To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 13 January 2004
Subject: T10/03-334r3 ALIGN insertion clarifications

Revision History
Revision 0 (25 September 2003) first revision
Revision 1 (9 October 2003) keep the ALIGN(n) existing rotation rule rather than weaken it. Updated the ALIGN and NOTIFY sections.
Revision 2 (3 November 2003) incorporated comments from November SAS WG plus added text on when STP initiator throttling starts and stops.
Revision 3 (13 January 2004) incorporated comments from January SAS WG.

Related Documents
sas1r02 - Serial Attached SCSI 1.1 revision 2
04-032 – ALIGNs through expanders

Overview
This addresses a few issues with ALIGN and NOTIFY insertion for clock skew management and rate matching.

1. No need for consecutive ALIGNs in STP
SAS-1.0 requires that STP initiator phys send out 2 consecutive ALIGNs or NOTIFYs within each 256 dwords (like SATA). However, the STP target phy is not guaranteed to receive them consecutively, because one could be deleted for clock skew management (there is no difference between the 1/2048 and the 2/256 ALIGNs to a receiver).

The STP initiator phy should be required to insert them at the same rate (2/256 dwords), but not be required to send them consecutively, since it doesn't help the receiver.

2. Rate matching interaction
There has been some confusion about how rate matching and clock skew management ALIGNs interact.

- SAS 1/2048 clock skew management ALIGNs are not rate matched.
- SAS 2/256 STP initiator ALIGNs are rate matched. The requirement is to throttle the 1.5 Gbps stream. If these ALIGNs are not rate matched, the dword stream is too fast for the STP/SATA bridge to insert 2/256 on the SATA physical link. Of each set of 512 dwords (ignoring the 1/2048 clock skew management ALIGN)
  - 256 of them are rate match ALIGNs
  - 2 of them are STP initiator ALIGNs
  - 254 are SATA dwords (data dwords or SATA primitives).
- SAS 1/2 rate matching ALIGNs do not fulfill the role of either the 1/2048 or 2/256 ALIGNs.

This proposal restricts the clock skew management section (7.3) to discussing the 1/2048 rule and moves the 2/256 rule to a new section, after connections (7.12) and rate matching (7.13). It calls the 2/256 ALIGNs "STP initiator phy throttling" ALIGNs to avoid confusing them with the SAS clock skew management ALIGNs. They are related to clock skew management when they reach the SATA physical link, but that's outside the scope of this standard.

3. ALIGN before connection starts or ends
Before a connection, an expander phy A is sourcing 1/2048 ALIGNs per clock skew management rules. It might have sent 2047 dwords since the last ALIGN when it forwards an OPEN_ACCEPT; it then starts forwarding dwords received from the remote expander phy B. Phy B might just have received an ALIGN before the OPEN_ACCEPT, so might receive 2047 dwords until it receives

another ALIGN. This results in phy A transmitting 2047+2047=4094 dwords between ALIGNs, violating the 1/2048 clock skew management rule on that physical link.

An expander phy needs to output an ALIGN before forwarding an OPEN_ACCEPT or the clock skew management ALIGN frequency might drop below 1/2048.

Similarly, at the end of a connection, the expander phy is changing from forwarding ALIGNs to sourcing them itself. It might have forwarded 2047 non-ALIGN dwords before the CLOSE. Before sending a CLOSE, it needs to send an ALIGN or NOTIFY to meet the 1/2048 rule on behalf of the forwarded dword stream. (it also must start its 1/2048 counter then, not after the 3 CLOSE dwords, or it would 2050 dwords through next).

Similarly, if the expander chooses to send a BREAK over an open connection, it needs to send an ALIGN or NOTIFY first and start its 1/2048 counter.

Overall, the expander must guarantee ALIGN rules while originating dwords and while changing between originating and forwarding.

4. Rolling windows

Some have questioned whether the 1/2048 requirement allows two back-to-back 2048 dword windows, where the first window has the ALIGN as the first dword and the last window has the ALIGN as the last dword (allowing 2047+2047=5094 dwords in between). This is incorrect; a note is proposed to try to make it clear.

