Power Supply Noise-Aware Workload Assignments for Homogeneous 3D MPSoCs with Thermal Consideration

Yuanqing Cheng, Aida Todri-Sanjial, Alberto Bosio, Luigi Dilillo, Patrick Girard, Arnaud Virazel
LIRMM – University of Montpellier 2 / CNRS, Montpellier, France
{yuanqing.cheng, aida.todri, bosio, dillio, girard, virazel}@lirmm.fr

Abstract—In order to improve performance and reduce cost, multi-processor system on chip (MPSoC) is increasingly becoming attractive. At the same time, 3D integration emerges as a promising technology for high density integration. 3D homogeneous MPSoCs combine the benefits of both. However, high current demand and large on-chip switching activity variations introduce severe power supply noises (PSN) for 3D MPSoCs, which can increase critical path delay, and degrade chip performance and reliability. Meanwhile, thermal gradient should also be considered for 3D MPSoCs to avoid hot spots. In the paper, we investigate the PSN effects of different workloads and propose an effective PSN estimation method. Then, a heuristic workload assignment algorithm is proposed to suppress PSN under the given thermal constraint. The experimental results show that PSNs can be reduced significantly compared with thermal-balanced workload assignment scheme, and the system performance can be improved as well.

I. Introduction

Due to aggressive technology scaling, billions of transistors can be fabricated on a single die. However, interconnects on chip can not scale very well with transistors, and it is difficult to increase frequency further when processor operating frequency approaches several giga Hertz. Moreover, power consumption also becomes a bottleneck of processor design. As a result, single processor can not afford to increasing performance and functionality requirements. Multiprocessor System-on-Chip (MPSoC), which integrates IP module, FPGA, hardware accelerator and processors together, provides a cost-effective way to tackle this problem [1]. At the same time, three dimensional integration technology emerges to reduce interconnect delay and power consumption [2]. Combining both technologies can provide more functionalities and higher performance within given power budget. In addition, 3D MPSoC can integrate modules fabricated by disparate processes effectively. Therefore, 3D MPSoCs draw much attention from both academia and industrial fields.

3D MPSoCs, however, also bring several new challenges. Among others, power supply noise (PSN) is a big concern threatening signal integrity, performance and reliability, especially for shared power delivery networks [3]. Compared with 2D counterparts, 3D MPSoC power supply current becomes much larger, and causes more severe IR drop. Moreover, due to power/ground TSV parasitics, power supply noise manifests large variations among different tiers, and can propagate from one tier to other ones through TSVs [4]. These distinguish characteristics require further investigations to suppress PSN magnitudes for 3D MPSoCs. On the other hand, thermal dissipation has become a prominent issue for 3D MPSoCs. Stacking structure prevents effective heat removal of bottom tiers from the heat sink. High temperature can degrade system performance, cause thermal runaway, and increase package cost [5].

There are many existing research works involving suppressing PSN and thermal optimization for both 2D and 3D ICs. Todri et al. investigated PSN interactions for 2D CMPs and proposed some useful workload assignment guidelines [3]. Gupta et al. explored the voltage variation impact on chip multiprocessors using a distributed power delivery network model [6]. Xu et al. claimed that PSN can cause severe variations of skitter of clock networks and degrade system performance dramatically [7]. Lung et al. explored the online task scheduling technique to maximize system throughput with the thermal constraint for 3D MPSoCs [8]. Coskun et al. proposed both static and dynamic task scheduling schemes to reduce thermal gradient and thermal cycles. Zhou et al. takes advantage of strong thermal correlations among vertical tiers, and proposed a thermal-aware task scheduling algorithm for 3D CMPs, which can reduce both peak temperature and occurrences of thermal emergency [9].

However, these works are either on purely circuit level or high system level, and can not fill the gap between different design levels and capture the close relations among power supply noise, thermal gradient and workload characteristics. In this paper, we integrate circuit-level PSN estimation, system level thermal evaluation and workload assignment together effectively. Our contributions are listed as follows:

- To the best of our knowledge, it is the first work to
PDNs residing on different tiers are connected by TSVs as illustrated in Fig. 1. Both decoupling and intrinsic capacitances are modeled as a lump decap in parallel with a current source representing the workload. Workload switching activities are represented by triangular waveforms as in [12], which can be characterized by several shape parameters, i.e., peak current, leakage, delay time, rise/fall time and period. The workload current waveform is shown in Fig. 2 along with the TSV model. The physical parameters of the PDN model are listed in Table I, and the workload current waveform can be derived by architecture level power simulation tools as done in [6].

