Flash Memory Summit

Addressing Memory Bottleneck with CXL™ Type 3 Memory Controllers

Presenter: Ranjit Gupte, Sr. Technical Staff Engineer

Microchip Technology Inc.

MICROCHIP

Agenda

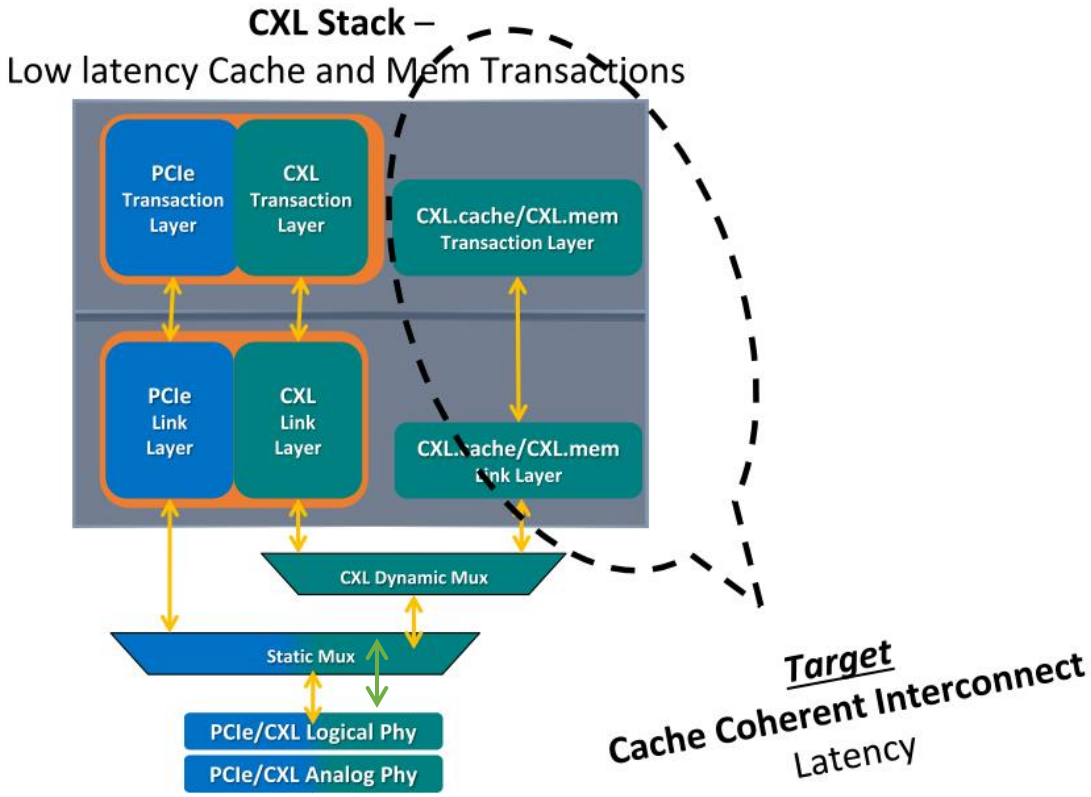
- Introduction to CXL™
- Define RAS
- Supported RAS Features
- CXL Error Handling
- Data Poison and Viral
- CXL Error Injection
- Conclusion

Introduction to CXL™

- CXL is an emerging open Industry standard based on PCIe® 5.0 infrastructure
  - High bandwidth, low latency interconnect for connectivity between host processors and accelerators/memory devices/smart NIC
  - Targets high-performance computational workloads like AI, Machine Learning, Comms, etc.
  - CXL 1.0 --> CXL 1.1 --> CXL 2.0 --> CXL 3.0
- Dynamic multi-protocol capability that enables new support models
  - **CXL.io** – Mandatory PCIe based protocol, initialization, discovery, register access, interrupts, I/O virtualization, DMA
  - **CXL.mem** – Protocol to support memory semantics, optimized for latency
  - **CXL.cache** – Protocol to support caching semantics, optimized for latency
- CXL runs on PCIe 5.0 PHY.Primary Data Rate: 32 GT/s (128b/130b)
- Plug and Play: Either PCIe or CXL card can be plugged in
  - Link negotiation for PCIe or CXL occurs during link training

