PAPER

Temperature-Aware Leakage Estimation Using Piecewise Linear Power Models*

Yongpan LIU[†a)], ***Member and*** **Huazhong YANG**[†], ***Nonmember***

SUMMARY Due to the superlinear dependence of leakage power consumption on temperature, and spatial variations in on-chip thermal profiles, methods of leakage power estimation that are known to be accurate require detailed knowledge of thermal profiles. Leakage power depends on the integrated circuit (IC) thermal profile and circuit design style. Here, we show that piecewise linear models can be used to permit accurate leakage estimation over the operating temperature ranges of the ICs. We then show that for typical IC packages and cooling structures, a given amount of heat introduced at any position in the active layer will have a similar impact on the average temperature of the layer. These two observations support the proof that, for wide ranges of design styles and operating temperatures, extremely fast, coarse-grained thermal models, combined with piecewise linear leakage power consumption models, enable the estimation of chip-wide leakage power consumption. These results are further confirmed through comparisons with leakage estimates based on detailed, time-consuming thermal analysis techniques. Experimental results indicate that, when compared with a leakage analysis technique that relies on accurate spatial temperature estimation, the proposed technique yields a 59,259× to 1,790,000× speedup in estimating leakage power consumption, while maintaining accuracy.
key words: *temperature-aware, power models, leakage estimation*

1. Introduction

As a result of continued integrated circuit (IC) process scaling, which reduces transistor threshold voltages, channel lengths, and gate oxide thicknesses, the importance of leakage power consumption estimating is increasing [2]. Presently, leakage accounts for 40% of the power consumption of modern 65 nm high-performance microprocessors [3]. Without leakage reduction techniques, this ratio will increase with further technology scaling. Due to the increasing impact of leakage on IC performance, power consumption, temperature, and reliability, the leakage power must now be considered and optimized throughout the entire IC design flow, within which leakage power analysis may be invoked tens of thousands of times. Therefore, it must be both accurate and fast. Researchers have developed a variety of techniques to characterize IC leakage power consumption, ranging from the architectural level to the device level [4]–[9]. We now survey leakage analysis work spanning these design levels.

Device-level leakage power estimation generally relies on models for individual transistor leakage mechanisms. Transistor leakage power consumption is a function of the physical properties and fabrication processes of the devices. For bulk CMOS, the main control variables for leakage are the dimensions of the device(feature size, oxide thickness, junction depth, etc.) and doping profiles in transistors [9]. Based on these physical characteristics, leakage models can be developed to predict the components of leakage, e.g., subthreshold leakage, gate leakage, and junction leakage. Generally, technology constants provided by the foundry can be used in such models [10]. Transistor-level simulators [11] incorporating these models can accurately predict leakage; however, they are computationally expensive as a result of iteratively solving complex leakage formulas. Furthermore, statistical leakage analysis techniques [12] should be adopted due to the increasing process variation phenomena in the nanoscale CMOS technology.

In addition to its dependence on device parameters, IC leakage power consumption is affected by a number of circuit level parameters, e.g., the distribution of device types (NMOS and PMOS), geometries (channel width and length) and control voltages. Numerous circuit-level leakage estimation techniques have been proposed. Sirichotiyakul et al. presented an accurate and efficient average leakage calculation method for dual-threshold CMOS circuits that is based on graphical reduction techniques and simplified nonlinear simulations [13]. Lee et al. proposed fast and accurate state-dependent leakage estimation heuristics using circuit block level look-up tables, targeting both subthreshold and gate leakage [14]. To conduct full-chip leakage estimation accurately, it is possible to model and sum the leakage currents of all the gates [15], [16]. However, this is too computationally-intensive for use in the earlier design stages of very large scale integrated circuits.

For architectural leakage models [4], design parameters characterizing microarchitectural design styles and transistor sizing strategies can be extracted from typical logic and memory circuits. Do et al. proposed high-level dynamic and leakage power models to accurately estimate physically-partitioned and power-gated SRAM arrays [17]. Given a set of inputs, Gopalakrishnan et al. used a bit-slice cell library to estimate the total leakage energy dissipated in a given VHDL structural datapath [18]. Kumar et al. presented a state-dependent analytical leakage power model for FPGAs

Manuscript received March 31, 2010.
Manuscript revised July 10, 2010.
[†]The authors are with the Faculty of the Electronic Engineering Department of Tsinghua University, Beijing, 100084, P.R. China.
*This paper was presented at the conference of Design, Automation and Test in Europe [1]. This work was supported in part by the NSFC under grant 60976032, National Science and Technology Major Project under contract 2010ZX03006-003-01, and High-Tech Research and Development (863) Program under contract 2009AA01Z130.
a) E-mail: ypliu@tsinghua.edu.cn
DOI: 10.1587/transele.E93.C.1679

Copyright © 2010 The Institute of Electronics, Information and Communication Engineers

[19]. The techniques described in this paragraph provide reasonable accuracy for early design stage leakage estimation, as long as the temperature is fixed. However, they do not consider temperature variations.

IC leakage power consumption is a strong function of temperature, e.g., subthreshold leakage increases superlinearly with chip temperature. In modern microprocessors, power density has reached the level of a nuclear reactor core, causing high chip temperatures and hence high leakage power consumptions. Due to time-varying workloads and operating states with different power consumption levels (up to 25 power states in the Core™ Duo processor [20]) and uneven on-chip power density distribution (ranging from 170 to 0 Watts/cm^2 in a realistic processor [21]), large on-chip temperature variations and gradients are common in high-performance ICs. For example, SoCs may have larger than 40 °C temperature differences [22], causing high on-chip leakage variation. In summary, increasing chip temperature and on-chip temperature variation significantly affect IC leakage power consumption [23]. Therefore, accurate leakage power analysis requires consideration of the temperature.

Some researchers have developed temperature-dependent architectural leakage power models. Zhang et al. developed HotLeakage, a temperature-dependent cache leakage power model [24]. Su et al. proposed a full-chip leakage modeling technique that characterizes the impact of temperature and supply voltage fluctuations [25]. Liao et al. presented a temperature-dependent microarchitectural power model [26]. These models generally assume operation in a flow similar to that depicted in Fig. 1.

One can be confident of accurate temperature-dependent leakage estimation using a fine-grained thermal model. However, this is computationally intensive. As explained in detail in Sect. 4, existing techniques either resort to accurate thermal analysis (with costly high analysis times) or sacrifice confidence in leakage power consumption accuracy. Previous work has not demonstrated that this permits accurate leakage estimation. Without an understanding of the requirements for accurate leakage prediction, conservative designers are forced to use slow, fine-grained thermal models. This hinders the use of accurate IC leakage power estimation during IC synthesis.

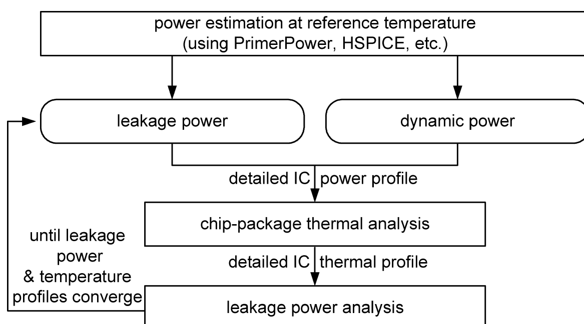

Fig. 1 Thermal-aware power estimation flow.

In this paper, we propose a fast, accurate method of estimating IC leakage power consumption.

1. We demonstrate that, within the operating temperature ranges of ICs, using a piecewise-linear leakage model with only a few segments for each functional unit results in accurate thermal-aware leakage estimation.
2. We demonstrate that IC packages and cooling structures have the useful property that a given amount of heat produced within the active layer will have a similar impact on the average temperature of the active layer, regardless of its distribution.
3. We use the preceding two properties to prove that within regions of uniform design style in a specific manufacturing process, knowledge of the average temperature is sufficient to accurately estimate leakage power consumption. Based on this result, we show that total leakage in a region can be predicted using a simple, but carefully designed, coarse-grained model without sacrificing accuracy.
4. We validate the proposed technique via analysis of potential sources of error and simulation results. We demonstrate that for a wide range of ICs, a simplified thermal model in which only one thermal element is used for each functional unit permits a speedup in leakage estimation of 59,259× to 1,790,000× while maintaining accuracy, when compared with a conventional approach that uses a thermal model of sufficient detail to permit accurate thermal analysis.

