ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

Power Scaling in CMOS Circuits by Dual-Threshold Voltage Technique

P.Sreenivasulu, P.khadar khan, Dr. K.Srinivasa Rao, Dr. A.Vinaya babu
[1]Research Scholar, ECE Department, JNTU Kakinada, A.P, INDIA.
[2] M.Tech(DECS) Student, Dr.S.G.I.T,Marakapur
[3]Principal 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*—. Reducing power dissipation has become an important objective in the design of digital circuits. One common technique for reducing power is to reduce the supply voltage. For CMOS circuits the cost of lower supply voltage is lower performance. Scaling the threshold voltage can limit this performance loss somewhat but results in increased static power dissipation. In modern digital integrated circuits, power consumption can be attributed to three main components: short circuit, leakage, and dynamic switching power *In fact, for modern submicron technologies, this simple analysis suggests optimal energy efficiency at supply voltages under 0.5 V. Other process and circuit parameters have almost no effect on this optimal operating point. If there is some uncertainty in the value of the threshold or supply voltage, however, the power advantage of this very low voltage operation diminishes. Therefore, unless active feedback is used to control the uncertainty, in the future the supply and threshold voltage will not decrease drastically, but rather will continue to scale down to maintain constant electric fields[4].

***Index Terms*— Threshold scaling, dual threshold voltage, low power, MTCMOS.**

I. INTRODUCTION

Using a first-order model of the energy and delay of a CMOS circuit, we show that lowering the supply and threshold voltage is generally advantageous, especially when the transistors are velocity saturated and the nodes have a high activity factor. In fact, for modern submicron technologies, this simple analysis suggests optimal energy efficiency at supply voltages under 0.5 V. Other process and circuit parameters have almost no effect on this optimal operating point. If there is some uncertainty in the value of the threshold or supply voltage, however, the power advantage of this very low voltage operation diminishes. There- fore, unless active feedback is used to control the uncertainty, in the future the supply and threshold voltage will not decrease drastically, but rather will continue to scale down to maintain constant electric fields. Scaling and power reduction trends in future technologies will cause sub threshold leakage currents to become an increasingly large component of total power dissipation. This paper presents several dual-threshold voltage techniques for reducing standby power dissipation while still maintaining high performance in static and dynamic combinational logic blocks. MTCMOS sleep transistor sizing issues are addressed, and a hierarchical sizing methodology based on mutual exclusive discharge patterns is presented. REDUCING power dissipation has become an important objective in the design of digital circuits. One common technique for reducing power is to reduce the supply voltage. For CMOS circuits the cost of lower supply voltage is lower performance. Scaling the threshold voltage can limit this performance loss somewhat but results in increased static power dissipation. Burr *et al.* [1], [2] have shown that if one optimizes for minimum energy, then operating in the sub threshold region is advantageous. Since minimum energy solutions are generally low performance solutions, we look instead at both energy and delay during optimization and use the energy-delay product as a measure of the efficiency of the circuit. In this paper we examine the effects of lowering the supply and threshold voltages on the energy efficiency of CMOS circuits. First-order model of the energy- delay product (EDP) of CMOS circuits. Using this model, one can find the optimal operating point, that is the value of supply and threshold voltage for which the EDP is minimum, as well as how this optimal point will change as circuit and process parameters change. For a modern 0.25- m technology the optimal operating point is a supply voltage of 250 mV and a Threshold voltage of 120 mV. The importance of operating near the minimum is set by how steep the curve, or surface, is near the minimum point. As I^{b} the curve becomes steeper the benefits of being near the optimal point increase. The performance cost of operating at this point is the ratio of the gate speed at this point to the original gate speed. We numerically solved the model described in Section II as a function of both V and V_{th} to determine the shape of the energy, delay, and EDP surfaces. We show that when transistors are velocity saturated, the EDP surface is pretty steep, and thus one wants to operate near the minima, but gates at this point are significantly slower than current operating conditions. In modern digital integrated circuits, power consumption can be attributed to three main components: short circuit, leakage, and dynamic switching power. Dynamic switching power is the dominant component of power consumption in modern integrated circuits, and results from the charging and discharging of gate capacitances during signal switching given by

