

Thermal issues in test: an overview of the significant aspects and industrial practice

J. Alt¹, P. Bernardi², A. Bosio³, R. Cantoro², H. Kerkhoff⁴, A. Leininger¹,
W. Molzer¹, A. Motta⁵, C. Pacha¹, A. Pagani⁵, A. Rohani⁴, R. Strasser¹

¹INTEL
Germany

²Politecnico di Torino
Italy

³LIRMM
France

⁴University of Twente
The Netherland

⁵STMicroelectronics
Italy

Abstract*— Thermal phenomena occurring along test execution at the final stages of the manufacturing flow are considered as a significant issue for several reasons, including dramatic effects like circuit damage that is leading to yield loss. This paper tries to redeem those bad guys in order to exploit them to improve the test quality, reducing the overall test cost without affecting the yield

Keywords—component; thermal-aware test; yield; functional test; automotive

I. INTRODUCTION

There are several perceptions of what thermal issues in test means. Thermal phenomena occurring during manufacturing test are mainly perceived as significant concerns as they are considered the root cause of dramatic effects like circuit damage resulting in yield loss. With silicon technology scaling, however, VLSI circuits operate very often at high temperature, which has negative impact on reliability, performance, power-efficiency and testability [1].

This is a subject frequently discussed in the literature, where approaches for thermal management during testing play a key role in reducing the cost while increasing yield and performance. Changes in packaging technology and the rapid increase in processor power and power density, however, are presenting unique thermal challenges requiring ad-hoc solutions.

Several thermal-aware test-scheduling researches were proposed so far to achieve a sustainable temperature distribution. They are computed a priori as a static or on-the-fly as dynamic schedule [2] by using an estimation of the power consumed by each test. The power consumption estimation is often derived by logic observations (i.e., switching activity) or use more accurate models. Considering microprocessor-based systems, innovative solutions have to be developed to meet thermal management challenges [3].

On the other hand, temperature management could also be useful to exacerbate the occurrence of faults, which shall not be detected in normal conditions. A point of view is that forcing an

excessive thermal configurations may lead to temperature-induced timing faults [4]; in contrast, some other works claim that delay defects under high temperature are one of the most critical factors to affect the reliability of computer systems [5] and propose test methods to address this problem properly.

How to create thermal conditions that lead to better testing is a strongly discussed topic. Approaches like [6] analyze the non-trivial issues associated with calculating the heat spreading from a heat source into a thermally conducting body, underlining that the popular series resistance approaches has severe limitations.

This paper encompasses three relevant subjects and proposes techniques that are related to thermal management and its implications. Section 2 discusses the creation detection conditions useful for stigmatize intermittent faults. Section 3 describes a model to estimate critical thermal configurations. Section 4 introduces a methodology for quickly estimating temperature distribution over chip surface produced by functional programs.

II. EVOKING INTERMITTENT RESISTIVE FAULTS BY MEANS OF ON-CHIP TEMPERATURE CYCLING

The emergence of Intermittent Resistive Faults (IRFs) is considered one of the challenges of modern VLSI circuits [7]. These faults pass routine tests and only have an enhanced chance that they can be activated during high temperature tests. However, imposing a temperature on a chip in a controlled procedure can be quite challenging. Most of the methods to raise temperature in a chip involve a thermal oven that causes the whole temperature of the chip to rise. In that method, it is very hard to control the local temperature of each functional block of the chip.

The purpose of this study is to introduce a mechanism to control the temperature profile of a VLSI chip. Tailored software workloads will be used to raise the local temperature in a chip, which means some blocks of the chip are hotter than the rest. By doing so, each block of the chip can be tested separately which helps to locate the IRFs induced by the temperature.

The first step in our study is to develop a model to extract the local temperature of each block. Our temperature model uses the

* This paper is summarizing the content of a hot topic session given at the Intl. IEEE VLSI Test Symposium 2016 and organized in the context of the LIA-LAFISI project.

well-known duality between heat transfer and electrical phenomena. In this model, heat is considered as a “current” which passes through a thermal resistance and creates temperature differences analogous to “voltage”. This principle is shown in Figure 1. The switching activity delivered to each block is considered as a current source in this model. This switching activity is directly dependent on the workload and is extracted for each couple workload/per block.

Figure 2 shows our mechanism to control temperature in a design. The information extracted from power analysis tool and place-and-route tool will be used to construct the thermal model of the system. This equation needs to be solved by an Electrical Simulator, such as Spice. The temperature profile of a block at the end of Figure 2 represents the thermal cycle of a block during execution of a workload. This thermal cycle will be reproduced again after manufacturing so that potential IRF faults which are evoked at a certain temperature will be located.

