Next Generation Architecture for NVM Express SSD

Dan Mahoney

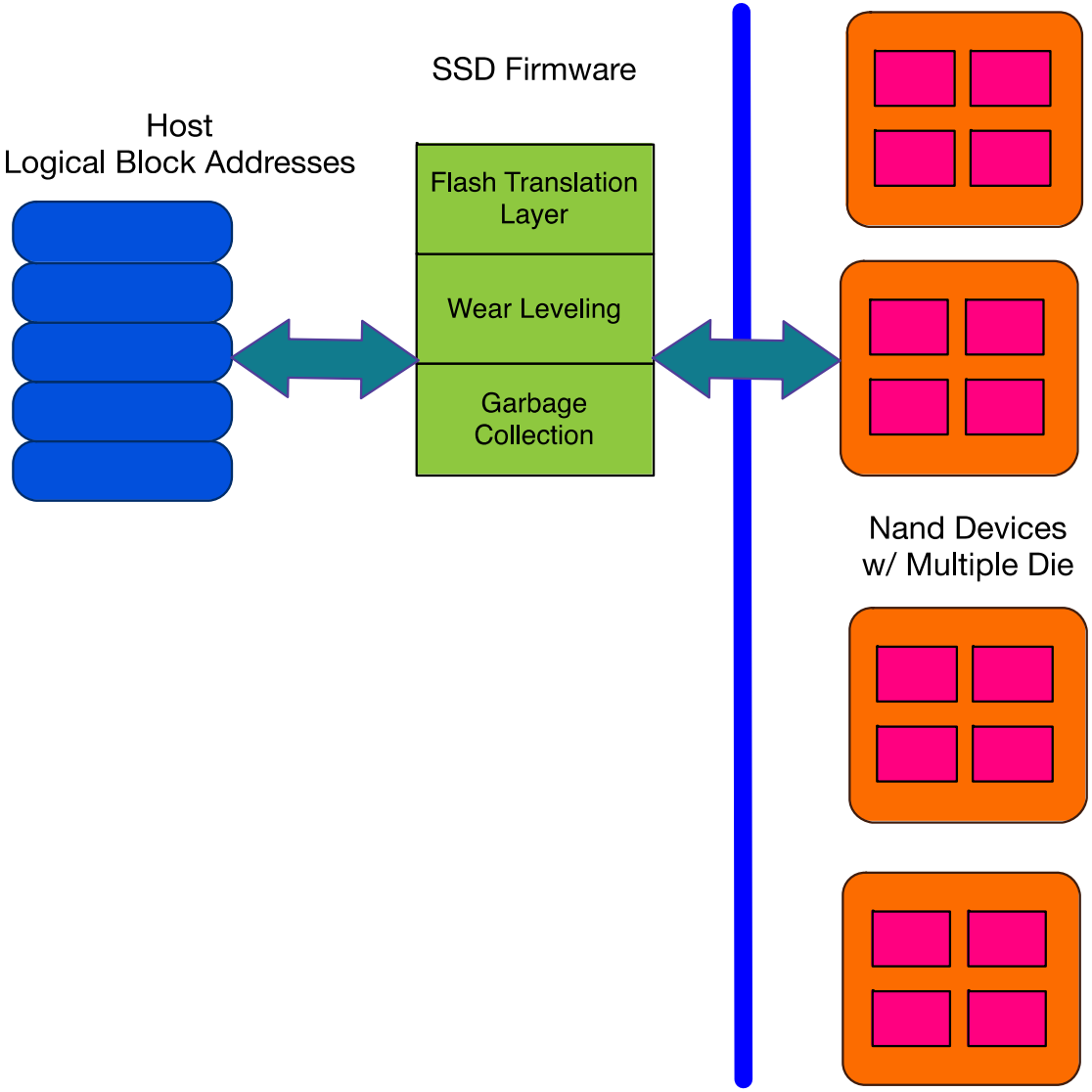
CEO

Fastor Systems

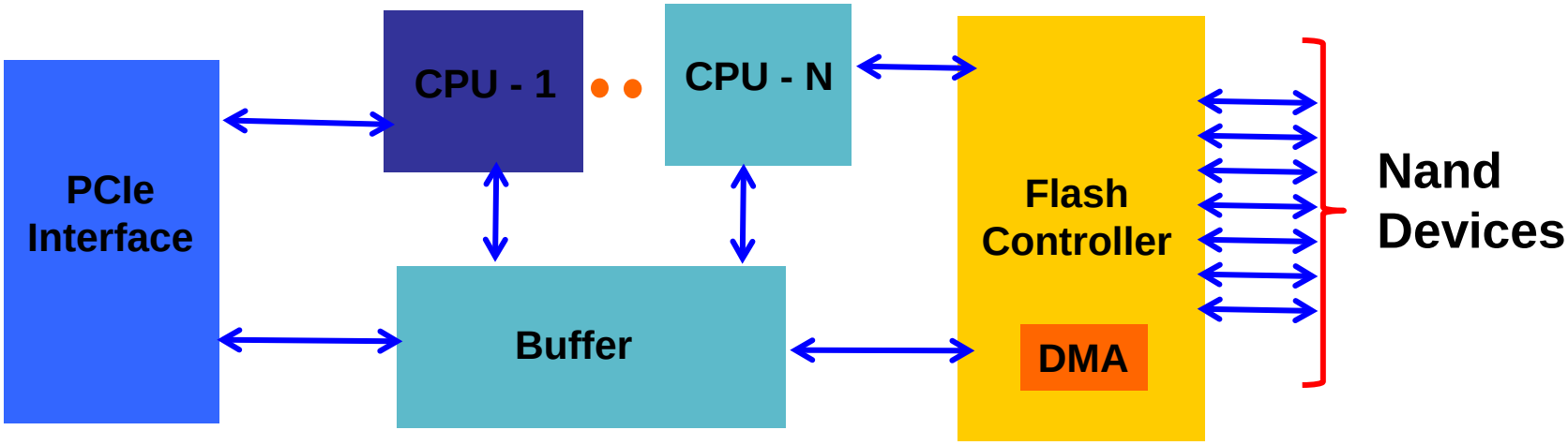
NVMExpress Key Characteristics

- Highest performance, lowest latency SSD interface protocol today
- Architected from ground up for NVM scaling from client to enterprise.
- Large number of parallel commands
- Wide queuing interface
- This protocol begs for unconstrained access to the Flash media

