

Serial Attached SCSI

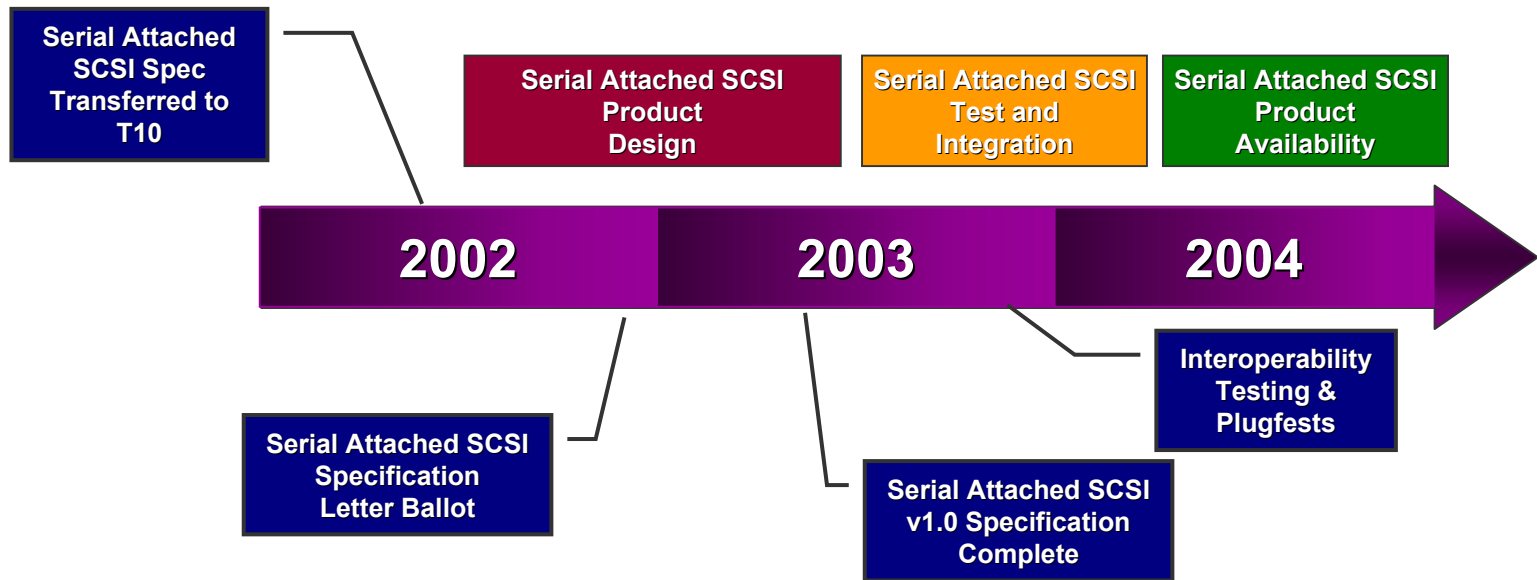
Serial Attached SCSI Logo Copyright ©2003 STA

Marty Czekalski

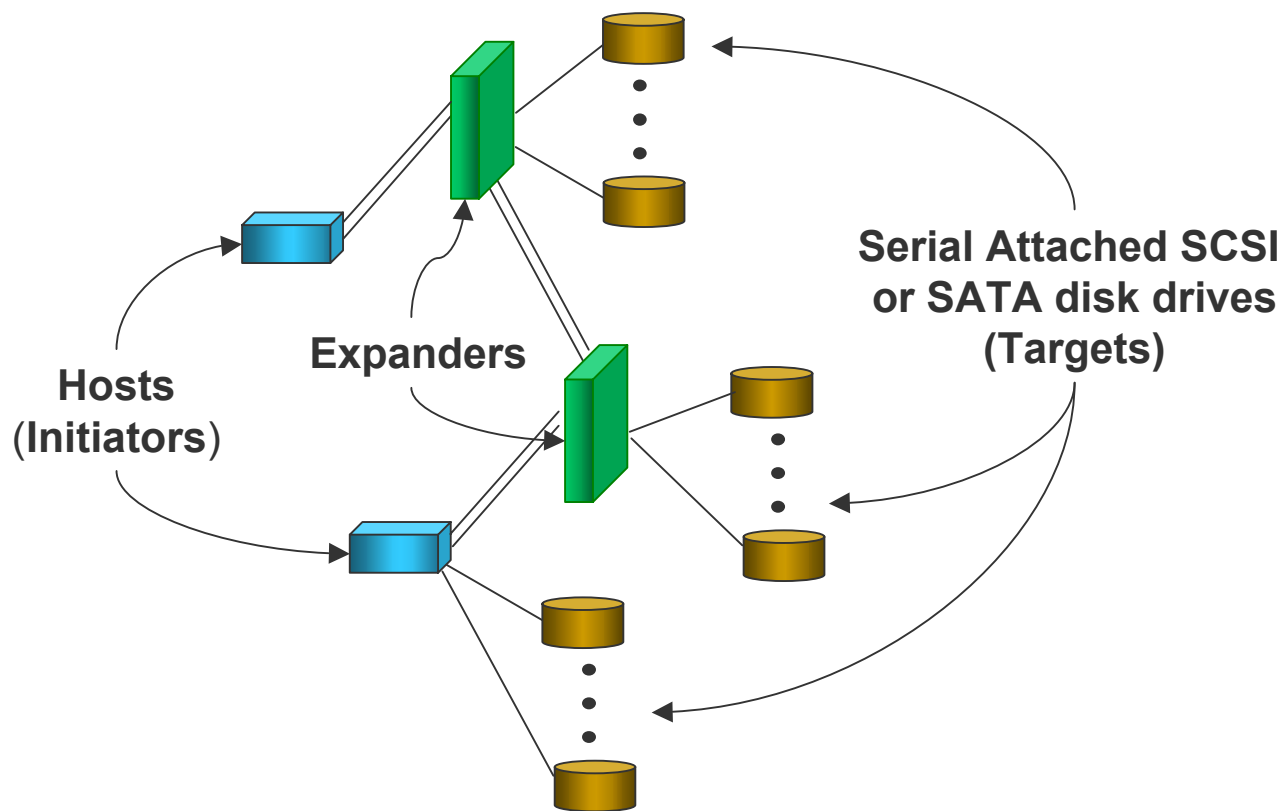
Vice President, Director – SCSI Trade Association

Technical Marketing Manager – Maxtor Corporation

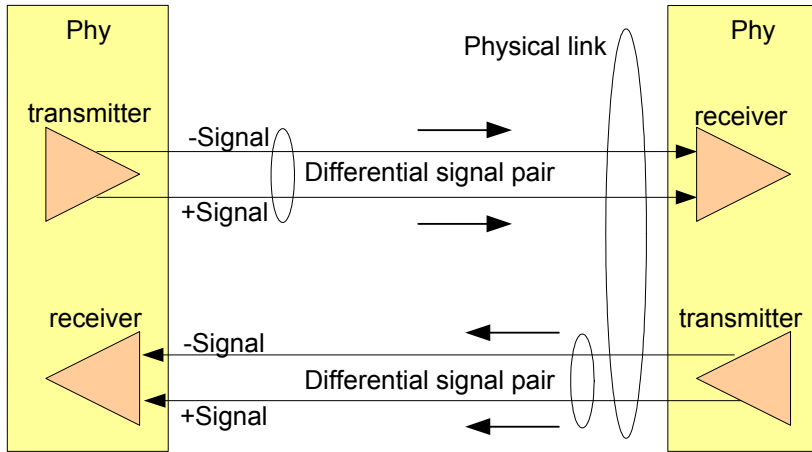
- Serial Attached SCSI is the next evolution of SCSI beyond U320
- SAS has been designed to be a device and near cabinet interface only, not a network interface
- Leverages an enhanced Serial ATA PHY while adding support for a second port
- Serial SCSI utilizes features of Fibre Channel AND compatibility with SATA drives in a point-to-point, switched architecture
 - Up to *16k physical links per domain* in a switched, point-to-point configuration for scalable performance.
 - Scalability with wide ports and future speed increases
 - Data frame based on FCP
 - Includes *rate matching* to optimize device transfer speeds
 - World wide name addressing
 - Uses a *simplified protocol* that will minimize interoperability issues between controllers & drives
 - Dual port drives for high availability



- Twenty-six companies have come together since August 2001 to develop a new industry standard addressing customer's enterprise needs
- SCSI Trade Association has assumed ownership of SAS marketing.
- Specification transferred to T10 in May 2002
- Specification sent to letter ballot on Nov. 22, 2002.
 - Letter ballot closed and approved with comments on Dec. 22, 2002
 - Editorial Board working to resolve comments.

