Georgia Tech

From **Domain-Specific Architectures** to **Systems** in the Chiplet Era: Packaging-Aware Co-Design Across Compute and Interconnect

Jiho Kim

Advisor: Cong (Callie) Hao

Georgia Institute of Technology

Sharc Lab

DoSSA?

Workshop on Domain-Specific System Architecture

For Specific applications

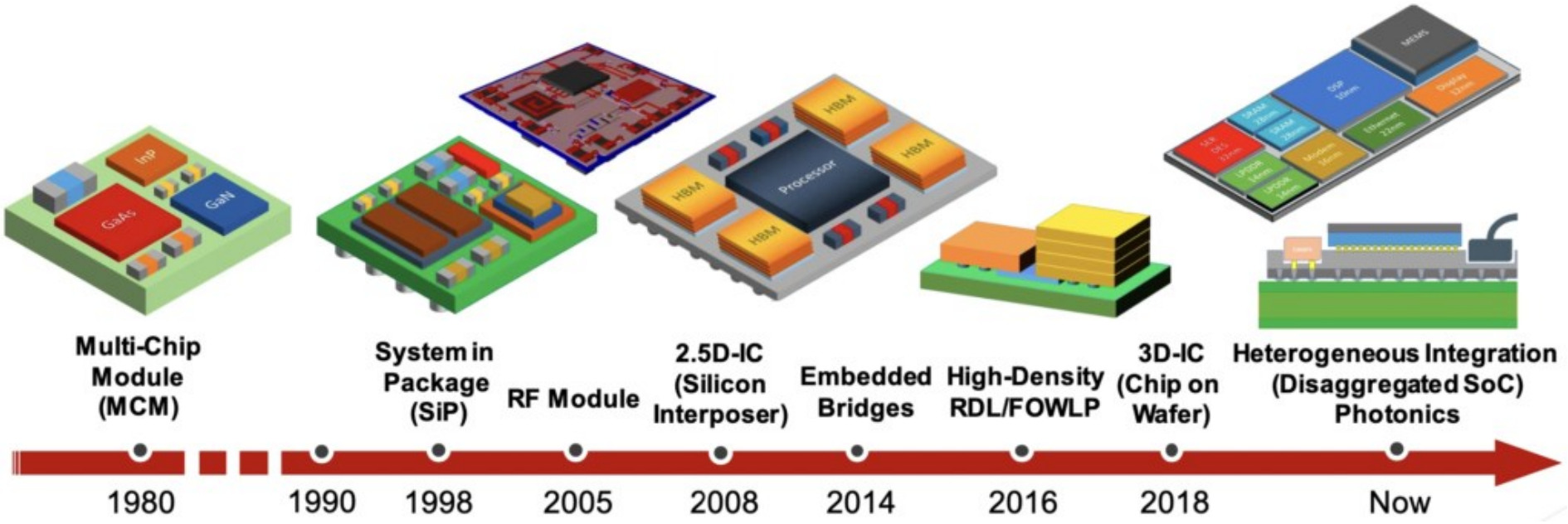
Hardware components & Software

CALL FOR PAPERS

Domain specific systems are an increasingly important computing environment for many people and businesses. As the information technologies emerge into various real world applications such as autonomous driving, IoT (Internet of Things), CPS (Cyber physical systems) and health care applications in the 4th industrial revolution era, the interest in the specialized domain specific computing systems is increasing significantly. In addition to the conventional computing platforms, domain specific computing systems have a lot of design challenges including specialized hardware components like hardware accelerator, optimized library and domain specific languages. This workshop focuses on domain specific system design in both hardware and software aspects and their interaction in order to improve the availability and efficiency in the emerging real world applications. The main theme of this workshop in this year is the HW/SW components for domain specific systems. Topics of particular interest include, but are not limited to:

"System" is Changing!

Workshop on Domain-Specific System Architecture



Moving towards 2.5D/3D Chiplet System!

"System" is Changing!

Workshop on Domain-Specific System Architecture

How does ***Domain-specific Architecture*** *alter in* ***Chiplet System****?*

Moving towards 2.5D/3D Chiplet System!

What is Chiplet?