Essential SSD Key Operations



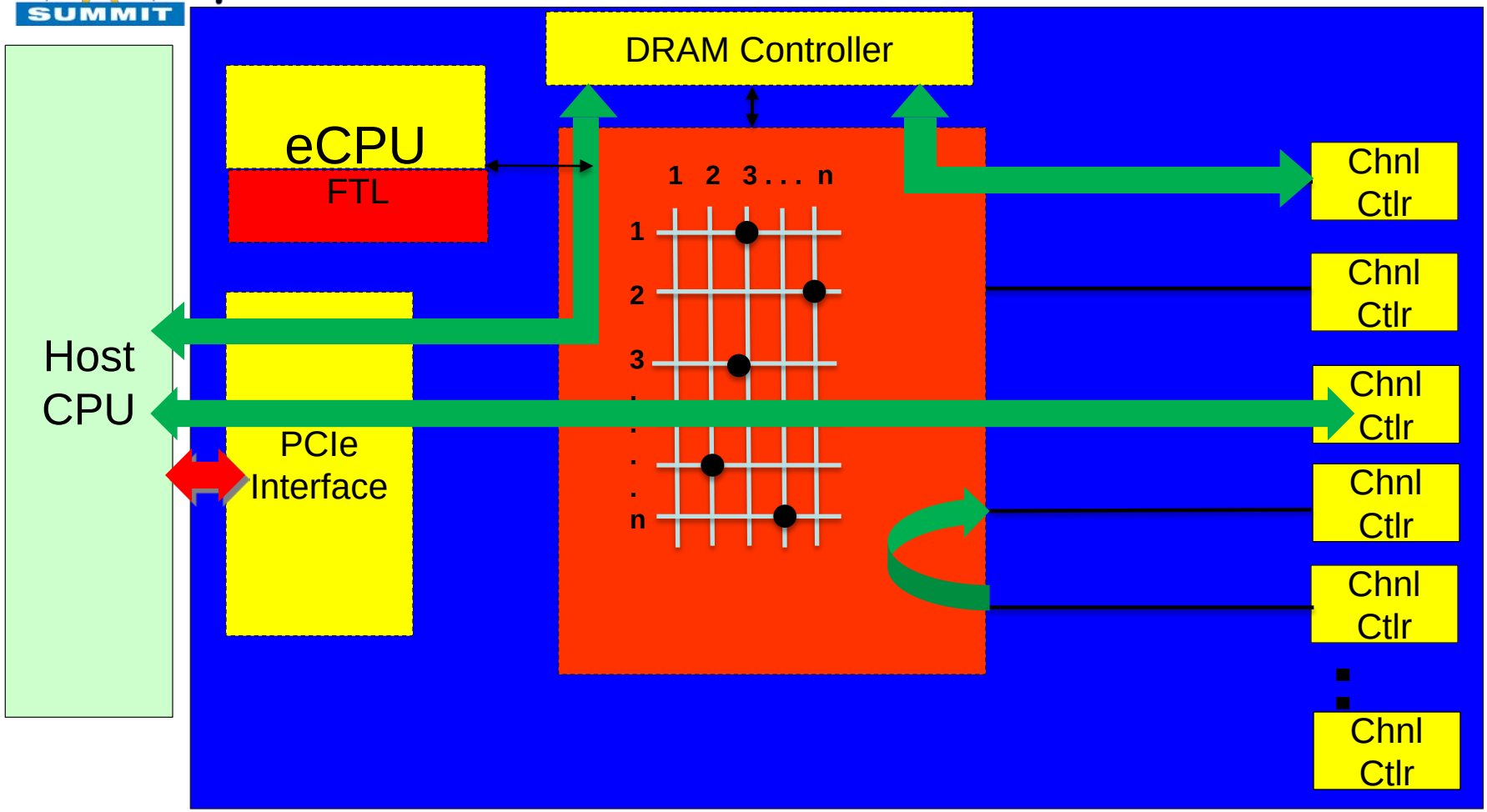
Typical SSD Controller



Fabric Based SSD Architecture

- Utilizes a Non-Blocking Fabric to connect:
  - Multiple Channel Controllers
  - PCIe Interface
  - Memory Controller
  - Embedded CPU
- Exploits PCIe high performance low latency protocol to connect to array of Flash devices
- Permits Concurrent Data Flow across all modules