The rest of this article is organized as follows. Section 2 presents the temperature-aware leakage power model for ICs. Derivative and piecewise-linear leakage models are described and their accuracies are evaluated. Section 3 describes the proposed analysis acceleration technique and proves that, given certain cooling structure properties and a constant power profile, the area-temperature product for an IC is constant. Based on these models, we propose a fast and accurate leakage estimation method in Sect. 4. Experimental results are reported in Sect. 5. We present our conclusions in Sect. 6.

2. Proposed Leakage Model

This section reviews past work in detailed IC leakage power consumption modeling and explains how to derive an accurate piecewise-linear leakage model. Finally, the accuracy of the proposed model is evaluated.

2.1 IC Leakage Sources

IC leakage current consists of various components, including subthreshold leakage, gate leakage, reverse-biased junction leakage, punch-through leakage, and gate-induced drain leakage [9], as shown in Fig. 2. Among these, subthreshold leakage and gate leakage are currently dominant, and are likely to remain dominant in the near future [2]. They will be the focus of our analysis.

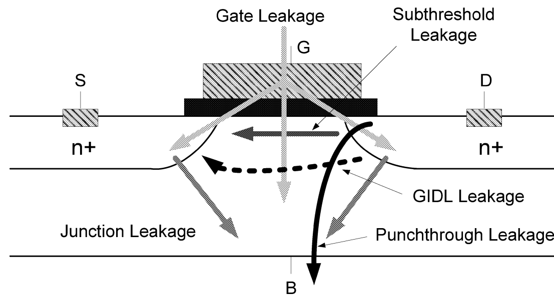

Fig. 2 Leakages current components in a MOS transistor.

Considering the weak inversion drain-induced barrier lowering and body effect, the subthreshold leakage current of a MOS device can be modeled as follows [27]:

$$I_{subthreshold} = A_s \frac{W}{L} {\upsilon_T}^2 \left(1 - e^{\frac{-V_{DS}}{\upsilon_T}}\right) e^{\frac{(V_{GS}-V_{th})}{n\upsilon_T}} \tag{1}$$

- where A_s is a technology-dependent constant,
- V_{th} is the threshold voltage,
- L and W are the device effective channel length and width,
- V_{GS} is the gate-to-source voltage,
- n is the subthreshold swing coefficient for the transistor,
- V_{DS} is the drain-to-source voltage, and
- υ_T is the thermal voltage.

$V_{DS} \gg \upsilon_T$ and $\upsilon_T = \frac{kT}{q}$. Therefore, Eq. (1) can be reduced to

$$I_{subthreshold} = A_s \frac{W}{L} \left(\frac{kT}{q}\right)^2 e^{\frac{q(V_{GS}-V_{th})}{nkT}} \tag{2}$$

Equations (1), (2) demonstrate that subthreshold leakage depends primarily on temperature, supply voltage, and body bias voltage. Gate leakage, in contrast, is primarily affected by supply voltage and gate dielectric thickness, but is insensitive to temperature. Using the Taylor series expansion at a reference temperature T_{ref}, the total IC leakage current of a MOS device can be expressed as follows:

$$\begin{aligned} I_{leakage}(T) &= I_{subthreshold} + I_{gate} \\ &= A_s \frac{W}{L} \left(\frac{k}{q}\right)^2 T^2 e^{\frac{q(V_{GS}-V_{th})}{nkT}} + I_{gate} \\ &= I_{linear}(T) + I_{high_order}(T) \end{aligned} \tag{3}$$

where the linear portion $I_{linear}(T)$ is

$$\begin{aligned} I_{linear}(T) &= I_{gate} + A_s \frac{W}{L} \left(\frac{k}{q}\right)^2 e^{\frac{q(V_{GS}-V_{th})}{nkT_{ref}}} \\ &\times \left({T_{ref}}^2 + (2T_{ref} - \frac{q(V_{GS}-V_{th})}{nk})(T - T_{ref})\right) \end{aligned} \tag{4}$$

and the high-order portion of $I_{high_order}(T)$ is

$$I_{high_order}(T) = I''_{leakage}(T_{ref})(T - T_{ref})^2 + O((T - T_{ref})^3) \tag{5}$$

Therefore, the estimation error resulting from the truncation of the super-linear terms is bounded as follows:

$$Err_{dev} = \left|\frac{I_{high_order}(T)}{I_{leakage}(T)}\right| \tag{6}$$

Equations (5), (6) demonstrate that the estimation error of the linear leakage power model is a function of $|T - T_{ref}|$, i.e., the difference between the actual circuit temperature T and the reference temperature T_{ref} at which the linear model is derived. Therefore, to minimize the estimation error, the linear leakage model should be derived as close as possible to the actual sub-circuit temperature.

2.2 Piecewise-Linear Model Extraction

To build the piecewise-linear leakage model, we can characterize IC leakage power at different temperatures using simulation or measurement. We have developed a gate-level leakage analysis flow to estimate IC leakage at different temperatures. The flow contains two steps:1) Setup a library characterizing the leakage of each cell under different parameters, such as manufacture process, temperature, input patterns, supply voltage and body bias voltage, and 2) Count the IC cell number and sum the total leakage current.

Step 1 is time consuming because HSPICE simulation must be used for each combination of the leakage-related parameters. Fortunately, this step is only necessary once per cell library. After the library is built, the number and type of cell or block can be extracted very quickly using Synopsys Design Compiler or SRAM compiler in Step 2. An IC is divided into regions within which all cells have similar leakage characteristics. For example, logic and memory would be divided into different regions. Since input vectors have a great influence on leakage current, we assign a specific probability to each cell or block for each input vector and use the input vector probabilities to estimate their leakage power consumption. For other applications, we can extract the switching information, denoted as input factors, from design tools, such as VCS and PrimePower by running corresponding benchmarks. Detailed information on input patterns can improve estimation accuracy. However, the proposed method is independent of any specific input factor. Part I in Fig. 3 summarizes the simulation-based leakage estimation flow.

Part II in Fig. 3 illustrates the proposed approach to extract a piecewise-linear leakage model. As illustrated in Fig. 3, the method uses linear least squared error curve fitting on each line segment. The proposed piecewise linear leakage model can be built based on both the simulated circuits leakage-temperature pairs, as well as the measured circuits leakage-temperature pairs. We have also demonstrated this flow based measured data from Freescale Semiconductor in Fig. 7. In this approach, users must decide the number of line segments and determine each segment's temperature range. Usually, the curve is nearly-linear over the operating temperature range of an integrated circuit, as shown in

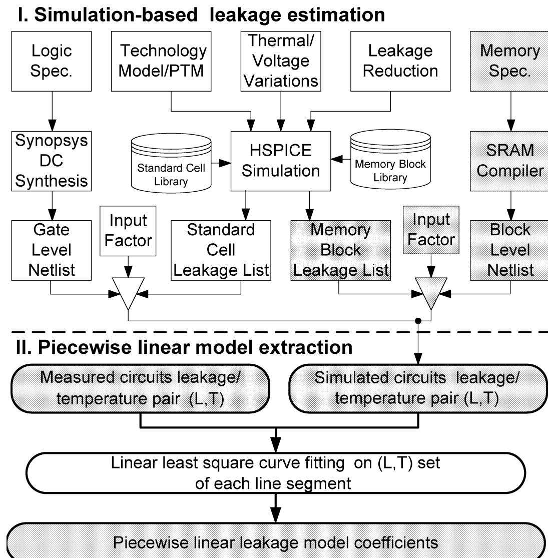

Fig. 3 Piecewise-linear leakage model extraction flow.

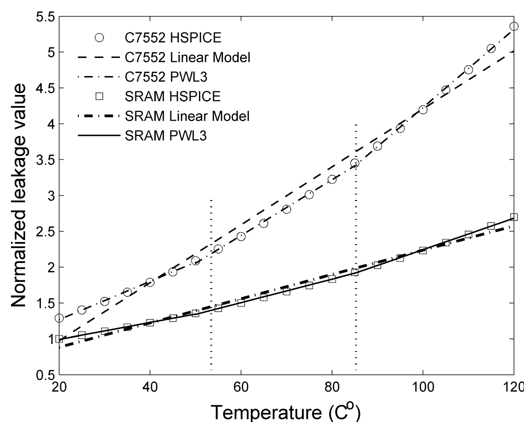

Fig. 4 Normalized leakages for HSPICE, piecewise-linear, and linear models using the 65 nm process for c7552 and SRAM.

