SERVER: COMPOSABLE MEMORY SYSTEMS (CMS)



CXL in Hyperscale Environment

Fleet-Scale Challenges, Failure Modes & the Path to Better Debuggability

Markandeya Mudigonda *Production Systems Engineer, Meta*

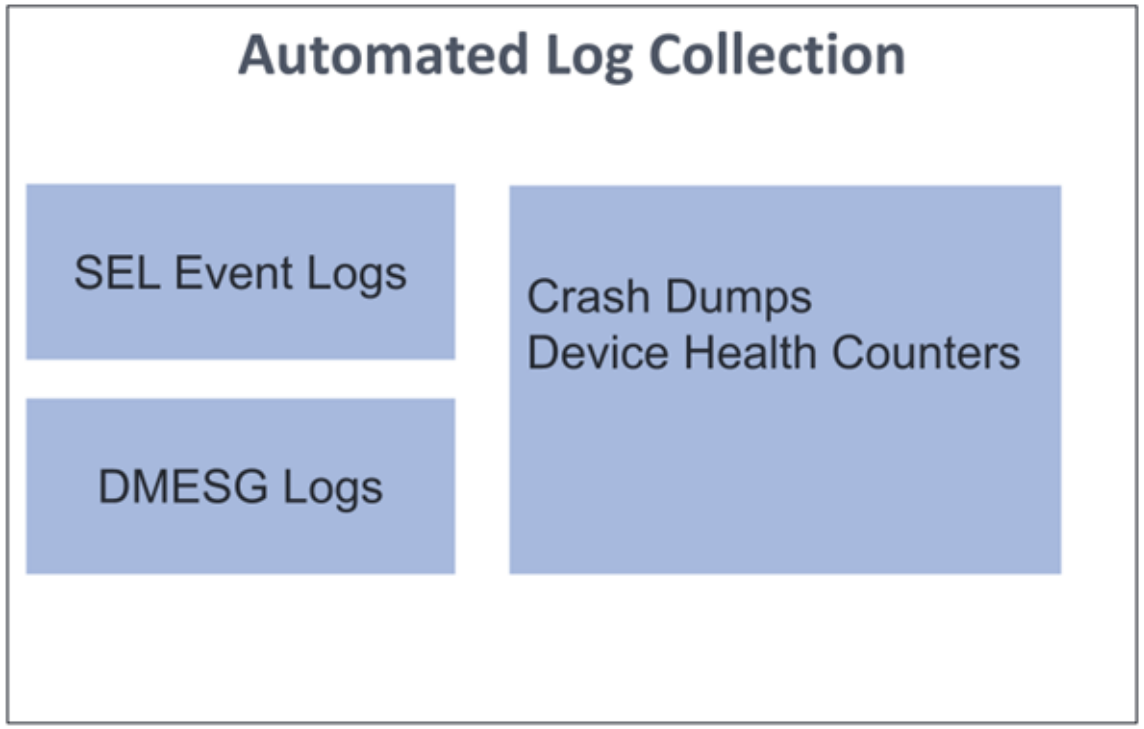
Vineet Parekh *Production Systems Engineer, Meta*



Failure Detection in Fleet



CXL At-Scale Telemetry Collection



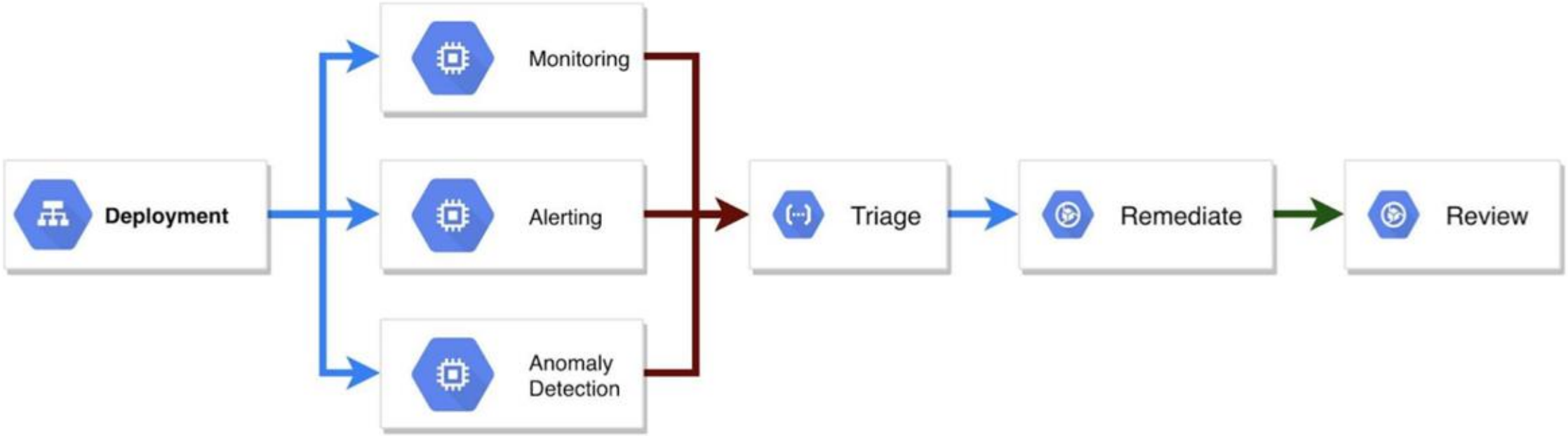
SEL event logs: BMC system event logs, hardware events with timestamps

