Serial Attached SCSI Technical Overview

by Rob Elliott, Compaq Computer Corporation
Based on T10/02-157r0 Proposed Serial Attached SCSI working draft 29 April 2002

Outline

- Introduction
- General (devices, domains, …)
- Physical layer (cables, connectors, electrical specs, …)
- Phy layer (8b10b, OOB, …)
- Link layer (primitives, connections, …)
- Transport layer (SSP, STP, and SMP frames)
- Application layer (SCSI mode pages)
- Further information

Introduction

- Serial Attached SCSI (SAS)
  - ***Serial SCSI Protocol (SSP)***
    - SCSI over Serial ATA physical layer
  - ***Serial ATA Tunneling Protocol (STP)***
    - Enhancement to Serial ATA adding addressing
  - ***Serial Management Protocol (SMP)***
    - Expander management

Introduction - General

- Expanders
  - Simple virtual circuit switches
  - STP to SATA protocol conversion
  - edge expanders - simple subtractive decode
  - fanout expanders - routing table - max. one per domain
  - 64 devices per expander
- 4096 total devices in a SAS domain

Introduction - Physical layer

- Connectors
  - Disk drive/backplane - Based on SATA connectors
    - Dual port - extra port on other side of SATA signals, between signal and power
    - SATA or SAS disk drive can plug into SAS backplane
  - External - Based on InfiniBand™ 4-wide connector (SFF-8470)
    - Special keying for SAS
  - Being standardized in SFF
- Electrical specs
  - 1.5 Gbps, 3.0 Gbps
  - Based on SATA 1.0 and XAUI

Introduction - Phy layer

- 8b10b like all other serial protocols
- OOB compatible with Serial ATA

Introduction - Link layer

- SAS primitives use K28.5; SATA use K28.3
- Address frames
- WWN addressing
- Connections
- Scrambling

Introduction - Transport layer

- SSP
  - SCSI frames are based on FCP
  - COMMAND, XFER_RDY, DATA, RESPONSE
  - TASK, AEN, and AEN_RESPONSE added
- STP
- SMP
  - Functions for expanders

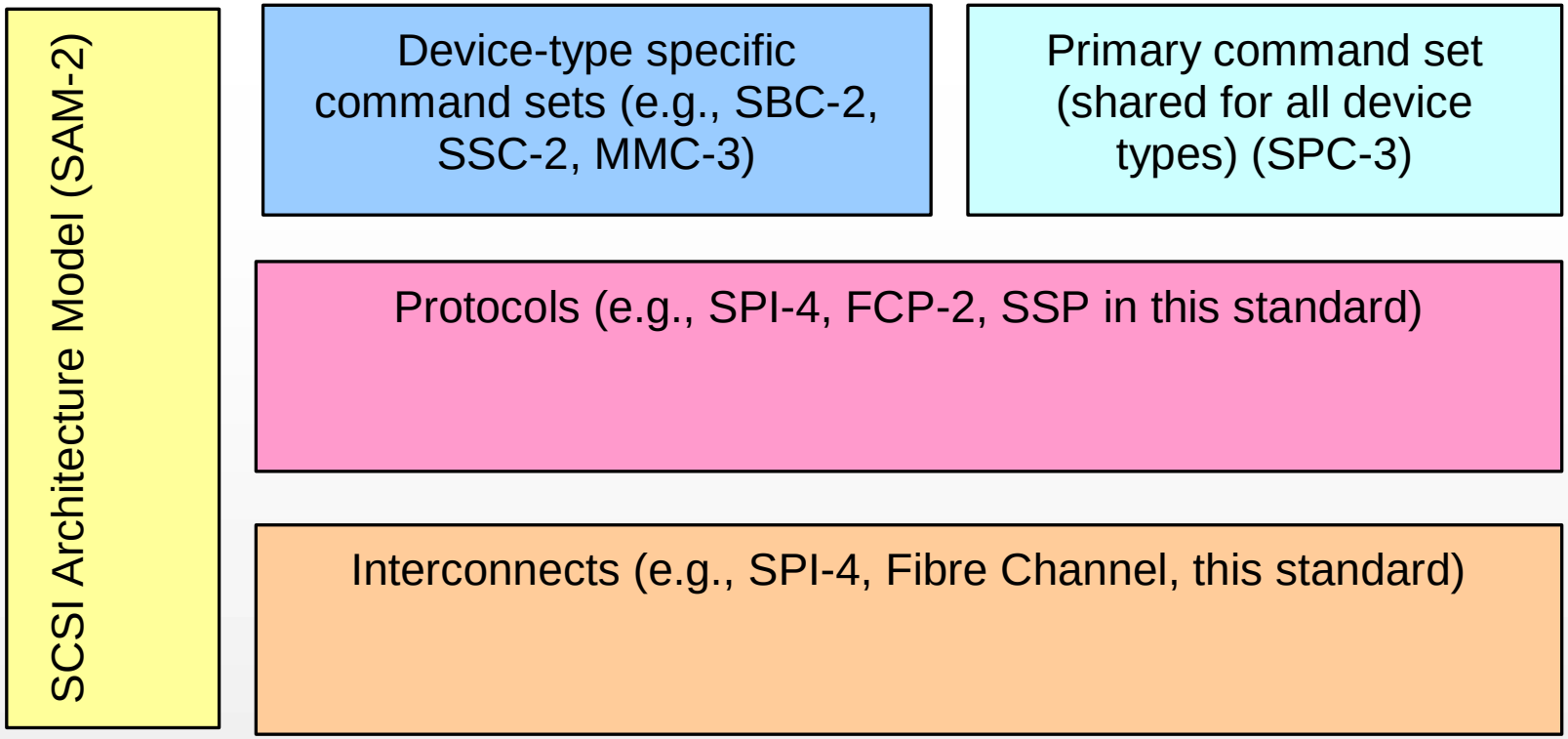
Introduction - Application layer

- SCSI
  - Disconnect-Reconnect mode page
  - Protocol-Specific mode page
- ATA
  - Addressing added

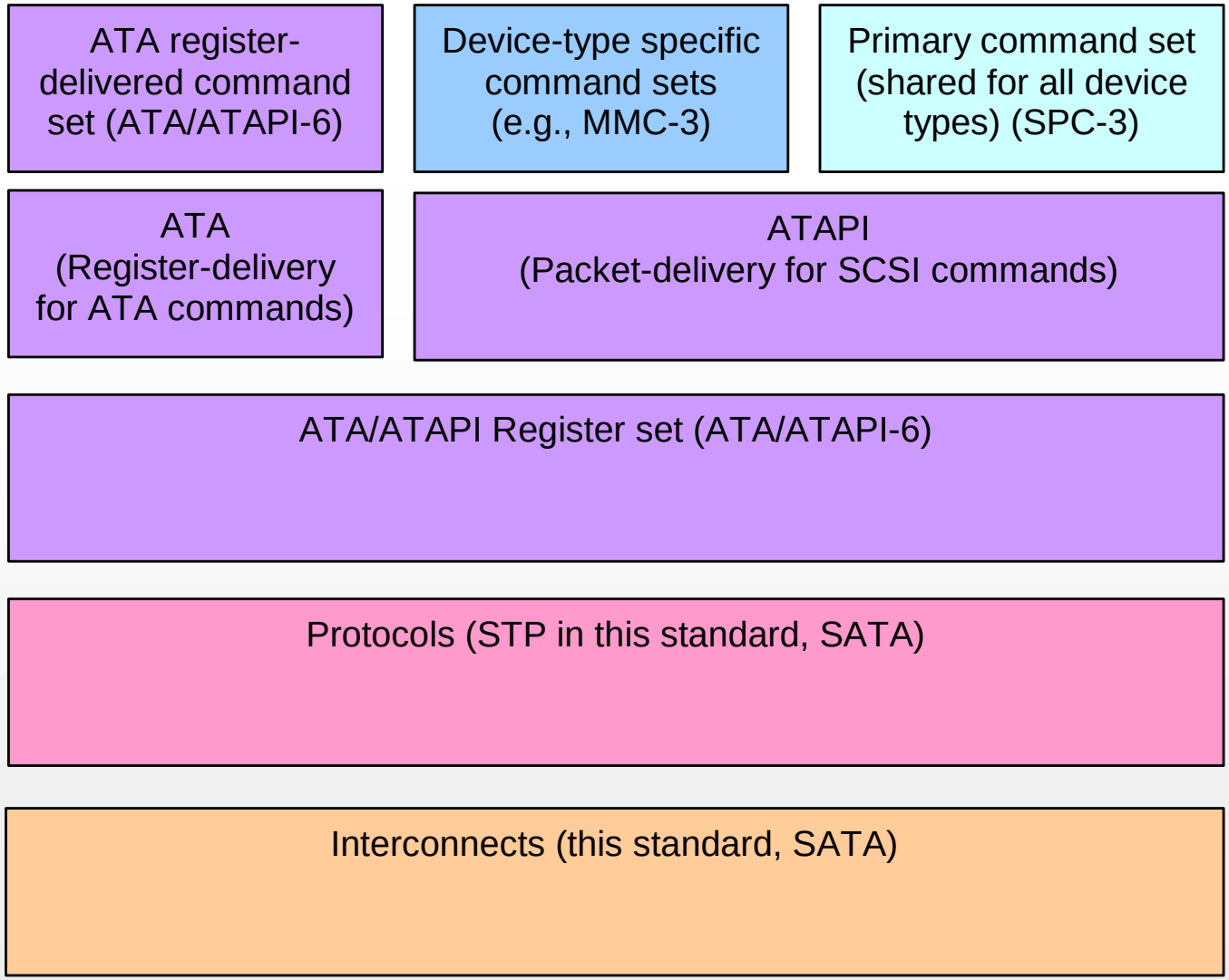
General outline

- SCSI standards
- ATA standards
- Serial ATA overview
- Protocol layers
- SSP (SCSI), STP (ATA), and SMP
- Initiators, targets, and expanders
- Phy
- Ports and wide links
- Domains
- Sample topologies
- Possible configurations
- Pathways
- Device names
- Transmit data path
- Resets
- Expander model

General - SCSI standards



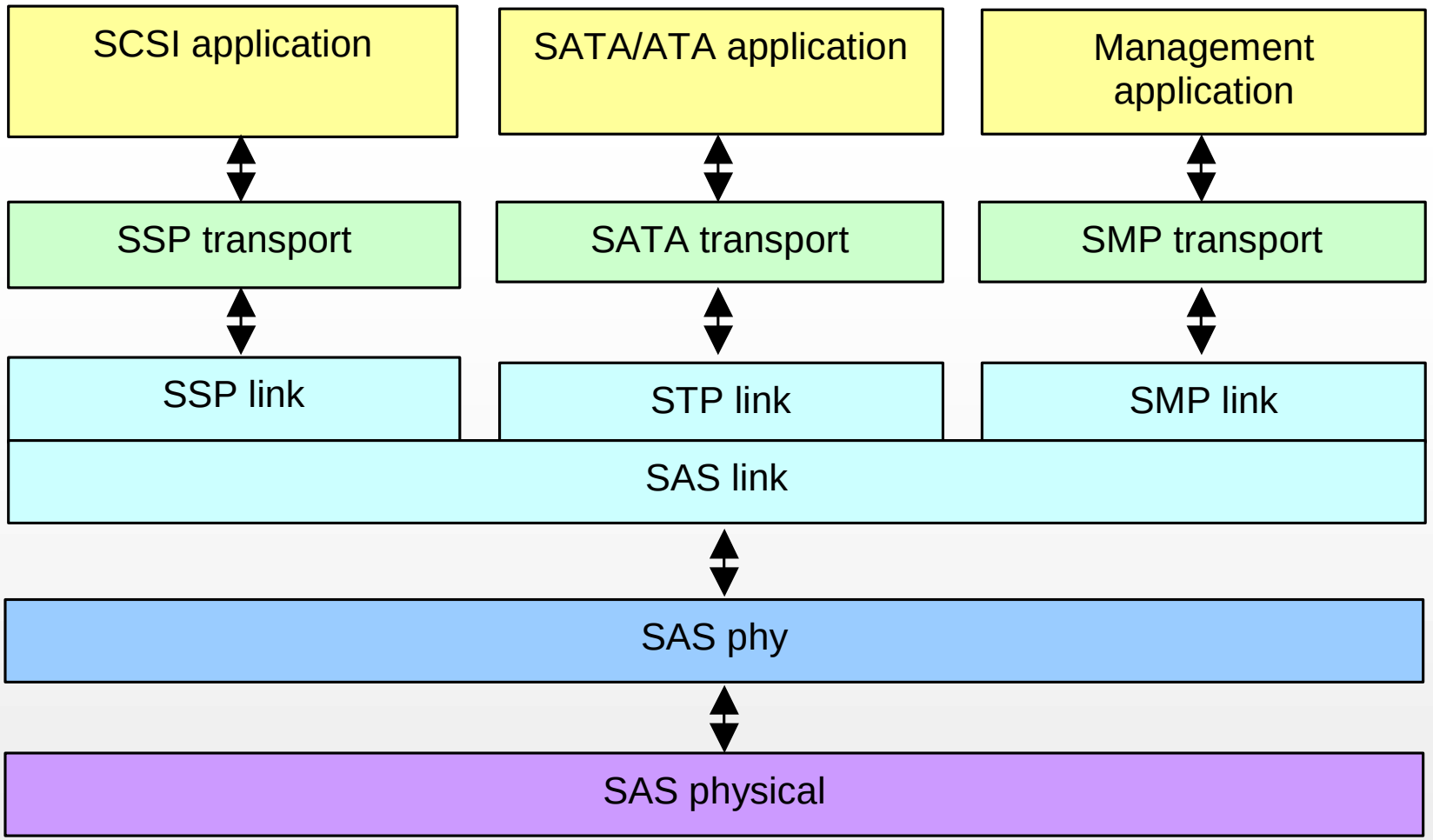
General - ATA standards



General - Serial ATA overview

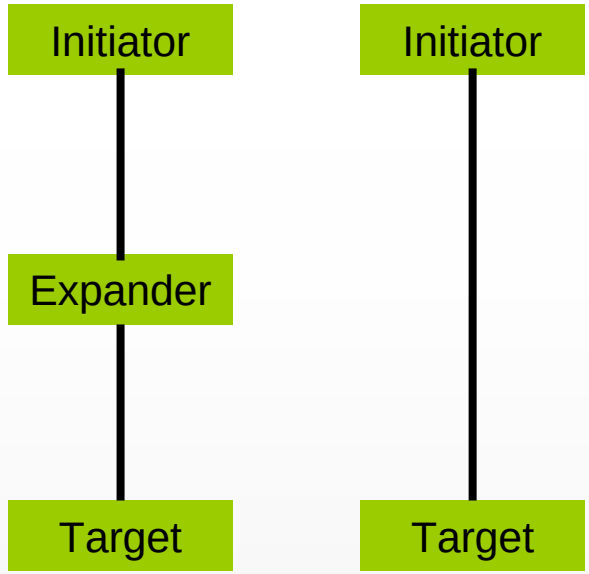
- Physical
  - Point-to-point links
  - 1.5 Gbit/sec transfer rate; 3.0 Gbit/sec and 6.0 Gbit/sec in the future
  - spread-spectrum clocking
  - Device connector, cables, backplane connectors
- Link
  - Out-of-band (OOB) reset sequence (includes speed negotiation)
  - 8b10b coding, repeated primitive scrambling, frame data scrambling, power management, half duplex
- Transport
  - ATA/ATAPI-6 transport protocol - PIO, DMA, DMA queuing, PACKET
  - Frame Information Structure (FIS) with CRC-32
  - No addressing; little-endian

General - Protocol layers

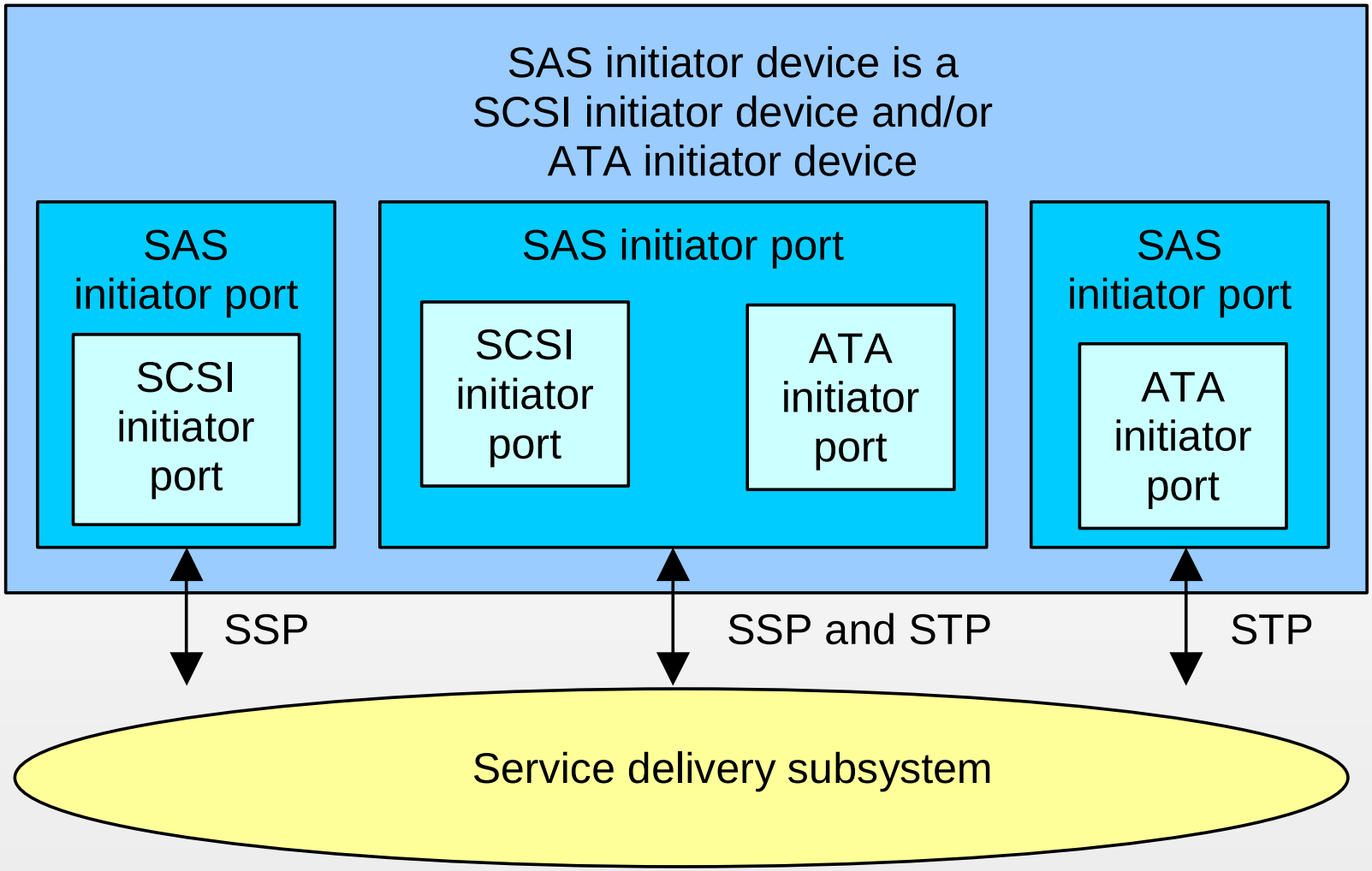


General - Initiators, targets, and expanders

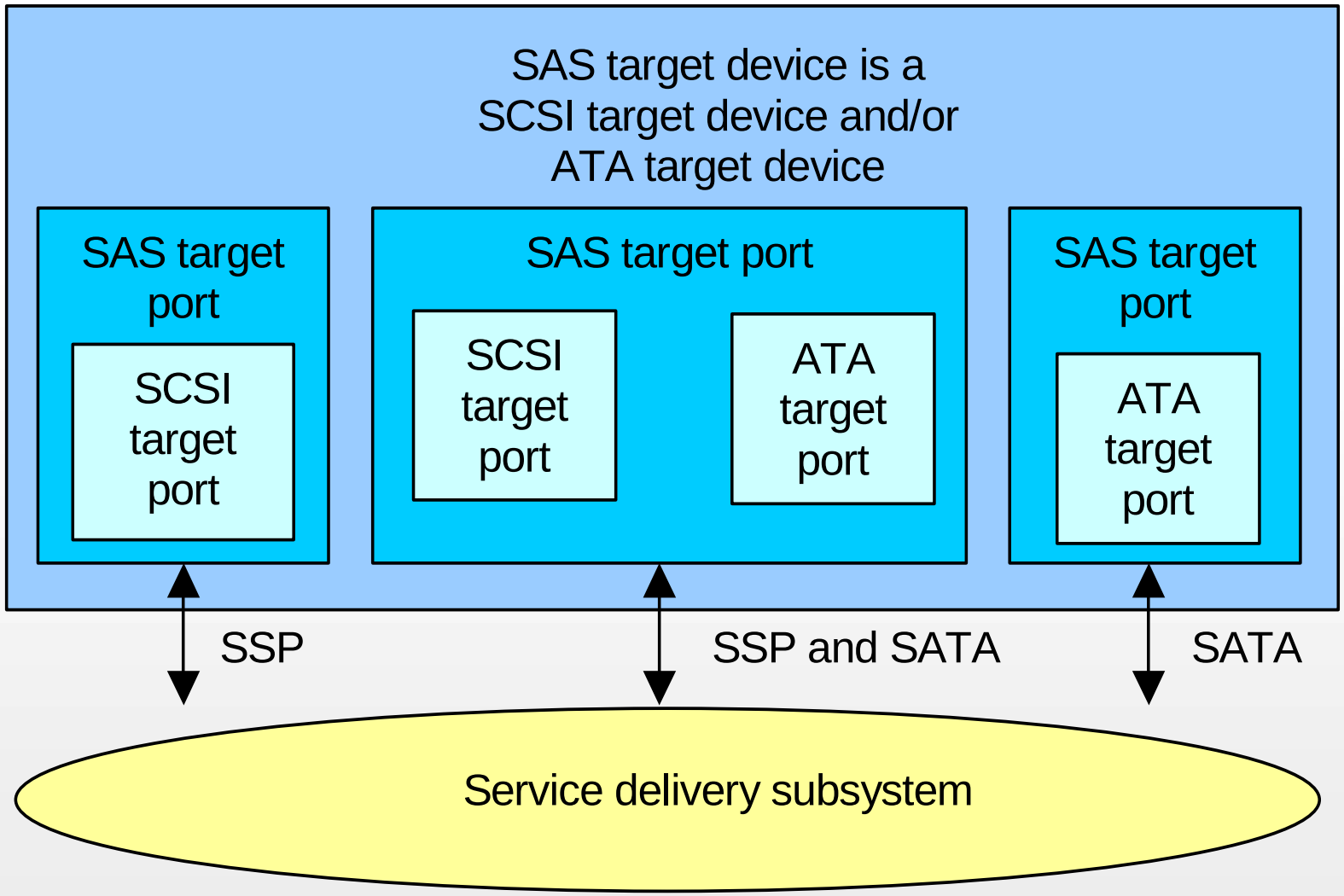
- Initiator (HBA) protocols
  - SSP (SCSI)
  - STP (ATA)
  - SMP
  - SATA (ATA)
- Target (disk or tape drive) protocols
  - SSP (SCSI)
  - SATA (ATA)
  - SMP
- Expander protocols
  - Initiator side
    - SSP (SCSI)
    - STP (ATA)
    - SMP
  - Target side
    - SSP (SCSI)
    - SATA (ATA)
    - SMP