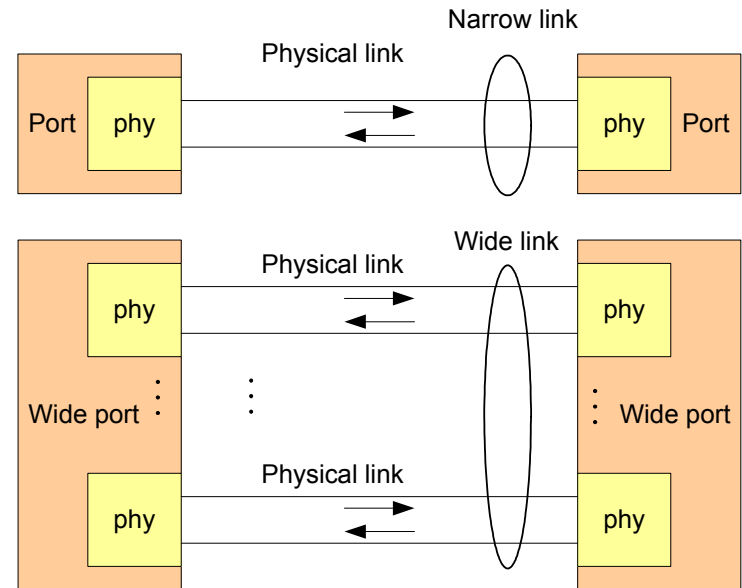


	Serial ATA	•Serial Attached •SCSI	Fibre Channel AL
Performance	Half-duplex	•Full-duplex with •Link Aggregation •(Wide ports)	Full Duplex
	•1.5 Gb/sec •(3.0 Gb/s announced)	3.0 Gb/sec(at intro.) (6.0 Gb/s planned)	2.0 Gb/sec (4Gb/s announced)
Connectivity	•1m internal cable	>6m external cable	>15m external cable
	•One device •(fan-out devices •demonstrated)	>128 devices Expanders (16k Phys. max)	127 devices Loop or loop switch
	SATA only	SAS and SATA	Fibre Channel only
Availability	•Single-port HDDs	Dual-port HDDs	Dual-port HDDs
	Single-host Point to point	Multi-initiator Point to point	Multi-initiator Shared media or point to point
•Driver •Model	•Software transparent •with Parallel ATA	•Software transparent •with SCSI	•Software transparent •with SCSI

