To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 26 April 2006
Subject: 06-188r0 SAS-2 Support multiple STP affiliations

Revision history
Revision 0 (26 April 2006) First revision

Related documents
sas2r03 - Serial Attached SCSI - 2 (SAS-2) revision 3

Overview
STP target ports (particularly STP/SATA bridges) can currently report support for two different modes in the SMP REPORT PHY SATA response:

a) Affiliations supported. The STP target port allows only one STP initiator port to access it at a time; connections from others are rejected with OPEN_REJECT (STP RESOURCES BUSY).
b) Affiliations not supported. The STP target port must have the ability to track all SATA commands by the STP initiator ports' SAS addresses, and allows simultaneous access to the task file registers. This model assumes the STP/SATA bridge supports essentially infinite STP initiator ports (SATA NCQ supports only 32 commands at a time which provides some relief).

It is possible to create a hybrid STP/SATA bridge that supports a limited number of STP initiator ports, but not an infinite number. This can be viewed as supporting multiple simultaneous affiliations.

Changes are proposed to expand the REPORT PHY SATA function to report more than one active affiliation and STP I_T nexus loss, and to modify the rest of the text in the standard to support the concept of multiple affiliations.

Editor's Note 1: A way to clear an affiliation for another I_T nexus may also be needed. If initiator A and B have affiliations, initiator B breaks and initiator C wants to take over, it shouldn't be required to send HARD RESET and abort all of initiator A's commands to do so. This will be handled by a separate proposal if it is wanted.

Suggested changes

7.2.5.11 OPEN_REJECT

All of the OPEN_REJECT versions defined in table 88 shall result in the originating port abandoning the connection request.

Table 88 — OPEN_REJECT abandon primitives

Primitive	Originator	Description
...	...	...
OPEN_REJECT (STP RESOURCES BUSY)	Destination phy	STP target port with destination SAS address exists but the STP target port supports affiliations and is not able to establish has an affiliation with anotherthis STP initiator port (e.g., because it has reached its maximum number of simultaneous affiliations) or the STP target port does not support affiliations and all of the available task file registers have been allocated to other STP initiator ports (see 7.17.5). Process the same as OPEN_REJECT (WRONG DESTINATION) for non-STP connection requests.

...

7.14.4.4 SL_CC2:Selected state

7.14.4.4.1 State overview

...

3) If the OPEN address frame PROTOCOL field is set to STP, the STP target port supports affiliations and the source SAS address is not that of thean STP initiator port with an affiliation established or the STP target port does not support affiliations and the source SAS address is not that of an STP initiator port with task file register set resources (see 7.17.5), this state shall send a Transmit OPEN_REJECT (STP Resources Busy) message to the SL transmitter (see 7.14.4.4.2);

...

7.17.5 Affiliations

Coherent access to the SATA task file registers shall be provided for each STP initiator port. STP target ports that do not track all commands by the STP initiator ports' SAS addresses shall implement affiliations to provide coherency. STP target ports that track commands for a limited number of STP initiator ports shall implement one affiliation context per supported STP initiator port to provide coherency. STP target ports that track all commands by the STP initiator ports' SAS addresses shall not implement affiliations.

An affiliation is a state entered by an STP target port where it refuses to accept connection requests from STP initiator ports other than the one that has established an affiliation.

An STP target port that supports affiliations shall establish an affiliation whenever it accepts a connection request from an STP initiator port that does not already have an affiliation. When anthe maximum number of simultaneous affiliations have been is established (i.e., all affiliation contexts are in use), the STP target port shall reject all subsequent connection requests from other STP initiator ports with OPEN_REJECT (STP RESOURCES BUSY).

