Thermal-Aware Mapping Algorithm for Reducing Peak Temperature of an Accelerator Deployed in a 3D Stack

Mehdipour, Farhad
E-JUST Center, Kyushu University

Nunna, Krishna Chaitanya
Department of Advanced Informatics, Kyushu University

Gauthier, Lovic
Department of Advanced Informatics, Kyushu University

Inoue, Koji
Department of Advanced Informatics, Kyushu University

他

出版情報：IEEE 3DIC 2011, 2012-02. IEEE
バージョン：
権利関係：

A Thermal-Aware Mapping Algorithm for Reducing Peak Temperature of an Accelerator Deployed in a 3D Stack

Farhad Mehdipour[1], Krishna Chaitanya Nunna[2], Lovic Gauthier[2], Koji Inoue[2], Kazuaki Murakami[2]
[1] E-JUST Center, Kyushu University, JAPAN
Email: farhad@ejust.kyushu-u.ac.jp
[2] Department of Advanced Informatics, Kyushu University, JAPAN

***Abstract*- Thermal management is one of the main concerns in three-dimensional integration due to difficulty of dissipating heat through the stack of the integrated circuit. In a 3D stack involving a data-path accelerator, a base processor and memory components, peak temperature reduction is targeted in this paper. A mapping algorithm has been devised in order to distribute operations of data flow graphs evenly over the processing elements of the target accelerator in two steps involving thermal-aware partitioning of input data flow graphs, and thermal-aware mapping of the partitions onto the processing elements. The efficiency of the proposed technique in reducing peak temperature is demonstrated throughout the experiments.**

I. INTRODUCTION

Augmenting a reconfigurable data-path accelerator to a base processor and memory elements is one of the solutions to gain higher performance through running computing-intensive parts of applications on the accelerator [6][7]. Critical parts of various applications are extracted in the form of data flow graphs (DFGs) and mapped onto the accelerator architecture by means of a sophisticated mapping tool. These DFGs are run on the accelerator, while the rest of application is run on the base processor. There are some additional benefits in 3D integration of the components of this computing system, however there are some challenges as well.

Three-dimensional integration can achieve higher transistor densities and shorter interconnect lengths than two-dimensional, which can work as a platform to integrate different layers and provide good substrate isolation among them [1]. In spite of 3D integration's advantages over 2D, thermal effects are expected to be significantly higher in 3D chips due to higher power density and greater thermal resistance of the insulating dielectric, and this can cause greater degradation in device performance and chip reliability [4][5]. The thermal problem is worse in 3D integration for two main reasons: a rapid increase in power density and existing of a dielectric layer between each tier to provide insulation [3]. The thermal conductivity of the dielectric layers is very low compared to silicon and metal [8].

Leakage power also plays a dominant role in resulting these higher temperatures. Leakage power consumption and temperature influence each other: increasing temperature increases leakage and vice versa. Localized heating occurs much faster than chip-wide heating; since power dissipation is spatially non-uniform across the chip, this leads to "hot spots". These effects ultimately results in excessive heat dissipation which is one of the most critical challenges in 3D design Therefore, it is necessary to consider the solutions for the thermal issues during every stage of 3D designs, including the placement and routing stages.

In this research we intend to devise a mapping algorithm for placing data flow graphs (DFGs) on a certain accelerator architecture so that the peak temperature is reduced through evenly temperature distribution. It is assumed that the target accelerator is augmented to other components such as a general-purpose processor, memory and etc. Further it is deployed in a 3D stack along with other components in one of the planes. Uniform thermal distribution is promising for preventing hot spots, hence effective for reducing temperature over the 3D stack.

In the rest of paper, in Section 2 our motivation for pursuing this research will be given. Section 3 explains about the proposed mapping algorithm and Section 4 will highlight the results of experiments. Finally, Section 5 concludes the paper.

