PRATHAM: A Power Delivery-Aware and Thermal-Aware Mapping Framework for Parallel Embedded Applications on 3D MPSoCs

Nishit Kapadia, Sudeep Pasricha
Department of Electrical and Computer Engineering
Colorado State University, Fort Collins, CO, U.S.A.
nkapadia@colostate.edu, sudeep@colostate.edu

ABSTRACT

In emerging 3D-ICs, thermal hotspots and high IR-drops in the power delivery network (PDN) can significantly limit overall system performance. The high core counts required to support parallel embedded applications in these 3D-ICs also notably increases communication energy. As inter-core communication patterns, IR-drop distributions, and 3D thermal profiles all influence system performance and power, it is critical that a system-level application mapping framework consider their combined effect. In this paper, for the first time, we propose a system-level parallel embedded application allocation framework (PRATHAM) that integrates a holistic solution evaluation methodology while considering the impact of network-on-chip (NoC) communication, drops in supply voltage, and the on-chip thermal profile on overall system performance, and co-optimizes on-chip IR-drop, thermal, and communication profiles, for improved overall system performance and lower energy consumption. Our experimental results indicate that PRATHAM reduces energy-delay-squared product (ED^2P) by up to 43.6% over recently proposed PDN-aware and thermal-aware 3D-mapping frameworks, respectively.

1. INTRODUCTION

In emerging 3D multiprocessor system-on-chip (MPSoC) architectures, network-on-chip (NoC) fabrics enable tens to hundreds of cores to communicate with each other and memory at the intra- and inter-layer levels. As the power dissipated in the NoC is a significant portion of the total on-chip power in these designs, optimizing communication power, in addition to computation power, has become a critical step in design methodologies for 3D MPSoCs.

Yet another critical component in 3D MPSoCs is the power delivery network (PDN), which is required to deliver a stable power supply across the chip that is within a desired voltage range and can tolerate large variations in load currents [1]. However, with increasing on-chip device densities and decreasing voltage levels, supply currents have risen but the scaling of PDN impedance has not kept up with this trend [2]. This has led to worsening IR-drops in the PDN. As circuit delay in modern CMOS technologies has a super-linear relationship to the supply voltage drop [3], high IR-drops in the PDN have increased circuit delay, limiting operating frequencies of cores and thus reducing MPSoC performance. This problem is even more severe in 3D MPSoCs, as the number of I/O pins on an *n*-layered 3D-IC is about *n* times smaller than its 2D counter-part, thus exacerbating the problem of a degraded voltage supply in 3D MPSoCs [4].

With much higher power densities in high-performance 3D MPSoCs, another challenge is to remove the heat generated in the 3D stacked dies. In addition, the circuit delay is also strongly related to temperature. Increasing temperatures can limit the operating frequency on the chip, thereby degrading system performance [5], [6]. It is also commonly known that leakage power has a super-linear relationship with temperature and that temperature in turn depends on the power profile of the chip.

As a result of these close interactions between the PDN, die temperature, circuit performance, and power dissipation, performing thermal analysis (or PDN analysis) in isolation, without considering the influence of temperature (or supply voltage) on power and performance could lead to a grossly inaccurate estimation of system metrics. Prior works [7]-[9] have motivated a joint thermal-power-performance-PDN analysis to capture the inter-dependencies between the various design metrics. However, to the authors' knowledge, none of these prior works have considered such co-analysis for accurately evaluating the merit of any given design solution as part of an MPSoC design-time synthesis framework for 2D or 3D ICs.

This work is motivated by the interesting insight that inter-core communication patterns, IR-drop distribution in the PDN, as well as the on-chip thermal profile can vary significantly with different mapping configurations of computation and communication resources on an MPSoC. Even though, prior works [10], [21], take into consideration the PDN costs (such as IR-drops, number of voltage regulator modules required, or external supply current drawn) in their NoC synthesis frameworks, they fail to account for the effects of IR-drops on power and performance of the system. The novel contributions of our framework are summarized as follows:-

- We propose a design-time CAD/synthesis approach to map cores running parallel embedded applications to the tiles on a 3D die while co-optimizing communication, PDN design, and thermal objectives to generate an optimized 3D MPSoC design;
- We integrate the effects of temperature and supply voltage drops in the performance and energy modeling of the compute cores as well as the communication resources; and show that considering such awareness in the solution evaluation framework results in significantly improved design solutions.

