

Thermal Coupling in Technologies Based on Tri-gate Transistors

M. Janicki, P. Zajac, M. Szermer and A. Napieralski

Department of Microelectronics and Computer Science, Lodz University of Technology
Zeromskiego 116, 90-924 Lodz, Poland, janicki@dmcs.pl

ABSTRACT

This paper discusses the problem of thermal coupling among various microsystem components in future tri-gate transistor technologies. The considerations are based on the results of thermal simulations performed for a test IC which was designed specifically to mimic the thermal behavior of microprocessors manufactured in various technologies. The power trace data for thermal simulations were obtained combining the standard Wattch program with the BSIM-CMG and the predictive technology models.

Keywords: multi-core architectures, IC thermal modeling.

1 INTRODUCTION

The ever increasing density of power dissipated in the state-of-the-art microsystems has forced engineers to search for new ways of distributing the heat generation within semiconductor dies. This led to the development of many-core microprocessor architectures and various multi-tasking and multi-threading scheduling techniques, which helped to spread the power dissipation all over the die surface and consequently reduce the hot spot temperature values.

However, previous research carried out by the authors in [1] indicated that when migrating to newer technologies the dynamic thermal coupling among microprocessor cores would increase and very soon all these techniques might cease to be effective since each system component would instantaneously heat also its neighbors. Another important threat identified there was the continuous increase of the leakage power, which affected the overall static temperature rise. This observation remained true even for the planar technologies with high-k gate dielectrics.

Possible improvements in this area have been brought by the introduction of the fully depleted tri-gate transistors, however this issue still needs to be investigated. Thus, the authors decided to embark upon this subject by analyzing both static and dynamic thermal coupling in circuits based on these transistors. The approach proposed by the authors is to combine the standard methodology to predict power dissipation in microprocessors using the Wattch software with the BSIM-CMG tri-gate transistor and the predictive technology models.

The next section of this paper describes in detail the power modeling methodology. This is followed by the brief presentation of the test ASIC used as a thermal benchmark and the thermal modeling method employed later in the analyses.

2 POWER MODELING

In our approach we considered an architecture based on the well-known Alpha21364 processor [2] scaled down to 20 nm and 10 nm technology nodes. Consequently, the simulated processor parameters were configured according to those for the Alpha21364. The supply voltage was set to 0.9 V and 0.75 V for 20 nm and 10 nm technologies respectively. The power data for the scaled Alpha21364 chip is obtained using the well-known Wattch power model [3] integrated into the cycle-accurate performance simulator SimpleScalar [4]. The model parameters were calculated using the standard scaling method employed in many other publications [5-6]. Moreover, taking into account that in modern technologies leakage power is a major factor and it cannot be neglected, the leakage power was calculated in our model based on the HotLeakage tool [7]. The entire power modeling methodology is summarized in Fig. 1.

The inclusion of the new technologies required several enhancements to be made in the Wattch and Hot Leakage tools. In order to estimate the parameters of leakage model for modern technologies, we performed the following steps. First, we run the Cadence Spectre circuit-level simulations of NMOS and PMOS transistors and calculated the leakage current for various temperature values and variable number of fins. In the simulations we used the 20 nm and 10 nm Predictive Technology model [8], based on the recent BSIM-CMG model destined for multi-gate FETs. Then, the leakage current in function of temperature and the number of transistor fins was found. Next, an iterative procedure was applied to fit the leakage current data obtained from the HotLeakage model to the data from the Spectre program. Finally, after fitting all the functions, the estimated model parameters were included into the leakage model.

Taking into account that the HotLeakage allows only the calculation of leakage for memory-like structures, the leakage power for ALUs was only roughly estimated based on their area. The average power dissipated in Alpha21364 processor under heavy workload was estimated with the SimpleScalar tool executing various standard SPEC2000 benchmarks. In order to make sure that all the most computationally-intensive parts of each benchmark were executed the simulator is initially fast-forwarded several hundred million instructions and then several hundred million instructions are simulated with detailed power data statistics. In this way, the average dynamic and static power dissipation for every processor unit is estimated. A more detailed description of the power modeling methodology can be found in [9].

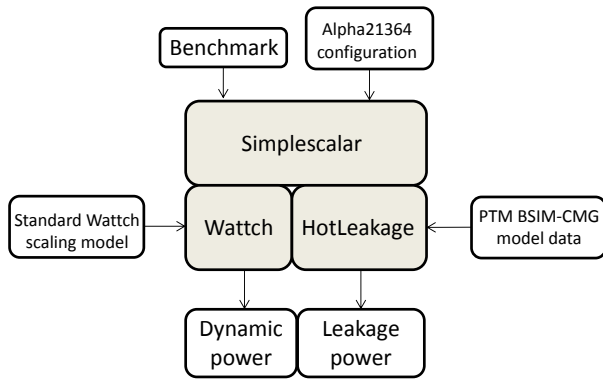


Figure 1. Power modeling methodology.