As stated in [13], critical path delay and system performance are affected by average PSN instead of peak PSN. Thus, we calculate average voltage drop as follows [14]:

\[
V_{\text{pnoise}} = \int_{t_s}^{t_e} \max\{V_{dd} - V_p(t), 0\} dt / t_{\text{switching}}
\]

\[
V_{\text{gnoise}} = \int_{t_s}^{t_e} \max\{V_{ss}(t), 0\} dt / t_{\text{switching}}
\]

\[
V_{\text{noise}} = V_{\text{pnoise}} + V_{\text{gnoise}}
\]

where \(t_s/t_e\) denotes the start/end of switching window. \(V_{dd}\) is the normal power supply voltage. \(t_{\text{switching}}\) denotes the length of switching window. \(V_p(t)\) and \(V_{ss}(t)\) capture the power supply variation and ground bounce respectively.

### B. Workload performance modeling

We use the formula proposed by [13] to model the relationship between average PSN and critical path delay, i.e.,

\[
\frac{D}{D_0} = 1 - k_1 \frac{\Delta V}{V_{dd} - V_t} + k_2 \left( \frac{\Delta V}{V_{dd} - V_t} \right)^2
\]

where \(V_{dd}\) is nominal supply voltage. \(V_t\) is the transistor threshold voltage. \(k_1\) and \(k_2\) are process dependent constants. \(D_0\) is the ideal critical path delay. Using Eq. (4), we can derive the clock frequency taking PSN into account. Then, MIPS (million instructions per second) is used to evaluate the system performance:

\[
MIPS_i = \frac{IPC_i \times f_i}{10^6}
\]
\[ Performance = \sum_{i=1}^{p} MIPS_i \]  

In the above formula, \( f_i \) is the running frequency of workload \( i \) considering PSN effect. \( IPC_i \) is instructions per cycle of workload \( i \), and \( p \) is the number of workloads.

C. Thermal evaluation

Thermal analysis generally takes advantage of duality between thermal and electrical properties [8]. Chip temperature can be derived by solving differential thermal equations. However, it is too time-consuming to be applied for iterative workload assignment optimization. In this paper, we adopt the \textit{stack power gradient} concept proposed by [9] to estimate thermal gradient across the chip. Assume that the MPSoC has \( t \) tiers, and each tier has \( m \times n \) cores. To compute stack power gradient, we need to obtain stack power firstly as follows:

\[
P_{i,j}^{stack} = \sum_{k=1}^{t} P_{k,i,j}
\]

\[ 1 \leq i \leq m, 1 \leq j \leq n \]

Then, the stack power gradient of the chip can be calculated as:

\[
P_{Gradient} = \max \{|P_{i,j}^{stack} - P_{i',j'}^{stack}|\}
\]

\[ 1 \leq i, i' \leq m, 1 \leq j, j' \leq n, (i,j) \neq (i',j') \]

In the paper, we will use this metric as the thermal constraint of workload assignment problem.

III. Motivation

As heat sink is often attached to the top tier via thermal interface material, workloads should be intuitively allocated to the upper tier from the heat removal perspective. On the other hand, package is attached to the bottom tier through TSVs. As a result, upper tiers may suffer larger PSNs, and may not be optimal positions for workload assignment from the PSN perspective. We use an illustrative example to show that with the same stack power gradient, PSN can be suppressed significantly by adjusting workload positions.

In the example, we assume the 3D MPSoC has 2 layers. There are 2 × 2 cores on each layer. The two layers are integrated using face-back bonding technique. The PDN model used for PSN estimation is presented in Section II. There are two types of workloads waiting for assignment, and their current profiles are: workload type I: \{0.15A, 1.5A, 75ps, 75ps, 75ps, 300ps, 0.2W\}, workload type II: \{0.12A, 1.2A, 325ps, 325ps, 325ps, 1.3ns, 0.1W\}. The meanings of the first six parameters are described in Section II. The last parameter denotes the workload’s average power consumption. Assume that the workload set consists of 4 type I and 2 type II workloads. In this paper, we ignore power consumptions and time overheads of data communications among running workloads because data transfer can be done very efficiently by TSVs.

First of all, we perform workload assignment using scheme 1 shown in the left part of Fig.3. If we exchange task 2 and 5, and task 3 and 6, the stack power gradient remains the same but the power supply noise reduces as shown in Fig. 3(b). Since workloads with high activities are assigned near the package, the PSNs are reduced. Then, we use Hotspot [15] to evaluate the MPSoC peak temperature. The simulations results show that peak temperature of scheme 1 is 319.42 while that of scheme 2 is 319.62, which are almost the same. This example indicates PSN can be reduced significantly without affecting thermal dissipation through deliberate task assignment strategy. It also validates the effectiveness of taking stack power gradient as the thermal evaluation metric.

