Different Perspectives of Low Power Design for CMOS VLSI Circuits
Vikram Singh1, Sandeep K. Arya2, Manoj Kumar1
1,2Department of Electronics & Communication Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India
3USICT, Guru Gobind Singh Indraprastha University, New Delhi, India

Abstract: Recently, due to the continuous scaling of feature size, a massive development has been taken place in the field of portable devices. Present electronics market is full of battery operated highly complex gadgets and multifunctional portable devices. The power consumption in these devices has become the pivotal constraint which imposes huge restrictions in the development of low power portable devices in the near future. The key thrust of this paper is to focus on the various perspectives of low power design for CMOS VLSI circuits.

Keywords: DTMOS, Leakage Power, Low power design, MTCMOS, Sleep Transistor

Introduction
In today’s life, every human being is surrounded by lots of portable electronic devices. Usages of these portable devices have been increased day by day. The increased demand of portable devices led the researchers to concentrate on low power design, rather than in the field of VLSI design to meet out the emerging challenges of twenty first century electricity scenario which slogans as, “Save resources for the next generations”. Low power consumption and high performance of portable devices have been reported at the top four priorities by ITRS (International Technology Roadmap for Semiconductors) 2012 [1]. The power consumption in VLSI circuits can be explored at different levels of abstraction starting from the lowest i.e., device level, circuit level, logic level, block level, architecture level and up-to system level. Power optimization techniques at the higher level requires, dedicated study of the application considered for optimization. Whereas power optimization techniques applied at the lower levels (circuit and device level) can be generally applied to any circuit. At the device level, due to the advancements in the semiconductor technology, power consumption is reduced, but the fabrication cost is increased. Several techniques have been proposed for low power design at circuit level, including Clock Gating [22-25], Transistor Sizing [17-20], Gate Reorganization [2], Pre-computation logic [21] and for circuits operating in subthreshold region techniques which have been proposed by several researchers includes, Multiple Threshold CMOS (MTCMOS) [37-39], Sleep Transistor [31-32], Stacked Transistor [27-30], etc., .

Need for Low Power Design
Power dissipation is one of the leading and limiting factors in designing of low power portable devices and increasing power consumption level in integrated circuits has become a major issue of the semiconductor industry. The demand of low power portable devices which consumes very less power is increasing day by day. Limited battery life, higher operating frequency, increased chip density and packaging problems are some main factors which causes the researchers to shift their mind from high speed to low power design. Some main factors for this diversion have been described briefly in the below mentioned subsections.

A. Performance Requirement
In present scenario, portable devices are requiring high system performance with very low power dissipation, such as mobiles, laptops, and personal digital assistants (PDAs).

B. Increased Chip Density
High chip density with high performance has also forced to VLSI designers to design high density circuits. But, with the increase in the chip density, the number of transistors fabricated on the same chip has been increased which causes the total capacitance to be increased. Consequently the power consumption is increased as it is directly proportional to the overall capacitance of the circuit [2]. Devices with high performance and throughput need high operating frequency, which again increases the power consumption because power consumption is directly proportional to operating frequency. This creates a great challenge for the designers to maintain the tradeoff between power consumption and system performance.
C. Battery Power Limitations
As all of the portable devices are powered by battery and, with the increasing demand of the portable devices the demand of large backup with a long lifetime battery, with less weight has been increased these days. Old Nickel Cadmium (Ni-Cd) batteries has already been replaced by high energy density (energy stored per unit weight) Nickel Metal-Hydride (Ni-MH) and Lithium-Ion batteries [3]. However, revolutionary increase in the battery capacity cannot be expected in the near future because it has already been reached to the severe level explosive chemicals and consumer safety is given a priority at this time. It concludes that, battery imposes a strong limit on the low power VLSI design and increasing the battery capacity cannot come under the category of low power circuit design, as battery design itself needs a special attention towards it.

D. Packaging Technology
Packaging problem is another major challenging issue for high performance microprocessors where it becomes necessary to mount cooling fan directly on the chip.