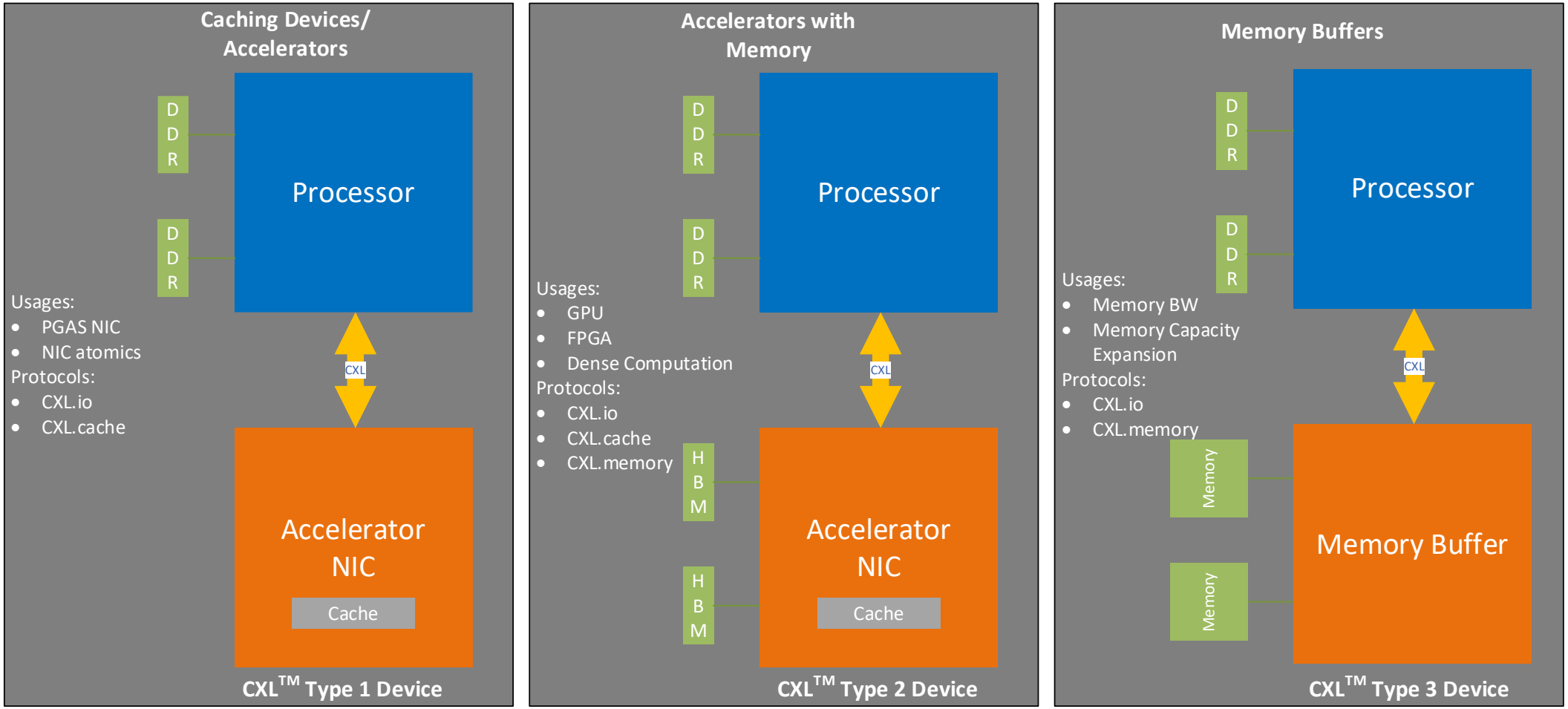
CXL™ Stack – Designed for Low Latency

- CXL IO transactions pass through a stack that is largely identical to PCIe stack
- CXL cache and memory stacks are optimized for latency, use separate transaction and link layers from CXL IO
- Based on the QoS mechanism, static ARB/Mux decides which protocol to schedule



CXL™ Device Types

- CXL consortium has defined three classes of devices to support various CXL based applications



What is RAS?

- Reliability, Availability and Serviceability (RAS) are major considerations in data center server designs
- Reliability: Provide reliable service by detecting and correcting faults early
- Availability: Provide service without interruption
- Serviceability: Diagnose and repair faults that may have caused errors
- The CXL™ standard defines multiple RAS features for emerging end-user requirements in the CXL ecosystem

