An Energy-Efficient Scan Chain Architecture to Reliable Test of VLSI Chips

M. Saeedmanesh [1], E. Alamdar[1], E. Ahvar[2]

***Abstract*—Scan chain (SC) is a widely used technique in recent VLSI chips to ease the test process of these chips. SCs enhance the observability and controllability of memory elements such as latches and flip-flops used in sequential parts of the chips. In this paper we propose a novel scan chain architecture which can effectively tolerate bit flip errors due to particle strikes. The proposed architecture is based on the use of redundant transistors in memory elements of the chip. Redundant transistors are added in a way which enables the memory elements to recover the stored value in the case of bit flip error occurrence in both test and normal modes of operation. Fault tolerance capability and energy consumption of the proposed scan architecture are compared with the previously proposed scan chains by the means of Spice simulations. Results of the simulation reveal that the proposed architecture has at least 40% improvement in reliability of test process while its power consumption overhead is at most 10% in both normal and test modes of operation.**

***Index Terms*—scan chain, testing, design for testability.**

I. Introduction

Ever increasing complexity of advanced VLSI technologies may produce several defects in manufacturing of recent chips. Manufacturing test is highly recommended to validate that the produced hardware contains no defects that could otherwise adversely affect correct functioning of the chip. Design for Test (DFT) is a design style which adds certain testability features into the VLSI chips in order to ease manufacturing tests for the designed hardware. DFT often calls design modifications that provide improved access to internal chip elements such that the local internal state can be controlled and/or observed more easily. The design modifications can be either adding a physical probe point to an internal node or adding active circuit elements to facilitate controllability/observability of the node. DFT plays an important role in the development of test programs since it eases the test application and diagnostics in VLSI chips.

The most common method for delivering test data from chip inputs to internal nodes of the chip, and observing their outputs, is called scan-chain design. In scan-chain architecture, memory elements (flip-flops or latches) of the design are connected in one or more chains, which are used to gain access to internal nodes of the chip. Test patterns are shifted in via the chain(s), and the results are then shifted out to output pins and compared to the expected golden run results. Straightforward application of scan chain architecture may result in a large set of long test vectors which in turn results in highly time- and energy-consuming test process[1][2]. Several versions of scan chains have been proposed in literature to mitigate the required time and energy of the scan chains during the test process. Test compression techniques address this problem, by decompressing the scan input on chip and compressing the test output [1][3][4][5][6][7]. Large gains are possible since any particular test vector usually only needs to set and/or examine a small fraction of the scan chain bits. The output of a scan design may be provided in forms such as Serial Vector Format (SVF), to be executed by test equipment.

Static compaction of scan vectors which is proposed in [10] provides a noticeable reduction in power consumption for several ISCAS benchmark circuits. However, the scheme presented in [10] only addresses scan-in power and it does not consider power dissipation during the scan-out operation.

A number of data compression techniques have been proposed for reducing volume of test [8][9][10]. In this approach, test vectors are calculated before the test application, after that the pre-calculated test set is encoded to a much smaller test set and stored in Test Equipment memory. An on-chip decoder embedded in the chip under test decodes the test vectors and the decoded vectors are applied to the chip. It was shown in [9][10] that compressing a "difference vector" sequence results in smaller test sets and reduces testing time.

Although several works have been proposed to reduce the power consumption of scan architectures in both normal and test modes of operation, to the best of our knowledge there is no work to deal with the reliability of scan architectures. Since the results of the test process will be used to decide whether the chip is defective or not, reliability of test process is of decisive importance. This paper proposes a novel scan chain architecture which can

[1] Computer Engineering Department, Azad University, Iran, m.saeedmanesh@iau-aligudarz.ac.ir, e.alamdaar@iau-sari.ac.ir.
[2] Information Technology and Communication Department Payame Noor University, Iran, ehssana2000@yahoo.com, IAENG member.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

effectively tolerate particle strikes which are among the main faults in advanced VLSI chips. The proposed architecture is based on the use of redundant transistors in memory elements of the chip. Redundant transistors are added in a way which enables the memory elements to recover the stored value in the case of bit flip error occurrence. Fault tolerance capability and energy consumption of the proposed scan architecture is compared with the previously proposed scan chains by the means of Spice simulations. Results of the simulation reveal that the proposed architecture has at least 40% improvement in reliability of test process while its power consumption overhead is at most 10% in both normal and test modes of operation.