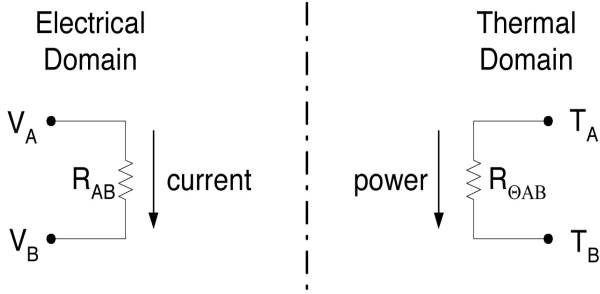


Figure 1. Fundamental relationships in the electrical and thermal domains (T_A and T_B means Temperature, $R_{\Theta AB}$ represents the thermal resistance)

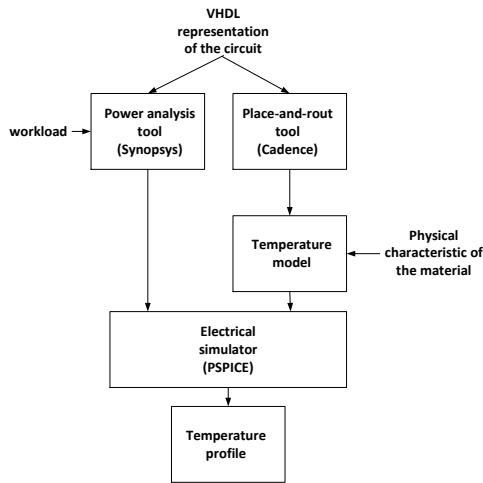


Figure 2. Dataflow to extract the temperature profile

To show the applicability of our method, the mechanism shown in Figure 2 has been applied to a thirty-two bit full-adder. The adder runs three different workloads in which every workload imposes different switching activities on different parts of the chip. The temperature profile of the first full-adder is shown in Figure 3.

As can be seen in Figure 3, different workloads cause different temperature raises in the first full adder of this system. Workload 1 (represented by W1) causes a temperature raise of 12.5°C degrees while workload 3 (represented as W3) imposes 10°C on this adder. The temperature follows an exponential raise until it is stable. If the power is stopped, the block would start to cool down in an exponential manner. Knowing this temperature cycle for every block of the chip will enable the designer to apply a controlled mechanism to evoke temperature-induced faults.

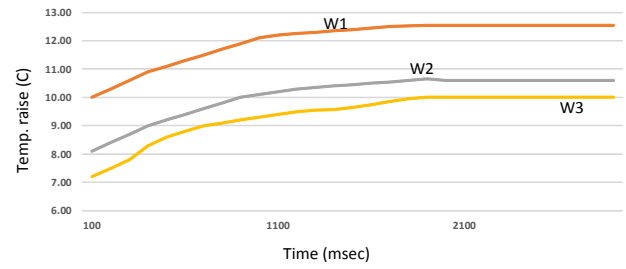


Figure 3. Temperature profile in the first full-adder for three different workloads

III. THERMAL INSTABILITY – NEW CHALLENGE FOR MANUFACTURING TEST AND SYSTEM OPERATION OF HIGHLY INTEGRATED VLSI DEVICES

With the ubiquity of mobile computing, System On-Chip (SOC) devices integrate an increasing number of functions into a single piece of silicon. SOC solutions integrate functions ranging from general purpose CPU, graphics, security, connectivity functions to cellular modems and many others. Along with this ultra-high integration comes a trend of growing power consumption. The progression into advance technology nodes aims to counteract this trend by leveraging the virtue of MOS device scaling. Although that worked to a high degree for the past 40 years, physical limitations like oxide thickness and limitations in the metallization make it increasingly hard to fully compensate the power increase by technology scaling. As a result, the total power consumption of SOC's has reached a level leading to non-negligible self-heating; the silicon die heats up several degrees of Celsius with respect to its ambient temperature. A convenient model uses a thermal resistance R_{Th} to quantify the corresponding temperature increase

$$T_J - T_A = R_{Th} \times P$$

The power consumption of SOC devices can be divided into a temperature independent component P_{act} , related to the switching activity, as well as a temperature dependent leakage component $P_{lkg}(T)$ as shown in Figure 4.

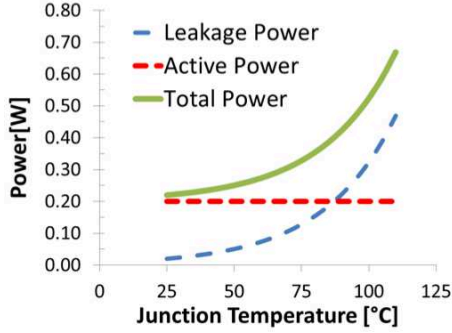


Figure 4. Power Consumption Components

Thus, a feedback from device temperature to leakage power is in effect as illustrated in Figure 5, leading to a non-linear system.

