Two-stage Thermal-Aware Scheduling of Task Graphs on 3D Multi-cores Exploiting Application and Architecture Characteristics

Zuomin Zhu[1] Vivek Chaturvedi[2] Amit Kumar Singh[3] Wei Zhang[1] Yingnan Cui[2]

[1]Hong Kong University of Science and Technology

[2]Nanyang Technological University

[3]University of Southampton

January 12, 2017

Outline

Outline

Challenges faced by 3D CMP

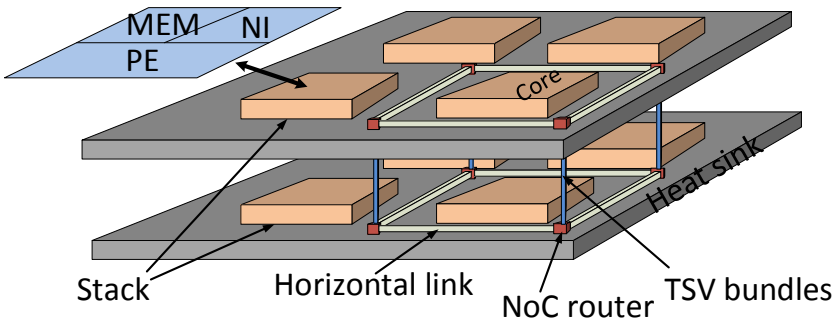

Figure 1: Two-layer 3D CMP Architecture

Benefits:

1. Multiple active layers stacked.
2. Energy efficiency and scalability.
3. Short interconnect delay.

Challenges:

1. High power density.
2. Thermal emergency.
3. Reliability threat, device aging.

How to address thermal challenges?

Thermal-aware task mapping techniques for 3D CMP

Based on thermal-aware or power-aware scheduling approaches:

Targeting independent tasks

1. Stack power balance[1]
2. Incremental update of thermal simulation[2]

Targeting task graphs

1. Swap tasks between hot and cold stacks[3]
2. Move low power tasks to top layer [4]
3. Schedule tasks in the order of priority, thermal simulations for mapping[5]

Drawbacks of the previous works

- Neglect influence of neighboring cores and leakage power on temperature[1]
- High computational complexity[2]
- Incur huge swapping overhead[4, 3]
- High thermal simulation overhead at run-time [5]

Thus we propose a novel decoupling scheduling algorithm for peak temperature minimization while optimizing the makespan of the application.

Outline

Two-stage scheduling algorithm

Decouple the optimization of makespan and temperature into two stages and optimize them separately.

Communication-aware group (CAG) stage: Design-time stage

1. Bind individual tasks to super tasks assuming different number of available cores.
2. Brute-force or genetic algorithm
3. Minimize the makespan considering the worst-case communication.

Thermal-aware scheduling (TAS) stage: Run-time stage

1. Super tasks mapped to real available cores.
2. Different strategies to handle different layers.
3. Thermal rank model & Combined power model.
4. For peak temperature minimization.

Design-time makespan optimization

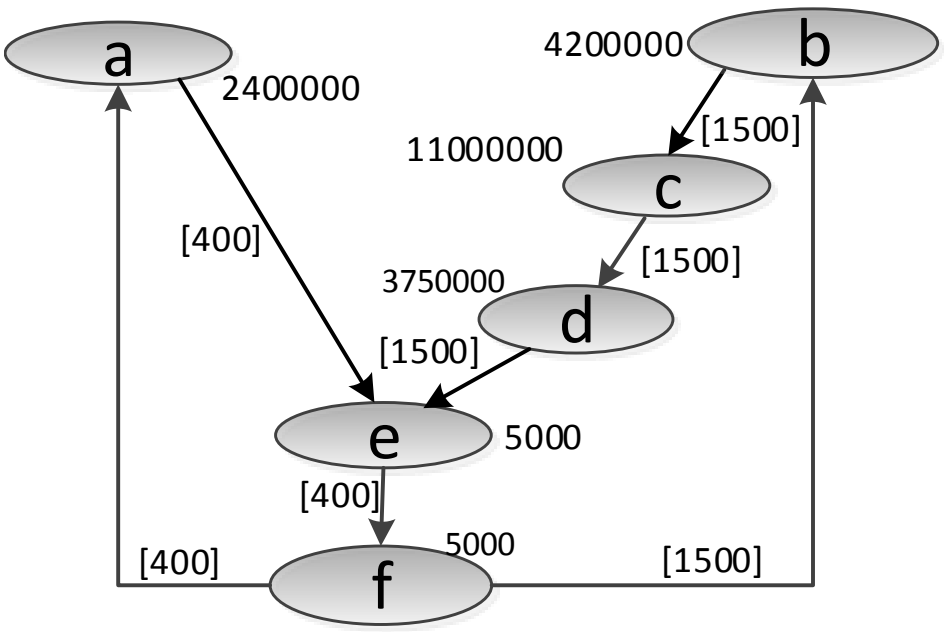

Figure 2: An example DG model: Autoindust2

- Search the best mappings utilizing different number of cores.
- Exhaustive exploration when task number is small.
- GA is applied when the task number is bigger.
- Consider the maximum hop (communication delay) between cores.

