

Serial Attached SCSI Comparison to Fibre Channel with FCP



by Rob Elliott

HP Industry Standard Servers

Server Storage Advanced Technology

elliott@hp.com <http://www.hp.com>

30 September 2003

- These slides are freely distributed by HP through the SCSI Trade Association (<http://www.scsita.org>)
- STA members are welcome to borrow any number of the slides (in whole or in part) for other presentations, provided credit is given to the SCSI Trade Association and HP
- This compilation is © 2003 Hewlett-Packard Corporation



Overview



- SAS borrows its lower layers from SATA
- SAS borrows its upper layers from Fibre Channel and FCP (the SCSI over Fibre Channel mapping)

FC layer	Standard(s)	Equivalent SAS layer(s)	Standards
ULP	FCP (and various SCSI)	SCSI application	SAS & others
FC-4 Mapping	FCP	Transport	SAS
FC-3	FC-FS	Port layer (roughly)	SAS
FC-2 Protocol	FC-FS	Link and some transport	SAS
FC-1 Code	FC-FS	Phy	SAS
FC-0 Physical	FC-PI	Physical	SAS

- SAS also supports ATA and management upper layers
- FC also supports VI, TCP, and others

Physical/phy layers



FC	SAS
<p>Media: Copper and fiber Up to 100 km with optical</p>	<p>Media: Copper only 1 m internal cable or 18" backplane trace 10 m external cable (4 wide)</p>
<p>Speed: 1.0625, 2.125, 4.25, and 10.51875 or 12.75 Gbps</p>	<p>Speed: 1.50, 3.00 Gbps</p>
<p>Both use 8b10b coding Both use dwords (4 bytes) as base quantities Both are big-endian (SATA is little-endian)</p>	

Disk drive connectors



FC	SAS
SCA-2 (40 pins)	Dual-ported version of the SATA connector (7+7+15 pins)
5 V, 12 V	3.3 V, 5 V, 12 V
START pins for spinup control	NOTIFY (ENABLE SPINUP) primitives
FC includes ESI (enclosure services interface) pins for access to attached enclosure management through a disk drive	Enclosure management must be accessed elsewhere (e.g. via expanders)
Both have the serial lines (Tx +, Tx -, Rx +, Rx -) for two ports	
Both have a READY LED pin	

Basic end device objects

Fibre Channel object	Equivalent SAS object	Equivalent SCSI architecture object
Platform	Device	Device
Node	<none>	<none>
Port	Port	Port
<none>	Phy	<none>

Basic fabric objects

Fibre Channel object	Equivalent SAS object	Equivalent SCSI architecture object
Cable or backplane	Cable or backplane	Service delivery subsystem
Port Bypass Circuit	<none>	
Hub	<none>	
Switch	Expander	

Addressing (part 1)



- Each SAS port has a **SAS address**
 - 64-bit NAA=5 format as defined by SPC-3 (and FC-FS)
 - Set by the factory
 - Used for addressing
- Each FC node has a **node name** and each FC port has a **port name**
 - Same NAA format as SAS
 - additional NAA types are allowed (e.g. NAA=3)
 - Set by the factory
 - Exchanged during logins
 - Not used for addressing (FC 24-bit fabric address is assigned)

Addressing (part 2)



- FC uses a port address for frame routing
 - 24-bit **Port address** assigned during initialization
 - 8-bit **Arbitrated Loop Physical Address (AL_PA)** is lower eight bits of a 24-bit address; used for loop arbitration
- FC has many **well-known addresses** for fabric services
 - Broadcast, fabric login, fabric controller, directory server, time server, management server, quality of service facilitator, alias server, security key distribution server, clock synchronization server, multicast server
 - Services can be provided by one or more switches or by management appliances
 - In SAS, expanders are addressable like any other device
 - No special addresses

Topologies (part 1)

