Low Power VLSI Design Techniques: A Review

Ketan J. Raut[1*], Abhijit V. Chitre[2], Minal S. Deshmukh[3] and Kiran Magar[4]

[1,2,3,4] Dept. of E&TC Engineering, Vishwakarma Institute of Information Technology, Pune, India

[1] Ketan.raut@viit.ac.in, [2] abhijit.chitre@viit.ac.in, [3] minal.deshmukh@viit.ac.in, [4]kiran.219m0014@viit.ac.in

Abstract: Since CMOS technology consumes less power it is a key technology for VLSI circuit design. With technologies reaching the scale of 10 nm, static and dynamic power dissipation in CMOS VLSI circuits are major issues. Dynamic power dissipation is increased due to requirement of high speed and static power dissipation is at much higher side now a days even compared to dynamic power dissipation due to very high gate leakage current and subthreshold leakage. Low power consumption is equally important as speed in many applications since it leads to a reduction in the package cost and extended battery life. This paper surveys contemporary optimization techniques that aims low power dissipation in VLSI circuits.

Keywords: Power dissipation, dynamic power, static power, clock gating, adiabatic logic

1. Introduction

In the past, IC designers were concerned about chip performance (speed), area, and cost. But in recent years semiconductor industries are more concerned about the power consumption of the VLSI ICs along with speed. Chip area and cost are not a major concern today due to scaling down of the MOSFETs in nanometer technologies and mass production respectively.

Dynamic power dissipation and static power dissipation are two main sources of power dissipation in CMOS circuits. The major contribution in dynamic power consumption comes from the switching of the input, output, and internal nodes of the circuit from logic 0 to logic 1 and vice versa. Power dissipation due to short-circuit current flow from supply to the ground when both PMOS and NMOS transistors are ON has minor contribution in the dynamic power consumption. Second source i.e. static power dissipation occurs due to the leakage current flow in the OFF state MOS transistors. Now a days a static power dissipation is more dominating because of billions of OFF state transistors on the chip.

---

* Corresponding Author

Leakage current is a critical part to be considered when designing low power VLSI circuits. Low power consuming components along with low power design has add on advantages. In the past, the real focus for VLSI design was on performance, area, and cost. But now low power is as important as these factors since there is scaling down of technology along with increasing complexity. Scaling down leads to leakage current which pose a major challenge in VLSI design. Many researchers presented that leakage power dissipation is up to 40 percent of total power consumption in deep sub-micron technologies [1]. Reducing power consumption varies from application to application. For example, in mobile phones, which fall in the class of small scale fueled battery applications, the main objective is to keep the battery life long enough along with a low cost.

Researchers has proposed many power optimization techniques at different levels of design abstraction. In the following sections we are describing some of the important and popular low power design techniques at different levels of design abstraction. These techniques are listed in Table 1. Remaining portion of this paper is organized as follows: Section 2 describes the principle of power dissipation in the CMOS circuits. Transistor level techniques to reduce power dissipation are discussed in section 3 followed by circuit and logic level power optimization techniques are presented in section 4 and 5 respectively. Section 6 describes the architecture level low power methods while some advanced techniques are discussed in section 7.

Table 1. Low Power Techniques at Different Design Levels.

Design level	Techniques
Transistor	Threshold voltage change, SOI transistors, and use of high-K materials
Circuit	Transistor sizing, Pin ordering, Gate network reorganization, Multi-threshold CMOS, and transistor stacking
Logic	State machine encoding, Bus invert encoding and clock gating
Architecture	Pipelining, and Parallel processing
Software	Low power compiler design, instructions scheduling etc.
Operating System	Power down and Partitioning

2. Power dissipation in CMOS circuits

There are two kinds of power dissipation in CMOS circuits: dynamic and static. Further dynamic power dissipation can be classified in two more categories: switching power dissipation and short-circuit power dissipation. The main source of switching power dissipation is charging and discharging of node capacitances (parasitic capacitances) of the circuit. Switching power dissipation component is represented by first term in the equation 1. Higher operating frequencies leads to frequent charging and discharging of node

capacitances. It results in increased dynamic power dissipation. Hence to reduce switching power dissipation, switching activity of circuit nodes has to be reduced.

The source of short-circuit power dissipation is short-circuit current flow from V_{DD} to ground in CMOS circuits when both PMOS and NMOS transistors are simultaneously ON. The amount of short-circuit power dissipation depends on short-circuit current flow duration. Higher current duration leads to higher quantity of charge per transition, Q_{sc}. Short-circuit power dissipation component is represented by second term in the equation 1.