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

$$P_{\text{switching}} = C_{\text{switched}} V_{\text{CC}}^2 f_{\text{clk}} \quad (1)$$

Where C_{switched} is the total effective switched capacitance, V_{CC} is the supply voltage, and f_{clk} is the switching frequency. However, as scaling trends continue in future generations and as low-power voltage scaling becomes more aggressive, sub- threshold leakage currents will become a larger, and potentially a dominant, component of overall power dissipation. Sub- threshold leakage currents vary exponentially with threshold voltage and are given by

$$I_{\text{leakage}} = \frac{W}{W_0} I_0 e^{(Vgs - V_t)/nV_{\text{th}}} = \frac{W}{W_0} I_0 10^{(V_{gs} - V_t)/S} \quad (2)$$

where V_{th} is the thermal voltage, is width, is a constant, and S ln 10 is the sub threshold slope. Thus, for a typ- ical technology with a sub threshold slope of 100 mV/decade, each 100-mV decrease in V_t will cause an order of magnitude increase in leakage currents.

$$T_{pd} \propto \frac{CV_{\text{CC}}}{(V_{\text{CC}} - V_t)^{\alpha}} \quad (3)$$

III. ENERGY AND DELAY IN CMOS CIRCUITS

The two main sources of power dissipation in CMOS circuits are static current, which results from resistive paths between power supply and ground, and dynamic power, which results from switching capacitive loads between different volt- age levels. There is a third source of power dissipation in CMOS circuits, short-circuit current, which results from both transistors in a CMOS inverter being on at the same time while the input switches. The short-circuit component is small [3], [13], therefore we ignore it throughout this paper. Static power is due to current sources and to leakage current when a transistor is nominally off.

For a CMOS gate, the dynamic power is

$$P = aCV^2 f \quad (1)$$

Where activity factor of the output node, C is the total capacitance of the output node, V is the supply voltage, and f is the operating frequency. If the circuit performs one operation per cycle, then the energy per operation is

$$E = aCV^2, \quad (2)$$

For a complex chip, the total dynamic power is simply the sum of the dynamic power of all the gates. The resulting equation has the same form as (1); the only difference is that C is now the total capacitance of all the loads, and the activity factor is the average activity factor. The leakage current for a gate can be written as

$$I_l = W I_s e^{(V_{\text{th}}/V_o)} \quad (3)$$

Where the effective transistor width1 of the cell is, I_s is the zero-threshold leakage current, V_{th} is the threshold voltage, and V_o is the sub threshold slope. We ignore the dependence of I_l on drain voltage, and also the leakage current in the reverse biased diodes. The leakage current for a complete chip is simply the sum of the leakage currents of all the gates. The total energy per operation of a chip thus can be written as

$$E = \sum_i a_i C_i V^2 + \sum_i W_i I_s e^{(V_{\text{th}}/V_o)} V T_c \quad (4)$$

Where T_c is the cycle time and is an index that runs over all gates in the circuit. The circuit dissipates static current throughout the cycle, but each gate dissipates dynamic energy for a short period of time while it switches. Notice that this equation is very similar to the energy consumed by a simple inverter (with the "correct" average activity and load C and assuming W is the total transistor width of the gate), so optimizing the energy of this average inverter will yield an optimal operating point for the chip. In fact, the optimal point remains unchanged if we further normalize this equation by the width of this average inverter, yielding the average energy consumed per micron of transistor width

$$E = C_{\text{eff}} V^2 + I_s e^{(V_{\text{th}}/V_o)} V T_c$$

Where average capacitance switched every cycle per micron of transistor width. This parameter is different for every design, depending on the types of circuits used. For the Strong ARM processor. Since caches—which have very low activity factors—occupy about 50% of the area of this chip, we expect other designs to have larger values of Later on we show the location of the optimal point is highly insensitive to the value of Leakage power is more important when the effective switched capacitance is small. Thus, we use a value of 1 fF, which is relatively high. This will make lower voltages seem more attractive.

