

Advancing the thermal stability of 3D-IC's using logi-thermal simulation

Gergely Nagy¹, Péter Horváth¹, László Pohl¹, András Poppe*¹

¹ Budapest University of Technology, Budapest, Hungary

* Corresponding Author: poppe@eet.bme.hu

Abstract

3D-ICs have emerged in the past few years. While they solve a large number of problems related to scaling, they also create new ones. Removing the heat from the layers far from the cooling facilities is a great challenge still under intensive research. This paper shows how logi-thermal simulation can be used to predict the operation parameters of large digital systems realized in 3D-ICs. The method can be effectively used to guide place-and-route algorithms and to find the thermal bottlenecks.

1 Introduction

In recent years the technology enabling 3D-ICs has emerged [1]. Scaling in the traditional chip-design technologies have become difficult due to power density constraints and the fact that although transistors are becoming smaller and faster but this does not apply to the wiring.

3D-ICs have the advantage of making large scale integration possible at a small footprint which reduces prices both by simplifying testing and making PCB's decrease in size. The average wire length is also reduced, although the interconnections between layers have a considerably larger capacitance than in-die wires [2]. By keeping signals on-chip the power consumption is also cut down while the operating frequency is higher. A typical application where the latter can be utilized is a 3D stack of a processor and a memory.

3D-ICs also make it possible to integrate circuit layers created with highly different processes. This enables the optimization of the components to a much higher degree than if they were built on a single wafer. Moreover system components requiring incompatible processes can also be integrated.

Although the problem of high power densities due to large scale integration is partially alleviated by stacking wafers, but new difficulties arise: the generated heat has to be taken away from every layer which can become an issue for those that are far from the heat sink. Simulation results in [3] showed that power dissipation in a 3D stacked structure results in a 3 times higher maximum temperature increase compared to the 2D case. Thus, efficient floor-planning techniques are essential for 3D ICs.

The vias (TSV – Through Silicon Via) connect layers in a 3D stack [4, 5]. They transmit power and signals but also the heat towards the heat sink in the package. Thus the placement and density of vias is vital for efficient cooling in 3D stacks [6].

2 Methodology

Hot spots and large temperature gradients in high scale integrated circuits can break signal integrity. Such failures are extremely hard to predict as simulations performed at

room temperature or at an evenly spread elevated temperature might not indicate them. Large systems have to be simulated using the exact temperature distribution that occurs when they have reached their thermal steady state.

This kind of investigation is of great importance in 3D stacks where the thermal issues experienced in 2D integrated circuits reappear in an even more complex fashion. Logi-thermal simulation is a very effective method that enables fast and effective analysis of digital circuits with their thermal behaviour taken into account.

Furthermore the thermal interaction of the chips can be explored. Logi-thermal simulation calculates the thermal feedback within the chips and between them as well. The TSVs not only play a role in removing the heat from the layers, they also contribute to the thermal coupling between the stacked chips. These effects are completely unique to 3D-ICs and need attention at design time.

The effect of TSVs placed right where hot spots are formed is also investigated.

3 Logi-thermal simulation

Logi-thermal simulation performs the logic and the thermal simulation of a digital system in a close coupling allowing designers to investigate the thermal behaviour of a digital circuit already during the design process. The advantage of using a logi-thermal simulator over electro-thermal simulations is that the digital behaviour is determined by a logic simulation which is orders of magnitude faster than an analog simulation.

A logi-thermal simulator performs a simulation cycle in the logic engine and determines the set of components that were active (and thus dissipated power) in that cycle. It then forwards this information to a thermal simulator together with the actual dissipations which calculates the heat map of the circuit. The thermal simulator is aware of the physical layout of the circuit and will continuously update the heat map in each simulation cycle. The calculated temperatures are fed back to the logic simulator which updates the delay of the components, which is temperature dependent. The resulting simulation gives an accurate waveform showing the system's operation altered by the changing temperatures

and also a heat map that indicates the hot spots forming as the system heats itself.