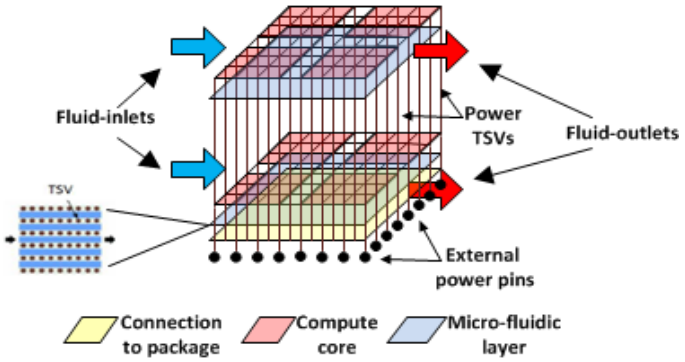

Figure 1. Example of a 3D package for 8-core MPSoC (2×2×2 3D-mesh) with: *(i)* a regular 3D power-grid with 64 external power-pins and 64 grid-points per tier (16 grid-points supplying to each core); *(ii)* a micro-fluidic cooling layer under each device layer

2. PROBLEM FORMULATION

In this work, we consider a 3D MPSoC platform with a 3D mesh of tiles, where there is a one-to-one mapping of cores to these tiles. We assume a 3D package with cores, PDN, and micro-fluidic cooling, as shown in figure 1. We assume the following inputs to our problem:

- A 3D-IC with a regular 3D mesh NoC, with dimensions (dim_x, dim_y, dim_z) and number of tiles $T = dim_x \times dim_y \times dim_z$ with each tile containing a compute core and a NoC router;
- A core graph $G(T,E)$ (corresponding to the given set of parallel embedded applications) with a set of T vertices representing homogenous compute cores on which application tasks have already been mapped, and a set of E edges that represent communication volumes between communicating cores;
- A triplet $\{V_{nom}, f_{nom}, t_{nom}\}$ constituting the nominal values of supply voltage, operating frequency, and temperature, representing nominal operating condition for all T cores;
- A set of nominal supply current values $\{I_{1\_nom}, I_{2\_nom},\ldots,I_{T\_nom}\}$ for the T cores corresponding to nominal operating conditions;
- A set of T nominal computation-times $\{CT_{1\_nom}, CT_{2\_nom}, \ldots, CT_{T\_nom}\}$ corresponding to workloads executed on each core;
- A regular 3D power grid, with $n \times n$ grid-points supplying to each core, and a chip-wide maximum IR-drop constraint Γ in the PDN (relative to V_{nom});
- A fixed liquid flow-rate Ω in the micro-fluid cooling setup, with a chip-wide maximum temperature constraint Ψ.

Problem Objective: Given the above inputs, the goal of our proposed framework is to obtain a core-to-die mapping on the 3D-IC and a mapping of communication flows to a regular 3D mesh NoC for a given set of parallel embedded applications, such that energy-delay-squared product (ED^2P) of the system is minimized, while the design constraints Ψ and Γ are satisfied. In our ED^2P calculations, we define system-energy as (computation + communication) energy dissipation in the 3D die, and system-delay as the aggregate of computation and communication times for all cores. As performance is considered the principal design metric in most high-performance chips, we choose system ED^2P as the optimization metric, as motivated in prior work [20], with an upper bound on the peak temperature of the chip.