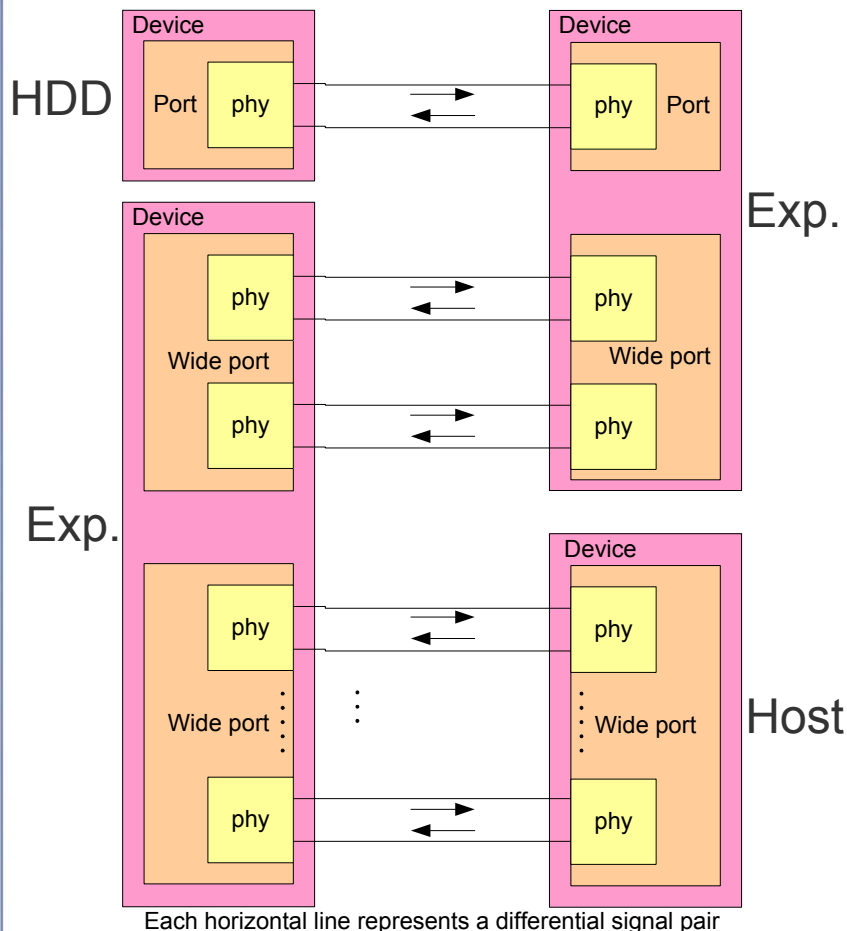


Links are full duplex differential pair connections

One or more links can be configured to be a port

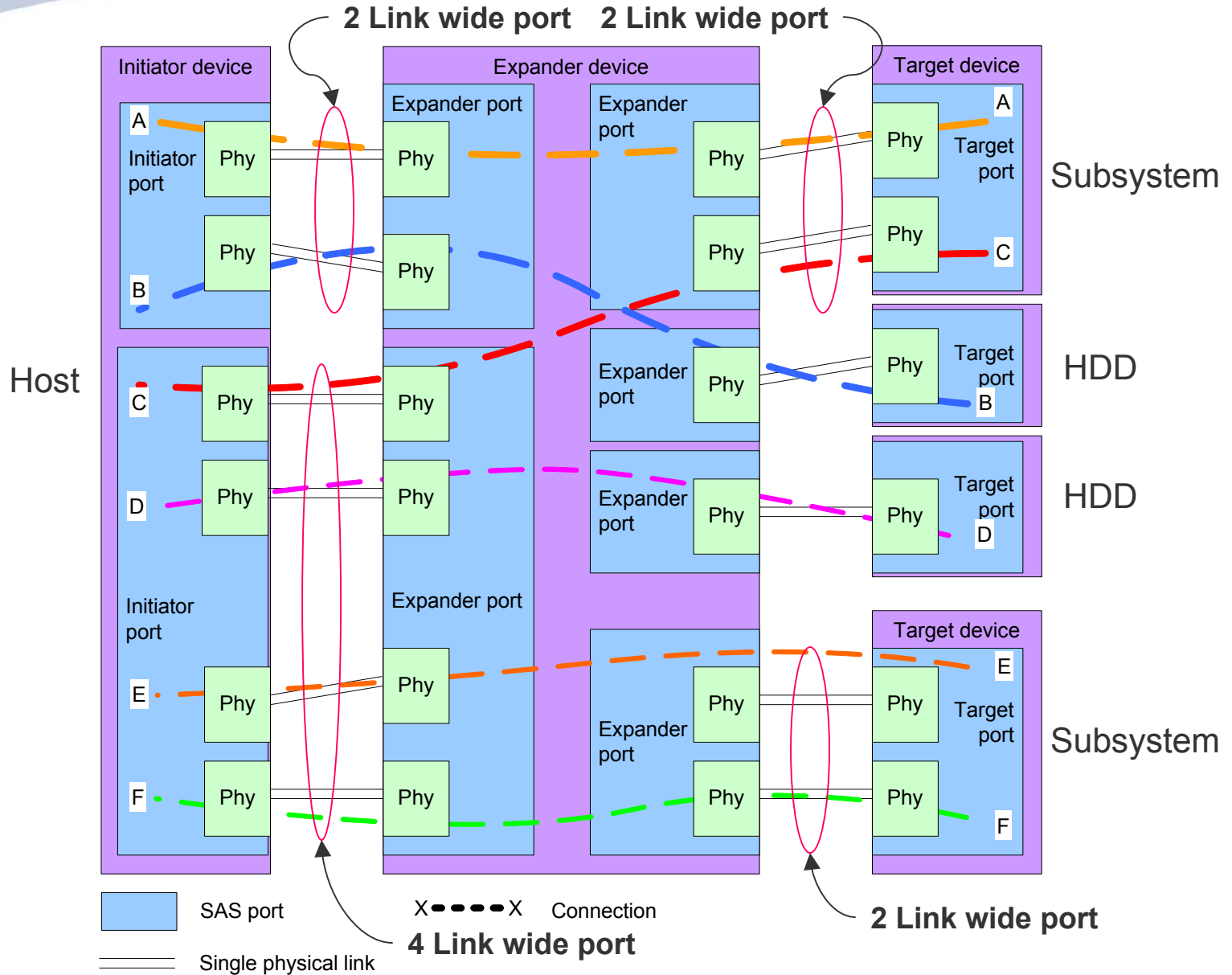


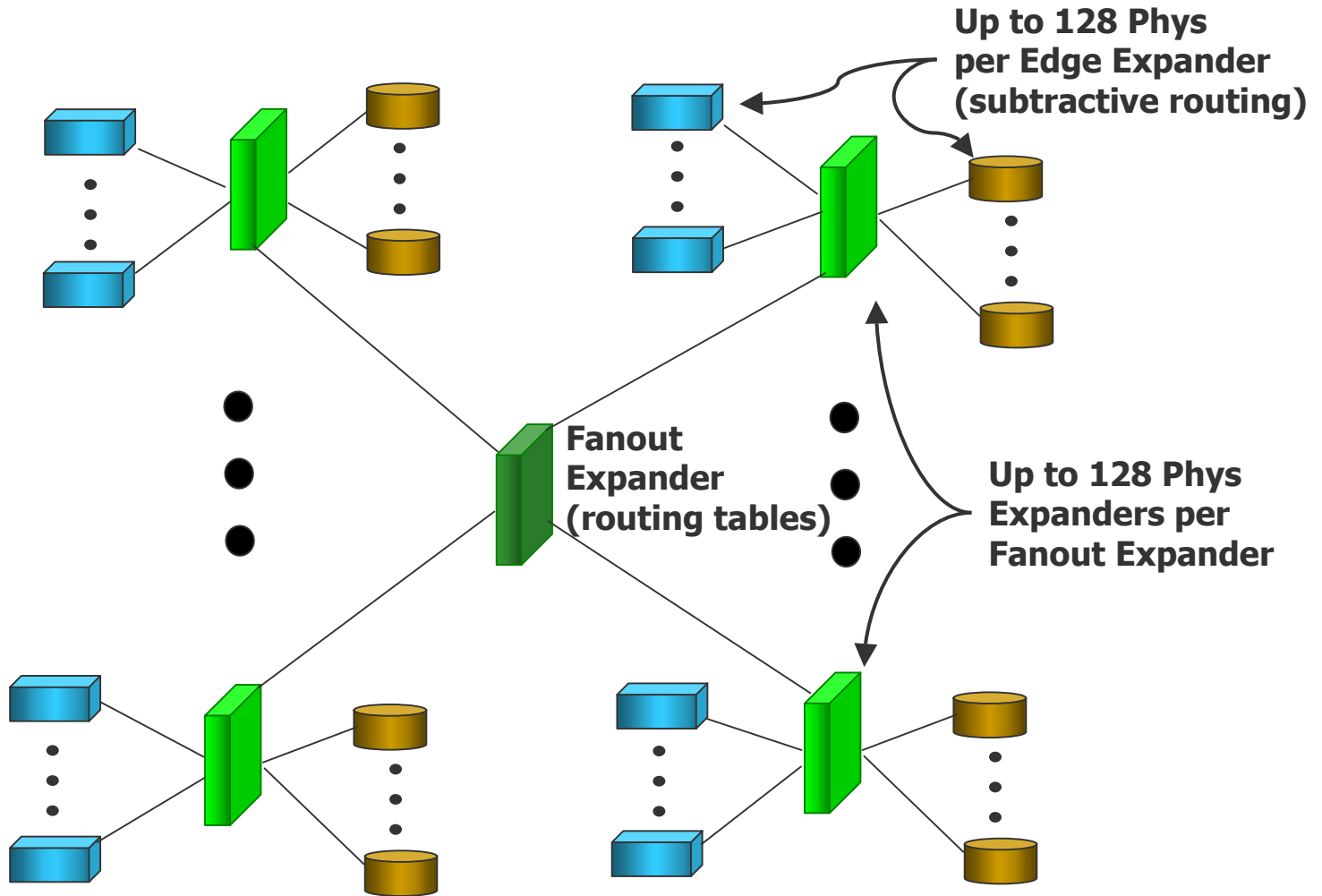
Each horizontal line represents a differential signal pair



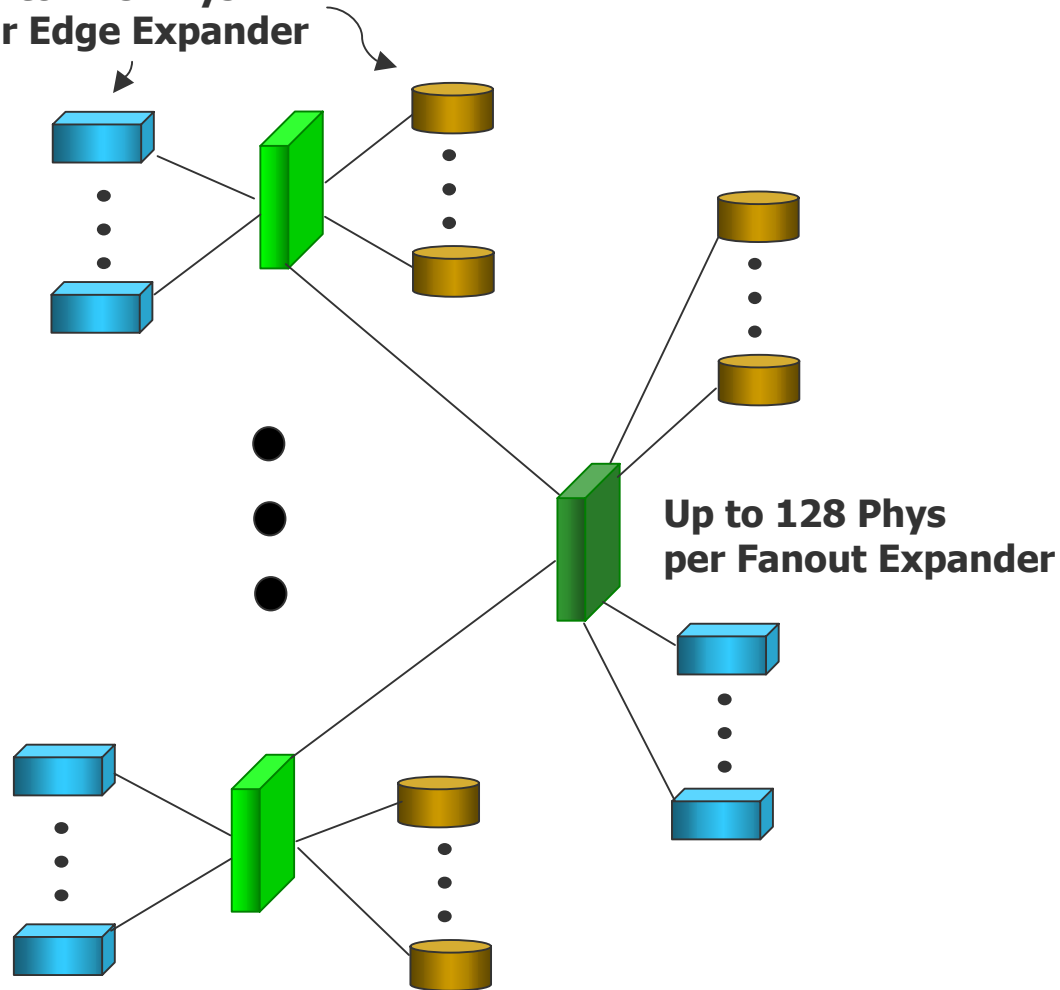
- Devices can have one or more ports, each configured as narrow or wide ports
- Initiator to expander links may be composed of multiple links
- Aggregates bandwidth
- One initiator address shared by all the links
- Command sent down one link; data may be returned on another (in a separate connection)

Multiple Pathways Allow Concurrent Operation

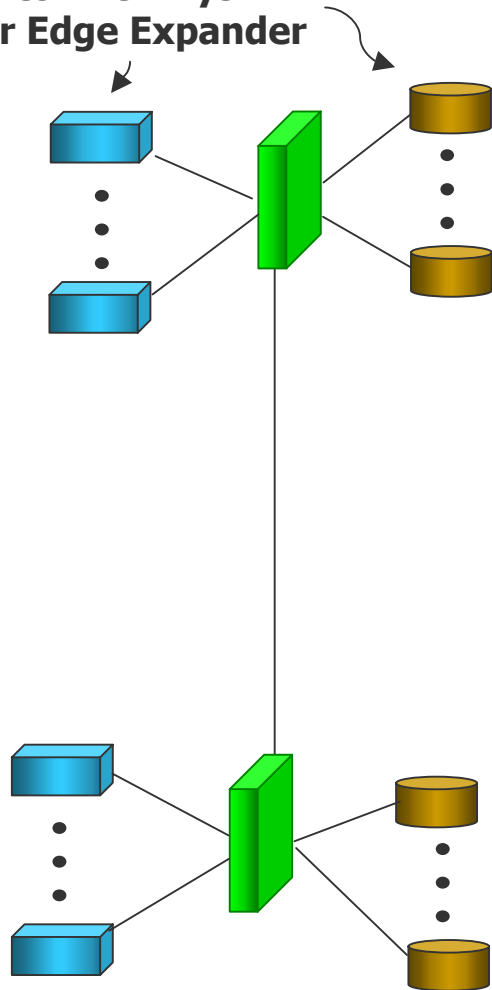




**Up to 128 Phys
per Edge Expander**

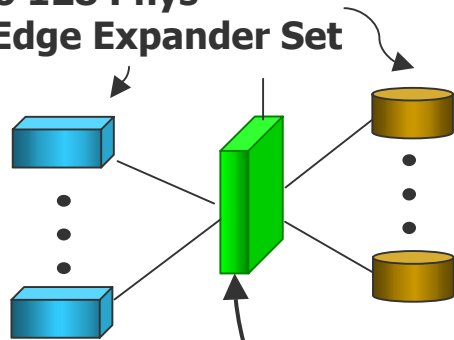


Up to 128 Phys
per Edge Expander



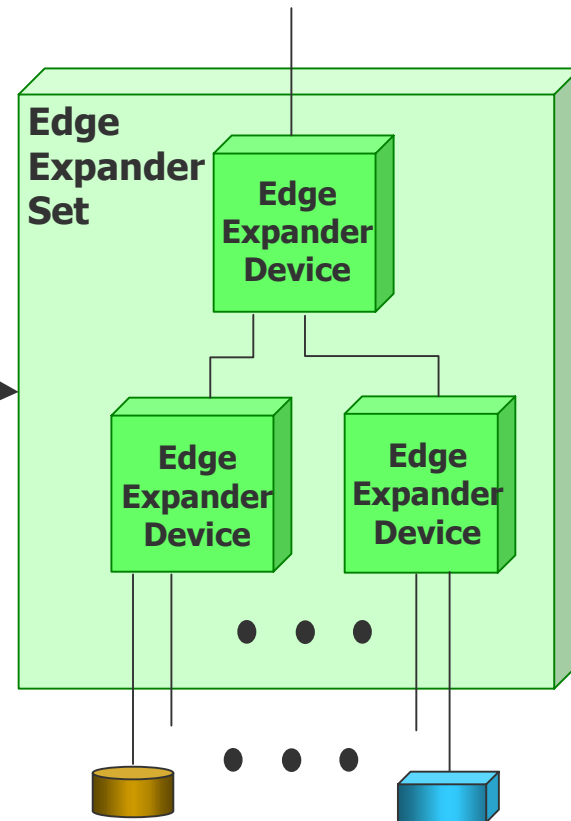
A maximum of two
Edge Expanders can be
connected without the
use of a Fanout
Expander