An STP target port shall maintain an affiliation until any of the following occurs:

a) power on;
b) the SAS target device receives an SMP PHY CONTROL request specifying the phy with the affiliation and specifying a phy operation of HARD RESET (see 10.4.3.13) from any SMP initiator port;
c) the SAS target device receives an SMP PHY CONTROL request specifying the phy with the affiliation and specifying a phy operation of TRANSMIT SATA PORT SELECTION SIGNAL (see 10.4.3.13) from any SMP initiator port;
d) the SAS target device receives an SMP PHY CONTROL request specifying the phy with the affiliation and specifying a phy operation of CLEAR AFFILIATION (see 10.4.3.13) from the same SAS initiator port that has the affiliation.

   If a connection is already established to the STP target port on one phy while an SMP PHY CONTROL request specifying a phy operation of CLEAR AFFILIATION is processed by an SMP target port on another phy, the affiliation shall be cleared and the STP target port shall respond to new connection attempts with:

   A) AIP (WAITING ON CONNECTION) and/or OPEN_REJECT (RETRY), if the STP target port is in an expander device; or
   B) OPEN_REJECT (RETRY), if the STP target port is in a SAS device;

   rather than OPEN_REJECT (STP RESOURCES BUSY);

e) a connection to the phy with the affiliation is closed with CLOSE (CLEAR AFFILIATION); or
f) the STP target port is part of a STP/SATA bridge and a link reset sequence is begun on the SATA physical link that was not requested by an SMP PHY CONTROL request specifying the phy and specifying a phy operation of LINK RESET (see 10.4.3.13).

An affiliation established when a command is transmitted shall be maintained until all frames for the command have been delivered. An STP initiator port implementing command queuing (see SATAII-EXT) shall maintain an affiliation while any commands are outstanding to avoid confusing SATA devices, which only understand one SATA host. STP initiator ports may keep affiliations for longer tenures, but this is discouraged.

2

An STP target port that implements affiliations shall implement at least one affiliation context per STP target port. Multiple phys on the same STP target port shall use the same set of affiliation contexts. Support for affiliations is indicated in the SMP REPORT PHY SATA function response (see 10.4.3.7).

7.17.6 Opening an STP connection

If no STP connection exists when the SATA host port in an STP/SATA bridge receives a SATA_X_RDY from the attached SATA device, the STP target port in the STP/SATA bridge shall establish an STP connection to the appropriate STP initiator port before it transmits a SATA_R_RDY to the SATA device.

Wide STP initiator ports shall not request more than one connection at a time to an STP target port. Wide STP target ports shall not request more than one connection at a time to an STP initiator port.

While a wide STP target port is waiting for a response to a connection request or has established a connection to an STP initiator port, it shall:

    a)   reject incoming connection requests from that STP initiator port with OPEN_REJECT (RETRY); and
    b)   if affiliations are supported and the maximum number of simultaneous affiliations have been established (i.e., all affiliation contexts are in use), reject incoming connection requests from other STP initiator ports that do not have affiliations with OPEN_REJECT (STP RESOURCES BUSY).

> Editor's Note 2: no need to list all the "accept" cases - those reject cases are complete

---

While a wide STP initiator port is waiting for a response to a connection request to an STP target port, it shall not reject an incoming connection request from that STP target port because of its outgoing connection request. It may reject incoming connection requests for other reasons (see 7.2.5.11).

If a wide STP initiator port receives an incoming connection request from an STP target port while it has a connection established with that STP target port, it shall reject the request with OPEN_REJECT (RETRY).

The first dword that an STP phy sends inside an STP connection after OPEN_ACCEPT that is not an ALIGN or NOTIFY shall be an STP primitive (e.g., SATA_SYNC).

7.17.9 STP (link layer for STP phys) state machines

The STP link layer uses the SATA link layer state machines (see ATA/ATAPI-7 V3), modified to:

    a)   communicate with the port layer rather than directly with the transport layer;
    b)   interface with the SL state machines for connection management (e.g., to select when to open and close STP connections, and to tolerate idle dwords between an OPEN address frame and the first SATA primitive); and
    c)   implement affiliations support a single affiliation, support limited task file register contexts and support multiple affiliations, or support separate task file register contexts for every STP initiator port and not support affiliations (see 7.17.5).

