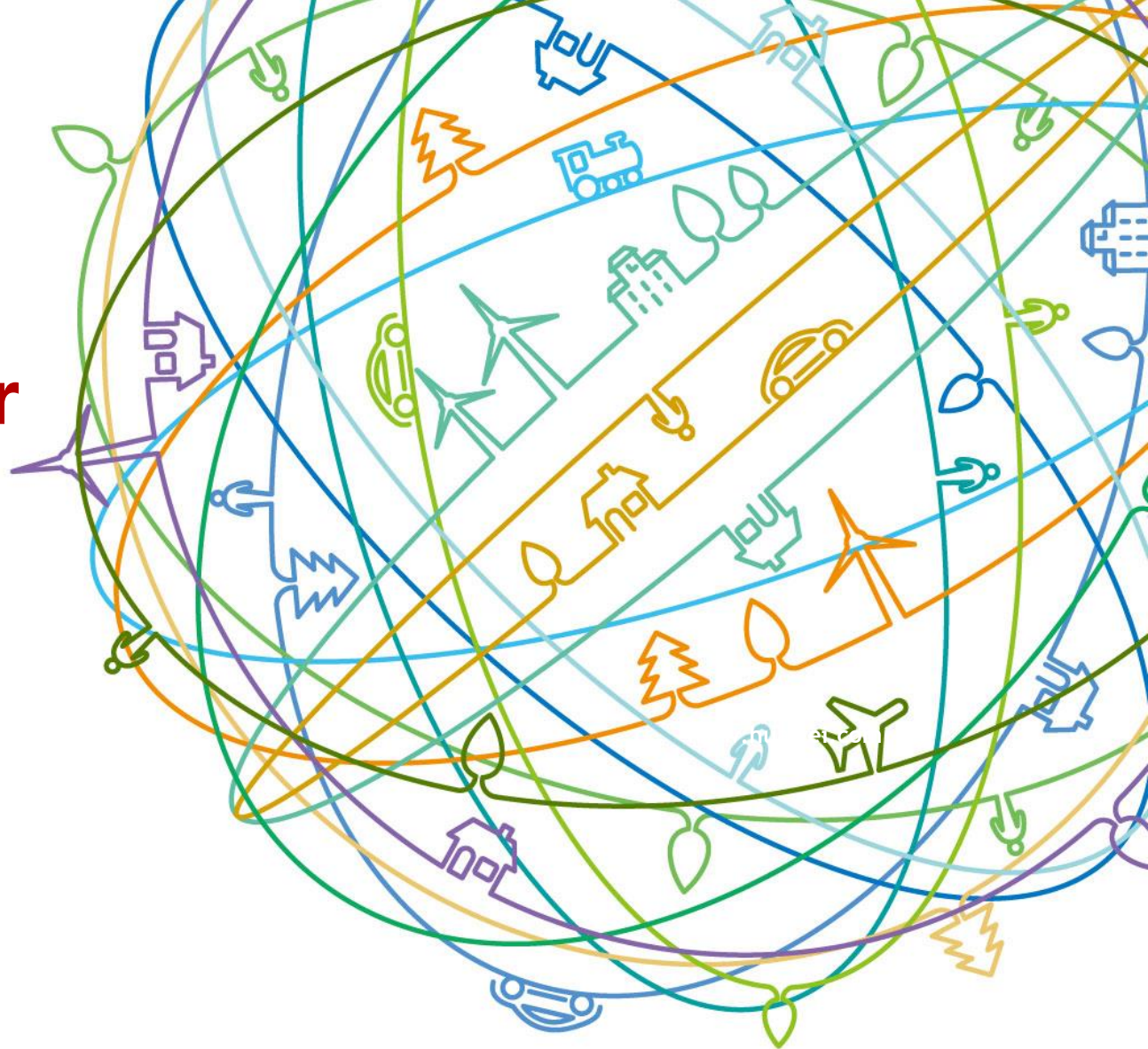
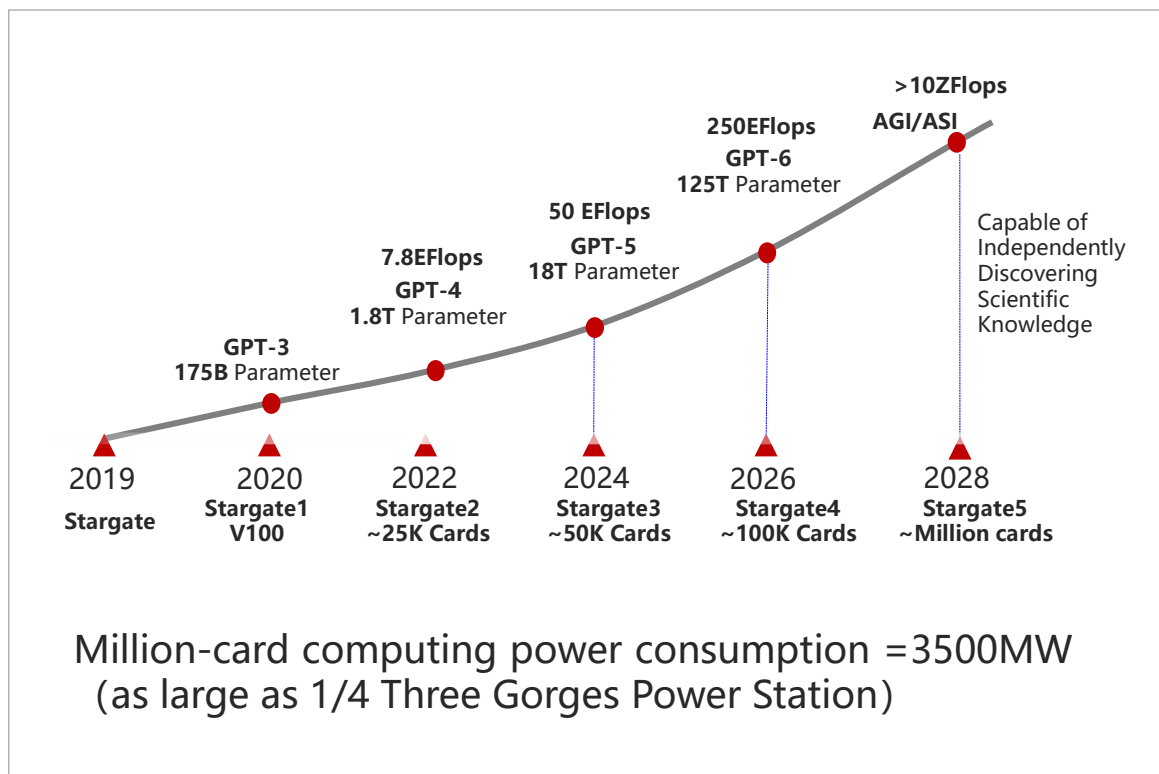


