

# Efficient Design Methodologies for Optimizing the Leakage Power in FPGAs

O. Venkata Krishna<sup>1</sup> and B. Janardhana Rao<sup>2</sup>

<sup>1</sup>CVR College of Engineering/EIE Department, Hyderabad, India

Email: venkatakrishna.odugu@gmail.com

<sup>2</sup>CVR College of Engineering/ECE Department, Hyderabad, India

Email: janardhan.bitra@gmail.com

**Abstract:** The scaling of the CMOS technology has precipitated an exponential increase in both sub-threshold and gate leakage currents in modern VLSI designs. Consequently, the contribution of leakage power to the total chip power dissipation for CMOS designs is increasing rapidly, which is estimated to be 40% for the current technology generations. In FPGAs, the power dissipation problem is further aggravated when compared to ASIC designs because FPGA uses more transistors per logic function when compared to ASIC designs. Consequently, in the nanometer technology, the leakage power problem is pivotal to devising power-aware FPGAs. This paper focuses on the architectural techniques for leakage power reduction in FPGAs. In this paper the multi-threshold CMOS (MTCMOS) techniques are introduced to FPGAs to permanently turn OFF unused resources of the FPGA to reduce the power dissipation and the leakage power reduction technique in FPGAs on the use of input dependency is developed. This paper focuses on the reduction of leakage power with respect to CAD tools for the implementation of FPGAs.

**Index Terms**—CMOS, FPGA, CAD, MTCMOS, leakage power, power dissipation.

## I INTRODUCTION

The continuous scaling of the CMOS process has attracted FPGA vendors to integrate more and more devices on the same chip to increase the chip functionality. As a result, the power dissipation of modern FPGAs increased significantly. Much of this increase in power dissipation is attributed to the increase in leakage power dissipation which is expected to exceed 50% of the FPGA power dissipation as modern FPGAs start using the 65nm CMOS process. In addition, the excessive scaling of the MOS gate oxide thickness  $t_{ox}$  resulted in a significant increase in the gate oxide tunneling current, thus exacerbating the leakage problem. In recent experiments, it was found that both the sub-threshold and gate leakage power dissipation increase by about 5X and 30X, respectively, across successive technology generations [1].

This paper will provide architectural modifications to FPGA designs to reduce the impact of leakage power dissipation on modern FPGAs. Firstly, multi-threshold CMOS (MTCMOS) techniques are introduced to FPGAs to permanently turn OFF the unused resources of the FPGA. The FPGAs are characterized with low utilization percentages that can reach 60%. Moreover, such architecture enables the dynamic shutting down of the FPGA idle parts, thus reducing the standby leakage significantly. Employing the MTCMOS technique in

FPGAs requires several changes to the FPGA architecture, including the placement and routing of the sleep signals and the MTCMOS granularity.

Secondly, a new technique for leakage power reduction in FPGAs based on the use of input dependency is developed. Both sub-threshold and gate leakage power are heavily dependent on the input state. In FPGAs, the effect of input dependency is exacerbated due to the use of pass-transistor multiplexer logic, which can exhibit up to 50% variation in leakage power due to the input states.

## II FPGA OVERVIEW

Generally, FPGAs consists of logic blocks, routing resources and input-output pads. However, the advancement in FPGAs including of embedded memory Phase Locked Loops (PLL), DSP blocks, embedded processors and special feature blocks. The complete architecture is shown in the figure.1 Based upon these features the FPGA is considered as attractive design method.

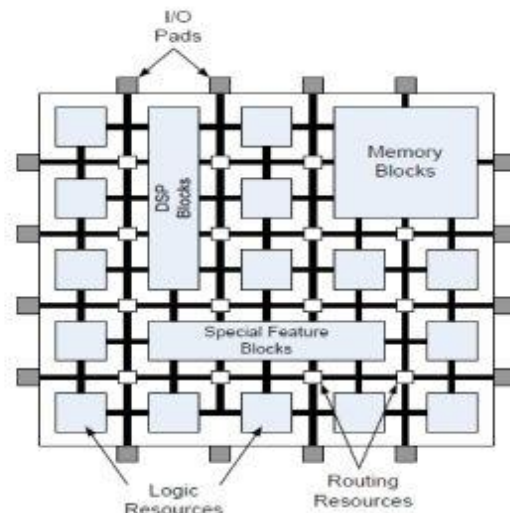


Figure 1: Modern FPGA fabric.