These modifications are not described in this standard.

9.3 STP transport layer

9.3.1 Initial FIS

...

After the Register - Device to Host FIS is accepted, if affiliations are supported and if the SATA device sends a SATA_X_RDY before an affiliation is established, the expander device shall not send SATA_R_RDY.

...

10.4.3.7 REPORT PHY SATA function

The REPORT PHY SATA function returns information about the SATA state for a specified phy. This SMP function shall be implemented by SMP target ports that share SAS addresses with STP target ports and by SMP target ports in expander devices with STP/SATA bridges. This SMP function shall not be implemented by any other type of SMP target port.

Table 89 defines the request format.

Table 89 — REPORT PHY SATA request

Byte\Bit	7	6	5	4	3	2	1	0
0	SMP FRAME TYPE (40h)							
1	FUNCTION (12h)							
2	Reserved							
3	REQUEST LENGTH (02h)							
4	Reserved							
8								
9	PHY IDENTIFIER							
10	~~Reserved~~AFFILIATION CONTEXT NUMBER							
11	Reserved							
12	(MSB)							
15	CRC							(LSB)

The SMP FRAME TYPE field shall be set to 40h.

The FUNCTION field shall be set to 12h.

The REQUEST LENGTH field shall be set to 02h. For compatibility with previous versions of this standard, a REQUEST LENGTH field set to 00h specifies that there are 2 dwords before the CRC field.

The PHY IDENTIFIER field specifies the phy (see 4.2.7) for which information shall be reported.

The AFFILIATION CONTEXT NUMBER field specifies the affiliation context for which information shall be reported.

The CRC field is defined in 10.4.3.1.

Table 90 defines the response format.

Table 90 — REPORT PHY SATA response (part 1 of 2)

Byte\Bit	7	6	5	4	3	2	1	0
0	SMP FRAME TYPE (41h)							
1	FUNCTION (12h)							
2	FUNCTION RESULT							
3	RESPONSE LENGTH (0Fh)							
4	Reserved							
8								
9	PHY IDENTIFIER							
10	~~Reserved~~AFFILIATION CONTEXT NUMBER							

4

Table 90 — REPORT PHY SATA response (part 2 of 2)

Byte\Bit	7	6	5	4	3	2	1	0
11	Reserved					STP I_T NEXUS LOSS OCCURRED	AFFILIATIONS SUPPORTED	AFFILIATION VALID
12	~~Reserved~~MAXIMUM AFFILIATION CONTEXT NUMBER							
13 ... 15	Reserved							
16 ... 23	STP SAS ADDRESS							
24 ... 43	REGISTER DEVICE TO HOST FIS							
44 ... 47	Reserved							
48 ... 55	AFFILIATED STP INITIATOR SAS ADDRESS							
56 ... 63	STP I_T NEXUS LOSS SAS ADDRESS							
n - 3	(MSB)				CRC			
n								(LSB)

The SMP FRAME TYPE field shall be set to 41h.

The FUNCTION field shall be set to 12h.

The FUNCTION RESULT field is defined in 10.4.3.2.

The RESPONSE LENGTH field shall be set to 0Fh. For compatibility with previous versions of this standard, a RESPONSE LENGTH field set to 00h specifies that there are 13 dwords before the CRC field.

The PHY IDENTIFIER field indicates the phy (see 4.2.7) for which information is being reported and is the same as the PHY IDENTIFIER field in the request frame.

The AFFILIATION CONTEXT NUMBER field indicates the affiliation context for which affiliation-related information (i.e., the AFFILIATION VALID bit, the AFFILIATED STP INITIATOR SAS ADDRESS field, and the STP I_T NEXUS LOSS SAS ADDRESS field) is being reported and is the same as the AFFILIATION CONTEXT NUMBER field in the request frame.

