Total Power Consumption in Modern VLSI Circuits

Bojan B. Jovanović and Milun Jevtić[1]

Abstract – **In this paper various design methods applicable on different stages of system design and aimed to minimize the power consumed by the system are presented. Along with the overview of the power minimization techniques, as an example, the circuit of binary divider is implemented in various FPGAs to demonstrate technology as well as placement and routing influence on total power consumption.**

Keywords – **Power consumption, Power minimization techniques, binary dividers.**

I. INTRODUCTION

During the past years the progress of silicon process technology marches on relentlessly. According to Moore's law, silicon process technology continues improvement at an astonishing pace. Every 2 years the number of transistors that can be integrated on a single IC approximately doubles. Despite a strong relationship between the consumed power and the performance of CMOS circuits, the power of early circuits remained within the allowable power envelope due to various heat dissipation techniques. In this scenario, designers focused primarily on achieving the needed performance. However, as the density and size of the chips and systems continue to increase, the difficulty in providing adequate cooling might either add significant cost to the system or provide a limit on the amount of functionality. Also, having in mind many high-volume battery-powered portable and wireless consumer devices, power minimization should be treated very carefully. Starting from 0.18μm technologies, static power consumption due to leaky "off" transistors, is now a non negligible source of power dissipation even in running mode. Thus, the total power consumption (i.e. dynamic plus static power) has to be optimized instead of simply reducing dynamic power. Design methods that explore true power optimization need to work in a large dimension search space, where power and performance of different solutions are compared. This includes system architecture optimization (outer loop), block-level optimization (intermediate loop), and fixed topology optimization (inner loop).

II. THE SOURCES OF POWER CONSUMPTION

The two main sources of power dissipation in CMOS VLSI's are the dynamic power dissipation due to charging and discharging of load capacitance, and the static power dissipation due to subthreshold leakage. There may be short-circuit power dissipation (V_{DD} to ground) as the third source of power dissipation. This power source is due to non-zero rise and fall time of input waveforms and it is less than 10% in total power dissipation.

The expression for dynamic power consumption is widely known. It depends on squared power supply, V_{DD}, operating frequency, load capacitance of the node, and the average number of $0 \rightarrow 1$ transitions within one clock cycle.

Leakage currents consists of two main components: subthreshold leakage (I_2), and gate leakage current (I_3) as shown in Fig. 1. There are some other leakage current components that have started to gain interest recently due to an excessive scaling of the transistor dimensions. They occur due to the short channel-lenght (I_4, I_6), the thinner oxide thickness (I_5), and due to high doping concetracions (I_1) [1].

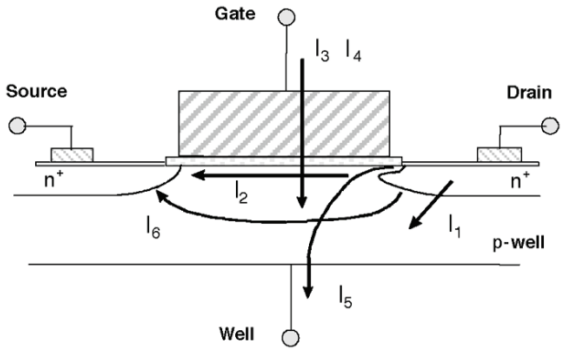

Fig. 1. Leakage current components

However, the largest amount of static power is still owed to subthreshold leakage current. It is the most temperature-dependent leakage component, and thus, every increase in dynamic power produces an increment of the chip temperature, which in turn, increase the leakage current. This leakage component is also one of the main reasons why the scaling process is facing difficulties. Four tunneling mechanisms (the gate to channel, bulk, source, and drain) as well as analytical expressions for gate leakage current can be found in [1,2].

III. FIXED TOPOLOGY OPTIMIZATION

Power optimization techniques in this level do not alter the circuit topology, so the principle variables they affect are transistor sizes, supply voltages, and the threshold voltages. Some of the authors investigate the impact of single variable on circuit power consumption and delay while other perform thorough analysis considering mutual influence of two or even more variables on design power consumption. There are a few commonly used power minimization techniques: gate-sizing, variable supply-voltage, variable threshold-voltage, multi voltage design, power gating, clock gating, stack forcing, on-chip optical interconnect, nano devices etc.