The source of static power dissipation is numerous leakage current flow in the OFF state transistor. The most dominant leakage currents are leakage current flow in the reversed biased PN junction and subthreshold leakage current. The total leakage current from various sources is denoted by I_{leak} in equation 1 and third term in this equation represents static power dissipation.

We can sum up all above discussed power consumption components to get the total power consumption in CMOS circuit by following equation [1]:

$$P = \frac{1}{2}(C \cdot V_{DD}^2 \cdot f \cdot N) + (Q_{SC} \cdot V_{DD} \cdot f \cdot N) + I_{leak} \cdot V_{DD} \qquad (1)$$

3. Transistor level techniques

3.1 Threshold voltage change

Higher as well as lower threshold voltage of MOS transistor both can benefit the reduction in the power dissipation. Higher threshold voltage transistors has lower subthreshold leakage current. This results in lower static power dissipation. On the other side, with lower threshold voltage transistors one can lower the supply voltage of the circuit which results in lower dynamic power dissipation.

3.2 SOI transistors

For many years we were using bulk transistor technology. Recently SOI (Silicon on Insulator) transistor technology has been developed. The main difference between bulk and SOI transistor is the presence of buried oxide layer below active silicon layer. Therefore, each transistor can be electrically isolated from the others [2].

Significant power reduction can be achieved by using SOI transistors because of lower parasitic capacitances and lower threshold voltage [2]. By using SOI transistors up to 50% active power can be reduced [3][4].

3.3 High-K materials for gate dielectric

In deep sub-micron technologies, gate dielectric thickness is very small just in the range of 2 to 3 atomic layers due to technology scaling. This results in significant gate leakage current if we are using low permittivity (K) insulators like SiO_2. Most of the industries now replaced

SiO$_2$ by High-K insulators such as HfO$_2$ and HfSiON to reduce the gate leakage even though these insulators has drawbacks like mobility degradation and increased short-channel effects.

4. Circuit level techniques

4.1 Transistor Sizing

Transistor size in a combinational circuit impacts the gate delay and the power dissipation. Wider transistors in a logic gate will have smaller gate delay, however switching power dissipation of the gate will increase. For a particular delay constraint, it is computationally difficult to find the size of the transistor that minimizes the power dissipation. One method to this problem is figuring the slack at each gate in the circuit. Positive slack provides us the margin by which gate can be slowed down so as its critical path delay is not affected. Circuits with positive slacks are processed, and the size of the transistors is reduced until the slack becomes zero, or unit the size of all transistors is smallest [1].

We can do transistor sizing for leakage power reduction also. Shorter channel length transistor will have more subthreshold leakage and thus leads to the more static power dissipation. However, on the other side, short channel transistors are faster because of higher saturation current. This leads to the trade-off between speed of the circuits and the static power dissipation [5].

4.2 Pin ordering

Figure. 1 shows NAND gate circuit designed using CMOS technology. A simple qualitative analysis of this circuit shows that for low power dissipation less charge transfer should happen in parasitic capacitances C$_{out}$ and C$_i$. This can be achieved by applying high transition signal to the input A and low transition signal to the input B. This technique is known as pin ordering [5].
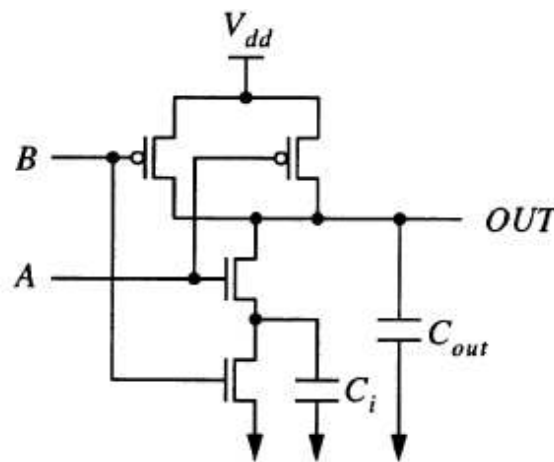

Figure 1. 2-Input CMOS NAND Gate

4.3 Gate network reorganization

We can construct multiple gate-level circuits for the given function which are logically equivalent. Although these different implementations have same functionality, they may differ in delays and power consumption. Figure 2 shows two different gate-level networks with same functionality. It is obvious, network shown in Figure 2(a) will have more glitches than network shown in Figure 2(b). Thus network shown in Figure 2(b) will have less power dissipation compared to network shown in Figure 2(a).