Logic simulator engines need to be modified to be able to perform a logi-thermal simulation. Every logic component needs to remember its temperature and be able to calculate its delay and its dissipation based on the temperature value. These are properties and algorithms that average logic engines lack. Therefore a custom logic engine has been created in which components can be described in pure C++ very much like the way it is done in SystemC but inheriting the necessary knowledge for logi-thermal simulation.

The engine allows for modelling components at an arbitrary level of abstraction, from gate-level and RTL to system-level. As long as the components communicate through their ports, their description can take advantage of the entire toolset of the C++ language. The datatypes of the ports are not limited either so basically any system can be modelled (even analog circuits at system-level).

4 The 3D-capable Sunred engine

The thermal simulator used in a logi-thermal simulation can be an ordinary simulator engine, nothing needs to be modified for this operation.

To gain temperature distribution in the chip we use a thermal field simulator engine linked to the program as a library. The simulator solves the 3D heat equation by discretizing the field (the chip) by Finite Differences or Finite Volumes Method to get ordinary differential equations (ODEs) from the heat equation which is a partial differential equation. The discretization is done by a 3D orthogonal mesh (see Fig. 1) where each cell is modelled with homogeneous material. The gained ODEs can be interpreted as an electrical network. In case of the logi-thermal simulator each cell is modelled with the network presented in Fig. 1. The electrical network is solved by the Successive Network Reduction (SUNRED) algorithm. Discretization of the field, SUNRED algorithm and their application in case of one-layer logi-thermal simulation were presented in our earlier publications [7, 8, 9].

In case of a monolithic one layer chip the thermal problem could be modelled as a homogeneous silicon block. Modelling a stacked die establishes new requirements: between two silicon layers there is an insulating layer of other material, and there can be thermal/electrical vias between/through layers [10]. Both new requirement can be handled efficiently by applying non-equidistant mesh pitch thus, different layer thicknesses can be used and via sizes can be set appropriately. Setting the material individually for each cell was also important for flexible via placing.

An interface class (*stackedInterface*) has been defined in C++ for the thermal engine to set up the simulation. The constructor of this class requires x , y and z resolution (=cell number in x , y and z direction) of the thermal model, and number of the applied materials. Maximum 256 materials can be used but in most cases three materials are enough (silicon, insulator, via). Next steps of the thermal engine set up are (a) the setting of pitch sizes, (b) setting thermal

parameters (heat conduction and heat capacitance) of the materials, (c) association of cells and materials and (d) the setting of the boundary conditions.

After the initialization the simulation loop starts: temperature distribution is calculated at each time step. Input data for the thermal engine is the dissipation map generated by the logic engine for the actual time step. Dissipation can be applied on the top of the layers. In the cell model it is represented by a heat source, marked $P_{dissipated}$ in Fig. 1. After the simulation step the thermal engine provides a layer-top temperature map for the logic engine.

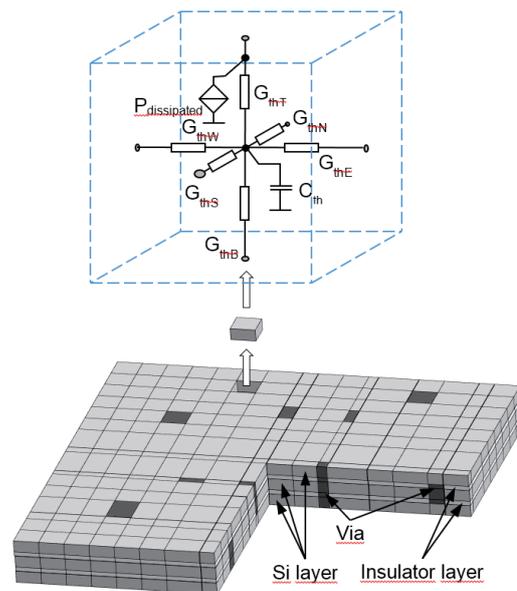


Figure 1. The model representing the 3D IC's structure in Sunred

5 TSV-placement optimization with logi-thermal simulation

In this section we present the initial simulation results achieved with the logi-thermal simulation engine. The simulated model included a simple microprocessor core, a 256-byte data memory and a 256-entry instruction memory. The memories and the core were placed onto two different dies and the interconnections were modeled with TSVs. The microprocessor implements a 3-address instruction-set with 12 machine instructions encoded in 16-bit instruction words. The core has an 8-bit, non-pipelined datapath with an internal register file including 16 entries. An iterative division algorithm was running during the logi-thermal simulations.

