A Survey on Power Management Techniques

Padmini G.Kaushik, Sanjay M.Gulhane, Athar Ravish Khan

Abstract — These Clock-gating and power-gating have proven to be very effective solutions for reducing dynamic and static power, respectively. The two techniques may be coupled in such a way that the clock-gating information can be used to drive the control signal of the power-gating circuitry, thus providing additional leakage minimization conditions w.r.t. those manually inserted by the designer. This conceptual integration, however, poses several challenges when moved to industrial design flows. Although both clock and power-gating are supported by most commercial synthesis tools, their combined implementation requires some flexibility in the back-end tools that is not currently available. This paper presents the survey on integration of clock and power gating schemes for power management.

Index Terms— Clock gating, Dynamic power, FPGA, Power gating, Power management, Static power, Sub-threshold leakage.

1 INTRODUCTION

The increasing demand for low power mobile computing and consumer electronics products has refocused VLSI design in the last two decades on lowering power and increasing energy efficiency. Power reduction is treated at all design levels of VLSI chips, from architecture through block and logic levels, down to gate-level, circuit and physical implementation. One of the major dynamic power consumers is the system’s clock signal, typically responsible for up to 50% of the total dynamic power consumption [31].

Clock network design is a delicate procedure, and is therefore done in a very conservative manner under worst case assumptions. It incorporates many diverse aspects such as selection of sequential elements, controlling the clock skew, and decisions on the topology and physical implementation of the clock distribution network. Advances in CMOS technology have made possible the integration of millions of transistors onto a small area. CMOS technology scaling allowed the reduction of the gate delay and the increasing the operating frequency, the increase of the transistor density and, finally, the decrease of the energy per transistor. Technology trends show that, in every circuit generation, delay and supply voltage have scaled down by 30%, performance and transistor density have doubled every two years and, finally, transistor’s threshold voltage has reduced by almost 15% [31],[34].

In this scenario, power consumption represents one of the most important concerns in modern systems, since the battery power increases by about 15% per year and the chip power requirements by about 35% [5]. For CMOS technologies down to 130nm, dynamic power was considered as the main source of circuit’s power consumption. However, due to the CMOS scaling process, leakage power dissipation has become a significant contributor to the total power. Leakage power is dissipated during both active and standby mode. Different physical phenomena can contribute to the leakage currents causing static consumption when one or more transistors are in a steady state. The four main components of leakage current are sub-threshold leakage ($I_{SUB}$), gate leakage ($I_{GATE}$), gate induce drain leakage ($I_{GIDL}$) and reverse bias junction leakage ($I_{JAB}$).

However, in today’s technologies, the major sources of leakage consumption are represented by the sub-threshold and gate currents, with the former that greatly dominates. For some classes of circuits manufactured with the 65nm process, dynamic and leakage power are equally important. And the situation is expected to worsen further as technology scaling will continue. In the last few years, several techniques have been proposed in order to reduce the main sources associated to the power consumption in CMOS circuits [7]. Regarding dynamic power consumption[6],[10], the clock distribution network represents around 40% of the total power budget of a CMOS circuit, since the clock nets operate at the highest switching frequency compared to any other signal and they drive a large capacitive load. A number of examples of handcrafted power-management schemes can be found in the literature [44] and are actually implemented in well-known commercial devices. Automatic detection of good power management opportunities and synthesis of the required circuitry is a relatively new discipline.

The key idea in power management is that unused parts of a complex design can be shut down during system operation. Shut-down can be achieved by zeroing the voltage supply or, in static synchronous CMOS logic, by stopping the clock. From a technology standpoint, supply shutdown has several serious drawbacks, such as current spikes on power and ground lines, loss of information stored in memory cells, and power-up delays. In this context, Clock Gating (CG) [1,3,4,6,7] is considered as one of the most effective techniques to reduce dynamic power. Clock-gating has a reduced technological impact, and can be successfully employed in a wider range of cases. Regarding leakage power, the main contributor to overall leakage can be considered the drain-to-source leakage through transistors that are nominally off, even though a number of new leakage sources are emerging as transistors do shrink (e.g., gate tunneling currents, junction reverse currents).

In this scenario, Power Gating (PG) represents one of the most popular approaches for leakage power reduction and relies on the introduction of sleep transistors which are placed between the logic circuit and the ground rail (i.e., gnd), thus creating an...
intermediate virtual rail (i.e., virtual gnd)[5], [8], [2],[14],[27]. An even lower-impact approach to power management requires the use of enabled flip-flops, sequential primitives where an additional input signal determines whether the stored values must be updated or held constant. Supply shutdown, clock-gating and flip-flop disabling; span the trade-off between achievable savings and technology complications. On one extreme, power supply turn-off is the hardest to implement, but it completely eliminates power dissipation during shut-down. On the other extreme, flip-flop disabling is minimally intrusive, but it does not reduce clock power and leakage power.