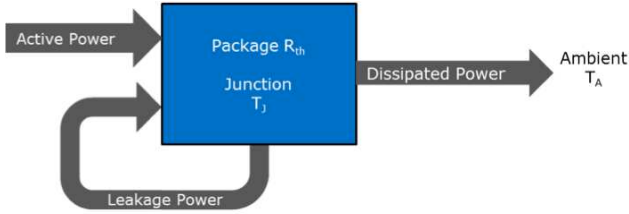


Figure 5. Leakage Induced Device Self-Heating

As long as the increase in leakage power per degree Celsius can be compensated by the accompanying increase in dissipated power through the chip package, the system is stable. Beyond that point, more thermal energy is produced than the ambient is able to absorb. The result is a so called *thermal runaway*.

It turns out that leakage power can be modelled by an exponential temperature dependence with reasonable accuracy as follows

$$P_{lkg} = P_{lkg,25} e^{\frac{T_J - 25^\circ C}{T_L}}$$

This allows for an analytical solution of thermal stability limit:

$$T_{J,crit} = 25^\circ C + T_L \ln \frac{T_L}{R_{th} \times P_{lkg,25}}.$$

The knowledge of $T_{J,crit}$ is essential for stable system operation during application as well as system test. $T_{J,crit}$ represents the highest junction (i.e. silicon die) temperature under which the system can operate without drifting into instability. Its value is defined by the temperature coefficient T_L ,

the thermal resistance of the system R_{th} as well as the leakage power at room temperature $P_{lkg,25}$. While T_L is defined by the properties of silicon, thermal system properties (R_{th}), as well as room temperature leakage ($P_{lkg,25}$) are a result of design choices. Interestingly, active power P_{act} does not have an impact on the highest achievable stable junction temperature.

Having this in mind, we can formulate guidelines for stable test at high temperatures: (1) for a given thermal resistance it is important to maintain leakage power consumption below a given limit. Circuit partitioning, including switch-off capability, is a feasible technique to achieve this. (2) Reducing active power will not improve thermal stability (3) the thermal design of the test system (socket, etc.) has a strong impact on the highest achievable stable test die temperature.

IV. CHIP SURFACE TEMPERATURE DISTRIBUTION ESTIMATION FOR FUNCTIONAL PROGRAMS RUNNING ON AUTOMOTIVE MICROCONTROLLERS

High reliability standards are required by automotive manufacturers that ask their electronic suppliers to guarantee a defect level lower than 1ppm. Thus, a major goal is to screen out defective parts in the earliest stages of production, anyhow before a defective device reaches the customer.

Test During Burn-in (TDBI) plays a key role during back-end phase because it is aimed to give rise to infant mortalities (early life latent failures); It is a process during which electronic components are exposed to high temperatures [8] in a climatic chamber prior to being placed in service, and devices that stop working at this step are discarded.

The usage of functional programs for stress and test purposes [9] along TDBI is a methodology for adding a dynamic temperature component to the static contribution of the climatic chamber [10]. Thus, we propose a temperature modeling methodology able to quickly predict the temperature over the surface of a chip under the execution of a functional program based only on logic simulation of the gate level of the device. In particular, it aims at extrapolating the temperature matrix by elaborating a matrix of Weighted Switching Activity (WSA) obtained by joining layout information with the switching activity produced by the execution of a functional program. Figure 6 visualizes the result of the technique.

The WSA of the i-th cell is computed as following:

$$WSA_i = \frac{1}{M} \cdot \sum_{k=1}^M SA_k \cdot FO_k$$

SA_k is the Switching Activity per clock cycle measured by logic simulation and FO_k is the fan-out of the k-th gate inside of the considered cell; M is the number of gates included in the cell.

To perform the estimation we resort to a Temperature Estimator Model θ obtained by feeding a Regressive Algorithm

with the physical measurements performed on a set of sample functional programs as shown in figure 7.

