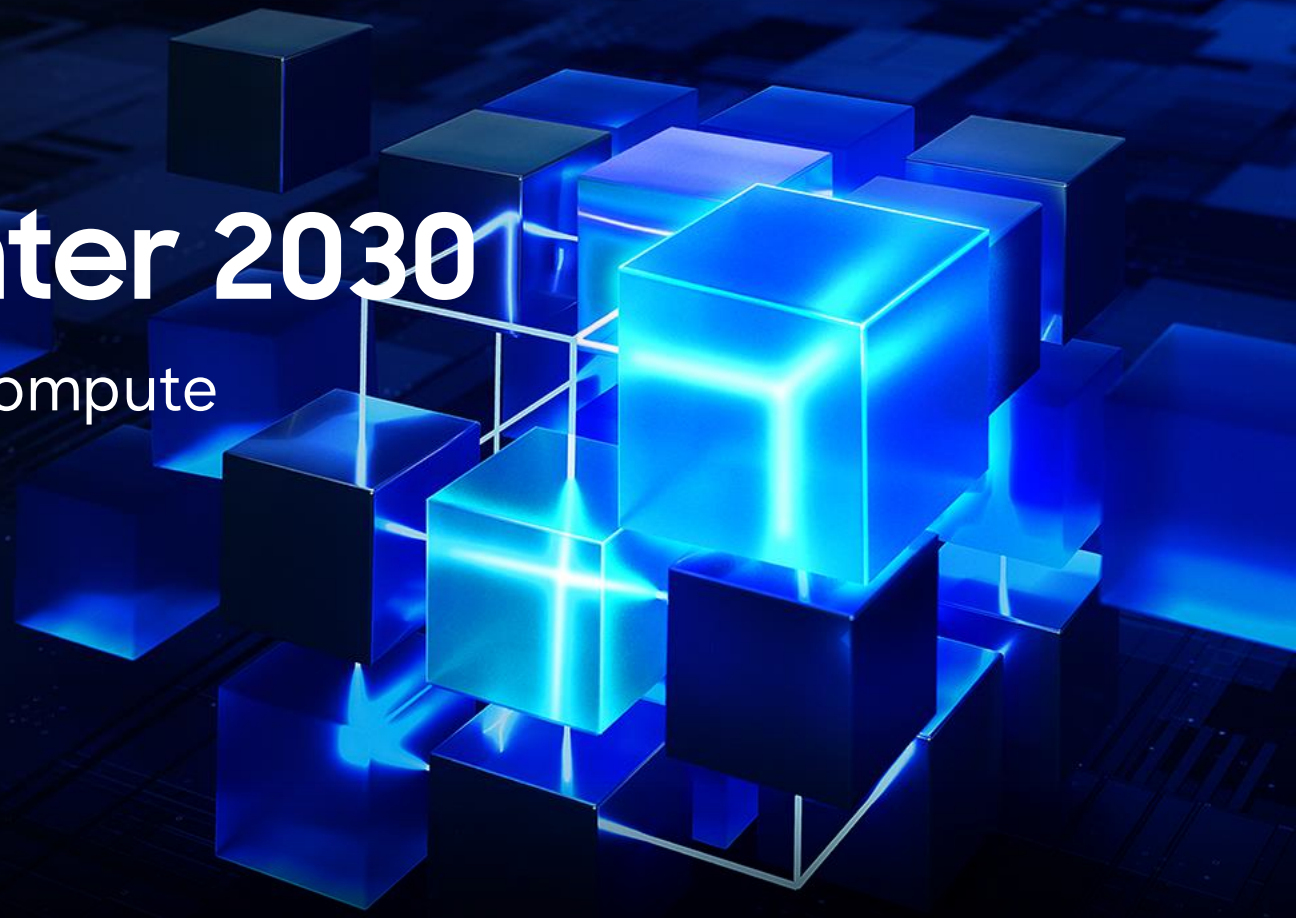
SAMSUNG

Inference Data Center 2030

Memory Pods Fabric with Rightsized Compute



Mayank Saxena
Chief Technologist
Memory Solutions Lab
Samsung Semiconductor

Data Center Divergence

Today, AI Is Still in Its “Brute Force Era”

Training

Token Manufacturing

A manufacturing cost to develop the product (models) that yields top quality tokens.



