooling arrangements, which would result in escalation of cost. Overall, low power design is a critical expertise needed in semiconductor industry today.

II. Power dissipation in Integrated Circuits

The total power dissipation in a circuit conventionally consists of two components, namely, the static and dynamic power dissipation. Many circuit techniques have been proposed to reduce these components in VLSI circuit design. Some of these techniques are suitable for reducing the static components while some techniques are efficient to reduce the dynamic power dissipation of the circuit.

A. Dynamic power

For dynamic power dissipation there are two components one is switching power due to charging and discharging of load capacitance. The other is the short circuit power due to the nonzero rise and fall time of input waveforms. The switching power of a single gate can be expressed as

$$P_{\mathrm{D}} = \alpha C_{\mathrm{L}} V_{\mathrm{DD}}^2 f \qquad (1)$$

where α is the switching activity, f the operation frequency, C_L the load capacitance and V_{DD} the supply voltage. The short circuit power of an unloaded inverter can be approximately given by

$$P_{\mathrm{SC}} = \frac{\beta}{12}(V_{\mathrm{DD}} - V_{\mathrm{th}})^3 \frac{\tau}{T} \qquad (2)$$

where β is the transistor coefficient, t the rise/fall time and T (1/f) the delay.

B. Static power

There are three dominant components of leakage in a MOSFET in the nanometer regime.

(i) Subthreshold leakage, which is the leakage current from drain to source I_{sub} (ii) direct tunneling gate leakage, which is due to the tunneling of electron (or hole) from the bulk silicon through the gate oxide potential barrier into the gate, and (iii) the source/substrate and drain/substrate reverse-biased p–n junction BTBT leakage. This leakage component is expected to be large for sub-50nm devices [1]. Other components of leakage current such as GIDL, impact ionization, etc. are not expected to be large for regular CMOS operations [2].

1. Subthreshold leakage

Subthreshold or weak inversion conduction current between source and drain in a MOS transistor occurs when gate voltage is below V_{th} [3]. Weak inversion typically dominates modern device off-state leakage due to the low V_{th} that is used. The weak inversion current can be expressed based on the following equation.

$$I_{subth} = Ae^{(q/nkT)(V_{GS}-V_{TH0}-\gamma' V_{SB}+\eta V_{DS})}(1-e^{(-qV_{DS}/kT)}) \quad (3)$$

$$A = \mu_0 C'_{ox} \frac{W}{L_{eff}} \left(\frac{kT}{q}\right)^2 e^{1.8} \quad (4)$$

where V_{GS}, V_{DS}, and V_{SB} are the gate voltage, drain voltage and body voltage of the transistor, respectively. Body effect is represented by the term γV_{SB}, where γ is the linearized body effect coefficient. η is the DIBL coefficient, representing the effect of V_{DS} on threshold voltage. C_{ox} is the gate oxide capacitance. μ_0 is the zero bias mobility and n is the sub threshold swing coefficient of the transistor. This equation shows the exponential dependency of sub threshold leakage on V_{th0}, V_{GS}, V_{DS} (due to DIBL), and V_{SB}. Each of the leakage reduction techniques described in the latter sections utilized these parameters in a MOSFET to achieve a low leakage state.

2. Gate leakage

Gate direct tunneling current is due to the tunneling of electron (or hole) from the bulk silicon through the gate oxide potential barrier into the gate. The direct tunneling is modeled as

$$J_{DT} = A(V_{ox}/T_{ox})^2 \exp\left(\frac{-B(1-(1-V_{ox}/\phi_{ox})^{3/2})}{V_{ox}/T_{ox}}\right) \quad (5)$$

where J_{DT} is the direct tunneling current density, V_{ox} is the potential drop across the thin oxide, ϕ_{ox} is the barrier height of tunneling electron and t_{ox} is the oxide thickness [4]. The tunneling current increases exponentially with decrease in oxide thickness. It also depends on the device structure and the bias condition [5].