3 BENCHMARK ARCHITECTURE

The power trace data computed in the way described in the previous section was used then in thermal simulations of an ASIC [10], which is supposed to mimic the behavior of a microprocessor manufactured in different technologies. This ASIC, sketched in Fig. 2, contains a large 16 x 24 matrix of transistor heat sources which can be individually switched on and off at a desired power level at a specified time instant. The dimensions of heat sources were chosen so that an integer or a floating point arithmetic unit in a real microprocessor would correspond in this circuit, as shown in the figure with the black squares, to a single heating cell in the 10 nm technology node and 4 cells in the 20 nm technology node. Similarly a 2 MB cache memory unit will correspond to 20 and 80 heating cells respectively. Owing to this solution it is possible to analyze thermal phenomena occurring in the state-of-the-art technologies experimenting with a circuit manufactured in a much older technology, which is much more cost effective.

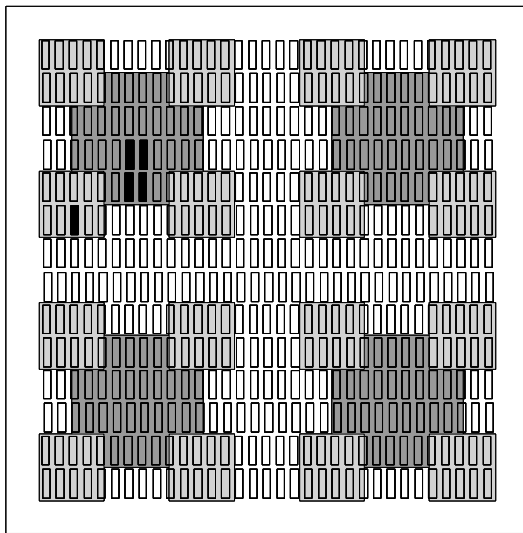


Figure 2. Thermal test ASIC floorplan.

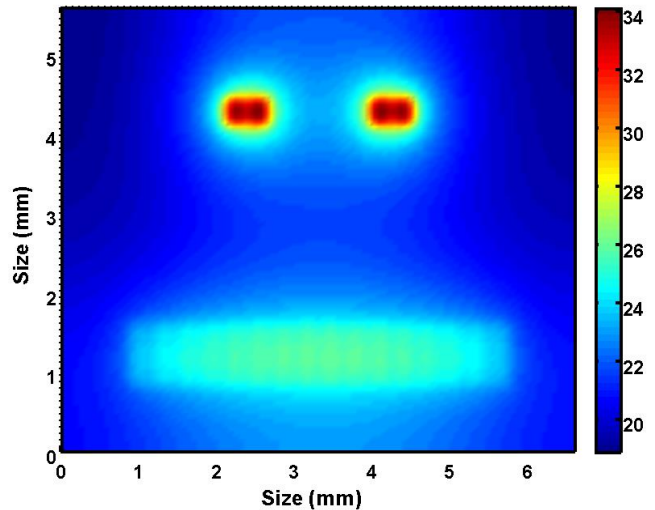


Figure 3. Temperature rise map for the 20 nm node with 2 ALUs and 2 MB cache memory active.

4 THERMAL SIMULATIONS

Thermal simulations of the benchmark ASIC are carried out based on the previously computed power trace data with a Green's function based solver [11]. During the simulation two different configurations were considered. In the 20 nm technology node the circuit was supposed to have a 2 MB cache memory (occupying the surface 80 heating cells) and ALUs (integer or floating point ones) occupying 4 heating cells. In the 10 nm technology, the circuit occupying the same area was assumed to have an 8 MB cache memory, hence occupying the same number of heating cells, and its ALUs occupy only one heating cell.

The power trace data obtained from the Wattch program computed for the 3.5 GHz operating frequency and at 75 °C indicated that the leakage power is almost negligible in the ALUs and for the cache memory it constitutes 25 % of the total power in the 20 nm node and 65 % in the 10 nm one. Even for such an important number of heat sources the temperature map at 10,000 points with the accuracy better than 0.1 K were carried out in less than 10 minutes on the Core i7 computer.

4.1 Static Analysis

The steady state thermal simulation results are provided in Table 1 and in Figure 3. The data in the table contain for both technology nodes considered here the information on the minimal, maximal and average temperature values when two neighboring ALUs executing the same SPEC benchmarks are placed as close as it is allowed by the technology rules, i.e. they are separated by the space large enough to fit between them another ALU. Moreover, the temperature rise values due to the power dissipation only in the cache memory are also given.

