Thermal optimal task allocation algorithm for multi-core 3D IC with interlayer cooling system

Wei Guo^{1a)}, Minxuan Zhang^{1,2}, Chaoyun Yao¹, Peng Li¹,
Chaochao Feng¹, and Hongwei Zhou¹

¹ College of Computer, National University of Defense Technology,
Changsha, 410073, P.R. China

² National Laboratory for Parallel and Distributed Processing,
National University of Defense Technology, ChangSha 410073, China

a) wineer-guowei@nudt.edu.cn

Abstract: Multi-core 3D IC is a new solution for greater quantity of cores, but with critical challenge of internal thermal. The interlayer cooling system is introduced for this problem, and expands the design space of task allocation. This work proposes a thermal optimal task allocation algorithm for multi-core 3D IC with interlayer cooling system, integrated with simulated annealing method. The results show that the maximal temperature and temperature gradient reduced by 19.4 °C and 5.1 °C respectively, compared to the randomized policy, and obtain better balance between normal and extreme cases than traditional ranking policy, in a design of 3 active device layers.

Keywords: 3D IC, interlayer cooling, task allocation, simulated annealing, thermal

Classification: Integrated circuits

References

- [1] International technology roadmap for semiconductors (ITRS) (2013 Ed.). <http://www.itrs.net/Links/2013ITRS/Home2013.htm>.
- [2] M. Koyanagi: IEICE Electron. Express **12** (2015) 20152001. DOI:10.1587/elex.12.20152001
- [3] D. Cuesta, J. L. Risco-Martin and J. L. Ayala: Microprocess. Microsyst. **36** (2012) 344. DOI:10.1016/j.micpro.2012.02.012
- [4] B. Su, J. L. Gu, L. Shen, W. Huang, J. L. Greathouse and Z. Y. Wang: IEEE/ACM MICRO (2014) 445. DOI:10.1109/MICRO.2014.17
- [5] J. L. Ayala, A. Sridhar and D. Cuesta: Integration **43** (2010) 327. DOI:10.1016/j.vlsi.2010.06.002
- [6] R. Abdollahi and A. Jahanian: IEICE Electron. Express **9** (2012) 1849. DOI:10.1587/elex.9.1849
- [7] S. X. Qu, M. X. Zhang, G. H. Liu and T. Liu: J. Central South Univ. **19** (2012) 193. DOI:10.1007/s11771-012-0991-8
- [8] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunswiler and D. Atienza: IEEE/ACM ICCAD (2010) 463. DOI:10.1109/ICCAD.2010.5653749

- [9] T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, W. Cesar, Y. Leblebici, B. Wunderle and H. Reichl: IEEE ITherm (2010) 1. DOI:10.1109/ITHERM.2010.5501254
- [10] J. J. Guo, K. Dai and Z. Y. Wang: IEEE HPCC (2006) 300. DOI:10.1007/11847366_31
- [11] M. Poongothai, A. Rajeswari and V. Kanishkan: IEICE Electron. Express **11** (2014) 20130975. DOI:10.1587/elex.11.20130975
- [12] A. A. El-Saleh, M. Ismail, R. Vikneesh, C. C. Mark and M. L. Chan: IEICE Electron. Express **6** (2009) 1219. DOI:10.1587/elex.6.1219
- [13] E. K. Ardestani and J. Renau: ACM HPCA (2013) 448. DOI:10.1109/HPCA.2013.6522340

1 Introduction

According to the ITRS roadmap [1], the process feature size is becoming smaller but closer to the physical extreme limit, which is the most important signal of the develop of integrated circuit technology. In traditional 2D processor, more circuits means larger area and longer wires, but the limit of long wire delay has been seem as an unsolved challenge [2]. To keep improving the performance of micro-processors and extend the capabilities of a single processor, three-dimensional integrated circuits (3D IC) is proposed [3].

The power consumption in processor will be translated into internal temperature finally. When the temperature increases, the mobility of carriers decays and the resistance of metal wires increases, which both result the descending of the speed of processor [4].

Moreover, the reliability will also be influenced by too high temperature, which has more direct relationship with temperature. The time to failure of integrated circuits can be predicted as $e^{Ea/kT}$, where T is the absolute temperature, k is the Boltzmann's constant, and Ea is the activation energy of the failure mechanism being accelerated by the increased temperature [5]. Furthermore, the electro-migration of metal wires is influenced by the temperature gradient, can also reduce the reliability of integrated circuits.

In 3D IC, the power density increases significantly compared to 2D IC for the reduced distances between units, the thermal problem becomes more serious. The thermal conductivity of bond and insulation layers between the silicon layers is too low compared to silicon and metal [6]. The external heat removal solutions, like heat sink with fans or even fluid cooling, have little influence on the internal temperature gradient [7].

Interlayer cooling can remove the internal heat flux of processor effectively, and is the only feasible heat removal solution for multiple active layers in 3D ICs so far [8]. In this scheme, the coolant pass through the fluid micro-channels between the active layers, and remove the internal heat flux as Fig. 1 shows.