[1]Bojan B. Jovanović and Milun Jevtić are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia,
E-mail: {bojan.jovanovic, milun.jevtic}@elfak.ni.ac.rs.

A. Gate Sizing and Supply Voltages in Power Minimization

Decreasing transistor sizes enables higher densities of transistors on a chip. In order to control the power of the circuit, the power supply voltage is also reduced with each transistor scaling. Due to quadratic relationship between dynamic power consumption and power supply, this supply voltage reduction is the most effective way to lower the dynamic power. For CMOS circuits, a lower supply voltage means lower performance. This problem is solved by reducing the threshold voltage (V_{TH}) of a transistor. V_{TH} is defined as a gate-source voltage of MOSFET transistor, above which, the transistor is turned on. Ideally, if the gate voltage is below the threshold voltage, the transistor is not conducting any current. However, in practice there is still some current flowing from the drain to the source of a transistor. This is the subthreshold current. Its most important feature is that it increase exponentially with any V_{TH} decrease. That's why this leakage current is one of the main limiting factors to scaling process. SIA Roadmap [3] forecast supply voltage as low as 0.8 to 0.5 V in year 2018. Predicted threshold voltages are up to 0.1 V.

It is sure that between all the combinations of V_{DD}/V_{TH} guaranteeing the desired speed only one will result in the lowest power consumption. The location of this optimal working point and its associated total power consumption are tightly related to architectural and technology parameters. Authors in [4] give the equation of total power consumption for circuits working at their optimal supply and threshold voltages, while in [5] authors present closed-form formula for optimum supply and threshold voltages that minimize power dissipation when technology parameters and required speed are given.

The techniques of variable supply or/and threshold voltages continuously adapt their values in the design (during the design run on) so the power dissipation is reduced and the critical-path delay is not changed. Instead of using variable voltages power minimization techniques with dual or multiple V_{DD} or/and V_{TH} can be applied. The gates on critical paths operate at higher V_{DD} or lower V_{TH}, while those on non-critical paths operate at the lower V_{DD} or higher V_{TH}, thereby reducing overall power consumption without performance degradation.

Among the leakage power reduction techniques power gating is commonly used to disconnect idle logic blocks from power network to curtail sub-threshold leakage [6]. The similar clock gating technique is used to prevent clock signal to give a pace to non-active gates. Stack forcing is another technique to tackle the ever-increasing leakage power. It has been shown that stacking of two off transistors can significantly reduce leakage power than a single one [7]. In 90nm technologies and beyond gate oxide leakage current has become comparable to subthreshold leakage. It is, therefore, necessary to develop methods for oxide leakage reduction, which unlike subthreshold leakage, occurs only in transistors that are ON as shown in Fig. 2. Increasing the oxide thickness will decrease gate oxide leakage current but this will be payed with substantial transistor delay. So, thick-oxide transistors in non-critical path will not speed down the circuit but will reduce static power consumption.

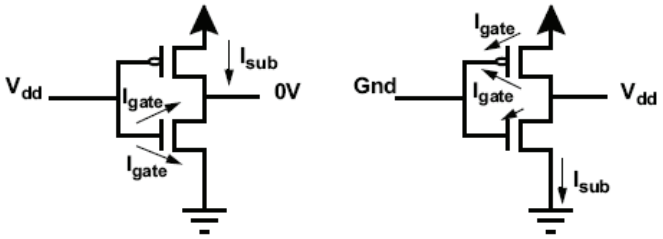

Fig. 2. Gate oxide leakage in CMOS inverter

B. On-chip Optical Interconnections for Power Minimization

Global interconnect performance required for future generations of ICs cannot be achieved with metal. Using optical instead of electrical interconnections lead to decrease in power consumption, enormous bandwidth increase, immunity to electromagnetic noise, and reduced sensitivity to temperature variations. However, there are some difficulties in obtaining a large enough optical-electrical conversion efficiency [8].

