New Thermal-Aware Voltage Island Formation for 3D Many-Core Processors

Hyejeong Hong, Jaeil Lim, Hyunyul Lim, and Sungho Kang

The power consumption of 3D many-core processors can be reduced, and the power delivery of such processors can be improved by introducing voltage island (VI) design using on-chip voltage regulators. With the dramatic growth in the number of cores that are integrated in a processor, however, it is infeasible to adopt per-core VI design. We propose a 3D many-core processor architecture that consists of multiple voltage clusters, where each has a set of cores that share an on-chip voltage regulator. Based on the architecture, the steady state temperature is analyzed so that the thermal characteristic of each voltage cluster is known. In the voltage scaling and task scheduling stages, the thermal characteristics and communication between cores is considered. The consideration of the thermal characteristics enables the proposed VI formation to reduce the total energy consumption, peak temperature, and temperature gradients in 3D many-core processors.

Keywords: Many core, task scheduling, thermal aware, three-dimensional integration, voltage island.

I. Introduction

The use of multiple supply voltages through the design of voltage islands (VIs) has been introduced into system-on-chip (SoC) design to minimize power consumption levels. A VI in a many-core processor is a set of cores that are physically contiguous and are powered by the same supply voltage. Timing-critical tasks are assigned to the cores in the VI that are powered by higher supply voltages so that the performance constraint can be met. Meanwhile, tasks that are not timing-critical are assigned to the cores in the VI that are powered by lower supply voltages; thus, these cores run more slowly, thereby reducing the total power consumption [1]–[2].

Three-dimensional (3D) integration technology has been accepted as a solution to the problems faced by traditional two-dimensional (2D) integration technology. In 3D ICs, the global wire length is reduced by a factor of $\sqrt{k}$, where k is the number of stacked layers [3]. Recently, microprocessor design has been shifting from multi-core to many-core configurations due to the power wall that multi-core processors are facing [4]. The performance of many-core processors can be improved using 3D integration technology because the on-chip communications are significantly shortened. Integrated multiple core dies are accepted as a promising alternative to be used in future high-performance computing systems [5].

Although 3D stacking of core layers offers a lot of advantages, some existing problems may be exacerbated. One of the challenges is the thermal crisis. The power density per unit volume considerably increases compared to 2D technology, so the peak temperature may soar. It was shown that the peak temperature of a 3D chip made of two layers increased by more than 20°C without any modifications to mitigate the thermal problems [6]. The temporal and spatial

Manuscript received Mar. 1, 2014; revised Sept. 17, 2014; accepted Oct. 6, 2014.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A2A1A03006255).

Hyejeong Hong (hjhong@soc.yonsei.ac.kr), Jaeil Lim (limji@soc.yonsei.ac.kr), Hyunyul Lim (lim8801@soc.yonsei.ac.kr), and Sungho Kang (corresponding author, shkang@yonsei.ac.kr) are with the Department of Electrical & Electronic Engineering, Yonsei University, Seoul, Rep. of Korea.

temperature gradients also become larger due to the thermal characteristics of stacked layers — it is easier for the layers that are closer to the heat sink to release the heat than it is for those that are further away [7]. The elevation of the temperature may cause reliability crisis or performance degradation. Thus, thermal-aware methods should be added to the 3D design flow [8].

Power delivery is also one of the critical challenges. Multiple layers are stacked, but the number of I/O pins is limited. 3D ICs have to deliver k times more current with the same power supply. Adopting on-chip voltage regulators to 3D many-core processors can support a stable supply power across the layers. Moreover, it supports the fine-grained voltage and frequency management and better power delivery efficiency. There have been attempts to embed voltage regulators on chips to support fine-grained voltage management [9], and a voltage regulator stacked in 2.5D has been demonstrated [10].

In this paper, a thermal-aware VI formation for 3D many-core processors is proposed. A VI formation for many-core processors includes a series of procedures to decide on which core a task is mapped and a proper supply voltage for a core. The proposed VI formation is based on homogeneous 3D many-core processor architectures, which are restricted by the power delivery network design. A set of cores is tied to an on-chip voltage regulator to form a voltage cluster (VC), which is the unit of a VI. The VCs are prioritized based on a steady-state temperature analysis and are selected to form a VI to minimize the communication and computation energy consumption. The remainder of this paper is organized as follows: Section II introduces the previous works related to this paper. Section III presents the target 3D many-core architecture, and the thermal-aware VI formation for the architecture is explained step by step in Section IV. Section V provides the experimental setup and an analysis of the experimental results.

II. Previous Work

A lot of studies have been done to optimize the energy consumption of many-core processors. The use of multiple supply voltages and dynamic voltage frequency scaling (DVFS) are popular techniques that reduce the computation energy by lowering the supply voltage. The early studies focused on the reduction of the power consumption without performance loss, but the more recent techniques aim to address other design issues at the same time. Managing thermal-management problems through DVFS in 2D ICs has been proposed [11]–[12].

With aggressive scaling, variations become severe. Each core's maximum frequency may differ due to the inter-core and intra-core variations. A communication, process, voltage, and temperature (PVT) variation-aware VI formation voltage selection method was proposed in [13]. It had the flexibility to change the voltage of each core, and it was shown that the cloud-shaped VI formation minimized the impact of PVT variations. However, the per-core level voltage management exacerbates the complexity of the power delivery network, so this approach cannot be easily extended to 3D-stacked many-core design.

There have been some attempts to tackle the thermal problems of 3D many-core processors. Thermal-aware task scheduling for 3D multi-core was proposed in [14]. There is a strong thermal connectivity between vertically adjacent cores. Based on that, a set of vertically adjacent cores is defined as a "supercore." Accordingly, a set of tasks that are scheduled together is bound as a "supertask." The task scheduling for 3D multi-core processors can be regarded as the task scheduling for 2D. A runtime optimization policy for 3D multi-core processors was proposed in [15]. The target architecture is specified as a two-layer 3D multi-core processor: a DRAM layer is stacked on a multi-core layer. The policy selects either the low-power mode or the turbo mode under the power and thermal constraints.

