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

Ayse Coskun¹, Furkan Eris¹, Ajay Joshi¹, Andrew B. Kahng²,³, Yenai Ma¹, Saiful Mozumder¹ and Tiansheng Zhang¹

¹ECE Department, Boston University, Boston, MA, USA; ²ECE and ³CSE 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 packaging 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 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 [8].

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 \( \alpha \) and \( \beta \) have no units. All notations are listed in Table I.

\[
\text{Minimize:} \quad \alpha \times \frac{\text{IPS}_{2.5D}}{\text{IPS}_{2D}} + \beta \times \frac{C_{\text{2D}}}{C_{\text{2.5D}}}
\]

Subject to:

\[
T_{\text{peak}}(f, r, p, s_1, s_2, s_3) \leq T_{\text{threshold}}
\]

\[
w_{s_1} \leq 50, \quad b_{s_1} \leq 50
\]

\[
w_c = \frac{w_{s_2}}{r} = \frac{b_{s_2}}{r}
\]

\[
w_{s_2} = w_c \times r + 2 \times s_1 + s_3 + 2 \times t_c, \quad h_{s_3} = h_c + r + 2 \times s_1 + s_3 + 2 \times t_c
\]

\[
2 \times s_1 + s_3 + 2 \times t_c > 0
\]

\[
N_{\text{CMOS}} = \frac{\pi \times (d_{\text{bump}}/2)^2}{A_{\text{CMOS}}}, \quad d_{\text{bump}} = \frac{\pi \times (d_{\text{bump}}/2)^2}{\sqrt{2 \times A_{\text{CMOS}}}}
\]

\[
Y_{\text{CMOS}} = (1 + \frac{C_{\text{CMOS}}}{V_{\text{CMOS}}} \cdot \frac{1}{\gamma})^{-1}
\]

\[
C_{\text{CMOS}} = C_{\text{CMOS}} \cdot \frac{V_{\text{CMOS}}}{V_{\text{CMOS}} / Y_{\text{CMOS}}}
\]

Where:

- \( \text{IPS}_{2.5D} \) and \( \text{IPS}_{2D} \) are the IPS of the 2.5D system and the baseline system, respectively.
- \( C_{\text{2D}} \) and \( C_{\text{2.5D}} \) are the costs of the 2D system and the 2.5D system, respectively.
- \( T_{\text{peak}} \) is the peak temperature of the system.
- \( T_{\text{threshold}} \) is the temperature threshold.
- \( w_{s_1} \) and \( b_{s_1} \) are the maximum width and height of the chiplet, respectively.
- \( w_c \) is the width of the interposer.
- \( h_c \) is the height of the interposer.
- \( s_1, s_2, s_3 \) are the spacing between the chiplets.
- \( d_{\text{bump}} \) is the diameter of the bump.
- \( V_{\text{CMOS}} \) is the voltage of the CMOS system.
- \( \gamma \) is the power exponent.

Fig. 1: Cross-sectional view of a 2.5D system.
Table 1: Notation used in Equations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
<th>Assumed Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D_0)</td>
<td>Defect density</td>
<td>0.25 (\text{mm}^2) [6]</td>
</tr>
<tr>
<td>(N_{\text{CMOS}},A_{\text{int}})</td>
<td>CMOS and interposer die area</td>
<td></td>
</tr>
<tr>
<td>(C_{\text{bond}})</td>
<td>Bonding cost of a chiplet</td>
<td></td>
</tr>
<tr>
<td>(y_{\text{bond}})</td>
<td>Chiplet bonding yield</td>
<td>99% [6]</td>
</tr>
<tr>
<td>(C_{\text{2SD}})</td>
<td>Cost of the 2SD system</td>
<td></td>
</tr>
<tr>
<td>(y_{\text{2SD}})</td>
<td>2SD System yield</td>
<td></td>
</tr>
<tr>
<td>(w_{\text{hpccg}}, h_{\text{hpccg}})</td>
<td>Width and height of the interposer (in mm)</td>
<td></td>
</tr>
<tr>
<td>(w_{\text{int}}, h_{\text{int}})</td>
<td>Width and height of the interposer (in mm)</td>
<td></td>
</tr>
</tbody>
</table>

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 \((s_1, s_2, s_3)\) 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 and 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 2SD system for all \((f, p, C_{\text{2SD}})\) combinations
2) sort \((f, p, C_{\text{2SD}})\) combinations based on obj. func. from low to high
3) foreach \((f, p, C_{\text{2SD}})\) combination in the sorted order do
   generate random start points of \((x_1, x_2, x_3)\)
   evaluate peak temperature \(T_{\text{peak}}\)
   repeat
      generate a random neighbor placement \(S_{\text{neighbor}}\)
      evaluate peak temperature \(T'_{\text{peak}}\)
      if \(T' < T_{\text{peak}}\) then
         output \(S_{\text{neighbor}}\) and \((f, p, C_{\text{2SD}})\) combination and exit
      end if
   until \(T < T_{\text{peak}}\) of all the neighbor placements
end for

Eq. (10)

\[
C_{\text{2SD}} = C_M + \sum_i (C_{\text{CMOS}} + C_{\text{bond}})
\]

C. Evaluation Methodology

Our evaluation framework is shown in Fig. 2(b). We use Sniper [11] for performance evaluation, McPAT [12] for power calculation, and HotSpot-6.0 [13] for thermal simulation [14]. There is a closed loop between chiplet optimizer, floorplan generator, and HotSpot. The chiplet optimizer is implemented using the multi-start greedy algorithm as discussed in Sec. II-B.

III. Evaluation Results

Fig. 3 shows examples of optimal chiplet organization and the workload allocation for \(\alpha = 1\) and \(\beta = 0\) under an 85°C 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 hpcgcg, 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