INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH, DINDIGUL
Volume 1, No 3, 2010

RESEARCH ARTICLE ISSN - 0976-4259

Low Power CMOS Inverter design at different Technologies

Vijay Kumar Sharma[1], Surender Soni[2]
1- Department of Electronics & Communication, College of Engineering, Teerthanker Mahaveer University, Moradabad (U.P.)
2- Department of Electronics & Communication, National Institute of Technology, Hamirpur
vijay.buland@gmail.com

ABSTRACT

The increasing prominence of portable systems and the need to limit power consumption and hence, power dissipation in very high density VLSI chips have led to rapid and innovative developments in low power design recently. Leakage control is becoming critically important for deep sub-100nm technologies due to the scaling down of threshold voltage and gate oxide thickness of transistors. Reports indicate that 40% or even higher percentage of the total power consumption is due to the leakage of transistors. This percentage will increase with technology scaling unless effective techniques are introduced to bring leakage under control. This article focuses on circuit optimization and design automation techniques to accomplish this goal.

Keywords: Low Power, CMOS inverter, Stacking, Leakage Power

1. Introduction

With the rapid progress in semiconductor technology, chip density and operation frequency have increased, making the power consumption in battery-operated portable devices a major concern. High power consumption reduces the battery service life. The goal of low-power design for battery-powered devices is thus to extend the battery service life while meeting performance requirements. Reducing power dissipation is a design goal even for non-portable devices since excessive power dissipation results in increased packaging and cooling costs as well as potential reliability problems.

IC power dissipation consists of different components depending on the circuit operating mode. First, the switching or dynamic power component dominates during the active mode of operation. Second, there are two primary leakage sources, the active component and the standby leakage component. The standby leakage may be made significantly smaller than the active leakage by changing the body bias conditions or by power-gating. Voltage scaling is perhaps the most effective method of saving power due to the square law dependency of digital circuit active power on the supply voltage. Regrettably, scaling V_{DD} also reduces the circuit speed since the gate drive, $V_{GS} - V_{th}$, is reduced. To deal with this, systems may exploit dynamic voltage scaling to allow the lowest V_{DD} necessary to meet the circuit speed requirements while saving the energy used for the computation.

The current trend of lowering the supply voltage with each new technology generation has helped reduce the dynamic power consumption of CMOS logic gates. Supply voltage scaling increases the gate delays unless the threshold voltage of the transistors is also scaled down. The unfortunate effect of decreasing the threshold voltage is a significant increase in the leakage current of the transistors. Therefore, there is a clear tradeoff between the off-state

RESEARCH ARTICLE ISSN - 0976-4259

leakage and the active power for a given application, leading to methodical selection of V_{th} and V_{DD} for performing a fixed task.

1.1 Sources of leakage power

There are four main sources of leakage current in a CMOS transistor (Figure 1):
1. Reverse-biased junction leakage current (I_{REV})
2. Gate induced drain leakage (I_{GIDL})
3. Gate direct-tunneling leakage (I_G)
4. Sub threshold (weak inversion) leakage (I_{SUB})

Figure 1: Leakage current components in an NMOS transistor

2.1 Junction Leakage

The junction leakage occurs from the source or drain to the substrate through the reverse biased diodes when a transistor is OFF. A reverse-biased p-n junction leakage has two main components: one is minority carrier diffusion/drift near the edge of the depletion region; the other is due to electron-hole pair generation in the depletion region of the reverse-biased junction [7]. For instance, in the case of an inverter with low input voltage, the NMOS is OFF, the PMOS is ON, and the output voltage is high. Subsequently, the drain-to-substrate voltage of the OFF NMOS transistor is equal to the supply voltage.

This results in a leakage current from the drain to the substrate through the reverse-biased diode. The magnitude of the diode's 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 n and p regions are heavily doped, band-to-band tunneling (BTBT) dominates the pn junction leakage. Junction reverse-bias leakage components from both the source-drain diodes and the well diodes are generally negligible with respect to the other three leakage components.