Key Technical Challenges for AI Cluster Networks



The rapid development of AGI and AI inference application rises tremendous requirement of AI computing infrastructure

Supercomputing power is the key to superintelligence(AGI/ASI)



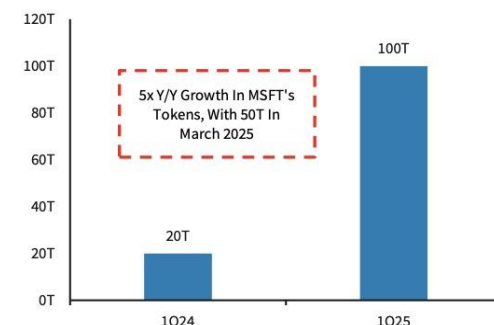
Massive AI computing power becomes essential to achieving widespread AI reference applications

FIGURE 1. GOOGL Monthly Tokens Accelerating Post AI Overview Expansion...



Source: Barclays Research, Company Disclosures

FIGURE 2. ...While MSFT Tokens Grew 5x in a Similar Time Period



Source: Barclays Research, Company Disclosures

- Google: 1Q25 634T tokens vs 1Q24 10T, **annual growth of 64 times**
- Microsoft: 1Q25 100T tokens vs 1Q24 20T, **annual growth of 5 times**

AI cluster network is essential to handle the tremendous requirement of computing power, but it also faces significant challenges

- LLM is the computing application with the most frequent computing, memory-access and communication, demanding extremely high standards for chip manufacturing and technologies.
- In the case of single-chip computing power lagging behind, it is inevitable to build larger-scale computing infrastructure through more extensive intelligent computing networks.

Total Effective Computing Power of Cluster = Single-chip computing power x Cluster scale x Computational Efficiency

- The larger and more hierarchical the network, the worse the cost, power consumption, reliability, and performance will be;
- The worse the communication performance, the lower the computational efficiency will be;
- The network itself is also constrained by Moore's Law.

- Network is the way to break the situation, but it also faces multiple contradictions and huge technical challenges:

1. High-bandwidth

2. High-capacity switching

3. Large-scale fabric

4. High-throughput

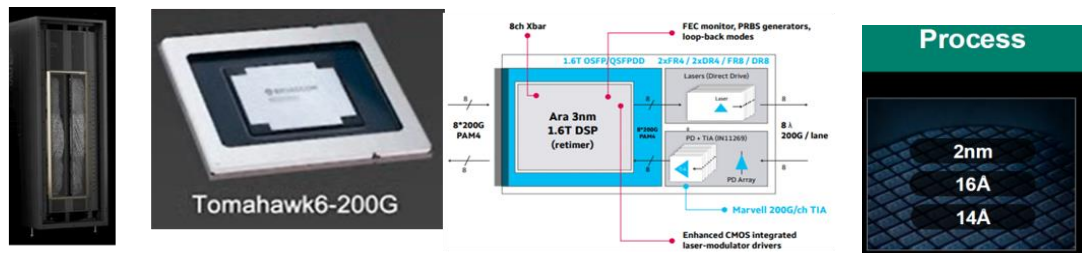
5. Short-path Data Movement

6. Long-distance connection

1. High-Bandwidth Connections: the challenges of 448G is huge, which requires researching on high edge-density optical interconnect technologies

High bandwidth is the foundation of super-nodes, which are the primary driving force behind the development of connection technologies. The rapid evolution of SerDes rates faces challenges in chip technology and driving distance.