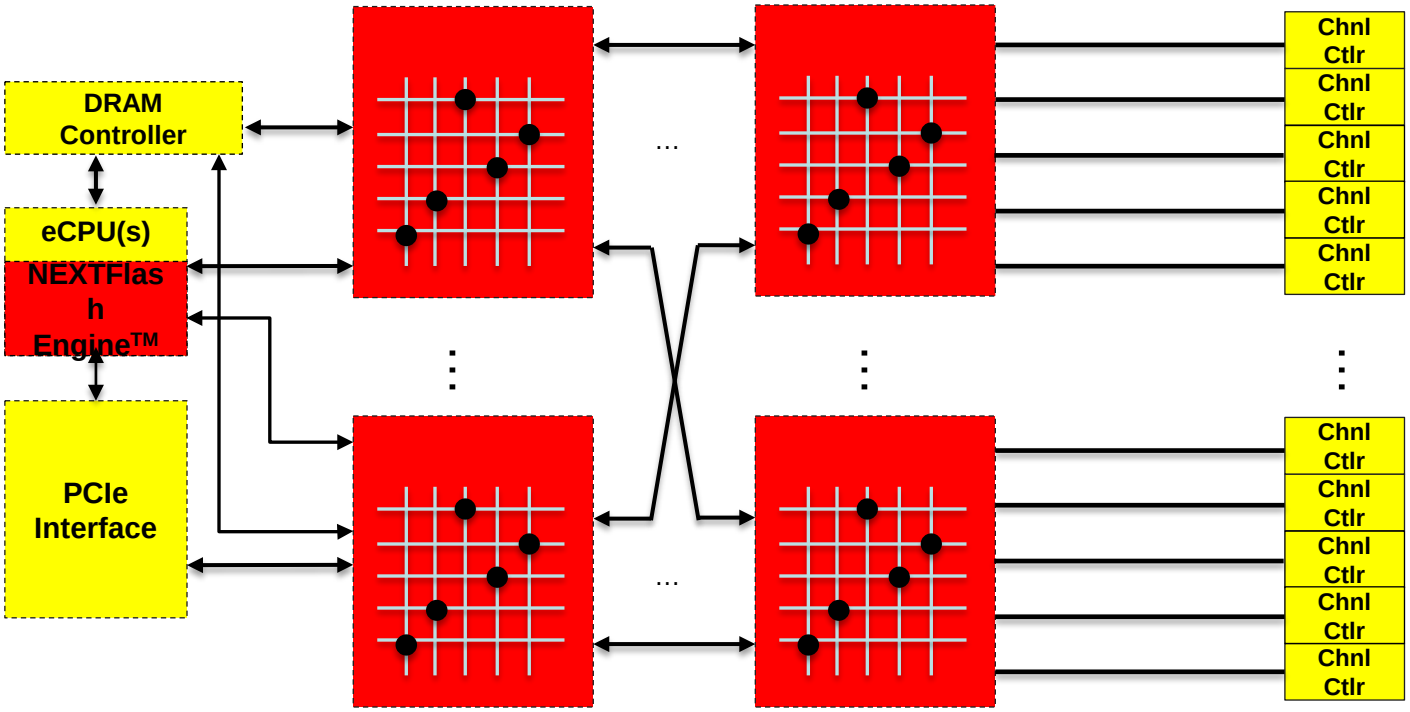
Fabric Based SSD Architecture



- Non-Blocking Fabric coupled to Individual Channel Controllers

A Highly Scalable Architecture

The data plane (i.e. Fabrics) Scales Independently of the Control Plane

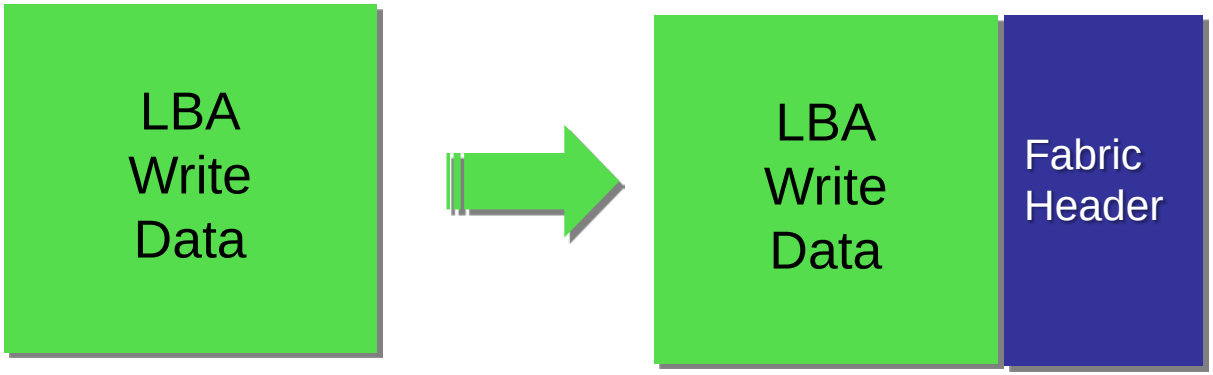


