# Break Memory Wall Through Silicon Photonics

Jiang Xu

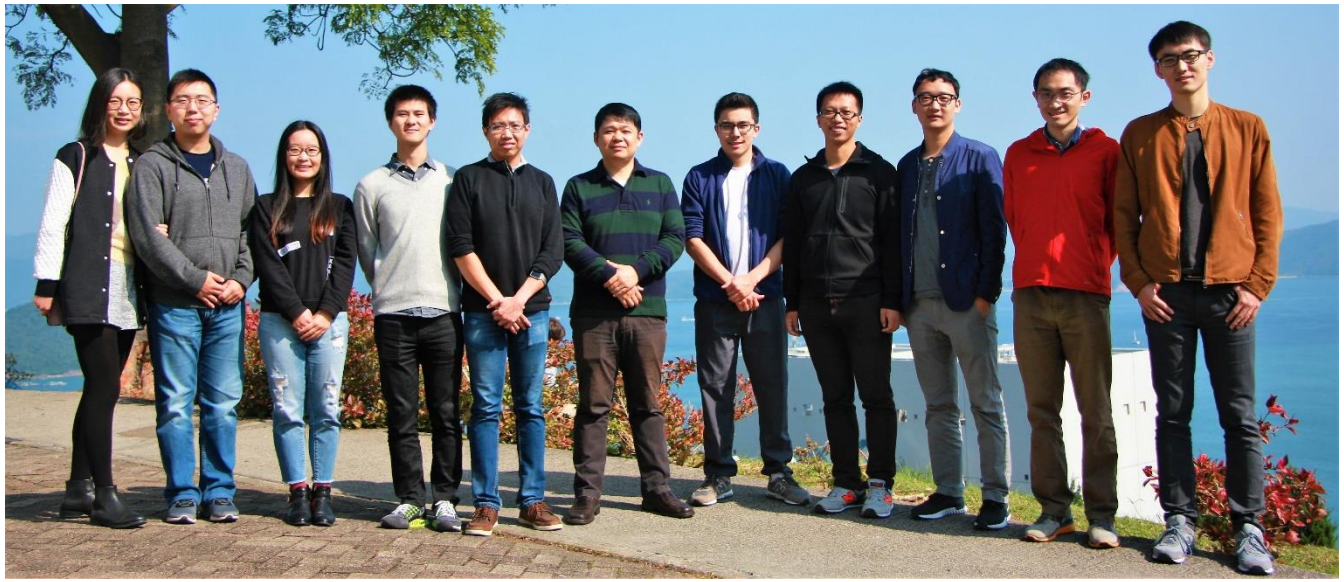
BIG DATA SYSTEM LAB

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

OPTICS Lab

# Acknowledgement


- Current PhD students
  - Zhehui Wang, Peng Yang, Zhifei Wang, Duong Huu Kinh Luan, Haoran Li, Rafael Kioji Vivas Maeda, Xuanqi Chen, Zhongyuan Tian
- Past members
  - Mahdi Nikdast, Yaoyao Ye, Zhe Wang, Weichen Liu, Xiaowen Wu, Xuan Wang, Xing Wen, Kwai Hung Mo, Sébastien Le Beux, Yu Wang, Yiyuan Xie, Huaxi Gu

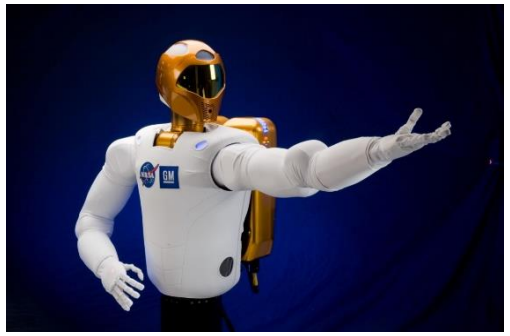
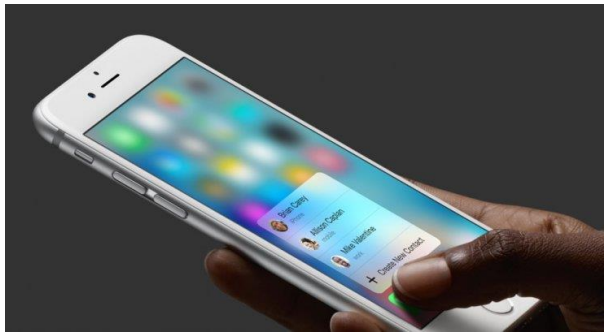

UGC 大學教育資助委員會 University Grants Committee

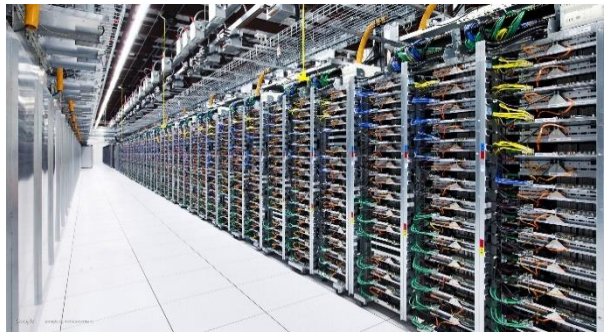
intel

XILINX

# PERFECT Computing Systems for Information Age

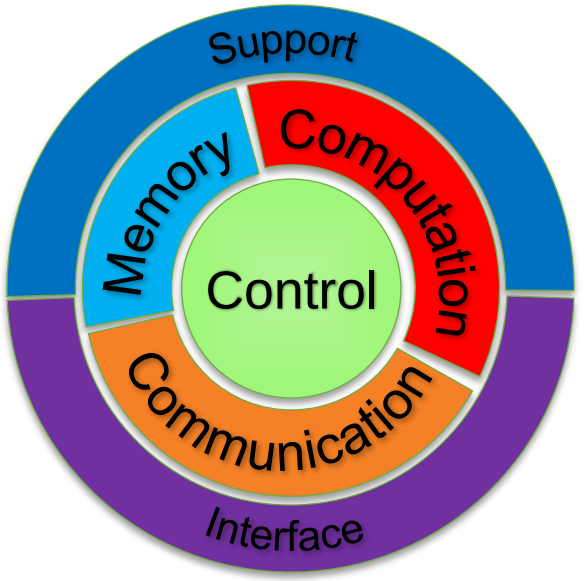
- Performance
- Energy efficiency
- Reliability
- Functionality
- Extensibility
- Cost
- Testability



# Memory Wall is a Major Issue

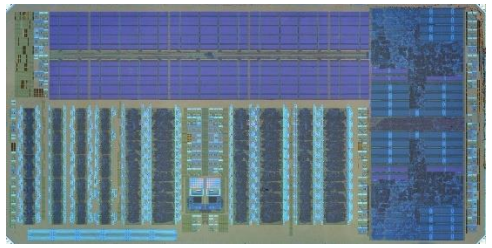
- Computation and memory are physically separated and distributed
  - At node, rack, and system levels
- Mismatch of computation, memory, communication, and support subsystems
  - More/faster processors require more and/or faster memory accesses
  - Distributed subsystems require efficient synchronization schemes
- Worsen by the ending of Moore's Law
  - Cores and caches compete for limited chip area and TDP




Subsystems of a computation system

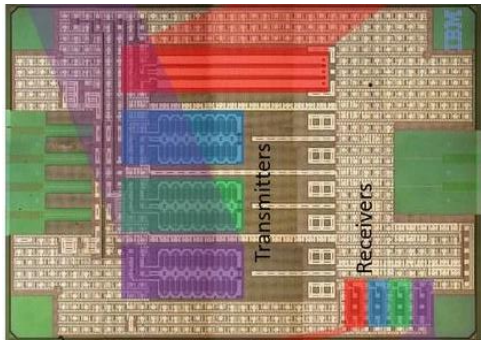
# Silicon Photonics is a Potential Solution

- Benefit from silicon-based technologies and fabs
  - Micron-scale nanosecond-level devices are widely demonstrated
- Active commercialization
  - IBM, Intel (Omni-Path), HPE (Machine), Oracle (UNIC), Cisco, Finisar, Mellanox, ST, NTT, NEC, Fujitsu (PECST), Huawei (DC3.0), ZTE …
  - Startups: Luxtera-ST, Lightwire/Cisco, Kotura/Mellanox, Caliopa/Huawei, Aurrion/Juniper, Rockley, Acacia, OneChip, Skorpios, Ayar, Sicoya, Elenion…
  - PEDA (photonic-electronic design automation): Cadence-PhoeniX-Lumerical, Mentor Graphics-Lumerical-Luceda, RSoft/Synopsys …
  - ST, GF, TSMC, TowerJazz …



