Accelerating Cache Coherence in Manycore Processor through Silicon Photonic Chiplet

Chengeng Li[1], Fan Jiang[1], Shixi Chen[1], Jiaxu Zhang[1], Yinyi Liu[1], Yuxiang Fu[1], Jiang Xu[1,2,*]
[1]The Hong Kong University of Science and Technology
[2]Microelectronics Thrust, The Hong Kong University of Science and Technology(GZ)
*Corresponding author: *jiang.xu@ust.hk*

ABSTRACT

Cache coherence overhead in manycore systems is becoming prominent with the increase of system scale. However, traditional electrical networks restrict the efficiency of cache coherence transactions in the system due to the limited bandwidth and long latency. Optical network promises high bandwidth and low latency, and supports both efficient unicast and multicast transmission, which can potentially accelerate cache coherence in manycore systems. This work proposes a novel photonic cache coherence network with a physically centralized logically distributed directory called PCCN for chiplet-based manycore systems. PCCN adopts a channel sharing method with a contention solving mechanism for efficient long-distance coherence-related packet transmission. Experiment results show that compared to state-of-the-art proposals, PCCN can speed up application execution time by 1.32x, reduce memory access latency by 26%, and improve energy efficiency by 1.26x, on average, in a 128-core system.

1 INTRODUCTION

Over the past decades, with the rapid development of big data, machine learning, high-performance computing, etc., parallel workloads have faced an increasing demand for computation capacity [23]. Many commercial high-performance manycore processors have been released, such as AMD's 64-core Threadripper 3990X, Intel's 72-core Xeon Phi (KNL) and Mellanox's 100-core TILE-Mx100, and in the future, the number of cores in manycore processors even reaches to the hundreds or thousands [3]. However, integrating more cores into a monolithic silicon die encounters many challenges such as integration intensity, cost, and yield [18]. Besides, the cache coherence is a significant obstacle that diminishes performance returns for increasing the number of cores.

Currently, industry and academia have viewed the chiplet as a promising approach to extending the system scale, and various chiplet-based manycore designs have been proposed [2, 18, 24, 25]. At both circuit and system levels, a manycore processor can be divided into several small modules, each of which is implemented into a single die. Chiplets bring many benefits to design and fabrication, such as higher yield, lower cost, and greater flexibility. However, chiplets also introduce new challenges, among which inter-chiplet communication is one of the most urgent problems.

With more cores integrated into a single chip and working in parallel, there is a significant increase in memory read and write activities. Therefore, to maintain cache coherence, many extra cache coherence messages are generated, such as invalidation, forwarding, and acknowledgment, forcing heavy pressure on the interconnect network. Additionally, the coherence related messages are sensitive to the transmission latency, thus, the longer data transmission latency caused by the increasing network scale exacerbating the impact of cache coherence and resulting in higher memory access latency. Researchers have proposed new protocols or modified existing protocols [6, 10] to reduce coherence related messages and transactions. However, there lacks systematic understanding of the impacts of the interconnect network on cache coherence, which poses a fundamental constraint on the efficiency of cache coherence protocol.

Traditional electrical networks such as Mesh and Ring are commonly used in commercial manycore processors. However, the latency and bandwidth of these traditional networks can not meet the requirement for efficient cache coherence and the scalability is limited when the number of nodes increases to hundreds even thousands [29]. The optical interconnect with low latency, high energy efficiency and a large bandwidth [4, 26] is a promising technology to solve the two aforementioned problems: inter-chiplet communication and cache coherence overhead. Optical interconnect supports long-distance point-to-point communication without the need for repeaters. Thus, the optical signal can quickly travel anywhere on the chip and the energy consumption is largely independent to the distance [29]. In addition, optical interconnect shows higher bandwidth density than electrical interconnect especially when wavelength division multiplexing (WDM) technology is applied, which can provide large bandwidth and alleviates the chip pin constraint on the package [26]. Further, different from electrical interconnects, optical interconnects are suitable for both effective unicast and multicast transmission because of their special wavelength routing mechanism. Thus, optical networks can effectively transmit multicast coherence messages such as invalidation.

Some optical networks have been proposed in the past. However, these designs target high-performance networks ignoring the non-uniform nature of the on-chip traffic pattern caused by memory hierarchy and the cache coherence protocol. Thus, these works can not properly alleviate cache coherence overhead and may even consume extra optical resources. There are also a few works focusing on accelerating the transmission of specific coherence related messages such as invalidation and forwarding [1, 14]. Nevertheless, these works for optimization of coherence message transmission focus on accelerating a small portion of cache coherence messages and the improvement in the system performance is limited.

Our work aims to accelerate cache coherence for manycore processors through silicon photonic chiplets. We propose a photonic cache coherence network, called PCCN, which highly accelerates coherence transactions and reduces average memory access latency, leading to better performance and higher energy efficiency in manycore system. Specifically, our contributions are:

ICCAD '22, October 30-November 3, 2022, San Diego, CA, USA © 2022 Copyright is held by the owner/author(s). Publication rights licensed toACM.
ACM 978-1-4503-9217-4/22/10…$15.00
https://doi.org/10.1145/3508352.3549338

- We systematically analyze the traffic pattern from the perspective of the cache coherence protocol and memory hierarchy in manycore processor.
- We propose a chiplet-based manycore processor with a physically centralized and logically distributed directory, which can improve the utilization rate of optical interconnect and effectively transmit cache coherence related packets through the optical interconnect network.
- We design a WDM-based channel-sharing optical interconnect network-based on cache coherence traffic pattern and an opto-electrical router architecture.
- We quantitatively compare the performance, memory access latency, and energy consumption of the PCCN-based system with the previously proposed optical network-based system and the traditional electrical network-based system.

The rest of the paper is organized as follows. Section 2 introduces optical interconnect background. Section 3 analyses the traffic pattern in manycore processors. Section 4 discusses our proposed PCCN-based manycore system. Section 5 quantitatively evaluates the PCCN-based system. Section 6 gives a survey. Section 7 makes a conclusion.