DMESG logs: kernel messages: AER errors, MCEs, CXL driver events

Crash dumps: captured when a fatal error takes down a process or the system

Device health counters: device telemetry: correctable error counts, temperature, media health

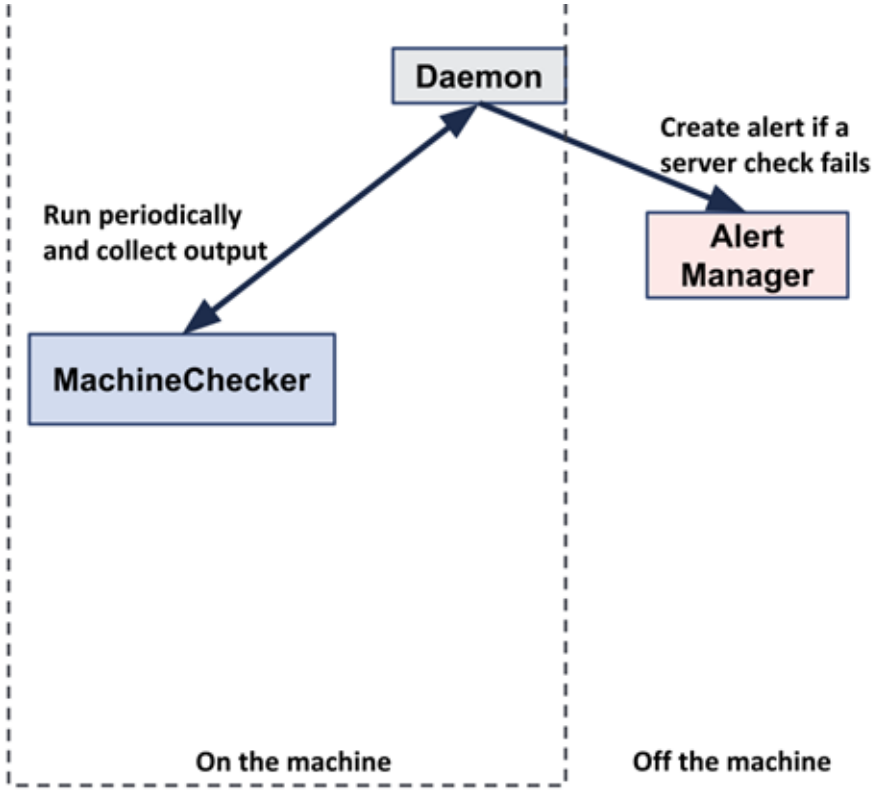
Fleet Failure Debug Overview



Hardware Remediation @ Scale

Failure Detection – MachineChecker

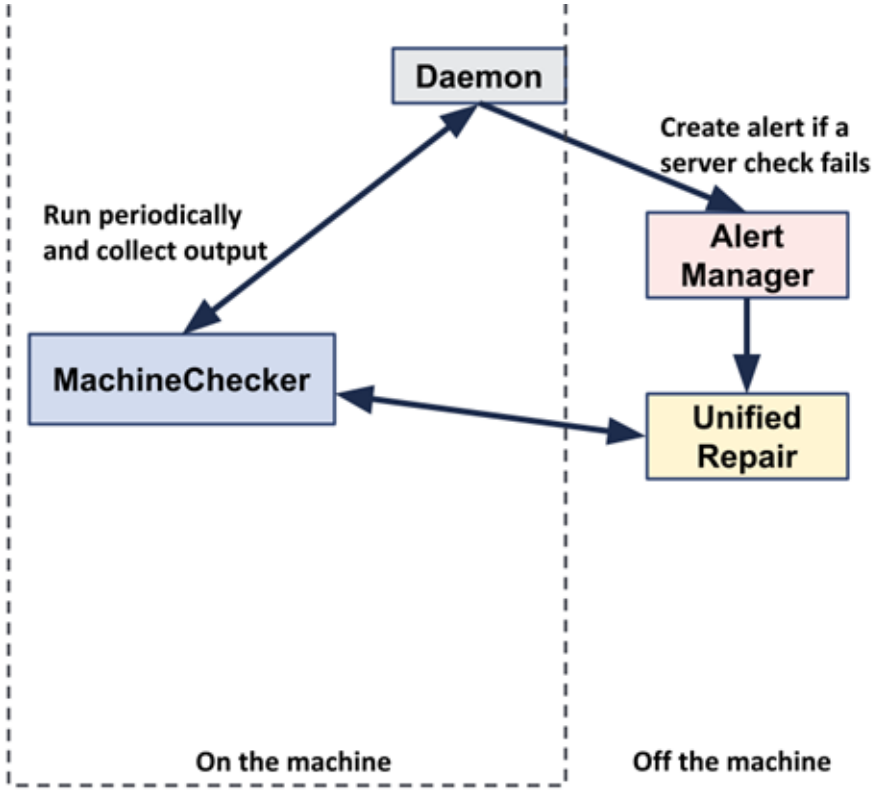
- Runs hardware checks periodically