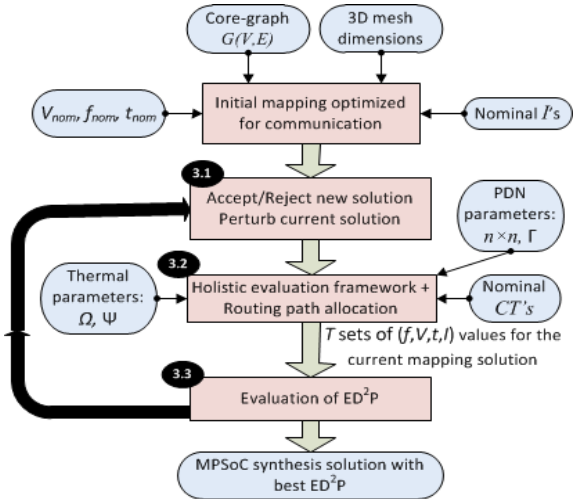

Figure 2. Design flow of the PRATHAM synthesis framework

3. PRATHAM FRAMEWORK: OVERVIEW

We first present a brief overview of our 3D MPSoC synthesis framework, before describing more details later in this section. The PRATHAM framework, shown in Figure 2, uses a simulated annealing (SA) based search algorithm (discussed in section 3.1), where the minimization objective is the system ED^2P. The initial mapping solution, which is optimized for communication traffic, is generated by an efficient core mapping approach based on [19]. A holistic solution evaluation stage and a routing path allocation heuristic are integrated within the SA-based search algorithm. The holistic evaluation stage computes the set of T quadruplets $\{(f_1,V_1,t_1,I_1), (f_2,V_2,t_2,I_2), \ldots, (f_T,V_T,t_T,I_T)\}$ for each new mapping solution (discussed in section 3.2). In this manner, computation times $(CT_1,CT_2,\ldots,CT_T)$ and computation energies of cores can be readily evaluated for any given mapping solution. Finally, the framework generates the best MPSoC synthesis solution with the least ED^2P.

3.1 SA-based Search Algorithm

Broadly speaking, the optimization objectives of our 3D MPSoC synthesis framework are three-fold: (*i*) minimize application communication latencies and energy (by mapping cores such that the high-volume communication flows have shorter path-lengths in the NoC); (*ii*) minimize the IR-drops in the PDN (by mapping cores with higher supply current requirements to tiles on the 3D mesh closer to the external input power pins), and (*iii*) facilitate more efficient cooling of the 3D chip (by mapping cores with higher power dissipation closer to the water inlet of the microfluidic-cooling system, while at the same time maintaining a reasonably uniform power profile across all the device layers).

The single minimization objective of our SA-based search, system ED^2P, is representative of the combination of the above objectives. In the SA search process, the solution is perturbed by swapping two arbitrarily chosen cores on the 3D mesh. In addition to the ED^2P value, the number of violations (at per-core granularity) of the PDN constraint Γ and the thermal constraint Ψ are also considered for the cost-evaluation of the current solution. The SA cost function used by the search is given below, where c is the violation-penalty coefficient with a high integer value.

$$cost = ED^2P + c \times \{num.\ of\ (PDN\text{-}violations + thermal\text{-}violations)\}$$

3.2 Holistic Solution Evaluation Framework

In this section, we first explain the various dependencies between different design metrics at the system-level. Later, we discuss details of our holistic evaluation framework, and the necessity of considering these inter-dependencies during solution evaluation.

It is well known that leakage power depends exponentially on temperature. At the same time, temperature directly depends on the power-profile of the system. Increasing temperatures contribute to significant circuit slow-down. Additionally, circuit delay is also related to the supply voltage. Supply voltage in turn depends on the IR-drop distribution in the PDN. Therefore, *the maximum frequency (f) that a core can be clocked at depends on both the supply voltage and the operating temperature of the core.* Moreover, the operating core frequency, along with the supply voltage, determines the total power dissipation $(V \times I)$, which in turn determines the supply current requirement of the given core. Finally, the supply currents flowing in the PDN determine the IR-drop distribution in the PDN. This inter-dependence among power, thermal, performance, and PDN metrics is depicted pictorially with a metric dependency graph in figure 3(a).