$$T_g = K \frac{V}{(V - V_{\text{th}})^{\alpha}}$$

We use a similar technique to model the minimum operating cycle time, or critical path, of the chip. The critical path normally goes through a variety of gates, each with a different delay. Luckily, changes in supply voltage, temperature, and threshold voltage affect all gates in the same way so delay of any gate remains roughly proportional to the delay of an inverter, as is shown in Fig. 1. This figure shows the delay of different circuit elements normalized to the delay of an inverter. Solid lines show the delay at high temperature (125 C), dashed and doted lines show the delay at lower temperatures (25 C and 25 C, respectively). Thus, we can normalize the critical path by dividing the cycle-time by the delay of the average inverter described above. We call this quantity the logic depth since it represents how many inverters are in a ring oscillator which has the same frequency as the maximum operating frequency of the chip. For modern microprocessors the logic depth is usually around

30 equivalent inverters. The cycle time is then just

$$\text{EDP} = \frac{K^2 I_s L_d V^3}{(V - V_{\text{th}})^{\alpha}} \left(K_2 + \frac{e^{V_{\text{th}}/V_o)}}{(V - V_{\text{th}})^{\alpha}} \right) \quad (7)$$

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

To determine how the delay of an inverter varies with operating conditions we use a simple power model for MOS current.

Where is a proportionality constant specific to a given technology. The power accounts for the fact that the transistors may be velocity saturated. It can be anywhere between one, complete velocity saturation, and two, no velocity saturation. For a 0.25- m technology, is likely to be 1.3–1.5.

Combining (5) with (6), the energy-delay product can be written as

$$K_2 = \frac{C_{\text{eff}}}{I_s K L_d}. \tag{8}$$

Where K_2 is a constant for the given technology and is given by

$$V = \frac{3V_{\text{th}}}{3-\alpha} + \frac{3\alpha}{3-\alpha}V_o \tag{9}$$

To find the optimal supply and threshold voltage we differentiate (7) with respect to V and V_{th} and set the equations to zero. Solving for V and V_{th}, one gets

$$e^{-n} = \frac{K_2\left[\frac{\alpha}{3-\alpha}(n+3)V_o\right]^{\alpha}(3-\alpha)}{(n+2\alpha-3)}$$

$$n = -\alpha \ln\left(\frac{\alpha}{3-\alpha}(n+3)V_o\right) = \ln(K_2) + \ln(n+2\alpha-3)$$

IV. SIMULATION RESULTS

MTCMOS (multithreshold CMOS) is a dual- technology that is very effective at reducing leakage currents in the standby mode. This technique involves using high- transistors to gate the power supplies of a low- logic block as shown in Fig. 1. When the high- transistors are turned on, the low- logic is connected to virtual ground and power, and switching is performed through fast devices. However, by introducing an extra series device to the power supplies, MTCMOS circuits will incur a performance penalty compared to CMOS circuits, which worsens if the devices are not sized large enough. When the circuit enters the sleep mode, the high- gating transistors are turned off, resulting in a very low sub threshold leakage current from to ground [10] [11]. Although both pMOS and nMOS gating transistors are shown in Fig. 1, only one polarity sleep device is actually required to reduce leakage if the logic block is purely combinational.

NMOS sleep transistors are typically more effective because they have lower "on" resistances, and subsequently can be made smaller for the same current drive. MTCMOS circuits can achieve several orders of magnitude reduction in leakage currents through two effects. First, the total effective leakage width of the original CMOS circuit is reduced to the width of the single "off" nMOS transistor (provided it is smaller than the original pull down width), and second, the increased threshold voltage results in an exponential reduction in leakage currents. If the sleep transistor is turned off even more strongly (reversed bias), even further leakage reduction can be achieved.