Dynamic thermal management is efficient since the heat generation is mainly dependent on the power consumption, consequently the workload. Unlike 2D ICs, 3D ICs have unique static thermal characteristics, and the problems caused by the static characteristics can be statically addressed. The configuration of the cores and on-chip memories in a layer may affect the thermal characteristics of 3D many-core processors. In [16] and [17], various configurations of cores and on-chip memories in a layer were studied in terms of the thermal issues. In general, it is better to locate cores on the bottom layer and to place cores and on-chip memories to be vertically adjacent.

Our work is not a straightforward extension of the 2D VI formation. Traditionally, the VI formation has aimed to reduce the power and energy consumption. Reducing the impact of some emerging issues, such as process variations, has recently become a new subject. Thermal and power delivery problems are the critical issues that are aggravated in 3D integration technology. We present the thermal-aware VI formation for 3D many-core processors, the architecture of which is restricted by the power delivery network.

III. 3D Many-Core Processor Architecture

Using multiple VIs incurs extra costs, such as the multiple power grid design overhead and additional supporting blocks. The higher the extra costs are, the higher the power efficiency is obtained due to the finer granularity of the VI formation. The design of multiple VIs can be categorized according to the

granularity of VIs. Per-chip VI has the coarsest granularity, which means that every core in the many-core processor operates according to the same DVFS schedule offered by the power grid. The opposite extreme is per-core VI, in which every core operates at an individual supply voltage according to its own DVFS schedule. In many-core processors, it is impractical to design the power delivery network to support per-core VI, as the number of integrated cores in a processor drastically increases. The reasonable alternative between the two extremes is per-cluster VI, which is where a set of cores form a VI. In [18], the trade-off between the number of cores in a cluster and the energy gain in a 2D many-core processor comprised of 80 cores was studied. Clustering four cores showed 70% improvement of energy gain compared with the case with no clustering, while reducing the extra cost by approximately 50%. In particular, 3D die stacking severely increases the complexity of the power delivery network design, so core clustering in pre-fabrication is inevitable.

We consider homogeneous 3D many-core processors, which consist of identical layers with VCs. This assumption of homogeneous 3D ICs is reasonable since stacking identical layers can reduce both design efforts and manufacturing costs [19]. The 3D many-core processor has either 128 cores or 256 cores. Each layer has 64 cores; that is, the 128-core processor consists of two layers, and the 256-core processor consists of four layers. A group of cores is connected to a voltage regulator module (VRM), which is called a VC. The VC is the unit to form a VI. The number of cores in a VC is set as either four or eight. Figure 1 illustrates the floorplan of a layer of the 3D many-core processor with 16 VCs. There are 64 cores in a layer, and four cores are tied to a VC that is connected to a single VRM. The colors of the cores in a VC are the same, which means that they work at the same supply voltage and frequency settings. The shapes of VCs can be either square or rectangular. The impact of the shapes of the VCs will be presented later on in Section V-2.

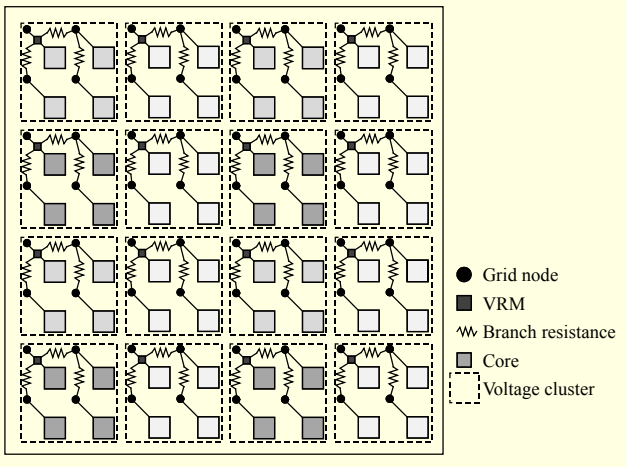

Fig. 1. Floorplan of layer of 3D many-core processors with 2×2 VCs.

The on-chip communication is based on the network-on-chip (NoC). NoC was introduced to support efficient on-chip interconnection as on-chip communication increases and the size of chips grows. Moreover, the concept of NoC, transmitting data through switches and routers, matches well with the use of multiple supply voltages. Mesh topology is the most popular 2D NoC topology due to its regularity. Based on that, a stacked 8 × 8 mesh topology NoC is assumed. The structure of the router of the stacked mesh topology is a straightforward extension of that of the 2D mesh topology: two physical ports, one for up and one for down, are added to support inter-layer data transmission [20].

IV. Thermal-Aware VI Formation

In this section, we present the process of the thermal-aware VI formation for the 3D many-core processors described in Section III. Figure 2 illustrates the overall flow. It aims to minimize the energy consumption while mitigating the thermal problems.

1. VC Thermal Analysis

The thermal dissipation of ICs is dependent on the power consumption, so it depends on the workloads. Therefore, dynamic approaches are needed to effectively solve the thermal problems. In 3D ICs, however, there are some features

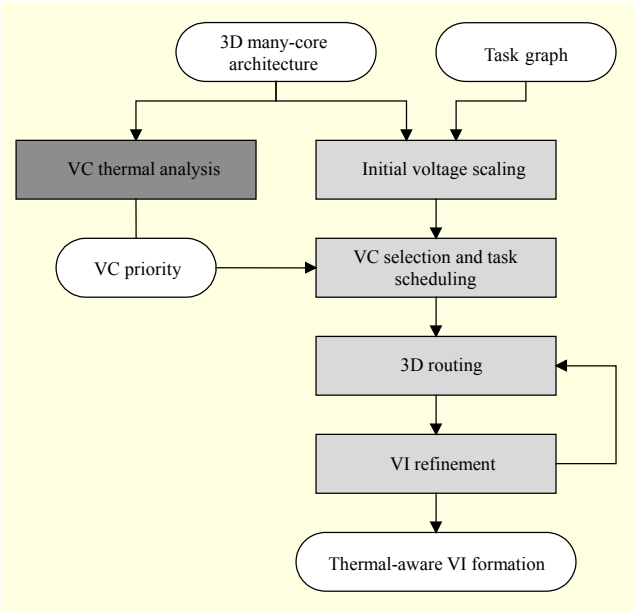

Fig. 2. Proposed thermal-aware VI formation flow.