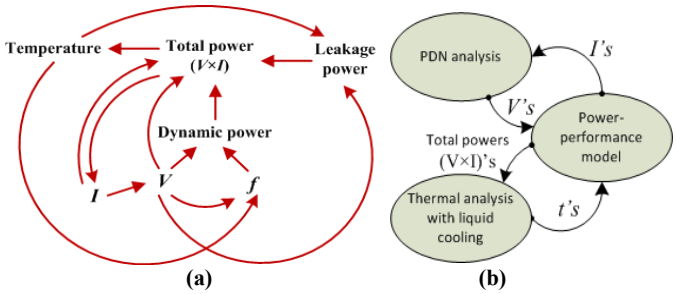

Figure 3. Holistic solution evaluation framework (a) metric dependency graph (b) control-flow of the evaluation framework

Our evaluation framework is applied to each new mapping candidate generated in the SA-based search. To capture the cyclic inter-dependencies among different design metrics as shown in figure 3(a), our holistic solution evaluation framework is executed iteratively until convergence to final values. Figure 3(b) shows a graphical representation (control-flow graph) of the three phases in our framework that are iteratively executed. The details of the three phases (*i*) Power-performance model (*ii*) Thermal-analysis (*iii*) PDN-analysis are discussed below:

Power-performance model: The impact of PDN IR-drops and chip thermal profile on performance and power dissipation is captured by the power-performance model. In each iteration of execution, updated values of core-temperatures (t's) and supply voltages (V's) are first utilized to update the core-frequencies (f's). The f's in turn are utilized (along with V's) to compute the dynamic power of cores. The t's and the V's are used to compute the leakage powers of cores. The individual sums of dynamic and leakage powers result in total powers of the T cores, which are finally used to update the supply currents (I's). The power-performance modeling is based on [9] (with parameters being technology-scaled based on [3] and [5]).

Thermal-analysis: To perform thermal evaluation of any given mapping solution in our framework, we utilize the open-source thermal emulator 3D-ICE 2.2.5 [13] which supports steady-state thermal analysis of 3D ICs with inter-layer liquid cooling. For the given power-profile, the tool outputs the core-temperatures (t's) on the 3D die, for a fixed micro-fluid flow-rate of Ω.

PDN-analysis: For any given candidate mapping solution (with the supply current requirements of the T cores on 3D mesh), we use a linear programming (LP) based formulation to solve for the grid-point voltages and currents flowing in the 3D regular power grid. We validated the LP formulation using HSPICE simulations and used the lp_solve solver [14] to obtain results. In our LP formulation, we assume a 2D grid of $n \times n$ grid-points supplying to each core on the 3D die (figure 1). This phase in the evaluation framework evaluates the IR-drops in the 3D resistive grid, to enable the updating of supply voltages of cores in the 3D MPSoC. Details of the design variables and constraints of our LP formulation are omitted for brevity.

Algorithm 1:Iterative holistic solution evaluation framework

Inputs: mapping-solution, nominal CT's, and Ω, Ψ, n×n, Γ

1: do {
2: PDN analysis computes V's for the current set of I's
3: Power-performance model updates f's, dynamic and leakage power values, and I's, for the updated set of V's
4: Thermal analysis computes the t's for the current set of power values
5: Power-performance model now updates f's, power values, and I's for the updated set of t's
6: exit do loop if *convergence-condition* achieved
7: go to step 2 **}**
8: update computation-times of cores based on the scaled (updated) f's
9: calculate core computation-energies using computation-times and power $(V \times I)$ values
10: output the *num. of violations* if Ψ or Γ constraints have been violated

outputs:{ computation-energies, computation-times, updated (f,V,t,I)'s of T cores} and number of violations

Algorithm 1 presents the pseudo-code for our holistic evaluation framework. The three phases of the framework: PDN analysis, Power-performance model, and Thermal analysis are iteratively executed **(steps 2-5)** until the assumed *convergence* condition is met. When all of core-temperatures (t's) change by less than 0.5°C in successive iterations of thermal analysis, we say that *convergence* is achieved **(step 6)**, and end the iterative evaluation process. We empirically chose the value of 0.5°C as a reasonable trade-off between accuracy and execution-time. Once temperature convergence is achieved, the power-profile can be assumed to have stabilized, and thus V's and I's of cores are also assumed to have converged to their final values.

