Breaking Through the AI Network Wall with High-radix Switching

OCP Educational Webinar Series

July 28, 2026

OPEN Compute Project®

650 GROUP MARKET INTELLIGENCE RESEARCH

ERIDU

Community-driven hyperscale innovation for all

Breaking Through the AI Network Wall with High-radix Switching

Cliff Grossner, Ph.D.

Chief Innovation Officer

OCP Foundation

Alan Weckel

Founder, Technology Analyst

650 Group

Promode Nedungadi

Chief Technology Officer

Eridu

Jay Gill

Sr. Director of Product Marketing

Eridu

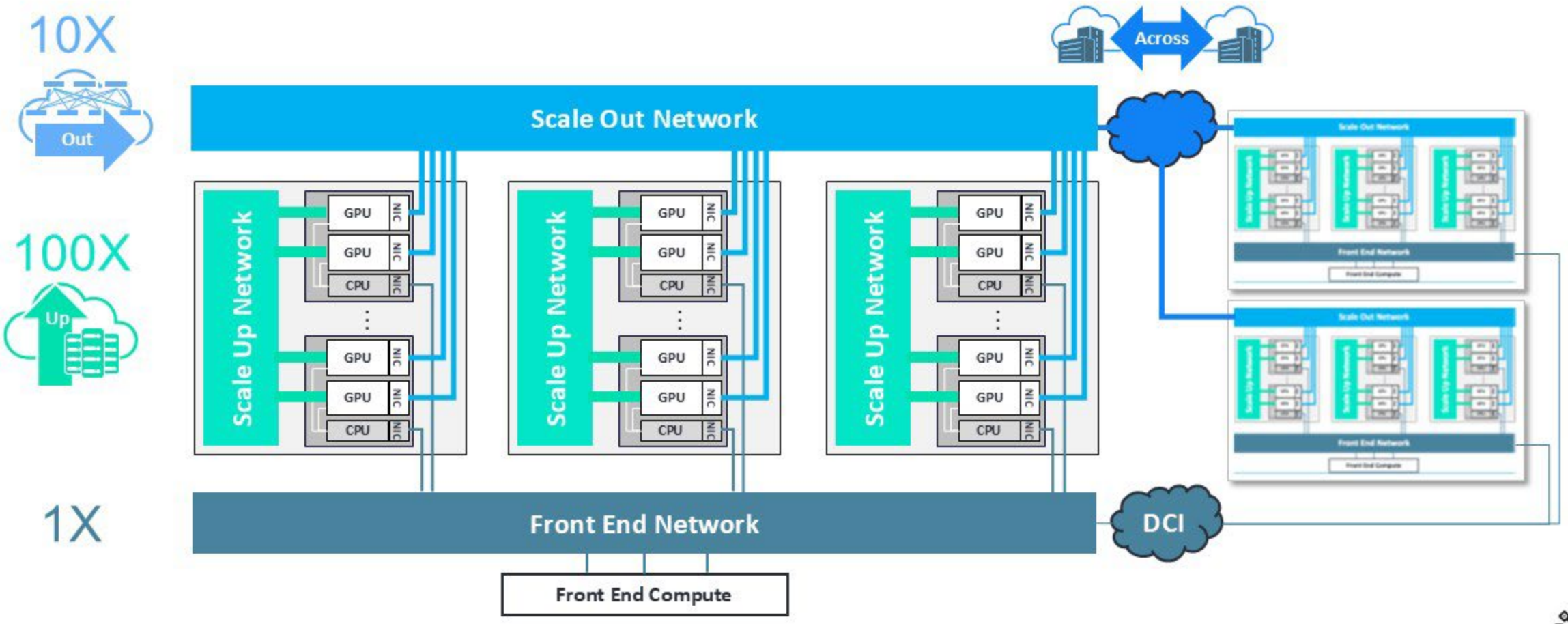
Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