- Host ping, memory, CPU, NIC, dmesg, S.M.A.R.T., power supply, SEL, etc.



Hardware Remediation @ Scale

Failure Detection – MachineChecker
Failure Digestion – Unified Repair

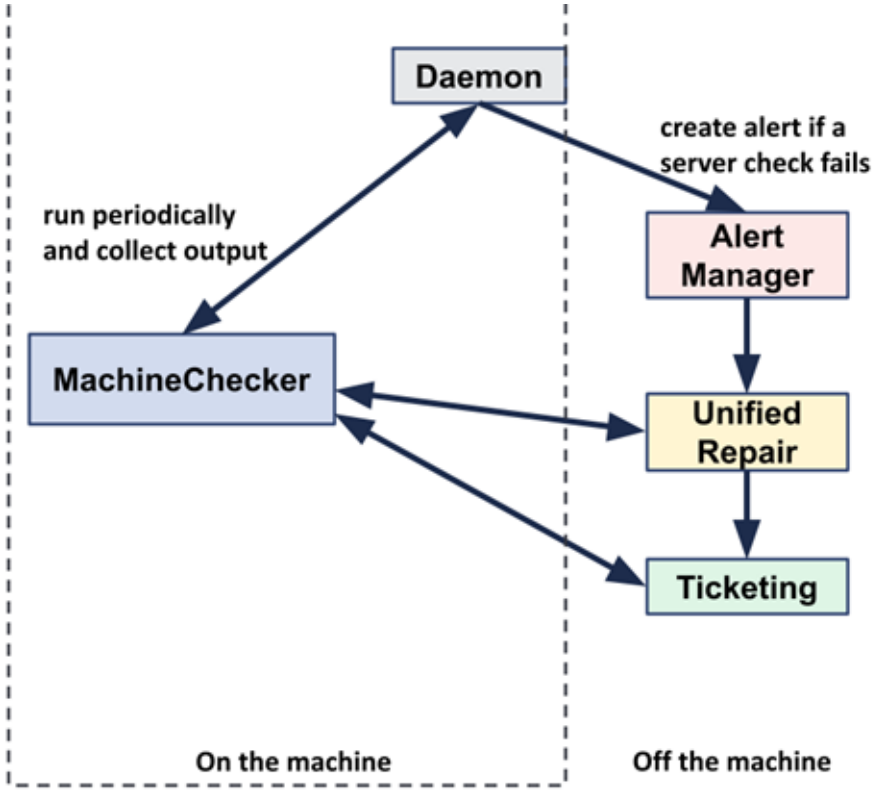
- Auto Remediation
- Picks up hardware failures, processes logged information, and executes custom-made remediation accordingly



Hardware Remediation @ Scale

Failure Detection – MachineChecker
Failure Digestion – Unified Repair
Low-Level Software Fix – Ticketing

