

To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 17 January 2006
Subject: 06-054r1 SAS-2 Expander issues resolutions

Revision history

Revision 0 (9 January 2006) First revision
Revision 1 (17 January 2006) Incorporated comments from the January 2006 SAS protocol WG.

Related documents

05-373r2 SAS-1.1 Expander issues (Craig Stoops, Expert I/O)

Overview

This addresses the issues raised by 05-373r2.

Issue #1: SAS Expander additional qualification in Arb Won

This is resolved by #4; no changes requested.

Issue #2: Phy Status clarification in XL1:Request Path

In 3.6.3, add "The state machine description text for each state wholly defines the messages sent while the state machine is in that state. If a state machine in one state repeatedly sending a message transitions to another state, it stops repeatedly sending that message unless stated otherwise in the new state."

Issue #3: Specification clarification for Phy Status (Partial Pathway)

In 4.6.6.3 table 11, change the last part of a) in both the Phy Status (Partial Pathway) and (Blocked Partial Pathway) rows as follows.

Change the Partial Pathway text to "... in the XL3:Open_Confirm_Wait state and the last AIP transmitted was not AIP (WAITING ON PARTIAL)), or the expander phy is in the XL6:Open_Response_Wait state and the last AIP received was not AIP (WAITING ON PARTIAL))"

Change the Blocked Partial Pathway text to "... in the XL3:Open_Confirm_Wait state and the last AIP transmitted was AIP (WAITING ON PARTIAL)), or the expander phy is in the XL6:Open_Response_Wait state and the last AIP received was AIP (WAITING ON PARTIAL))"

Issue #4: <see below>

Issue #5: End device connection rate matching algorithm causes expander overflow

In 7.15.10.1 (XL7) change "The state shall send Forward Dword requests..." to "If this state has not sent a Forward Close request to the ECR, this state shall send Forward Dword requests..."

In 4.6.4 (ECR), add:

"When forwarding dwords during a connection from a source phy with a higher physical link rate to a destination phy with a lower physical link rate, rate matching (see 7.13) ensures the dwords are at a connection rate equal to or less than the lower physical link rate. The ECR may be requested to forward more dwords than the destination phy is able to accept if:

- a) an invalid dword occurs during an ALIGN or NOTIFY;
- b) an invalid dword occurs during a CLOSE; or
- c) multiple invalid dwords occur during a BREAK.

The ECR may discard dwords if needed and count them as receive elasticity buffer overflows."

Issue #6: Backoff Reverse Path OPEN address frame connection rate checking

Not an issue; Reverse Path is only sent if the rates are equal.

Issue #7: Receiving Forward Open indications while in XL4:Open Reject

XL4 should go directly to XL5 and include the Forward Open indications as a hated argument to the transition. Enhanced BPP support to recognize that a Transmit Broadcast indication could also show up during XL5 is deferred until a future proposal (right now there is no interlock to ensure BPP only makes its requests at the appropriate times).

Add:

7.15.6.n Transition XL4:Open_Reject to XL5:Forward_Open

This transition shall occur if a Forward Open indication is received. This transition shall include an OPEN Address Frame Received argument containing the arguments received in the Forward Open indication.

Issue #4: Conflicting grants and Transition XL1:Request Path to XL5:Forward Open

Make the changes detailed below.

7.12.4.1 Arbitration and resource management in an expander device overview

The ECM shall arbitrate and assign or deny path resources for Request Path requests from each expander phy.

Arbitration includes adherence to the SAS arbitration fairness algorithm and path recovery. Path recovery is used to avoid potential deadlock scenarios within the SAS topology by deterministically choosing which partial pathway(s) to tear down to allow at least one connection to complete.

Several of the Request Path arguments are used for arbitration. The Arbitration Wait Time, Source SAS Address, and Connection Rate arguments are filled in from the received OPEN address frame and are used to by the ECM to compare Request Path requests. The Retry Priority Status argument is used to prevent the Arbitration Wait Time argument from being considered during an arbitration which occurs after a Backoff Retry response is sent by an expander phy (see 7.15.4).