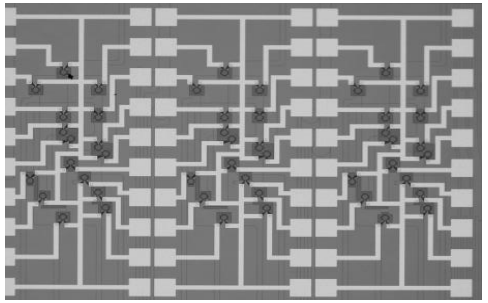
**Integrated OE Interfaces & Processor**

C. Sun *et al.*, "Single-chip microprocessor that communicates directly using light", Nature 2015




**Integrated OE Interfaces**

D.M. Gill, *et al.*, "Demonstration of Error Free Operation Up To 32 Gb/s From a CMOS Integrated Monolithic Nano-Photonic Transmitter", CLEO 2015



**Integrated Optical Switches**

R. Ji, *et al.* "Five-Port Optical Router Based on Microring Switches for Photonic Networks-on-Chip", IEEE Photonics Technology Letters 2013

# Photonics is Different from Electronics

- Advantages
  - Ultra-high bandwidth
  - Low propagation delay
  - Low propagation loss
  - Low sensitivity to environmental EMI
- Disadvantages
  - Electrical/optical conversion
  - Thermal sensitivity
  - Crosstalk noise
  - Process variation
  - Difficult to “buffer”

Differences bring opportunities and challenges

Solkan Bridge, Slovenia 1906
Stone 85/220m

Cold Spring Bridge, USA 1963
Steel 210/371m

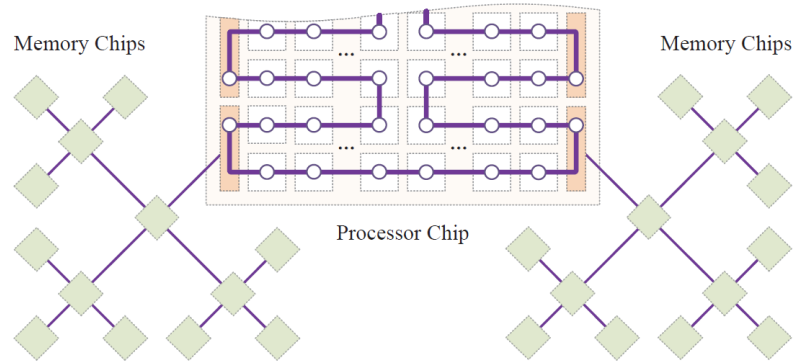
Tsing Ma Bridge, Hong Kong 1997
Steel 1377/2160m

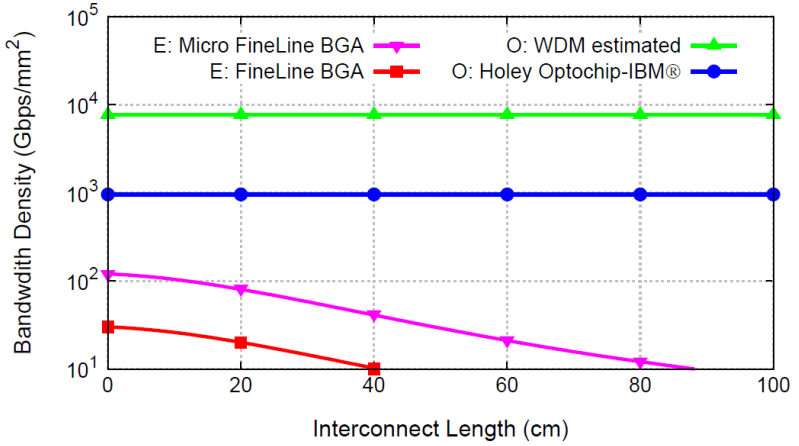
# Outline

- Introduction
- OMIN: optical memory interconnection network
- MOCA overview
- Evaluation and analysis
- Conclusions

# Optical Memory Interconnection Network (OMIN)

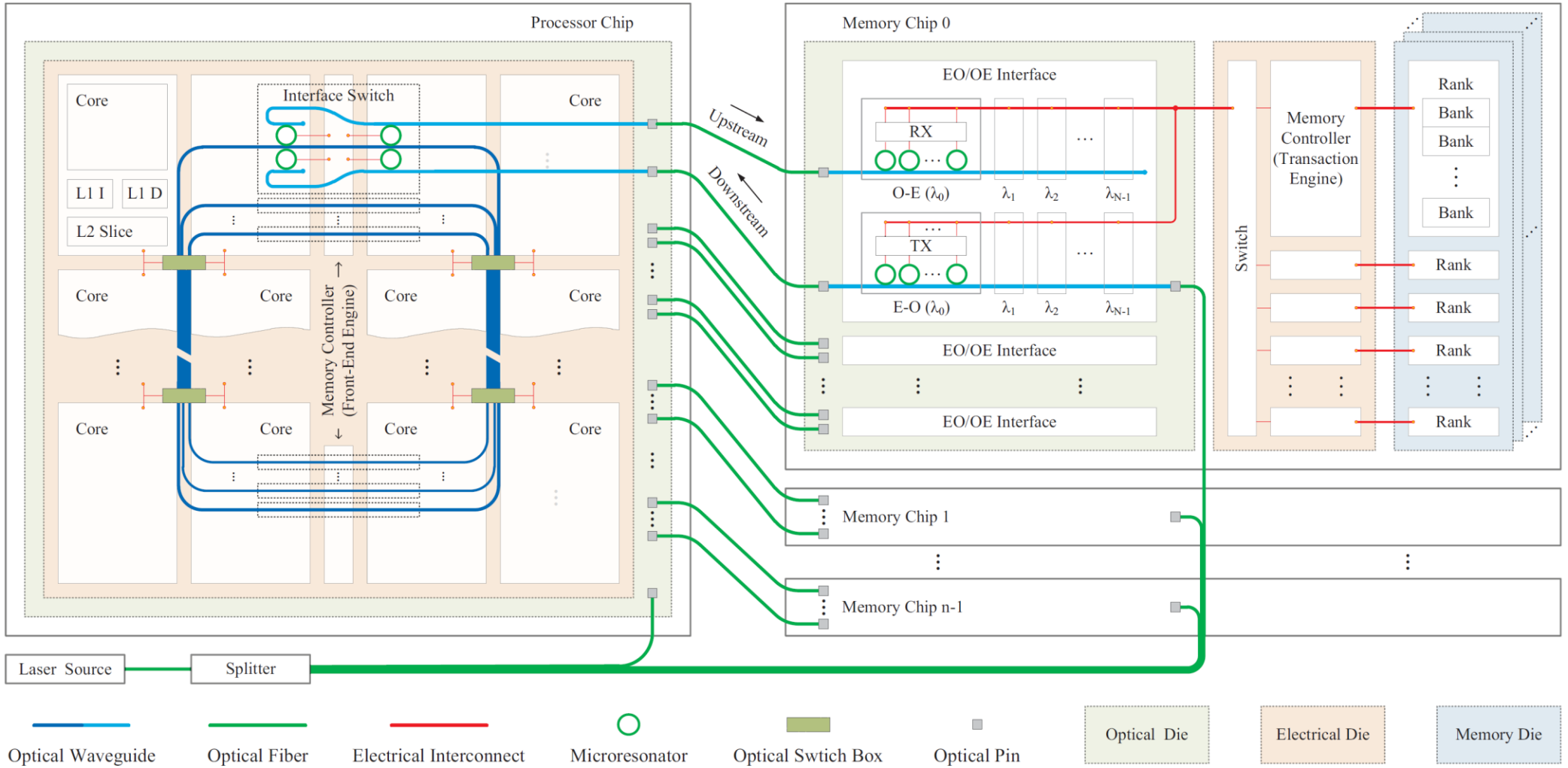
- OMIN connects cache, local memory, and remote memory
  - Based on unified inter/intra-chip and inter-node optical network
- Inter- and intra-chip electrical interconnects are separately designed
  - Limited and expensive chip pins create a sharp chip boundary
  - Different on-chip and off-chip constraints
  - Maximize design flexibility and allow third-party system integration
- Co-design inter/intra-chip and inter-node network to take the full advantages of optical interconnects
  - Avoid buffering and reduce OE/EO conversions
  - Optical chip pins offer 100X~1000X higher bandwidth than electrical chip pins*




* Z. Wang, *et al.*, "Alleviate Chip I/O Pin Constraints for Multicore Processors through Optical Interconnects", ASP-DAC 2015.