ETRI Journal, Volume 37, Number 1, February 2015
http://dx.doi.org/10.4218/etrij.15.0114.0257

resulting from the nature of multiple die stacking that can be statically addressed. One side of a chip is contiguous to the heat sink, and the other side is contiguous to the board in 2D ICs, so it is relatively easy to release the generated heat. However, releasing the heat becomes difficult in 3D ICs. Although every layer is identical in the homogeneous 3D many-core processors, the thermal dissipation of each layer differs. The closer a layer is to the heat sink, the greater its rate of heat release. Moreover, the cores at the boundary of a layer tend to release heat more easily than those at the center.

The shapes of VCs are fixed in pre-fabrication, and the floorplan is known. HotSpot [21] is a widely used temperature modeling tool that is basically for 2D systems, but it can be extended to model 3D stacked dies. We utilize the extended HotSpot [15] for the steady-state temperature simulations. We assume that every core runs at the same supply voltage and frequency setting without any idle time for a certain period of time. By making all cores run the same workload, the results of the steady-state temperature simulation show solely the static thermal characteristics of the cores. The characteristics are due to the physical location of the cores; thus, they are deterministic after fabrication. The core is modeled based on the SPARC core, which is relatively small and simple. The power delivery network of VCs in a many-core processor is a standard design; thus, we can assume that all VCs generate the same amount of heat.

Figure 3 shows the two layers with sixteen 4 × 1 VCs — "4 × 1" indicates that the number of cores in the VC is four and that the shape of the VC is a column with four elements. For the sake of simplicity, the power delivery network is omitted in the figure. We number the VCs from left to right and from bottom to top in the floorplan. The first number after the term "VC" represents the layer — "0" represents the top layer, which is the farthest layer from the heat sink. As a result of the steady-state temperature analysis, the VCs with the lowest temperature are given the highest priority. The priority is strongly dependent on the location of the VC. Here, VC1_0, VC1_7, VC1_8, and VC1_15 are the VCs with the highest priority, and VC0_0, VC0_7, VC0_8, and VC0_15 are the VCs with the second-highest priority. The priority is to be used in the VC selection stage to bind the proper VCs to improve the thermal dissipation.

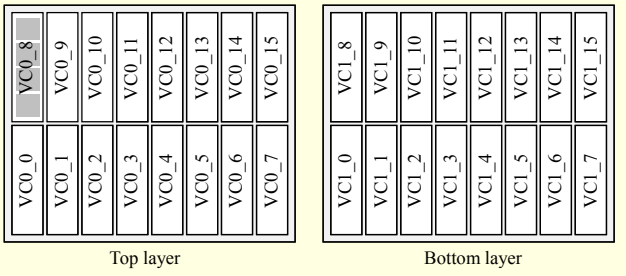

Fig. 3. Floorplan of two layers with 4 × 1 VCs that comprise the 3D many-core processors.

2. Initial Voltage Scaling

The initial voltage scaling is the first step of the dynamic part of the proposed thermal-aware VI formation. The former stage, VC thermal analysis, utilizes only information that is related to the 3D many-core processor architecture as input, but this stage additionally requires the information related to the workloads. A task graph, TG(V, E), is a directed acyclic graph that represents the workloads to be executed on the 3D many-core processor. The vertices represent the tasks, and the edges represent the communication between the connected vertices. A computation time is defined for each vertex, and a communication weight, which is the amount of data transferred, is defined for an edge. Also, the timing constraints are given to each task graph.

First, the initial voltage for each task is selected under the timing constraints for the given task graph. The routing is not yet known in this stage, so the worst-case routing is assumed; it is assumed that all communications between the cores are simultaneously conducted using a single routing path.

We assume that there are m available supply voltages, $\{V_{dd1}, V_{dd2}, \ldots, V_{ddm}\}$, where V_{dd1} is the highest. The operating frequency is defined to be scaled down according to the supply voltage. That is, the supply voltage and operating frequency setting is given as $\{(V_{dd1}, f_1), (V_{dd2}, f_2), \ldots, (V_{ddm}, f_m)\}$. Tasks are slowed down to eliminate their slacks by selecting one of the m available supply voltages and frequency settings. First, the critical path of the task graph is found, and then the supply voltage is selected not to violate the timing constraint but to reduce the slack of the critical path as much as it can. For example, if the selected supply voltage is V_{dd2}, then all the tasks are now scaled down to V_{dd2}. Although the supply voltage is lowered, a lot of tasks still have slacks. The supply voltages of these tasks are then scaled down again to the next lowest, V_{dd3}. This is iterated until there are no remaining tasks having slacks.

Once the initial voltage scaling is completed, all the tasks in the task graph are given one of the supply voltages. Although the physical contiguity cannot be known at this stage, what is certain is that we know which tasks compose a certain VI. This is called the initial VI formation. The number of cores in a VI is now known; thus, the required number of VCs to form the VI is also known.

3. VC Selection and Task Scheduling

After the initial VIs are formed, tasks are physically assigned

to the cores and scheduled. This step brings the thermal awareness into the VI formation by exploiting the results of the VC thermal analysis, VC priority. VCs are sorted according to the VC priority, but there may be several VCs with the same priority. There are four VCs with the highest priority in the 3D many-core platform in Fig. 3; VC1_0, VC1_7, VC1_8, and VC1_15. Moreover, simply using VC priority may increase the total energy consumption considerably. For example, in the case in which three VCs are required for a certain VI, VC1_0, VC1_7, and VC1_8 are selected if the VC priority is followed. However, VC1_7 is physically far from VC1_0 and VC1_8, so this selection elevates the communication overhead if tasks assigned to VC1_7 communicate with tasks on either VC1_0 or VC1_8. If VC0_0, from the second-highest priority, is selected instead of VC1_7, then it results in larger energy savings. To take care of this issue, the VC priority is elaborated by taking communication overhead into account.

