Reclaiming Dark Silicon Using Thermally-Aware Chiplet Organization in 2.5D Integrated Systems

Ayse Coskun[1], Furkan Eris[1], Ajay Joshi[1], Andrew B. Kahng[2,3], Yenai Ma[1], Saiful Mojumder[1] and Tiansheng Zhang[1]
[1]ECE Department, Boston University, Boston, MA, USA; [2]ECE and [3]CSE Departments, UC San Diego, La Jolla, CA, USA
{acoskun, fe, joshi, yenai, msam, tszhang}@bu.edu, abk@cs.ucsd.edu

***Abstract*—As on-chip power densities of manycore systems continue to increase, one cannot simultaneously run all the cores due to thermal constraints. This phenomenon, known as the 'dark silicon' problem, leads to inactive regions on the chip and limits the performance of manycore systems. This paper proposes to reclaim dark silicon through a thermally-aware chiplet organization technique in 2.5D manycore systems. The proposed technique adjusts the interposer size and the spacing between adjacent chiplets to reduce the peak temperature of the overall system. In this way, a system can operate with a larger number of active cores at a higher frequency without violating thermal constraints, thereby achieving higher performance. To determine the chiplet organization that jointly maximizes performance and minimizes manufacturing cost, we formulate and solve an optimization problem that considers temperature and interposer size constraints of 2.5D systems. We design a multi-start greedy approach to find (near-)optimal solutions efficiently.**

I. Introduction

Over the past decade, CMOS technology scaling has slowed down, and as a result, it is difficult to sustain the historic performance improvements in CMOS-based VLSI systems. To address this challenge, the computing industry has moved towards packing an increasing number of cores on a single die and using thread-level parallelism to continuously improve performance. At the same time, the on-chip power density has risen with shrinking transistor feature size. This increasing power density has led to 'dark silicon' [1] on a chip. As a result in manycore systems not all cores can be operated at the highest frequency or even turned on simultaneously due to thermal constraints. Thus, there is a significant amount of performance that is 'left on the table' in today's manycore systems.

Solutions have been proposed to address the dark silicon problem at both hardware level [2] and system management level [3] for single-chip systems. These techniques help balance the heat dissipation across the chip, thereby improving system energy efficiency under thermal constraints. However, these techniques are not able to maximize the performance in manycore systems persistently.

In tandem with technology scaling and the move to manycore systems, die-stacking technologies such as 2.5D and 3D integration have emerged to improve system performance [4]–[6]. 3D integration, which stacks dies vertically to form a system, reduces system footprint and increases memory bandwidth [5], but exacerbates the thermal issues [4]. 2.5D integration, which integrates small chiplets on a silicon interposer, is less prone to the thermal challenges observed in 3D stacking [6]. Moreover, it provides additional routing resources through the interposer, and is more cost-effective [5], [6]. Currently, 2.5D integration technology is being extensively investigated by both academia and industry [5], [7].

In 2.5D integration, the general approach to arrange chiplets is to integrate them as close as possible on an interposer to save cost. There is however an opportunity here to solve the 'dark silicon' problem by organizing the chiplets in a thermally-aware fashion such that we can lower the overall

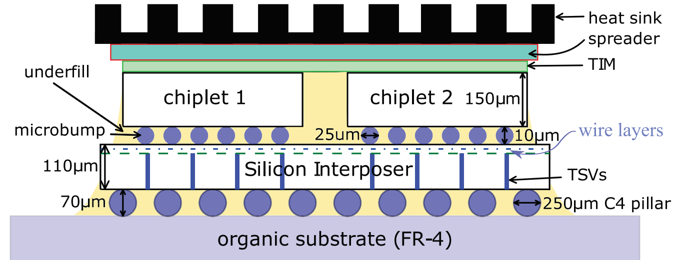

Fig. 1: *Cross-sectional view of a 2.5D system.*

manycore system temperature and in turn improve performance (by having more active cores operating at higher frequency) without significantly increasing the cost. In this paper, we propose a thermally-aware chiplet organization strategy to address the dark silicon problem in 2.5D manycore systems. We strategically insert spacing between the chiplets to lower the system temperature. This reduction enables higher operating frequency and/or more active cores in the 2.5D manycore system under the same temperature threshold, which in turn improves the overall system performance. We design a multi-start greedy approach to efficiently find the (near-)optimal thermally-aware chiplet organization that jointly maximizes the manycore system performance and minimizes the system manufacturing cost.

