AI OPEN
DATA CENTER



Large-Scale Deployment of 800Gb/s Active Electrical Cables (AEC) in Meta Data Centers for AI Applications

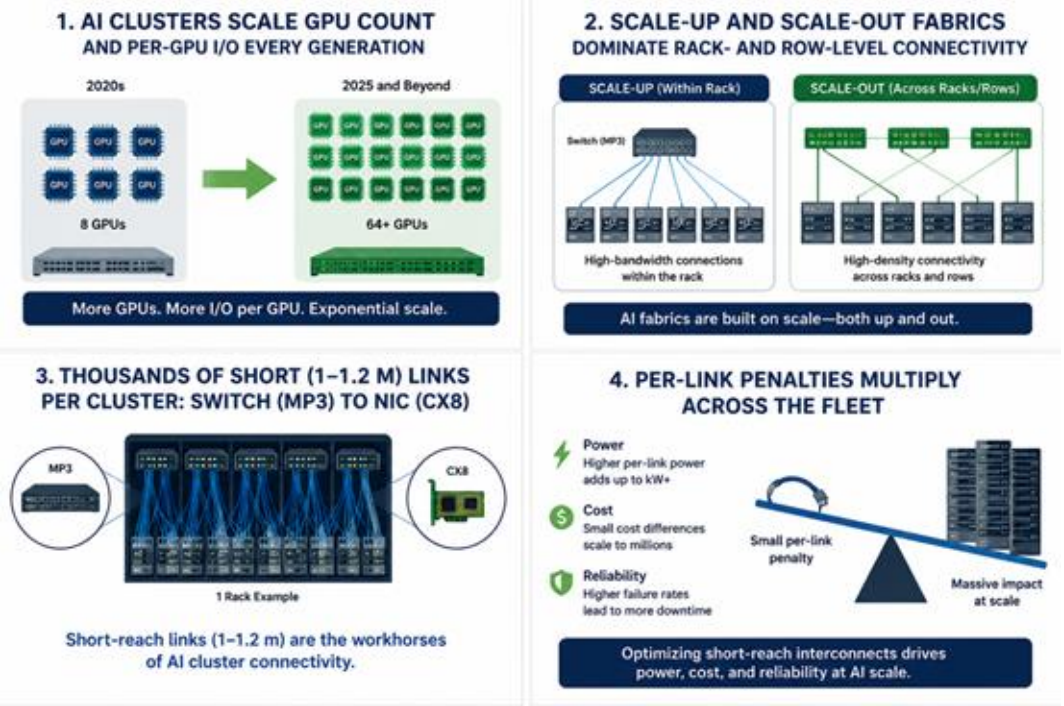
Jingyu (Kevin) Yu — Component Quality Engineer, Meta

Abhijit Chakravarty — Quality Head - Optics & Interconnects, Meta



Why Short-Reach Interconnects Matter for AI

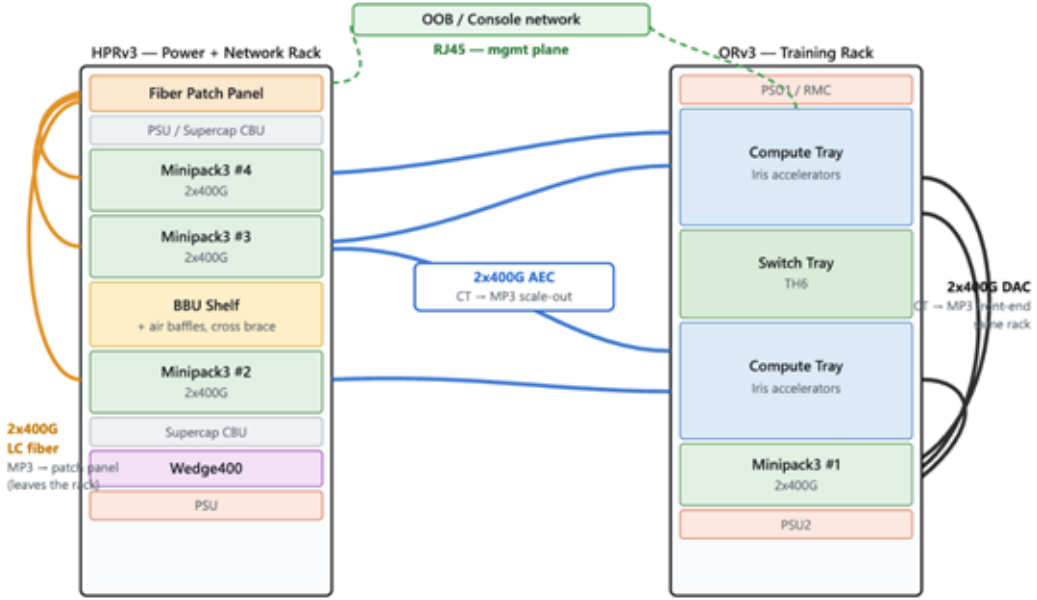
The economics of AI networking are set at the short-reach link.



