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
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

*Slide 10*
General - SCSI standards

- Device-type specific command sets (e.g., SBC-2, SSC-2, MMC-3)
- Primary command set (shared for all device types) (SPC-3)
- Protocols (e.g., SPI-4, FCP-2, SSP in this standard)
- Interconnects (e.g., SPI-4, Fibre Channel, this standard)
General - ATA standards

ATA register-delivered command set (ATA/ATAPI-6)

Device-type specific command sets (e.g., MMC-3)

Primary command set (shared for all device types) (SPC-3)

ATA (Register-delivery for ATA commands)

ATAPI (Packet-delivery for SCSI commands)

ATA/ATAPI Register set (ATA/ATAPI-6)

Protocols (STP in this standard, SATA)

Interconnects (this standard, SATA)
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 - 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)
  - **Target side**
    - SSP (SCSI)
    - SATA (ATA)
    - SMP

*Slide 15*
SAS initiator device is a SCSI initiator device and/or ATA initiator device.

Service delivery subsystem
General - Target device

SAS target device is a SCSI target device and/or ATA target device.

- SAS target port
  - SAS target device
    - SCSI target port
    - ATA target port

- SAS target port
  - SCSI target port
  - ATA target port

- SAS target port
  - ATA target port

Service delivery subsystem
General - Expander device

Expander device

expander port

SAS target port

expander port

SSP and STP, or SATA

Initiator ports, target ports, or expander ports
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
Signal pair
- Signal
+ Signal
Signal pair
General - Ports and wide links

- Physical link (1-wide physical link)
- 2-wide physical link
- n-wide physical link
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

- Initiator port (SSP)
- Initiator port (STP)
- Target port (SSP)
- Target port (SATA)
- Expander port
- SAS domain
- SATA domain

STP

SATA

02-158r0 SAS
Technical Overview
General - Initiator device in 2 domains

- Initiator device
  - Initiator port
- Expander devices with expander ports
- Target port
- Target device
- Target port
- Target device
- SAS domain
- SAS domain
General - Target device in 2 domains

Slide 25
General - Pathways

- Single physical link
- Pathways

Initiator Port -> phy -> Target Port

Expander device

Slide 27
General - Possible configurations 1

Internal drives

Server
SAS RAID Controller

Internal drives and External JBOD

Server
SAS RAID Controller

Storage
Expander
Expander

Internal drives and External JBOD

Server
SAS RAID Controller

Expander
Expander
Expander

SAS or SATA drives
General - Possible configurations 2

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

Internal drives and external RAID as a 4 Node Cluster
Each device has a 64-bit FC style Worldwide Name (WWN), used for addressing.

<table>
<thead>
<tr>
<th></th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NAA (5h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IEEE Company ID (24 bits)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Vendor-Specific Identifier (40 bits)
General - Transmit data path SAS portion

Phy layer state machine

OOB signals, D10.2, ALIGN

Phy layer
Insert ALIGNs

SAS Link layer
Connection management

Address frames, OPEN_ACCEPT, CLS, OPEN_REJECT, BREAK

SSP Link layer

SMP Link layer

STP Link layer

OOB

ALIGN

Pause

Ready

other

other

Ready

Connected

SSP dword

SMP dword

STP dword
General - Transmit data path - SSP portion

- **SCSI Application layer** runs SCSI commands
- **SSP Transport layer** builds and parses SSP frames
- **Port layer** Transport layer to multiple link layers arbiter
- **SSP Link layer**
  - SSP Transmit frame state machine
  - SSP Done state machine
  - SSP idle dword generator
  - SSP Link layer
    - SSP Receive frame state machine
    - SSP Receive credit counter state machine

Vendor-specific to send ACK/NAK/RRDY inside frames

- **SSP Link layer** SSP state machine
- **DONE**
- **idle dword**
- **ACK/NAK**
- **RRDY**

**SOF / dwords / EOF**

**SSP dword**
General - Transmit data path - STP portion

**SATA Application layer state machines**
- **Device**: Device
- **Command layer (D)**
- **Host**: Vendor-specific

**SATA Transport layer state machines**
- **Host transport**: HT (transmit FIS, decompose FIS)
- **Device transport**: DT (transmit FIS, decompose FIS)