An STP I_T NEXUS LOSS OCCURRED bit set to one indicates that the STP target port encountered an I_T nexus loss for the STP initiator port whose SAS address is indicated in the STP I_T NEXUS LOSS SAS ADDRESS field. An STP I_T NEXUS LOSS OCCURRED bit set to zero indicates that:

    a) an I_T nexus loss has not occurred;

    b) an I_T nexus loss has occurred and been cleared by the SMP PHY CONTROL function CLEAR STP I_T NEXUS LOSS phy operation (see table 4 in 10.4.3.13); or

    c)   the STP target port has successfully established a connection with the indicated STP initiator port.

An AFFILIATIONS SUPPORTED bit set to one indicates that ~~affiliations are~~ the specified affiliation context is supported by the STP target port containing the specified phy. An AFFILIATIONS SUPPORTED bit set to zero indicates that ~~affiliations are~~ the specified affiliation context is not supported by the STP target port containing the specified phy.

An AFFILIATION VALID bit ~~be~~ set to one indicates that the STP target port is currently maintaining an affiliation in the specified affiliation context and the AFFILIATED STP INITIATOR SAS ADDRESS field is valid ~~and the STP target port containing the specified phy is maintaining an affiliation (see 7.17.5)~~. An AFFILIATION VALID bit set to zero indicates that the STP target port is not currently maintaining an affiliation in the specified affiliation context and the AFFILIATED STP INITIATOR SAS ADDRESS field is not valid~~that no affiliation is being maintained~~.

The MAXIMUM AFFILIATION CONTEXT NUMBER field indicates the maximum affiliation context number of the affiliation contexts supported by the STP target port (e.g., a value of 00h indicates support for one affiliation, and a value of 01h indicates support for two simultaneous affiliations). Affiliation contexts are numbered sequentially starting with 00h. The selection of how affiliation contexts are assigned to new affiliations is vendor-specific.

The STP SAS ADDRESS field contains the SAS address (see 4.2.2) of the STP target port that contains the specified phy.

The REGISTER DEVICE TO HOST FIS field contains the contents of the initial Register - Device to Host FIS. For an STP/SATA bridge, this is delivered by the attached SATA device after a link reset sequence (see ATA/ATAPI-7 V3 and SATAII-EXT). For a native STP target port in an end device, this is directly provided.

The FIS contents shall be stored with little-endian byte ordering (i.e., the first byte, byte 24, contains the FIS Type).

For an STP/SATA bridge, the first byte of the field (i.e., the FIS Type) shall be initialized to zero on power on and whenever the phy has restarted the link reset sequence after losing dword synchronization (see 6.9)(i.e., the SP state machine transitioned from SP22:SATA_PHY_Ready to SP0:OOB_COMINIT (see 6.8)) to indicate the field is invalid and the attached SATA device has not delivered a Register – Device to Host FIS. The first byte of the field shall be set to 34h when the attached SATA device has delivered the initial Register – Device to Host FIS. The remaining contents of the REGISTER DEVICE TO HOST FIS field shall remain constant until a link reset sequence causes the attached SATA device to deliver another initial Register – Device to Host FIS.

An STP/SATA bridge that receives a connection request for a SATA device that has not successfully delivered the initial Register – Device to Host FIS shall return an OPEN_REJECT (NO DESTINATION).

    NOTE 1 - If there is a problem receiving the expected initial Register - Device to Host FIS, the STP/SATA bridge should use SATA_R_ERR to retry until it succeeds. In the DISCOVER response, the ATTACHED SATA DEVICE bit is set to one and the ATTACHED SAS ADDRESS field is valid, but the ATTACHED DEVICE TYPE field is set to 000b (i.e., no device attached) during this time.

If the AFFILIATION VALID bit is set to one, t~~T~~he AFFILIATED STP INITIATOR SAS ADDRESS field contains the SAS address (see 4.2.2) of ~~the last~~ the STP initiator port that ~~had~~has an affiliation in the specified affiliation context with the STP target port that contains the specified phy. If the AFFILIATION VALID bit is set to zero, the AFFILIATED STP INITIATOR SAS ADDRESS field may contain the SAS address (see 4.2.2) of an STP initiator port that previously had an affiliation in the specified affiliation context with the STP target port that contains the specified phy.