This scheme can control the maximal junction temperature increase is 54.7 °C at a 1 bar pressure drop with water as coolant, compared to the unacceptable 203 °C with traditional back-side heat removal scheme [9]. In Section 4, the detail construction of interlayer cooling system will be proposed.

(a) Completed circuit board (b) Cross-section view of stacking sequence with fluid connection

Fig. 1. Interlayer cooling system

Traditional task allocation algorithms decide the binding relationship between programs and cores, such as randomized and ranking policy, which have little sense about thermal problem. But different programs have different features of power consumption, while different cores have different features of thermal removal [10]. The results of task allocation will influence the distribution of internal temperature, especial hot-spots [11].

This work proposes a thermal optimal task allocation algorithm for multi-core 3D IC with interlayer cooling system, integrated with simulated annealing method, considering to reduce the temperature of multi-core 3D IC according to the features of different programs and cores, and obtain a balance between normal and extreme cases in the design with 3 active device layers.

2 Analysis of benchmarks

The power consumption of programs changed during the run time, and the thermal features of cores which programs bound changes accordingly. In multi-core 3D IC, the temperature of one core is not only determined by the program bound, but also influenced by the neighbour cores and programs, including horizontal and vertical direction [12]. This section will make an analysis of the power consumption of different benchmarks, which influence the distribution of temperature in multi-core 3D IC.

The power information of benchmarks used in this work is from ESESC, a fast microprocessor architecture simulator [13]. The benchmarks set include SPEC-2000 and SPEC-2006, and these benchmarks can be classified into 4 categories as follow:

- **Sustained stable:** Power consumption during whole run time keeps around a value within little variety range, like 300.twolf and 473.astar;
- **Burst hotspot:** Power consumption during most of run time changes within little range, but becomes high significantly at some moments, like 403.gcc and 462.libquantum;
- **Periodical stable:** Power consumption keeps around a value during a period of run time, then change to another value, like 254.gap and 410.bwaves;
- **Wide-range variation:** Power consumption during whole run time changed with huge variety range, like 179.art and 401.bzip2.

Fig. 2. Power consumption of benchmarks

Fig. 2 shows power trace of 8 benchmarks selected from the whole set according the categories mentioned above, where the green line labeled with “99%” means the maximal power value during most of whole run time (99% here).

When the benchmarks are allocated by task allocation algorithm, the thermal feature of multi-core 3D IC is determined by the power consumption of the benchmarks in each core, the cooling feature of each core and the cross-influence between neighbour cores [2].

In normal case, most of the benchmarks are running below the “99%” value simultaneously at any moment. However, it’s possible to face an extremal case

that all benchmarks get the maximal power value simultaneously, the whole processor is no doubt to catch the thermal peak, which is seriously harmful for the processor.

The task allocation algorithm must consider the cooling feature of each core and the binding relationship between benchmarks and cores. Meanwhile, the gaps between the maximal and “99%” power of benchmarks are usually various, make it necessary to consider both the normal and extremal cases simultaneously.

3 Description of algorithm

Randomized policy is widely used in task allocation algorithm for its simple construction, the benchmarks are allocated randomly in this strategy. Another popular strategy is ranking policy, all benchmarks are sorted by power feature and the cores are sorted by cooling feature, then the hottest benchmark is allocated in the coolest core, the coolest benchmark is allocated in the hottest core and so on. In practice, the reference standard is usually the maximal power or the “99%” power as mentioned above [7]. Both policies above can’t reach the best thermal optimal, no matter under the normal or extremal case.

The thermal optimal task allocation algorithm proposed in this work is designed with 2 steps:

- **Step 1:** Generate an initial result by randomized policy;
- **Step 2:** Obtain the optimized result using simulated annealing method.

The result achieved by step 1 has some limits, and the design space has not been explored sufficiently.

Table I. Pseudo-code of simulated annealing in step 2

```

 $k \Leftarrow 0, S \Leftarrow S_0, T \Leftarrow T_0$ 
 $Cost \Leftarrow Cost\_calc(S)$ 
while( $k < k_{max}$  and  $T > T_{cold}$ )
     $S_{new} \Leftarrow Random\_exchange(S)$ 
     $Cost_{new} \Leftarrow Cost\_calc(S_{new})$ 
    if( $Cost_{new} < Cost$  or  $Random\_frac() < e^{-(\Delta Cost)/T}$ )
         $S \Leftarrow S_{new}$ 
        if ( $Cost_{new} < Cost$ )
             $S_{best} \Leftarrow S_{new}$ 
     $T \Leftarrow T \times Ratio_{cool}, k++$ 

```

Simulated annealing method is widely used as optimization solution. It can obtain a good approximation to the global optimum result in large search space. During the searching for the better solutions, a worse one can be accepted by an acceptance probability to explore the solution space.