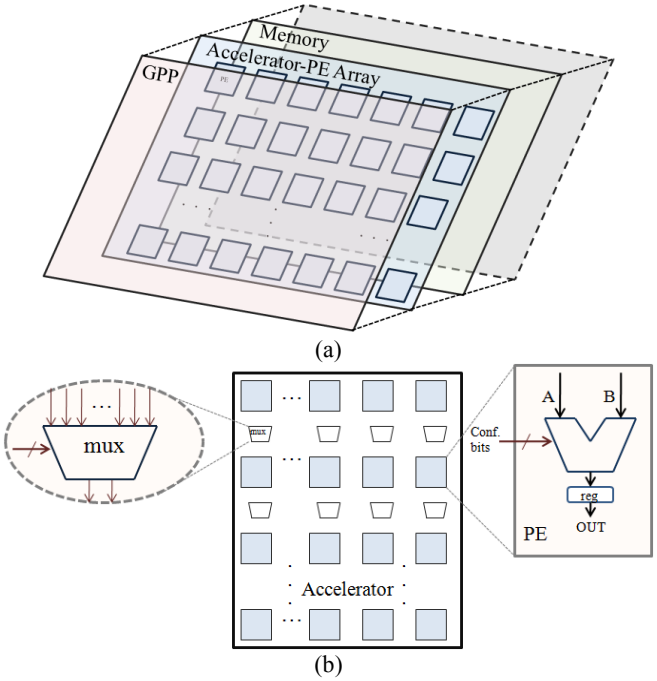


Fig. 1. (a) an integration of a DFG accelerator in a 3D stack (b) detailed architecture of the pipelined data-path accelerator

II. MOTIVATION

In this research we intend to reduce the peak temperature of an accelerator which is integrating to a base processor, memory components and etc. in a 3D stacking (Fig. 1(a)). The accelerator is a pipelined data-path processing architecture, comprising a matrix of processing elements (PEs) as well as multiplexers among them to provide unidirectional flow of data (Fig. 1(b)). *H* and *W* represent the accelerator dimensions including the number of rows (PE stripes) and columns. Multiplexers as routing resources have limited size to save the area and power, which requires mapping tool to minimize connection length between PEs, so that connections can be routed through multiplexers. Example of such architecture can be found in [7].

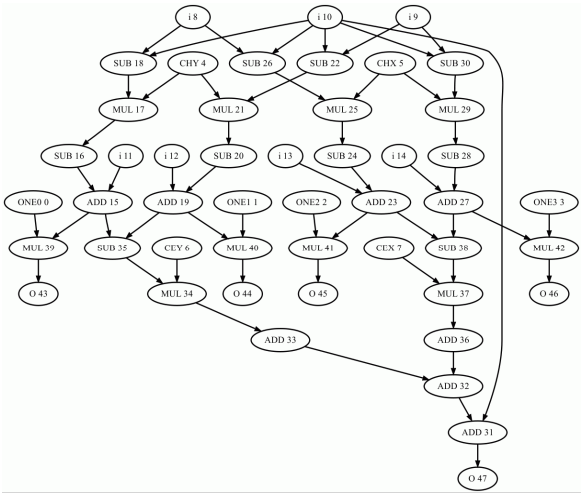
Data flow graphs (DFGs) extracted from various applications are mapped onto the accelerator for the sake of gaining higher performance. Fig. 2(b) shows the result of mapping a sample data flow graph (Vib-4x2 as displayed in Fig. 2(a)) on the PE array. Red connections denote the longest wirelengths between two consecutive PE stripes which impacts the multiplexer sizes being utilized as interconnection resources. A non-uniform distribution of DFG operations can be easily observed in the mapping result. The density of operations and interconnections in some parts of array is high, whereas there is a large number of unused resources in other parts. This may cause hot spots in condensed area of the accelerator as displayed in Fig. 2(c) which is indicating the peak temperature around 90°C and minimum temperature of 56°C for unused spaces. Evenly distribution of DFG operations over the accelerator PEs can be helpful in reducing these peak temperatures and avoiding hot spots. It should be noted that the limitation on the interconnection resources (such as availability of only limited-length unidirectional interconnections) is the main reason in constraining mapping and routing algorithms which may result in non-uniform distribution of DFG operations on the accelerator.