AI Networks:
Front End + Scale Up/Out/Across



ERIDU

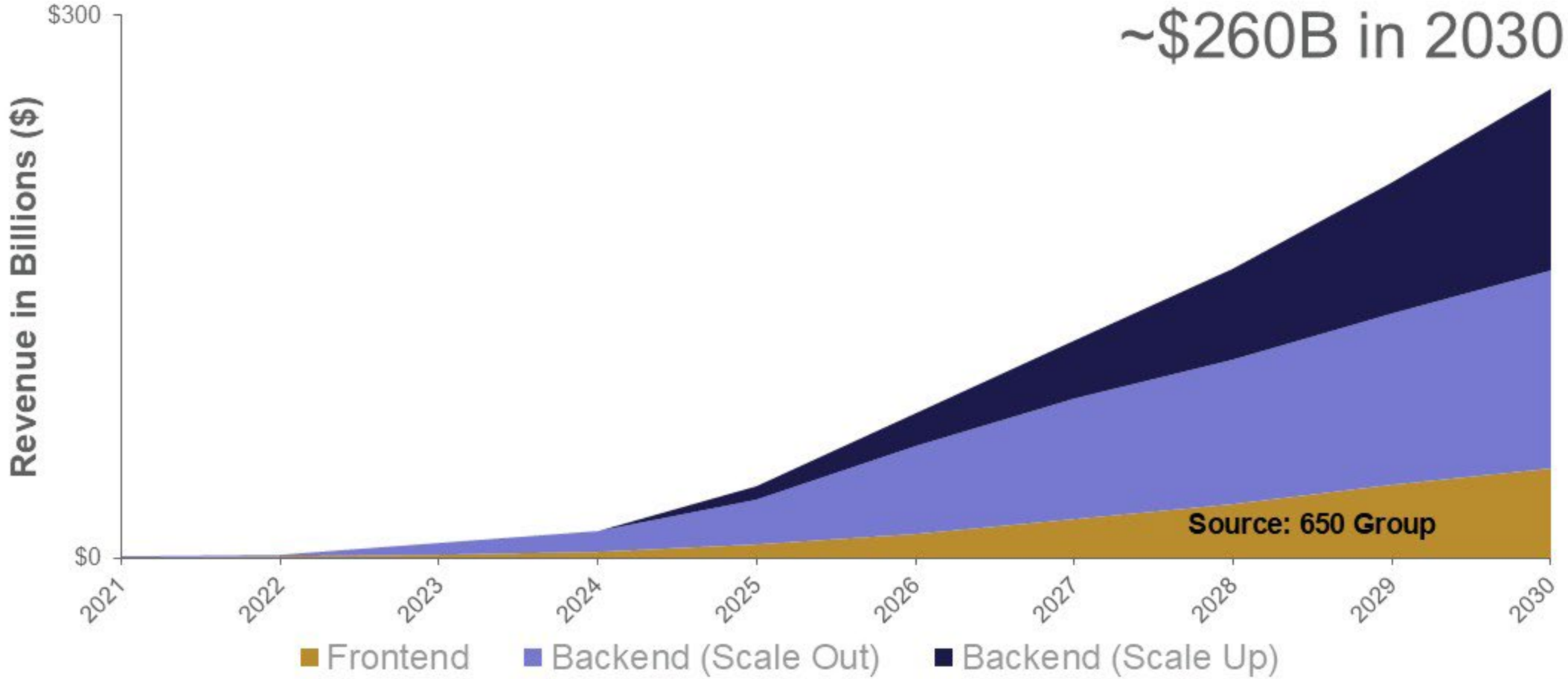
DC Evolution: Traditional Compute to AI Campus

Traditional Data Center	Volume/Size
Number of Buildings	1
Servers per Building	100K
CPUs per Server	2
GPUs per Server	0
NIC Ports	100K
NIC Port Bandwidth	25-100Gbps
Management Ports	100-120K
DC Switch Ports	300-400K
Cables/Fibers	300K-1.2M

AI Campus (1M XPUs)	Volume/Size
Number of Buildings	4-12
Servers per Building	100-200K
CPUs per Server	2-4
GPUs per Server	8
NIC Ports	1.1-1.4M
NIC Port Bandwidth	400Gbps – 1.6Tbps
Management Ports	1.2-2.5M
DC Switch Ports	3.4-6M
Cables/Fibers	6-24M

650 GROUP
MARKET INTELLIGENCE RESEARCH

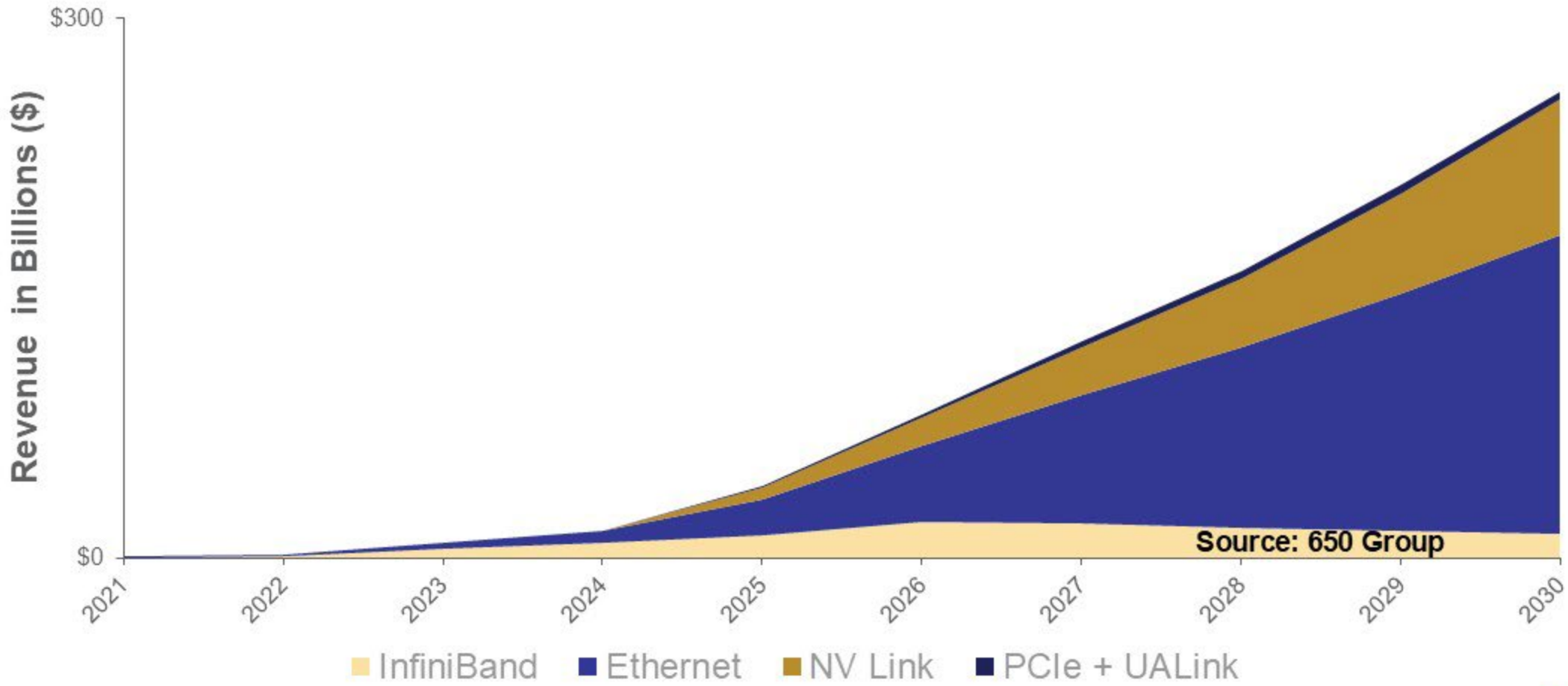
Networking for AI Total Market
Backend dominates, scale up grows fastest



