Support for Open-Channel SSDs

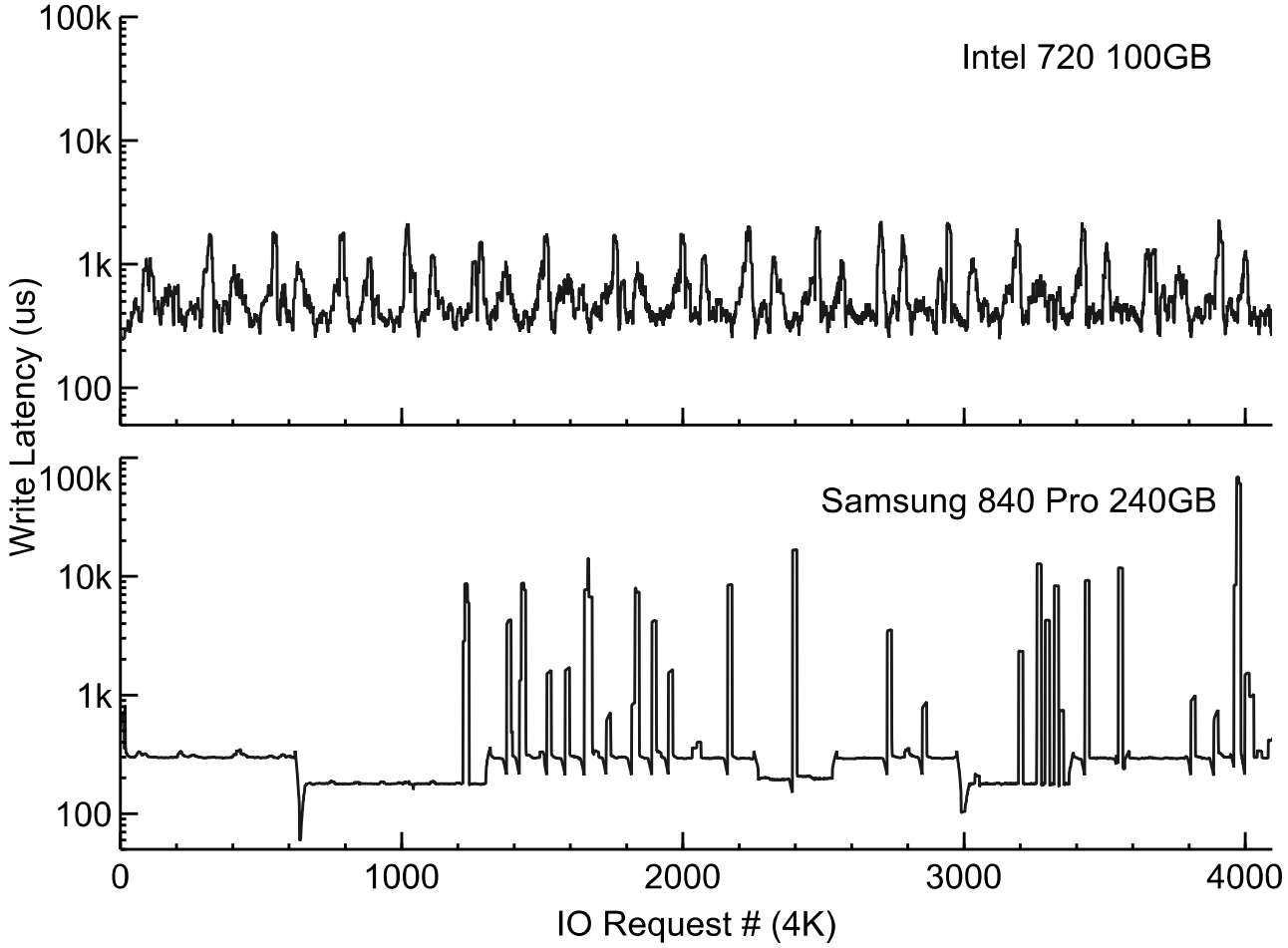
Matias Bjørling, Jesper Madsen, Javier González