One of the main contributions of this work is to highlight the significance of considering these dependencies during design-time 3D MPSoC synthesis. A holistic evaluation approach as discussed in this section is inherently more accurate, but more importantly, when integrated within the synthesis framework, it produces significantly improved design solutions by facilitating the solution search space to be explored more intelligently (corroborated with experimental results in section 4.2). The updated values of operating frequencies and voltages from this stage are utilized by our routing path allocation step, based on [10] (discussion omitted for brevity), which optimizes path latencies of communication flows, to produce a complete synthesis solution for the current mapping solution.

3.3 ED^{2}P Evaluation

Once all the communication flows have been routed, the aggregate communication power dissipation in the NoC is computed taking into consideration the link loads, router sizes, and corresponding voltage/frequency values. Note that as each core operates at its own frequency (each router operates at the same frequency/voltage as the core it is associated with), the communication link operates at the lesser frequency of the pair of routers it connects. The power and latency overheads of MCFIFO based frequency converters (that are required to ensure correctness when crossing frequency domains) are included in our analysis. Path latencies are computed for each flow and used to determine the total communication time for each core. We define the core-time of the i^{th} core with q number of outgoing communication-flows as: $core\text{-}time_i$ = $computation\text{-}time_i$ + $communication\text{-}time1_i$ + $communication\text{-}time2_i$ + ... + $communication\text{-}timeq_i$. The average of all *core-times* is termed as *system-delay*, which is finally used to evaluate the ED^{2}P of the synthesis solution: ED^{2}P = (communication + computation) energy × $(system\text{-}delay)^2$. After the completion of this step, a mapped and routed 3D MPSoC is obtained for the given set of applications.

4. EXPERIMENTAL STUDIES

4.1 Experimental Setup

In our experiments, ARM Cortex A-9 processors [16] are used as the baseline MPSoC compute cores, at 32-nm technology node. For the nominal operating condition, values of V_{nom} = 1.1V, f_{nom} = 1800MHz, and t_{nom} = 45°C are assumed. We assume leakage power to be 30% of total power under nominal operating conditions, based on our analysis at 32nm. We consider 64-core and 144-core 3D MPSoC platforms, with dimensions 4×4×4 and 6×6×4 ($dim_x \times dim_y \times dim_z$). The cores are inter-connected using a 3D mesh NoC topology. Under nominal operating conditions, the maximum supply current values for the 64-core MPSoC, are assumed to be between 1A and 4A (0.4A and 1.7A for 100-core) , based on the computation requirements of tasks assigned to the respective cores.

We use eight parallel application benchmarks from the SPLASH-2 and PARSEC benchmark suites [17], [18]. Our core-graphs are modeled based on inter-core communication characterizations given in [12]. We consider four combinations of compute-intensive and memory-intensive benchmarks implemented on 64-core and 144-core platforms for a total of eight workloads. We assume equal number of threads for all concurrently executing benchmarks. The eight different workloads considered are: wld-1A and wld-1B: combination of compute-intensive and memory-intensive PARSEC benchmarks {*vips+dedup+streamcluster+blackscholes*}; wld-2A and wld-2B: combination of compute-intensive and memory intensive SPLASH2 benchmarks *{fft+lu+cholesky+ radix};* wld-3A and wld-3B: combination of compute-intensive benchmarks *{streamcluster+ blackscholes+cholesky+radix};* wld-4A and wld-4B: combination of memory-intensive benchmarks {*vips+dedup+fft+lu*}. The workloads with suffix "A" are executed on the 64-core MPSoC platform and those with suffix "B" are executed on the 144-core MPSoC platform.

