Modeling Temperature Distribution in Networks-on-Chip using RC-Circuits

Andreas Tockhorn, Claas Cornelius, Hagen Saemrow and Dirk Timmermann
Institute of Applied Microelectronics and Computer Engineering
University of Rostock, Germany
andreas.tockhorn@uni-rostock.de

Abstract—**As transistor dimensions are shrinking into regions of only a few atomic layers, designers are faced with various problems including increased reliability and power issues. Since these problems are amplified by higher circuit temperatures, this paper proposes an approach for the fine-grained modeling of temperature distribution in many-core systems based on Networks-on-Chip. With this model, algorithms can be developed that consider the significant impact of temperature – e.g. on performance, power or reliability. To simulate the dynamic nature of temperature, the thermal properties of according integrated systems are modeled through the instantiation of equivalent RC-circuits. This approach exploits the dualism between electrical and thermal flows of energy. Finally, an application with system control for task mapping and power management exemplifies the proposed simulation methodology.**

Keywords- Network-on-Chip, Temperature Modeling, Reliability, System control, Integrated circuit

I. INTRODUCTION

The continuous scaling of transistor dimensions has resulted in an ever increasing complexity and computing power. Though, some new issues arise from this development. On the one hand, there exist challenges how to utilize this increasing number of transistors efficiently. On the other hand, problems like reduced reliability and high power consumption gain in importance. Hence, this paper introduces an approach for modeling temperature distributions in Networks-on-Chip (NoC) in order to enable designers to explore algorithms tackling temperature-indicated problems.

A. Network-on-Chip

The first issue, how to handle the increasing amount of transistors, is tackled through the application of an ever increasing number of functional modules in so called Systems-on-a-Chip (SoC). Conventional bus-based systems though are limited through scalability issues of their communication and extensive synchronization overhead. However, a new on-chip communication paradigm called Networks-on-chip was introduced promising to solve many of these challenges [1]. The approach is to replace buses by an interconnection network based on modern data networks connecting computers. Furthermore, large global on-chip connections are replaced by short links connecting distributed routers.

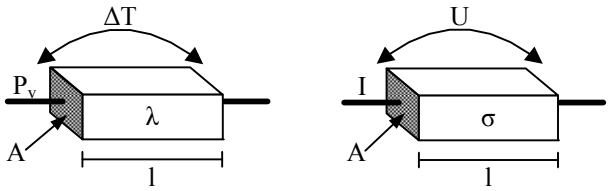
B. Temperature-indicated issues

Admittedly, NoCs are not able to solve the issues that are caused by transistor shrinking itself. However, a comprehensive observation holds true for the influence of temperature on performance and power consumption as well as on reliability. A strongly simplified but illustrative explanation can be given by the Arrhenius model [12]. This model describes the exponential influence of temperature on the velocity of chemical reactions. This depiction though emphasizes the great importance of thermal management for microelectronic circuits. According issues that are amplified by high temperatures are decreased reliability, performance loss and increased power consumption.

Thereby, the impact of temperature on reliability is hardly considered. Srinivasan et al. considered that raising temperature exponentially accelerates wear-out failures [11]. Thus, two mechanisms concerning wear-out effects are described in more detail for clarifying the correlations. These are electromigration and Time Dependent Dielectric Breakdown (TDDB). Electromigration affects the electrical connections and can result in interrupts of the physical wires and shorts between neighboring wires. The physical cause is the transport of material, more precisely of metal ions [7]. The second stated example is TDDB, which identifies another physical damage of microelectronic circuits. TDDB is a failure of the insulator between the gate and the channel of a transistor. As a result of TDDB, an unintentional electrical connection arises between gate and channel. First of all, this current path deteriorates the transistor characteristics. Over time though, this current path grows and may result in an incorrect logical function. High temperatures drastically accelerate both of the exemplarily mentioned defects. Furthermore, delays of gates in microelectronic circuits increase because the threshold voltage of transistors is a function of temperature. Hence, the maximum clock frequency of circuits decreases correspondingly [8]. The third mentioned issue is the higher power consumption under worse thermal conditions. Static power mainly consists of leakage in standard CMOS. Unfortunately, leakage increases exponentially depending on increasing temperature [5]. Large leakage currents also result in additional reliability issues due to nodes losing their charge and therefore their logical value.