2 OPTICAL INTERCONNECT BACKGROUND

A typical optical interconnect consists of three main parts: E-O interface, transmission channel, and O-E interface, as is shown in Fig. 1. At the sender end, parallel low-frequency electrical signals are modulated into serial high-frequency optical channels through E-O interface. We adopt optical waveguides to be the transmission channel for optical light in our design, which can be manufactured in a standard CMOS process [9] with the losses of 0.2dB/mm. Optical waveguides support long-distance point-to-point data transmission which eliminates the need for multiple hops between two nodes in large-scale system, leading to lower transmission latency than its electrical link counterpart. In addition, multiple wavelengths can be transmitted in a single waveguide simultaneously without communication contention using Wavelength Division Multiplexing (WDM) technology, which can highly improve network bandwidth. At the receiver end, the optical filter extracts the optical light of a specific wavelength and transfers it to the photo detector. The filter is implemented by micro-rings that can switch optical light with specific wavelength from one waveguide to another waveguide. Finally, the photo detector in O-E interface converts the optical signal into the electrical signal. The speed of the traditional E-O/O-E

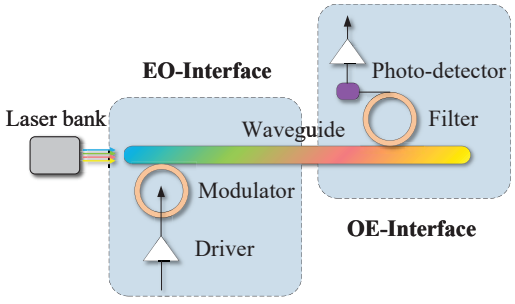


Figure 1: The structure of optical interconnect

Table 1: Optical Device Parameter

Parameter	Value
Microresonator passing loss	0.05 dB
Microresonator insertion loss	2 dB
Microresonator heat tuning power	0.65 mW
Waveguide propagation loss	2 dB/cm
Optical pin coupling loss	1.5 dB
Receiver sensitivity	-20 dBm
Laser power conversion efficiency	30%
Data rata per wavelength	32 Gbps

interface is limited when working in high frequency, which is a bottleneck in optical networks. We apply the optical weaving E-O and O-E interface [27], the conversion latency of which can be limited within one cycle in the electrical domain. Table 1 shows the important parameters of the optical devices used in PCCN [5, 22, 31].

3 MOTIVATION

In this section, we introduce the motivation of our work from the perspective of traffic distribution in manycore processors. In a typical manycore processor, each L1 cache is private, each L2 cache is private or shared by its neighbouring L1 caches, and there is a shared last-level cache containing several distributed L3 cache slices. The cache coherence problem exists between two adjacent layers in cache hierarchy such as the L1, L2 caches and L2, L3 caches. Thus, a two-level hierarchical cache coherence protocol is applied. Because there are only several L1 caches sharing a home L2 cache and the L1 caches are close to their home L2 cache, the cache coherence problem between the L1 cache and L2 cache is not a heavy burden. By contrast, the last-level cache is shared by all the L2 caches and many L2 caches are far away from the last-level cache. Thus, the large portion of long-distance cache coherence related packets will cause large memory access latency. If the transmission distance of a packet is more than three hops in a mesh-based electrical network, the packet is defined as a long-distance packet in this work. Transmission latency of long-distance packets can be extremely reduced using optical interconnect compared to electrical interconnect. Fig. 2 shows the proportion of long-distance packets in a 128-core system running Splash3 benchmark, where each core has a private 64KB L1 I/D cache and 512KB L2 cache, and there are 16 8MB L3 slices distributed in the system and shared by all

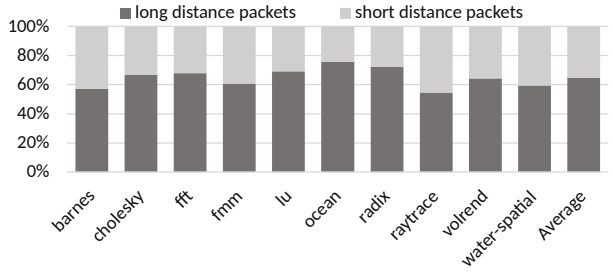


Figure 2: The portion of long-distance packets in a 128-core system running Splash3 benchmark

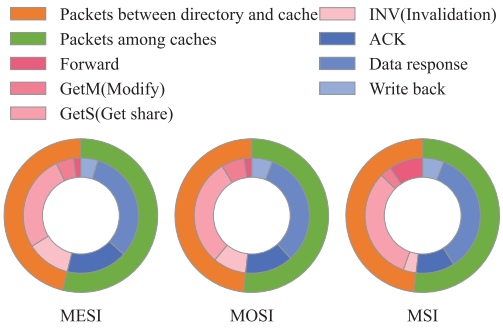


Figure 3: The traffic distribution in manycore processor from the perspective of cache coherence.

the cores. It can be observed that about 63 % of packets are long-distance packets on average and this percentage can even reach 78 % in some applications such as ocean. Thus, if all the long-distance transmission is optimized by optical interconnect, the cache coherence overhead can be extremely alleviated leading to a great improvement in performance.

The directory-based cache coherence protocol is commonly used in manycore processors to avoid broadcast overhead. The directory keeps the data sharing information and owner information. But extra directory access latency is introduced in coherence transactions. To design a manycore processor architecture applicable to most cache coherence protocols, we analyse the traffic pattern of three typical directory-based cache coherence protocols (MSI, MOSI, MESI) in this section. Although different protocols can lead to different traffic patterns, they have some common characteristics. Fig. 3 shows the traffic distribution of Splash benchmark in the 128-core system. In terms of packet size, there are two kinds of packets: control packets and data packets. The data packet is normally large, containing multiple flits, while the control packet is normally small, containing only one flit. Among all kinds of cache coherence packets, data response and write back are data packets and the other packets are control packets. Invalidation packets demand both multicast and unicast transmission, and other packets only demand unicast transmission. In terms of the source and the destination of packets, traffic can be divided into two parts: packets among caches and packets between caches and the directory. All the packets between the directory and caches are control packets. Most packets among caches are data packets and have a high demand for transmission bandwidth.

