A comprehensive thermal-aware power mangement system with block-level optimization in 100nm CMOS technology

W.-M. Chan, H. Chiueh

Belgirate, Italy, 28-30 September 2005

A COMPREHENSIVE THERMAL-AWARE POWER MANAGEMENT SYSTEM WITH BLOCK-LEVEL OPTIMIZATION IN 100NM CMOS TECHNOLOGY

Wei-Min Chan and Herming Chiueh

SoC Design Lab, Department of Communication Engineering,
College of Electrical and Computer Engineering, National Chiao Tung University
Hsin-Chu 30050, Taiwan

ABSTRACT

Modern SoC integrations and mobile systems have emphasized low power techniques due to shortage of battery life. Conventional power management designs focused on the reduction of dynamic power consumption, recent designs begin to take leakage power into consideration since it becomes an important factor in nano-scale CMOS technology. Latest development has taken advantage of modularity in SoC design methodology to develop the block-level control technique for power reductions. However, thermal gradient over the system and its impacts to SoC designs are barely discussed. In this research, a block-level optimization of comprehensive thermal aware power management is presented. The proposed design applies several low power techniques to control different power sources and handles thermal impacts to provide performance coherence. As a result, optimal powerreductions and performance coherence can be guaranteed within the whole system. The simulation results show a significant improvement in stability and leakage power reduction for most circuitries. The results are based on TSMC 100nm CMOS technology.

1. INTRODUCTION

Power dissipation in modern VLSI designs has become the most critical issue in System-on-Chip era. The ever increasing on-chip integrations in recent decade have enabled a dramatically increase in system performance and scale. Unfortunately, accompanied with the performance improvement, a significantly increase in power dissipation and heat density is introduced [1]. However, in modern VLSI circuitry for mobile systems, such as handheld audio and video applications, low power considerations are becoming an important issue as battery life and geometry of mobile systems are limited. Therefore, power and thermal issues have become the major limitation of such systems. Low-power circuitry designs and architecture-level power-reduction techniques have become more important in modern system-on-chip implementation.

Generally, power consumption is categorized to three portions in VLSI circuitry: short circuit, leakage, and dynamic switching power. Traditionally, dynamic switching power is the dominant component of power consumption. However, as scaling trends continue in future generations, leakage power has become comparable to the dynamic power, and potentially a dominant component of overall power consumptions. Therefore, unlike previous power reduction methods, modern power management techniques for SoC designs have to deal with dynamic switching power as equally important as static leakage power.

Several power reduction techniques can be adopted for modern power management designs, such as dynamic voltage scaling (DVS) [2], clock gating [3], power gating [1, 4], body bias [5-7], and voltage islands [8]. DVS and clock gating can effectively reduce dynamic power consumption, while power gating and reverse body bias reduce static leakage power. Voltage islands can provide block-level control of power management, which makes control mechanism more flexible.

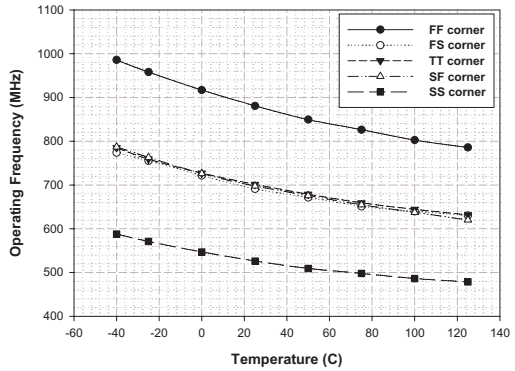
Besides battery life, thermal impact is another major reason to utilize power management. Power consumption of the chip contributes thermal impact, such as frequency sacrifice, leakage increase, or even circuit breakdown. Because of high costs of packaging and cooling requirements for reliability, power reduction is also addressed to face thermal impact. However, system temperature is not temporally and spatially constant, so adaptive power control doesn't imply adaptive thermal control.