Since temperature has such a wide influence on various parameters of integrated circuits and especially on their reliability and lifetime, several approaches exist to tackle this issue. Exemplarily, dynamic voltage scaling (DVS) and dynamic frequency scaling (DFS) are mentioned. Both approaches lower the power consumption of an integrated system with a penalty in performance resulting from lower switching frequency.

NoCs are well qualified to be managed during runtime concerning thermal issues using these two technologies because their architecture offers a fine granularity of processing elements, which can be monitored and controlled independently from each other. On the one hand, it is possible to consider



(a) Thermal resistance (b) Electrical resistance

Figure 1. Dualism of thermal and electrical resistance.

thermal implications during the mapping of applications on NoC-based systems. Therefore, an appropriate temperature model has to be used. On the other hand, a NoC-based system can be monitored during runtime using thermal sensors implemented within the chip. Therefore, an accurate model describing the temperature distribution in NoC is needed in order to evaluate runtime management algorithms in an early design stage.

C. Existing approaches to model temperature

There exist different approaches in order to model the thermal behavior of microelectronic devices. Many of them are based on the equivalency of thermal and electrical energy flows [2][3][10][13]. These approaches use electrical circuits utilizing resistors and capacitors to model the spreading of heat within the material. These circuits are mapped on appropriate differential equations, which are resolved by numerical approximations. Though, all these approaches model the temperature distribution of processors. Another drawback of these algorithms is their granularity. NoCs are well qualified to be managed aiming at a uniform temperature distribution. Therefore, a more fine-grained temperature model is needed, in order to investigate influences of mapping and control algorithms in NoC-based systems. For instance, the methodology proposed by Dhodapkar et al. treats a processor as a single component and therefore only investigates the resulting highest temperature and not the distribution within the chip [3].

Shang et al. proposes an algorithm based on the heat spreading angle at which thermal energy dissipates through the material of microelectronics [9]. This approach has to model different kinds of heat spreading. For example, it distinguishes heat transfer between neighboring routers and further away routers. This results in a more complex and less generic model.

Another completely different methodology to measure the temperature distribution of a microelectronic system is SIMP, which stands for Spatially resolved Imaging of Microprocessor Power [4]. There, an infrared camera measures temperature. Primary, this was introduced to derive the power consumption of chip components by their thermal image but it can directly be used to monitor temperature as well. The main drawback of this approach is the need for an already manufactured chip. Thus, investigations concerning task mapping cannot be done in advance.

II. MODELING OF TEMPERATURE

Since the proposed methodologies do not meet our demands concerning investigations on NoCs, we utilize the often used dualism of electrical and thermal energy flows to model thermal behavior of NoCs. Therefore, this section presents the analogies of thermal and electrical simulations and clarifies the necessary parameters. Subsequently, the applied assumptions are discussed before the implemented flow of steps for the simulation is introduced in more detail.

TABLE I. DUALISM OF THERMAL AND ELECTRICAL PARAMETERS WITH THEIR SYMBOLS AND THEIR UNITS OF MEASUREMENT.

Thermal Model	Unit of Measurement	Electrical Model	Unit of Measurement
Power, Heat Flow	P [W]	Current	I [A]
Temperature	T [K]	Voltage	U [V]
Resistance	R_{th} [K/W]	Resistance	R [V/A]
Capacity	C_{th} [J/K]	Capacity	C [As/V]
Time constant	τ [s]	Time constant	τ [s]

A. Dualism of thermal and electrical parameters

In order to model the heat flow within as well as between the different materials of an integrated system, a definition of the following three components is necessary:

- Heat sources
- Thermal resistors
- Thermal capacitances

Heat sources are all components of a circuit where electrical current flows and thus where electrical energy is converted into thermal energy. All materials feature a thermal resistance, which interferes with the spreading of heat. Components store thermal energy depending on their materials. This behavior is described by their thermal capacitance.