Today's fabric-based architecture: 12TB & 3.2M IOPS

The architecture of the future: 100TB & 25M IOPS !

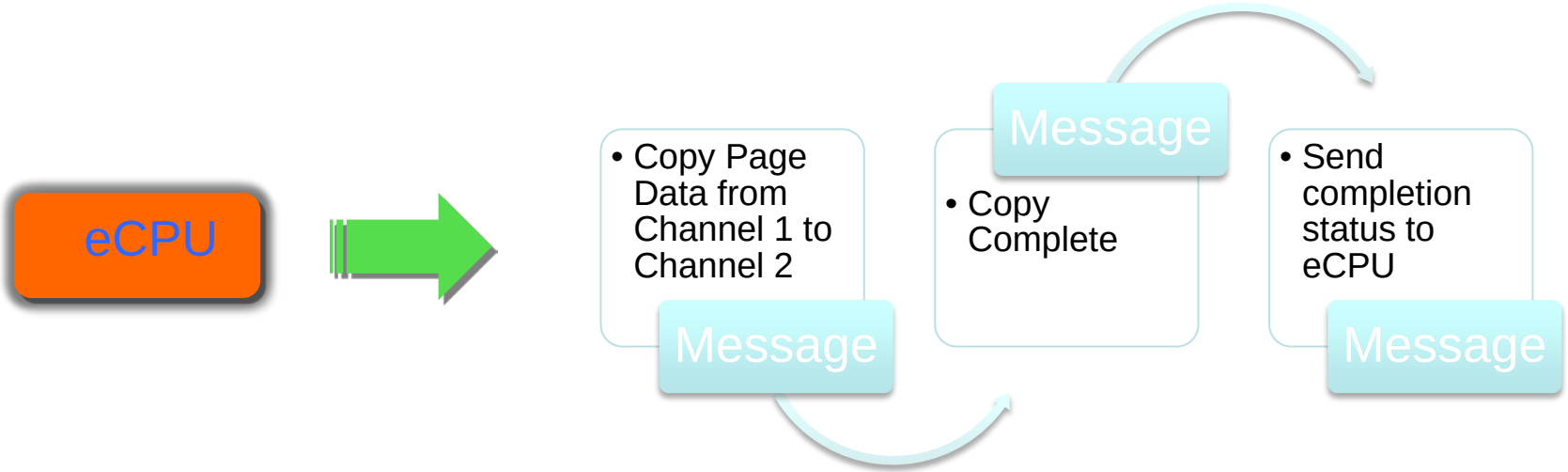
Command Processing

- Fetch command and data from host
- Assemble internal datagram with Fabric header



