



Thermal-aware design rules for nanometer ICs

J.-L. Rossello, S. Bota, M. Rosales, A. Keshavarzi, J. Segura

► To cite this version:

J.-L. Rossello, S. Bota, M. Rosales, A. Keshavarzi, J. Segura. Thermal-aware design rules for nanometer ICs. THERMINIC 2005, Sep 2005, Belgirate, Lago Maggiore, Italy. pp.1-6. <hal-00189444>

HAL Id: hal-00189444

<https://hal.science/hal-00189444v1>

Submitted on 21 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Belgirate, Italy, 28-30 September 2005

THERMAL-AWARE DESIGN RULES FOR NANOMETER ICs

Jose L Rosselló[†], Sebastià Bota[†], Marcos Rosales[†], Ali Keshavarzi[‡] and Jaume Segura[†]

[†]Electronic Technology Group, Universitat Illes Balears, Campus UIB, 07122, Spain

[‡]Microprocessor Research Labs., Intel Corp., Portland (OR)

ABSTRACT

The ever-aggressive increase in performance and integration of CMOS ICs is leading to higher power dissipation and power density, with a consequent increase of the circuit junction temperature together with the generation of non-uniform thermal maps and hot-spots. Moreover, it has been predicted that technology scaling will bring static power to a significant fraction of the total power, complicating further this analysis due to the exponential dependence of leakage with temperature. This scenario represents an important challenge in circuit design due to the lack of accurate and compact thermal models capable of describing not only the impact of non-uniform thermal maps on circuit performance, but also on static leakage for an accurate power modeling. In this paper we analyze the influence of considering the leakage power contribution to the circuit thermal map and therefore on its overall power dissipation and performance. Based on the analysis of two key thermal parameters (the mean and the maximum junction temperatures) we derive design guidelines to reduce the impact of hot-spots on circuit performance and reliability using an entropy-based cost function for thermal optimization.

Index Terms—Thermal-aware design, circuit modeling.

1. INTRODUCTION

Thermal estimation has become a major concern in IC design due to the elevated power density in high-performance ICs. High operating temperatures degrade circuit performance and impact reliability. Both power dissipation and operating frequency are worsened at high temperatures due to the increase of leakage currents and the reduction of carrier's mobility. Technology scaling rules based on constant field scaling dictate voltage supply reduction from generation to generation, imposing a threshold voltage reduction to maintain performance. The exponential dependence of leakage current with threshold voltage leads to a dramatic increase of power with technology scaling. For the 90nm process generation node, subthreshold leakage power can contribute as much as 42% of the total power [1]. As a consequence, the impact of

temperature increase on the overall circuit power consumption will be more relevant in the future.

Non-uniform circuit thermal maps, and specifically hot spots, may also impact circuit timing parameters by slowing down non-critical paths or increasing clock skew. In addition, hot-spots also impact circuit reliability since thermally-activated failure mechanisms are significantly accelerated due to the exponential relationship between mean time to failure (MTF) and temperature for these type of mechanisms.

The inclusion of thermal effects in the design cycle requires the development of fast and accurate power–delay–thermal (PDT) estimation tools, and the adoption of thermal-aware design rules. Moreover, optimal design decisions require an accurate estimation of the IC thermal characteristics to be performed at different levels of abstraction and applied at each step of the design cycle. A PDT estimation tool must iterate the solutions of the three parameters until convergence as they are closely inter-related. Power is estimated initially, as the starting point for the thermal profile computation after which the initial thermal map is obtained. Delay is then estimated, and used to recompute power since it has an impact on short-circuit power and therefore on temperature. After a few iterations, an accurate thermal map is obtained.

The estimation of the thermal profile of ICs [2] can be performed at different levels of abstraction depending on the required granularity. At the lower level the temperature is estimated considering the contribution of each transistor in the circuit modeled as the elemental heat source. At a higher level of abstraction an entire circuit block can be considered as the basic heat source. These models have been presented in [3] where we develop a methodology to estimate the IC thermal profile generated from power sources of sizes $W \times L$. Using the thermal model developed in [3] we can estimate the operating temperature of a square block 'j' with sizes $L \times L$ dissipating a power P_j as:

$$T_j \approx T_{amb} + R_{th} P + 3.74 \left(\frac{P_j}{L} + 0.283 \sum_{i=1}^N \frac{P_i}{r_{ij}} \right) \quad (1)$$

Thermal-Aware Design Rules for Nanometer ICs

where T_{amb} is the ambient temperature, R_{th} is the effective thermal resistance of the packaging system, P is the total power dissipated by the IC, and P_i is the power dissipated by the i -th block that is located at a distance r_{ij} from the block considered.