Suggested Changes

3.1.xx throttling: Reducing the rate at which an STP initiator phy is sourcing dwords. See 7.17.x.

7.2.5.2 ALIGN

ALIGNs are used for:

a) OOB signals;
b) character and dword alignment during the speed negotiation sequence;
c) clock skew management after the phy reset sequence (see 7.3); ~~and~~
d) rate matching during connections (see 7.13)~~.~~; and
e) STP initiator throttling during STP connections (see 7.17.x).

Table 56 defines the different versions of ALIGN primitives.

Phys shall use ALIGN (0) to construct OOB signals as described in 6.5. Phys shall use ALIGN (0) and ALIGN (1) during the speed negotiation sequence as described in 6.6.4.2. Phys shall rotate through ALIGN (0), ALIGN (1), ALIGN (2), and ALIGN (3) for all ALIGNs sent after the phy reset sequence ~~for clock skew management (see 7.3) and rate matching (see 7.13)~~.

Phys receiving ALIGNs after the phy reset sequence shall not verify the rotation and shall accept any of the ALIGNs at any time.

Phys shall only detect an ALIGN after decoding all four characters in the primitive.

NOTE 14 - SATA devices are allowed to decode every dword starting with a K28.5 as an ALIGN, since ALIGN is the only primitive defined starting with K28.5.

For clock skew management ~~and~~ , rate matching, and STP initiator throttling, ALIGNs may be replaced by NOTIFYs (see 7.2.5.9).

7.2.5.9 NOTIFY

NOTIFY may be transmitted in place of any ALIGN (see 7.2.5.2) being transmitted for clock skew management (see 7.3) ~~or~~ , rate matching (see 7.13), or STP initiator throttling (see 7.17.x). Substitution of a NOTIFY may or may not affect the ALIGN ~~sequencing~~ rotation (i.e., the NOTIFY may take the place of one of the ALIGNs in the rotation through ALIGN (0), ALIGN (1), ALIGN (2), or ALIGN (3) or it may delay the rotation). A specific NOTIFY shall not be transmitted a second time until at least three ALIGNs or different NOTIFYs have been transmitted.

NOTIFY shall not be forwarded through expander devices. Expander devices shall substitute an ALIGN for a NOTIFY if necessary.

SAS target devices are not required to detect every transmitted NOTIFY.

...

7.3 Clock skew management

The internal clock for a device is typically based on a PLL with its own clock generator and is used when transmitting dwords on the physical link. When receiving, however, dwords need to be latched based on a clock derived from the input bit stream itself. Although the input clock is nominally a fixed frequency, it may differ slightly from the internal clock frequency ~~due to accepted manufacturing tolerance and, for SATA physical links, due to spread spectrum clocking~~. up to the physical link rate tolerance defined in table 24 (see 5.3.2). Over time, if the input clock is faster than the internal clock, the device may receive a dword and not be able to forward it to an internal buffer; this is called an overrun. If the input clock is slower than the internal clock, the device may not have a dword when needed in an internal buffer; this is called an underrun.

To solve this problem, transmitting devices insert ALIGNs or NOTIFYs in the dword stream. Receivers may pass ALIGNs and NOTIFYs through to their internal buffers, or may strip them out when an overrun occurs. Receivers add ALIGNs or NOTIFYs when an underrun occurs. The internal logic shall ignore all ALIGNs and NOTIFYs that arrive in the internal buffers.

Elasticity buffer circuitry, as shown in figure 74, is required to absorb the slight differences in frequencies between the SAS initiator phy, SAS target phy, and expander phys. The frequency tolerance for a phy is specified in 5.3.2.

[**Figure 74 — Elasticity buffers**]

A phy that is the original source for the dword stream (i.e., a phy that is not an expander phy forwarding dwords from another expander phy) shall ~~periodically~~ insert ~~ALIGNs or NOTIFYs into the dword stream as shown in table 65~~one ALIGN or NOTIFY for clock skew management within every 2 048 dwords (i.e., every overlapping window of 2 048 dwords).

