Microelectronics Journal 52 (2016) 40–48

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

Impact of floorplanning and thermal vias placement on temperature in 2D and 3D processors

Piotr Zajac*, Melvin Galicia, Cezary Maj, Andrzej Napieralski

Department of Microelectronics and Computer Science, Lodz University of Technology, Wolczanska 221/223, 90-924 Lodz, Poland

ARTICLE INFO

Article history:
Received 29 April 2015
Received in revised form
27 January 2016
Accepted 13 February 2016
Available online 28 March 2016

Keywords:
Multi-core processors
3D integrated circuits
Thermal vias
Thermal simulation

ABSTRACT

In modern integrated circuits, manufactured in nanometer technologies, reducing the hotspot temperature even by several degrees may lead to significant advantages. In particular, in high performance processors, lower temperature translates into fewer reliability concerns, lower cooling costs, the possibility of increasing the operating frequency and extending the device's lifetime. Therefore, in this paper we investigate how the positioning of particular processor units in the floorplan (i.e. floorplanning) affects the chip temperatures. We take into consideration 8-and 6-core processors manufactured in 14 nm technology and simulate the temperature distribution for various floorplan designs. It is shown that the difference in maximal temperature for various floorplans can reach even 7.2 K for a typical case. Moreover, the idea for thermal buffers is presented. While it is shown that thermal buffers may not be of great significance in 2D integrated circuits, obtained results indicate that in 3D ICs the combination of thermal buffers and vertical thermal vias may considerably reduce the temperature of the hottest areas.

1. Introduction

In modern nanometer technologies, transistor could theoretically run at much higher frequency than 3.5–4 GHz which is currently common in high-performance processors. However, for several years it is the power density which has limited the operating frequency of modern processors [8]. The Thermal Design Power (TDP) for quad-core i7 Intel processors was around 100 W and 85 W for 32 nm and 22 nm technology, respectively. Higher power dissipation would require a more complex cooling system which would be impractical from the commercial point of view. Therefore, it may be expected that although processors with more and more cores will be produced, the TDP will have to be maintained at the same level. Although the dynamic power per core is reduced when migrating to the next technology node, more cores are added on chip and, moreover, leakage power increases with each technology node. As a result, the total power dissipation increases even if operating frequency is kept constant. Thus, processor designers face a considerable challenge: which parameters future processors (manufactured at 14 nm technologies and below) should have, so that the processor under heavy workload does not overheat? Is it better to design a six-core processor with a larger L3 cache and larger graphics unit or it is possible to produce an eight-core processor while keeping the same chip area? Is it better to have fewer cores but higher operating frequency or more cores operating at lower speed? Moreover, recent research on 3D stacked architectures shows that soon it will be possible to integrate several silicon layers in one chip. Especially interesting is the possibility to stack DRAM memory in the same chip as processor [10], which would reduce the memory access latency and therefore mitigate the problem of memory wall [11]. However, while it certainly offers many advantages, from the thermal point of view 3D stacking actually aggravates the thermal problems due to increased ratio of power per area [12]. One potential solution for removing the excessive heat in 3D architectures is the use of thermal vias [13,14], however the question is how effective they are and how/where to place them inside the chip.

This paper is organized as follows: in the second section the methodology for obtaining power data is explained, as well as our approach to the thermal simulation of both 2D and 3D chip packages. The next section presents the results obtained for several floorplan types of 2D architectures. In Section 4 the idea of thermal buffers with thermal vias for 3D stacked architectures is presented and analyzed. It also includes simulations results for 3D chips. In Section 5 a summary of findings concludes the paper.

2. Power and thermal simulation methodology

In this paper we investigate how various processor floorplan configurations and the operating frequency influence the temperature distribution in 14 nm technology processors. Additionally, for 3D stacked ICs we analyze how the implementation of thermal vias

* Corresponding author.
E-mail address: pzajac@dmcs.pl (P. Zajac).

