BROADCOM

ENABLING AI Infrastructure

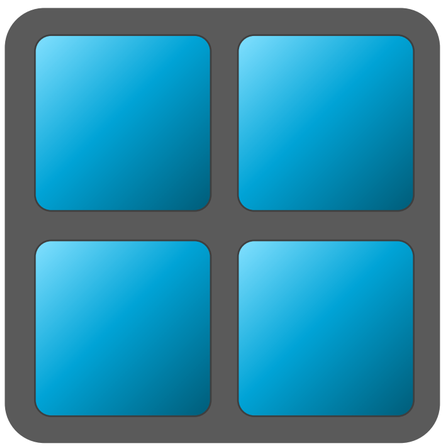
Ethernet for AI @ Scale

Hasan Siraj

Head of Software and AI Infrastructure Products, Broadcom

OPEN // SCALABLE // POWER EFFICIENT

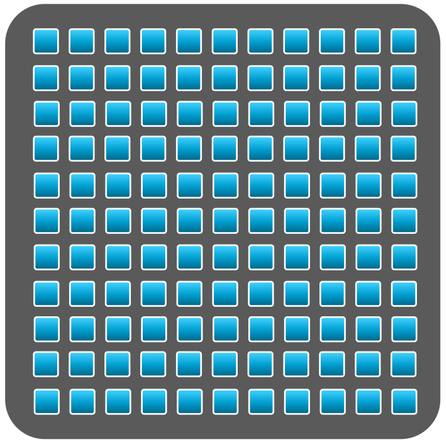
Exponential Acceleration of Compute for AI