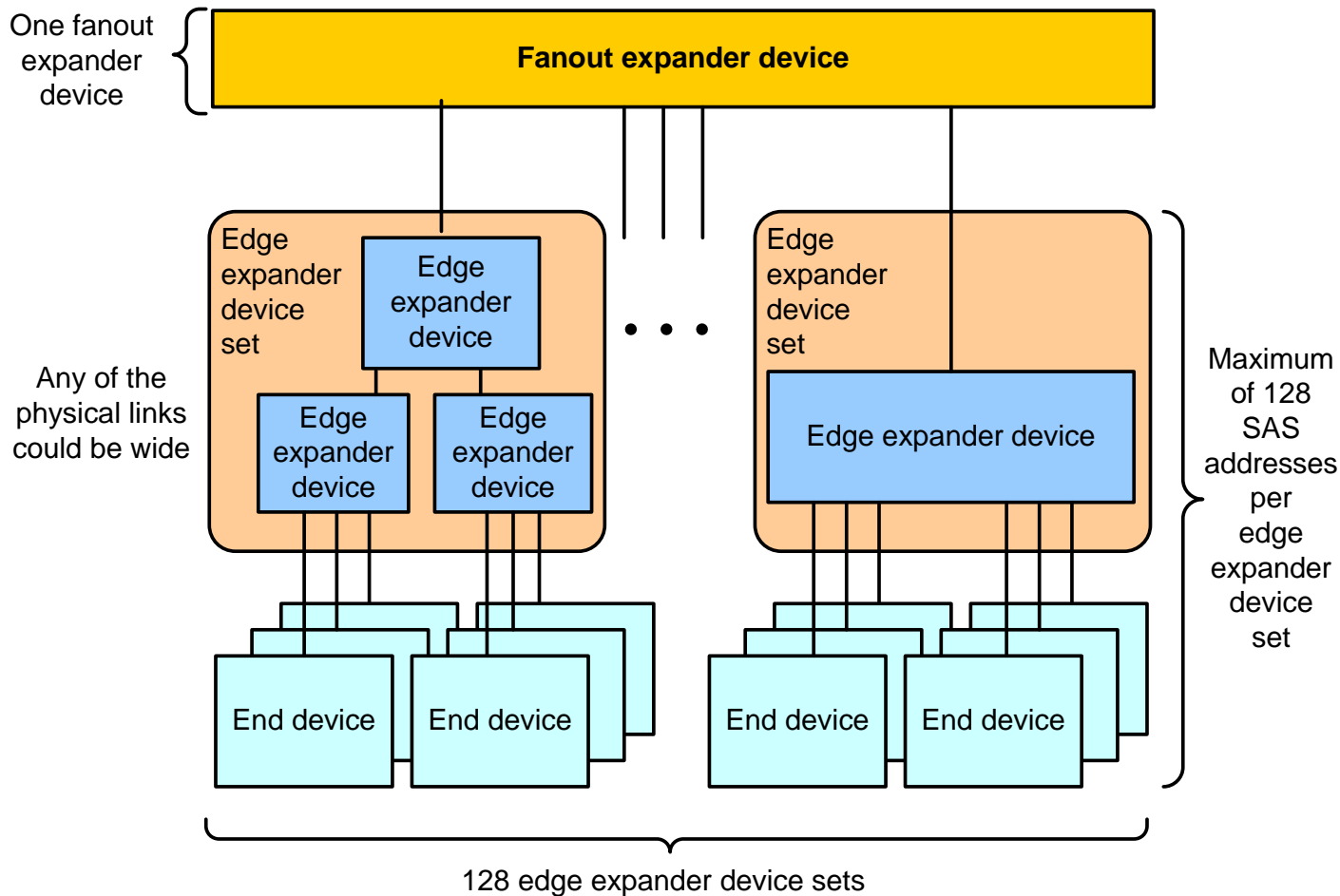


FC	SAS
<p>Simple or complex fabrics:</p> <ul style="list-style-type: none">• Fabric Shortest Path First algorithm• Cross-connected switches• $2^{24} = 16 \text{ M}$ devices in a fabric• Dual fabrics for no-single-points-of-failure	<p>Simple fabrics:</p> <ul style="list-style-type: none">• Tree of expanders• No loops allowed (complicates routing)• Initiators perform discovery and program routing tables• Expander sets have maximum of 128 phys• $128 * 128 = 16 \text{ K}$ maximum addresses in a SAS fabric• Dual fabrics for no-single-points-of-failure
<p>FC supports arbitrated loops</p> <ul style="list-style-type: none">• Often used for disk drive connections• 127 devices in a loop• Dual loops for no-single-points-of-failure	<p>No loops</p>
<p>Both are point-to-point serial interfaces, not multidrop busses</p>	