The rest of the paper is organized as follows. Section II presents the related work. Sections III and IV present the proposed scan architecture and simulation results respectively. Finally Section V concludes the paper.

II. Related work

A. Blocked using NOR gate

For preventing the increasing scan wave effects upon logical gates, using the blocked logic in scan flip-flops stimulus path is a simple and effective method sake decrease test energy that is independent from test set.

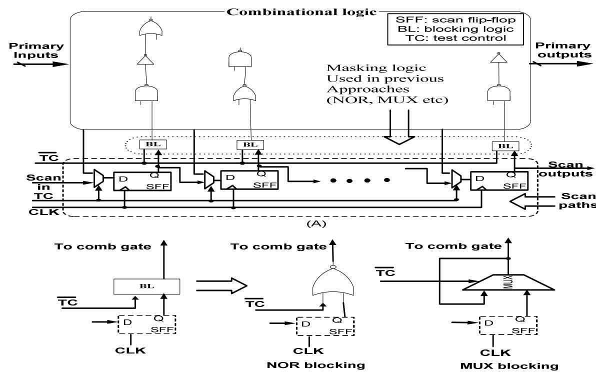


Fig.1. Scan architecture (A) with blocking circuits (B) toward energy reduction.

A research team proposed blocked method according to either NOR or NAND gates. Blocked gates from type of NAND or NOR will be control with test activation signal (Fig1-b) and stimulus paths will be stationary among whole time of scan shift operation even in 0 or 1 logic.

Zhang and his cooperators [17] used the multiplexer in scan cells outputs that maintain the previous scan states among the shift, hence prevent activity in combinational logic.

B. Source blocked transistor toward active energy decrement in scan mode

Another signal block technique called First Level Supply Gating or (FLS) is proposed. It is work by putting a source block transistor in the first level of connected logic to the scan cells outputs that are VDD or GND necessarily. The proposed approach is effective in whole of energy wastage and energy reduction among scan test. Yet whereas we put only one transistor in the first level of charge/discharge path, Latency penalty mainly will be reduce in the comparison of other blocked methods that creates logical extra level in signal propagation path.

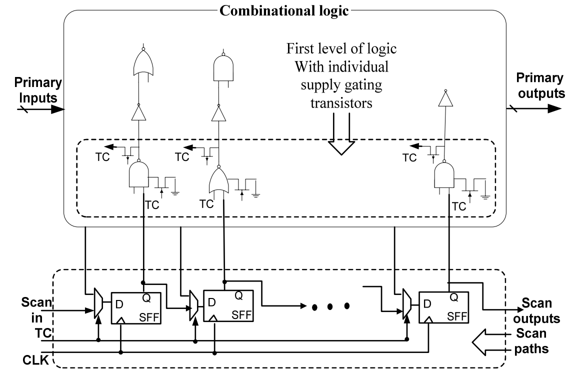


Fig.2. Low energy scan architecture using FLS

We survey the first level source blocked technique that only block the first level connected gates to the scan flip-flops using source block transistors. Entering the blockade transistor into the first level logic separate other combinational logic from input transmissions (scan input) except one transmission 1 to 0 in NMOS source and one transmission from “0” to “1” in PMOS source. It may induct from Figure 2 that the first input change from “1” to “0” will charge the output. This transmission will be distributing among reverser chain. Yet whereas it is impossible to depletion circuit's output among scan shift work-time, another input changes (e.g. from 0 to 1) will not propagate and it effectively decrease the activities during the scan shift in circuit. The outputs of first level gates are float when they are in “0” logic. Float output voltage is determined by leakage equilibrium among pull-up PMOS and pull-down NMOS. Moreover overhearing noise or transient effects caused soft fault may change float output voltage. If the output of first level gate is not VDD or GND exactly, it may cause static shortcut in the logical gate derived by first level gate. This problem has a main concern in the sub-micron technology because noise and leakage increment. For prevent this, first level gates outputs should be adjust on VDD or GND. If GND blocked as Figure 3-a, first level gates outputs may get to VDD by pull-up PMOS derived with GC control signal. If VDD

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

blocked as Figure 3-b, first level gates outputs may get to GND by pull-down NMOS network derived with GC control signal. The generic schematic for proposed source blocking illustrated in Figure 2. For both NOR and NAND circuits, the blocked circuits with GND are rather than the blocked circuits with VDD, because the NMOS transistors are rather than the PMOS transistors with same scale.