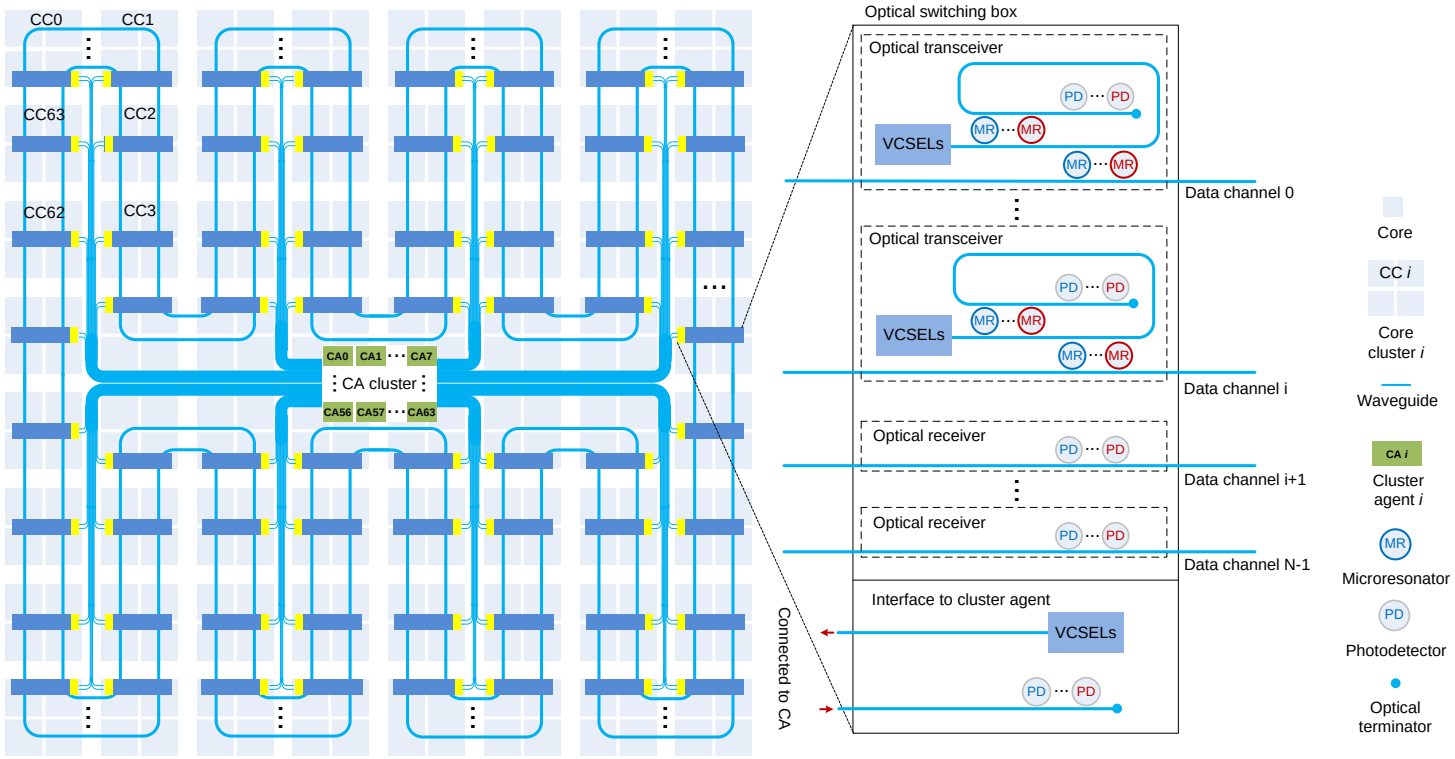
# MOCA: Memory Optical Communication Architecture

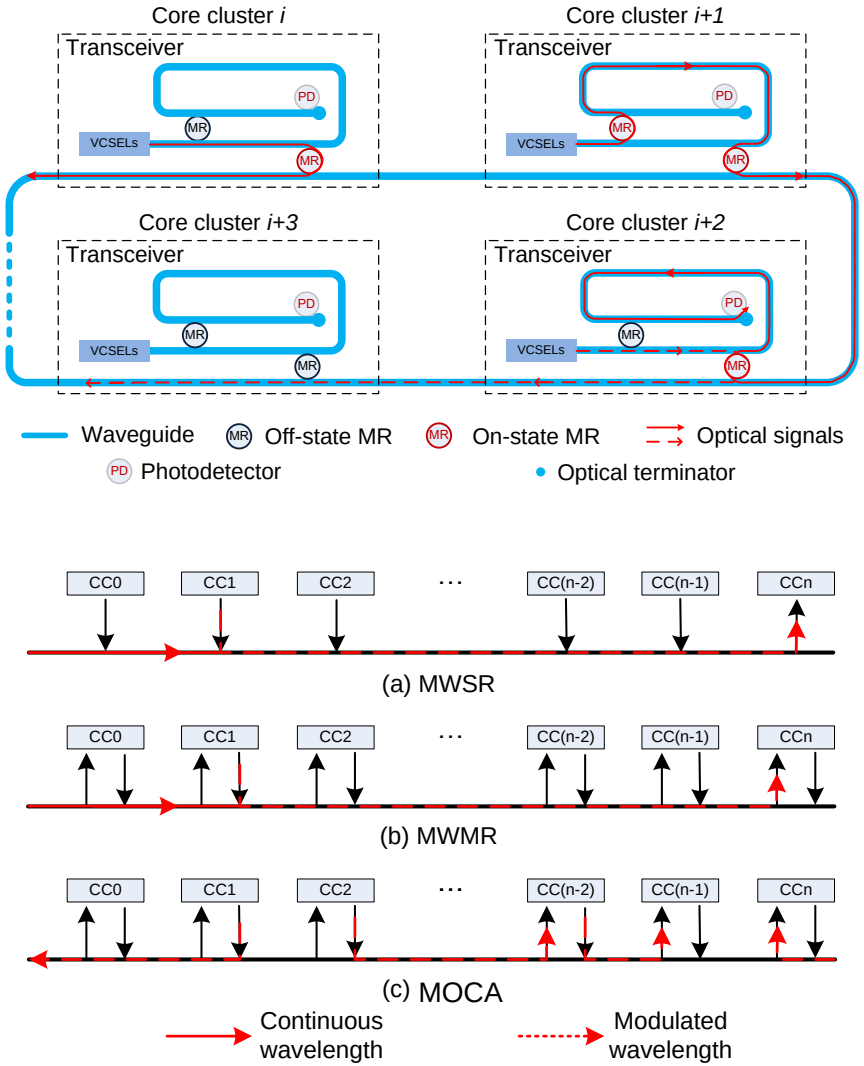



* Z. Wang, Z. Pang, P. Yang, J. Xu, *et al.*, "MOCA: an Inter/Intra-Chip Optical Network for Memory," DAC 2017

# Intra-Chip Optical Network

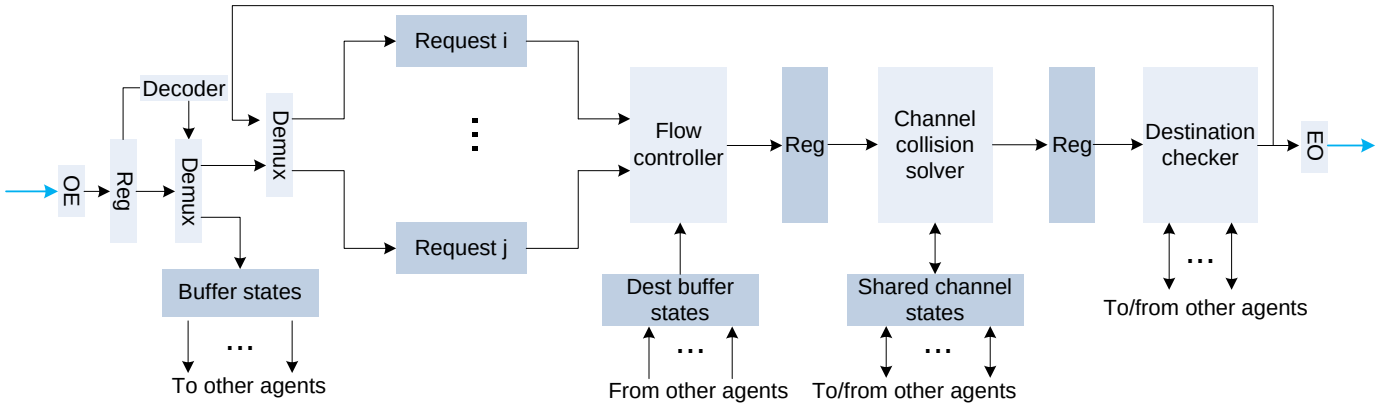
- Segmented bidirectional optical ring
- Multiple simultaneous transactions on one data channel

