System Level Methodology for Interconnect Aware and Temperature Constrained Power Management of 3-D MP-SOCs

Sumeet S. Kumar, *Student Member, IEEE*, Arnica Aggarwal, Radhika Sanjeev Jagtap, Amir Zjajo, *Member, IEEE*, and Rene van Leuken, *Member, IEEE*

***Abstract*— Modern 3-D multiprocessor systems-on-chip (MP-SoC) incorporate processing elements (PEs) and memories within die-stacks interconnected using through-silicon vias (TSVs). The resulting power density of these systems necessitates the inclusion of thermal effects in the architecture space exploration stage of the design process. The number and placement of TSVs influences the thermal conductivity in the vertical direction in die-stacks, and consequently these must be considered during thermal analysis. However, the special requirement of keep out zones (KOZs) for TSVs due to mechanical stress considerations complicates the design of the vertical interconnect, potentially impacting its electrical performance as well. This paper presents an integrated methodology that allows for TSV topology exploration to evaluate the best vertical interconnect structure while considering crosstalk, area overheads, and KOZ requirements using an initial system floorplan. After incorporating feedback from the exploration, the resulting vertical interconnect is included within a temperature-power simulation that estimates the thermal profile of the 3-D stack. Within this methodology, a novel power management scheme for 3-D MP-SoCs that considers both temperature as well as positional information and thermal relationships between PEs, while performing dynamic voltage-frequency scaling (DVFS), is introduced. The scheme effectively maintains smooth temperature profiles, decreases fluctuations in voltage-frequency levels, and increases the aggregate frequency of operation at a lower total power dissipation. Further, the scheme is applied to a stack partitioned into voltage islands, where it is shown to match the conventional per-core DVFS schemes in its performance.**

***Index Terms*— 3-D integrated circuits, design methodology, system-level design, thermal management, through-silicon vias (TSVs).**

Manuscript received July 13, 2012; revised January 29, 2013 and June 12, 2013; accepted June 19, 2013. This work was supported by the CATRENE programme's Computing Fabric for High Performance Computing Project CA104.

S. S. Kumar, A. Zjajo, and R. van Leuken are with the Circuits and Systems Group, Delft University of Technology, Delft 2628CD, The Netherlands (e-mail: s.s.kumar@tudelft.nl; amir.zjajo@ieee.org; t.g.r.m.vanleuken@tudelft.nl).

A. Aggarwal was with the Circuits and Systems Group, Delft University of Technology, Delft 2628CD, The Netherlands. She is now with ASML Holding N.V., Veldhoven 5504DR, The Netherlands (e-mail: arnica.aggarwal@asml.com).

R. S. Jagtap was with the Circuits and Systems Group, Delft University of Technology, Delft 2628CD, The Netherlands. She is now with ARM Holdings, Cambridge CB1 9NJ, U.K. (e-mail: radhika.jagtap@arm.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2013.2273003

I. Introduction

THE progression toward smaller technology nodes has enabled an increase in integration density of modern silicon dies. The reduction in feature sizes has also exposed issues, such as process variation, leakage power consumption, and the limitations of interconnect performance [1]. 3-D integration is an emerging solution that targets these challenges through die-stacking and the use of through-silicon via (TSV)-based vertical interconnects. In the context of multiprocessor systems-on-chip (MP-SoC), die-stacking improves the system scalability by allowing the integration of a larger number of processing elements (PEs), without the associated increase in the chip's overall area footprint. The increased integration density, however, exposes multiple design challenges on account of the incorporation of logic, memory and the TSV-based vertical interconnect within the same die-stack [2], [3]. The design of the vertical interconnect, for instance, is complicated by the keep out zone (KOZ) requirement, which serves to insulate circuit elements from the mechanical stress induced by the thermal expansion and contraction of TSVs. The choice of KOZ also determines the area, the electrical noise, and the delay characteristics of the vertical interconnect. It is essential that these parameters and their effects be considered during early 3-D architecture space exploration to yield a vertical interconnect design that achieves the desired electrical performance, within the available silicon area.

State-of-the-art high-performance MP-SoCs contain a large number of general and special purpose PEs that significantly increase the power density when integrated within a single die-stack. As a consequence, thermal issues are observed especially in the lower tiers of the die-stack [4]– [6]. The vertical interconnect structure reduces the magnitude of these issues to some extent as it improves the number of heat transfer paths in the stack [7], and thus the thermal conductance to the higher tiers. During conventional architecture space exploration, PEs, and memory blocks are placed based on the simulation results at locations that yield the best system performance. However, such a technology-oblivious approach may aggravate thermal issues and inadvertently reduce system performance in 3-D stacked designs. Hence, initial system floorplans must be evaluated in terms of their thermal performance alongside conventional

system performance during a 3-D architecture space exploration.

While the thermal performance analysis provides critical feedback on the floorplan of the system, variations in the behavior of different target applications may necessitate multiple iterations of the analysis. Even so, an optimal solution satisfying all applications may remain elusive. In such cases a runtime power management scheme provides the degree of adaptability required to maintain thermal performance even with dynamic application behavior.

This paper addresses these design challenges through a system-level methodology that enables architecture space exploration based on the performance and cost of vertical interconnect structures, and the thermal behavior of die-stacks. In addition, it presents a runtime power management scheme to maintain the thermal performance of such stacks despite variations in workload behavior.

This paper is organized as follows. Section II provides a high-level overview of the proposed conceptual flow. Section III elaborates on the TSV model and method used for the topology exploration. Section IV focuses on the thermal modeling, temperature constrained power management scheme and temperature-power simulation method. Section V describes the experimental setup using a 3-D MP-SoC, and presents results from TSV topology exploration. It also introduces the temperature-constrained power management scheme as well as the temperature-power simulation methodology. Finally, Section VI lists the conclusions of this paper.

II. Overview of Methodology

A number of studies have investigated the challenges of stacked-die architectures, and have attempted to address the need for an analysis and exploration methodology for 3-D designs. MEVA3-D [8] is one such exploration tool that enables automated floorplanning, routing, placement of TSVs, and the thermal profiling of stacked-die architectures. A demonstration of the tool applied to the 3-D design of a conventional processor core illustrated the performance benefits of stacked-die architectures and their associated thermal issues. However, this paper does not describe the planning of the 3-D TSV network, nor the KOZ considerations considered in their placement. In addition, while MEVA3-D includes support for thermal via insertion, it does not support the use of a runtime power manager alongside the performance simulation. Although thermal vias can reduce the severity of thermal issues at tiers far from the heatsink, a runtime power management strategy can suitably manage the temperature profile of the stack, thereby reducing the number of such vias required. Inclusion of the power management strategy during the analysis of thermal performance is therefore critical toward preventing the insertion of vias where they are not necessary. Hung *et al.* [9] presented a thermal-aware floorplanner for the 3-D stacked processor cores that considers the power dissipation of the interconnect during floorplan exploration. Despite its merits, this methodology too does not describe KOZ considerations, nor the placement of TSVs within the floorplan. In addition, it does not include support for a runtime power manager in its analysis. An example of how such exploration tools benefit application performance in MP systems can be found in [10]. In this paper, Ozturk *et al.* investigated the optimal topology for an application specific 3-D MP, in terms of placement options for PEs and memory blocks. Through an exploratory simulation, multiple topologies are evaluated in terms of their average data access cost, and whether the consequent temperature of logic blocks remains within the imposed design constraints. Based on this, an optimal topology is found for the 3-D MP.

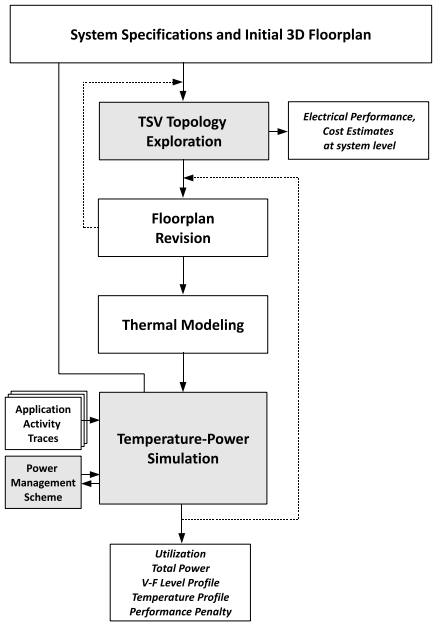


Fig. 1. Conceptual flow of the integrated methodology incorporating the TSV topology exploration, thermal modeling, temperature-power simulation, and the runtime power management scheme.

We present a system-level methodology that incorporates both vertical interconnect exploration and thermal performance analysis in a single SystemC flow along with a runtime power management scheme to enable the 3-D architecture space exploration. The conceptual flow is shown in Fig. 1.

Vertical interconnects may contain TSVs arranged in several topologies. For instance, they may be organized as bundles, or be placed along the boundaries of the vertical interconnect area. Each topology exhibits a different electrical performance and a distinct area penalty. Thus, the first step of the flow consists of a novel methodology to explore TSV placement topologies for multitier die-stacks. Topologies are analyzed on the basis of their electrical performance and area penalty using parameterized TSV models according to the system specifications and the initial floorplan [11]. The results from this exploration allow the initial floorplan to be revised to incorporate the TSV topology found superior in terms of electrical performance and cost, and better achieve target specifications.

The revised floorplan may differ from the initial in several ways, especially in the number of TSVs that constitute the vertical link on each tier of the die-stack. Since the TSVs essentially act as vertical heat transfer paths within the die-

stack, a significantly different thermal conductance can be expected when compared with the initial floorplan. These characteristics of the vertical interconnect are considered in the thermal modeling stage in which a mesh of thermal cells is generated for each device tier to determine its thermal relationship with others in the stack. The resulting thermal model provides a comprehensive set of effective thermal relationships between blocks in the 3-D floorplan.