CPU

Optimized for
Serial Tasks

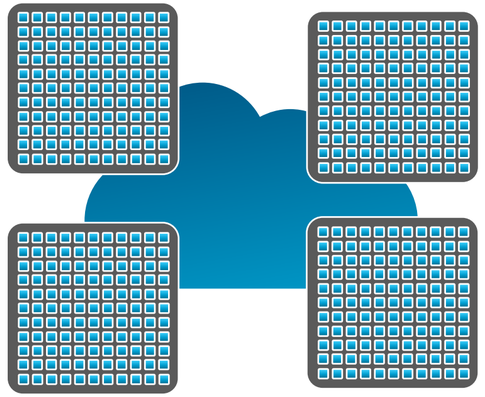
Multiple Cores



GPU

Optimized for
Parallel Tasks

Thousands of Cores



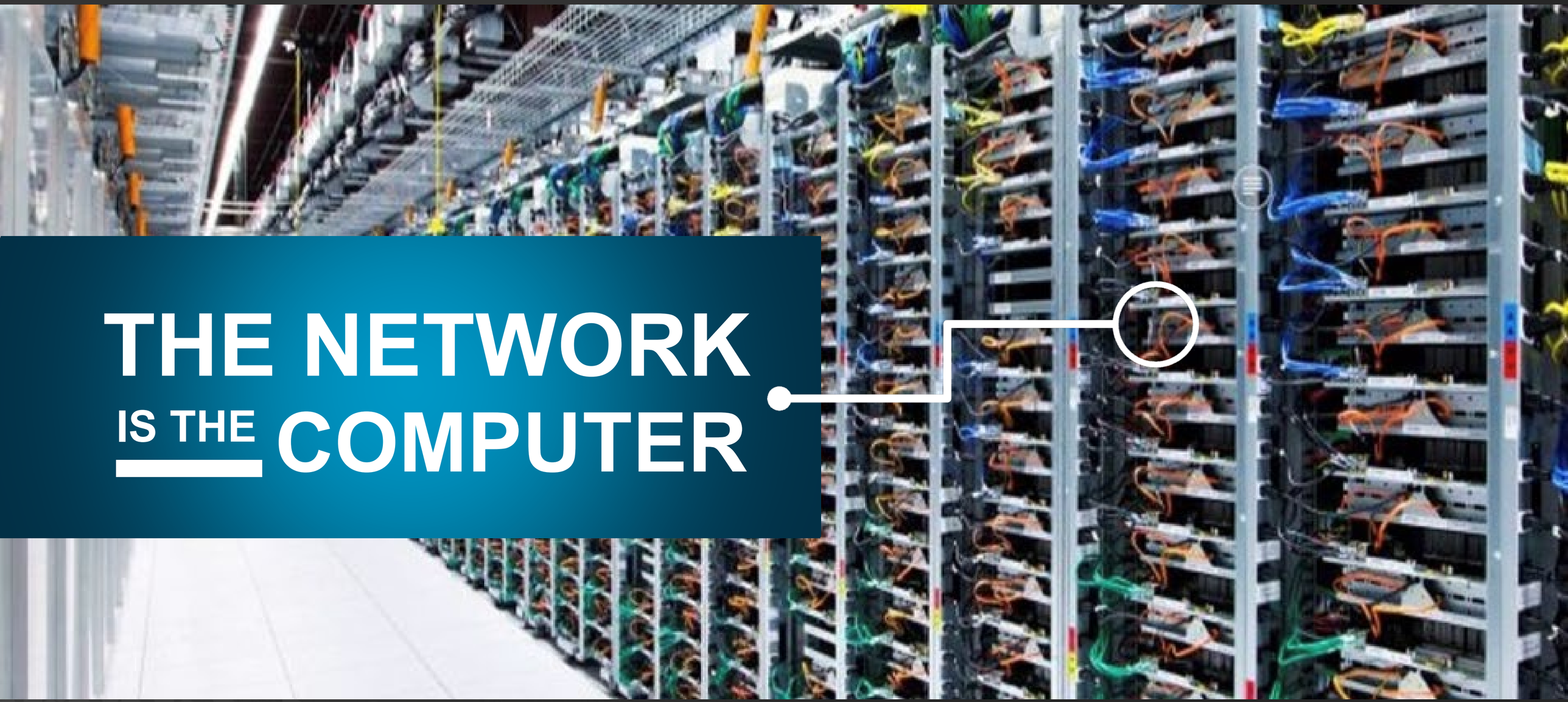
GPU Clusters

Scale-up GPU Network
for AI Workloads

Tens of Thousands of GPUs

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM



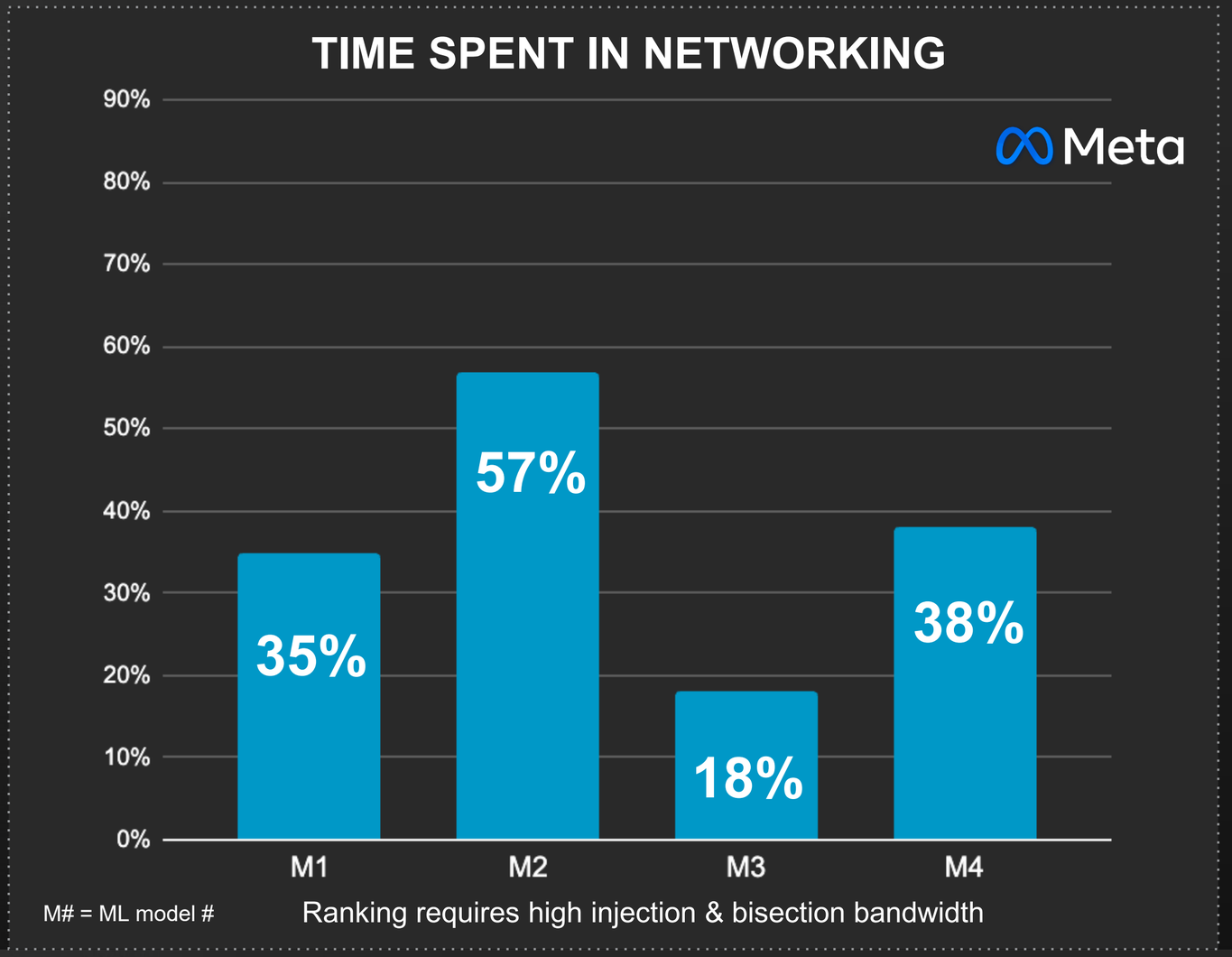
THE NETWORK IS THE COMPUTER

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM®

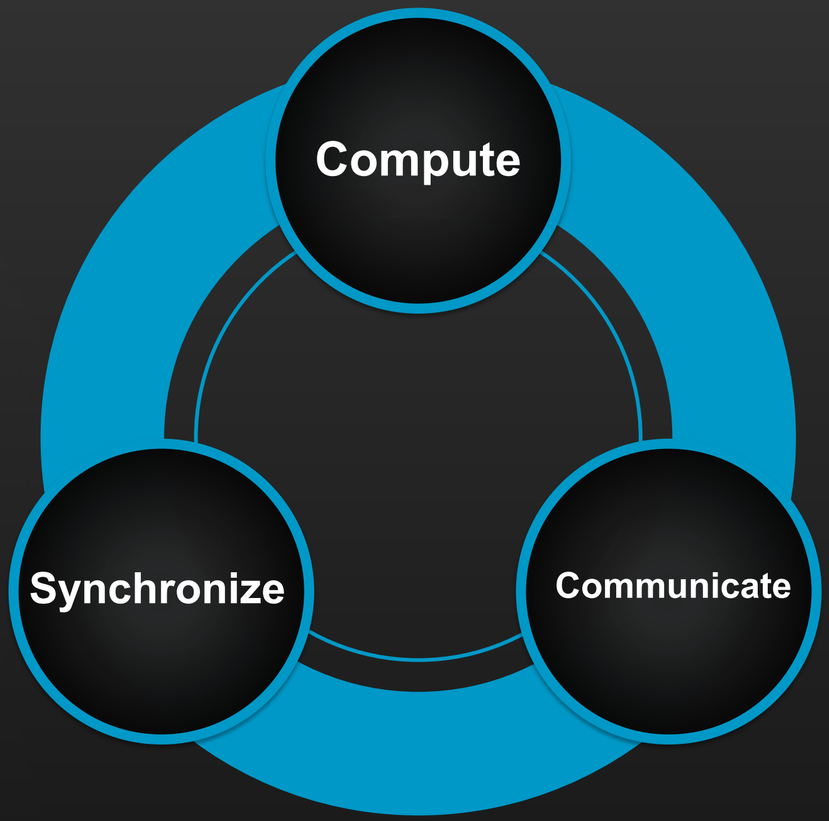
> **Network I/O** is Key for Recommendation Workloads.

OCP keynote by Alexis Bjorlin
at 2022 OCP Global Summit



What Makes AI Networks Unique?