- ***System's building block***

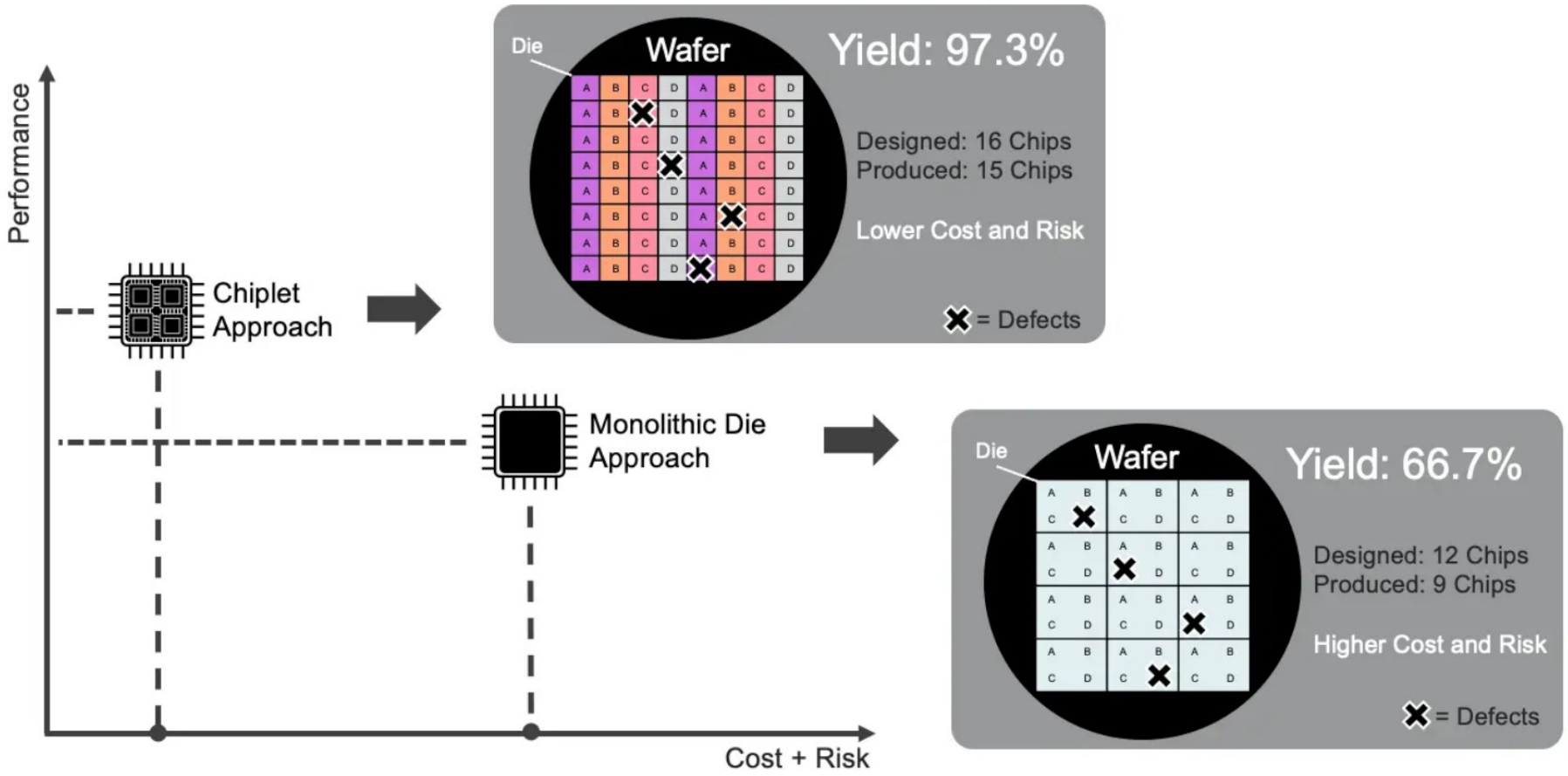
System-on-chip (SoC)

Chiplet-based System-in-package (SiP)

Source: IDTechEx (https://www.idtechex.com/en/research-article/chiplet-technology-unlocking-new-potential-in-semiconductor-industry/32352)

Why Chiplet?

- ***Squeezing 100 billion transistors into one single die?***



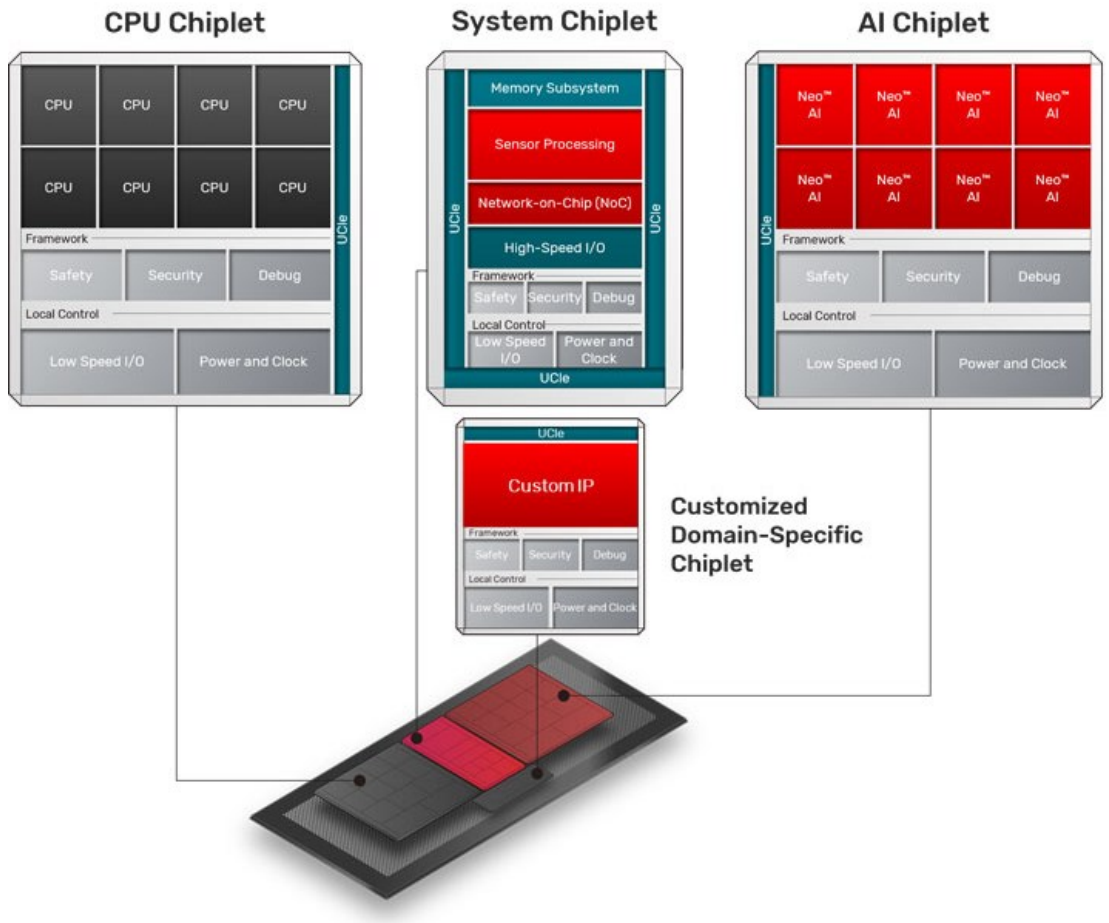
Source: Semiconductor Engineering (https://semiengineering.com/what-is-a-chiplet-and-why-should-you-care/)

Open Chiplet Ecosystem

- **Mix-and-match chiplets!**

OPEN Compute Project®

UCIe Universal Chiplet Interconnect Express



Source: Cadence (https://www.cadence.com/en_US/home/solutions/chiplets.html#chiplet-based-physical-ai-platform)