Fig. 4. Using three equal-size segments results in an error less than 1% for all of the circuits we evaluated. The user can adjust the number of segments based on the required accuracy. To obtain the coefficients, linear least-squared error curve fitting is done for each segment. This approach requires a manual trade off between accuracy and model complexity.

2.3 Model Accuracy

We used simulation and comparisons with measured data to evaluate the accuracy of the proposed piecewise-linear model. Figure 4 presents the normalized leakage power consumptions of two circuits (a combinational circuit c7552 [28] and SRAM [29]) as functions of temperature. For each circuit, we compare linear and three-segment piecewise-linear (PWL 3) models with HSPICE simulation results for a 65 nm predictive technology model process [30]. Within the normal operating temperature ranges of many ICs, 55 °C–85 °C, even a linear model is fairly accurate. This accuracy can be further improved using a piecewise-linear model. Accuracy improves with segment count although, in practice, only a few segments are needed.

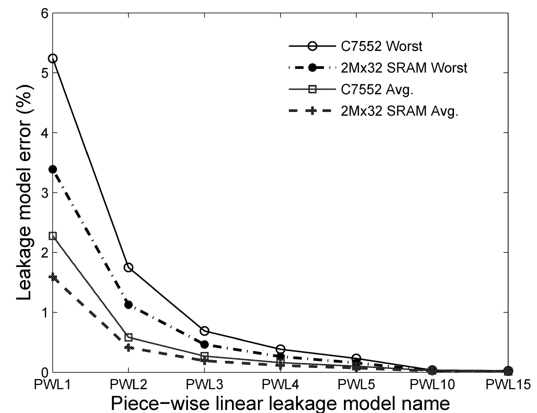

Fig. 5 Linear leakage model error trend for c7552 and SRAM under different segment configurations.

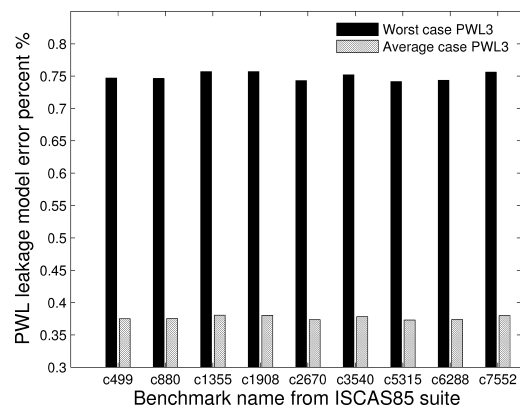

Fig. 6 Linear leakage model errors for ISCAS85 benchmark suites using three-segment piecewise-linear model.

Figure 5 shows average and maximum leakage power model errors as functions of the piecewise-linear model segment count for the same two circuits considered in Fig. 4. The error is reported relative to the HSPICE simulation. Leakage was modeled in the temperature range of 40 °C–110 °C, i.e., a typical normal operating temperature range. Within each piecewise-linear region, a linear leakage model is derived at the average temperature of this region using Eq. (4). The accuracy permitted by the piecewise-linear model is determined by the granularity of the regions. Figure 5 shows that the modeling error decreases as the number of linear segments increases. For three or more segments, the maximum errors are less than 1% for both c7552 and SRAM. Furthermore, Fig. 6 also indicates average and worst-case errors for different benchmarks in the ISCAS85 suite using PWL3 models. The worst-case errors are always below 0.8% and the average error is less than 0.4%. These results indicate that coarse-grained piecewise-linear models permit good leakage estimation accuracy. Finer granularity

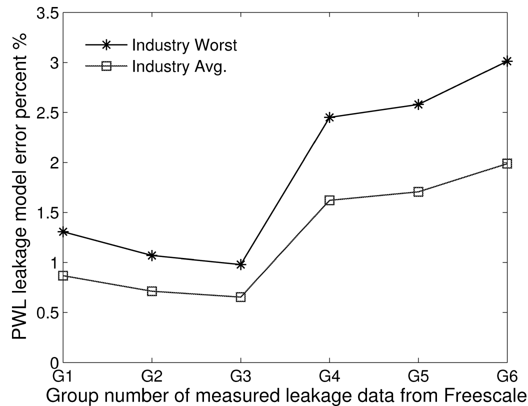

Fig. 7 Piecewise-linear model error with measured leakage power data.

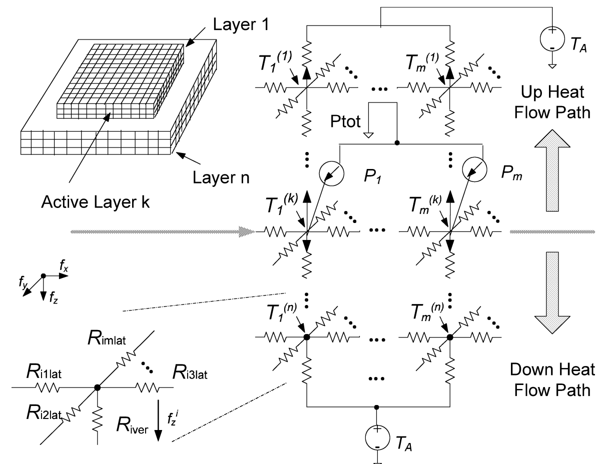

Fig. 8 Heat flow in a typical IC thermal package.

or differentiation of curve fitted continuous functions will generally further improve accuracy, at the price of increased complexity.

To determine the accuracy of the piecewise-linear model for industrial circuits, we applied the proposed modeling process to the measured leakage power consumption values provided by industrial collaborators at Freescale Semiconductor. The linear least-squared error method was used to extract the piecewise-linear model parameters. The resulting modeling errors are plotted in Fig. 7. Freescale provided six groups of measured leakage power data for an embedded microprocessor design. Each group corresponds to a 60 °C temperature range. For the three low-temperature groups (G1, G2, G3), the leakage power is measured from 25 °C to 85 °C. For the three high-temperature groups (G4, G5, G6), the leakage power is measured from 45 °C to 105 °C†. As Fig. 7 shows, the accuracy of linear modeling decreases as temperature increases. The average and worst-case errors are below 2% and 3%. Both simulated and measured experimental results show that the proposed piecewise-linear model accurately describes the relationship between leakage power consumption and temperature within the operating temperature range of the IC.

3. Thermal Model and Properties

This section introduces a thermal modeling technique commonly used in detailed temperature-aware IC leakage estimation and explains the properties of IC cooling solutions that enable the proposed leakage analysis technique. There are alternative thermal modeling techniques but the trade off between modeling accuracy and computation time is typical for thermal modeling.

3.1 Thermal Model Introduction

To conduct numerical thermal analysis, the IC chip and package are partitioned into numerous isothermal elements. This permits heat flow to be modeled in the same manner as electrical current in a distributed *RC* network.

$$\mathbf{C}\frac{\mathrm{d}\vec{T}(t)}{\mathrm{d}t} = \mathbf{A}\vec{T}(t) - \vec{p}\,U(t) \tag{7}$$

where

- $\mathbf{C}$ is an $\lambda \times \lambda$ diagonal thermal capacitance matrix,
- $\mathbf{A}$ is an $\lambda \times \lambda$ thermal conductance matrix,
- $\vec{T}(t) = [T_1 - T_A, T_2 - T_A, \cdots, T_\lambda - T_A]^{\mathrm{T}}$ is the temperature vector in which T_A is the ambient temperature,
- $\vec{p} = [p_1, p_2, \cdots, p_\lambda]^{\mathrm{T}}$ is the power vector, and
- $U(t)$ is the unit step function.

In steady-state thermal analysis, the thermal profile does not vary with time. Therefore, we can denote $\lim_{t\to\inf} \vec{T}(t)$ as $\vec{T}$, allowing Eq. (7) to be simplified as follows:

$$\vec{p} = \mathbf{A} \times \vec{T} \tag{8}$$

The thermal resistance matrix $\mathbf{R}$ is the inverse of the thermal conductance matrix, i.e., $\mathbf{R} = \mathbf{A}^{-1}$.

3.2 Useful Property of IC Cooling Structure

In this section, we list several reasonable assumptions to show that most realistic cooling configurations have a property, i.e. the sum of products area-temperature conservation property.