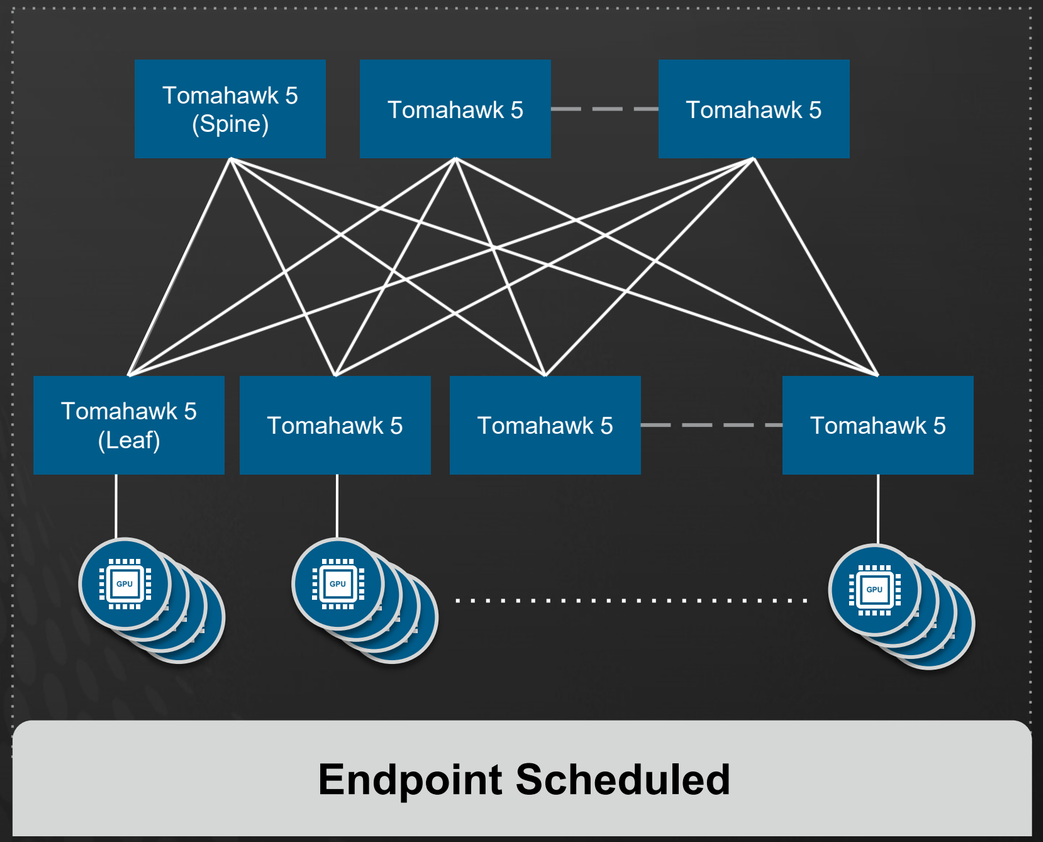
- Very high bandwidth
- RDMA traffic ➔ Bulk data transfers
- Intermittent data surges
- Straggler data significantly impacts job completion time
- Training jobs run over long durations (hours, days)

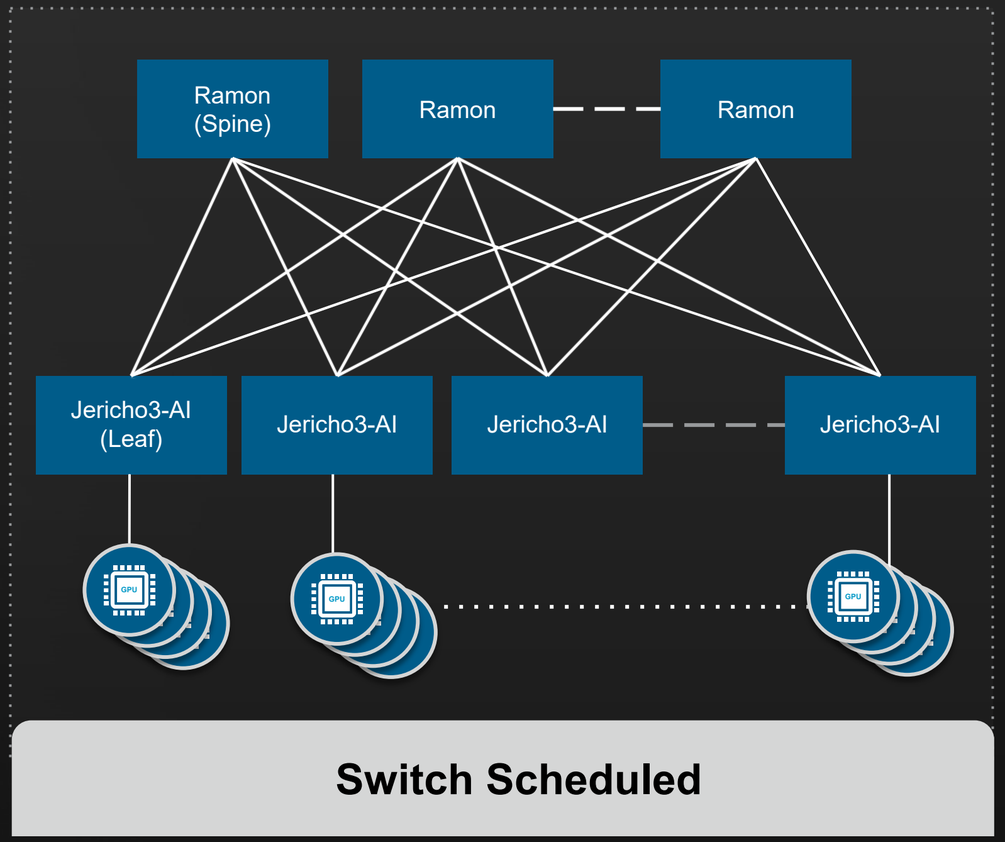


OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Broadcom’s AI Network Solutions



