POLITECHNIKA ŁÓDZKA

Wydział Elektrotechniki, Elektroniki,

Informatyki i Automatyki

Melvin Estuardo Galicia Cota

Praca doktorska

Modelling of multicore processors for the investigation of temperature reduction methods

Modelowanie procesorów wielordzeniowych w celu badania metod obniżania maksymalnej temperatury

Nr albumu: 800615

Opiekun pracy:

prof. dr hab. Andrzej Napieralski

Promotor pomocniczy:

dr Cezary Maj

Łódź, 2016

Modelling of multicore processors for the investigation of temperature reduction methods

ABSTRACT

The reduction of hotspot temperatures reached in modern microprocessors fabricated in nanometer technologies, even by a few degrees, can potentially generate significant improvements on the processor overall performance. In particular for high performance processors, lower temperature has direct impact into lowering cooling expenditures, less reliability issues, higher operating frequencies and extending their lifetime. All these advantages have direct impact on the enhancement of the processor's computational power.

Therefore, the investigation of temperature reduction techniques has become very important area. This thesis initiates the temperature behaviour study by analysing the differences that exist on the thermal models used nowadays to predict the temperatures inside modern processors. Thus, in this work, it is described and compared the two main used approaches, a detailed Finite Element Method (FEM) based simulation and a simpler architectural compact model. Results show that both models output similar results when predicting the processor maximum temperature, however, there are also non-negligible discrepancies when predicting thermal gradients within a chip. Furthermore, results show some differences in temperature profile during microprocessor heating process.

This work continues the investigation on microprocessor temperature, by studying how the positioning of particular processor units inside the floorplan impacts the final chip temperatures in 2D and 3D ICs. For this, it is taken into consideration a processor manufactured in 14 nm technology containing eight and six cores to simulate the temperature distribution for several floorplan designs. Results show that the difference in maximum temperature in various floorplans can achieve notorious reduction. Moreover, the idea for thermal buffers is introduced. But despite results obtained for 2D processors showing that thermal buffers may not provide sufficient thermal alleviation, results also show that, for 3D chips, thermal buffers combined with vertical thermal vias can effectively reduce the temperature of the chip hottest areas.

Following those transcendent results, this work focuses more on 3D stacking of microprocessors. Technology that promises the accomplishment of high performance processors while allowing

the increment of transistors density inside a given volume. It is well known that 3D stacking major challenge is overcoming thermal issues due to excessive power density and, therefore, this must be overcome before all the advantages that 3D processor promises can be exploited. The use of vertical thermal vias, by assuming they can be placed anywhere in the chip, has been widely suggested in literature for the mitigation of thermal problem as they facilitate heat transfer between stacked layers. Instead, in this work, it is proposed and investigated the use of localized thermal vias. They are inserted only into dedicated silicon areas around cores and they do not dissipate power. Result shows that without vias the heat flows vertically to the heat sink, almost uniformly over entire chip. But with localized thermal via regions the effect is favourable since an important amount of heat passes through them although no power is dissipated in these regions.

Moreover and in order to verify those findings, the thermal behaviour of an eight core processor simulated as a 2D chip and as a 3D architecture with two layers was analyzed by using an commercially available processor as an example. Localized thermal vias were utilized for improving the thermal behaviour of the 3D stack. Results confirm that the processor maximum temperature can be significantly reduced due to the implementation of thermal vias.

Additionally in this work, the liquid cooling technique was explored by analysing the performance of heat removal from a 3D stacked chip using microchannels. Thus, detailed stacked layers composing a 3D chip were modelled and used for the development of a coupled thermo-fluidic finite element simulation for several microchannel designs. Results show the relationships between key chip parameters and microchannel parameters.

1.1 Theses

Based on the current strong urgency of finding mitigations or even a definite solution to the processor thermal problem, explained in the background section, this dissertation brings forward the following theses:

Thesis 1:

> **A substantial temperature reduction of the 3D processor's hotspots can be obtained by using through silicon vias implemented in localized thermal via regions, located next to chip cores.**

Thesis 2:

> **The drawback of the implementation of localized thermal via regions next to 3D processor's cores is the formation of an additional thermal gradient in the chip, whose magnitude increases the more the hotspot temperature is reduced.**

Both theses are addressed during the development of Chapter 3 and 4, presenting results that provide support to these claims.