*Scaleup portion is for standalone switch hardware only and does not include embedded devices within the server or portions of accelerators.

650 GROUP
MARKET INTELLIGENCE RESEARCH

Networking for AI Total Market
Increasing Ethernet share across all networks



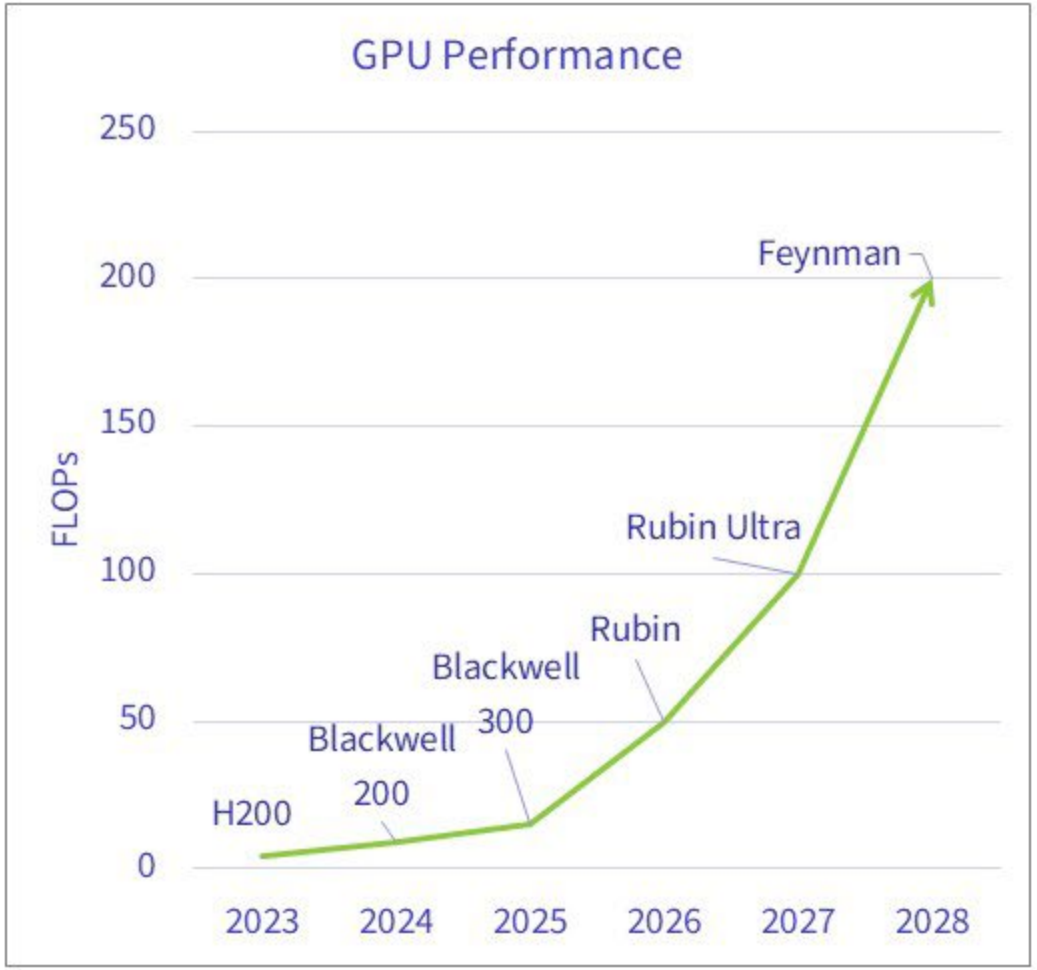
*Scaleup portion is for standalone switch hardware only and does not include embedded devices within the server or portions of accelerators.

650 GROUP
MARKET INTELLIGENCE RESEARCH

Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

AI Scaling Pushes Networks Relentlessly



Larger high-bandwidth domains, clusters
Scale up to 1k+, Scale out to 1M+

Evolving model architectures
MoE, more experts, active experts

Larger context windows
Memory, personalization, multimodal

Disaggregated inference, heterogeneous compute
Prefill & decode in separate racks