The VCs that have the same VC priority as a result of the steady-state thermal analysis are categorized according to their adjacency. VCs are grouped together if they are physically contiguous. As a result, there may be several groups of VCs with identical VC priorities.

Task scheduling is based on an as-soon-as-possible approach, which gives the top priority to the tasks that have the smallest flexibility in terms of timing constraints. Thus, tasks on the critical path are dealt with first. The highest voltage and frequency setting has been selected for the tasks on the critical path in the initial voltage scaling. The VIs with the highest voltage are dealt with first. The first VC selection is straightforward: it randomly selects one of the VCs in any of the groups with the highest VC priority. From the second VC selection, we utilize the elaborated priority (EP), which is given by

$$\mathrm{EP}(\mathrm{VC}_{l\_c}) = \alpha \Delta T_{\mathrm{steady}} + \sum_{w=1}^{W} \mathrm{CW}_w \times D_w, \tag{1}$$

$\mathrm{EP}(\mathrm{VC}_{l\_c})$ includes two factors: static thermal characteristics and communication overhead. The variable $\Delta T_{\mathrm{steady}}$ denotes the difference in the steady-state temperatures between the VC with the lowest steady-state temperature and VC_{l_c}. A smaller $\Delta T_{\mathrm{steady}}$ indicates a more appropriate location for tasks with higher supply voltages. The amount of communication between VC_{l_c} and the predecessor VCs is denoted by CW_w. The Manhattan distance between the center of VC_{l_c} and that of the predecessor VCs is denoted by D_w. The weighting factor, α, is the empirical value that maximizes the thermal improvements for each 3D many-core platform. The VC that has the lowest $\mathrm{EP}(\mathrm{VC}_{l\_c})$ is selected as the next VC.

Once the VCs are selected, the tasks in the VI are assigned to the cores in the VCs and are scheduled. The root task, which is the task that has no inputs but has outputs, is the first one that is assigned. The rest of the tasks are assigned to the cores that are close to those where their predecessors were assigned, thereby minimizing the communication overhead.

4. 3D Routing

Many routing algorithms for many-core processors with NoCs have been proposed, but most of them are for 2D architectures. In [22], a thermal-aware routing algorithm for 3D NoCs was proposed. This algorithm also exploits the fact that the lower layers tend to be easily cooled down. The downward routing gives the *z*-axis routing top priority so that the majority of the data transmissions happen in the lower layers. We modified the algorithm for our target architecture, but we still perform the *z*-axis routing first. The *west-first* routing algorithm is used to prevent deadlock.

5. VI Refinement

As a result of the 3D routing, the actual routing is now known. The routing costs are calculated, and the task graphs are updated with the actual routing costs. The energy that is consumed in the generated VI is calculated. The dynamic power of each core is calculated as

$$P_{\mathrm{dynamic}} = C_{\mathrm{eff}} V_{\mathrm{dd}}{}^2 f, \tag{2}$$

where C_{eff} is the effective switching capacitance, V_{dd} is the supply voltage, and f is the operating frequency. The static power is modeled as

$$P_{\mathrm{static}} = I_0 e^{-\frac{qV_t}{nkT}} V_{\mathrm{dd}}, \tag{3}$$

where I_0 and n are technology-dependent constants, and T is the temperature. The total energy that is consumed by the cores is therefore

$$E_{\mathrm{core}} = (P_{\mathrm{dynamic}} + P_{\mathrm{static}}) \times t_{\mathrm{active}} + P_{\mathrm{static}} \times t_{\mathrm{idle}}, \tag{4}$$

where t_{active} is the time period during which the core is active, and t_{idle} is the time period during which the core is not working. The cores to which no tasks are assigned are completely powered off, so the energy consumption is zero.

The energy consumed by the NoC can be obtained by modeling the communication energy consumed to transfer data from a source to a destination. The communication energy per bit is modeled as

$$E_{\mathrm{bit}} = \left(n_{\mathrm{hops}} \times e_{\mathrm{S_{bit}}}\right) + \left(n_{\mathrm{v\_hops}} \times e_{\mathrm{V_{bit}}}\right) + \left[(n_{\mathrm{hops}} - 1 - n_{\mathrm{v\_hops}}) \times e_{\mathrm{H_{bit}}}\right], \tag{5}$$

where n_{hops} is the number of switches to the destination, and $n_{\mathrm{v\_hops}}$ is the number of inter-layer transmissions. The energy consumed by the switch, per bit, is denoted by $e_{\mathrm{S_{bit}}}$, and $e_{\mathrm{V_{bit}}}$ and

ETRI Journal, Volume 37, Number 1, February 2015
http://dx.doi.org/10.4218/etrij.15.0114.0257

$e_{H_{bit}}$ are the energy consumed per bit by the inter-layer link and the intra-layer link, respectively. The inter-layer link is much shorter than the intra-layer link in a 3D architecture, and $e_{H_{bit}}$ is even greater than $e_{V_{bit}}$.

After the calculation, the next step is to determine whether there are any timing violations or chances for further energy reductions by lowering any task's voltage level. Because we assumed the worst-case routing at the beginning, hardly any timing violations are found. If a timing violation is found, then the root task in the VI is moved to a VI with a higher voltage level, and then the routing has to be done again, and the energy consumption is calculated. On the other hand, if there is slack, then the task that is farthest from the root task is moved to a VI with a lower voltage level. Moving the task may increase the total energy consumption due to the increase in the communication energy. If the new energy savings is not larger than the original VI formation, then the new formation is not taken. Once the calculation is completed, power traces of cores, switches, and links can be obtained. The power traces are utilized to extract temperature traces in the modified HotSpot.