The STP I_T NEXUS LOSS SAS ADDRESS field contains the SAS address (see 4.2.2) of the last STP initiator port for which the STP target port experienced an I_T nexus loss (see 4.5) in the specified affiliation context.

The CRC field is defined in 10.4.3.2.

10.4.3.13 PHY CONTROL function

...

Table 91 defines the PHY OPERATION field.

Table 91 — PHY OPERATION **field** (part 1 of 2)

Code	Operation	Description
00h	NOP	No operation.
01h	LINK RESET	If the specified phy is not a virtual phy, perform a link reset sequence (see 4.4) on the specified phy and enable the specified phy. If the specified phy is a virtual phy, perform an internal reset and enable the specified phy. See 7.11 for BROADCAST (CHANGE) requirements related to this phy operation in an expander device. ~~Any~~All affiliations (see 7.17.5), if any, shall continue to be present. The phy shall bypass the SATA spinup hold state, if implemented (see 6.8.3.9). The SMP response shall be returned without waiting for the link reset to complete.
02h	HARD RESET	If the specified phy is not a virtual phy, perform a link reset sequence (see 4.4) on the specified phy and enable the specified phy. If the attached phy is a SAS phy or an expander phy, the link reset sequence shall include a hard reset sequence (see 4.4.2). If the attached phy is a SATA phy, the phy shall bypass the SATA spinup hold state. See 7.11 for BROADCAST (CHANGE) requirements related to this phy operation in an expander device. If the specified phy is a virtual phy, perform an internal reset and enable the specified phy. ~~Any~~All affiliations (see 7.17.5), if any, shall be cleared. The SMP response shall be returned without waiting for the hard reset to complete.
03h	DISABLE	Disable the specified phy (i.e., stop transmitting valid dwords and receiving dwords on the specified phy). The LINK RESET and HARD RESET operations may be used to enable the phy. See 7.11 for BROADCAST (CHANGE) requirements related to this phy operation in an expander device.
04h	Reserved	
05h	CLEAR ERROR LOG	Clear the error log counters reported in the REPORT PHY ERROR LOG function (see 10.4.3.6) for the specified phy.
06h	CLEAR AFFILIATION	Clear ~~an~~the affiliation (see 7.17.5), if any, from the STP initiator port with the same SAS address as the SMP initiator port that opened this SMP connection. If there is no such affiliation, the SMP target port shall return a function result of SMP FUNCTION FAILED in the response frame.

Table 91 — PHY OPERATION **field** (part 2 of 2)

Code	Operation	Description
07h	TRANSMIT SATA PORT SELECTION SIGNAL	This function shall only be supported by phys in an expander device. If the expander phy incorporates an STP/SATA bridge and supports SATA port selectors, the phy shall transmit the SATA port selection signal (see 6.6) which causes the SATA port selector to select the attached phy as the active host phy and make its other host phy inactive. See 7.11 for BROADCAST (CHANGE) requirements related to this phy operation in an expander device. ~~Any~~All affiliation~~s~~ (see 7.17.5), if any, shall be cleared. If the expander phy does not support SATA port selectors, then the SMP target port shall return a function result of PHY DOES NOT SUPPORT SATA. If the expander phy supports SATA port selectors but is attached to a SAS phy or an expander phy, the SMP target port shall return a function result of SMP FUNCTION FAILED.
08h	CLEAR STP I_T NEXUS LOSS	Clear the STP I_T nexus loss conditions reported in the REPORT PHY SATA function. The STP I_T NEXUS LOSS OCCURRED bit in the REPORT PHY SATA function for each affiliation context (see 10.4.3.7) shall be set to zero.
All others	Reserved	

...