Topologies (part 2)



- SAS limited topology



- SAS has no logins
- FC has three layers of logins
 1. **Fabric login** (FLOGI): announce presence to the fabric
 2. **Port login/logout** (PLOGI/LOGO): establish a session with another port
 3. **Process login/logout** (PRLI/PRLO): establish a session from an upper level process with another upper level process
 - Third-party process logout (TPRLO) also available

Classes of service



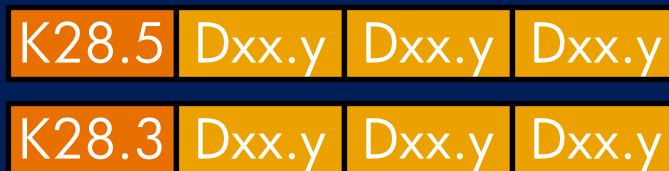
- SAS only has one class of service, similar to Fibre Channel class 1

FC class of service	Description
1	Virtual circuit, full bandwidth (how a telephone network seems to operate)
2	Frame-routed, confirmed delivery (ACKs), out-of-order frames possible
3	Frame-routed, unconfirmed delivery, out-of-order and lost frames possible (though they cause problems in practice; most switches provide in-order delivery to be safe). Almost all FC implementations use class 3 .
4	Virtual circuit, fractional bandwidth
6	Multicast

Primitives/ordered sets

- FC Ordered Set = SAS Primitive

- All FC ordered sets are K28.5 based
- SAS uses K28.5 too; but SATA uses K28.3



- SAS primitives can start with positive or negative disparity
 - FC primitives always start with negative disparity
 - FC primitives never sent inside data frames

Primitives/ordered sets



FC ordered set	Description	SAS equivalent?
ARB(nn)	Arbitration	OPEN
IDLE	Idle	Idle dwords
SOFnn, EOFnn	Start/end of frame	SOF, EOF
SYN	Clock synchronization	<none>
MRK	Mark	<none>
OPN(nn)	Open	OPEN
CLS	Close	CLOSE
DHD	Dynamic Half-Duplex	DONE
R_RDY, VC_RDY	Receiver Ready	R_RDY
LR	Link reset	HARD_RESET primitive, OOB sequences
LIP	Loop initialization	<none>
LPB, LPE	Loop port bypass/enable	<none>
NOS, OLS	Not operational/offline	<none>

Exchanges and sequences



- SAS doesn't have Exchanges or Sequences and their associated IDs
- No Sequence Initiative to transfer after certain frame types
- No low-level exchange or sequence aborts

Frame transmission



- Both use R_RDY primitive to grant credit to transmit a frame
- Frames start with an SOF and end with an EOF
 - FC has 10 SOFs and 6 EOFs
 - SOFs for class 1, 2, 3, and 4
 - EOFs to manage sequences and indicate outbound errors (normal, terminate, abort, normal invalid, etc.)
 - SAS has one of each
- SAS allows primitives inside frames
 - SAS allows ACK, NAK, R_RDY, and deletable clock skew management ALIGN primitives inside frames
 - FC locates deletable IDLE primitives only outside frames
 - 6 IDLES transmitted between frames by the source (some may not arrive at the destination)

Frame acknowledgement

- SAS acknowledges frames with ACK or NAK primitives
 - FC uses ACK, BSY, and RJT frames
 - These frames themselves require R_RDY credits
- SAS requires each frame be ACKed or NAKed
 - Like FC Class 1 or Class 2
 - FC has an ACK_0 and ACK_N frame to aggregate ACKs
 - SAS does not
 - Assumes round trip is short enough the extra latency doesn't matter
 - FC Class 3 doesn't use ACK at all
 - Ship-and-pray (although most switches provide in-order delivery)
 - ULP has to handle out-of-order and lost frames
 - FCP command reference numbers, FCP_CONF