RESEARCH ARTICLE ISSN - 0976-4259

2.2 Gate-Induced Drain Leakage

The gate induced drain leakage (GIDL) is caused by high field effect in the drain junction of MOS transistors. For an NMOS transistor with grounded gate and drain potential at V_{DD}, significant band bending in the drain allows electron-hole pair generation through avalanche multiplication and 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 worse by high drain to body voltage and high drain to gate voltage. Transistor scaling has led to increasingly steep halo implants, where the substrate doping at the junction interfaces is increased, while the channel doping is low. This is done mainly to control punch-through and drain-induced barrier lowering while having a low impact on the carrier mobility in the channel. The resulting steep doping profile at the drain edge increases band to band tunneling currents there, particularly as V_{DB} is increased. Thinner oxide and higher supply voltage increase GIDL current. As an example, with a V_{DG}=3V and t_{ox} of 4nm, there is roughly a 10 fold increase in the GIDL current when V_{DB} is increased from 0.8V to 2.2V.

2.3 Gate Direct Tunneling Leakage

The gate leakage flows from the gate through the “leaky” oxide insulation to the substrate. In oxide layers thicker than 3–4 nm, this kind of current results from the Fowler-Nordheim tunneling of electrons into the conduction band of the oxide layer under a high applied electric field across the oxide layer. For lower oxide thicknesses (which are typically found in 0.15µm and lower technology nodes), however, direct tunneling through the silicon oxide layer is the leading effect. Mechanisms for direct tunneling include electron tunneling in the conduction band (ECB), electron tunneling in the valence band (EVB), and hole tunneling in the valence band (HVB), among which ECB is the dominant one.

The magnitude of the gate directs tunneling current increases exponentially with the gate oxide thickness t_{ox} and supply voltage V_{DD}. In fact, for relatively thin oxide thicknesses (in the order of 2-3 nm), at a V_{GS} of 1V, every 0.2nm reduction in t_{ox} causes a tenfold increase in I_G. Gate leakage increases with temperature at only about 2x/100^{o}C. Note that the gate leakage for a PMOS device is typically one order of magnitude smaller than that of an NMOS device with identical t_{ox} and V_{DD} when using SiO_2 as the gate dielectric.

As transistor length and supply voltage are scaled down, gate oxide thickness must also be reduced to maintain effective gate control over the channel region. Unfortunately this results in an exponential increase in the gate leakage due to direct tunneling of electrons through the gate oxide. An effective approach to overcome the gate leakage currents while maintaining excellent gate control is to replace the currently used silicon dioxide gate insulator with high-K dielectric material.

Use of the high-k dielectric will allow a less aggressive gate dielectric thickness reduction while maintaining the required gate over drive at low supply voltages. According to the 2003 International Technology Roadmap for Semiconductors (ITRS-03), high-K gate dielectric is required to control the direct tunneling current for low standby power devices in process technology nodes below 90 nm.

RESEARCH ARTICLE ISSN - 0976-4259

2.4 Sub threshold Leakage

The sub threshold leakage is the drain-source current of a transistor operating in the 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 the 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 the temperature, supply voltage, device size, and the process parameters out of which the threshold voltage (V_{th}) plays a dominant role.

In current CMOS technologies, the sub threshold leakage current, I_{SUB}, is much larger than the other leakage current components. This is mainly because of the relatively low V_{th} in modern CMOS devices. I_{SUB} is calculated by using the following formula:

$$Ids = \mu_{oCox} \frac{W}{1}(m-1)\,(Vt)^2\, e(Vgs-Vth)/mVt_{*}\,(1\text{-}e\text{-}Vds/Vt) \quad (1)$$

Where:

$$m = 1 + \frac{Cdm}{Cox} = 1 + \frac{3tox}{Wdm} \quad (2)$$

Where:

Vth = The threshold voltage
Vt = The thermal voltage
Cox = Gate oxide capacitance
μ_0 = Zero bias mobility
m = The sub threshold swing coefficient (also called body effect coefficient)
Wdm= The maximum depletion layer width
tox = The gate oxide thickness
Cdm = The capacitance of the depletion layer

where w and l denote the transistor width and length, μ denotes the carrier mobility, Vt =kT/q is the thermal voltage at temperature T. Cdm = Cdep + Cit denotes the summation of the depletion region capacitance and the interface trap capacitance both per unit area of the MOS gate. m is the slope shape factor.

where Cox denotes the gate input capacitance per unit area of the MOS gate. When a long channel transistor with V_{DS} larger than a few vth is in the OFF state (V_{GS}=0).