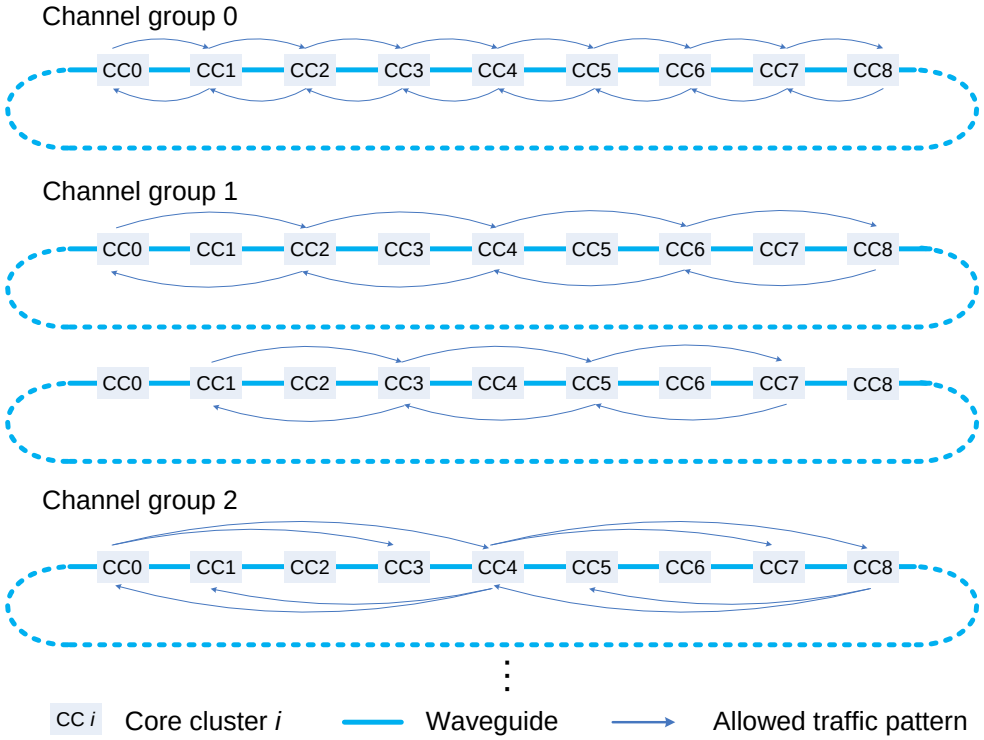


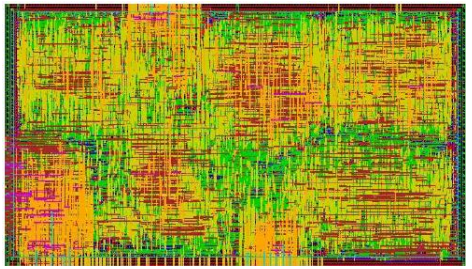



# Physically-Centralized Logically-Distributed Control

- Cluster agents are in the chip center
  - Optically or electrically connected with clusters
  - Electrically connected among each other
- Distributed control algorithm
  - Computational complexity is O(1)
  - 0.0035mm$^2$ and 43μW/GHz at 16nm

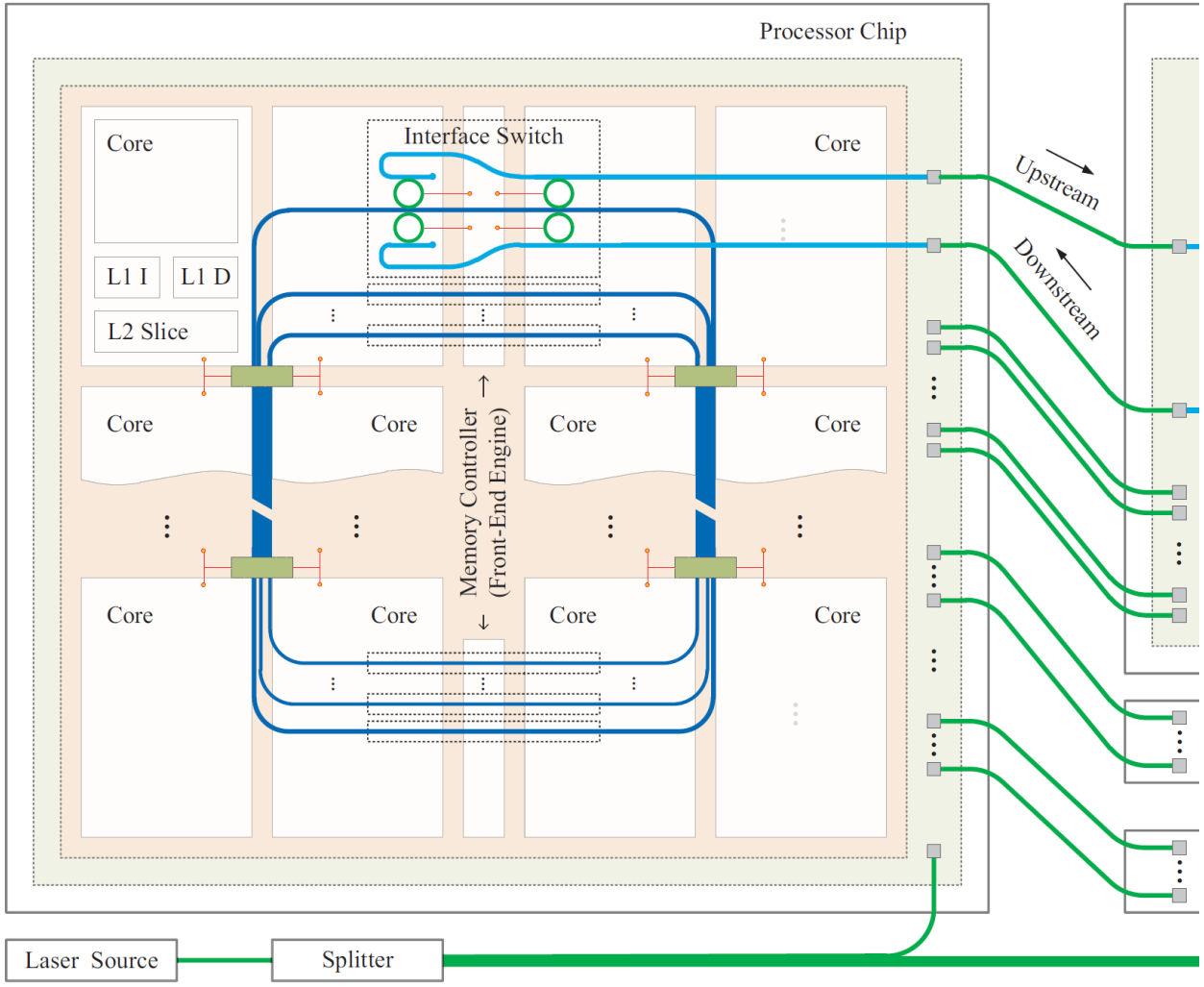


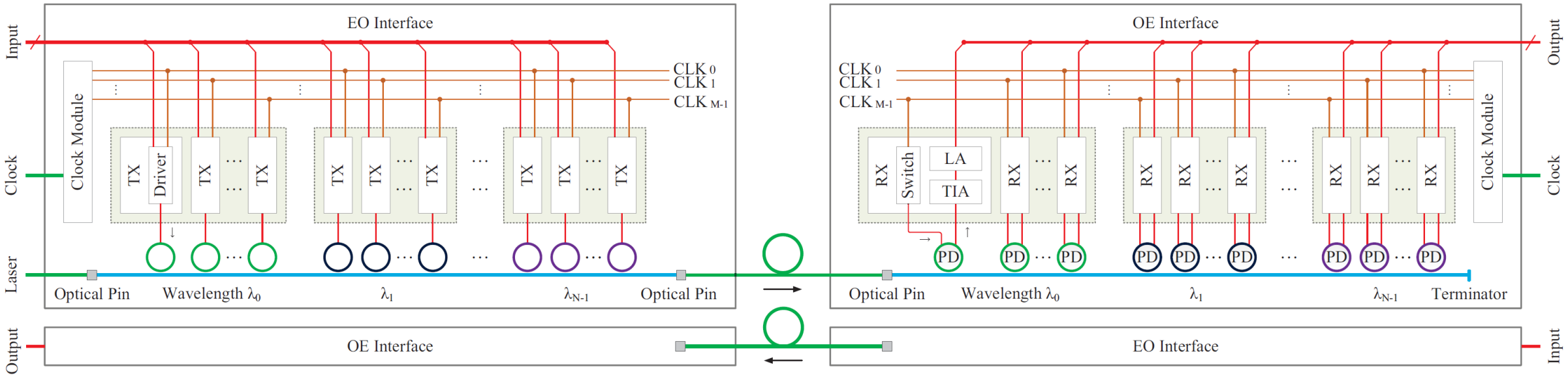



# Laser Source and Clock

- Off-chip laser source
  - Shared by processors and memories
  - On-board centralized
  - Replaceable
  - Improve thermal control
  - Better energy efficiency
- Optical clock distribution
  - Synchronize processors and memories
  - Optical fibers distribute reference clock
  - Low power
  - Low skew