V. Experiments

In this section, we present the experimental setup and analyze the proposed technique in terms of energy saving and thermal management.

1. Experimental Setup

To examine the impact of the size and shape of VCs comprehensively, energy savings and temperature changes were evaluated while varying the size and shape of VCs. The shape of a VC is represented as $n_{row} \times n_{column}$, where n_{row} is the number of core rows in the VC and n_{column} is the number of core columns in the VC. The shapes of VCs used in the experiments are 2×2, 4×1, and 4×2. We also varied the layer count; 128-core processor of two layers and 256-core processor of four layers are considered. Therefore, the proposed technique was evaluated in six 3D many-core platforms in total.

We assume that the on-chip voltage regulator can offer fine-grained voltage management so that the voltage levels used vary from 0.7 V to 1.3 V, in 0.1 V increments. To model the simple and low-performance core in a many-core processor, the maximum frequency is set to 0.5 GHz, while the NoC operates at 1 GHz. The technology used is assumed to be 45 nm process technology.

We have evaluated the energy consumption and temperature using task graphs. Twenty task graphs were randomly generated in the standard task graph [23] format. The three real task graphs, the robot control program, sparse matrix solver, and SPEC95 fpppp kernel, were also used. Table 1 shows the characteristics of the task graphs used in the experiments. The numbers of tasks of robot and sparse are not large enough to evaluate many-core processors, so we tripled them: three identical task graphs are used to compose a task graph. The critical path length is the total computation time of the tasks on the critical path. The parallelism is defined as the sum of the computation times of all tasks divided by the critical path length. The last column of Table 1 shows the range of the twenty generated task graphs.

Table 1. Characteristics of task graphs.

	Robot	Sparse	Fpppp	Random TGs
Number of vertices	88	96	334	300–500
Number of edges	131	67	1,145	827–3,472
Critical path length	569	122	1,062	985–2,623
Parallelism	4.36	15.87	6.71	6.10–25.42

There have been no previous works on the VI formation for 3D many-core architectures yet. In addition, the key objective of the proposed thermal-aware VI formation is to utilize the static thermal characteristics of 3D architectures in making dynamic decisions, which 2D architectures do not have; thus, it is impossible to compare with any other 2D techniques. To demonstrate the advantages of the proposed thermal-aware (TA) VI formation, a non-thermal-aware (non-TA) VI formation was also built. A non-TA VI formation follows the same flow as TA, but it skips the VC selection stage, which uses EP. Instead, it selects the first VC randomly, and the next VCs are selected to minimize the communication overhead. A non-TA VI formation reduces the energy consumed by communication but neglects thermal management. The priority for non-thermal VC selection is given by

$$P(\mathrm{VC}_{\mathrm{l\_c}}) = \sum_{w=1}^{W} \mathrm{CW}_w \times D_w, \tag{6}$$

which is the same as (1) but with α equal to zero.

2. VC Thermal Analysis

In the first step of the proposed VI formation, a steady-state temperature analysis is performed for each of the six many-core platforms. We modified the extended HotSpot to model our 3D many-core processors. The parameters used in the thermal simulation are shown in Table 2. It was assumed that 2% of the die area is occupied by through-silicon vias (TSVs) and that they are evenly spread. The modeling of TSVs is

Table 2. 3D thermal simulation parameters.

Parameter	Value
Die thickness	0.15 mm
Silicon thermal conductivity	100 W/mK
Silicon specific heat capacity	1750 kJ/m^3 K
Interface material thickness	0.02 mm
Interface material conductivity	4 W/mK
Spreader thickness	1 mm
Spreader thermal conductivity	400 W/mK
Heat sink thickness	6.9 mm
Heat sink resistance	0.1 k/W

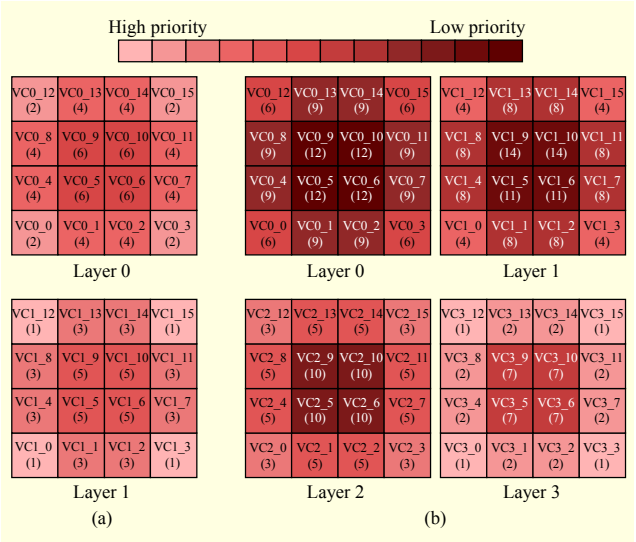

Fig. 4. VC priority maps: (a) 3D many-core processor of 128 cores with 2 × 2 VCs and (b) 3D many-core processor of 256 cores with 2 × 2 VCs.

provided into the modified HotSpot so that the Hotspot can include the heat generated from TSVs in the thermal analysis. The parameters that are not related to 3D die stacking follow the default values of the original HotSpot.

Figure 4 visualizes the VC priority of the 3D many-core processors with 2 × 2 VCs: (a) and (b) are the VC priority maps of the two-layer processors and the four-layer processors, respectively. A small square represents a four-core VC. The darker red color indicates a VC with a higher steady-state temperature; that is, a lower priority. Conversely, the core in lighter red is cooler and has higher priority. The numbers in the parentheses after the VC names are the VC priorities, where "1" means the highest priority. The VCs with the highest priority will be selected first to assign power-consuming jobs. The VC priority maps clearly show the static thermal characteristics of homogeneous 3D many-core processors. The four VCs that are located at the centers of the layers have lower priorities in every layer, and the VCs on the bottom layer have higher priorities.

