International Journal of Modern Electronics and Communication Engineering (IJMECE) ISSN: 2321-2152
Volume No.-7, Issue No.-3, May, 2019

Power Optimization Techniques Targeting FPGA – A Survey

V.Prasanth
Department of Electronics and communication
Pragati engineering College
Surampalem, India
Prasanth.v@pragati.ac.in

Ch.satish
Department of Electronics and communication
Pragati engineering College
Surampalem, India
chembrolusatish@gmail.com

P.Vijay Gopal
Department of Electronics and communication
Pragati engineering College
Surampalem, India
Vijaygopal.p@pragati.ac.in

B.Prakruthi
Department of Electronics and communication
Pragati engineering College
Surampalem, India
prakruthi.79@gmail.com

Abstract: Silicon area, testability and performance are the major design constraints to be met during the development of digital circuits. In recent years, power consumption was given more importance compared to other design parameters. Field Programmable Gate Arrays (FPGAs) are highly desirable for implementation of digital system due to their flexibility, programmability and low end product cycle. In this paper, we presented an efficient methodology for dynamic and static power reduction of FPGAs based designs. Significant reduction in power consumption is achieved by handling both circuit design and optimization issues at circuit, logic, architecture and system level. Unfortunately, the advantages of FPGAs are offset in many cases by the high power consumption and area. This paper gives an overview of power optimization techniques to reduce the power consumption without sacrificing much performance or increasingly a large area so that FPGA applications can be increased

Keywords: Field Programmable gate arrays, power, optimization, low-power deign, power reduction.

I. INTRODUCTION

Field Programmable gate arrays (FPGAs) are Programmable logic devices that can be used to implement any digital circuit. FPGAs have become a popular technology in VLSI design due to its low manufacturing cost and fast design cycle. According to the current surveys, about 60-70% of power dissipation in FPGA device is due to dynamic power. High power consumption is a important concern factor for FPGA vendor and their users. Optimizing power is the key to lower cooling costs and improving device reliability. Studies on power consumption on Xilinx Virtex^{TM} devices reveal that interconnect power dominates the total power consumption due to their internal structure which consists of prefabricated wire segment with used and unused switches attached to each segment. This dominance of interconnect in total FPGA power consumption makes a high leverage target for power optimization [7]. Other dominant power is leakage power which is accounting for as much as 20%-25% of total active power at 85C in FPGAs implemented in 90nm technology. Low-power FPGA design is rapidly becoming an area of research interest. Software solutions for power reduction have been used to optimize FPGA power by performing power-aware technology mapping, placement, routing and look up tables (LUT) reprogramming [1]-[3]. In the area of hardware design low power techniques such as low-swing interconnect, heterogeneous interconnect, multi- V_t and mutli-V_{dd} , clock gating and fine grained power gating have been used to improve FPGA power consumption[1][3]-[10]. As multi-million-gate FPGAs became a reality, increasing design complexity and need to reduce the design time require early design decisions, especially for the FPGA customers because they care more about time-to-market [11]. As a result, we need to estimate the power consumption at a high level of abstraction, before the low level details of the circuit have been finalized . We presented survey on various power optimization methods to reduce power consumption.

II. POWER CONSUMPTION IN FPGA

Designing for the low-power devices is not a small task. Designers use several methods to reduce power in FPGA designs. Various types of FPGAs, different design methodologies, numerous intellectual property (IP) cores, assorted system design methods and diverse software algorithms and power tools all contribute to power used in a design[12]-[16].

The two primary sources of power consumption in FPGAs are: Static power dissipation due to leakage currents during device standby mode. Using better semiconductor process geometries has caused an increase in static power consumption in FPGAs. As transistor size shrinks and lower voltages are utilized, a greater sub-threshold leakage current occurs in the transistor channel when the transistor is in the off state.