IV. Problem formulation

Given a 3D MPSoC architecture \( \textbf{Ar}(t, m, n) \), which has \( t \) tiers. Each tier has \( m \times n \) cores. There are \( p \) workloads waiting for assignment. Every workload is characterized by a set of parameters, i.e., \{\( L, P, D, R, F, T, W \)\}. The objective is to assign \( p \) workloads to \( \textbf{Ar} \) such that their performance can be maximized under the given thermal constraint, i.e.,

\[
\text{Maximize : } \text{Performance} = \sum_{i=1}^{p} MIPS_i
\]

s.t.

\[
P_{Gradient}(\textbf{Ar}) \leq C
\]
where formula (12) specifies stack power gradient can not exceed the given threshold value $C$ to avoid local hot spots. Eq. (11), (4) and (5) indicate that to calculate performance, we must obtain the PSN magnitude firstly. Although PSN analysis can be performed by HSPICE simulation or other power grid analysis methods as in [16], the solving procedure is very time-consuming due to the large scale of the power grids. In the next section, we propose an effective and efficient method to estimate PSN impacts through investigating the PSN characteristics of different type of workloads.

V. PROPOSED METHOD

In the paper, we assume a $3 \times 3 \times 3$ 3D homogeneous MPSoC platform, and there are three type workloads for assignment, i.e., high, middle and low activity workloads. Their current profiles are listed as follows: High Activity: \{0.1A, 1A, 50ps, 50ps, 50ps, 200ps, 2.09W\}, Middle Activity: \{0.07A, 0.7A, 250ps, 250ps, 250ps, 1ns, 1.375W\}, Low Activity: \{0.03A, 0.3A, 1ns, 1ns, 1ns, 4ns, 0.55W\}. The idle power consumption is assumed to be 0.1W. In the rest of the section, we introduce our PSN estimation method at first. Then, we propose a heuristic algorithm to optimize PSN under the thermal constraint.

A. PSN estimation method

Our PSN estimation consists of two stages. First, we perform off-line pre-characterization for each type of workload using HSPICE simulation. In the first stage, to account for the PSN impact of the idle core, we obtain the PSN distribution when no workload is assigned, and denote it as $PSN_{idle}$. Then, we assign a specific type of workload individually to a specific position each time. As shown in Fig. 1, due to symmetry, there are three distinct positions within one tier (i.e., position 1, 2, and 3). The PSN increase due to the workload can be calculated as,

$$\Delta PSN^{k,i,j} = PSN^{k,i,j} - PSN_{idle}$$

where the triple $k$, $i$, $j$ specifies the core position within 3D MPSoC, $c$ denotes the type of workload, $t$ is the number of tiers, $m$ is the number of rows of cores within one tier, and $n$ is the number of columns of cores within one tier. Using Eq. (13), we can obtain the PSN magnitude increase caused by a single workload. The pre-characterization results are stored in the MPSoC memory as a table for reference during online task assignment.

During online task assignment stage, we estimate the PSN distributions induced by multiple running workloads by the following formula:

$$PSN^{(k,i,j)} = PSN_{idle} + \sum_{t=1}^{p} \Delta PSN^{k,i,j}$$

The equation implies that the PSN impacts caused by multiple workloads are the sum of the PSN impacts when no workload is assigned and the PSN impact increase caused by each single workload. Note that our method is different from the common superposition rule in electronic circuits. We just use the formula to estimate the relative PSN impacts caused by multiple workloads, and do not attempt to obtain the accurate absolute value of PSN magnitudes.

To validate the correctness of our method, we plot the PSN distributions predicted by our method and HSPICE simulations respectively in Fig. 4. As shown in the figure, the left part shows the workload assignment scenarios, and the right part shows the PSN distributions corresponding to the dashed line rectangle area. The PSN distributions on different tiers are derived by our estimation method and HSPICE respectively. It shows that although our method has some errors on predicting the PSN magnitudes, it can predict the hot spot area correctly from the power supply noise perspective, and provide enough guidelines for our online task assignment algorithm. We simulate different workload assignment scenarios and get the similar results. The prominent advantage of our method is that it only involves very simple calculations based on the pre-characterization results, and PSN impact evaluation can be performed very efficiently for online workload assignment. The detailed algorithm is described next.

B. Proposed Workload Assignment Algorithms

As workload assignment problem is well known as a NP-hard problem [17], we propose a simulated annealing (SA) based heuristic algorithm to attack it. The inputs of the algorithm are the number of cores($n$), initial assignment ($S^*$), PSN impact of each kind of workload derived from subsection A ($PSN_{Table}$), iteration count set for simulated annealing procedure($IterCnt$), the given stack power gradient constraint ($P$) and instruction per cycle (IPC) value of each workload ($IPC$). The pseudo-code of
the algorithm is shown in Algorithm 1. Line 2-6 is the initialization of the algorithm, including computing performance corresponding to the initial assignment (based on Eq. (5) and Eq. (6)) and initialization of simulated annealing parameters. Then the simulated annealing procedure is invoked, which is based on the algorithm proposed by [18]. It exchanges the workload positions in the neighborhood and evaluate the performance corresponding to the exchange to find the optimal solution. During the procedure, the power gradient constraint is checked consistently. If the new assignment violate the constraint, it is dropped. At last, the obtained solution is further improved by TABOO search technique [18].