4 PCCN ARCHITECTURE

This section introduces the architecture of PCCN, including the architecture overview, physically centralized and logically distributed directory, chiplet architecture, optical interconnect network and opto-electrical router.

4.1 Architecture Overview

PCCN is a chiplet-based manycore processor with a physically centralized and logically distributed directory. Fig. 4 shows the overview of PCCN. Cores, caches, and memory controllers are integrated into several core chiplets. All the directory slices are integrated into one separated chiplet. Electrical routers and opto-electrical routers within the same chiplet are connected by an electrical network, and all the opto-electrical routers are connected by two optical networks.

4.2 Physically Centralized and Logically Distributed Directory

Traditional manycore systems use a physically distributed directory consisting of many distributed directory slices to get achieve high throughput, each of which keeps the corresponding L3 cache slice coherent. However, there are two bottlenecks in this design. First, with the increase in system scale, the long distance between the directory and some L2 caches leads to a large access latency. In addition, there are some multicast control packets, but multicast transmission in the traditional packet-switched electrical network is not efficient. Thus, the frequent communication exists between the directory and L2 caches because of the large amount of coherence transactions, which limits the system performance. To address these problems, we design a physically centralized and logically distributed directory, as is shown in Fig. 4, which achieves effective data transmission and has higher throughput than transitional centralized directory. All the directory slices are put together and there is an optical network connecting the directory and caches. The optical signal can quickly travel between caches and the directory through the optical network with low energy consumption, which is highly suitable for long-distance data transmission. Additionally, the optical signal can be received by many nodes at the same time, which achieves effective multicast transmission. The transmission latency between the directory and L2, L3 caches in a 128-core manycore system is no more than 19 cycles and 28 cycles respectively, which shows great improvement compared to electrical networks. Further, the directory slices can share the optical channels in our design, which further improves the utilization of the optical channel and reduces energy consumption.

4.3 The Chiplet Architecture

In PCCN, there are two kinds of chiplets, namely core chiplet and directory chiplet, as is shown in Fig. 4. Each core with its private L1 cache and private L2 cache are grouped into a cluster that is attached to a electrical router. Each core chiplet contains 16 clusters, two L3 cache slices, and two memory controllers. There are two kinds of routers in a core chiplet, namely the electrical router and the opto-electrical router, and all routers are connected by an electrical network. The distance between any node and the opto-electrical router is no more than three hops in the electrical network. Directory slices communicate with caches through the optical network. To alleviate the congestion and queuing delay, four opto-electrical routers are integrated into the directory chiplet to handle the traffic for all the directory slices. All chiplets are connected by two optical networks, namely "S" shaped optical network and ring-shaped optical network. Off-chip laser bank is used as light resources, which can emit lights in different wavelengths. The communication between two nodes within the same chiplet is achieved by the electrical network, while the communication between two nodes in different chiplets is achieved by both the

electrical network and the optical network. In inter-chiplet communication, packets are sent to opto-electrical router from the source node through the electrical network first, then transmitted to the opto-electrical router in the destination chiplet, and sent to the destination node through the electrical network finally.

4.4 "S"-shaped Optical Network and Ring-shaped Optical Network

We design two inter-chiplet optical networks with three transmission modes: C-D (Cache to Directory), D-C (Directory to Cache), and C-C (Cache to Cache), for different packets based on their characteristics such as size, source, destination, and unicast or multicast. The "S"-shaped optical network with two transmission modes: C-D and D-C, is designed for communication between core chiplets and the directory chiplet. The ring-shaped optical network with C-C transmission mode handles the traffic among the core chiplets. We group several laser lights as one optical channel and the number of lights in each optical channel is determined by the required bandwidth. We apply the wavelength-routed technique to achieve point-to-point transmission in the two optical networks. Fig. 5 shows the optical channel allocation between opto-electrical routers in core chiplets and one opto-electrical router in the directory chiplet in the "S"-shaped optical network. C-D mode handles the traffic from the core chiplets to the directory chiplet. All the packets in C-D mode are unicast control packets. Thus, each core chiplet is allocated with a four-wavelength optical channel to send packets to the directory chiplet to avoid contention in C-D mode. D-C mode handles the traffic from the directory chiplet to the core chiplets. Packets in D-C mode include both unicast and multicast control packets. The optical channel is suitable for multicasting because the light can be easily filtered by the micro resonators (MRs) with

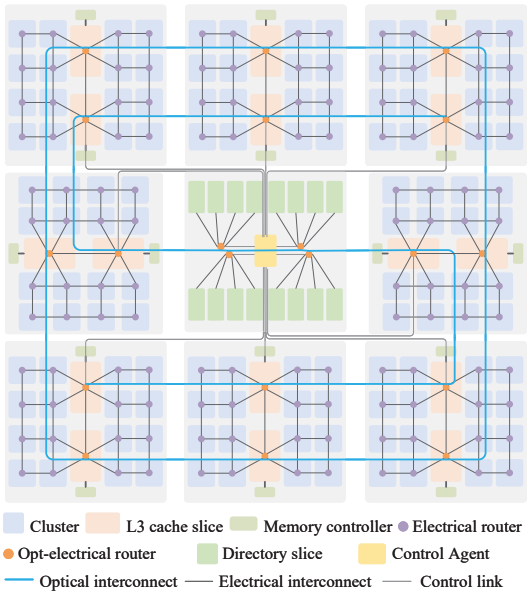


Figure 4: Proposed 128-core PCCN system with a physically centralized and logically distributed directory.

