第28卷 第4期 2007年4月
半 导 体 学 报
CHINESE JOURNAL OF SEMICONDUCTORS
Vol.28 No.4 Apr.,2007

A Thermal-Conscious Integrated Circuit Power Model and Its Impact on Dynamic Voltage Scaling Techniques*

Liu Yongpan†, Yang Huazhong, and Wang Hui

(*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*)

Abstract: We propose a novel thermal-conscious power model for integrated circuits that can accurately predict power and temperature under voltage scaling. Experimental results show that the leakage power consumption is underestimated by 52% if thermal effects are omitted. Furthermore, an inconsistency arises when energy and temperature are simultaneously optimized by dynamic voltage scaling. Temperature is a limiting factor for future integrated circuits, and the thermal optimization approach can attain a temperature reduction of up to 12℃ with less than 1.8% energy penalty compared with the energy optimization one.

Key words: CMOS integrated circuits; power model; temperature; DVS

EEACC: 2570A; 1265

CLC number: TN47 **Document code**: A **Article ID**: 0253-4177(2007)04-0530-07

1 Introduction

Dynamic voltage scaling (DVS) techniques are effective for reducing both power and energy consumption of integrated circuits. Due to the leakage power's super-linear relationship with temperature in nanometer integrated circuits (ICs), the thermal independent power models presented in previous DVS research[1~5] are inaccurate. Even worse, the steady increase of leakage power with the downscaling of feature size makes those models' errors larger. Therefore, a thermal-conscious power model for ICs is necessary to deploy dynamic voltage scaling. Recently, temperature dependent power models[6,7] have been developed for ICs. However, neither energy nor thermal optimization has been covered. In this paper, a novel thermal-conscious power model is presented for ICs, and its impact on energy and thermal optimization by DVS techniques is discussed.

2 Thermal-conscious modeling methodology

The power consumption of modern CMOS ICs mainly consists of dynamic, leakage, and short circuit components. The short circuit currents occur when both nMOS and pMOS blocks conduct simultaneously for a short period of time. It is generally determined by the switching activities and input signals. Due to its relatively small contribution to the total power budget, we will focus on the former two components in this paper. Dynamic power consumption is insensitive to temperature[8], while leakage current is strongly affected by it. Therefore, we first derive the temperature dependent leakage power model. Dynamic power, delay and thermal models are then introduced. Due to the interaction between leakage and temperature, an iterative method is presented to estimate the leakage power for integrated circuits.

2.1 Preliminary

As Figure 1 shows, subthreshold leakage, gate leakage, and junction leakage are three main leakage components in modern digital circuits. Since the former two will dominate in the near future[9], our discussion will focus mainly on them.

Based on the BSIM4 CMOS model[10], subthreshold leakage is given as

$$I_{subn} = A_s \frac{W}{L}(v_T)^2(1 - e^{-V_{DS}/v_T})e^{(V_{GS}-V_{th})/nv_T} \quad (1)$$

where V_{GS} and V_{DS} are the gate-source and drain-source voltages, respectively, $v_T = kT/q$ is the thermal voltage, V_{th} is the threshold voltage, A_s is

* Project supported by the National Natural Science Foundation of China (Nos. 90207001, 60506010)

† Corresponding author. Email: ypliu99@mails.tsinghua.edu.cn

Received 13 October 2006, revised manuscript received 27 November 2006

©2007 Chinese Institute of Electronics

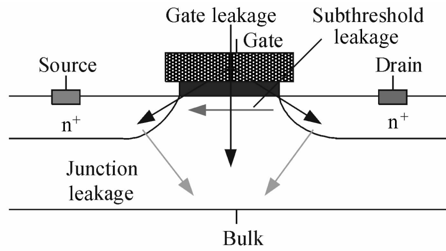

Fig.1　Main leakage components

a technology-dependent constant, and W and L are the effective channel width and length. The gate leakage can be obtained from the equation

$$I_{gate} = WLJ_{gate} \tag{2}$$

where J_{gate} is the gate current density. As Figure 1 shows, gate leakage consists of tunneling current between gate and substrate, current between gate and channel, and current between the gate and source diffusion region. All of them have the following uniform expression[10]:

$$J_{gate} = A_{gate}\left(\frac{T_{oxref}}{T_{ox}}\right)^{nt_{ox}} \times \frac{V_g V_{aux}}{T_{ox}^2} e^{-\phi T_{ax}(\xi - \tau|V_{ox}|)(1+\upsilon|V_{ox}|)} \tag{3}$$

Equation (3) shows that gate leakage depends on oxide thickness and supply voltage exponentially.