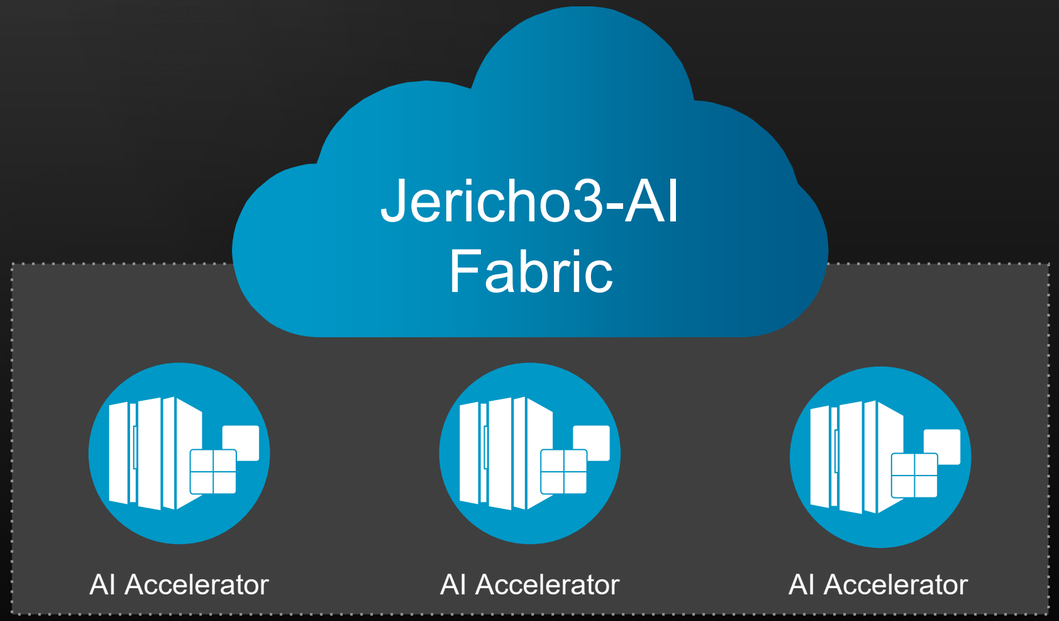


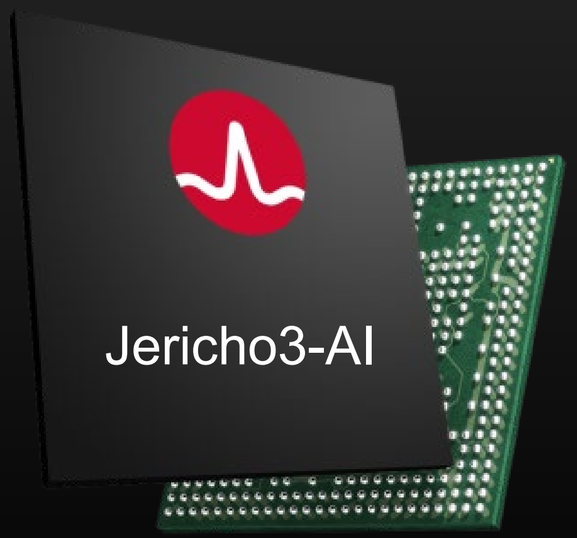
OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Jericho3-AI Fabric: Switch Scheduled Ethernet Network

10% Performance improvement = network more than pays for itself



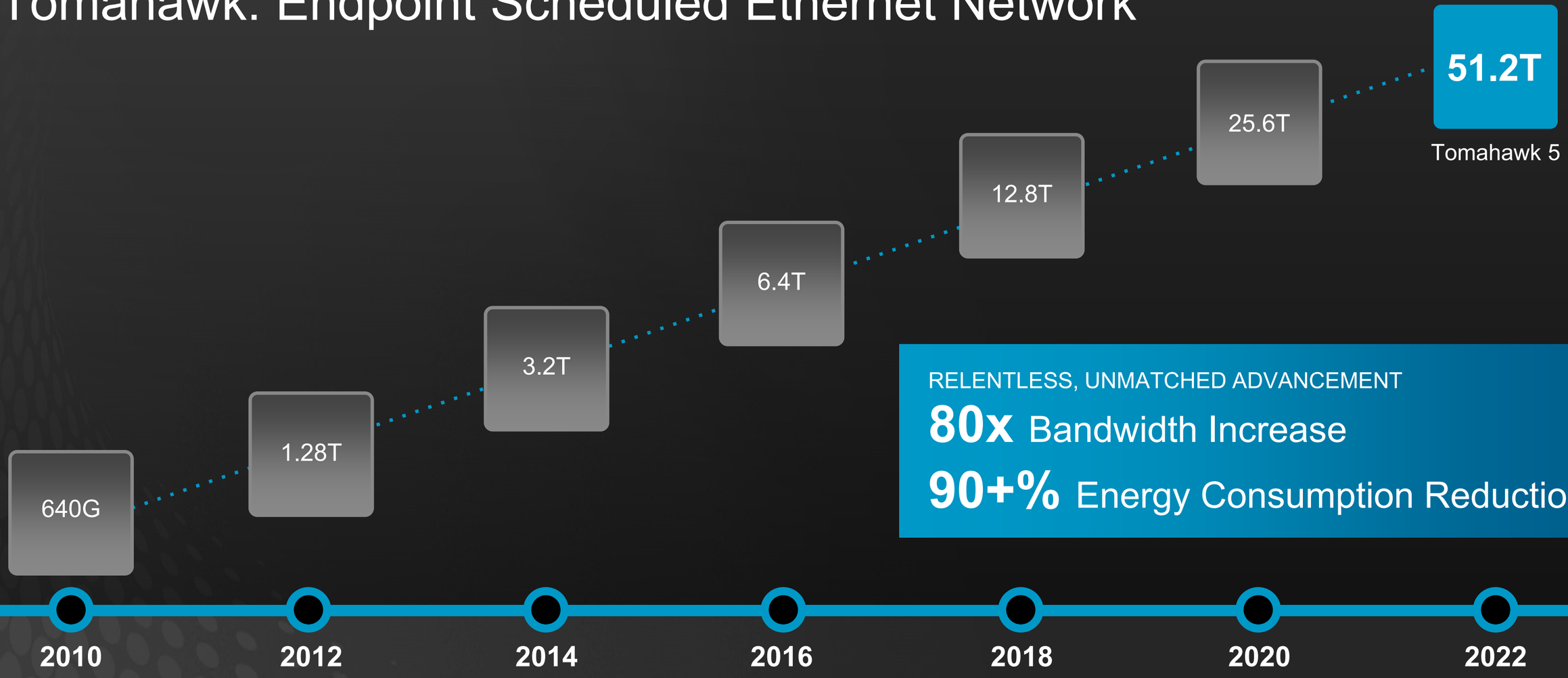


32,000 AI Accelerators** at **800Gbps each
Lowest time spent in networking

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Tomahawk: Endpoint Scheduled Ethernet Network



RELENTLESS, UNMATCHED ADVANCEMENT

80x Bandwidth Increase

90+% Energy Consumption Reduction

* G=Gbps, T=Tbps

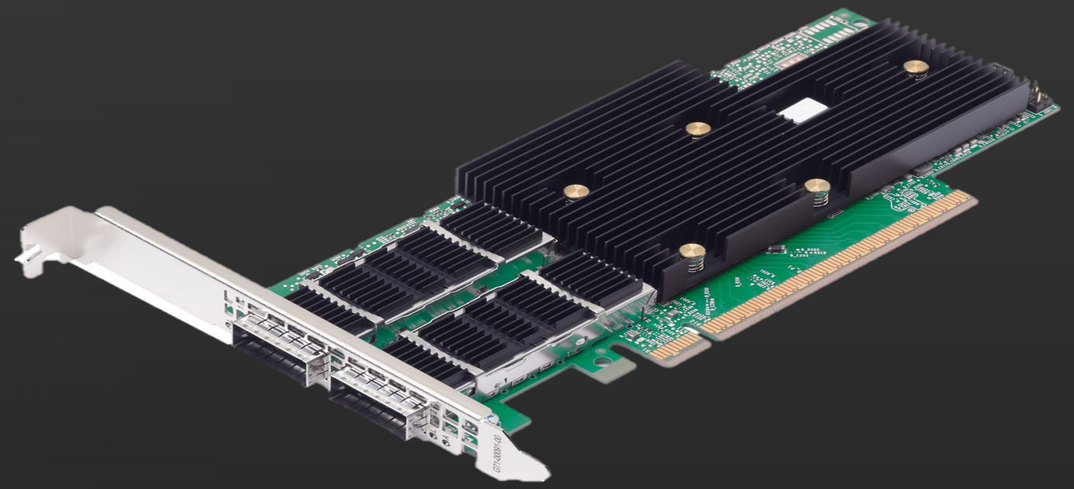
OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Introducing THOR2: AI Optimized NIC

