香港中文大學
The Chinese University of Hong Kong

Chiplet Architecture Heterogeneous Interface

Shixin CHEN
Department of Computer Science & Engineering
Chinese University of Hong Kong
sxchen22@cse.cuhk.edu.hk

December 14, 2023

Outline

Chiplet Architecture

- most current multi-chiplet systems are based on one uniform die-to-die interface;
- The uniform interface does not cope well with flexible workloads, especially for large-scale systems.

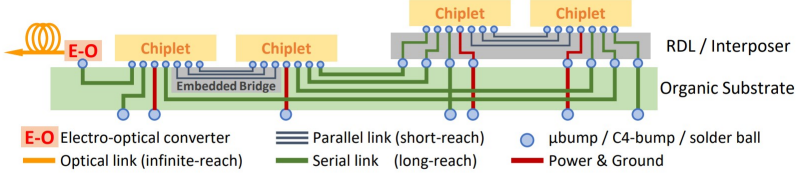


Figure 1: Multi-chiplet system: Advanced packaging technologies provide abundant interconnection possibilities.

Caption

Die-to-Die Interface

Heterogeneous Interface (Hetero-IF)[1]

- Pros: The interconnection allows chiplets to use two different interfaces (parallel IF and serial IF) at the same time.
- Challenges: The microarchitecture, scheduling, interconnection, and routing issues have not been discussed so far.

[1] Yinxiao Feng, Dong Xiang, and Kaisheng Ma (2023). "Heterogeneous Die-to-Die Interfaces: Enabling More Flexible Chiplet Interconnection Systems". In: *Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 930–943.

Motivation

- Parallel interface: Advanced Interface Bus (AIB),low latency and low-power, short reach, small scales
- Serial interface: large latency and power consumption, long-reach
- "Universal": Trade-off, on one-size-fits-all
- The Uniform interface is not flexible enough to handle complex and mixed network traffic.
- handshake, synchronization, coherence protocols
- heavy network traffic: all-reduce operation

Challenges of Heterogeneous Interfaces

- Lack of discussion
- problem: out-of-order delivery, heterogeneous router
- complex scheduling
- requires efficient deadlock-free interconnection solutions, avoid congestion

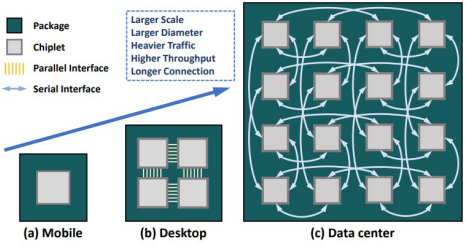
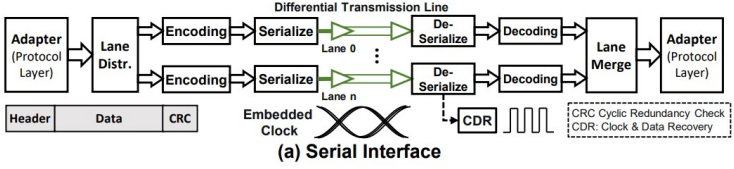


Figure 2: Chiplet reuse in systems of different scales. For different scenarios, though using the same chiplet, systems have very different integration and interconnection architectures.

Steps

Internal communication protocol: AXI; External protocols: Compute Express Link (CXL);

The interface can be divided into two layers: PHY and Adapter:



Interfaces

- **Serial Interface**: high-data-rate, long-reach, high-latency, high-power. *e.g.*,Serializer/Deserializer (SerDes): Anti-interference: double-terminated differential lines, CDR (clock & data recovery), and FEC (forward error correction)
- **Serial Interface**: low-power, low-latency, short-reach, low-data-rate, and high-port-count (costly). *e.g.*, Advanced Interface Bus (AIB);High Bandwidth Memory (HBM);OpenHBI
- **Compromised Interface**: e improvements based on advanced packaging technologies, *e.g.*, Bunch of Wires (BoW), Universal Chiplet Interconnect Express (UCIe).

Table 1: Specification of typical die-to-die interface

Specification	SerDes [2, 4, 5, 40]	AIB [3, 40, 60]	BoW [8, 10, 11, 27]	UCIe [7, 58, 59]
Data Rate (Gbps)	112 ✔	6.4 ✘	32	32
Latency (ns)	5.5+L_D+FEC ✘	3.5 ✔	3+L_D+FEC	2+L_D
Power (pj/bit)	2 ✘	0.5 ✔	0.7	0.3 / 1.25
Reach (mm)	50 ✔	10 ✘	50 ✔	2 / 25

Interface-based Application

Different workloads have different requirements for interfaces. The parallel interface is suitable for the frequent local movement of small data, and the serial interface is suitable for the long-distance movement of large data.

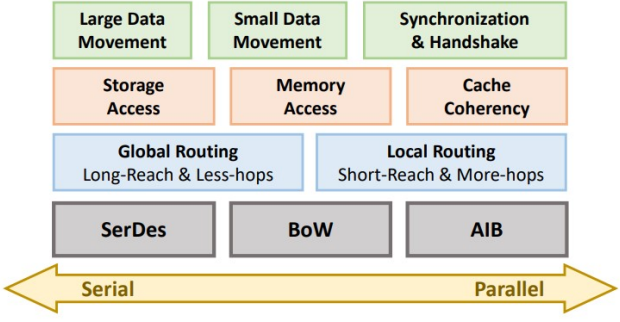


Figure 4: Different workloads have different requirements for interfaces.

Heterogeneous Interfaces Architecture

Heterogeneous interfaces make interconnection architectures and routing algorithms more complex

- Hetero-PHY
- Hetero-Channel

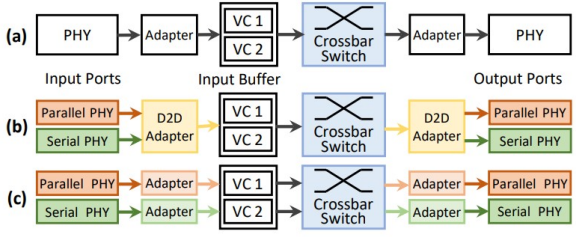


Figure 5: Heterogeneous interface architectures. (a) Uniform Interface; (b) Heterogeneous PHY; (c) Heterogeneous Channel.

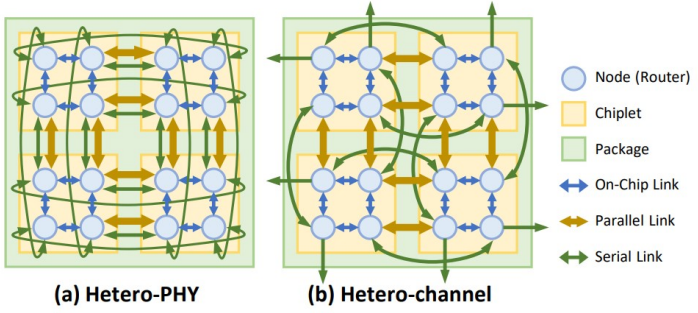


Figure 6: Interconnection networks of chiplets based on heterogeneous interfaces.

THANK YOU!