In this work we analyze the thermal characteristics of deep-submicron CMOS ICs and propose a method based on the estimation of the circuit power entropy that can be used for thermal optimization during logic synthesis. Finally, we present a basic design rule for cell placement. The rest of this paper is organized as follows: in section 2 we show the basis of an efficient thermal-aware design flow. In section 3 we present the basic thermal characteristics of deep-submicron CMOS ICs while section 4 presents an entropy-based thermal model for logic synthesis and a thermal-aware design rule for cell placement. Finally, we expose the conclusions in sections 5.

2. THERMAL-AWARE DESIGN FLOW

An efficient thermal-aware design flow [4] of deep-submicron ICs must consider at least two thermal parameters: the mean and the maximum temperature of operation (the latter parameter related to the presence of hot spots). We define ΔT_{SH} as the mean value of the circuit junction temperature increase, while δT is defined as the difference between the maximum and the mean temperatures. Therefore, $T_{mean} = T_{amb} + \Delta T_{SH}$, and $T_{max} = T_{amb} + \Delta T_{SH} + \delta T$, where T_{mean} and T_{max} are the mean and maximum junction temperatures and T_{amb} is the room temperature.

Circuit reliability impasses a maximum value for T_{max} through the whole die. For any point in the circuit $\Delta T_{SH} + \delta T < \Delta T_{ref}$ where ΔT_{ref} is the temperature increase over the ambient temperature that gives an operating temperature T_{MTF} causing a given Mean Time to Failure equal to $t_{MTF} = M_0 \exp(E_a / k_B T_{MTF})$, where E_a is and activation potential, k_B the Boltzman constant, M_0 a technology-dependent parameter and $T_{MTF} = T_{amb} + \Delta T_{ref}$.

The mean value of self-heating is proportional to the total dissipated power where R_{th} is the thermal resistance of the circuit package. This gives a second restriction for the thermal parameters such that, $\Delta T_{SH} < P_{max} \cdot R_{th}$ where P_{max} is the maximum allowed power dissipation. This would be the maximum power limit of the circuit.

As was previously mentioned, hot spots may change the timing of correlated paths inside the circuit. Dynamic architectures are very sensible to these changes since they have more complex timing requirements than static CMOS logics. Therefore, it is important to control the possible presence of hot spots in the circuit that may degrade the delay of critical control signals in the logic, thus leading to a possible circuit malfunction. Parameter δT is a good

indicator of hot spots intensity, and therefore can be used to monitor their presence in the circuit.

A thermal-aware design flow must consider three important thermal restrictions: the reliability limit, the maximum power limit, and the hot-spots limit, as shown in Fig.1 with respect parameters $(\Delta T_{SH}, \delta T)$. The optimal operating point in the allowed region is that providing the highest performance. An efficient thermal-aware design flow is proposed in Fig. 2, and is based on the estimation of the circuit thermal behavior (parameters δT and ΔT_{SH}) at each step of the design process. If the thermal conditions are outside the allowed region (or the timing requirements are not satisfied), the design step must be repeated. Therefore, each step in the design must have its own thermal estimation model according to its degree of abstraction.

3. BASIC THERMAL CHARACTERISTICS OF DEEP-SUBMICRON CMOS ICs

As the feature size of transistors is scaled down, subthreshold currents and therefore the thermal impact on total power increases drastically. The temperature of circuit blocks dissipating a power $P = P_D + P_S(T)$ increases due to self-heating (P_D and P_S being the dynamic and the static power respectively), therefore increasing the block power until $P + \delta P = P_D + P_S(T + \Delta T_{SH})$. The exponential dependence of static power with temperature makes self-heating to be more severe in blocks having a higher switching activity. Therefore for future technologies the presence and intensity of hot-spots is predicted to increase.

This effect is shown in Fig. 3 where the thermal characteristics of a deep-submicron IC under different operating conditions are plotted. Diamonds represent the circuit thermal behavior neglecting the static power impact on temperature, while circles correspond to the thermal behavior considering both static and dynamic power components. Not only the mean temperature increases due to the contribution of leakage but also the differences between the maximum and the mean temperature (δT) increase. This means that, as static power increases due to technology scaling, the intensity of hot spots due to non-uniform switching activity will also increase. Fig. 4 shows the correlation between δT with and without taking into account the static power $\delta T(\text{dynamic})$. The increment in δT due to the static power ($\Delta \delta T$) is greater for those designs with larger values for $\delta T(\text{dynamic})$. This means that the intensity of hot-spots would be greater in those designs without a thermal-aware design (and probably with larger asymmetries in the switching activity distribution along the circuit).