2.2　Temperature dependent leakage model

Summarizing Equations (1) ~ (3), we can conclude that subthreshold leakage increases exponentially with the temperature, supply voltage, and body bias voltage, while gate leakage strongly depends on supply voltage. After simplifying these formulas, the following expression can be used to calculate the total leakage current:

$$P_{leakage} = V_{dd} I_s(T_{ref}, V_{ref})(AT^2 e^{(\alpha V_{dd} + \beta V_{bs} + \gamma)/T} + Be^{\mu V_{dd}}) \tag{4}$$

Here, the leakage current under reference temperature and voltage is denoted as I_s, the temperature of the IC is defined as T, the empirical constants, A, B, α, β, γ, μ, are acquired by curve fitting methods for different circuit types, processes, and designs. We have developed a leakage power estimation flow[11] to extract those constants for different ICs. Constant values for ISCAS85[12] and SRAM circuits[13] under a 65nm PTM[14] process are shown in Table 1. Compared with the HSPICE simulation results, the average and worst-case modeling errors are 1.3% and 6%.

From Table 1, we can see that integrated circuits, synthesized on the same standard cell library, have a very similar leakage variation relationship with control parameters, such as supply voltage, body bias voltage, and temperature. Thus we can use the same leakage parameters in Eq. (4) to estimate the leakage power for integrated circuits with similar design styles.

Table 1　Leakage parameters for ISCAS85 and SRAM circuits

ICs	$A/10^{-4}$	α	β	γ	$B/10^{-4}$	μ
C5315	5.406	1127	1670	2224	6.597	5.69
C6288	5.447	1122	1670	2223	6.770	5.69
C7552	5.467	1123	1671	2224	6.769	5.69
16Kx32	2.867	1177	1593	2163	20.037	5.68
2Mx32	2.824	1178	1592	2161	20.422	5.68

2.3　Dynamic power and delay model

The dynamic consumption of ICs can be estimated by the formula

$$P_{dyn} = C_{eff} V_{dd}^2 f \tag{5}$$

where C_{eff} is the effective switch capacitance, and f is the operating frequency. To estimate the circuit's operating frequency under variable supply voltages, the IC's delay is determined by the alpha-power formula[1],

$$f^{-1} = t = \frac{K}{(V_{dd} - V_{th})^{\sigma}} \tag{6}$$

where the constant K is determined by logic depth and process, and σ is a measure of velocity saturation.

2.4　Thermal model

According to the heat transfer theory, the thermal phenomena can be modeled as an equivalent RC thermal circuit, as shown in Fig.2.

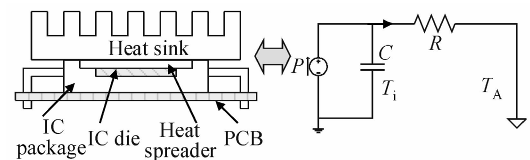

Fig.2　Equivalent thermal RC circuits

Suppose the ambient temperature of an IC is denoted as T_A, and T_t is the IC temperature at time t. The temperature after time Δt can be calculated as below:

$$T_{t+\Delta t} = T_A + (T_t - T_A)e^{\frac{-\Delta t}{RC}} + PR(1 - e^{\frac{-\Delta t}{RC}}) \tag{7}$$

Here the IC's power consumption is P during Δt, and the thermal resistance and capacitance are de-

noted as R and C, which are estimated using the thermal parameters in Table 2. The values of vertical thermal resistance and capacitance are calculated by the formula

$$R_{\text{vertical}} = t/kA, C_{\text{vertical}} = ctA \quad (8)$$

where A is the IC area. Although this model tends to underestimate the temperature at hotspots, it provides a good first-order estimation of the average temperature, which is widely used for leakage power estimation in the industry[15]. The first order thermal model can be extended into a fine-grained grid-based one, and it will enable us to deploy the accurate thermal analysis and figure out the hotspots on the chip. However, there is a tradeoff between the modeling efficiency and accuracy at an early design stage.

