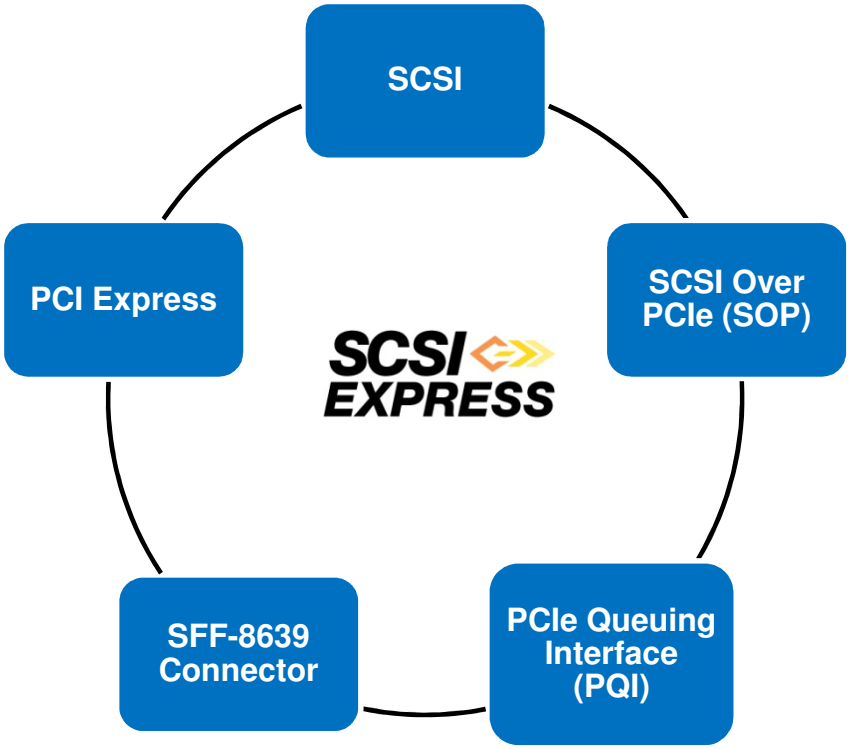
Optimized Solid State SCSI Initiative

Joe Foster
Member, SCSI Trade Association
Distinguished Technologist, HP

Express Overview

- **What is SCSI Express?**
  - Proven SCSI protocol combined with PCIe creating an industry standard path to PCIe-based storage
- **Why do we need SCSI Express?**
  - Deliver proven enterprise storage for PCIe-based storage devices in a standardized ecosystem
  - Take advantage of lower latency PCIe to improve performance
  - Unified management and programming interface

SCSI Express Components



- The SCSI storage command set
- Packages SCSI for a PQI queuing layer
- Flexible, high-performance queuing layer
- Accommodates PCIe, SAS, and SATA drives
- Leading server I/O interconnect

SCSI Express Value Proposition

Performance and Innovation

- Increased performance through lower latency for emerging advanced technologies
- Enables new storage architectures

Reliability

- Proven enterprise SCSI ecosystem
- Architected for nonstop availability

Investment Protection

- Coexistence with SAS via Express Bay and common command set
- Leveraging robust middleware ecosystem

SCSI Express Hardware/Software

SCSI Express Controllers

- Supports SOP-PQI driver functionality on the controller to the target device on the PCIe lanes

SCSI Express Drive/Device

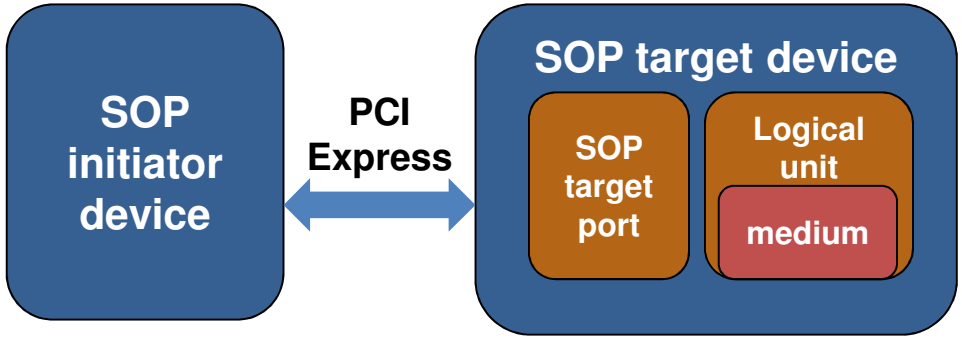
- SOP-PQI protocol
- Connects to SFF-8639
- PCIe up to x4 interface