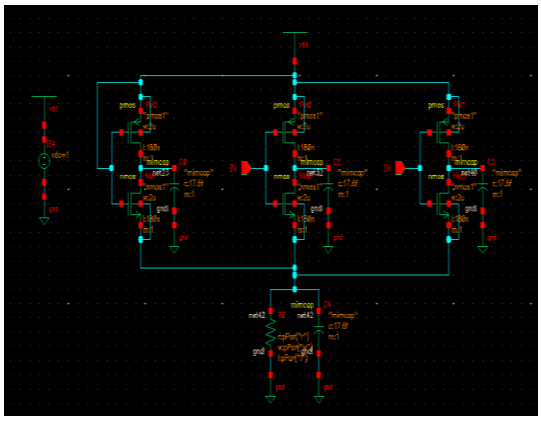

Fig. 1. MTCMOS block illustrating equivalent resistance, capacitance, and reverse conduction effects

The parasitic capacitance due to wiring and junction capacitances on the virtual ground line shown in Fig. 2 actually helps reduce the virtual ground line bounce by serving as a local charge sink or reservoir for current. However, having a large capacitance in itself does not offset the effects of a poorly sized sleep transistor. Since current is constantly switching through the sleep resistance of a complicated logic block, the parasitic capacitance would have to be prohibitively large to prevent an IR drop from developing over time. With a large time constant, it will also take longer for the virtual ground node to discharge back to ground if it does reach a large value. While capacitance on the virtual power does help reduce transient spikes in MTCMOS circuits, proper sleep transistor sizing is still of utmost importance.

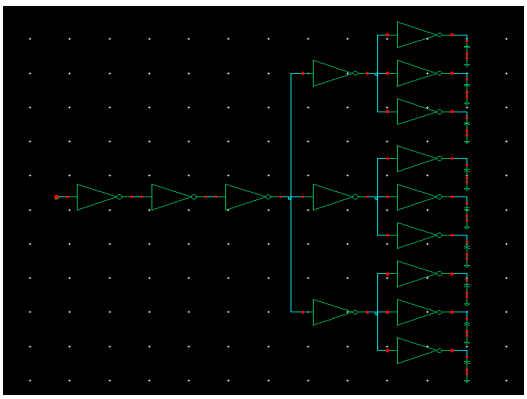

Fig. 2 MTCMOS inverter tree.

Once the MTCMOS circuit is sized with individual sleep transistors, one can then systematically merge the sleep transistors together because they can be shared among mutually exclusive gates, where no two gates can be discharging current at the same time. Finally, these sets of sleep transistors can then be combined to make a single sleep transistor for the whole circuit that

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

guarantees that for any input vector, the MTCMOS circuit performance will be within the specified range of the corresponding CMOS circuit.

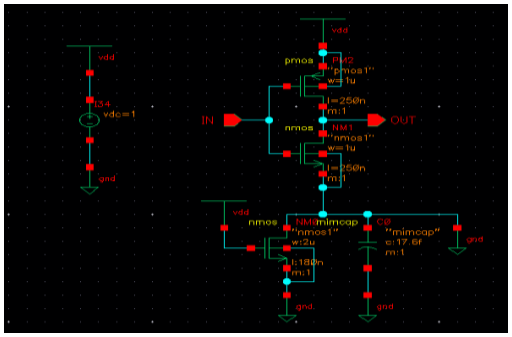

Fig. 3. Simulation circuit A

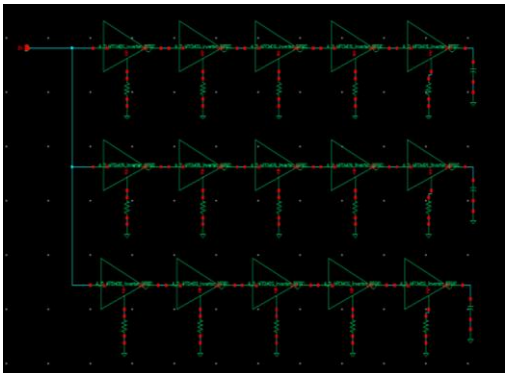

Fig. 4. . Inverter Chain Example Individual Sleep Resistors for Each Gate.