- Both use a 24 byte frame header
- SAS has no optional frame headers
 - FC defines several optional headers
 - expiration/security, network, association, and device
- SAS maximum data size 1 KB
 - FC maximum 2 KB data
- SAS uses primitives for link control (e.g. ACK primitive)
 - FC uses frames for link control (e.g. ACK_0 frame)

Frame header



- SAS drops a few fields (and many F_CTL bits)

Byte	Field(s)		
0	Frame Type (FC: R_CTL)		
1 to 3	Hashed Destination SAS Address (FC: Destination_ID)		
4	Reserved (FC: CS_CTL)		
5 to 7	Hashed Source SAS Address (FC: Source_ID)		
8	Reserved (FC: Type)		
9	Reserved (FC: F_CTL)		
10	Reserved (FC: F_CTL)	Retransmit	Rsvd
11	Reserved (FC: F_CTL)	Number of Fill Bytes	
12 to 15	Reserved (FC: Sequence ID, DF_CTL, Sequence Count)		
16 to 17	Tag (FC: OX_ID)		
18 to 19	Target Port Transfer Tag (FC: RX_ID)		
20 to 23	Data Offset (FC: Parameter)		
24 to m	Information Unit		
m to (n-3)	Fill bytes, if needed		
(n-3) to n	CRC		

SCSI transport layer frames



- Similar
 - COMMAND, XFER_RDY, RESPONSE, DATA
- Different
 - SAS has a TASK frame for task management
 - FCP overloads the COMMAND frame
 - FCP implements ABORT TASK as a lower level "abort sequence" link service rather than a separate COMMAND frame
 - Since SAS ACKs its RESPONSE frames, it does not support a CONF frame

Name	FCP	SAS
COMMAND	✓	✓
TASK		✓
XFER_RDY	✓	✓
DATA	✓	✓
RESPONSE	✓	✓
CONF	✓	

SCSI COMMAND frame



- FC has Command Reference Number to counter out-of-order delivery problems
- FC handles task management functions with this frame
 - SAS uses separate TASK frame
- FC tags commands as reads or writes
- FC includes a Data Length for the command

Byte	Field(s)	
(24 bytes)	Frame header	
0 to 7	Logical Unit Number	
8	SAS: Reserved FC: Command Reference Number	
9	Reserved	Task Attribute
10	SAS: Reserved FC: Task Management Flags	
11	Additional CDB Length	SAS: Reserved
		FC: Rd FC: Wr
12 to 27	CDB	
28 to m	Additional CDB bytes (if needed)	
m + 1 to m + 4	SAS: non-existent FC: Data Length	
(0 bytes)	Fill bytes NOT needed	
(4 bytes)	CRC	

SCSI XFER_RDY frame



- Identical layout
- SAS-1 does not allow offsets that are not continuously increasing during the command
- SAS-1.1 will allow going back to the last point for retries

Byte	Field(s)
(24 bytes)	Frame header
0 to 3	Requested Offset
4 to 7	Write Data Length
8 to 11	Reserved
(0 bytes)	Fill bytes NOT needed
(4 bytes)	CRC

SCSI DATA frame



- Data frames are the same
- FCP maximum of 2048 data bytes
- SAS maximum of 1024 data bytes

Byte	Field(s)
(24 bytes)	Frame header
0 to n	Data
(0 to 2 bytes)	Fill bytes, if needed
(4 bytes)	CRC

SCSI RESPONSE frame



- Since the FCP COMMAND frame includes an expected data length, overflow or underflow is reported here (along with the residual amount)
- SAS ACKs can be used to keep track of overflow and underflow

Byte	Field(s)			
(24 bytes)	Frame header			
0 to 9	Reserved			
10	Reserved	SAS: Reserved	FC: Residual Overflow Underflow	DataPres
11	Status			
12 to 15	SAS: Reserved FC: Residual			
16 to 19	Sense Data Length			
20 to 23	Response Data Length			
24 to (23+m)	Response Data			
(24+m) to (23+m+n)	Sense Data			
(0 bytes)	Fill bytes NOT needed			
(4 bytes)	CRC			