The approach to model temperature, described in this paper, is based on the well known dualism between heat flow and the behavior of electrical circuits [6]. First, the duality of thermal and electrical resistance shall be described whereas both parameters depend on the physical properties of the underlying component. As depicted in Figure 1, there are two components of length l and the cross section of the area A. These two parameters describe the geometry of the component. Beyond that, the electrical and thermal properties depend on the material of the component. The corresponding parameters are the specific electrical conductance σ and the specific thermal conductance λ, respectively. Equation 1 specifies how these parameters determine the electrical and thermal resistance (i.e. R_Ω and R_{th}):

$$R_{th} = \frac{l}{\lambda \cdot A} \quad \Leftrightarrow \quad R_\Omega = \frac{l}{\sigma \cdot A} \qquad (1)$$

Based on these physical correlations an electrical current, respectively a heat flow, is described flowing orthogonally through the cross section A and along the length l of the component. This current I results in a voltage drop U across the component, which is equivalent with a difference in temperatures ΔT in the thermal model (see Figure 1). The analytical relationships are described by equation 2:

$$\Delta T = R_{th} \cdot P_v \quad \Leftrightarrow \quad U = R_\Omega \cdot I \qquad (2)$$

Similar to Ohm's law for the electrical current, there exists this corresponding equation for thermal energy describing relations in terms of heat flow P_v and differences concerning temperature. As the equations indicate, the basic relations between heat flow and electrical current are identical. The correlations of the individual parameters with their units of measurement are summarized in table 1. However, the dualism holds also

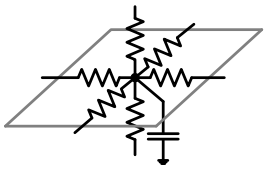


Figure 2. Structure of a basic RC-tile

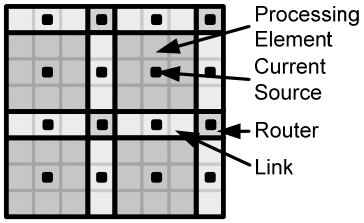


Figure 3. A 2x2 NoC mapped on a grid of 8x8 RC-tiles

true for the combination of several thermal resistances and capacitances. Thus, thermal components in serial or parallel connection can be transformed to one equivalent component depending on the same regularities as for electrical circuits.

B. Flow of steps for temperature simulations

The fundamental principle of the supposed methodology is to divide the NoC into a regular grid of basic RC-elements. These RC-elements model the thermal behavior of a certain area of the NoC. Hence, the resolution of the temperature distribution across the NoC can be adapted by the number of these RC-tiles, respectively by their size. Before describing the flow of the thermal simulation in more detail, the basic structure of an RC-tile is depicted in Figure 2. It consists of a capacitance to model the transient behavior of heat flow. Moreover, there are up to six resistors whereas the single upper resistor models the connection of the chip to cooling equipment like the heat spreader. The single lower resistor models heat transfer to a printed circuit board or the chip carrier as well. The remaining four resistors in the plane form connections to the neighboring tiles. This basic RC-element models the passive part of the NoC that only functions as a thermal conductor depending on the material it consists of. Moreover, there also exists the possibility to connect a current source to the central node of this basic element in order to model a heat source, which is still to be discussed.

Based on the underlying RC-grid, a simulation flow was developed (see Figure 4). The main steps are the following:

- Mapping the NoC components and their power consumption to the RC-grid
- Simulate the generated netlist
- Generate a graphical representation of temperature distribution based on the voltages from the simulation

At the beginning of a simulation run, the RC-grid must be generated. The number of tiles depends on the desired accuracy (1). Following (2), the NoC components are automatically mapped onto this homogeneous structure of tiles. Therefore, only geometric information is needed, which defines the dimensions of the basic components like routers, links and processing elements as well as the dimensions of the whole NoC. Based on this information, it is determined to which kind of NoC component a single tile belongs. The parameters for the resistors and the capacitance of the tiles are set according to the thermal properties of the material of the NoC component. Subsequently (3), the heat sources have to be mapped according to the actual current sources and are attached to the network of RC-tiles. Processing elements, routers and links are assumed to be sources of thermal energy. Basically, each RC-tile can be extended by a current source in order to inject heat generated

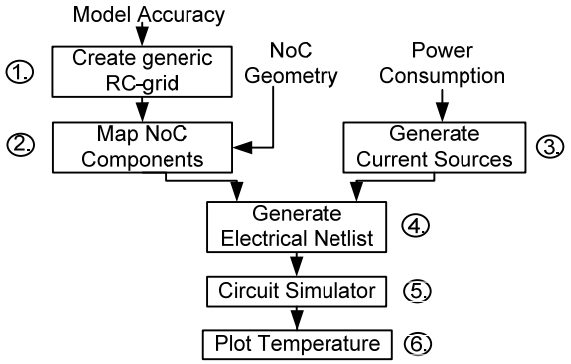


Figure 4. Flow of simulating temperature distribution based on the geometry and activity of the NoC components.