Fig. 4 shows a simple circuit consisting of three chains of five low- transistors and illustrates how individually sized sleep transistors can be combined into a common power switch for a larger block of logic. Fig. 5 shows the first step in the transistor sizing procedure, where individual sleep resistors (which model sleep transistors in the "on" state) are sized to ensure that no gate degrades by more than a fixed percentage. The overall degradation of the series degenerated gates will be less than the individual gate degradation because the low-to-high transitions of and are not degraded by the nMOS sleep transistor. Fig. 6 shows how the virtual ground lines and for this circuit will fluctuate as a result of a rising step function applied to the input.

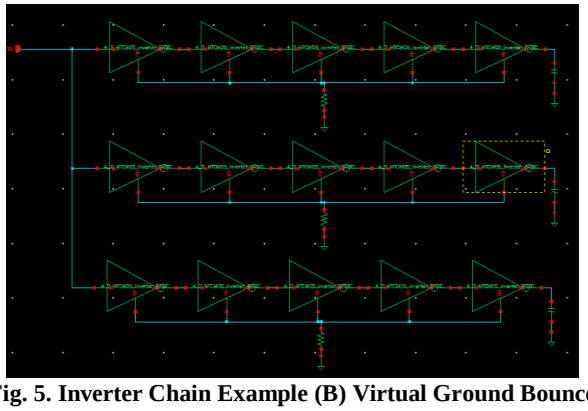

Fig. 5. Inverter Chain Example (B) Virtual Ground Bounce for (A) R = 2, 4, 6, 8, and 10K.

In Fig. 6, the three separate sleep resistors from Fig. 4 can be replaced by a single resistor with three times the conductance that now gates the entire circuit.

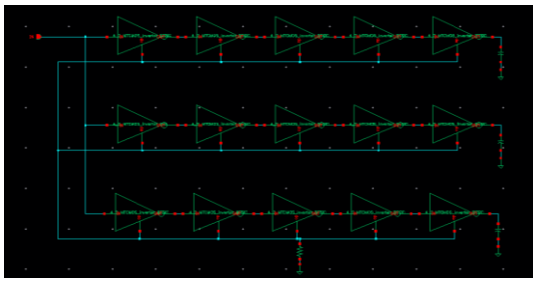

Fig. 6. Sleep Resistor Sharing For Mutual Exclusive Gates.

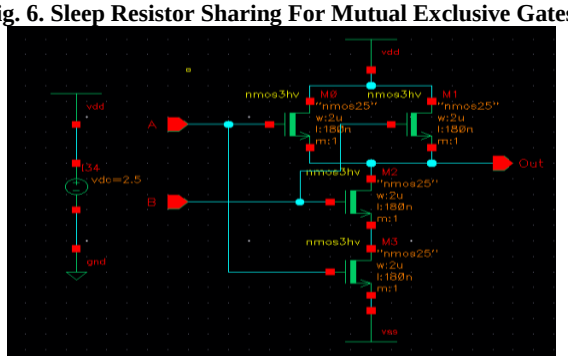

Fig. 7. Embedded neither Dual-V NOR Gates with Low-V

Fig. 5 and 6 shows comparisons of the delay versus sleep resistor size for these two cases and illustrates how the resistance must be lowered by one-third in order to achieve the same performance. Dual-threshold voltage domino provides the performance equivalent of a purely low- design with the standby leakage characteristic of a purely high- implementation [15]. Because of the fixed transition directions in domino logic, one can easily place the dual- domino gate into a low leakage state, and can imbed high- devices in noncritical transition directions without impacting performance. In effect, the dual-domino gate allows one to trade-off reduced precharge time for lower standby leakage currents. Dual- domino methodology utilizes low threshold voltages for all transistors that can switch during the evaluate mode and utilizes high threshold voltages for all transistors that can switch during the precharge modes. Fig. 10 shows a typical dual- domino stage, consisting of a pull down network, inverter, leaker device and clock drivers, with the low- devices shaded.