800G AEC in for X-rack Connectivity

One rack pair, four cable types — reach decides

Compute Tray → Minipack3 links: copper inside the rack, retimed copper across racks, fiber out of the rack.



CABLE	TYPICAL REACH	WHERE IT IS USED HERE
DAC PASSIVE COPPER	≤ ~2 m within a rack	**2x400G** CT → MP3 front-end Plus low-speed 100G, 10G and Y-cable links. Zero power, lowest cost, no FW — default whenever reach allows.
AEC ACTIVE ELECTRICAL	~2–7 m rack to rack	**2x400G** CT → MP3 scale-out The only high-speed link crossing between the two racks. Retimers hold the 400G eye over the longer, tightly-bent run — still copper, so far below optics in power and cost.
Optics LC FIBER	~7 m – 500 m out of the rack	**2x400G** MP3 → fiber patch panel Plus LC to Wedge400. Mandatory past copper reach; highest power per port and highest cost.
RJ45 MANAGEMENT	in-rack + to OOB	**RMC, OOB and console** links Control plane only, never in the data path.

Interconnect Selection: Optics vs DAC vs AEC

	OPTICS	PASSIVE DAC	AEC
	Optical Transceiver + Fiber	**Passive Copper Cable**	**Active Electrical Cable**
Reach	Up to 10 km+ (Longest reach)	Up to ~2–3 m (Short reach only)	Up to < 2 m (Short reach)
Power (per end)	**~12–15 W+**	~1–2 W (Lowest power)	< 12 W (Low power)
Cost	**Highest**	**Lowest**	**Low / Medium**
112G PAM4 Budget (at required length)	**Meets easily**	Cannot meet (Fails budget)	Meets with margin (Re-conditions signal)
Cable Management	More complex (Fiber, patch panels)	Simple (Lightweight, flexible)	Simple (Similar to DAC)
Field Replaceable	**Yes**	**Yes**	**Yes**
Fiber Handling Discipline	Required (Strict fiber handling)	Not required (Copper cable)	Not required (Copper cable)

800G AEC in Meta fleet

Cable Spec

- 800G OSFP AEC (112 Gb/s PAM4 lanes)
- IHS (switch) to RHS (NIC) connectivity

Physical & Power

- 32 AWG twinax, meshed sleeve (~7.5 mm OD)
- Low power design: < 12 W per end

Port & Software

- Flex modes: 2x400G, 1x800G, 8x100G
- CMIS 5.1 compliant management

Compatibility

- 1.0 m to 3.2 m (intra-/adjacent-rack)
- POR for Minipack3 <-> Multiple NICs

How Meta Qualifies AEC: From Lab Validation to Production Scale

Component Qualification	→	Cluster Qualification
Product Qualification	→	Process Qualification
Single Platform	→	Cross-Platform Multi-Supplier
Lab Validation	→	Production Deployment

Five Lessons from deploying Millions of AI Interconnects

01 **Component qualification ≠ Cluster qualification**

02 **Interop failures often originate from firmware rather than SI**

03 **Manufacturing variation becomes visible only at hyperscale**

04 **Thermal cycling exposes failure modes hidden by steady-state stress**

05 **Supplier qualification is a process, not an event**

Lesson #1: Component Qualification ≠ Cluster Qualification

AI-scale deployment introduces failure modes that are invisible at the component level.