It is highly desirable to have a sub threshold swing as small as possible since this is the parameter that determines the amount of voltage swing necessary to switch a MOSFET from OFF to ON state. This is especially important for modern MOSFETs with supply voltages reaching sub-one volt region. To minimize sub threshold slope (S), the thinnest possible gate

RESEARCH ARTICLE ISSN - 0976-4259

oxide (since it increases Cox) and the lowest possible doping concentration in the channel (since it decreases Cdep) must be used. Higher temperature results in larger S value, and hence, an increase in the OFF leakage current.

In long channel devices, the influence of source and drain on the channel depletion layer is negligible. However, as channel lengths are reduced, overlapping source and drain depletion regions cause the depletion region under the inversion layer to increase. The wider depletion region is accompanied by a larger surface potential, which attracts more electrons to the channel. Therefore, a smaller amount of charge on the gate is needed to reach the onset of strong inversion and the threshold voltage decreases. This effect is worsened when there is a larger bias on the drain since the depletion region becomes even wider. More precisely, when a high drain voltage is applied to a short-channel device, it lowers the barrier for electrons between the source and the channel, resulting in further decrease of the threshold voltage. The source then injects carriers into the channel surface (independent of gate voltage), causing an increase in IOFF. This phenomenon, which can be thought of as a lowering of V_{th} as V_{DS} increases, is the DIBL effect. There is yet another phenomenon known as the "V_{th} Roll off" where by the V_{th} of a MOSFET decreases as the channel length is reduced. In such a case, the sub threshold swing parameter degrades and the impact of drain bias on V_{th} increases. Finally, there is the well-known "body effect," which causes an increase in V_{th} as the body of the transistor is reverse-biased (i.e., V_{SB} of an NMOS transistor is increased). Clearly, decreasing the threshold voltage increases the leakage current exponentially. In fact decreasing the threshold voltage by 100mV increases the leakage current by a factor of 10. Decreasing the length of transistors increases the leakage current as well. Therefore, in a chip, transistors that have smaller threshold voltage and/or length due to process variation contribute more to the overall leakage.

2. Stack Effect based Method

If the value of the input of a circuit in STANDBY mode is known, some NMOS and PMOS transistors can be added in series with gates to increase the stack effect and reduce the leakage current as a result. In Figure 2, the output of gate (a) is high when it is in the STANDBY mode. This means that the pull down network is OFF. Therefore, putting a transistor in series with the pull down network and keeping it off in the STANDBY mode will not change the value of the output. However, it will increase there resistance between the supply and ground (Figure 2 (b)). Therefore, the leakage of the logic gate is reduced. Notice that if the output of the gate was low, then adding the transistor and turning it off would make the output of the gate float. This could potentially create a problem as short circuit current flows. This is an important consideration any time a logic cell that has been put to sleep is driving some other logic cells that are not in sleep. The authors report that an average of 65% reduction in the leakage can be achieved by using this method. Higher savings may be achieved if high threshold sleep transistors are used.

The leakage current of a logic gate is a strong function of its input values. The reason is that the input values affect the number of OFF transistors in the NMOS and PMOS networks of a logic gate.
The minimum leakage current of the gate corresponds to the case when both its inputs are zero. In this case, both NMOS transistors in the NMOS network are off, while both PMOS

RESEARCH ARTICLE ISSN - 0976-4259

transistors are on. The effective resistance between the supply and the ground is the resistance of two OFF NMOS transistors in series. This is the maximum possible resistance. If one of the inputs is zero and the other is one, the effective resistance will be the same as the resistance of one OFF NMOS transistor. This is clearly smaller than the previous case. If both inputs are one, both NMOS transistors will be on. On the other hand, the PMOS transistors will be off. The effective resistance in this case is the resistance of two OFF PMOS transistors in parallel. Clearly, this resistance is smaller than the other cases. Therefore, when in the STANDBY state, if, by some means, values of the internal signals are also controlled, even higher leakage savings can be achieved.