THOR2

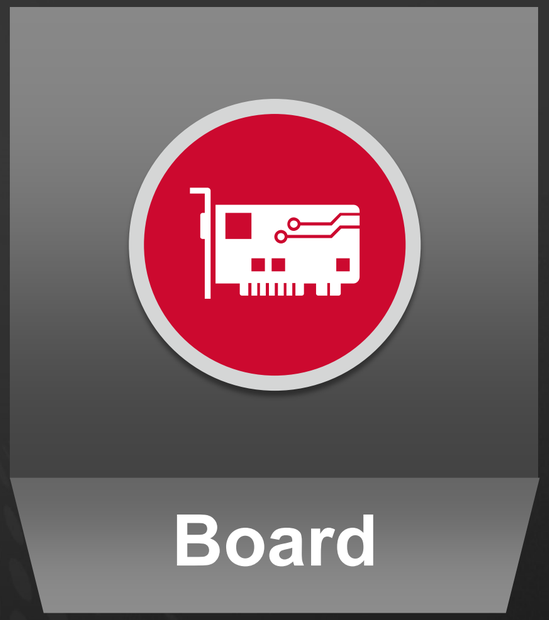
- 400G high-performance NIC
- High-scale RDMA
- Industry’s lowest power
- Longest reach 100G Serdes

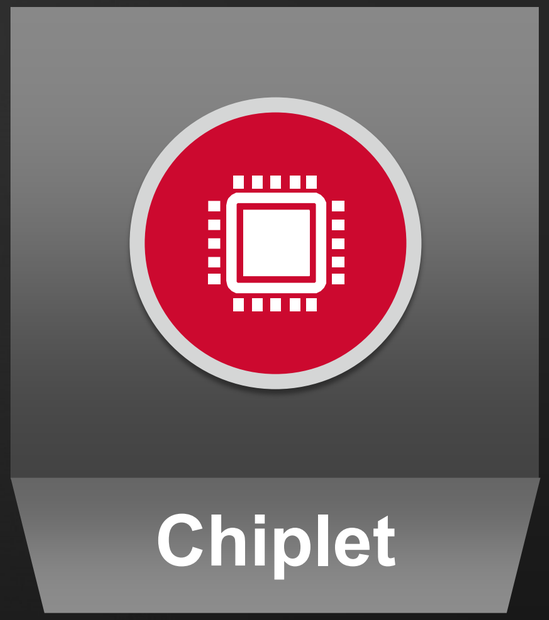


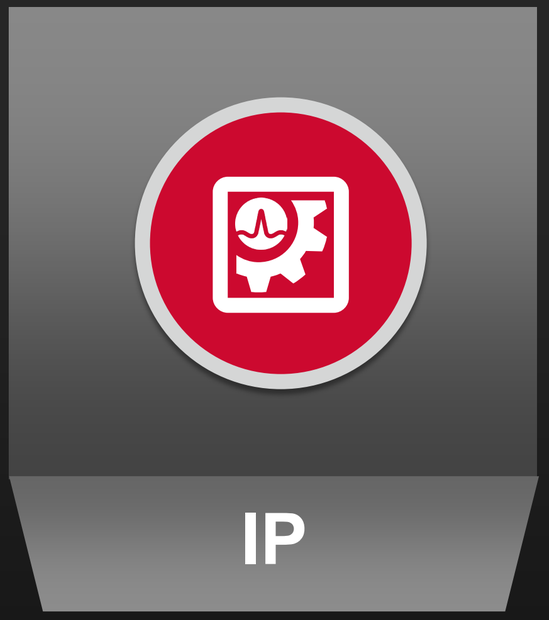
OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