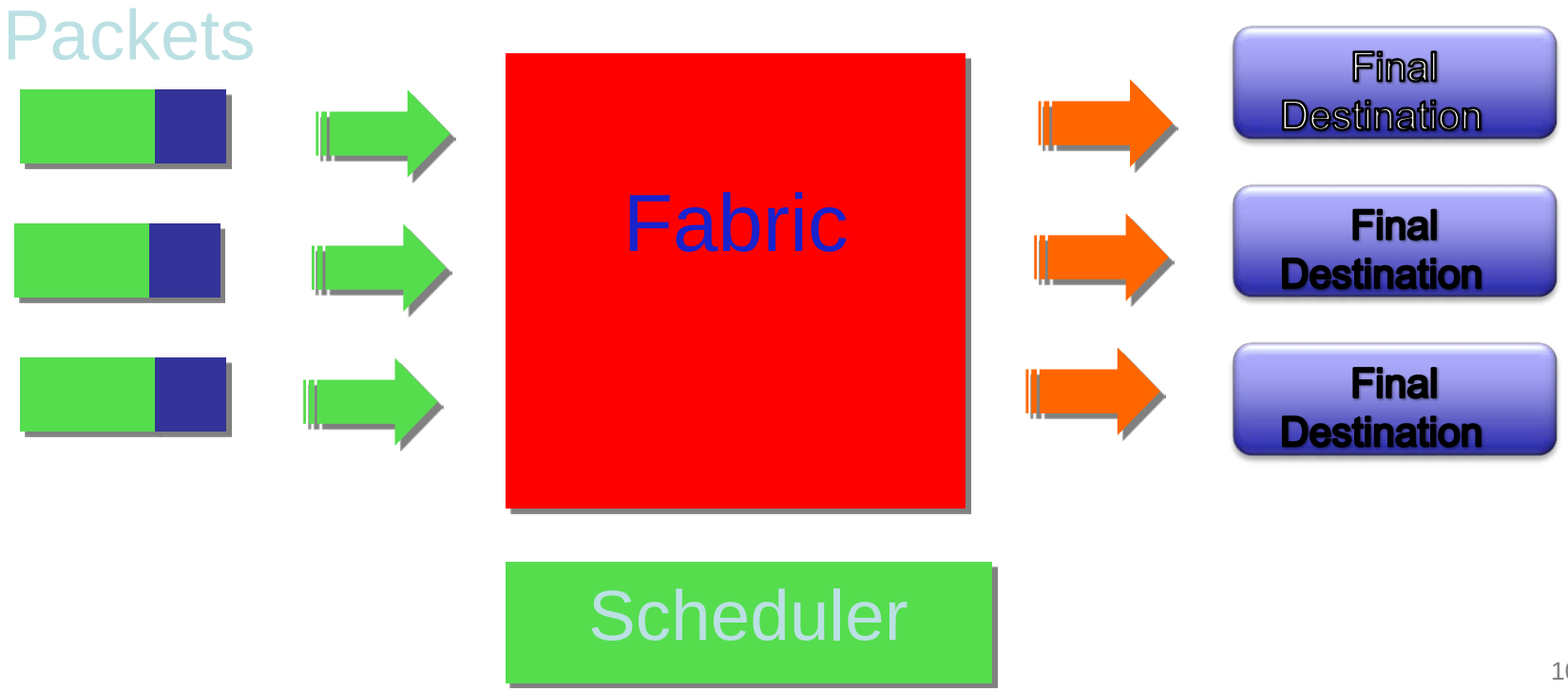
Message Based Architecture

- Messages sent to Channel Controllers for Data Movement



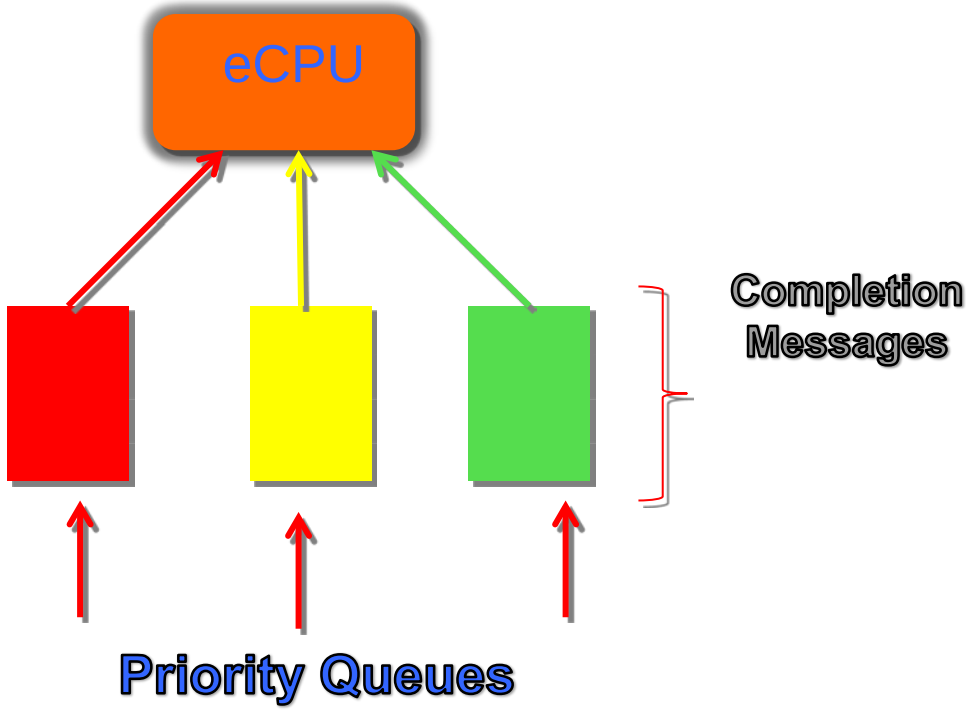
Packet Forwarding

- Schedule Packets for Ingress into Fabric