Dynamic power dissipation during charging and discharging of internal capacitances in the logic array and interconnect networks of an active device. Dynamic power is affected in two ways by process scaling. First, the use of smaller feature sizes and lower voltages significantly reduces dynamic power consumption. As higher the device operating frequencies there is drastic increase in dynamic power.

As power consumption increase for smaller process geometries, semiconductor manufacturers use various

techniques to optimize power consumption, both in the dynamic and static domains

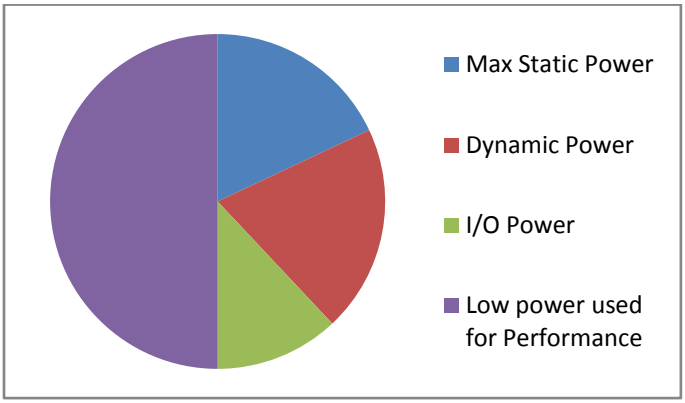

Figure 1: Break down of power on 28nm Xilinx Devices

Figure 1: shows the results on 28nm Xilinx devices where the Power optimizations method in FPGA has reduced Static, Dynamic and I/O Power to 50% of total power. The subsections shows the methodology and techniques adopted by FPGA vendor to optimize the power at different abstraction levels

.

A. *Dynamic power dissipation*

Dynamic power dissipation can be further subdivided into three mechanisms: switched, short circuit, and glitch power dissipation. All of them more or less depend on the activity, timing, output capacitance, and supply voltage of the circuit. The repeated charging and discharging of the output capacitance is necessary to transmit information in CMOS circuits. This charging and discharging causes for the switched power dissipation. The power consumption of a CMOS digital circuit can be represented as

$$P = CV_{dd}f + I_{Short}V_{dd} + I_{leak}V_{dd} \quad (1)$$

. Where f is the clock frequency, C is the average switched Capacitance per clock cycle, V_{dd} is the supply voltage, I_{short} is the short circuit current and I_{leak} is the leakage current. In a well optimized low power VLSI circuits, the first term of equation (1) is by far the dominant. The stand by power consumption is accounted for by the 3rd term. Using a lower V_{dd} is an effective way to reduce the dynamic power consumption since I[st] term is proportional to the square of V_{dd} .It should also be noted that the short circuit and leakage power dissipation are also strongly dependent on V_{dd}. However, using a lower V_{dd} degrade device performance

Because of programmability of FPGA the dynamic power is design-dependent and the factors that influence the design the effective capacitance of resources, the resources utilization, and the switching activity of resources. Since FPGA architecture usually provides more resources than required to implement a particular design, some resources are not used after chip configuration and they do not consume the dynamic power. Switching activity represents the average number of signal transitions in a clock cycle

B. *Static power dissipation*

The static power components become important when the circuits are at rest, i.e. when there is no activity in the circuits and they are all biased to a specific state. The static power dissipation includes sub threshold and reversed - biased diode leakage currents .Due to the necessary but risky, down - scaling of threshold voltages, the sub threshold leakage is becoming more and more prominent. Below the threshold voltage, in weak inversion, the transistors are not completely off.

1. Reverse biased pn junction current

Diode leakage occurs when a transistor is turned off and another active transistor charges up/down the drain with respect to the former's bulk potential. For Inverter example, The output is low, and the NMOS is on. The PMOS transistor will be turned off, but it's drain to bulk voltage will be V_{dd} since the output voltage is at 0V and the bulk for PMOS is at V_{dd} . For the p-well to bulk diode, the leakge current is given by

$$I_D = I_S(e^{V/V_T} - 1)$$

Where I_S is the reverse saturation current, V is the diode voltage, and V_T is the thermal voltage and is equal to KT/q . This current is especially significant for an application which spends much of its time idle, since this power is always being dissipated even when there is no switching.