Design-time makespan optimization

Super task: a group of tasks (nodes) which are mapped to a particular core in a given scheduling interval.

#Cores	C1	C2	C3	C4	C5	C6	Period(cycles)
6	e	d	c	b	f	a	19018400
5	*	e	d	c	bf	a	19004900
4	*	*	de	c	bf	a	18991400
3	*	*	*	cde	bf	a	18977900
2	*	*	*	*	bcdef	a	18960000
1	*	*	*	*	*	bacdef	21360000

Figure 3: Best mappings generated at design-time for Autoindust2

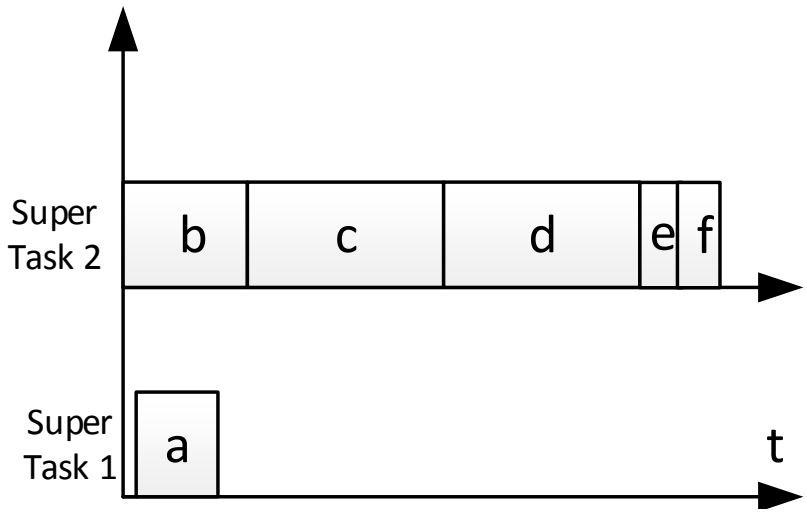

Figure 4: Two super tasks with the minimum makespan

Run-time peak temperature optimization

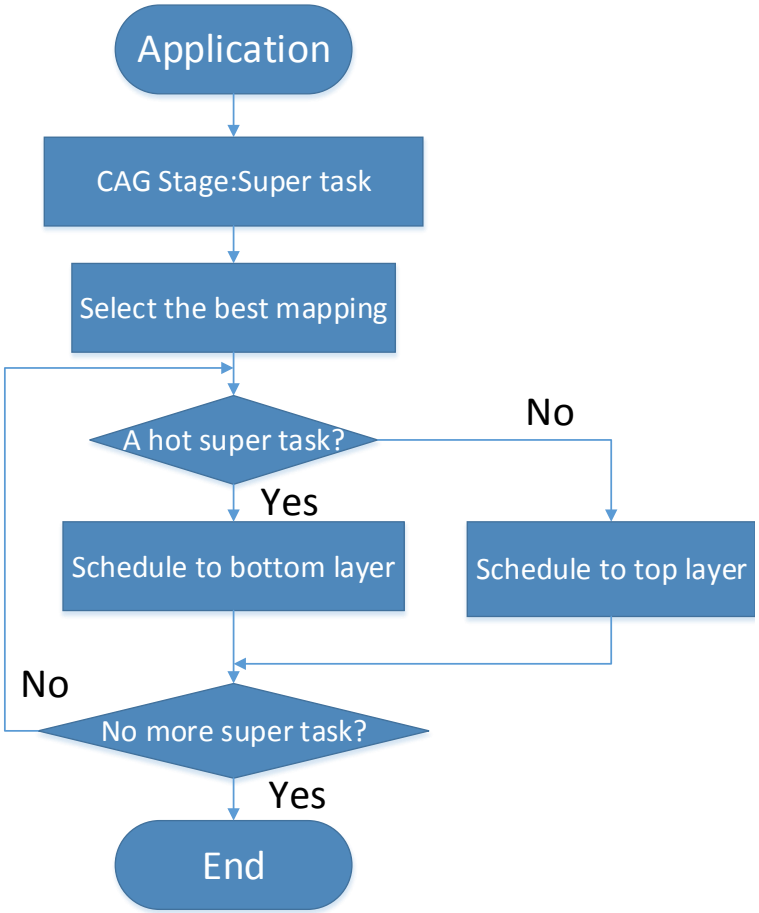

Figure 5: Flow chart of TSS

- Select the best mapping depending on the number of available cores.
- Classify super tasks based on their power.
- Different core selection strategy in different layers.
- A combination of two heuristics:
  1. **Thermal rank model in bottom layer** (the layer closest to heat sink).
  2. **Combined power model in top layer** (the layer far away from heat sink).