Traditional Approach

Component
Reliability
SI / BER
PASS

- Component electrical performance
- Environmental reliability
- Connector durability
- SI / BER compliance

✓ Sufficient for product qualification, but not for hyperscale AI deployment.

PASS in Lab

≠

PASS in AI Cluster

Meta Deployment Approach

Component
Firmware Validation
Cross-NIC Interoperability
System Traffic Validation (RDMA / NCCL / RoCE)
Thermal Cycling
Production Cluster Validation

- Multi-vendor interoperability
- CMIS / firmware compatibility
- Production-representative traffic
- Rack-scale thermal behavior
- Manufacturing consistency

Example

Component Qualification	Cluster Deployment	Root Cause	Solution
✓ BER clean ✓ SI margin pass ✓ Environmental qualified	✗ Link training failed ✗ Firmware mismatch ✗ Cross-NIC interoperability issue	CMIS application mode inconsistency	Standardized firmware configuration across suppliers

Key Takeaway: At hyperscale AI deployment, component qualification is necessary—but not sufficient. System- and cluster-level validation is required before production release.

Lesson #2 Interoperability Lesson: Cross-supplier

01. Failure

Bring-Up Stall

After reset cycling, NICs never completes datapath bring-up.

NIC waits indefinitely, leading to a complete link bring-up failure.

02. Root Cause

Firmware Race Condition

AEC MCU I2C pre-fetch + firmware latched-status "unerase" logic.

Stale reg 0xCA reads 0x1C ("config in progress") instead of 0x11 ("config complete").

03. The Fix

New Firmware Update

Assert "config in progress" for < 1 ms to prevent stale reads.

Ensure a clean and reliable transition to state 0x11.

Key Learning & Test Plan Integration

Surfaced only under high-iteration reset cycling added from a prior NIC-AEC investigation — we must carry these critical field learnings directly into the future test plan.

Lesson #3: Manufacturing variation becomes visible only at hyperscale

A process that works in one factory or one build does not automatically scale to millions of AI interconnects.

1. Traditional Assumption

Factory A
Engineering Build
PASS ✓
Release

Traditional Assumption

- Qualified once
- One production line
- One factory
- Small sample size

2. What Actually Happens at Scale

Factory A	Factory B	Factory C
Different equipment		
Different operators		
Different material lots		
Different process windows		
⚠ Variation appears		

Examples of Variation

- Connector insertion force
- Solder quality
- Glue consistency
- AOI sensitivity
- Test correlation
- Firmware programming

3. Meta Approach: Process Qualification

Factory Qualification
Line Qualification
Correlation Study
SPC Monitoring
MES Traceability
On-site Audit
Mass Production

Meta wants to ensures manufacturing consistency, not just product functionality.

Key Takeaway: A qualified design does not guarantee a qualified factory.
Production readiness must validate manufacturing consistency—not just product functionality.

Statistical Process Control (SPC):

Process Capability Beyond Product Qualification

Why do we need SPC?

- **Limited sampling**
- **Snapshot only**
- **Hidden process drift**
- **Variation discovered after deployment**

Example of SPC (conductor width)



Date	1/10			1/17			1/24			3/2			
Time	9:40			9:40			9:40			9:40			
Recorder	QC58			QC58			QC58			QC58			
Test value X1	0.061	0.079	0.102	0.095	0.070	0.081	0.087	0.071	1.305	0.104	0.078	0.102	0.064	0.076	0.072
Test value X2	0.062	0.085	0.058	0.064	0.100	0.047	0.091	0.060	0.042	0.087	0.062	0.080	0.042	0.107	0.054
Test value X3	0.057	0.055	0.078	0.088	0.065	0.042	0.051	0.059	0.058	1.303	0.052	0.058	0.103	0.052	0.065
Test value X4	0.064	0.046	0.047	0.023	0.080	0.062	0.079	0.071	0.070	0.000	0.075	0.070	0.043	0.045	0.076
Test value X5	0.076	0.105	0.065	0.039	0.077	0.107	0.091	0.071	0.070	0.000	0.064	0.070	0.080	0.065	0.071

What process we are watching:

Process	Monitoring
Hot-bar conductor-to-pad	SPC
Tensile force	SPC
Conductor width	SPC
Solder paste volume	SPC
SMT critical dimensions	SPC
BER trend (SPC-like)	Continuous monitoring

The goal shifted from "Can we build it?" to "Can we build it consistently every day?"