3. Source/substrate and drain/substrate PN junction leakage

Drain and source to well junctions are typically reverse-biased causing pn junction leakage current. A reverse bias p-n junction leakage has two main components: One is minority carrier diffusion/drift near the edge of the depletion region and the other is due to electron–hole pair generation in the depletion region of the reverse-biased junction. Figure 1 shows gate leakage and junction currents. In the presence of a high electric field (4106 V/cm) electrons will tunnel across a reverse-biased p–n junction. A significant current can arise as electrons tunnel from the valence band of the p-region to the

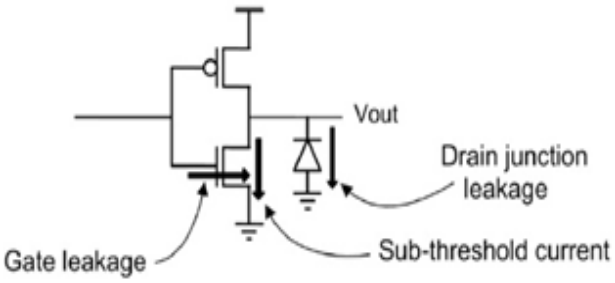

Figure 1: Leakage currents

conduction band of the n-region [3]. Tunneling occurs when the total voltage drop across the junction is greater than the semiconductor band-gap. Since silicon is an indirect band-gap semiconductor the BTBT current in silicon involves the emission or absorption of phonons.

III. Leakage Optimization Techniques

Many circuit techniques targeting the reduction of leakage power have appeared in the literature. Most of these techniques target the circuits during the standby mode and some target the circuits during the active mode of operation. Also, some of these techniques mainly aim to reduce sub-threshold leakage current while others tend to reduce gate leakage current. The leakage reduction circuit techniques can be categorized into the following classes.

i. Transistor stacking techniques

All these techniques are based on the fact that when there are two or more stacked transistors which are switched OFF. Transistor stacking has a dual effect in reducing the sub-threshold leakage current. It increases the source bias of upper transistors in the stack and also lowers the gate-to-source voltages of these transistors.

ii. Multi-Vth techniques

This is one of the most common approaches to reduce leakage currents where two different types of transistors are fabricated on the chip, a high V_{th} to lower sub-threshold leakage current and a low V_{th} to enhance circuit performance by increasing its speed.

Multi-threshold voltage CMOS: The leakage is reduced by inserting high-threshold devices in series to low V_{th} circuitry. A sleep control scheme is introduced for efficient power management. In the active mode, sleep is set low and sleep control high V_{th} transistors are turned on and vice versa.

Super Cutoff CMOS (SCCMOS) technique: MTCMOS can only reduce the standby leakage power, and the large inserted MOSFETs can increase the area and delay. Instead of using high V_{th} sleep control transistors as MTCMOS, Super Cutoff CMOS (SCCMOS) technique uses low V_{th} transistors with an inserted gate bias generator.

Clustered MTCMOS/SSCMOS technique: Clustered sleep transistor insertion implies addressing a number of issues, including the granularity of the insertion. For large CMOS blocks, the size of the sleep transistors and the driving strengths of sleep signals may become large and for small CMOS blocks the number of sleep transistors and the size of the control logic may become large.

Smart Series Switch (Triple-S) technique: In this technique two parallel switches are connected in series with a leaky device. A low V_{th} transistor switch as a function of the operation mode (active/standby) and a high V_{th} transistor switch as a function of the state of the leaky device. The Triple-S technique suffers from an area overhead ranging from 20% to 40%.

Dual-threshold voltage CMOS: Dual-threshold CMOS (DTCMOS) has the same critical delay as the single low CMOS circuit, but the transistors in non-critical paths can be assigned high to reduce leakage power. Dual threshold technique is good for leakage power reduction during both standby and active modes without delay and area overhead.