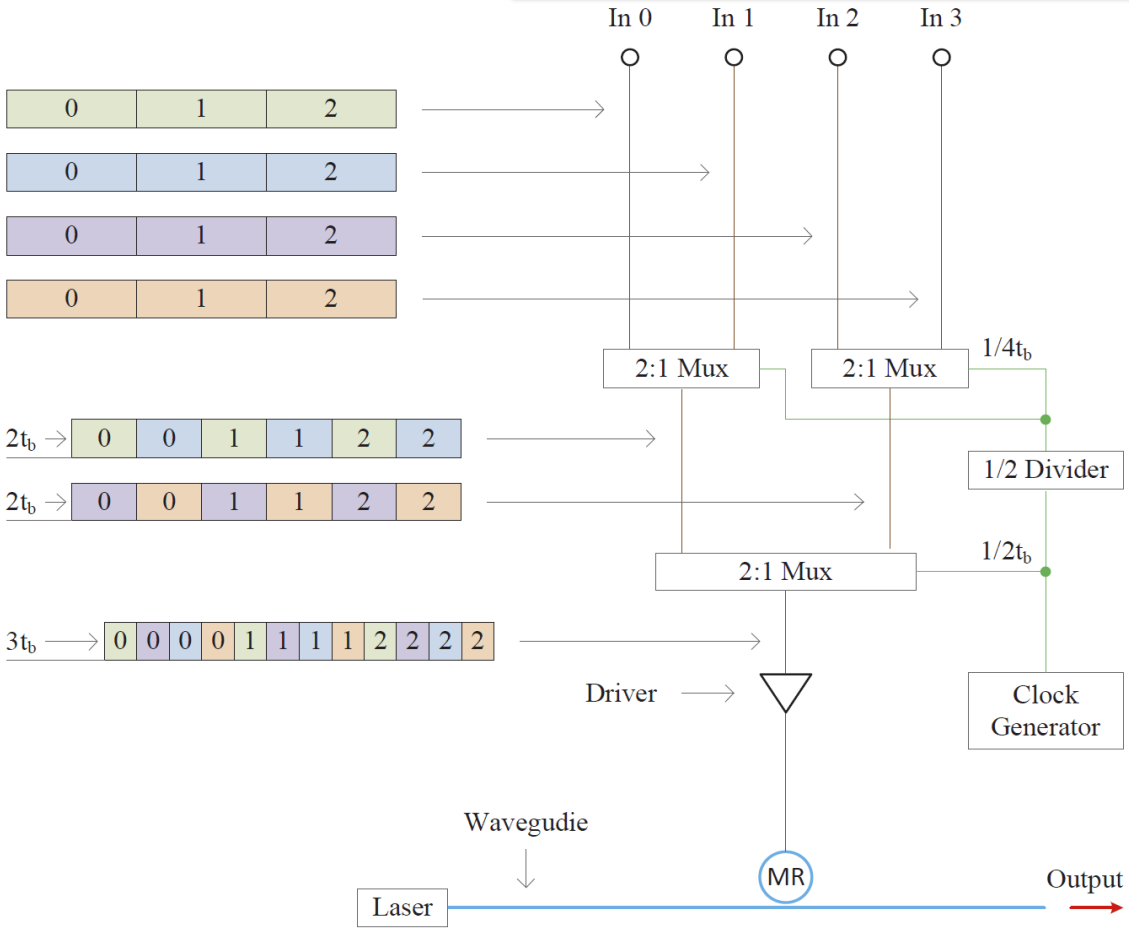
# Optical Weaving OE Interface

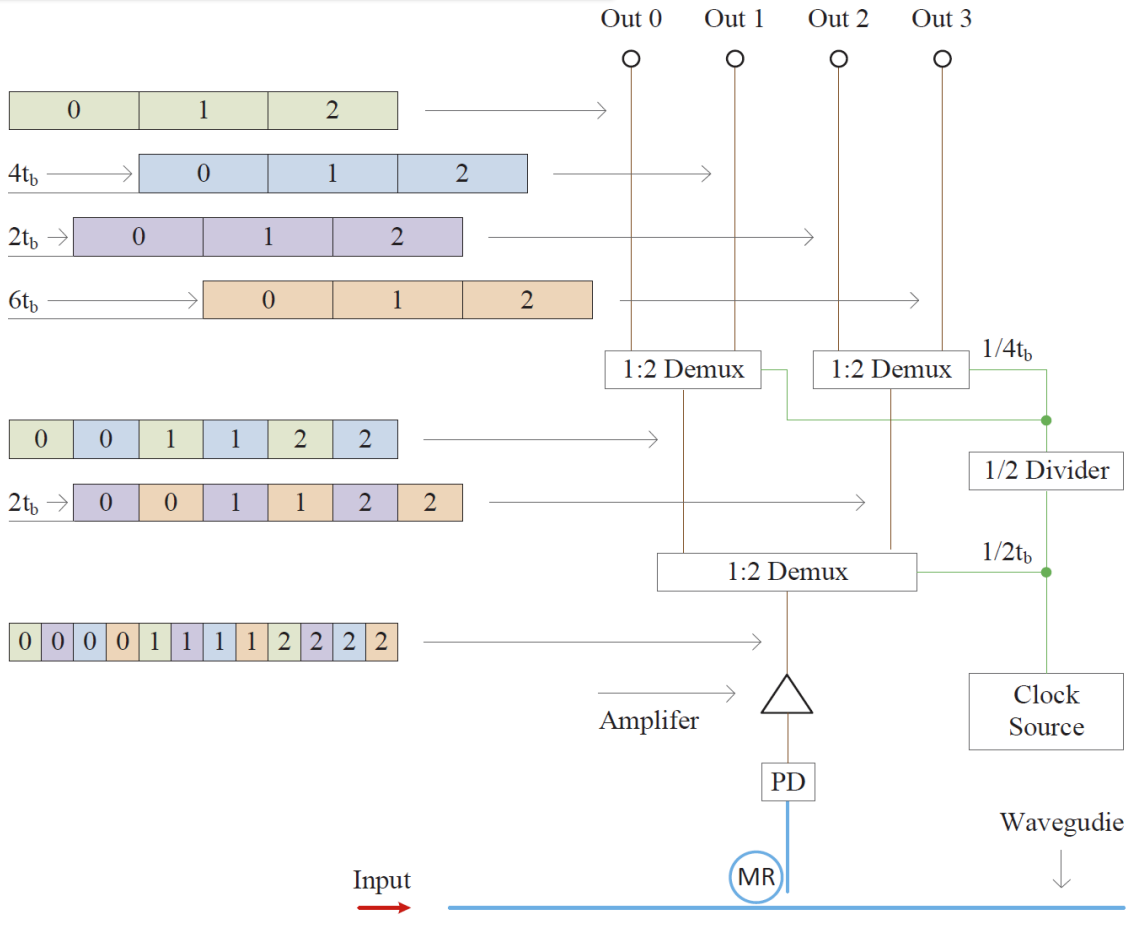



- Based on a novel optical-electrical SerDes
- High energy efficiency
- Low latency