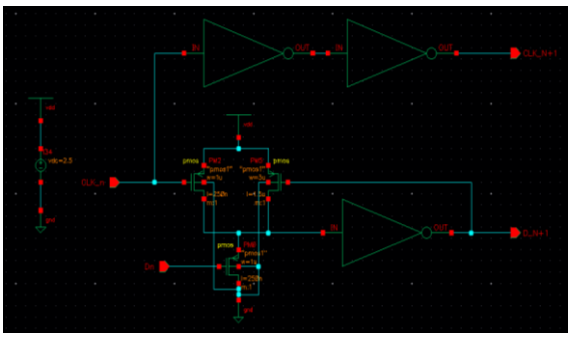

Fig. 8. Dual-V Domino Logic Gate with Low-V Devices Shaded

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

To verify the functionality and benefit of dual- domino logic, simulation were performed on a representative pipeline stage modeled as an inverter chain with four dynamic NOR gates and four accompanying static inverters in an aggressive 0.18um technology. The NOR gate has eight inputs, each driving a fanout of 3 load. These wide gates are a good representative of domino circuits, because domino technology is most effective for gates with wide, rather than deep, pull down networks. The experimental circuit has the exact same structure as shown in Figs. 7 and 8.

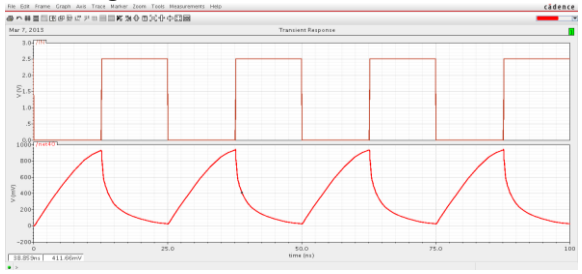

Fig. 9. Waveforms of Simulation circuit A.

In domino circuits, the noise margin is directly related to the threshold voltage of the nMOS pull down tree, so there is definitely a limit to how low be can scale. Furthermore, active leakage in large fan in gates, if large enough, can affect functionality when a domino gate tries to hold an internal node high.

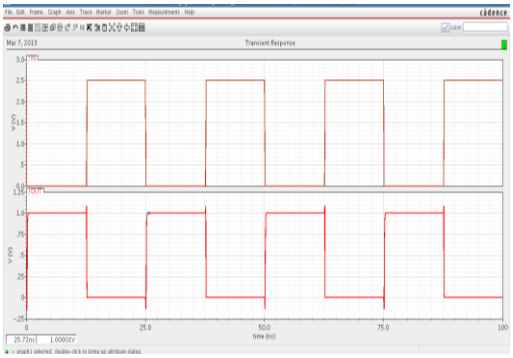

Fig. 10. Waveforms of Simulation circuit A

A large keeper device helps, but this will directly affect performance, and active leakage power dissipation still remains a problem. However, research has shown that domino gates can be made to function at low voltages and lows. With careful attention to noise, the use of keeper devices, and improved device characteristics, domino logic will likely continue to be used in future technologies. As long as low and low dynamic logic can be made to work, then it will be beneficial to use the dual- domino methodology. Although it has little effect on active leakage power, dual- domino significantly reduces standby leakage, which can play an important role in many applications where waiting times are long. Furthermore, switching to standby mode using this methodology has low overhead because one only needs to gate the clocks and then assert the initial inputs into the pipeline. As a result, this power down mode can also be effective at fine grain control such as for inactive modules within a chip like a multiplier or divider.