THOR2 Consumption Models



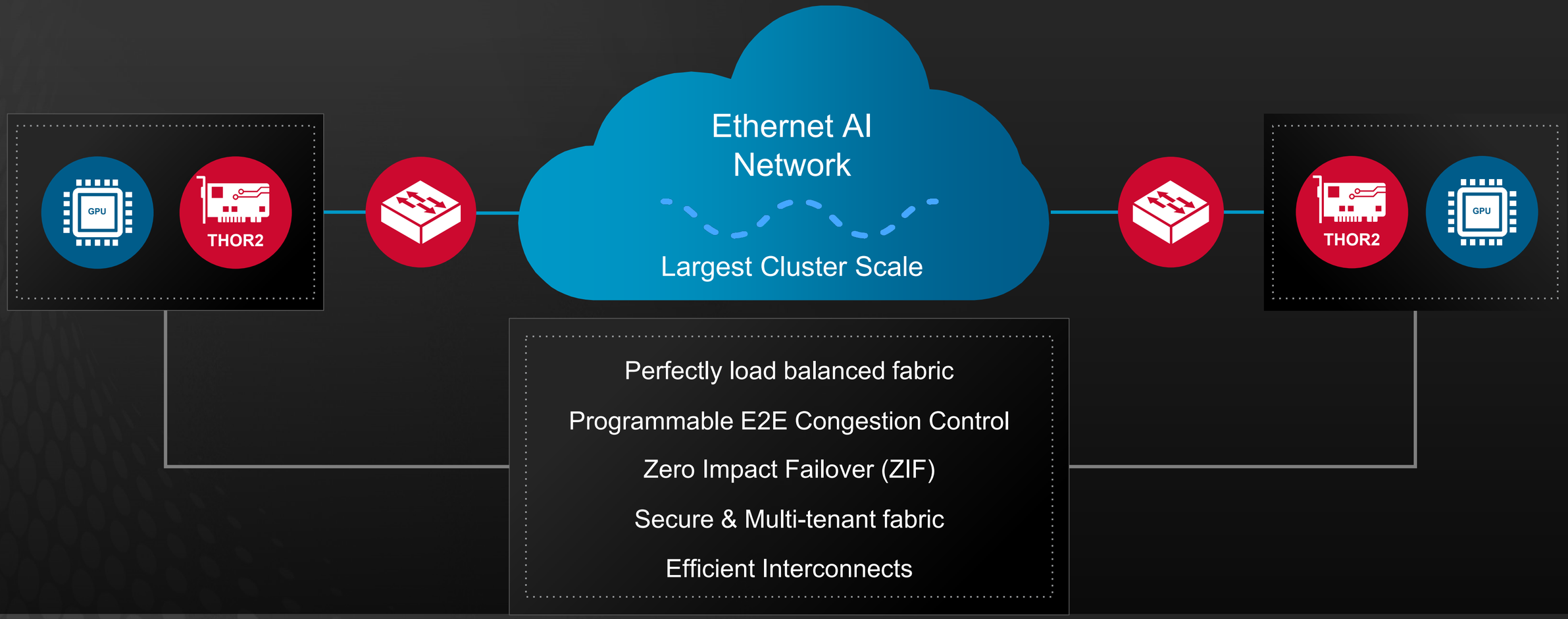




OPEN // SCALABLE // POWER EFFICIENT

BROADCOM®

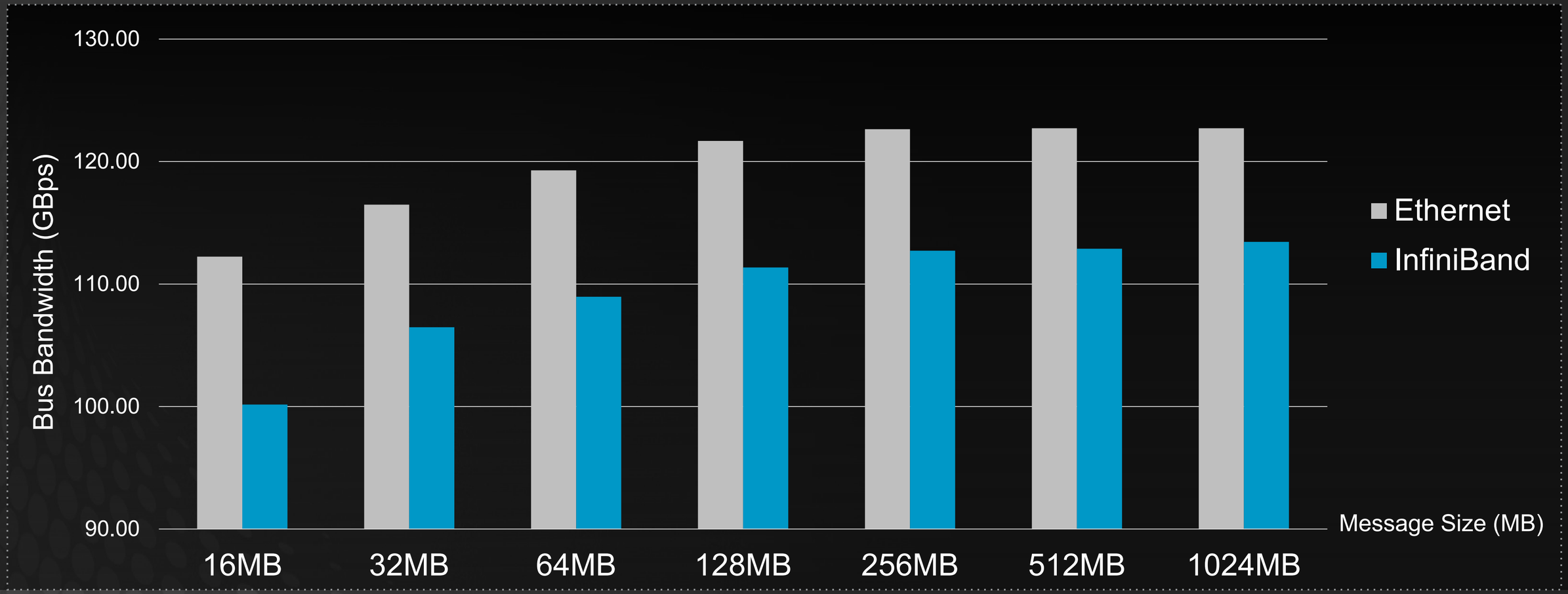
End-to-End High Performance Ethernet AI Network



OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Ethernet Beats InfiniBand: 10+% Improvement in Job Completion Time



OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

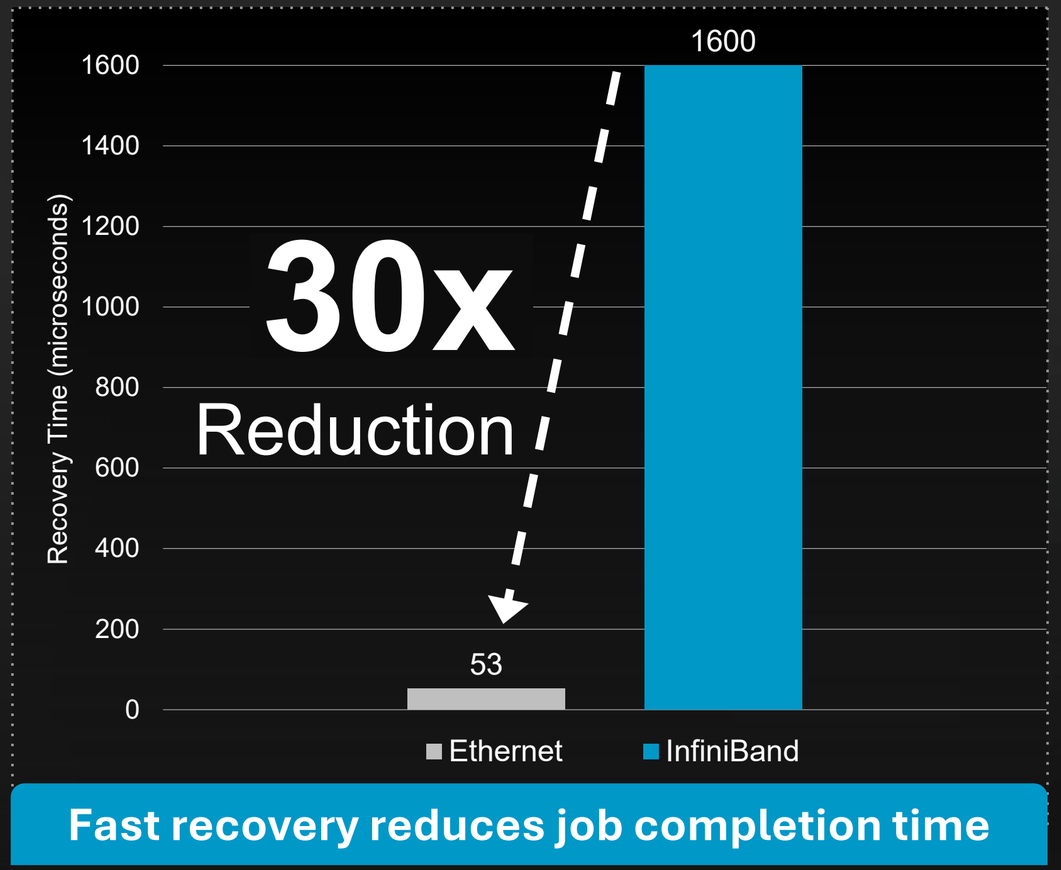
Ethernet Provides 30x Faster Failover than InfiniBand