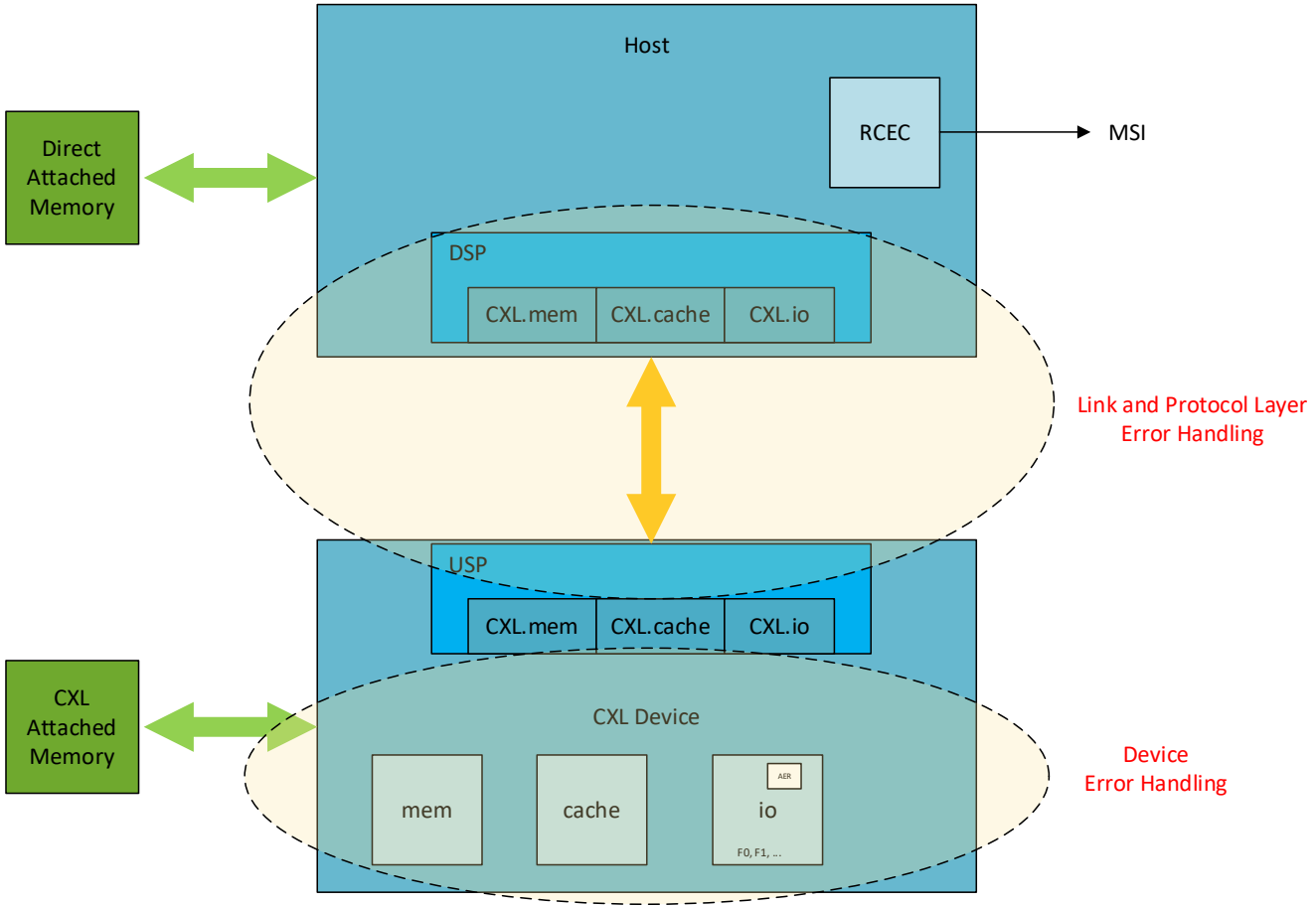
MICROCHIP

CXL™ RAS Features

Feature	CXL.io	CXL.cache and CXL.mem
Link CRC and Retry	Mandatory	Mandatory
Link Retraining and Recovery	Mandatory	Mandatory
eDPC	Optional	CXL.mem errors may be signaled via ERR_FATAL or ERR_NONFATAL and may trigger eDPC.
ECRC	Optional	N/A
Hot-Plug	Supported in CXL 2.0	Supported in CXL 2.0
Data Poisoning	Mandatory	Mandatory
Viral	N/A	Mandatory