III. MAPPING ALGORITHM

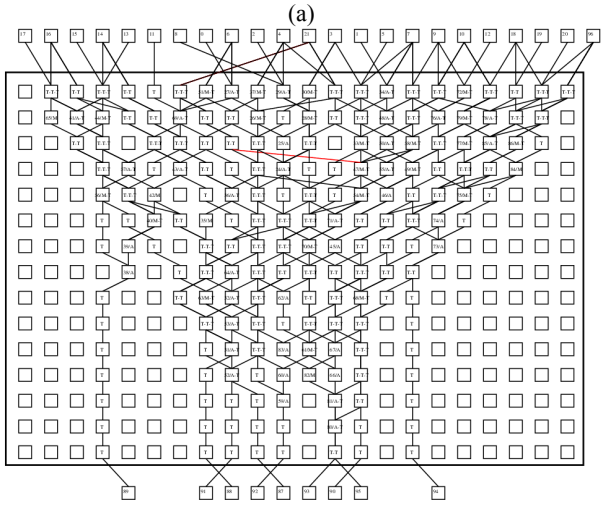
To reduce the peak temperature and avoid hot spots, we propose a DFG mapping algorithm in which there is an attempt to distribute DFG operations more uniformly over the PE array. The main intuition behind that is locating two hot blocks adjacent to each other may cause hot spots, while surrounding a hot block by several colder blocks helps in cooling down the peak temperature. The proposed mapping algorithm is a partitioning-based one and performs in two steps: (a) thermal-aware partitioning of input DFG and creating a number of partitions associated with PE stripes, and (b) placing the DFG operations of each partition onto the PEs of corresponding PE stripe considering temperature distribution.

A. *Thermal-Aware (TA) Partitioning*

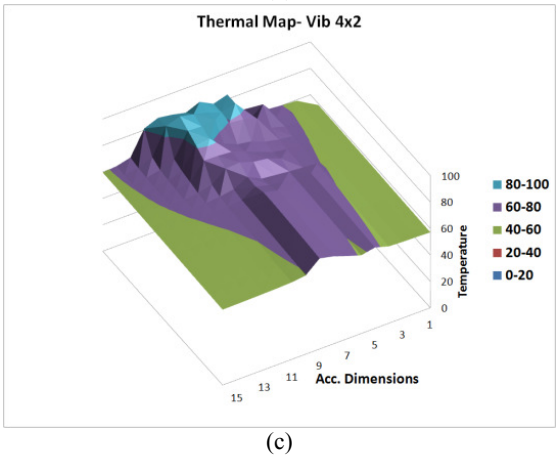
Firstly, initial partitions are created after list scheduling of DFG (by means of ASAP scheduling). The initial number of partitions is equal to the highest ASAP level of the DFG, whereas the final number of partitions is equal to *H* (the number of PE stripes). Node n_i of DFG is located in *k*-th partition, assuming that the ASAP level of the node is *k*. Extra partitions are created during the partitioning procedure, if the initial number of partitions is less than the number of available PE stripes in PE array.



(a)



(b)



(c)

Fig. 2. (a) A sample DFG, (b) its corresponding map on the PE-array with non-uniform distribution (c) thermal map of the accelerator after mapping

Afterward, a dependency partitioning graph (referred as DPG) is created so that hyper nodes in DPG are associated with initial partitions. There are two types of nodes in the DPG: ordinary nodes associating with the original ones in DFG, and hyper nodes corresponding to the partitions. There is an edge from a node to a hyper node if all its children belong to the hyper node or succeeding hyper nodes. For each edge (n_i, n_j) connecting two nodes of DFG n_i and n_j a weight value ($w_{i,j}$) is assigned which it is equal to the distance of associated hyper nodes. Also, below variables are defined and used within partitioning algorithm:

$p(n_i)$: the partition number associated with the hyper node, in which the node n_i is located.