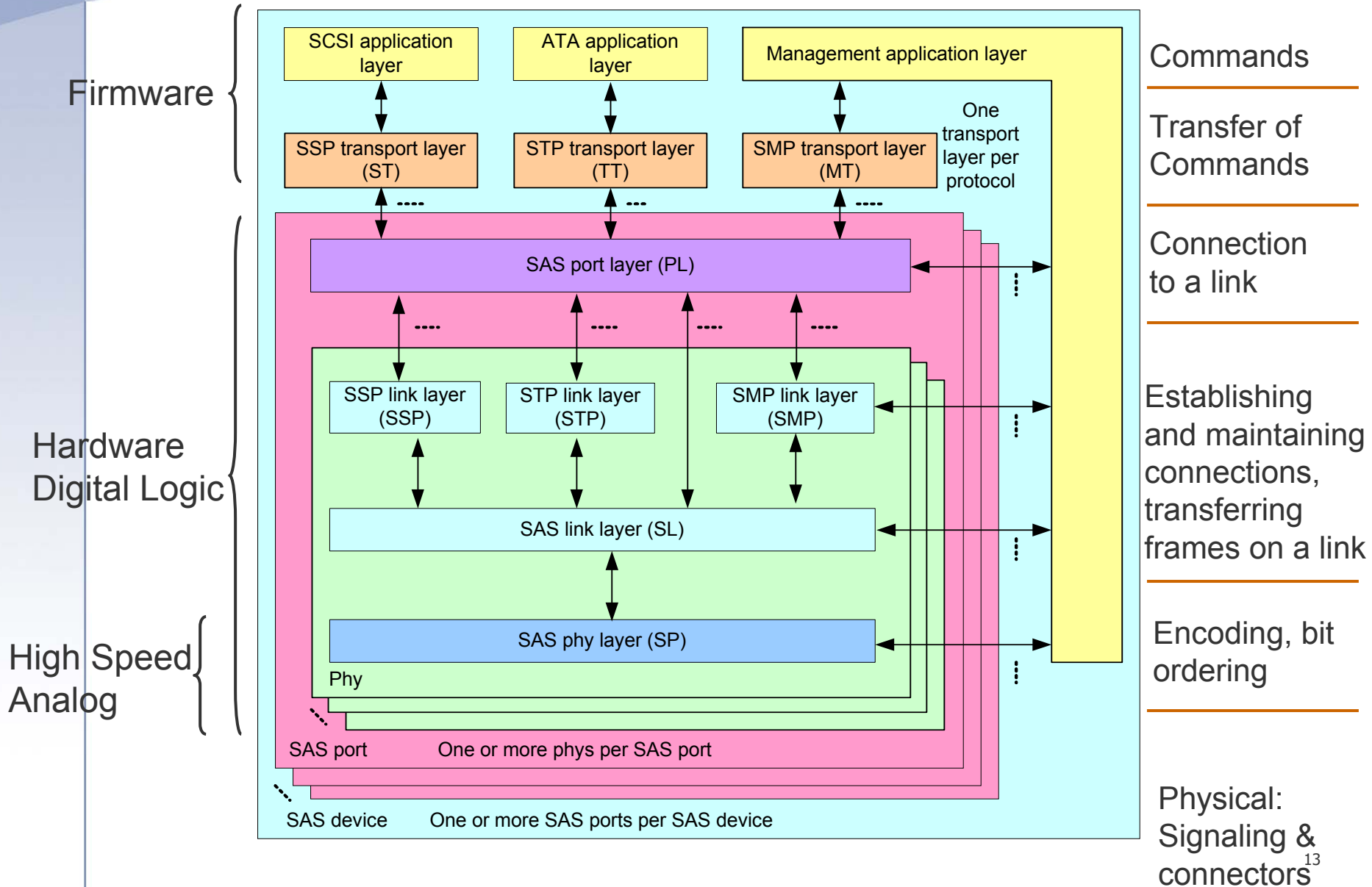
Up to 128 Phys
Per Edge Expander Set



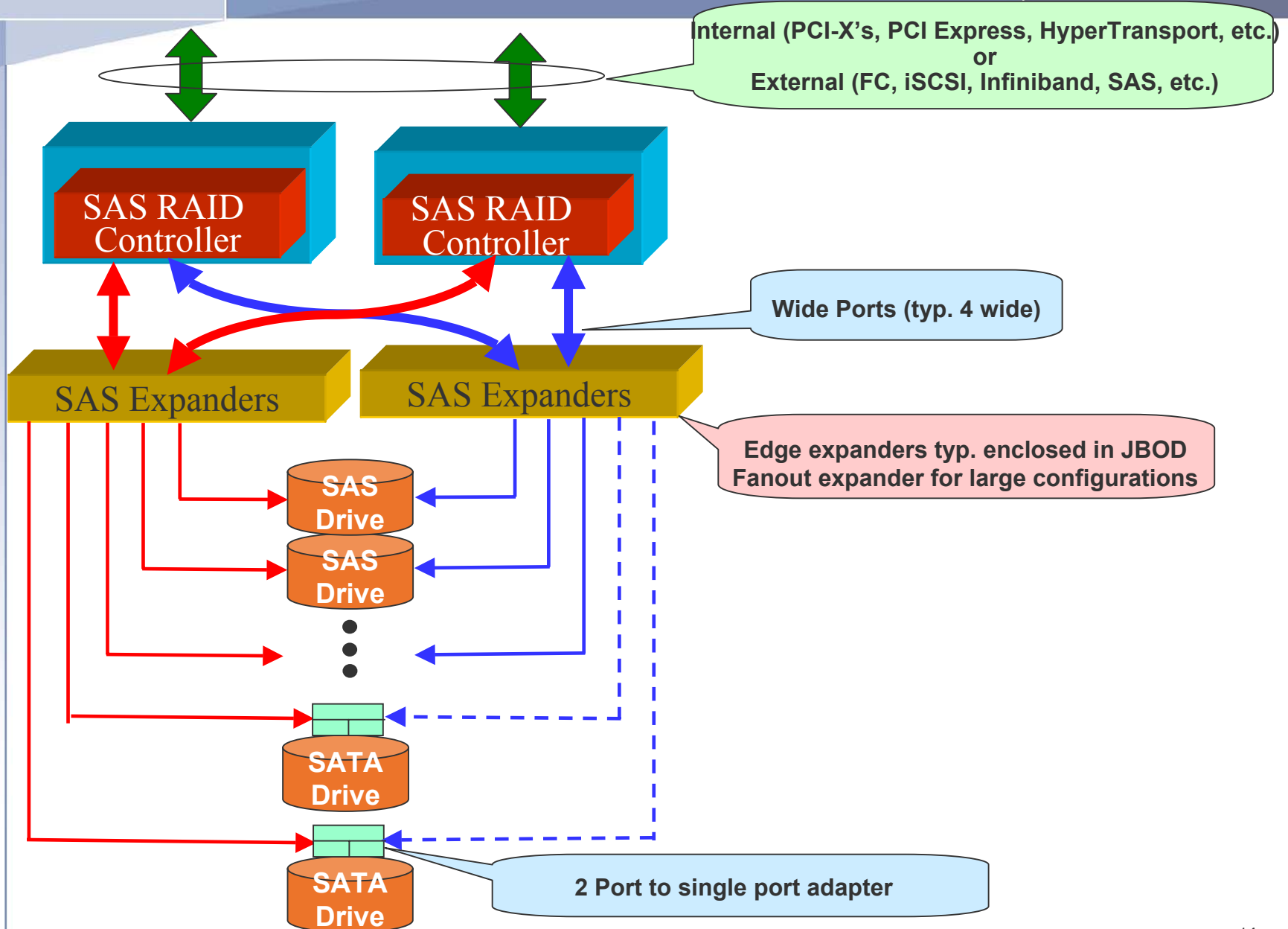
Initial implementations are expected to support 40-80 devices per “Edge Expander Set”

To facilitate cost effective implementations, Edge Expanders can be comprised of multiple devices in an ‘Edge Expander Set’