**STP Link layer machine** based on SATA Link state machine
- **L** = Link idle
- **LT** = Link transmit
- **LR** = Link receive
- **LPM** = Link power mode

SATA state machines are defined by SATA and are shown here only for reference.
General - Reset terminology

SATA

- Time
  - Link reset sequence
  - Phy reset sequence
  - SATA power-on sequence
    - SATA OOB sequence
    - SATA speed negotiation sequence

SAS

- Time
  - Link reset sequence
  - Phy reset sequence
    - SAS OOB sequence
    - SAS speed negotiation sequence
  - Identification sequence
    - Phy layer state machine enters Ready
    - Link layer state machine starts
    - Link layer state machine enters Idle

Slide 34
General - Expander model

Expander Port

SAS Target Port

SMP Target Port

SSP Target Port (optional)

SSP Link

internal expander port contains SAS Target Port

Expander Port

SAS Phy

SSP Link

Request

Confirm

Response

Indicate

Expander Port

SAS Phy

STP Link

Request

Confirm

Response

Indicate

Expander Port

SAS Phy

SSP Link

Request

Confirm

Response

Indicate

Expander Port

SAS Phy

SSP Link

Request

Confirm

Response

Indicate

Slide 35

02-158r0 SAS Technical Overview
...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

- Internal cabled environment
  - Host
  - Power supply
  - SATA cable
  - Device plug connector

- Power cable receptacle connector

- Internal backplane environment
  - Host
  - SATA device
  - Host receptacle connector
  - Device plug connector

Slide 38
Physical - SAS external environment

- External cable environment
  - Initiator device
    - SAS external plug connector (4 physical links)
  - SAS external cable (1-4 physical links)
  - Expander device
    - SAS external plug connector (4 physical links)
    - SAS external cable receptacle connectors (4 physical links)
Physical - SAS internal environments

Internal cabled environment

SAS plug connector (dual-port)

SAS cable (1 physical link plus power)

SAS target device

SAS plug connector (dual-port)

SAS internal cable receptacle connectors (single-port)

Internal backplane environment

SAS initiator device or expander device

SAS plug connector (dual-port)

SAS target device

SAS backplane receptacle connector (dual port)
## Physical - Connectors

<table>
<thead>
<tr>
<th>Connector</th>
<th>Attaches to</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS plug</td>
<td>SAS internal cable receptacle</td>
</tr>
<tr>
<td></td>
<td>SAS backplane receptacle</td>
</tr>
<tr>
<td>SAS internal cable receptacle</td>
<td>SAS plug</td>
</tr>
<tr>
<td></td>
<td>SATA device plug (single-port)</td>
</tr>
<tr>
<td>SAS backplane receptacle</td>
<td>SAS plug</td>
</tr>
<tr>
<td></td>
<td>SATA device plug (single-port)</td>
</tr>
<tr>
<td>SAS external cable receptacle</td>
<td>SAS external plug</td>
</tr>
<tr>
<td>SAS external plug</td>
<td>SAS external cable receptacle</td>
</tr>
</tbody>
</table>
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

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

- Amplitude and time based on eye diagrams

Slide 45
## Physical - Transmit signal characteristics

<table>
<thead>
<tr>
<th>Compliance point</th>
<th>Characteristic</th>
<th>SATA</th>
<th>1.5 Gbps</th>
<th>3.0 Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drive connector</td>
<td>Maximum drive strength (B1-B2)</td>
<td>N/A</td>
<td>1200 mV</td>
<td>1600 mV</td>
</tr>
<tr>
<td></td>
<td>Minimum drive strength (A1-A2)</td>
<td>N/A</td>
<td>600 mV</td>
<td>800 mV</td>
</tr>
<tr>
<td>External connector</td>
<td>Maximum drive strength (B1-B2)</td>
<td>N/A</td>
<td>1600 mV</td>
<td>1600 mV</td>
</tr>
<tr>
<td></td>
<td>Minimum drive strength (A1-A2)</td>
<td>N/A</td>
<td>800 mV</td>
<td>800 mV</td>
</tr>
<tr>
<td>Initiator or expander attached to SATA</td>
<td>Maximum drive strength (B1-B2)</td>
<td>900 mV</td>
<td>1200 mV</td>
<td>1600 mV</td>
</tr>
<tr>
<td></td>
<td>Minimum drive strength (A1-A2)</td>
<td>600 mV</td>
<td>600 mV</td>
<td>800 mV</td>
</tr>
</tbody>
</table>
## Physical - Receive signal characteristics

