NVMe Over Fabrics Agenda

❑ **High-level Overview**

➢ Steve Sardella – EMC

❑ **Performance and Emerging NVMs**

➢ Zvonimir Bandic – HGST

❑ **Real World Use Cases**

➢ Idan Burstein - Mellanox

NVMe Over Fabrics

Overview

Steve Sardella

Distinguished Engineer, EMC

August 11, 2015

Topics

Why NVMe over Fabrics?

What is NVMe over Fabrics?

Status

A Brief History of NVMe over Fabrics

NVMe (over PCI Express) is an extremely successful standard, that continues to grow in popularity

But PCI Express is not without limitations
- Number of SSDs that can be supported
- Robustness and error handling
- Maturity as a Data Center "fabric"

The NVMe Specification did such a good job of defining an efficient queuing interface for storage, there is now a desire to extend it to other protocols
- These mature protocols are already established within the Data Center

Without standardization, there could be many disparate implementations, by protocol or by silicon vendor

NVMe in a non-PCIe environment
## (SCSI approach)

- Creating an NVMe Flash appliance containing hundreds of SSDs using SCSI based protocols requires protocol translation

- This adds latency and defeats a major benefit of NVMe
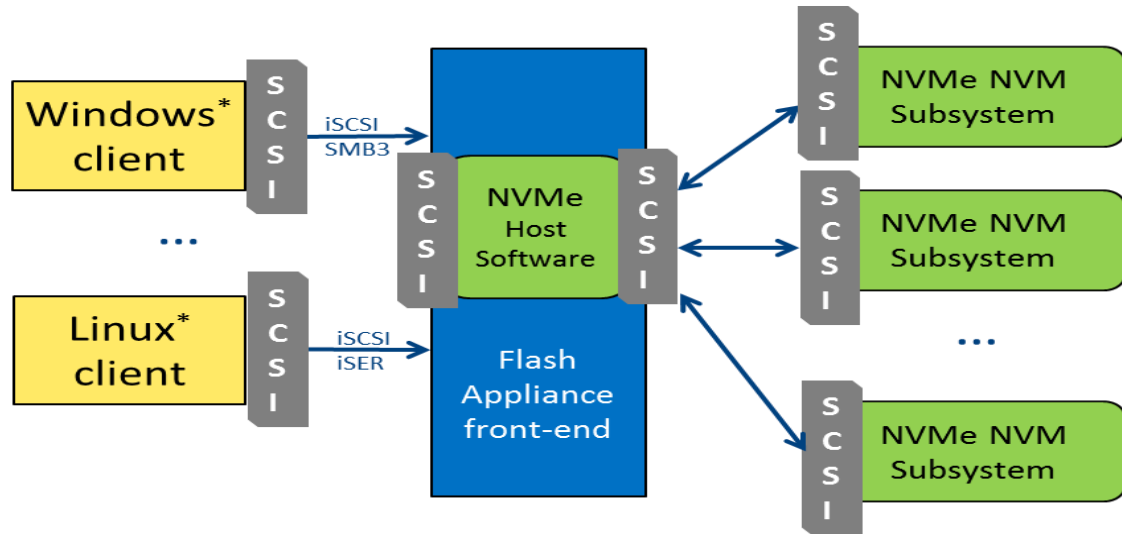
Concern: Low latency of next gen NVM lost in (SCSI) translation
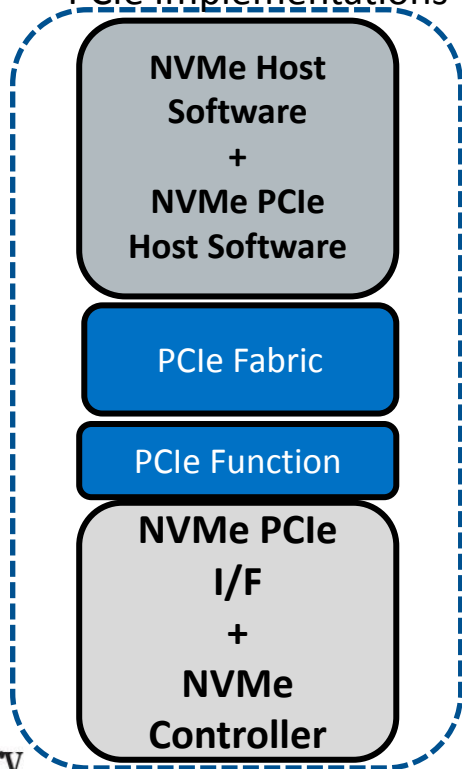
Introducing NVMe over Fabrics