### A) CAD for FPGAs

Generally, all the FPGAs are implemented by huge number of programmable switches. By using those switches only logic functions are designed and implemented. The Computer Aided Design tools of FPGAs transform the design in to stream of binary bits 1's and 0's only. The

design is either schematic entry or a hardware description language. These binary streams of 0's and 1's are used to program the FPGA by proper configuration. The Figure 2 represents the flow diagram of the CAD tools for FPGA design.

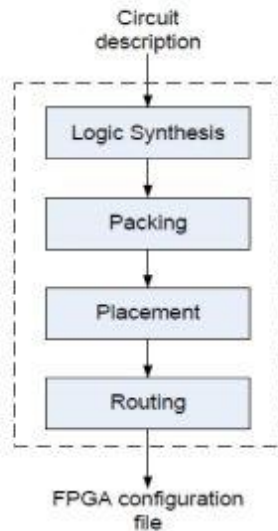


Figure 2: CAD tool flow for FPGA.

### III LEAKAGE POWER REDUCTION IN FPGAs USING MTCMOS TECHNIQUES

In this section, the different supply gating methods using Multi- threshold CMOS techniques for FPGA are presented. These are mainly used to reduce sub- threshold leakage power in FPGAs. For leakage reduction a modified architecture of FPGA using sleep transistors and corresponding CAD algorithms are proposed. In the CAD flow, a new activity profile phase is introduced to identify the blocks of FPGA exhibiting the similar idleness. During the idle times, those idle blocks are turn off by this activity profile algorithm. Next , some new packing methods are introduced to pack the idle blocks with same activity profiles for turning to off them.

In FPGAs, the crucial one is leakage power reduction compared to Full-custom and Semi-custom ASIC designs because in FPGAs the unused resources are more. This standby leakage power is almost 40% due to unutilized resources, thus 60% of resources are only utilized in FPGAs [2]. For wireless communications, the idle mode period is very long [3]. In such FPGA applications, some utilized blocks are forcibly put into standby mode during the idle time and then leakage power is reduced.

In an MTCMOS implementation of FPGA, the sleep transistors are designed by high threshold voltage (HVT) devices. These transistors are connected to the pull-down network. This pull-down network is designed by low threshold voltage (LVT) devices, to which circuit is connected to ground as shown in the figure 3(a). When the sleep transistor is turned off, the leakage current of the circuit is reduced, but has very low saving. Hence, the circuit gets benefit from the Low VT pull down network

when the sleep transistor is on, otherwise the sub threshold leakage current is limited, because sleep transistor is turned OFF.

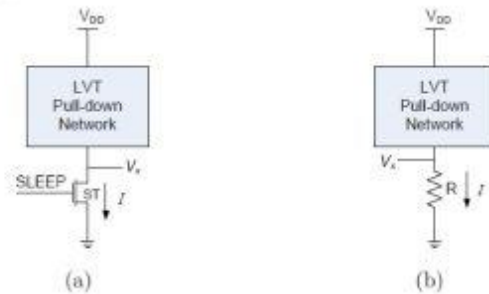


Figure 3: Architecture of MTCMOS. (a) Architecture of general MTCMOS, (b) Equivalent ST circuit in the active mode.

In the above figure.3, the equivalent circuit for the sleep transistor was shown as a finite resistance 'R' between the ground in the case of the sleep signal is high with a small voltage at the virtual GND line Vx. However, the sleep transistor finite resistance R incurs a penalty of performance due to the reduction of driving potential in the circuit VDD-Vx [3]. If the sleep signal of the sleep transistor is low, the entire circuit goes into a standby mode. In this case, the voltage Vx rises to 0 Volts and VDD Volts. Next, the sleep transistor acts as a very high resistance. This resistance is used to reduce the sub-threshold leakage current. In FPGAs, sleep transistor circuits can reduce sub-threshold power leakage by (i) Power down the unutilized resources of the chip during configuration permanently (ii) turning off and on the unused resources of the chip dynamically based upon the activity of the circuits, and (iii) powering down the entire FPGA during the idle time.

In this paper, the changes at the CAD level were developed for FPGA design using the MTCMOS technique to get the full advantage of leakage power saving. All these changes are integrated into the Versatile Place and Route (VPR) flow [4]. The figure 4(a) and figure 4(b) shows the flow diagrams of the typical VPR CAD tool and proposed modifications respectively.