Dynamic threshold CMOS: For dynamic threshold CMOS (DTMOS), the threshold voltage is altered dynamically to suit the operating state of the circuit. A high threshold voltage in the standby mode gives low leakage current, while a low threshold voltage allows for higher current drives in the active mode of operation.
Double-gate dynamic threshold SOI CMOS: The double-gate dynamic threshold voltage (DGDT) SOI MOSFET combines the advantages of DTMOS and double-gate FD SOI MOSFETs without any limitation on the supply voltage.

iii. Dynamic V_{th} techniques

Dynamic threshold voltage scaling is a technique for active leakage power reduction. This scheme utilizes dynamic adjustment of frequency through back-gate bias control depending on the workload of a system. When the workload decreases, less power is consumed by increasing V_{th}. Two varieties of dynamic V_{th} scaling (DVTS) have been proposed.
V_{th}-hopping scheme: Using the control signal, which is obtained from software, the power control block generates select signals, V_{th}–low-Enable and V_{th}–high-Enable, which in turn control the substrate bias for the circuit. When the controller asserts V_{th}–low-Enable, V_{th} in the target processor reduces to V_{th}–low.
Dynamic Vth-scaling scheme: A clock speed scheduler, embedded in the operating system, determines the (reference) clock frequency at run-time. The DVTS controller adjusts the PMOS and NMOS body bias so that the oscillator frequency of the voltage-controlled oscillator tracks the given reference clock frequency. The error signal, which is the difference between the reference clock frequency and the oscillator frequency, is fed into the feedback controller.

iv. Supply voltage scaling techniques

There are three key components for implementing dynamic voltage scaling (DVS) in a general-purpose microprocessor. An operating system that can intelligently determine the processor speed, a regulation loop that can generate the minimum voltage required for the desired speed, and a microprocessor that can operate over a wide voltage range.

v. Input vector control techniques:

The basic idea behind input vector control techniques for leakage reduction is to force the combinational logic of the circuit into a low-leakage state during standby periods. Determining the best input vector that needs to be fed into the circuit depends on circuit structure, complexity, and size.

vi. Body biasing techniques

Adaptive body biasing is an effective way of reducing active leakage, as well as standby leakage through its effect in increasing the threshold voltages of MOS transistors. In this technique a high reverse body bias is applied so as to increase the threshold voltage of transistors to reduce sub-threshold leakage currents. Body biasing is also effective in reducing the negative effect of DIBL and V_{th}-Roll off to further reduction of leakage currents and improve circuit performance. One example of using the body bias technique is the Variable Threshold CMOS (VTMOS) technique, where a deep reverse body bias is applied during standby [6,7,8].

IV. MTCMOS

MTCMOS method is shown in Figure 2. Low-V_{th} transistors are used to implement the logic block and high-V_{th} transistors are used to gate the logic block from the power supply. Here, virtual power supply rails are created instead of direct connection from the actual power rails. By employing high-V_{th} transistors in this manner, leakage power is greatly reduced. In active mode, the circuit functions normally and in standby mode high-V_{th} transistors called sleep transistors are switched off enabled by a sleep signal input. Sleep device can be

a PMOS transistor which is placed in between pull up network and V_{DD} or an NMOS transistor that is placed in between pull down network and ground [4].

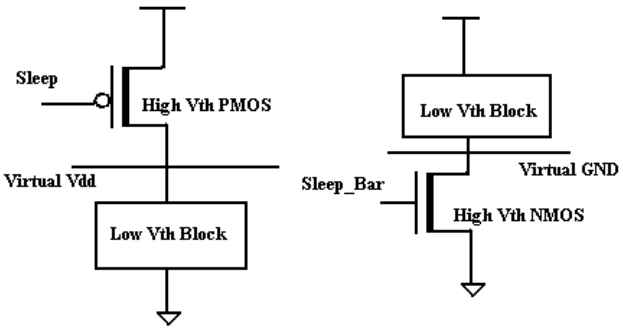

Figure 2: MTCMOS Method

MTCMOS reduces the leakage by inserting high threshold sleep control devices in series to low V_{th} as shown in Figure 3. In the active mode, the sleep control signal SL is set low and the control transistors are tuned on. Since the on resistances of high-V_{th} sleep is low, V_{DD} and V_{SS} act like power supply lines. In the standby mode, SL is set high, the high threshold sleep control transistors are tuned off, resulting in low leakage current.