Shown in Fig.1 is the thermal impact to operating frequency, power consumption, and leakage power of a ring oscillator. We notice that there are about 13% frequency decrease, 18% total power consumption increase in the active mode, and 9X leakage power increase in the standby mode from 0 to 125°C. Functional blocks will not keep performance coherence at the same target frequency due to temperature difference. Leakage problems will become worse and worse as temperature rises. System will become asynchronous and power will not be controlled within specifications if we ignore these changes. Therefore, power management of SoC chip should take temperature variation into consideration.

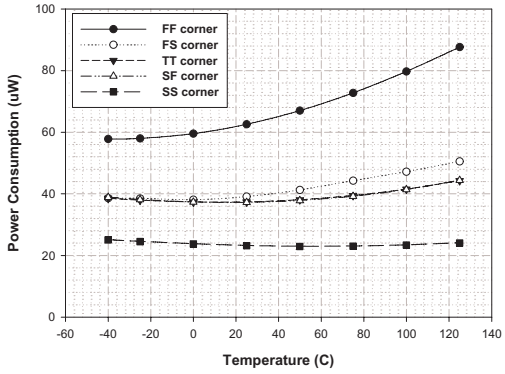
In this paper, we present a comprehensive thermal-aware power management architecture and emphasis thermal optimization in block-level control. The proposed design not only deal with both dynamic switching power and static leakage power, but also deal with temperature variation to dynamically adjust each block's performance and power. Our goal is to make power and performance scalable as well as performance coherence between functional blocks in the meantime.

The rest of the paper is organized as follows: Section II presents a comprehensive thermal-aware power management architecture and relative low power techniques. The procedure to develop the control parameters for block optimization is presented in Section III. In Section IV summarizes, conclusions and future works are discussed.

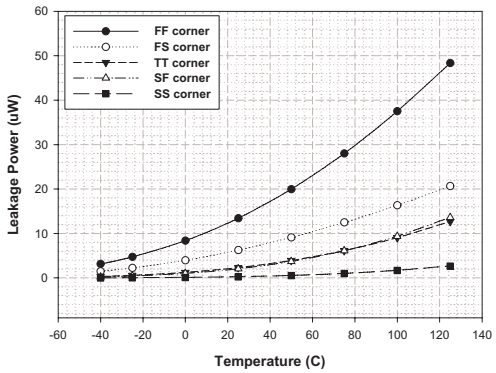
ISBN 2-916187-01-4



(a)



(b)



(c)

Figure 1. Thermal impact to (a) operating frequency, (b) dynamic power, and (c) leakage power consumption of a ring oscillator in TSMC 100nm technology. Supply voltage is 1.0V.

2. ARCHITETURE DESIGN

A comprehensive thermal-aware power management unit (PMU) architecture is shown in Fig. 2. Techniques including dynamic voltage scaling, clock gating, power gating, and body bias are combined to control system power and adapt to thermal impact. These techniques can be applied at the full-chip level, or they can be applied on a finer block-level granularity. As the result, the concept of voltage islands is also included in power management design.

The functionality and role of different circuitry block will be described in the following paragraphs:

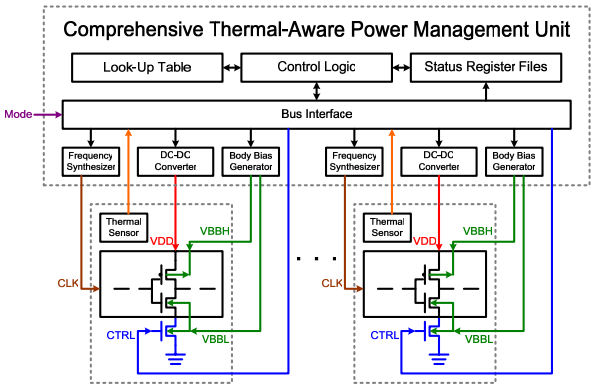


Figure 2. Architecture of proposed comprehensive thermal-aware power management. Peripheral circuits, such as DC-DC converter and temperature sensor, are required.

Dynamic voltage scaling is the most efficient technique to reduce dynamic power. It also has benefits for leakage power saving. When the required performance of the target system is lower than the maximum performance, supply voltage can be dynamically reduced to the lowest possible extent that ensures proper operation. Significant power reduction is possible, since dynamic power of CMOS circuits is proportional to the square of the supply voltage.