Thermal rank model in bottom layer

- Dissipating heat to ambient faster, consider thermal efficiency.
- Thermal rank determines the likelihood of a core to receive new tasks for execution.
- Smaller thermal rank, higher thermal efficiency.
- $TR = (T + Lw \cdot T_l + Vw \cdot T_v) \cdot \rho \cdot P_f$

T: temperature of the candidate core;
T_l, T_v: temperature of the lateral and vertical neighboring cores;
ρ: the proximity to the heat sink;
P_f: position factor, the influence of absolute position of a core;
L_w, V_w: lateral and vertical thermal conductance.

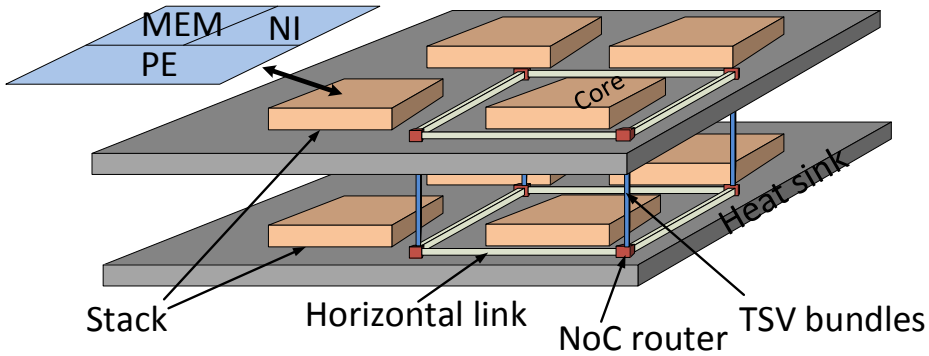

Figure 6: 3D CMP Architecture

Combined power model in top layer

$$P_{combined} = (R_{0\_amb} \cdot P_0 + R_{1\_amb} \cdot P_1) \cdot P_f$$

- Hotspot and temperature emergency in this layer.
- Limiting the total power in a core stack.
- Consider the stacked power in the top layer.

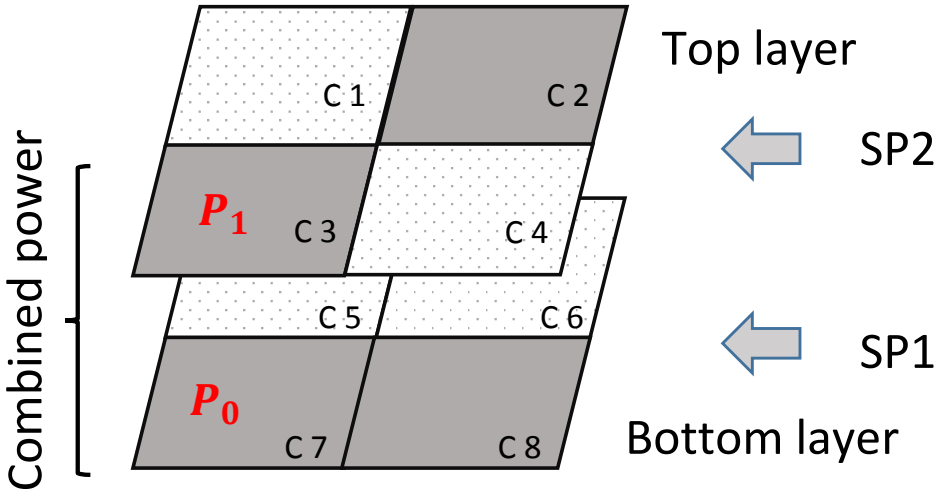

Figure 7: An example of target 3D CMP

Extension to Multi-layer 3D CMP

- Consider thermal efficiency and the stacked power for different layers;
- Hot super task to hyper-bottom layer, cool super task to hyper-top layer.

For a three-layer 3D CMP and four-layer 3D CMP.