We exploit clock-gating and power gating for implementing our power management strategy because we believe that it represents a good trade-off between aggressiveness and achievable power savings.

2 REVIEW OF LITERATURE

Clock gating is an effective method of reducing power dissipation of a high-performance circuit. However, deployment of gated cells increases the difficulty of optimizing a clock tree. Clock network design is a delicate procedure, and is therefore done in a very conservative manner under worst case assumptions [3],[4],[6],[44],[43]. It incorporates many diverse aspects such as selection of sequential elements, controlling the clock skew, and decisions on the topology and physical implementation of the clock distribution network[12]. Clock gating is employed at all levels: system architecture, block design, logic design, and gates [6]. Clock enabling signals are usually introduced by designers during the system and block design phases, where the interdependencies of the various functions are well understood. In contrast, it is very difficult to define such signals at the gate level, especially in control logic, since the interdependencies among the states of various flip-flops (FFs) depend on automatically synthesized logic. Mathematically the system can modeled as a probabilistic model of the clock gating network that allows us to quantify the expected power savings and the implied overhead [6, 27]. Expressions for the power savings in a gated clock tree are presented and the optimal gater fan-out is derived, based on flip-flops toggling probabilities and process technology parameters [6].

Author proposes a delay-matching approach to addressing this problem in [1]. Delay-matching uses gates cells whose timing characteristics are similar to that of their clock buffer (inverter) counterparts. It attains better slew and much smaller latency with comparable clock skew and less area when compared to type-matching[29]. The skew of a delay-matching gated tree, just like the one generated by type-matching, is insensitive to process and operating corner variations [1]. With the technology scaling to 65nm and below, leakage power is becoming a dominant component of the total power dissipation. Leakage, however, can be a significant drain of power during active mode[33], [27], i.e. when the digital circuit is doing useful work [8]. The leakage power can be reduced using power-gating (PG) techniques, based on the insertion of a power switch with high-V_{th} and thick T_{ox} between the logic and the ground, in order to electrically isolate the circuit in the standby mode.

In [8] a new technique, called sub-clock power gating, for reducing leakage power in digital circuits is presented. This technique works concurrently with voltage and frequency scaling and power reduction is achieved by power gating within the clock cycle during active mode unlike traditional power gating which is applied during idle mode.

For FPGA based systems, to cut the power consumption of clock network and detect the activity of the cell efficiently, asynchronous architecture is proposed in [7],[17].

Power gating technique is usually driven by a predictive control, and frequent mis-predictions can counter-productively lead to a large increase in energy consumption. This energy vulnerability could be exploited by malicious applications such as a power virus, or it may be exposed by regular applications containing repetitive mis-predictions patterns. In [21] author proposes a technique to counteracting this vulnerability by using a guard mechanism to prevent power overruns.

The basic idea is to reduce the timing overhead while integrating clock gating (CG) to reduce the dynamic power and power gating (PG) for the reduction of leakage power, an exhaustive design space exploration of the cost-function is carried out. In this way, it is possible to understand the relationship that may exist between the selected variables as a function of the energy-delay product. Based on the results of the exploration, a novel concept of critical-interconnects is introduced and exploited for reducing the timing overhead of the CG/PG designs [14].

Another method which combines CG and PG, in order to achieve leakage and dynamic power savings as well as to reduce the overhead associated to the insertion of control logic by its sharing. Ideally, the same control signal could be used for managing CG and PG, but in practice the two signals are characterized by different timing behaviors and consequently combining the two approaches may result even infeasible as the overlapping of the gating conditions may tend to become null. In order to evaluate the effective feasibility of the proposed approach, author implemented an analysis tool which is able to provide a suitable evaluation of an RTL design in order to determine whether CG and PG integration is convenient or not. Assuming a circuit featuring CG, the tool is able to verify whether the insertion of the sleep transistors necessary to activate the PG strategy may lead to an overall reduction of both types of circuit power.