2. Sub-threshold leakage

This occurs when the gate-source voltage, V_{gs}, has exceeded the weak inversion point but is still below the threshold voltage V_{th}. In this region, the MOSFET behaves similar to a bipolar transistor, with it's exponential characteristics. The current in the sub-threshold region is given by:

$$I_{SUB} = K\left(\frac{W}{L}\right)e^{\frac{V_{gs}-V_{th}}{nV_T}}(1 - e^{-V_{ds}/V_T})$$

Where n and K are technology paramaters, and V_{ds} is the drain-source voltage. Scaling down the supply voltage in CMOS requires also to scale down the threshold voltage, V_{th} , in order to maintain the performance of the scaled down logic. From the equation above, it becomes clear that the reduction of the threshold voltage increases the sub-threshold leakage current significantly. Sub-threshold leakage current along with reverse biased pn junction current are currently the most important components of leakage current.

3. *Gate induced drain leakage (GIDL)*

Gate induced drain leakage (GIDL) current (I_{GIDL}) arises in the high electric field under the gate/drain overlap region causing deep depletion. GIDL occurs at low V_G and high V_D and generates carriers into the the substrate and drain from surface traps or band-to-band tunneling.

4. Punch through

Punch through occurs when the drain and source depletion region approach each other and electrically "touch" deep in the channel. Punch through current (I_{PT}) varies quadratically with drain voltage.

5. Gate tunneling

Gate oxide tunneling current (I_G) is present when the electric field at the gate is high enough to tunnel through the gate oxide layer. This phenomenon is common in scaled down devices with reduced oxide thickness

C. Low-Power Design Techniques

An integrated low power methodology requires optimization at all design abstraction layers as mentioned below [13].

- System :Partitioning and Power down
- Algorithm l: Complexity, Concurrency, and Regularity.
- Architecture l: Parallelism, Pipelining, Redundancy, Data Encoding Circuit.
- Logic: Logic Styles, Energy Recovery, Transistor Sizing.
- Technology: Threshold Reduction, Multi threshold Devices.

Dynamic power varies as square of supply voltage. So reducing the supply voltage reduces power dissipation. Also selective frequency reduction technique can be used to reduce dynamic power. Multi threshold voltage can be used to reduce leakage power at system level [13]-[17]. Transistor resizing can be used to speed-up circuit and reduce power. Sleep transistors which we will discuss in following tutorials can be used effectively to reduce standby power. Parallelism and pipelining in system architecture can reduce power significantly. Clock disabling, power-down of selected logic blocks, adiabatic computing, software redesign to lower power dissipation are the other techniques commonly used for low power design..

III. POWER OPTIMIZATION TECHNIQUES

There are two primary types of power consumption in FPGAs : 1) Static Power and 2) Dynamic Power. The techniques that used to reduce these two types of power consumptions have been explored in following subsections.

A. Static power reduction

Many Circuit and Device level techniques have be evolved to reduce static power dissipation

a) Power Gating

As the transistor feature sizes and thresold voltages decrease, the standby power due to leakage current becomes compareble to dynamic power. Especially, in FPGAs, the standby power is serious problem because it has an enormously large number of transistor to achieve its programmaility. Power gating has emerged as the most effective design technique to achieve low standby power[20]. Power gating techniques are based on selectively setting the functional units into a low leakage mode when they are inactive.