by an underlying component. As a result, the thermal characteristics of complex modules like CPUs can be modeled by adding several current sources of various strengths to the RC-tiles corresponding to one NoC component. The currents of the miscellaneous current sources depend on the power consumption of the modeled NoC component and the power distribution within each component. Based on the described mapping of resistors, capacitors and current sources, a netlist is generated (4). This netlist implements the described structures including the adapted values of the electrical elements (resistance and capacitance as well as current and timing of the current sources). Following, the netlist is evaluated using a circuit simulator like SPICE (5). As a result, voltage values of all nodes in the netlist are generated. These voltages are interpreted as temperature. These temperatures are used as input for further algorithms like mapping of tasks or evaluation of control algorithms. Furthermore, a last step is applied to enable researchers analyzing the resulting thermal distribution. As shown in phase (6) of Figure 4, the temperature values are used to generate a graphical representation.

Following, two important parameters of the simulation flow shall be discussed. First, the total number of RC-tiles determines the spatial resolution. Simulation time heavily depends on this parameter since the number of electrical components and nodes in the netlist is directly proportional to it. Second, there also exists a temporal resolution. If control algorithms, which react on temperature, shall be evaluated this resolution will be of importance because of the dynamic nature of such an application scenario. These two kinds of resolutions can be adapted to any needs. However, a trade-off between resolution and simulation speed has to be made.

III. APPLICATION SCENARIO

The following section exemplifies the proposed methodology for a specific application scenario. Before that, a simple example is briefly discussed for clarification. In Figure 3, a possible mapping of a 2x2 NoC on a coarse grid is shown. For convenient handling, a router is assumed to be quadratic and mapped on one single tile. As it is shown there, all processing elements are mapped to nine tiles each. Thus, in this example the dimensions of a link directly depend on the width of processing elements and routers –i.e. it is mapped to three tiles. However, the NoC components do not exactly have to fit into the grid because it is also possible to map two different components onto one RC-tile. Thus, the electrical parameters have to

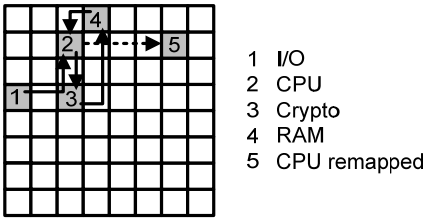


Figure 6. Mapping of processing elements and their associated communication paths within the application scenario.

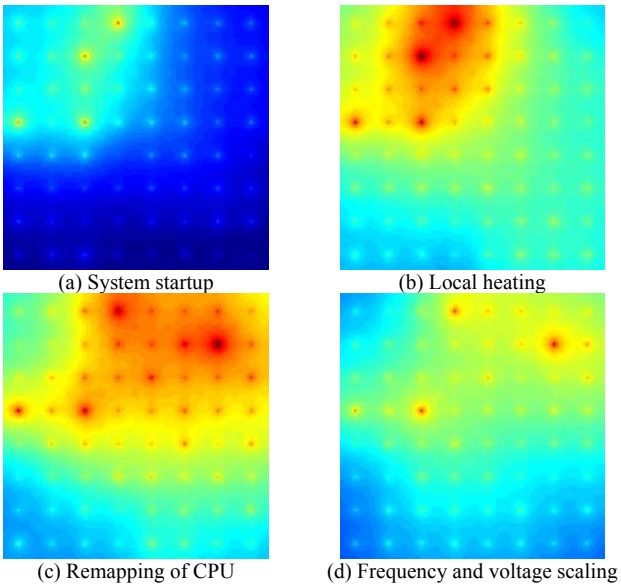
be adjusted respectively. In order to model the injection of thermal energy, each NoC component injects its thermal power using one current source connected to its central RC-tile.

This being said, an application scenario exemplifies the benefits of the developed methodology. Given is an 8x8 NoC. For clarification of temperature distributions, a simple application with only four main resources is mapped onto the NoC. Moreover, a slight uniform traffic pattern is applied to all other processing elements. Figure 5 depicts the main communication paths and the mapping within the NoC. There are an input-output-module (I/O) from which a central processing unit (CPU) gets a stream of data that has to be processed. To speed up the processing of the encrypted data stream, the CPU sends data to another functional module accelerating cryptographic (crypto) calculations. This co-processor transmits its results to the main memory (RAM). From there, the CPU retrieves the decrypted data. Communication relationships between the four components are depicted by arrows in Figure 5. Depending on the activity and therefore on the power consumption of each of the four components, various temperature distributions evolve. The resulting distributions shown in Figure 6 were obtained by using the proposed methodology (see section II B). The NoC components' activities were determined running functional simulations of this application scenario.