The final stage of the flow is a temperature-power simulation that incorporates a thermal simulator using the model from the previous step, as well as a power estimation function that computes the power dissipation of logic blocks based on the initial system specifications and their activity rate. Based on this, the temperature-power simulator determines the effective thermal profile for the 3-D stack [12]. In the case of MP-SoCs, the activity rate is replaced by a cycle-accurate trace of each PEs execution, indicating the cycles during which computational operations were performed, and those during which it remained idle.

The voltage and frequency levels of PEs are controlled by a custom power management scheme that enables the investigation of the thermal implications of various power management techniques on 3-D stacks. Based on the analysis of a conventional dynamic voltage-frequency scaling (DVFS) technique, a novel temperature-constrained power management scheme is presented that controls the voltage and frequency levels of PEs based on their temperature and physical position in the stack, as well as the thermal model of the die-stack. This scheme is described in Section IV-C.

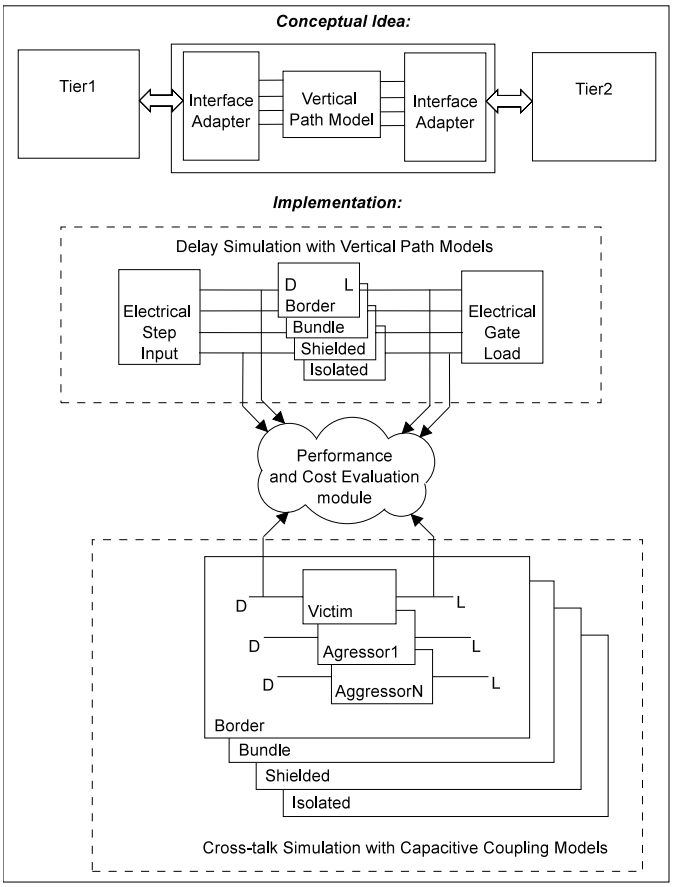


Fig. 2. TSV topology exploration (above) conceptual idea and (below) implementation. Note the concurrent simulation of the four topologies used in the exploration: border, bundle, shielded, and isolated.

III. TSV Topology Exploration

A. Background and Related Work

In [13] and [14], 3-D architectures are shown to have high performance, assuming a negligible delay through the TSV bundle without considering the signal integrity aspect of vertical interconnects. Recent papers on the characterization of TSVs [15], [16], however, indicate that the noise induced in signal TSVs due to capacitive coupling through the substrate is not negligible. In [15], Liu *et al.* performed a TSV-to-TSV noise analysis for a 45-nm two-tier 3-D integrated circuits (ICs) which shows a significant impact on full chip timing performance and total noise voltage. They present two effective ways to mitigate noise, namely, buffer insertion and shielding TSVs, both of which result in an appreciable increase in area penalty. Significant research in the area of EDA tools for 3-D physical design has led to the development of 3-D floorplanning, placement and routing algorithms that consider TSV technology constraints and thermal effects in developing an optimized design, although in the final stages of the design flow [8], [17]. In this paper, we present a system-level methodology that examines the physical effects like capacitive coupling of a TSV-based vertical interconnect and allows them to dictate aspects of the 3-D architecture to be employed. Additionally, the TSV KOZ [18] which is a critical technology constraint is also considered.

B. Exploration Method

The conceptual idea of the simulation of a two-tier stack with a vertical interconnect is shown in Fig. 2. Here, functional blocks on two tiers communicate through a vertical interconnect. This comprises of interfacing adapters and parameterized driver-to-load (D to L) vertical path models. We do not perform a functional simulation of blocks during TSV topology exploration, but only an electrical simulation using the vertical path model. As shown in Fig. 2, the D to L path model is instantiated multiple times to simulate several TSV topologies simultaneously. A builtin performance and cost evaluation module reads signals within the path model and computes the delay and coupling noise across the vertical interconnect for each topology. The cost is estimated in terms of the area penalty of TSVs, including KOZ and the total capacitance of vertical interconnect which contributes to dynamic power.

Although the methodology allows for custom TSV topologies to be defined, only four topologies are implemented and analyzed in this paper, namely, border, bundle, shielded, and isolated. We determined these as the corner points in our design space in terms of noise voltage, area penalty and implications on the placement and routing of the design, and therefore consider them as representative topologies. The border topology is based upon the rationale that if TSVs are placed along the periphery like I/O pins then a rectangular

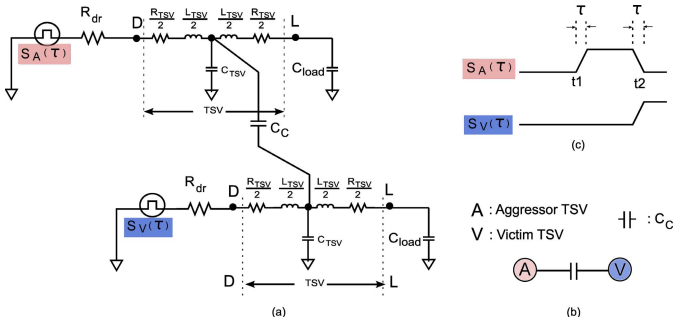


Fig. 3. (a) Circuit diagram for simulation of capacitive coupling noise using D-to-L path models for two TSVs. (b) Top-view schematic. (c) Input signal waveforms for the aggressor and victim TSVs.

area can be used for placement of devices in its entirety. Thus, existing planar design practices and algorithms can be retained. With a closely packed bundle of TSVs, an enclosure around the bundle including the KOZ can be blocked out for placement of devices causing minimal disruption to existing place and route tools. The shielded topology is a form of the bundle topology but with only two aggressors, thereby improving signal integrity at the cost of increased capacitance to ground and with twice the area overhead. In the last topology, TSVs are spaced out such that stress components do not add up and it is thus termed as isolated. The structure, implications, and utilization of these topologies is explained and illustrated in Section V.

C. Modeling With Physical Effects

The RLC model from [19] is used to model TSVs as it has been shown to be reasonably accurate in the dimension range of the ITRS 3D-SOC and 3-D-SIC taxonomies [20]. The model expresses TSV self parasitic resistive (R_{TSV}), inductive (L_{TSV}), and capacitive (C_{TSV}) components as a function of TSV geometry, material properties, and the frequency (clock edge rate $1/\tau$). To simulate signal integrity, the coupling capacitor C_c is connected between two TSVs as shown in Fig. 3(a). A top-view schematic of the same is shown in Fig. 3(b). The input signal waveforms for the aggressor $[S_A(\tau)]$ and victim $[S_V(\tau)]$ TSVs are shown in Fig. 3(c). The worst case capacitive coupling induced noise normalized to the supply voltage, observed on a victim line due to the switching of all aggressors [t1 in Fig. 3(c)], and the maximum increase in delay resulting from the switching of all aggressors in a direction opposite to that of the victim [t2 in Fig. 3(c)] are estimated by simulation.

It should be noted that the coupling noise observed in TSVs is strongly influenced by the structure of the vertical interconnect. For instance, in the case of a vertical interconnect with TSVs arranged in a closely packed bundle of 3×3, the victim TSV in the center of the bundle is coupled with eight neighboring aggressors. If the same interconnect were organized such that TSVs formed a border around the logic block, each victim would only have two aggressors. Therefore, the electrical performance of vertical links is highly dependent upon how TSVs are placed. In addition, this placement has also been found to influence the KOZ requirements for TSVs. For a digital 3-D chip, the KOZ is been defined as the area around a TSV where change in saturation current ΔIdsat for MOSFETs is greater than 5% [18]. In [18], Mercha *et al.* analyzed the impact of thermomechanical stresses induced during TSV formation on device integrity. They observed that for TSVs arranged in a row (border) or in a bundle, the stress components add up and thus propagate larger distances into the surrounding silicon, implying the need for a larger KOZ. These results form the basis of the KOZ guidelines derived in Section V-A.

The dependence of signal integrity as well as the KOZ requirements on the placement of TSVs forms the primary motivation behind creating TSV placement topologies. Our exploration methodology provides estimates for TSV area after accounting for the KOZ, and also provides delay and noise information for each topology. In addition, it enables the selection of a suitable vertical interconnect structure early in the architecture space exploration, and provides feedback allowing initial floorplans to be revised to better achieve target system specifications. The TSV area estimates and revised floorplan from this stage are used as inputs by the subsequent thermal modeling step shown in the system-level methodology in Fig. 1.

IV. Temperature-Power Simulation

A. Related Work

Thermal issues arising from the high density of integration in 3-D architectures necessitates the use of aggressive thermal management techniques, and the inclusion of thermal effects in the architecture space exploration stage of the design flow. Recent studies, such as [10] illustrate the performance benefits of defining 3-D MP-SoC architecture based on the thermal simulation results. However, given the gravity of thermal issues encountered deep within die-stacks, a runtime power management strategy is essential toward ensuring a reliable design. It is also prudent for such runtime schemes to be included within the simulation setup to better understand the thermal performance of 3-D architectures.