(a)

(b)

Figure 2. Two Different Gate-level Networks for Same Function

4.4 Multi-threshold (V_{th}) CMOS (MTCMOS)

This technique uses both low-V_{th} and high-V_{th} MOS transistors in the circuit. In the standby mode of the circuit both sleep transistors (MP and MN) are in OFF state to disconnect V_{DD} and ground from the circuit. Both of these transistors are high-V_{th} transistors to minimize the leakage current. In the active mode of the circuit both sleep transistors are in ON state connecting V_{DD} and ground to the circuits. Circuit can be implemented using low-V_t transistors to get the benefit of higher speed. Even in the circuit non-critical paths can have high-V_t transistors to minimize the power dissipation as delay is not the major concern on these paths [6][7]. One such example MTCMOS circuit is shown in the Figure 3 [6].

4.5 Transistor stacking

This is a technique in which two OFF transistors are stacked. This reduces the subthreshold leakage in active mode as compared to a single transistor. Two or more series connected OFF transistors will have less leakage current compared to single OFF transistor.

There are two types of stacking: forced stacking and sleepy stacking. In forced stacking, a single transistor of width 'W' can be split into two transistors of width 'W/2' as shown in Figure 4 [8]. In sleepy stacking, first forced stacking is performed and then sleep transistor is connected in parallel with one of the stacked transistor. This helps to reduce the propagation delay.
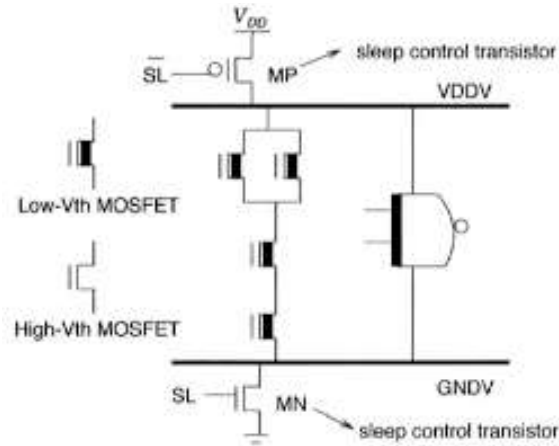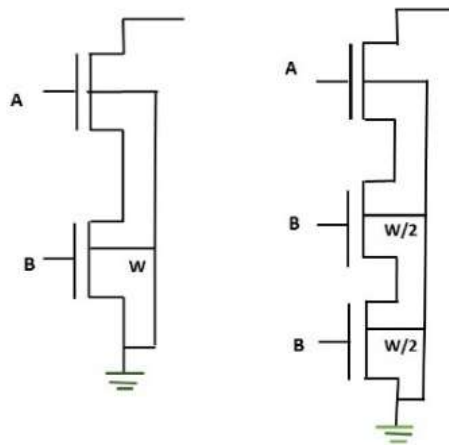

Figure 3. MTCMOS Circuit

Figure 4. Forced Transistor Stacking

5. Logic level techniques

5.1 State machine encoding

Two-bit binary counter will have states 00, 01, 10, 11, and again 00. Thus in four clock cycles there will be six switching activities. Now consider two-bit gray code counter which has 00, 01, 11, 10, and again 00 states. Here with gray code encoding, there are only four switching activities. Thus gray code encoding will have less switching activities and hence consume less switching power and more power efficient compared to binary encoding.

5.2 Bus invert encoding

It is a low power encoding technique first proposed in [9]. This technique is best suited to reduce the power consumption from the switching of data on off-chip buses. Figure 5 illustrates the architecture for this technique [5].

At each clock cycle, current and next bit at source end is observed and decision is made whether to invert next bit or not so that switching should not happen. If current and next bit are not same then next bit is inverted to avoid switching. A polarity status bit is send to the output side XOR gate to re-invert the inverted bits at source end. This technique is very efficient to reduce huge amount of dynamic power dissipation in case of transmission of high switching data on the buses.
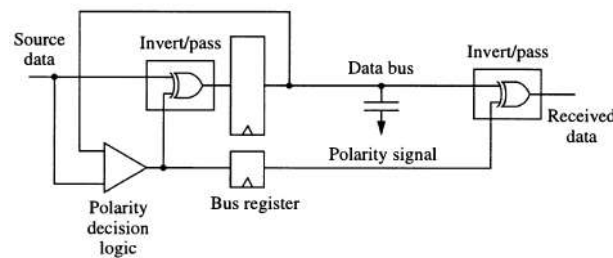

Figure 5. Bus Invert Encoding Architecture

5.3 Clock gating

One of the most efficient and popular clock signal power reduction technique is clock gating. When a clock signal to the functional block like memories, ALUs, co-processors etc. are not required for extended amount of time then we can mask the clock signal to these modules by gating. Normally NAND or NOR gate is used to stop the clock signal reaching to functional units. Figure 6 illustrates one such general clock gating scheme using NAND gate [5].

