Exploring Benefits and Designs of Optically Connected Disintegrated Processor Architecture

Yan Pan, Yigit Demir, Nikos Hardavellas, John Kim[†], Gokhan Memik
Northwestern University [†]KAIST

1. Introduction

Power density and off-chip bandwidth are the two major limitations for the scalability of the performance of future Chip-Multiprocessors (CMP). While advanced silicon fabrication technology allows for an increasing number of transistors to be integrated on a single chip, an increasing part of them cannot be powered on at the same time, resulting in a "dark silicon era". At the same time, the limited pin count and low efficiency in off-chip communication limits the off-chip bandwidth. Thus, the performance of future CMPs is not likely to scale, even if the workload itself is parallelized [3]. However, with the introduction of silicon nanophotonics, both limitations can be alleviated. The low latency and high bandwidth density of optical signaling can be utilized for efficient off-chip communication, thus bringing physically separated chips effectively closer together. This makes it possible to build a large scale many-core "macrochip" with multiple optically connected chiplets [1][2], as shown in Figure 1.

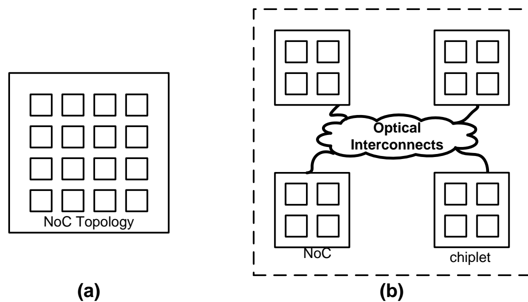

Figure 1. (a) Conventional CMP. (b) Optically-connected disintegrated processor (OCDP).

There are many important questions to be answered about this novel optically connected disintegrated processor (OCDP) architecture, including the potential benefits that can be harvested, the organization of the optical channels, the interfacing of on-chip/off-chip networks, and the design space search for the chiplet sizing and composition. In this work, we first use first-order analytical models to estimate how far this architecture can push back the power and bandwidth wall for scalable workloads. We then explore the design concerns for chiplets and propose a possible topology based on the new trade-offs in the multi-chiplet scenario.

2. Benefits of OCDP

Optical signaling provides low latency and high bandwidth density compared to conventional electrical signaling. As the main memory can be placed on one of the chiplets, the memory access latency in OCDP can be reduced. The total off-chip bandwidth can also scale up. More importantly, the disintegrated architecture alleviates the power budget of the processor by reducing the power density and enabling more effective cooling. The die size of conventional CMP is limited due to manufacturing yield. By having multiple smaller chiplets, the total die area that can be summoned for a single macro-processor can be increased.

With these intuitive motivations for adopting OCDP, it is critical to estimate how the improved latency, bandwidth and power/area budget are going to affect the overall performance of a future many-core processor. We adopted the analytical model described by Hardavellas et al. [3], which estimates the *optimal* performance of a homogeneous CMP with Niagara-like cores at 20nm technology node on highly parallelizable workloads. For this evaluation, we assume a 99% parallelizable TPC-C workload. We pick a baseline CMP design with memory latency of 25.1ns, off-chip bandwidth of 76GB/s, a die size of 310mm^2 and a power budget of 130W, based on ITRS projections. We then vary these parameters to reflect the changes brought about by optical signaling and the OCDP design and evaluate their impact on the overall performance.

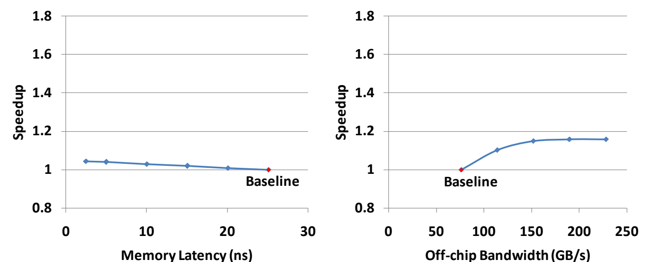

Figure 2. Estimated TPC-C execution speedup with varied (a) memory latency (b) provided off-chip bandwidth.

We first evaluate the impact of low latency and improved off-chip bandwidth by replacing electrical pins with optical links. As shown in Figure 2(a), the reduced memory latency has minimal performance benefit for the TPC-C benchmark, since the caches hide much of the memory latency. On the other hand, improving off-chip bandwidth has more significant benefits, but also tapers off at 1.18x speedup, as the system hits power wall and cannot sustain excessive communication. Both results show that simply replacing electrical links with optical links provides limited performance gain, as future processors are heavily power bounded.

Thus, scaling up the power budget is critical for the scalability of processor performance. Figure 3 shows the TPC-C performance with increased total die area of the macro-processor. For conventional single chip CMPs, the total die area cannot be easily increased due to yield limit, but emerging technologies like 3D integration, with its low latency and high bandwidth across stacked layers, can effectively increase the total die area of a processor. Unfortunately, the power budget of a 3D stack cannot proportionally scale up. We approximate this limitation by the fixed power

budget line in the figure. Obviously, the extra silicon area cannot be all powered on at the same time, and the overall performance benefit is limited at 5%. OCDP, on the other hand, spreads out components (e.g., cores) of a macro-processor and hence enables scalable total power budget. This allows the performance to improve by up to 29%, until the system becomes bandwidth limited. If the high bandwidth density of the optical interconnect is leveraged at the same time, up to 60% speedup can be reached.