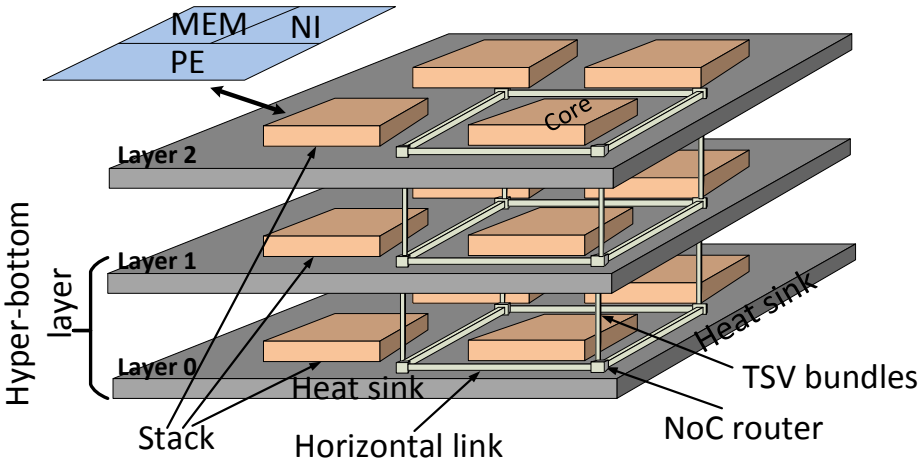

Figure 8: A three-layer 3D CMP

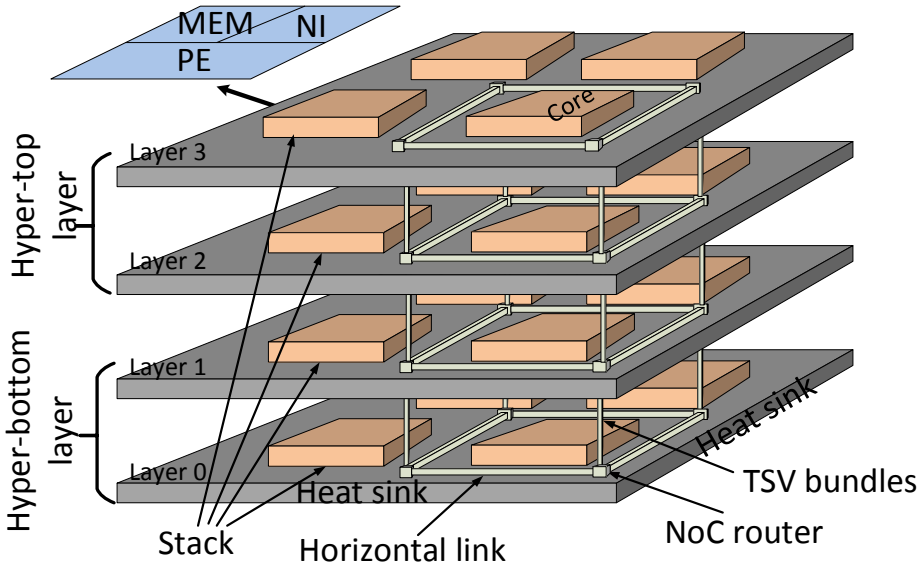

Figure 9: A four-layer 3D CMP

Run-time Mapping Algorithm

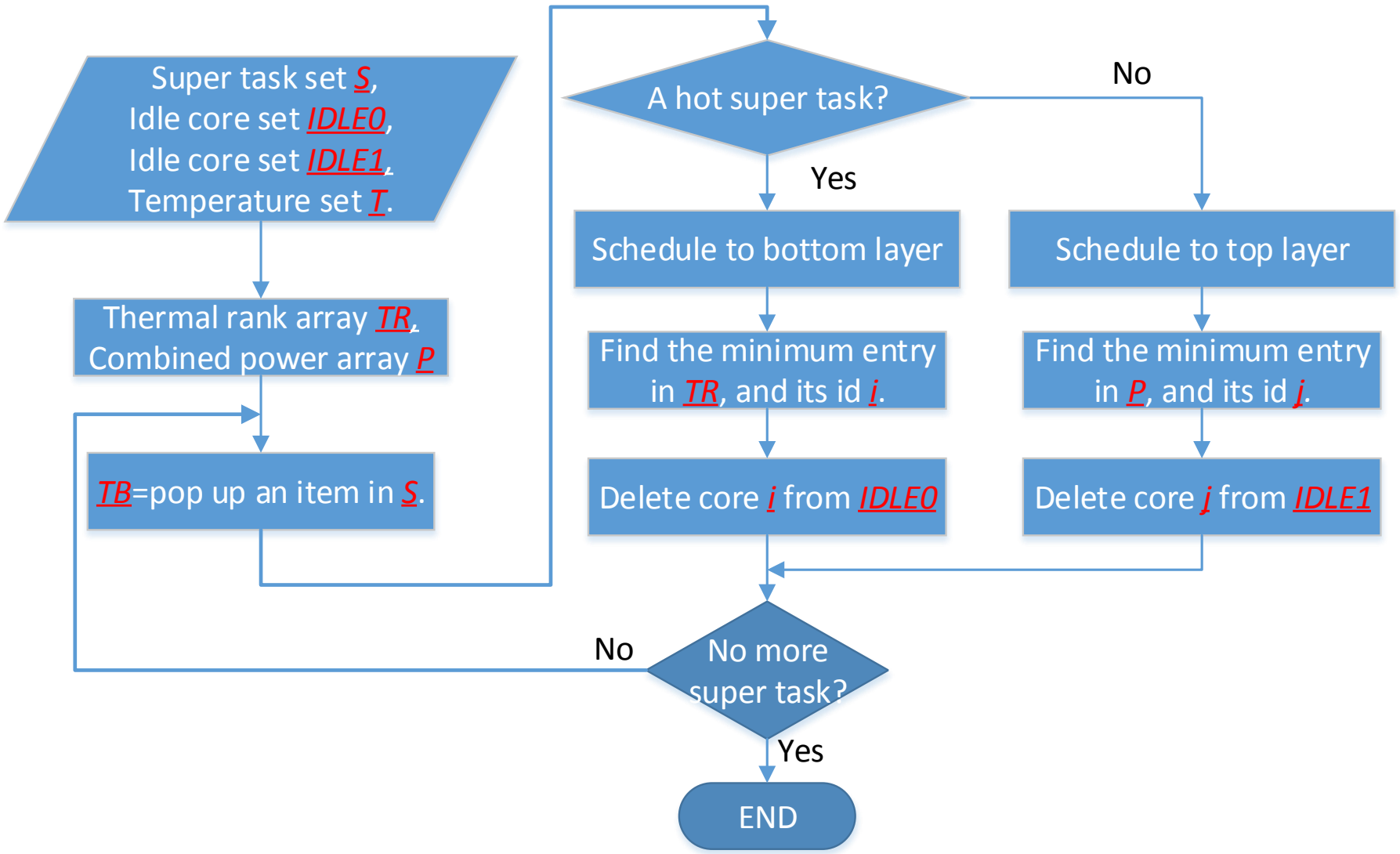

Figure 10: Flow chart of the run-time algorithm

Outline

Parameters and Overall Experimental Flow

Steps for comparison:

1. CF and TR: targeting independent tasks. Temperate comparison to validate the run-time TAS algorithm.
2. PTLS: targeting task graphs. Both temperature and performance comparison to validate our TSS algorithm.
3. Vary scheduling interval from 1ms to 2ms to explore the scenarios under different workload pressure.

Methods	Abbreviation	References
Coolest First	CF	[4]
Thermal Profiling	TR	[1]
Peak Temperature List Scheduling	PTLS	[5]
Two-stage scheduling	TSS	proposed