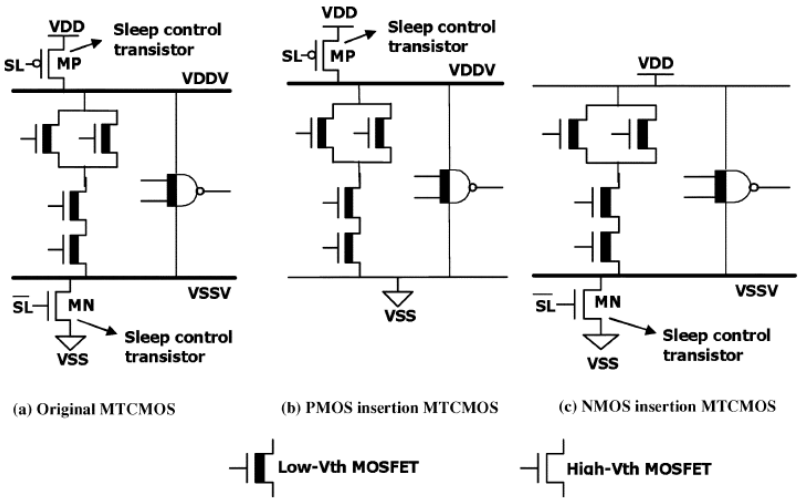

Figure 3: MTCMOS technique

V. Dimensioning of MTCMOS circuits

The MTMCOS technique is a well-known way to combine high switching speed with low standby current, by using low V_{th} transistors for the logic part and high V_{th} transistors for the sleep transistors. However, a practical analytic formula, how to correctly dimension the sleep transistor for a demanded performance, has not been provided.

The logic transistors, however will work in the saturation region. Since the current through logic and sleep transistor must be identical, the following equation describes the resulting" ground shift" V_H due to the sleep transistor.

$$V_H = \frac{q-1}{q} \cdot (V_0 - V_{thL}) . \quad (6)$$

The factor q (q>1) is used to describe the specified delay time factor of the MTCMOS circuit in comparison to a standard CMOS configuration. With the help of Equation (1) it is possible to calculate the

necessary width W_H of a sleep transistor, with W_L as the accumulated width of all low-V_{th} logic transistors that are controlled by the sleep transistor.

$$W_H = 0.625 \cdot \frac{W_L}{2} \cdot \frac{V_0 - V_{thL} - V_H}{(V_0 - V_{thL}) \cdot V_H - \frac{V_H^2}{2}} \cdot \quad (7)$$

The high threshold sleep transistor is controlled using the Sleep signal and limits the leakage current to a low value in the standby mode. The load dependent delay d_i of a gate i in the absence of a sleep transistor can be expressed as

$$d^i \propto \frac{C_L V_{dd}}{(V_{dd} - V_{tL})^\alpha} \quad (8)$$

where C_L is the load capacitance at the gate output, Vt and L is the low voltage threshold = 350 mV, V_{dd} = 1.8 V and α is the velocity saturation index (≈ 1.3 in 0.18-µm CMOS technology). In the presence of a sleep transistor, the propagation delay of a gate can be expressed as

$$d^i_{sleep} = \frac{K C_L V_{dd}}{(V_{dd} - 2V_x - V_{tL})^\alpha} \quad (9)$$

Where V_x is the potential of the virtual rails as shown in figure 1 and K is the proportionality constant. Let us suppose I_{sleep} ON is the current flowing in the gate during active mode of operation. During this mode, the sleep transistor is in the linear region of operation. Using the basic device equations for a transistor in linear region, the drain to source current in the sleep transistor (which is the same as I_{sleep} ON) is given by

$$\begin{aligned} I_{sleep_{ON}} &= \mu_n C_{ox} (W/L)_{sleep} ((V_{dd} - V_{tH}) V_x - \frac{V_x^2}{2}) \\ &\simeq \mu_n C_{ox} (W/L)_{sleep} (V_{dd} - V_{tH}) V_x \end{aligned} \quad (10)$$