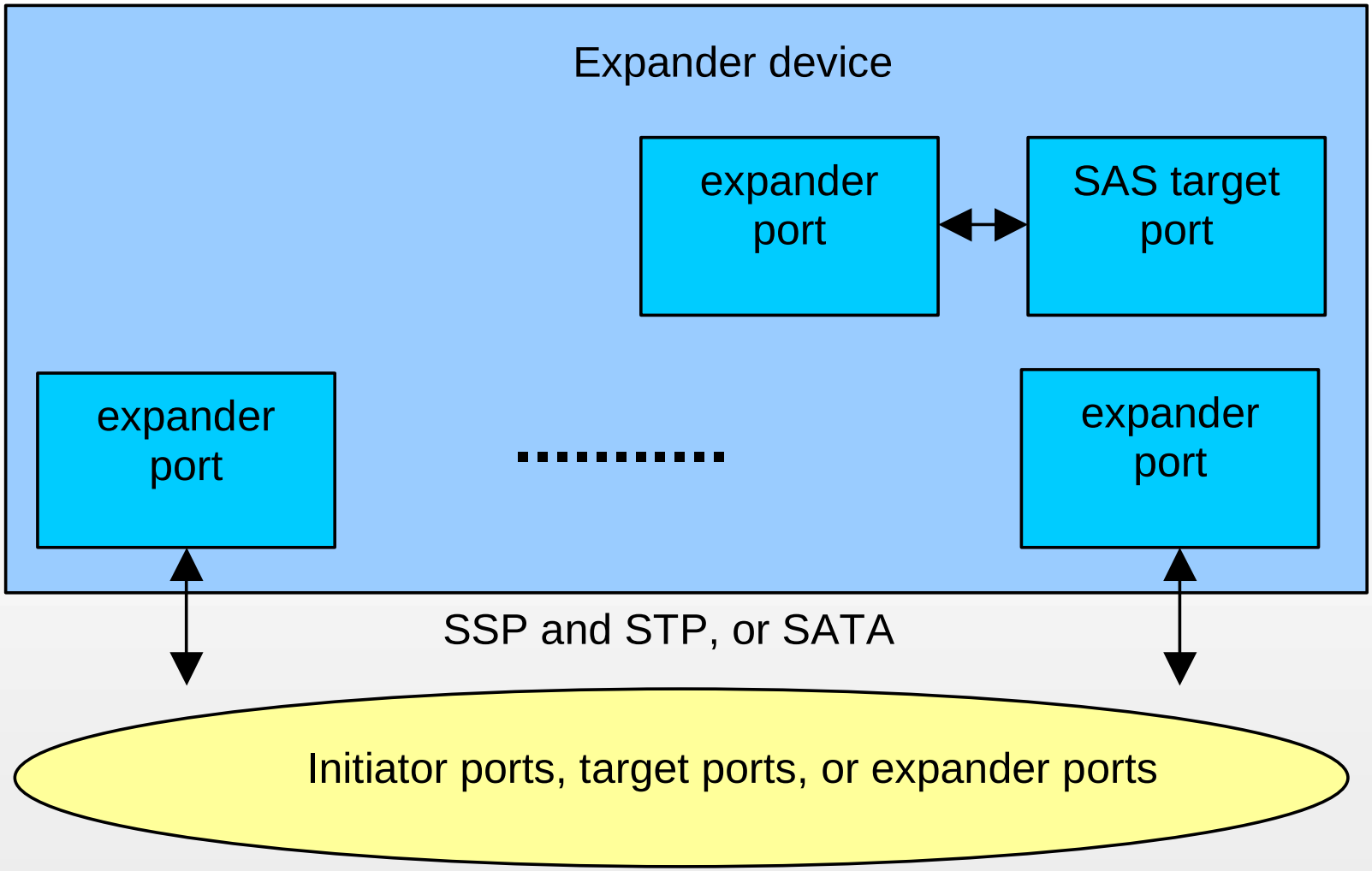
General - Initiator device



General - Target device



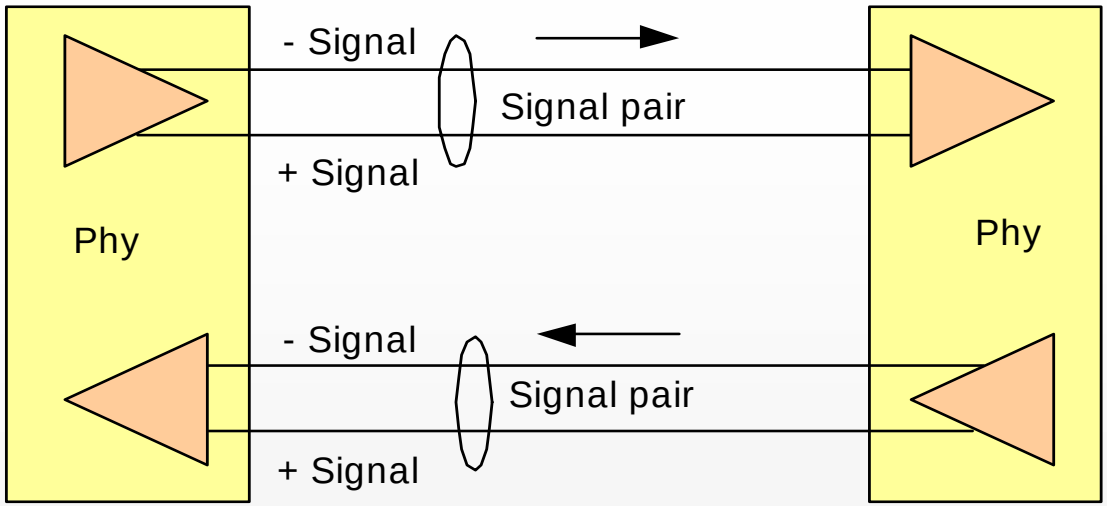
General - Expander device



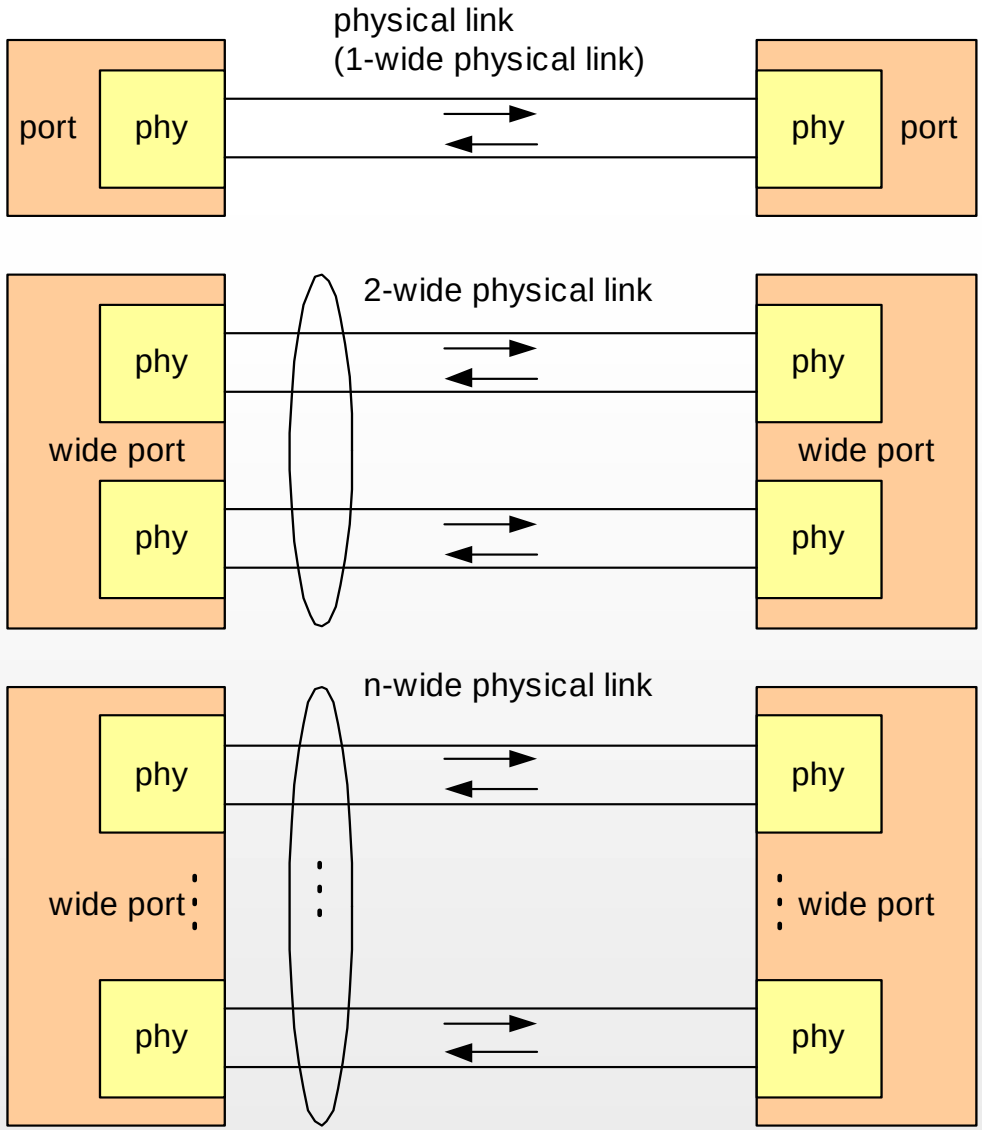
General - Expander types

- Fanout expander
  - Up to 64 phys
  - One per domain
  - Maintains routing table for whole domain
  - Attaches to edge expanders, initiators, and targets
- Edge expanders
  - Up to 64 phys
  - Subtractive routing
  - Attaches to initiators, and targets, and one other expander

General - Phy


- Signal
Signal pair
+ Signal
Phy
Phy
- Signal
Signal pair
+ Signal

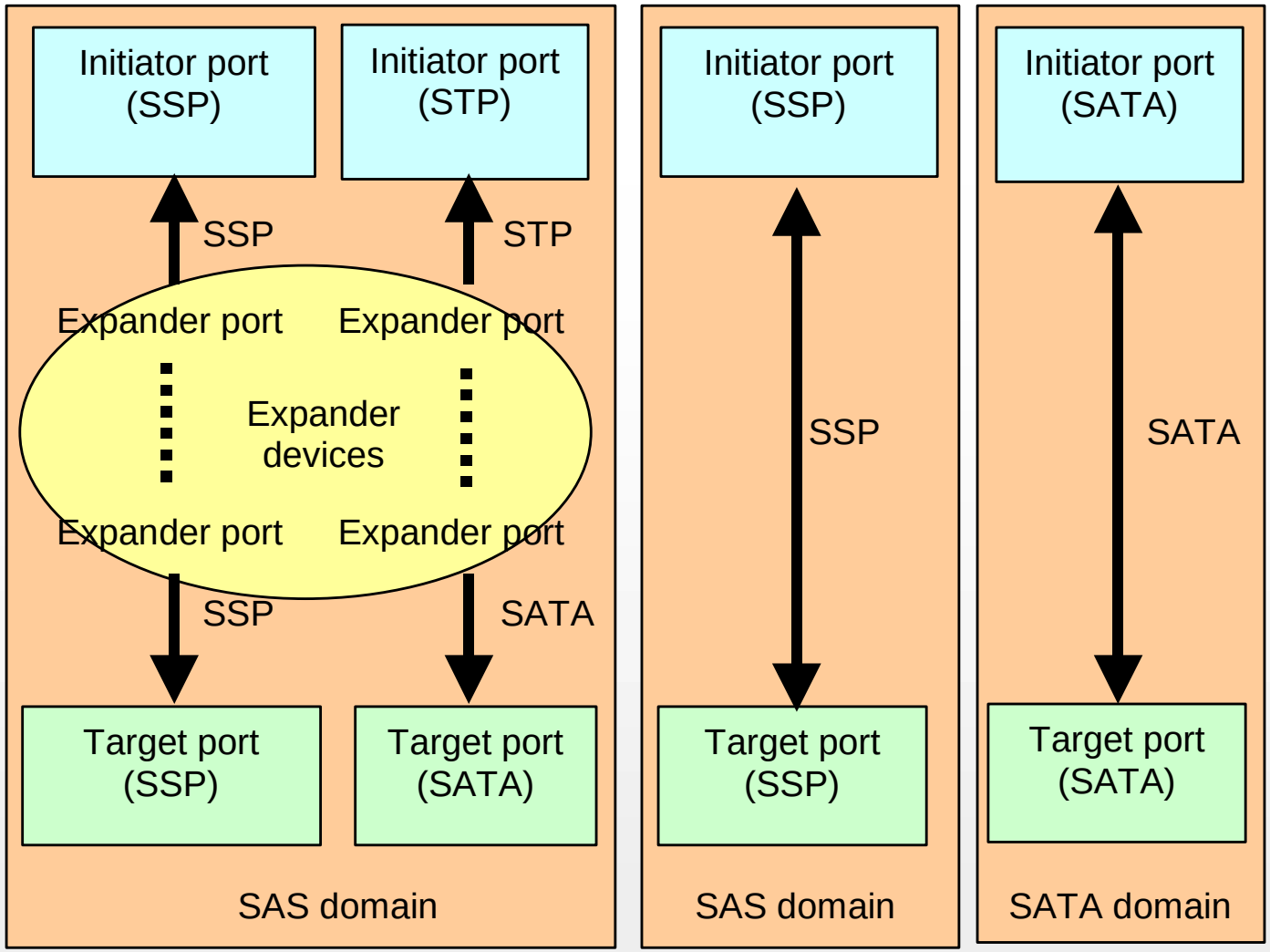
General - Ports and wide links



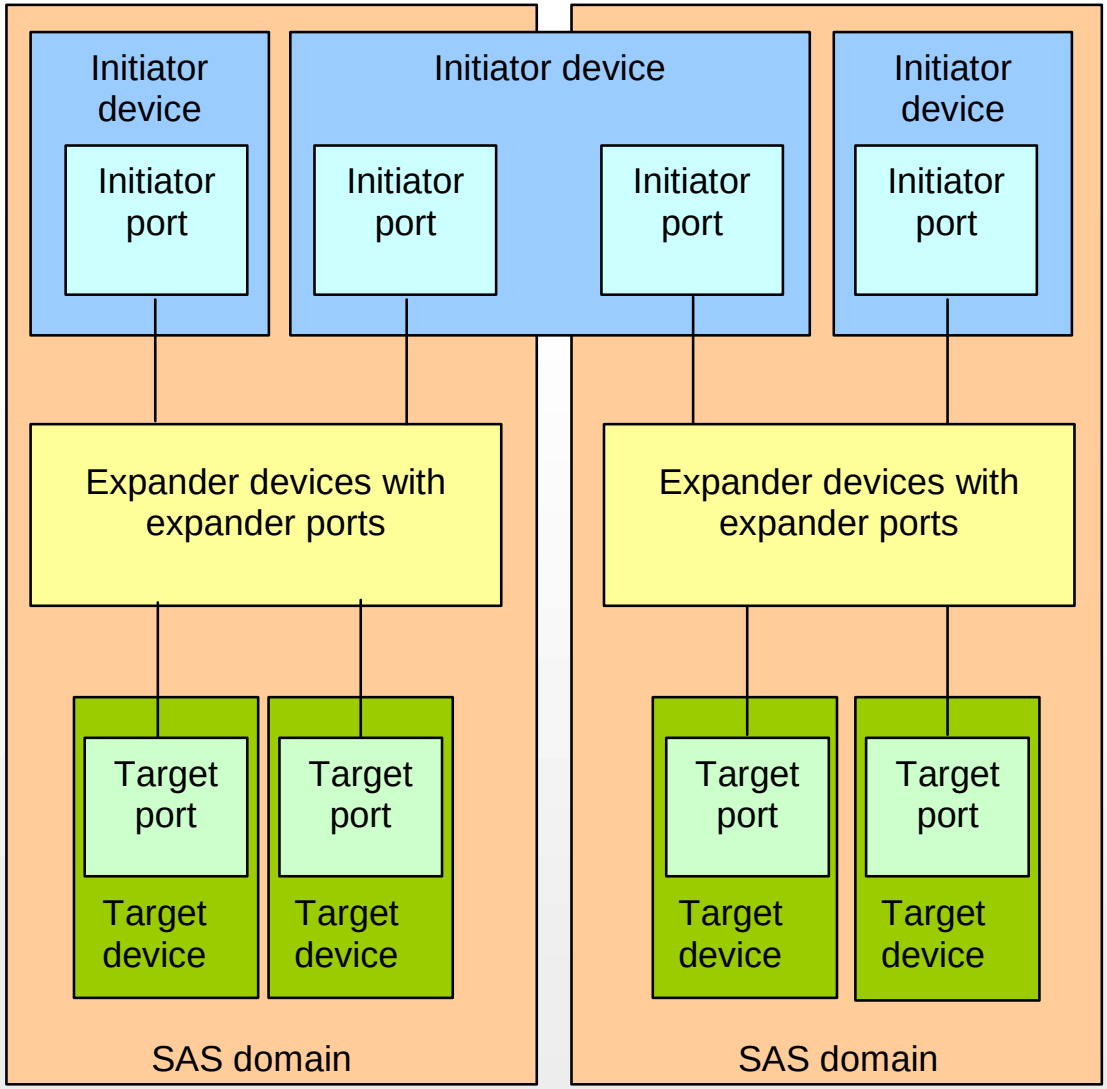
General - Wide links

- Aggregates bandwidth
- Different connection may be open on different link
- Expected usage
  - Common: Wide HBA to wide expander
    - External 4-wide cables common
  - Possible: Wide HBA to wide RAID controller
  - Unlikely: Wide disk drives
    - Dual ports only for use in separate domains, not more bandwidth