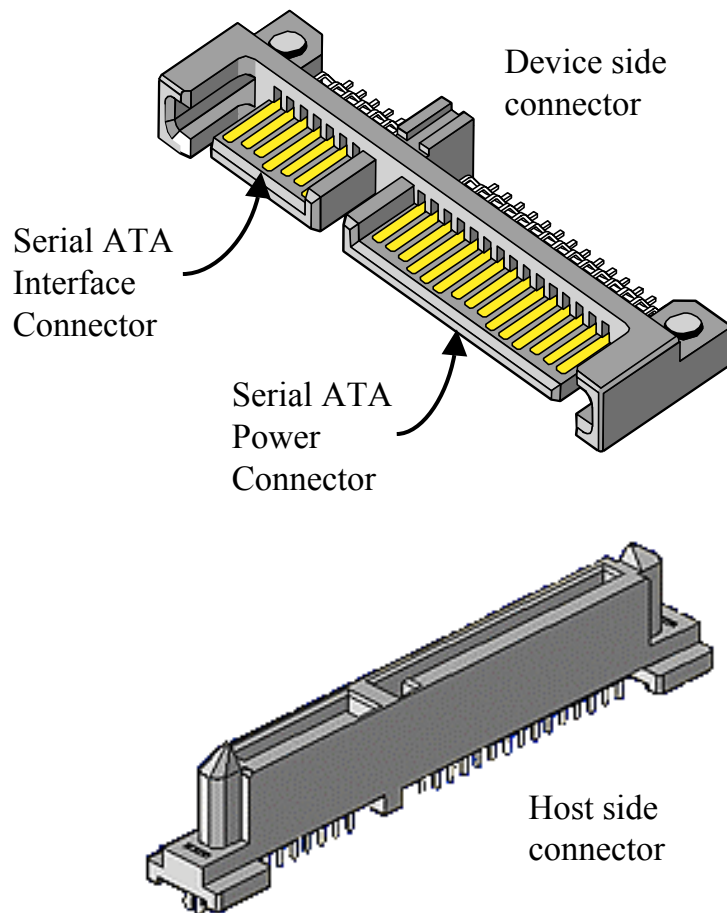
Typical SAS Configuration



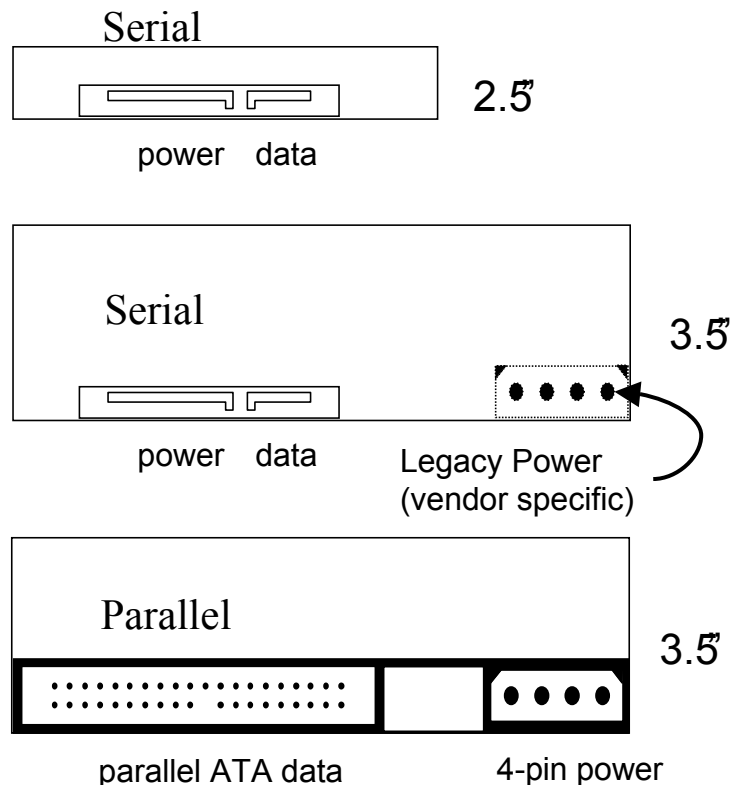
- Phy
 - Cables, Connectors, and Backplanes
 - Link rates
 - Out of Band signaling
 - Encoding
- Link
 - Primitives
 - Scrambling
 - Wide and narrow links
 - Connections
 - Rate matching
- Protocols and Expanders
 - SMP (Expander Management)
 - SSP (SAS)
 - STP (SATA)

Appearance of Serial ATA Connectors

(Drawing courtesy of Molex)



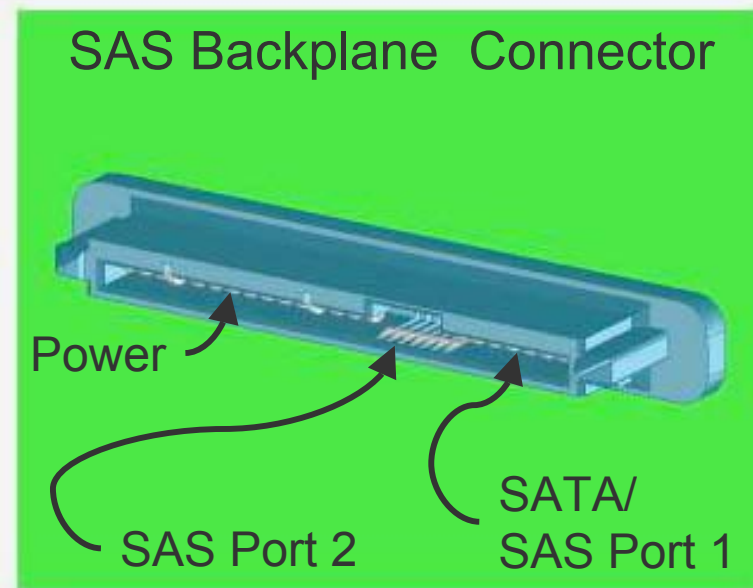
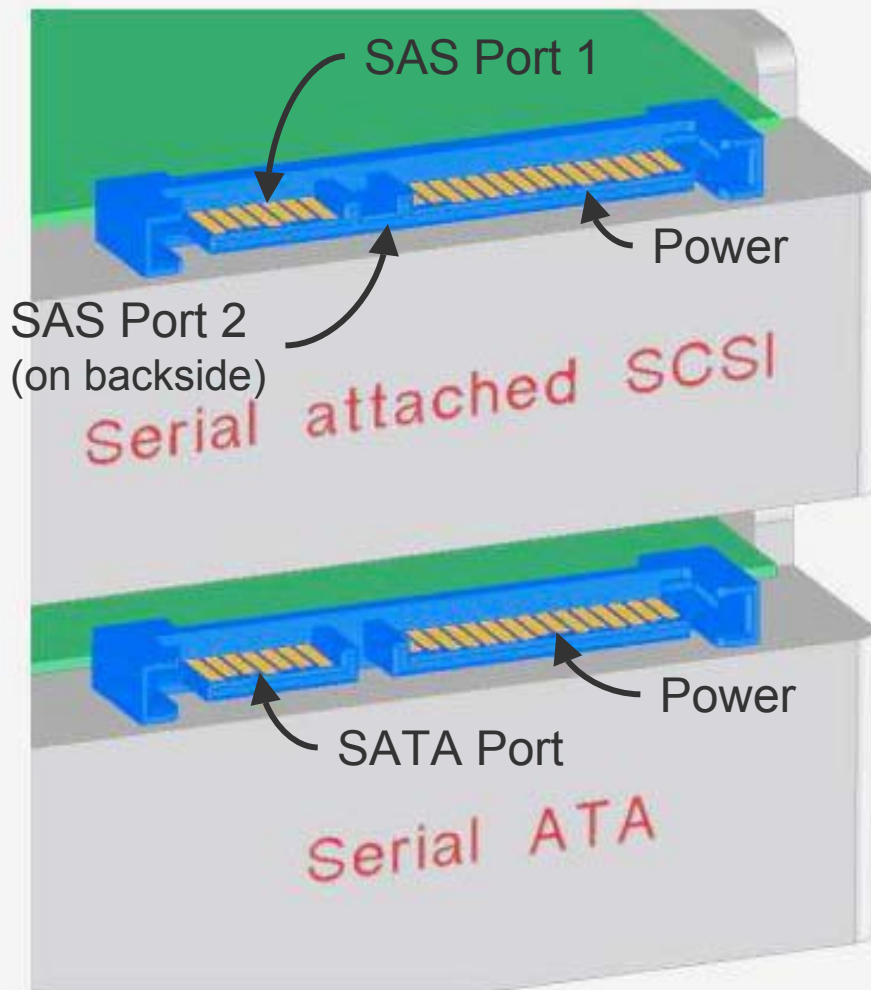
Device Connector Sizes and Locations



Note: SATA defines the exact connector location on a drive, AND, the connector location is identical between 2.5" & 3.5" HDDs

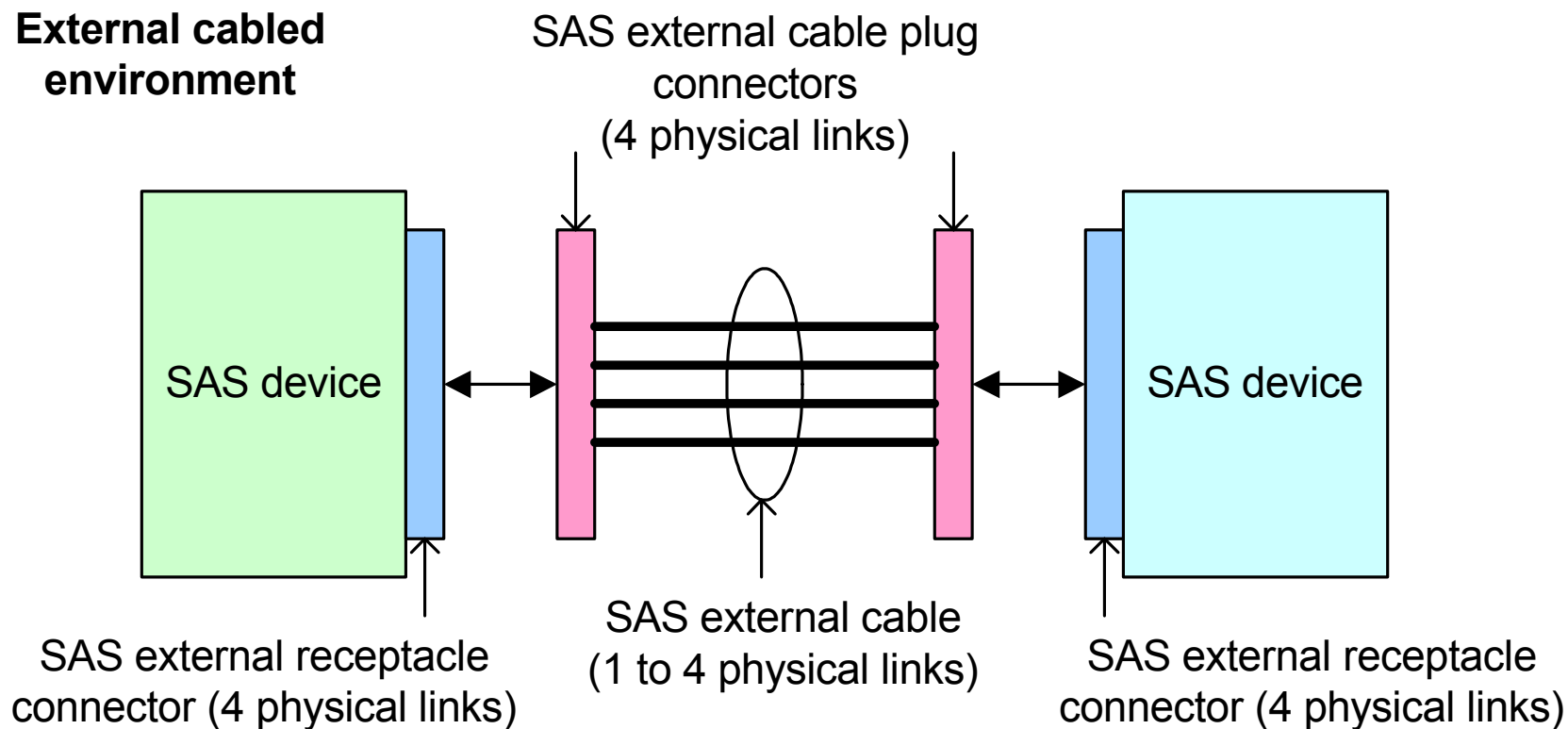
SAS Backplane Connector accepts both SATA and SAS drives