Open Chiplet Ecosystem

- **Mix-and-match chiplets!**

The real question is,

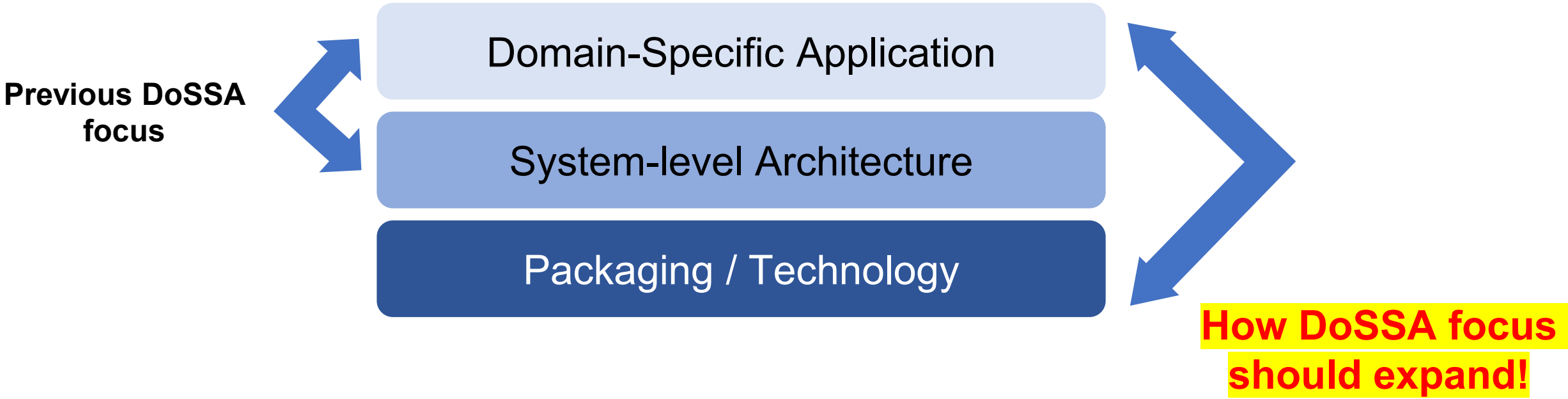
How are we ***mix-and-matching***?

Source: Cadence (https://www.cadence.com/en_US/home/solutions/chiplets.html#chiplet-based-physical-ai-platform)

DoSSA in Chiplet System?

- **Reimagining Computer Architecture's Role**

  ➔ Beyond *processor-centric* designs to *system-level* orchestration



DoSSA in Chiplet System?

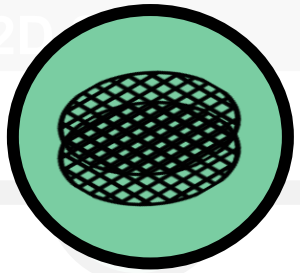
- **Reimagining Computer Architecture's Role**

➔ Beyond *processor-centric* designs to *system-level* orchestration

Domain-Specific **Chiplet** System Architecture

DoSSA → DoSCSA!

2D → 2.5D/3D



Packaging
Chip assembly

2.5D/3D?

Packaging is leading 2.5D/3D for sure!

But is Packaging the only thing evolving?

Hierarchical Interconnection Network

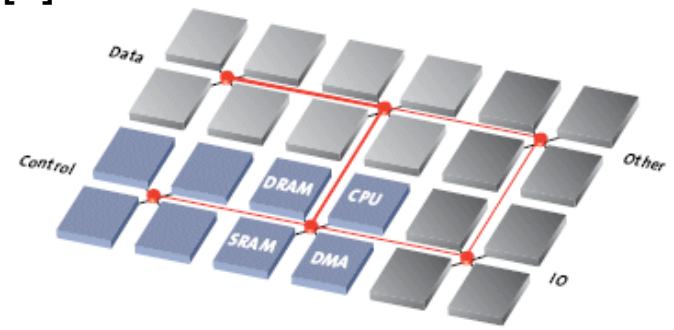
- **Monolithic system on chip (2D SoC)**

  → Network on chip(NoC)

- **Chiplet-based 2.5D/3D architecture (2.5D/3D SiP)**

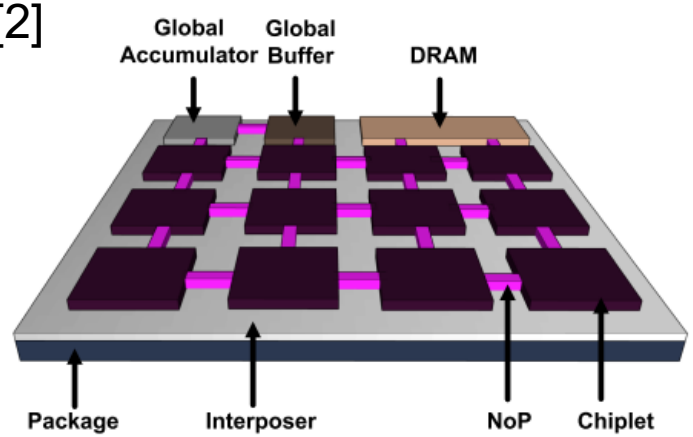
  → Network on Interposer (NoP/NoI)

[1]