II. Thermally-Aware Chiplet Organization

A. 2.5D System Overview

We use a 256-core homogeneous system operating at $1GHz$ as an example manycore system in this work. In the example 2.5D system (Fig. 1), we split a single chip into chiplets, place the chiplets onto a passive TSV interposer, and use microbumps to connect the chiplets and the interposer. We place the interposer on top of a substrate using C4 bumps for connection. Our evaluation uses the conventional 2D single-chip system as a baseline, where the 256-core chip is placed directly on top of an organic substrate with C4 bumps for connection.

B. Optimization of Chiplet Organization

To determine the optimal thermally-aware chiplet organization (including chiplet count, chiplet placement, active core count, and operating frequency), we formulate an objective function that maximizes system performance while minimizing system cost (see Eq. (1)). In Eq. (1), 2.5D system performance (in terms of instructions per second (IPS)) and cost are normalized to the baseline single-chip system, and the user-specified weight factors α and β have no units. All notations are listed in Table I.

$$Minimize: \quad \alpha \times \frac{IPS_{2D}}{IPS_{2.5D}(f,p)} + \beta \times \frac{C_{2.5D}(n,s_1,s_2,s_3)}{C_{2D}} \tag{1}$$

$$Subject\ to: \quad T_{peak}(f,p,n,s_1,s_2,s_3) <= T_{threshold} \tag{2}$$

$$w_{int} <= 50, \quad h_{int} <= 50 \tag{3}$$

$$w_c = \frac{w_{2D}}{r}, h_c = \frac{h_{2D}}{r} \tag{4}$$

$$w_{int} = w_c \times r + 2 \times s_1 + s_3 + 2 \times l_g, h_{int} = h_c \times r + 2 \times s_1 + s_3 + 2 \times l_g \tag{5}$$

$$2 \times s_1 + s_3 - 2 \times s_2 > 0 \tag{6}$$

$$N_{CMOS} = \frac{\pi \times (\phi_{wafer}/2)^2}{A_{CMOS}} - \frac{\pi \times \phi_{wafer}}{\sqrt{2 \times A_{CMOS}}}, \quad N_{int} = \frac{\pi \times (\phi_{wafer_{int}}/2)^2}{A_{int}} - \frac{\pi \times \phi_{wafer_{int}}}{\sqrt{2 \times A_{int}}} \tag{7}$$

TABLE I: Notation used in Equations

Notation	Definition	Assumed Value
$\phi_{wafer}, \phi_{wafer_{int}}$	Diameter of CMOS and interposer wafer	$300mm$
N_{CMOS}, N_{int}	CMOS and interposer dies per wafer	Eq. (7)
D_0	Defect density	$0.25/mm^2$ [6]
γ	Defect clustering parameter	3 [6]
Y_{int}	Yield of an interposer	98% [8]
Y_{CMOS}	Yield of a CMOS chiplet	from Eq. (8)
C_{wafer}	CMOS wafer cost	$5000 [9]
$C_{wafer_{int}}$	Interposer wafer cost	$500 [9]
$C_{int}, C_{CMOS}, C_{2D}$	Chiplet, interposer, and single chip cost	from Eq. (9)
Y_{bond}	Chiplet bonding yield	99% [6]
$C_{2.5D}$	Cost of the 2.5D system	from Eq. (10)
l_g	Guard band along each interposer edge	$1mm$
w_{2D}, h_{2D}	Width and height of the baseline single chip	$18mm$
w_{int}, h_{int}	Width and height of the interposer (in mm)	from Eq. (5)
w_c, h_c	Width and height of the chiplets (in mm)	from Eq. (4)