The sub-threshold leakage current I_{leak} in the sleep mode will be determined by the sleep transistor and is expressed as given by

$$I_{leak} = \mu_n C_{ox} (W/L)_{sleep} e^{1.8} V_T^2 e^{\frac{V_{gs} - V_{th}}{n V_T}} (1 - e^{\frac{-V_{ds}}{V_T}}) \quad (11)$$

Where μ_n is the N-mobility, C_{ox} is the oxide capacitance, V_{th} is the high threshold voltage (= 500 mV), V_T is the thermal voltage = 26mV and n is the sub-threshold swing parameter.

Equation 2 establishes a relation between delay of a gate d_i sleep and V_x. By replacing V_x in equation 4 in terms of di_{sleep} (using equation 2), we get a dependence between (W/L) sleep and di_{sleep} (assuming the ON current is constant for each gate). Thus, a range of (W/L) sleep for the sleep transistor would correspond to a range of gate delays. Finally, (W/L) sleep in equation 5 can be replaced in terms of di_{sleep}, hence establishing a relationship between gate delay and gate leakage. The final relation between leakage and delay can be expressed as

This relationship exists for only those gates that have a sleep transistor assigned to them. Note that the moment a sleep transistor is assigned, some delay penalty is incurred. The range of delay that a gate can have is decided by the range of the acceptable (W/L) sleep. The objective of sleep transistor sizing is to decide the best values of (W/L) sleep for all sleep transistors such that the global delay constraint is satisfied and the total leakage is minimized [9].

$$I_{leak} = \mu_n C_{ox} e^{1.8} V_T^2 e^{\frac{V_{gs}-V_{th}}{nV_T}} (1 - e^{\frac{-V_{ds}}{V_T}}) \frac{I_{sleep_{ON}}}{\mu_n C_{ox}(V_{dd} - V_{tH})} \frac{{d_{sleep}}^{1/\alpha}}{(V_{dd} - V_{tL}){d_{sleep}}^{1/\alpha} - (KC_L V_{dd})^{1/\alpha}} \quad (12)$$

V. STANDARD CELL LIBRARY DEVELOPMENT

A standard cell is a group of transistor and interconnect structures that provides Boolean logic functions AND, OR, NOT, NOR, NAND, EXOR and EXNOR or a storage function (flip-flop or latch). The simplest cells are direct representations of the elemental NAND, NOR, and XOR Boolean function, although cells of much greater complexity are commonly used. The cell's Boolean logic function is called its logical view. Functional behavior is captured in the form of a truth table or Boolean equation for combinational logic, or a state transition table for sequential logic [10,11].

A standard cell library is a collection of low-level logic functions such as AND, OR, INVERT, flip-flops, latches, and buffers. These cells are realized as fixed-height, variable-width full-custom cells. The key aspect with these libraries is that they are of a fixed height, which enables them to be placed in rows, easing the process of automated digital layout. The cells are typically optimized full-custom layouts, which minimize delays and area[12-16]. The Standard cell library developed consists of three sets of standard cells like:

1. Basic standard Cells : NOT, NAND, NOR, OR, AND, XOR, XNOR.
2. Low power Standard cells using
   a) nMOS hvt insertion mechanism
   b) pMOS hvt insertion mechanism

All these cells are designed using the Cadence Virtuoso tools. The library also consists of layouts of the standard cells developed using the Cadence Virtuoso Layout Editor.