# Traditional Electrical Funneling OE Interface

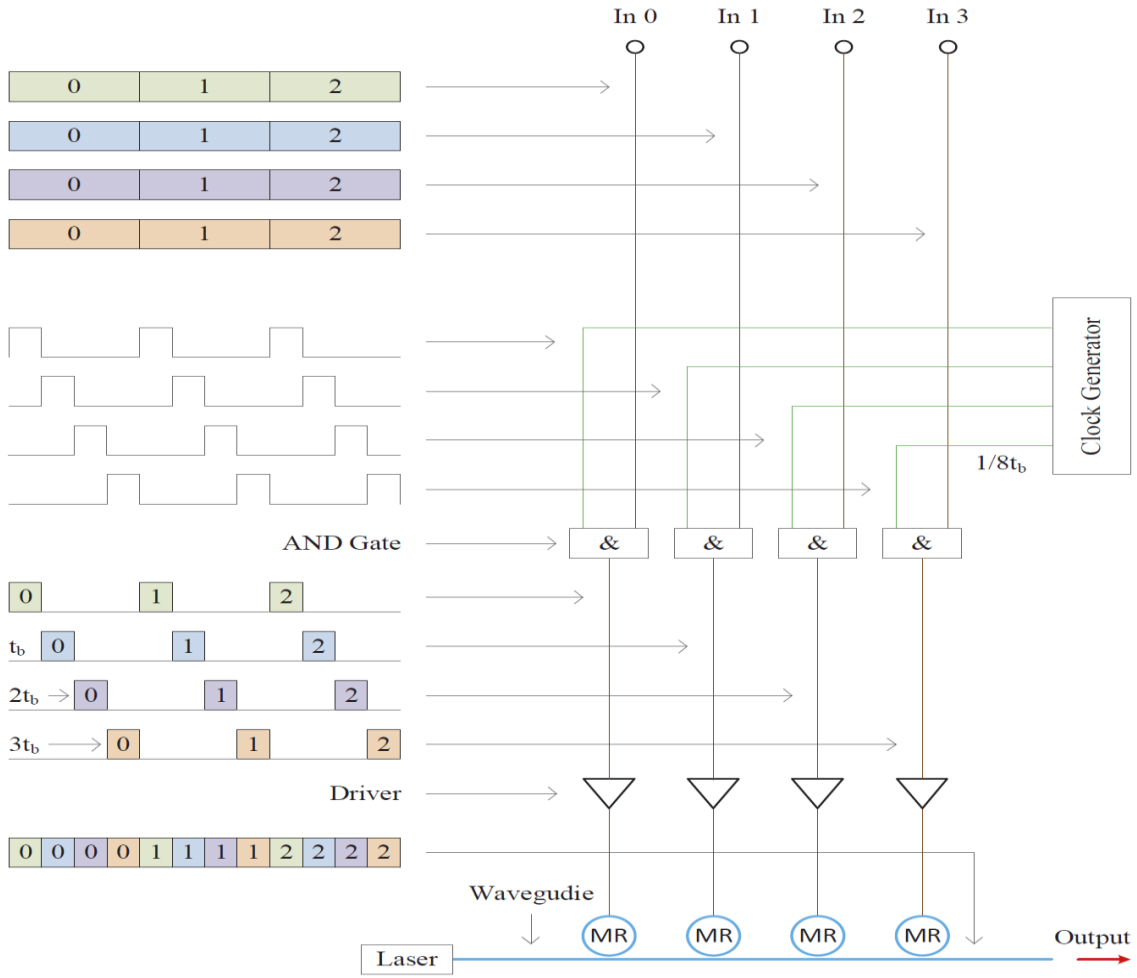
## Electrical SerDes + OE conversion

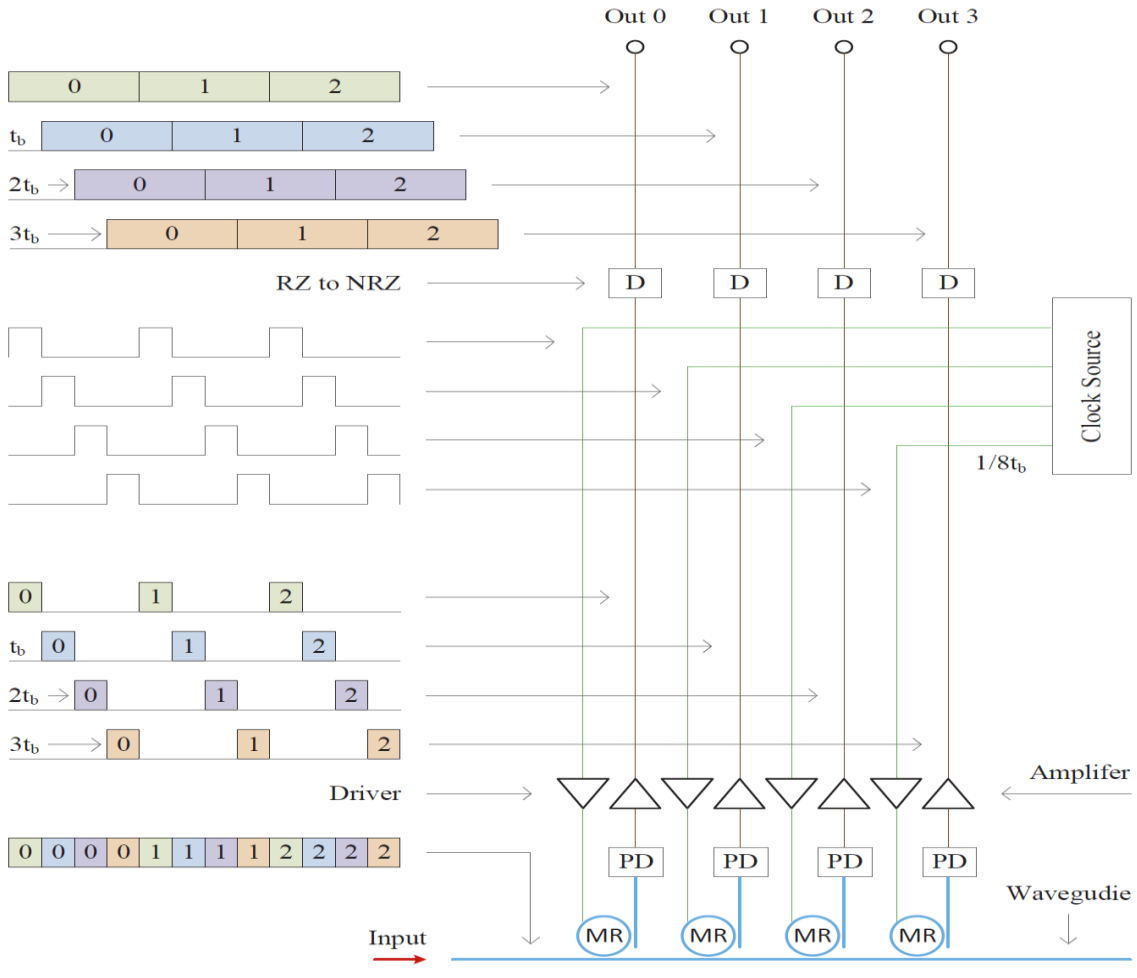


# Optical-Electrical SerDes

- High-speed photonic circuit with low-speed electronic circuit



# Simulation Environment and Setup

- Simulation environment: JADE [15]
- Benchmarks
  - COSMIC [25]
  - STREAM [16]
- Component parameters
  - JADE [15]
  - OEIL [20]
  - Micron memory power model [19]

Processor

| Parameter | Value |
|---|---|
| Core | 32~256 ARM-v7 cores @3GHz |
| I/D $ | 32KB/core, private |
| L2 $ | 128~512KB/core, shared |
| Cache line size | 64B |
| Cache coherence | directory-based MOSI |
| NoC topology | ring |
| Technology | 7nm electronic die<br>65nm photonic die |

# Simulation Environment and Setup

## Memory

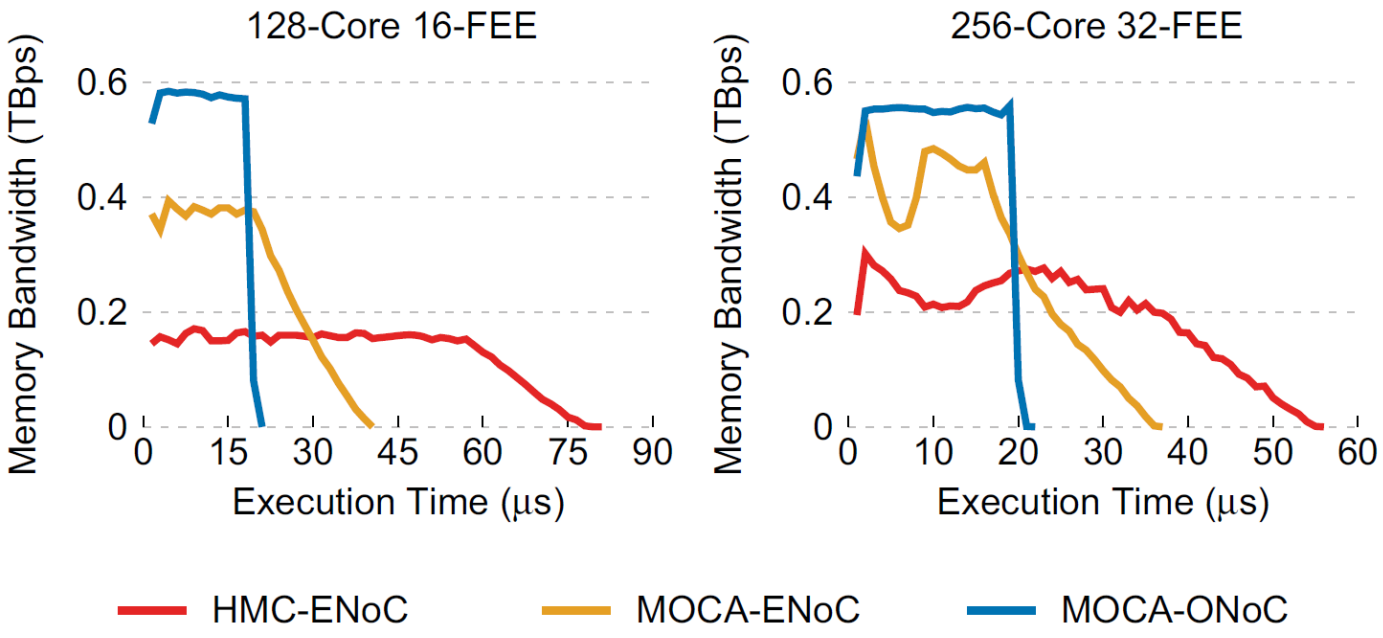
| Parameter | Value |
|---|---|
| Organization | 2 bank/rank<br>x8 rank/transaction engine<br>x32 transaction engine/chip<br>x32 chip/front-end engine<br>x32 front-end engine |
| Frequency | 800MHz |
| Memory Size | 8GB/chip |
| Schedule Policy | FR-FCFS |
| Page Policy | close-page policy |
| Technology | 14nm logic die<br>22nm memory die<br>65nm photonic die |