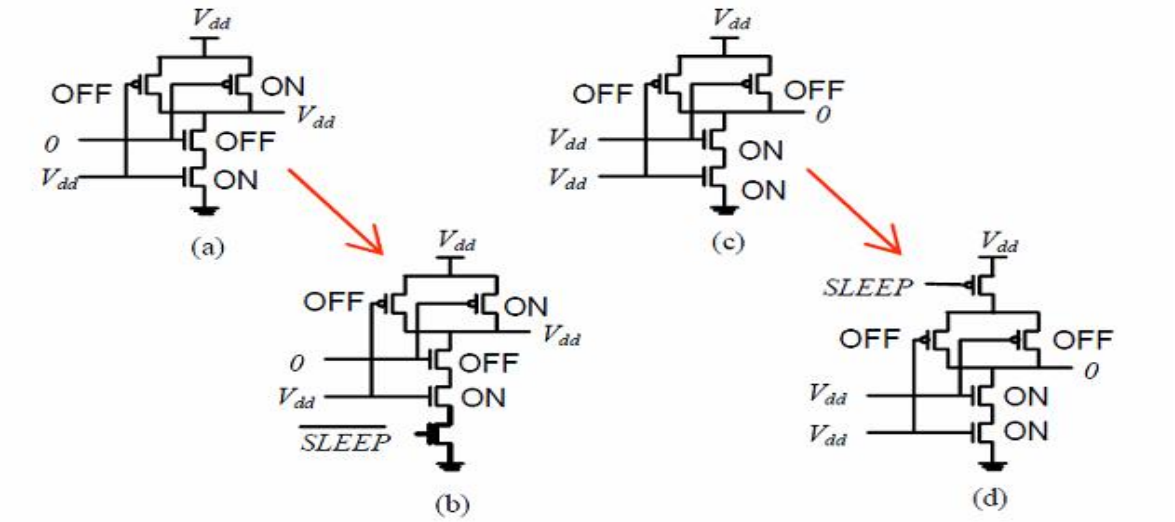

Figure 2 Reducing the leakage using stacking

3. Results and discussion

The implemented circuits are simulated to measure total leakage Power. Technique using 70 nm, 100 nm, 130, 180 nm technology has been simulated on TANNER tool. Table shows the simulation results of each circuit at different temperature. Using leakage monitoring and the balance between sub threshold leakage and BTBT (band to band tunneling) leakage, the new optimal body bias technique gives reduction in total leakage power. Since the optimal bias results in the minimum leakage current for nano-scale device in STANDBY mode, the power in active mode is also improved by applying the optimal body bias. The circuits for CMOS inverter are shown below and their corresponding total leakage powers vs. Technology, total leakage power vs. temperatures are listed in the table.

From the graphs we see that leakage power increased with temperature and when we move towards the sub 100 nm technologies. Subthreshold current is also the function of thermal voltage. Thermal voltage is related to the temperature proportionally. When temperature is increased then thermal voltage also increase and thus subthreshold leakage current.
In stacking of transistors we vary the threshold voltage of the transistors by providing bulk to source biasing negative. This increase the threshold voltage of the device, more threshold

RESEARCH ARTICLE ISSN - 0976-4259

voltage means less sub threshold current, this cause less total leakage power. The threshold voltage of a device is given as:

$$V_{th} = V_{fb} + 2\psi B + \frac{\sqrt{2\varepsilon siqN_a(2\psi B + V_{bs})}}{C_{ox}} \quad (3)$$

Where:

Vfb = The flat-band voltage
Na = The doping density in the substrate
ψB = The difference between the Fermi potential and the intrinsic potential in the substrate

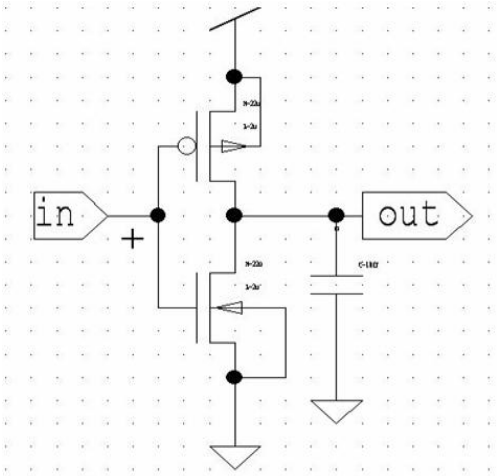

Figure 3: CMOS Inverter

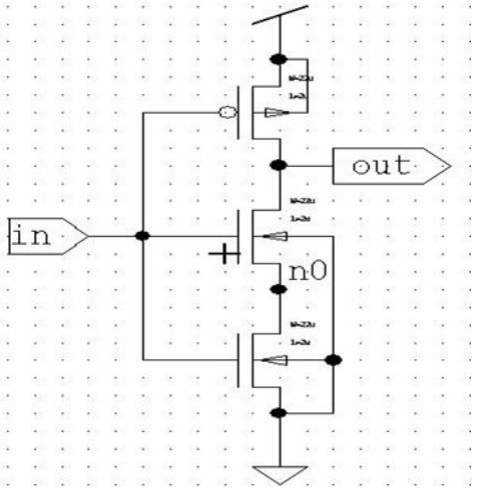

Figure 4: CMOS forced NMOS Inverter

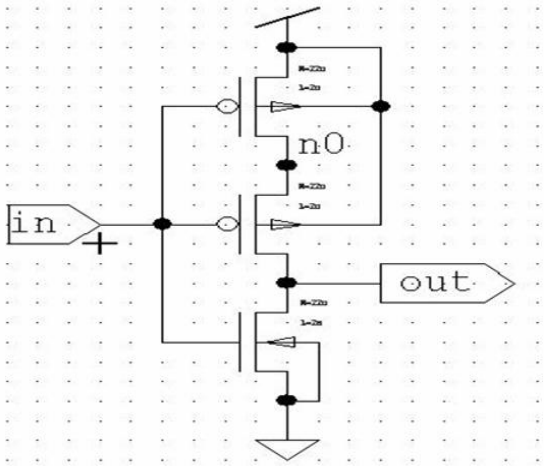

Figure 5: CMOS forced PMOS Inverter

RESEARCH ARTICLE ISSN - 0976-4259

TABLE 1:
TOOL:- TANNER 9.13, WHEN V_{IN} = 0V

Technique	Technology (nm)	Temp. (^{o}C)	Leakage Power (W)
Conventional CMOS Inverter	70	40	9.9013n
		50	14.3421n
		60	20.287n
	100	40	2.2773n
		50	3.4957n
		60	5.2177n
	130	40	470.097p
		50	766.947p
		60	1.2190n
	180	40	385.9924p
		50	677.776p
		60	1.1453n
Forced PMOS Inverter	70	40	6.1351n
		50	835.44p
		60	1.10p
	100	40	1.021n
		50	1.093p
		60	1.3678p
	130	40	445.21p
		50	766.94p
		60	1.213n
	180	40	289.73p
		50	677.77p
		60	1.149n
Forced NMOS Inverter	70	40	1.247n
		50	1.324p
		60	1.6218p
	100	40	153.203p
		50	1.625p
		60	2.002p
	130	40	117.037p
		50	198.74p
		60	324.84p
	180	40	25.83p
		50	2.744p
		60	3.383p