Table 1: Methodologies considered for comparison

Validate the TAS algorithm

1. TR in [1] and CF in [4] are considered for only temperature comparison, validating the TAS stage.
2. For fair comparison, we assume that the design time makespan optimization results are available to the approaches targeting independent tasks.
3. Seven benchmarks at fixed scheduling interval of 1ms are mapped to $4 \times 4 \times 2$ CMP.
4. Reduce peak temperature up to 5.0°C compared to CF, and up to 4.4°C compared to TR.

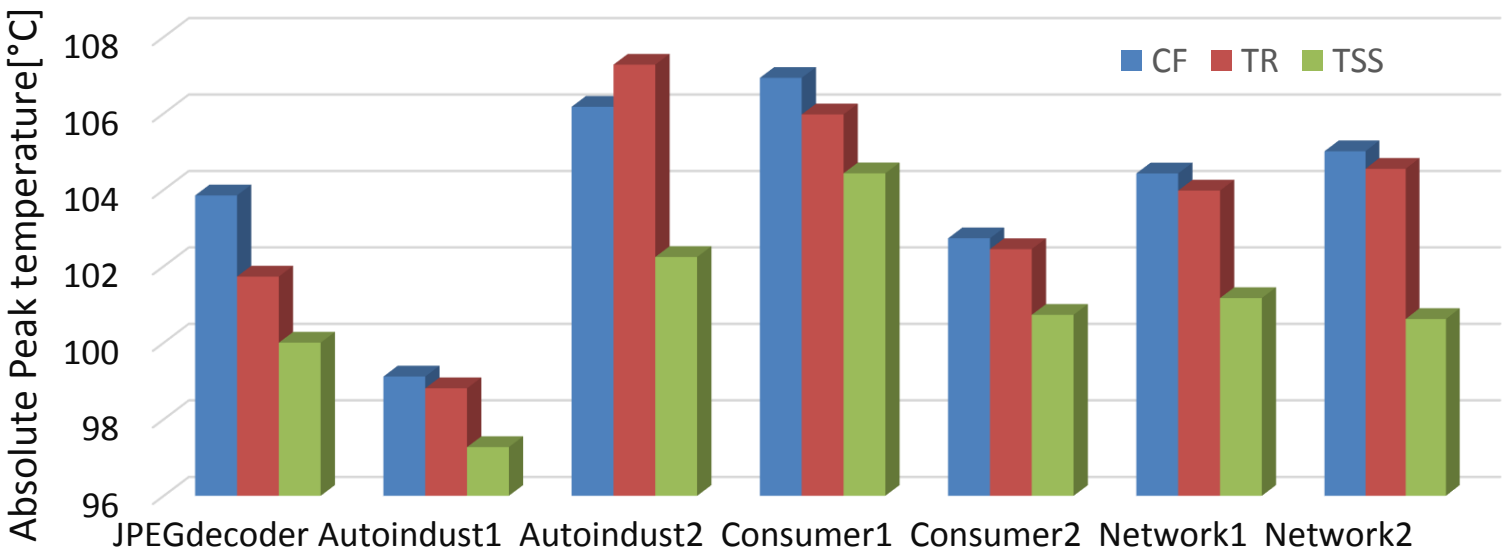

Figure 11: Absolute peak temperature of individual benchmarks

Varying the scheduling interval

1. Varying the scheduling interval to explore the scenarios under different workload pressure.
2. Average reduction of 3.6°C, maximum reduction of 6.1°C.