## Optical devices

| Parameter | Value |
|---|---|
| Waveguide propagation loss | 1 dB/cm |
| Waveguide crossing loss | 0.1 dB |
| Fiber propagation loss | $5x10^{-6}$dB/cm |
| 32-way splitter excess loss | 4 dB |
| Optical pin coupling loss | 2 dB |
| Receiver sensitivity | -20 dBm |
| Laser power conversion efficiency | 10 dB |
| Laser power extinction ratio | 10 |
| Microresonator passing loss | 0.2 dB |
| Microresonator insertion loss | 1 dB |
| Microresonator heat tuning power | 0.05 mW |

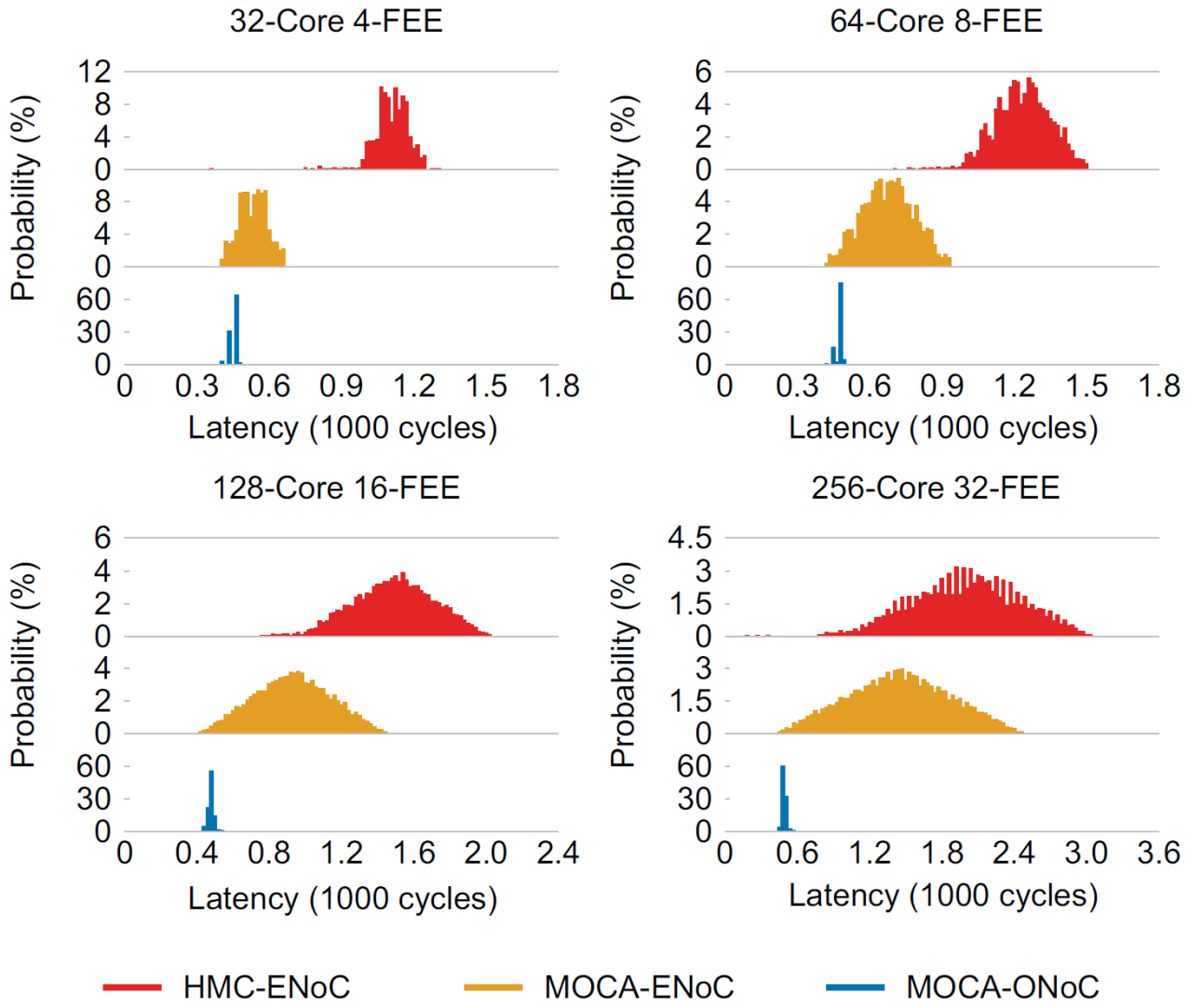
# Delivered Memory Bandwidth and Performance

- MOCA delivers 162% higher memory bandwidth compared to HMC for 256-core processors
- MOCA speedups 2.6X compared to HMC for 256-core processors
- Under STREAM benchmark [16]



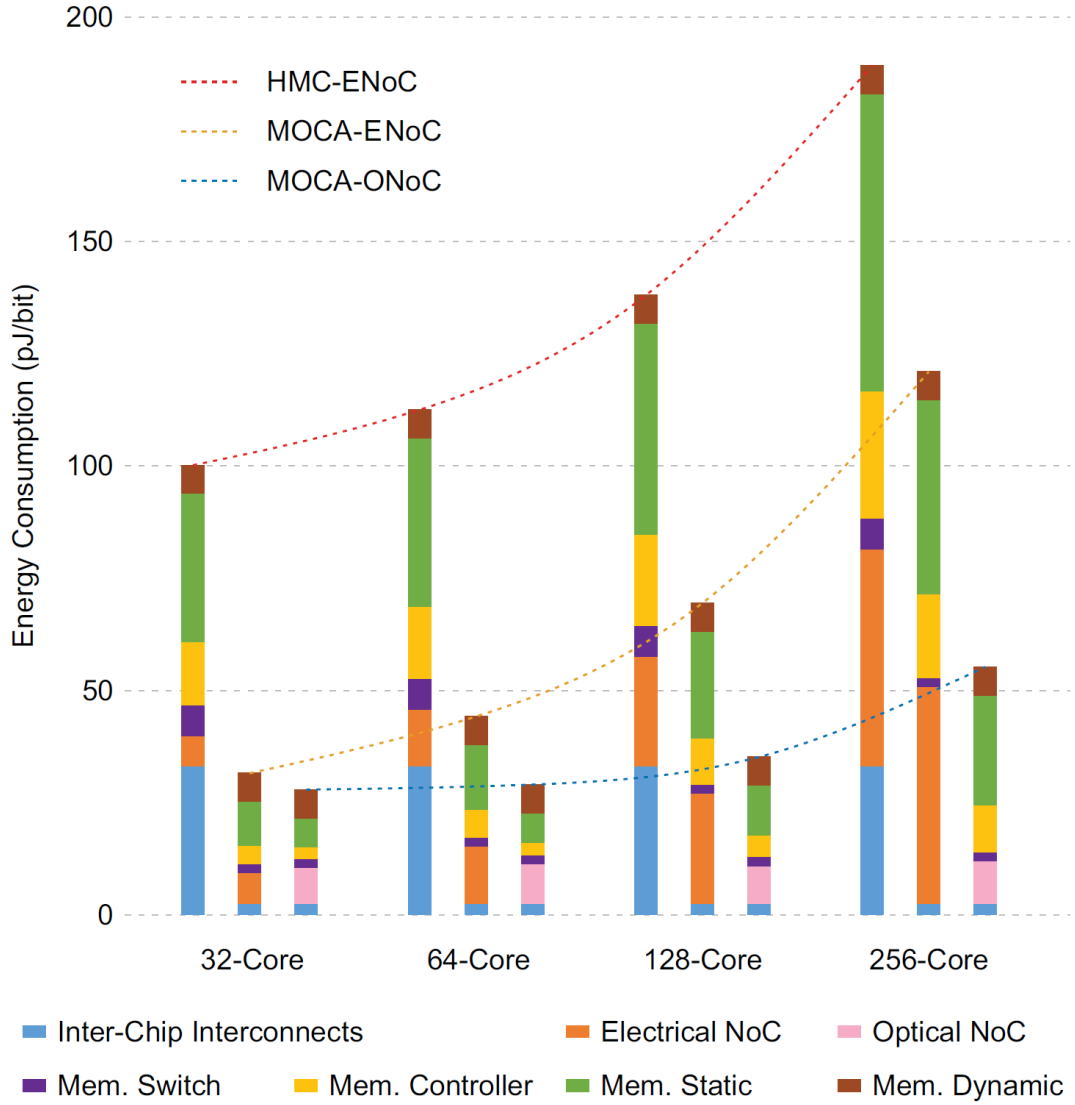
# Latency

- MOCA reduces latency by 75% compared to HMC for 256-core processors
- Narrowly distributed
- Under STREAM benchmark [16]