RESEARCH ARTICLE ISSN - 0976-4259

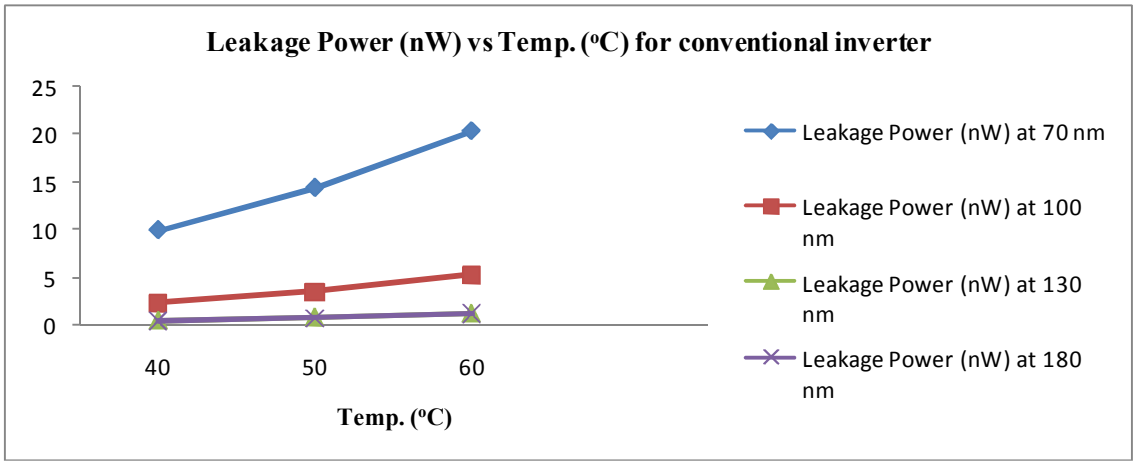

Figure 6: Leakage Power vs. Temperature at different technologies curve

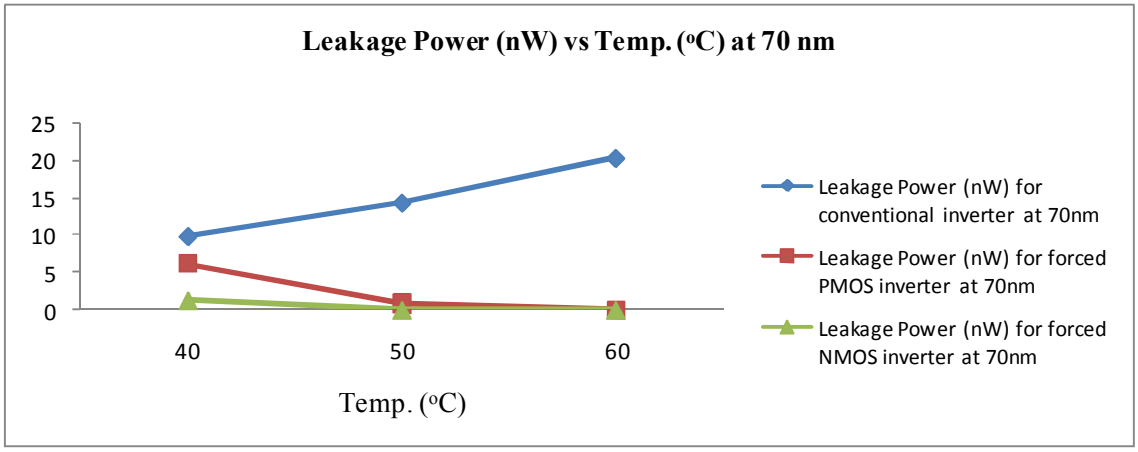

Figure 7: Leakage Power vs. Temperature at 70nm curve

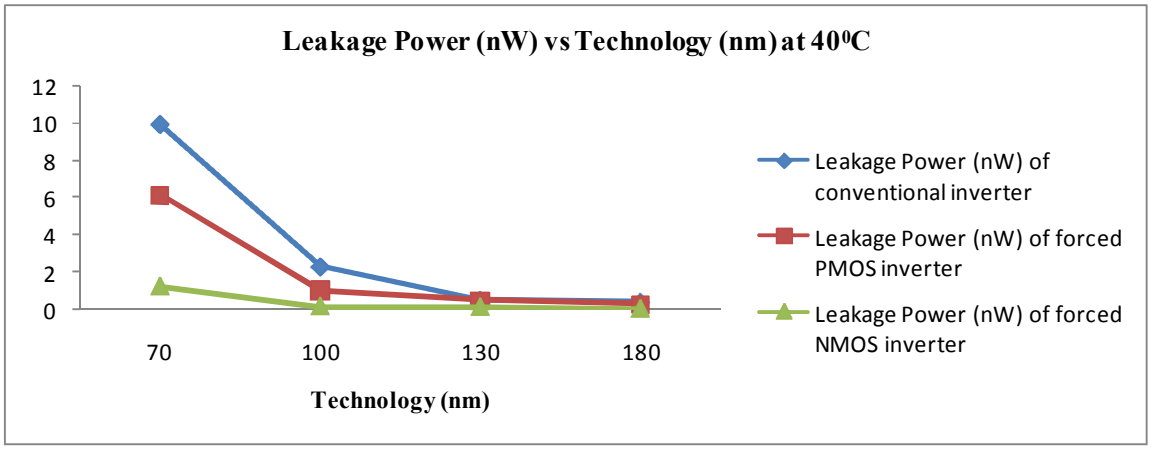

Figure 8: Leakage Power vs. Technology curve

RESEARCH ARTICLE ISSN - 0976-4259

4. Conclusion

As CMOS devices continue to scale, leakage becomes an even more important contributor to the total power consumption. In current technologies, sub threshold and gate leakage are the dominant sources of leakage and are expected to increase with technology scaling. In advanced devices, band-to-band tunneling **is** also likely to be a concern. To manage these leakage currents it will he necessary to consider leakage management at both the process technology and circuit levels. At the process technology level, well engineering techniques such as retrograde and halo doping are used to reduce leakage and improve short channel characteristics. At the circuit level, transistor stacking can effectively reduce the leakage current in high performance logic and memory designs.

References

1. A. Abdollahi, et al., “Minimizing leakage current in VLSI circuits,” Technical Report, Department of Electrical Engineering, University of Southern California, No. 02-08, May 2002.
2. B. Sheu, at al., "BSIM: Berkeley short-channel IGFET model for MOS transistors," IEEE Journal of Solid State Circuits, Vol. 22, August 1987, pp. 558-566. 20