(Drawing courtesy of Molex)



Note: SATA backplane connectors Will NOT accept SAS drives

External cabled environment

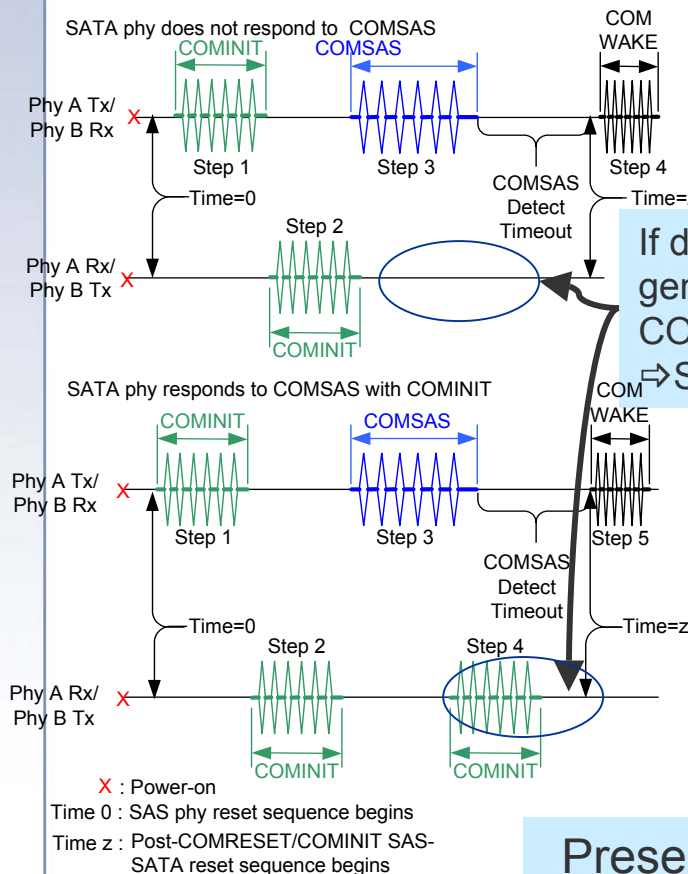


(SAS external cable connects Tx to Rx on each physical link)

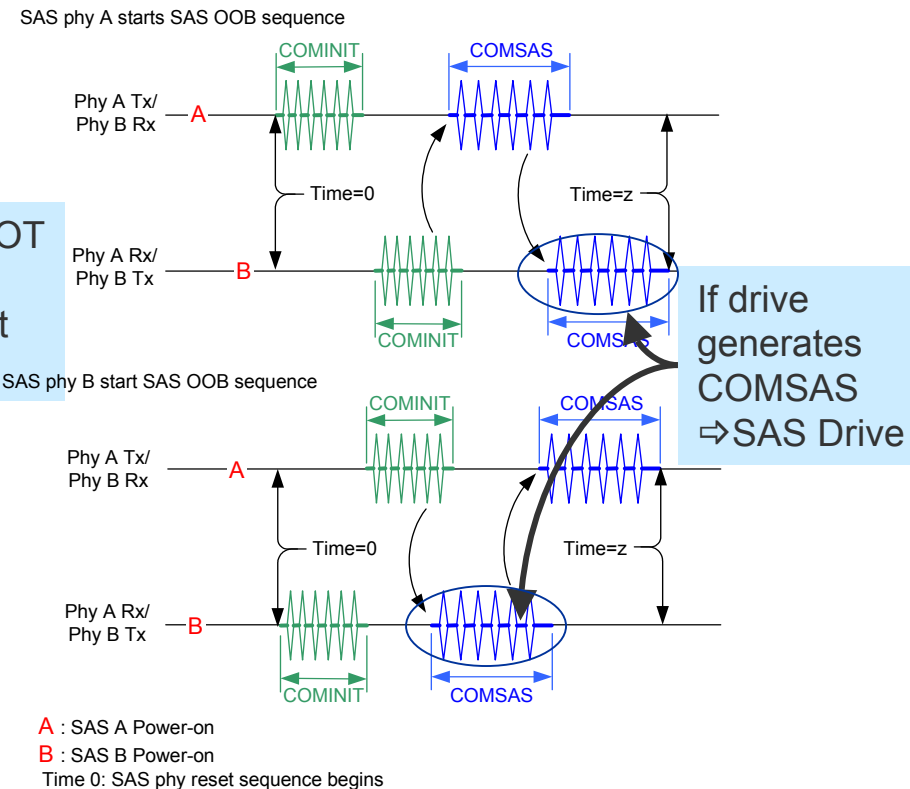
- SAS will be introduced at 3Gbit/sec
 - Devices will also support 1.5Gbit/sec
- SATA rates supported
 - 1.5 Gbit/sec
 - 3.0 Gbit/sec (targeted)
- Auto-detect link rate during ALIGNs after COMWAKE

- SATA out of band (OOB) special patterns (“signals”)
 - Signals are sent after power-up to initialize the link
 - Signal is a burst of ALIGN primitives, then idle time; repeated 6 times
 - Detected by squelch detector and frequency comparators
- SATA’s COMRESET, COMINIT, and COMWAKE signals are unchanged
- COMSAS signal added
 - Inserted after calibration sequence before COMWAKE
 - Initiator: SAS (SSP and/or STP)
 - Target: SAS protocol (SSP)
 - Expander: always asserted (not present on SATA initiator-target)
- If both sides assert COMSAS, then the link is a SAS link rather than a SATA link

- Out of band signals are slow speed bursts used to initialize a link



If drive does NOT generate a COMSAS burst ⇒ SATA Drive



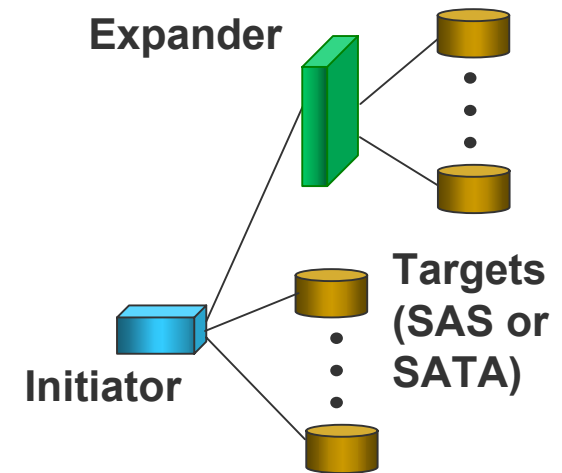
Presence or absence of COMSAS OOB signal easily identifies a HDD as SATA or SAS

- 8b10b coding
 - As used in SATA, Fibre Channel, et al.
 - ↪ Provides embedded clocking and DC balance
 - ↪ Once a link is established, characters are continuously transmitted
 - **Character = 10 bits** as transmitted on the wire
 - Control characters Kxx.y - special uses
 - Data characters Dxx.y - represent 8 bit data bytes
 - Running disparity
- Dword = 4 characters
 - Everything in SAS is based on dwords
- Primitive = dword starting with a control character

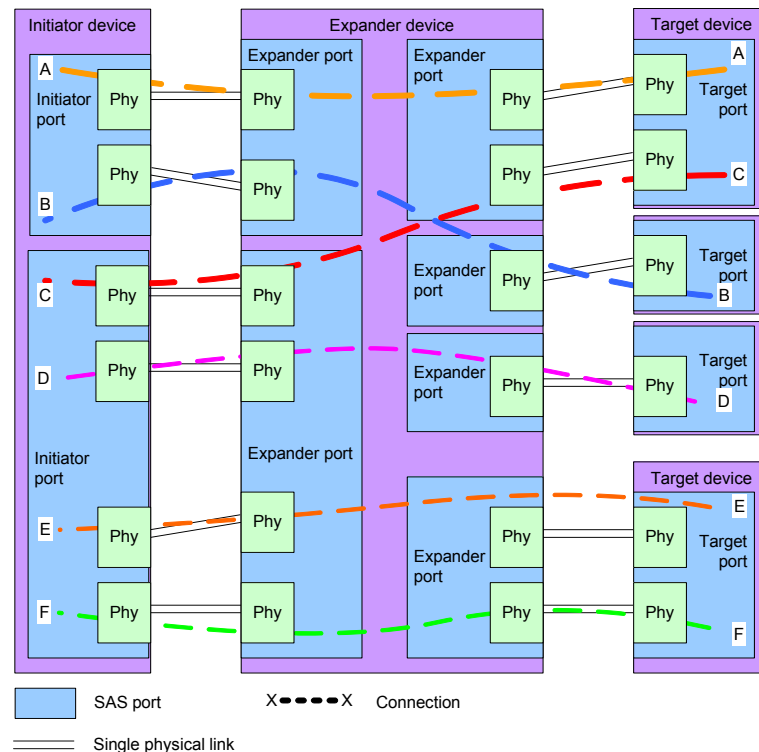
- Primitive is a dword starting with a control character
- Primitives have no endianness; just first, second, third, and last bytes
- ALIGN starts with K28.5
- All other SATA primitives start with K28.3
- All SAS primitives start with K28.5
- Primitives may start/end with any disparity