Table 2 Thermal parameters for typical configuration

Variable	Value
Die thickness, t	250μm
Die thermal conductivity, k_d	157W/(m·K)m^3
Die heat capacity, c_d	1.638×10^6J/(m^3·K)
Cu heat sink thermal conductivity, k_s	400W/(m·K)
Cu heat sink heat capacity, c_s	3.55×10^6J/(m^3·K)

2.5 Iterative modeling

Given initial dynamic and static power values, the IC's temperature is calculated from Eq.(7). After that, the power values are updated based on the latest temperature according to Eqs. (4) and (5). By this way, the temperature and power values are calculated iteratively until both temperature and total power are invariable. That is, the IC's thermal equilibrium is reached under its thermal configuration. Figure 3 shows an iterative way to estimate the steady power and the temperature of an IC. In our experiments, the IC's power and temperature variations are less than 0.1% after 5 iterations.

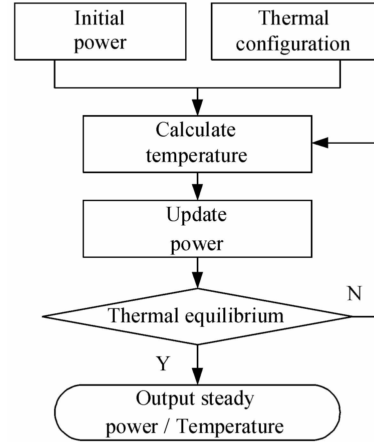

Fig.3 Iterative thermal and power modeling

3 Experiments and discussion

Using the proposed thermal-conscious modeling methodology, we first show its impact on power estimation under voltage scaling. Furthermore, the traditional energy efficient DVS optimizing problem is revisited by special experiments.

3.1 Impact on power estimation

Figure 4 illustrates a processor's leakage power at different supply voltages based on the proposed models. The processor's dynamic power parameters are calculated using published data on the 65nm CMOS Core™ Duo processor[16], while thermal parameters are estimated using Hotspot[17]. The leakage parameters are adapted from the PTM models for the same process in Section 2. All of them are shown in Table 3. As we can see in Fig. 4, the cross and square curves give out traditional leakage power trends under constant temperature. The dot curve gives out the accurate iterative leakage power profile. A power difference between the traditional trend and the iterative one of up to 52% is observed in our experiments, which underscores the necessity of a thermal-conscious power model.

Table 3 Experimental setup

Variable	Value	Variable	Value	Variable	Value
V_{dd}	0.9~1.4V	C_{eff}	5nF	K	0.551×10^{-9}
V_{th}	0.40V	f_{max}	2.16GHz	I_s	1.238A
σ	1.2	R	1.2℃/W		

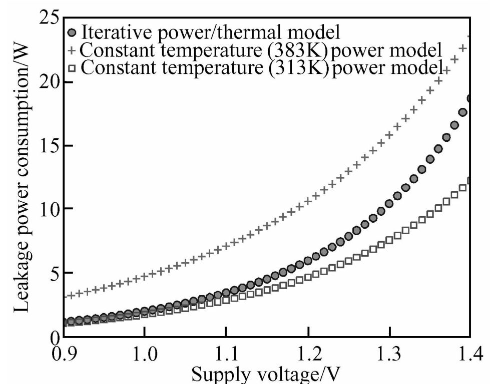

Fig. 4 Thermal-aware leakage power versus supply voltage

3.2 Impact on traditional DVS techniques

Next, we explore the traditional energy efficient optimization by DVS. Beside energy, temperature has become a limiting factor for future ICs due to increasing power density, shrinking volume, and dense integration. Most reliable challenges, such as electromigration, thermal cycling, time dependent dielectric breakdown, and stress migration[18], have an exponential relationship with temperature. Therefore, the temperature is considered as well as the energy in our experiments.

Suppose two tasks, task 1 and task 2, are to be executed on a processor, and there is a real-time deadline DL, with task 2, which has to wait until the completion of task 1. Using the models in Section 2, the power and energy consumption of each task can be given by the following formulas (the formulas are used in an iterative way to auurately estimate the power and the temperature in our implementations.):

$$E_{tot} = P_{tot}\mathrm{EC}/f \tag{9}$$

$$P_{tot} = \theta C_{eff} V_{dd}^2 f + \omega V_{dd} I_s (A T^2 e^{(\alpha V_{dd} + \beta V_{bs} + \gamma)/T} + B e^{\mu V_{dd}}) \tag{10}$$

Here EC is the instruction cycle number, θ is the probability of switching capacitance, and ω is the leakage factor, which describes the effects of input patterns on leakage current. These parameters characterize the task property and can be extracted using the simulation method in Ref. [19]. Their values in our experiment are shown in Table 4. The constant values for the processor are calculated using the datasheet for a 65nm CMOS high performance Pentium™ processor[20].