The VC priority is simple when the layer count is two. The VCs are categorized into six groups in this platform. VC1_0, VC1_3, VC1_12, and VC1_15 have first priority; VC0_0, VC0_3, VC0_12, and VC0_15 have second priority; and VC1_1, VC1_2, VC1_4, VC1_7, VC1_8, VC1_11, VC1_13, and VC1_14 have third priority. By doubling the layer count, the maximum ΔT_{steady} is nearly tripled and shows the static thermal characteristics much more clearly. The VCs are categorized into twelve groups in this platform. The VCs at the vertices of the bottom layer, VC3_0, VC3_3, VC3_12, and VC3_15, have top priority regardless of the layer count. Interestingly, VC3_1, VC3_2, VC3_4, VC3_7, VC3_8, VC3_11, VC3_13, and VC3_14 have second priority, and VC2_0, VC2_3, VC2_12, and VC2_15 have third priority in this platform. This explains that it becomes much more difficult to remove the heat in upper layers as the number of stacked dies grows. This tendency is also shown in the other platforms with 4 × 1 VCs and 4 × 2 VCs.

3. Energy Saving and Thermal Management

The proposed technique basically aims to minimize the total energy consumption by the use of multiple supply voltages. The energy consumption was calculated using (4) and (5) at the last stage of the proposed VI formation. The energy savings are

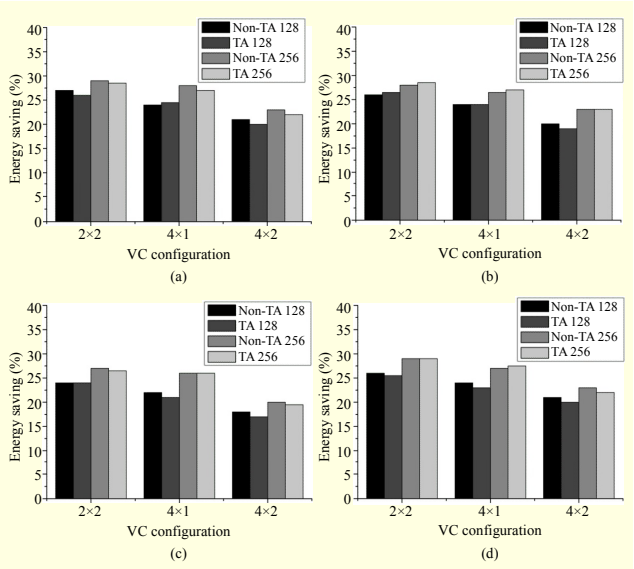

Fig. 5. Energy savings over the case without the use of multiple supply voltages: (a) robot control program, (b) sparse matrix solver, (c) fpppp kernel, and (d) average of 20 randomly generated tasks.

ETRI Journal, Volume 37, Number 1, February 2015
http://dx.doi.org/10.4218/etrij.15.0114.0257

illustrated over the case where multiple supply voltages are not supported (see Fig. 5). In Fig. 5, the two darker bars represent the energy savings of the many-core processor with 128 cores, and the remaining two lighter bars indicate the many-core processor with 256 cores. Generally, the energy savings of the 256-core configurations are larger than those of the 128-core configurations. There are more chances to exploit the parallelism, and in addition, the energy consumption by the inter-layer transmission is significantly smaller, so the total energy consumption is reduced.

What we focus on is the size and shape of the VCs. The energy savings are the least significant when the VC has eight cores, as expected. This is due to the reduced flexibility in the VI formation. However, it reduces the cost for the power delivery network, so the size of the VCs has to be decided depending on the given cost constraints. The impact of the shape of the VCs can be seen by comparing the 2 × 2 and 4 × 1 shapes. For the platforms with 4 × 1 VCs, the communication distance may be lengthened unless the VCs that are vertically adjacent are bound to form a VI. Therefore, the energy efficiency of 4 × 1 VCs is lower in the two-layer many-core processors. However, the energy efficiency of 4 × 1 VCs is improved in the four-layer many-core processors, showing unnoticeable differences compared to 2 × 2 VCs. Overall, the VI formation results in energy savings over the configuration with no multiple supply voltages by 17% to 29%. The differences in the energy savings made by the non-TA and TA VI formations are not significant.

The advantage of the proposed VI formation is demonstrated through the thermal simulations. The power traces were extracted for each task graph in every platform. They are input into the modified HotSpot to generate temperature traces. Since running a task graph does not take long enough to obtain meaningful temperature traces, four scenarios that run a number of task graphs in a specific order were built. The power traces in each core and NoC were extracted and inputted into the modified HotSpot to generate temperature traces. The resulting temperature traces were named T trace 1, T trace 2, T trace 3, and T trace 4. Each scenario has a different combination of the three real task graphs and the twenty generated task graphs, in terms of the number of task graphs and the order in which to run them.

The peak temperatures of the temperature traces when VIs are formed by the proposed TA technique are listed in Table 3. The peak temperatures by non-TA were also obtained, and the reductions on the peak temperatures by TA over non-TA VI formations are listed in Table 4. Although non-TA and TA lead to similar levels of energy saving, the differences in the reduction of the peak temperature clarify the benefit of TA. By introducing the thermal-awareness into the VI formation, the peak temperatures were reduced by 6.18°C to 16.31°C for all of the platforms we considered.

Table 3. Peak temperatures by TA (°C).

	128-core two-layer			256-core four-layer		
	2×2	4×1	4×2	2×2	4×1	4×2
T trace 1	95.02	93.78	95.13	105.81	104.27	105.18
T trace 2	93.89	94.57	95.24	103.67	103.23	105.71
T trace 3	94.37	92.28	94.56	104.43	105.37	104.87
T trace 4	95.57	94.45	94.46	106.22	105.44	106.74

Table 4. Reductions on peak temperature by TA over non-TA (°C).