SCSI frame interlocking



- SAS interlocks all but DATA frames
 - Must wait for ACK or NAK before sending another frame
 - Only DATA frames with same tag can be pipelined together
 - Only allows one COMMAND or TASK frame at a time from an initiator phy
 - Only allows one XFER_RDY or RESPONSE frame at a time from a target phy
- Avoids complications due to frames in flight that complicate Fibre Channel and iSCSI
 - Low latency environment is not as concerned with this issue

Mode pages/log pages



- Mode pages are a dangerous way to control protocol features
 - Must be used carefully in multi-initiator environments
- Both support the **Disconnect-Reconnect mode page**
 - Both support the Bus Inactivity Limit, Connect Time Limit, and Maximum Burst Size fields
- Both define the **Protocol-Specific Port mode page**
 - FC Resource Recovery Timeout (RR_TOV) similar to SAS I_T Nexus Loss Time
- FCP defines a **Protocol-Specific Logical Unit Control mode page** to turn on command reference numbers (enable precise delivery ordering)
 - SAS-1.1 may have one to enable transport layer retries
- SAS defines a **Protocol-Specific log page** for error counters

- FC defines State Change Notification and Registered State Change Notification
 - SAS just has a BROADCAST (CHANGE) primitive
 - SAS initiator must search through entire fabric to figure out what changed



Wrap up

Serial Attached SCSI tutorials



- General overview (~2 hours)
- Detailed multi-part tutorial (~3 days to present):
 - Architecture
 - Physical layer
 - Phy layer
 - Link layer
 - Part 1) Primitives, address frames, connections
 - Part 2) Arbitration fairness, deadlocks and livelocks, rate matching, SSP, STP, and SMP frame transmission
 - Upper layers
 - Part 1) SCSI application and SSP transport layers
 - Part 2) ATA application and STP/SATA transport layers
 - Part 3) Management application and SMP transport layers, plus port layer
 - SAS SSP comparison with Fibre Channel FCP

Key SCSI standards

- Working drafts of **SCSI** standards are available on <http://www.t10.org>
- Published through <http://www.incits.org>
 - Serial Attached SCSI
 - SCSI Architecture Model – 3 (SAM-3)
 - SCSI Primary Commands – 3 (SPC-3)
 - SCSI Block Commands – 2 (SBC-2)
 - SCSI Stream Commands – 2 (SSC-2)
 - SCSI Enclosure Services – 2 (SES-2)
- **SAS connector** specifications are available on <http://www.sffcommittee.org>
 - SFF 8482 (internal backplane/drive)
 - SFF 8470 (external 4-wide)
 - SFF 8223, 8224, 8225 (2.5", 3.5", 5.25" form factors)
 - SFF 8484 (internal 4-wide)

Key ATA standards



- Working drafts of **ATA** standards are available on <http://www.t13.org>
 - Serial ATA 1.0a (output of private WG)
 - ATA/ATAPI-7 Volume 1 (architecture and commands)
 - ATA/ATAPI-7 Volume 3 (Serial ATA standard)
- **Serial ATA II** specifications are available on <http://www.t10.org> and <http://www.serialata.org>
 - Serial ATA II: Extensions to Serial ATA 1.0
 - Serial ATA II: Port Multiplier
 - Serial ATA II: Port Selector
 - Serial ATA II: Cables and Connectors Volume 1

For more information



- International Committee for Information Technology Standards
 - <http://www.incits.org>
- T10 (SCSI standards)
 - <http://www.t10.org>
 - Latest SAS working draft
 - T10 reflector for developers
- T13 (ATA standards)
 - <http://www.t13.org>
 - T13 reflector for developers
- T11 (Fibre Channel standards)
 - <http://www.t11.org>
- SFF (connectors)
 - <http://www.sffcommittee.org>
- SCSI Trade Association
 - <http://www.scsita.org>
- Serial ATA Working Group
 - <http://www.serialata.org>
- SNIA (Storage Networking Industry Association)
 - <http://www.snia.org>
- Industry news
 - <http://www.infostor.com>
 - <http://www.byteandswitch.com>
 - <http://www.wwpi.com>
 - <http://searchstorage.com>
- Training
 - <http://www.knowledgetek.com>



i n v e n t