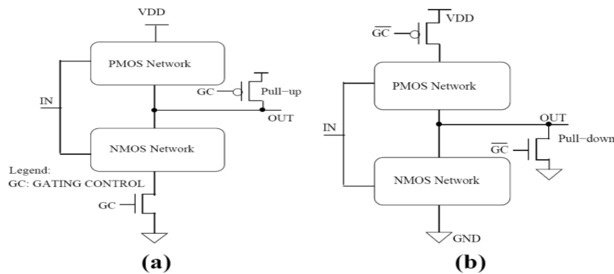


Fig.3. The proposed source blocking schematic (a) GND blocking (b) VDD blocking.

Blocking transistors are reduces the leakage in whole idle gates. It called active leakage because it is occur in the active mode of idle gates. The leakage from compliment circuit is a submission of scan inputs [18] [19]. So with select the best input vector for a combinational circuit in the standby mode, leakage energy waste will be minimized. Many algorithms are proposed for finding the best input vector so far [18] [19]. Current blocking techniques hold inputs status stable during shift. Yet this input status may not be the best input vector that can minimize the whole leakage energy in the combinational block. The circuit status may transform to the best input vector during scan test with selective use of blocking GND for special inputs in order to minimizing the leakage energy waste in combinational circuit. It is noteworthy that the blocking transistor sharing is still possible. For prevent probably short cut status, the sharing must take place among logic gates with same blocking. For example whole NMOS transistors with GND blocking, must share among first level gates with blocked GND and whole PMOS transistors with VDD blocking, must share among first level gates with blocked VDD. Herein an inverse blocking signal needs for VDD blocking transistors. The best input vectors may gain with proposed algorithms in [18].

C. Multiple scan chains for minimizing energy during sequential circuit test

This section introduces another method using multiple scan chains for minimizing energy during sequential circuits test. This method reduce liar transmissions in the under test circuit. For easily reduce liar transmission the proposed method taxonomy scan latch into compatible, incompatible and independent groups. Based on this taxonomy scan latch divide into multiple scan chains and an individual extra test vector calculate with each scan chain, this extra test vector goes into the primary inputs, whereas changing the shift response in each chain it reduce energy waste in combinational part. It is one of latest proposed methods based on scan chain that dose not reduce performance but minimize the energy waste with low effect on area overhead and test data. A path is a set connected gates and wires. Two faults are compatible when they can detect with a single test vector so they are incompatible when a single test vector unable to detect them. In the case that available test vectors cannot detect the least fault, another test vector is needed.

The proposed DFT architecture using multiple SC_0, …, SC_{k-1} scan chains illustrated in Fig-4. Scan in goes into all scan chains but scan out selected from different chains based on controlling structure. The SC0,…,SCk-1 scan chains acts using noninterference state that enable clk0,...,clkk-1 clock signals. This function implements with a scan control register illustrated in figure 4.

Primary inputs adjusts by extra test vector so that reduce the produced liar transmission in the SCi under scan chains scan latchs When shifting existent test response into SCi scan chain.

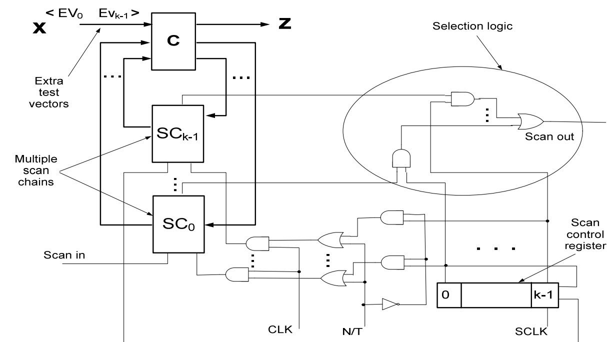


Fig.4. The proposed DFT architecture based multiple scan chains.