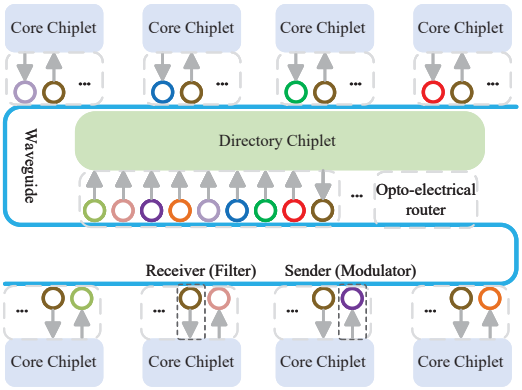


Figure 5: The core chiplet and directory chiplet is connect by an "S"-shaped optical network. The MR modulator and filter in the same color correspond to a sender-receiver pair of a wavelength channel.

the same resonance wavelength, and the resonance wavelength of the MRs can be tuned to be "On/Off" by the electrical signal. Therefore, we design an optical channel supporting both unicast and multicast for each opto-electrical router in the directory chiplet in mode D-C, which is labeled in the color brown in Fig. 5. There is a control agent in the directory chiplet to control the MRs in mode D-C, which can change the resonance wavelength of the MRs. For the data transmission in mode D-C, the opto-electrical router sends destination information to the control agent first and then the control agent sends the configuration message to the destination routers through the control link[8, 19] to "turn on" the MRs in the receiver. In this way, only the targeted core chiplets will receive the packet. The latency over 4 inches of electrical link in 90 nm CMOS is about 2.5 ns in [8, 19]. In our design, the latency over 6 cm of control link in 7 nm CMOS can be no more than 1 ns based on our estimation. To achieve fast and energy-efficient data transmission among core chiplets, we design a channel sharing mechanism in the ring-shaped optical network, which can improve the utilization rate of the optical channels. The eight chiplets are divided into four regions and each region contains two chiplets. The two chiplets in the same region are connected by an optical channel. The four regions are fully connected by eight-wavelength optical channels, and two chiplets in the same region share the optical channels. Fig.

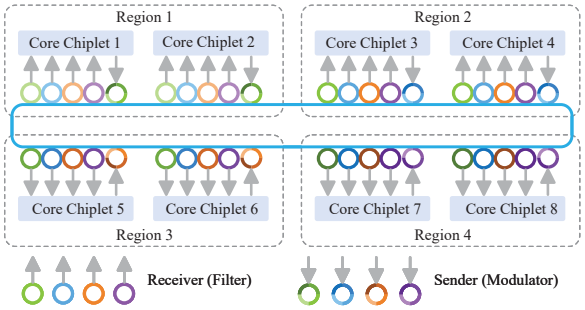


Figure 6: The optical channel allocation in the ring-shaped optical network.

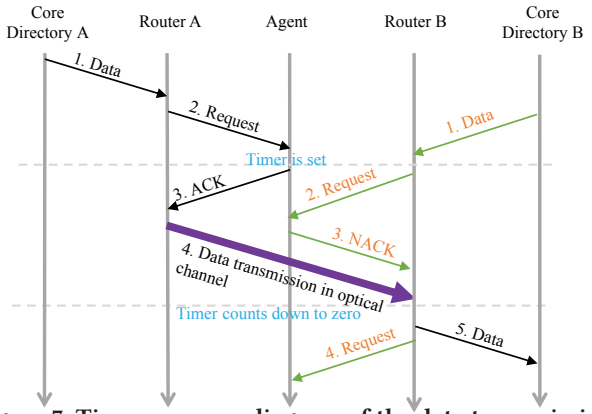


Figure 7: Time-sequence diagram of the data transmission process when contention exists.

6 shows the optical channel allocation in the ring-shaped optical network. Due to the channel sharing between the two chiplets in the same region, channel contention for the optical channels exists between the two chiplets.

To solve the contention problem, we propose a control and arbitration mechanism for the ring-shaped optical network. There is a control and arbitration agent in each region integrated into the opto-electrical router. Fig. 7 shows the data transmission process with channel contention. If a core chiplet A wants to send a packet through opto-electrical router A, the packet is sent to the waiting buffer in optical router A first. Then, a request is sent to the agent to check whether the optical channel is available. If the channel is available, the agent sends an ACK message to the waiting buffer and sends configuration information to set up the sender at the same time. The timer in the agent will also be set to show that the optical channel is occupied. The occupation time depends on the laser wakeup latency, EO/OE conversion latency and packet transmission latency. The laser wakeup latency and EO/OE conversion latency are fixed. The packet transmission latency is determined by the packet size and the bandwidth of the optical channel. Before the timer counts down to zero, any request on this optical channel will be rejected. In Fig. 7 Router B sends a request to the agent when the optical channel is unavailable. The agent will send back a NACK message containing time information about the rest cycles of the timer to router B. After the given time, router B will send the request to the agent again.

4.5 Opto-electrical Router Architecture

We propose an opto-electrical router architecture to support the PCCN. The opto-electrical router has two functionalities: 1. transmit packets in the electrical network. 2. send and receive inter-chiplet packets from the optical network. Fig. 8 shows the structure of an opto-electrical router in the ring-shaped optical network. Each router has four receivers (O-E interface) and one sender (E-O interface), which is based on [28]. Each receiver consisting of a group of MRs and photodetectors can receive data from a specific optical channel, and the sender consisting of a group of MRs can modulate the parallel electrical signals to the specific optical channel by tuning the resonance wavelength of MRs in the sender. The architecture of opto-electrical router in the "S"-shaped optical network is similar. In addition to the hardware design, there is a small change in the routing algorithm.

- The packets are classified into intra-chiplet packets and inter-chiplet packets based on the source and destination information in the packet head.
- Intra-chiplet packets are transmitted through the electrical network. Inter-chiplet packets are sent to the opto-electrical router through the electrical network and then transmitted to the destination chiplet through the optical network.

5 RESULT

In this section, we compare the PCCN-based system to the two most representative manycore system: mesh-based electrical network system and the latest proposed PhotoBNoC-based system [1]. Mesh-based electrical network system with great performance and scalability is widely applied to commercial processors, which is the baseline in our experiment to outline the benefit of implementing optical links. PhotoBNoC-based system comes closest to our goal of designing an optical-electrical hybrid network for cache coherence acceleration, which combines an SWBR (single write broadcast read) optical network with a mesh-based electrical network. To get a fair comparison, these three systems are implemented with the same core number, core type, cache configurations, and cache coherence protocol. Four wavelengths are grouped as one optical channel for control packets and optical weaving E-O/O-E interface is applied in PhotoBNoC-based system, which are the same as PCCN-based system. We quantitatively evaluate the performance, memory access latency, and energy consumption of these three systems.