- Through encapsulation, NVMe over Fabrics transfers the vast majority of existing NVMe Commands, Responses, Structures and Concepts, end-to-end across a Fabric
- Maximizes HW/SW reuse
- Goal is to add less than 10 microseconds additional latency between a local and remote SSD
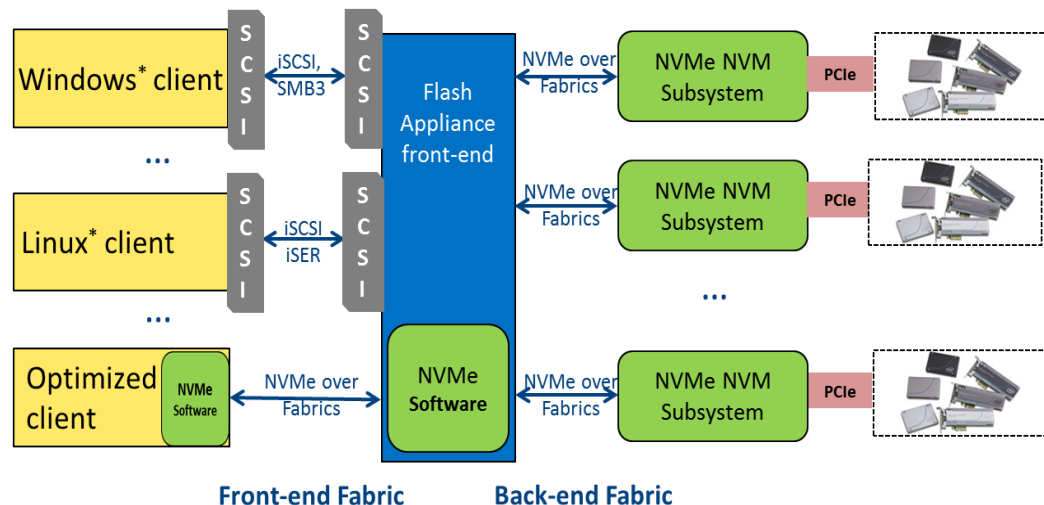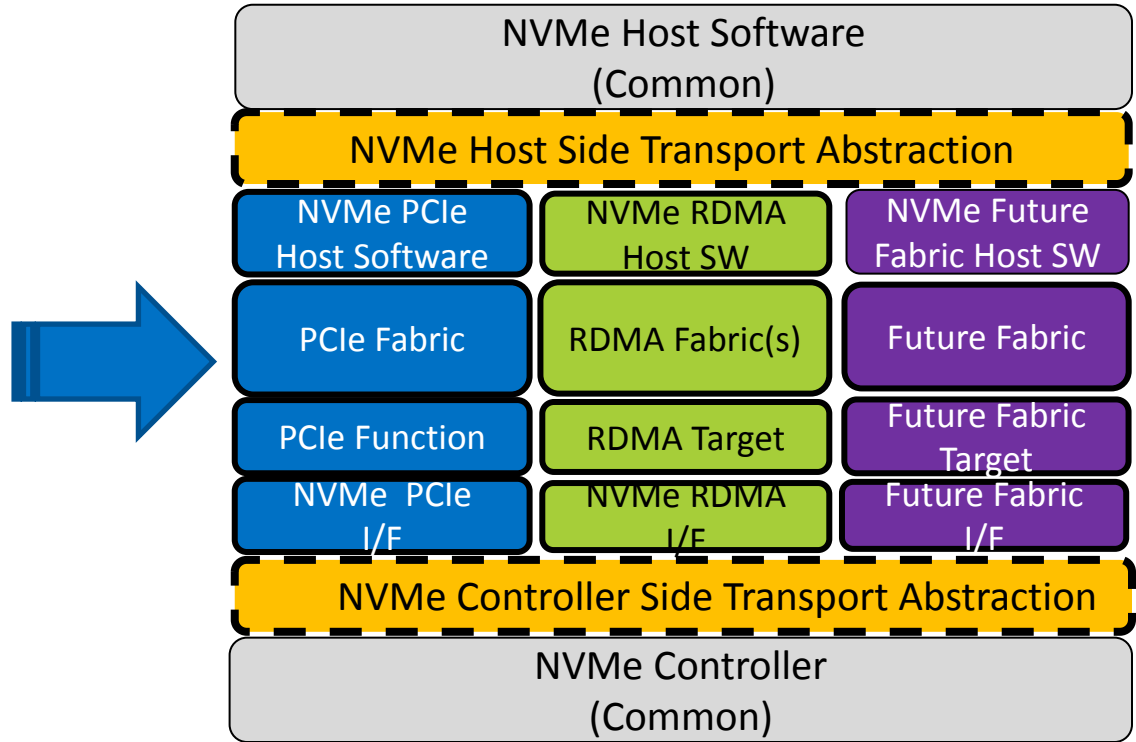