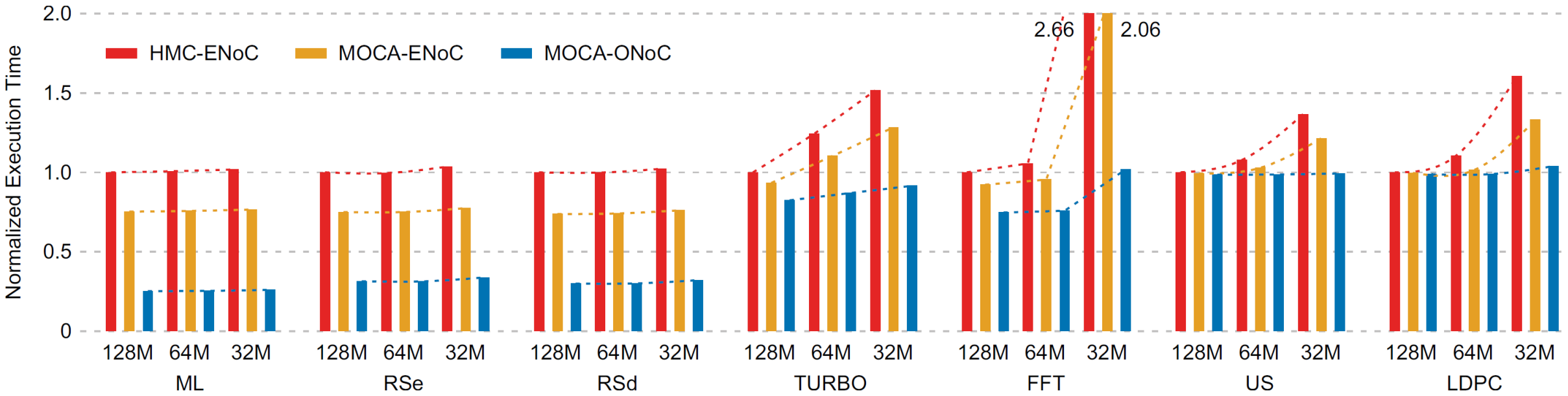


# Energy Efficiency

- MOCA helps to save 71% energy compared to HMC in 256-core processors
- ENoC-based MOCA can save 37% energy
- Better scalability in term of energy efficiency



# Execution Time under Different LLC Sizes




- 256-core, 128~32MB shared LLC
- MOCA offers 59% higher performance than HMC
- MOCA can support memory-intensive applications with smaller LLC

# Conclusions

- MOCA is an example of OMIN
  - Unified inter/intra-chip optical network
  - Physically-centralized logically-distributed control
  - Off-chip central laser source
  - Optically distributed clock
  - Optical weaving OE interface
- Help to reduce on-chip cache sizes
- Significantly improve performance and energy efficiency

# Publically Released Tools

- Bibliography for inter/intra-chip optical networks
- JADE heterogeneous multiprocessor simulation environment
- COSMIC heterogeneous multiprocessor benchmark suite
- CLAP optical crosstalk and loss analysis platform
- OTemp optical thermal effect modeling platform
- OEIL optical and electrical interface and link analysis environment
- MCSL realistic network-on-chip traffic patterns
- PowerSoC power delivery system analysis platform

www.ece.ust.hk/~eexu

# Forums and Journal Special Issue

- OPTICS Workshop
  - **O**ptical/**P**ho**T**onic **I**nterconnects for **C**omputing **S**ystems
  - Annually since 2015, base in Europe

    www.ece.ust.hk/~eexu/OPTICS
- PHOTONICS Workshop
  - **PH**otonics-**O**ptics **T**echnology **O**riented **N**etworking, **I**nformation and **C**omputing **S**ystems
  - Annually since 2017, base in US

    www.ece.ust.hk/~eexu/PHOTONICS
- ACM Journal of Emerging Technologies in Computing Systems
  - Special issue on silicon photonics

    http://jetc.acm.org/announcements.cfm

# Reference


[1] A. Hadke et al., “Design and Evaluation of an Optical CPU-DRAM Interconnect," in ICCD, 2008.
[2] A. N. Udipi et al., “Combining Memory and a Controller with Photonics Through 3D-stacking to Enable Scalable and Energy-efficient Systems," in ISCA, 2011.
[3] S. Beamer et al., “Re-architecting DRAM Memory Systems with Monolithically Integrated Silicon Photonics," in ISCA, 2010.
[4] C. Batten et al., “Building Manycore Processor-to-DRAM Networks with Monolithic Silicon Photonics," in HOTI, 2008.
[5] D. Brunina et al., “An Energy-Efficient Optically Connected Memory Module for Hybrid Packet- and Circuit-Switched Optical Networks," JSTQE, vol. 19, March 2013.
[6] S. L. Beux et al., “Potential and Pitfalls of Silicon Photonics Computing and Interconnect," in ISCAS, 2013.
[7] W. Y. Tsai et al., “A Novel Low Gate-Count Pipeline Topology with Multiplexer-Flip-Flops for Serial Link," TCSI, vol. 59, pp. 2600-2610, Nov 2012.
[8] G. Kim et al., “Memory-centric System Interconnect Design with Hybrid Memory Cubes," in PACT, 2013.
[9] T. Krishna et al., ”Smart: Single-Cycle Multihop Traversals over a Shared Network on Chip," in MICRO, 2014.
[10] “Hybrid Memory Cube Specification 2.0," Technical Publication, 2014.
[11] J. Zhan et al., “A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers," in MICRO, 2016.
[12] S. Li et al., “McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures," in MICRO, 2009.
[13] X. Wu et al., “An Inter/Intra-Chip Optical Network for Manycore Processors," TVLSI, vol. 23, pp. 678-691, April 2015.
[14] C. Schow et al., “A 24-Channel, 300 Gb/s, 8.2 pJ/bit, Full-Duplex Fiber-Coupled Optical Transceiver Module Based on a Single "Holey" CMOS IC," JLT, vol. 29, pp. 542-553, Feb 2011.
[15] R. K. V. Maeda et al., ”JADE: A Heterogeneous Multiprocessor System Simulation Platform Using Recorded and Statistical Application Models," in AISTECS, 2016.
[16] J. D. McCalpin, “Memory Bandwidth and Machine Balance in Current High Performance Computers," TCCA, pp. 19-25, Dec. 1995.
[17] P. Rosenfeld et al., “DRAMSim2: A Cycle Accurate Memory System Simulator," CAL, vol. 10, pp. 16-19, Jan 2011.
[18] O. Naji et al., “A High-level DRAM Timing, Power and Area Exploration Tool," in SAMOS, 2015.
[19] “Calculating Memory System Power for DDR3," Technical Publication, 2014.
[20] Z. Wang et al., “Improve Chip Pin Performance Using Optical Interconnects," TVLSI, vol. 24, pp. 1574-1587, April 2016.
[21] “Corning R Single-Mode Optical Fiber," Technical Publication.
[22] Q. Xu et al., “Micrometre-scale Silicon Electro-optic Modulator," Nature, vol. 435, no. 7040, pp. 325-327, 2005.
[23] Y. Zhang et al., “Towards Adaptively Tuned Silicon Microring Resonators for Optical Networks-on-Chip Applications,” JSTQE, vol. 20, pp. 136-149, July 2014.
[24] A. Supalov et al., Optimizing HPC Applications with Intel Cluster Tools. Apress, 2014.
[25] Zhe Wang et al., “A Case Study on the Communication and Computation Behaviors of Real Applications in NoC-based MPSoCs,” IEEE Computer Society Annual Symposium on VLSI, Florida, July 2014.