5.1 Experiment Setup

We evaluate the performance of three systems by a cycle-accurate simulator JADE5.0 [15] with the splash3 benchmark. We modified JADE5.0 to implement PCCN and PhotoBNoC. Table 2 summarizes the detailed configurations of manycore systems in our experiment. These three manycore systems apply a three-level inclusive cache hierarchy based on the TILE-Gx72 manycore processor. In PCCN and PhotoBNoC, there is a 0.65mW heater placed close to each opto-electrical router to stabilize the temperature and tolerate variations in the optical channel. To estimate power dissipated by the

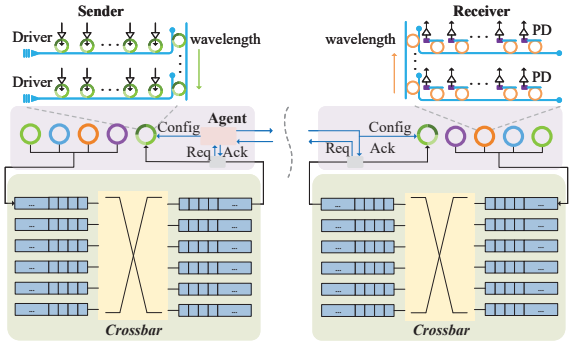


Figure 8: The opto-electrical router in the ring-shaped optical network.

processor cores and cache hierarchy of the many-core system when running the application workloads, JADE5.0 comes interfaced with the latest McPAT 1.0 tool [12]. To evaluate power dissipation under 7 nm technology, we use standard technology scaling rules to scale down the power value provided by McPAT 1.0 at 22 nm to target the 7 nm technology node [16]. In addition, the latency and energy consumption of the electrical interconnect and the optical interconnect are analyzed by OEIL [27] based on the parameters in Table 1. The power consumption of optical links includes the power of laser sources, E-O/O-E interfaces, heaters and thermal tuning devices. We calculate the power consumption of laser sources by finding the worst-case power loss of any possible optical channels. The energy consumption of memory is calculated based on Micron's DDR4 power model [17]. To prove that our design is suitable for various cache coherence protocols, we implement manycore processors based on three typical directory-based protocols: MSI, MESI, MOSI. To outline the great scalability of PCCN, we also evaluate three different system scales: 64-core, 128-core, and 256-core. In all of the following results, we use the electrical network-based system as a baseline to get the normalized results for all applications.

5.2 Performance and Scalability

The normalized performance of ten applications when running with the 128-core PCCN-based system and the PhotoBNoC-based system respectively is plotted in Fig. 9. The performance is inversely proportional to the execution time. We can observe that both the PCCN system and the PhotoBNoC outperform the traditional electrical network-based system. The main reason is that both PCCN system and PhotoBNoC system utilize optical interconnect to transmit some coherence related packets, which significantly accelerates the coherence transactions and alleviates the cache coherence overhead. It can also be noticed that there is more improvement in the applications with a large portion of remote shared data and high communication intensity, such as Lu, Ocean, and Radix [13, 30]. In these applications, the shared data is read or modified frequently leading to heavy long-distance data transmission, which can highly benefit from optical networks. Further, compared to the electrical network system, the PCCN system shows a 1.32x speedup while the PhotoBNoC system shows a 1.05x on average. Two main reasons limit the performance improvement of PhotoBNoC. First, PhotoBNoC focuses on accelerating the communication of two kinds of packets: Forward and Invalidation, which only takes approximately 15% portion of coherence related traffic, as is shown in Fig. 3 and the rest of the packets are still transmitted by electrical network. In addition, PhotoBNoC builds a fully connected point-to-point optical network among L2 caches and L3 caches without considering the distance. Optical interconnect shows higher latency than electrical interconnect for short-distance transmission due to the E-O/O-E conversion latency of optical interconnect. Therefore, optical interconnect in PhotoBNoC can slow down the data transmission in some cases such as communication between two neighbouring nodes. In PCCN, we carefully analyze the traffic pattern and take the sequential character of the data transmission in coherence transactions into consideration. Thus, all kinds of packets are transmitted properly and the coherence transactions can get maximum acceleration. Further, in PCCN, we apply optical interconnect to

Table 2: System Configuration

Parameter	Value
Processor core	ARMv8 cores, @4 GHz
L1 I/D cache	64 KB/core, private,
L2 cache	512 KB/cache, private
L3 cache	8 MB/L3 slice, shared by all cores
Cache coherence	MSI/MESI/MOSI, keep L2 coherent
Memory	8 GB/module
Router parameter	16 Byte flit size, 4-stage pipeline
Technology	7 nm

Table 3: Network Configuration

Parameter	Value
Flit size	16 Bytes
Routing	X–Y dimension ordered in E-network wavelength routing in O-network
Router	4-stage pipeline, 4 cycles latency
Packet Size	72 Bytes (Data), 8 Bytes (Control)
Control Flow	Virtual-channel, Credit-based backpressure
E-Link Latency	1 cycles
O-Link Latency	3 cycles (E/O+O/E conversion-2, traversal-1)

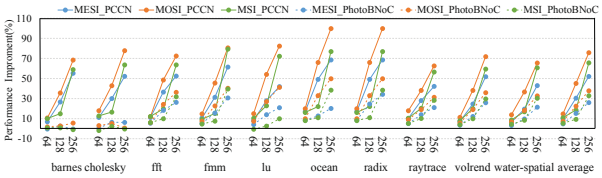


Figure 10: The performance improvement of the 64-core, 128-core, and 256-core PCCN-based system and PhotoBNoC-based system.