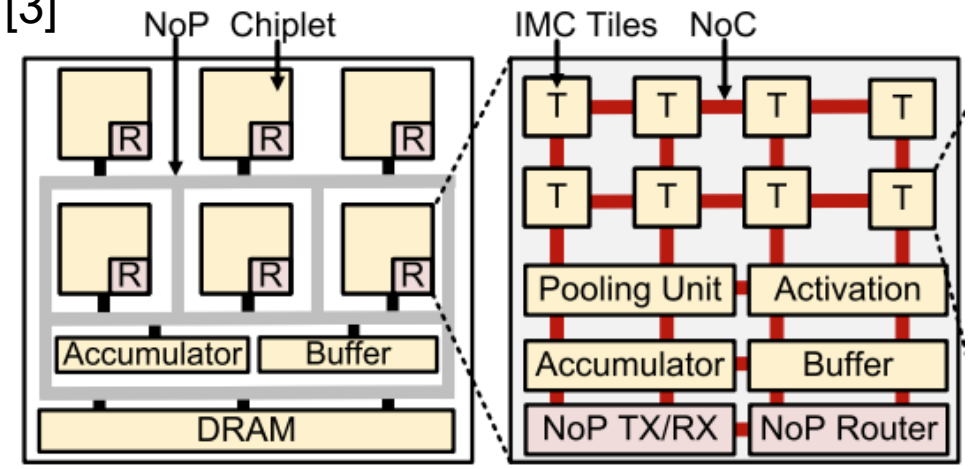
<Network on chip>

[2]



<Network on Interposer>

[3]



<Network on chip & Interposer>

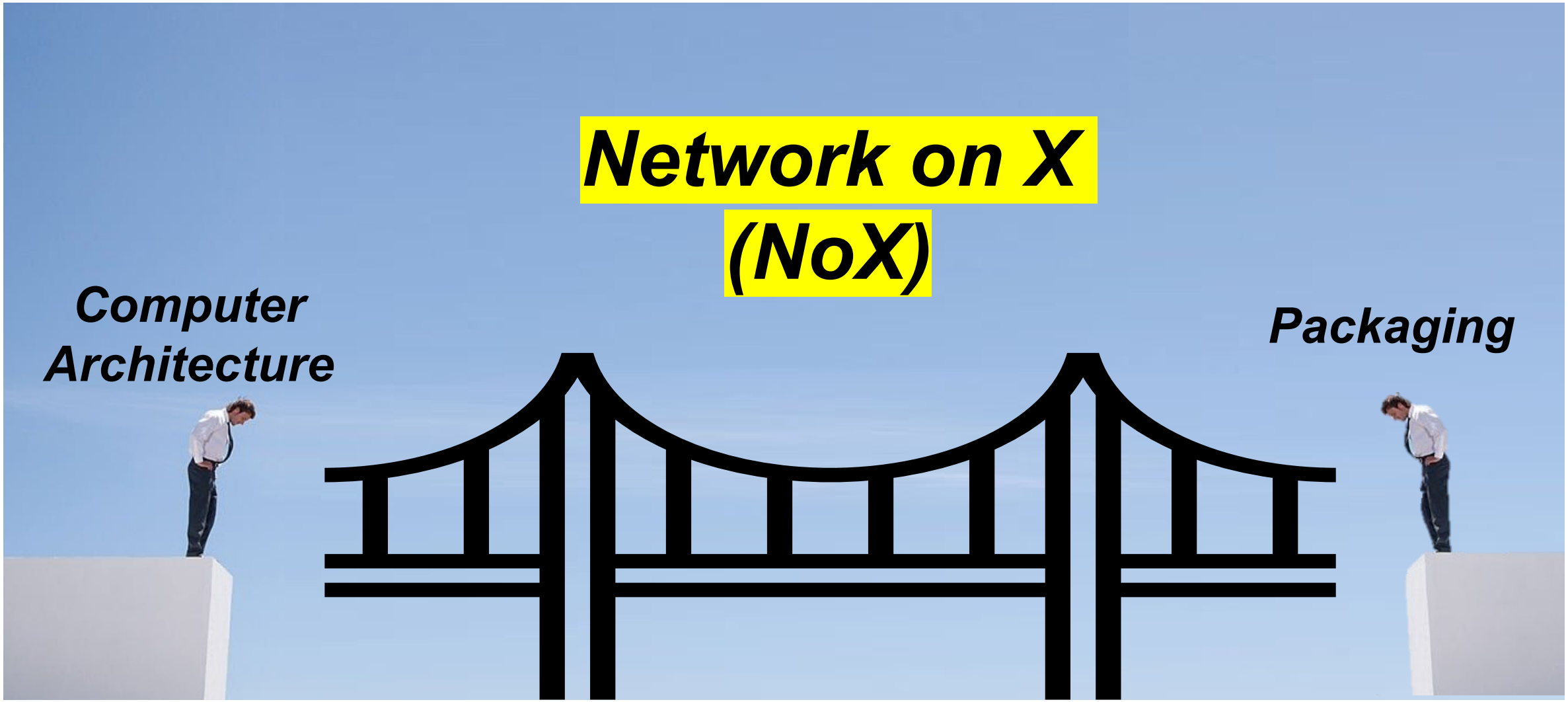
[1] Arteris, “A comparison of Network-on-Chip and Busses”, 2005, https://www.design-reuse.com/articles/10496/a-comparison-of-network-on-chip-and-busses.html
[2] Gokul Krishnan, et al. “SIAM: Chiplet-based Scalable In-Memory Acceleration with Mesh for Deep Neural Networks,”, ACM Trans. Embed. Comput. Syst. 20, 5s, Article 68, October 2021
[3] Shao, Yakun Sophia, et al. "Simba: Scaling deep-learning inference with multi-chip-module-based architecture." *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*. 2019.