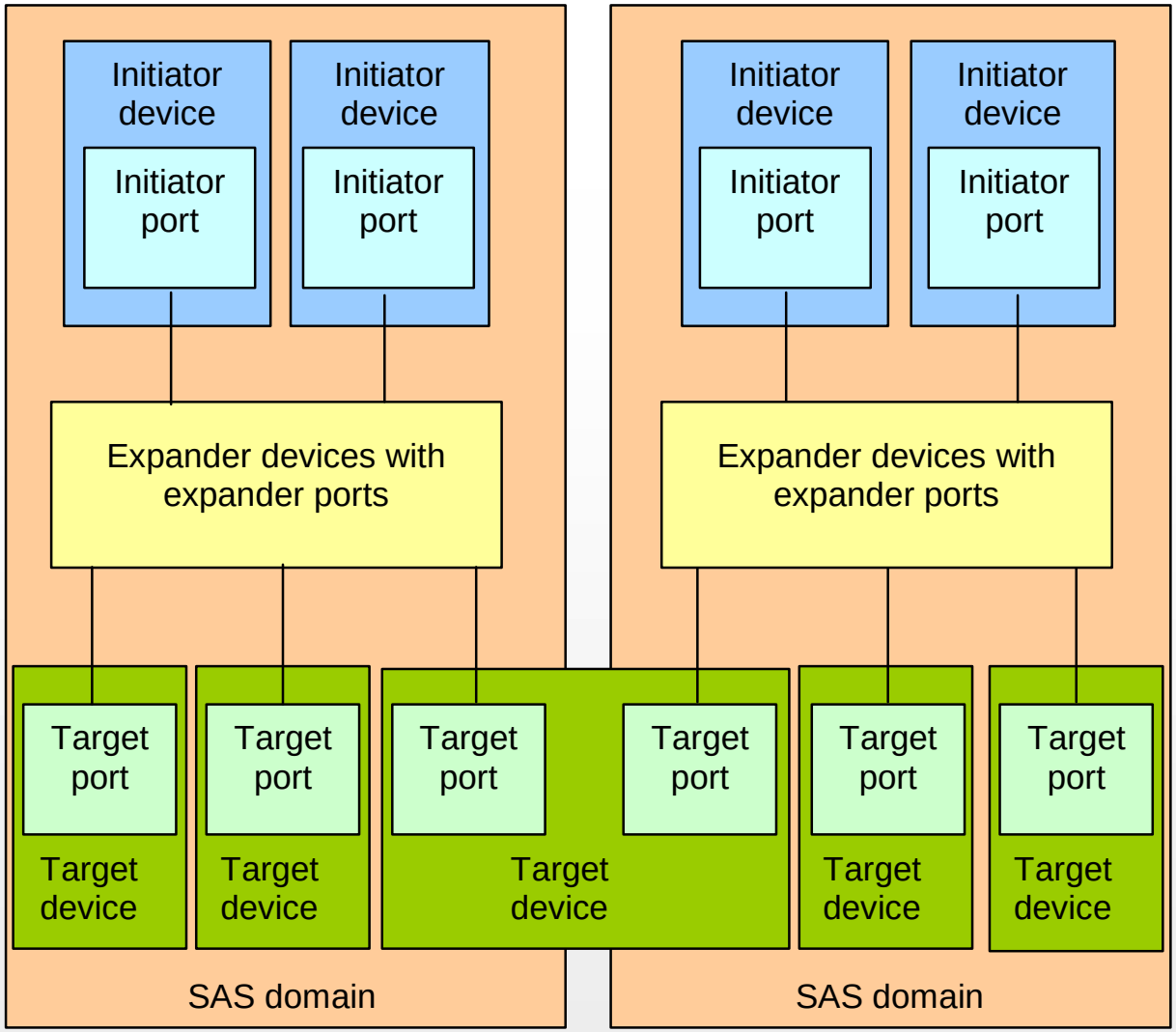
General - Domains



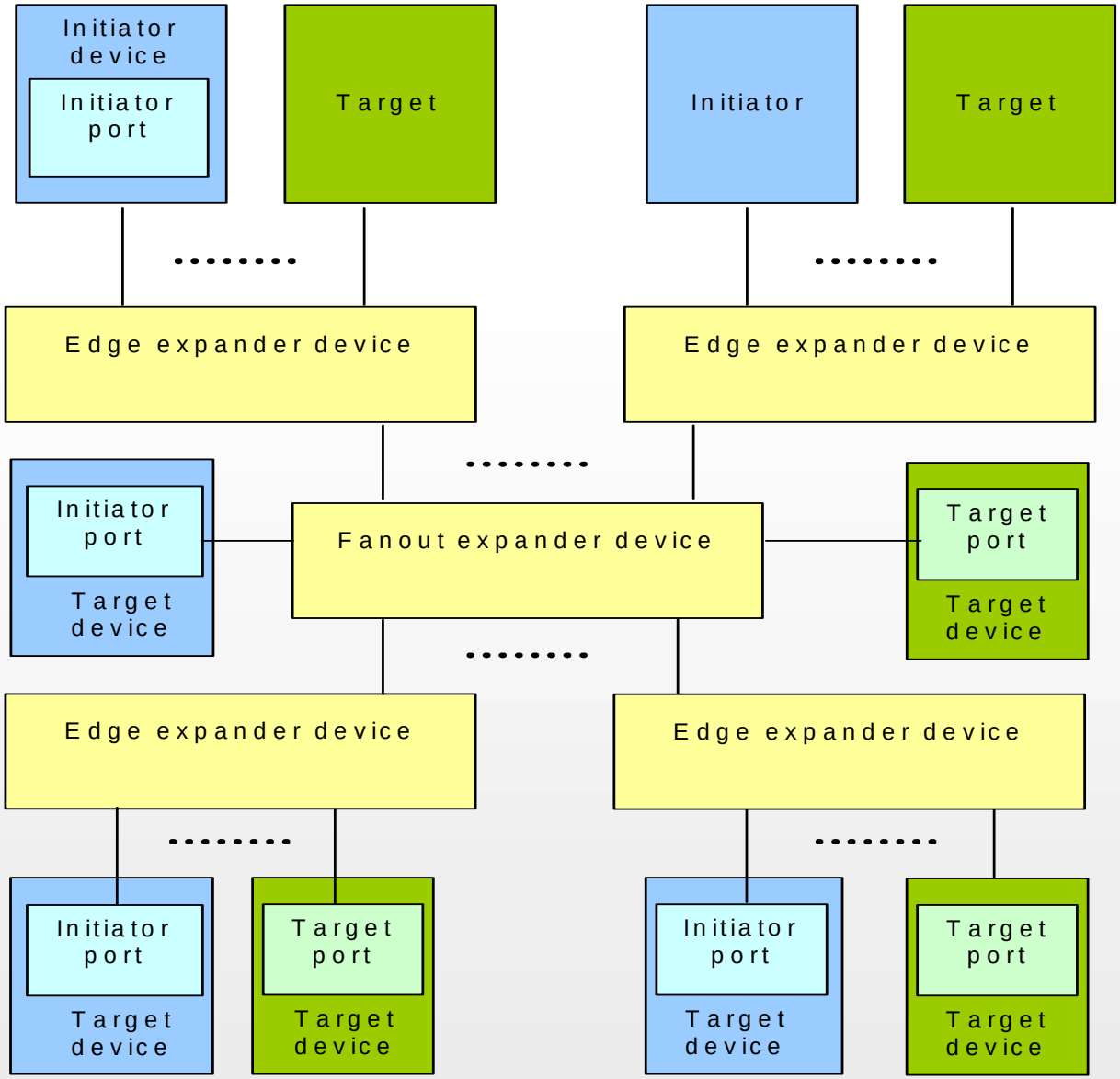
General - Initiator device in 2 domains



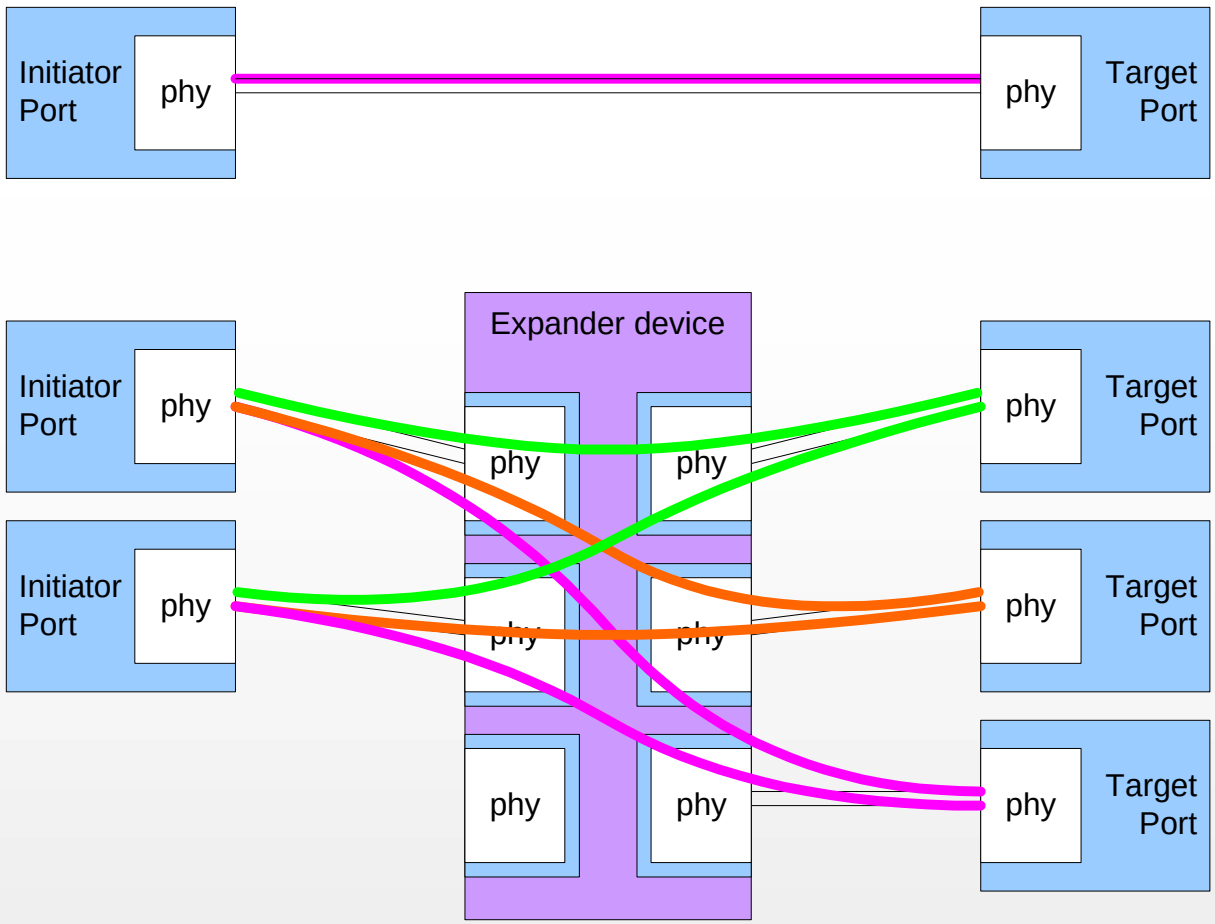
General - Target device in 2 domains



General - Maximum configuration

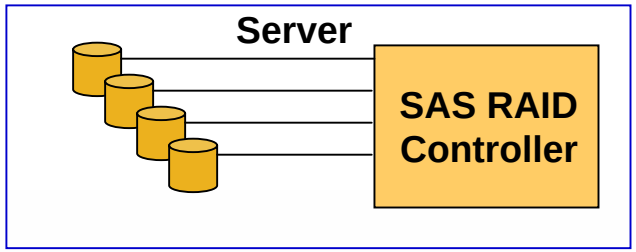


General - Pathways

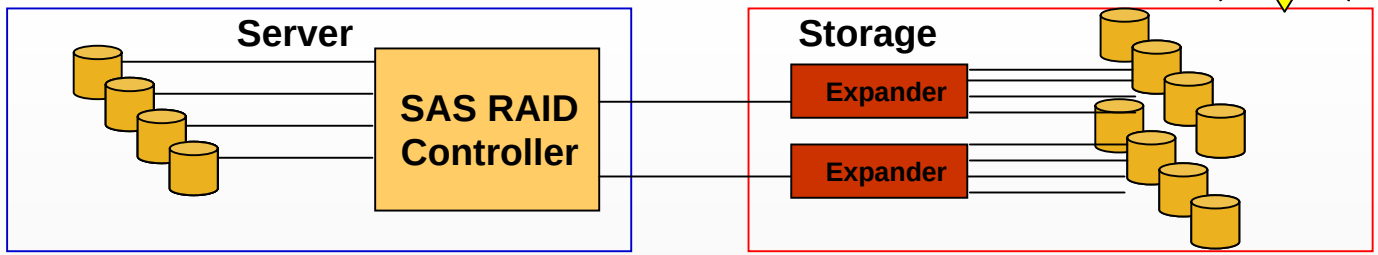


General - Possible configurations 1

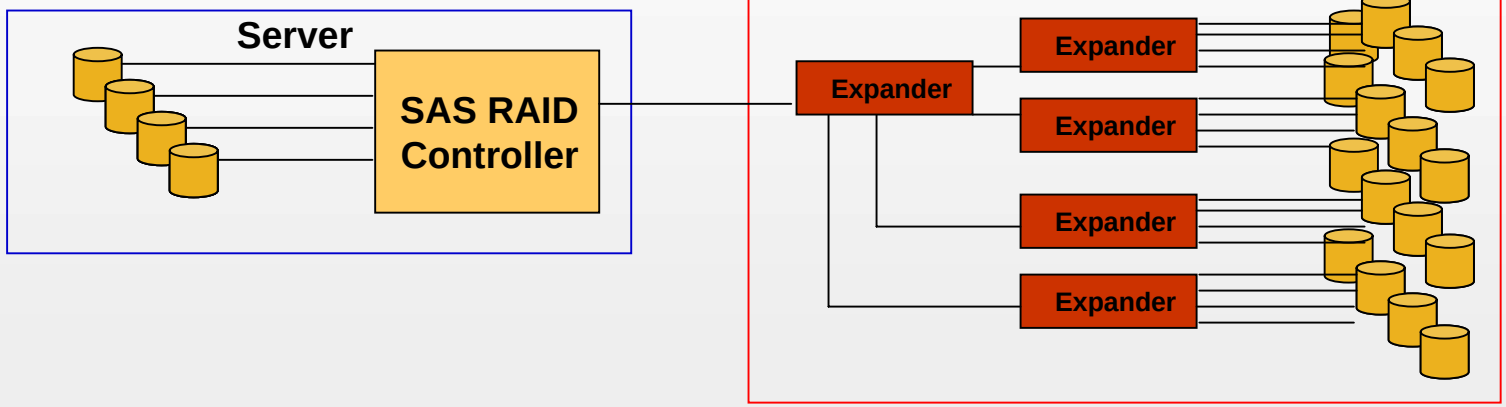
Internal drives



Internal drives and External JBOD

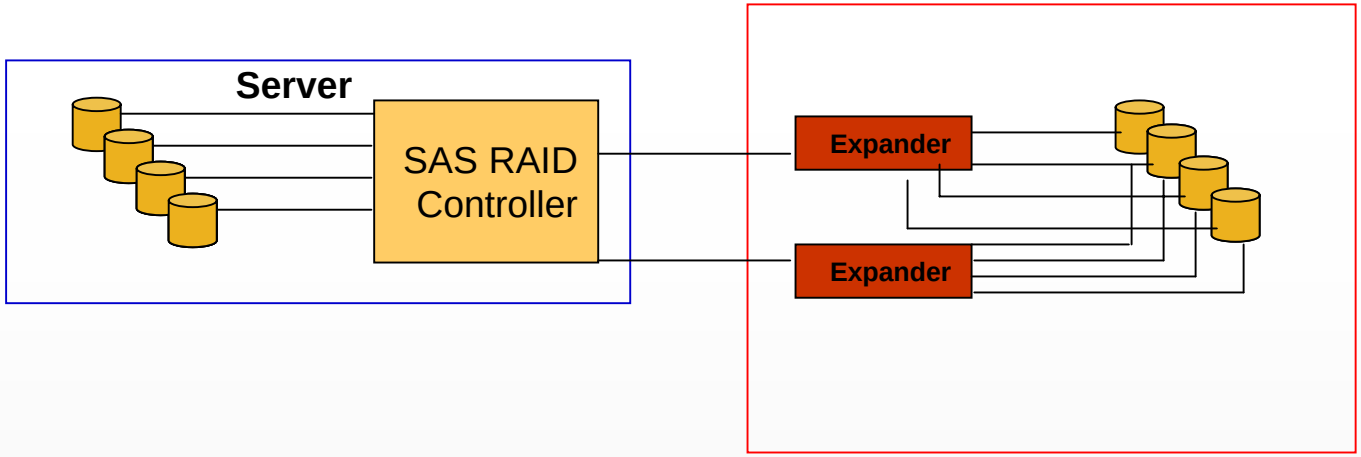


Internal drives and External JBOD

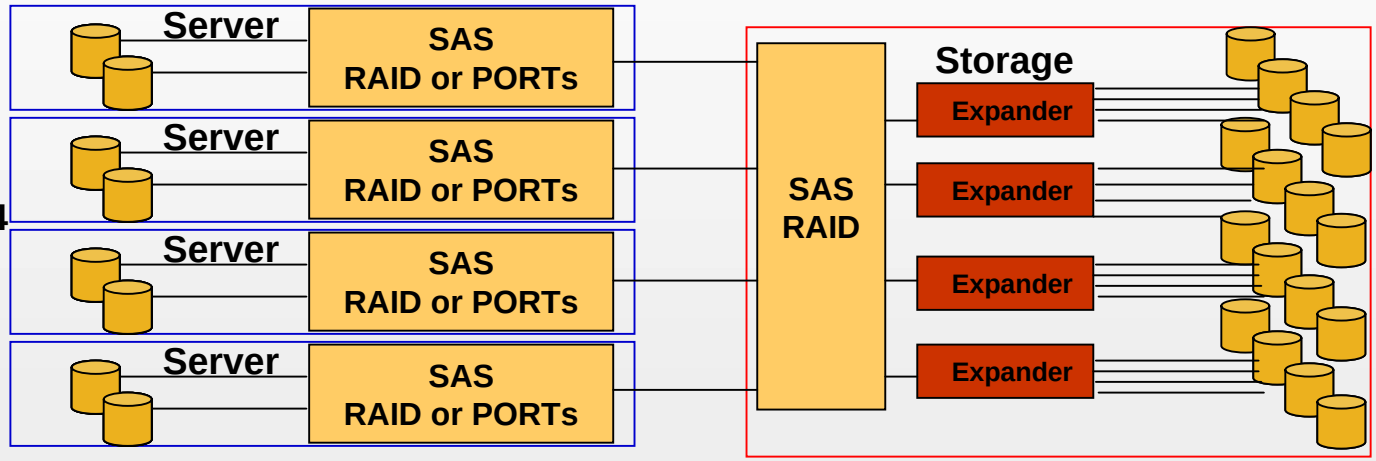


General - Possible configurations 2

Internal drives and External JBOD with dual port drives (future)