The pseudo-code of simulated annealing method used in step 2 is shown in Table I, S_{new} is the new solution, which is generated based on the previous accepted solution, through an exchange between 2 benchmarks selected randomly. The cost of a solution with n cores allocated with benchmarks is calculated by the equation:

$$Cost = \sum_{i=0}^{n-1} \alpha_{cooling,i} \times (\lambda_1 \cdot P_{max,i} + \lambda_2 \cdot P_{max_avg_neighbour,i} + \lambda_3 \cdot P_{99\%,i} + \lambda_4 \cdot P_{99\%_avg_neighbour,i}) \quad (1)$$

Where $\alpha_{cooling,i}$ is the cooling coefficient of core i . $P_{max,i}$ and $P_{99\%,i}$ are the maximal and “99%” power of the benchmark allocated at core i separately. Accordingly, $P_{max_avg_neighbour,i}$ and $P_{99\%_avg_neighbour,i}$ are the average values of maximal and “99%” power of the benchmarks allocated at the neighbours of core i . λ_1 , λ_2 , λ_3 and λ_4 are the corresponding weight coefficients, respectively. After the whole simulated annealing process, S_{best} refers to the best solution with minimum cost.

4 Experimental results

The designs in our work are realized in the simulation platform named 3D Interlayer Cooling Emulator (3D-ICE), which simulates the transient and steady thermal behaviour of 3D IC with interlayer cooling system, to perform thermal analysis of 3D IC in the early design stages of processor [8].

Fig. 3. Floorplan for layer_0

The design for 24 cores in our work are distributed in 3 active layers, 8 cores for each layer using same floorplan as Fig. 3 shows, and total area is $10 \times 10 \text{ mm}^2$.

The 3D stack structure for our work is shown in Fig. 1(b), the chip starts with a $10 \mu\text{m}$ thick PCB layer at bottom, followed by the first micro-channel cooling layer, which is $100 \mu\text{m}$ thick. A $50 \mu\text{m}$ thick silicon substrate layer is placed above the micro-channel, and below the bottom active device layer, which is $2 \mu\text{m}$ thick. There is a $10 \mu\text{m}$ thick BEOL (Back End Of Line) layer for bonding between the bottom active device layer and the second micro-channel cooling layer. Then, the second and third micro-channel cooling layer, silicon substrate layer, active device layer and BEOL layer are placed above. At the top of chip, there are the top micro-channel layer and a silicon layer to finish the stack.

There are 24 benchmarks used in our work, the maximal and “99%” power of each benchmark are showed in Table II.

Fig. 4 shows the temperature map of 3 layers in normal case for the allocation result obtained by randomized policy, and Fig. 5 is for extremal case, where the green arrows are the directions of coolant flow. Fig. 6 and Fig. 7 are the temper-

Table II. List of benchmarks [13]

No.	Bench name	$Power_{max}$ (W)	$Power_{99\%}$ (W)
0/12	300.twolf	13.9538	11.1037
1/13	473.astar	12.6271	9.9739
2/14	171.swim	15.1578	11.7921
3/15	175.vpr	11.8115	10.8271
4/16	453.povray	15.9494	13.8400
5/17	403.gcc	23.6230	13.7040
6/18	462.libquantum	25.2214	10.1648
7/19	433.milc	14.9669	14.7325
8/20	410.bwaves	16.1982	11.6419
9/21	254.gap	14.0311	12.4467
10/22	179.art	20.7296	18.7328
11/23	401.bzip2	18.2379	15.3972

Fig. 4. Temperature maps for layer 0–2 with randomized policy in normal case

Fig. 5. Temperature maps for layer 0–2 with randomized policy in extremal case

Fig. 6. Temperature maps for layer 0–2 with simulated annealing in normal case

Fig. 7. Temperature maps for layer 0–2 with simulated annealing in extremal case

Fig. 8. Contrast of T_{max} and $T_{gradient}$

ature maps in normal and extremal cases for the allocation result obtained by simulated annealing method. In latter result, the maximal temperatures of the whole processor are reduced by 19.4 °C in normal case and 16.4 °C in extremal case, the maximal temperature gradients are reduced by 2.3 °C in normal case and 5.1 °C in extremal case, which is showed in Fig. 8, where @Power_{99%} means in normal case and @Power_{max} means in extremal case.

Also in Fig. 8, Rank_{99%} is refer to the allocation result obtained through ranking policy with “99%” power of benchmarks as references, and Rank_{max} is refer to the allocation result obtained through ranking policy with maximal power of benchmarks as references. The first ranking result performs better in normal case, but worse in extremal case. While the second ranking result performs in the contrary. The allocation result obtained through the simulated annealing proposed in this work can gain a better balance both in normal and extremal cases.

5 Conclusion

This work proposes a thermal optimal task allocation algorithm for multi-core 3D IC with interlayer cooling system, integrated with simulated annealing method. The allocation result achieved can reduce the maximal temperature of whole processor by 19.4 °C and reduce the maximal temperature gradient by 5.1 °C, compare with the randomized policy. Meanwhile, the algorithm proposed can gain a better result both in normal and extremal cases, unlike the ranking policy.

Acknowledgements

The research is supported by National Natural Science Foundation of China with Grant No. 61303066 and 61303069.