Overview on Low Power SoC Design Technology

Kimiyoshi Usami

Dept. of Information Science and Engineering
Shibaura Institute of Technology
Tokyo 135-8548, Japan
Tel : +81-3-5859-8510
Fax : +81-3-5859-8501
e-mail : usami@sic.shibaura-it.ac.jp

Abstract - So far, low power design for SoC has mainly focused on techniques to reduce dynamic power and standby leakage power. In further scaled devices, design technology to reduce active leakage power at the operation mode becomes indispensable. This is because the share of leakage power in the total operation power continues to increase as the device gets scaled. This paper gives a brief overview on the conventional leakage reduction techniques and describes novel approaches to use run-time power gating for active leakage reduction.

I. Introduction

Power dissipation in SoC becomes more and more a serious problem as MOS transistors get scaled. As shown in Fig.1, dynamic power per unit area of a chip increases as the transistor is scaled. But more importantly, leakage power increases exponentially with device scaling. If the leakage power continues to increase at this pace, it becomes equal to dynamic power at 20nm technology node [1]. Since sub-threshold leakage current increases with temperature, it is also pointed out that leakage power gets comparable to dynamic power at 50nm technology node at high temperature (100C). So far, low power design has mainly focused on techniques to reduce dynamic power and standby leakage power. In further scaled devices, leakage reduction even at the operation mode becomes indispensable because the share of leakage power in the total power dissipation gets larger.

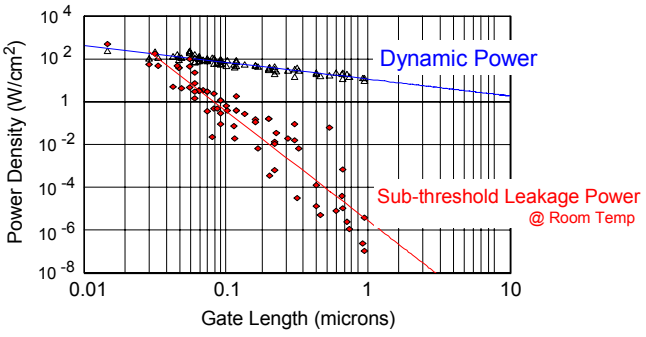

Fig. 1. Trend on power dissipation in device scaling.

In this paper, first I give an overview on major techniques to reduce leakage power, and then I present novel approaches to reduce active leakage power at the operation time. The paper is organized as follows: Section II describes the conventional design techniques to reduce leakage. Section III presents novel power gating approaches to reduce active leakage power.

II. Conventional Design Techniques to Reduce Leakage Power

Leakage power is dissipated due to leakage current flowing at MOS transistors. Sub-threshold leakage and gate leakage are two major components in sub-0.1μm devices. The sub-threshold leakage current flows between the drain and the source of an MOS transistor when the transistor is OFF. The gate leakage current flows through the thin gate oxide due to tunneling. Techniques to reduce gate leakage have been explored not only at the design community but also at the device/process community. Development of high-k dielectric is a key to solve this problem. In contrast, techniques to reduce sub-threshold leakage rely greatly on design level approaches.

Sub-threshold leakage current is strongly related to the threshold voltage Vth of an MOS transistor. To reduce the sub-threshold leakage, Vth should be raised. However, raising Vth reduces the drain current of the transistor, leading to increasing the delay. Thus, there is a trade-off between the leakage and the performance of a transistor. Design techniques to reduce leakage while keeping high performance have been developed so far. Dual-Vt, body-biasing and power gating are most important techniques that are actually used in real chips. The dual-Vt is a technique to use high-Vth and low-Vth depending on the path criticality. Low-Vth is used at the critical paths of a chip to keep high performance, while high-Vth is used at non critical paths to reduce leakage. Body-biasing is a technique to dynamically change the effective Vth by applying the body bias. A chip is fabricated with low Vth and runs at high performance in the active mode. In the standby mode, reverse body-bias is applied and the effective Vth of the transistor is raised, resulting in reducing leakage.

In contrast to these techniques, the power gating technique reduces leakage current by electrically disconnecting the circuit from the power or the ground using a power switch. A power gating technique to use high-Vth for the power switch and low-Vth for logic gates is refereed as MTCMOS (Multi Threshold-voltage CMOS). The structure of an MTCMOS

1-4244-0630-7/07/$20.00 ©2007 IEEE.

circuit is depicted in Fig.2. In the standby mode, the high-Vth NMOS power switch is turned off. Since logic circuit and the ground are electrically disconnected, the leakage is reduced. In the active mode, the power switch is turned on and low-Vth logic gates run at high speed. The power gating technique has a big advantage over other techniques stated above: the gate leakage can be reduced as well by using a power switch with thicker gate oxide in power gating.

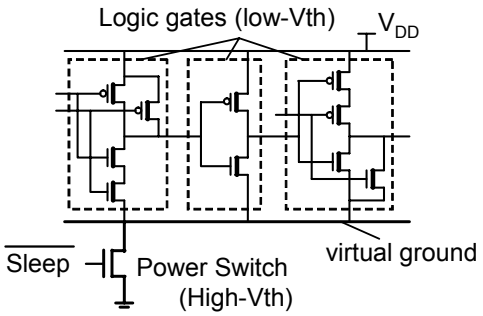

Fig. 2. Structure of MTCMOS circuit.