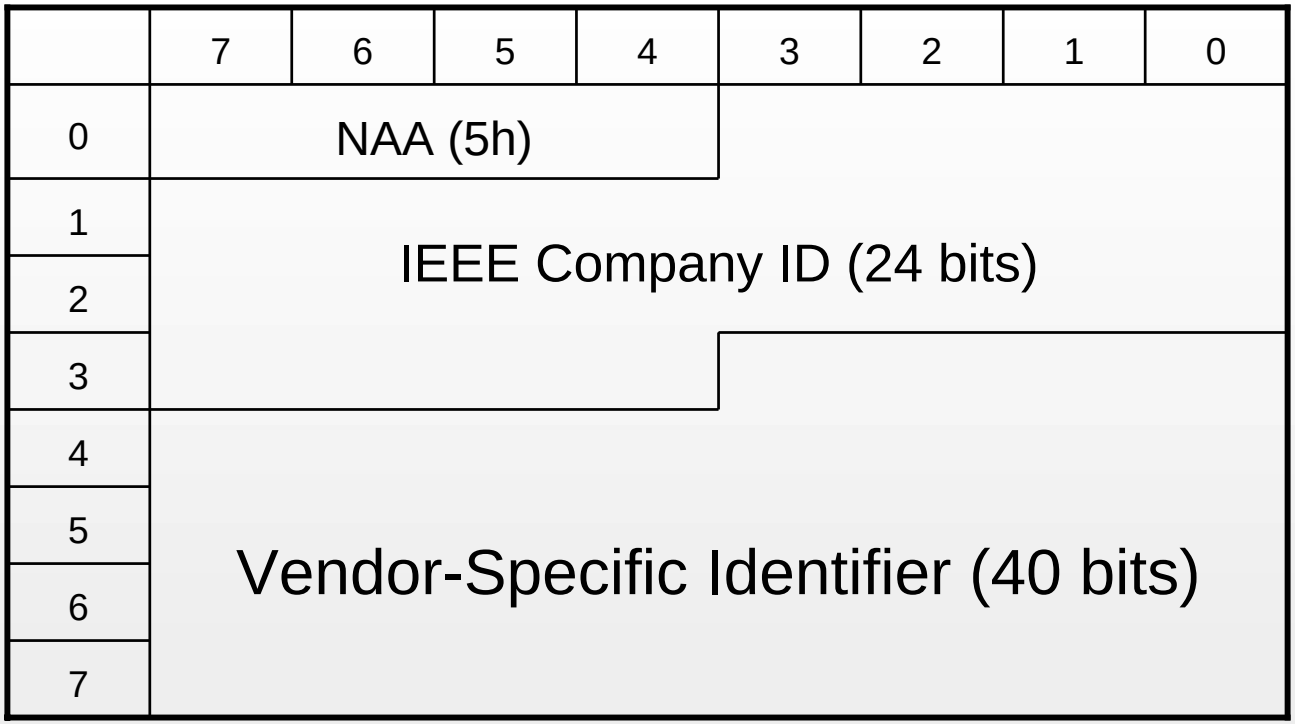


Internal drives and external RAID as a 4 Node Cluster

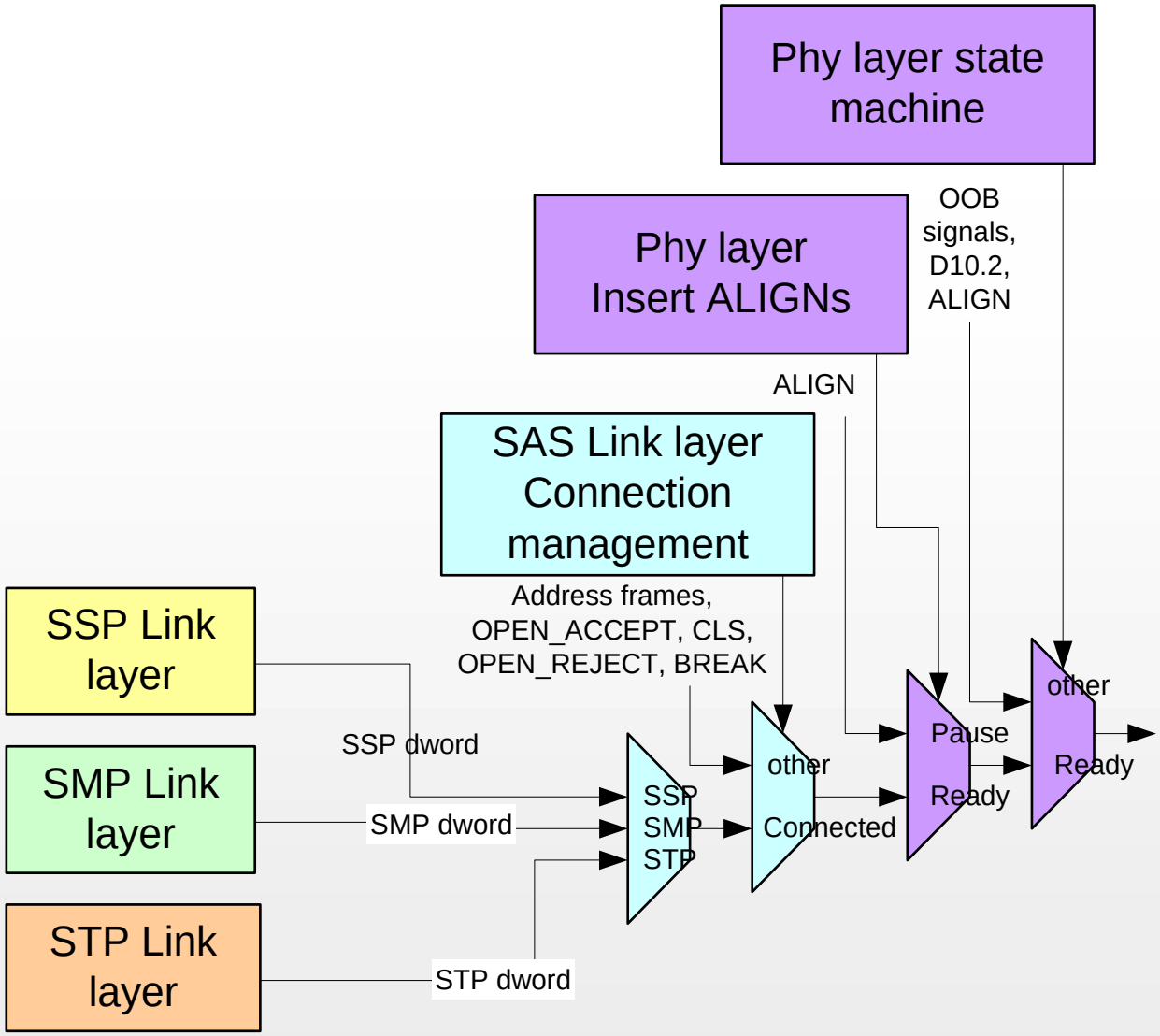


General - Device names

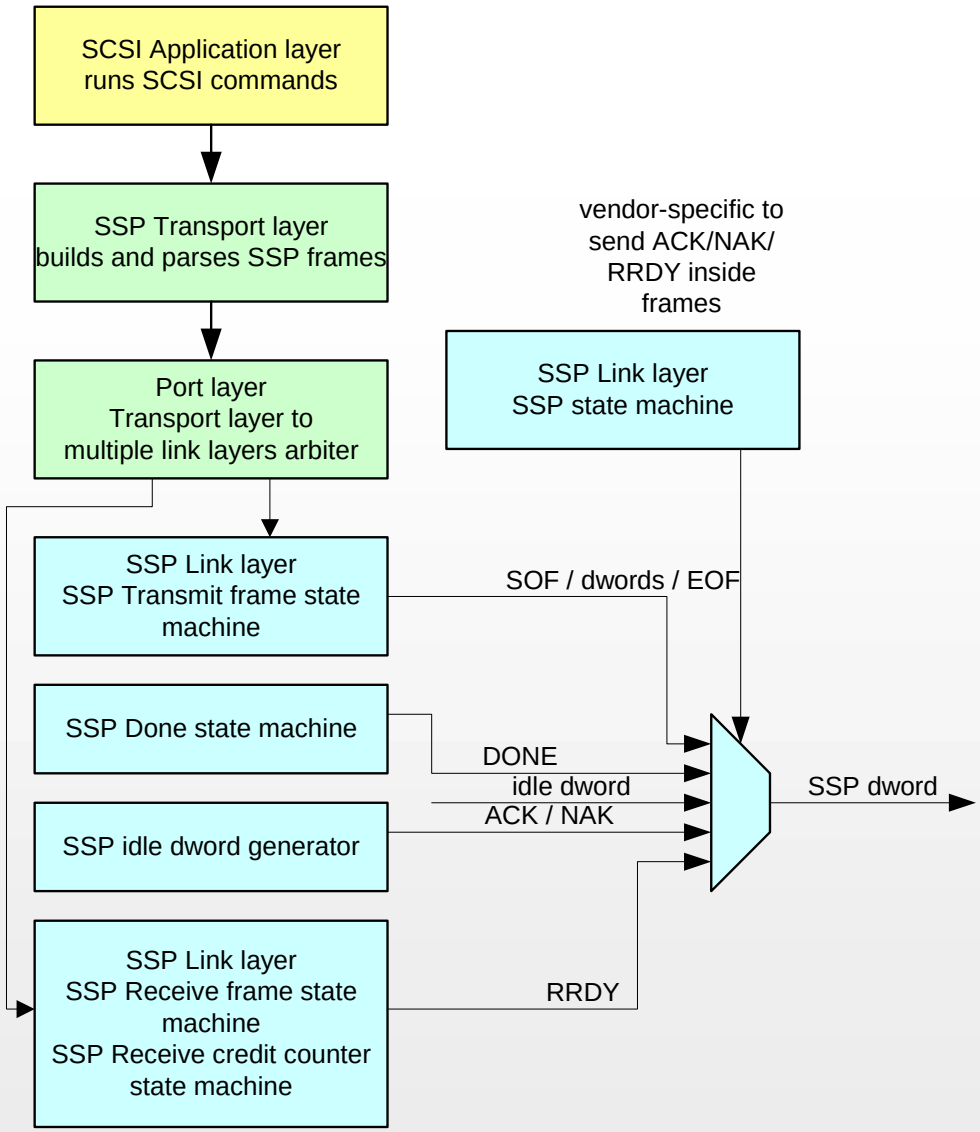
- Each device has a 64-bit FC style Worldwide Name (WWN), used for addressing



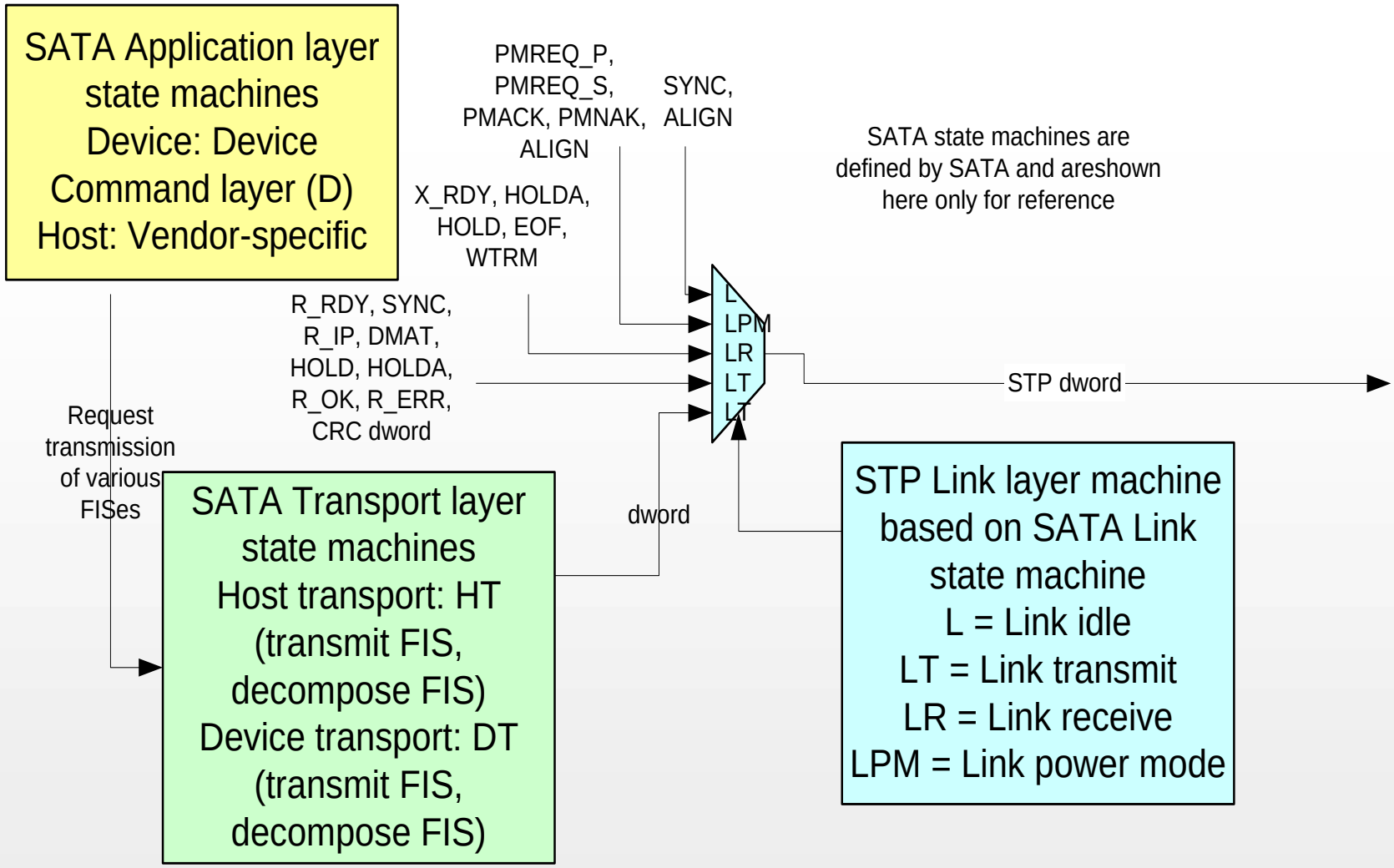
General - Transmit data path SAS portion



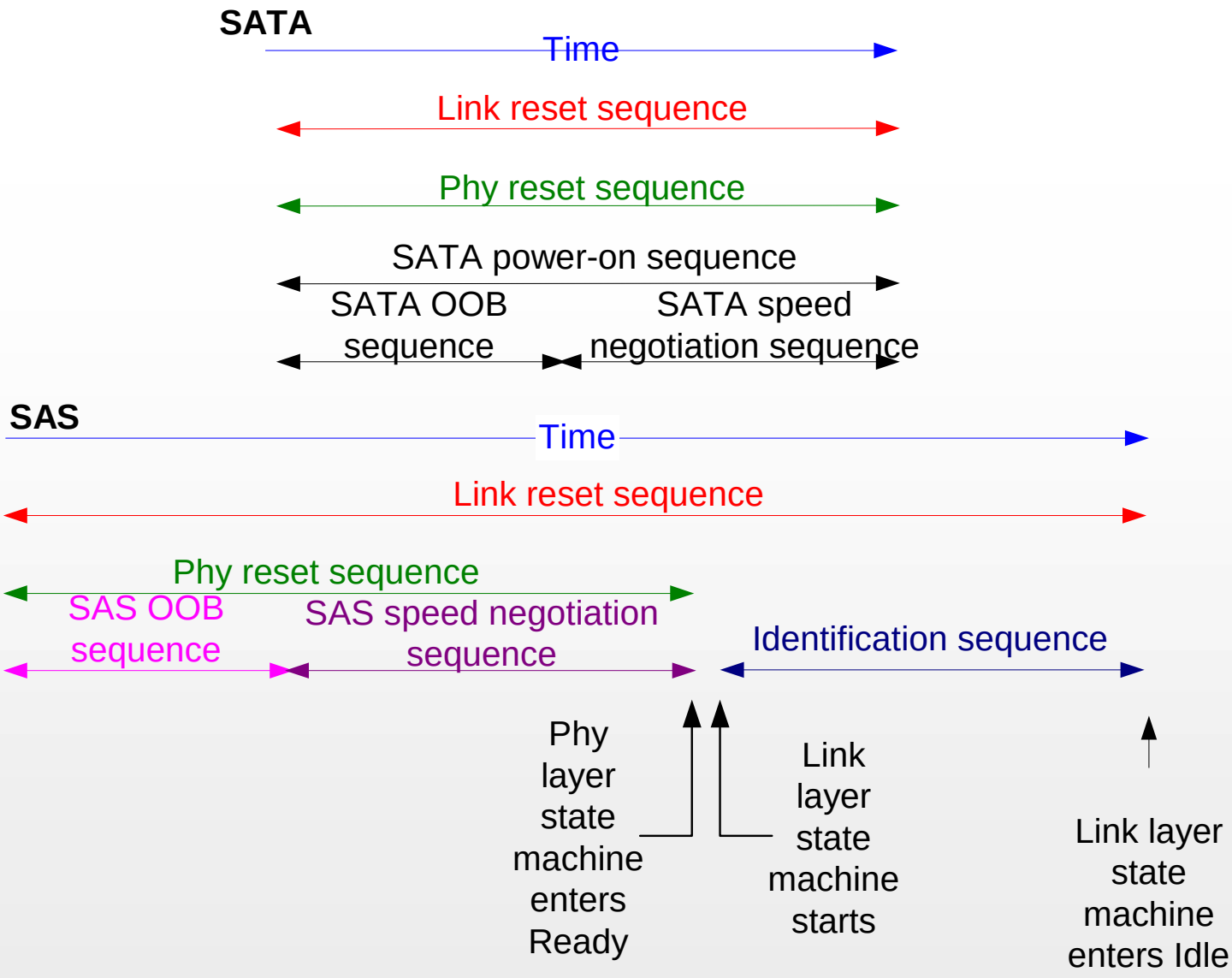
General - Transmit data path - SSP portion



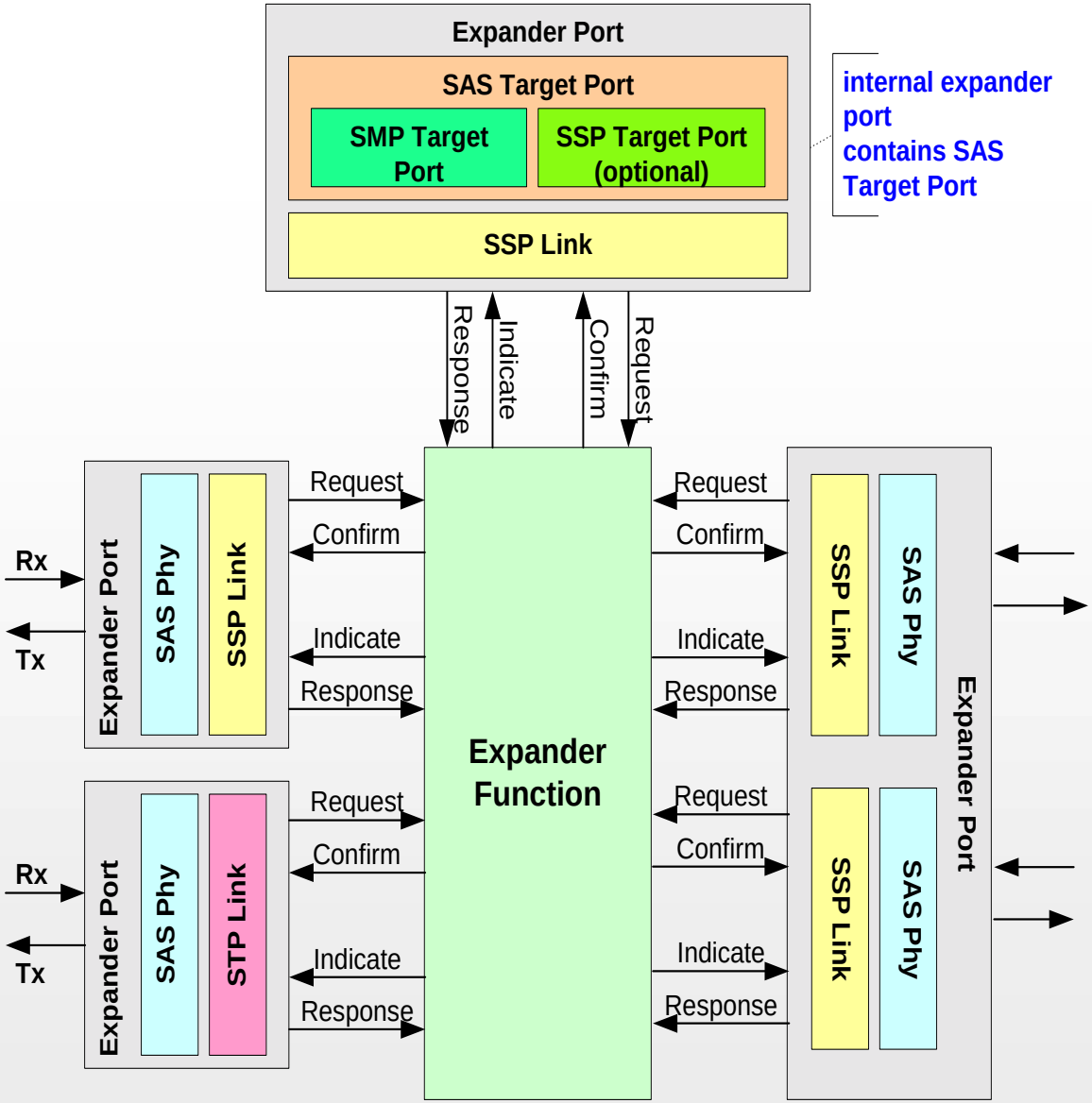
General - Transmit data path - STP portion