Sources of Power Dissipation
Power dissipation in CMOS circuits [4] can be divided into two categories: Dynamic power consumption and Static power consumption i.e.,

\[ P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}} \]  

(1)

Figure 1: Major sources of power dissipation in CMOS circuits

A. Dynamic Power
There are three main components of dynamic power dissipation. First is switching power second is short circuit power and the third is power consumption due to unwanted glitches presents in digital circuits.

\[ P_{\text{dynamic}} = P_{\text{switching}} + P_{\text{shortcircuit}} + P_{\text{glitching}} \]  

(2)

Figure 2: Basic CMOS inverter, (a) charging mode and (b) discharging mode
B. Switching Power

The dynamic power consumption occurred due to the switching [2] activities of the logical inputs from logic HIGH to LOW or vice-versa. All logical switching activities provide the foundation for internal capacitive nodes to either charge from 0V to \( V_{dd} \) through pull-up network or to discharge the nodes from \( V_{dd} \) to ground through pull-down network. This charging and discharging action is shown in figure 2 using basic CMOS inverter. Charging activity draws current from the power supply, whereas discharging action flows current from node to ground and hence the power is consumed. This power is dissipated in the charging and discharging resistor (parasitic or intentionally fabricated) in the form of heat. The average power being delivered by charging and discharging circuitry is given by

\[
P = CV_{dd}^2f.
\]

Where, \( C \) = Node capacitance (parasitic or intentionally fabricated capacitive load), \( V_{dd} \) = Supply voltage, \( f \) = charging and discharging frequency.

C. Short Circuit Power

When input voltage of CMOS inverter is greater than \( V_{th} \) of NMOS then, NMOS transistor will be in ON state and when input voltage is less than \( (V_{dd} - V_{th}) \) then, PMOS transistor will be in ON state. Short circuit power in CMOS circuits represents the power consumed during a condition in which both PMOS and NMOS logic is in ON state. When input switches (from LOW to HIGH or vice versa), there is a short duration for which input level lies between \( V_{th} \) and \( (V_{dd} - V_{th}) \) and both transistors turned ON. A short circuit path is established between \( V_{dd} \) and ground so that, short circuit current will flow which causes power consumption. The short circuit power is given by [5]

\[
P_{shortcircuit} = \frac{\beta}{12} (V_{dd} - 2V_{th})^2 \frac{\tau}{T_p}
\]

Where, \( \beta \) is transistor coefficient, \( \tau \) is rise/fall delay and \( T_p \) is the period of the input waveform.

D. Glitching Power

The glitching power [6-7] is another form of dynamic power consumption, which includes both power consumption due to switching and power consumption due to short circuit. Power consumption due to glitch generally occurs when output of a logic circuit changes momentarily changes its state before going into steady state. This is a condition occurs when more than two inputs has changes their state. For example in the figure 3, if input changes from “111” to “010”, output jumps to logic HIGH for a moment and then come back to logic LOW level. Presence of glitch at the output is the result of unequal rise and fall delay of logic gates at different level. Problem of glitch can be resolved by applying proper logic optimization technique to the circuit i.e., by latching the outputs and synchronized it with a clock signal, so that inputs to the next block are synchronized. The glitching power is given by

\[
P_{glitch} = V_{dd}^2 C_L f_{glitch}
\]

Where, \( f_{glitch} \) is the average frequency of the glitch.

E. Leakage power

Leakage power [8-10] is due to the sum of all the currents which is flowing through the MOS device in standby mode, that’s why leakage power dissipation is also known as standby power dissipation. Leakage power is another important challenging parameter for the designers. Due to the regular scaling of supply voltage, the threshold voltage is also decreased, where the design enters into the sub-threshold region. Leakage power is a result of leakage current that flow from the gate oxide and p-n junctions. With the
continuous scaling of feature size, leakage power becomes a dominating factor in the low power VLSI design. The three main components of leakage current in MOS are sub-threshold leakage, gate leakage and source to body or drain to body reverse biased leakage current. Diagrammatic representation of components of leakage currents is shown in figure 4.