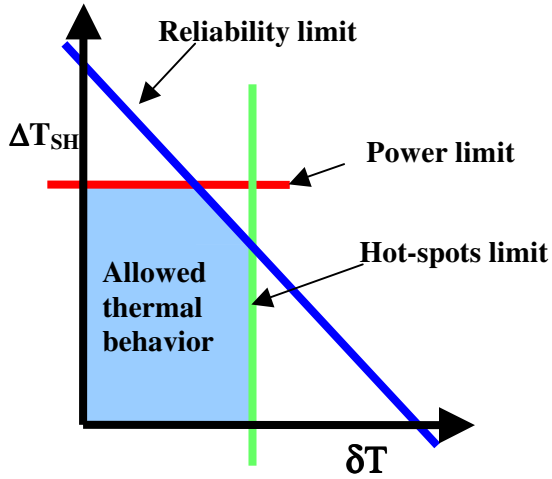


Fig. 1. Design space restrictions. The thermal behavior of the circuit must be inside the allowed region to accomplish the reliability, power and hot-spots requirements.

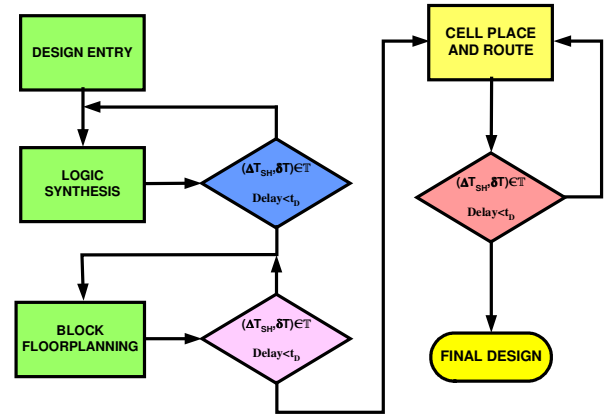


Fig. 2 Basic Thermal-aware design flow

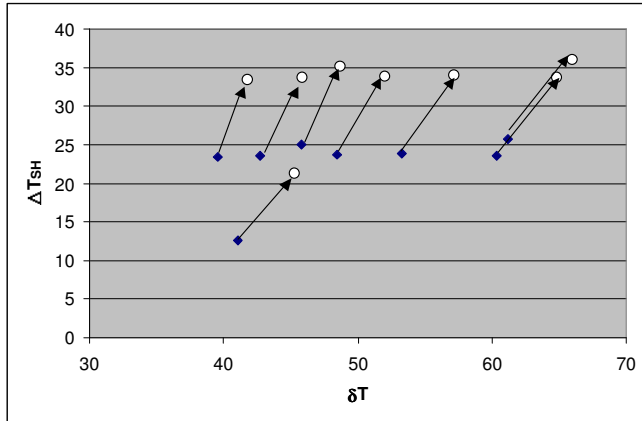


Fig. 3 Thermal influence of static power. Both self heating ΔT_{SH} and the difference $T_{max}-T_{mean}$ are increased significantly due to the thermal dependence of static power.

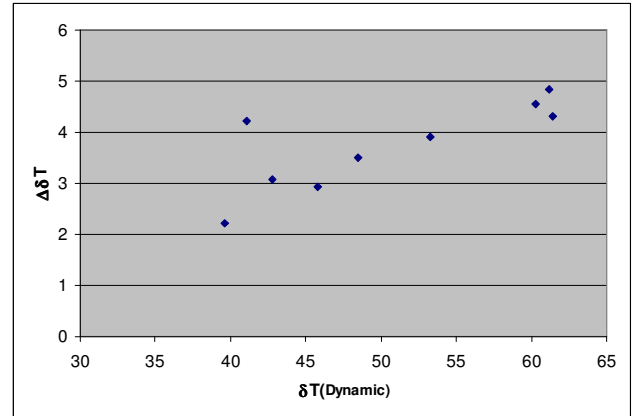


Fig. 4 The increment in δT due to static power is larger for those designs with larger values of δT when only the dynamic component is considered.