- Forward to Destination

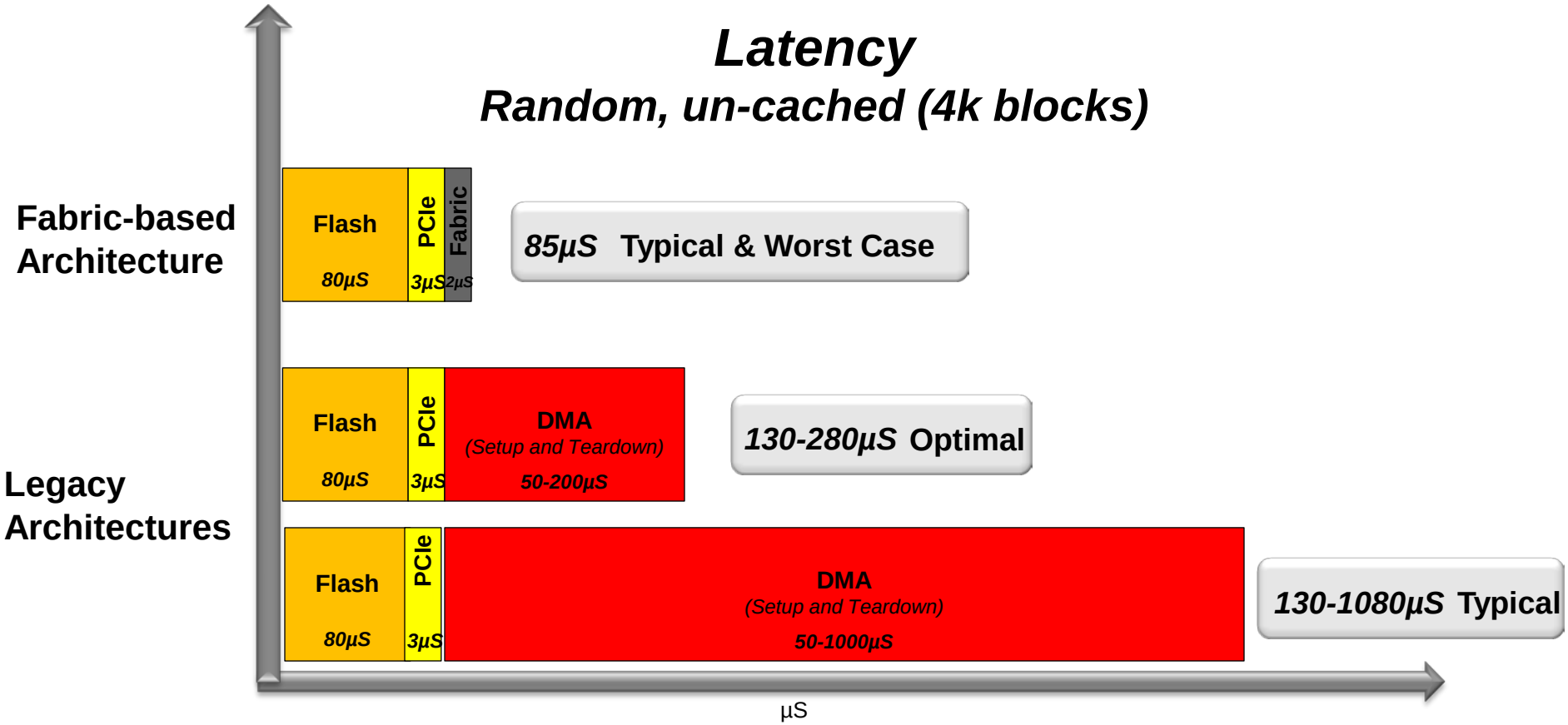


Packet Completions

- After packets are processed by channel controller, completions are sent to eCPU
- Permits Asynchronous completions