General - Reset terminology



General - Expander model



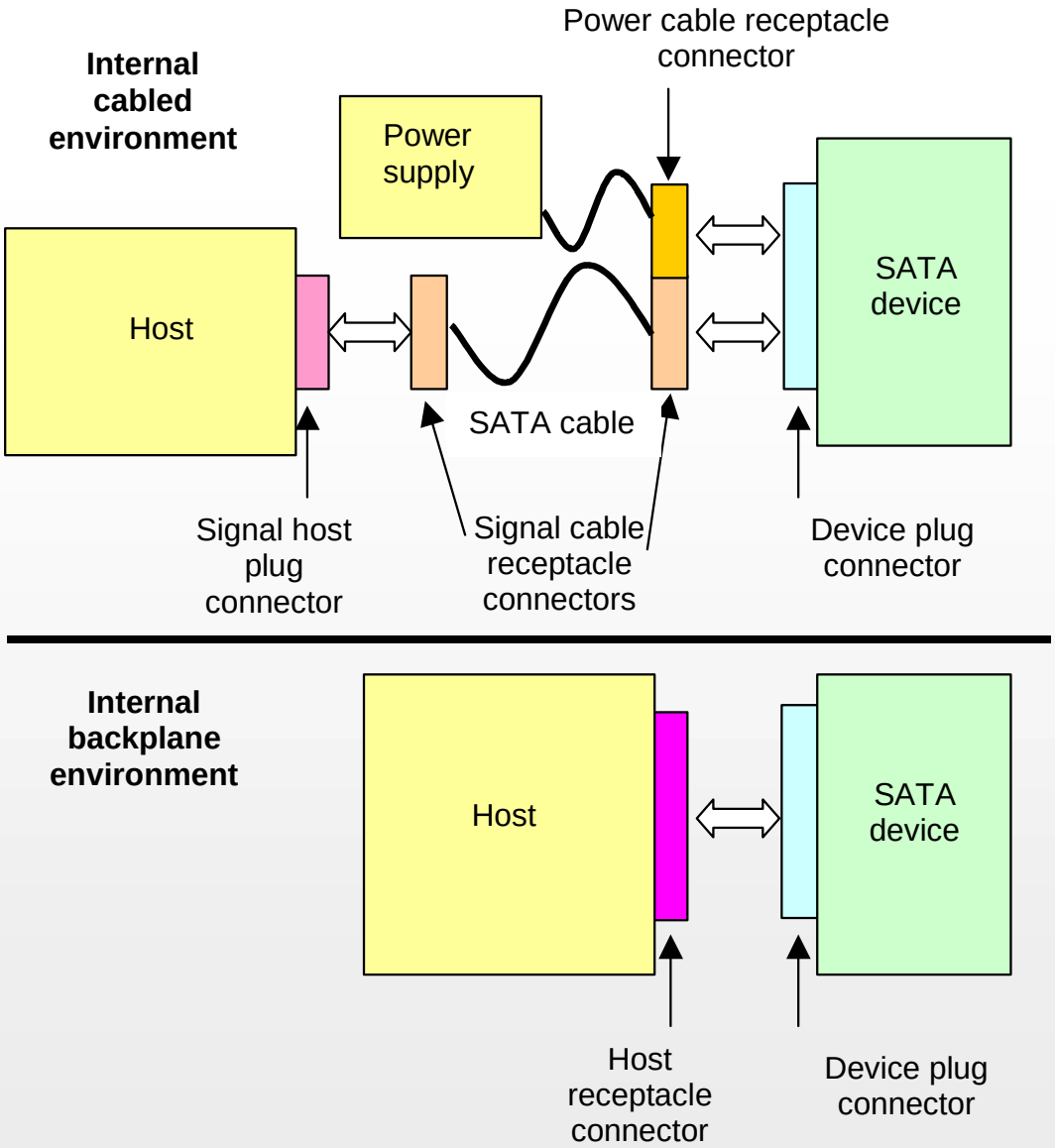
…Outline…

- Introduction
- General (devices, domains, …)
- ***Physical layer (cables, connectors, electrical specs, …)***
- Phy layer (8b10b, OOB, …)
- Link layer (primitives, connections, …)
- Transport layer (SSP, STP, and SMP frames)
- Application layer (SCSI mode pages)
- Further information

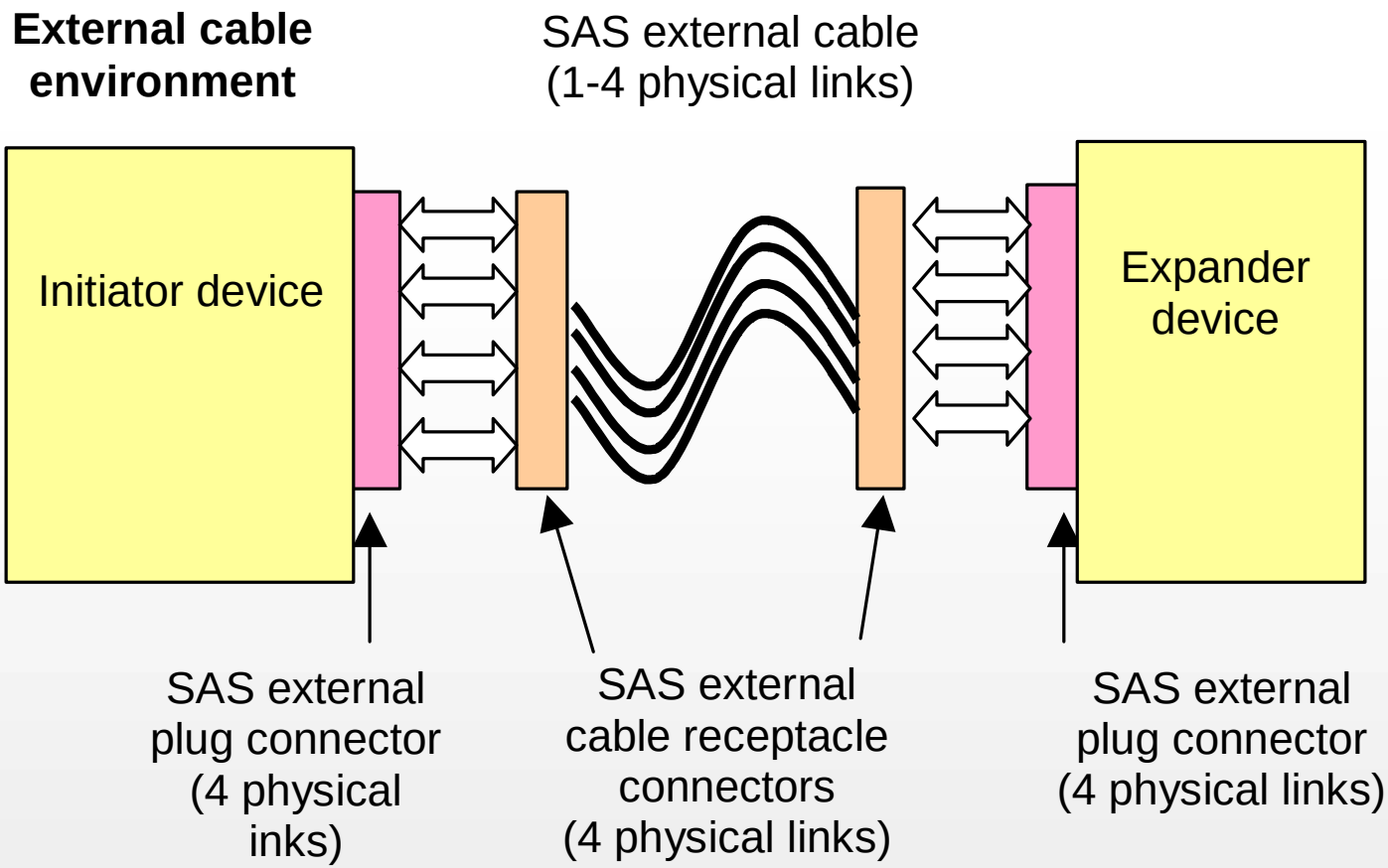
Physical layer outline

- SATA cables and connectors
- SAS external environment
- SAS internal environments
- Cables and connectors
- Compliance points
- Electrical characteristics
- Eye diagrams
- Transmit and receive electrical characteristics
- Other highlights

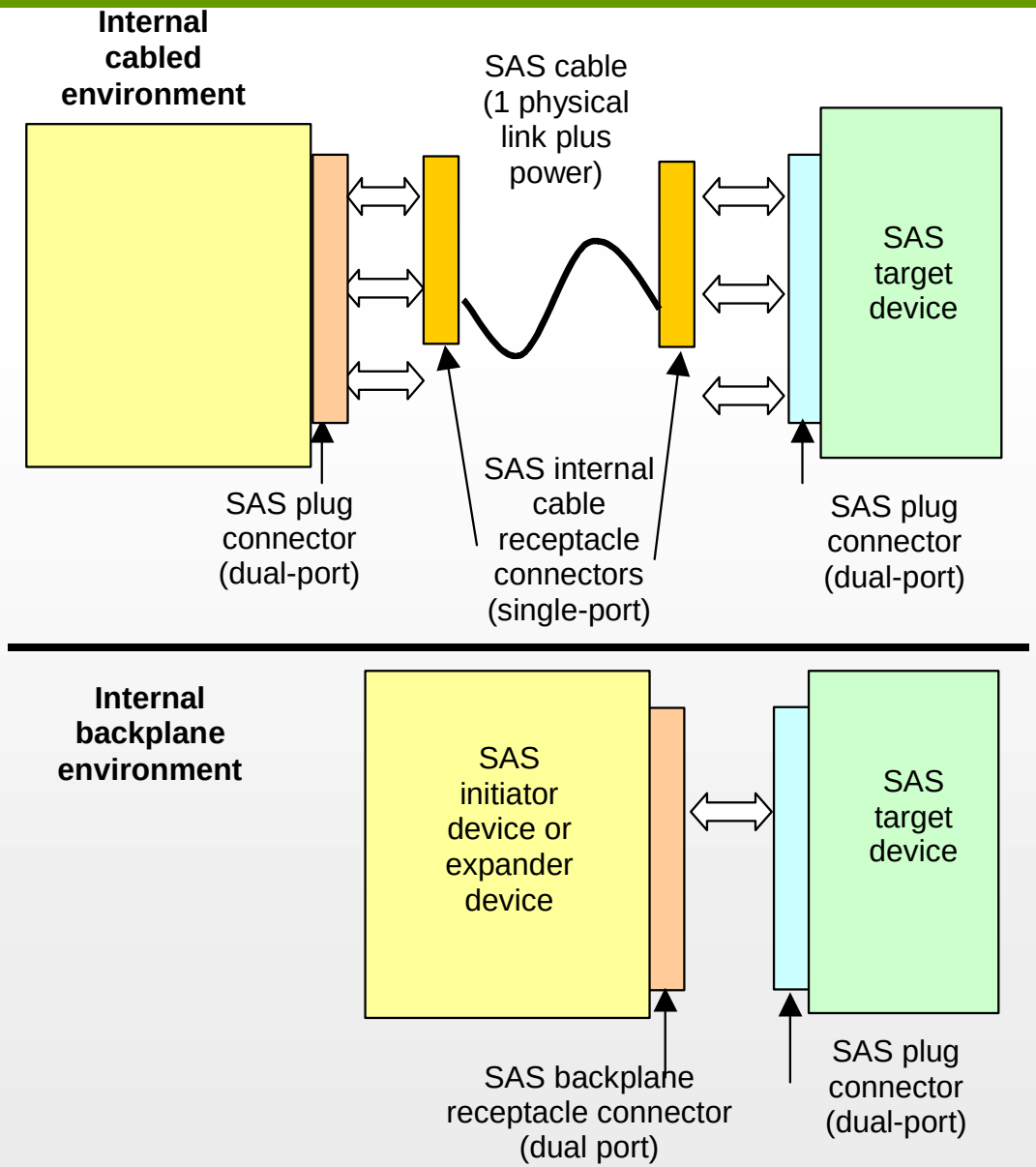
Physical - SATA cables and connectors



Physical - SAS external environment



Physical - SAS internal environments



Physical - Connectors

Connector	Attaches to
SAS plug	SAS internal cable receptacle
	SAS backplane receptacle
SAS internal cable receptacle	SAS plug
	SATA device plug (single-port)
SAS backplane receptacle	SAS plug
	SATA device plug (single-port)
SAS external cable receptacle	SAS external plug
SAS external plug	SAS external cable receptacle

Physical - Cables and connectors

- READY LED pin added to device connector
  - Disk drive output indicating activity
- InfiniBand™ connectors and cables for the external environment
- 10 meter external cable length
- 500 plug events on device connector

Physical - Compliance points

- Compliance points
  - Dt, Dr - SAS disk drive connector
    - Attaches to backplane connector leading to SAS initiator or SAS expander ASIC
  - Ct, Cr - SAS external connector
    - Attaches to other external connectors
- Optional compliance points
  - Xt, Xr - SAS expander ASIC
  - It, Ir - SAS initiator ASIC
    - May be attached to SATA drives or SAS drives

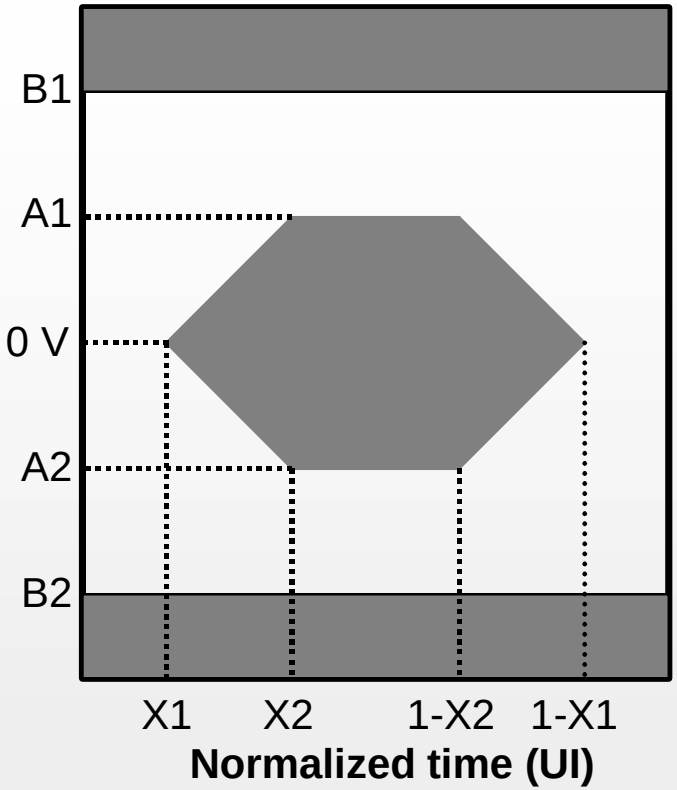
Physical - General electrical characteristics

Characteristic	1.5 Gbps	3.0 Gbps
Data rate	150 Mbps	300 Mbps
Unit interval (UI)	666.667 ps	333.333 ps
Frequency stability for initiator ASICs and expanders supporting SATA 1.0 device with SSC	+350/-5150 ppm	+350/-5150 ppm
Frequency stability for SAS-only compliance points (SAS drives, external connectors)	+100/-100 ppm	+100/-100 ppm
Media impedance	100 ohm	100 ohm

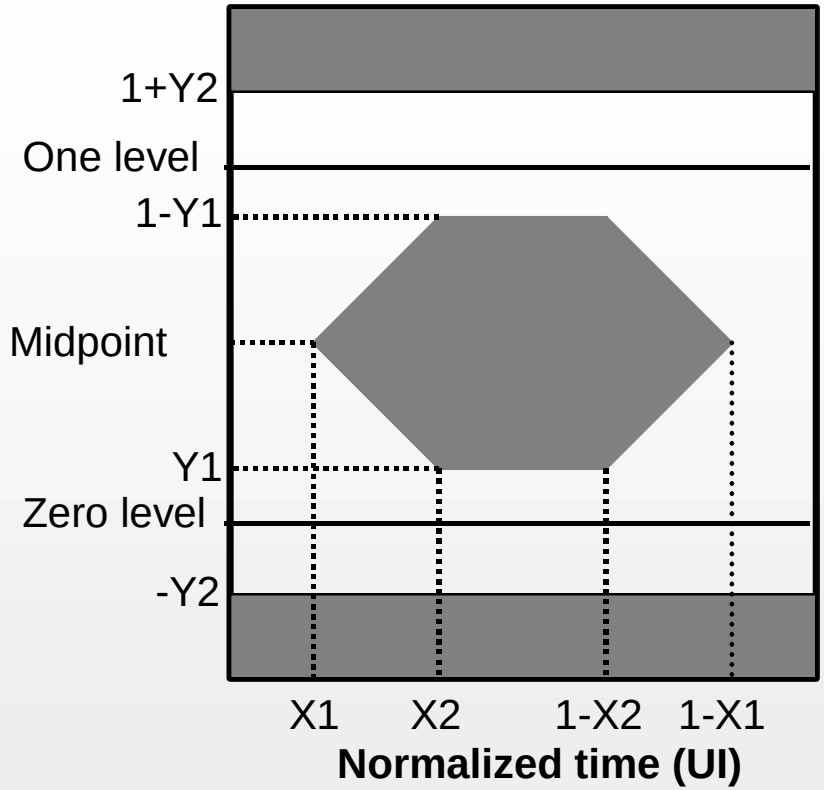
Physical - Eye diagrams

- Amplitude and time based on eye diagrams

Absolute amplitude



Normalized amplitude



Physical - Transmit signal characteristics

Compliance point	Characteristic	SATA	1.5 Gbps	3.0 Gbps
Drive connector	Maximum drive strength (B1-B2)	N/A	**1200 mV**	**1600 mV**
	Minimum drive strength (A1-A2)	N/A	**600 mV**	**800 mV**
External connector	Maximum drive strength (B1-B2)	N/A	**1600 mV**	**1600 mV**
	Minimum drive strength (A1-A2)	N/A	**800 mV**	**800 mV**
Initiator or expander attached to SATA	Maximum drive strength (B1-B2)	**900 mV**	**1200 mV**	**1600 mV**
	Minimum drive strength (A1-A2)	**600 mV**	**600 mV**	**800 mV**

Physical - Receive signal characteristics

Compliance point	Characteristic	SATA	1.5 Gbps	3.0 Gbps
Drive connector	Maximum drive strength (B1-B2)	N/A	1200 mV	1600 mV
	Minimum drive strength (A1-A2)	N/A	**325 mV**	**275 mV**
External connector	Maximum drive strength (B1-B2)	N/A	1600 mV	1600 mV
	Minimum drive strength (A1-A2)	N/A	**275 mV**	**275 mV**
Initiator or expander attached to SATA	Maximum drive strength (B1-B2)	600 mV	1200 mV	1600 mV
	Minimum drive strength (A1-A2)	**225 mV**	**325 mV**	**275 mV**

Physical - Other highlights

- Jitter specs
- 10^{-12} system bit error rate
- Impedance requirements
- AC coupled
- No spread spectrum clocking (SSC)
  - SSC slightly varies the frequency of the transmit clock
  - This reduces EMI at one peak frequency but spreads the emissions over multiple frequencies
  - expander and initiator must tolerate a SATA drive transmitting with SSC, but no SAS component will transmit with SSC
- Non-tracking clock architecture
  - Each device runs off its own internal PLL

...Outline...