DVFS is a commonly used runtime power management technique that operates PEs at different voltage and frequency levels according to their workload [21]. Wang *et al.* reported improvements in application performance as well as the effective utilization of power budget in [22] using a temperature constrained DVFS-based power management scheme for planar chip multiprocessors (CMPs). The scheme controls the voltage and frequency levels of individual PEs based on their local operating temperature, and the available chip power budget. However, it cannot be applied to 3-D architectures since it does not consider thermal coupling between adjacent PEs–a significant factor in die-stacks [23]. Sabaryz *et al.* highlighted the inefficacy of the conventional DVFS approaches applied to 3-D architectures in [24] by analyzing the variation in thermal conditions between the extremities of deep stacks, that resulted in PEs on lower tiers turning off more often than others. Their solution, however, used a thermal management policy requiring the use of an inter-tier liquid cooling system. A comprehensive thermal

management policy for 3-D CMPs incorporating temperature aware workload migration and runtime global power-thermal budgeting is presented in [23]. Within the policy, PEs with available temperature budgets executing high instructions per cycle workloads are scaled to higher voltage and frequency levels to improve performance after weighing the potential performance benefits of such scaling against the consequent thermal implications for neighboring PEs.

Our flow integrates a runtime power manager with a thermal simulation engine to yield a methodology for temperature-power simulation of 3-D architectures. This enables the exploration and refinement of 3-D floorplans, and their evaluation in presence of a runtime power management strategy. A key contribution that resulted from our methodology is a novel temperature-constrained power management scheme for 3-D MP-SoCs that uses instantaneous temperature monitoring coupled with information on the physical structure of the die-stack to determine operating voltage-frequency levels for PEs. The scheme uses a weighted policy while implementing scaling decisions, thereby preventing PEs on deeper tiers from reaching critical temperatures and being turned off. The scheme outperforms the conventional 2-D DVFS both in its ability to maintain the temperatures of all PEs stable, as well as in its improvement of performance by increasing the aggregate system frequency.

B. Thermal Modeling

The complexity of the interconnection and TSV structures increases the complexity of the conductive heat transfer paths in die-stacks. Dummy vias and intertier connections can be used to increase the vertical heat transfer through the stack and reduce the temperature peaks in the die [25]. Even so, the thermal conductivity of the dielectric layers inserted between device layers for insulation is very low as compared with silicon and metal [26], leading to temperature gradients in the vertical direction of the 3-D chip. In the case of hot spots, these thermal effects are even more pronounced. Continuous thermal monitoring is therefore necessary to reduce thermal damage and increase reliability of the chip. Successful application of 3-D integration requires the development of an analytical model for heat transport in 3-D ICs, and thus a set of thermal design rules governing the feasibility of integration options. A number of such models are found in literature, such as [27] which presents a thermal analysis of heterogeneous 3-D ICs with various integration schemes. The analysis of temperature distribution on an inhomogeneous substrate layer can be performed by employing the finite-difference method [28], the green's function and the fast Hankel transform of the green's function [29], or mesh-based methods [30]. However, existing thermal simulation methods when applied to a full-chip reduce the computational complexity of the problem by homogenizing the materials within a layer, limiting the extent of an eigenfunction expansion, or ignoring the source's proximity to boundaries. These simplifications render their results less accurate at fine length-scales, on wires, vias, and individual transistors.

To accurately estimate the on-chip thermal gradients, we use a temperature profile estimation methodology [31] with the capability to include a layout geometry of individual circuit blocks in a chip. The simulator uses layout geometry, boundary conditions, and physical thermal parameters as initial values to formulate a system of partial differential equations, which are then approximated into a system of ordinary differential equations (ODEs) with the discontinuous Galerkin method. The ODEs are subsequently numerically integrated in a self-consistent manner using the modified Runge–Kutta method. The electrothermal couplings are also embedded in the core of the simulator which simultaneously estimates temperature-dependent quantities for each simulation step.

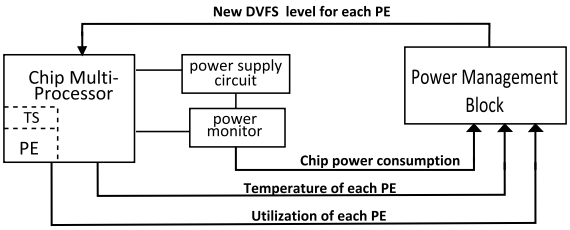


Fig. 4. Control loop for power management scheme.

C. Power Management Scheme

The temperature-constrained power management scheme for 3-D MP-SoCs is implemented within the customizable power management block (PMB) which is responsible for controlling the voltage and frequency of PEs within the temperature-power simulation. Fig. 4 shows the conceptual control loop for the PMB. The block reads the utilization or activity rate of each PE and its temperature, and the total chip power computed through a power measurement circuit within the power supply, to set a new voltage and frequency levels for PEs at regular intervals. For such a scheme to be effective, it is important to model the dynamics of the controlled system, i.e., establish the relationship between the manipulated and the controlled variables. In this case, the operating voltage-frequency level is used as a manipulated variable to control power and temperature of the system.

The range of DVFS in MP-SoCs is usually limited, and within this small range [32], [33] observe that the relationship between power and DVFS level can be approximated with a linear function. Hence, the power dissipation of PEs is modeled as

$$P = A * V^2 * F + B \tag{1}$$

where A and B are the constants, P is the power, and V and F are the voltage and frequency corresponding to a DVFS level. The value of A depends on the characteristics of the workload being executed on the PE, and in cases where the target workload is known, this may be set to a generalized value. Equation (1) strongly resembles

$$P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}}. \tag{2}$$

Constants A and B are hence representative of activity factor and static (leakage) power, respectively. To develop a dynamic model equation, the difference equation is considered

$$\Delta P = A * \Delta(V^2 * F). \tag{3}$$

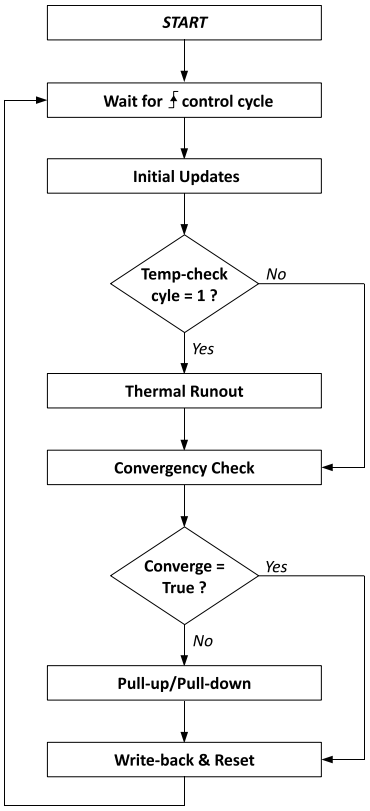


Fig. 5. Flowchart illustrating the various stages of the power management scheme.

While the thermal conductance between two PEs is calculated by using conductance equations, due the complex nature of heat flow, additional information, such as the possible heat transfer paths, as well as the impedance along each such path are necessary to establish a direct relation between the temperature and voltage-frequency levels.

The temperature of a PE in a 3-D stack is primarily determined by its power dissipation, physical location within the die-stack, and its area. The power management scheme considers these parameters in determining appropriate voltage-frequency levels to keep the total chip power below a set power budget value, while keeping the temperature of PEs under critical temperature values. A temperature margin (T_{margin}) is considered to maintain the temperature of PEs at a safe distance from the critical limit even under unexpected circumstances, such as noise in the power supply or a sudden increase in their workload.

DVFS decisions are made by the algorithm shown in Fig. 5. Note that control period defines the interval at which the PMB samples the three earlier mentioned inputs and computes new voltage-frequency levels. Temperature inputs to the PMB are made available at intervals defined by the temperature check period.

The algorithm shown in Fig. 5 is divided into several stages, namely, initial updates, thermal run-out, convergence check, and pull-up/pull-down, and write-back and reset. No scaling decisions are made in the first two control periods after system startup. The system is initialized at maximum voltage-frequency levels, and begins execution with the maximum power dissipation.

Initial Updates: At the beginning of a new control period, the difference between the total chip power and the local power budget value is computed. In the event that a new temperature check cycle has started, the difference between the actual and the critical temperatures of each PE is updated.

Thermal Runout: This step ensures that the temperature of each PE is maintained within the safety margin. Each PE is assigned a weight

$$a * (1 - \text{Util}) + b * (\text{normalized } R_{\text{eff}} \text{ [victimPE] } [i]) \quad (4)$$

where a and b are the coefficients whose values are obtained through a linear regression [34], once the range of voltage-frequency values for the algorithm are specified. A PE that is close to its critical temperature is referred to as a victim. From the equation, a less active PE bearing a strong thermal relation with the victim is considered to have the heaviest weight, and is thus the prime candidate for voltage-frequency scale down. If required, the next candidate PE is selected and scaled down, and this process continues until the victim's temperature is brought under the critical value. In the event that the victim remains at or exceeds this critical temperature, it is clock gated. Repeated fluctuations between voltage-frequency levels may, however, be observed in certain cases, incurring large performance and power penalties. As a means to avoid this, the voltage-frequency levels of PEs that were scaled down due to a victim are prevented from being reinstated until the victim is within the safe temperature margin. This is implemented by means of a special flag that can only be reset in the initial updates stage of the algorithm.

Convergence Check: To prevent frequent fluctuations in the voltage-frequency levels, the algorithm considers the power value as converged when the total chip power is between 98% and 100% of the power budget value. If this is not the case, voltage-frequency levels may be scaled in the pull up/pull down stage of the algorithm.

The power value is considered as converged if total chip power is between 98% and 100% of power budget value. If this is not the case, V-F levels are scaled in the next stage of the algorithm–pull up/pull down.