A typical IC thermal model is shown in Fig. 8. To accurately model spatial temperature variation, an adequate number of layers of thermal elements is generally necessary between the active layer and heat sink. We denote the layer number as n. Assuming an IC floorplan within which the active layer is divided into m isothermal blocks, blk_i, $i \in \{1, 2, \cdots, m\}$, the temperature, area, and power consumption of blk_i are expressed as T_i, s_i, and p_i. The total power consumption of the chip is $P_{tot} = \sum_{i=1}^{m} p_i$. The matrix, $\mathbf{S}$, holds the values of the vector $\vec{s}$, $[s_1, s_2, \cdots, s_m]$ along its diagonal. We now present a useful property of IC cooling solutions that permits the use of the proposed leakage

†Other information is not available due to restrictions imposed by a non-disclosure agreement limits with Freescale™ company.

estimation technique.

First of all, we list the assumptions held for the property:

1. The thermal element is made vertically isothermal in each layer by arbitrarily reducing the thickness of the thermal element.
2. Each layer has uniform material conductivity and thickness and does not have holes.
3. All heat generated in the active layer flows eventually to the ambient through the top of the heatsink and the bottom of the package.

The above assumptions do not strictly hold in real cooling solutions. However, Sect. 5 contains numerical simulations indicating that under various realistic cooling configurations, the following Theorem 1 is a good approximation to reality.

Theorem 1 (Thermal Property of Cooling Solution): While the following assumptions are satisfied, for the active layer of an integrated circuit divided into discrete elements, the sum of the products of element temperatures and areas

$$\sum_{i \in L_0} s_i T_i^{L^0} \text{ is constant} \tag{9}$$

as long as the total power in the active layer L^0 is constant.

Table 1 defines terms used in the proof. In order to simplify the proof, we transform two heat paths into a single equivalent one heat flow path. We have proved and validated the transformation is correct by simulations in [31]†.

We shall start from L^n, which represents the nth layer and the farthest layer from the active layer. Based on previous assumptions, the chip and package can be discretized as illustrated in Fig. 8. The temperature T_i^n of each element i in the layer L^n is decided by the vertical heat flow and can be expressed as follows:

$$T_i^n = T^A + f_i^z s_i R_{iver} = T^A + f_i^z s_i \frac{t^n}{k^n s_i} \tag{10}$$

As Eq. (10) has shown, the horizontal heat flow does not contribute to the temperature due to Assumption 3. Since we assume that each element in one layer has uniform thermal conductivity and thickness, using Eq. (10), Eq. (11) can be obtained as follows:

$$\sum_{i \in L^n} s_i T_i^n = \sum_{i \in L^n} s_i \left(T^A + \frac{f_i^z t_i^n}{k_i^n} \right) = T^A \sum_{i \in L^n} s_i + \frac{t^n}{k^n} \sum_{i \in L^n} s_i f_i^z \tag{11}$$

It is straightforward to see that the first term $T^A \sum_{i \in L^n} s_i$ is constant. The second term $\frac{t^n}{k^n} \sum_{i \in L^n} s_i f_i^z$ is also a constant as long as the total vertical heat flow f^z through the layer is a constant under the homogeneous layer partition (Assumption 2). The condition is obviously satisfied, since we assume no heat can flow through the sides of the chip package.

Next, consider layer L^{n-1}. For each element i in layer L^{n-1}, which contains m elements in total, we have the following expression according to Fig. 8.

$$\begin{aligned} \sum_{i \in L^{n-1}} s_i T_i^{n-1} &= \sum_{i \in L^{n-1}} s_i \left(T_i^n + \frac{f_i^z t^{n-1}}{k^{n-1}} \right) \\ &= \sum_{i \in L^{n-1}} s_i T_i^n + \frac{t^{n-1}}{k^{n-1}} \sum_{i \in L^{n-1}} s_i f_i^z \end{aligned} \tag{12}$$

Equations (10), (11) show that the term $\sum_{i \in L^{n-1}} s_i T_i^n$ is constant. Term $\frac{t^{n-1}}{k^{n-1}} \sum_{i \in L^{n-1}} s_i f_i^z$ is also constant because all heat generated in the active layer must flow through layer L^{n-1} (Assumption 3). Induction can be used to push heat to the active layer. Therefore, the sum of the products of element temperatures and areas $\sum_{i \in L_0} s_i T_i^{L^0}$ remains constant as long as the total power input remains constant.

4. Fast Temperature-Aware Leakage Estimation

In this section, we start from the conventional fine-grained thermal-aware leakage analysis, from which we propose fast and accurate temperature-dependent leakage estimation methods for chips with uniform and non-uniform leakage characteristics. We then discuss the use of the proposed method in real designs.

4.1 Proposed Method

Assume the IC is divided into n isothermal homogeneous grid elements, blk_i, $i \in 1, 2, \cdots, n$. The temperature, area, and power consumption of each element, blk_i, are expressed as T_i, s_i, and p_i. Using the linear or piecewise-linear leakage model developed in Sect. 2, the leakage power of blk_i is expressed as follows:

$$p_{leak}^{blk_i}(T_i) \simeq V_{DD} I_{linear}^{blk_i}(T_i) \tag{13}$$

For a subcircuit with uniform design style, the leakage current is linearly proportional to the temperature, yielding the following formula:

Table 1 Terms and definitions in the proof.

Term	Definition
s_i	area of thermal element i
T_i^j	temperature of thermal element i in layer j
T^A	ambient temperature
f_i^x	x-axis horizontal heat flow for thermal element i
f_i^y	y-axis horizontal heat flow for thermal element i
f_i^z	z-axis vertical heat flow for thermal element i
L^0	thermal elements in active layer
L^n	thermal elements in layer farthest from active layer
L_t^i	terminal thermal elements in layer i
L_n^i	non-terminal thermal elements in layer i
k^l	material thermal conductivity in layer l
t^l	the thickness of layer l

†For the sake of clarity, we will provide a direct inductive proof based on heat transfer equation below. The first author's Ph.D. dissertation provides a longer, less intuitive, but more general indirect matrix-based proof [31].

$$I^{blk_i}_{linear}(T_i) = I_i(T_0)(\eta_i T_i + \xi_i) \tag{14}$$

where $I_i(T_0)$ is the leakage current per block at the reference temperature T_0. This value depends on the manufacturing technology, design style, supply voltage, and input pattern. Since input vectors influence leakage current, it should be weighted by the input vector probabilities. η_i and ξ_i are parameters obtained by curve fitting in the piecewise-linear model. Collectively, $I_i(T_0)$, η_i, and ξ_i are referred to as *leakage coefficients*.

Uniform Case: Without considering thermal factors, the leakage coefficients $I_i(T_0)$, η_i, and ξ_i are decided only by the circuit design style, supply voltage, and input pattern. For an IC with uniform design style and supply voltage, such as SRAM and field-programmable gate arrays (FPGAs), these values are the same under specific input patterns for all portions of the IC in a linear power model and can be denoted as $I_{tech}(T_0)$, η, ξ. Theorem 1 can be used to show that:

$$\sum_{i=1}^{n} I^{blk_i}_{linear}(T_i) = I_{tech}(T_0) \sum_{i=1}^{n} (\eta T_i + \xi) = nI_{tech}(T_0)(\eta T_{avg} + \xi) \tag{15}$$

Therefore, as long as the conditions necessary to use Theorem 1 are well satisfied and the linear power model is used, we will show later that only one thermal element is needed to calculate the T_{avg} of the entire IC. This permits highly-efficient leakage estimation.

Nonuniform Case: Many ICs are composed of regions with different design styles, e.g., logic and memory, or with different supply voltages. These regions have different $I_i(T_0)$, η_i, and ξ_i values. Furthermore, even if in the uniform design style ICs with quite large temperature variations, e.g, in Sect. 2 over 30 °C temperature variations and less than 1% leakage power modeling error will require adoption of piecewise linear modeling, in which leakage coefficients would be different in different temperature regions. In this case, we divide the chip into p regions, within which the leakage coefficients are consistent. In fact, proper partition in leakage curves are needed to build the piecewise-linear model and to keep the leakage coefficients consistent in each region. Therefore, the IC leakage current is expressed as follows:

$$\sum_{k=1}^{p} \sum_{i=1}^{q_k} I^{blk_i}_{linear}(T_i) = \sum_{k=1}^{p} I_k(T_0) \sum_{i=1}^{q_k} (\eta_k T_i + \xi_k) = \sum_{k=1}^{p} q_k I_k(T_0)(\eta_k T^{reg}_k + \xi_k) \tag{16}$$

where T^{reg}_k is the average temperature of region k. By summing the leakage current of all regions, the total leakage current is obtained. In contrast with other traditional fine-grained thermal analysis methods, we will show that the use of only one, or a few, thermal elements for each region allows extremely fast and accurate thermal analysis for the average temperature T^{reg}_k.