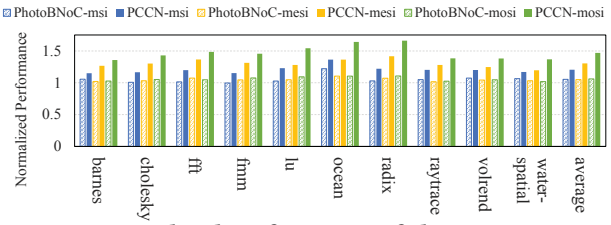


Figure 9: Normalized performance of the 128-core PCCN-based system and PhotoBNoC-based system.

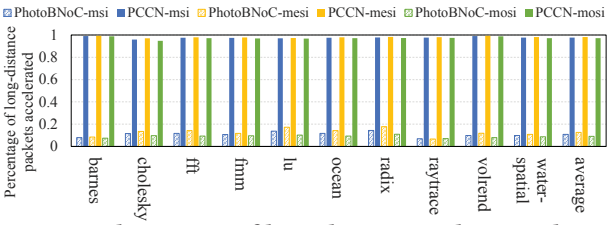
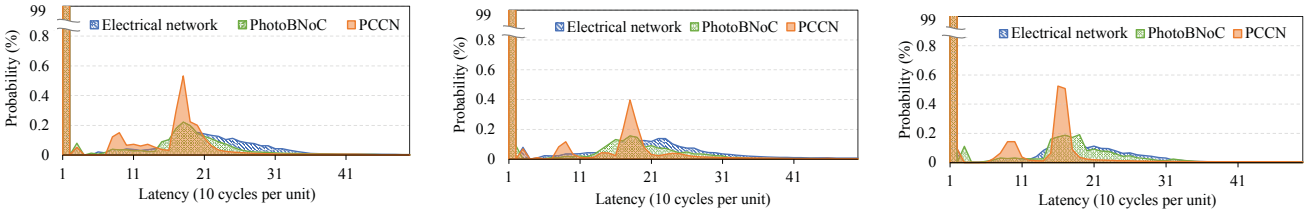


Figure 11: The portion of long-distance packets accelerated in PCCN-based system and PhotoBNoC-based system



(a) MSI (b) MESI (c) MOSI

Figure 12: The probability distribution of memory access latency of the PCCN-based system, PhotoBNoC-based system, and the electrical network-based system

inter-chiplet communication (long-distance) and electrical interconnect to intra-chiplet communication (short-distance), which can fully exert their advantages.

Fig. 10 shows the performance improvement of PCCN and PhotoBNoC compared with the electrical network on a 64-core, 128-core, and 256-core system. Among the three system scale, the 256-core PCCN-based system shows the most considerable 0.65X performance improvement on average and for some communication intensive applications such as radix and ocean, the improvement can even reach 1X. With the number of cores rising from 64 to 256, the PCCN-based system shows a more obvious advantage than PhotoBNoC-based system. The main reason is that with the system scale rising, the data transmission latency accounts for a larger proportion of the total execution time. Thus, manycore systems benefit more from PCCN with the core number increasing, which indicates PCCN has great scalability.

5.3 Average Memory Access Latency

To understand the performance differences explained among the three systems, we analyze the average memory access latency combined with the portion of long-distance packets accelerated by optical interconnect in PCCN and PhotoBNoC, as shown in Fig. 13 and Fig. 2. Memory access latency measures the time for the processor cores getting data from the memory system, which directly affects system performance. Fig. 13 shows the normalized average memory access latency of the PCCN-based system and the PhotoBNoC-based system. On average, the reduction in average memory access latency of the PCCN and PhotoBNoC are 26% and 10% respectively compared to the electrical network system. Fig. 12 shows the distribution of memory access latency. The latency distribution of the PCCN-based system is more concentrated compared to that of the electrical network system and PhotoBNoC-based system because the latency of packet transmission in the electrical network is sensitive to physical distance. If the source is far away from the destination, packets need to pass many hops in the electrical network. Different from the electrical network, packets are transmitted in a point-to-point way in optical network whose latency is relatively independent of distance and more fixed. Thus, PCCN brings an additional benefit to system designers that the latency of PCCN is more predictable. There are four peaks in the latency distribution of PCCN, which represents the average latency of the processor cores getting data from the L1, L2, L3 caches, and memory modules. Compared to the electrical network and PhotoBNoC, PCCN effectively optimizes the access latency from the L3 cache and memory module. Fig. 11 shows that PCCN accelerates almost all the long-distance packets while the packets accelerated by PhotoBNoC only take a small percentage, which leads to differences in the memory access latency between PCCN and PhotoBNoC.

5.4 Energy Consumption and Energy Delay Product

Fig. 15 shows that PCCN-based system can save about 21% energy and PhotoBNoC-based system consumes 6.5% more energy compared to electrical network-based system. About 80 percent of energy is consumed by processor cores and network. PCCN can effectively reduce the energy consumption of the processors, caches, and memory compared to the electrical network because the energy consumption of these parts is related to execution time and PCCN can greatly improve the system performance. Fig. 16 shows that PCCN consumes 42% less energy than PhotoBNoC because the optical network in PCCN is more energy efficient. There are two main reasons. Considering the low utilization rate of the optical interconnect, PCCN applies the channel sharing mechanism, which reduces the network scale and improves the energy efficiency. In addition, PCCN designs two optical networks with the physically centralized and logically distributed directory to effectively transmit different kinds of packets, which further reduces the energy

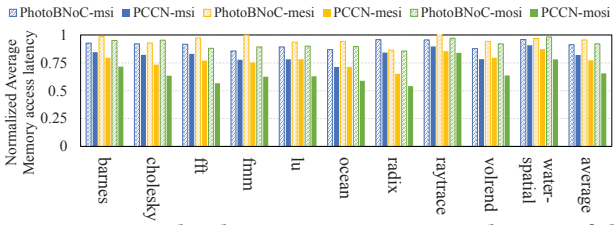


Figure 13: Normalized average memory access latency of the PCCN-based system and the PhotoBNoC-based system.