- Introduction
- General (devices, domains, …)
- Physical layer (cables, connectors, electrical specs, …)
- ***Phy layer (8b10b, OOB, …)***
- Link layer (primitives, connections, …)
- Transport layer (SSP, STP, and SMP frames)
- Application layer (SCSI mode pages)
- Further information

Phy layer outline

- Encoding
- Out-of-band (OOB) signaling
- Reset sequences
- State machines
- Spin-up

Phy - Encoding

- 8b10b coding
  - As used in SATA, Fibre Channel, et al.
  - **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

Phy - Out-of-band (OOB) signaling

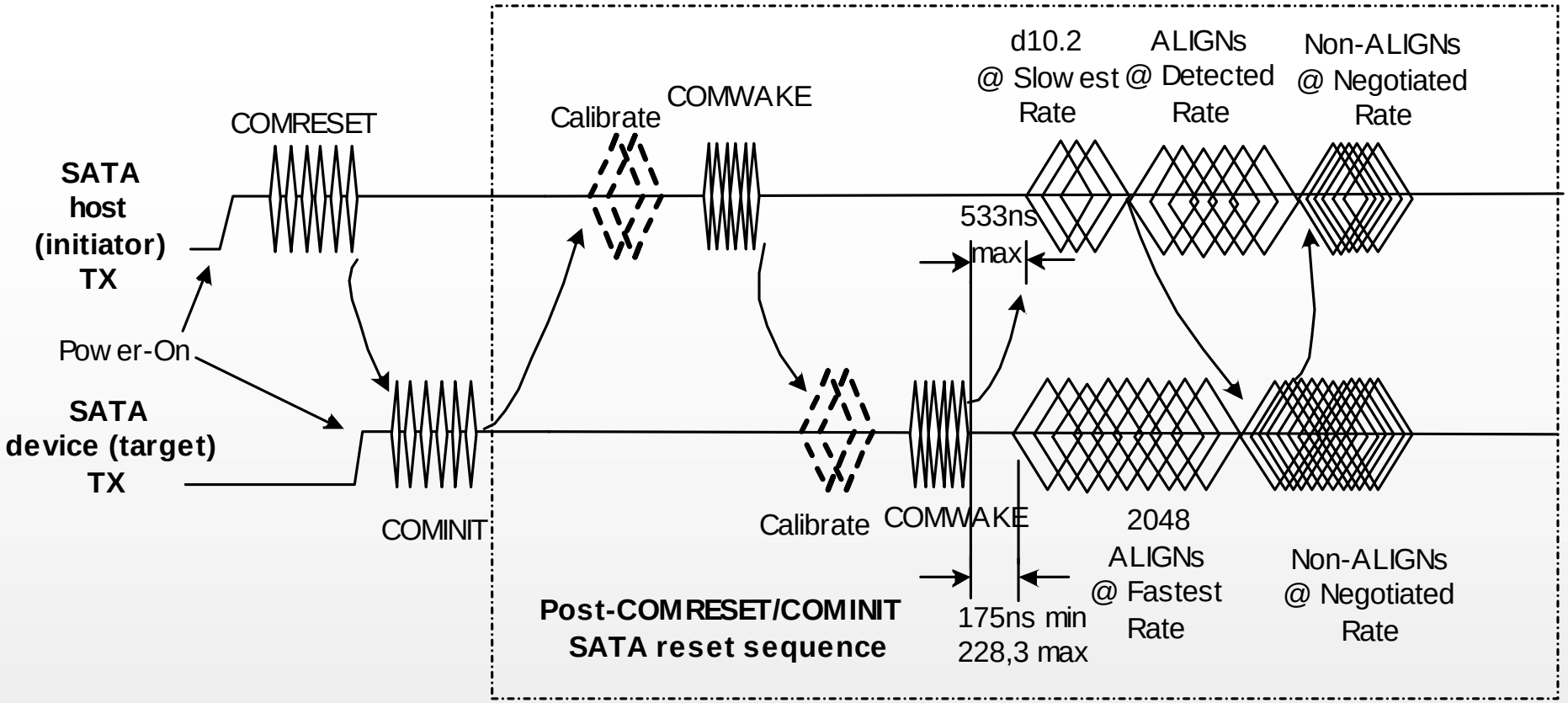
- SATA out of band (OOB) special patterns (“signals”)
  - Signals are sent after power-up to initialize the link
  - Signal is a burst of ALIGN primitives, then idle time; repeated 6 times
  - Detected by squelch detector and frequency comparators
- SATA’s COMRESET/COMINIT, and COMWAKE signals are unchanged
- COMSAS signal added for SAS devices
  - Inserted after calibration sequence before COMWAKE
- If both sides assert COMSAS, then the link is a SAS link rather than a SATA link

Phy - OOB signals

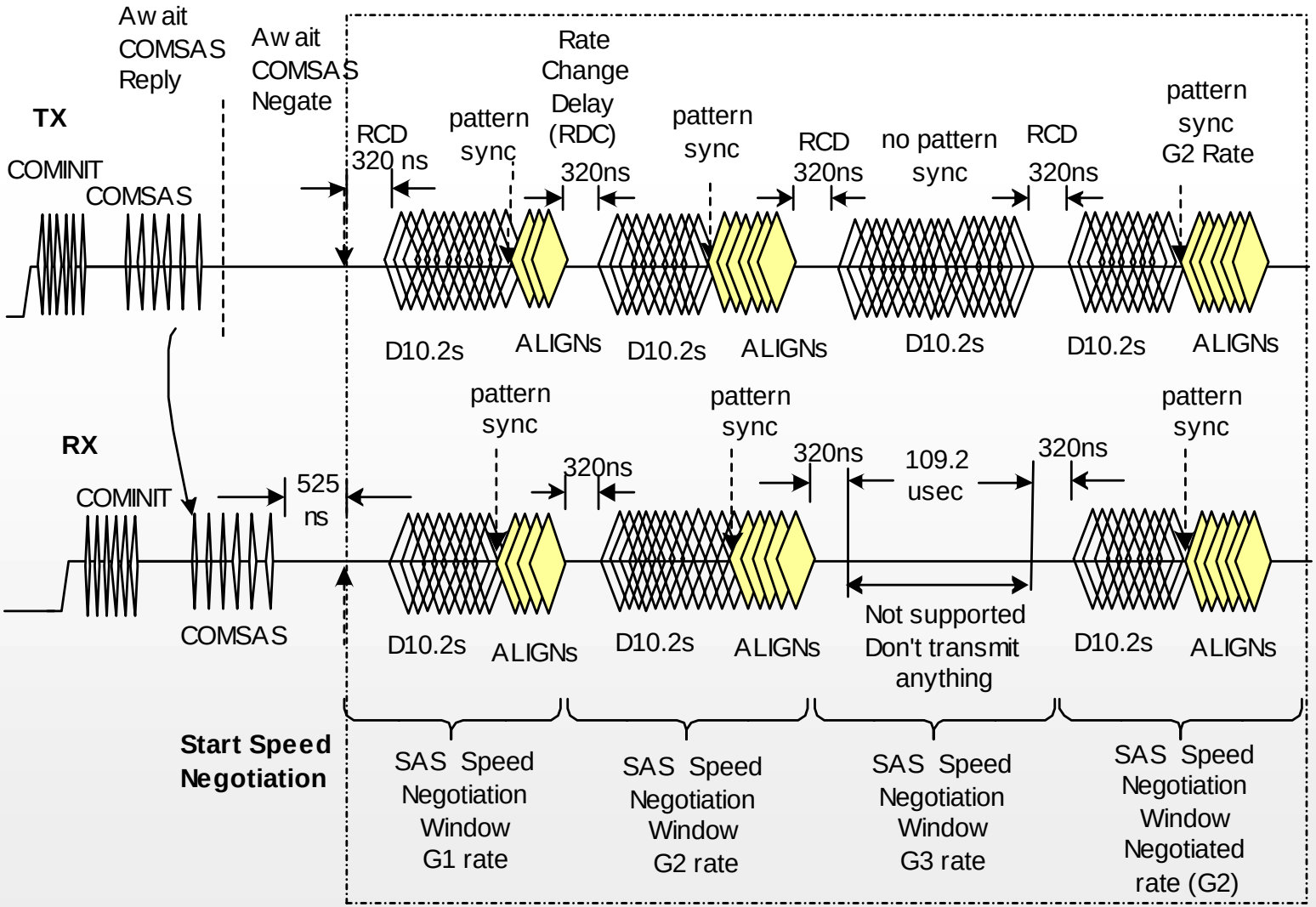
OOB signals

Signal	Nominal burst time	Nominal idle time
COMINIT/ COMRESET	107 ns	320 ns
COMWAKE	107 ns	106.7 ns
COMSAS	***214 ns***	***320 ns***

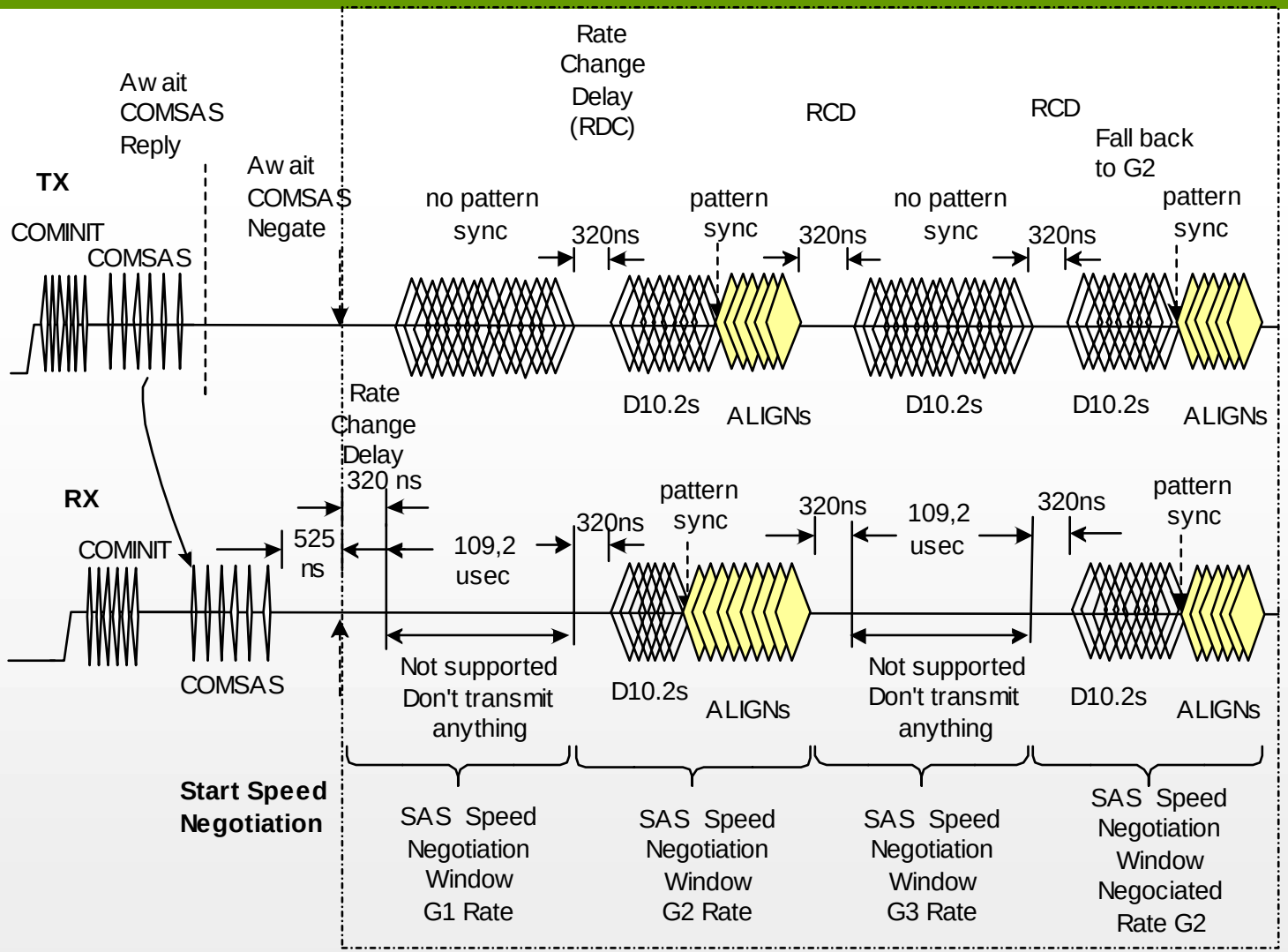
Phy - SATA reset sequence



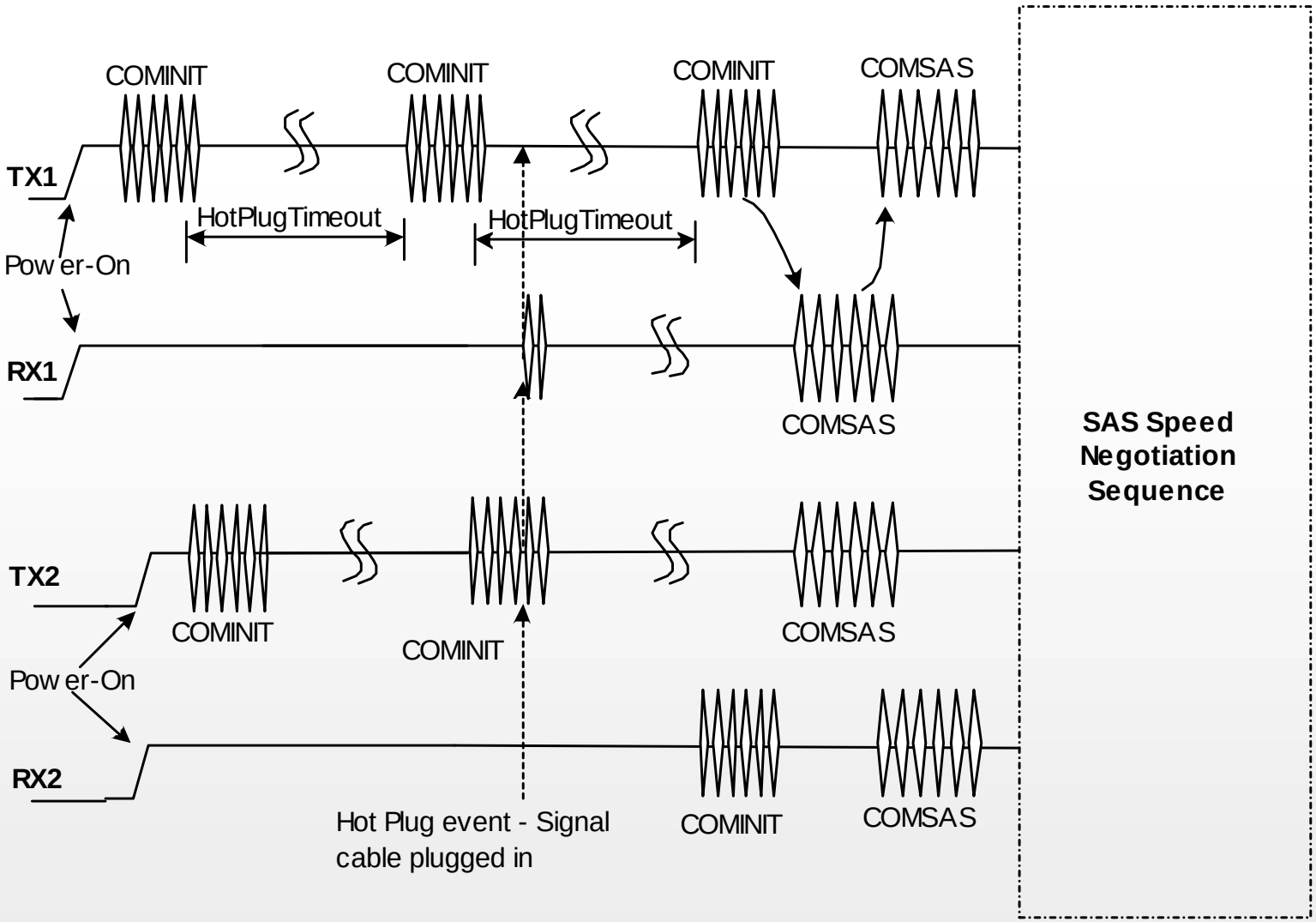
Phy - SAS reset sequence



Phy - SAS reset sequence 2



Phy - Hot plug and the reset sequence



Phy - State machines

- Phy state machine
  - SATA and SAS reset sequence
- Dword synchronization state machine
  - Determines when link is gone bad

Phy - Spin-up

- Desktop (ATA) goal - boot quickly, spin-up ASAP
  - ATA Power Up in Standby feature rarely implemented
- Enterprise goal - stagger spin-up to avoid excessive power drain
  - Delayed start feature with SCA-2 connector
- Rack of SATA drives may overwhelm power supplies
- SAS rules/recommendations:
  - SAS-capable SATA devices shall spin-up only after reset sequence
  - SATA devices should spin-up only after reset sequence
  - SAS devices shall not spin-up until START STOP UNIT is run

…Outline…

- Introduction
- General (devices, domains, …)
- Physical layer (cables, connectors, electrical specs, …)
- Phy layer (8b10b, OOB, …)
- ***Link layer (primitives, connections, …)***
- Transport layer (SSP, STP, and SMP frames)
- Application layer (SCSI mode pages)
- Further information

Link layer outline

- Primitives
- Idle links
- Power management
- SATA loopback tests
- Tests
- Wide links
- Multiplexing
- Domain management
- Rate matching
- Elasticity buffers
- Scrambling
- Fabric management
- Connections
- Frame transmission
- Flow control
- SSP flow control
- STP flow control
- Asynchronous event notification

Link - Primitives

- 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

Link - SAS primitives

- SAS primitives
  - AIP
  - ALIGN()
  - BREAK
  - CHANGE
  - CLOSE
  - EOAF
  - HARD_RESET
  - OPEN_ACCEPT
  - OPEN_REJECT()
  - SOAF
- SSP/SMP primitives
  - ACK
  - DONE()
  - EOF
  - NAK()
  - RRDY
  - SOF

Link - SATA primitives

- SATA primitives
  - SATA_CONT
  - SATA_DMAT
  - SATA_EOF
  - SATA_HOLD
  - SATA_HOLDA
  - SATA_PMACK
  - SATA_PMNAK
  - SATA_PMREQ_P
  - SATA_PMREQ_S
  - SATA_R_ERR
  - SATA_R_IP
  - SATA_R_OK
  - SATA_R_RDY
  - SATA_SOF
  - SATA_SYNC
  - SATA_WTRM
  - SATA_X_RDY

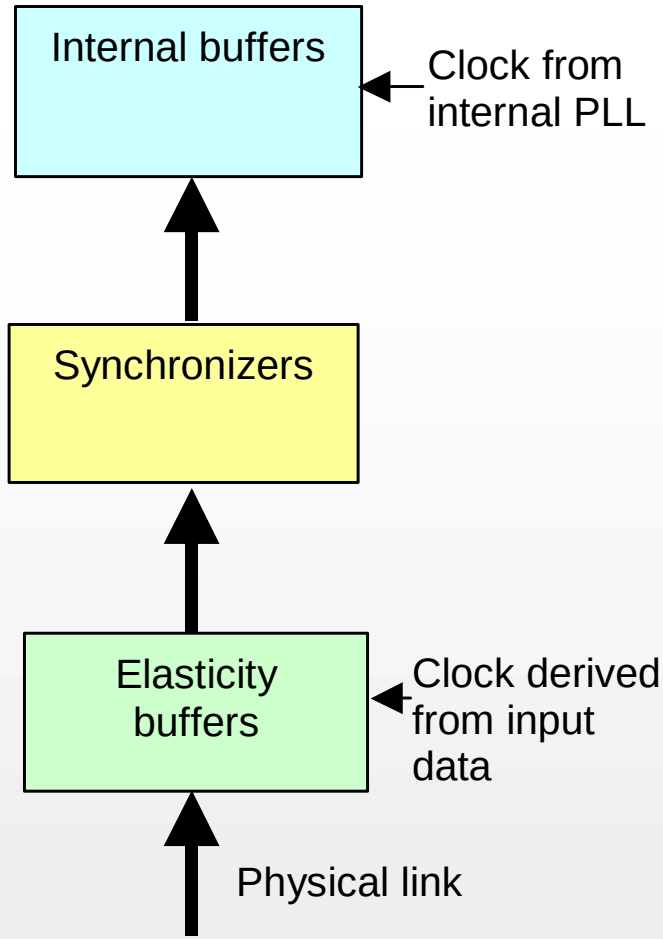
Link - Primitive sequences

- Some primitives are sent more than one for more reliable delivery - tolerate single bit errors
- “Repeated” is for SATA primitives

Type	Send	Detect
Single	1	1
Repeated	2	1
Triple	3	3
Redundant	6	3

Link - Clock skew management

- Input data clock does not exactly match the internal clock
  - Overflow = sender faster than receiver
  - Underflow = Sender slower than receiver
- ALIGN primitives added to data stream
  - Receiver throws them out
  - 1 per 2048 dwords
  - 2 per 256 extra for STP



Link - Idle links

- Between connections, or within an SSP or SMP connection between frames, idle dwords are sent
  - Idle dword = random scrambled data
- During an idle STP connection, SATA_SYNC is sent
  - Usually followed by SATA_CONT and random scrambled data

Link - IDENTIFY address frames

- IDENTIFY address frame sent after reset sequence

	7	6	5	4	3	2	1	0
0	Reserved				FRAME TYPE (0h)			
1	PHY IDENTIFIER							
2	Reserved				MAX LINK RATE			
3	DEV TYPE		SSP_I	STP_T	SSP_I	SSP_T	SMP_I	SMP_T
4 19	Reserved							
20 27	DEVICE NAME							
28 31	CRC							

Link - OPEN address frame

	7	6	5	4	3	2	1	0
0	INIT		PROTOCOL		FRAME TYPE (1h)			
1	Reserved				LINK RATE			
2 3	INITIATOR CONNECTION TAG							
4 9	Reserved							
10 11	ARBITRATION WAIT TIME							
12 19	DESTINATION DEVICE NAME							
20 27	SOURCE DEVICE NAME							
28 31	CRC							