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

---

*Slide 47*
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
- 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
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

<table>
<thead>
<tr>
<th>Signal</th>
<th>Nominal burst time</th>
<th>Nominal idle time</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMINIT/COMRESET</td>
<td>107 ns</td>
<td>320 ns</td>
</tr>
<tr>
<td>COMWAKE</td>
<td>107 ns</td>
<td>106.7 ns</td>
</tr>
<tr>
<td><strong>COMSAS</strong></td>
<td><strong>214 ns</strong></td>
<td><strong>320 ns</strong></td>
</tr>
</tbody>
</table>
Phy - SATA reset sequence

SATA host (initiator) TX
SATA device (target) TX
Power-On

COMRESET
Calibrate
COMWAKE

aligns
533ns max
2048
aligns
@ Fastest Rate
Non-aligns @ Negotiated Rate

Post-COMRESET/COMINIT
SATA reset sequence

aligns
175ns min
228.3 max
2048
aligns
@ Negotiated Rate
Non-aligns @ Negotiated Rate
Phy - SAS reset sequence

Start Speed Negotiation

- SAS Speed Negotiation Window G1 rate
- SAS Speed Negotiation Window G2 rate
- SAS Speed Negotiation Window G3 rate
- SAS Speed Negotiation Window Negotiated rate (G2)
Phy - SAS reset sequence 2

- Wait COMSAS Reply
- Await COMSAS Negate
- Rate Change Delay (RDC)
- RCD
- Fall back to G2
- Pattern sync
- Not supported
- Don't transmit anything
- 320ns
- 109.2 usec
- SAS Speed Negotiation Window G1 Rate
- SAS Speed Negotiation Window G2 Rate
- SAS Speed Negotiation Window G3 Rate
- SAS Negotiation Window Negotiated Rate G2

Slide 56
Phy - Hot plug and the reset sequence

Slide 57
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
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
SAS primitives
- AIP
- ALIGN()
- BREAK
- CHANGE
- CLOSE
- EOAF
- HARD_RESET
- OPEN_ACCEPT
- OPEN_REJECT()
- SOAF

SSP/SMP primitives
- ACK
- DONE()
- EOF
- NAK()
- RRDY
- SOF
## 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
Some primitives are sent more than one for more reliable delivery - tolerate single bit errors

“Repeated” is for SATA primitives

<table>
<thead>
<tr>
<th>Type</th>
<th>Send</th>
<th>Detect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Repeated</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Triple</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Redundant</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>
- 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

<table>
<thead>
<tr>
<th>Bit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-0</td>
<td>Reserved</td>
</tr>
<tr>
<td>1</td>
<td>PHY IDENTIFIER</td>
</tr>
<tr>
<td>2</td>
<td>Reserved</td>
</tr>
<tr>
<td>3</td>
<td>MAX LINK RATE</td>
</tr>
<tr>
<td>3-0</td>
<td>DEV TYPE, SSP_I, STP_T, SSP_I, SSP_T, SMP_I, SMP_T</td>
</tr>
<tr>
<td>4</td>
<td>Reserved</td>
</tr>
<tr>
<td>19</td>
<td>Reserved</td>
</tr>
<tr>
<td>20</td>
<td>DEVICE NAME</td>
</tr>
<tr>
<td>27</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>CRC</td>
</tr>
</tbody>
</table>
## Link - OPEN address frame

<table>
<thead>
<tr>
<th>Bit</th>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>INIT</td>
<td>Frame Initiator</td>
</tr>
<tr>
<td>6</td>
<td>PROTOCOL</td>
<td>Protocol Identifier</td>
</tr>
<tr>
<td>5</td>
<td>FRAME TYPE</td>
<td>Frame Type (1h)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Reserved</td>
<td>LINK RATE</td>
</tr>
<tr>
<td>1</td>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>INITIATOR CONNECTION TAG</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ARBITRATION WAIT TIME</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Destination Device Name</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Source Device Name</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CRC</td>
<td></td>
</tr>
</tbody>
</table>