4.2 Fast Average Temperature Calculation

This paragraph shows how to calculate the average temperature based on efficient coarse-grained thermal analysis.

For the uniform design style IC, Theorem 1 shows that the sum of the products of element temperatures and areas

$$\sum_{i \in L_0} s_i T_i^{L^0} \text{ is constant} \tag{17}$$

as long as the total power in the active layer L^0 is constant. Therefore, we can assume the temperature is evenly distributed across the chip and a single equivalent thermal resistance can be used to calculate T_{avg}. The thermal resistance can be obtained using the conventional analytical method, which can be expressed by the following equation:

$$T_{avg} = P_{tot} \cdot R_{ver} = \frac{P_{tot} \cdot t_{blk}}{k_{blk} \cdot A_{blk}} \tag{18}$$

where t_{blk} is the thickness of the block, k_{blk} is the thermal conductivity of the material of that block, and A_{blk} is the cross-sectional area of the block. For the nonuniform design style IC, we also need to find an efficient way to calculate the average temperature for each region with different design styles. According to the relationship between heat flow and electrical current, we can extract the vertical thermal resistance using Eq. (18). The lateral thermal resistance between regions can be obtained using the spreading resistance approach in Hotspot [32]. It builds an equivalent thermal resistance network to estimate the average temperature of each region. Since the granularity of region based partitions is much smaller than the conventional fine-grained approach, the thermal analysis speed is extremely fast. Our experimental results show that the average temperature calculated by such a coarse-grained thermal model is generally consistent with the results of the fine-grained model for different real chip layouts, even under extremely imbalanced power distributions.

4.3 Discussion on Usage

Based on our conclusions in Sects. 2 and 3, we can provide upper and lower bounds on the IC leakage power consumption. In general, there is a difference in $I_i(T_0)$ between blocks with different design styles. Therefore, according to Eq. (15), the upper and lower bounds follow:

$$\sum_{i=1}^{n} L^{blk_i}(T) \leq I_{max}(T_0) \sum_{i=1}^{n} (\eta_{max} T_i + \xi_{max})$$
$$\sum_{i=1}^{n} L^{blk_i}(T) \geq I_{min}(T_0) \sum_{i=1}^{n} (\eta_{min} T_i + \xi_{min}) \tag{19}$$

where $I_{max}(T_0)$, $I_{min}(T_0)$, η_{max}, η_{min}, ξ_{max}, and ξ_{min} are $max(I_i(T_0))$, $min(I_i(T_0))$, $max(\eta_i)$, $min(\eta_i)$, $max(\xi_i)$, and $min(\xi_i)$. We can use Eq. (19) to bound the total leakage power of the chip. These bounds are not tight enough for

ICs with large on-chip thermal gradients or dramatically-different design styles. However, they are useful for ICs with moderate temperature gradients and design style variations, such as low power FPGAs, SRAMs, or other ICs with regular structures. In reality, users can calculate leakage coefficients for different regions under estimated thermal gradients to see if the bounds are good enough for their purpose.

Furthermore, Sect. 3 draws the conclusion that, if the regions in an IC have differing leakage characteristics, physical designs that reduce the temperature in the high-leakage coefficient regions will reduce the total leakage power, even if the temperatures of the low-leakage coefficient regions increase by the same amount. Thus, proper physical design can reduce leakage power by changing the power and thermal distribution. The optimized results should be bounded by Eq. (19).

In practice, most block-based designs can be partitioned into a few functional blocks for thermal-aware leakage estimations. Both our experimental results in Sect. 2 and other independent work [26] showed that blocks constructed with similar standard cells and structures have consistent leakage coefficients. Therefore, according to its function partitioning, the chip can be divided into regions with different design styles. It should be noted that users can further partition a functional unit into smaller regions based on its detailed circuit structure if more consistent leakage coefficients are needed. However, such partitioning is not beneficial without detailed structural information, which is often not available in the early design stages.

The importance of process variation is likely significant in future deep-submicron processes. We will briefly discuss its implications on efficient temperature-aware leakage modeling. A thorough treatment is beyond the scope of this article. Process variations can be divided into two categories: inter-die variation and intra-die variation. Inter-die variation can be considered directly, using Eq. (19) to bound the leakage power for different corner cases. Given the knowledge of the detailed variation probability distribution functions, more detailed analysis would be possible. Intra-die variation requires a more sophisticated treatment. We recall that separate thermal elements are necessary for each region of uniform leakage coefficients. As a consequence, the magnitudes and distance scales over which the leakage-related process variation affects a particular process, will influence the required number of discrete thermal elements. We intend to analyze this in more detail in the future.

5. Experimental Results

In this section, we evaluate the accuracy and efficiency of the proposed temperature-dependent leakage estimation technique, which consists of piecewise-linear leakage modeling and coarse-grained thermal analysis. We characterize the two sources of leakage estimation error introduced by this technique: truncation error as a result of using a linear leakage model and temperature error as a result of using a coarse-grained thermal model. The base case for comparison is conventional temperature-aware leakage estimation using a super-linear leakage model and fine-grained thermal analysis [25], where millions of thermal elements are used. Our experiments demonstrate that for a set of FPGA, SRAM, microprocessor, and application specific integrated circuit (ASIC) designs, the proposed leakage modeling technique is accurate and permits great increases in efficiency. All analysis runs used an AMD Athlon-based Linux PC with 1 GB of RAM.

5.1 Experimental Setup

We use the 65 nm predictive technology model [30] for leakage modeling. This model characterizes the impact of temperature on device leakage. We first derive the super-linear leakage model using HSPICE simulation. The piecewise-linear leakage model is then derived using the method described in Sect. 2, i.e., partitioning the temperature range into uniform segments and using least-squared error fitting for each segment.

We use HotSpot 3.0 [32] for both coarse-grained and fine-grained steady-state thermal analysis. HotSpot 3.0 supports both block-based coarse-grained and grid-based fine-grained stead-state thermal analysis. Previous work demonstrated that the coarse-grained block-based method is fast [33]. In contrast, fine-grained grid-based partitioning is slower but permits more accurate thermal analysis. In this work, coarse-grained thermal analysis uses the block-based method, as only the average block temperature is required. For fine-grained thermal modeling, we partition the IC active layer into 100×100 elements. This resolution is necessary; as decreasing resolution to 50×50 resulted in a 6 °C error in peak temperature for the Alpha 21264 for the SPEC2000 gcc power trace programs. A resolution of 100×100 elements is also sufficient for our benchmarks. We have used resolutions up to $1{,}000 \times 1{,}000$ to validate our results and have found that increasing resolution beyond 100×100 has little impact on temperature estimation accuracy.

5.2 Leakage Power Estimation

This section gives experimental results for leakage power estimation using piecewise-linear power modeling and coarse-grained thermal analysis. We consider a number of different power profiles and design styles. First of all, we show the accuracy of the proposed approach in an FPGA under uniformly distributed random power profiles. It represents the accuracy of the method for general cases. Second, a quite unbalanced power profile is used to show the accuracy of the method under pessimistic conditions. Third, to validate the proposed method in chips with different design styles, a microprocessor is used to show its accuracy. Table 2 shows the accuracy and speedup resulting from using the proposed leakage estimation technique on an FPGA [34]. We used six sets of 30 random power profiles, con-

Table 2 Leakage error for FPGA.

T_{avg} (°C)	P_{tot} (W)	DM error Avg. (%)	DM error Max. (%)	CPU time SF (s)	CPU time DM (μs)	Speedup (million ×)
40	10	0.003	0.005	16.1	10	1.60
50	40	0.039	0.092	14.7	10	1.47
60	70	0.122	0.258	16.1	10	1.61
70	110	0.300	0.650	16.2	10	1.62
80	150	0.505	0.960	16.2	9	1.79
90	180	0.731	1.205	16.0	9	1.78

sidering both dynamic and leakage power. Six different total power consumptions (Column 2) resulting in different average temperatures (Column 1) were considered. Power profiles were generated by assigning uniformly-distributed random samples ranging from [0, 1] to each cell in a 5 × 5 array overlaying the IC and then adjusting the power values to reach the target total IC power while maintaining the ratios of power consumptions between cells.