III. PROPOSED SCAN ARCHITECTURE

This section proposes a new energy-fault aware scan cell architecture to reliable test of VLSI chips; Fig.5. In normal mode, cell speed has a direct relation to circuit switching speed; circuit-switching optimization can improve it. For this reason, the changed data must be written in output without considering effect of added NOTs. Therefore, we send clock signal to NOT gates with low delay. In this case these gates are open circuit in switching time. They confirm output after switching time with low delay.

When a fault is occurred in the circuit, the first NOT goes to filter mode. The filter mode causes the output goes to the open circuit state. But in this case NOT gates (are

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

shown in red rectangle) receive the signal clock with delay. Delay causes the output data to be held constant in previous value till next clock is received or soft fault is finished. Therefore, the output voltage is protected and noise capacitors can not acceptable affect on output charge. On the other hand, the open-circuit output will be susceptible for faults. Also, these NOT gates can prevent the output to be converted open circuit.

Another benefit of our propose method is its scale. TMR latch consist of 42 transistors but our proposed method has only 30 transistors. It means scale improves about 25% with same speed. Triple Modular Redundancy (TMR) is a well-known fault circuit. The TMR scheme uses three identical logic blocks performing the same task in tandem with corresponding outputs being compared through majority voters.

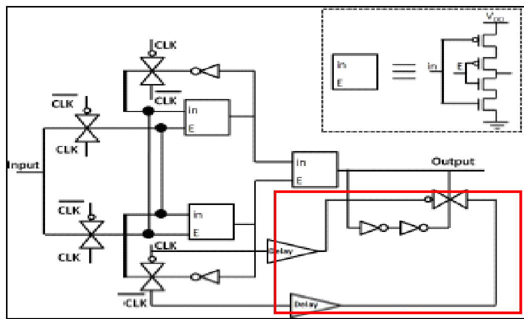


FIG.5. THE PROPOSED SCAN CELL.

IV. EVALUATIONS

A. *Simulation platform*

All blocking methods try to minimize or delete the switching operation in the combination part during sequential part testing. The first method used NOR gates for this purpose, during test NOR gates enters to test state using control line (control=1) and these outputs stuck to 0 so during test take place many switching operations in the sequential part but these operations does not enter to combinational part hence energy consumption is reduced.

The second blocking method uses two PMOS and NMOS transistors for blocking the output of sequential part. Figure 7 illustrates this method. The transistor that is connected to NAND and NOR gates cutting the connection of the first level of combinational part from the ground (or power supply) but in this case these gates outputs are open circuit so they will be susceptible for faults, for prevent this the second transistor that is connected to this gates outputs cause NOR and NAND gates outputs connect to ground or VDD.

Third blocking method is based on type of scan cells outputs. We consider the best division state in this method for simulation and comparison it with other methods so in the specific circuit it is not impossible to do optimize division the energy consumption will be increased. Fig.8 illustrates this method for a template simulated circuit.

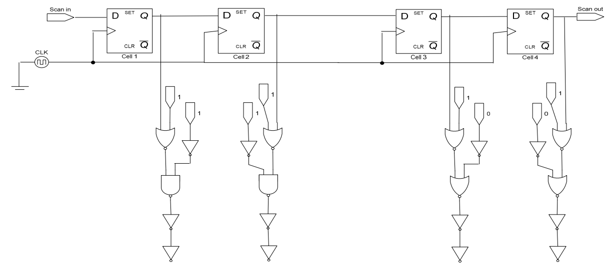


Fig.6. Blocking using the NOR gates.

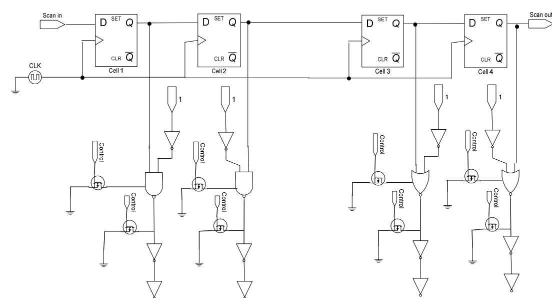


Fig.7. Blocking using PMOS and NMOS transistors.