- Trends:** future single XPU bandwidth will reach 10Tbps level; SerDes rate 112G->224G, 448G under preliminary research.
- Chip technology challenge:** 224G adopts advanced chip technology, while 448G heavily relies on chip technology. Through architectural and algorithmic innovation, achieving a breakthrough in the 224G generation.

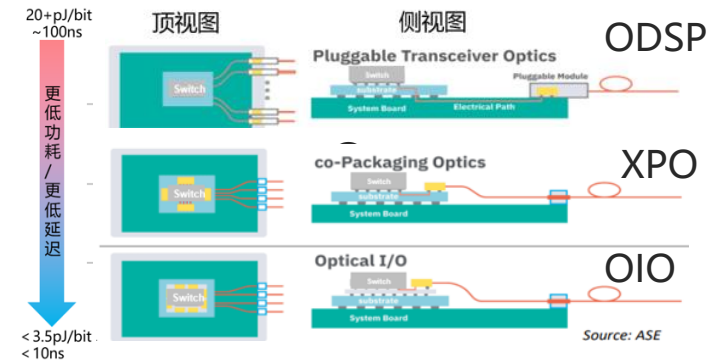


英伟达 NVL72 博通 TH6芯片 Marvell ODSP芯片 448G Serdes

- Distance Challenge:** The 448G electric drive distance may be as short as $\leq 1\text{m}$, posing a challenge for full electrical interconnection within the cabinet (requiring $\sim 1.5\text{m}$), making inter-cabinet connections impossible.

For the further, need to explore high-density optical interconnects to address the challenges of process, distance, power consumption, and latency.

- Explore high-edge density XPO/OIO technology to reduce the requirements for single-channel rates.**
- By utilizing multiple low-speed parallel channels or wavelength division multiplexing, the requirements for bandwidth, distance, power consumption, and latency are met, while reducing the demand for advanced manufacturing processes.



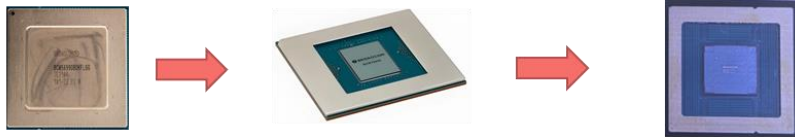
- Focusing on ①High-Density EIC, ② High-Density PIC, ③ High-Density optical connector, ④ High-Density electrical connector, ⑤Advancing 2.5D/3D encapsulation .etc**



2. High-capacity switching: The evolution of chip switching capacity has reached over 100T, and new switching architectures are key research directions

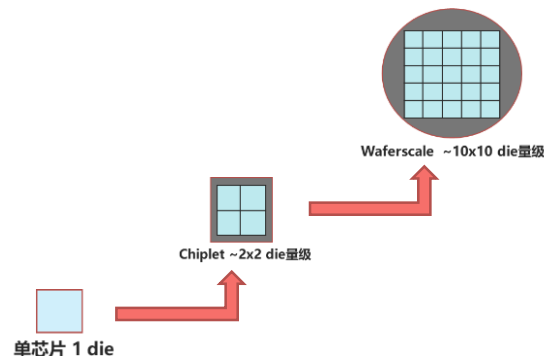
Intelligent computing drives rapid evolution of switching capacity (50T->100T) Single-Die capacity encounters a bottleneck, and the industry moves towards the chiplet architecture.

- The capacity of mainstream switching chips has rapidly increased from 51.2T to 102.4T. Transistor density is approaching the Moore's limit, and the industry is moving towards a Chiplet architecture.



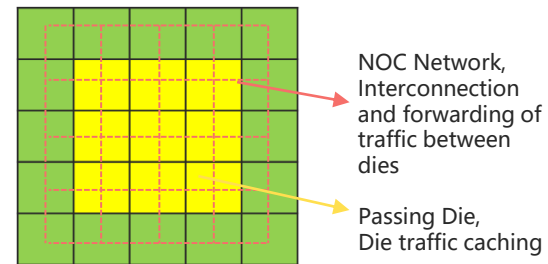
- TH4 25.6Tbps
IO: 512 x 50Gbps
单Die架构
- TH5 51.2Tbps
IO: 512 x 112Gbps
单Die架构
- TH6 102.4T
IO: 512x224Gbps / 1024x 112Gbps
1主Die+8 IO Die

- With limited chip technology, if adopting the Chiplet approach, it will require moving towards larger-scale integration and more chiplets.



In 100T+ era, Chiplet Architecture Evolution and New Switch Architecture Exploration Are Key Directions

The key challenge in building larger-scale Chiplets lies in reducing the "interconnect tax"



- The interconnection and switching overhead between internal Dies grow super-linearly with scale. How can we innovate in architecture and reuse resources to reduce the Chiplet tax?
- The complexity of routing, balancing, flow control, and cache management between internal dies increases super-linearly with scale. How can we reduce the speedup ratio?