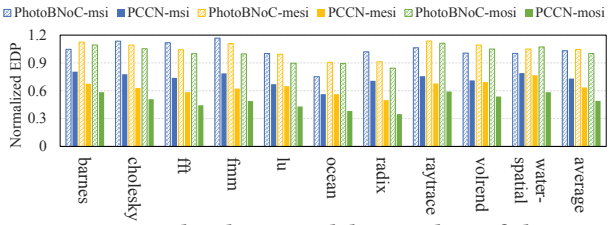


Figure 14: Normalized energy delay product of the PCCN-based system and the PhotoBNoC-based system.

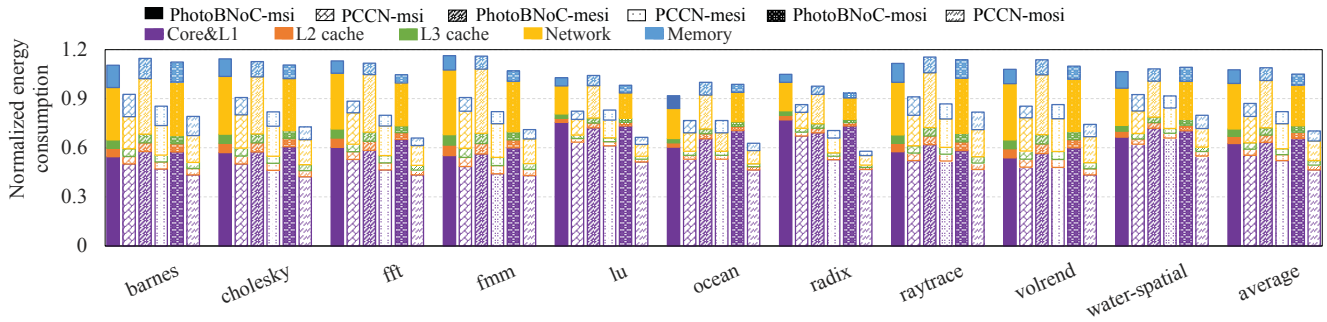


Figure 15: Normalized energy consumption of the PCCN-based and PhotoBNoC-based systems including the energy of processors, caches, network and memory.

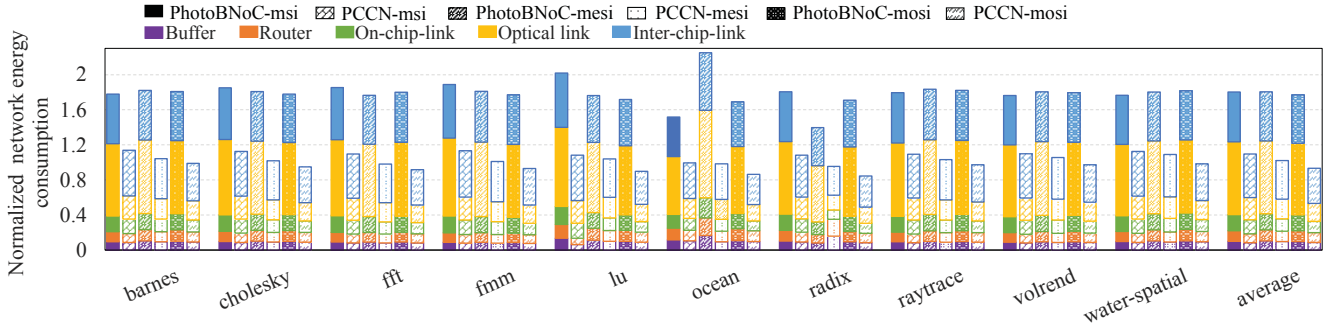


Figure 16: Normalized energy consumption of the PCCN and the PhotoBNoC including the energy of the routers, buffers, on-chip links (on-chip electrical network), inter-chip links (interconnect between processor chip and memory chip) and optical links.

consumption. From Fig. 14, it can be noticed that PCCN-based system has the lowest energy-delay-product (EDP) among the three systems and the EDP of PCCN-based system is even 2X better than that of electrical network-based system. All these results show that PCCN can largely accelerate the cache coherence with high energy efficiency in manycore processors, which can be applied to various protocols.

6 RELATED WORKS

Cache coherence protocol A major obstacle that restricts the efficiency of cache coherence protocol is long-distance transmission during coherence transactions. Many optimizations have been proposed to accelerate cache coherence to reduce coherence related messages such as invalidation, ack, and forwording [11, 20, 21, 33]. However, these works reduce a small portion of packets and a large portion of long-distance packets still exist, which results in the limited improvement. Some work focus on reducing the transmission latency of coherence-related messages by moving the data close to the most active requestors [10]. Abhishek Das, et al propose a dynamic directory to colocate directory entry with the most active requestors of the corresponding cache blocks, which eliminates unnecessary network traversals [7]. Extra data movement and control overhead is introduced in these work, which may cause congestion in the network running communication-intensive applications. In our work, we adopt optical interconnect to optimize the transmission of all kinds of coherence-related packets based on their characteristics, fundamentally improving the efficiency of cache coherence protocol, which can be widely applied to various cache coherence protocols.

Optical interconnect A few works have been proposed to accelerate the cache coherence using optical interconnect [1, 11, 32]. These works adopt a SWMR (Single-Write-Multiple-Read) crossbar topology for all transmission ignoring the characteristics of coherence-related packets such as the amount, size, and multicast/unicast. Thus, some optical channels are underutilized but congestion is existing in some other optical channels, which influences the energy efficiency and system performance. PCCN adopts channel-sharing mechanism to improve the utilization of optical channels and carefully allocate optical resources for different kinds of packets based on traffic analysis to improve energy efficiency and performance.

7 CONCLUSION

In this work, we propose PCCN, a silicon photonic chiplet-based manycore processor with a physically centralized and logically distributed directory to alleviate cache coherence overhead. Compared to state-of-the-art proposals, PCCN shows a 1.32x speedup in performance, 26% reduction in memory access latency and 1.26x improvement in energy efficiency.

8 ACKNOWLEDGMENTS

This work is partially supported by ACCESS.