Hierarchical Interconnection Stack for 2.5D/3D

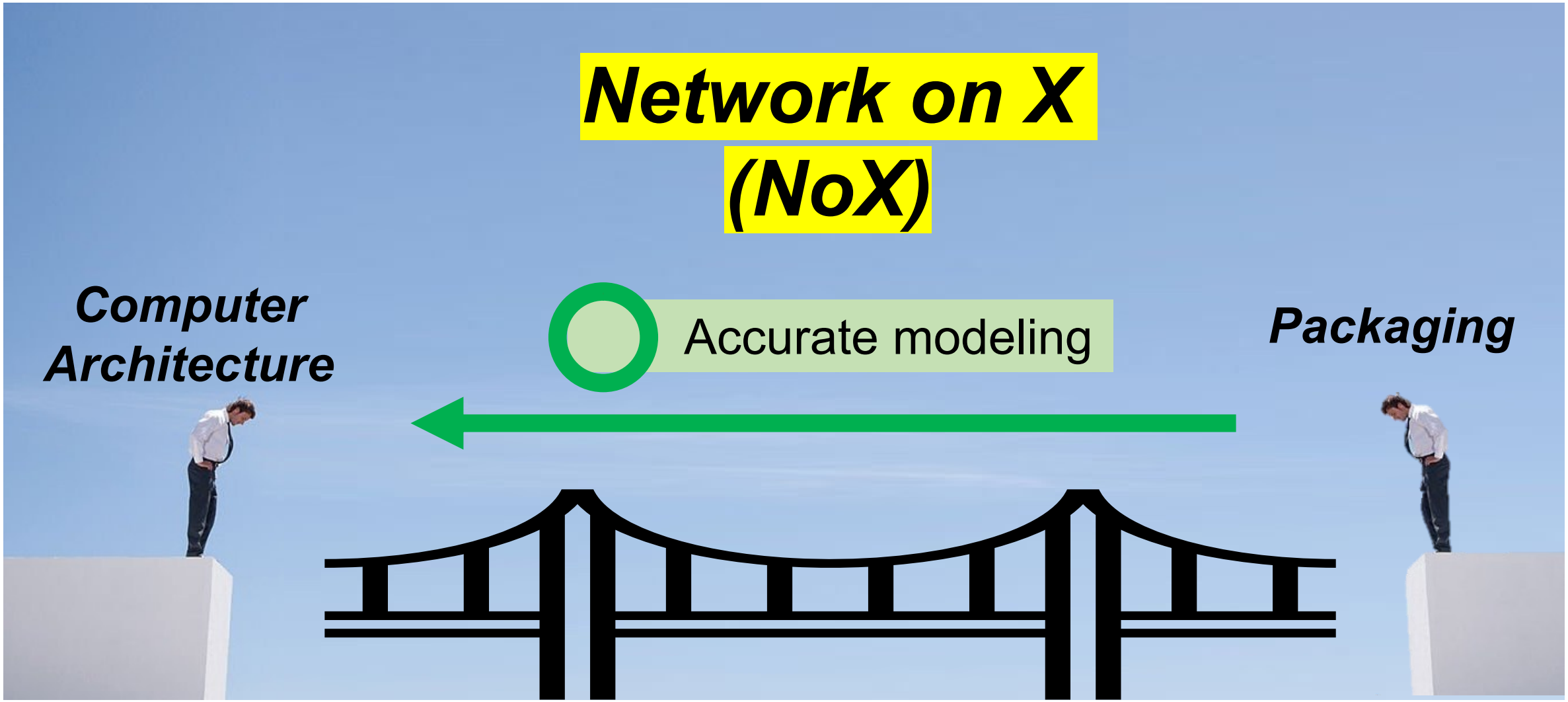
Do we have to be a packaging expert?

Computer Architect

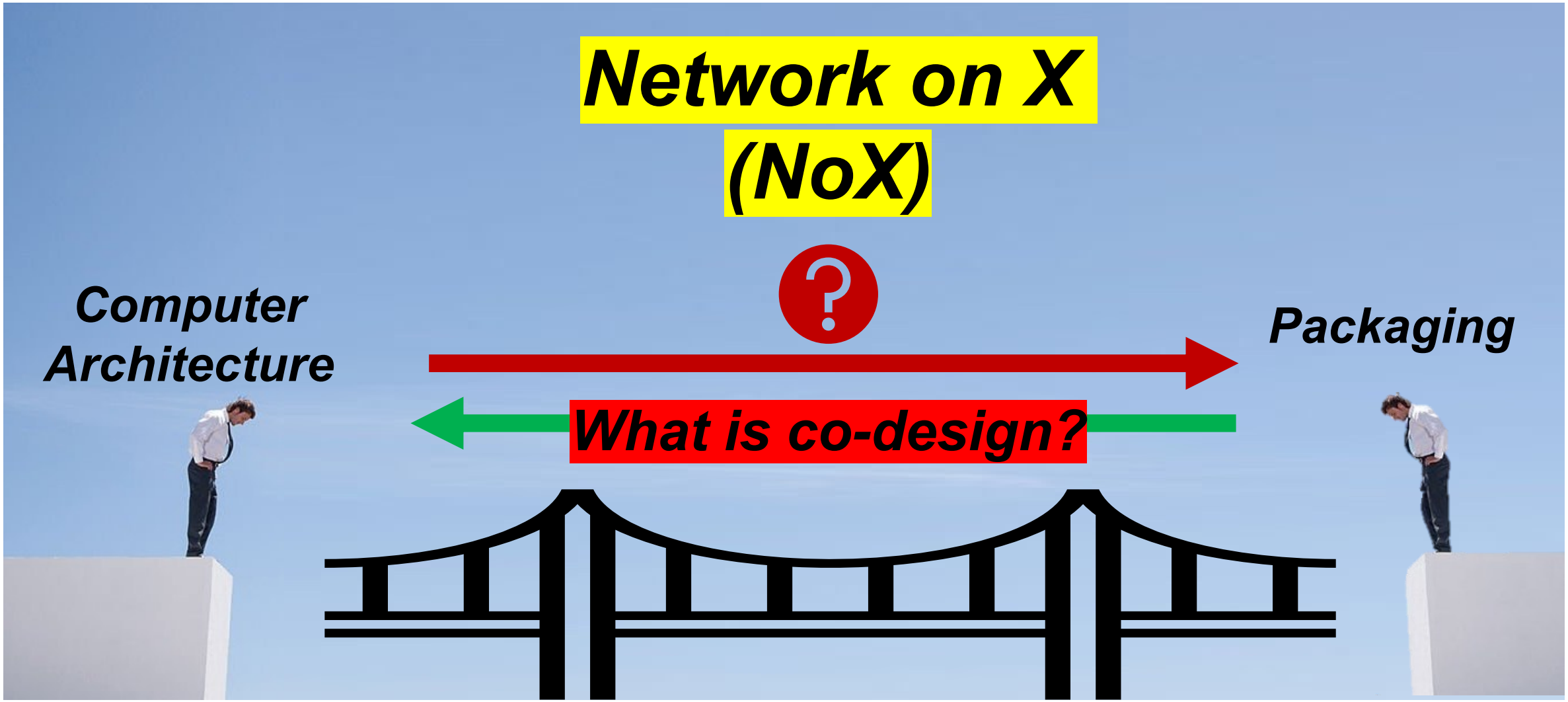
Bridging Architecture-Packaging



Bridging Architecture-Packaging



Bridging Architecture-Packaging



Are We Ready to explore 2.5D/3D Architecture? NO!