Pull Up/Pull Down: In this stage, the voltage-frequency level of PEs is pulled up or down based on their allocated power budget using the weighted equation

$$(c * \text{Util}) + (d * \text{normalized\_temp\_margin}) + (e * \text{normalized\_height}) + (f * \text{normalized\_area}) \quad (5)$$

where c, d, e, and f are the weights whose values are determined through an exploratory simulation at design time. The weights serve to establish the impact of these parameters on the choice of PE for scaling. Since the height of the stack and area of PEs may be expected to remain constant even through floorplan revisions, only the utilization and temperature margin are considered to be variable in this equation. In addition, since the value of utilization may be generalized for a homogeneous MP-SoC, an exploratory simulation is only required once to determine

the value of d, corresponding to the temperature margin. This equation may be applied to both island-based as well as per-core schemes. The per-core scheme may simply be considered as an island scheme in which each island contains only one PE. The weight of an island is thus the average weight of all PEs within it. From the equation, a highly active PE that is cooler, situated close to the heatsink and has a larger area, is the preferred choice for voltage-frequency upscaling. However, this is performed only if the projected temperature after scaling is found to be below the safety margin.

The PE with the largest weight is chosen for voltage-frequency upscaling. This upscaling is performed iteratively until no more PEs can be pulled up or if the total power reaches the 98% window of convergence with the budget value. In the event that the budget has been exceeded, the pull down stage is invoked to achieve convergence. For the voltage-frequency downscaling, the PE with the smallest weight is selected and the pull down is performed iteratively until no more PEs can be pulled down or until the total power falls below the budget value. At each instance of pull up and pull down, the difference between the PE's actual and critical temperatures is updated. It is recommended that the range of voltage-frequency values supported by the algorithm be set keeping in mind the power budget value. This ensures that even in the extreme case where all PEs are pulled down to their minimum voltage-frequency level, their power dissipation falls well within the power budget, thereby allowing the temperature of the critical PE to be brought within the safe margin.

Write-Back and Reset: Finally, the voltage-frequency level determined for each PE is actuated along with the ON/OFF state signals if required. At this stage, internal parameters are reset, and the algorithm is suspended until the next control cycle.

D. Exploration Method

The complete simulation setup is shown in Fig. 6. The revised floorplan generated after the TSV topology exploration is used to compute the percentage area occupied by the TSVs. This determines the effective thermal conductance of various layers (active layer, metal layer, and bond layer) of the 3-D stacked IC. A constant thermal conductance is assumed throughout the layer, ignoring the nonuniform density of TSVs. The effective thermal resistance matrix for the PMB is then obtained using the geometrical and material properties of the 3-D stack. Activity traces for PEs are generated by simulating a benchmark application on a system and power simulator, with the trace recording each cycle as either a computation, or as a memory operation. This trace is then used to calculate the activity factor of a PE in each control period. The power values for each voltage-frequency level are then put through a linear regression to determine the values of the parameters A and B in (1).

When the PMB sets the operating voltage and frequency for PEs inside a control period, the power dissipation value for each is computed and recorded. The total power dissipated by the stack is subsequently calculated using these values, and is provided to the PMB as an input in the next control

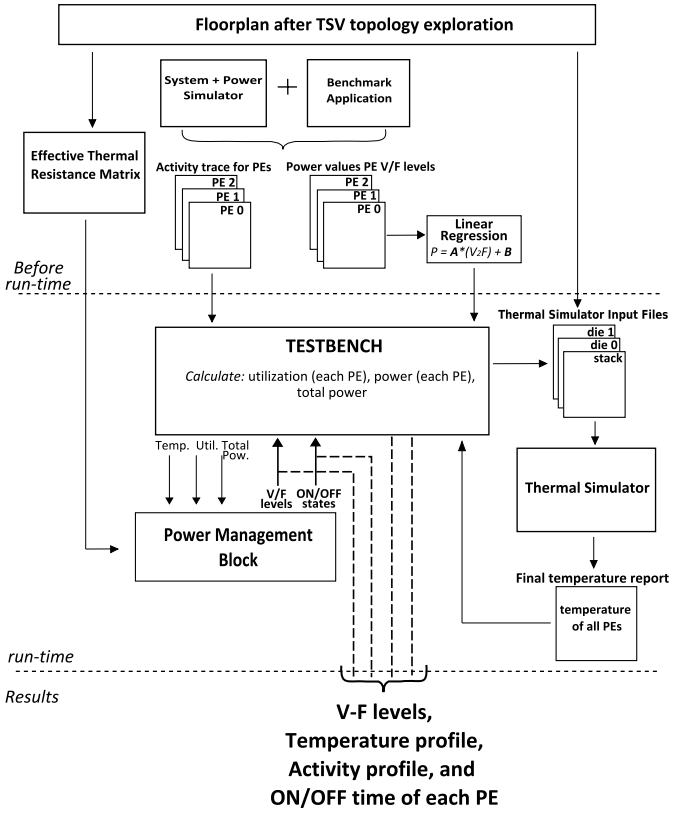


Fig. 6. Complete simulation environment.

period. At the invocation of the temperature check cycle, the average power value for each temperature cycle is written to the thermal simulator, which is then triggered with timing information identical to that of the temperature check cycle. This simulator generates a thermal profile for the stack, and writes the final temperature values for each PE back to the PMB. Performance analysis is enabled by the testbench which captures information, such as temperature, power, operating voltage-frequency levels, actual execution time, and the off time of each PE. Additionally, it also records the total power dissipation within each control period.

The use of DVFS implies that some PEs may run at frequencies lower than their maximum, resulting in reduced execution performance. This penalty is estimated using the expression

$$\text{Perf. Penalty } (\%) = \frac{\text{increase in exec. time}}{\text{exec. time with max freq}} * 100\%. \quad (6)$$

Voltage and frequency switching take a finite amount of time to complete, during which PEs remain stalled. These switching times are however not included in our models. Their implications are discussed in the next section.

V. Experimental Setup and Results

As a test case, a many-core clustered processor architecture targeted toward high performance computing applications is implemented within our flow. The Naga architecture [35] addresses the challenges of planar many-core scalability by incorporating PEs and caches on separate tiers in a 3-D stack

to improve integration density without increasing overall area footprint. The architecture is composed of multiple single-issue in-order PEs with private L1 instruction and data caches within tiles that are arrayed over multiple tiers of the stack. Tiles are interconnected through a wormhole routed 3-D mesh network consisting of seven-port routers with two TSV-based vertical links. Alongside enabling stacking, the use of a 3-D mesh results in lower end-to-end packet latencies when compared with planar meshes with the same number of nodes and under identical traffic conditions [36]. L2 cache banks traditionally placed alongside the processing array are now placed on independent tiers of the stack enabling larger cache sizes, and separating them from PEs which are characteristically different in their power dissipation. This paper only considers the PEs within the 3-D stack for TSV topology exploration and temperature-power simulation, and does not include L2 cache banks. Nonetheless, the methodology and results remain valid even with their inclusion.

The TSV topology exploration and the subsequent temperature-power simulation methodologies are implemented in SystemC, facilitating interoperability with C/C++− based system simulators. In addition, the interface adapters between the digital signals from functional blocks and electrical signals of the path model shown in Fig. 2 can also be realized using the analog mixed signal extension of SystemC. Support for functional simulation during TSV topology exploration may also be added through the SystemC description of functional blocks.

A basic algorithm for TSV placement is used that tries to maintain the required topology while keeping the wire length minimum and avoiding overlaps while placing TSVs. The run time for the TSV topology exploration is less than 4 min for up to 1000 TSVs on a workstation with an Intel dual core CPU 6600 at 2.40 GHz, with 2-GB RAM. To estimate power figures corresponding to execution, the SimpleScalar processor simulator [37] is used with an online power estimator [38] at different voltage-frequency levels. Activity traces are generated for the basicmath application from the MiBench benchmark suite [39] for an in-order 32-bit PISA configuration. Basicmath is used since it contains a mix of varied computations and load-stores, and represents the activity patterns of a typical computation kernel.

A. Implementation and Setup

Four geometries that form the limits of the TSV dimension range for 3-D-SOC [20] are considered to derive the set of TSV parameters that result in the minimum and the maximum delay through the TSV. Although the TSV model trends observed by sweeping the values of material properties and the frequency indicated a large percentage change [40], the impact of this change on the path delay is not significant once the overall path circuit is considered. These values are thus maintained fixed to $S_{\text{gnd}} = L$, $\tau = 10$ ps, $N_A = 1 \times 10^{21}$ /m^3 for a high resistivity Si substrate, Cu TSV fill, and SiO^2 as TSV dielectric. The dimensions and TSV RLC values computed for the four TSV geometries as per [19] are shown in Table I.

TABLE I
TSV GEOMETRIES CONSIDERED VALUES OF THEIR PARASITIC COMPONENTS AND PERFORMANCE CORNERS

	Thick TSV		Thin TSV	
	t_{diel}	$2 \times t_{diel}$	t_{diel}	$2 \times t_{diel}$
Dimensions				
Diameter (D μm)	8	8	4	4
Pitch ($P\mu m$)	16	16	8	8
Aspect Ratio (AR)	5	5	10	10
Length (L μm)	40	40	40	40
t_{diel} (μm)	0.5	1	0.5	1
Parasitic Components				
R_{TSV} (Ω)	0.20	0.20	0.50	0.50
L_{TSV} (pH)	15.00	15.00	20.20	20.20
C_{TSV} (fF)	52.30	26.10	30.90	15.40
C_c (F)	5.88	5.88	5.87	5.87
Performance				
50% Delay (ps)	4	3	3	2
Corner Case	*Worst*			*Best*

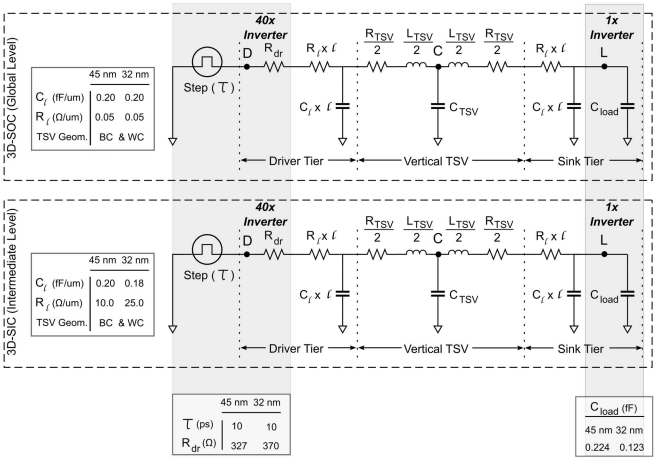


Fig. 7. D to L path model for 3-D-SOC and 3-D-SIC.