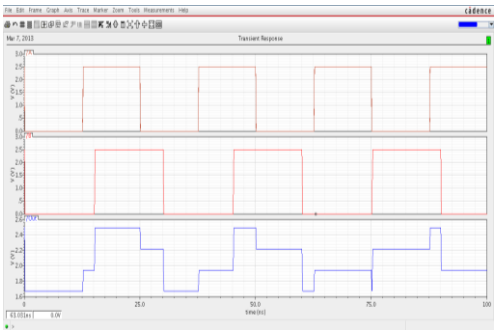

Fig. 11. Waveforms of Simulation Circuit A

VII. CONCLUSION

The supply and threshold voltage is generally advantageous, especially when the transistors are velocity saturated and the nodes have a high activity factor. In fact, for modern submicron technologies, this simple analysis suggests optimal energy efficiency at supply voltages under 0.5 V. Since sub threshold leakage currents will become an increasingly dominant component of overall power consumption in future technologies, dual-threshold voltage circuit techniques will play an important role in future circuit design.

REFERENCES

[1] V. De and S. Borkar, “Technology and design challenges for low power and high performance,” in Proc. Int. Symp. Low Power Electronics and Design, 1999, pp. 163–168.

[2] T. Sakurai and R. Newton, “Alpha-power law MOSFET model and it’s applications to CMOS inverter delay and other formulas,” IEEE J. Solid-State Circuits, vol. 25, pp. 584–594, Apr. 1990.

[3] A. Chandrakasan, I. Yang, C. Vieri, and D. Antoniadis, “Design considerations and tools for low-voltage digital system design,” in ACM/IEEE Design Automation Conf., June 1996, pp. 113–118.

[4] R. Gonzalez, B. Gordon, and M. Horowitz, “Supply and threshold voltage scaling for low power CMOS,” IEEE J. Solid-State Circuits, vol. 32, pp. 1210–1216, Aug. 1997.

[5] M. Horiguchi, T. Sakata, and K. Itoh, “Switched-source-impedance CMOS circuit For low standby sub threshold current giga-scale LSI’s,” IEEE J. Solid-State Circuits, vol. 28, pp. 1131–1135, Nov. 1993.

[6] T. Kawahara, M. Horiguchi, Y. Kawajiri, G. Kitsukawa, and T. Kure, “Sub threshold current reduction for decoded-driver by self-reverse biasing,” IEEE J. Solid-State Circuits, vol. 28, pp. 1136–1144, Nov. 1993.

[7] Y. Ye, S. Borkar, and V. De, “A new technique for standby leakage reduction in high-performance circuits,” in 1998 Symp. VLSI Circuits, June 1998, pp. 40–41.

[8] T. Kuroda and T. Fujita et al., “A 0.9V, 150MHz, 10mW, 4mm , 2-DCT core processor with variable VT scheme,” IEEE J. Solid-State Circuits, vol. 31, pp. 1770–1778, Nov. 1996.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

[9] W. Lee et al., "A 1V DSP for wireless communications," in ISSCC, Feb. 1997, pp. 92–93.

[10] S. Mutoh, T. Doeskin, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Ya-mada, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," IEEE J. Solid-State Circuits, vol. 30, no. 8, pp. 847–854, August 1995.

[11] S. Mutoh, S. Shigematsu, Y. Matsuya, H. Fukada, and J. Yamada, "1V multi-threshold CMOS DSP with an efficient power management tech- nique for mobile phone application," in IEEE ISSCC, 1995/1996, pp. 168–319.

[12] J. T. Kao, A. P. Chandrakasan, and D. Antoniadis, "Transistor sizing issues and tool For multi-threshold CMOS technology," in ACM/IEEE Design Automation Conf., June 1997, pp. 409–414.

[13] J. Kao, S. Narendra, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns," in ACM/IEEE Design Automation Conf., June 1998, pp. 495–500.

[14] T. Sakuta, W. Lee, and P. Balsara, "Delay balanced multipliers for low power/low voltage DSP core," in IEEE Symp. Low Power Electronics, 1995, pp. 36–37.

[15] J. Kao, "Dual threshold voltage domino logic," in 25th Eur. Solid-State Circuits Conf., Sept. 1999, pp. 118–121. VLSI Circuits,