Clock gating is an efficient technique to reduce clock switching power when the system is idle or data remain unchanged, which is widely used in modern designs [3].

Power gating is the most efficient technique to reduce leakage power. In the active mode, the power gating devices act as minor power-rail resistance which does not affect the correctness of the function. In the standby mode, however, the power gating devices act as open switches that cut off the direct connection between power supply and internal circuits so as to reduce leakage current [6].

Reverse body bias can be applied for leakage power reduction although its effectiveness reduces with technology scaling [6, 7]. In the standby mode, reverse body bias is applied to raise threshold voltage to save leakage power.

Voltage Islands provide block-level control of the system. In the active mode, local supply voltage can scale down to reduce dynamic power while maintaining the same performance for islands at non-critical paths. In the standby mode, leakage savings are possible when the island is not being used, regardless of whether or not the rest of others are operating. Block-level control makes it easier to spatially optimize the system.

As shown is Fig. 2, the look-up table is among the most important design parameter of the PMU. It provides frequency-voltage information for software feedback of the system, either in operation mode transitions or toward temperature variation. In this paper, we focus on the look-up table design for the block-controlling parameters, especially for adjust frequency-voltage relationship under temperature variation. Beside power control mechanism, we utilize DVS on the temperature axis to meet performance requirements. We hope that power consumption can be controlled by the PMU but also deal with thermal impact to performance coherence in the meantime.

3. EXPERIMENTAL RESULTS

To verify the concepts, a 101-ring oscillator is chosen as test vehicle. The simulation results of different circuitry techniques utilized in proposed power management system such as power gating insertion, DVS, and body-bias is presented. The simulation of clock gating method is skipped since it is a mature technique. The experimental results are based on TSMC 100nm CMOS technology.

3.1. Test Vehicle

A 101-stage ring oscillator is chosen, as shown in Fig. 3, to be the test vehicle for following two reasons: First, the ring oscillator is composed by inverters, which consumes the largest static leakage power because of the least stacking effect. Second, the scale of the 101-stage ring oscillator is comparable to a 32-bit adder, meaning that the experimental results are practicable.

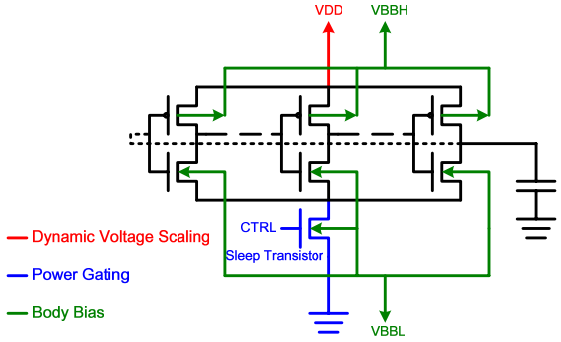


Figure 3. A 101-stage ring oscillator is used as the test vehicle. Several power reduction concepts are applied to this circuit.

3.2. Power Gating Insertion

Sleep transistors, as power gating devices, can be PMOS (header-type) or NMOS (footer-type). Because PMOS sleep transistors require larger silicon area to be capable of sourcing the maximum instantaneous current in the active mode, we choose NMOS to be power gating devices. Sleep transistors sizing is the most important and hardest due to frequency loss, area overhead, and other trade-offs. As shown in Fig. 4, 3X minimum size of NMOS sleep transistor is chosen for our test vehicle, which brings about 4% frequency loss, 9% dynamic power saving, and small virtual ground raising. Leakage saving of power gating technique will be discussed later with body bias.

3.3. DVS Simulation

Fig. 5 shows relationships between operating frequency and supply voltage in 8 different temperature levels. We notice that supply voltage range can be divided into high, middle, and low VDD zones. When operating in high VDD zone, frequency decreases as temperature increases, which causes the problem of performance coherence between functional blocks. In middle VDD zone, frequency decreases slightly as temperature increases. However, frequency increases as temperature increases in low VDD zone, resulting in no performance coherence problem at all.