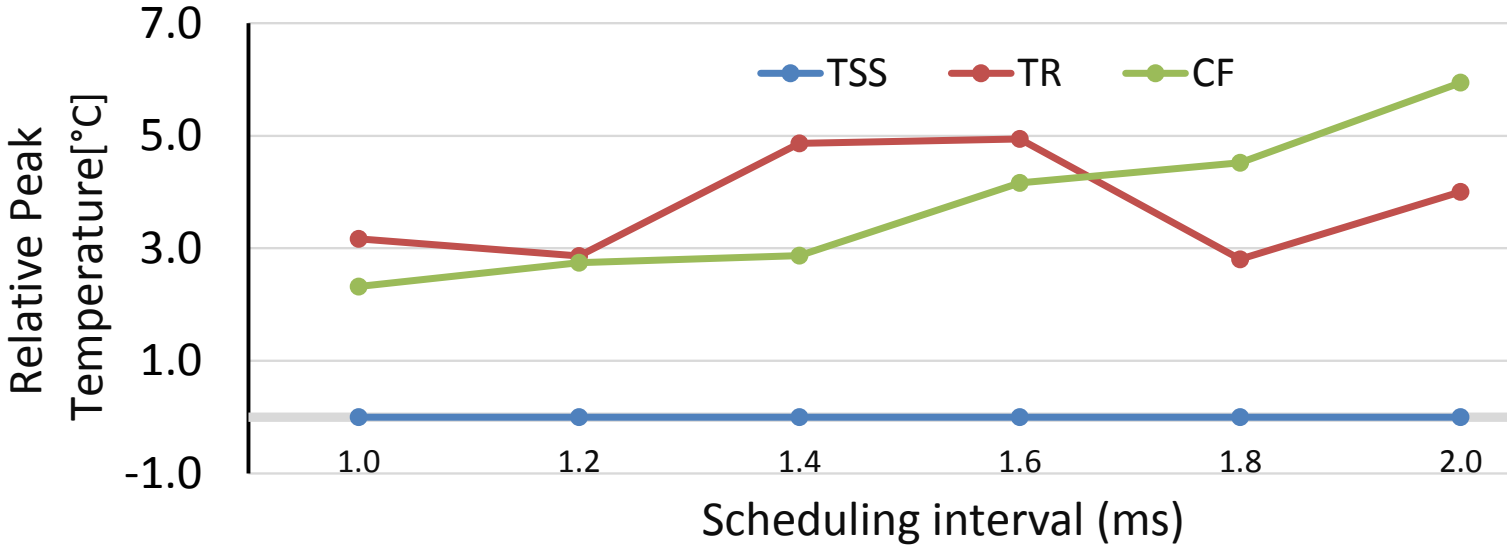

Figure 12: Relative peak temperature on two layer 3D CMP

Comparing to PTLS on Two Layers

1. 7 benchmarks on a two-layer 3D CMP arranged in the $4 \times 4 \times 2$ pattern.
2. CAG stage optimizes the makespan, and TAS stage reduces peak temperature.
3. An average of 4.9% performance improvement, and 6.3°C peak temperature reduction.