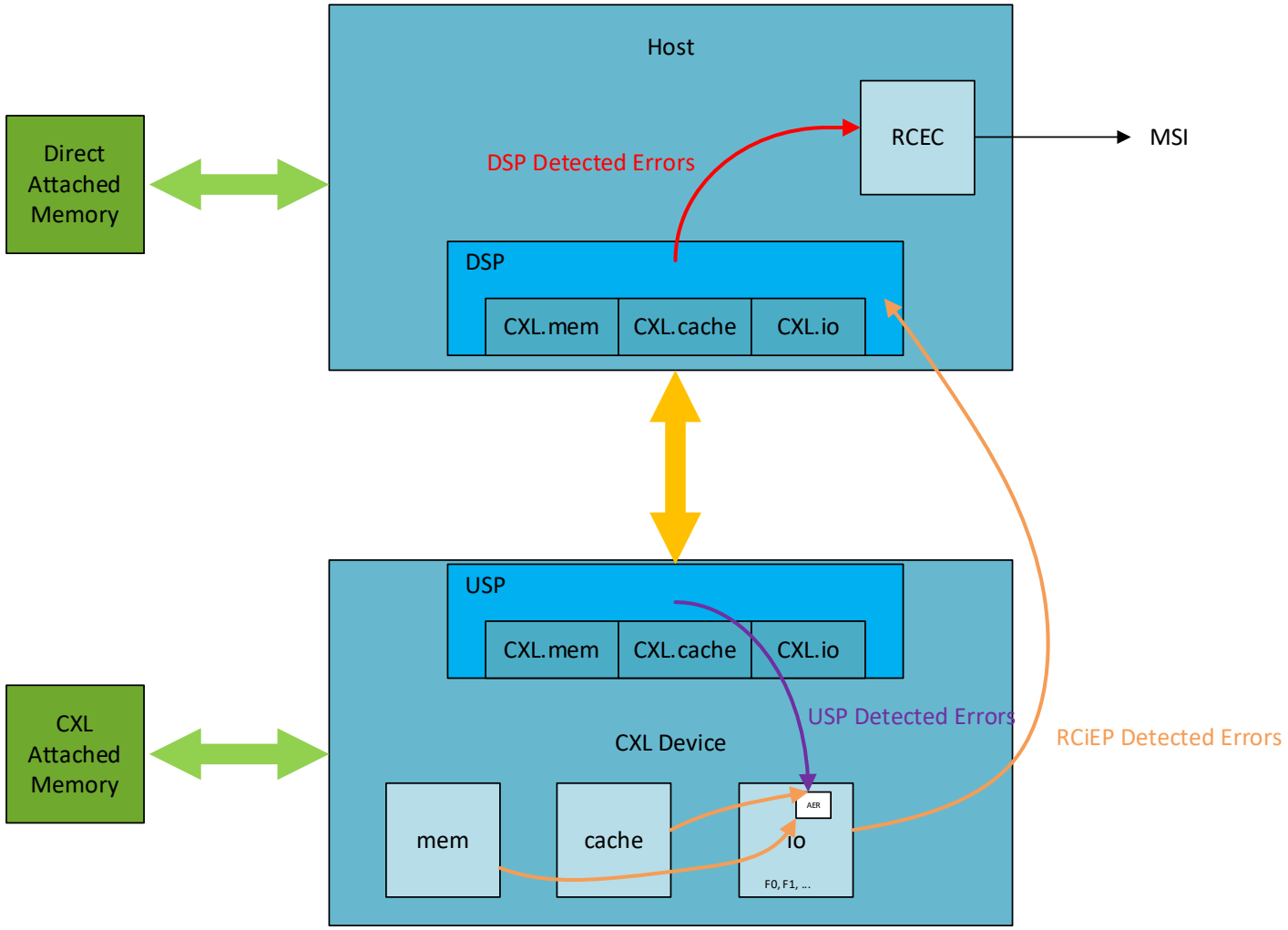
MICROCHIP

CXL™ Error Handling Mechanism



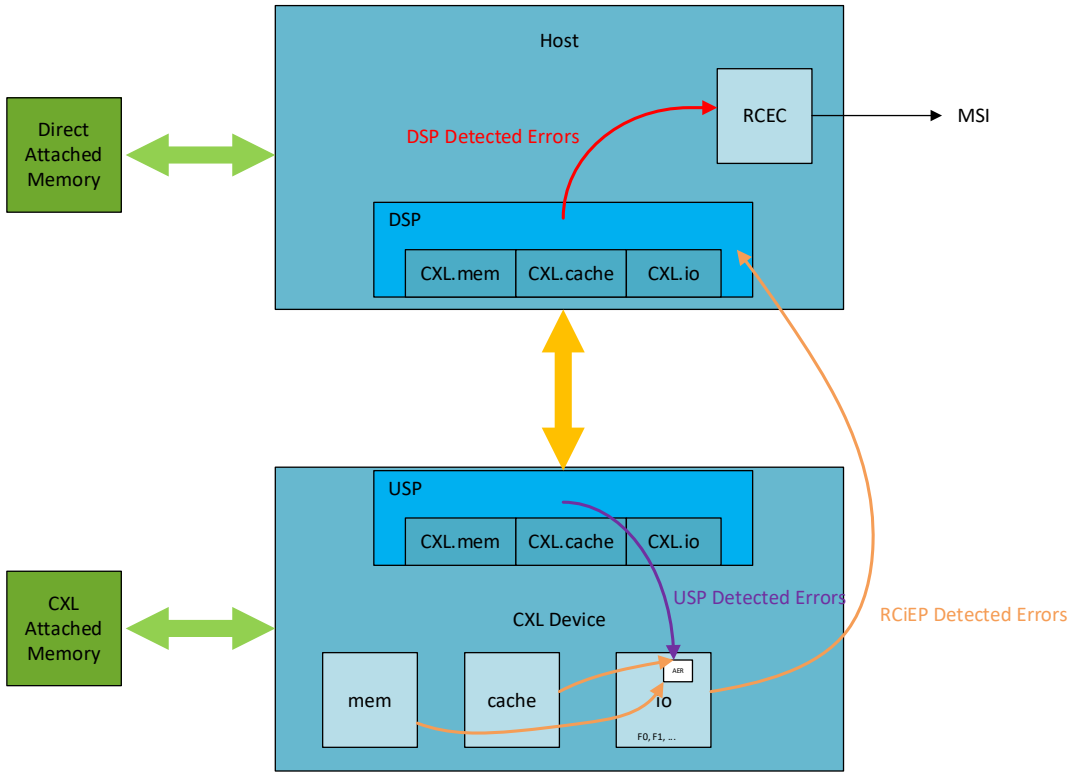
- Link and Protocol Layer Error Handling
  - Logged in RCEC AER
  - Communicated to host over CXL.io as PCIe® error message
  - Errors are classified and reported as Uncorrectable (UIE) or Correctable internal errors (CIE)