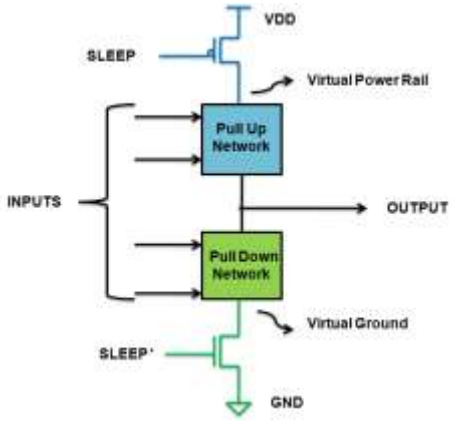

Figure 2: Power Gating

Power gating techniques are classified into two types: coarse-grain power gating and fine-grain power gating.

In coarse-grain power gating, a large number of lookup tables (LUTs) share a single sleep controller so the area and power overheads of the sleep controller are relatively small.

On the other hand Fine-grain power gating, each LUT has its own sleep transistor and related sleep controller, so when any LUT are inactive, they can be set to the sleep mode immediately. This result in much lower standby power compared to coarse-grain power gating. In Fine grain power gating, each LUT has its sleep controller, the number of the sleep controllers is much larger than that of coarse-grain power gating results in large area and dynamic power overheads.

Due to these overheads, fine-grain power gating is commonly assumed to be less efficient than coarse-grain power gating, although it has the potential to cut most of the standby power

b) Multi Voltage

power is proportional to the square of the supply voltage, lowering the supply voltage can reduce the power consumption significantly. But, lowering the supply voltage also increases the delay of the block. Blocks on the criticalpath are run at the highest voltage since its performance determines the system performance. The rest blocks can be run at a lower voltage without system performance degradation *[1]*. Thus, a large proportion of blocks can be run at the lower voltage. This approach can save the power consumption significantly.

Multi-voltage techniques can mainly be classified as follows:

Static Voltage Scaling (SVS): different blocks are given different, fixed supply voltages.

Dynamic Voltage and Frequency Scaling (DVFS): blocks are dynamically switched between two or more voltage levels follow changing workloads.

DVFS is a more efficient technique for power reduction.

c) Dual Threshold Voltage

In modern process technology, multiple threshold voltages are provided for each transistor. A dual V_T Process provides the

designer with transistors that are either fast(with a high leakage) or slow (with a low leakage) [11][4]. Therefore, a circuit can be portioned into high and low threshold voltage gates or transistors, which is a tradeoff between performance and reduced leakage currents.

Figure 3: the critical path within the circuit should be implemented with a low V_T to maximize the performance, whereas non-critical paths should be implemented with high V_T devices to minimize the leak age currents.

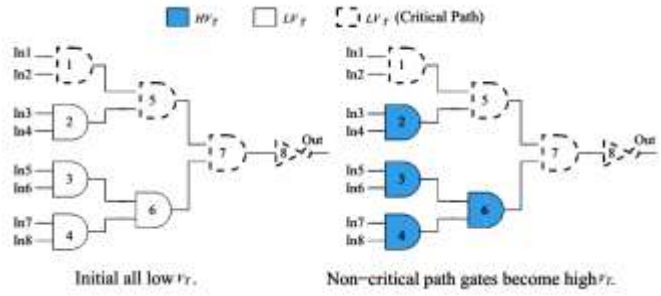

Figure 3: Dual V_T Partitioning scheme

As a result, the leak age currents are significantly reduced in both the standby and active modes, compared to an all low V_T implementation. At the same time, circuit performance is maintained at low supply voltages.

A limitation of this technique is that CAD tools need to be developed and integrated in to the design flow to optimize the partitioning process. A gate-level method for assigning dual threshold voltage is described in [10]. Another method, based on a transistor-lever for a dual V_T assignment, achieves a better leakage reduction because the individual transistors within the gates themselves are optimized to have multiple threshold voltage options[11][12].