**Slide 69**
### SAS support for power management

<table>
<thead>
<tr>
<th>ATA device power management</th>
<th>SATA interface power management</th>
<th>SCSI power conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle (yes)</td>
<td>Partial (no)</td>
<td>Idle (yes)</td>
</tr>
<tr>
<td>Standby (yes)</td>
<td>Slumber (no)</td>
<td>Standby (yes)</td>
</tr>
<tr>
<td>Sleep (no)</td>
<td></td>
<td>Sleep (no)</td>
</tr>
</tbody>
</table>
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

- Signal

SAS loopback test modes

![Diagram of SAS loopback test modes](image)

- Signal pair

- Signal

Near-end analog loopback

Synchronizing register

Far-end retimed loopback

Output register

Phy

Phy
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
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
Break - 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

- Breaking a connection
- Expander device
- Originating device
- Destination device

(last) BREAK OPN (first)
(first) AIP BREAK (last)

idle

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

AIP BREAK

(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
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
- 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

Slide 84
- Full duplex
- SOF, frame dwords, CRC, EOF
- Each frame acknowledged with ACK, NAK
- Credit with RRDY
- SSP link layer state machine
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

Non-interlocked frame tag A
Non-interlocked frame tag A
Non-interlocked frame tag A
Non-interlocked frame tag A
Non-interlocked frame tag A

ACK or NAK
ACK or NAK

Non-interlocked frame tag A
Non-interlocked frame tag A

ACK or NAK
ACK or NAK
ACK or NAK
ACK or NAK

Non-interlocked frame tag B

ACK or NAK

Non-interlocked frame tag C
Non-interlocked frame tag C

. . .
- 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
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

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION UNIT TYPE</td>
<td></td>
</tr>
<tr>
<td>HASHED DESTINATION DEVICE NAME</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>HASHED SOURCE DEVICE NAME</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>COMMAND ID</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>TAG</td>
<td></td>
</tr>
<tr>
<td>TARGET PORT TRANSFER TAG</td>
<td></td>
</tr>
<tr>
<td>Reserved</td>
<td></td>
</tr>
<tr>
<td>INFORMATION UNIT</td>
<td></td>
</tr>
<tr>
<td>Fill bytes</td>
<td></td>
</tr>
<tr>
<td>CRC</td>
<td></td>
</tr>
</tbody>
</table>
## Transport - SSP information units

<table>
<thead>
<tr>
<th>IU</th>
<th>Originator</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMAND</td>
<td>Initiator port</td>
</tr>
<tr>
<td>TASK</td>
<td>Initiator port</td>
</tr>
<tr>
<td>XFER_RDY</td>
<td>Target port</td>
</tr>
<tr>
<td>DATA</td>
<td>Initiator port (writes)</td>
</tr>
<tr>
<td></td>
<td>or target port (reads)</td>
</tr>
<tr>
<td>RESPONSE</td>
<td>Target port</td>
</tr>
<tr>
<td>AEN</td>
<td>Target port</td>
</tr>
<tr>
<td>AEN_RESPONSE</td>
<td>Initiator port</td>
</tr>
</tbody>
</table>
Transport - SSP Task Management sequence

Initiator port

TASK

RESPONSE

Target port
Transport - SSP Write sequence

- **COMMAND**
- **XFER_RDY**
- **DATA**
- **RESPONSE**

Reply to XFER_RDY with one or more DATA IUs

Repeat XFER_RDY until all write data is transferred
Transport - SSP Read sequence

Repeat until all read data is transferred
Transport - SSP Bidirectional sequence

**Initiator port**
- COMMAND
- XFER_RDY
- DATA
- RESPONSE

**Target port**
- Repeat XFER_RDY until all write data is transferred
- Repeat DATA until all read data is transferred
- Reply to XFER_RDY with one or more DATA IUs
- Read (DATA) and write (XFER_RDY and DATA) IUs may be interleaved in any order

**Slide 99**
Transport - SSP AEN sequence

Initiator port

AEN

AEN_RESPONSE

Target port
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

Slide 102
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
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

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