- Device Error Handling
  - Data poisoning mechanism for uncorrectable data errors
  - Errors that can't be handled with poison are signaled back to host as PCIe error messages

CXL™ 1.1 (RCD) Detection and Reporting



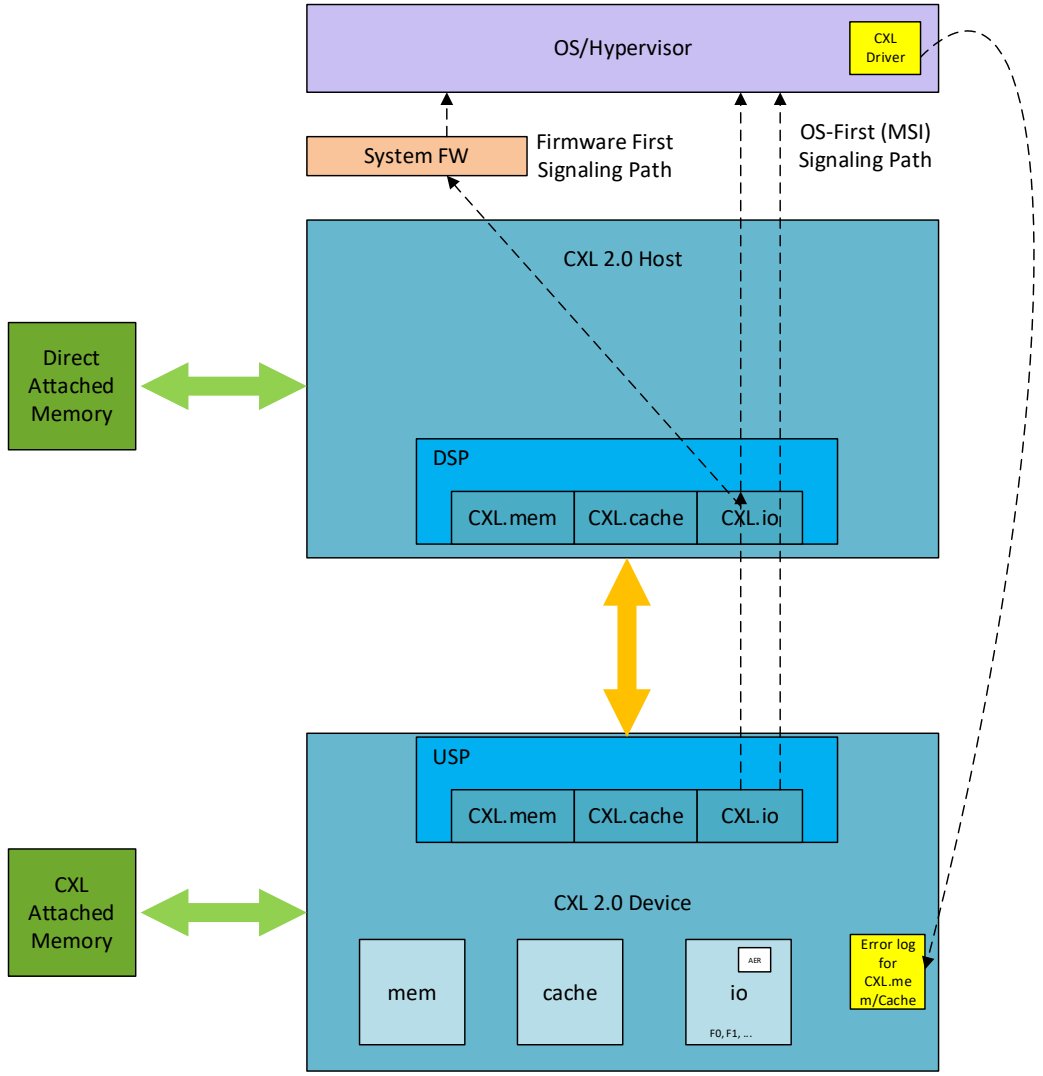
- Downstream detected errors
- Upstream detected Errors
- RCiEP detected Errors