![Diagram of Leakage Current Components in MOS](image)

**Figure 4: Leakage Current Components in MOS**

F. Subthreshold Leakage

Subthreshold current is also known as weak inversion conduction current, which flows between drain to source when gate voltage is below $V_{th}$ [11-12]. Due to regular scaling of MOS devices subthreshold leakage current leads the total power consumption due to leakage. The subthreshold leakage current can be expressed by [11]

$$I_{sub} = A e^{\frac{1}{n} \left[ V_{sd} - V_{th} + q V_{dr} \right] \left[ 1 - e^{-V_{ds}/V_{th}} \right]}$$  \hspace{1cm} (6)

Where,

$$A = \mu_0 C_{ox} \left( \frac{W}{L_{eff}} \right) V_{th}^2 e^{1.8},$$

$$V_{th} = \frac{K T}{q},$$

$n$ is the sub-threshold swing coefficient, $V_{sd}$, $V_{dr}$, $V_{ds}$ and $V_{th}$ are the bulk to source voltage, gate to source voltage and drain to source voltage and threshold voltage, respectively. $\gamma$ is the body bias effect coefficient and $\eta$ is the Drain Induced Barrier Lowering (DIBL) coefficient. The above equation shows that sub-threshold leakage current exponentially depends on $V_{sd}$, $V_{dr}$ and $V_{th}$. Sub-threshold leakage is a major source of leakage power dissipation in nanoscale MOS devices.

G. Gate Leakage

Gate leakage current is due to the direct tunneling of electrons or holes from substrate to the gate terminal. This direct tunneling is modeled in [14-15] as gate direct tunneling current density ($J_{dt}$) and is given by

$$J_{dt} = A \left( \frac{V_{ox}}{T_{ox}} \right)^2 \exp \left[ \frac{-B \left( 1 - \left( \frac{V_{ox}}{T_{ox}} \right)^{2.5} \right)}{V_{ox}/T_{ox}} \right].$$  \hspace{1cm} (7)

Where,

$$B = \frac{8 \pi \sqrt{2m^*}}{3qK},$$

$m^*$ is the effective mass of electron, $V_{ox}$ is the potential drop across thin oxide, $\Phi_{ox}$ is the barrier height of the tunneling electron and $T_{ox}$ is the oxide thickness.
Low Power Design Methodologies

A. Dynamic Power Reduction Techniques
Dynamic power consumption is directly proportional to the capacitance \( C \), square of the supply voltage \( V_{dd} \) and operating frequency \( f \). Hence for minimizing the dynamic power consumption, it is needs to minimize the parameters \( C \), \( V_{dd} \) and \( f \). Merely decreasing these parameters is not a solution to power consumption problem. It is big a challenging issue for the designers, as it degrades the system performance and leading to increasing delay [15].

(i). Device Scaling
The basic idea of scaling is to reduce the dimensions of mos devices and the connecting wires in the circuits. Scaling achieves the same electric field pattern in the smaller transistors by reducing the applied voltage, also with the thickness of the oxide layer; this is known as voltage scaling.

In constant E-field scaling (electric field) one has to increase the doping concentration of the smaller devices. This reduces the size of the depletion region. In constant E-field scaling, the dimensions, voltage and doping concentration are all modified by a common factor \( \alpha \). Constant E-field scaling results in increase in chip density by \( \alpha^2 \), as the length of the connecting wires & device dimensions are reduced. The speed of the small device is improved by a factor of \( \alpha \) because capacitance \( C \) of the every node is now decreased. Therefore the power dissipation per circuit is reduced by a factor \( \alpha^2 \).

Constant E-field has a disadvantage of increase in the leakage current, which limits the practical scaling of power supply. However, to decrease this leakage current, one has to increase the electric field which is known as generalized scaling. Increased electric field is limited by reliability concepts during the long term use such as device degradation resulting from hot carrier mechanism or gate insulation failure. However, in constant voltage scaling, the device dimensions are scaled by the scaling factor \( \alpha \), while the supply voltages and terminal voltages are kept constant. Constant voltage scaling results in the increase in drain current same factor and decrease in the delay by \( \alpha^2 \) with an increase in the power dissipation of the device [16].

(ii). Transistor Sizing
Transistor sizing is an effective and efficient circuit level low power design technique. Sizing the transistor up or down to achieve desired circuit/system performance is one of the important techniques of low power design. The transistor sizing [17-20] can be used to reduce dynamic power consumption by reducing the junction as well as overall gate capacitance of the circuit. Transistor sizing is also important to minimize the circuit area with minimized power consumption and high performance level. Transistor sizing can be characterized in path based transistor sizing and global transistor sizing. The path based transistor sizing is applied to the transistors occurring in critical paths. These critical path transistors are generally upsized so that desired performance can be achieved. Whereas transistors occurring in non-critical paths are sized down to reduce the overall power consumption. However, in global transistor sizing [18], all the transistors are globally sized down to achieve given power and delay product.