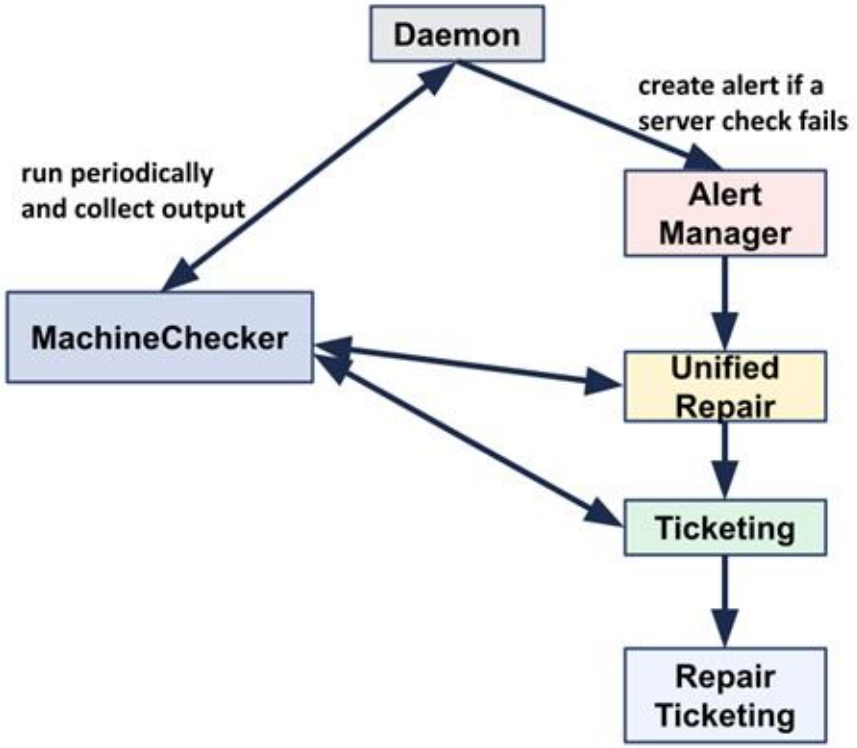
- Handles low-level software fixes such as firmware update and reimaging



Hardware Remediation @ Scale

Failure Detection – MachineChecker
Failure Digestion – Unified Repair
Low-Level Software Fix – Ticketing
Manual Fix – Repair Ticketing

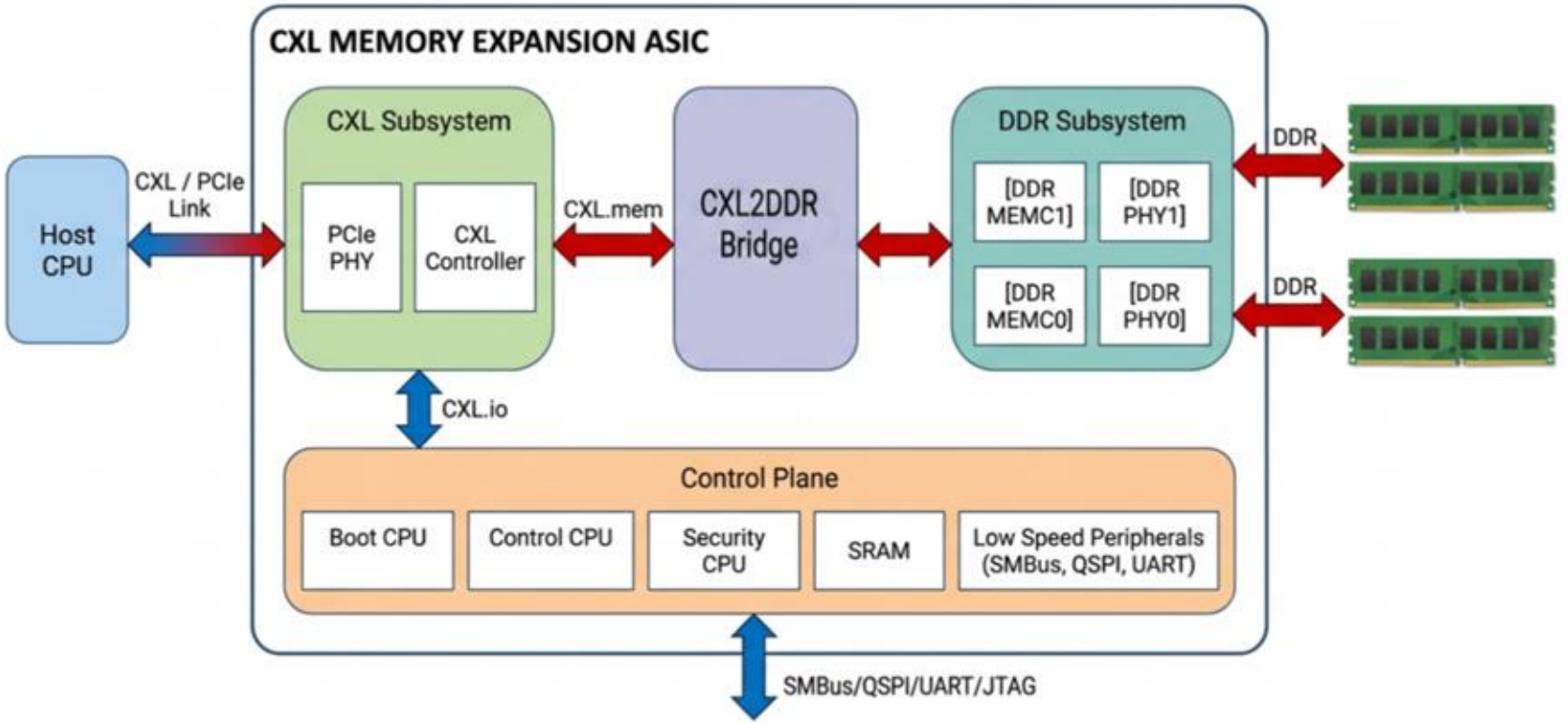
- Creates repair tickets for DC technicians to swap CXL ASIC / DIMMs
- Provides detailed logs throughout the auto-remediation
- Logs repair actions for further analysis