CXL™ 1.1 (RCD) Detection and Reporting



- Downstream Detected Errors
  - Reported via UIE/CIE mechanism
  - Error messages sent to RCEC for logging
  - RCEC generates MSI
- Upstream Detected Errors
  - Reported to RCEC via CXL.io functions and across the link to RCEC
- RCiEP Detected Errors
  - Reported by CXL.mem and CXL.cache via CXL.io and across the link to RCEC

CXL™ 2.0 Memory Reporting Enhancements



- OS-First (MSI) continues to be a traditional Method of reporting errors
- Firmware-first signaling path added for better servicing and handling of the device errors

Device Specific Error Reporting and Nomenclature Guidelines

Error Severity	Example	Signaling Option (SW determined)	Logging	Host HW/FW/SW Response
Corrected	Memory Single Bit Error corrected via ECC	MSI or MSI-X to device driver	Device specific registers	Device specific flow in device driver
Uncorrected Recoverable	UC errors that are device recoverable	MSI or MSI-X to device driver	Device specific registers	Device specific flow in device driver
Uncorrected NonFatal	Equivalent to PCIe® UCNF, contained by the device (e.g., write failed, memory error that affects many computations)	MSI/MSI-X, PCIe® AER internal error	Device specific registers + PCIe AER	Device specific, System FW
Uncorrected Fatal	Equivalent to PCIe® UCNF, contained by the device (e.g., write failed, memory error that affects many computations)	PCIe® AER internal error, viral	Device specific registers + PCIe AER	System FW/SW, Viral flow

MICROCHIP

CXL™ Data Poison and Viral

- CXL defines two containment mechanisms
  - Poison:
    - Return Data on CXL.io, CXL.mem or CXL.cachemem gets tagged as poison. Poison Bit gets set on the flit in either direction to notify that the data is poisoned. For example, CXL device may notify the host that the memory read data is poisoned by setting an appropriate field in the S2M DRS field.
  - Viral:
    - Strong error containment feature. Upon detecting Viral status, CRC error is forced on next ongoing flit by the viral device. Receiver returns RETRY.Req to the viral device, which return RETRY.Ack with Viral bit set.
    - Instant reporting of viral ensured that fatal errors do not propagate through the system and cause data integrity issues.

CXL™ Error Injection

- CXL error injection mechanism is recommended to test various error scenarios and error handling flow. Following error injection hooks are recommended:
  - One type of CXL.io UC error
  - One type of CXL.mem UC error
  - One type of CXl.cache UC error
  - Link correctable errors
  - Returning poison on a read specified address

MICROCHIP

Conclusion

- Reliability, Availability and Serviceability (RAS) are key features for large scale data center server design and deployment
- The CXL™ standard defines multiple RAS features to address error containment and reporting
- CXL is an emerging technology. Further evolution of the standard, as well as more advanced RAS features and capabilities, should be expected