$w(n_i, n_j)$: is the weight of a node connecting ni and nj is calculated as $p(n_j)$- $p(n_i)$- 1. This calculation is based on the impact of limitations on the interconnections.

$e(n_i)$: is the eligibility of node n_i for moving to another partition in the $e(n_i)$ vicinity, and it is calculated as: $Min\ (w(n_i, n_j))$ for any j.

$PD(i,j)$: power density of PE_j when running operation n_i.

DPG nodes are assigned with movement eligibilities which are representing the maximum distance for a DFG node that can be moved to another succeeding hyper node or partition. Constraints on available interconnection length due to limitation of multiplexer size impact the movement eligibility of DFG nodes. The more flexible interconnection resources, the more eligibility for movement can be assigned to the nodes. The first phase of algorithm starts with moving eligible nodes over the initial partitions and tries to make a balance among partitions in terms of the total power densities of partitions. Any movement making a better balance of power densities is committed and movement eligibilities of the affected nodes are updated. 0 explains partitioning algorithm. The thermal cost is calculated as in Eq. 1 which is denoting the standard deviation of power densities in partitions.

$$Cost=(\Sigma_{i=1..H}(PD(i,j)-PD_{avg})/H)^{sqrt} \quad (1)$$

while, H is the number of partitions (accelerator height) and PD_{avg} is the average power densities of partitions.

B. *Mapping Partitions on PE Stripes (TA-Mapping)*

In the second phase, each partition is mapped on corresponding PE stripe such that a uniform distribution of power densities around each PE with its adjacent neighbors is obtained and also the connection length is maintained at the allowed range. Fig. 3(b) shows that in order to place a node (number 1) at a PE strip, all possible and unoccupied locations are tried. In every try, a window around the tentatively placed node is formed and summation of power densities is tried. The size of window is up to nine PEs which includes a PE that the placed node is located on, and eight other PEs in neighborhood as well. Any PE giving the minimal lumped sum of power densities is chosen, while constraint on wirelength is also met. The idea is inspired from matrix synthesis problem in [2] which models the thermal placement and suggests algorithms to solve it. The objective function is defined as in Eq. 2.

$$\text{Min } (\sum_{i=1..nn} PD(i,j)) \quad (2)$$
$$\text{subject to: } \text{Max}(CL_i) \leq CL_{max}$$

where, $\max(CL_i)$ is the maximum wirelength between n_i and the nodes connected to it, CL_{max} is the maximum allowed wirelength, and nn is the number of nodes in n_i's vicinity representing the window size. Window size is equal to eight or five when the PEs at left or right hand sides are examined. It could be limited to only three PE when a PE at corner is tried.