3. D. Shin, et al., “Low-energy intra-task voltage scheduling using static timing analysis,” Proc. of Design Automation Conf., June 2001, pp 438-443.
4. K. Cao, et al., “BSIM4 gate leakage model including source drain partition,” in Tech. Dig. Int. Electron Devices Meeting, 2000, pp 815–818.
5. K. Roy, et al., “Leakage Control with Efficient use of Transistor Stacks in Single Threshold CMOS”, Proc. Design Automation Conference, June 1999.
6. L. Wei, et al., “Low voltage low power CMOS design techniques for deep submicron ICs,” Proc. of Thirteenth International Conference on VLSI Design, 2000, pp 24-29.
7. L.T. Clark, et al., “Managing standby and active mode leakage power in deep submicron design,” Proc. International Symposium on Low Power Electronics and Design,, August 2004.
8. P. van der Meer et al., “Ultra-low standby-currents for deep sub-micron VLSI CMOS circuits: smart series switch,” Proc. IEEE Int. Symp. Circuits and Systems, vol. 4, 2000, pp 1-4.
9. S. Borkar, “Design Challenges of Technology Scaling”, IEEE Micro, Vol. 19, Issue 4, Jul.-Aug. 1999, pp 345-365.
10. S. Lee et al., “Run-time power control scheme using software feedback loop for low-power real-time applications,” Proc. of Asia-Pacific Design Automation Conf., January 2000, pp 381-386.
11. Y. Taur, et al., “CMOS scaling into nanometer regime”, Proc. of the IEEE, vol. 85, Apr. 1997, pp 486–504.