Agentic era
Hundreds of API calls per prompt

Managed agents with memory
AI workflows + storage appliances

ERIDU

Switches Are Not Keeping Up

Conventional Switch Architecture
Single-chip ASIC

Scaling 2x every ~2.5 years

GPUs sit idle up to 60% of the time

Networking uses >25% of the power

Network cost 20% of IT spend

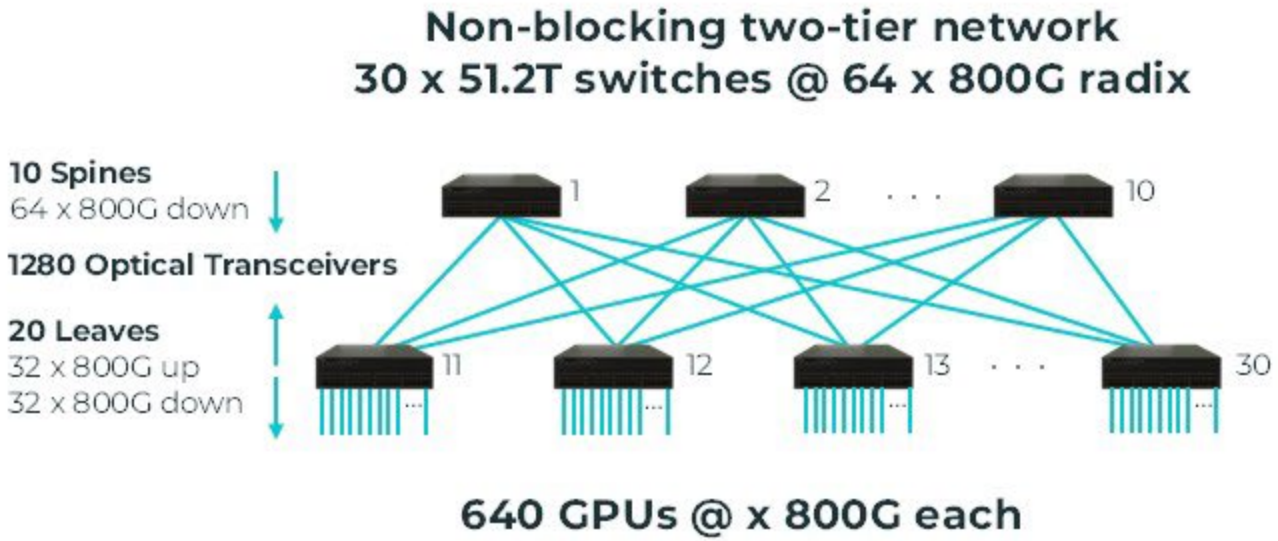
Network failures often stop training jobs

Revenue loss due to slow deployment, poor tokenomics

$BILLIONS in wasted capex and lost revenue

ERIDU

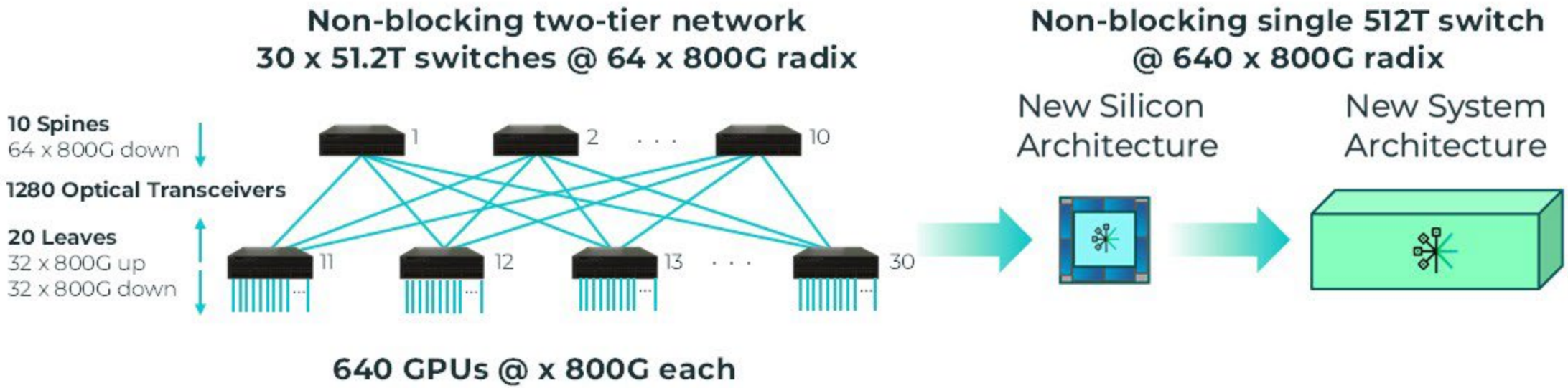
The Network Wall:
A Simple Example Connecting 640 GPUs



- 30 switches required to achieve non-blocking 512T
- Two tiers (1-3 hops)
- Variable latency and higher probability of congestion vs. single-tier network

ERIDU

10x Radix → 30x Reduction in Switches + Improved Power, Cost, Performance, Simplicity



- **>50%** reduction in power
- **30-50%** reduction in cost
- Single switch hop for **lowest latency**

- **Higher reliability**
- Massively **reduced cabling**
- **Faster deployment**

ERIDU

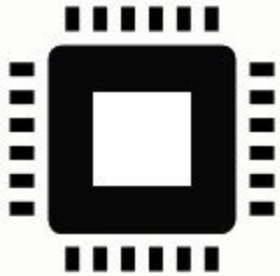
Switch ASIC Scaling History

ASIC Capacity	Process Node	Radix
3.2 Tbps	28nm	128 x 25G
6.4 Tbps	16nm	256 x 25G
12.8 Tbps	16nm	256 x 50G
25.6 Tbps	7nm	512 x 50G
51.2 Tbps	5nm	512 x 100G
102.4 Tbps	3nm	1024 x 100G 512 x 200G

- SERDES lanes continue to increase
- Higher radix switches significantly lower deployment costs
- Switch silicon has already moved from trailing to leading edge process nodes
- Innovation is needed beyond moving to lower process nodes

650 GROUP
MARKET INTELLIGENCE RESEARCH

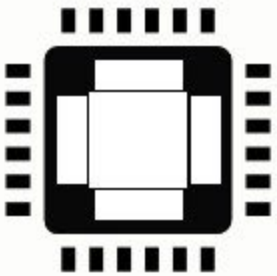
Switching Silicon
From Evolution to Revolution



51.2 Tbps
2022*

- Monolithic ASIC
- Embedded I/O
- 5 nm

2x



102.4 Tbps
2026*

- Monolithic core
- Separate I/O tiles
- 3 nm

10x

?