For many years clock gating technique has been widely used as dominant power reduction technique in processor design by researchers as well as by industries. Recently this technique is even used in the soft-core processor implementation on Artix-7 FPGA [10].
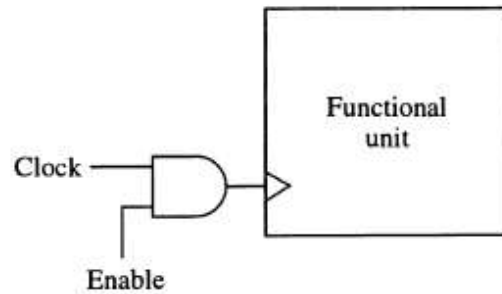
Figure 6. Clock Gating Scheme

6. Architecture level techniques

6.1 Pipelining for low power

Pipelining is normally used to increase the throughput of the system. But if there is no requirement of increase in the throughput, we can you pipelining to effectively reduce the power dissipation of the system. Two-stage pipeline architecture is depicted in Figure 7. We know that, the dynamic power dissipation of conventional CMOS circuit is given by

$$P_{conv} = C\,V_{DD}^2\,f \tag{2}$$

For an N-level pipelined system, its critical path is reduced to 1/N of its original length and the path capacitance is also reduced to $(1/N)^{th}$ of its original capacitance. If the same clock speed is maintained, then in the same amount of time, only $(1/N)^{th}$ capacitance has to be charged and discharged. This suggests that the supply voltage can be reduced to a new value (βV_{DD}). Here β is in the range of 0 to 1. Hence the power dissipation of the pipelined architecture is given by [11]

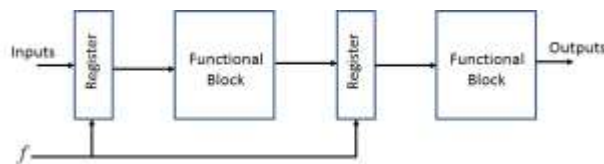$$P_{pip} = C\,\beta^2 V_{DD}^2\,f = \beta^2 P_{conv} \tag{3}$$

Figure 7. Two-stage Pipeline Architecture

6.2 Parallel architecture for low power

Parallel architecture, like pipelining can reduce the power dissipation by lowering the supply voltage. A two-stage parallel architecture is depicted in Figure 8. In an L-parallel architecture,

operating frequency can be reduced to (f/L) without affecting the throughput of the system. As we can operate the system now at much lower frequency, now the V_{DD} can be lowered to a new value βV_{DD} because now to charge the same capacitance more time is available. Here $0 <$ $\beta < 1$. Thus the power consumption of L-stage parallel architecture is given by [11]

$$
\begin{aligned}
P_{par} &= LC\,(\beta V_{DD})^2\,\frac{f}{L} \\
&= \beta^2 C\,V_{DD}^2 f \\
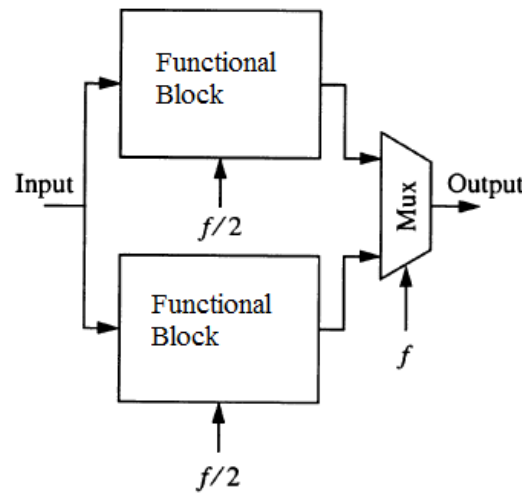&= \beta^2 P_{conv}
\end{aligned}
\tag{4}
$$