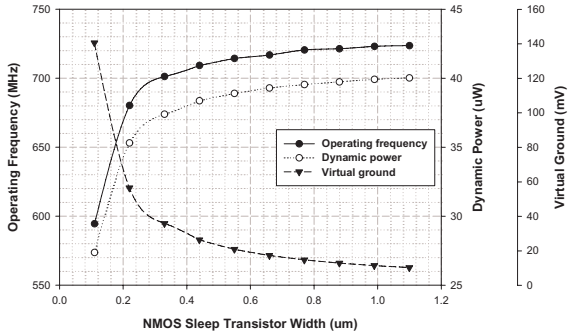


Figure 4. Relationships between operating frequency, dynamic power, virtual ground level, and sleep transistor width. The NMOS sleep transistor is designed in minimum length.

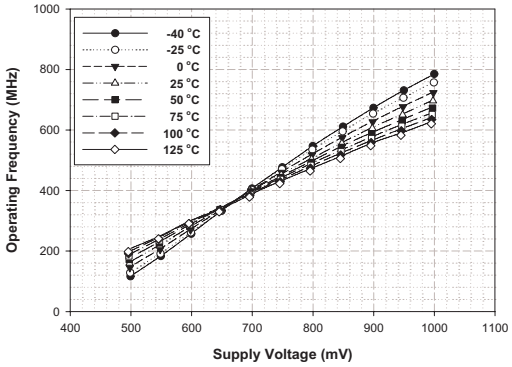


Figure 5. Relationships between operating frequency and supply voltage in 8 different temperature levels.

Therefore, we raise a margin of supply voltage in high VDD and middle VDD zone to keep performance coherence, but keep minimum required supply voltage in low VDD zone. Table 1 is simulation results of the DVS look-up table. Without utilizing DVS, supply voltage must be the highest one, 980mV in the table, to ensure normal operation. On the contrary, it only needs to be 500mV for 200MHz operation at 150°C if we utilize DVS on both frequency and temperature dimensions. It can lead to 73% dynamic power reduction and maintain performance coherence at the same time.

Table 1. The DVS look-up table.

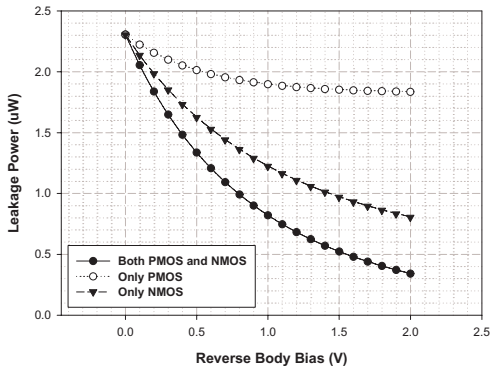
V_{DD} (mV)	200 MHz	300 MHz	400 MHz	500 MHz	600 MHz
25~0°C	560	630	700	780	870
0~25°C	550	620	700	790	890
25~50°C	530	620	710	800	910
50~75°C	520	610	710	810	930
75~100°C	510	610	710	820	940
100~125°C	500	600	720	830	960
125~150°C	500	600	720	840	980

3.4. Body Bias Simulation

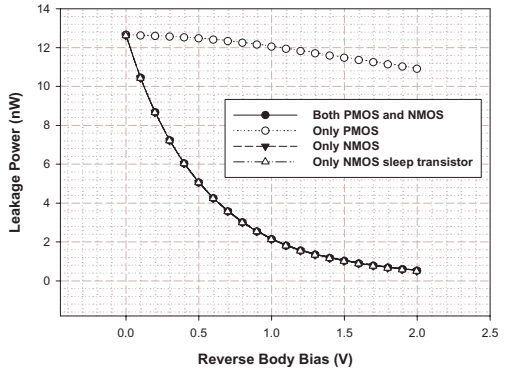
In the end of this section, we present experimental results about body bias technique. As shown in Fig. 6(a), applying reverse body bias on PMOS is the least effective since the leakage

current of OFF-state PMOS devices is much smaller than that of NMOS devices. Applying reverse body bias on both PMOS and NMOS leads to the largest leakage saving. However, applying reverse body bias on NMOS only is efficient due to lower power overhead of the charge-pump circuit, which is needed to provide body bias voltage.