As TSVs function both as data transmitting wires and heat spreaders, several layouts have been simulated with the TSVs positioned on different locations. The simulation results show that placing the TSVs where the hot spots develop helped reduce temperature peaks by 20% and thus decreased the risk of thermally induced errors.

Fig. 2 and 3 show the arrangement of the design entities on the dies, the TSVs and the simulation results achieved.

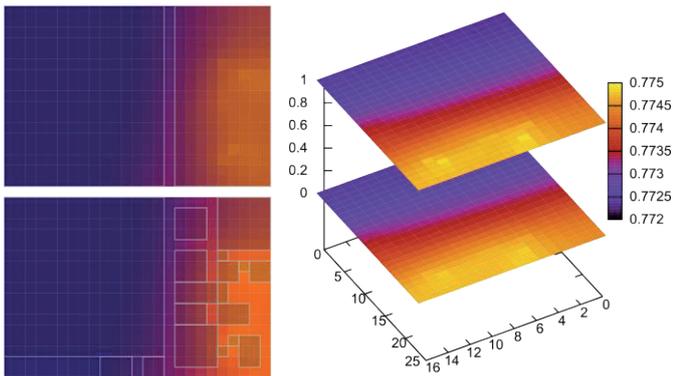


Figure 2. Logi-thermal simulation results without TSVs

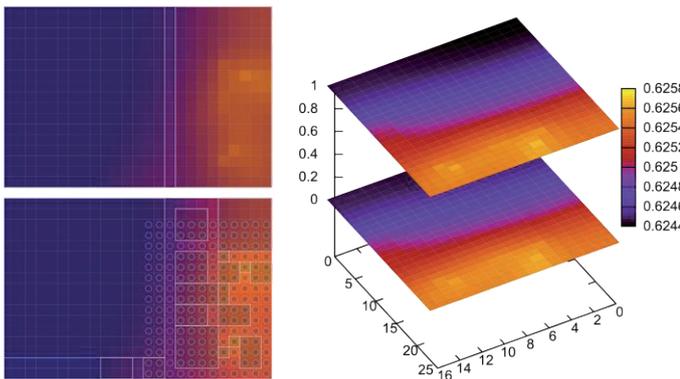


Figure 3. Simulation results with TSVs under the hotspots

The temperature range and the hot spots are quite low. The reason for this is that although the simulated system is realistic and relatively complex but it can still be considered small having a low power dissipation. The effectiveness and usefulness of the method are demonstrated by these simulations nonetheless.

6 Model selection and power-consumption estimation

The model granularity is the key concept, which has to be taken into consideration when selecting a model for logi-thermal simulation. A fine-grained model consists of a significant amount of design units, which can be handled independently in terms of thermal behaviour (that means that the power consumptions and areas of the design units can be determined independently), while a coarse-grained representation comprises more functionality in a single design unit. The model granularity leads to a trade-off between thermal resolution and the time-consumption of the simulation. The simulation models with different granularities usually represent different abstraction levels as well. The fact that the proposed logi-thermal simulation method is capable of handling arbitrary levels of abstraction makes it possible to investigate, which level is feasible to use for different purposes. We chose the Register-Transfer Level (RTL) as the baseline of this investigation, since it is the highest abstraction level, which can directly be correlated with the physical quantities used in logi-thermal simulation, such as power consumption, without any complicated power-estimation methods.

We prepared the simulation models presented in Section 5 using the implementation schemes called structural RTL and behavioural RTL defined by [10]. These two model types represent slightly different levels of abstraction but both of them are capable of underlying an automated RTL synthesis. Four instances of the models (core, instruction memory, data memory) were placed onto a single die-pair and a 250 ms logi-thermal simulation was performed. The simulation results including the temperature distribution are shown in Fig. 4.