Notation	Definition
A_{CMOS}, A_{int}	CMOS, interposer die area
C_{bond}	Bonding cost of a chiplet
r	Number of chiplets in a row or column
n	Number of chiplets $n = r \times r$, $n \in \{4, 16\}$
F V	Frequency set $\{1000, 800, 533, 400, 320MHz\}$ Corresponding voltage set $\{0.9, 0.87, 0.71, 0.63, 0.63V\}$
f	Operating frequency $f \in F$
p	Active core count $p \in \{32, 64, 96, 128, 160, 192, 224, 256\}$
$IPS_{2.5D}, IPS_{2D}$	Instructions per second (IPS) of 2.5D system and 2D system
s_1, s_2, s_3	Chiplet spacings (Fig. 2(a)). $s_1 = s_2 = 0$ for 4-chiplet case
$T_{peak}, T_{threshold}$	Peak operating temperature and Temperature threshold for safety

$$Y_{CMOS} = (1 + A_{CMOS} D_0/\gamma)^{-\gamma} \tag{8}$$

$$C_{CMOS} = C_{wafer}/N_{CMOS}/Y_{CMOS}, \quad C_{int} = C_{wafer_{int}}/N_{int}/Y_{int} \tag{9}$$

$$C_{2.5D} = \frac{C_{int} + \Sigma_{i=1}^{n}(C_{CMOS} + C_{bond})}{Y_{bond}^{n-1}} \tag{10}$$

Eq. (2) is the peak temperature constraint for a valid chiplet organization. Eq. (3) limits the interposer size to be no larger than $50mm \times 50mm$. Eq. (4) calculates the chiplet width and height. Eq. (5) calculates the interposer width and height as a function of chiplet spacings (s_1, s_2, and s_3 in Fig. 2(a), which vary **independently**). Eq. (6) ensures there is no overlap between center chiplets. The 2.5D system cost is calculated using Eqs. (7) to (10). Eqs. (7) through (10) [6] calculate CMOS dies per wafer and interposer dies per wafer, CMOS chiplet yield, CMOS per-chiplet cost and interposer cost, and the overall cost of a 2.5D system, respectively.

To solve the optimization problem, an exhaustive search approach would take $180k$ CPU hours. Hence, we use a multi-start greedy approach. We validated this approach against the exhaustive search. Our multi-start greedy approach determines the solution to the optimization problem $100\times$ faster.

```
Pseudocode: Multi-Start Greedy Approach
1)  calculate cost and performance of 2.5D system for all (f, p, C_2.5D) combinations
2)  input obj. func. weights (α,β)
    sort (f, p, C_2.5D) combinations based on obj. func. from low to high
3)  foreach (f, p, C_2.5D) combination in the sorted order do
      generate random start points of (s1, s2, s3)
      foreach start point (S_current) do
        evaluate peak temperature T of S_current
        repeat
          generate a random neighbor placement (S_neighbor)
          evaluate peak temperature T' of S_neighbor
          if T' < T_threshold then
            output S_neighbor and (f, p, C_2.5D) combination and exit
          if T' < T then
            update minimum peak temperature T ← T'
            update current placement S_current ← S_neighbor
        until T < peak temperature of all the neighbor placements
      end for
    end for
```

C. Evaluation Methodology

Our evaluation framework is shown in Fig. 2(b). We use Sniper [10] for performance evaluation, McPAT [11] for power calculation, and HotSpot-6.0 [12] for thermal simulation. There is a closed loop between chiplet organizer, floorplan generator, and HotSpot. The chiplet organizer is implemented using the multi-start greedy algorithm as discussed in Sec. II-B.