SCSI Express Driver

- Driver supplied by storage OEMs, IHVs or OSVs
- Open Source Linux driver and IHV drivers available beginning in Q1CY13

Simple Express Devices

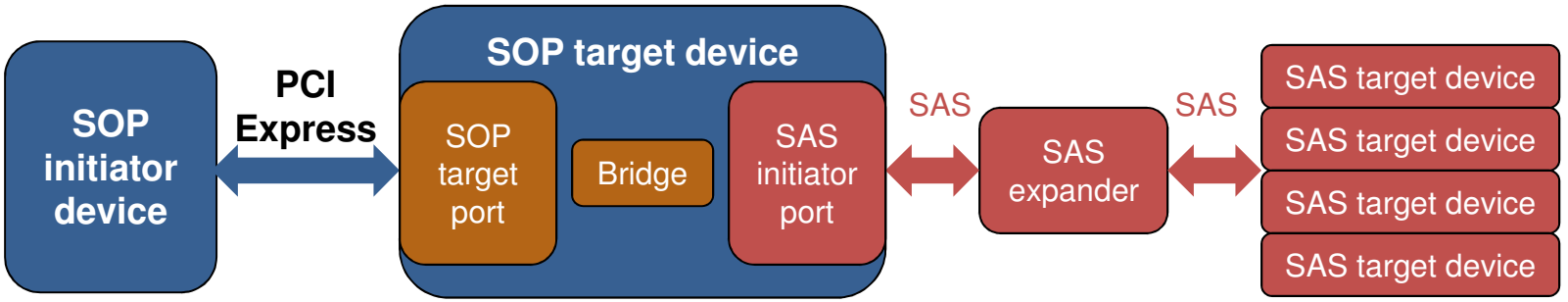
- SSDs, etc.
  - Usually just a single logical unit with LUN 0
  - Any SCSI device type is possible
  - SSD, tape drive, optical drive (CD/DVD/BluRay), etc.



SCSI Express Bridges

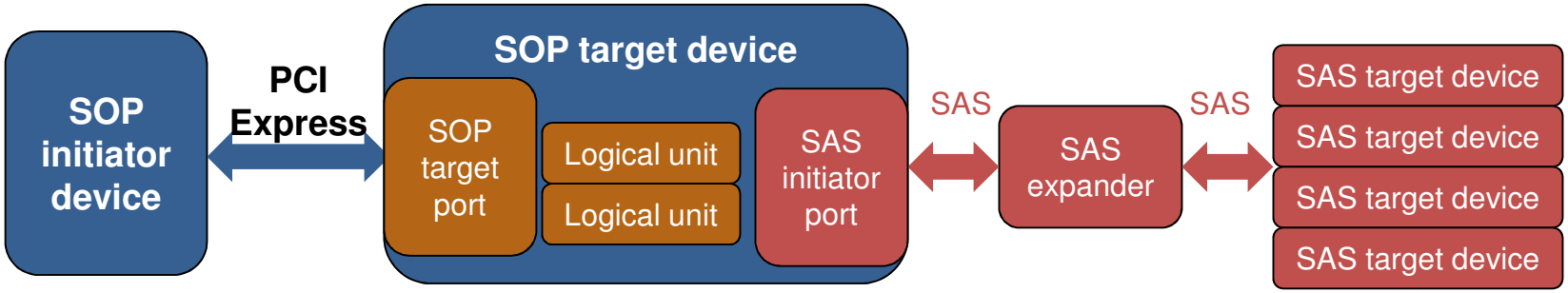
- Host Bus Adapter (HBA)
  - Bridges from PCI Express to another interconnect supporting SCSI
  - Maps SCSI target devices one-for-one
  - Usually referred to only by the back-end interconnect e.g. “SAS HBA”
  - Manage with SOP bridge management functions