How to build a switch with 10x radix?

* Year when switching systems based on this generation of ASICs began volume shipment (Source: 650 Group)

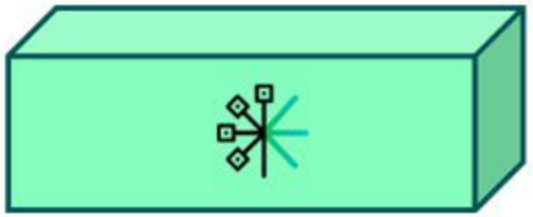
ERIDU

Technical Challenges of Increasing Radix

New Silicon Architecture

- Buffer architecture
- I/O density
- Interconnect technology
- ASIC power
- ASIC cooling

New System Architecture



- Front panel density
- Interconnect technology
- System power
- System cooling
- Software

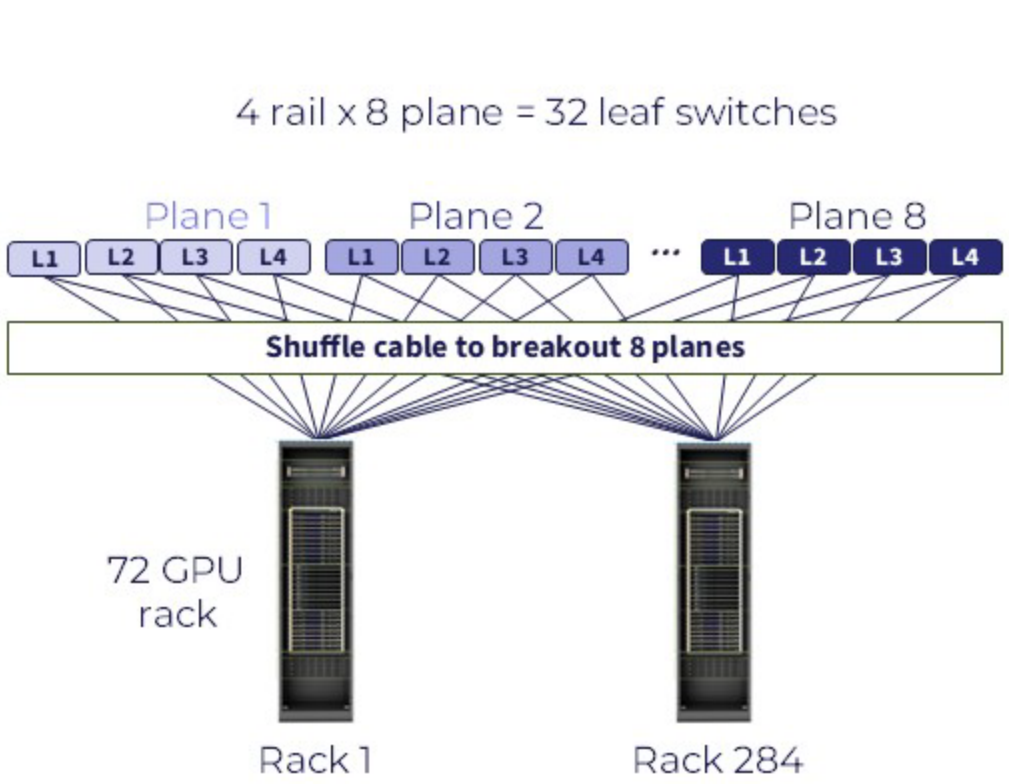
Co-design silicon + systems + software

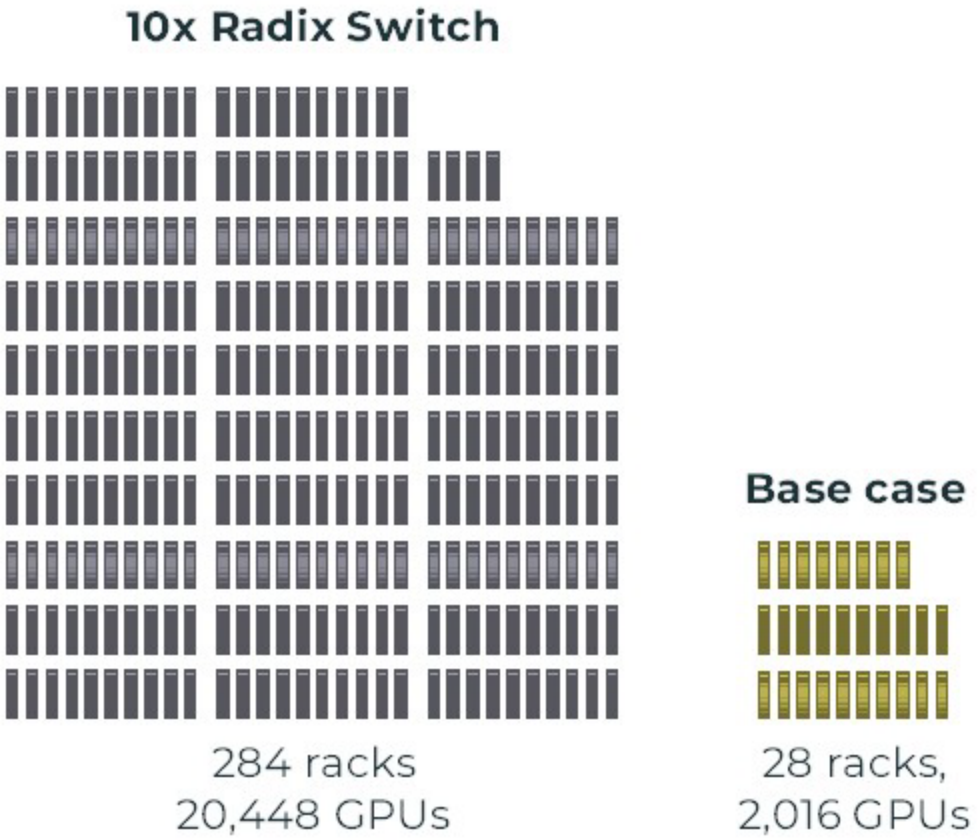
ERIDU

Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

High Radix Enables Large Single-tier Scale-out Clusters





ERIDU