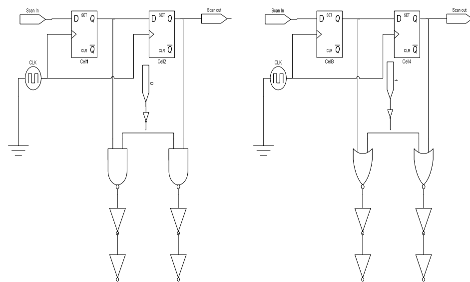


Fig.8. Scan cells division to multiple scan chains based on output type

B. . *Simulation results and analysis*

This section introduces the simulation results for above methods using HSPICE software and 130nm technology. We evaluate and compare the various schemes. The performance measures of interest in this study are: (a) energy consumption; (b) fault tolerance.

We consider the following scenarios for simulations. In Scenario I, we compare our propose method with another methods (described in related work).

In this scenario our measurement is *fault tolerance*. To estimate the fault tolerance of under test circuit the proposed method in [12] is used and circuit is examined in both operational and test states. Conflict effect of high-energy dusts to the sensitive points is simulated with the HSPICE software, in this section the results that are different for these methods are illustrate.

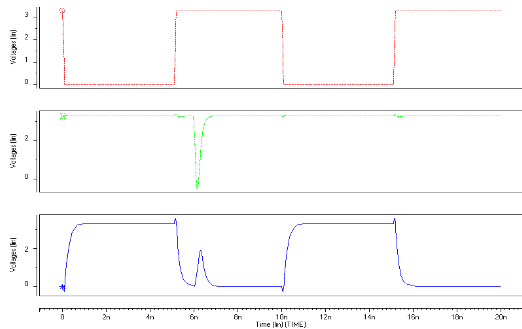
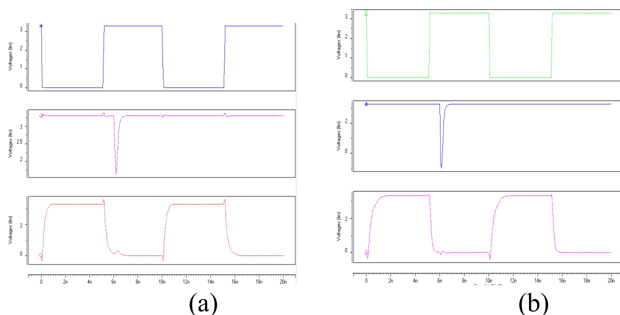


Fig.9. Fault injection simulation in the normal mode of operation (multiple chains method).

In the first parts of figures circuits input pulse is shown, in the second part of them the conflict effect of a high energy dusts is simulated, in the third part of them circuit reaction again dusts conflict is shown.



(a) (b)

Fig.10. Fault injection simulation in the operational state of circuit (a) using extra PMOS and NMOS gates(b)using NOR gate.

As showed in figures 10-11, these methods do not afoul into soft fault again high energy dusts conflict but they tolerance it, because in third parts of figures 10-11 that is showed the output there is no sensible change in the voltage, but figure 10 is shown multiple scan chains based blocking methods do not tolerance dusts effects, so there is soft error in their outputs. Figure 12 shows energy consumption of the proposed scan architecture is compared with the TMR scan cell by the means of Spice simulations. Results of the simulation reveal that the proposed architecture has at least 40% improvement in power consumption.

V. Conclusions

This paper addresses to design of an energy- fault aware scan cell architecture to reliable test of VLSI chips. As simulation results showed, in comparison with TMR method, we could improve the cell scale. Also simulation results showed that energy consumption of the proposed scan architecture has lower overheads in comparison with TMR method. We also compared the proposed method with three scan chain methods base on fault tolerance factors. Upon gained results the proposed method had a better fault tolerance than multiple scan chains again soft faults, also it had acceptable fault tolerant in comparison with other methods.

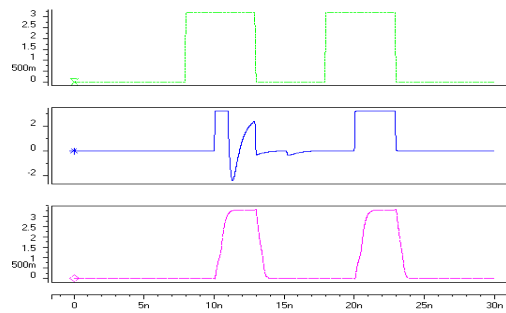


Fig.11. Simulation results for strong fault injection to circuit in operational mode (proposed method).