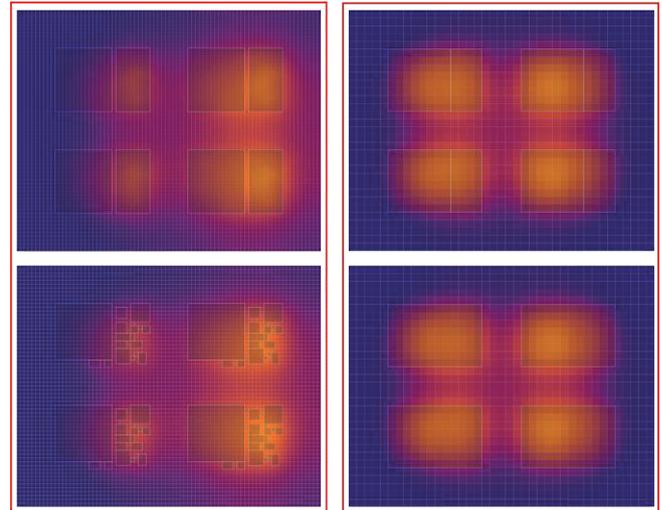


Figure 4. Fine-grained and coarse-grained simulation results of a four core microprocessor system

The thermal maps on the left side of Fig. 4 show a high-resolution simulation performed on the structural RTL models of the microprocessor system while the thermal maps on the right side show the results of the simulation performed on the behavioural RTL models. In the latter case the resolution of the thermal engine was adapted to the granularity of the RTL model. The grey frames indicate the RTL design entities. The high-resolution structural RTL simulation took 61.78 s to perform while the lower resolution behavioural RTL simulation ran for 2.95 s.

The simulation results show how the multi-abstraction nature of the simulation method may be exploited for simulation-time optimization in large systems. In such cases the main components, whose detailed structure is irrelevant in terms of thermal issues, may be represented by high-level models and the resolution of the thermal engine may be adapted to the coarse model granularity in order to decrease simulation time [12]. The limit of this abstraction increase is the technique of estimating the power-consumption of a single design unit; when performing a logi-thermal simulation, the power-consumption characteristics of the design components have to be determined as accurately as possible in order to achieve reliable results. The logic gate is the basic design component whose power-consumption can be precisely estimated based on its data sheet and size. Therefore, in the proposed simulation procedure the designer has to create the structural RTL representation of the system and, before executing a logi-thermal simulation, a pre-synthesis has to be performed in advance in order to

determine the type and the amount of logic gates constituting the elements of the structural RTL model. Then the power-consumption, and also the area properties, of the RTL design elements can be determined by summing up the power-consumptions of the well-known logic gates implementing them with regard to their data sheets and scaling factors. Finally a logi-thermal simulation of the structural RTL model can be performed.

The power-consumption data retrieved with the above described method can be used with higher-level models with an important restriction. The higher abstraction level models may also be able to underlie an automated RTL synthesis procedure, e.g. the behavioural RTL model, but the synthesis results can be significantly different from those reached by the synthesis of the structural RTL model. If a designer wants to perform a logi-thermal simulation based on a higher-level representation of the system in order to decrease the simulation time, and the power-consumption and area information are based on the structural RTL model then it has to be considered that the simulation results will only be highly correlated with the real thermal behaviour of the synthesized circuit if the RTL synthesis is not based on any other representations but the structural RTL model.

There is one interesting phenomenon that needs to be considered when using a high-level, coarse-grained model which is also shown in the figures above. When a set of low-level components is substituted with one high-level model, the dissipation of the latter is assumed to be homogeneous. In our case, the largest dissipators of the processor core were positioned on the right side of the layout and thus the hot spot formed right to the center of the core's bounding box. The four cores are placed in a symmetrical arrangement which results in a horizontal shift of the hot areas of the cores. Subsequently the right side of the chip will be hotter than the left hand side and so a hot spot of a higher temperature value develops on the right hand side than what could be determined from the behavioural simulation. Therefore when falling back to a high-level model, apart from a lower resolution, qualitative errors may also be expected.

The highest precision can be achieved by using a gate-level model. The dissipation of gates or standard cells can be exactly determined by analog simulations. In the case of standard cells these values together with their temperature dependence are usually given by the manufacturer. These details can be incorporated into some logi-thermal simulators directly [13] providing a very smooth integration of this new tool into the existing design flow.