In Sect. 4 we showed that the leakage power of an IC with uniform leakage coefficients depends only on total power consumption. To verify this claim, we compare the superlinear fine-grained model (SF) with the single-element linear derivative-based model (DM). At each total power setting, the average estimation error for the 30 randomized power profiles is shown in Column 3. As shown in Column 4, the maximum estimation error was never greater than 1.2%. As shown in Columns 5–7, the speedup permitted by our technique ranges from 1,470,000× to 1,790,000×. This speedup results from a reduction in thermal model complexity that greatly accelerates the thermal analysis portion of leakage estimation.

In addition to considering modeling accuracy for uniform leakage coefficients in the presence of randomized power profiles, we designed a power profile to determine the error of the proposed technique under pathological conditions. In this configuration, all of the power in the IC is consumed by a corner block and other blocks consume no power. The total power input is set to 117 W, leading to an extremely unbalanced thermal profile. Temperatures ranged from 52.85 °C to 106.85 °C. This case goes well beyond what can be expected in practice, but serves to establish a bound on the estimation error of the proposed approach. Figure 9 shows the leakage estimation error as a function of thermal modeling granularity for piecewise-linear thermal models with various numbers of segments and a linear model based on the derivative of the continuous leakage function at the predicted temperature of the block. Using the same one-segment linear model for all blocks (PWL 1) results in an approximately 2% estimation error. However, piecewise-linear models with five or more segments, and the derivative-based model, all maintain errors of less than 0.5%, as long as at least four thermal elements are used. Note that the derivative based model is not identical to a piecewise-linear model in which the number of segments approaches infinity because the piecewise-linear model is fitted to the leakage function using a least-squared error minimizer while the derivative based model

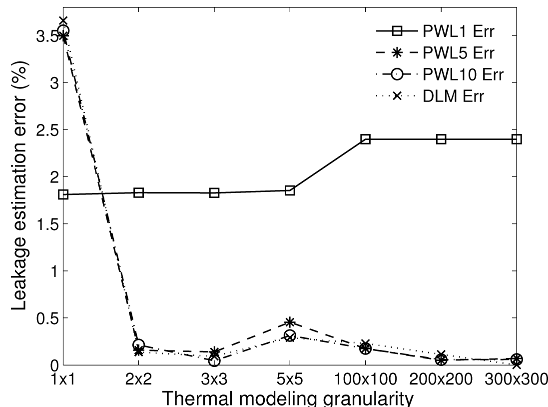

Fig. 9 Leakage estimation error of FPGA for worst-case power profile.

Table 3 Leakage error for alpha 21264.

Benchmark		gcc	equake	mesa	gzip	art	bzip2	twolf
Error (%)	PWL 5	0.52	0.71	0.53	0.42	0.34	0.45	0.65
	DM	0.54	0.64	0.51	0.48	0.56	0.47	0.57
Speedup (thousand ×)		59	67	65	81	66	67	66

uses a Taylor series expansion around a single temperature. Therefore, it is possible for the piecewise-linear model to result in higher accuracy in some cases. From these data, we can conclude that even when faced with extreme power profiles, only a few thermal elements are necessary to permit high leakage power estimation accuracy. In addition to considering ICs with uniform design styles, e.g., FPGAs, we have evaluated the proposed technique when used on the Alpha 21264 processor, an IC that has regions with different sets of leakage coefficients, e.g., control logic, datapath, and memory. Power traces were generated using the Wattch power/performance simulator [35] running SPEC2000 programs. One thermal element is used for each functional unit in the processor. Most existing architectural power models provide a uniform power distribution in each function unit. Recently-published measured results [36] indicate that the power profile of a microprocessor in normal operation is nearly-uniform over spatial ranges of 400 μm which is greater than the typical length scale of an on-chip functional unit. Table 3 shows results for five-segment piecewise-linear (PWL 5) and derivative-based (DM) leakage models. Row 4 shows that reducing thermal model complexity results in leakage estimation speedups ranging from 59,259× to 80,965×. As Rows 2 and 3 show, derivative-based and piecewise-linear model leakage estimation errors are less than 1% for all benchmarks, compared with a HSPICE-based superlinear leakage model used with fine-grained thermal analysis. This small error has two components: truncation error resulting from the piecewise-linear model and a slight deviation in the average temperature calculation.

5.3 Thermal Model Error Breakdown

In Sect. 3, we showed the necessary and sufficient condi-

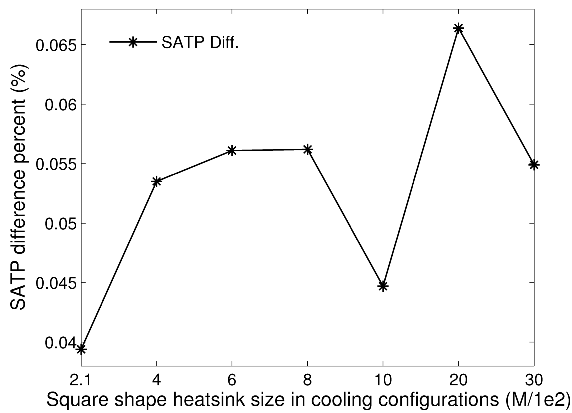

Fig. 10 Heatsink shape's effect on the SATP.

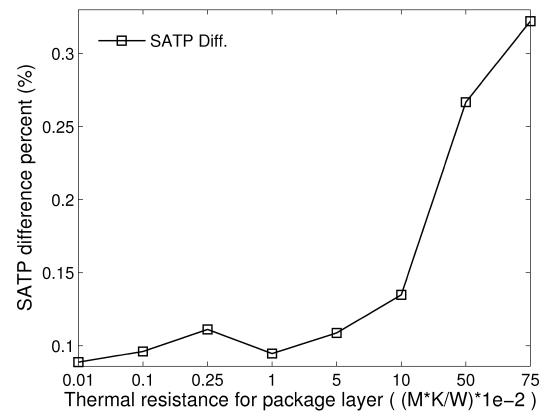

Fig. 11 Thermal conductivity's effect on the SATP.

tions for Theorem 1 to hold under reasonable assumptions. In this section, we first show that the conditions required for Theorem 1 to hold with small errors are satisfied by numerous real cooling configurations, e.g., plastic packages used for low-power circuit designs and ceramic packages with forced-air heatsink cooling solutions for high-performance designs. Second, we illustrate that for a specific cooling solution, Theorem 1 holds for several ICs with differing floorplans and uniformly-distributed random power profiles. Finally, we show that the average temperature of the uniform design style region of an IC can be accurately and efficiently estimated given either random or real power profiles for different ICs. We will then discuss the even power distribution limit on the leakage power estimation accuracy, where some extreme imbalanced power profiles in specific function units are used to evaluate the proposed method. These properties substantiated in this section, along with the piecewise-linear power model, yield the proposed efficient and accurate leakage estimation techniques.

Different cooling configurations have various heatsink sizes and package thermal resistances, which may have an effect on the validation of Theorem 1. Therefore, we consider these effects on the proposed method. Assume the IC has an 4 mm^2 area, and the total power consumption is 50 W. Figure 10 compares the sum of the area-temperature products (SATP) between worst-case and even distributed power profiles of an IC with different heatsink sizes. As we can see, the SATP error is the smallest when the heatsink has an equal size of the bulk silicon, which is predicted by Theorem 1. Though the error increases in a range with larger heatsink sizes, the maximum value is still below 0.07%. Figure 11 gives out the SATP difference by changing the thermal conductivity of the package layer material. As we can see, the difference increases, when poor thermal conductivity materials are used. However, the worst-case difference is below 0.4%. These results clearly indicate that Theorem 1 holds with trivial deviations for different regular cooling solutions. We now use several ICs with differing floorplans: FPGA, SRAM [37], Alpha 21264, and HP, an ASIC benchmark from the MCNC benchmark suite [38], to compare SATP values given different power profiles. For

Table 4 $\sum_{i=1}^{n} s_i T_i$ with different power profiles.