ALIGNs and NOTIFYs inserted for clock skew management are in addition to ALIGNs and NOTIFYs inserted for rate matching (see 7.13) and STP initiator throttling (see 7.17.x). See Annex L for a summary of their combined requirements.

~~**Table 65 — Clock skew management ALIGN or NOTIFY insertion requirements**~~

~~**Original source of dword stream**~~	~~**Clock skew management ALIGN or NOTIFY requirements**~~
~~Either: a) SSP initiator phy or SSP target phy in SSP connection; b) SMP initiator phy or SMP target phy in SMP connection; c) STP target phy in an STP connection; or d) any phy outside connections.~~	~~One ALIGN or NOTIFY within every 2 048 dwords~~
~~STP initiator phy in an STP connection~~	~~Two consecutive ALIGNs or NOTIFYs within each 256 dwords plus one ALIGN or NOTIFY within each 2 048 dwords~~

[changes in the rest of this section are new in r2]

An expander device that is forwarding dwords (i.e., is not the original source) is allowed to insert or delete as many ALIGNs and/or NOTIFYs as required to match the transmit and receive connection rates ~~(e.g., i.~~ It is not required to ~~ensure that it~~ transmit~~s one ALIGN or NOTIFY within~~

~~every 2 048 dwords~~any particular number of ALIGNs and/or NOTIFYs for clock skew management when forwarding to a SAS physical link~~)~~.

NOTE 17 - One possible implementation for expander devices forwarding dwords is for the expander device to delete all ALIGNs and NOTIFYs received and to insert ALIGNs and/or NOTIFYs at the transmit port whenever its elasticity buffer is empty.

The STP target port of an STP/SATA bridge is allowed to insert or delete as many ALIGNs and/or NOTIFYs as required to match the transmit and receive connection rates ~~(e.g., i.~~ It is not required to ~~ensure that it~~ transmit~~s one ALIGN or NOTIFY within every 2 048 dwords~~any particular number of ALIGNs and/or NOTIFYs for clock skew management when forwarding to a SAS physical link~~)~~. ~~The STP target port in an STP/SATA bridge is not required to insert ALIGNs or NOTIFYs in pairs when transmitting dwords.~~and is not required to ensure that any ALIGNs and/or NOTIFYs it transmits are in pairs.

NOTE 18 - Due to clock skew ALIGN and NOTIFY removal, the STP target port may not receive a pair of ALIGNs and/or NOTIFYs every 256 dwords, even though the STP initiator port transmitted at least one pair. However, the rate of the dword stream allows for ALIGN or NOTIFY insertion by the STP/SATA bridge. One possible implementation is for the STP/SATA bridge to delete all ALIGNs and NOTIFYs received by the STP target port and to insert two consecutive ALIGNs at the SATA host port when its elasticity buffer is empty or when 254 non-ALIGN dwords have been transmitted. It may need to buffer up to 2 dwords concurrently being received by the STP target port while it does so.

7.13 Rate matching

Each successful connection request contains the connection rate (see 4.1.10) of the pathway.

~~Every~~ Each phy in the ~~physical link~~ pathway shall insert ALIGNs and/or NOTIFYs between dwords ~~to match the connection rate~~if its physical link rate is faster than the connection rate as described in table xx~~.~~

Table xx — Rate matching ALIGN and/or NOTIFY insertion requirements

Physical link rate	Connection rate	Requirement
1.5 Gbps	1.5 Gbps	none
3.0 Gbps	1.5 Gbps	One ALIGN or NOTIFY within every 2 dwords that are not clock skew management ALIGNs or NOTIFYs (i.e., every overlapping window of 2 dwords)
3.0 Gbps	3.0 Gbps	none

~~Phys receiving ALIGNs and NOTIFYs delete them regardless of whether the ALIGNs and NOTIFYs were inserted for clock skew management (see 7.3) or for rate matching.~~

~~The faster phy shall rotate between ALIGN (0), ALIGN (1), ALIGN (2), and ALIGN (3) to reduce long strings of repeated patterns appearing on the physical link. NOTIFYs may be used to replace ALIGNs (see 7.2.5.9).~~

ALIGNs and NOTIFYs inserted for rate matching are in addition to ALIGNs and NOTIFYs inserted for clock skew management (see 7.3) and STP initiator throttling (see 7.17.x). See Annex L for a summary of their combined requirements.