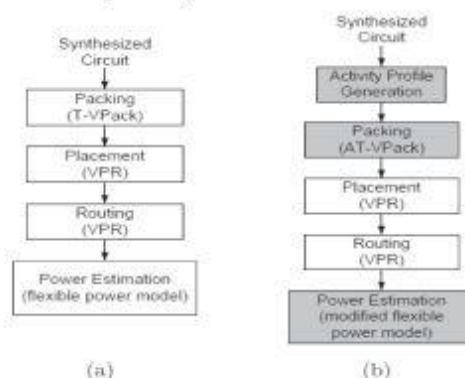


Figure 4: Process flow of CAD in FPGA (a) flow chart for VPR conventional method. (b) Proposed CAD flow diagram included in the VPR flow.



The Figure 4(b) shown as the considering the activity generation phase. In this stage, the identification of unutilized blocks with same activity is taken place. Blocks with similar activity profiles are forced into a standby mode together. The next stage is T-V Pack algorithm, which is used to integrate with activity profiles generation method and packing algorithm resulted algorithm is AT-V Pack as shown in figure 4(b). Finally, a modified power estimating model is introduced to properly calculate the power savings in this proposed MTCMOS FPGA architecture.

The conventional FPGA architecture is used in most of the modern FPGAs. This FPGA architecture consists of logic blocks, which are implemented by a four input Look Up Table (LUT), a Flip Flop and a 2X1 multiplexer as shown in figure 5.

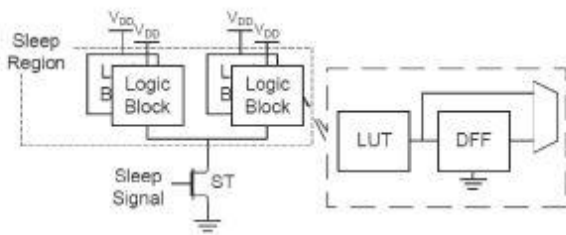


Figure 5: The MTCMOS type FPGA Architecture.

The logical blocks connected to one sleep transistor are called as sleep regions. The sleep signal controls each sleep region. Deactivation of the sleep signal moves the N clusters into low power mode during the inactive times. The output of the each logic block is stored in the Flip-Flop, before entering the sleep mode so it can recovered when the corresponding sleep region wakes up again. The sleep signals of the unused blocks or resources of the FPGA are deactivated during the OFF time permanently. In Modern FPGAs, during the runtime the sleep signals are dynamically generated using the partial reconfiguration logic [5], thus providing minimum area overhead. If the design is well- known in advance, the sleep signals easily can be generated [6].

The no. of clusters that can be determined by the size of the sleep transistor, leakage power savings area over head and the maximum permitted ground bounce about virtual ground lines. For this performance large sleep regions employ few number is larger transistors. As a result, the control circuitry needed to produce the less complex sleep signals. These signals consume less power and occupy less area when compared to small sleep regions.

The figure 6 shows the proposed FPGA fabric, in which the sleep transistors are pre-fabricated with fixed size in the FPGA Architecture. This figure shows the placement information. It provides the full connectivity between the sleep transistors and logic blocks and also provides minimum area overhead. During the FPGA fabrication, the sleep control signals of the each and every sleep transistor are hardwired [7]. As shown in the figure, the virtual ground  $V_{GND}$  line is connected to the pull-down networks of the logic blocks to the sleep transistor. The Virtual ground

$V_{GND}$  lines are hardwired to their corresponding sleep transistors. Many research methods are proposed to optimize area [8], and the average overhead of the MTCMOS architectures using fine granularity in FPGAs is around 5% [9, 10]. For sleep transistor implementation, two types of devices are considered: header or footer devices. The Pmos transistors are used as header devices to block the current path from the supply line and pull down network. Footer devices are implemented by Nmos transistors to block the ground path, as shown in figure.7.

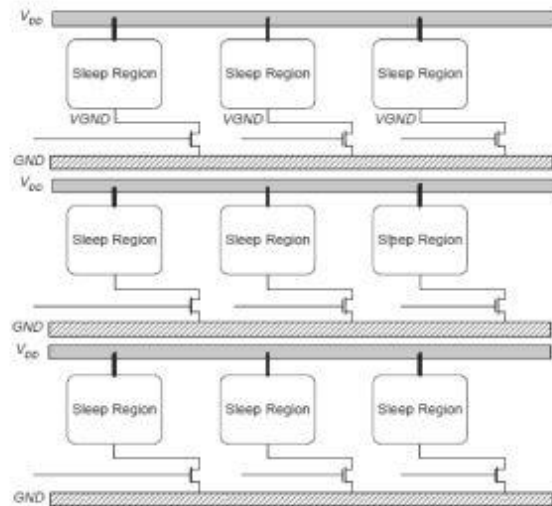


Figure 6: MTCMOS-based FPGA fabric with sleep transistors.