2 CONCLUSIONS

In this work, by using as reference a high-performance modern multicore processor, the two main used thermal modelling approaches were compared in order to describe their respective advantages and disadvantages. Results from steady-state analysis for three different cases of power inputs showed that the compact model, based on RC network, outputs similar results to the FEM analysis when predicting the maximum temperature inside the microprocessor. The use of compact model has the advantage of shortening the required simulation time by the order of magnitude. However as disadvantage, the compact model overestimates the temperature gradients generated within the processor by several degrees. This, possibly, due to, the simplification of lateral heat transfer between two blocks lying on different layers on the edge of the chip made by the compact model. Additionally, transient comparison showed that compact model may not accurately describe over time the heat transfer inside the processor package. The heating profiles of both methods show differences, usually due to the RC model underestimation of the temperature during initial and final phase of heating and overestimation in the middle. Nevertheless, both methods predict the same final steady-state result at the end of transient analysis.

Furthermore, results presented in this thesis showed that from the thermal point of view, an efficient floorplanning has significant advantages. Even though temperature reduction depends on several factors, by using an optimized floorplan design, for a typical power distribution case, the processor maximum temperature can be reduced by several degrees. Results have shown that the hottest microprocessor parts need to be placed near the center of the chip, but not be placed next to each other. Following this guidance can warrant the lowest maximal temperature for a given power dissipation and the lowest temperature gradient over the chip. However cannot be ignored the fact that thermally-optimized floorplan may represent suboptimal from other points of view. For instance, the mentioned guidance may have conflicts with the requirement of reducing interconnections length. Consequently, producing the need of trade-offs between thermally-optimized floorplan and other designing aspects.

Results in this work have additionally shown that thermal buffers located between processor cores can be used to implement localized vertical thermal vias insted of being distributed over the entire silicon chip layer. Implantation of thermal vias can reduce the maximum temperature in 3D processors by several degrees by providing a lower thermal resistance between stacked layers. This is possible due to the fact that the heat generated in lower layers is more efficiently transferred to the ambient. Thus, the analysis presented in this work also quantified the impact of the size and density of via regions on heat transfer. Results showed that improvement depends also on layers thickness, via region width and other important parameters that augment the thermal benefit of via region. Obtained results aim to provide a better understanding of the mechanism of heat flow in 3D ICs, helping to design analytical models for via regions, and the selection process of parameters for via region in future floorplans.

This work also presented a detailed analysis of the implementation impact of localized thermal vias for a designed 3D processor. Thermal vias were located inside regions where they do not disrupt an existing and optimized processor design. Results showed that, while dissipating the same amount of power, it is possible to reduce the maximum temperature of a 3D chip to almost the same maximum temperature of the same 2D processor. Additionally, it was found that the 3D chip peak temperature has an exponential relationship with the via region width. Such that at certain point, increasing via region width does not significantly reduce the temperature any longer. Also was found that thermal vias implementation brings also one disadvantage, the increment of local temperature gradients, which translate into potential reliability problem for the processor.

Finally, in this work was modelled a general 3D chip including detailed chip layers for the investigation of liquid cooling as temperature reduction technique. Simulations required a coupled thermo-fluidic analysis that considered several chip parameters as inputs and outputs. Results showed the relationships between parameters such as the peak temperature, the pressure drop, fluid flow rate and microchannel geometry involved on the chip overall temperature response. Obtained parameters relationships can help the search of optimal chip parameters. Although specific numerical results of the heat removal with the use of liquid cooling are presented, those values were obtained using software simulation and not measurements. However, they can provide a general vision on thermal efficiency of microchannels.

In conclusion, modelling of several microprocessors were carried out in this work in order to quantify the impacts of the currently investigated temperature reduction techniques. Results obtained in this thesis represent only the thermal point of view. Therefore, a combined effort of researchers from various domains is encourage in order to overcome current issues addressed on the investigated temperature reduction techniques, as they may become a major step on the mitigation of thermal problems in future microprocessors.