	128-core two-layer			256-core four-layer		
	2×2	4×1	4×2	2×2	4×1	4×2
T trace 1	9.56	9.20	7.69	14.21	15.75	13.01
T trace 2	9.23	9.07	6.82	14.53	16.31	12.58
T trace 3	9.04	8.36	6.18	14.33	16.02	14.20
T trace 4	10.12	9.61	7.22	14.82	15.96	13.59

Table 5. Maximum temperature gradients by TA (°C).

	128-core two-layer			256-core four-layer		
	2×2	4×1	4×2	2×2	4×1	4×2
T trace 1	12.20	11.87	12.79	20.33	20.84	21.14
T trace 2	11.78	12.29	13.47	19.05	18.51	20.87
T trace 3	13.13	13.04	13.66	20.04	19.71	21.71
T trace 4	12.54	13.26	13.64	19.64	20.01	22.14

Also, the impact of the shapes of the VCs can be seen if looking into the results of TA listed in Table 4. The impact is not clearly shown in the results of the two-layer many-core processors. In the four-layer many-core processors, it is clearly shown: using the 4 × 1 VCs results in further reductions of up to 1.11°C in the peak temperature compared to the 2 × 2 VCs. Although the 4 × 1 VC is not very helpful in minimizing the communication overhead due to its unbalanced shape, it can be beneficial for thermal management because the cores in the boundary in a layer are tied together. The 4 × 1 VCs that are vertically adjacent form a VI and take care of the timing-critical tasks.

The maximum temperature gradients of the temperature traces when VIs are formed by the proposed TA technique are listed in Table 5, and the reductions over non-TA are listed in Table 6. The temperature gradients considerably increased in the four-layer many-core processors compared to the two-layer

Table 6. Reductions on maximum temperature gradients by TA over non-TA (°C).

	128-core two-layer			256-core four-layer		
	2×2	4×1	4×2	2×2	4×1	4×2
T trace 1	2.43	3.08	2.22	5.31	6.24	4.99
T trace 2	2.59	2.67	2.16	6.58	6.89	5.68
T trace 3	2.12	2.08	2.02	6.21	6.77	5.48
T trace 4	2.69	2.32	2.18	5.89	6.14	5.24

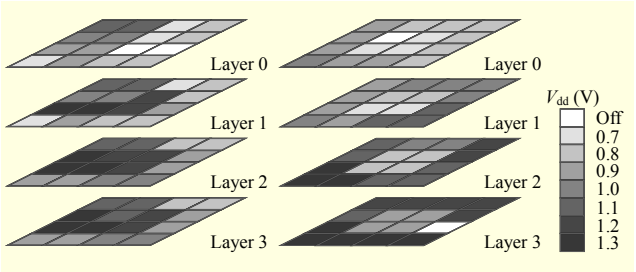

Fig. 6. VI formation for 3D many-core processor of four layers with 2 × 2 VCs, where the task graph is sparse: (a) non-TA VI formation and (b) proposed TA VI formation.

many-core processors. The TA VI formation reduced the maximum temperature gradients by 4.99°C to 6.89°C in the four-layer many-core processors.

Figure 6 compares the results of VI formation by non-TA and TA for 3D many-core processor of four layers with 2 × 2 VCs after the completion of running sparse. The trapezoid represents a VC, and the VCs, which are in the same color and physically contiguous, form a VI. Non-TA VI formation randomly selected VC2_4 as the first VC. The following VC selection is done to minimize the energy consumption by communications. The energy consumption by the inter-layer communication is much lower than that by the intra-layer communication in a 3D many-core processor. Therefore, the majority of the VIs are formed across multiple layers, thereby minimizing the energy consumption by communications. TA VI formation selected VC3_0 on the bottom layer as the first VC. By bringing the thermal-awareness into the VI formation, the result of the VI formation has regularity; the result is reminiscent of the thermal characteristics of 3D many-core processors with 2 × 2 VCs, as illustrated in Fig. 4. The VCs at the boundaries and at the centers are never tied together to form a VI. This is the key to the improvement of the thermal characteristics of 3D many-core processors.

VI. Conclusion

In this paper, we proposed a thermal-aware VI formation for 3D stacked many-core processors. A VC, which is a set of cores connected to a single on-chip voltage regulator, is the unit used to form a VI. This homogeneous floorplan can support stable power delivery across all layers. The VCs are prioritized through a steady-state temperature analysis, and then proper VCs are selected to form a VI. The communication overhead is also taken into account to reduce the communication energy consumption. The experimental results showed that the proposed TA VI formation resulted in energy savings ranging from 20% to 29% when compared to the case without the use of multiple supply voltages. The advantages of TA were remarkable in the four-layer many-core processors. The peak temperature is reduced by up to 16.31°C, and the maximum temperature gradient is reduced by up to 6.89°C.

References

[1] J. Hu et al., "Architecturing Voltage Islands in Core-Based System-on-a-Chip-Designs," *Int. Symp. Low Power Electron. Des.*, Newport, CA, USA, Aug. 9–11, 2004, pp. 180–185.

[2] W.-K. Mak and W. Chen, "Voltage Island Generation under Performance Requirement for SoC Designs," *Asia South Pacific Des. Autom. Conf.*, Yokohama, Japan, Jan. 23–26, 2007, pp. 798–803.

[3] J.W. Joyner, P. Zarkesh-Ha, and J.D. Meindl, "A Stochastic Global Net-Length Distribution for a Three-Dimensional System-on-a-Chip (3D-SoC)," *Int. ASIC/SOC Conf.*, Arlington, VA, USA, Sept. 2001, pp. 147–151.

[4] S. Borkar, "Thousand Core Chips: A Technology Perspective," *Des. Autom. Conf.*, San Diego, CA, USA, June 2007, pp. 746–749.

[5] A.K. Singh et al., "Mapping on Multi/Many-Core Systems: Survey of Current and Emerging Trends," *ACM/EDAC/IEEE Des. Autom. Conf.*, Austin, TX, USA, June 2013, pp. 1–10.