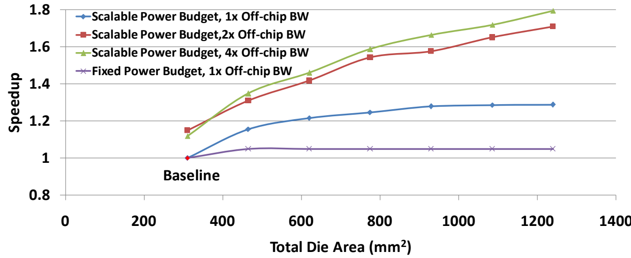

Figure 3. TPC-C speedup with increased total die area. Scalable power budget increases proportionally with area.

Even though the above results are based on a first order analytical performance model, the model takes into account various limiting factors for performance scaling in future many-core processors. Our analysis demonstrates that OCDP, if properly implemented, has the potential to push back not only the bandwidth wall, but also the power wall. This unique feature makes OCDP well worth careful study.

3. OCDP Architecture Design

There are two main aspects to the design of an OCDP architecture: the chiplet and the inter-chiplet network.

3.1 Chiplet Design

A chiplet may contain a set of cores, caches, memory controllers or even main memory. Its size depends on the fabrication yield, the efficiency of on-chiplet/off-chiplet communication, cost of assembly, functionality division and power budget. Smaller chiplets may have higher fabrication yield, leverage more optical communication, and have higher power budget; but they may also incur higher assembly cost and complicate the overall architecture of the OCDP. Since the chiplets are fabricated separately, they may even take different sizes according to their respective constraints and functionalities. In heterogeneous designs, pre-built special function chiplets can be re-used/mixed to lower design cost.

3.2 Inter-chiplet Network

To benefit from the increased total die area and power budget for an OCDP, the inter-chiplet communication has to be efficient and seamlessly interfaced with the on-chip communication fabric. Given the different nature of signaling technologies intra- and inter-chiplet, a hierarchical topology is likely to be feasible and more efficient. Unique to the OCDP, the extra optical loss introduced by couplers between off-chip optical fibers and on-chip waveguides has to be minimized. Also, the size of the chiplets may also impact the design of the interconnection network. Both point-to-point interconnects[2] and “hub-chip”[4] designs have been proposed.

Currently, we are investigating an “extended Dragonfly” topology for the OCDP, with Firefly topology on each chiplet, as shown in Figure 4.

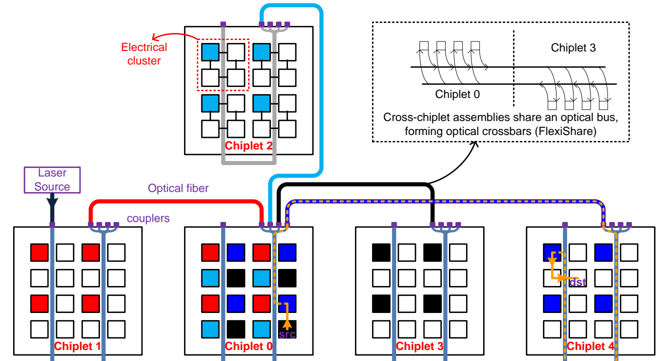

Figure 4. Proposed inter-chiplet architecture with 5 chiplets.

Internally, each chiplet employs the Firefly [6] topology. Neighboring routers are electrically connected into *clusters*, as shown in chiplet 2. A U-shape waveguide runs through each chiplet, and *assemblies* are formed across clusters as optical crossbars. Different assemblies are identified by the color of the nodes. Each assembly is further extended to a different chiplet using low-loss cross-chiplet optical fibers, as shown between chiplet 0 and 3. Thus, a Dragonfly [5] topology is formed at the top level, so that communication between any pair of routers only takes the optical link once.

This architecture is proposed based on several considerations. First, the amount of couplers on any optical path is limited to 3 (i.e., 1 from the laser source to the on-chip waveguide, and 2 between two chiplets) for lower optical loss. Second, the electrical cluster size can be varied to fit a range of chiplet sizes. Third, the optical links extend further into the chiplets to lower latency and improve communication efficiency. Fourth, varied amount of the inter-chiplet optical links can be used to form FlexiShare [7] optical crossbars to match the bandwidth demand. Lastly, the hybrid electrical/optical design improves the efficiency of the system.

A sample routing is shown between the *src* node in chiplet 0 and the *dst* node in chiplet 4. A packet first takes the electrical link to a node optically connected to chiplet 4, where it is converted into optical signals and traverses the optical link to the node in chiplet 4 that is nearest to its final destination. Further electrical hops are taken to reach *dst*.

REFERENCES

[1] S. Beamer et al., “*Designing Multi-socket Systems Using Silicon Photonics*”, ICS-09, Yorktown Heights, NY, June 2009.

[2] P. Koka et al., “*Silicon-Photonic Network Architectures for Scalable, Power-Efficient Multi-Chip Systems*”, ISCA-37, Saint-Malo, France, June 2010.

[3] N. Hardavellas et al., “*Power Scaling: the Ultimate Obstacle to 1K-Core Chips*”. Northwestern University Tech Report NWU-EECS-10-05, Mar. 2010.

[4] M. Cianchettie et al., “Leveraging Nanophotonics for System-in-Package Interconnect”, Workshop on Interaction between Nanophotonic Devices and Systems (in conj. With MICRO-43),Atalanta, GA, Dec 2010.

[5] J. Kim et al. “Cost-efficient Dragonfly Topology for Large-scale Systems”, IEEE Micro Top Picks, Jan 2009.

[6] Y. Pan et al. “Firefly: Illuminating Future Network-on-chip with Nanophotonics”, ISCA-36, Austin, TX, June 2009.

[7] Y. Pan et al. “FlexiShare: Channel Sharing for an Energy-Efficient Nanophotonic Crossbar”, HPCA-16, Bangalore, India, Jan 2010.