Figure 8. Two-stage Parallel Architecture

7. Advanced techniques

7.1 Adiabatic logic design

This logic design technique is known as adiabatic because it consumes zero power ideally. Static CMOS circuits are driven by a constant power supply. In this circuits there is significant amount of power loss as energy gets dissipated by channel resistance of MOSFETs during charging and discharging of load capacitance. To minimize this power dissipation energy-recovery adiabatic logic was originally proposed in [12][13]. Figure 9 shows a four-phase adiabatic inverter proposed in [12] and presented in [5].
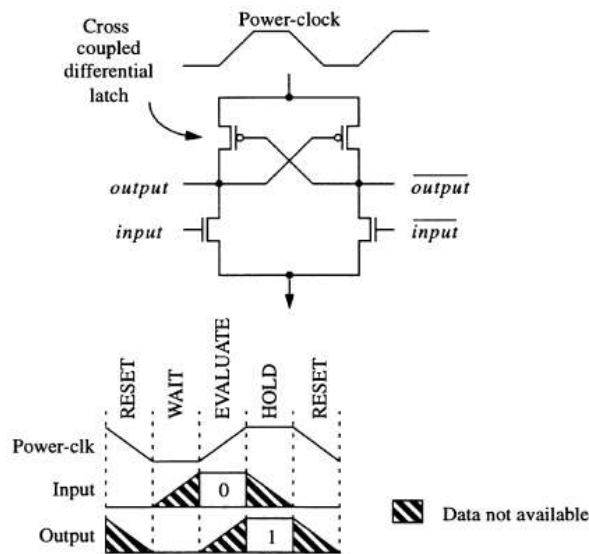
Figure 9. Four-phase Adiabatic Inverter

A differential input is applied to the circuit and output is also differential. This circuit requires four-phase power clock for its working. Supply voltage is not constant. However a variable power-clock is used to power the circuit. PMOS transistors are cross-coupled with differential outputs to provide positive feedback. NMOS transistors acts as evaluation transistors during logic computation. Logic is computed during the EVALUATION and the HOLD phase as shown in Figure 9. The other phases are required for synchronous step-by-step operation of the circuit.

The complexity of full adiabatic logic is high. The power-clock signal requires large number of phases for a moderately complex full adiabatic logic. A large amount of energy is wasted to generate many phases of power-clock signal. Hence quasi-adiabatic logic is more energy efficient at lower frequencies. A new technique to further reduce the power dissipation of quasi-adiabatic circuit is proposed in [14].

The power efficiency of adiabatic circuit is good only at lower frequencies. At higher frequencies energy consumption of adiabatic circuits became comparable to energy consumption of static CMOS circuits [5] as depicted by graphs in the Figure 10. This is the main drawback of adiabatic logic.
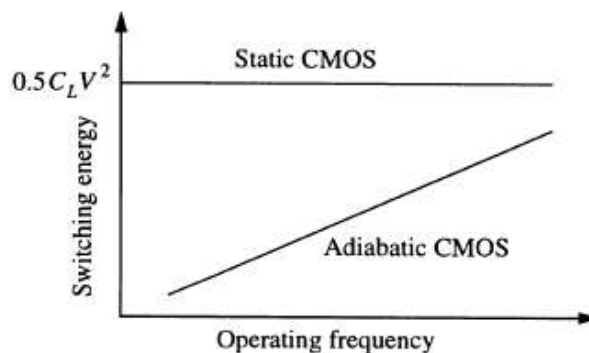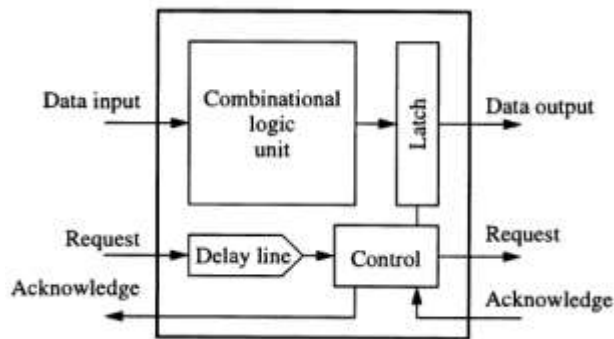
Figure 10. Switching Energy of Adiabatic Logic and Static CMOS

7.2 Asynchronous computation

Today, a major portion of dynamic power dissipation in the chip arises from high frequency clock signal applied to synchronous processing unit. If we succeed to design asynchronous processing unit then there will be huge power saving as clock signal is not required. The speed of asynchronous computation unit is limited by the inherent delay of the circuit components [5].