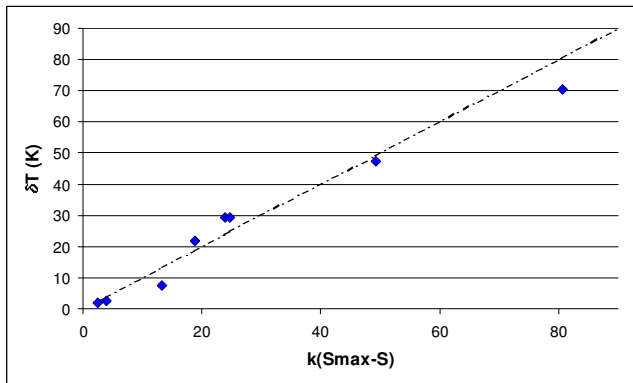


Fig. 5 Relationship between the thermal difference of the circuit $\delta T = T_{max}-T_{mean}$ and the power entropy.

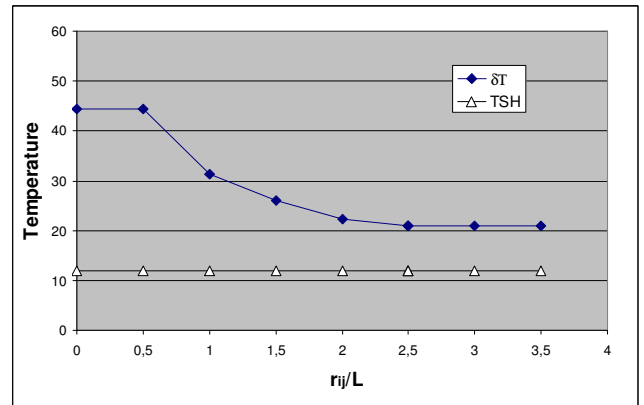


Fig. 6 Self heating ΔT_{SH} and thermal difference $\delta T = T_{max}-T_{mean}$ as a function of the distance between blocks.

4. THERMAL-AWARE DESIGN GUIDELINES

From section 3 we conclude that ΔT_{SH} and δT are key parameters that must be considered to obtain a successful thermal-aware design. In this section we present two guidelines in order to optimize these parameters at two different stages in the design flow: logic synthesis and physical design.

4.1 Thermal-aware design rules for logic synthesis

Logic synthesis is concerned with the generation and optimization of a circuit at the level of Boolean gates. In general a large library of distinct logic gates are available for the realization of digital circuitry. Taking into account that the number of all the feasible solutions of equivalent circuits (search space) compatible with the target library cell is high, logic synthesis tools use a cost function to assign a real number to each possible solution. The final solution consists to find a feasible solution with minimal cost. Usually, the goal is to minimize the area or power consumption while satisfying delay constraints. For a thermal aware design is important to include parameters ΔT_{SH} and δT in the optimization of the related cost function.

Assuming that each feasible solution is characterized by a number of logic gates, N , each one with area A_i and power consumption P_i , the mean value of self-heating can be easily computed using Fourier's law as $\Delta T_{SH} = R_{th} \sum_i P_i$ [5]. This model must be iterated with the power model (dynamic and static power estimation) until convergence. Note that this parameter is taken into account when power consumption is used as optimization parameter.

δT is more difficult to estimate. We assume that this magnitude is proportional to the mean value of the thermal gradient that may appear inside the circuit

$$\delta T \cong K \|\nabla T\|_{mean} \sqrt{A} \quad (2)$$

Where A is the total area, $A^{1/2}$ is the typical distance inside the circuit and $\|\nabla T\|_{mean}$ is defined as:

$$\|\nabla T\|_{mean} = \frac{1}{A} \int_{IC\ Area} \|\nabla T\| dA \quad (3)$$

To estimate $\|\nabla T\|_{mean}$ we realize that $\|\nabla T\|_{mean}$ is zero for uniform thermal maps (corresponding to a disordered distribution of P_i within the circuit, or equivalently to a maximum entropy in power distribution).

If we assume an ordered power distribution (i.e. there are a small number of cells with much more power consumption than others so that nearly all power is dissipated in these

cells) then the magnitude of $\|\nabla T\|_{mean}$ will increase, this configuration corresponds to a low entropy of the system (S). Therefore, we assume a linear relationship between $\|\nabla T\|_{mean}$ and S as:

$$\|\nabla T\|_{mean} = \frac{C}{Ak_{Si}} (S_{max} - S) \quad (4)$$

where C is a technology-dependent parameter, k_{Si} is the thermal conductivity of Silicon and $S_{max}-S$ is computed as [6]:

$$S_{max} - S = \sum_{i=1}^N P_i \ln\left(\frac{P_i}{A_i}\right) - P \ln\left(\frac{P}{A}\right) \quad (5)$$