Interconnect	SCSI transport protocol
Serial Attached SCSI (SAS)	Serial SCSI Protocol (SSP)
Fibre Channel (FC)	Fibre Channel Protocol (FCP)
Ethernet	Internet SCSI (iSCSI)
Universal Serial Bus (USB)	USB Attached SCSI (UAS)
InfiniBand	SCSI RDMA Protocol (SRP)
PCI Express	SCSI over PCI Express (SOP)



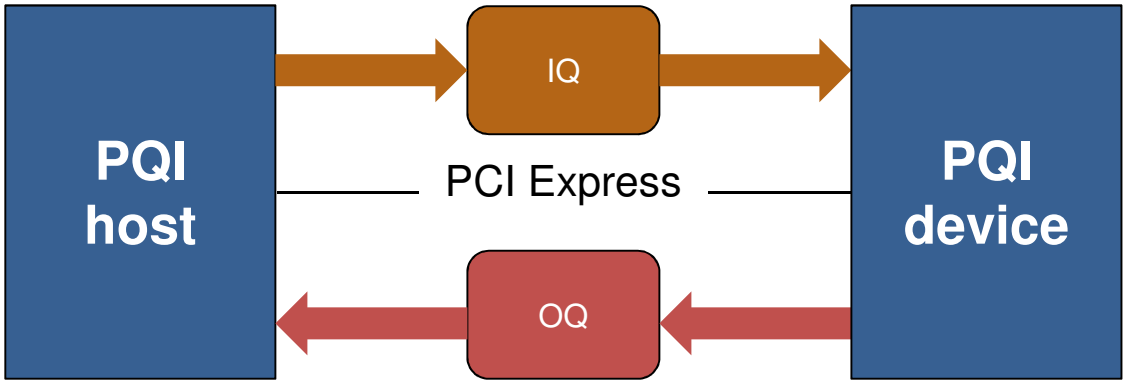
SCSI Express Controllers

- RAID Controllers
  - Indirectly bridges from PCI Express to another interconnect supporting SCSI
  - Not a one-to-one mapping of SCSI target devices
  - Presents logical drives over PCI Express created from physical drive
  - Manage with standard SCSI commands
  - REPORT LUNS reports the logical units that have been created



