State of the CXL Linux Ecosystem

FMS: Flash Memory Summit

Santa Clara, CA. August 2024.

Davidlohr Bueso (Samsung), Dan Williams (Intel)
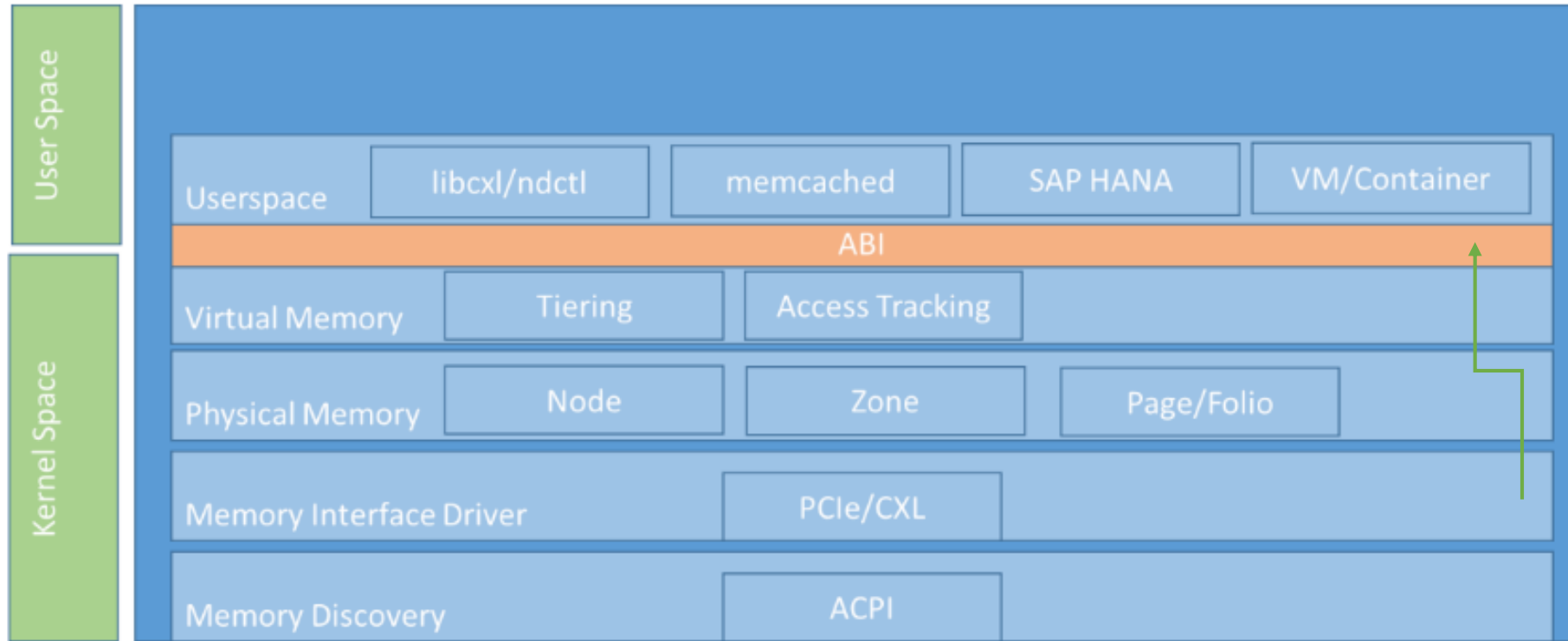
1000-ft View

- Hardware availability and momentum continues to increase.
    - CXL 1.1 Type3 devices, vendors announcing 2.0 devices.
    - Interest in Type2 devices.
    - Enterprise Linux coming into the picture.
        - VMs, databases, cloud/containers, etc.

- Concerns about pooling and sharing notwithstanding.
    - Linux community skepticism, *"A Case Against CXL Memory Pooling"*.
    - Shared memory (FAM) is very much an experimental effort.

CXL Linux Software Stack

Testing: qemu

Enables testing Linux support as well as OoB solutions.

- Support is a mix of 2.0+ feats.
- Type3 (pmem/volatile)
- Switch

- Dynamic Capacity Devices
- CCI/mbox
- OS irq/FW first
- Features
- CXL RAS
- FW Update
- Poison/Scan Media
- MHD

Driver Highlights

- FW-First (Events)
  - CXL to CPER reporting.

- RCH (CXL 1.1) Error Recovery.

- CXL QoS
  - Static ACPI tables cannot represent dynamic CXL configs.
  - Linux can now determine NUMA topology info.

- Error Injection (EINJ module).

- Sanitation, FW Update, Logs, Error Injection (EINJ module), etc.

Driver Opens

- DCD

- Accelerators
  - HDM-D, HDM-DB.

- CXL RAS
  - EDAC integration.
  - Maintanence, PPR, scrubbing, DDR5 ECS, etc.

- Switch CCI

- Dirty Shutdown, Reset, enum corner cases, RCH Link Status, etc.

Memory Management Highlights

- Less concern about introducing CXL ad-hoc changes to core MM.
  - Everything is in terms of NUMA nodes.
  - Remove general assumptions of CPU/compute in every node.
- Weighted-node memory allocations (better bandwidth).
- Memory tiering.
  - Promotion/demotion strategies are a hot topic.
- Hugepages (VMs, Databases).
  - Hugetlbfs unifications, multi-sized THP.

Fabric Management (special mention)

- Unconsolidated efforts. Lack of community, fragmented, etc.
  - Qemu (Switch CCI).
  - Something to improve for those interested in such topics.

- Open source efforts:
  - Jonathan Cameron / cxl-fmapi-tests
  - computexpresslink/libcxlmi: CXL Management Interface library
  - computexpresslink/cxl-fabric-manager: CXL FM infrastructure
  - JackrabbitLabs/jack: CXL Switch Emulation

Getting Involved (open collaboration)

- CXL Linux mailing list.
  - linux-cxl.vger.kernel.org

- Monthly upstream sync meetings.

- Linux Plumbers Conference.

Thank you.