REFERENCES

[1] José L Abellán et al. 2018. Photonic-based express coherence notifications for many-core CMPs. *J. Parallel and Distrib. Comput.* 113 (2018), 179–194.

[2] Noah Beck et al. 2018. 'Zeppelin': An SoC for multichip architectures. In *ISSCC*.

[3] Shekhar Borkar. 2007. Thousand core chips: a technology perspective. In *DAC*.

[4] Guoqing Chen et al. 2007. Predictions of CMOS compatible on-chip optical interconnect. *Integration* 40, 4 (2007), 434–446.

[5] Corning. 2014. Corning® Single-Mode Optical Fiber. *Technical Publication* (2014).

[6] Blas Cuesta et al. 2011. Increasing the effectiveness of directory caches by avoiding the tracking of noncoherent memory blocks. *TC* 62, 3 (2011), 482–495.

[7] Abhishek Das et al. 2012. Dynamic directories: A mechanism for reducing on-chip interconnect power in multicores. In *DATE*.

[8] Yigit Demir et al. 2014. Galaxy: A high-performance energy-efficient multi-chip architecture using photonic interconnects. In *International Conference on Supercomputing*.

[9] Randolph Kirchain and Lionel Kimerling. 2007. A roadmap for nanophotonics. *Nature Photonics* 1, 6 (2007), 303–305.

[10] George Kurian, Omer Khan, and Srinivas Devadas. 2013. The locality-aware adaptive cache coherence protocol. In *Proceedings of the 40th Annual International Symposium on Computer Architecture*. 523–534.

[11] George Kurian, Jason E Miller, James Psota, Jonathan Eastep, Jifeng Liu, Jurgen Michel, Lionel C Kimerling, and Anant Agarwal. 2010. ATAC: A 1000-core cache-coherent processor with on-chip optical network. In *2010 19th International Conference on Parallel Architectures and Compilation Techniques (PACT)*. IEEE, 477–488.

[12] Sheng Li. 2009. McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures. In *ISCA*. 469–480.

[13] Zheng Li et al. 2008. PARSEC vs. SPLASH-2: A quantitative comparison of two multithreaded benchmark suites on chip-multiprocessors. In *International Symposium on Workload Characterization*.

[14] Zheng Li et al. 2009. Spectrum: a hybrid nanophotonic-electric on-chip network. In *DAC*.

[15] Zheng Li et al. 2016. JADE: A heterogeneous multiprocessor system simulation platform using recorded and statistical application models. In *AISTECS*.

[16] Jie Meng, Chao Chen, Ayse Kivilcim Coskun, and Ajay Joshi. 2011. Run-time energy management of manycore systems through reconfigurable interconnects. In *Proceedings of the 21st edition of the great lakes symposium on Great lakes symposium on VLSI*. 43–48.

[17] Micron. 2017. TN-40-07: Calculating Memory System Power for DDR4 SDRAM. *Technical Publication* (2017).

[18] Samuel Naffziger et al. 2021. Pioneering Chiplet Technology and Design for the AMD EPYC™ and Ryzen™ Processor Families: Industrial Product. In *ISCA*.

[19] John Poulton et al. 2007. A 14-mW 6.25-Gb/s transceiver in 90-nm CMOS. *IEEE Journal of Solid-State Circuits* 42, 12 (2007), 2745–2757.

[20] Alberto Ros and Stefanos Kaxiras. 2012. Complexity-effective multicore coherence. In *Proceedings of the 21st international conference on Parallel architectures and compilation techniques*. 241–252.

[21] Alberto Ros and Stefanos Kaxiras. 2015. Callback: Efficient synchronization without invalidation with a directory just for spin-waiting. In *2015 ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 427–438.

[22] Clint L Schow et al. 2011. A 24-channel, 300 Gb/s, 8.2 pJ/bit, full-duplex fiber-coupled optical transceiver module based on a single "holey" CMOS IC. *Journal of Lightwave Technology* 29, 4 (2011), 542–553.

[23] Lisa Su. 2019. Delivering the future of high-performance computing. In *2019 IEEE Hot Chips 31 Symposium (HCS)*.

[24] Pascal Vivet et al. 2020. A 7-nm 4-GHz Arm[1]-core-based CoWoS[1] chiplet design for high-performance computing. *Journal of Solid-State Circuits* 55, 4 (2020), 956–966.

[25] Pascal Vivet et al. 2020. IntAct: A 96-core processor with six chiplets 3D-stacked on an active interposer with distributed interconnects and integrated power management. *Journal of Solid-State Circuits* 56, 1 (2020), 79–97.

[26] Zhehui Wang et al. 2015. Improve chip pin performance using optical interconnects. *TVLSI* 24, 4 (2015), 1574–1587.

[27] Zhehui Wang et al. 2016. A holistic modeling and analysis of optical–electrical interfaces for inter/intra-chip interconnects. *TVLSI* 24, 7 (2016), 2462–2474.

[28] Zhehui Wang et al. 2019. CAMON: Low-cost silicon photonic chiplet for manycore processors. *TCAD* 39, 9 (2019), 1820–1833.

[29] Sebastian Werner et al. 2017. Designing low-power, low-latency networks-on-chip by optimally combining electrical and optical links. In *HPCA*.

[30] Steven Cameron Woo. 1995. The SPLASH-2 programs: Characterization and methodological considerations. *SIGARCH computer architecture news* 23, 2 (1995), 24–36.

[31] Qianfan Xu et al. 2005. Micrometre-scale silicon electro-optic modulator. *nature* 435, 7040 (2005), 325–327.

[32] Yi Xu, Yu Du, Youtao Zhang, and Jun Yang. 2011. A composite and scalable cache coherence protocol for large scale CMPs. In *Proceedings of the international conference on Supercomputing*. 285–294.

[33] Hongzhou Zhao, Arrvindh Shriraman, Snehasish Kumar, and Sandhya Dwarkadas. 2013. Protozoa: Adaptive granularity cache coherence. *ACM SIGARCH Computer Architecture News* 41, 3 (2013), 547–558.