Linux Plumbers Conference
16th October 2014

IT University
of Copenhagen

SSDs are Unpredictable

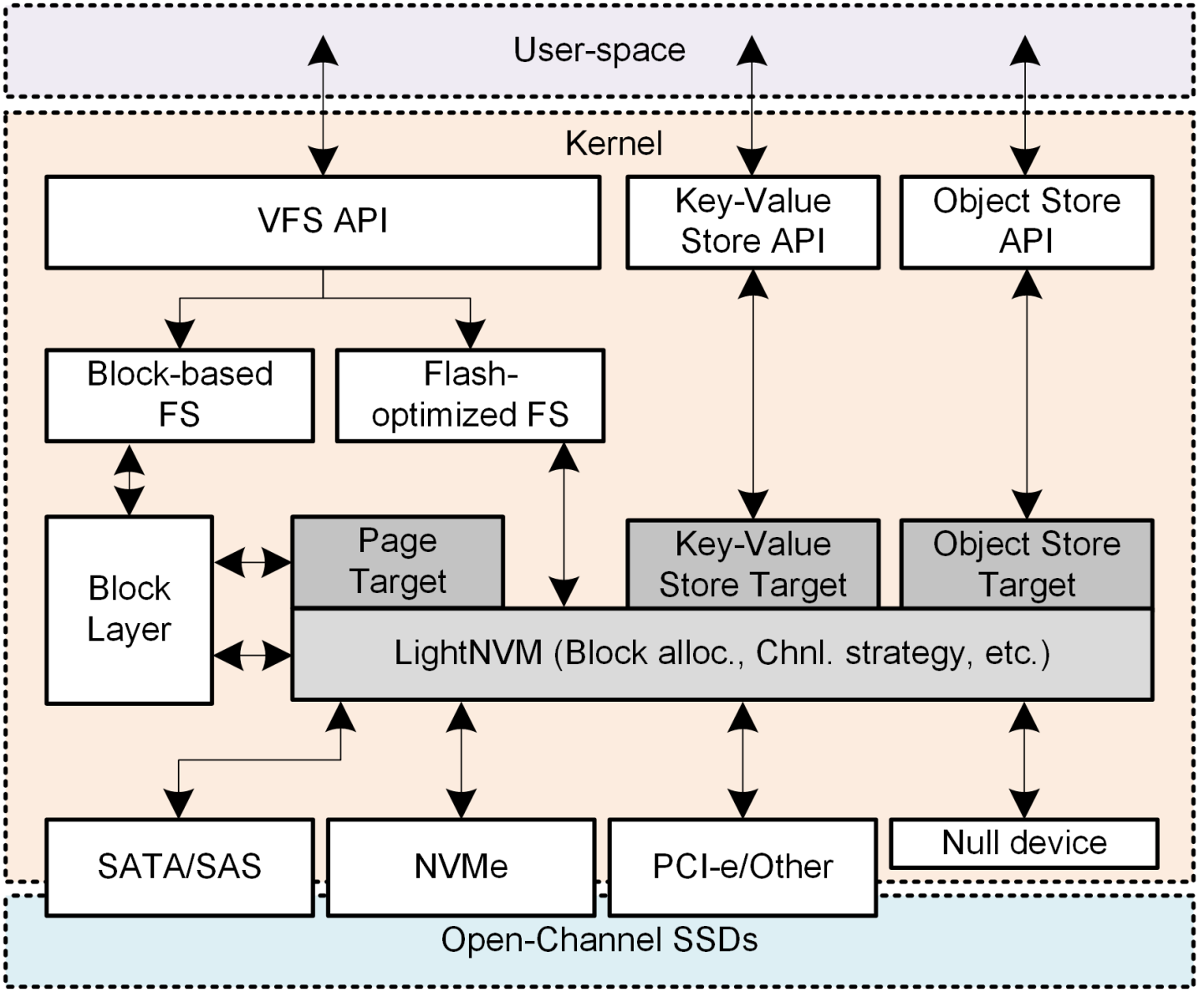
- Narrow block interface
- 1000x difference in latency



Open-Channel SSDs

- Allow host to control data placement and garbage collection
- Expose new commands to the host
  - Identification of SSD geometry and configuration
  - Flash Erase (Sync./Async.)
  - Flash health (piggy-backed onto read and write accesses)
- Offload as much as possible to hardware
  - Flash controller, bad block management, ECC, etc.
- Benefits
  - Control data placement → Control parallelism → Control latency
  - Align SSD algorithms to system workload → Optimal resource usage
  - Allow new interfaces to interact directly with flash → Single translation

Architecture



Platforms

- Availability
  - OpenSSD Jasmine (SATA)
  - OpenSSD Cosmos (PCI-e/NVMe)
  - IIT Madras, India – (NVMe/RapidIO)
- Keith's QEMU NVMe implementation
- Prototype SSD from commercial vendors

Community

- Local File Systems?
  - Atomic IOs, expose SSD configuration
- Database Systems? (ABIs)
  - Atomic IOs, simplify storage path, directly access physical flash pages
- Distributed File Systems? (ABIs)
  - Ceph (osd-ceph/LevelDB/RocksDB)
  - Gluster
- LightNVM Internals?
  - Efficient flash scheduling profiles
  - Low and high memory environment
- How can we make other drives work?
- Is this the right architecture for your drive?