Figure 85 shows an example of rate matching between a 3,0 Gbps source phy and a 3,0 Gbps destination phy, with an intermediate 1,5 Gbps physical link in between.

A phy shall start inserting ALIGNs and/or NOTIFYs for rate matching at the selected connection rate with the first dword that is not an ALIGN or NOTIFY inserted for clock skew management following:

a) transmitting the EOAF for an OPEN address frame; or

b) transmitting an OPEN_ACCEPT.

The source phy transmits idle dwords including ALIGNs and NOTIFYs at the selected connection rate while waiting for the connection response. This enables each expander device to start forwarding dwords from the source phy to the destination phy after forwarding an OPEN_ACCEPT.

A phy shall stop inserting ALIGNs and/or NOTIFYs for rate matching after:

a) transmitting the first dword in a CLOSE;
b) transmitting the first dword in a BREAK;
c) receiving an OPEN_REJECT for a connection request; or
d) losing arbitration to a received OPEN address frame.

If an STP initiator port discovers a SATA device behind an STP/SATA bridge with a physical link rate greater than the maximum connection rate supported by the pathway from the STP initiator port, the STP initiator port should use the SMP PHY CONTROL function (see 10.4.3.10) to set the MAXIMUM PHYSICAL LINK RATE field of the expander phy attached to the SATA device to the maximum connection rate supported by the pathway.

7.15 XL (link layer for expander phys) state machine

...

7.15.2 XL transmitter and receiver

The XL transmitter receives the following messages from the XL state machine indicating primitive sequences, frames, and dwords to transmit:

a) Transmit Idle Dword;
b) Transmit AIP with an argument indicating the specific type (e.g., Transmit AIP (Normal));
c) Transmit BREAK;
d) Transmit BROADCAST with an argument indicating the specific type (e.g., Transmit BROADCAST (Change));
e) Transmit CLOSE with an argument indicating the specific type (e.g., Transmit CLOSE (Normal));
f) Transmit OPEN_ACCEPT;
g) Transmit OPEN_REJECT, with an argument indicating the specific type (e.g., Transmit OPEN_REJECT (No Destination));
h) Transmit OPEN Address Frame; and
i) Transmit Dword.

The XL transmitter sends the following messages to the XL state machine:
a) OPEN Address Frame Transmitted.

The XL transmitter shall ensure clock skew management requirements are met (see 7.3) while originating dwords.

The XL transmitter shall ensure clock skew management requirements are met (see 7.3) during and after switching from forwarding dwords to originating dwords, including, for example:

a) when transmitting BREAK;
b) when transmitting CLOSE;
c) when transmitting an idle dword after closing a connection (i.e., after receiving BREAK or CLOSE);
d) while transmitting a SATA frame to a SAS physical link, when transmitting the first SATA HOLDA in response to detection of SATA_HOLD; and
e) while receiving dwords of a SATA frame from a SAS physical link, when transmitting SATA_HOLD.

NOTE: The XL transmitter may always insert an ALIGN or NOTIFY before transmitting a BREAK, CLOSE, or SATA_HOLDA to meet clock skew management requirements.

The XL transmitter shall insert an ALIGN or NOTIFY before switching from originating dwords to forwarding dwords, including, for example:

a) when transmitting OPEN_ACCEPT;

b) when transmitting the last idle dword before a connection is established (i.e., after receiving OPEN_ACCEPT);
c) while transmitting a SATA frame to a SAS physical link, when transmitting the last dword from the SATA flow control buffer in response to release of SATA_HOLD;
d) while transmitting a SATA frame to a SAS physical link, when transmitting the last SATA_HOLDA in response to release of SATA_HOLD (e.g., if the SATA flow control buffer is empty); and
e) while receiving dwords of a SATA frame from a SAS physical link, when transmitting the last SATA_HOLD.

NOTE: This ensures that clock skew management requirements are met, even if the forwarded dword stream does not include an ALIGN or NOTIFY until the last possible dword.