A simultaneous input vector con trol and dual V_T assignment approach is proposed in [13] , whereas a dual V_T partitioning technique in conjunction with a transistor sizing algorithm is explored in [14][15] . Dual V_T technology is becoming increasingly attractive in modern advanced CMOS circuit designing because of its cost of an additional threshold voltage is relatively inexpensive for a process technology, requiring only an extra implant step.

d)MTCMOS Technology

A specialized case of dual V_T technology that is more effective at reducing leakage currents in the standby mode is MTCMOS (Multi-Threshold CMOS) first described in [17]. This technique involves using high V_T transistors to gate the power supplies of a low V_T logic block, and is described in more detail in the next chapter. MTCMOS gates are especially effective at reducing standby leakage currents because the leakage currents can be shut off with high V_T gates during the standby state, yet during the active state the internal logic can operate at high speeds through low V_T devices

Although the MTCMOS circuit technique is effective for controlling leakage currents in combinational logic, a drawback is that the technique can cause the internal nodes to float and result in the loss of the stored state for the memory units and the flip-flops

e)Body Biasing Techniques

With a body biasing methodology, and a suitable triple well technology, it is very straight forward to place a circuit block into a low leakage standby mode. Sub threshold leakage reduction is to utilize the body effect to directly change the threshold voltage of individual devices through body biasing. One simply needs to apply maximum reverse body bias to all the PMOS and NMOS devices, which will raise the device threshold voltages and lower the sub threshold leakage current exponentially. This approach to sub threshold leakage reduction does not use any multiple threshold devices or extra series power switches, but instead directly manipulates the intrinsic threshold voltages of the existing devices in the circuit. As a result, there are no sleep transistor sizing issues or gate partitioning tools that need to ensure correct operation during the active mode.The only modification to an existing CMOS design is that standby circuitry must be provided to supply appropriate bias voltages (greater than VCC and less than ground) to be the device body terminals.

B. Dynamic power reduction

Dynamic power is consumed by switching activity as a function of voltage, frequency, and capacitance and is dissipated through charging and discharging of capacitances during the operation of the circuit. The dynamic power consumption is generally modeled as below

$$P = \sum_i C_i V_i^2 f_i$$

Where C,V and f represent capacitance, the voltage swing, and clock frequency of the resource i, respectively[2].

The total dynamic power consumed by a device is the summation of the switching activity at each node. The factors which influence the dynamic power because of programmability of FPGA are: capacitance, the resources utilization, and the switching activity [2].

The Total capacitance corresponds to the sum of parasitic effects due to interconnection wires and transistors. Generally FPGA architecture uses minimum resources to implement a particular design, some resources are not used after chip configuration and they do not consume the dynamic power. Switching activity represents the average number of signal transitions in a clock cycle. Most dependent factor of switching activity clock itself, but sometimes it may also depend on other factors (e.g. temporal patterns of input signals). Hence, the above equation can be rewritten as

$$P = V^2 f \sum_i C_i U_i S_i$$

Where V is the supply voltage, f is the clock frequency, and C , U , and S , are the effective capacitance, the utilization, and the switching activity of each resource, respectively

The following techniques will help you reduce dynamic power:

- Decrease the average logic-switching frequency.
- Reduce the amount of logic switching at each clock edge.
- Reduce the propagation of the switching activity.

*a)**Switching Activity***

Although leakage is becoming more vital in new processes, dynamic power is still the main source of power dissipation. The dynamic power dissipated in an FPGA is directly proportional to switching activity, which is the average number of transitions of all nodes in the circuit per unit time. In addition, the impact that the accuracy of the switching activity estimation has on the ability to obtain accurate [9]. Power reduction in FPGAs requires both power-aware CAD tools and accurate power estimation models [2]. An accurate method of estimating switching activity is a key part.

FPGA power models require precise switching activity information to generate accurate and useful power estimates. In Design pipelining stages and flip-flops are increased to reduce the average switching activity, because flip-flops filter out glitches. FPGA power model with accurate switching activity information, the designer can simply compare each implementation and choose the one that dissipates the least amount of power. Power-aware FPGA CAD tools To reduce power the distance between logic elements is minimized and connected by high-activity wires.