High Radix Delivers Large Scale-out Advantages

Network designs for ~1 GW building/campus using Vera Rubin NVL72

	102.4T switch	1P switch	
Topology	Three tier	Two tier	2 tiers vs. 3
Pod size, aka "scale unit"	1,008 GPU packages	10,224 GPU packages	Larger scale unit
Total GPU packages	258k	265k	More GPUs
Total switches	20,480	1,344	Fewer switches
Network power	81.9 MW (9% of total)	45.7 MW (5% of total)	Less power

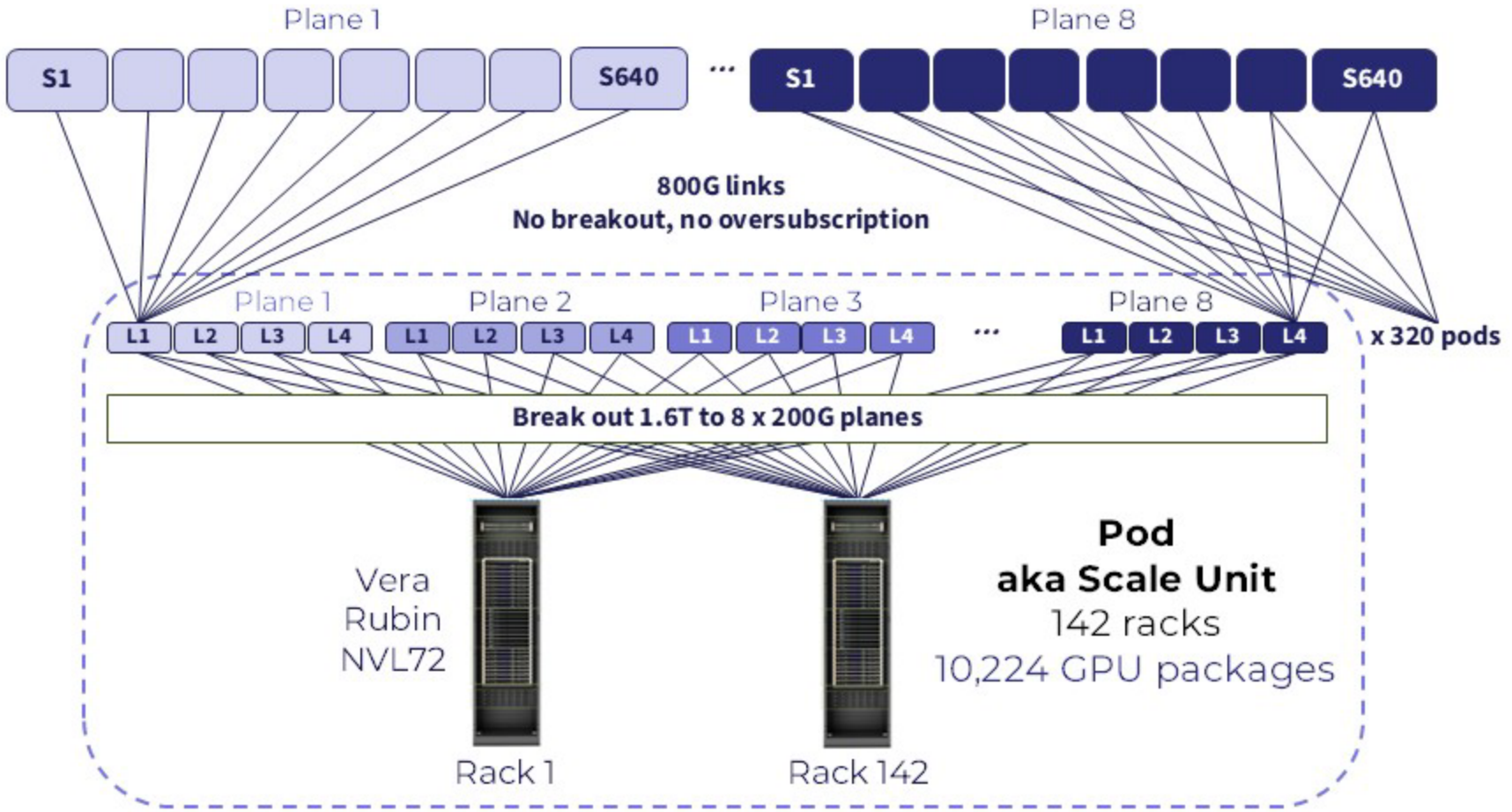
ERIDU

Extreme Scale-out/Across in Two Tiers: Up to 3.27M GPUs

5120 spine switches
Each spine
1280 x 800G down

10,240 leaf switches
Each leaf
640 x 800G up
2560 x 200G down

Total Cluster
320 Pods
~3.27M GPUs



Never need to go beyond 2 tiers. Never break out leaf to spine.