The header- PMOS approach has the drawback of area penalty when compared to the footer- NMOS approach. This is due to the low driving current of PMOS transistors because of low mobility of holes. Consequently, only NMOS footer sleep transistors are used in this work.

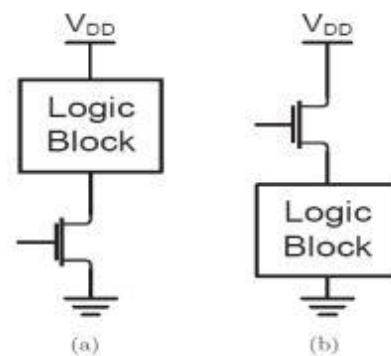


Figure 7: (a) NMOS footer sleep transistor implementation. (b) PMOS header sleep transistor implementation.

There are two types of sleep transistor architectures used in this paper. Those are local sleep transistors and global sleep transistors. The local sleep transistors are used at the logic block level, these are independently idle. Second, the global sleep transistor architecture consists of a single sleep

transistor for entire large block that consists again many local blocks [11].

**IV LEAKAGE POWER REDUCTION IN FPGAs THROUGH INPUT PIN REORDERING**

In FPGAs, the power dissipation is state dependent as the i/p signals are forced into substantial leakage power reduction technique. This happens due to the pass transistor logic in the design.

In this section, a new methodology is introduced, which is based on reordering of the i/p pins reordering to decrease the leakage power dissipation in all blocks and resources of the FPGAs without any performance penalties. This proposed methodology handles the logic and routing resources differently to achieve more leakage reductions. Next, a new method with some modifications is proposed and implemented to increase the performance along the critical path [12].

The proposed pin reordering algorithm consists of two phases: (1) Logic pin reordering (LPR)- it targets leakage power reduction in the logic blocks of FPGA and(ii) Routing pin reordering (RPR)- this phase targets the routing switches. Right after the synthesis the LPR phase is performed and previous to the Packing stage. The RPR phase is performed next to the routing stage only. This process stages are shown in the figure 8. The CAD flow corresponding to these pin reordering methodologies are changed. The T-V Pack is used for packing and VPR CAD tool for placement and routing [13].

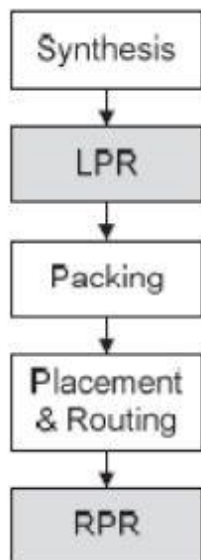


Figure 8: The proposed pin reordering algorithms with VPR CAD flow.

Figure 9 shows the different leakage savings due the LPR Phase. From the figure, we can understand the maximum power savings are generated due to the input swapping phase. The very small leakage power saving is

due to the unutilized block into a low leakage mode because of the absolute minimum unutilized logic blocks or resources.

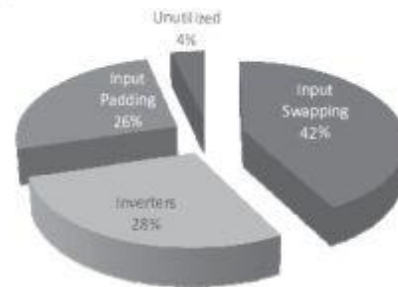


Figure 9: Leakage savings breakdown in logic blocks.

Figure 10 represents the RPR breakdown leakage power savings in the FPGAs. It can be understood from the figure that the average power savings of the unused routing resources is more than the unused logic resources. This is due to the percentage of unutilized routing resource is greater than that of the functional blocks. One more observation is that the leakage power savings in the inverters circuits is larger in the routing resources, which can be justified by the fact that the inverter circuits used in routing resources are of bigger in sizes than those used in the logic blocks. Hence the routing resources consume large leakage power [13, 14].

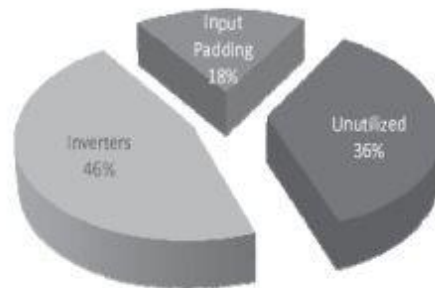


Figure 10: Leakage savings breakdown in the routing resources.