Computer architecture research depends on **simulation tools** to explore

But, existing simulators are ***oversimplified***

- Ignores circuit & packaging impact → Inaccurate
- Forces flat interconnect model → Rigid
- Lacks real-world validation → Unreliable
- Isolated architecture & packaging exploration → Suboptimal trade-offs

A New Approach is Needed for 2.5D/3D Simulation!

Proposing Network-on-X (NoX) framework!

01 Technology-aware (Packaging & Circuit) — Accurate

02 Hierarchical Communication for multi-tier — Configurable

03 End-to-end Validation — Reliable

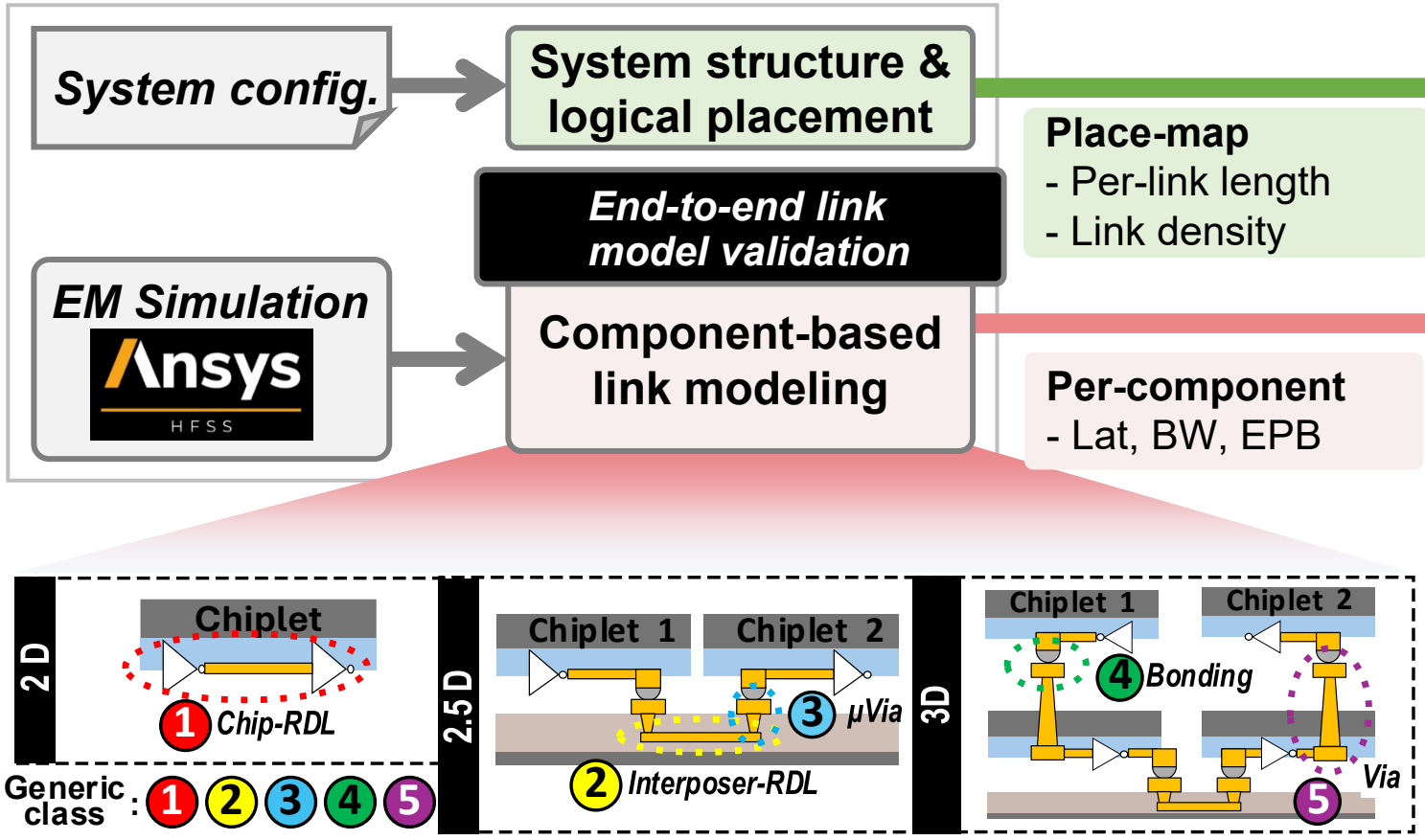
04 Architecture-Packaging Comprehensive DSE — Optimized trade-offs

NoX : Accurate hierarchical chiplet simulator

Network on X (NoX)