3. Large-Scale Fabric: Scaling Demand vs. The Cost of CLOS Topology

IF: 1. Compute chip capability lags by factor x . To match total FLOPS, **cluster size must scale by x**

2. Switch chip capacity lags by factor y . With 2-tier Clos, **scaling capability degrades by y^2**

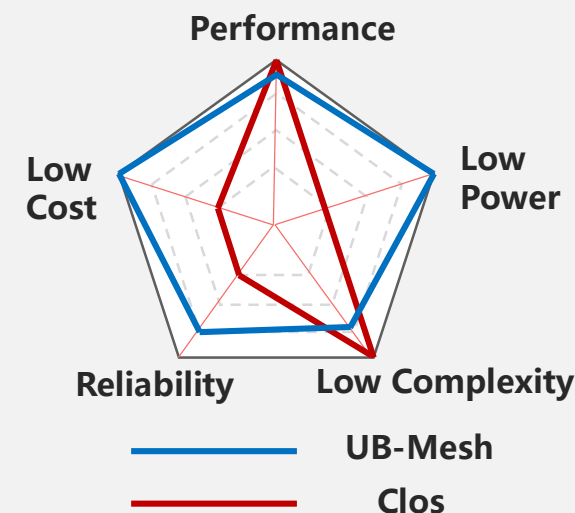
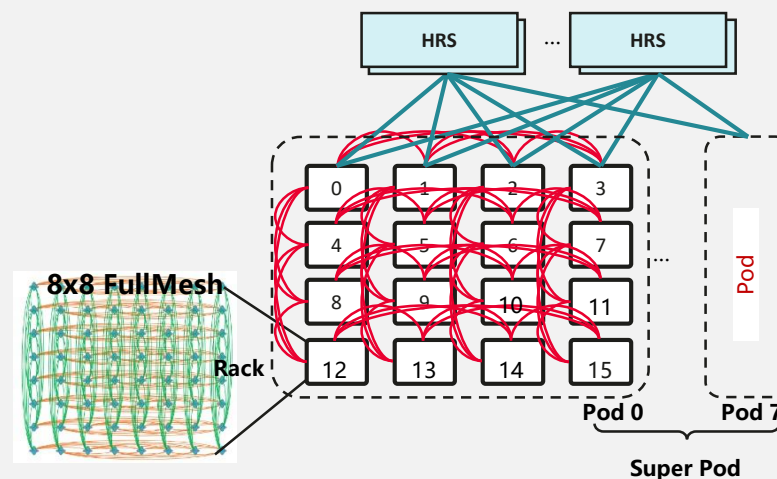
To match cluster FLOPS requires xy^2 larger network. Actual cost is **super linear** due to:

Increased switching tiers, interconnect cost, power, reliability system issues

Explore novel topologies to overcome " $>xy^2$ " scaling penalty of traditional Clos

Case Study : UB-Mesh

- Hybrid direct-connect + switched fabric
- Traffic locality: electrical (short), optical (long)
- High reliability, low cost; only ~1% training perf. loss

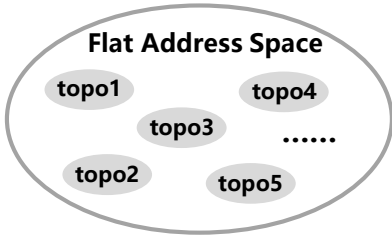


Research Direction: Cost-effective, high-performance, and reliable large-scale topologies tailored for AI traffic

Challenges: New topologies need full innovations in FPR, Deadlock-free flow control, LB, QoS & CC algorithms

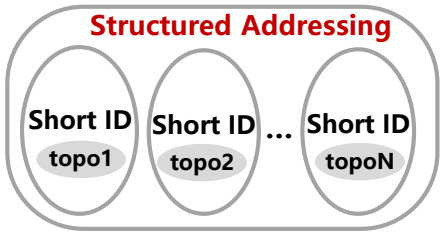
All Path Routing Technologies

1. High-perf. minimalist **Structured Addressing & Lookup**



Full Address lookup (LMP, EM)