**V CONCLUSIONS**

This paper proposed several methodologies for leakage power reduction in modern nanometer FPGAs. The use of supply gating using Multi-Threshold CMOS (MTCMOS) techniques was proposed to enable turning OFF the unused resources of the FPGA, which are estimated to be close to 30% of the total FPGA area. Moreover, the utilized resources are allowed to enter a sleep mode dynamically during run-time depending on certain circuit conditions. Several new activity profiling techniques were proposed to identify the FPGA resources that will share common idleness periods, such that these resources will be turned OFF together.

Another technique proposed in this paper for leakage power reduction in FPGAs is the pin reordering algorithm.



The pin reordering algorithm makes use of the input state dependency of leakage power to place as much as possible of the FPGA circuits in a low leakage mode without incurring any physical or performance penalties. The guidelines for finding the lowest leakage power dissipation mode were derived and it was shown how they vary with every process node depending on the relative magnitude of sub-threshold and gate leakage power components. The proposed pin reordering technique was applied to several FPGA benchmarks and resulted in an average of 50% leakage power savings. Furthermore, another version of the proposed algorithm is also developed that results in a performance that was improved by 2.5%, while achieving an average leakage power reduction of 48%. This paper presented new CAD methods for the reduction of power dissipation in FPGAs.

### REFERENCES

- [1] S. Borkar, "Design Challenges of Technology Scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23-29, 1999.
- [2] A. Gayasen, Y. Tsai, N. Vijaykrishnan, M. Kandemir, M. J. Irwin, and T. Tuan, "Reducing leakage energy in FPGAs using region constrained placement," in *Proc. of ACM Intl. Symp. on Field Programmable Gate Arrays*, 2004, pp. 51-58.
- [3] J. Kao and A. Chandrakasan, "Dual-Threshold Voltage Techniques for Low Power Digital Circuits," *IEEE J. Solid-State Circuits*, vol. 35, no. 7, pp. 1009-1018, July 2000.
- [4] Betz, J. Rose, and A. Marquardt, "Architecture and CAD for Deep Submicron FPGAs. Norwell, MA: Kluwer Academic Publishers, 1999.
- [5] Z. Hu, A. Buyuktosunoglu, V. Srinivasan, V. Zyuban, H. Jacobson, and P. Bose, "Micro architectural Techniques for Power Gating of Execution Units," in *Proc. of Intl. Symp. on Low Power Electronics and Design*, 2004, pp. 32-37.
- [6] S. V. Kosonocky, M. Immediato, P. Cottrell, T. Hook, R. Mann, and J. Brown, "Enhanced Multi-Threshold (MTCMOS) Circuits using Variable Well Bias," in *Proc. of Intl. Symp. on Low Power Electronics and Design*, 2001, pp. 165-169.
- [7] H.-O. Kim, Y. Shin, H. Kim, and I. Eo, "Physical Design Methodology of Power Gating Circuits for Standard-Cell-Based Design," in *Proc. Of IEEE/ACM Design Automation Conf.*, 2006, pp. 109-112.
- [8] Calhoun, F. Honore, and A. Chandrakasan, "A Leakage Reduction Methodology for Distributed MTCMOS," *IEEE J. Solid-State Circuits*, vol. 39, no. 5, pp. 818-826, May 2004.
- [9] T. Tuan, S. Kao, A. Rahman, S. Das, and S. Trimberger, "A 90nm Low-Power FPGA for Battery-Powered Applications," in *Proc. of ACM Intl. Symp. on Field Programmable Gate Arrays*, 2006, pp. 3-11.
- [10] R. S. Guindi and F. N. Najm, "Design Techniques for Gate-Leakage Reduction in CMOS Circuits," in *Proc. of IEEE Intl. Symp. on Quality of Electronic Design*, 2003, pp. 61-65.
- [11] Marquardt, V. Betz, and J. Rose, "Timing-Driven Placement for FPGAs," in *Proc. of ACM Intl. Symp. on Field Programmable Gate Arrays*, 2000, pp. 203-213.
- [12] M. Anis, S. Areibi, and M. Elmasry, "Design and Optimization of Multi threshold CMOS (MTCMOS) Circuits," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 10, pp. 1324-1342, Oct. 2003.
- [13] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep Sub micrometer CMOS Circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305-327, Feb 2003.
- [14] J. Anderson, F. N. Najm, and T. Tuan, "Active Leakage Power Optimization for FPGAs," in *Proc. of ACM Intl. Symp. on Field Programmable Gate Arrays*, 2004, pp. 33-41.