Full Custom Design

Full-custom design is a methodology for designing Integrated Circuits by specifying the layout of each individual transistor and the interconnections between them. It defines all the photolithographic layers of the device. Full-custom design is used for both ASIC design and for standard product design. Full-custom design potentially maximizes the performance of the chip, minimizes its area thereby reducing component cost.

Full custom integrates analog components and other pre-designed, thus fully verified components such as microprocessor cores that form a System-on-Chip (SoC) but is extremely labor-intensive to implement. For digital-only designs, "standard-cell" libraries together with modern CAD systems, can offer considerable performance/cost benefits with low risk. Automated layout tools are quick and easy to use and also offer the possibility to "hand-tweak" or manually optimize any performance-limiting aspect of the design [17].

VI. Results

The design of all the Low Power Standard Cells is done using the CADENCE ICFB 6.1 gpdk tool. The reduction in power dissipation from the basic cells to the low power cells designed using MTCMOS technique is 50%. Figure 7 shows the layout of NOR gate with the PMOS transistor as the sleep device. Figures 8,9 show the post layout simulation results of NOR with PMOS sleep transistor and NAND with NMOS sleep transistor. Figure 10 shows the simulation results of XNOR with PMOS sleep transistor. Figure 11shows the post layout simulation of NOR with NMOS as sleep transistor and Figure 12 shows the simulation Results of XOR with PMOS as sleep transistor Table 1 shows the comparison of power dissipation for different Logic gates.