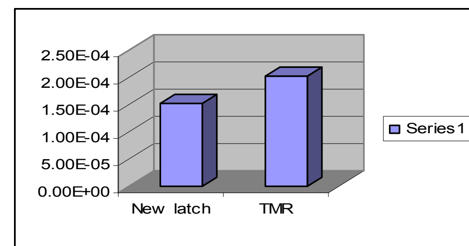


Fig.12. Energy consumption of our propose method and TMR method.

References

[1] Chandra and K. Chakrabarty, "Low-power scan testing and test data compression for system-on-achip," IEEE Trans. on CAD, pp. 597-604, May 2002.

[2] Y. Zorian, "A distributed BIST control scheme for complex VLSI devices," in Proc. IEEE VLSI Test Symp., 1993, pp. 4–9.

[3] S. Wang and S. K. Gupta, "LT-RTPG: A new test-per-scan BIST TPG for low heat dissipation," in Proc. Int. Test Conf., 1999, pp. 85–94.

[4] S. Gerstendörfer and H.-J.Wunderlich, "Minimized power consumption for scan-based BIST," in Proc. Int. Test Conf., 1999, pp. 77–84.

[5] S. Wang and S. K. Gupta, "ATPG for heat dissipation minimization during scan testing," in Proc. Design Automation Conf., 1997, pp. 614–619.

[6] V. Dabholkar, S. Chakravarty, I. Pomeranz, and S. M. Reddy, "Techniques for minimizing power dissipation in scan and combinational circuits during test application," IEEE Trans. Computer-Aided Design, vol. 17, pp. 1325–1333, Dec. 1998.

[7] R. Sankaralingam, R. R. Oruganti, and N. A. Touba, "Static compaction techniques to control scan vector power dissipation," in Proc. IEEE VLSI Test Symp., 2000, pp. 35–40.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[8] A. Jas, J. Ghosh-Dastidar, and N. A. Touba, “Scan vector compression/ decompression using statistical coding,” in Proc. IEEE VLSI Test Symp., 1999, pp. 114–120.

[9] A. Jas and N. A. Touba, “Test vector decompression via cyclical scan chains and its application to testing core-based design,” in Proc. Int. Test Conf., 1998, pp. 458–464.

[10] A. Chandra and K. Chakrabarty, “Test data compression for system-on-a-chip using Golomb codes,” in Proc. IEEE VLSI Test Symp., 2000, pp. 113–120.

[11] W J Dally and J W Poulton, Digital Systems Engineering, CUP, 1998.

[12] M. Fazeli, A. Patooghy, S. Gh. Miremadi, A. Ejlali, “Feedback Redundancy: A Power-Aware SEUTolerant Latch Design in DSM Technologies”, to appear in the Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07), June 25-28, Edinburg, UK.

[13] A. Ejlali, M. Al-Hashemi, P. Rosinger, S. G. Miremadi, “Joint consideration of fault-tolerance, energy-efficiency and performance in on-chip networks”, Proceedings of the conference on Design, automation and test in Europe, France, 2007, pp. 1647 – 1652.

[14] J. Kim, C. Nicopoulos, D. Park, “A Gracefully Degrading and Energy-Efficient Modular Router Architecture for On-Chip Networks”, 33rd International Symposium on Computer Architecture (ISCA'06) pp. 4-15.

[15] M. Omana, D. Rossi, C. Metra, “Novel transient fault hardened static latch”, Proceedings International Test Conference, 2003.. ITC 2003, pp. 886 – 892.

[16] J. M. Rabaey and M. Pedram, Low Power Design Methodologies, Kluwer Academic Publishers, 1996.

[17] X. Zhang and K. Roy, Power Reduction in Test-Per-Scan BIST, International On Line Testing Workshop, 2000, pp. 133-138.

[18] M.C. Johnson, D. Somasekhar, and K. Roy, Models and algorithms for bounds on leakage in CMOS circuits, IEEE Transactions on Computer- Aided Design of Integrated Circuits and Systems, Vol. 18 , no. 6, 1999, pp. 714-725.

[19] D. Lee and D. Blaauw, Static leakage reduction through simultaneous threshold voltage and state assignment, Design Automation Conference, pp. 191-194, 2003.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)