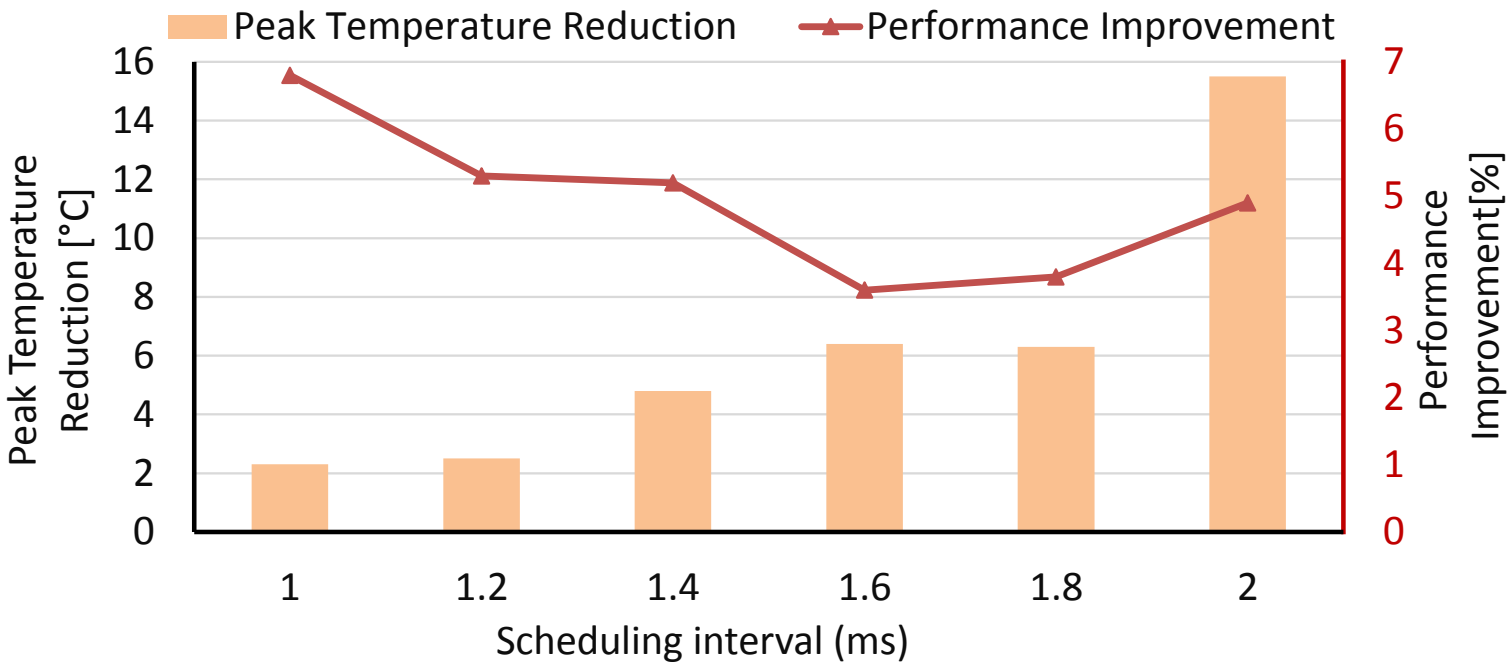

Figure 13: Improvement compared to PTLS for two layers

Comparing to PTLS on Four Layers

1. A four layer 3D CMP arranged in the $2 \times 4 \times 4$ pattern, the same total number of cores as two-layer 3D CMP.
2. An average of 6.78% performance improvement, and 10.0°C peak temperature reduction.
3. Greater improvement is achieved with the increasing number of layers.