Table 4　Experimental setup

Variable	Value	Variable	Value	Variable	Value
V_{dd}	0.9～1.4V	C_{eff}	15nF	K	0.344×10^{-9}
V_{th}	0.244V	f_{max}	3.46GHz	I_s	2.599A
σ	1.2	R	0.8℃/W	DL	2s
θ	0.6/0.7	ω	2.0/0.4	EC	3/3 (Billion)

Here, the focus is on the high-power density case, which can be encountered in emerging ICs, such as 3D ICs and multiprocessor systems-on-a-chip. The voltage relationship by real-time constraint between task 1 and task 2 is obtained as below:

$$\mathrm{DL} = \frac{\mathrm{EC}_1 K}{(V_1 - V_{th})^{\sigma}} + \frac{\mathrm{EC}_2 K}{(V_2 - V_{th})^{\sigma}} \tag{11}$$

Based on Eqs. (9)～(11), we draw an energy and power curve of task 1 and task 2 by the supply voltage of task 1 in Figs. 5 and 6, respectively.

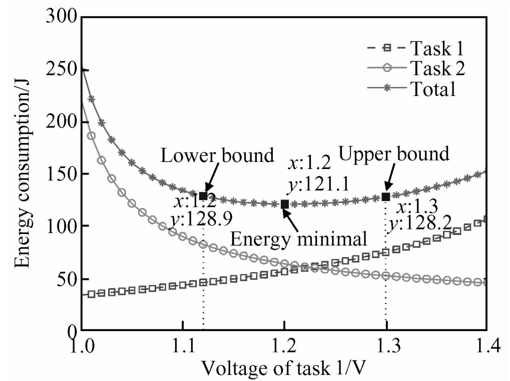

Fig.5　Energy curves by different supply voltages

In Fig. 5, the energy of task 1 increases with its supply voltage, while that of task 2 decreases. The total energy curve is flat in a large voltage range. A lower and upper bound of task 1's voltage can be found to define the range, in which the total energy overhead is less than 7% compared with the energy minimal value. In Table 5, energy consumption, peak power, peak temperature, and energy overhead are listed for some voltage settings in this range. They are energy minimal, lower bound, upper bound, and thermal optimal voltage settings. As all of the voltage settings in this range consume similar energy, we focus on their peak power and temperature values. According to Eq. (7), the IC's steady state temperature is decided by the power and thermal configurations.

Table 5　Experimental setup

Voltage settings	Energy /J	Peak power /W	Peak temp /K	Overhead /%
Energy min.	121.1	71	370	0
Lower bound	128.9	104	396	6.4
Upper bound	128.2	78	375	5.9
Thermal opt.	122.1	61	362	0.8

Therefore, for a processor with certain thermal configuration, the optimal thermal voltage settings require equal power consumption of both task 1 and task 2 to minimize the peak temperature. However, other voltage settings usually cause imbalanced power profiles (up to 60W power difference) between task 1 and task 2. As we can see in Fig. 6, the imbalanced profiles generate a large peak power and temperature. Furthermore, even if the minimal energy voltage settings are used, the power consumption of task 1 and

task 2 are also different. Therefore, the peak temperature in the energy optimization case can be much larger than that in the thermal optimal case.

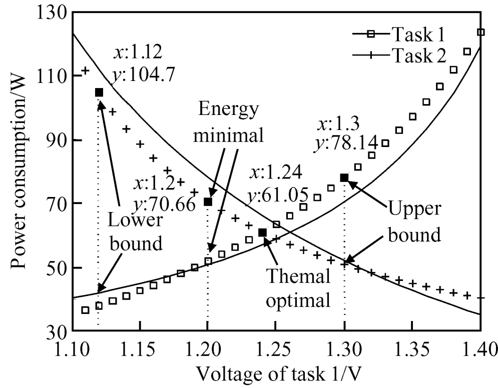

Fig.6 Power curves by different supply voltages

Due to deviation from the optimal energy voltage, voltage settings obtained by heuristic energy efficient voltage schedulers can be located anywhere in this range. They may not work well due to the high peak temperature. For example, if the lower bound voltage settings are used in Table 5, the peak temperature can reach as high as 396K (123℃), 34K higher than that of thermal optimal voltage settings, though its energy consumption is quite close to that of the energy minimal voltage settings. We conclude that energy-only optimization will encounter thermal challenges. Thermal optimization may introduce some energy overheads, but they reduce the peak temperature effectively. Furthermore, the energy overhead caused by the thermal optimization is usually small, because decreasing the peak temperature also helps to reduce energy to some extent. Therefore, the designer should decide the tradeoff between energy and temperature according to different demands.