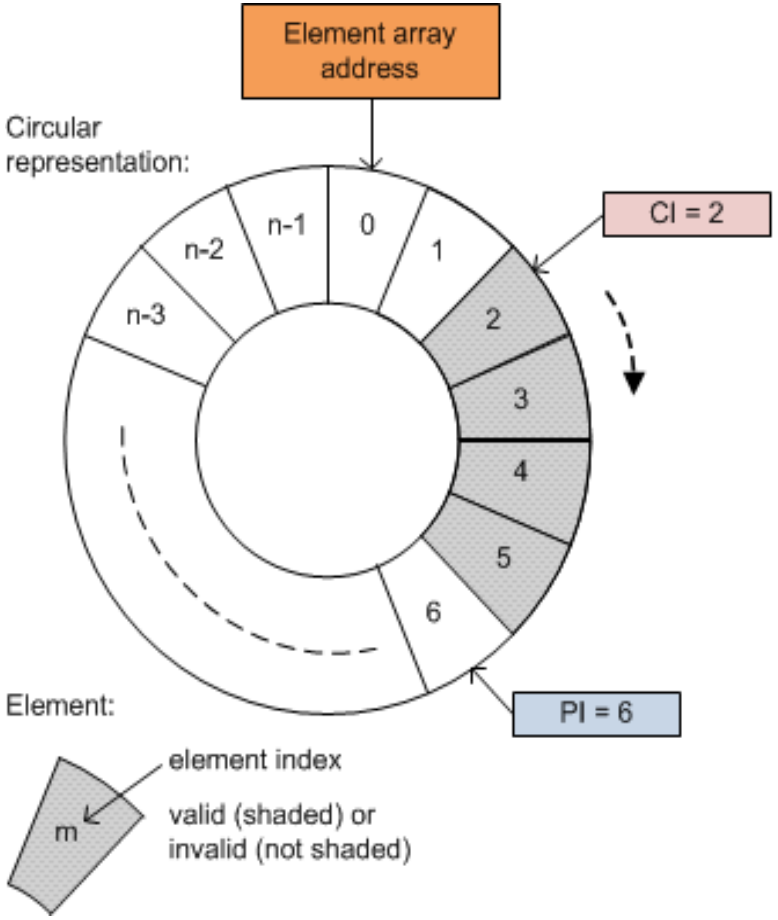
Inbound Queues and Outbound Queues

- SOP expects a queuing layer over PCI Express (PQI) to define
  - Inbound Queues (IQs) to transfer IUs from SOP initiator port to SOP target port
  - Outbound Queues (OQs)to transfer IUs from SOP target port to SOP initiator port



Circular Queue Basics

- Element array
  - Fixed size elements (e.g., 64 bytes)
- Producer index (PI)
  - Location to which producer writes elements
  - Write to element array [PI++]
  - Wrap at size of the element array
- Consumer index (CI)
  - Location from which consumer reads elements
  - Read from element array [CI++]
  - Wrap at size of the element array



Queue Types

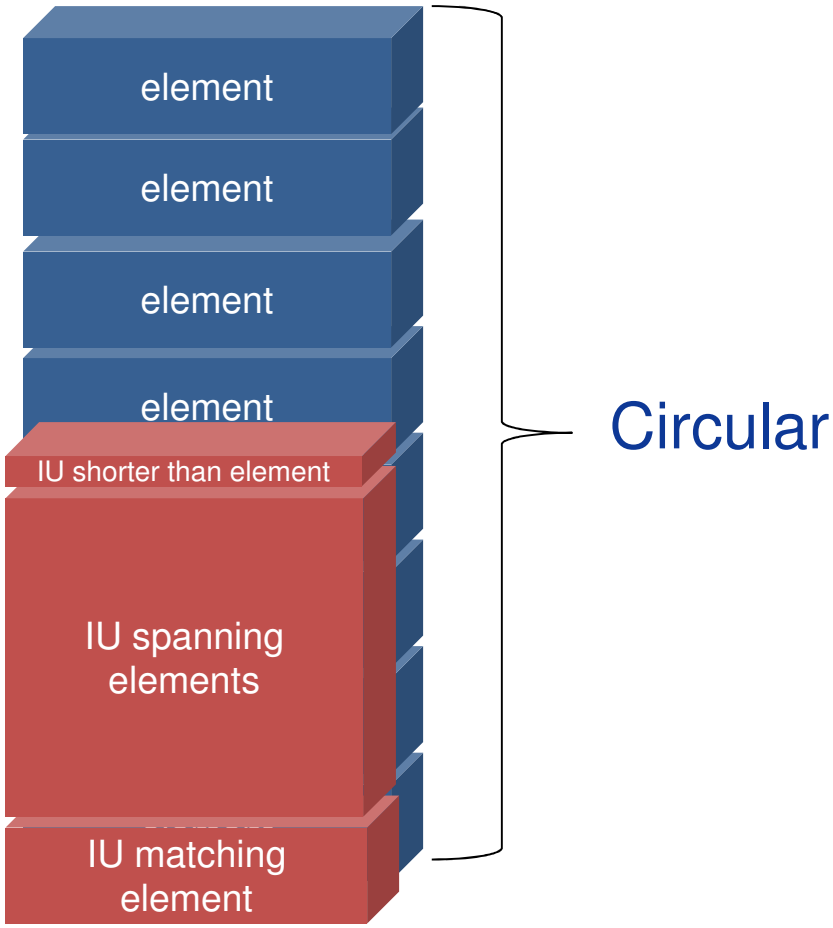
- Administrator queues
  - Created via PQI device registers
    - Located in PQI device memory space
  - Single administrator IQ and administrator OQ
  - IUs defined by PQI
- Operational queues
  - Created via PQI administrator functions
    - Delivered over the administrator queues
  - Any number of operational IQs and operational OQs
    - Not in pairs
    - Can be specific to different cores
  - IUs defined by the information unit layer standard
    - e.g., SCSI over PCI Express (SOP)