G.1 STP differences from Serial ATA (SATA)

Some of the differences of STP compared with SATA are:

 a) STP adds addressing of multiple SATA devices. Each SATA device is assigned a SAS address by its attached expander device with STP/SATA bridge functionality. The STP initiator port understands addressing more than one STP target port;
 b) STP allows multiple STP initiator ports to share access to a SATA device behind an STP/SATA bridge using affiliations and/or multiple task file register contexts (see 7.17.5);
 c) interface power management is not supported;
 d) far-end analog loopback testing is not supported;
 e) far-end retimed loopback testing is not supported;
 f) near-end analog loopback testing is not supported;
 g) use of SATA_CONT is required; and
 h) BIST Activate frames are not supported.

G.1 Affiliation policies

G.1.1 Affiliation policies overview

SATA is based on a model that assumes a SATA device is controlled by a single SATA host, and does not address the notion of multiple SATA hosts having the ability to access any given SATA device.

With STP/SATA bridges, SATA devices are cast into an environment where multiple STP initiator ports, by sharing the SATA host port of the STP/SATA bridge, have access to the same SATA device. The SATA protocol used inside STP connections does not account for the possibility that more than one STP initiator port might be vying for access to the SATA device. Affiliations provide a way to ensure contention for a SATA

device does not result in incoherent access to the SATA device when commands from different STP initiator ports collide at the SATA device.

To prevent a SATA device from confusing commands from one STP initiator port with commands from another, an STP initiator port needs a means to maintain exclusive access to a SATA or STP device for the duration of the processing of a command.

For example, consider the case where an STP initiator port establishes a connection to send a command (e.g., a read), and then closes the connection while the SATA device (e.g., a disk drive) retrieves the data (e.g., performs a seek operation to the track containing the data). If, after the connection is closed, another STP initiator port is allowed to establish a connection and send another command, the SATA device would no longer have a means to determine which STP initiator port should receive the data when the device requests the connection to send the data for the first command. This is because, unlike SCSI target devices, SATA devices have no notion of multiple SATA hosts.

The consequences are worse for write commands since the result could be wrong data written to media, with the original data being overwritten and permanently lost.

Affiliation provides a means for an STP initiator port to establish atomic access to a SATA device across the processing of a command or series of commands to the SATA device, without requiring the STP initiator port to maintain a connection open to the STP target port for the duration of command processing.

G.1.2 Affiliation policy for static STP initiator port to STP target port mapping

Affiliations should not be used to enforce policies establishing fixed associations between STP initiator ports and STP target ports.

G.1.3 Affiliation policy with SATA queued commands and multiple STP initiator ports

STP initiator ports using queued commands when other STP initiator ports may be accessing the same STP target port should, at vendor-specific intervals, allow commands to complete and release the affiliation to allow other STP initiator ports access to the STP target port.

G.1.4 Applicability of affiliation for STP target ports

Affiliation may or may not be necessary for STP target ports depending on whether the STP target port tracks the STP initiator port's SAS address on each command received. If the STP target port has the means to manage and track commands from each STP initiator port independently, then affiliations are not necessary because the STP target port is capable of associating each information transfer with the appropriate STP initiator port, and is capable of establishing a connection to the appropriate STP initiator port when sending information back for a command.

An STP target port capable of tracking commands may support a limited number of STP initiator ports (i.e., more than one, but less than one per command) and use multiple affiliations in order to manage that restriction.

An STP target port that behaves the same as a SATA device, in that it maintains only a single ATA task file register imagecontext to be shared among all STP initiator ports, supports a single affiliations in order to provide a way for STP initiator ports to maintain exclusive access to the STP target port while commands remain outstanding. In this model, an STP target port is capable of establishing connections to an STP initiator port, but is only capable of remembering the SAS address of the last STP initiator port to establish a connection, and therefore is only capable of requesting a connection back to that same STP initiator port.

See 10.4.3.7 for an explanation of how an STP target port reports support for affiliations.