- Scrambling tries to randomize data
 - XOR data with the contents of a linear feedback shift register at both sender and receiver
 - changes constant 000000... and 111111... patterns into pseudo-random patterns of 1s and 0s
 - Constant patterns occur more often than other patterns, including the worst case pattern that undoes the scrambling effect
- Reduces EMI peaks
 - Spread spectrum clocking addresses EMI for all patterns

- All I_T communication occurs within an SSP, SMP or STP connection
- Establishing connection through an expander involves arbitration
- OPEN address frame to make connection request
- Open timeout timer
- OPEN_ACCEPT means connection is active

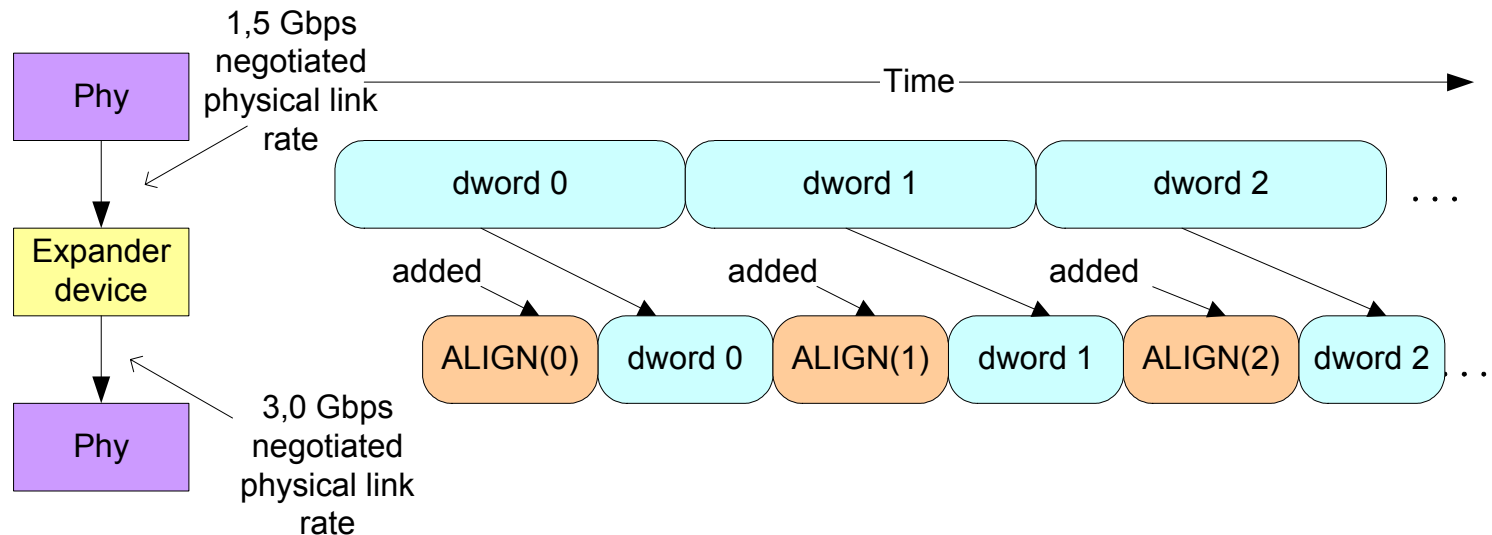


- Responses to open request
 - Arbitration in progress - AIP
 - ↪ reset open timeout timer and keep waiting
 - Cross on wire - OPEN address frame
 - ↪ Arbitration fairness dictates who wins
 - Accepted - OPEN_ACCEPT
 - Rejected - OPEN_REJECT
 - ↪ Numerous reasons - Retry, bad protocol, deadlock avoidance, etc.
 - Cancel - BREAK
 - No response - timeout and send BREAK

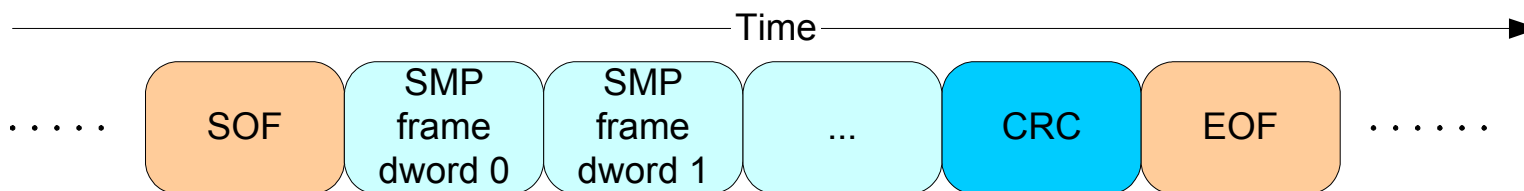


- Least recently used arbitration fairness
 - Arbitration wait timer (in OPEN)
 - ↪ Source maintained
 - ↪ Expander increments and forwards as part of OPEN
 - ↪ Reset rules and crossing requests
- Deadlock Prevention
 - Partial Pathway timer
 - Pathway Recovery
 - ↪ Priority based on Pathway Blocked Count
 - ↪ Equal counts resolved using SAS address
- Scales well to large configurations

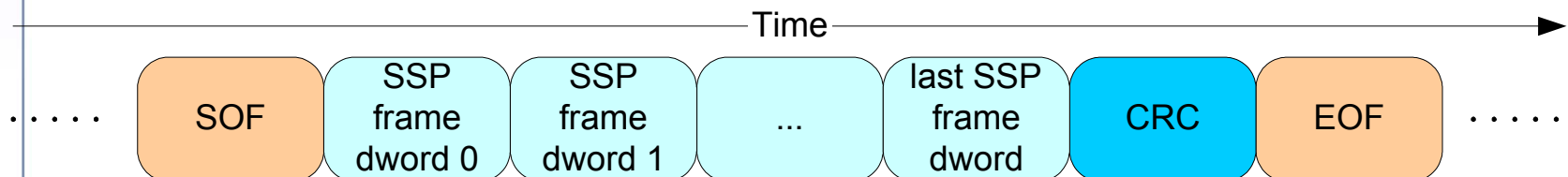
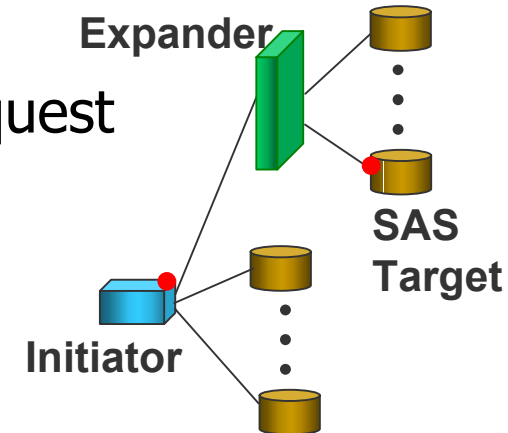
- When initiator port and target port are separated by an expander, their link rates may differ
 - E.g. Initiator to expander 3 Gbit/sec; expander to target 1.5 Gbit/sec
- Solution: insert ALIGNs on the faster links
 - E.g. every other dword is used