The path models for 3-D-SOC and 3-D-SIC for 45 and 32 nm are shown in Fig. 7. R_{dr} and C_{load} are estimated by assuming a 40 buffer driver and a 1× buffer load and using device models provided by Predictive Technology Model (PTM) [41]. A buffer typically comprises of a chain of inverters, and therefore for the driver buffer, R_{dr} is estimated for the last inverter (assumed as 40×) in the chain while for the load buffer C_{load} is estimated for the first inverter (assumed as 1×) in the chain. The resulting values of R_{dr} and C_{load} and assumed values of R_l, C_l, and τ are summarized in tables in the figure. Note that while different assumptions for driver and load devices change the resulting path delay, the methodology still remains valid.

To validate the path model, 40× and 1× buffers are built for 45 nm using the PTM models. Two circuits are simulated in *Cadence Spectre*, one with a capacitance of 40 fF (global wire of length 200 μm) and the other with an additional C_{TSV} of 35 fF in the path. The accuracy of the path model comprising of R_{dr} and C_{load} as compared with the circuit using inverters is 83% and 90% for the two circuits.

We base the KOZ guidelines on the analysis of the impact of TSVs on 40-nm devices [18].

1) KOZ is equal along both axes of TSV as ΔIdsat = 0% is observed at the center of the diamond structure.

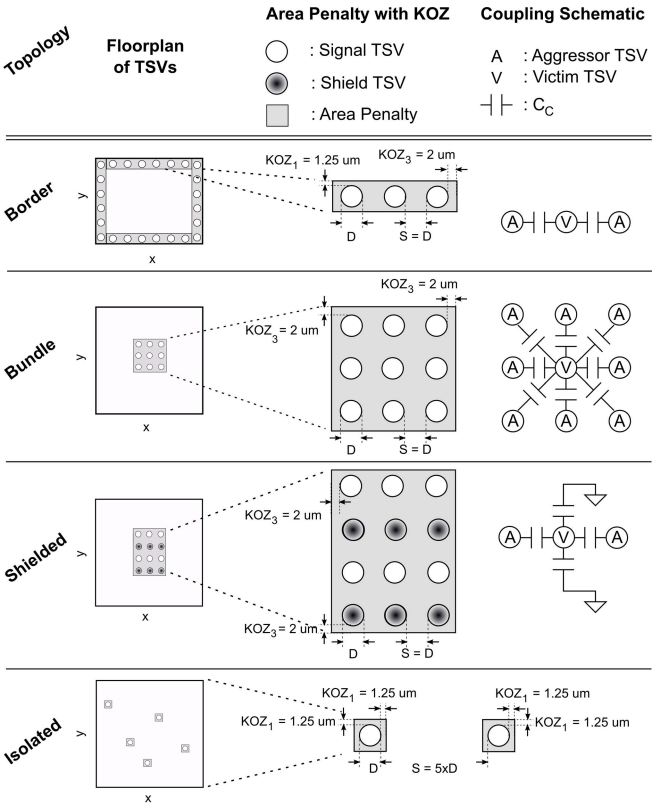


Fig. 8. TSV placement topologies. Note that the number of TSVs shown in each topology is only illustrative.

2) Devices cannot be placed between two TSVs when spacing $S = D$ as ΔIdsat $> 5\%$.
3) For $S \geq 4 \times D$, KOZ is minimum $\text{KOZ}_1 = 1.25$ μm.
4) When N TSVs are placed in a row with $S = D$, KOZ along the row is $\text{KOZ}_2 = 1.53$ μm for $N = 2$, $\text{KOZ}_3 = 2.0$ μm for $N = 3$ and $\text{KOZ}_4 = 2.125$ μm for $N \geq 4$.

For an isolated TSV, the aggregate area including the KOZ is about 2.6 times the area of the TSV itself, which is a significant overhead. In addition, this factor varies according to whether TSVs are arranged in a row or in a matrix, implying that the KOZ must be accounted for in the estimation of total TSV area penalty. Both capacitive coupling noises induced in a TSV and its KOZ requirement are highly dependent on the placement of other TSVs in its proximity. Our paper combines these two aspects in a novel way to explore TSV placement topologies, namely border, bundle, shielded, and isolated. These represent corner points in the design space in terms of their noise voltage and area penalty as well as in their implications on the placement and routing of the design. Other topologies that were considered included different forms of the bundle using a larger TSV pitch. However, these were omitted since their performance was found to fall within the envelope of the four previously mentioned topologies, except with a higher area penalty. The four selected topologies are therefore considered as representative, and are shown in Fig. 8 with their corresponding floorplan, area penalty and coupling schematic.

The topology exploration step offers exploration of TSV technology across the best and worst case performance corners. It also allows for the selection of an appropriate stacking level at the input, i.e., 3-D-SOC and 3-D-SIC, across the 32- and 45-nm CMOS technology nodes. Based on these, it enables the exploration of user-defined TSV topologies to determine the optimal in terms of electrical performance and area penalty.

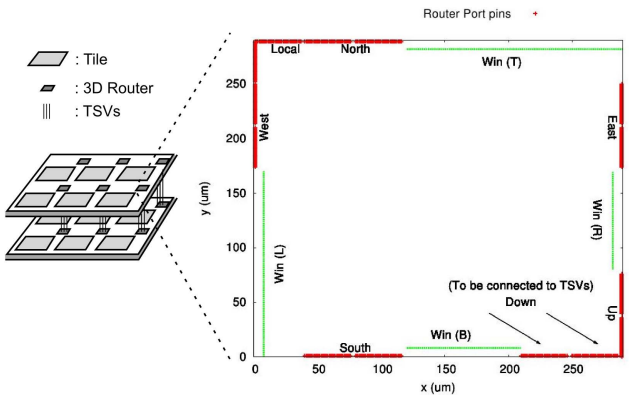


Fig. 9. Seven-port routers forming a 3-D mesh and floorplan of the router in the 45-nm technology node.

B. TSV Topology Exploration Results

Naga's interconnect is organized as a 3-D mesh topology-based network-on-chip [35]. Routers within the network consist of five planar ports, connecting to adjacent routers in the north, south, east, and west directions, as well as one port servicing the local tile. Two additional ports, namely up and down, connect to routers on the upper and lower tiers, respectively, through TSV-based vertical links, enabling the 3-D expansion of the network. The placement of TSVs for these vertical links is challenging due to the constraints imposed by the floorplan of the router itself. Firstly, the router's ports are situated along its edges, with outgoing horizontal wires implementing planar links. As shown in Fig. 9, these ports are mirrored about the horizontal and vertical axes such that the global wire length between routers is minimized. Secondly, to minimize the length of vertical links, the up and down ports must be aligned accurately. Thirdly, a certain amount of area is occupied by the arbitration logic for each port. This leads to a limited number of windows within which TSVs may be placed, as shown by the annotated bars in the 45-nm router floorplan.

The TSV topology exploration methodology is applied to the 3-D router for the 45-nm technology node, and topologies are compared on the basis of their performance and cost for the Best Case TSV geometry (shown in Table I). The unidirectional link width of the router is 37-bit [35] and hence 74 TSVs need to be placed for the down port as shown in Fig. 9. The stacking level used is 3-D-SOC.

Table II lists the exploration results for the 3-D router for the Best Case at 45 nm. The maximum and minimum capacitance, and hence delay, correspond to the longest and shortest nets, respectively, in the vertical path connecting two routers. The ratio of maximum to minimum delay for the border and the isolated topologies is seven while that for the bundle and the shielded topologies is two. However, the router

TABLE II
EXPLORATION RESULTS FOR THE 3-D ROUTER FOR BEST CASE AT 45 nm

	Border	*Bundle*	*Shielded*	*Isolated*
Performance metric				
Min. Capacitance (fF)	26	80	80	26
Max. Capacitance (fF)	184	152	176	186
Min. Delay (ps)	7	20	19	7
Max. Delay (ps)	47	38	44	47
Optimal		✔		
Max. Increase in Delay (ps)	5	23	8	1
Normalized Noise	0.16	0.42	0.11	0.05
Optimal				✔
Freq. (GHz)	13.9	15.8	14.5	13.9
Freq. With noise (GHz)	13.0	11.6	12.9	13.7
Percent decrease	7	27	11	1
Optimal				✔
Cost metric				
Total capacitance (fF)	8080	8276	9092	6970
Optimal				✔
Total TSV area (μm^2)	3328	5220	10422	3126
Percent TSV area	4.0	6.2	12.4	3.7
Optimal				✔

TABLE III
EXPLORATION RESULTS FOR THE 3-D ROUTER FOR BEST CASE AT 32 nm

	Border	*Bundle*	*Shielded*	*Isolated*
Performance metric				
Min. Capacitance (fF)	62	46	52	-
Max. Capacitance (fF)	112	118	148	-
Min. Delay (ps)	17	13	14	-
Max. Delay (ps)	31	32	41	-
Optimal	✔			
Max. Increase in Delay (ps)	7	26	9	-
Normalized Noise	0.17	0.43	0.11	-
Optimal			✔	
Freq. (GHz)	17.8	17.5	15.2	-
Freq. With noise (GHz)	15.8	12.0	13.3	-
Percent decrease	11	31	13	-
Optimal	✔			
Cost metric				
Total capacitance (fF)	5986	5760	7050	-
Optimal		✔		
Total TSV area (μm^2)	3854	5220	10422	-
Percent TSV area	9.2	12.4	24.8	-
Optimal	✔			

to router path in the planar mesh of a single tier consists of equal length horizontal wires that result in a uniform delay. If the same is desirable across the vertical path as well, the placement of the down port can be changed and the methodology runs again. With respect to the maximum delay, the bundle topology is found optimal. To provide a comparison with planar links in 2-D meshes, assume a global wire of length 1.5-mm interconnecting two routers placed on adjacent tiles. The capacitance of this horizontal interconnect link is estimated to be 300 fF, which is 1.8 times larger than that of the TSV-based vertical link.