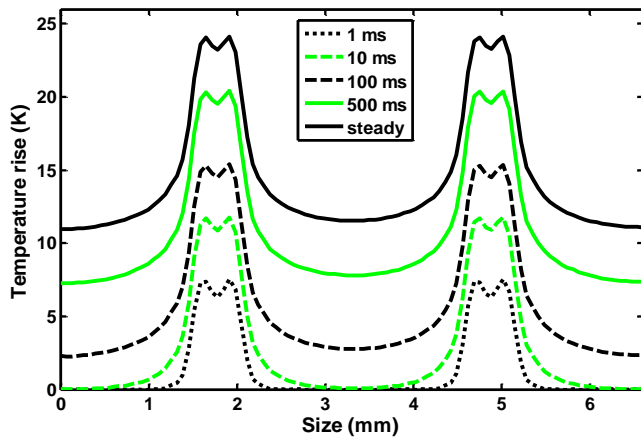


Figure 4. Temperature evolution for 20 nm node and 2 ALUs placed apart.

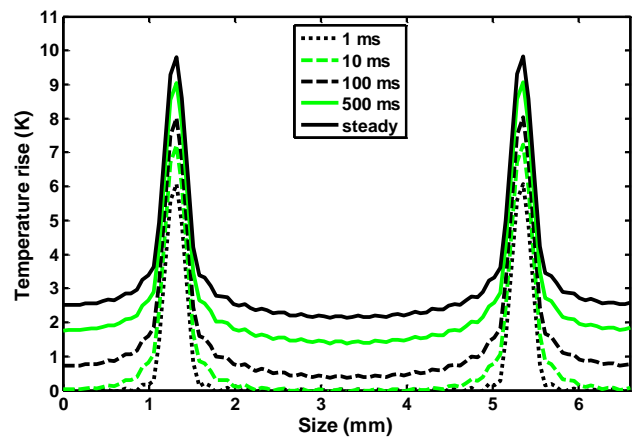


Figure 6. Temperature evolution for 10 nm node and 2 ALUs placed apart.

Case	Temperature rise (K)		
	minimal	average	maximal
20 nm ALUs close	18.9	21.9	34.1
20 nm cache only	8.5	12.0	18.4
10 nm ALUs close	12.5	16.7	24.0
10 nm cache only	10.4	14.7	22.5

Table 1: Simulated steady state data.

The data provided in the table prove that, compared to the traditional planar technologies, the tri-gate transistor technologies can be particularly beneficial from the thermal point of view. Namely, they demonstrate that the migration from the 20 nm to the 10 nm technology node might bring about the reduction of temperature rise by 24 % on average if the same performance is maintained. Another interesting observation is that in the 10 nm technology node much more temperature rise is due to the power dissipation in the cache memory, mostly due to its increased size.

4.2 Dynamic Analysis

The results of the dynamic simulations obtained for both of the considered technologies are presented in Figures 4-7 which show the evolution of temperature profiles along the line passing through the middle of power dissipating ALUs. The figures at the top show the temperature distribution when heat sources representing the ALUs are placed apart and the ones at the bottom when they are placed as close as possible.

The analysis of these figures leads to very interesting conclusions. First of all, it can be observed that the dynamic coupling between the cores is less important in the 10 nm technology node, even when the units dissipating power are closer to each other than in the 20 nm technology. Namely, in the latter node the coupling is around 15 % after 100 ms and 25 % in the steady state. For the 10 nm node these numbers amount to 10 % and 20 % respectively. When the ALUs are placed further from each other thermal coupling is decreased and these differences between the technology nodes are even more striking.

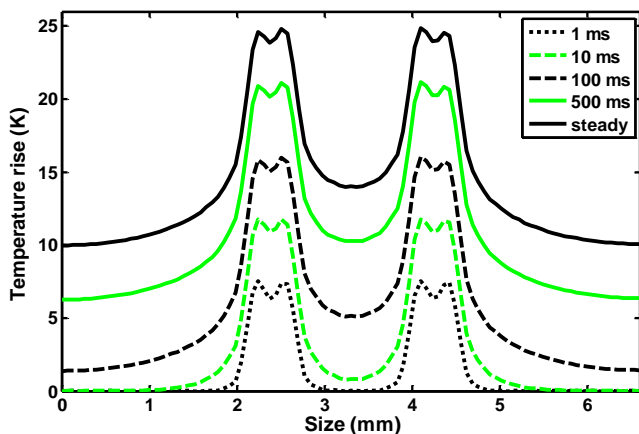


Figure 5. Temperature evolution for 20 nm node and 2 ALUs placed close to each other.