[6] Y. Xie et al., "Design Space Exploration for 3D Architecture," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 2, no. 2, Apr. 2006, pp. 65–103.

[7] X. Zhou, J. Yang, and Y. Zhang, "Thermal-Aware Task Scheduling for 3D Multicore Processors," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 1, Jan. 2010, pp. 60–71.

[8] J. Kong, S.W. Chung, and K. Skadron, "Recent Thermal Management Techniques for Microprocessors," *ACM Comput. Survey*, vol. 44, no. 3, June 2012.

[9] W. Kim et al., "A Fully-Integrated 3-Level DC/DC Converter for Nanosecond-Scale DVS with Fast Shunt Regulation," *Int. Solid-State Circuits Conf.*, San Francisco, CA, USA, Feb. 20–24, 2011, pp. 268–270.

[10] N. Sturcken et al., "A 2.5D Integrated Voltage Regulator Using Coupled-Magnetic-Core Inductors on Silicon Interposer Delivering 10.8 A/mm^2," *Int. Solid-State Circuits Conf.*, San

Francisco, CA, USA, Feb. 19–23, 2012, pp. 400–402.
[11] M. Bao et al., "On-Line Thermal Aware Dynamic Voltage Scaling for Energy Optimization with Frequency/Temperature Dependency Consideration," *Des. Autom. Conf.*, San Francisco, CA, USA, July 26–31, 2009, pp. 490–495.
[12] K. Kang et al., "Temperature-Aware Integrated DVFS and Power Gating for Executing Tasks with Runtime Distribution," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 29, no. 9, Sept. 2010, pp. 1381–1394.
[13] S. Majzoub et al., "Energy Optimization for Many-Core Platforms: Communication and PVT Aware Voltage-Island Formation and Voltage Selection Algorithm," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 29, no. 5, May 2010, pp. 816–829.
[14] X. Zhou et al., "Thermal Management for 3D Processors via Task Scheduling," *Int. Conf. Parallel Process.*, Portland, OR, USA, Sept. 9–12, 2008, pp. 115–122.
[15] J. Meng, K. Kawakami, and A.K. Coskun, "Optimizing Energy Efficiency of 3D Multicore Systems with Stacked DRAM under Power and Thermal Constraints," *ACM/EDAC/IEEE Des. Autom. Conf.*, San Francisco, CA, USA, June 3–7, 2012, pp. 648–655.
[16] D. Cuesta et al., "Thermal-Aware Floorplanning Exploration for 3D Multi-core Architectures," *Great Lakes Symp. VLSI*, Providence, RI, USA , 2010, pp. 99–102.
[17] A.K. Coskun, A.B. Kahng, and T.S. Rosing, "Temperature- and Cost-Aware Design of 3D Multiprocessor Architecture," *Euromicro Conf. Digital Syst. Des., Archit., Methods Tools*, Patras, Greece, Aug. 27–29, 2009, pp. 183–190.
[18] S. Dighe et al., "Within-Die Variation-Aware Dynamic-Voltage-Frequency-Scaling with Optimal Core Allocation and Thread Hopping for the 80-Core TeraFLOPS Processor," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, Jan. 2011, pp. 184–193.
[19] P.D. Franzon et al., "Design for 3D Integration and Applications," *Int. Symp. Signals, Syst. Electron.*, Montreal, Canada, July 30–Aug. 2, 2007, pp. 263–266.
[20] L.P. Carloni, P. Pande, and X. Yuan, "Networks-on-Chip in Emerging Interconnect Paradigms: Advantages and Challenges," *Int. Symp. Netw.-on-Chip*, San Diego, CA, USA, May 10–13, 2009, pp. 93–102.
[21] K. Skadron et al., "Temperature-Aware Microarchitecture," *Int. Symp. Comput. Archit.*, San Diego, CA, USA, June 9–11, 2003, pp. 2–13.
[22] C.H. Chao et al., "Traffic- and Thermal-Aware Run-Time Thermal Management Scheme for 3D Noc Systems," *Int. Symp. Netw.-on-Chip*, Grenoble, France, May 3–6, 2010, pp. 223–230.
[23] T. Tobita and H. Kasahara, "A Standard Task Graph Set for Fair Evaluation of Multiprocessor Scheduling Algorithms," *J. Scheduling*, vol. 5, no. 5, 2002, pp. 379–394.

Hyejeong Hong received her BS and PhD degrees in electrical and electronic engineering from Yonsei University, Seoul, Rep. of Korea, in 2006 and 2014, respectively. She is currently working for Samsung Electronics, Hwaseong, Rep. of Korea. Her research interests include power and thermal management of multicore processors and 3D multicore design.

Jaeil Lim received his BS degree in electrical and electronic engineering from Yonsei University, Seoul, Rep. of Korea, in 2010. He is currently pursuing his PhD degree in electrical and electronic engineering from the Department of Electrical and Electronic Engineering, Yonsei University. His current research interests include DRAM refresh management, low power design, reliability, and very-large-scale integration design.

Hyunyul Lim received his BS degree in electrical and electronic engineering from Yonsei University, Seoul, Rep. of Korea, in 2013. He is currently working toward his combined MS and PhD degree in electrical and electronic engineering from Yonsei University. His research interests include scan testing and low-power scan testing.

Sungho Kang received his BS degree in control and instrumentation engineering from Seoul National University, Seoul, Rep. of Korea, in 1986 and his MS and PhD degrees in electrical and computer engineering from the University of Texas at Austin, TX, USA, in 1988 and 1992, respectively. He was a research scientist with the Schlumberger Laboratory for Computer Science, Schlumberger Inc., Austin, Texas, USA and a senior staff engineer with Semiconductor Systems Design Technology, Motorola Inc., Austin, Texas, USA. Since 1994, he has been a professor with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Rep. of Korea. His main research interests include VLSI/SOC design and testing; design for testability; and design for manufacturability.