The bundle topology shows a worst case normalized capacitive coupling noise of 0.42, which violates the noise margin of 0.34 for the 45-nm technology node. The isolated topology on the other hand, exhibits a much lower normalized noise value of 0.05, well within the margin, on account of the wide spacing between its TSVs. Similarly, increased capacitance to ground with the shielded topology results in better noise performance than the border topology. The resulting frequency for vertical links in each topology is calculated by adding a setup time of 25 ps to the path delay, and any increase in this delay due to noise limits the maximum achievable operating frequency. This is highlighted in the case of the bundle topology where the relatively high coupling noise causes a 27% reduction in operating frequency compared with that of the same topology without noise. The wide spacing of TSVs in the isolated topology reduces the impact of capacitive coupling, resulting in a mere 1% difference between frequencies with and without noise. These results provide an early estimate of operating frequency for the vertical interconnect, and enable evaluation of TSV topologies based on their noise performance.

The cost of each topology is measured in terms of total capacitance and the area penalty of the TSVs including the KOZ. The isolated topology is found to be the least costly amongst the explored topologies, and is accordingly determined as the optimal. The flexibility in TSV placement for this topology results in short horizontal wires and hence a lower total capacitance. In addition, the total area is calculated as an aggregate of individual TSV areas in contrast to other topologies where an enclosure is considered. For the border topology, more TSVs are placed in the lower-right of the router (nearer to the down port) than in the upper-left. As a consequence, the total capacitance for border is lower than for the bundle or the shielded topologies. In summary, the results from the topology exploration for the 45-nm technology node indicate the isolated topology as optimal, in terms of performance as well as cost, for the Naga 3-D router.

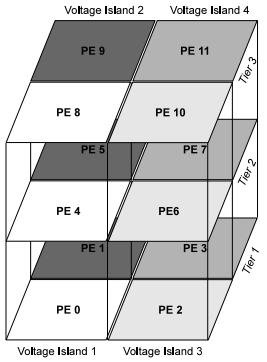


Fig. 10. Voltage island partitioning.

Table III lists the exploration results for the 3-D router at 32 nm using the same Best Case TSV geometry. The results indicate that the border topology is superior to the shielded topology in terms of electrical performance as well as area overhead. However, decreased feature sizes at this technology node result in a smaller router block, consequently reducing the area available for the placement of TSVs (same diameter as in the 45-nm technology node). The narrow placement windows resulting from the reduced area of the router cannot accommodate all 74 TSVs of the vertical link in the form of the border topology, leading to a number of these TSVs being placed outside of the router block. Such a placement is considered invalid as it violates the topological requirement of TSVs forming a border along the inner periphery of the router block. This indicates that the border topology is viable only if the area for TSV placement is increased. In the event that other design constraints restrict such a floorplan modification, the shielded topology forms the next feasible option in terms of both electrical performance as well as placement viability.

C. Temperature-Power Management Results

In the three-tier stack shown in Fig. 10, each PE is operated at one of six voltage-frequency levels as shown in Table IV.

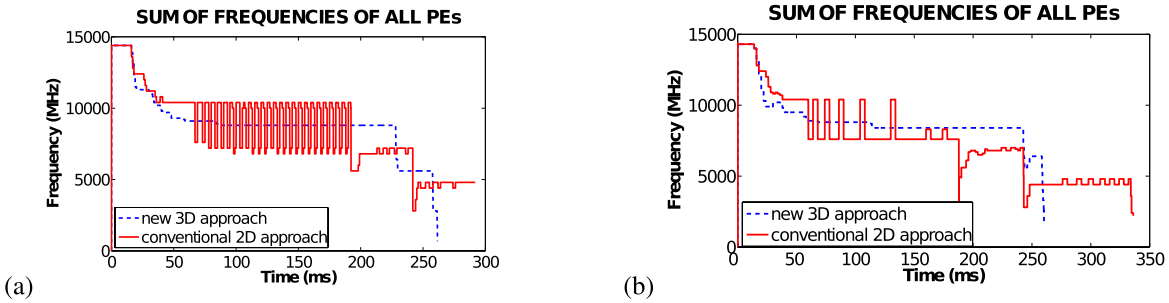


Fig. 11. Sum of frequencies of all PEs. (a) Shielded topology. (b) Isolated topology. Note that the sum of frequencies achieved with the shielded topology using the conventional 2-D approach is higher as compared with the isolated case, resulting in the reduction of execution time. The new 3-D approach achieves shorter execution times than the 2-D approach with both topologies.

TABLE IV
DVS LEVELS WITH CORRESPONDING DFS LEVELS

		Frequency 1	Frequency 2
Voltage 1	0.855V	700MHz	800MHz
Voltage 2	0.956V	900MHz	1000MHz
Voltage 3	1.048V	1100MHz	1200MHz

By supporting multiple frequency levels within each voltage, large islands can be prevented from completely switching between voltage levels by performing only dynamic frequency scaling (DFS) on their constituent PEs. Although DFS can quickly arrest a thermal runoff situation, in the event that it proves insufficient, dynamic voltage scaling (DVS) is performed. In the extreme case of both DVS as well as DFS proving insufficient in controlling the thermal runoff, the victim PE is clock gated, thereby reducing its power dissipation to only the leakage power. For the power management scheme, a control period of 60 000 cycles, corresponding to 50 μs at maximum frequency is used based on the time required for voltage transitions to complete. Since such transitions are of the order of tens of nanoseconds, the selected control period results in a negligible overhead during switching of the voltage-frequency levels [42]. These transitions are consequently ignored in the temperature-power simulation results. It must be noted that while the number and granularity of DVFS levels can be increased within the methodology at the cost of increased runtime, only six DVFS levels are considered in this paper. Two parallel simulation setups are used, each using the same convergence algorithm, similar conditions for the VFS and similar constraints on power and temperature. One of the setups implements our power management scheme for 3-D MP-SoCs while the other uses a conventional DVFS scheme for 2-D chips where the temperature of each PE is considered independently. To reiterate, a 2% window of convergence is considered within the scheme to maximize power budget utilization and reduce fluctuations in voltage-frequency levels. The temperature margin is dependent on temperature sensor accuracy, and is determined experimentally as 2 K. The critical temperature margin is set to 318 K.

Fig. 11 shows the sum of frequencies for the MP-SoC achieved with the shielded and isolated topologies, which were determined earlier as the optimal topologies during the TSV topology exploration. The shielded topology incorporates a larger number of TSVs within the same area as compared with the isolated topology, and thus offers a higher thermal conductance toward the heatsink. This has the effect of reducing temperatures at lower tiers, resulting in a reduction in the duration of PE stalls. This is shown in Fig. 11(a), where the aggregate frequency achieved during execution is higher as compared with that of the isolated topology. Execution, as a consequence, completes faster. However, even though the duration of PE stalls is decreased in the case of the shielded topology, their frequency of occurrence is increased, as evidenced by the repeated spikes in the aggregate frequency profile. The presented 3-D power management approach on the other hand, prevents such stalls from occurring by weighing the local performance benefits obtainable by voltage-frequency scaling, against the thermal implications of such an operation on neighboring PEs. Therefore, even though individual PEs operate at voltage-frequency levels lower than their maximum, the presented scheme achieves an overall higher aggregate frequency for the MP-SoC than the 2-D approach by preventing stalls from occurring. As similar execution performance is noted with both topologies, the isolated topology is determined to be optimal for the architecture due to its lower area penalty.

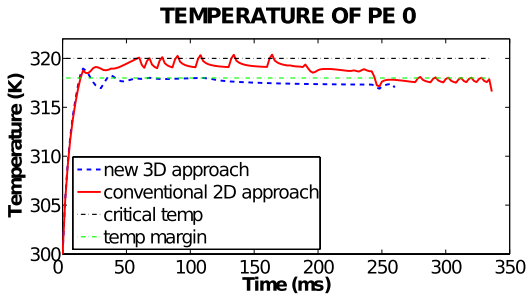


Fig. 12. Operating temperature of a PE on the lowest tier of the stack illustrating the effective temperature control with the new approach.

Fig. 12 shows the temperature of PE0, situated on the lowest tier of the three-tier stack with both the conventional 2-D as well as the presented 3-D approaches using the isolated topology. Conventional 2-D DVFS schemes consider the temperatures of PEs independently and are oblivious to the physical structure of the stack. Therefore, the temperature is observed to fluctuate within the margin, and occasionally breach the critical limit resulting in the PE being clock gated and its execution stalled. The new 3-D approach, however,

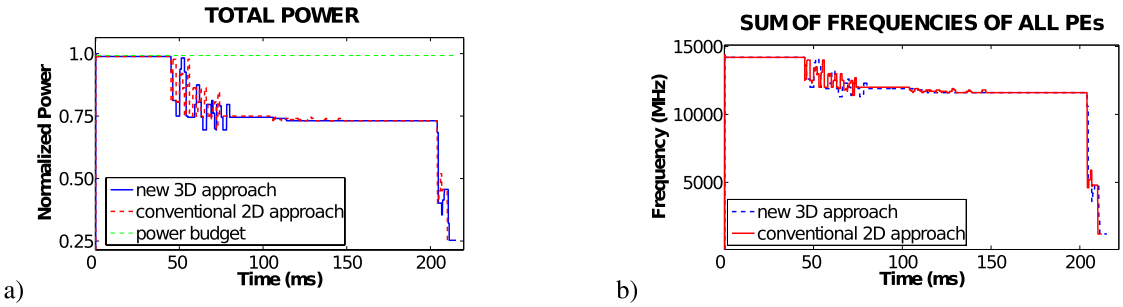


Fig. 13. (a) Total power dissipation of the stack. (b) Sum of frequencies with PEs grouped into islands.