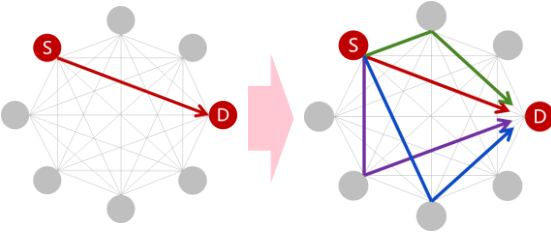
Table entry space ↓x100



Short ID[1,2,...N] Segment Linear Tables

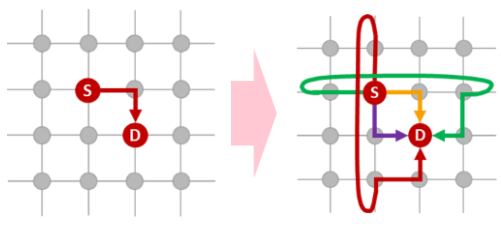
Minimized packing overhead

2. **All Path Routing**, maximize traffic bandwidth



Topology unfolding enables non-minimal paths

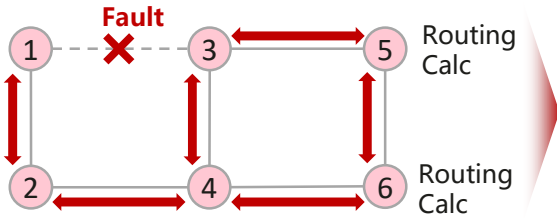
Doubles bandwidth



Exploits topo regularity for accelerated routing calc

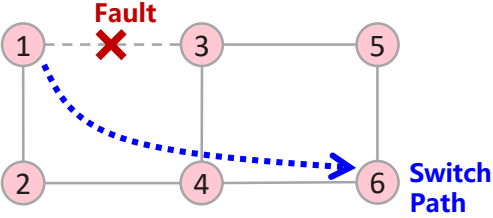
Algo complexity $O(N \log N) \rightarrow O(N)$

3. Topo-aware **Ultra-Fast Fault Handling**



Replaces "hop-by-hop flooding" with "direct delivery".

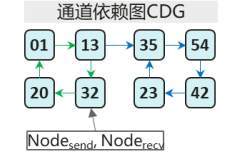
Ultra-fast routing convergence



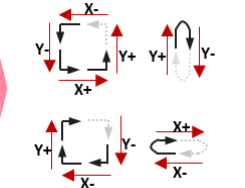
50%+ control plane load reduction

4. Topo & route-aware **Flow Control Channel Planning**

Models potential deadlocks via CDG



Applies cycle-breaking theory for channel shifting



Design table based on $C \times N \rightarrow C$

Input Channel	DstIP	Output Channel
0	0	0
0	1	1
...

Design table based on $C \rightarrow C$

Input	Port1	Port2
0		0.7

Reduces VC requirement 3x

Table entry from $O(N)$ to $O(1)$

4. High-throughput: Building Efficient and Reliable Transport Protocol for Scale-up Network

Fundamental Changes for Network Settings

Host Bus

small scale, high bandwidth,
ultra-low latency, high reliability

Latency	60~100 ns
Bandwidth	1 Tbps
Scale	1 host
Reliability	high
Semantics	load/store
Typical Technology	PCIe

Scale Up Network

large scale, high bandwidth,
low latency, low reliability

2~4 μs
~10 Tbps
256-2K host
low
load/store
NVLink, UB, UALink.....

Key Challenges to Transport Protocol for Scale-up Network

1. Bandwidth-delay product increased by two orders of magnitude, **AICore's outstanding window is too small to support full bandwidth transmission**

Outstanding window limits
of on-fly packets

1 2 3 4

1 2 3 4

Outstanding window bounded

2. Transport retransmission for memory semantics incurs high memory/processing overhead, leading to **extra cost in chip area and latency**

Message

Host HBM

Retry Buffer

IO-Die

Packet

command

Core buf/cache

Host HBM

Retry Buffer

Full Data

IO-Die

Packet

3. Single path → non-equivalent multi-path, emerging mesh topology poses **new challenges to bandwidth, latency, and ordering**

3. Switch Detour

Switch

D1

D2

D3

1. Direct path

2. Mesh Detour

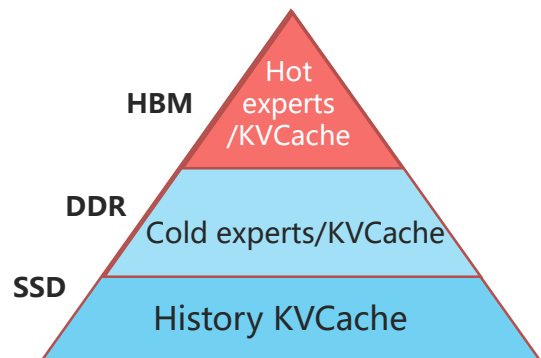
8

Huawei Confidential

5. Short-path Data Movement: fully P2P architecture and short-path direct access

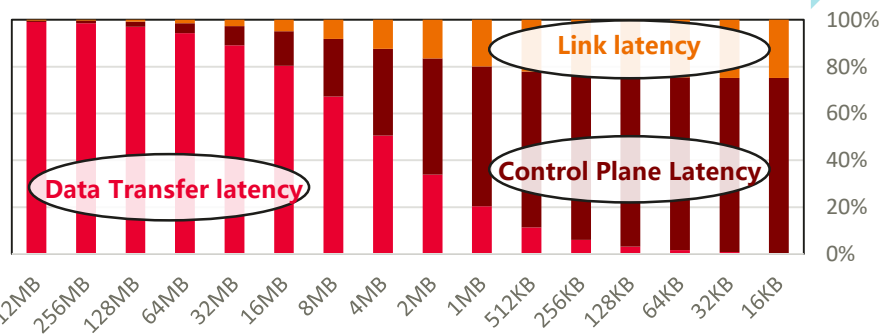
High BW transmission → Sparse Communication

Due to limitations in computational power and capacity, large model inference is gradually moving towards sparsity and hierarchical approaches.