Next, energy and thermal optimization are further compared in Table 6, which gives total energy consumption and peak temperature under different parameters. Both task parameters and processor's cooling configuration are considered and their values are also given. First of all, we keep the task parameters constant and change the cooling conditions. A larger temperature difference (up to 10℃) between energy and thermal optimization is observed under poor cooling conditions. It is due to the fact that the same peak power value causes higher temperature under poor cooling conditions. In this case, the maximum energy overhead is 1.6%.

Table 6 Comparison between energy and temperature optimization

Test case	Processor thermal resistance /(℃/W)	Task parameter				Total energy			Peak temperature		
		θ_1	ω_1	θ_2	ω_2	EO/J	TO/J	Overhead /%	EO/K	TO/K	Reduction /℃
Poor cooling	1.5	0.3	1.6	0.4	0.4	75.0	76.2	1.6	379	369	10
Typical cooling	0.8	0.3	1.6	0.4	0.4	69.1	69.53	0.62	344	341	3
Well cooling	0.5	0.3	1.6	0.4	0.4	67.1	67.6	0.75	332	330	2
High leakage	0.8	0.6	2.6	0.7	0.4	127.6	128.9	1.02	374	363	11
Medium leakage	0.8	0.6	1.0	0.7	0.3	108.8	109.2	0.37	361	356	5
Low leakage	0.8	0.6	0.2	0.7	0.1	97.6	97.7	0.10	353	352	1
Random case 1	0.6	0.3	0.5	0.4	0.1	55.9	56.1	0.36	331	329	2
Random case 2	0.8	0.8	0.4	0.9	0.3	132.2	132.2	0	368	365	3
Random case 3	0.8	0.7	1.6	0.8	0.5	134.5	134.9	0.30	374	368	6
Random case 4	1.1	0.4	1.5	0.7	0.3	101.8	103.6	1.77	382	370	12
Random case 5	1.2	0.6	1.0	0.6	0.7	111.4	111.5	0.09	382	379	3
Average								0.63			5.3

Second, keeping the thermal resistance fixed, we adjust the leakage ratio in the total power components. In the high leakage ratio case, a bigger temperature difference (up to 11℃) is observed, because a higher leakage ratio makes the power curve change more dramatically by the supply voltage due to stronger power temperature dependence. Therefore, the peak power difference becomes larger in the high leakage case, and so does the thermal difference.

Finally, random values are selected for task parameters and thermal resistances. As we can

see, the largest energy difference is smaller than 1.8%. However, the maximum temperature reduction is up to 12℃. The average energy overhead of all cases is 0.63%, while the average thermal reduction is 5.3℃. Based on these experimental results, we summarize the proposed model's impact on the DVS techniques:

(1) In reality, energy optimized voltage settings are usually not the thermal optimized solutions. However, the thermal optimized solution can be achieved with little energy overhead. In our experiments, the overhead is less than 1.8% with a up to 12℃ temperature reduction. The difference between energy optimization and thermal optimization is affected by the thermal configuration as well as task parameters.

(2) Thermal constraints must be integrated into the energy-efficient voltage scheduler for future ICs. Otherwise, the energy-efficient scheduler can generate solutions with unnecessary high peak power, which causes serious thermal problems.

4 Conclusion

In this paper, a novel thermal-conscious power model and related modeling methodology have been presented. Its impact on the power estimation and DVS scheduling techniques has also been discussed. It is shown that there is a difference between energy optimized and thermal optimized voltage selection. The thermal optimized settings can reduce the peak temperature effectively by up to 12℃ with less than 1.8% energy overhead. Next, we show that thermal metrics must be integrated in the energy-efficient scheduler for future ICs; otherwise thermal failure will invalidate the voltage scheduler's solution. Our future work will focus on implementing an energy-efficient voltage scheduler considering thermal constraints to deal with more complex benchmarks.