Link - Power management

SAS support for power management

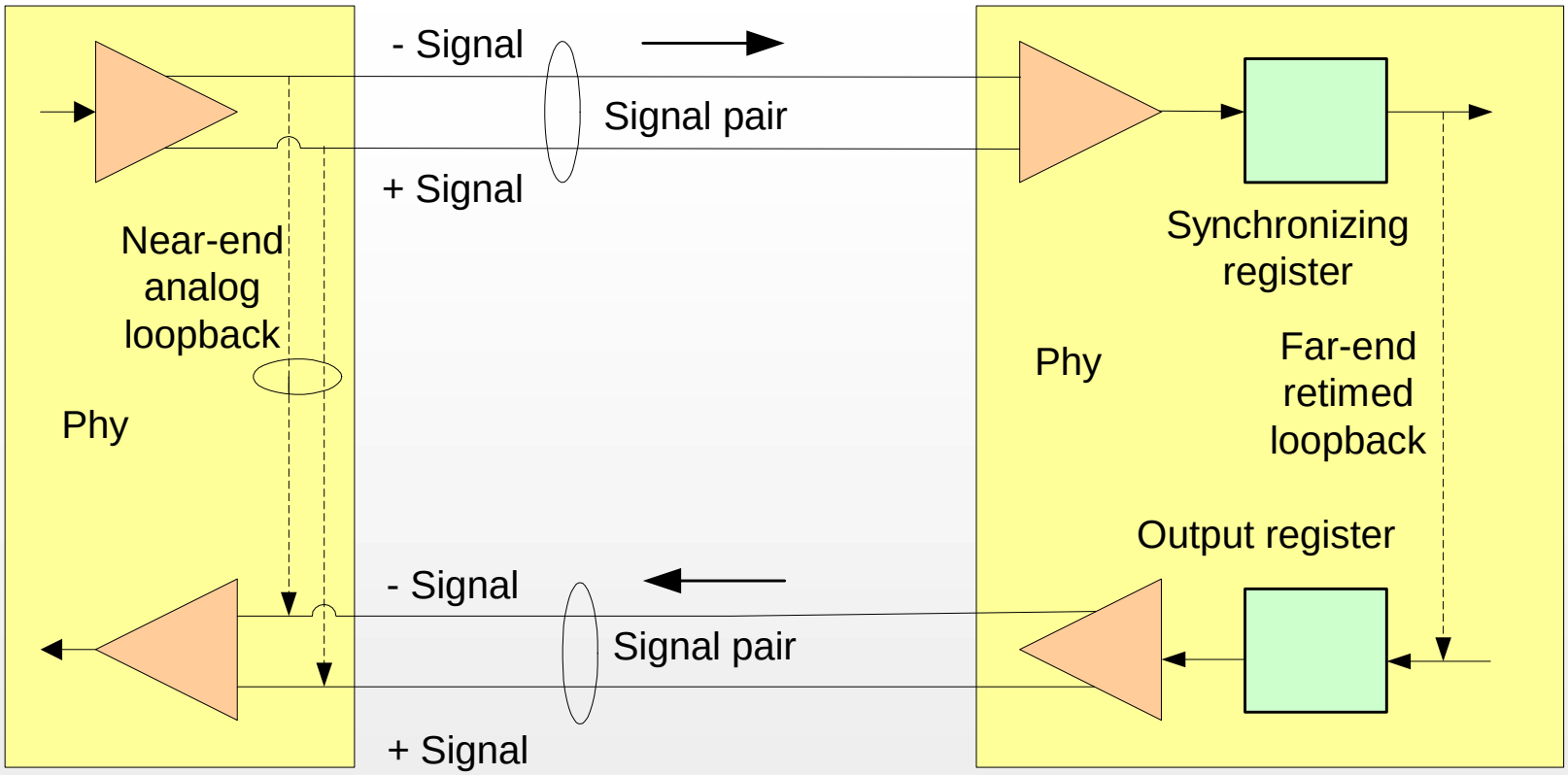
ATA device power management	SATA interface power management	SCSI power conditions
Idle (yes)	Partial (no)	Idle (yes)
Standby (yes)	Slumber (no)	Standby (yes)
Sleep (no)		Sleep (no)

Link - Power management 2

- ATA device power management
  - Idle, Standby, Sleep
  - SAS: Idle and Standby supported; Sleep shall not be requested
- SATA interface power management
  - Partial, Slumber states
  - PMREQ_P, PMREQ_S, PMACK, PMNAK
  - SAS: Initiator shall not request interface power management
  - SAS: Initiator shall reply with PMNAK to PMREQ_P and PMREQ_S
- SCSI power management
  - Idle, Standby, Sleep
  - SAS: Idle and Standby supported; Sleep not supported

Link - SAS loopback test modes

- Invoked with SMP



Link - Domain changes

- Expander broadcasts CHANGE primitive to notify initiators and expanders that
  - Phy has lost dword sync
  - Reset sequence completed on a phy
  - CHANGE received
- Initiators perform level-order traversal of domain
- Fanout expanders run DISCOVER function to fetch routing tables from edge expanders

Link - Domain management 2

- CHANGE broadcast from expanders to all initiator ports
  - Indicates some port has changed state (e.g. hot plug, loss of bit sync)
  - Initiators should rescan the fabric looking for changes
- Special primitives to configure expander devices
  - Start OOB reset sequence on a port
  - Disable a port (e.g. if a loop is detected with WWNs)
  - Enable rate matching
  - Set up multiplexing

Link - Scrambling

- 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 and helps DC balance
  - Spread spectrum clocking addresses EMI for all patterns
  - 8b10b coding addresses DC balance

Link - Scrambling 2

- Repeated primitives (STP)
  - Replaces repeated primitives
  - <prim>, <prim>, … SATA_CONT, <random data>, <random data>, …, <new prim>
  - ALIGNs may be inserted inside random data
  - Exit with any primitive except ALIGN
- Idle dwords (SAS, SSP)
- Frame data (SAS address frames, SSP, SMP, and STP frames)
  - All data dwords in frames are scrambled (between SOF and EOF)
  - Primitives inside frames are NOT scrambled
  - Polynomial reset every SOF or SATA_SOF
- SAS big-endian, SATA little-endian polynomial
  - Doesn't matter for random/idle data, does matter for frame data

Link - Connections

- 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

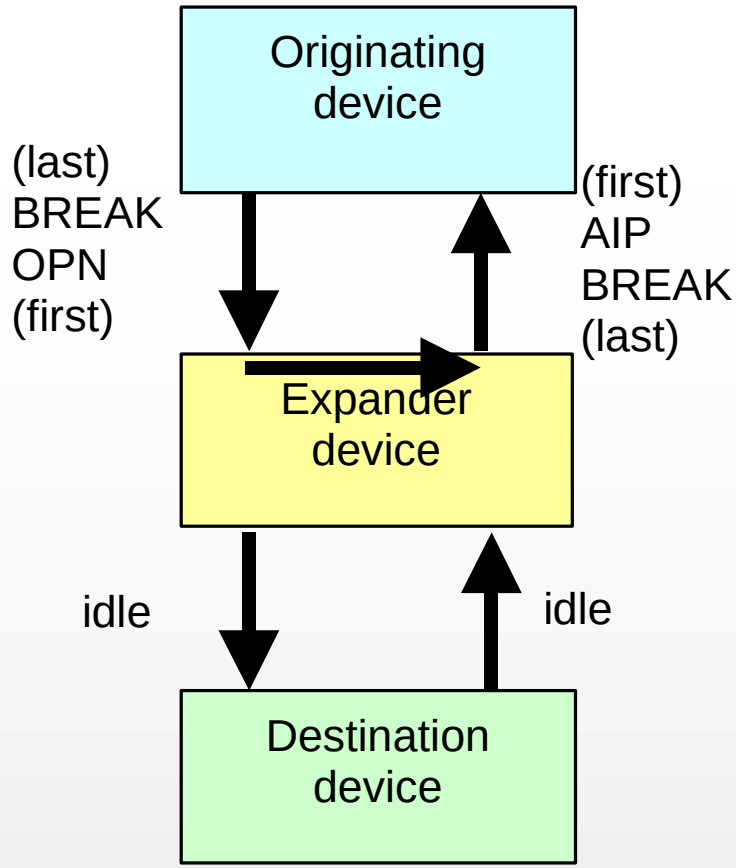
Link - Opening a connection

- 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

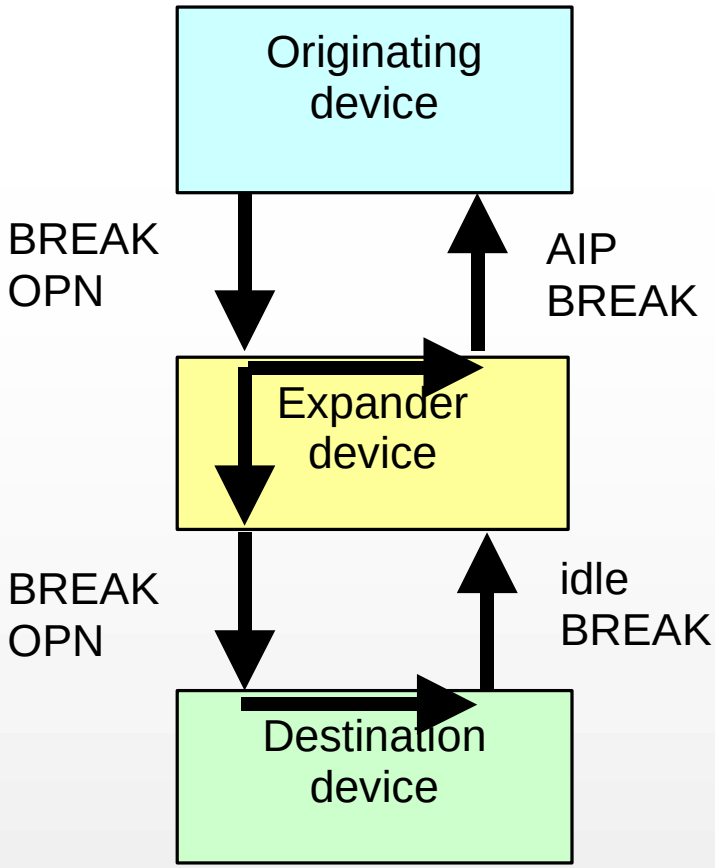
Link - Breaking a connection

- BREAK primitive signals a unilateral close of the connection
- BREAK response allowed but not required
- Expander tears down connection when it sees BREAK

Link - Breaking a connection 2



(if expander device has not forwarded the OPN to the destination)



(if expander device has forwarded the OPN to the destination)

Link - Closing a connection

- After no more data is being sent on the connection, it may be closed by either side
- CLOSE primitive sent; CLOSE received
- Expander tears down the connection when it sees CLOSE in both directions

Link - Closing an SSP connection

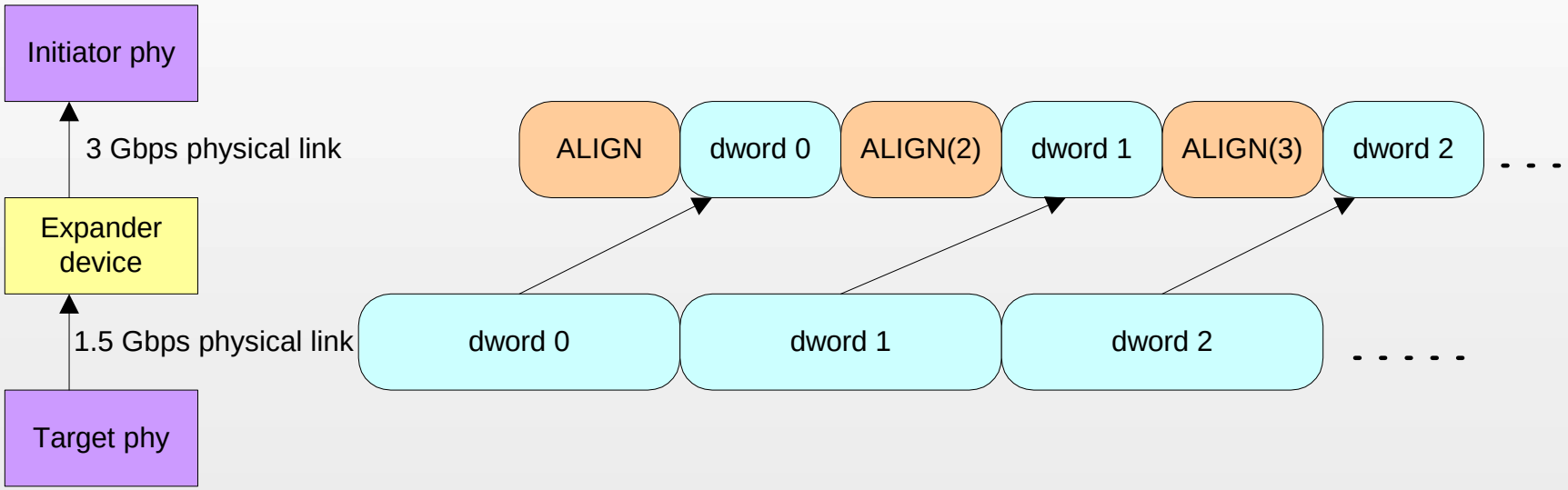
- Before sending CLOSE, must guarantee that the SSP traffic is finished
- DONE primitive indicates sender is done originating frames
- Works like FC-AL DHD (dynamic half duplex)
- Back channel may still be active
  - sender may send ACK, NAK, RRDY after DONE to keep the other direction active
- When both sides have sent DONE, the connection is idle and CLOSE can be exchanged

Link - State machines

- SAS endpoint connection management state machine (SL)
- SAS expander connection management state machines (XL)

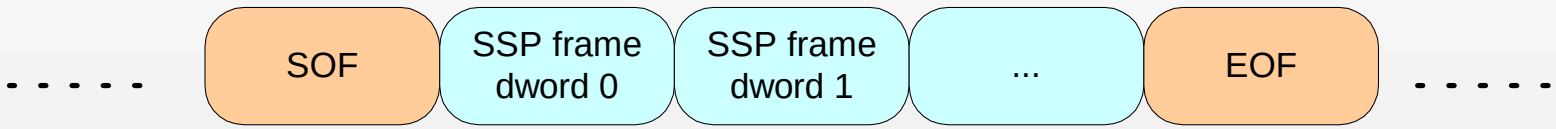
Link - Rate matching

- 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



Link - SSP (SCSI)

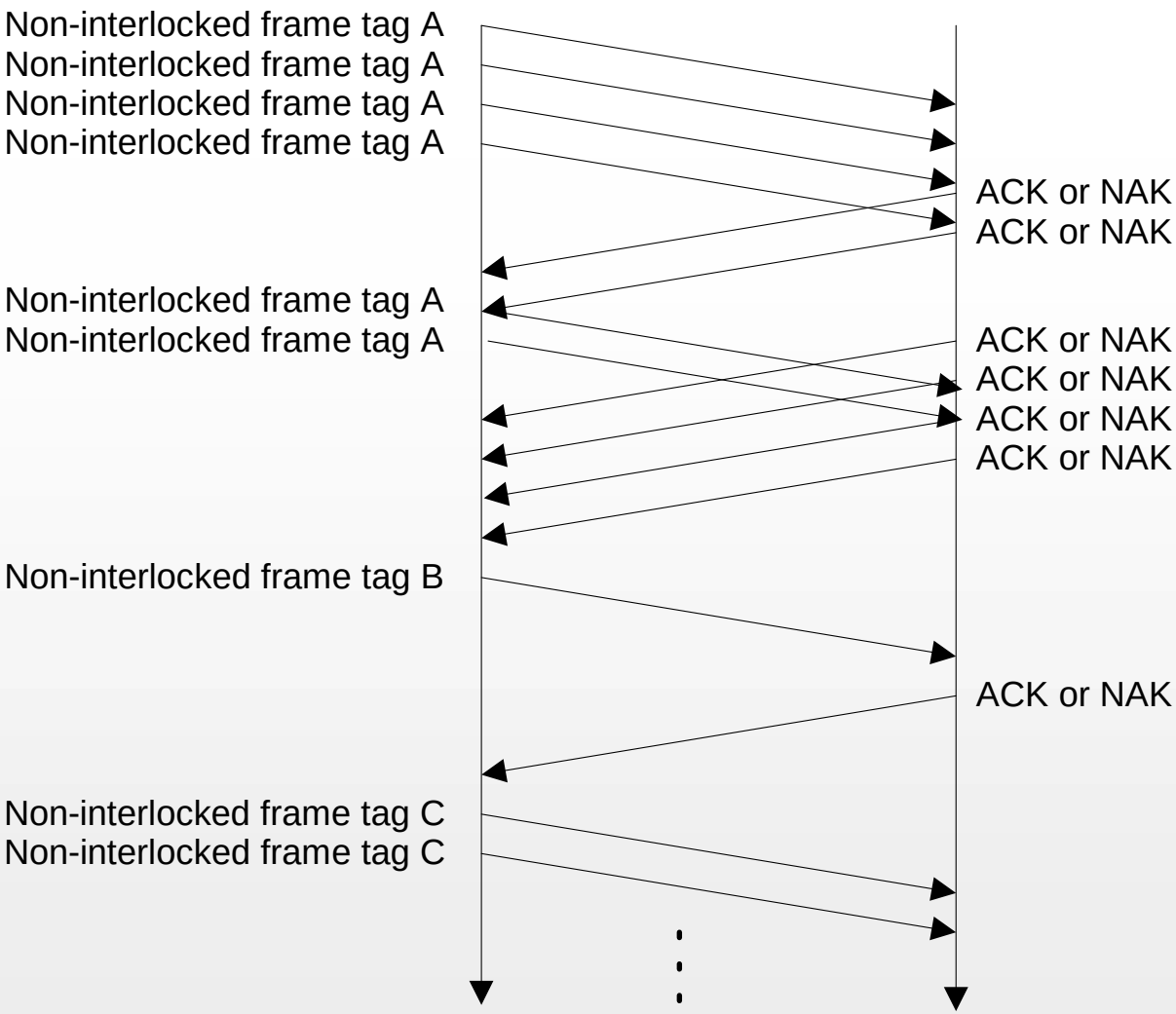
- Full duplex
- SOF, frame dwords, CRC, EOF
- Each frame acknowledged with ACK, NAK
- Credit with RRDY
- SSP link layer state machine



Link - SSP flow control

- Credit maintained by each connected port
  - Initialized to 0
  - Receive RRDY -> increase by 1
  - Transmit frame -> decrease by 1
  - Maximum 255
- Interlocked frames
  - COMMAND, TASK, XFER_RDY, RESPONSE, AER, AER_RESPONSE frames with different tags
  - Must receive ACK or NAK before sending another, regardless of credit
- Non-interlocked frames
  - DATA frames
  - If same tag, may send without waiting for ACK or NAK, provided credit is available

Link - SSP interlocking



Link - STP (ATA)

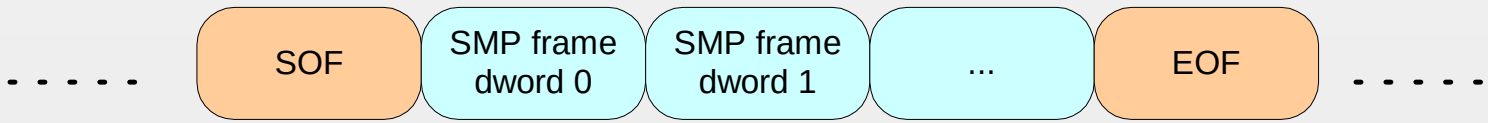
- STP (ATA) connection
  - STP from initiator to last expander
  - SATA from the expander to the SATA device
  - After an STP connection is opened, follow SATA rules
  - Frame sent as: SATA_SOF, SATA frame, SATA_EOF
  - Each frame receives SATA_R_OK or SATA_R_ERR
  - SATA_X_RDY/SATA_R_RDY for permission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

Link - STP flow control

- During an STP connection, SATA flow control operates as defined by SATA
  - SATA_HOLD and SATA_HOLDA primitives
- SATA targets accept 20 dwords after HOLD
- STP initiators shall accept 128 dwords after HOLD
- Expanders must insert HOLD/HOLDA themselves if they add latency
  - Must guarantee the 20 dword/128 dword rules

Link - SMP

- Simplified version of SSP
- No ACK, NAK, RRDY, or DONE
- Initiator opens and closes connection
- Send one AER_REQUEST
- Receive one AER_RESPONSE
- SMP link layer state machine



Outline

- Introduction
- General (devices, domains, …)
- Physical layer (cables, connectors, electrical specs, …)
- Phy layer (8b10b, OOB, …)
- Link layer (primitives, connections, …)
- ***Transport layer (SSP, STP, and SMP frames)***
- Application layer (SCSI mode pages)
- Further information