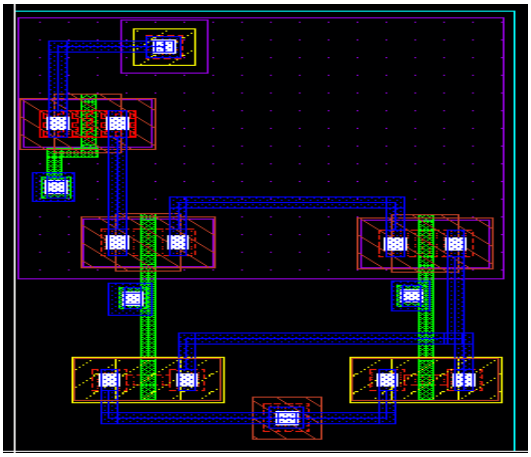

Figure 7: Layout of NOR with PMOS as sleep transistor

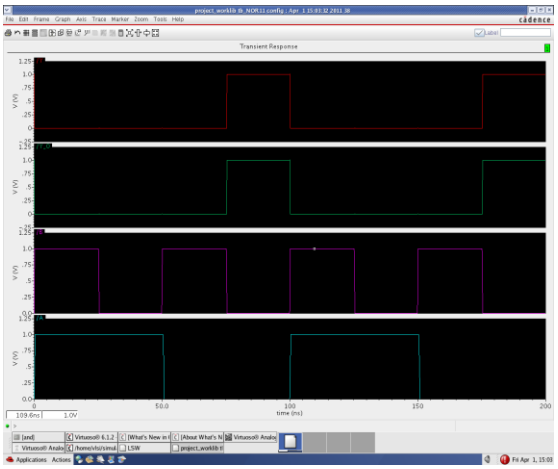

Figure 10: Simulation Results of XNOR with PMOS sleep transistor

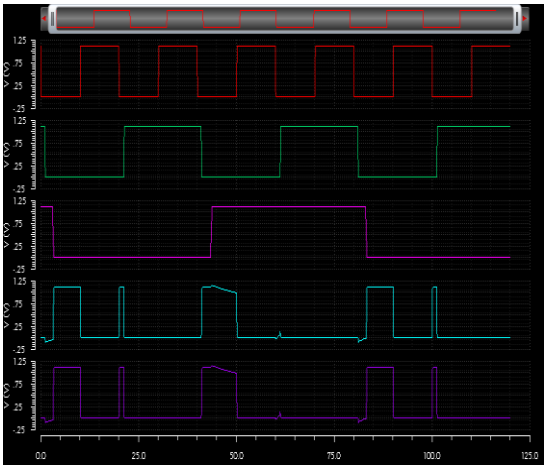

Figure 8: Post Layout simulation results PMOS sleep transistor

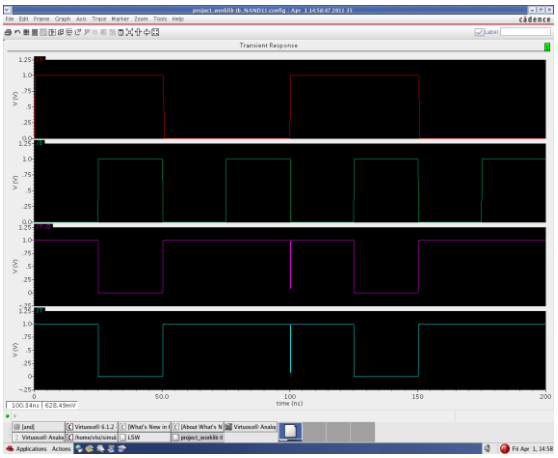

Figure 11: Post Layout Simulation of NOR with NMOS as sleep transistor

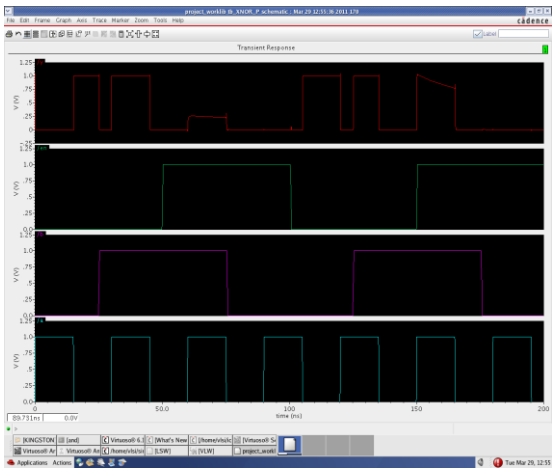

Figure 9: Post Layout Simulation of NAND with NMOS as sleep transistor

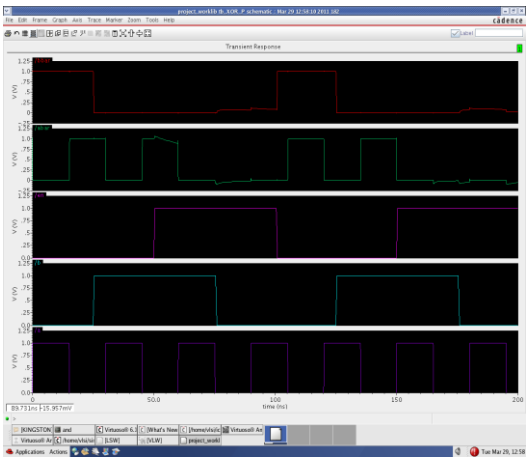

Figure 12: Simulation Results of XOR with PMOS as Sleep Transistor

Table 1: Comparison of Power Dissipation for Logic gates

Standard Cell	Basic	Using nMos high V_{th} cell	Using pMos high V_{th} cell
NOT	67.2 nW	31.9 nW	27.5 nW
NAND	76.7 nW	38.2 nW	35.2 nW
NOR	69.6 nW	73.5 nW	58.4 nW
OR	35.8 nW	44.2 nW	27.5 nW
AND	47.0 nW	22.0 nW	18.4 nW
XOR	252.5 nW	124.5 nW	112.6 nW
XNOR	303.5 nW	175.5 nW	163.5 nW

For a 16-bit register, the sleep device with different sizes is verified and the values of power while the clock is in active mode and the stand-by mode and are shown in Table 2. As the sleep device size is increased, the power and delay are reduced to more than 35% and 44% respectively. Sizing of the sleep device is done based on delay budgeting.