- Technology-aware & Hierarchical Interconnect Simulator for 2.5D/3D

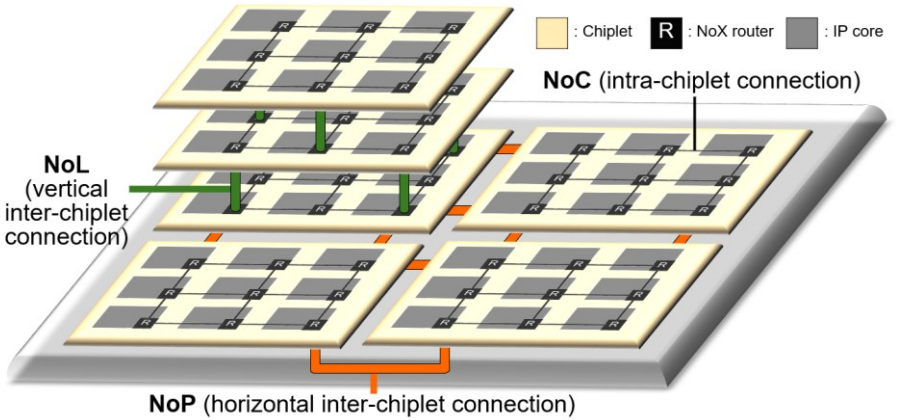
Technology-aware Modeling



Hierarchical Interconnect

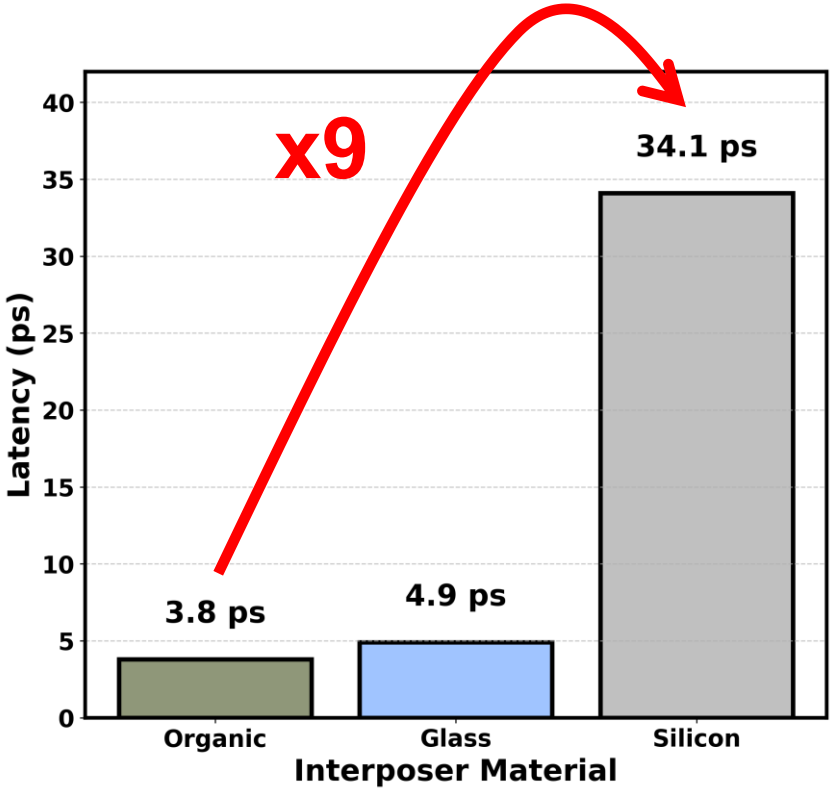
- ***Support 2D/2.5D/3D***
  - Physical structure-aware
- ***Interconnect heterogeneity***
  - Network-on-chip (NoC)
  - Network-on-interposer (NoI)
  - Network-on-layers (NoL)