(iii). Transistor/Gate Reordering
Transistor reordering is used together with the transistor sizing. In this technique, switching power is reduced by reordering the transistors connected in a serial mode. This technique is based behavior of the different logical inputs [2]. One has to do deep study of the probabilistic time-spam for which an input stays at a particular logic level (HIGH or LOW).

(iv). Pre-computation or Conditional Operation
In combinational and sequential circuits, there are number of possible combinations at input side for which internal nodes are kept charging and discharging but the final output is not changed. The basic principle of pre-computation logic is that the designer has to provide a block that selectively precompute the output logic values before one clock cycle prior to their requirement, and then use these precomputed values to reduce internal activities. This reduces the switching capacitance and hence the dynamic switching power is reduced [21].

(v). Clock Gating
Clock unit, which includes the clock generator, clock drivers, clock tree, etc., of the large synchronous circuit such as microprocessor is the main power consuming portion. Unnecessary switching activities of the input terminals/nodes increases dynamic switching power dissipation significantly. By the use clock gating, the dynamic switching power dissipation in digital circuits can be reduced [22-24]. Clock gating is used to disable
the idle blocks in the synchronous circuits. Clock gating has been proved as an efficient and effective technique for power saving in synchronous circuits. By shutting down the inactive portion of the large synchronous circuit, lots of power can be saved. Clock gating is also helpful in reducing the glitching power by stopping redundant inputs to apply to the circuit [25]. A simple clock gating digital Circuit is shown in figure 5.

Figure 5: Clock gating: (a) traditional digital circuit (b) Clock- Gated implementation of digital circuit [25]

The effectiveness of clock gating depends upon the ability of the circuit designer and the size of the circuit itself. Since an extra controlling circuitry is required to operate gate enable signal i.e., for enabling or disabling the clock signal when required. So clock gating is suitable for large digital circuits, as power is also consumed by controlling circuit. Several synthesis algorithms have been proposed in [23-25] for searching and grouping the circuit for clock gating and for proper clock routing so that switching capacitance can be minimized with large amount of power saving.

**Static Power Reduction Techniques**

With the technology scaling, due to the reduced channel length, static power consumption becomes more significant problem than dynamic power consumption. Lower threshold voltage in CMOS circuit results in increase in the leakage current because the transistors operating sub-threshold region cannot be turned off completely. Due to this, leakage power turned out to be an important portion of the total power consumption. Various techniques have been proposed by researchers to reduce this leakage power such as multiple thresholds CMOS (MTCMOS) [37-39], transistor stacking and use of sleep transistor. However, different tradeoff between chip area, cost and runtime performance had been noticed depending upon the technique used. In this section, we briefly reviewed various leakage power reduction techniques at circuit level.

A. Forced Stack

Forced stacking [27-28] is done by replacing every transistor with a series combination of two transistors as shown in figure 6(a), so that for every input, two half sized transistors in ON and OFF state are present which reduces leakage current significantly. The disadvantage of this technique is that the current driving capability of the circuit is reduced which in turn increases the delay, as the number of transistors are doubled [29]. Similar dual stack approach using two or more stacked transistors and a sleep transistor has been also described in [30].
The main drawback of this sleepy stack technique is that it incurs additional delay, power and area costs in this technique. The operation of the sleep transistors in sleepy stack is same as that of the sleep transistors in sleepy stack technique shown in figure 6(d). Due to the parallel connection of sleep transistor and the stacked transistors as shown in figure 6(d). Due to the parallel connection of sleep transistor and the stacked transistors, leakage power is reduced while delay has been increased. However, tradeoff between delay, power and area occurs in this technique. The operation of the sleep transistors in sleepy stack is same as that in the sleep transistor technique.

B. Power Gating using Sleep Transistor

Power gating is a useful technique to reduce standby or leakage power in CMOS circuits. In this technique, extra NMOS and PMOS transistors which are inserted in series with the pull-up or pull-down network between the network and power supply or ground. These extra transistors are known as sleep transistors [31-32]. The example circuit (CMOS inverter) is shown in figure 6(b). For standby leakage power reduction, sleep transistors are made to switch from ON to OFF or vice versa depending upon the condition, whether the main circuit is in active mode or in standby mode. These sleep transistors are turned ON when the main circuit is in active mode and are turned off when the main circuit is in standby mode. The main circuit is disconnected from the power supply or ground when not in use and the leakage power is reduced. However, extra controlling circuitry is needed to run these sleep transistors. So, power gating using sleep transistors is useful when there is an overall power reduction.