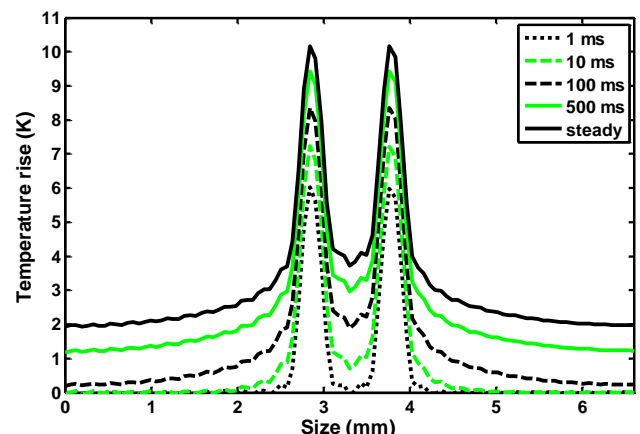


Figure 7. Temperature evolution for 10 nm node and 2 ALUs placed close to each other.

5 CONCLUSIONS

This paper discussed an important problem of thermal coupling between individual components of microsystems manufactured in technologies based on tri-gate transistors. An interesting conclusion from the simulations presented here is that, unlike in the case of standard planar MOS technologies, when scaling down the tri-gate devices the thermal coupling between particular microsystem blocks might decrease what could be beneficial for performance improvement.

The results presented here were obtained in a novel way combining the standard cycle accurate microprocessor simulators with the predictive technology transistor models. This allowed the prediction of dynamic and leakage power dissipation in individual microprocessor components during the execution of standard benchmarks. These power trace data were used then in the thermal simulator to compute the temperature distribution across the silicon die. Another original idea presented in this paper is that the migration to next generation technology nodes is emulated by the test integrated circuit whose individual heat sources or their blocks correspond to functional blocks of microprocessor components.

Obviously, it should be clearly said that the simulated values presented in this paper are only indicative, however in the absence of reliable power data currently it is the only possibility to compute processor temperature. This remark particularly concerns the electrical model parameters of tri-gate transistors which are hard to predict and still require further calibration.

ACKNOWLEDGMENTS

The research presented in this paper has been supported by the grant of Polish National Center of Science No. N515 5091 40.

REFERENCES

- [1] M. Janicki, J. Collet, A. Louri and A. Napieralski, "Hot spots and core-to-core thermal coupling in future multi-core architecture", in Proc. 26th SEMI-THERM, pp. 205-209, 2010.
- [2] R. E. Kessler, E. J. McLellan and D. A. Webb, "The Alpha 21264 microprocessor architecture", in Proc. 16th ICCD, pp.90-95, 1998.
- [3] D. Brooks, V. Tiwari and M. Martonosi, "Wattch: a framework for architectural-level power analysis and optimizations", in Proc. 27th ISCA, pp. 83-94, 2000.
- [4] T. Austin, E. Larson and D. Ernst, "SimpleScalar: An Infrastructure for Computer System Modeling", Computer, vol.35, pp. 59-67, 2002.
- [5] K. Skadron, M. Stan, Sankaranarayanan, W. Huang, S. Velusamy, D. Tarjan, "Temperature-Aware Micro-architecture: Modelling and Implementation", ACM Trans. on Arch. and Code Optim., vol. 1, pp. 94-125, 2004.
- [6] R. Mukherjee and S. Ogrenci Memik, "Systematic temperature sensor allocation and placement for microprocessors", in Proc. 43rd DAC, pp. 542-547, 2006.
- [7] Y. Zhang, D. Parikh, Sankaranarayanan, K. Skadron, M. Stan, "HotLeakage: an Architectural, Temperature-Aware Model of Subthreshold and Gate Leakage", in University of Virginia Dept. of Computer Science Technical Report CS-2003-05, 2003.
- [8] S. Sinha, G. Yeric, V. Chandra, B. Cline and Y. Cao, "Exploring sub-20nm FinFET design with predictive technology models", in Proc. 49th DAC, pp. 283-288, 2012.
- [9] M. Szermer, P. Zajac, L. Kotynia, C. Maj, P. Pietrzak, M. Janicki and A. Napieralski, "New Methodology for Thermal Analysis of Multi-Core Architectures Based on Dedicated ASIC", article in press to be published in Microelectronics Journal.
- [10] M. Szermer, C. Maj, P. Pietrzak, M. Janicki, P. Zajac, and A. Napieralski, "Test ASIC for the investigation of thermal coupling in many-core architectures", in Proc. 28th SEMI-THERM, pp. 135-138, 2012.
- [11] Janicki, G. De Mey, and A. Napieralski, "Thermal analysis of layered electronic circuits with Green's functions", Microelectronics Journal, vol. 38, pp. 177-184, 2007.