Table 2: Power and delay values with different sizes of sleep device to handle different loads

Size of sleep device	Active power (µm)	Standby power (nm)	Delay (pS)
520 nm	29.49	77	69
960 nm	30.53	77	43
2 µm	19.75	77	40
4 µm	19.57	77	39

VII. Conclusion

The Low Power Standard cell Library is successfully developed using the MTCMOS technique. The layouts for all the basic cells are developed and obtained the post layout simulation results for the standard cells. The power dissipation for the basic standard cells is calculated and compared with the power dissipation of MTCMOS standard cells. It is observed that there is a significant reduction in power dissipation. The power dissipation in pMOS insertion scheme is less than that of nMOS insertion scheme. But, due to pMOS transistor size, at fine grain level nMOS may be preferred. This work can be further advanced with extended designs using the Standard Cell Library and providing a complete library for developing any System on Chip (SoC).

ACKNOWLEDGMENT

The author would like to thank the UGC for funding and the management of CVR College of Engineering for providing Cadence tools and laboratory facilities.

REFERENCES

[1] S. Mukhopadhyay, K. Roy, Accurate modeling of transistor stacks to effectively reduce total standby leakage in nano-scale CMOS circuits, in: Symposium of VLSI Circuits, 2003.
[2] K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, Leakage current mechanisms and leakage reduction techniques in deep-submicron CMOS circuits, Proc. IEEE (2003).
[3] Y. Taur, T.H. Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, New York, 1998.
[4] K. Schuegraf, C. Hu, Hole injection SiO2 breakdown model for very low voltage lifetime extrapolation, IEEE Trans. Electron Devices 41 (1994) 761.
[5] C. Choi, K. Nam, Z. Yu, R.W. Dutton, Impact of gate direct tunneling current on circuit performance: a simulation study, IEEE Trans. Electron Devices 48 (2001) 2823.
[6] K. Roy, S.C. Prasad, Low-Power CMOS VLSI Circuit Design, Wiley Interscience Publications, New York, 2000.
[7] Bipul C. Paul, Amit Agarwal, Kaushik Roy ,"Low-power design techniques for scaled technologies," INTEGRATION, the VLSI journal 39 (2006)64–89.
[8] Kaushik Roy, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits" Proceedings of the IEEE 03/2003; 91(2):305 – 327
[9] Magdy A. Bayoumi. "Self-sleep Circuit for Distributed MTCMOS Design", 21st International Conference on VLSI Design (VLSID 2008), 01/2008
[10] Siri Uppalapati, "Low Power Design of Standard Cell Digital VLSI circuits"
[11] Mano, M. Morris and Charles R. Kime. Logic and Computer Design Fundamentals, Third Edition. Prentice Hall, 2004. p. 73
[12] Boole, George (2003) [1854]. An Investigation of the Laws of Thought. Prometheus Books. ISBN 978-1-59102-089-9.
[13] "The name Boolean algebra (or Boolean 'algebras') for the calculus originated by Boole, extended by Schröder, and perfected by Whitehead seems to have been first suggested by Sheffer, in 1913.
[14]" E. V. Huntington, "New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell's Principia mathematica", in Trans. Amer. Math. Soc.**35** (1933), 274-304; footnote, page 278.
[15] Givant, Steven; Halmos, Paul (2009). Introduction to Boolean Algebras. Undergraduate Texts in Mathematics, Springer. ISBN 978-0-387-40293-2.
[16] J. Michael Dunn; Gary M. Hardegree (2001). Algebraic methods in philosophical logic. Oxford University Press US. p. 2. ISBN 978-0-19-853192-0
[17] Narayan, K. Lalit (2008). Computer Aided Design and Manufacturing. New Delhi: Prentice Hall of India. p. 3. ISBN 812033342X.