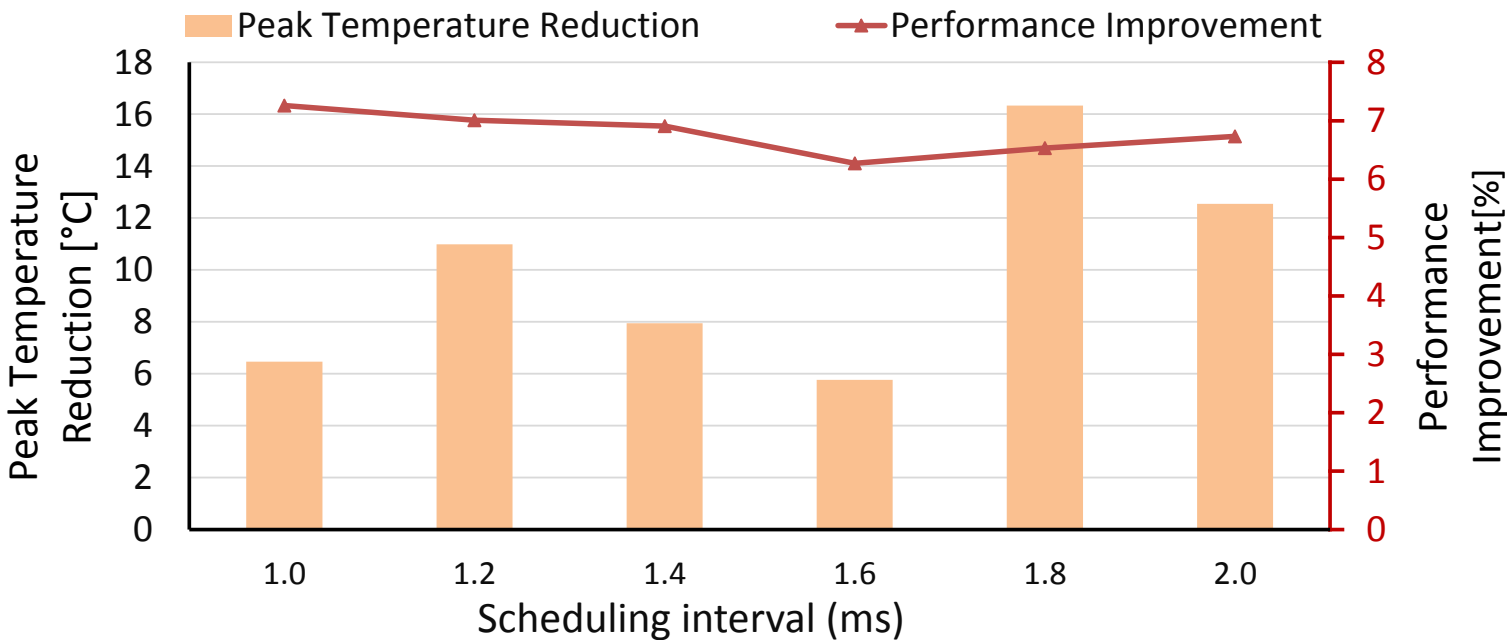

Figure 14: Improvement compared to PTLS for four layers

TAS VS. Thermal rank only and Combined power only

1. Compare our TAS approach which incorporates the consideration of both thermal efficiency and power stack against the ones considering only one factor.
2. Check the peak temperature at fixed scheduling interval of 1ms on two-layer, three-layer, and four-layer 3D CMPs.
3. Every layer is comprised of 4×4 identical cores.
4. TAS outperforms them by 7.3°C and 5.7°C reduction, respectively.

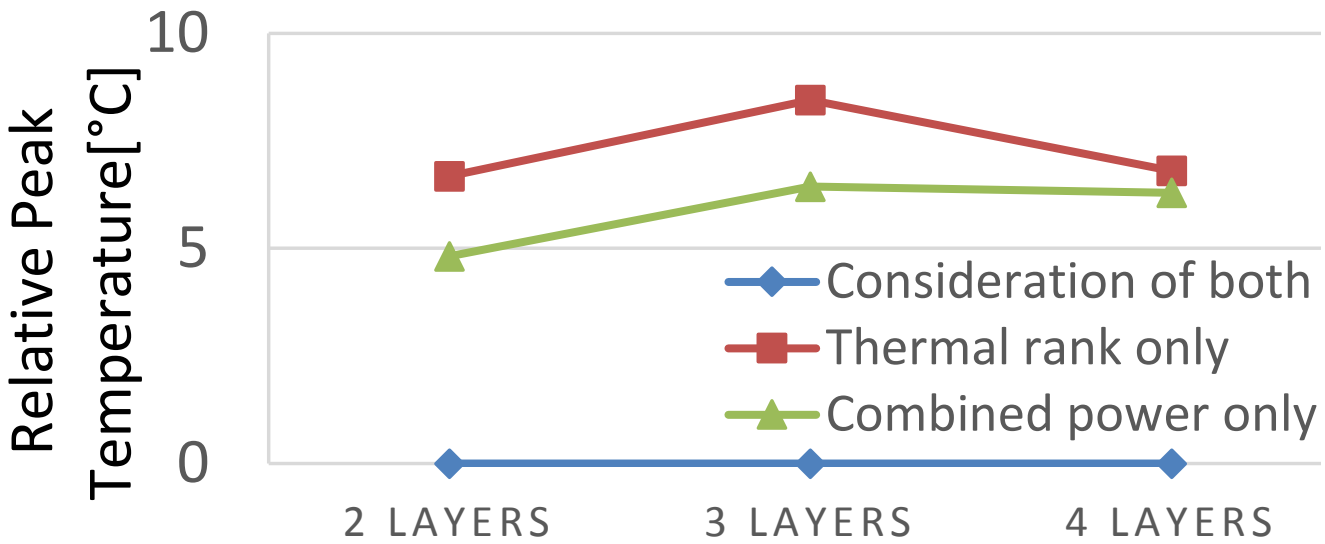

Figure 15: Relative peak temperature on 3D CMPs with different layers

Outline

Summary and Future Works

Summary

- A two-stage thermal-aware task scheduling algorithm.
- Two steps: the communication-aware group stage at design-time, and the thermal-aware scheduling step at run-time.
- Average 4.9% performance improvement and 6.3°C peak temperature reduction.

Future Works

Apply DVFS into the thermal-aware scheduling algorithm to further reduce the peak temperature.

For Further Reading


S. Liu *et al.*, "Thermal-aware job allocation and scheduling for three dimensional chip multiprocessor," in *Proc. of Int. Symp. on Quality Electronic Design*, pp. 390–398, 2010.


C. H. Yu *et al.*, "Thermal-aware on-line scheduler for 3-D many-core processor throughput optimization," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 5, pp. 763–773, 2014.


V. Chaturvedi *et al.*, "Thermal-aware task scheduling for peak temperature minimization under periodic constraint for 3D-MPSoCs," in *Proc. of Int. Symp. on Rapid System Prototyping*, pp. 107–113, 2014.


Y. Cui *et al.*, "Thermal-aware task scheduling for 3D network-on-chip: A Bottom-to-Top scheme," in *Proc. Int. Symp. on Integrated Circuits*, pp. 224–227, 2014.


J. Li *et al.*, "Thermal-aware task scheduling in 3D chip multiprocessor with real-time constrained workloads," *ACM Trans. on Embedded Computing Systems*, vol. 12, no. 2, p. 24, 2013.