In our thermal analysis, a fixed micro-fluid flow-rate Ω=10ml/min (constant cooling power is assumed) and maximum temperature constraint Ψ=90°C is used. The incoming coolant (de-ionized water) temperature is assumed to be 25°C. Our regular 3D-PDN power grid is modeled based on the guidelines provided in [4]. For the 64-core platform, with 16 cores on each tier, a total of 256 input power pins are used with $n^2=16$ grid-points for each core. For the 144-core platform, with 36 cores on each tier, $n^2=9$ grid-points per core are used. Values of 40mΩ and 83mΩ are used for horizontal and vertical branch resistances, based on [4]; and max IR-drop constraint Γ=6.5%.

The nominal power values of NoC routers and links (32-bit wide) for different voltages, frequencies, and router complexities at varying communication loads for the 32nm node are obtained from ORION 2.0 [15]. To integrate the effects of temperature and IR-drops in NoC routers, the leakage power of each router is scaled in proportion to the leakage power of the corresponding compute-core connected to it.

4.2 Experimental Results

To the authors' knowledge, [10] is the only prior work that has proposed a design-time MPSoC synthesis framework to co-optimize PDN costs and on-chip communication costs. Therefore, to evaluate the quality of solutions generated by our MPSoC synthesis framework (PRATHAM), we compare our results with the synthesis framework from [10]. Although, [10] integrates PDN-awareness (in terms of worst case IR-drop constraints) in the synthesis framework, it fails to account for the effects of IR-drops on power and performance of the system. Additionally, [10] does not consider thermal-awareness while evaluating the cost of any given MPSoC solution. PRATHAM overcomes the above mentioned drawbacks by integrating the evaluation of the combined impact of IR-drops and temperature on the system power/performance profile into our holistic evaluation methodology. We also compare our framework with a thermal and communication-aware mapping framework for 3D mesh-based MPSoCs, proposed in [11]. The framework in [11] utilizes an incremental mapping heuristic to minimize both the peak temperature and the communication traffic in the 3D MPSoC, but it fails to consider the effects of IR-drops in the PDN.

Table-I. Total number of (thermal/IR-drop) violations obtained from using [11], [10] and PRATHAM.

Num. of violations	wld-1A	wld-2A	wld-3A	wld-4A	wld-1B	wld-2B	wld-3B	wld-4B
[11]	0/9	0/10	0/9	0/12	0/19	4/24	0/26	0/17
[10]	1/0	7/0	5/0	8/0	0/0	16/0	10/0	1/0
PRATHAM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

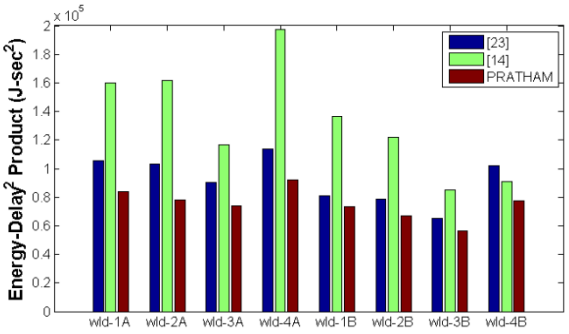

Figure 4. Improvements in system-ED^{2}P with PRATHAM over [11] and [10] for eight multi-application workloads

PRATHAM intelligently explores the overall solution search-space, co-optimizing communication, thermal profile, and IR-drop distribution in the PDN, while satisfying all PDN and thermal constraints, and thus: *(i)* produces average improvement of 7.8% and 2.7% in leakage energy (7.5% and 2.1% in terms of total energy) over [10] and [11]; and *(ii)* produces the highest average core-frequencies and the lowest network traffic; resulting in significant savings in system-delay of 21.4% and 8.7% over [10] and [11] respectively. Frameworks focusing primarily on thermal-awareness or PDN-awareness during 3D MPSoC synthesis produce infeasible solutions with IR-drop violations or thermal violations, respectively (Table-I). PRATHAM integrates thermal- and PDN-awareness to generate solutions with better overall optimality. As shown in Figure 4, the system-ED^{2}P is reduced by 18.5% and 43.6% on average, over [11] and [10] respectively. PRATHAM derives its improvements by combining the communication and computation domains of optimization within a single solution search space, while more accurately modeling the complex set of inter-dependencies among various system metrics.