http://dx.doi.org/10.1016/j.mejo.2016.02.013
0026-2692/

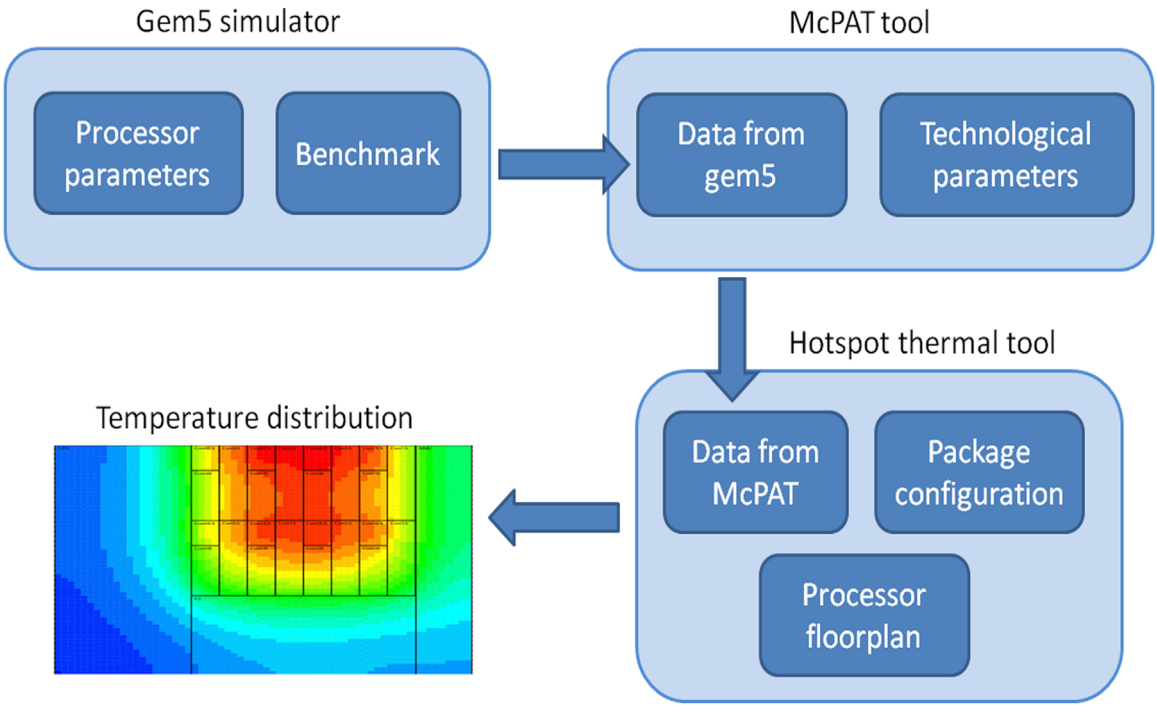

Fig. 1. Simulation methodology.

affects the temperature. To achieve these goals, our methodology consists of three main steps (see Fig. 1):

- We used a widely known processor simulator gem5 [1] to simulate the execution of benchmarks by the processor. Gem5 is able to simulate a modern out-of-order x86 processor at cycle-level accuracy and compute the number of accesses to all processor units during the execution of a benchmark.
- The data obtained from step 1 were transferred to McPAT [2], which was used to calculate the power dissipated in particular processor units during the execution of the benchmark. The processor configuration that we used was based on the parameters of Penryn microarchitecture configuration file provided by McPAT, which was appropriately modified to better resemble a new Intel Haswell [6] microarchitecture. Additionally, McPAT calculated the leakage power dissipation based on technological parameters.
- The power data obtained from step 2 were used as an input to Hotspot [3] thermal simulator, which computed the temperatures in the chip. Hotspot uses a dynamic compact thermal model and considers the processor package to be composed of four layers (chip die, thermal interface material, heat spreader and heat sink). In our study, processor package parameters (dimension and physical properties of each layer) were typical for modern processor packages and were chosen based on our previous experiments [4,9]. Although Hotspot has also support for simulating 3D architectures, in our study, we used an extension of Hotspot [15], which uses a similar approach, but with a more detailed 3D thermal model.

2.1. gem5 simulator

We simulated out-of-order x86 processor using full-system simulation mode (as opposed to simpler Syscall Emulation mode in which system calls are emulated). Gem5 model is extremely detailed (see the authors' publication [1]) and we configured its parameters according to the publicly available data for modern Intel processors. As a benchmark, we chose SHA cryptographic application from mibench suite [5]. All processor cores simultaneously executed the same benchmark in parallel. According to our simulations this benchmark caused the highest total power dissipation among all benchmarks that we tested.

2.2. McPAT power model

McPAT calculates power dissipated in particular chip units based on unit activity data obtained from gem5. However, it also needs information about the parameters of the technology in which the processor was manufactured. Although in the version that we used McPAT does not explicitly support technologies below 22 nm, there exists in the code the configuration for lower technology node. Consequently we unblocked this option and ran the simulations using this new configuration. McPAT provides power statistics for small processor units such as Instruction Fetch unit, Load Store unit, Execution unit etc. Using this information and knowing roughly the position of these units in processor's core, we were able to assign the appropriate power values to the units used in our thermal simulation.

2.3. Hotspot thermal tool

In the thermal model used in Hotspot [3], the package is composed of four layers (from bottom to top): die, thermal interface material (TIM), heat spreader and heat sink. The heat is generated in the die layer only. The heat dissipation from the top of the package to the ambient is modeled by one convection resistance (heat sink to ambient) which is a configurable parameter. The heat dissipation to ambient through other package sides is neglected. Each layer is divided into a 2D grid. Each grid element is modeled by one capacitance, four lateral resistances (resistance to adjacent elements in the same layer) and two vertical resistances to the corresponding element in the upper/lower layer.

Material properties used for simulation include thermal conductivity, specific heat and density for all layers. They were found assuming that the die is made of silicon while heat spreader and heat sink of copper. For the TIM layer, thermal conductivity was set to 4 W/mK (a typical value for thermal paste) and its thickness was