C. Sleepy Keeper

Sleepy keeper [33] uses sleep transistors with two additional transistors to save the state of the circuit output when the circuit is in standby mode. These additional transistors are placed in parallel with the previously placed sleep transistors.

D. Sleepy Stack

Sleepy stack [34] includes both sleep transistor as well as stack approaches and divides existing transistors into two half size transistors like the forced stack approach. Then sleep transistors are placed in parallel to one of the stacked transistors as shown in figure 6(d). Due to the parallel connection of sleep transistor and the stacked transistor in each network, leakage power is reduced while delay has been increased. However, tradeoff between delay, power and area occurs in this technique. The operation of the sleep transistors in sleepy stack is same as that of the sleep transistors in the sleep transistor technique. The main drawback of this sleepy stack technique is that it increases area a lot.

E. LECTOR and GALEOR Techniques

LECTOR technique [35] of leakage current reduction uses two extra transistors known as Leakage Control Transistors (LCTs) out of which one is PMOS and another is NMOS. These LCTs are connected in series and is inserted between pull-up and pull-down networks as shown in figure 7(a). The gate terminals of these LCTs are connected in such a way that at lease one LCT is in cutoff region and hence the resistance between power supply and ground has been increased and leakage current is decreased. However in GALEOR technique [36], the LCTs of LECTOR technique has been interchanged. LECTOR technique suffers from state voltage reduction for logic LOW output while in GALEOR technique (figure 7(b)) both logic LOW and logic HIGH suffers from state voltage reduction [16].

Figure 6: Inverters: (a) Forced Stacked, (b) Sleep Transistor, (c) Sleepy Keeper, (d) Sleepy Stack
The Leakage current has been significantly reduced due to the HIGH threshold sleep transistors and in active mode performance of the circuit is maintained by LOW threshold devices in logic circuit.

F. Drain Gating

Drain gating is another important technique of leakage current reduction. Different arrangements of drain gating have been described in [29]. In this technique, series connected PMOS and NMOS transistors acting as sleep transistor are inserted between pull-up and pull-down networks. The Leakage current has been significantly reduced as compared to LECTOR and GALEOR techniques. Problem of state voltage reduction is removed using drain gating.

G. Multithreshold CMOS (MTCMOS)

Multithreshold CMOS (MTCMOS) [37-39] technique to reduce leakage power involves the use of HIGH threshold MOS devices for sleep transistor, whereas the LOW threshold MOS devices are used to implement the logic circuit. Logic circuits with LOW threshold voltage are fast but have high subthreshold leakage current while logic circuits with HIGH threshold voltage are slower and have much reduced subthreshold leakage currents. The leakage power in standby mode is reduced due to the HIGH threshold sleep transistors and in active mode performance of the circuit is maintained by LOW threshold devices in logic circuit.

H. Dual Threshold MOS (DTMOS)

Use of dual threshold is another important technique to reduce leakage current. However this technique is based on the capability of circuit designer to search out the critical and non-critical paths between input and output. In dual threshold MOS [40-41] technique, low threshold voltage is used for transistors present in critical path to maintain the performance, while the high threshold is used for non-critical paths to reduce the leakage current and hence the leakage power.

Architecture level Low Power Design

At architecture level, it becomes important to reduce system clock frequency and to decrease the number of critical paths, so that number of transitions can be reduced. At this level, various techniques such as retiming, pipelining and parallel processing [4] are available to reduce overall system power consumption without sacrificing the target system performance. However, it is important to note that power optimization at higher levels of abstraction generally requires a fully devoted research on the specific application. Dynamic power management (DPM) is another low power design methodology that dynamically reorganizes the system such that the requested services and performance can be achieved with minimum number of active components [42].

Conclusion

Power dissipation in CMOS VLSI circuits has been considered as an important factor for the low power VLSI design by integrated circuit designers in modern IC design process. In this paper we presented many low-power design techniques at different abstraction levels. At circuit level and device level, transistor sizing and device scaling are the effective techniques to reduce power. Clock gating, Precomputation logic, pipelining and dynamic power management are the techniques to reduce power dissipation at gate/architecture-level design. The low power design techniques reviewed in this paper facilitate the students to understand the different low power design methodologies.

References