Another power minimization technique is on-chip wavelength division multiplexing. For example, a single waveguide could be used to replace a 64-bit bus, where each individual signal makes use of a distinct wavelength.

C. Technology Influence on Total Power Consumption

In order to practically demonstrate the influence on design's technology parameters on total power consumption the 12-bit binary divider logic circuit is described in VHDL and implemented in Xilinx FPGA devices from different families (Virtex-4, Virtex-5, Virtex-6, and Virtex-6 Lower Power). For binary division Radix-2 non-restoring algorithm with non-fractional remainder is used [9].

XPower CAD tool (within ISE 12.4 software) was utilized for power consumption measurements. Divider inputs were generated in MATLAB as signals with Gaussian distribution (1000 values for both dividend and divisor). Mean value, auto-correlation and cross-correlation of these signals are all equal to zero. Implementation results are presented in Fig. 3.

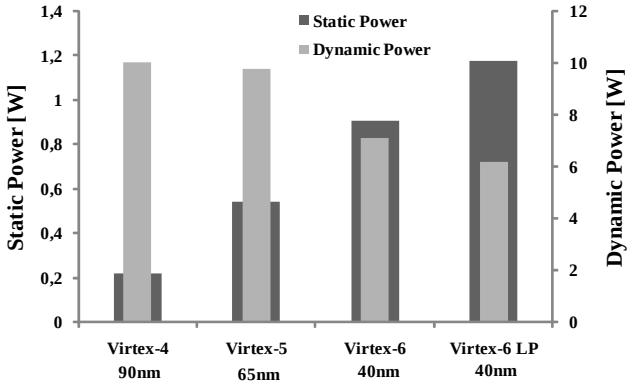

Fig. 3. Power consumption of binary dividers implemented in FPGAs

Technology process variations (from 90nm to 40nm), supply voltage variations (from 1.2 to 0.9 V), threshold voltage and transistor sizes variations obviously influence

both static and dynamic power consumption. There is a clear increase in static power consumption when moving toward newer generation FPGA families. Static consumption increases due to increase in leakage currents as a consequence of shrink in transistor sizes. The shorter channel lengths and thinner gate oxides generally used at the new process node make it easier for current to leak, either across the channel region or through the gate oxide of the transistor. Concerning dynamic power, the core FPGA supply voltage and node capacitance generally reduce with each new process node, providing substantial dynamic power savings over previous generation FPGAs. The new 28nm Xilinx-7 series FPGAs will enable a 50% overall power reduction compared to previous 40nm generation [10].

IV. Placement and Routing Influence on Power Consumption

Mandatory part of each integrated circuit (ASIC, FPGA, custom IC etc.) design flow is the process of physical implementation. This process is usually iterative and divided into several stages. In each iteration, design must be firstly partitioned into groups or blocks small enough to fit into a single unit (ASIC standard cells, FPGA configurable logic blocks etc). Secondly, these units are assigned specific location on the chip. This stage is usually called placement. Finally, placed blocks are interconnected by wires. Assigning paths, or routes, to the wires is usually done in two stages. After rough or global wiring, in detailed wiring (also called exact embedding) each wire is given a unique complete path. In the case of ASIC or custom IC design, from the detailed wiring results, masks can be generated and chips fabricated. At each stage of physical implementation process one tends to optimize the eventual performance of the system (to minimize chip area, power consumption and delay) without compromising the feasibility of the subsequent design stages. The major focus in placement and routing (PAR) is on minimizing the length of interconnections since this translates into the time required for propagation of signals and thus into the speed of the entire design. Also, minimal interconnection length leads to less capacitive interconnects and consequently to decrease in dynamic power consumption. However, the presence of regions in which the wiring is too congested for the packaging technology should be anticipated and minimized during the PAR process. Besides, the sources of noise such as crosstalk between adjacent wires should be eliminated.