TABLE V
VOLTAGE-FREQUENCY LEVEL TRANSITIONS WITH THE CONVENTIONAL 2-D APPROACH AND THE NEW 3-D APPROACH

		PER CORE		ISLAND	
		2D	NEW 3D	2D	NEW 3D
TIER 3	*PE 11*	1	5	16	23
	PE 10	1	6	18	5
	PE 9	1	6	19	5
	PE 8	1	7	22	11
TIER 2	*PE 7*	16	4	17	23
	PE 6	12	5	20	5
	PE 5	12	5	19	5
	PE 4	11	6	25	13
TIER 1	*PE 3*	33	6	29	33
	PE 2	31	7	28	9
	PE 1	31	5	26	9
	PE 0	33	7	31	17
AGGREGATE FREQ.		67	30	111	57

is observed to maintain temperatures well below the temperature margin without any fluctuations. Thus, PE0 completes execution of its task much sooner with the new approach, than with the conventional approach. This is shown in Fig. 12, with the temperature profile for the new approach terminating close to the 250 ms mark, well ahead of the conventional approach which extends up to 340 ms. Fig. 11(b) shows the aggregate frequency of the complete MP-SoC, i.e., the sum of frequencies of all PEs, for the isolated topology, in which the conventional 2-D approach is found to have a higher aggregate frequency than the 3-D approach up to 50 ms. Subsequently, however, the temperatures of some PEs, such as PE0 begin to approach the critical limit, causing them to be clock gated. At this point, the aggregate frequency is seen to fall, and once again rise a few milliseconds later. Such fluctuations are observed throughout the period of execution. In the new approach, the weighted equation used to determine candidates for scaling ensures that the temperature of PEs on the lowest tiers of the stack is considered before switching any voltage-frequency levels. As a consequence, all PEs are operated at voltage-frequency levels lower than the maximum, thereby reducing their power dissipation, and ensuring that they never cross the critical temperature margin. Execution stalls due to thermal runoff are therefore eliminated, and a performance improvement of 19.55% is observed for this case.

The stack is subsequently partitioned into voltage islands as shown in Fig. 10. The process essentially groups vertically adjacent PEs with strong thermal relationships into a single voltage domain. Similar to the previous experiment, the voltage island partitioned stack with the new scheme is compared with a stack with a per-core 2-D DVFS scheme. The total power dissipation for the two schemes is shown in Fig. 13. While both schemes maintain power dissipation well within the set power budget, the new 3-D approach is seen to result in a smoother profile with much fewer fluctuations than the 2-D approach. The sum of frequencies in Fig. 13(b) is also observed to be smoother with the new approach. Table V lists the total number of voltage-frequency transitions that occur during execution in the MP-SoC, with both the conventional 2-D as well as the new 3-D power management schemes. Our scheme is seen to result in a fewer transitions at each PE, as well as fewer fluctuations in the aggregate frequency of the system as compared with the conventional 2-D scheme.

With the 2-D scheme, PEs on the upper tiers are found to operate at the highest frequency level that their local power and temperature budgets allow. On the other hand, those on lower tiers run at much lower frequencies on account of their increased temperature. The new scheme considers the temperature of PEs on lower tiers while scaling the island voltage and the frequency of individual PEs on higher tiers. This effectively improves the performance of PEs deep within the stack. The island partitioning however, also restricts PEs in the upper tiers to voltage-frequency levels lower than that achievable with the per-core 2-D approach, i.e., only a part of the available performance from the upper PEs is utilized. The two approaches are noted to present similar results in terms of execution performance. While the 2-D approach offers this performance primarily from PEs in tiers closer to the heatsink, the new 3-D scheme achieves the same performance by ensuring that PEs even in the lower tiers remains active. This leads to a more uniform utilization of all PEs in the MP-SoC rather than the preferential utilization of only a few close to the heatsink. By grouping PEs into islands, the area overhead of level shifters and voltage converters is reduced, in this case by over 60%. The similarity in temperature and frequency profiles illustrates that the performance of 2-D per-core DVFS can be achieved even while using voltage islands. Since our power management scheme achieves this performance through the uniform utilization of all PEs in the MP-SoC, devices can be expected to wear more evenly than with the conventional scheme which results in the preferential utilization of the cooler PEs alone.

VI. CONCLUSION

While 3-D integration is seen as a promising solution toward sustaining the trend of increasing integration densities, a number of challenges arise from the stacking of silicon dies.

Two such critical issues include the KOZ, necessitated by mechanical stress considerations of TSVs, and the consequent thermal implications of integrating multiple layers of logic in die-stacks with different vertical interconnect topologies. This paper introduces an integrated system-level methodology that addresses these challenges by enabling the architecture space exploration based on the performance and cost of vertical interconnects, and thermal behavior of 3-D stacks. The methodology concurrently evaluates several custom TSV topologies for vertical interconnects, and reports the electrical performance and area overhead for each by means of a fast SystemC simulation. The results of this evaluation enables the selection of a vertical interconnect structure that meets both the design as well as the KOZ requirements within the architecture space exploration phase itself, and the determination of a prospective placement plan for TSVs well ahead of the long running IC design flow. In addition, these results can also be applied toward revising floorplans to accommodate specific TSV topologies that achieve certain critical design requirements.

The temperature-power simulation step of the introduced methodology provides early design time estimates of the thermal performance of the 3-D die-stack. It relies on a thermal model of the stack derived from the revised floorplan after TSV topology exploration and information on the TSV count and thus thermal conductance of each tier to generate an effective thermal profile. A unique aspect of our methodology lies in its ability to simulate the thermal behavior of stacks in the presence of a runtime power management strategy. The effectiveness of a conventional 2-D DVFS scheme in controlling temperatures of PEs in a 3-D stack was evaluated using this feature of the methodology. This analysis indicated the preferential utilization of PEs closer to the heatsink by the conventional power management strategy, and consequently, the under utilization of PEs farther away from the heatsink. As a result of this analysis, a novel temperature-constrained power management scheme that addresses the short-comings of the conventional DVFS approach was developed specifically for 3-D MP-SoC. The developed scheme is observed to improve execution performance by 19.55% by increasing aggregate frequency through the uniform and sustained utilization of all PEs within the stack, while reducing the number of voltage-frequency level transitions that occur. Finally, our scheme when applied to a voltage island partitioned 3-D stack, was found to perform at par with per-core 2-D DVFS, with a 60% lower area overhead in terms of level shifters and voltage converters.

REFERENCES

[1] J. Kim, J. S. Pak, J. Cho, E. Song, J. Cho, H. Kim, T. Song, J. Lee, H. Lee, K. Park, S. Yang, M.-S. Suh, K.-Y. Byun, and J. Kim, "High-frequency scalable electrical model and analysis of a through silicon via (TSV)," *IEEE Trans. Compon., Packag. Manuf. Technol.*, vol. 1, no. 2, pp. 181–195, Feb. 2011.

[2] T. Kgil, A. Saidi, N. Binkert, R. Dreslinski, S. Reinhardt, K. Flautner, and T. Mudge, "Picoserver: Using 3D stacking technology to enable a compact energy efficient chip multiprocessor," in *Proc. Int. Conf. Archit. Support Program. Lang. Operat. Syst.*, 2006, pp. 117–128.

[3] G. Loh, "3D-stacked memory architectures for multi-core processors," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2008, pp. 453–464.

[4] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza, "3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling," in *Proc. Int. Conf. Comput.-Aided Design*, 2010, pp. 463–470.

[5] A. Jain, R. E. Jones, R. Chatterjee, S. Pozder, and Z. Huang, "Thermal modeling and design of 3D integrated circuits," in *Proc. Intersoc. Conf. Thermal Thermomech. Phenomena Electron. Syst.*, 2008, pp. 1139–1145.

[6] C. Sun, L. Shang, and R. P. Dick, "Three-dimensional multiprocessor system-on-chip thermal optimization," in *Proc. Int. Hardw./Softw. Codesign Syst. Synthesis Conf.*, 2007, pp. 117–122.

[7] J. Cong, J. Wei, and Y. Zhang, "A thermal-driven floorplanning algorithm for 3D ICs," in *Proc. Int. Conf. Comput.-Aided Design*, 2004, pp. 306–313.

[8] J. Cong, A. Jagannathan, Y. Ma, G. Reinman, J. Wei, and Y. Zhang, "An automated design flow for 3D microarchitecture evaluation," in *Proc. Asia South Pacific Design Autom. Conf.*, 2006, pp. 384–389.

[9] W.-L. Hung, G. M. Link, Y. Xie, N. Vijaykrishnan, and M. J. Irwin, "Interconnect and thermal-aware floorplanning for 3D microprocessors," in *Proc. Int. Symp. Quality Electron. Design*, Mar. 2006, pp. 98–104.

[10] O. Ozturk, F. Wang, M. Kandemir, and Y. Xie, "Optimal topology exploration for application-specific 3D architectures," in *Proc. Asia South Pacific Design Autom. Conf.*, Jan. 2006, pp. 390–395.

[11] R. Jagtap, S. S. Kumar, and R. van Leuken, "A methodology for early exploration of TSV placement topologies in 3D stacked ICs," in *Proc. 15th Euromicro Conf. Digit. Syst. Design*, Sep. 2012, pp. 382–388.

[12] A. Aggarwal, S. S. Kumar, A. Zjajo, and R. van Leuken, "Temperature constrained power management schemes for 3D MPSoC," in *Proc. Int. Workshop Signal Power Integr.*, 2012, pp. 7–10.

[13] M. Daneshtalab, M. Ebrahimi, P. Liljeberg, J. Plosila, and H. Tenhunen, "CMIT—A novel cluster-based topology for 3D stacked architectures," in *Proc. Int. 3D Syst. Integr. Conf.*, Nov. 2010, pp. 1–5.

[14] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan, and M. Kandemir, "Design and management of 3D chip multiprocessors using network-in-memory," in *Proc. Int. Symp. Comput. Archit.*, 2006, pp. 130–141.

[15] C. Liu, T. Song, J. Cho, J. Kim, J. Kim, and S. K. Lim, "Full-chip TSV-to-TSV coupling analysis and optimization in 3D IC," in *Proc. Design Autom. Conf.*, Jun. 2011, pp. 783–788.