III. Power gating approaches to reduce active leakage

As presented above, the power gating technique was originally developed to reduce standby leakage. An attempt to use the power gating to reduce leakage power dissipated in the operation time ("active leakage power") has been reported in two papers [2] [3].

One approach is to partition an internal circuit of a processor into 20 power domains and to control the power switches to each domain independently at run time [2]. This approach was demonstrated at a mobile processor shown in Fig.3. When performing the video telephony, almost all the power domains are activated. In contrast, when waiting for calling, the power domains except for a couple of domains (e.g. PD13, PD14) are deactivated by turning off their power switches. It is reported that the leakage current is reduced to 35% when the usage scene is changed from the video telephony to the waiting for calling [4]. This is a good example of the module-level Run-Time Power Gating (RTPG).

Fig. 3. Mobile processor with module-level Run-Time Power Gating.

The other approach is to apply RTPG in much finer grain of circuits [3]. This approach uses sleep signals that are extracted locally within the design. By utilizing enable signals in gated clock, the design is partitioned into domains. By considering dynamic energy overhead due to turning on/off power switches, the domains that will achieve the gain in energy savings are finally chosen as a target of power gating.

Gated clock is originally a technique to reduce dynamic power of clock network. When data stored in flip-flops (F/F's) are not updated, clock toggling to the F/F's is stopped to reduce dynamic power. During this period, combinational logic gates located at the transitive fan-in of the F/F's are not required to compute new data to the F/F's. If outputs of the combinational logic gates are not used at anywhere else, the logic gates are considered as "idle". By detecting this idle period, the power switch provided to the combinational logic gates can be turned off. This results in reducing active leakage power of the combinational logic gates. Figure 4 shows the basic structure that implements this approach.

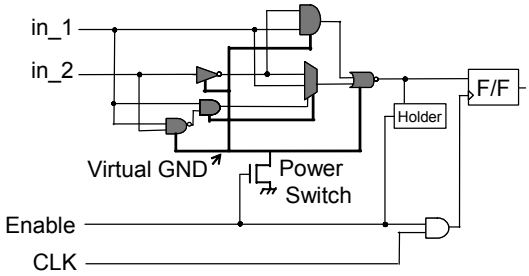

Fig. 4. Basic structure used for fine-grained Run-Time Power Gating.

In actual clock-gated designs, it is likely that more than one enable signals exist. To perform a fined-grained RTPG for these designs, an idea of "power gating domain" (PG-domain) was introduced. The PG-domain is defined as a group of circuits that are power gated with a unique enable signal. An example of the PG-domain is depicted in Fig. 5. In this circuit there are two enable signals EN_A and EN_B, controlling clock-gating for registers regA and regB, respectively.

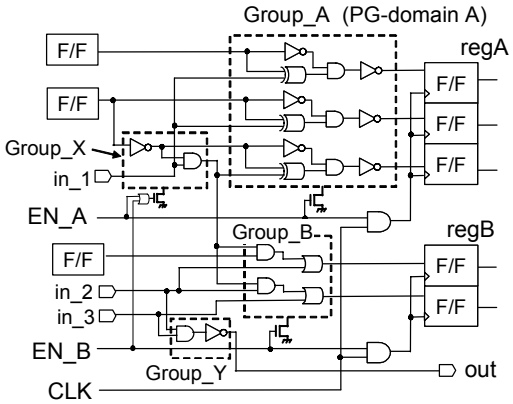

Fig. 5. Power gating domain for fine-grained Run-Time Power Gating.

Combinational logic gates enclosed with a dotted line and indicated as "Group_A" perform computation only for the register regA. In other words, the logic gates in Group_A

become idle if regA is not updated. This allows us to power gate the combinational logic gates in Group_A with the enable signal EN_A. Hence, Group_A is referred as the "PG-domain A". Similarly, logic gates indicated as "Group_B" can be power-gated using the enable signal EN_B. In contrast, combinational logic gates indicated as "Group_X" influence not only regA but also regB. These logic gates become idle only when neither regA nor regB are updated. Group_X is referred as the "PG-domain AB" and the domain is power-gated using both EN_A and EN_B.

This partitioning algorithm was implemented and incorporated into the design flow. This approach was experimentally applied to a datapath module of an embedded microprocessor and evaluated with a 90nm device model used in industry. It is reported that active leakage power was saved by 83% at the area penalty by 20% [3].

IV. Conclusions

Power dissipation in SoC gets more serious as the device is scaled. The share of leakage in the total operation power gets larger in further scaled devices. Techniques to reduce active leakage power become indispensable. Run-time power gating is a promising technique to achieve this goal. Design issues include timing analysis and power estimation that are more complicated in run-time power gating. Design tools and environment to support run-time power gating are required.

References

[1] D. Lackey, P. Zuchowski, J. Koehll, "Designing mega-ASICs in nanogate technologies," *Proc. of DAC'03*, pp.770-775, June 2003.

[2] T. Hattori, et al, "A power management scheme controlling 20 power domains for a single-chip mobile processor," *ISSCC Dig. Tech. Papers*, pp. 542-543*,* Feb. 2006.

[3] K. Usami and N. Ohkubo, "A design approach for fine-grained run-time power gating using locally extracted sleep signals," *Proc. of ICCD'06*, pp.155-161, Oct. 2006.

[4] T. Hattori, et al, "Hierarchical power distribution and power management scheme for a single chip mobile processor," *Proc. of DAC'06*, pp.292-295, July 2006.