2%

Optics Annual Failover Rate*

15

Failures per Month**



Fast recovery reduces job completion time

* Typical industry failure rate. ** Assuming 4K node cluster using 9.2K optic modules

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

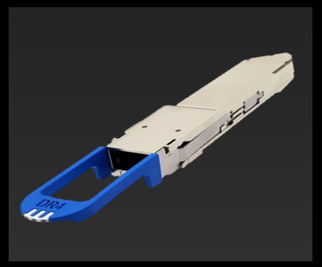
Reducing AI Interconnect Cost and Power

Extended Reach for Copper Cables

4m+ DAC (2x IEEE spec)

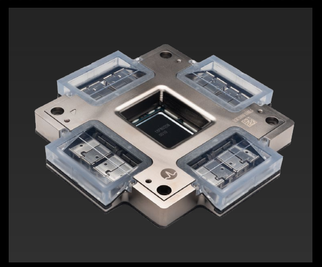
Linear Pluggable Optics

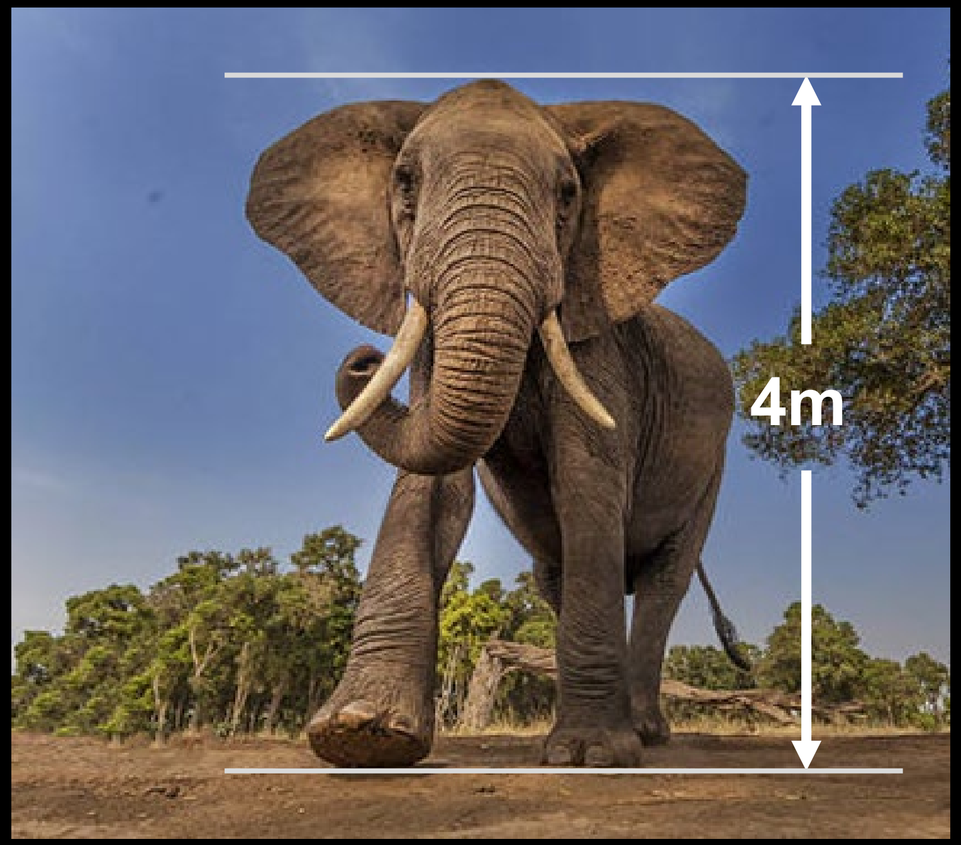
33% Lower Power



Co-Packaged Optics

Lowest Power & Cost

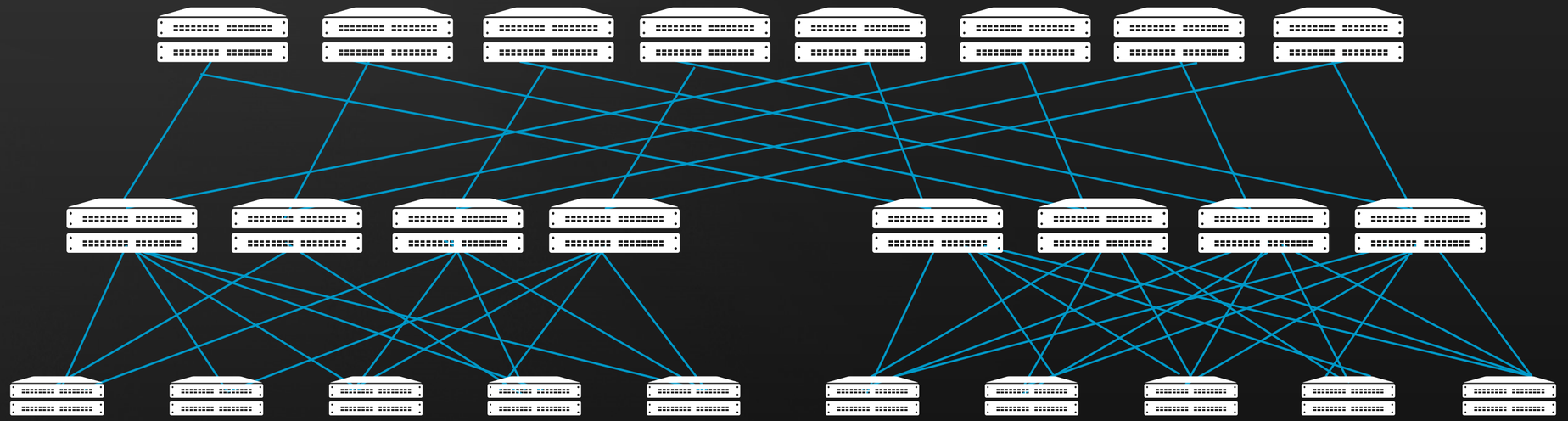




OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Ethernet is the De-facto AI Network