CXL Failures in Fleet



CXL Memory Expander Failure Domains



ASIC Enumeration Issues

ROOT CAUSES

- **CXL FW self-test failure**
- **SPI Flash Failure**
- **Link Training Failure**

ASIC Init failure

CXL Link Never Trains

DETECTION

- **PCIe scan (In-band)**
- **BMC polls ASIC heartbeat (OOB)**

Run time failures

ROOT CAUSES

PCIe Errors
Surprise link down, completion timeout

CXL Viral Error

Memory CE Storm

ASIC Unresponsive

Host Gets MCE

DETECTION

Fatal error → MCE
(In-band)

BMC detects
heartbeat loss (OOB)

CXL Memory Failures

TYPES OF MEMORY ERRORS

Correctable ECC Error (two Symbol Error)	Uncorrectable ECC Error (Multi-bit Error)	DDR Training Failures (Margin Issues)

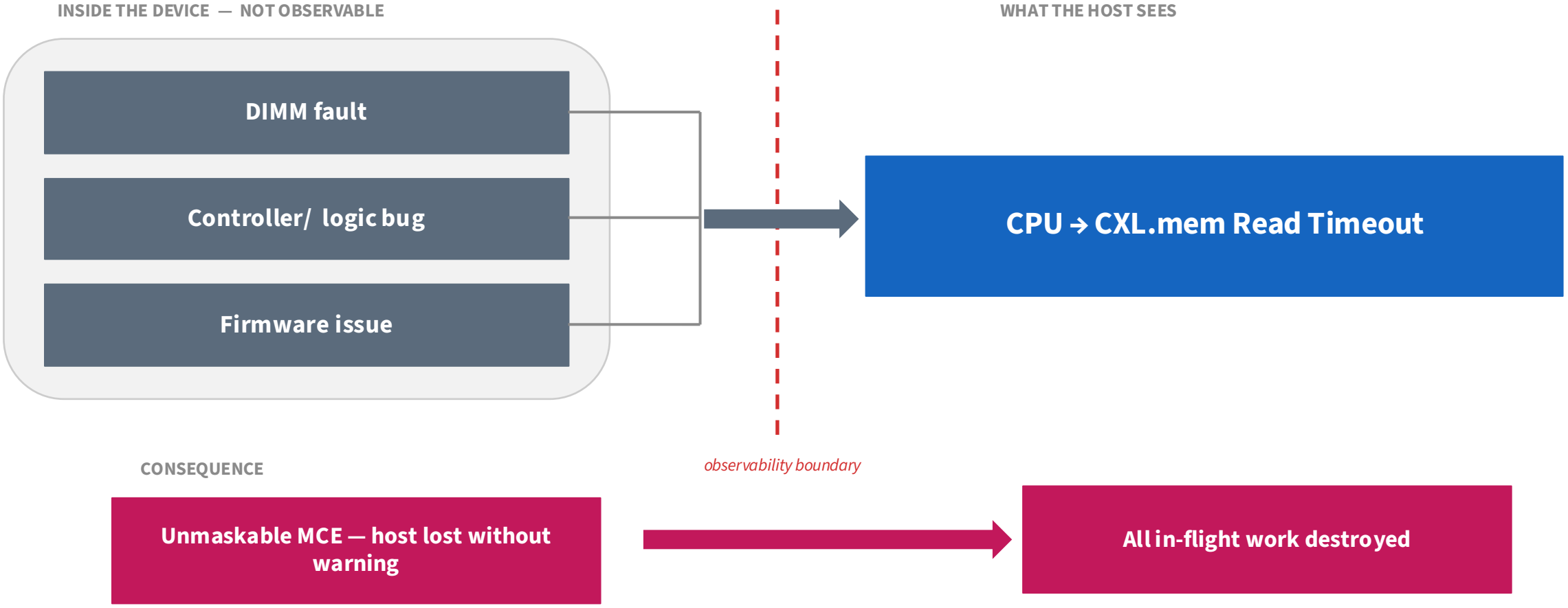
ROOT CAUSES

- **Defective DIMM**
- **SI/Power/Thermal issues**

DETECTION

System Event Log (OOB)

One Symptom. Several Possible Causes.



Why CXL: Workload Impact at Scale

Category	Primary Bottleneck	Key Improvement
Caching	Memory Capacity	Throughput +25%, Retention 5–10x
Distributed Caching	Network BW	Avg query latency −29%
Data Warehouse	Alloc Mem / CPU	Executors/server +25%
DevInfra (Devmachines)	Memory Capacity	VMs/server +33%, Fleet −15%
ML Parameter Server	Memory Capacity	Servers −25%, Throughput +12%

Call to Action

Problem to solve

- A CXL memory expander device can be fully spec-compliant and still be non debuggable at fleet scale

What should be addressed?

- Define and standardize requirements for CXL device debuggability and telemetry at Fleet Scale