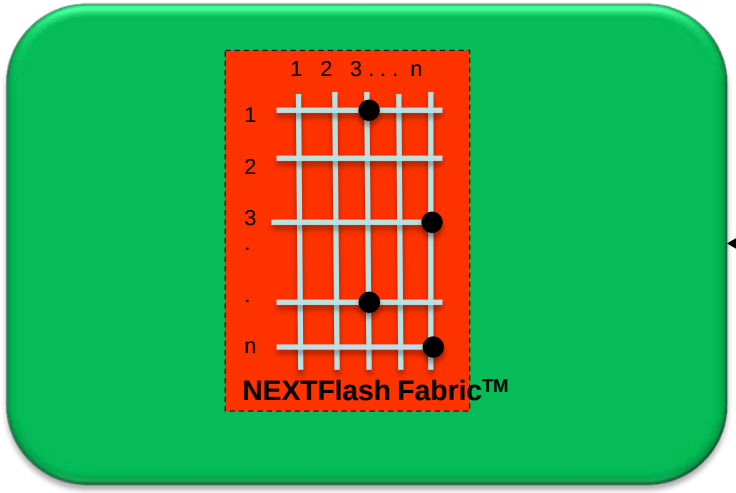


Concurrency Eliminates the Main Source of Latency



Determinism Enables SDS

Unmatched Throughput & Latency

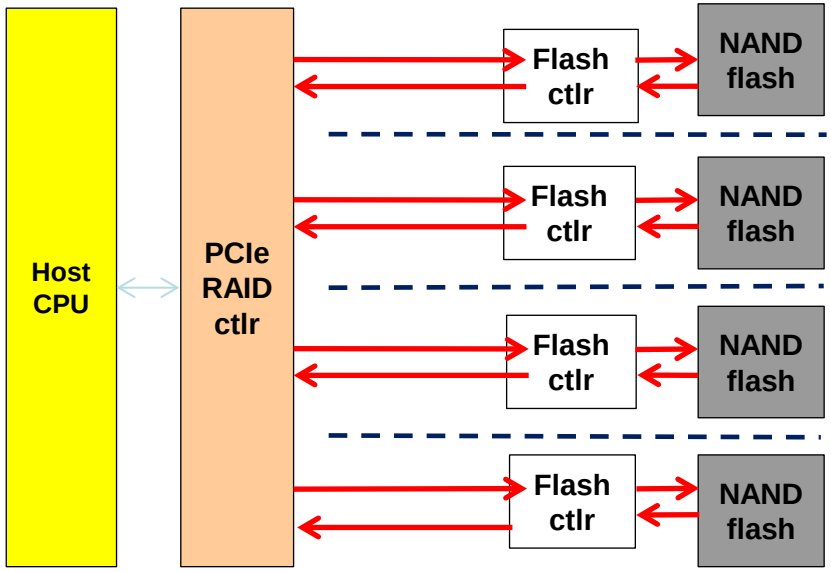


Enabling SDS

- Storage QoS
- VM storage SLAs
- Application Optimization
- IOPs Carving
- Capacity Carving

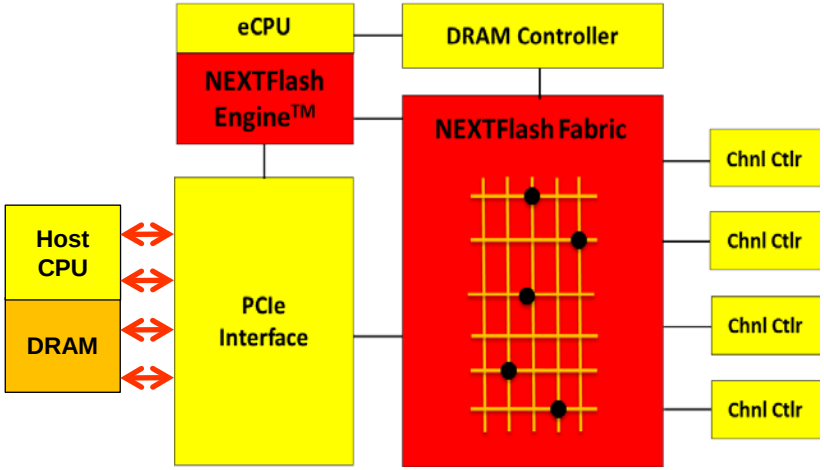
Legacy vs Fabric Summary

Scaling by Controller Replication



- Islands of Flash – no global wear leveling
- Control resources must be replicated to scale the Flash
- Comingled control and data planes make the controller a choke point for data movements

Fabric Operated



- Saturate Interface
- Deterministic
- Highly Scalable
- SDS enablement
- Low TCO

Comparison of Write Overhead

Controller-Centric

1. ID DMA channel
2. CPU issues start address
3. CPU hand over end address
4. CPU send start command to DMA controller
5. DMA transmits from memory to Flash