Fig. 6(b) is the case with power gating insertion. Because leakage current of the circuit is decided by the OFF-state NMOS sleep transistor in this situation, applying reverse body bias on this sleep transistor only is as effective as on all PMOS and NMOS devices. In addition, applying reverse body bias on NMOS sleep transistor only introduces the least loading and power overhead of charge-pump circuit.



(a)



(b)

Figure 6. Relationships between leakage power and reverse body bias. (a) is the case without power gating, and (b) is conversely. Supply voltage is 1.0V at 25°C.

What is worth to mention is that the magnitude of reverse body bias voltage has its limit; in fact, applying too large reverse body bias will bring about obvious second order effects (SCE), such as gate induced drain lowering (GIDL). Previous research indicates that beyond a certain optimal RBB voltage, which is about 0.3X VDD for 0.13um technology, the transistor OFF-state current starts to increase due to increased GIDL leakage [6]. We choose 0.3V as optical reverse body bias voltage to avoid the GIDL effect because the lack of BSIM4 model.

Utilizing power gating technique will reduce leakage power from 2.3uW to 12.7nW, and combining power gating with 0.3V reverse body bias can further reduce leakage power to 7.2nW. These simulation results show that power gating and reverse body bias are powerful to reduce leakage power. The experimental results shows the proposed architecture have potential to reduce more than 70% of dynamic and 90% of static power dissipation in regular CMOS systems.

4. CONCLUSION

This paper presented a block-level optimization of comprehensive thermal aware power management. The proposed design applies different low-power circuitry techniques to control different blocks of SoC designs to provide performance coherence with power reduction optimization. The simulation results show a significant improvement in stability and leakage power reduction for most circuitries. The results are based on TSMC 100nm CMOS technology.

5. ACKNOWLEDGEMENT

This research was partially supported by National Science Council, Taiwan. The contract numbers are 93-2220-E-009-016 and 94-2220-E-009-016. The authors would also like to acknowledge the design parameters provided by NCTU-TSMC joined project.

6. REFERENCES

[1] M. Powell, S.-H. Yang, B. Falsafi, K. Roy, and T. N. Vijaykumar, "Gated-Vdd: A Circuit Tecnhique to Reduce Leakage in Deep-Submicron Cache Memories," in International Symposium on Low Power Electronics and Design, 2000.

[2] S. Lee and T. Sakurai, "Run-Time Voltage Hopping for Low-Power Real-Time Systems," in 37th Conference on Design Automation, 2000.

[3] Q. Wu, M. Pedram, and X. Wu, "Clock Gating and Its Application to Low Power Design of Sequential Circuits," in IEEE Transactions on Circuit and Systems I: Fundamental Theory and Applications, vol. 47, March 2000, pp. 415-420.

[4] C.-H. Hua and W. Hwang, "A Power Gating Structure with Concurrent Data Retention and Intermediate Modes in 100nm CMOS Technology," in 15th VLSI Design/CAD Symposium, 2004.

[5] J. W. Tschanz, S. G. Narendra, Y. Ye, B. A. Bloechel, S. Borkar, and V. De, "Dynamic Sleep Transistor and Body Bias for Active Leakage Power Control of Microprocessors," IEEE Journal of Solid-State Circuits, vol. 38, pp. 1838-1845, November 2003.

[6] B. Chatterjee, M. Sachdev, S. Hsu, R. Krishnamurthy, and S. Borkar, "Effectiveness and Scaling Trends of Leakage Control Techniques for Sub-130nm CMOS Technologies," in International Symposium on Low Power Electronics and Design, 2003.

[7] A. Keshavarzi, S. Ma, S. Narendra, B. Bloechel, K. Mistry, T. Ghani, S. Borkar, and V. De, "Effectiveness of Reverse Body Bias for Leakage Control in Scaled Dual Vt CMOS ICs," in International Symposium on Low Power Electronics and Design, 2001.

[8] R. Puri, L. Stok, J. Cohn, D. Kung, D. Pan, D. Sylvester, A. Srivastava, and S. Kulkarni, "Pushing ASIC Performance in a Power Envelope," in 40th Design Automation Conference, 2003.