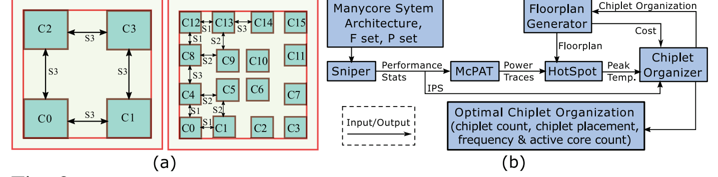

Fig. 2: *(a) Chiplet count and placement options. We vary the chiplet spacings* ***independently*** *to find the optimal chiplet placement. (b) Evaluation framework.*

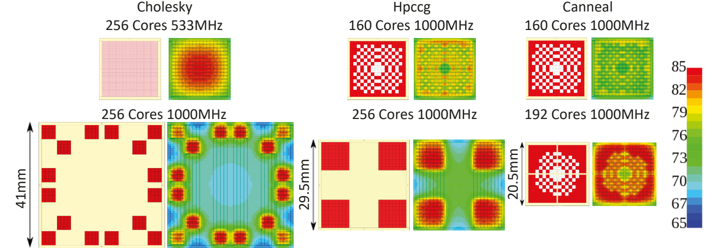

Fig. 3: *Choice of chiplet organizations that maximizes the performance under 85^oC for single-chip baseline (top) and 2.5D systems (bottom).*

III. Evaluation Results

Fig. 3 shows examples of optimal chiplet organization and the workload allocation for $\alpha = 1$ and $\beta = 0$ under an 85^oC constraint. For `cholesky`, our technique improves performance by 80% by increasing frequency from $533MHz$ to $1GHz$, while the cost is similar compared to the baseline. For `hpccg`, our 2.5D system achieves 40% higher performance by increasing active core count from 160 to 256 and lowers cost by 28%. For `canneal`, the performance benefit is 7% because its performance saturates at 192 active cores; however, our approach reduces the cost by 36%. These results demonstrate that our thermally-aware chiplet organization technique can reclaim dark silicon by having more active cores and/or operate the cores at a higher frequency without violating the temperature threshold.

IV. Conclusion

We propose a thermally-aware chiplet organization strategy to reclaim dark silicon in 2.5D manycore systems. We use a multi-start greedy approach to efficiently solve the optimization problem which jointly maximizes performance and minimizes manufacturing cost.

References

[1] M. Shafique *et al.*, "The EDA challenges in the dark silicon era:temperature, reliability, and variability perspectives," in *Proc. DAC*, 2014, pp. 1–6.

[2] G. Venkatesh *et al.*, "Qscores: Trading dark silicon for scalable energy efficiency with quasi-specific cores," in *Proc. MICRO*, 2011, pp. 163–174.

[3] S. Pagani *et al.*, "TSP: thermal safe power: efficient power budgeting for many-core systems in dark silicon," in *Proc. CODES+ISSS*, 2014, p. 10.

[4] G. H. Loh, Y. Xie, and B. Black, "Processor design in 3D die-stacking technologies," *IEEE Micro*, vol. 27, no. 3, 2007.

[5] A. Kannan *et al.*, "Enabling interposer-based disintegration of multi-core processors," in *Proc. MICRO*, 2015, pp. 546–558.

[6] D. Stow *et al.*, "Cost analysis and cost-driven IP reuse methodology for SoC design based on 2.5D/3D integration," in *Proc. ICCAD*, 2016, p. 56.

[7] "DARPA CHIPS," http://www.darpa.mil/news-events/2016-07-19, 2016.

[8] K. Tran *et al.*, "High-bandwidth memory white paper: Start your HBM/2.5D design today," Amkor Technology Inc., Tech. Rep., 2016.

[9] G. Parès, "3D technology for photonics silicon interposer," in *Green IT workshop - Leti Days*, 2013.

[10] T. E. Carlson *et al.*, "Sniper: exploring the level of abstraction for scalable and accurate parallel multi-core simulation," in *Proc. SC*, 2011, p. 52.

[11] S. Li *et al.*, "McPAT: an integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proc. MICRO*, 2009, pp. 469–480.

[12] R. Zhang *et al.*, "Hotspot 6.0: Validation, acceleration and extension," 2015.