Alibaba Group 阿里巴巴集团 amazon ByteDance Google Meta ORACLE Tencent 腾讯

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Large Ethernet AI Clusters

amazon	60,000+
ORACLE	30,000+
Meta	20,000+
ByteDance	10,000+

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

New AI Models → 100X+ Scale ➔ Distributed Computing

1 Million

AI Accelerators

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Ultra Ethernet: AI at Scale

AMD
ARISTA
BROADCOM
cisco
EVIDEN
an atos business
Hewlett Packard Enterprise
intel
Meta
Microsoft
ORACLE

Incredibly Strong Industry Reception: 55+ companies

BROADCOM

Modernizing RDMA…

Classic RDMA

- In-order packet delivery
- Go-back-n → inefficient
- No multipathing
- DCQCN → hard to tune

Ultra Ethernet

- Out-of-order placement, in-order message completion
- Selective Ack and retransmit
- Packet-level multipathing
- Configuration-free congestion control

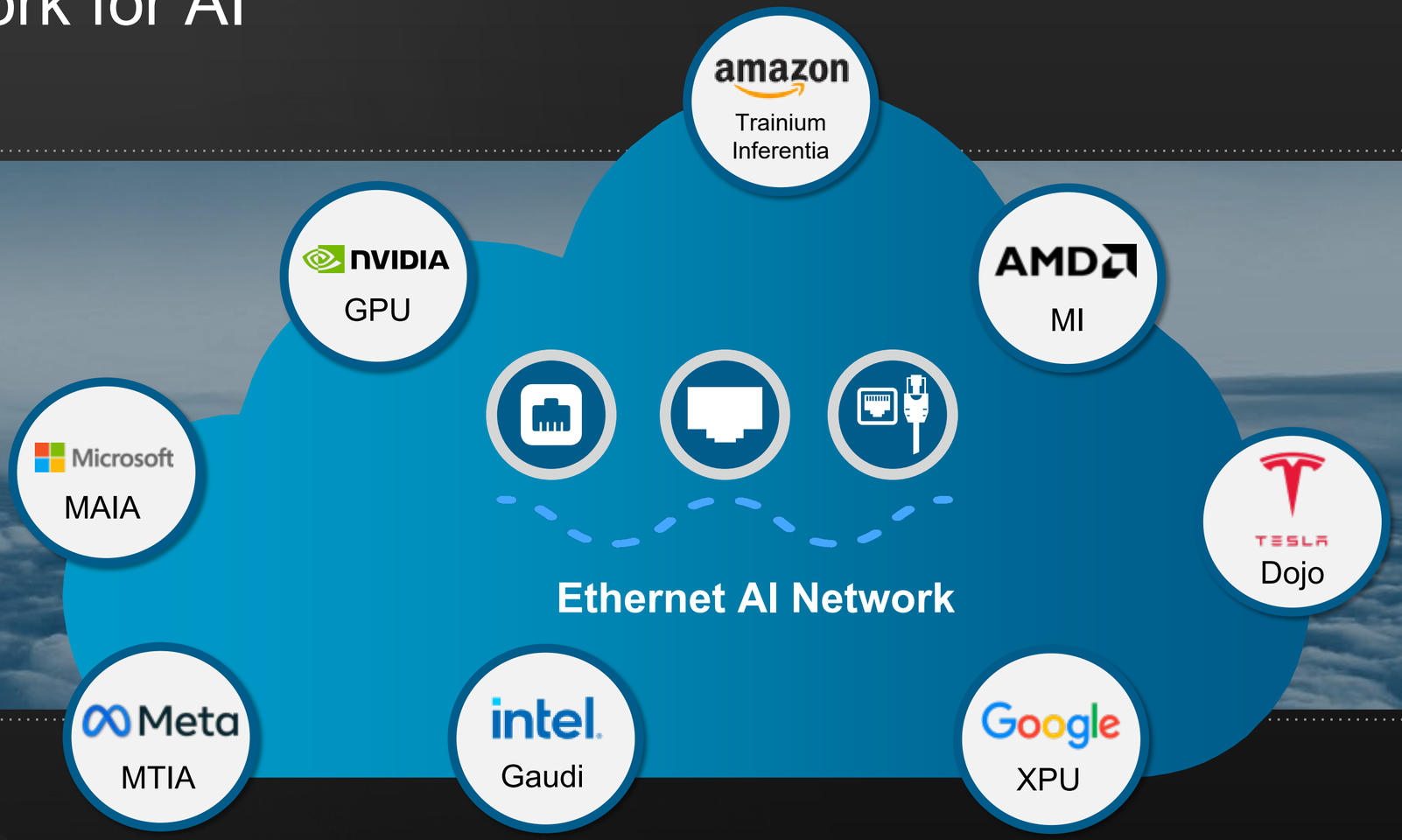
Ethernet for AI Networks

- Pervasively deployed, Open & standards-based technology
- Highest RDMA performance for AI fabrics
- Lowest cost compared to proprietary technologies
- Provides deployment consistency across front-end, back-end, storage and management networks … no technology islands
- Highly available, reliable and easy to use
- Broad silicon, hardware, software, automation, monitoring & debugging tools ecosystem … staffing and operational skills widely available

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Ethernet Network for AI



OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

Comprehensive Ethernet Portfolio and Ecosystem

Network Control & Automation

CHEF, apstra, BE, ANSIBLE, lagopus, NEC, nuagenetworks From Nokia, OPEN DAYLIGHT, puppet, openstack, vmware by Broadcom

Operation System

ARISTA, ARRCUS, CISCO, DELL, Extreme networks, H3C, Hewlett Packard Enterprise, Infoblox NEXT LEVEL NETWORKING, ipinfusion, JUNIPER NETWORKS, Ruijie, SONiC

Hardware Platform

Accton Making Partnership Work, ALPHA Networks, ARISTA, Celestica, CISCO, DELL, DNI A Delta Group Company, Hewlett Packard Enterprise, H3C, JUNIPER NETWORKS, OPEN Compute Project, QCT, Ruijie

Tomahawk | Trident | Jericho | Thor

OPEN // SCALABLE // POWER EFFICIENT

BROADCOM

BROADCOM®

ENABLING AI Infrastructure

OPEN // SCALABLE // POWER EFFICIENT