A general block diagram of asynchronous computation unit is depicted in the Figure 11. The acknowledge and request signals synchronize the computation sequence and hence acts like a clock signal. Like clock signal there is no need of routing the acknowledge and request signals throughout the IC. Absence of high speed clock signal significantly reduces the switching activities in the asynchronous computation unit. This results in much lower power dissipation in the circuit. In [15] one such error correction code integrated circuit designed using fully asynchronous processing is presented. It reports, an 80 percent power saving when equated to the synchronous design.

Figure 11. A General Asynchronous Computation Unit.

8. Conclusion

Power optimization techniques at various levels of design abstraction is presented. Static power dissipation on the chip is increasing day by day due to very high level of integration. New manufacturing and design methods and techniques are evolving now a days to limit very high leakage current in the CMOS circuits. Today, high speed clock signal is the major contributor of dynamic power dissipation compared to data processing on the chip. Out of various power optimization techniques presented in this paper, the clock gating technique is commercially more often used technique. Adiabatic logic design and asynchronous computation are better choices but has many practical limitations. Software and operating system level power optimization techniques are gaining more attention now a days.

References

[1]   Srinivas Devadas and Sharad Malik, "A survey of optimization techniques targeting low power VLSI circuits", Proceedings of the 32nd Design Automation Conference, San Francisco, CA, (1995) pp. 242-247.

[2]   Christian Piguet, Marc Belleville, Olivier Faynot "Low-Power CMOS Circuits – Technology, Logic Design and CAD Tools", 1st Ed., CRC press, (2006).

[3]   L. E. Thon et al., "250-600 MHz 12b digital filters in 0.8-0.25 μm bulk and SOI CMOS technologies", Proceedings of the Int. Symp. on Low-Power Electronics, (1996) August 12-14, pp. 89-92.

[4]   M. Itoh et al., "Fully depleted SIMOX SOI process technology for low power digital and RF device", Proceedings of the 10th Int. Symp. Electrochemical Society, Washington, D.C., (2001) March 25-29, pp. 331-336.

[5]   Gary K. Yeap, "Practical Low Power Digital VLSI Design", Springer, (1998).

[6]   K. Roy and S. Prasad, "Low-Power CMOS VLSI Circuit Design", 1st Ed., Wiley-Interscience, (2000).

[7]   M. Anis, S. Areibi and M. Elmasry, "Design and optimization of multithreshold CMOS (MTCMOS) circuits", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 22, no. 10, (2003) pp. 1324-1342.

[8]   M. Geetha Priya, K. Baskaran, D. Krishnaveni, "Leakage power reduction techniques in deep submicron technologies for VLSI applications", Procedia Engineering, Vol. 30, (2012) pp. 1163-1170.

[9]   M. R. Stan and W. P. Burleson, "Bus-invert coding for low-power I/O", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 3, no. 1, (1995) pp. 49-58.

[10]   B. Tan, W. Lee, K. Mok and H. Goh, "Clock gating implementation on commercial Field Programmable Gate Array (FPGA)", Proceedings of the 4th International Conference on Electrical, Electronics and System Engineering (ICEESE), Kuala Lumpur, Malaysia, (2018) pp. 102-106.

[11]   Keshab K. Parhi, "VLSI Digital Signal Processing Systems: Design and Implementation", Wiley-Interscience, 1st Ed., (1999).

[12]   J. Denker, S. Avery, A. Dickinson, A. Kramer and T. Wik, "Adiabatic computing with the 2N-2N2D logic family", Proceedings of International Workshop on Low Power Design, (1994) pp. 183-187.

[13]   A. Kramer, J. Denker, B. Flower and J. Moroney, "Second order adiabatic computation with 2N-2P and 2N-2N2P logic circuits", Proceedings of International Symposium on Low Power Design, (1995) pp. 191-196.

[14]   Prasad D. Khandekar, Shaila Subbaraman, and Rajendra S. Talware, "Ultra-low power quasi-adiabatic inverter", Proceedings of Int. Conference on VLSI and Communication Engineering, (2009) April 16-18.

[15]   K. Van Berkel, R. Burgess, J. L. W. Kessels, A. Peeters, M. Roncken and F. Schalij, "A fully asynchronous low-power error corrector for the DCC player", IEEE Journal of Solid-State Circuits, vol. 29, no. 12, (1994) pp. 1429-1439.