With small data volumes and sparse transmission, the collective communication time is dominated by link and control plane delays, shifting the conflict from data transmission to system architecture and structure.

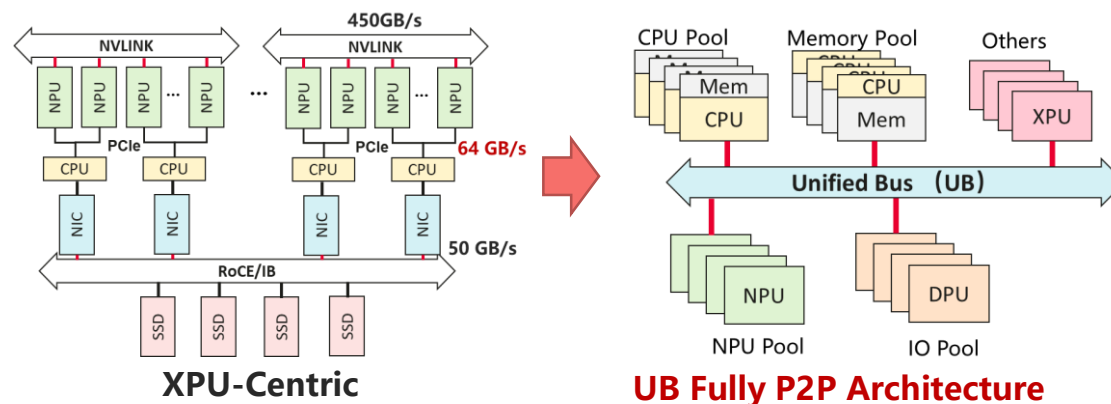
Large data volume → Sparse Collective Communication



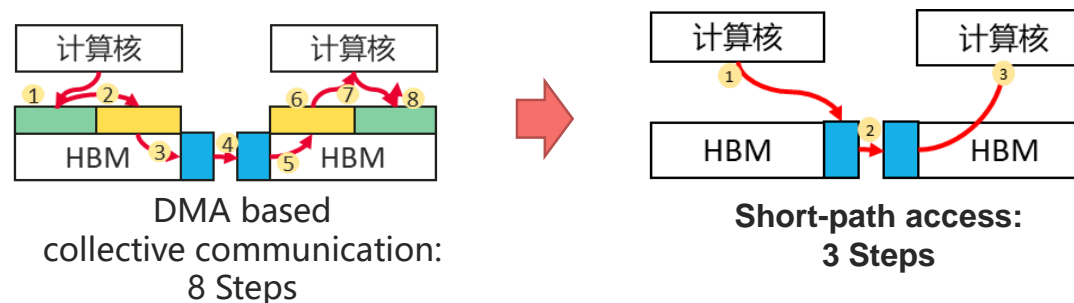
Estimated parameters: 15μs control plane latency, 5μs forwarding + link delay, 1.6Tbps bandwidth