Fabric-Based

1. Fire and forget!

Five steps – CPU cycles – compared to one

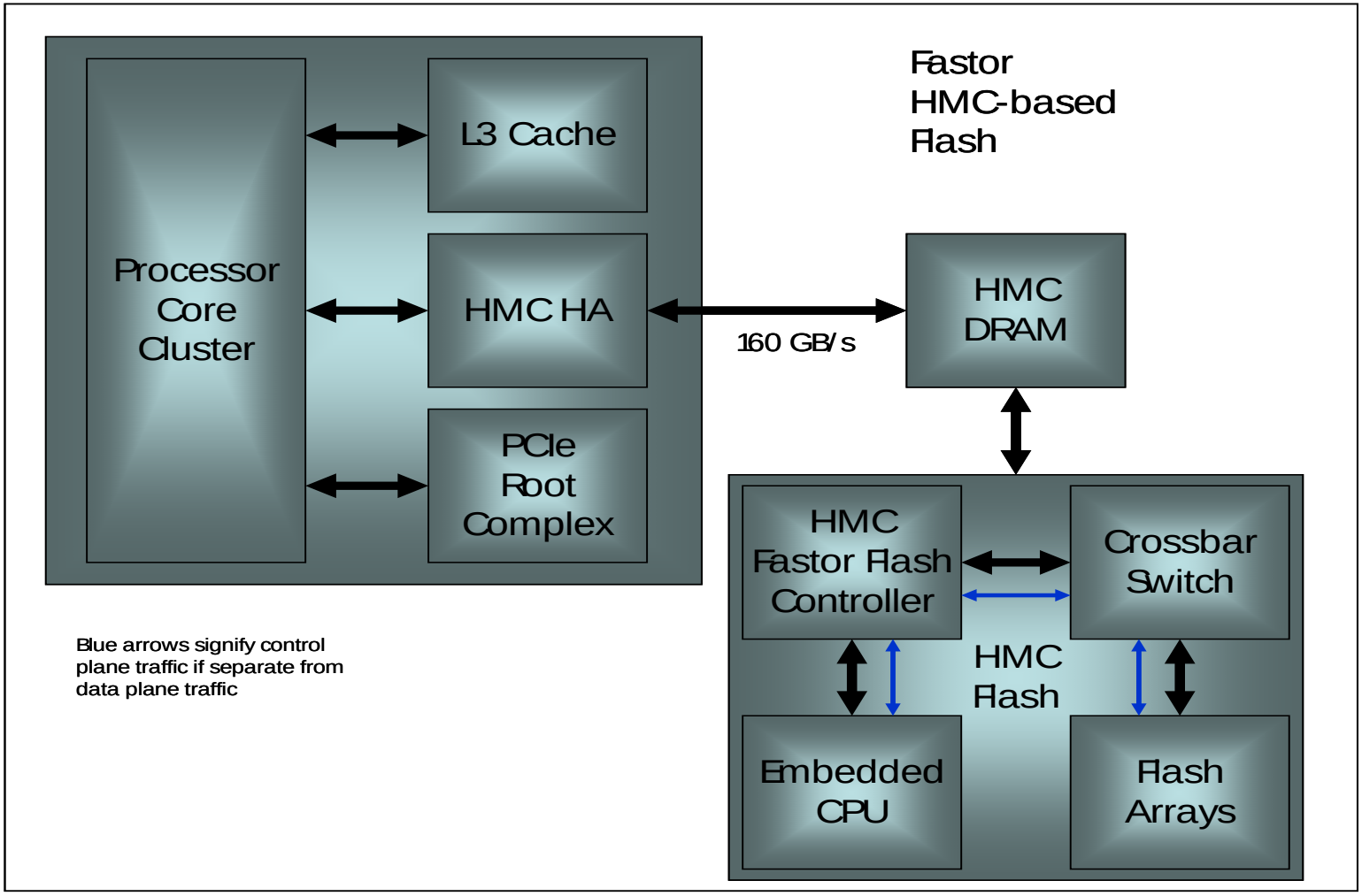
Advantages of Fabric Based Architecture

- Allows asynchronous movement of datagrams within the SSD – Concurrency!
- No Setup/Tear down of DMA channels
- Provide QoS within the SSD
- Separation of control and data planes offers scalable and deterministic performance
- Agnostic regarding interface and storage media – aligned with emerging technologies such as MRAM, PCM, etc..

Ultimate Architectural Potential

- This interface-agnostic architecture can support higher speed interfaces – i.e. HMC
- This would allow large scale Flash modules that look like DRAM to the host CPU
- HMC Flash modules attached to an HMC-enabled processor could ultimately scale to
  - 10X Bandwidth improvement with 2.5X or greater latency reduction compared to state of the art Flash-based solutions
  - Potential appliance capacity of 25% of an HDD-based appliance with 100X the bandwidth

HMC Attached Flash



Conclusion

- Better controllers are good, but if they remain in the SSD's data path, they remain a bottleneck
- Emerging storage software solutions will all benefit from faster, more deterministic, more scalable storage hardware