The first image shows the system during startup. Activities and thus temperatures are increasing (Figure 6 a). The continuous operation of the application results in a fortified local heating (Figure 6 b). Hence, the system management decides to shift CPU functionality to another NoC element to avoid exceeding a given temperature limit. Following, a new temperature distribution evolves (Figure 6 c). The remapping reduces the maximum on-chip temperature and leads to a more balanced temperature across the chip. If the temperature is still too high the dynamic frequency scaling in combination with dynamic voltage scaling can be applied, which further reduces temperature. The effect of a combined frequency reduction of the four active processing elements is that the temperature across the whole system is significantly decreased (see Figure 6 d). Even though this rough reaction decreases the overall system performance, it still guarantees system functionality and might preserve the system from total, permanent failure.

IV. Conclusion

This paper presented a methodology, which enables designers to conveniently model the temperature distribution in NoC. Thereto, advantage is taken of the analogy between electrical and thermal energy flows. The proposed methodology gives the designer a flexible, generic temperature model. In addition, the paper exemplified the use of the approach for a specific application scenario. Concluding, the presented methodology im-



(a) System startup (b) Local heating

(c) Remapping of CPU (d) Frequency and voltage scaling

Figure 5. Visualization of temperature distribution of an 8x8 NoC without or with dynamic frequency scaling and the remapping of one task.

plies several starting points for ongoing research and future improvements. The main focus will lie on investigations concerning the adaption of the electrical components modeling the accurate thermal behavior of the NoC and the resulting accuracy of this approach. Subsequently, temperatures obtained from an adapted implementation have to be verified against known results. Investigations concerning the speed of this modeling approach depending on its resolution will also have to be made. Finally, the integration into existing NoC simulators based on VHDL or SystemC is aspired. This could be accomplished using the AMS extensions of the specific HDLs.

References

[1] Benini, L.; De Micheli, G.; "Networks on chips: a new SoC paradigm," *Computer*, vol.35, no.1, pp.70-78, Jan 2002

[2] Brooks, D., "Power, Thermal, and Reliability Modeling in Nanometer-Scale Microprocessors," *Micro, IEEE* , vol.27, no.3, pp.49-62, 2007

[3] Dhodapkar, A, et al."TEM2P2EST: A Thermal Enabled Multi-model Power/Performance ESTimator" *Proc. first international Workshop on Power-Aware Computer Systems,* 2008

[4] Hamann, H.F.; et al., "Spatially-resolved imaging of microprocessor power (SIMP): hotspots in microprocessors," *ITHERM, 10th Conf,*. 2006

[5] Kim, N.S.; et al.; "Leakage Current: Moore's Law Meets Static Power", *Computer,* vol.36, no.12, 2003

[6] Krum, A., "Thermal management", Frank Kreith, editor, The CRC handbook of thermal engineering, CRC Press, Boca Raton, FL, 2000

[7] Lienig, J., "Introduction to Electromigratino-Aware Physical Design", Proceedings of ISPD'06, 2006

[8] Roy, K.; et al., "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proc. IEEE* , vol.91, no.2, 2003

[9] Shang, L.; et al., "Thermal Modeling, Characterization and Management of On-Chip Networks" *Proc. 37th IEEE/ACM Symp. on Microarchitecture*, 2004

[10] Skadron, K.; Abdelzaher, T.; Stan, M.R., "Control-theoretic techniques and thermal-RC modeling for accurate and localized dynamic thermal management," *High-Perf. Comp. Arch., Proc. 8th Symp.*,Feb. 2002

[11] Srinivasan, J.; et al., "The case for lifetime reliability-aware microprocessors,"*Comp. Arch., Proc. 31st Symp. Comp. Arch.*,2004

[12] Srinivasan, J.;et al.; "RAMP: A Model For Reliability Aware MircoProcessor Design", IBM Research Report, RC23048, 2003

[13] Yonghong Yang; et al.; "Adaptive Chip-Package Thermal Analysis for Synthesis and Design," *DATE '06. Proceedings*, March 2006