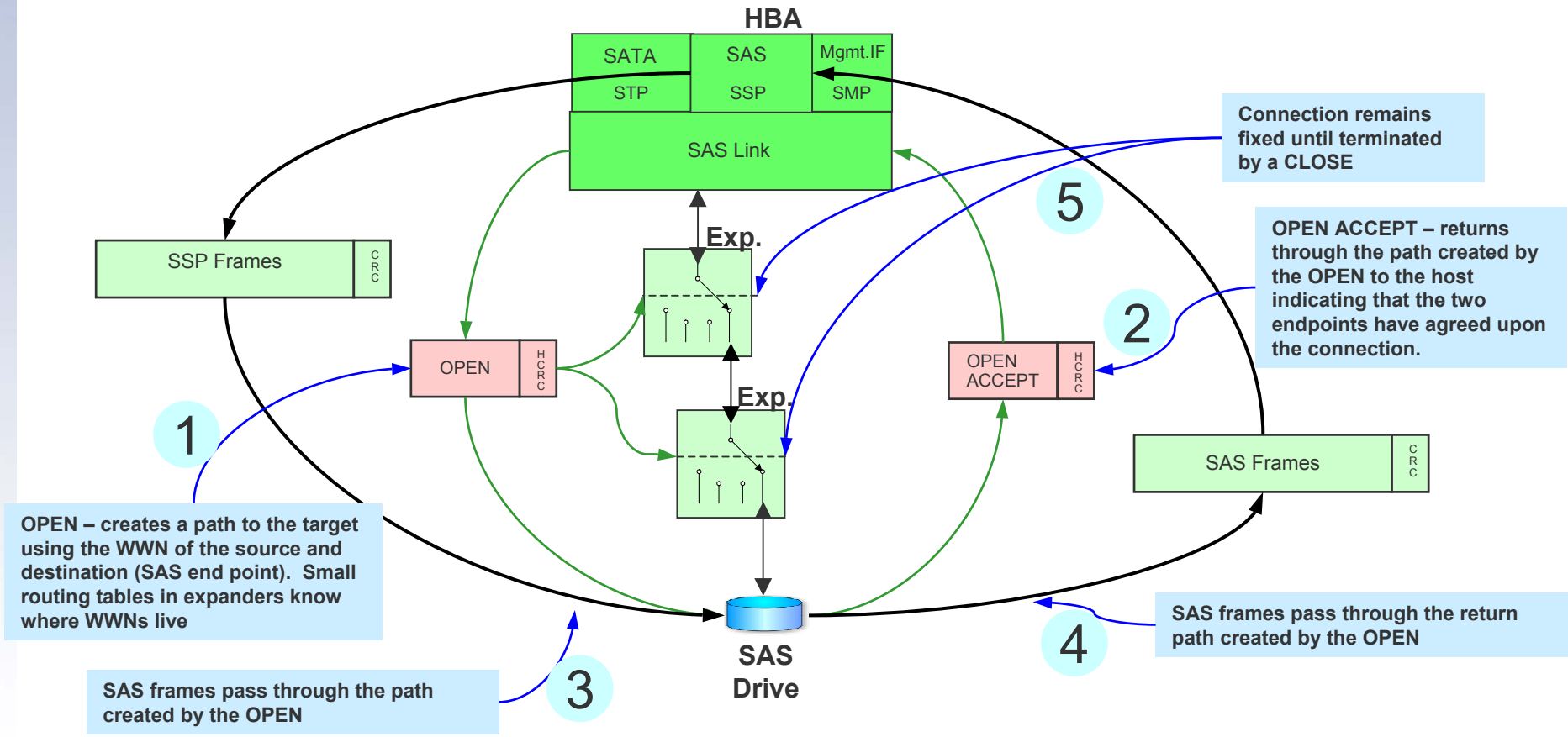


- Used to manage Expanders
 - Expanders are required to support SMP
 - Target do not support SMP
 - Functions
 - REPORT (GENERAL, SATA CAPABILITIES, MANUFACTURER INFORMATION, PHY ERROR LOG, PHY SATA, ROUTE INFORMATION), PHY CONTROL, CONFIGURE ROUTE INFORMATION, DISCOVER
- OPEN address frame to make connection request
- 1KB (1024) maximum frame size



- Full duplex
- OPEN address frame to make connection request
- SOF, frame dwords, CRC, EOF
- Each frame acknowledged with ACK, NAK
- Credit with RRDY
- SSP link layer state machine
- 1KB (1024) max frame length
- Error recovery defined for COMMAND TASK, RESPONSE frames
- Errors in DATA frames results in termination of the command

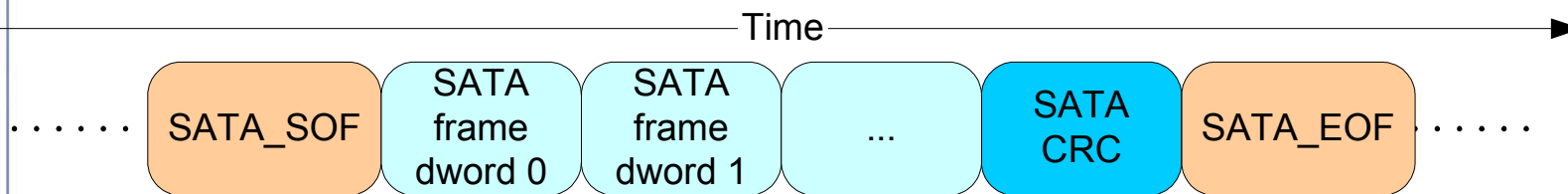
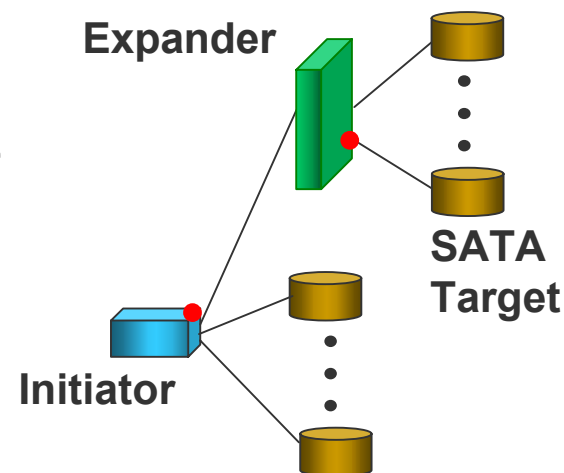




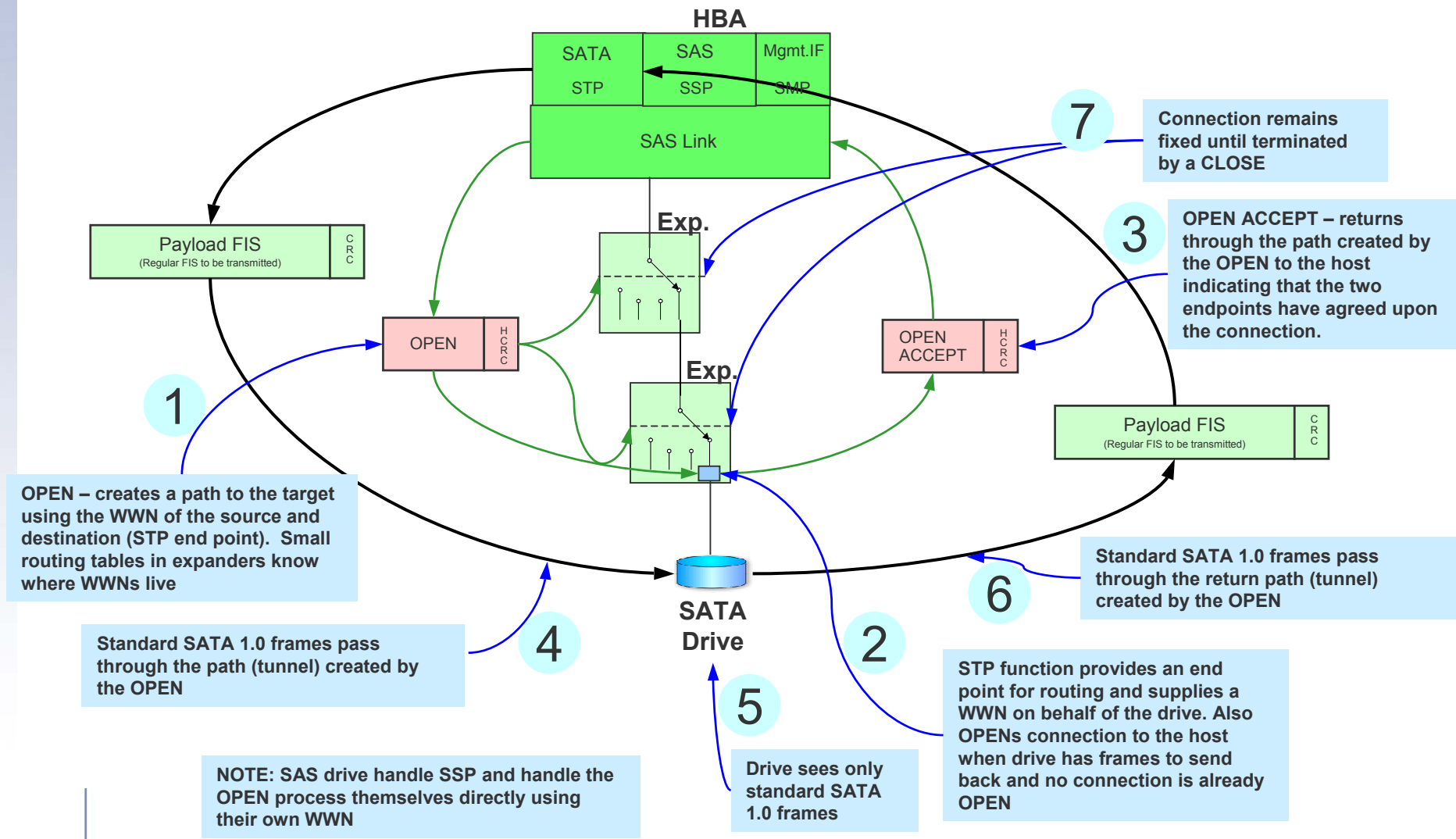
NOTE: SAS drive handle SSP and handle the OPEN process themselves directly using their own WWN

- STP (ATA) connection

- STP from initiator to last expander
- SATA from the expander to the SATA device
- After an STP connection is opened, follow SATA rules
- Frame sent as: SATA_SOF, SATA frame, SATA_EOF
- Each frame receives SATA_R_OK or SATA_R_ERR
- SATA_X_RDY/SATA_R_RDY for permission to send another frame
- Max SATA frame is 8KB (8192)



SAS Expander Operation (with SATA drives)



- Serial Attached SCSI meets the needs of mainstream enterprise class storage systems
 - Its point to point architecture will provide performance, configuration capability and scalability beyond what is available from today's device level interconnects
 - Provides a choice of devices use
 - ↪ Low cost Serial ATA disks or
 - ↪ High performance, dual port Serial Attached SCSI drives
 - ↪ Simplifies purchasing and deployment to meet changing application needs
 - Provides the flexibility to meet the cost, availability, scalability, and performance requirements for future generations of storage systems and applications

www.scsita.org
www.maxtor.com