5. CONCLUSION

PRATHAM represents one of the first efforts to tightly integrate PDN design, thermal, power, and performance objectives as part of a design-time exploration effort for 3D MPSoCs. Our experimental results indicate that PRATHAM improves system-ED^{2}P by up to **43.6%** over prior work. Solutions generated by PRATHAM are also more amenable to efficient physical design because of considering PDN design issues and the impact of IR-drops and temperature on system power and performance early in the design flow.

ACKNOWLEDGMENTS

This research is supported by grants from SRC, NSF (CCF-1252500, CCF-1302693), and AFOSR (FA9550-13-1-0110).

REFERENCES

[1] B. Amelifard, M. Pedram, "Optimal design of the power-delivery network for multiple voltage-island system-on-chips," IEEE TCAD 28(6), pp. 888-900, May 2009.

[2] P. Jain, T. Kim, J. Keane, C. Kim, "A multi-story power delivery technique for 3D integrated circuits," ISLPED pp. 57-62, Aug. 2008.

[3] T. Okumura et al., "Gate delay estimation in STA under dynamic power supply noise," ASP-DAC, pp. 775-780, Jan. 2010.

[4] N. Khan et al., "Power delivery design for 3-D ICs using different through-silicon via (TSV) technologies," IEEE TVLSI 19(4), Apr. 2011.

[5] S. Ganapathy et al., "Circuit propagation delay estimation through multivariate regression-based modeling under spatio-temporal variability," DATE, pp. 417-422, 2010.

[6] A.T. Winther et al., "Temperature dependent wire delay estimation in floorplanning," NORCHIP 2011.

[7] H. Su et al., "Full chip leakage-estimation considering power supply and temperature variations," ISLPED 2003, pp. 78-83.

[8] M. Pedram et al., "Thermal Modeling, Analysis, and Management in VLSI Circuits: Principles and Methods," Proc. IEEE, 94(8), 2008.

[9] W. Liao et al., "Temperature and supply voltage aware performance and power modeling at microarchitecture level," IEEE TCAD, 24(7), 2005.

[10] N. Kapadia, S. Pasricha, "A co-synthesis methodology for power delivery and data interconnection networks in 3D ICs," ISQED, 2013.

[11] M. Arjomand, et al., "Voltage-frequency planning for thermal-aware, low-power design of regular 3-D NoCs," VLSID, pp. 57-62, Jan. 2010.

[12] N. Barrow-Williams, C. Fensch, S. Moore, "A communication characterization of Splash-2 and Parsec," IEEE IISWC, Oct. 2009.

[13] 3D-ICE open-source tool: http://esl.epfl.ch/3d-ice.html

[14] lp_solve 5.5.2.0, http://lpsolve.sourceforge.net/5.5/

[15] A. Kahng et al., "ORION 2.0: A fast and accurate NoC power and area model for early-stage design space exploration," DATE, Apr. 2009.

[16] ARM Cortex-A9, http://www.arm.com/products/processors/selector.php

[17] S.V. Woo et al., "The SPLASH-2 programs: characterization and methodological characterization," ISCA, pp. 24-36, May 1995.

[18] C. Bienia, S. Kumar, J.P. Singh, K. Li, "The PARSEC benchmark suite: characterization and architectural implications," PACT, Oct. 2008.

[19] N. Kapadia, S. Pasricha, "VISION: A framework for voltage island aware synthesis of interconnection networks-on-chip," GLSVLSI, 2011.

[20] A. J. Martin, "Towards an energy complexity of computation," Information Processing Letters 77(2-4), pp. 181-187, Feb. 2001.

[21] N. Kapadia, S. Pasricha, "A Power Delivery Network Aware Framework for Synthesis of 3D Networks-on-Chip with Multiple Voltage Islands," Proc. VLSID, pp. 262-267, Jan 2012.