Example: System level EMI Compliance

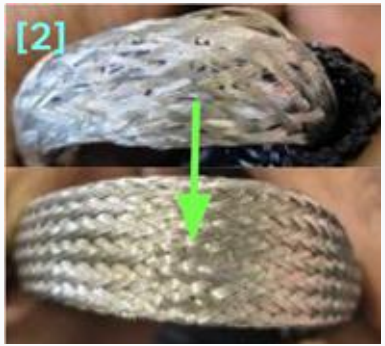
01 System-Level Qualification
System-level radiated-emission (RE) qualified: IHS-to-RHS

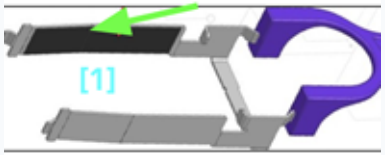
02 Design Margin Goal
10 dB margin design goal, set for high-cable-count AI racks

03 Iterative Improvements
Connector, PCB, and shielding improvements from initial failing states

04 Qualification Results
Sources ranged ~5.7 to ~9.0 dB margin; low-margin source improving via silent rolls







Lesson #4 Thermal cycling exposes failure modes hidden by steady-state stress

AI clusters experience dynamic thermal conditions that cannot be reproduced by constant-temperature testing alone.

Typical Qualification

- High-temperature operating life (HTOL)
- Constant chamber temperature
- Stable workload
- Component-focused reliability

check_circle Useful for long-term aging

cancel Does not represent real AI cluster operation

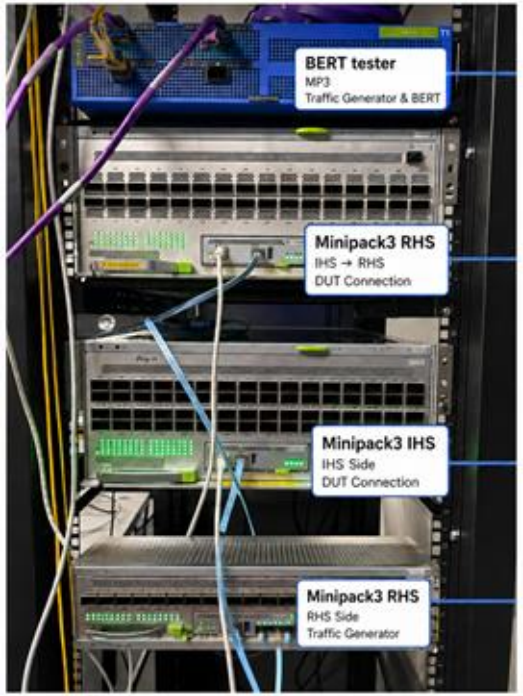
Actual AI Cluster Behavior

Driven by

- GPU workload
- Fan speed control
- Traffic burst
- Power management

Reliability is determined not only by temperature, but by repeated temperature transitions experienced during real AI workloads.

Meta established a deployment methodology



Meta AEC LTR:
In-house Long-term Reliability (LTR)

HW Quality Lab Tests Method: LTR

Sample Size:
- 800G AEC OSFP (IHS to RHS)
- Switch: MP3 with DiagOS, Frame size & Load: RFC2544 (Frame sizes: 1024, 1518)
- Load: 69%, 79%, 89%, 99%

Traffic generator & BERT tester
MP3

Target: Continuous traffic with temp cycling of ~2–3 months (Ramp rate: ~0.7°C/min)

Goal: Monitor link stability & failures under long-term stress

Sample Populations
Supplier A 64, Supplier B 64, Supplier C 32 units

Hardware & Traffic Setup
3x MiniPack3 (DiagOS, 400G), IHS-to-2xRHS; 800G traffic gen (Spirent / Xena Freya)