A recent work in [16] suggests that dynamic power in FPGAs is 7 to 14 times higher than in custom ASICs for implementing a given circuit. Large strides are needed on the power front to enable deployment of programmable logic in the handheld battery powered devices pervasive in society today. Research on FPGA architecture goes hand-in-hand with research on computer-aided design (CAD) tools, as proper architectural evaluation demands CAD tools that leverage architectural enhancements. A placement approach presented in [16] is to reduce routing capacitance on the clock network, and placing design objects in a manner that takes advantage of the proposed clock gating architectures [16].
3 LIMITATIONS AND SCOPE

Literature survey given in section-2, gives brief idea on the techniques which are proposed for dynamic and leakage power reductions. Though the proposed work in the literature addressed various issues in power reduction they suffer from one or the other limitations. A delay-matching approach is presented in [1] to optimize a gated clock tree for power reduction. Delay-matching achieves smaller latency and better slew but requires larger area and skew. In addition this method requires reengineering a cell library to include delay-matching cells, larger area and skew. In [6] a probabilistic model of the clock gating network has been proposed that allows quantifying the expected power savings and the implied overhead. Expressions for the power savings in a gated clock tree are presented and the optimal gate fan-out is derived, based on flip-flops toggling probabilities and process technology parameters. The question of how to combine FFs into groups whose optimal size is known has also been formalized. Though the logic aspect of FFs toggling correlations is clear, the difficult question of how to account for layout considerations in such grouping needs further study. Reference [14] presents a solution for reducing the timing overhead that may occur when the integration of clock gating and power gating is performed. In particular, a new multilevel partitioning heuristic that increases the efficiency of the clustering phase is proposed. Although promising from the point of view of the achievable results, the combination of clock-gating and power-gating is not mature enough to be applicable to real-life circuits. In particular, several aspects related to the support of integrated CG/PG in EDA frameworks are still open. Reference [16] presents a placement algorithm that incorporates clock power and locates clock loads accordingly to take advantage of clock gating architectural enhancements. However, incorporating the clock gating aware placement algorithm into a complete power-aware FPGA CAD flow, allowing trade-offs between other power-aware tools phases need to be explored. Also modeling the clock gating to permit extraction of accurate post-routing critical path delay and interconnect capacitance is to be addressed.

A novel wasting-toggle-rate (WTR) based clock power reduction technique is introduced in [24]. It compares the WTR of a stimulus with the pre-computed threshold WTR of the circuit, and clock gating is applied to the circuit only if there is power-saving benefit. This methodology can be extended to stimulus independent methodology by using activity pre-computation method like probabilistic activity analysis. By exploiting the flexibility between the clock-gating conditions and the next state function, an iterative optimization technique to minimize the overall timing is proposed in [26]. If critical paths of a circuit are located in the paths to primary outputs, timing optimization technique will not be efficient.

Reference [31] presents an analysis methodology and a prototype CAD tool that support the designer in understanding when the joint application of Clock Gating and Power Gating may result in significant power savings. However, issues such as the size of the logic implementing the activation function and the duration of the idle periods affect the effectiveness of the proposed technique. Reference [36] is concerned with minimization of total power and energy when power gating is applied to circuit blocks that alternate between idle mode and active operation. The reference control circuit fails to prevent the overdrive voltage from exceeding the maximum limits and, thus, this generator has to be designed with a safeguard (i.e. it has to be resized) so that the overdrive voltage does not exceed the maximum limits at the worst-case process corner. In [38], the conventional square wave clock signal is replaced by a sinusoidal clock generated by a resonant circuit. Such a modification in clock signal prevents application of existing clock gating solutions. As opposed to square wave clocking, the clock gating cannot be implemented by insertion of masking logic gates at any arbitrary node on the clock network. That is because insertion of such logic gates on a sinusoidal clock network destroys the shape of the clock and eliminated the energy recovery property in the downstream fan-out capacitances of the clock network. The efficient implementation of the algorithm across all levels of hierarchy, optimizing the clock gating function and refining the power savings estimate to obtain better savings for smaller circuits need to be addressed.

4 CONCLUSION

The joint application of clock-gating and power-gating, although appealing from the theoretical stand-point, is quite problematic, mainly because of the lack of support by the existing tools and flows. In this paper, we have discussed the methods which are using individual schemes for reducing dynamic and static power and integration of both the techniques to improve the power consumption in digital circuits.

ACKNOWLEDGMENT

The authors wish to thank god.

REFERENCES

[3] Ting-Hao Lin, Chung-Yang (Ric) Huang” Using SAT-Based Craig Interpolation to Enlarge Clock Gating Functions” DAC’11San Diego, California, USA ACM 2011
[5] Li Li, Ken Choi, Haiping Nan “Effective Algorithm for Integrating Clock Gating and Power Gating to Reduce Dynamic and Active Leakage Power Simultaneously” Quality Electronic Design (ISQED), 2011 12th International Sympo-
sium on , pp- PPI – 6,2011.
1126, SEPTEMBER 2008.