System-level innovation: Short-path Data Movement

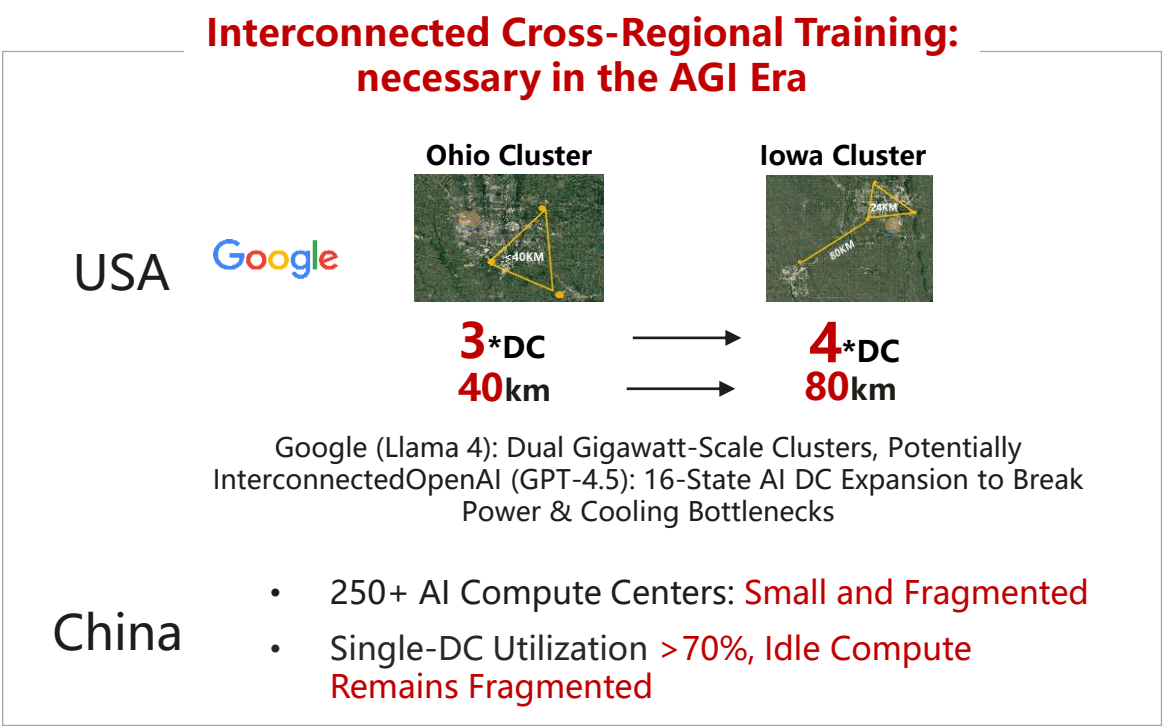
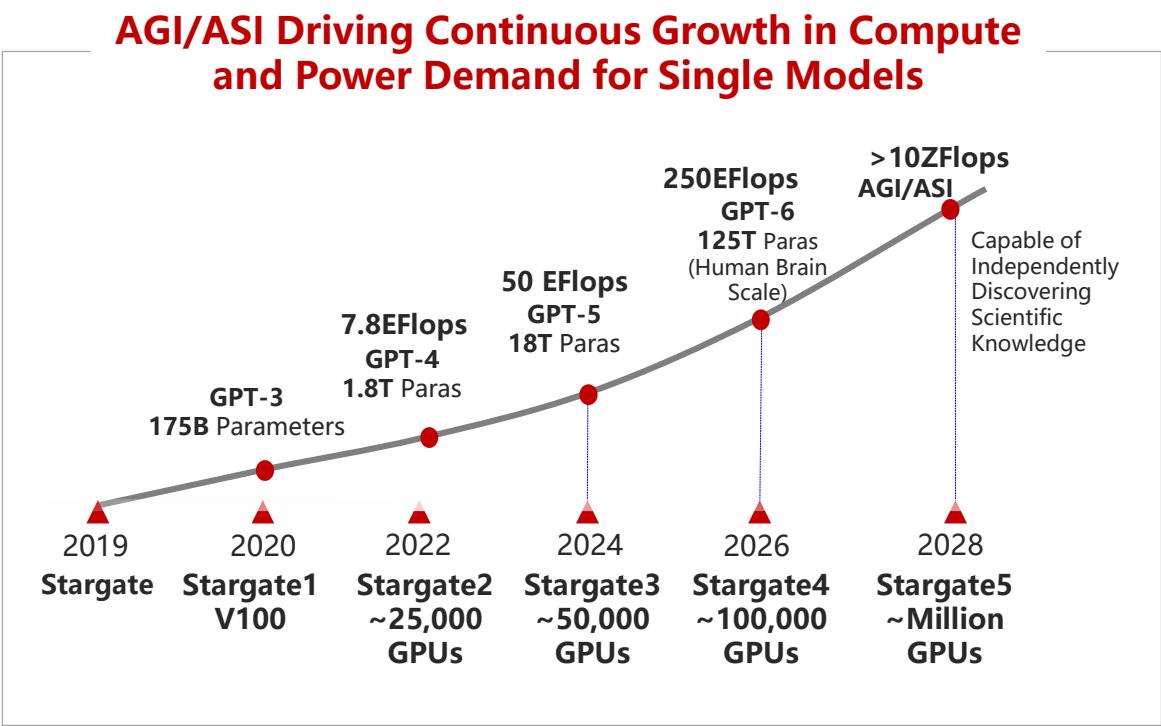
Challenge 1: In a centralized architecture, cross-layer data migration faces multiple protocol conversions and significant control plane overhead. It is necessary to establish a **fully equivalent interconnection architecture to eliminate migration bottlenecks**.



Challenge 2: DMA semantics across XPU data transfers involve multiple HBM data read and write operations, resulting in significant data plane overhead. **It is necessary to establish a short-path direct semantics to achieve HBM bypass**.



6. Long distance connection: huge computing power requirement rises distributed AI training



1. High-Throughput Challenge: How to Scale RoCE Across WANs?

- How to reach **90%+ DCN RoCE throughput** over 1,000 km?
- How to minimize collective communication time via **load balancing**?
- How to ensure task-level **guarantees and isolation**?