[16] R. Weerasekera, M. Grange, D. Pamunuwa, H. Tenhunen, and L.-R. Zheng, "Compact modelling of through-silicon vias (TSVs) in three-dimensional (3-D) integrated circuits," in *Proc. Int. Conf. 3D Syst. Integr.*, Sep. 2009, pp. 1–8.

[17] P. Falkenstern, Y. Xie, Y.-W. Chang, and Y. Wang, "Three-dimensional integrated circuits (3D IC) floorplan and power/ground network co-synthesis," in *Proc. Asia South Pacific Design Autom. Conf.*, 2010, pp. 169–174.

[18] A. Mercha, G. Van der Plas, V. Moroz, I. De Wolf, P. Asimakopoulos, N. Minas, S. Domae, D. Perry, M. Choi, A. Redolfi, C. Okoro, Y. Yang, J. Van Olmen, S. Thangaraju, D. S. Tezcan, P. Soussan, J. H. Cho, A. Yakovlev, P. Marchal, Y. Travaly, E. Beyne, S. Biesemans, and B. Swinnen, "Comprehensive analysis of the impact of single and arrays of through silicon vias induced stress on high-k/metal gate CMOS performance," in *Proc. IEEE IEDM*, Dec. 2010, pp. 2.2.1–2.2.4.

[19] I. Savidis and E. Friedman, "Closed-form expressions of 3-D via resistance, inductance, and capacitance," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1873–1881, Sep. 2009.

[20] (2010). *International Technology Roadmap for Semiconductors* [Online]. Available: http://www.itrs.net/links/2010itrs/home2010.htm

[21] S. Herbert and D. Marculescu, "Analysis of dynamic voltage/frequency scaling in chip-multiprocessors," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2007, pp. 38–43.

[22] X. Wang, K. Ma, and Y. Wang, "Adaptive power control with online model estimation for chip multiprocessors," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 10, pp. 1681–1696, Oct. 2011.

[23] C. Zhu, Z. Gu, S. Li, R. P. Dick, and R. Joseph, "Three-dimensional chip-multiprocessor run-time thermal management," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 8, pp. 1479–1492, Aug. 2008.

[24] M. M. Sabryz, D. Atienza, and A. K. Coskuny, "Thermal analysis and active cooling management for 3D MPSoCs," in *Proc. Int. Symp. Circuits Syst.*, 2011, pp. 2237–2240.

[25] Y. Z. J. Cong and J. Wei, "A thermal-driven floorplanning algorithm for 3D ICs," in *Proc. Int. Conf. Comput. Aided Design*, 2004, pp. 306–313.

[26] A. M. Ionescu, G. Reimbold, and F. Mondon, "Current trends in the electrical characterization of low-k dielectrics," in *Proc. Int. Semicond. Conf.*, 1999, pp. 27–36.

[27] T.-Y. Chiang, S. Souri, C. O. Chui, and K. Saraswat, "Thermal analysis of heterogeneous 3D ICs with various integration scenarios," in *IEDM Tech. Dig.*, 2001, pp. 31.2.1–31.2.4.
[28] T.-Y. Wang and C. C.-P. Chen, "3-D thermal-ADI: A linear-time chip level transient thermal simulator," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 21, no. 12, pp. 1434–1445, Dec. 2002.
[29] B. Wang and P. Mazumder, "Fast thermal analysis for VLSI circuits via semi-analytical Green's function in multi-layer materials," in *Proc. Int. Symp. Circuits Syst.*, vol. 2. May 2004, pp. 409–412.
[30] N. Allec, Z. Hassan, L. Shang, R. Dick, and R. Yang, "Thermalscope: Multi-scale thermal analysis for nanometer-scale integrated circuits," in *Proc. Int. Conf. Comput.-Aided Design*, Nov. 2008, pp. 603–610.
[31] A. Zjajo, N. van der Meijs, and R. van Leuken, "Thermal analysis of 3D integrated circuits based on discontinuous Galerkin finite element method," in *Proc. Int. Symp. Qual. Electron. Design*, Mar. 2012, pp. 117–122.
[32] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No 'power' struggles: Coordinated multi-level power management for the data center," in *Proc. Int. Conf. Archit. Support Program. Lang. Operat. Syst.*, 2008, pp. 48–59.
[33] X. Wang and M. Chen, "Cluster-level feedback power control for performance optimization," in *Proc. Int. Symp. High Perform. Comput. Archit.*, Feb. 2008, pp. 101–110.
[34] A. Aggarwal, "Temperature-constrained power management scheme for 3D MPSoC," M.S. thesis, Dept. Microelectron. Comput. Eng., Delft Univ. Technol., Delft, The Netherlands, Dec. 2011.
[35] S. S. Kumar and R. van Leuken, "A 3D network-on-chip for stacked-die transactional chip multiprocessors using through silicon vias," in *Proc. Int. Conf. Design Technol. Integr. Syst. Nanoscale Era*, Apr. 2011, pp. 1–6.
[36] V. Pavlidis and E. Friedman, "3-D topologies for networks-on-chip," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 10, pp. 1081–1090, Oct. 2007.
[37] T. Austin, E. Larson, and D. Ernst, "Simplescalar: An infrastructure for computer system modeling," *IEEE Comput.*, vol. 35, no. 2, pp. 59–67, Feb. 2002.
[38] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A framework for architectural-level power analysis and optimizations," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2000, pp. 83–94.
[39] M. R. Guthaus, J. S. Ringenberg, D. Ernst, T. M. Austin, T. Mudge, and R. B. Brown, "MiBench: A free, commercially representative embedded benchmark suite," in *Proc. Int. Workshop Workload Characterizat.*, Dec. 2001, pp. 3–14.
[40] R. Jagtap, "A methodology for early exploration of TSV interconnects in 3D stacked ICs," M.S. thesis, Circuits Syst. Group, Faculty Electr. Eng., Math. Comput. Sci., Delft Univ. Technol., Delft, The Netherlands, Sep. 2011.
[41] (2008, Sep. 30). *Predictive Technology Model* [Online]. Available: http://ptm.asu.edu/
[42] W. Kim, M. S. Gupta, G.-Y. Wei, and D. Brooks, "System level analysis of fast, per-core DVFS using on-chip switching regulators," in *Proc. Int. Symp. High Perform. Comput. Archit.*, 2008, pp. 123–134.

Sumeet S. Kumar (S'08) received the B.E. degree in electronics and communications engineering from Visvesvaraya Technological University, Belgaum, India, in 2008, and the M.Sc. degree in electrical engineering with a specialization in microelectronics from the Delft University of Technology, Delft, The Netherlands, in 2010, where he is currently pursuing the Ph.D. degree in electrical engineering with the Circuits and Systems Group.

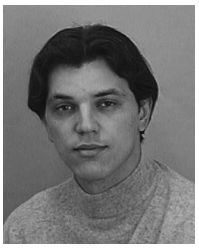
He was an Intern with Indrion Technologies, Bangalore, India, in 2008, where he worked on developing energy-efficient instruction set extensions for a sensor control network processor. His work focuses on the design of a high-performance 3-D many-core processor architecture that is scalable, dependable, and easy to program.

Arnica Aggarwal received the B.Tech. degree in electronics and communication engineering from the Amity School of Engineering and Technology, Noida, India, in 2009, and the M.Sc. degree in electrical engineering with a specialization in microelectronics from the Delft University of Technology, Delft, The Netherlands, in 2011.

She joined ASML Netherlands B.V., Veldhoven, The Netherlands, in 2012, and is currently an E-Integrator with the Electronics Development Department. Her current research interests include VLSI design, power aware and low power, and energy digital system design.

Radhika Sanjeev Jagtap received the B.Tech. degree in electronics and telecommunication engineering from the College of Engineering, Pune, India, in 2007, and the M.Sc. degree in electrical engineering with a specialization in microelectronics from the Delft University of Technology, Delft, The Netherlands, in 2011.

She was with the IBM Systems and Technology Laboratory, Bangalore, India, from 2007 to 2009, as a Research and Development Engineer, where she was involved in the physical design of the Z-series processor. She joined ARM Holdings, Cambridge, U.K., in 2012, and is currently a Graduate Engineer with the Research and Development Engineer Division. Her current research interests include computer architecture, full system simulation, trace-driven simulation, performance measurement, and power modeling.



Amir Zjajo (M'02) received the M.Sc. and D.I.C. degrees from Imperial College London, London, U.K., in 2000, and the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2010, all in electrical engineering.

He joined Philips Research Laboratories, Eindhoven, in 2000, as a Research Staff Member with the Mixed-Signal Circuits and Systems Group. From 2006 to 2009, he was with Corporate Research of NXP Semiconductors, Eindhoven, as a Senior Research Scientist. In 2009, he joined the Delft University of Technology, Delft, The Netherlands, as a Faculty Member with the Circuit and Systems Group. He has published more than 60 papers in referenced journals and conference proceedings, and holds more than ten U.S. patents or patents pending. He is the author of *Low-Voltage High-Resolution A/D Converters: Design and Calibration* (Springer, 2010, Chinese translation, 2012). His current research interests include mixed-signal circuit design, signal integrity and timing, and yield optimization of VLSI.

Dr. Zjajo serves as a member of Technical Program Committee of the IEEE Design, Automation and Test in Europe Conference, the IEEE International Symposium on Circuits and Systems, and the IEEE International Mixed-Signal Circuits, Sensors and Systems Workshop.

Rene van Leuken (M'80) was born in The Netherlands in 1955. He received the Ph.D. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 1988.

He is a Professor with the Circuit and Systems Group, Delft University of Technology. He has been involved in many research projects, including ESPRIT, FP6, FP7, JESSI, MEDEA, and recently in MEDEA+ and ENIAC/CATRENE projects. He has published papers in all major conferences and workshops proceedings. His current research interests include high level system design, design automation, system design optimization, and DSP engines.

Dr. van Leuken is a member of the PATMOS Program Committee.