From the above mentioned one can conclude that PAR are very important phases in system design influencing overall system performance. Furthermore, placement and routing phases have both equal influences on the system features. Consequently, for the best results, both place and route phases need to be considered since the benefits from high quality placement might impose low quality routing decreasing overall system performances [11].

In theory, PAR problem is an NP-complete problem and require iterative algorithms capable of efficiently searching for a near optimal solution in a large solution space. Some of widely used ones are generic arithmetic, tabu search and simulated annealing [12]. Many authors investigate power-aware placement and routing. Vorwerk et al. introduce two techniques for minimizing power during the FPGA placement [13]. The first aspect discussed in their work is a power-aware objective function for placement. In particular, a capacitance model for global nets allowing net power reduction is described. The second technique permits area and power reduction by optimizing the number of combinational and sequential cells. The results are quantified across a suite of 119 industrial benchmarks targeting Actel Igloo FPGA architecture. Power is reduced by 13% on average with a 6.7% average improvement in timing performance across the suite. Cheon et al. [14] present power-aware placement method that simultaneously performs activity-based register clustering for clock power reduction and activity-based net weighting for net switching power reduction. In [15] Xilinx's engineers consider dynamic power dissipation and present CAD techniques for dynamic power reduction in Xilinx Virtex-5 FPGA. The proposed techniques, comprising power-aware placement, routing, and a novel post-routing transformation are applied to optimize the industrial designs power consumption. Board level measurements show that the techniques reduce power by 10%, on average. There are also researches on leakage-aware PAR. Gupta et al. [16] introduce LEAF – a novel tool for leakage-aware PAR for SoCs. Up to 190% difference in the leakage power between leakage-unaware and leakage aware PAR is observed. The similar tool called TPAP intended for FPGAs is presented in [17]. It is also known that temperature variations across a chip (thermal non-uniformities) are an issue that threatens chip performance and reliability. The correlation between the total power consumption and the temperature variations across a chip is investigated in [18]. As a result, PAR guidelines are proposed that uses the correlation to efficiently optimize the chip's total power. It is demonstrated that optimizing a floorplan to minimize either the leakage or the peak temperature can lead to a significant increase in the total power consumption. The experimental results show that lowering the temperature variations across a chip not only addresses performance degradation and reliability concerns, but also significantly contributes to chip power reduction.

To demonstrate PAR influence on overall system performance already mentined 12-bit binary divider circuitry is implemented in Xilinx Virtex-5 FPGA device. In each implementation of the same logic design the only changing parameter was PAR design goal. Three different design goals were Area Reduction, Power Optimization and Timing Performance. Xilinx PAR tool runs in ten iterations and at different effort levels (1-5). This effort levels indicate the amount of time the tool spends searching for a better quality solution. In the fourth case Xilinx IP core divider (using the same division algorithm) [19] was implemented. Fig. 4 shows implementation results.

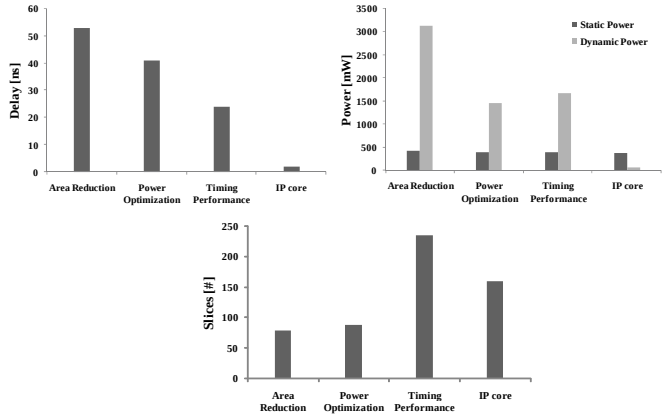

Fig. 4. Delay, power consumption and size of the same design with the different PAR goals

Implementation of Xilinx IP core divider ensures the best design performances. It is expected, having in mind utilization of hard IP blocks (circuitry dedicated to commonly used functions) which contain only the transistors necessary to implement the required function. Furthermore, there are no programmable interconnects, so routing capacitance is as small as possible. On the other hand, three remaining designs are implemented in the general-purpose FPGA logic. Changing of the PAR objective function can lead to obvious design performance variation. So, speed and dynamic power fluctuations are about 55 % while design size varies about 65 % (compared to the worst case). Static power consumption insignificantly changes.