References

[1] Martin S M, Flautner K, Mudge T, et al. Combined dynamic voltage scaling and adaptive body biasing for lower power microprocessors under dynamic workloads. ICCAD, 2002: 721

[2] Jejurikar R, Pereira C, Gupta R. Leakage aware dynamic voltage scaling for real-time embedded systems. DAC, 2004: 275

[3] Wu D, Al-Hashimi B M, Schmitz M, et al. Power composition profile driven co-synthesis with power management selection for dynamic and leakage energy reduction. DSD, 2005: 34

[4] Andrei A, Schmitz M, Eles P, et al. Overhead-conscious voltage selection for dynamic and leakage energy reduction of time-constrained systems. IEE Proceedings, 2005, 152(1): 28

[5] Yan L, Luo J, Jha N K. Joint dynamic voltage scaling and adaptive body biasing for heterogeneous distributed real-time embedded systems. IEEE Trans CAD, 2005, 24(7): 1030

[6] Liao W P, He L, Lepak K M. Temperature and supply voltage aware performance and power modeling at microarchitecture level. IEEE Trans CAD, 2005, 24(7): 1042

[7] ZhangY, Parikh D, Sankaranarayanan K, et al. Hotleakage: a temperature-aware model of subthreshold and gate leakage for architects. Dept Comput Sci, Univ Virginia, Charlottesville, VA, Tech Rep CS-2003-05, Mar 2003

[8] Chandrakasan A, Bowhill W, Fox F. Design of high-performance microprocessor circuits. IEEE Press, 2001

[9] International technology roadmap for semiconductors. http://public.itrs.net

[10] BSIM4 home page. http://www-device.eecs.berkeley.edu/～bsim3/bsim4.html

[11] Zhang F, Luo R, Liu Y, et al. Leakage power modeling method for SRAM considering temperature, supply voltage and vias voltage. ICSICT, 2006: 1180

[12] ISCAS 8 5 benchmarks suite. http://www.visc.vt.edu/～mhsiao/iscas85.html

[13] Zhang F. Power modeling for microprecessors. BS Thesis, Tsinghua University, 2006

[14] Zhao W, Cao Y. New generation of predictive technology model for sub-45nm design exploration. Proc Int Symp Quality of Electronic Design, 2006: 585

[15] Dhodapkar A, Lim C, Cai G. TEM^{2}P^{2}EST: a thermal enabled multimodel power/performance estimator. PACS, Cambridge, MA, 2000

[16] Intel ® Core™ Core Processor T2600 on 65nm Process, DataSheet

[17] Skadron K, Stan M R, Huang W, et al. Temperature-aware microarchitecture. Proc Int Symp Computer Architecture, 2003: 2

[18] Srinivasan J, Adve S V, Bose P, et al. Exploiting structural duplication for lifetime reliability enhancement. Proc Int Symp Computer Architecture, 2005: 520

[19] Tsai Y F, Duarte D, Vijaykrishnan N, et al. Characterization and modeling of run-time techniques for leakage power reduction. IEEE Trans Very Large Scale Integration Systems, 2004, 12(11): 1221

[20] Intel® Pentium® Processor Extreme Edition 955$^{\Delta}$ on 65nm Process, DataSheet

一种热感集成电路功耗模型及其对动态电压调整技术的影响*

刘勇攀† 杨华中 汪 蕙

(清华大学电子工程系，北京 100084)

摘要： 提出了一种新型热感功耗模型，该模型能够准确估计出电压调整情况下的功耗和温度.实验结果表明如果忽略热效应，泄漏功耗将会被低估最高达 52%.使用电压调整技术对电路的能量消耗和温度进行协同优化时，两者具有不一致的优化方向.温度是未来集成电路发展的一个重要限制因素，而温度优化方法可以降低电路温度最高达 12℃，同时其能耗的增长低于最优解的 1.8%.

关键词： CMOS 集成电路；功耗模型；温度；动态电压调整

EEACC： 2570A；1265

中图分类号： TN47 **文献标识码：** A **文章编号：** 0253-4177(2007)04-0530-07

* 国家自然科学基金资助项目(批准号:90207001，60506010)

† 通信作者.Email:ypliu99@mails.tsinghua.edu.cn

2006-10-13 收到，2006-11-27 定稿

©2007 中国电子学会