The XL transmitter shall ~~insert ALIGNs and NOTIFYs needed for~~ ensure rate matching requirements are met (see 7.13) during a connection.

The XL transmitter shall ensure STP initiator port throttling requirements are met (see 7.3) when:

a) transmitting dwords in the direction of an STP target port while originating dwords (e.g., while transmitting HOLD and unloading the SATA flow control buffer);
b) switching from forwarding dwords to originating dwords; and
c) switching from originating dwords to forwarding dwords.

...

7.17 STP link layer

7.17.x STP initiator phy throttling

On a SATA physical link, phys are required to transmit two consecutive ALIGN (0)s within every 256 dwords. To ensure an STP/SATA bridge is able to meet this requirement, an STP initiator phy has to reduce (i.e., throttle) the rate at which it is sourcing dwords by the same amount.

During an STP connection, an STP initiator phy shall insert two ALIGNs or NOTIFYs within every 256 dwords (i.e., within every overlapping window of 256 dwords) that are not ALIGNs or NOTIFYs for clock skew management or rate matching. They are not required to be inserted consecutively, because a phy in the pathway may delete one of them for clock skew management since STP initiator throttling ALIGNs and NOTIFYs are indistinguishable from clock skew management ALIGNs and NOTIFYs.

STP target phys are not required to insert extra ALIGNs and/or NOTIFYs, because SATA hosts are not supported by SAS domains. STP initiator phys, the only recipients of data from STP target phys, do not require extra ALIGNs or NOTIFYs.

ALIGNs and NOTIFYs inserted for STP initiator phy throttling are in addition to ALIGNs and NOTIFYs inserted for clock skew management (see 7.3) and rate matching (see 7.13). See Annex L for a summary of their combined requirements.

[Specific start/stop times new in r2:]

A phy shall start inserting ALIGNs and NOTIFYs for STP initiator throttling after:

a) transmitting an OPEN_ACCEPT; or
b) sending the first SATA primitive after receiving an OPEN_ACCEPT.

A phy shall stop inserting ALIGNs and NOTIFYs for STP initiator throttling after:

a) transmitting the first dword in a CLOSE; or
b) transmitting the first dword in a BREAK.

Annex L [all new]

(informative)

ALIGN and/or NOTIFY insertion summary

Table L.1 shows all the possible combinations of ALIGN and/or NOTIFY insertion rates for clock skew management (see 7.3), rate matching (see 7.13), and STP initiator throttling (see 7.17.x).

Table L.1. ALIGN and/or NOTIFY insertion rate examples

Physical link rate	Connection rate	Type of dword stream	ALIGN and/or NOTIFY insertion rate (per dword)	Minimum number of ALIGNs and/or NOTIFYs within any 2 048 dword window
3.0 Gbps	3.0 Gbps	all but to STP target	1 per 2 048	1
3.0 Gbps	3.0 Gbps	to STP target	1 per 2 048 + 2 per 256	17
3.0 Gbps	1.5 Gbps	all but to STP target	1 per 2 048 + 1 per 2	1024 or 1025 (a)
3.0 Gbps	1.5 Gbps	to STP target	1 per 2 048 + 1 per 2 + 2 per 256	1032 or 1033 (b)
1.5 Gbps	1.5 Gbps	all but to STP target	1 per 2 048	1
1.5 Gbps	1.5 Gbps	to STP target	1 per 2 048 + 2 per 256	17

(a) There are 2 047 dwords left after the clock skew management ALIGN or NOTIFY. These alternate between rate matching ALIGNs and other dwords. These requirements alternate every 2048 running windows.

(b) There are 2 047 dwords left after the clock skew management ALIGN or NOTIFY. These alternate between rate matching ALIGNs and/or NOTIFYs and other dwords, leaving 1 024 or 1 025 dwords that are neither clock skew management nor rate matching ALIGNs and/or NOTIFYs. Of these, 2 per 256 (i.e., 8 per 1024) is an STP initiator phy throttling ALIGN and/or NOTIFY. These requirements alternate every 2048 running windows of 2048 dwords.