V. CONCLUSION

Some power-aware system design methods are presented in this paper. These methods take into account static power consumption as more and more dramatic issue in very deep submicron technologies. The influence of technology parameters on design power consumption is demonstrated through implementation of binary dividers in FPGAs with the different technological properties. With the improvements of the technology parameters, there is an obvious trend in decreasing dynamic and increasing static power consumption of the design. Various placement and routing design goals during the design implementation also influence design power consumption.

ACKNOWLEDGEMENT

This paper is supported by Project Grant III44004 (2011-2014) financed by Ministry of Education and Science, Republic of Serbia.

REFERENCES

[1] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-submicrometer CMOS circuits", in Proc. of the IEEE, vol. 91, no. 2, pp. 305-327, 2003.
[2] R. Inagaki, N. Sadachika, D. Navarro, M.M. Mattasch, and Y. Inoue, "A Gate-Current Model for Advanced MOSFET Technologies Implemented in HiSIM2", ICCCAS, pp. 1057-106, Kokura, Japan, 2007.
[3] http://public.irts.net
[4] C. Schuster et al.,"Architectural and Technology influence on the Optimal Total Power Consumption", DATE, pp. 13-19, Munich, Germany, 2006.
[5] K. Nose, T. Sakurai, "Optimization of V_{DD} and V_{TH} for Low-Power and High-Speed Applications", ASP-DAC, pp. 469-474, Yokohama, Japan, 2000.
[6] S.Y. Chen et al.,"Power gating design for standard-cell-like structured ASICs", DATE, pp. 514–519, Germany 2010.
[7] S. Narendra et al.,"Scaling of Stack Effect and its Application for Leakage Reduction", ISLPED, pp. 195–200, USA, 2001.
[8] C. Piguet et al.,"Extremely Low-Power Logic", DATE, pp. 656–661, Paris, France 2004.
[9] R. Jevtić, B. Jovanović, C. Carreras,"Power Estimation of Dividers Implemented in FPGAs", GLSVLSI, Lausanne, 2011.
[10] Xcell Journal, Fourth Quarter 2010, Issue 73, pp. 65, www.xilinx.com/xcell
[11] C. Mulpuri, S. Hauck, "Runtime and Quality Tradeoffs in FPGA Placement and Routing", ASM/SIGDA, pp. 29–36, San Jose, USA, 2001.
[12] M. R. Minhas, "Iterative Algorithms for Timing and Low-Power Driven VLSI Standard Cell Placement", Master's thesis, Department of Computer Engineering, King Fahd University of Petroleum and Minerals, Saudi Arabia, Dhahran 31261, 2001.
[13] K. Vorwerk et al., "Power minimization during field programmable gate array placement", IET Journ. on Computers&Digital Techn., vol. 4, no. 3, pp. 170–183, 2010.
[14] Y. Cheon et al., "Power-aware placement", ASP-DAC, pp. 795–800, Shanghai, China, 2005.
[15] S. Gupta, J. Anderson, L. Farragher, Q. Wang,"CAD Techniques for Power Optimization in Virtex-5 FPGAs", CICC pp. 85–88, San Jose, USA, 2007.
[16] A. Gupta et al., "LEAF: A System Level Leakage-Aware Floorplanner for SoC", ASP-DAC, pp. 294–297, Yokohama, Japan, 2007.
[17] B. Salami, "Chip-dependent leakage Power-Aware Placement algorithm for FPGAs", ELECO, pp. 350–360, Bursa, 2010.
[18] K. Haghdad, M. Anis, Y. Ismail, "Floorplanning for low power IC design considering temperature variations", Microelectronics Journal, vol. 42, no. 1, pp. 89–95, 2011.
[19] Xilinx IP Core Divider Datasheet, www.xilinx.com, Jan. 2006