T_{avg} (°C)	FPGA SATP Error (‰)		SRAM SATP Error (‰)		EV6 SATP Error (‰)		HP SATP Error (‰)	
	Avg.	Max.	Avg.	Max.	Avg.	Max.	Avg.	Max.
40	0.016	0.019	0.013	0.018	0.002	0.003	0.202	5.407
50	0.057	0.075	0.097	0.131	0.099	0.115	0.085	1.458
60	0.099	0.113	0.189	0.247	0.180	0.204	0.139	2.116
70	0.145	0.169	0.280	0.361	0.263	0.302	0.168	2.093
80	0.178	0.217	0.337	0.472	0.338	0.389	0.177	1.788
90	0.224	0.282	0.424	0.570	0.421	0.514	0.215	1.913

each IC, SATP is calculated for 30 randomized power profiles, which are generated in the same way as those for Table 2. Each IC has a different area ranging from 8 mm^2 to 400 mm^2. The block number in each chip ranges from 1 to 18. Therefore, total power consumption values were chosen to produce each of the six reported average temperatures. Table 4 shows maximum and average differences between the SATP values for the random power profiles and the SATP value for a uniform power profile. From these results, we can conclude that the SATP error is less than 0.6% for all four benchmark ICs. We also computed the SATP error for the unbalanced worst-case power FPGA profile used in Fig. 9. The worst-case error is smaller than 0.015% for all thermal model granularities. We conclude that the conditions required to use Theorem 1 are well-satisfied for a wide range of ICs.

Although we have shown that the properties required to use Theorem 1 are well-approximated for a number of ICs, we have yet to show the implications of this observation upon average temperature estimation for uniform design style blocks. We partitioned the IC into blocks, each of which corresponds to a region with uniform leakage coefficients, and compare the average block temperatures with those calculated by using a fine-grained thermal model. Figure 12 shows the maximum temperature estimation error as a function of average IC temperature for the same set of benchmarks as shown in Table 4. The error is computed on the Kelvin scale. Figure 12 shows that the maximum temperature estimation error over all power profiles is less than 1.1%. Figure 13 illustrates the average temperature errors for the Alpha21264 processor when used with power traces extracted from the SPEC2000 benchmarks using the Wattch

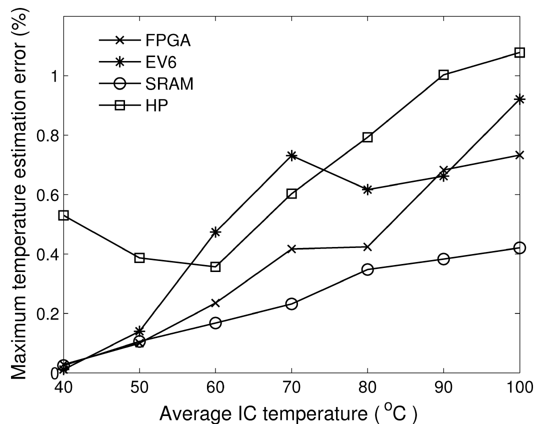

Fig. 12 Thermal error breakdown among different types of ICs.

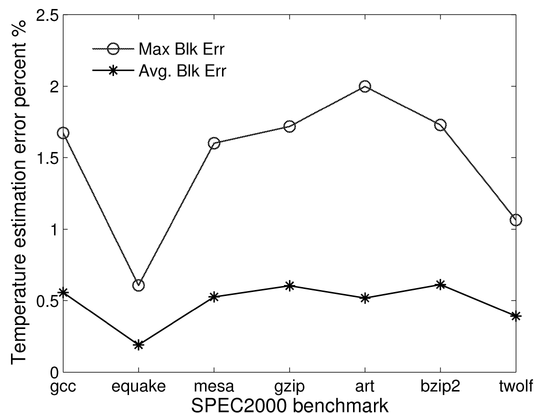

Fig. 13 Thermal error breakdown in alpha processor.

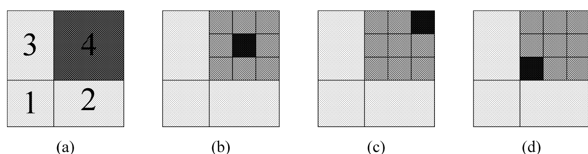

Fig. 14 Impact on the average temperature by difference power distribution in a region.

[35] power simulator. The circle curve and star curve describe the maximum and average temperature error between all the blocks. In all cases, the average block error is less than 0.61%, which affects the leakage estimation accuracy.

In the above experiments, we assume even power distributions in all function blocks because no more information is available in the architecture level design stage. We developed the following case study to justify the applicability of using a coarse-grained thermal model to calculate the average temperature given a heterogeneous power condition. Figure 14(a) assumes a uniform power distribution in a region, and Figs. 14(b)–(d) show a severely heterogeneous power distribution (26× maximum–minimum power ratio and over 60°C thermal gradient) in the same region. All of these power distributions have the same total power in the region. We first calculate the average temperature under the even power distribution in (a). After that, we use fine-grained thermal analysis to calculate the average temperatures under different heterogeneous power distributions. The results are listed in Table 5.

Table 5 Average temperature under different power distribution with the same total power in a region.

Region	T_{avg} Dist. (a)	T_{avg} Dist. (b)	T_{avg} Dist. (c)	T_{avg} Dist. (d)
Region 1	322.5	322.4	322.4	323.1
Region 2	323.2	322.4	322.5	324.4
Region 3	323.2	322.4	322.5	324.4
Region 4	332.3	333.4	333.3	330.5

As we can see, the average temperature of Region 4 under power distribution (d) is lower than distribution (a), because more heat flows to other regions since a source with high power density is near the boundary. It can also be seen that the average temperature of Region 4 under power distributions (b) and (c) is higher than those in distribution (a), because the source with high power density is far from the boundary. In those cases, the vertical heat flow inside Region 4 is larger than distribution (a). As we can see, the average temperature difference in each region is below 1%. This shows that the total vertical heat flow in one region does not change greatly even under several extreme power distribution patterns. This can be explained by the low-pass filter effect in the spatial frequency domain described in a previous work [36]. This effect limits lateral heat flow.

It is apparent our approach has few errors even if there is extreme variation in power density within a thermal element. In reality, if more detailed power profiles are available, the user can partition functional units with the same design style into even smaller regions such that each region has a mostly-uniform power distribution. However, the user can also use a single thermal element for the whole functional unit and use the above method to verify that this partitioning is proper for their accuracy requirements. Based on our experimental results, there is generally no need to use a region smaller than a functional unit when calculating the average temperature.

6. Conclusion

This article has presented a fast and accurate method of estimating temperature-dependent IC leakage power consumption during design and synthesis. The proposed technique allows a speedup of 59,259× to 1,790,000× while maintaining accuracy compared with a conventional temperature-aware leakage estimation technique using a detailed thermal model. The accuracy of the proposed technique is proven base on two observations: (1) leakage may be accurately modeled as a linear function of temperature over the operating temperature ranges of real functional units and (2) given a fixed total power consumption, the average temperature of an IC active layer is mostly independent of the power distribution. Its accuracy is further validated via numerous comparisons with results from detailed thermal modeling and by comparison with measurements from industry. The pro-

posed technique can easily be used in commercial or academic synthesis and design flows in order to accelerate accurate temperature-dependent leakage power consumption estimation. Our future work will focus on investigating the impact of process variations on the proposed technique in nanoscale CMOS technology.

Acknowledgments

We would like to acknowledge Prof. Robert Dick from University of Michigan Ann Arbor and Prof. Shang Li from University of Colorado, Boulder with their helpful discussions and suggestions; Dr. Hangsheng Wang from Freescale Semiconductor for his assistance with leakage measurement data and Dr. Huang wei and Karthik Sankaranarayanan from University of Virginia for sharing their power traces.

References

[1] Y. Liu, R. Dick, L. Shang, and H. Yang, “Accurate temperature-dependent integrated circuit leakage power estimation is easy,” Proc. Design, Automation & Test in Europe Conf., pp.1531–1536, 2007.

[2] “International technology roadmap for semiconductors,” 2006. http://public.itrs.net

[3] S. Naffziger, B. Stackhouse, T. Grutkowski, D. Josephson, J. Desai, E. Alon, and M. Horowitz, “The implementation of a 2-core, multi-threaded itanium family processor,” IEEE J. Solid-State Circuits, vol.41, no.1, pp.197–209, Jan. 2006.

[4] J.A. Butts and G.S. Sohi, “A static power model for architects,” Proc. Int. Symp. Microarchitecture, pp.191–201, Dec. 2000.