S is the natural logarithm of the possible number of ways of distributing the total power P through the N cells that are dissipating a power P_i . Equation (4) is a first approach to $\|\nabla T\|_{mean}$, and therefore δT can be described as:

$$\delta T = \frac{K'}{k_{Si} \sqrt{A}} \left[\sum_{i=1}^N P_i \ln\left(\frac{P_i}{A_i}\right) - P \ln\left(\frac{P}{A}\right) \right] \quad (6)$$

where K' is a process-dependent parameter. In Fig. 5 we show the relationship between δT and the entropy-based expression for δT (6) for different values of the power dissipation of a high-density circuit using a deep-submicron CMOS technology. As can be appreciated, as the entropy of the circuit power decrease (due to different power distributions through the IC blocks), the δT value increase proportionally. Therefore, expression (6) is found to be an appropriate method for the estimation of the thermal asymmetry within an IC. In these way, Both parameters ΔT_{SH} and δT can be computed and included as parameters of the cost function.

5. Thermal-aware design rules for cell placement

The presence of hot spots is related with the position of each cell in the layout. Cell placement methods are based on min-cut partitioning and clustering that try to minimize the length of the nets between different cells. Therefore the probability to form a cluster of high-activity cells is high if there are some correlation between these cells in the input netlist.

From expression (1) is evident that local junction temperature will be very dependent on cell placement (with large temperature values as the distance r_{ij} between hottest blocks decreases). Hot spots can be produced if clustering of hot cells is not avoided during this design step. From expression (1) we can obtain a minimum distance between critical cells (hottest cells) such that the thermal influence between them is not enhanced. We state that the

Thermal-Aware Design Rules for Nanometer ICs

temperature increment of critical cell 'j' due to an adjacent critical cell 'i' must be lower than the 10% of the expected self-heating temperature of cell 'j'. That is:

$$0.283 \frac{P_i/P_j}{r_{ij}/L} < 0.1 \quad (8)$$

where r_{ij} is the distance between the two cells. Therefore r_{ij} must satisfy:

$$r_{ij} > \frac{P_i}{P_j} 2.8L \quad (10)$$

where L is the typical length of the hottest cell (note that r_{ij} is related with δT). In Fig. 6 we show parameters δT and ΔT_{SH} when varying the distance between two critical cells r_{ij} (the two hottest cells dissipating the same power). The mean value of self-heating in all the circuit is nearly constant (power is always the same) while δT increases as the distance between the two critical blocks decreases. For this situation, the optimal distance is $r_{ij} > 2.8L$ as stated by (10). Note that δT goes from 20° to more than 40° when $r_{ij}=0$ (both cells integrated within the same area).

This supposes that in order to obtain a thermal-aware placement, is necessary the use of tools that compute and include the value of this safety distance r_{ij} during the placement step.

6. CONCLUSIONS

Since the thermal behavior of ICs has become a major concern in IC design due to their great impact on the static

power, thermal-aware design rules and cost functions must be developed.

Thermal computations must be performed at each step in the design flow. The thermal behavior of the circuit must be within an allowed thermal region that takes into account power, reliability and hot spots considerations. These design techniques would be vital for next technology generations since thermal asymmetries would be larger due to the increasing importance of leakage.

In this work we presented an accurate entropy-based thermal-aware cost function for an optimal logic synthesis. At a lower level of abstraction, a simple cell-placement design rule is provided for hottest blocks (those blocks with larger values of power density).

7. REFERENCES

- [1] J. Jao, S. Narendra, A. Chandrakasan, "Subthreshold Leakage Modeling and Reduction Techniques" In *Proc. International Conference on Computer-Aided Design (ICCAD'02)*.
- [2] M. Rencz, V. Székely and A. Pope "A Fast Algorithm for the layout based Electro-thermal Simulation" In *Proc. Design, Automation and Test in Europe, (DATE) 2003*, pp.1032-1037.
- [3] J.L.Rossello, V. Canals, S.A. Bota, A. Keshavarzi and J. Segura "A Fast, Concurrent Power-Thermal Model for Sub-100nm Digital ICs" *Design, Automation and Test in Europe, (DATE) March 2005*, pp.206-211.
- [4] W. Huang, M.R. Stan, K. Skadron, K. Sankaranarayanan, S. Ghosh, S. Velusamy, "Compact Thermal Modeling for Temperature-Aware Design", in *Proc. Design Automation Conference, 2004*, pp. 878-883
- [5] K. Kanda, K. Nose, H. Kawaguchi, T. Sakurai, "Design impact of positive temperature dependence on drain current in sub-1-V CMOS VLSI", *IEEE J. Solid-State Circuits*, vol. 36, no. 10, pp.1559-1564 Oct. 2001
- [6] Pathria, R. *Statistical Mechanics*, Pergamon Press, 2nd edition (1998).