GPU Cluster

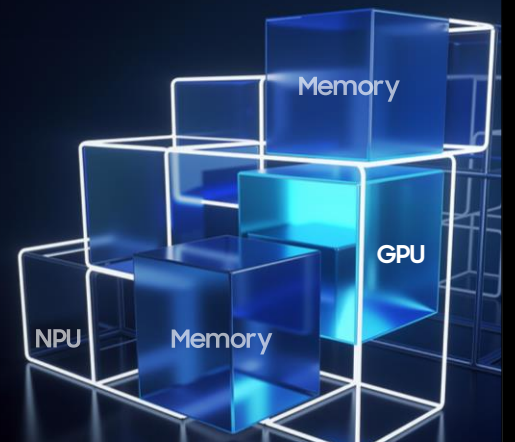
Compute Centric

Models

Inference

Token Factory

A revenue-generating factory serves the right token to the right customer at speed.



Memory Centric

Today: Compute-Centric AI

Pick Compute → Right Size Everything Else

Invert

Future: Memory-Centric Supercomputing

Pick the Memory Fabric → Right Size Compute

Inference AI Factories

Interconnected Data Ocean Linked by Photonics

Model Data Becoming Fluid

Memory Pods

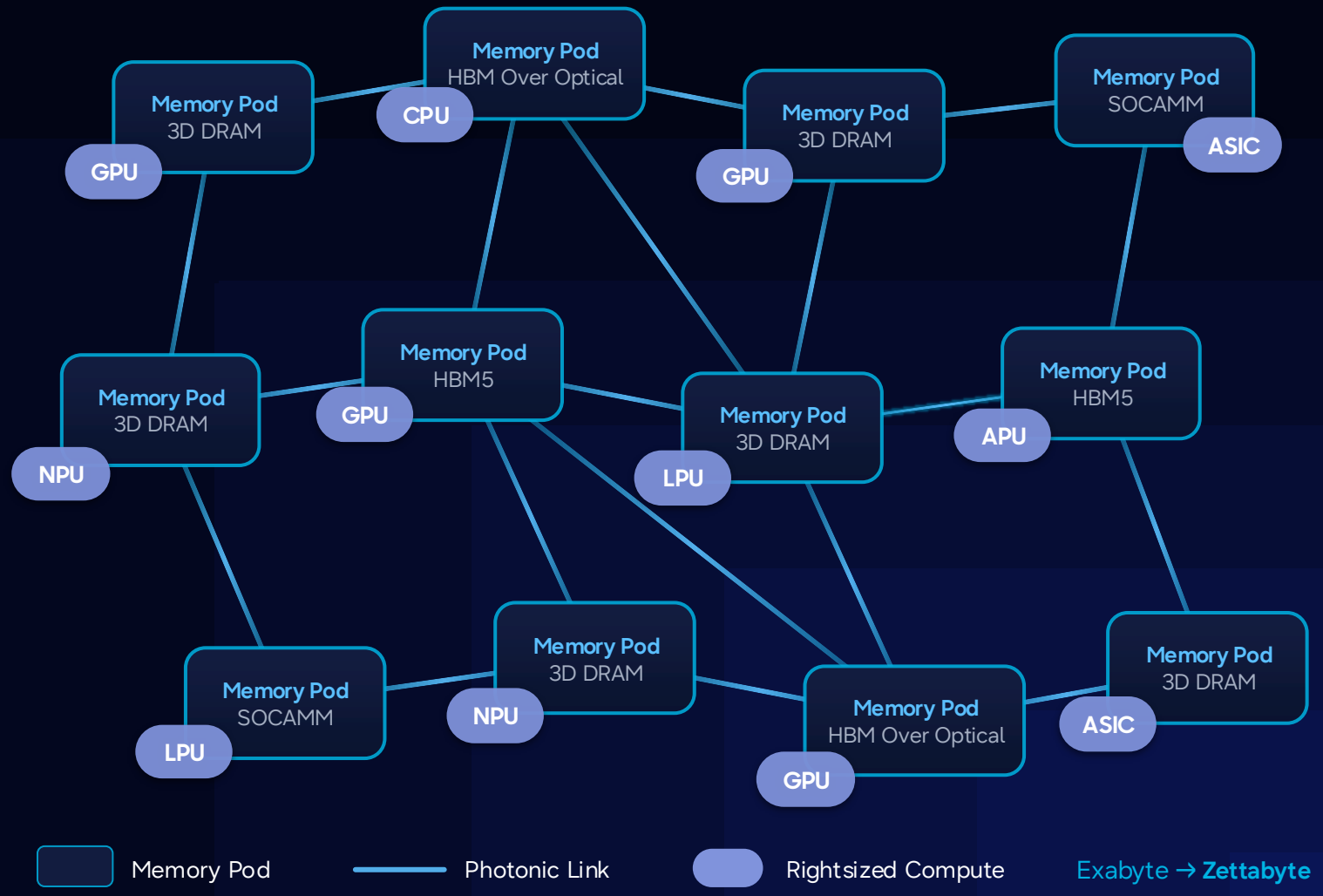
Disaggregated 3D DRAM Tiers

Scale-Free Photonic Fabric

Exabyte → Zettabyte-Scale Access