2. Parallelism Challenge: How to Maximize Efficiency in Large-Scale Compute/Communication Parallelism?

- How to **optimally partition and deploy compute tasks** across domains for 90%+ scalability?
- Optimizing collective communication** for high-convergence, high-latency networks

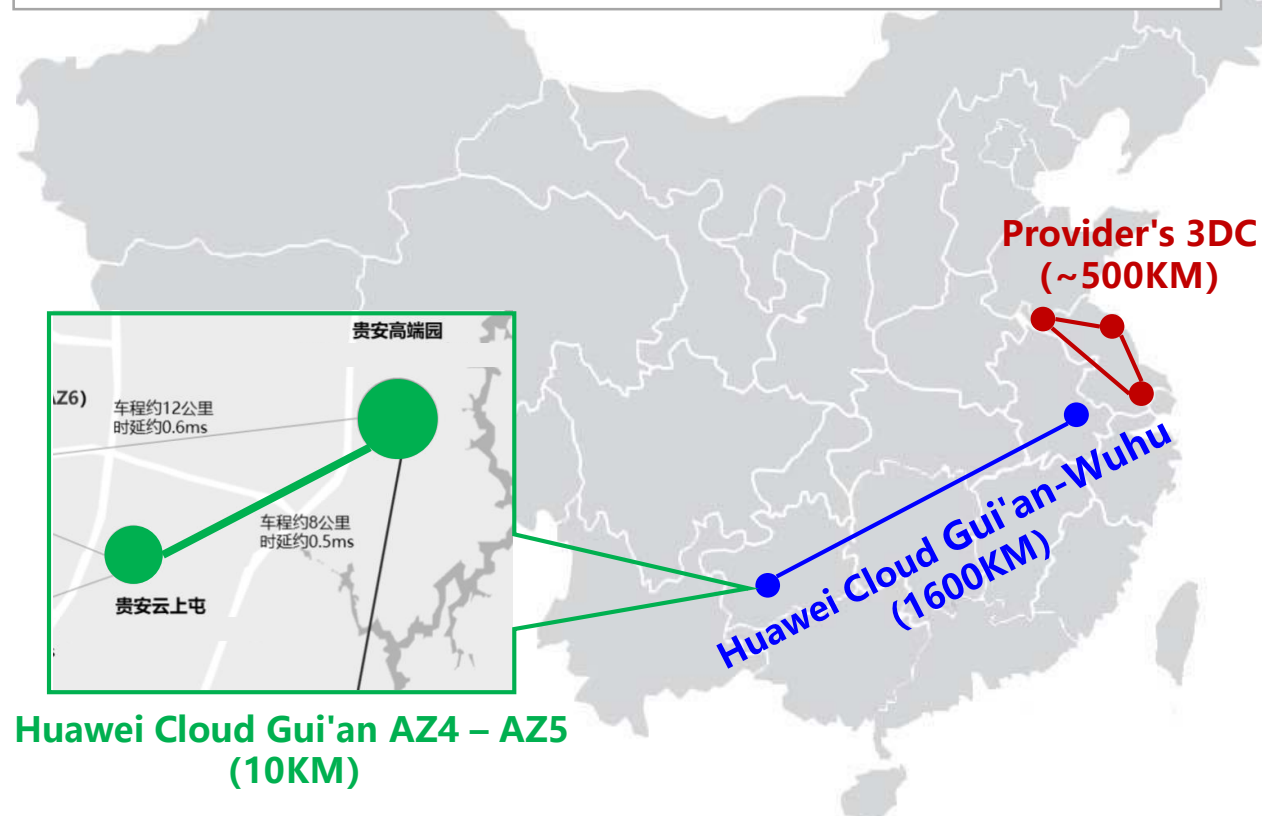
3. Resource Scheduling Challenge: How to Flexibly Orchestrate Compute, Network, and Storage for High Efficiency?

- Elastic scheduling** of compute and network resources to balance efficiency and utilization
- Breaking storage silos to enable **unified data view and free data flow**

Distributed AI Training Practice, achieving less than 5% performance loss @1000KM

Long-Distance Training (incl. PP/DP parallelism):

1. 500km, Provincial 3DC: GPT3-175B training, 2.8% performance drop
2. 1600km, Gui'an-Wuhu 2DC (Huawei Cloud): LLaMA2-70B training, 2.5% drop
3. 10km, Gui'an AZ4-AZ5 (Huawei Cloud): Pangu 8x8B MoE training, 3.5% drop



Technical Stack

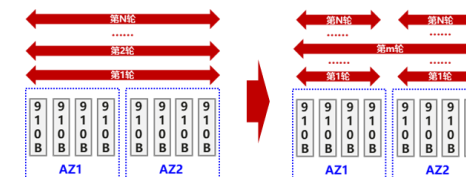
Scheduling

Multi-plane hierarchical parallel comms **hides 30% comm time**

Adaptive model partitioning, **accelerates training by 30%**

Collective Comms.

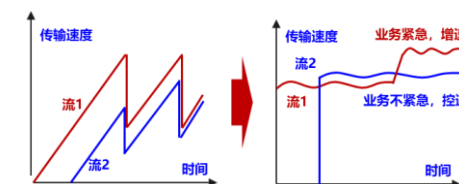
AHC Hierarchical Asymmetric Collective Communication



Compresses inter-AZ traffic by **factor N** (N = #machines)

Transport

FlatRate congestion control



Reduces cross-AZ P99 latency by **>50%**

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and
organization for a fully connected,
intelligent world.

**Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