ERIDU

Audience Poll Question

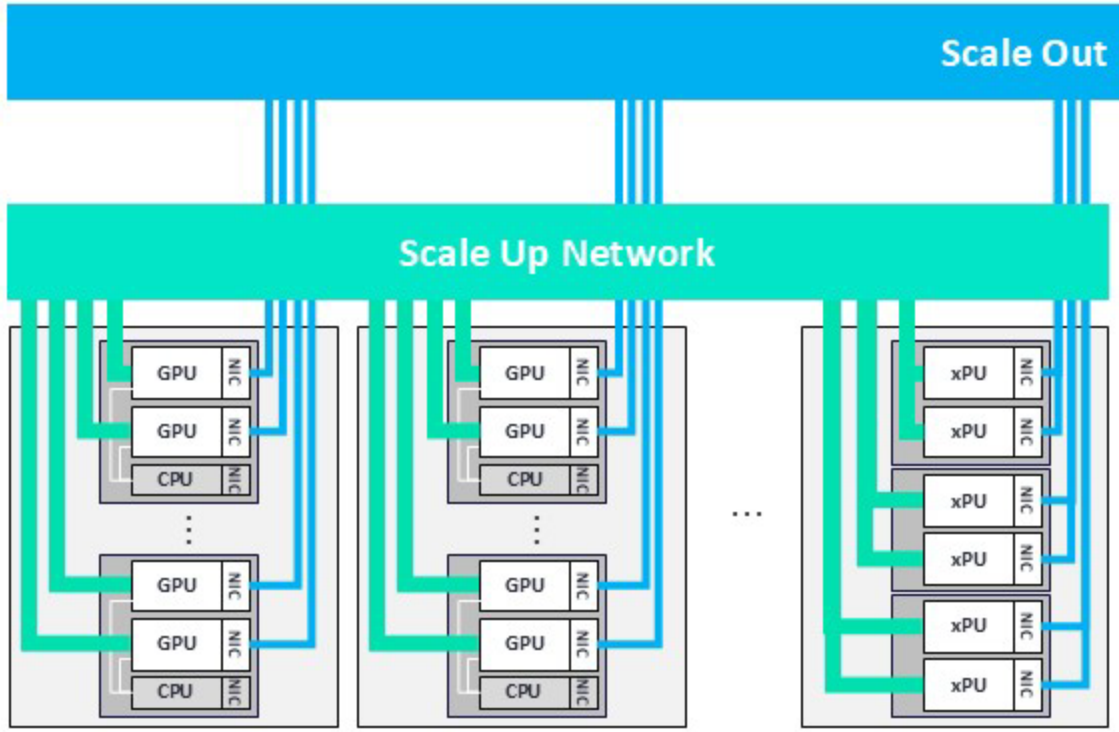
What is the largest scale-out/scale-across cluster that you plan to deploy or use within the next 3 years?

- <1000 GPUs
- 1000 – 10,000 GPUs
- >10,000 – 100,000 GPUs
- >100,000 GPUs

Audience Poll Results

- Some discussion facilitated by host and analyst of the results with optional comments from sponsors

Scale Up Evolution:
Multi-rack, Disaggregated, Heterogeneous, Open



- **Multi-rack**: grow world size beyond 72 xPUs, maintain high-bandwidth, single-hop connectivity
- **Disaggregated**: mix/match compute racks, separate networking racks
- **Heterogeneous**: GPUs, CPUs, xPUs for diverse training, inference, and agentic workloads
- **Open**: Ethernet scale-up networking (ESUN), open racks (e.g. ORv3), optical MSAs (e.g. OCI)

ERIDU

Growing Scale-up Domains

Strong demand for “world sizes” beyond rack-scale...
... but how do you build the network?