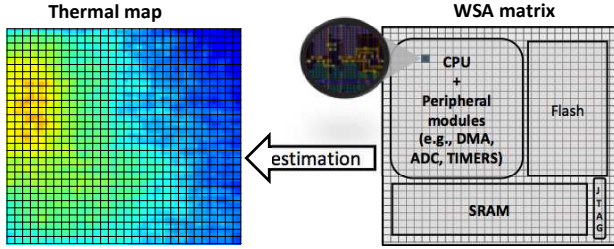


Figure 6. Temperature estimation technique

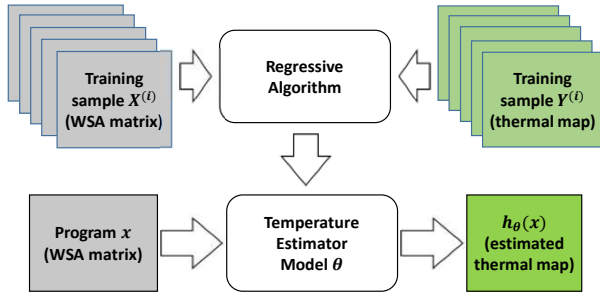


Figure 7. Training and usage of the Temperature Estimator Model

Thermal maps used to train the regressive algorithm are obtained by recording the temperature distribution functionally produced by ad-hoc programs on a scrubbed chip, as depicted in figure 8.



Figure 8. Experimental setup

Experimental results were gathered on a 32-bit automotive microcontroller to demonstrate that the estimated temperature distribution is strongly correlates with the real measurement.

The size of a cell in the WSA matrix was determined by the precision of the thermo-camera and each cell contains about 2K logic gates.

Overall, 33 programs were selected as training sample to build the temperature estimator model. These programs target several functionalities of the microcontroller, including arithmetic-logic modules, register file, pipeline-fetch unit and peripheral cores such as timers and DMA controllers.

As different set of programs were used to validate the estimation model; for validation sakes the temperature

distribution produced by these programs was estimated and then compared with the measurement by the thermo-camera.

The calculated Pearson correlation index [11] returns 0.96 as final value, while the highest temperature estimation have an accuracy of about $\pm 0.29^{\circ}\text{C}$.

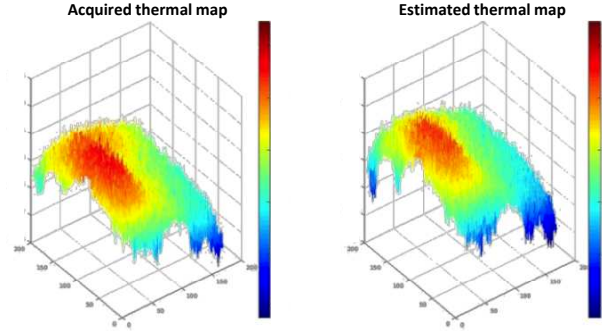


Figure 9. Comparison among acquired and estimated thermal map

V. CONCLUSIONS

This paper shown that thermal phenomena occurring during test may also be exploited to improve the test quality without affecting the yield. Discussion about the benefits and techniques that are related to thermal management and its implications are given.

REFERENCES

- [1] C. Yao, K. K. Saluja and P. Ramanathan, "Thermal-Aware Test Scheduling Using On-chip Temperature Sensors," VLSI Design (VLSI Design), 2011, pp. 376-381.
- [2] D. R. Bild et al., "Temperature-aware test scheduling for multiprocessor systems-on-chip," Computer-Aided Design, 2008. ICCAD 2008, pp. 59-66.
- [3] Z. Peng, "Thermal challenges to building reliable embedded systems," VLSI Technology, Systems and Application (VLSI-TSA) 2014 pp. 1-2.
- [4] P. Tadayon, "Thermal Challenges During Microprocessor Testing, " Intel Technology Journal Q3, 2000.
- [5] Y. Zhang, Z. Peng, J. Jiang, H. Li and M. Fujita, "Temperature-aware software-based self-testing for delay faults," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015, pp. 423-428.
- [6] C. J. M. Lasance, "The practical usefulness of various approaches to estimate heat spreading effects," Thermal Issues in Emerging Technologies, ThETA '08. pp. 149-158.
- [7] H. G. Kerkhoff and H. Ebrahimi, "Intermittent Resistive Faults in Digital CMOS Circuits," IEEE 18th International Symposium on Design and Diagnostics of Electronic Circuits and Systems, pp. 211-216, 2015
- [8] A. Birolini, "Reliability Engineering, Theory and Practice", SpringerVerlag, 3rd edition, 1999
- [9] N. Aghaee, Z. Peng; P. Eles,"An efficient temperature-gradient based burn-in technique for 3D stacked ICs," in Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014, pp.1-4
- [10] A. Benso et al., "ATPG for Dynamic Burn-In Test in Full-Scan Circuits", IEEE ATS, 2006, pp. 75-82
- [11] M. Ross Sheldon, Introduction to Probability Models, Elsevier 2010, ISBN 978-0123756862