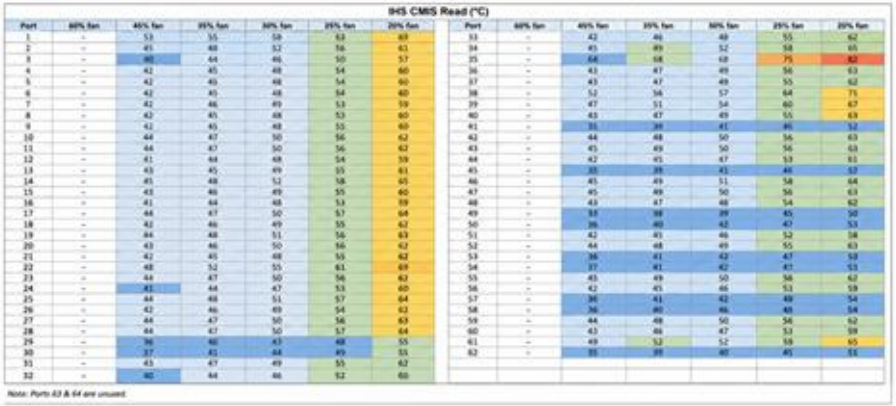
Thermal Cycling Profile
Temperature-cycled, ~0.7 C/min ramp; target case temp ~40 / 50 / 70 C

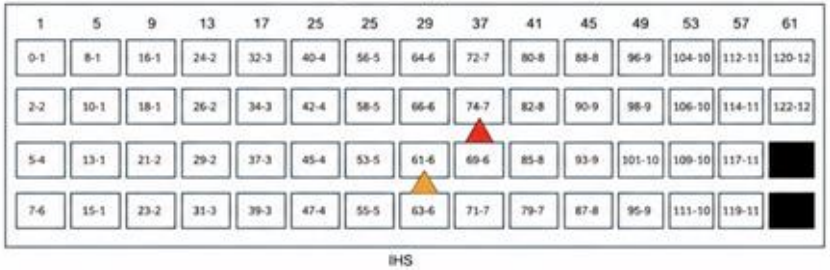
Staged Deployment
Pre-test 36-72 hr 1-week 1-month (30d) 3-month (90d) Temp chamber

LTR with Thermal Cycling

Cage Temp Measured from System					
Condition	Min (°C)	Max (°C)	Mean (°C)	P95 (°C)	P99 (°C)
60% fan	–	–	–	–	–
45% fan	33	64	42.6	≈49	≈52
35% fan	38	68	46.2	≈52	≈56
30% fan	39	68	48.4	≈55	≈57
25% fan	45	75	54.3	≈61	≈64
20% fan	50	82	60.8	≈65–67	≈71

Temp (°C) | ≤ 40 | 41 – 50 | 51 – 60 | 61 – 70 | > 70





IHS port map (subgroup / lane)

- Case temperature controlled via switch fan PWM
- 45% fan -> max 64 C (cool) ; 35% -> 68 C (nominal)
- 25% fan -> max 75 C (hot) ; 20% -> 82 C (overstress, not continuous)
- Daily 24 h cycling profile + periodic overstress pulses
- Runs halted if measured case temp > 70 C in standard operation

→ Overstress pulses expose marginal behavior without continuous over-temperature stress.

Lesson #5

Supplier qualification is a process, not an event

Phase 01

Initial Evaluation

01 Supplier Selection

02 Design Qualification

03 Factory Qualification

Phase 02

Production Launch

04 Line Qualification

05 Pilot Build

06 Mass Production

Phase 03

Ongoing Lifecycle

07 Production Monitoring

08 Corrective Action

09 Continuous Improvement

buildORT & SPC Foundation

"Qualification is not the end of supplier management—it is the beginning of production readiness."

Call to Action

01

Deployment Qualification

Move beyond component qualification toward deployment qualification.

02

Representative Workloads

Validate interoperability, reliability, and manufacturing consistency under production-representative AI workloads.

03

Continuous Lifecycle

Treat supplier qualification as a continuous lifecycle—not a one-time milestone.

04

OCP Collaboration

Work together through OCP to establish common qualification methodologies for next-generation AI interconnects.

Thank You!