Thousands of GPUs

Hundreds of GPUs

You Are Here

72 GPUs

Disaggregated compute + network
Optics + high radix switching required

8 GPUs

Single hop, low-latency between every GPU pair

Server

Rack

Multi-rack

Multi-row

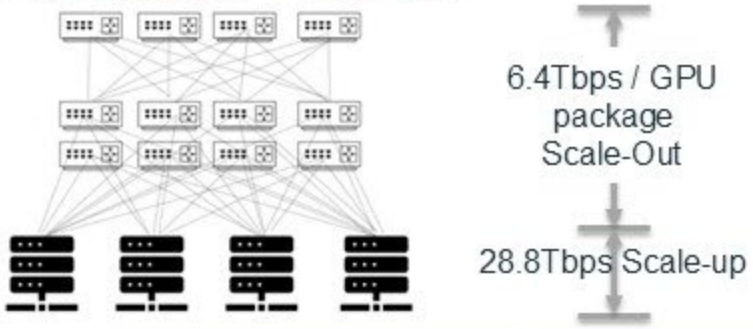
ERIDU

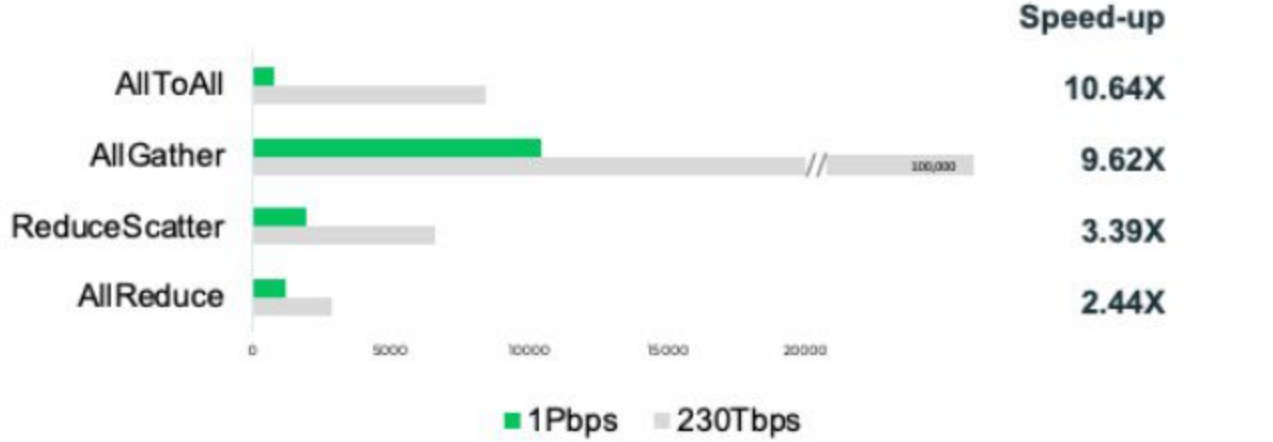
5.26x Speed-up in Job Completion

Training Simulation with Large Scale-up Domain

5,120 GPU with 1 Pbps switch
scale-up

5,120 GPU with 230 Tbps switch
72 GPU scale-up + scale-out *





Training Model	Parameters
Open source MoE	500 billion parameters
Number of Experts	256
Expert Parallelism	64
Pipeline Parallelism	2
Data Parallelism	40
Sequence length	131,072
Global Batch size	38,400

- Packet level simulations (NS3) using Scala Computing Platform
- Mixture-of-Experts (MoE) workload using Symbolic Tensor Graph (STG) Generator: https://github.com/astra-sim/symbolic_tensor_graph
- Transport on scale-out network: RoCEv2 with PFC
- Profile used on scale-up network: ESUN/SUE-T

* Building a comparable 5,120 GPU scale-up domain with 230 Tbps switches is impractical, requiring >10,000 switches

ERIDU

Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

Who is Eridu?

ERIDU
FASTER AI

- Founded 2024, exited stealth March 2026
- Developing high radix switching solutions
- Raised over $300M
- Leadership team with several multi-$B networking product lines and multi-$B startup exits

ERIDU

Eridu's Vision for a New Generation of AI Networking

The next generation of networking for AI will be defined by:

- ✓ An order of magnitude leap in radix.
- ✓ A clean-sheet approach with extreme co-design: silicon + systems + software.
- ✓ The ability to support scale-out clusters of hundreds of thousands of GPUs and scale-up domains of thousands of GPUs.

Stay tuned for more details!

ERIDU

Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

Key Benefits of Higher Radix Switching

Larger scale-out clusters with flatter, simpler networks

Larger high-bandwidth scale-up domains

Economics
Higher GPU utilization, lower cost & power

Performance
Lower latency & congestion, higher reliability

Operations
Simpler cabling, deployment & management

Better AI outcomes

More model intelligence & flexibility

Faster training

Lower cost per token

ERIDU

Take Aways

Networking is increasingly becoming a bottleneck to unlock XPU/GPU value

- Evolving DC architectures require more bandwidth, more tiers and thousands of switches
- Increased cost, power, and operational complexity along with slower deployment, lower reliability
- Ultimately, the networking wall throttles collective communications resulting in GPU underutilization

Larger Scale Up Domains support larger and more intelligent models

- Scale up domains must stay in a single tier, thus world size is limited by switch radix
- Larger radix switches and disaggregated, optically connected ORv3 racks can deliver smarter models and increased GPU utilization

The speed of networking ASIC innovation needs to increase

- XPU/GPU product cycle is 1-2 years with each generation demanding more bandwidth, larger world sizes
- Higher radix switch ASICs required to break the status quo and keep pace with AI accelerators

650 GROUP
MARKET INTELLIGENCE RESEARCH

Agenda Overview

- AI Data Center Networking Trends
- Understanding the Network Wall
- How High-radix Switching Breaks Through the Network Wall
- Use Cases: Scale Up, Out and Across
- Eridu's Mission
- Conclusions
- Live Q&A

Breaking Through the AI Network Wall with High-radix Switching

Cliff Grossner, Ph.D.

Chief Innovation Officer

OCP Foundation

Alan Weckel

Founder, Technology Analyst

650 Group

Promode Nedungadi

Chief Technology Officer

Eridu

Jay Gill

Sr. Director of Product Marketing

Eridu

Thank you for attending

Breaking Through the AI Network Wall with High-radix Switching

OCP Educational Webinar Series

July 28, 2026

OPEN Compute Project®

650 GROUP MARKET INTELLIGENCE RESEARCH

ERIDU

Community-driven hyperscale innovation for all