Extend efficiency of NVMe over front and back-end fabrics
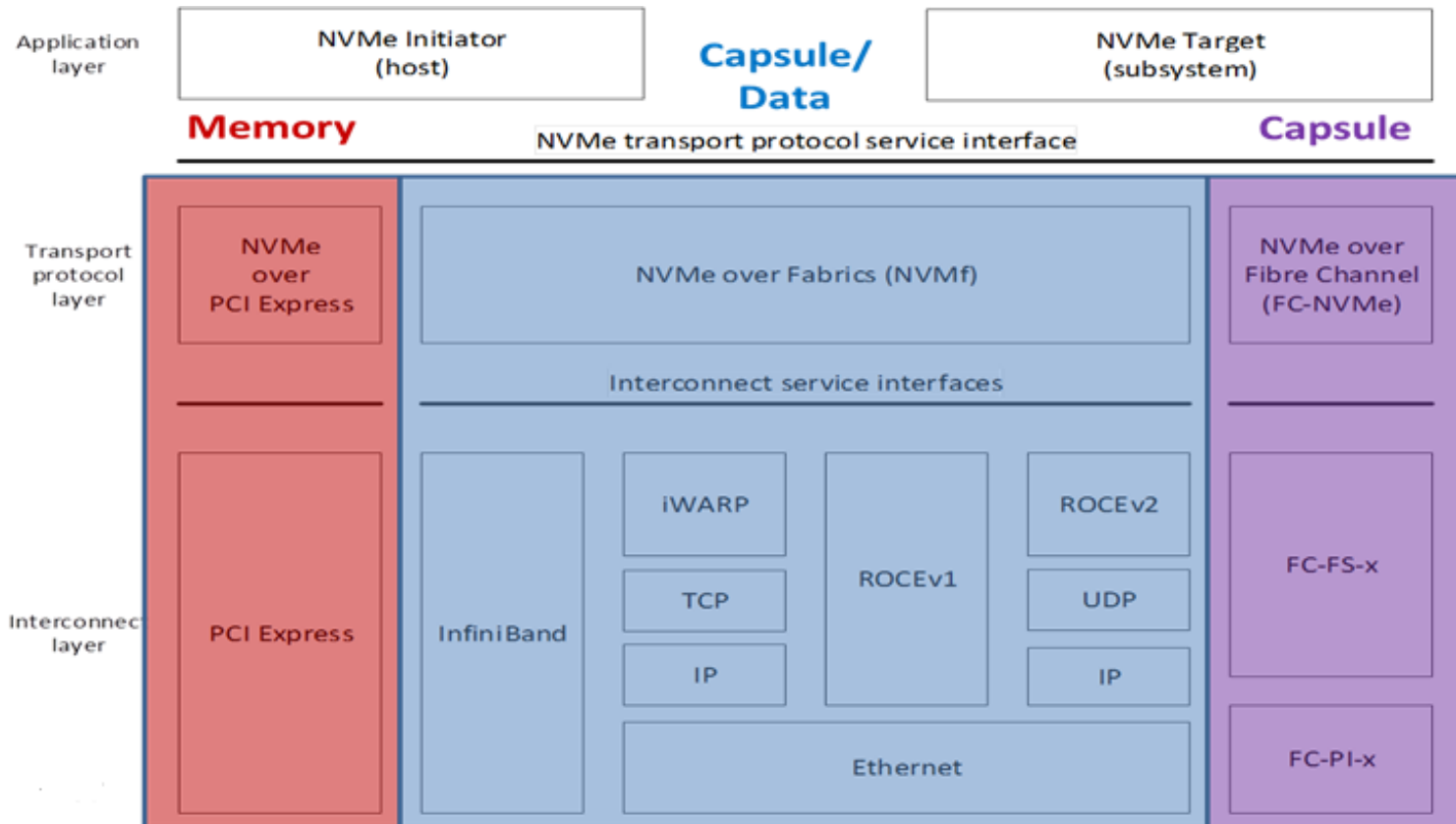
NVMe Multi-Fabric Model