Information Unit (IU) Types

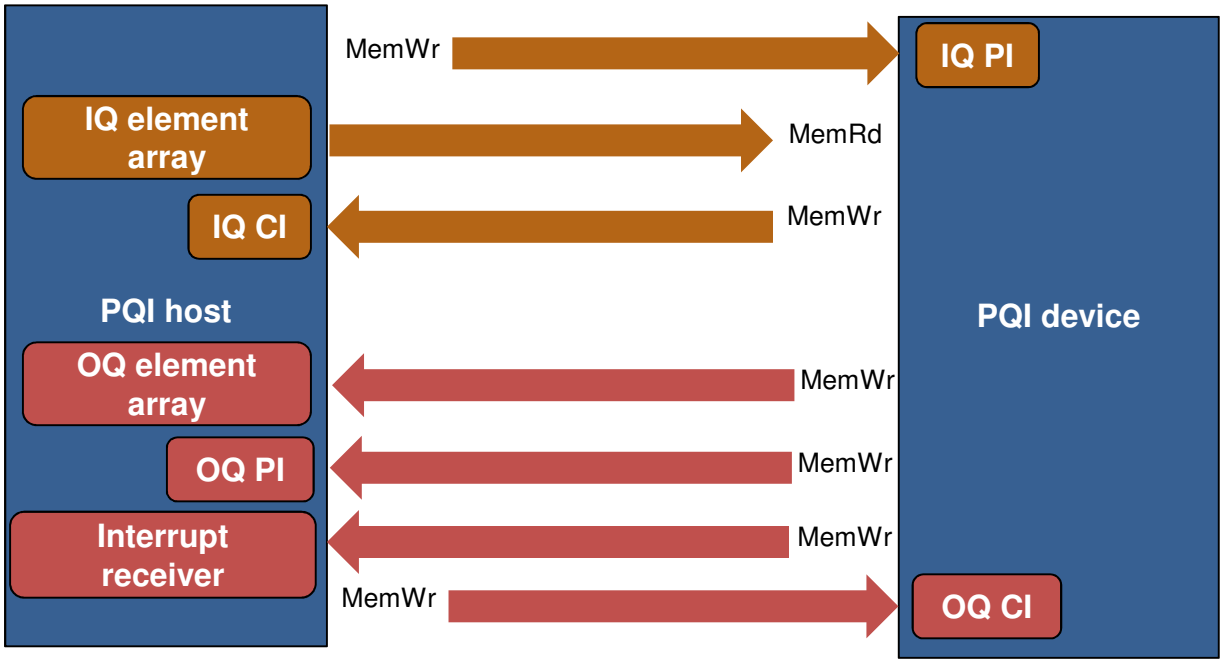
- SCSI Request/Response IUs
  Commands, Task Management, Success, Command Response, Task Management Response, etc.
- General Management Request/Response IUs
  Report General, Report Configuration, Set Configuration, Report Event Configuration, Management Response, Event, Event Acknowledge
- Bridge Management Request/Response IUs
- Administrator Request/Response IUs
- Other
  Null IU, etc.

IUs and Queues

- IU smaller, equal, or larger than queue element



Avoiding Downstream PCI Express Memory Reads



KEY PQI Features

- NUMA support
  - Directed MSI-X
  - Flexible queue organization
- Interrupt Coalescing
  - Single interrupt for multiple queue entries
  - Tuning via; count, min and max times
- Queue Element Spanning
- Scatter Gather Lists (SGL)
  - Describes a data buffer and how it is distributed across non contiguous chunks of memory
  - Can be embedded in IUs or a separate list
  - Widely supported method across multiple OSs

Drivers

- Open Source Linux driver
  - *Https://github.com/HPSmartStorage/scsi-over-pcie*
  - Block driver in development
    - Bypasses native Linux SCSI stack for maximum performance
    - Performance enhancements have not started
  - Low level SCSI driver is dormant
    - SCSI path enhancements proposed (2013 Kernel Summit)
- Windows
  - Proprietary Storport drivers available today