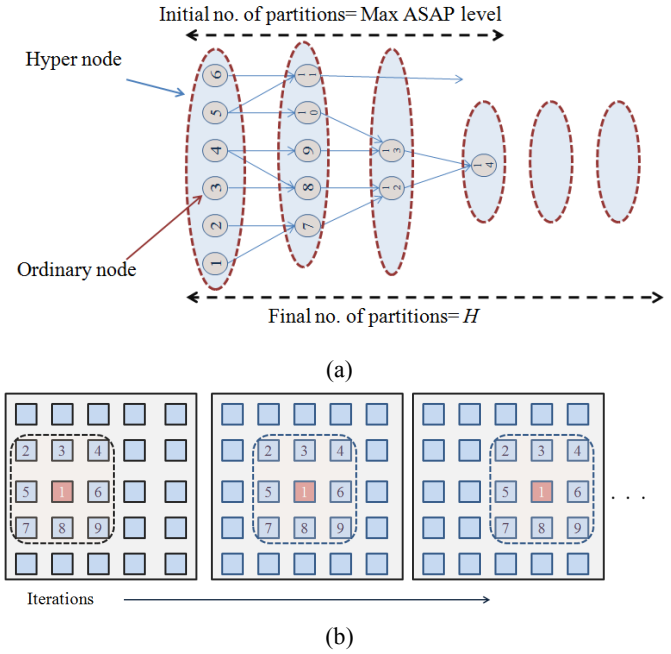


Fig. 3. The proposed algorithm runs two steps (a) firstly a DPG as well as initial partitions and then, final partitions are created, (b) later, to place a node, all PEs of a stripe are tried to find a location with minimum summation of power densities with neighbor PEs.

Thermal-aware partitioning algorithm:

1. Apply list scheduling algorithm (ASAP)on input DFG.
2. Construct initial partitions based on the ASAP levels of nodes (Assign node n_i to k-th partition assuming that the ASAP level of node n_i is k).
3. Calculate the values of $p(n_i)$ for every node, $w(n_i, n_j)$ for every edge and $e(n_i)$ the eligibility of movements for any node of DFG.
4. Construct DPG based on ASAP values.
5. For each partition p_i starting from the first partition:
   5-1. sort all nodes based on descending order of the values of eligibility of movements.
   5-2. for node $e(n_i)$ in the sorted list, move it to partition $(p(n_i)+ e(n_i))$-th succeeding partition and calculate the thermal cost.
   5-3. repeat step 5-2 for all succeeding partitions until $p(n_i)+1$ and choose partition number giving the minimum cost.
   5-4. move n_i to the partition resulting in minimum thermal cost and update all affected $p(n_i)$, $e(n_i)$ and $w(n_i, n_j)$ values.
   5-5. repeat steps 5-2 to 5-4 for all nodes in the partition
6. repeat step 5 until no further improvement is achieved.

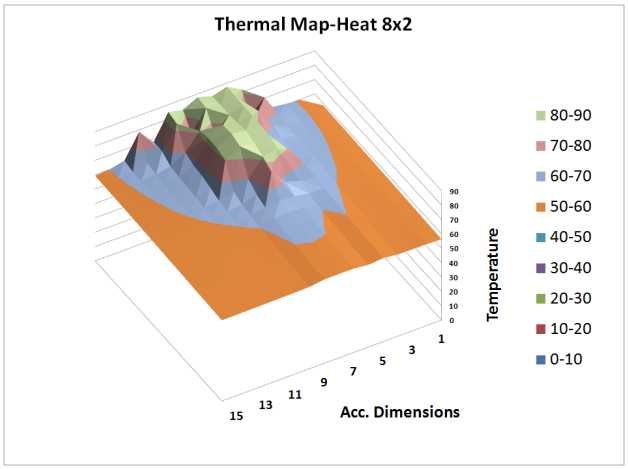
Fig. 4. Thermal-aware DFG partitioning algorithm

IV. Experiment Results

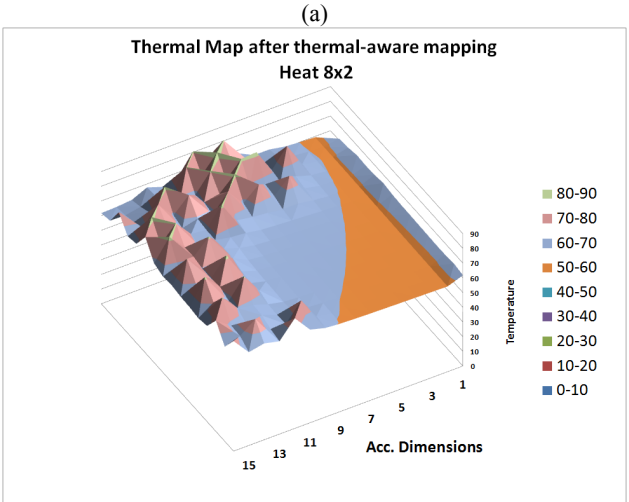
The introduced algorithm has been evaluated using HotSpot tool [3] for various applications. Accelerator dimensions including *W* and *H* are 16 and 16, respectively. A number of DFGs were attempted within experiments (from [6]), their specifications, initial number of partitions, the number of partitions after thermal-aware partitioning, and peak temperatures without and with thermal-aware mapping are displayed in Table 1. Obviously, since the height of accelerator which is indicting number of the PE stripes as well is equal to 16, final number of partitions for all attempted DFGs is the same.