7 Conclusions

This paper shows how logi-thermal simulation can be used to ease the job of the designers of large digital systems to be realized in 3D-ICs. Guidelines for the optimal density of Through Silicon VIAs and for the place-and-route algorithms can be determined by logi-thermal simulations. We propose that digital 3D-ICs be investigated by logi-thermal simulation as the effective heat conductivity of the

interconnect layers and heating effect of neighbouring layers affect signal transmission in completely different ways as in 2D chips. Lower lateral thermal gradients can be expected as a result of the thermally conductive layers between the layers compared to 2D chips but higher temperatures are foreseen on chips on top of a stack cooled from the bottom and on chips in the middle in case cooling from both sides.

References

- [1] Dick James, "3D ICs in the Real World", Proceedings of the 25th Advanced Semiconductor Manufacturing Conference, ASMC 2014, Saratoga Springs, NY, May 19-21 2014, pp. 113-119
- [2] Paul Falkenstern, Yuan Xie, Yao-Wen Chang, Yu Wang, "Three-Dimensional Integrated Circuits (3D IC) Floorplan and Power/Ground Network Co-synthesis", Proceedings of the 15th Asia South Pacific Design Automation Conference, ASP-DAC 2010, Taipei, Taiwan, January 18-21, 2010. IEEE 2010 ISBN 978-1-60558-837-7
- [3] Geert Van der Plas, Paresch Limaye, Igor Loi, Abdelkarim Mercha, Herman Oprins et al., "Design Issues and Considerations for Low-Cost 3-D TSV IC Technology", IEEE Journal of Solid-State Circuits, Vol. 46, No. 1, January 2011
- [4] H. C. Cheng, C. Y. Yang, Alan Cheng, Karl Cheng, "Micro Bridge Technology for 3D-IC Interconnection Could Benefit the 3D-IC Test Strategy", Proceedings of the 7th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT), 2012, IEEE Catalog Number: CFP1259B-ART, ISBN: 978-1-4673-1638-5
- [5] John H. Lau, "Evolution, challenge, and outlook of TSV, 3D IC integration and 3d silicon integration", Proceedings of International Symposium on Advanced Packaging Materials, 2011
- [6] Z. Li, et al., Thermal-aware P/G TSV planning for IR drop reduction in 3D ICs, INTEGRATION, the VLSI journal (2012), <http://dx.doi.org/10.1016/j.vlsi.2012.05.002>
- [7] Székely V: SUNRED a new thermal simulator and typical applications. 3rd THERMINIC Workshop (1997), pp. 84-90.
- [8] L. Pohl, Zs. Kohári, V. Székely, Fast field solver for the simulation of large-area OLEDs, Microelectronics Journal, Vol.41, No.9, pp. 566-573 (2010)
- [9] G. Nagy, L. Pohl, A. Timár, A. Poppe, Yield enhancement by logi-thermal simulation based testing. THERMINIC'12. (2012) pp. 196-199.
- [10] H. Oprins, et al., "Fine grain thermal modeling and experimental validation of 3D-ICs", Microelectron. J (2010), doi:10.1016/j.mejo.2010.08.006
- [11] P. Horváth, G. Hosszú, F. Kovács (2014) A Proposed Novel Description Language in Digital System Modeling, In: Mehdi Khosrow-Pour (ed) Encyclopedia of Information Science and

- Technology, 3rd edn., IGI Global, New York, pp. 22-37. doi: 10.4018/978-1-4666-5888-2.ch687
- [12] A. Poppe, G. Horváth, G. Nagy, M. Rencz, V. Szekely, „Electro-thermal and Logi-thermal Simulators aimed at the Temperature-aware Design of Complex Integrated Circuits”, Twenty-fourth Annual Semiconductor Thermal Measurement and Management Symposium, 2008. pp. 68-76.
- [13] Andras Timar, Marta Rencz, Temperature dependent timing in standard cell designs, Microelectronics Journal, Volume 45, Issue 5, May 2014, Pages 521-529, ISSN 0026-2692, <http://dx.doi.org/10.1016/j.mejo.2013.08.16>.