NVMe 1.0->1.2
PCIe Implementations

NVMe over Fabrics Architecture Model

NVMe Host Software + NVMe PCIe Host Software		
PCIe Fabric		
PCIe Function		
NVMe PCIe I/F + NVMe Controller		

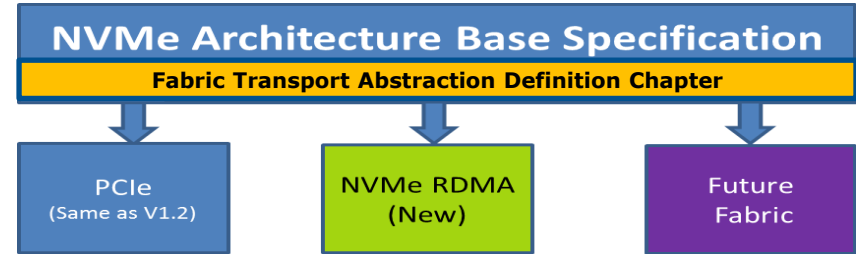| NVMe Host Software (Common)		
NVMe Host Side Transport Abstraction		
NVMe PCIe Host Software	NVMe RDMA Host SW	NVMe Future Fabric Host SW
PCIe Fabric	RDMA Fabric(s)	Future Fabric
PCIe Function	RDMA Target	Future Fabric Target
NVMe PCIe I/F	NVMe RDMA I/F	Future Fabric I/F
NVMe Controller Side Transport Abstraction		
NVMe Controller (Common)		

NVMe Multi-Fabric Model

Specification Strategy and Breakdown of Work

Do not create a standalone specification

- Initial goal is to minimize changes to existing specification
- Cleanly separate out the non-PCIe NVMe Transport layers through separate chapters/sections
  - Fabrics Core (concepts and RDMA binding)
  - Fabrics Base Differences (SGL changes, etc.)
- Long-term goal is to create a Transport agnostic base spec

NVMe Architecture Base Specification

Fabric Transport Abstraction Definition Chapter

PCIe (Same as V1.2)	NVMe RDMA (New)	Future Fabric

Break the work into functional sub-sections

- Capsules
- Discovery
- Connections
- Flow Control
- Naming
- Binding
- Error Handling

NVMe over Fabrics in the News

- Press Release and Intel iWARP demo at IDF 2014
- SNIA Webcast last November, which included NVMe and NVMe over Fabrics
  - Record breaking attendance
- Chelsio December announcement
- Mellanox/Mangstor demo announcement
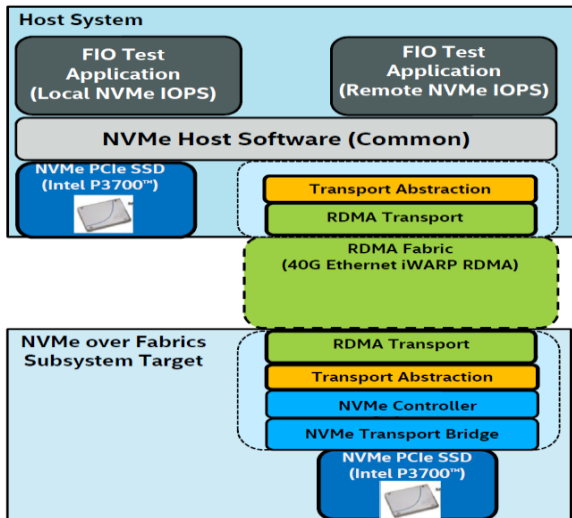
- Extending NVMe's capabilities and benefits to Fabrics is the next evolutionary step

- The team has achieved many milestones, and there is more work ahead

- The completed specification is scheduled to be released by the end of 2015

- The principles that make NVMe so popular today will continue to guide the Work Group for NVMe over Fabrics

  - Performance and efficiency
  - Low latency and low overhead