[5] S.M. Martin, K. Flautner, T. Mudge, and D. Blaauw, “Combined dynamic voltage scaling and adaptive body biasing for lower power microprocessors under dynamic workloads,” Proc. Int. Conf. Computer-Aided Design, pp.721–725, Nov. 2002.

[6] S. Narendra, V. De, S. Borkar, D.A. Antoniadis, and A.P. Chandrakasan, “Full-chip subthreshold leakage power prediction and reduction techniques for sub-0.18 CMOS,” IEEE J. Solid-State Circuits, vol.39, no.2, pp.501–510, Feb. 2004.

[7] Y.F. Tsai, D. Duarte, N. Vijaykrishnan, and M. Irwin, “Characterization and modeling of run-time techniques for leakage power reduction,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol.12, no.11, pp.1221–1232, Nov. 2004.

[8] A. Abdollahi, F. Fallah, and M. Pedram, “Leakage current reduction in CMOS VLSI circuits by input vector control,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol.12, no.2, pp.140–154, Feb. 2004.

[9] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, “Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits,” Proc. IEEE, vol.91, no.2, pp.305–327, Feb. 2003.

[10] “BSIM4.” http://www-device.eecs.berkeley.edu/˜bsim4/bsim4.html

[11] “HSPICE.” http://www.synopsys.com/products/mixedsignal/hspice/hspice.html

[12] K. Meng, F. Huebbers, R. Joseph, and Y. Ismail, “Modeling and Characterizing Power Variability in Multicore Architectures,” IEEE International Symposium on Performance Analysis of Systems & Software, pp.146–157, 2007.

[13] S. Sirichotiyakul, T. Edwards, C. Oh, R. Panda, and D. Blaauw, “Duet: An accurate leakage estimation and optimization tool for Dual-Vt circuits,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol.10, no.2, pp.79–90, April 2002.

[14] D. Lee, W. Kwong, D. Blaauw, and D. Sylvester, “Analysis and minimization techniques for total leakage considering gate oxide leakage,” Proc. Design Automation Conf., pp.175–180, June 2003.

[15] R.M. Rao, J.L. Burns, A. Devgan, and R.B. Brown, “Efficient techniques for gate leakage estimation,” Proc. Int. Symp. Low Power Electronics & Design, pp.100–103, Aug. 2003.

[16] A. Rastogi, W. Chen, A. Sanyal, and S. Kundu, “An efficient technique for leakage current estimation in sub 65 nm scaled cmos circuits based on loading effect,” Proc. Int. Conf. VLSI Design, Jan. 2007.

[17] M.Q. Do, M. Drazdziulis, P. Larsson-Edefors, and L. Bengtsson, “Leakage-conscious architecture-level power estimation for partitioned and power-gated sram arrays,” Proc. Int. Symp. Quality of Electronic Design, pp.185–191, March 2007.

[18] C. Gopalakrishnan and S. Katkoori, “An architectural leakage power simulator for vhdl structural datapaths,” Proc. Int. Symp. VLSI Circuits, pp.211–212, Feb. 2003.

[19] A. Kumar and M. Anis, “An analytical state dependent leakage power model for fpgas,” Proc. Design, Automation & Test in Europe Conf., pp.612–617, March 2006.

[20] A. Naveh, E. Rotem, A. Mendelson, S. Gochman, R. Chabukswar, K. Krishnan, and A. Kumar, “Power and thermal management in the Intel® Core™ duo processor,” Tech. Rep., Intel Technology Journal Q2, Intel Corporation, 2006.

[21] Y. Zhan and S.S. Sapatnekar, “Fast computation of the temperature distribution in VLSI chips using the discrete cosine transform and table look-up,” Proc. Asia & South Pacific Design Automation Conf., Jan. 2005.

[22] T. Sato, J. Ichimiya, N. Ono, K. Hachiya, and M. Hashimoto, “On-chip thermal gradient analysis and temperature flattening for SoC design,” Proc. Asia & South Pacific Design Automation Conf., pp.1074–1077, Jan. 2005.

[23] S.C. Lin and K. Banerjee, “An electrothermally-aware full-chip substrate temperature gradient evaluation methodology for leakage dominant technologies with implications for power estimation and hot-spot management,” Proc. Int. Conf. Computer-Aided Design, pp.568–574, Nov. 2006.

[24] Y. Zhang, D. Parikh, K. Sankaranarayanan, K. Skadron, and M. Stan, “HotLeakage: A temperature-aware model of subthreshold and gate leakage for architects,” Tech. Rep., CS-2003-05, Univ. of Virginia, May 2003.

[25] H. Su, F. Liu, A. Devgan, E. Acar, and S. Nassif, “Full chip leakage estimation considering power supply and temperature variations,” Proc. Int. Symp. Low Power Electronics & Design, pp.78–83, Aug. 2003.

[26] W.P. Liao, L. He, and K.M. Lepak, “Temperature and supply voltage aware performance and power modeling at microarchitecture level,” IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol.24, no.7, pp.1042–1053, July 2005.

[27] A. Chandrakasan, W. Bowhill, and F. Fox, Design of High-Performance Microprocessor Circuits, IEEE Press, 2001.

[28] “ISCAS85 benchmarks suite.” http://www.visc.vt.edu/˜mhsiao/iscas85.html

[29] F. Zhang, “System-level leakage power modeling methodology,” bachelor’s degree thesis, Dept. of Electronics Engg., Tsinghua University, July 2006.

[30] W. Zhao and Y. Cao, “New generation of predictive technology model for sub-45 nm design exploration,” Proc. Int. Symp. Quality of Electronic Design, pp.585–590, March 2006.

[31] Y. Liu, Power thermal joint analysis and optimization in nano-scale integrated circuits, Ph.D. Dissertation, Electronic Engineering Department, Tsinghua University, April 2007.

[32] K. Skadron, M.R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, “Temperature-aware microarchitecture,” Proc. Int. Symp. Computer Architecture, pp.2–13, June 2003.

[33] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. Stan, “HotSpot: A compact thermal modeling methodology for early-stage VLSI design,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol.14, no.5, pp.501–524, May 2006.

[34] I.C. Kuon, Automated FPGA Design Verification and Layout, Ph.D.

Thesis, Dept. of Electrical and Computer Engg., University of Toronto, July 2004.

[35] D. Brooks, V. Tiwari, and M. Martonosi, “Wattch: A framework for architectural-level power analysis and optimizations,” Proc. Int. Symp. Computer Architecture, pp.83–94, June 2000.

[36] K. Etessam-Yazdani, M. Asheghi, and H.F. Hamann, “Investigation of the impact of power granularity on chip thermal modeling using white noise analysis,” IEEE Trans. Compon. Packag. Technol., vol.31, pp.211–215, March 2008.

[37] “SRAM layout.” http://www.eecs.umich.edu/UMichMP/Presentations

[38] “MCNC.” http://www.cse.ucsc.edu/research/surf/GSRC/MCNCbench.html

Yongpan Liu was born in Henan Province, P.R.China. He received his B.S., M.S. and Ph.D. degrees from Electronic Engineering Department, Tsinghua University in 1999, 2002, and 2007. He worked as a research fellow in Tsinghua University from 2002 to 2004. Since 2007, he became an assistant professor in Electronic Engineering Department, Tsinghua University. He has published over 30 peer-reviewed conference and journal papers, supported by NSFC, 863, 973 Program. His main research interests include embedded systems, power-aware architecture and VLSI design and electronic design automation. Specifically, his projects consist of ultra-low power wireless sensor network and heterogeneous MPSoCs for software defined radio. He is an IEEE member and served as a reviewer and TPC of several IEEE conferences and TVLSI.

Huazhong Yang was born in Sichuan Province, P.R. China, on Aug. 18, 1967. He received B.S., M.S., and Ph.D. Degrees in Electronic Engineering from Tsinghua University, Beijing, in 1989, 1993, and 1998, respectively. Now, he is a Professor and Head of the Institute of Circuits and Systems in Electronic Engineering Department, Tsinghua University, Beijing. His research interests include CMOS radio-frequency integrated circuits, VLSI system structure for digital communications and media processing, wireless sensor network, low-voltage and low-power circuits, and computer-aided design methodologies for system integration. He has authored and co-authored over 30 patents, 7 books, and over 200 journal and conference papers. He was granted National Palmary Young Researcher Fund of China.