The switching activity of a node has two components. One is functional transitions that occur when performing the computation. Other one is glitches, which are transitions that occur unnecessarily when the values at the inputs of a logic element arrive at different times. After placement and routing, detailed information about the arrival times of signals to each lookuptable is known.

*b)**Clcok Gating***

Power efficiency is one of the critical concerns in clock network design. The clock network has a significant impact on power since it connects to each flip-flop on the FPGA in a flexible way and toggles every clock cycle. The clock network in a current 28 nm FPGA accounts for 18% of dynamic power dissipation. Moreover, depending on how flexible it is, the clock network can impose constraints on the placement of logic within the FPGA; which affects power dissipation in other parts of the FPGA.

Clock gating is a way of reducing switching activity on circuit signals, and it entails temporarily disabling the clocking of specific registers when the outputs of those registers are inconsequential to circuit outputs. The more flexible the clock network, the more parasitic capacitance on the clock nets, and the more routing switches traversed by each clock signal; this leads to increased power dissipation.

The clock distribution network also has an impact on the ability of the computer-aided design (CAD) tools to minimize the power and maximize the clock frequency of a user circuit. FPGA clock networks typically are not flexible enough to supply any clock signal to any flipflop. This imposes additional constraints on the placement algorithm, as well as the clustering algorithm that groups logic elements into clusters. If the clock network is not flexible enough, these constraints could result in increased power dissipation and delay in a user circuit. Again, this balance must be considered by FPGA vendors as they design their clock distribution networks.

Clock Gating at Chip Level

The easy and popular system-level clock-gating stops the clock for an entire FPGA, effectively disabling all functionality. It prevents the logic from switching. Flash-based FPGAs use Freeze mode, which is more flexible than system-level clock gating, controlling input and output states and freezing clocks.

Clock Gating at Design Level

Clock gating at RTL level is a commonly used power -saving technique. There are two types of clock gating: latch-based and latch-free clock gating types. The latch-free clock gating style uses a simple AND gate. However, this imposes a requirement on the circuit that the gating signal be held constant from the active (rising) edge of the clock until the inactive (falling) edge of the clock to prevent clock glitches. The latch-based clock gating style adds a level-sensitive latch to the design to hold the enable signal from the active edge of the clock until the inactive edge of the clock, making it unnecessary for the circuit itself to enforce that requirement..

IV. CONCLUSION

Trends in technology scaling tend to change drastically which increases leakage power and also a steady increase in dynamic power with each successive process generation. Field programmable gate arrays (FPGAs) require considerable hardware overhead to offer programmability, making them less power-efficient than custom ASICs for implementing a given logic circuit. Power management in FPGAs will be mandatory to ensure correct functionality, provide high reliability, and to reduce packaging costs. This paper summarizes the different methods that have been carried out and various techniques used to reduce the power consumption of FPGAs. Still there are ways to reduce power in FPGA even after lot of significant improvements in FPA designs. Continued advances in low power software and hardware will open the door for FPGAs entering new power sensitive markets

REFERENCES

[1] Tim Tuan and Bocheng Lai, "Leakage power analysis of a 90nm FPGA," in Proceedings of the IEEE 2003 Custom Integrated Circuits Conference, 2003., 2003, pp. 57-60.

[2] Tim Tuan, Sean Kao, Arif Rahman, Satyaki Das, and Steve Trimberger, "A 90nm low-power FPGA for battery-powered applications," in Proceedings of the 2006 ACM/SIGDA 14th international symposium on Field programmable gate arrays, 2006, pp. 3-11.

[3] Li Shang, Alireza S. Kaviani, and Kusuma Bathala, "Dynamic power consumption in Virtex™-II FPGA family," in Proceedings of the 2002 ACM/SIGDA tenth international symposium on Field-programmable gate arrays, 2002, pp. 157-164.

[4] Jan M. Rabaey and Massoud Pedram, Low power design methodologies.: Springer Science & Business Media, 2012, vol. 336.