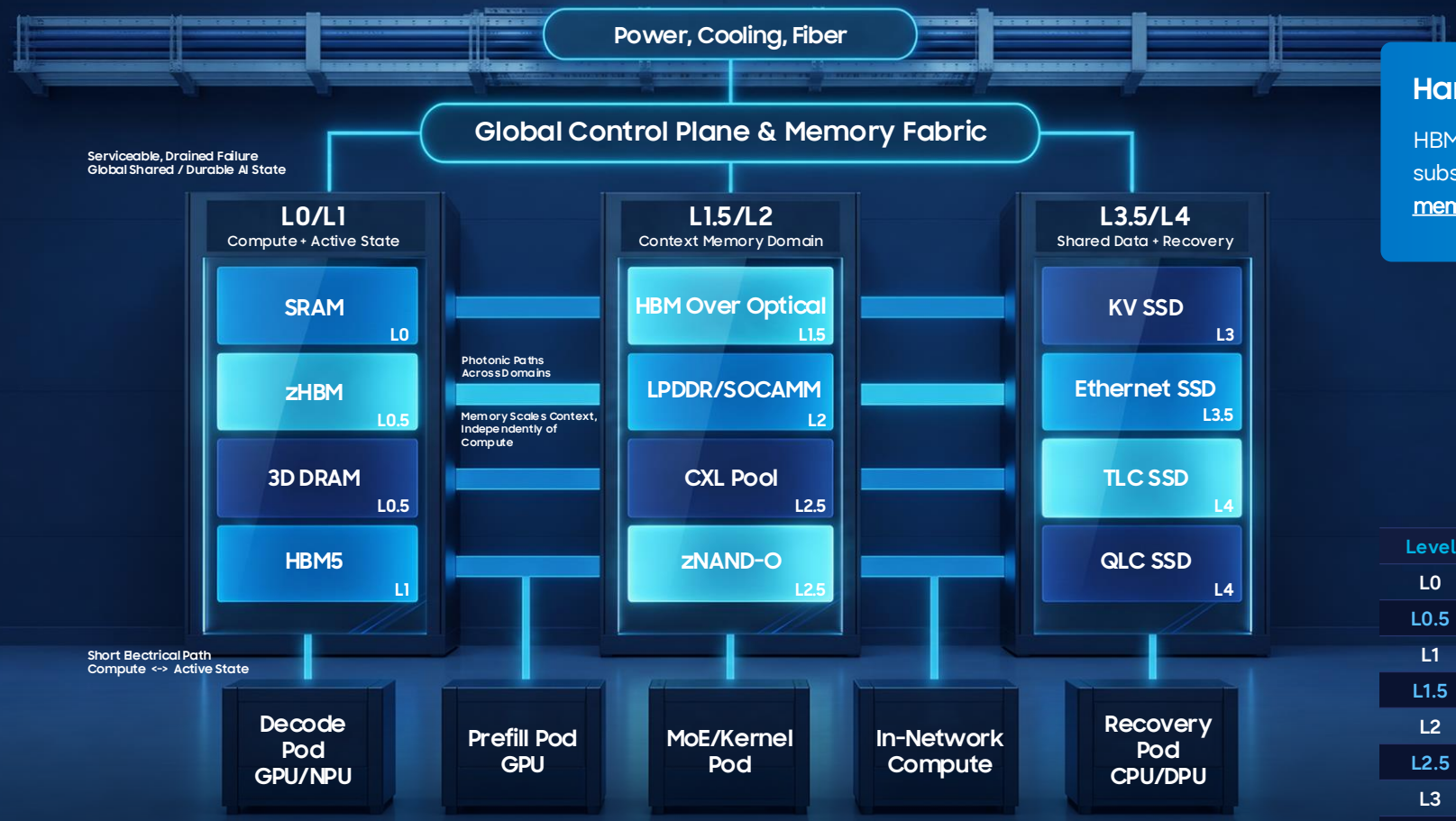
Rightsized Compute

Attaches to the Fabric by Workload



The Inference Centers of 2030

How Our Innovation Makes This Possible



Memory Bandwidth Management Across All Layers of the AI Memory Hierarchy

Hardware Evolution

HBM5, SOCAMM and glass substrates create high-density memory tiers

Deployment Tradeoff

Memory-first architecture with composable compute, high-radix fabric with localized domains

Decision You Make Today

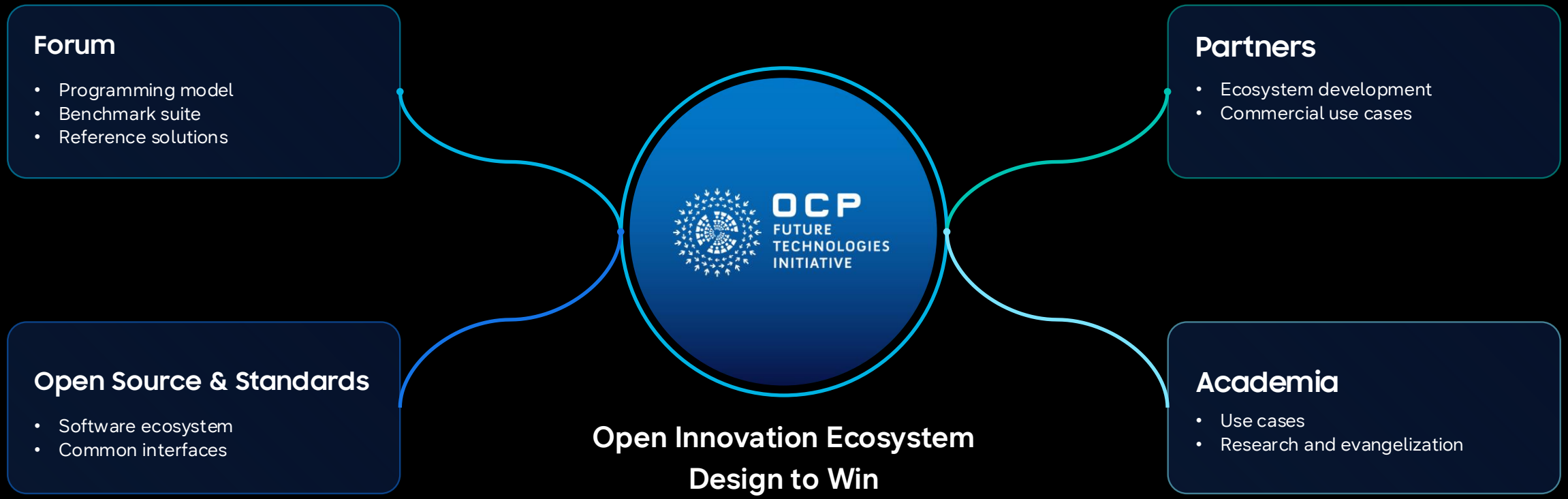
Define locality contracts and design fungible placement instrumentation per token path, not generic disaggregated memory

Level	Latency	Bandwidth	Capacity
L0	< Tens of ns	> Tens of TB/s	Few MBs
L0.5	TBD	> Few TB/s	Few GBs
L1	~Tens of ns	Few TB/s	Tens of GBs
L1.5	TBD	Few TB/s	Hundreds of GBs
L2	~Tens of ns	Hundreds of GB/s	Hundreds of GBs
L2.5	Hundreds of ns	Hundreds of GB/s	Hundreds of GBs ~ 1 TB
L3	~Few µs	~Few GB/s	Few TBs
L3.5	~Tens of µs	Hundreds of GB/s	Tens of TBs
L4	~Few ms	~MB/s	Tens of TBs

Blue = Memory Tiers Newly Introduced in the AI Era

Building the Future on Today's Decisions

Architecting AI Systems and Innovations Together



Reconsider Your Strategy, Shift Innovation Focus, and Design To Win

Thank You

SAMSUNG