The time complexity of the algorithm is mainly determined by simulated annealing procedure, which is \( O(IterCnt \times InnerCnt \times n^2) \). IterCnt, InnerCnt are outer and inner iterative counts of SA algorithm respectively. \( n \) is the number of cores of the 3D MPSoC.

### VI. Experimental Results

The simulation platform used in the experiments is a \( 3 \times 3 \times 3 \) homogeneous 3D MPSoCs. The workload PSN distributions are obtained by the PDN model described in Section II. Our workload assignment algorithm is implemented by C++ and runs on a machine with Intel 2.60G dual core and 4GB memory. Our simulation results are compared with thermal-balanced algorithm proposed by [9]. This algorithm can obtain the workload assignment with the minimum stack power gradient. To make the paper self-contained, we briefly describe its main idea. First, all workloads are sorted by their power consumptions in descending order. Each time, a workload is selected and assigned to the stack with the minimum power. Within each stack, the workload with high power consumption is assigned to the position near the heat sink. This procedure is repeated until all workloads are assigned. In the rest of paper, we call it thermal-balanced algorithm. In the comparisons, we firstly apply thermal-balanced algorithm to obtain the minimum stack power gradient and take it as the stack power gradient constraint. Then we apply our proposed method to optimize the PSN and system performance.

For simplicity, we assume the IPC of each type workload is 1. The performance degradation due to PSN is derived by Eq. 4. If the PSN reduces the high activity frequency by \( \alpha (\alpha > 1) \), then it will decrease the frequencies of low and middle workloads by \( (\alpha + 4)/5 \) and \( (\alpha + 19)/20 \) respectively.

#### A. Case I: 3 high, 3 middle and 3 low activity workloads

In case I, we assume the workloads set consisting of 3 high, 3 middle and 3 low activity workloads. The minimum stack power gradient is 1.54W, which is set as the stack power gradient constraint. Fig. 5 plots the PSN comparisons of all workloads when use different workload assignment schemes. As the figure shown, PSNs introduced by thermal-balanced algorithm are larger for almost all workloads compared with results of our method, especially for high activity workloads, which have the largest effects on system performance. On average, our method can reduce PSN by 12.42% compared to the thermal-balanced algorithm. For high activity workloads, our method can reduce 19.63% PSN on average. The left two bars of Fig. 6 shows the system performance (MIPS) comparisons considering PSN. Our method can increase system performance by 262 MIPS.
B. Case II: 5 high, 2 middle and 2 low activity workloads

In this scenario, we assume there are 5 high, 2 middle and 2 low activity workloads to be assigned to 3D MPSoCs. The minimum stack power gradient is still 1.54W. The PSN comparisons of all workloads are plotted in Fig. 7. Our method can reduce PSN by 15.6% compared to thermal-balanced method on average. For high activity workload, our method can reduce PSN by 17.14%. The system performance comparisons are shown with the middle two bars in Fig. 6. As shown in the figure, our method can increase 366 MIPS compared with thermal-balanced method.

C. Case III: 8 high, 4 middle and 4 low activity workloads

In this case, we assume there are 8 high, 4 middle and 4 low activity workloads to be assigned on 3D MPSoCs. The minimum stack power gradient is 1.275W. The PSN comparisons of all workloads are plotted in Fig. 8. As the figure shown, our method can reduce PSN by 6.14% on average. For high activity workloads, our method can reduce PSN by 14.23% on average. The system performance comparisons are shown as the right two bars in Fig. 6. Although our method increases PSN of some workloads, the system performance increases 582 MIPS compared with thermal-balanced method because our method can reduce PSNs of high activity workloads effectively, which dominate the system performance.

VII. Conclusions

3D MPSoCs combine the benefits of 3D integration and many-core technologies, and are becoming more attractive. However, due to dense stacking structure and high current delivery demand, the PSN and thermal issues become big concerns. In the paper, we investigate the PSN effects induced by different workloads. Then, we propose an effective PSN-aware heuristic to suppress PSN under given thermal constraint. For $3 \times 3 \times 3$ 3D homogenous MPSoCs, the experimental results show that the proposed algorithm can reduce PSN significantly compared with thermal-balanced algorithm. Due to reduced PSNs, system performance can also be improved as well.

REFERENCES