[5] Fei Li, Yan Lin, and Lei He, "FPGA power reduction using configurable dual-Vdd," in Proceedings of the 41st annual design automation conference, 2004, pp. 735-740.

[6] Fei Li, Deming Chen, Lei He, and Jason Cong, "Architecture evaluation for power-efficient FPGAs," in Proceedings of the 2003 ACM/SIGDA eleventh international symposium on Field programmable gate arrays, 2003, pp. 175-184.

[7] Julien Lamoureux and Steven J. E. Wilton, "On the interaction between power-aware FPGA CAD algorithms," in Proceedings of the 2003 IEEE/ACM international conference on Computer-aided design, 2003, p. 701.

[8] Julien Lamoureux and Steven J. E. Wilton, "Activity estimation for field-programmable gate arrays," in 2006 International Conference on Field Programmable Logic and Applications, 2006, pp. 1-8.

[9] Tanay Karnik et al., "Total power optimization by simultaneous dual-Vt allocation and device sizing in high performance microprocessors," in Proceedings of the 39th annual Design Automation Conference, 2002, pp. 486-491.

[10] James T. Kao, "Subthreshold leakage control techniques for low power digital circuits," Massachusetts Institute of Technology, Department of Electrical Engineering …, Ph.D. dissertation 2001.

[11] Shota Ishihara, Masanori Hariyama, and Michitaka Kameyama, "A low-power FPGA based on autonomous fine-grain power gating," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 19, pp. 1394-1406, 2010.

[12] Varghese George, Hui Zhang, and Jan Rabaey, "The design of a low energy FPGA," in Proceedings. 1999 International Symposium on Low Power Electronics and Design (Cat. No. 99TH8477), 1999, pp. 188-193.

[13] Vijay Degalahal and Tim Tuan, "Methodology for high level estimation of FPGA power consumption," in Proceedings of the 2005 Asia and South Pacific Design Automation Conference, 2005, pp. 657-660.

[14] Derek Curd, "Power consumption in 65 nm FPGAs," Xilinx white paper, vol. 711, 2007.

[15] Jason Helge Anderson and Farid N. Najm, "Power-aware technology mapping for LUT-based FPGAs," in 2002 IEEE International Conference on Field-Programmable Technology, 2002.(FPT). Proceedings., 2002, pp. 211-218.

[16] Jason Helge Anderson and Farid N. Najm, "Low-power programmable routing circuitry for FPGAs," in IEEE/ACM International Conference on Computer Aided Design, 2004. ICCAD-2004., 2004, pp. 602-609.

[17] Jason Helge Anderson and Farid N. Najm, "Active leakage power optimization for FPGAs," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 25, pp. 423-437, 2006.

AUTHOR'S BIOGRAPHIES

V.Prasanth is a Associate professor of Electronics and communication engineering in Pragati Engineering College, Andhra Pradesh, India. He has 14 year of Teaching Experience. He is a Ph.D Scholar of JNTUK, Kakinada, Andhra Pradesh, India. He completed Masters in VLSI System Design. His area of Interest is Low power VLSI and Embedded Systems. He published total of 14 Papers.

Ch.Satish is a Assistant professor of Electronics and communication engineering in Pragati Engineering College, Andhra Pradesh, India. He has 8 year of Teaching Experience. He completed Masters in Embedded Systems. His area of Interest is Signal Processing and Embedded Systems. He published total of 1 Papers.

P.Vijay Gopal is a Assistant professor of Electronics and communication engineering in Pragati Engineering College, Andhra Pradesh, India. He has 6 year of Teaching Experience. He completed Masters in Communications & Signal Processing. His area of Interest is Audio & Video Processing, Communication Systems ad DSP Processors.

B.Prakruthi is a Assistant professor of Electronics and communication engineering in Pragati Engineering College, Andhra Pradesh, India. She has 2 years of Teaching Experience. She completed Masters in Digital Electronics and Communication Systems. Her area of Interest is Image Processing and Communication Systems.