NoX

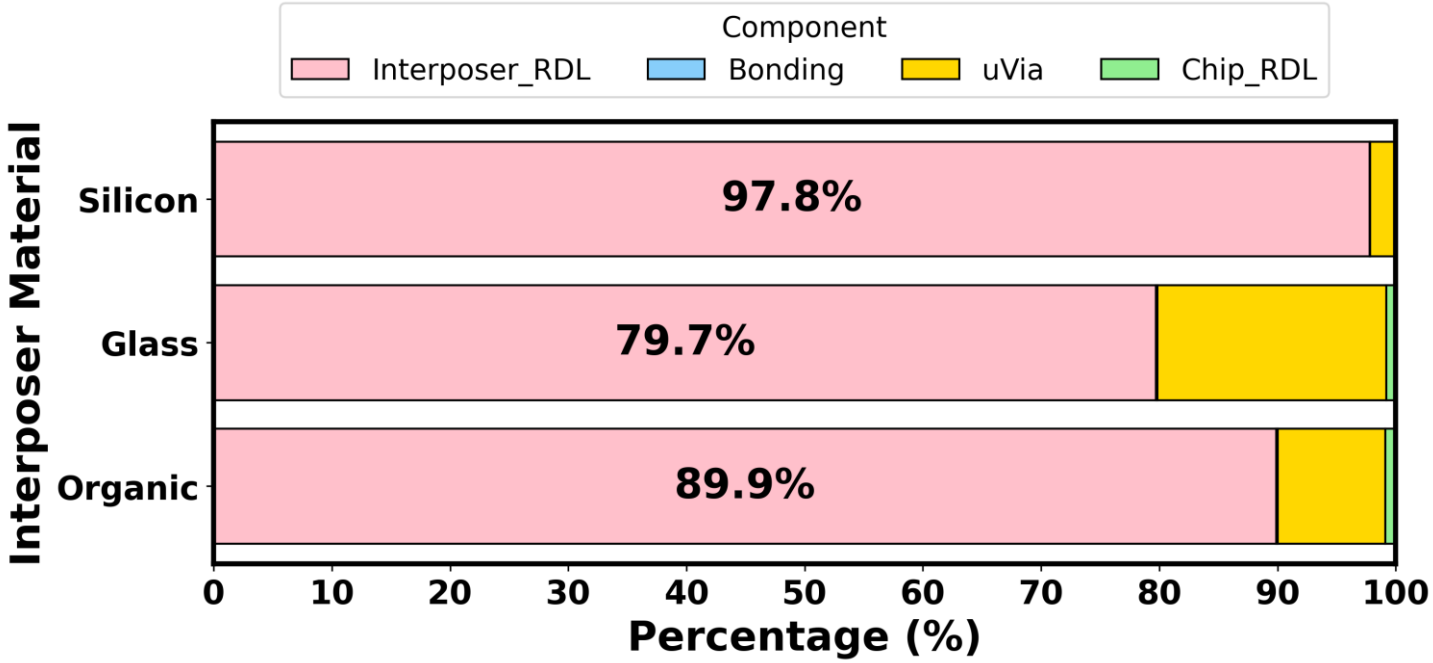


System-level impact of packaging

- *Interposer material impact (w/ diff. wire pitch)*
  - *Latency*

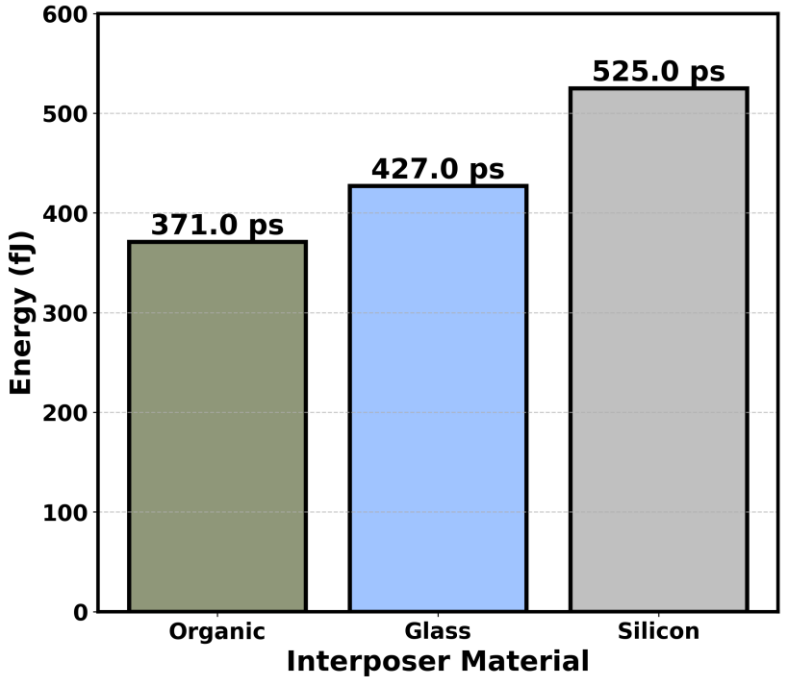


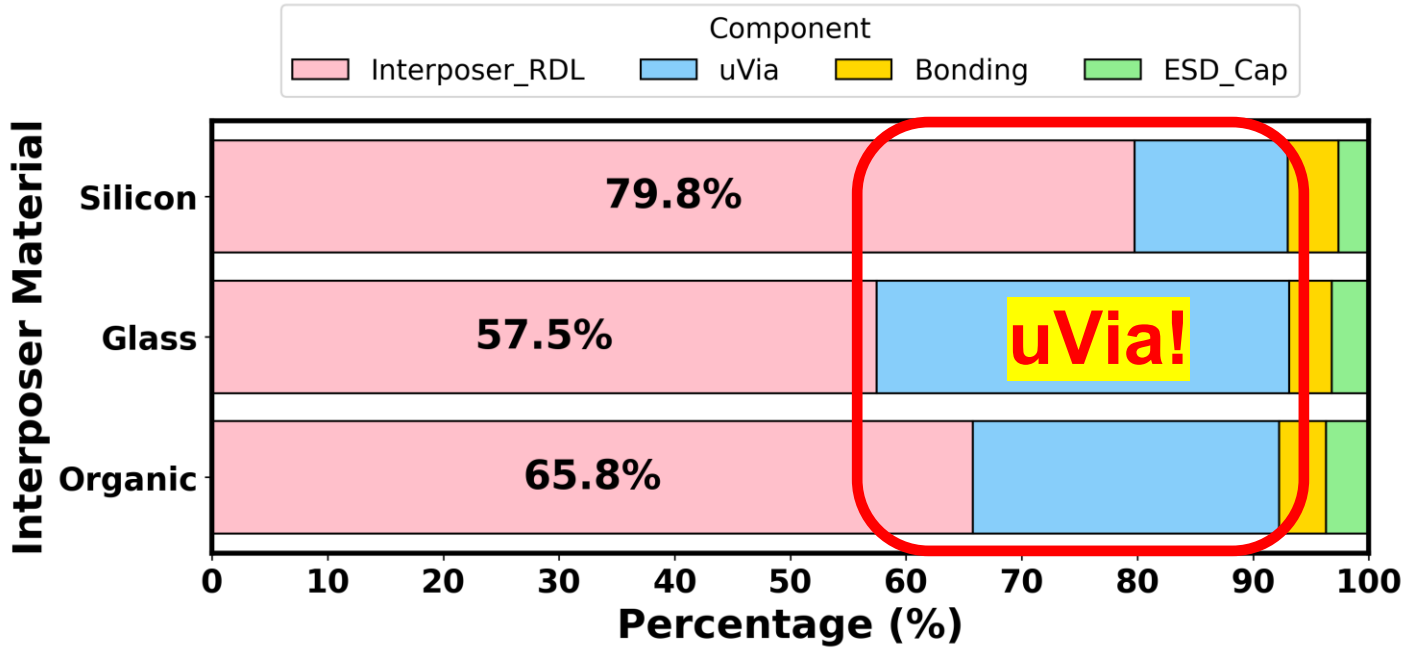
  - *Latency breakdown*



System-level impact of packaging

- *Interposer material impact (w/ diff. wire pitch)*
  - *Power*
  - *Power breakdown*





Key Takeaways

✅ Packaging *significantly* impact **system performance in 2.5D/3D**

❌ Overlooking physical constraints leads to **suboptimal system**

✅ Traditional simulators limit architecture exploration for emerging systems

❌ Technology-agnostic → Lack of realism & flexibility

☝️ NoX offers ***structure-aware***, ***technology-informed*** for 2.5D/3D+ systems

⚙️ Delivers accurate, physically grounded simulations

🔬 Push-button exploration across interconnects, networks, and layout

🔓 Unlock hidden Architecture-Packaging co-optimization opportunity

NoX bridges the gap for 2.5D/3D co-design!