Table 1. Attempted DFGs and the result of partitioning

DFG	Heat-8x2	Vib-4x2	Vib-8x2	Poi-3x3
no of nodes	32	8/4	72	33
no. of inps/outs	8/4	24	16/12	18/1
initial no. of partitions	6	8	8	10
no. of partitions after TA-partitioning	16	16	16	16
peak temperature (ºC)	88.4	89.3	104.1	83.5
peak temp. after TA-mapping (ºC)	82.0	84.9	98.3	79.2



(a)



(b)

Fig. 5. Thermal maps of the accelerator after and before applying the thermal-aware mapping algorithm

According to Table 1, the results of experiment show temperature reduction after applying the proposed mapping technique (referred as TA-mapping) compared with one without thermal considerations. Fig. 4 displays thermal maps for one of the DFGs (Heat-8x2) before and after applying the thermal-aware mapping algorithm. The distribution is more uniform in the latter case as DFG nodes have moved from upper PE stripes to lower ones, therefore peak temperature decreases from 88.4oC to 82oC without violating constraints on the available interconnection length.

V. Conclusion

A thermal-aware partitioning algorithm is proposed to divide input DFG into a number of evenly power density distributed partitions and then placing them on the target accelerator. In the target architecture, a data-path accelerator is deployed in a 3D stack along with the other components. The proposed algorithm is effective in reducing the peak temperature of the accelerator. We intend to expand this algorithm for a 3D mutli-tier accelerator and also, take the effect of interconnections into account in thermal evaluations.

Acknowledgements

This research was supported in part by New Energy and Industrial Technology Development Organization and Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Corporation (JST).

References

[1] K. Banerjee, S. J. Souri, P. Kapur and K. C. Saraswat, "3-D ICs: A novel chip design for improving deep submicrometer interconnect performance and system-on-chip integration," Proc. of IEEE, 89(5), pp. 602-633, May 2001.

[2] C. N. Chu and D. F. Wong. Matrix synthesis approach to thermal placement. Proc. Int. Sym. on Physical Design, pp. 163-168, 1997.

[3] HotSpot, http://lava.cs.virginia.edu/HotSpot.

[4] S. Im and K. Banerjee, "Full chip thermal analysis of planar (2-D) and vertically integrated (3-D) high performance ICs," in IEDM Tech. Dig., pp. 727–730, 2000.

[5] S. A. Kuhn, M. B. Kleiner, P. Ramm, andW.Weber, "Thermal analysis of vertically integrated circuits," in IEDM Tech. Dig., pp. 487–490, 1995.

[6] F. Mehdipour, H. Honda, K. Inoue, H. Kataoka, and K. Murakami, A design scheme for a reconfigurable accelerator implemented by single-flux quantum circuits, J. Syst. Architect. Embedded Systems Design 57(1), pp.169-179, Jan. 2011.

[7] N. Takagi, K. Murakami, A. Fujimaki, N. Yoshikawa, K. Inoue, H. Honda, Proposal of a desk-Side Supercomputer with reconfigurable data-paths using rapid single flux quantum circuits, IEICE Trans. Elec. E91-C(3), pp. 350–355, 2008.

[8] P. Wilkerson, M. Furmanczyk, and M. Turowski, Compact thermal modeling analysis for 3D integrated circuits, 11th International Conference Mixed Design of Integrated Circuits and Systems, pp.277-282, 2004.