- Keen to collaborate with the ecosystem and CXL Consortium

Thank You!

Click here to review OCP Trademark Guidelines

Please use the appropriate Global Track name and icon.

AI CLUSTERS



AI COMPUTING



AI OPEN DATA CENTER



AI OPEN DATA CENTER: Power Distribution



COOLING ENVIRONMENTS



COOLING: COLD PLATE



COOLING: IMMERSION



MARKET IMPACT & ADOPTION



NETWORKING



OPEN PLATFORM FIRMWARE (OPF)



Please use the appropriate Global Track name and icon.

RACK & POWER



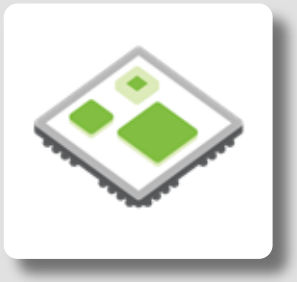
SECURITY



SERVER: COMPOSABLE MEMORY SYSTEMS (CMS)



SERVER: OPEN CHIPLET ECONOMY



STORAGE



SUSTAINABILITY



SYSTEMS MANAGEMENT



TEST & VALIDATION



TIME APPLIANCES (TAP)

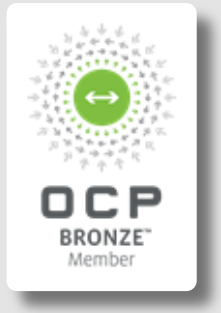


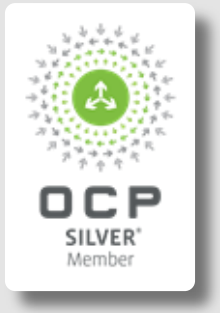
OCP SPECIAL FOCUS: PHOTONICS



Please use one of these membership logos to designate your company’s membership level.











Please use one of these logos if your facility is an OCP Ready™ facility or you're hosting an Experience Center.





Please use this logo if your product has been approved through the OCP S.A.F.E.™ program.



Please use one of these logos if your product has been recognized as OCP Accepted ™



Please use this logo if you are hosting an area in the Innovation Village.



Please use this logo if you are an OCP Solution Provider.



Please use the appropriate icon representing the Regional Project Group.