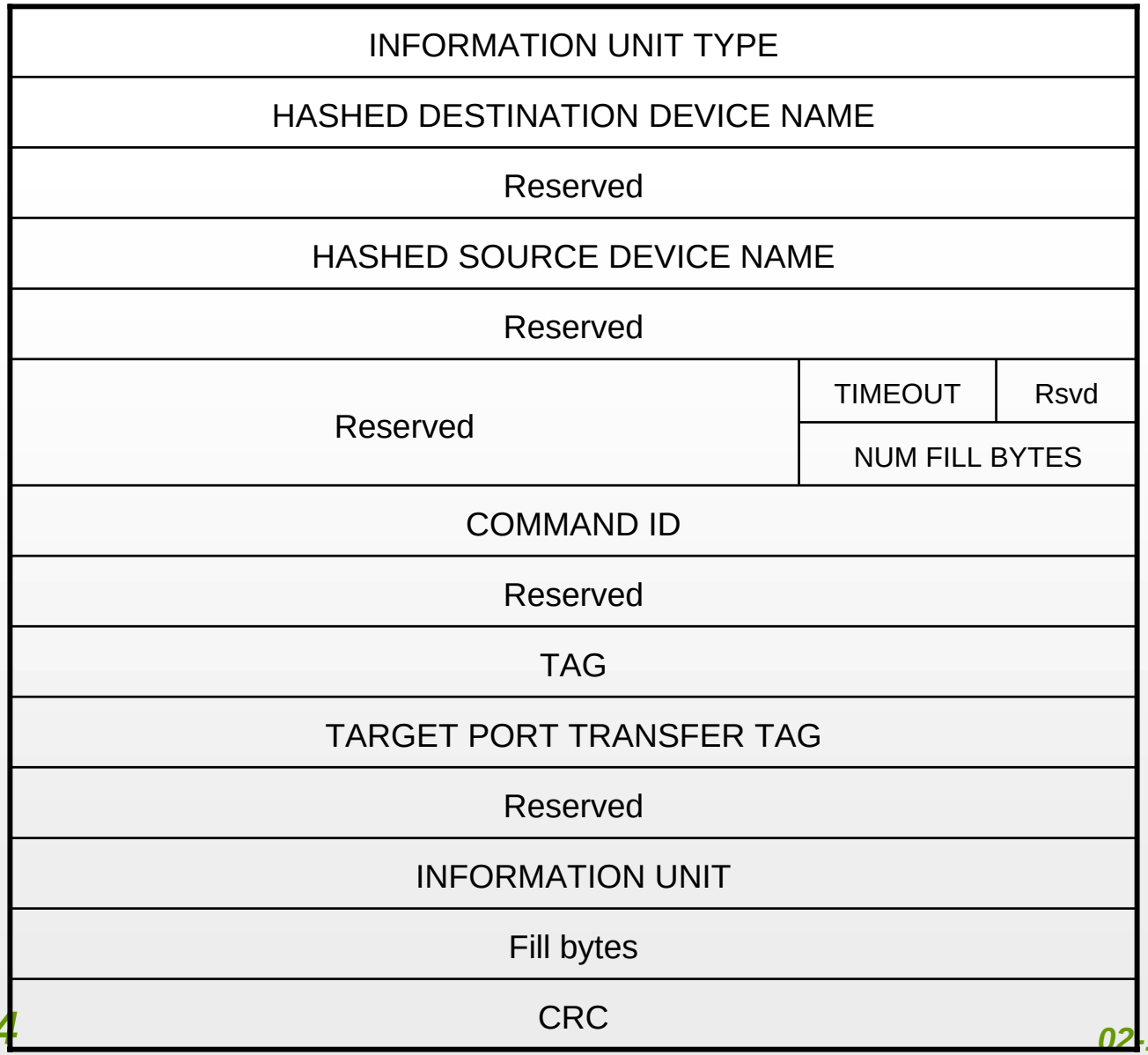
Transport layer outline

- SSP (SCSI) frame format
- SSP information units
- SSP information unit sequences
- SSP TASK IU notes
- SSP information unit notes
- STP (ATA)
- SMP
- Port Control state machine

Transport - SSP frame format

- Based on Fibre Channel and FCP-2
- Lots of reserved fields
  - No exchanges
  - No sequences
  - Ack, Nak, etc. handled with primitives, not frames
- The *only* fields used in the outer frame:
  - Frame type (FCP calls this R_CTL)
  - Fill bytes
  - Tag (FCP calls this OX_ID)
  - CRC
  - Maybe the Source_ID and Destination_ID (TBD)
- Frame payload carries information units

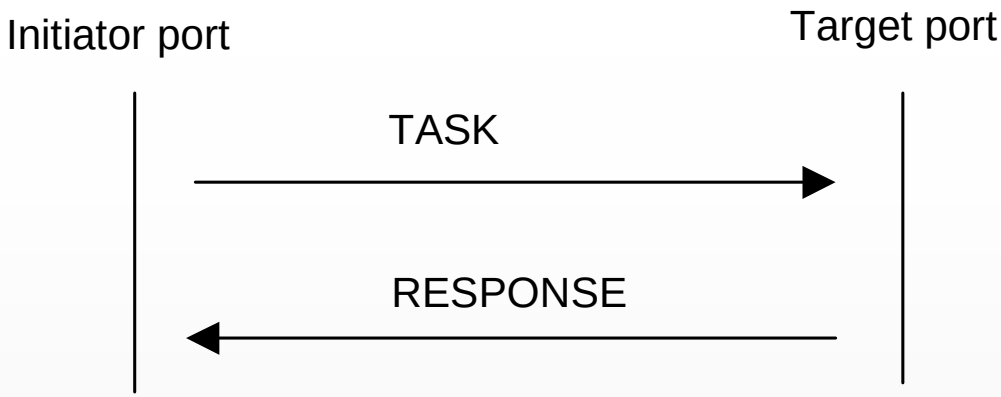
Transport - SSP frame format 2



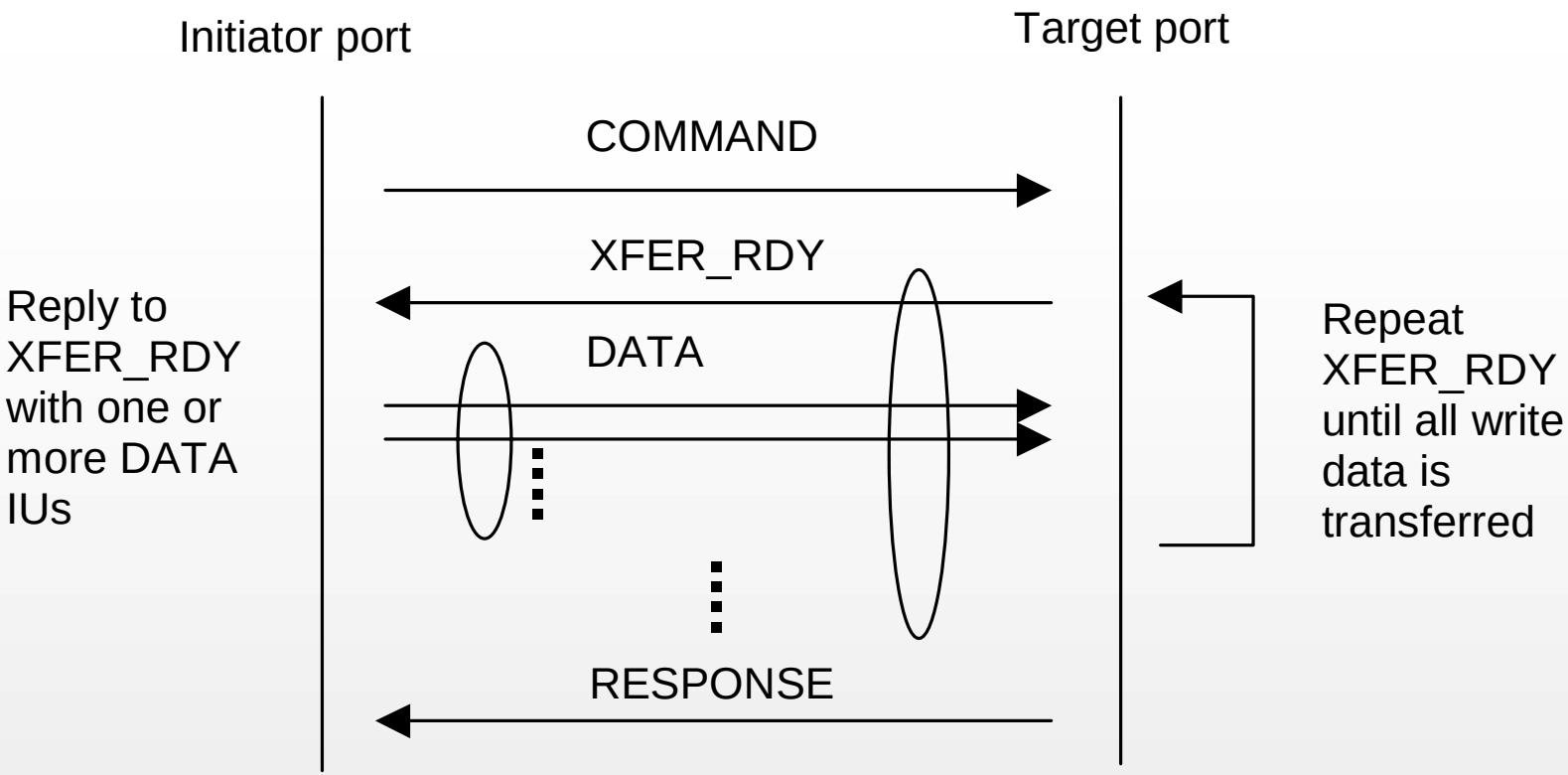
Transport - SSP information units

IU	Originator
COMMAND	Initiator port
TASK	Initiator port
XFER_RDY	Target port
DATA	Initiator port (writes) or target port (reads)
RESPONSE	Target port
AEN	Target port
AEN_RESPONSE	Initiator port

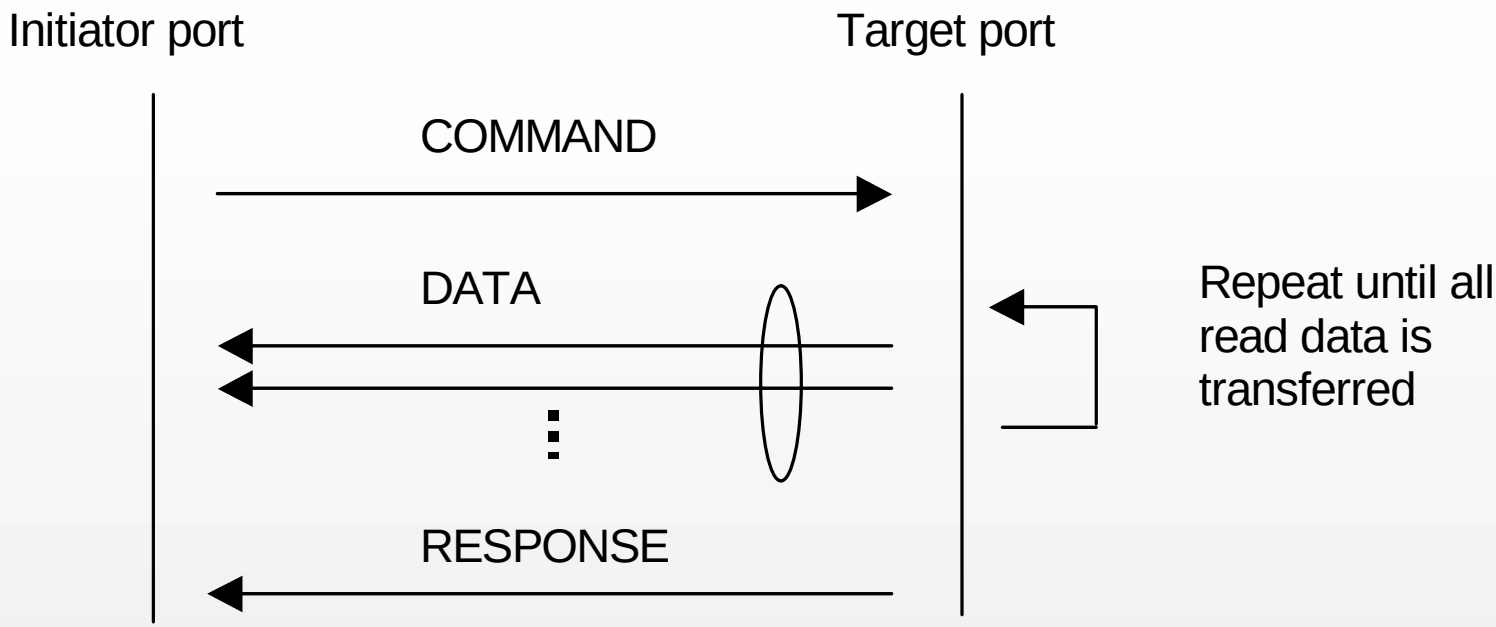
Transport - SSP Task Management sequence



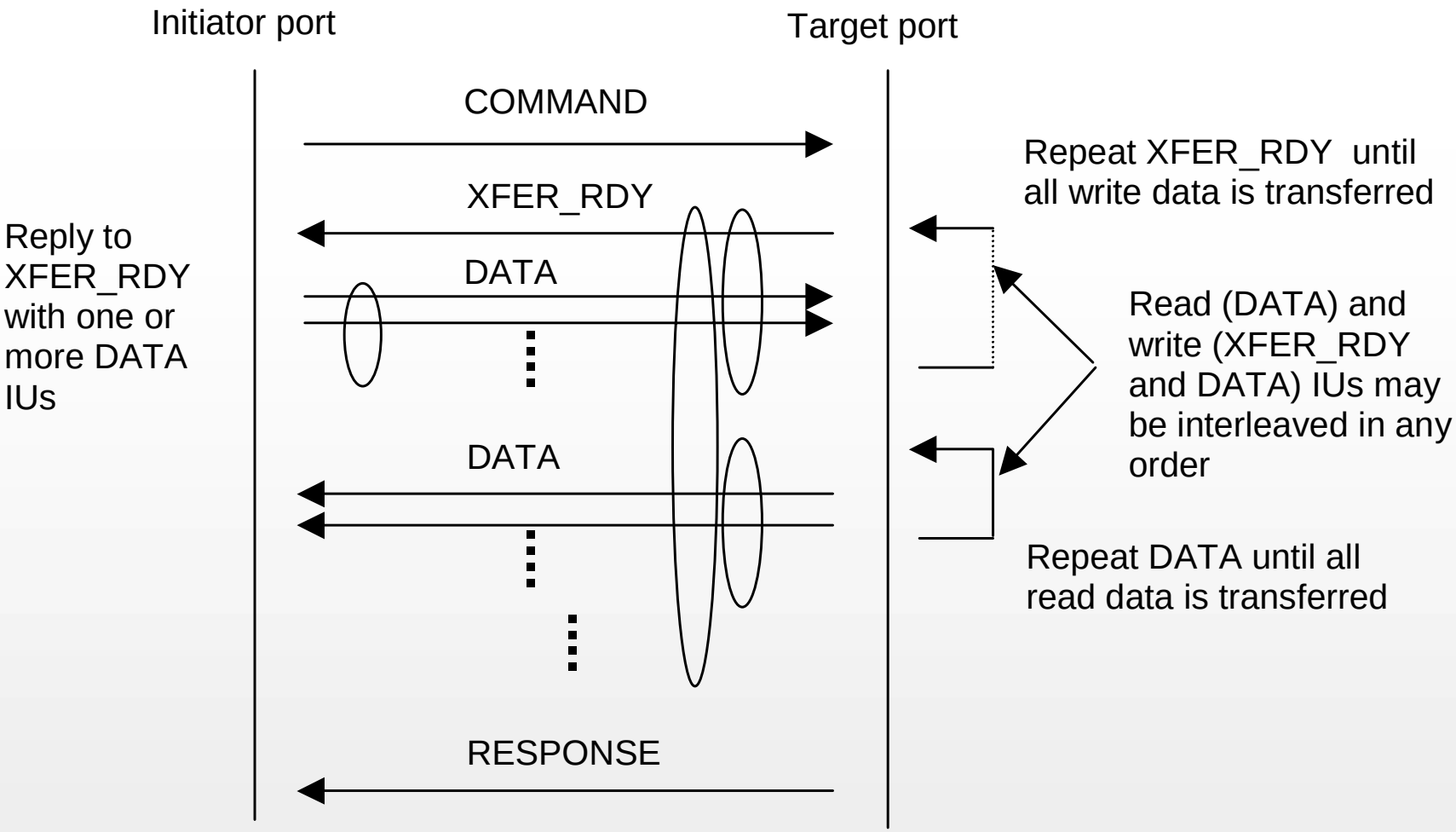
Transport - SSP Write sequence



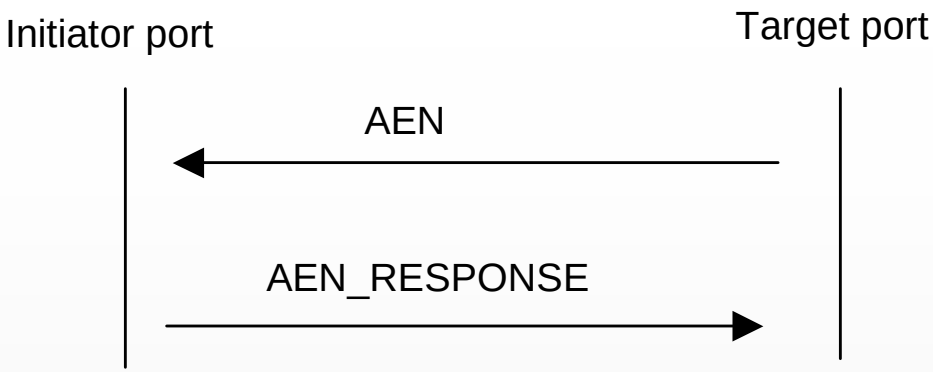
Transport - SSP Read sequence



Transport - SSP Bidirectional sequence



Transport - SSP AEN sequence



Transport - SSP TASK IU rationale

- ABORT TASK references the tag of the task to be aborted
  - No other task management functions reference tags
- FCP uses Abort Sequence to implement ABORT TASK so didn't face this issue
  - SAS does not have Sequences
- The ABORT TASK itself could be rejected or have errors; needs its own tag
- Two tags don't fit in COMMAND IU cleanly
- Separate IU implemented like iSCSI and SRP

Transport - SSP Information unit notes

- Only COMMAND and TASK contain the LUN
  - Tag must be target-wide, no reused for different LUNs
- No residuals in RESPONSE
- COMMAND supports bidirectional commands and variable length CDBs
- No TARGET RESET or WAKEUP task management functions
  - HARD RESET primitive is the low-level debug reset

Transport - SSP Asynchronous event notification

- SCSI asynchronous events
  - Initialization complete unit attentions (after power on) ***NOT SUPPORTED***
  - All other unit attentions (e.g. SMART events)
  - Deferred errors (e.g. cached write failed)
- Target port sends AEN to all initiators it knows when an asynchronous event occurs
  - Contains LUN - optional for initiator to use
  - REPORT LUNS enhanced to identify which logical units have asynchronous events pending if LUN is ignored
- REQUEST SENSE to retrieve the sense data

Transport - SSP Asynchronous event notification 2

- The “official” way in pSCSI and FCP:
  - disk drive becomes an initiator
  - HBA becomes a target with a “processor” device type
  - Disk drive uses SEND command to send sense data
  - Nobody implements this
  - Workarounds
    - Leave a vendor-specific command outstanding forever
    - Run unnecessary commands that can carry the sense data
- AEN and REPORT LUNs much more efficient

Transport - STP (ATA)

- SATA Target sends frame
  - Expander detects SATA_X_RDY
  - Expander arbitrates and generates OPN to initiator
  - Expander passes through SATA until it sees SATA_WTRM in one direction and SATA_IDLE/SATA_SYNC in the other
  - Expander may close connection
- STP initiator sends frame
  - Wraps frame in OPN/CLS
  - May leave connection open to send more frames

Transport - SMP

- SMP_REQUEST frame
- SMP_RESPONSE frame
- SMP state machine

Transport - SMP functions

Function	Description
DISCOVER	Used by fanout expander
REPORT GENERAL	General info
REPORT SATA CAPABILITIES	STP/SATA support info
REPORT PHY	Phy-related info
REPORT PHY ERROR LOG	Counters of # errors detected
REPORT PHY SATA	STP/SATA phy-specific state
REPORT PHY DEVICE NAMES	Topology management
PHY CONTROL	Request loopback test modes, reset sequence, HARD RESET

Transport - Port control state machine

- Port control (PC) state machine
- Sits between multiple transport layer state machines (e.g. SSP, STP, and SMP) and multiple link layer state machines (e.g. for wide links)

...Outline...

- Introduction
- General (devices, domains, …)
- Physical layer (cables, connectors, electrical specs, …)
- Phy layer (8b10b, OOB, …)
- Link layer (primitives, connections, …)
- Transport layer (SSP, STP, and SMP frames)
- ***Application layer (SCSI mode pages)***
- Further information

Application layer outline

- SCSI
- ATA

Application - SCSI

- Disconnect-reconnect mode page protocol-specific fields
  - Supported fields
    - Bus inactivity limit - n * 100ms
    - Maximum connect time limit - n * 100 ms
    - Maximum burst size - n * 512 bytes - devices may burst this much
    - First burst size - n * 512 bytes - implicit XFER_RDY for each new command
  - Not supported
    - Buffer full/empty ratios - no - devices decide on their own
    - Enable modify data pointers - no - all transfers must be in order

Application - SCSI

- Protocol-specific mode page
  - I_T Nexus loss time
- No protocol-specific log pages
- No protocol-specific commands

Application - ATA

- SATA targets must work without changes
- STP initiators add the concept of addressing to ATA
  - Initiator may present a standard ATA register interface over PCI-X, one per target
  - Alternate interfaces are also possible

Further information

- INCITS T10 (SCSI)
  - http://www.t10.org
  - Home of the SAS standard
- INCITS T13 (ATA)
  - http://www.t13.org
- Serial ATA Working Group
  - http://www.serialata.org
- SCSI Trade Association
  - http://www.scsita.org
- Original Serial Attached SCSI Working Group
  - http://www.serialattachedscsi.com

End of technical overview

Serial Attached SCSI