An expander phy shall set the Retry Priority Status argument to IGNORE AWT when it requests a path after:

- a) it has forwarded an OPEN address frame to the physical link;
- b) an OPEN address frame is received with higher arbitration priority (see 7.12.3); and
- c) the destination SAS address and connection rate of the received OPEN address frame are not equal to the source SAS address and connection rate of the transmitted OPEN address frame (see 7.15.4 and 7.15.9).

Otherwise, the expander phy shall set the Retry Priority Status argument to NORMAL.

The ECM responds to ~~connection~~each Request Path requests by returning an Arb Won, Arb Lost, or Arb Reject confirmation to the requesting expander phy.

If two or more Request Path requests contend and all of the Request Path requests include a Retry Priority Status argument set to NORMAL, the ECM shall select the winner by comparing the OPEN address frame contents described in table 1.

Table 1 — Arbitration priority for contending Request Path requests in the ECM when all requests have Retry Priority Status arguments of NORMAL

Bits 83-68 (83 is MSB)	Bits 67-4	Bits 3-0 (0 is LSB)
ARBITRATION WAIT TIME field value	SOURCE SAS ADDRESS field value	CONNECTION RATE field value

If two or more Request Path requests contend and one or more of the Request Path requests include a Retry Priority Status argument set to IGNORE AWT, the ECM shall select the winner from the set of Request Path

requests with Retry Priority Status arguments ~~offset to~~ IGNORE AWT by comparing the OPEN address frame contents described in table 2.

Table 2 — Arbitration priority for contending Request Path requests in the ECM among requests with Retry Priority Status arguments of IGNORE AWT

Bits 67-4 (67 is MSB)	Bits 3-0 (0 is LSB)
SOURCE SAS ADDRESS field value	CONNECTION RATE field value

The ECM shall generate the Arb Won confirmation when all of the following conditions are met:

- a) the ~~connection~~Request Path request maps to a ~~a~~destination expander phy that:
 - A) supports the connection rate; and
 - B) is not reporting a Phy Status (Partial Pathway), Phy Status (Blocked Partial Pathway), or Phy Status (Connection) response, unless that expander phy is arbitrating for the requesting expander phy ~~making this connection request~~;
- b) there are sufficient routing resources to complete the connection request;
- c) no higher priority ~~connection~~Request Path requests (see table 101 and table 102) are present with ~~this~~the requesting expander phy as the destination; and
- d) the ~~connection~~Request Path request is ~~chosen as~~ the highest priority Request Path request (see table 101 and table 102) ~~connection request in the expander device~~ mapping to the ~~specified~~ destination expander phy (i.e., only send one Arb Won for concurrent connection requests to the same destination phy).

The ECM shall generate the Arb Lost confirmation when all of the following conditions are met:

- a) the ~~connection~~Request Path request maps to a ~~a~~destination expander phy that:
 - A) supports the connection rate; and
 - B) is not reporting a Phy Status (Partial Pathway), Phy Status (Blocked Partial Pathway), or Phy Status (Connection) response unless that expander phy is arbitrating for the requesting expander phy ~~making this connection request~~;
- b) there are sufficient routing resources to complete the connection request; and
- c) one of the following conditions are met:
 - A) the destination expander phy ~~of this connection request has received a higher priority OPEN address frame with this~~ is making a Request Path request with the requesting expander phy as its destination (i.e., when two expander phys both receive an OPEN address frame destined for each other, the ECM shall provide the Arb Lost confirmation to the expander phy that received the lowest priority OPEN address frame); or
 - B) the ECM is sending an Arb Won confirmation to another expander phy that is using the requesting expander phy as the destination.

The ECM shall generate the following Arb Reject confirmation when any of the following conditions are met and all the Arb Won conditions are not met:

- 1) Arb Reject (No Destination) or Arb Reject (Bad Destination) if the ~~connection~~Request Path request does not map to a ~~a~~destination expander phy that is not part of the same expander port as the requesting expander phy (i.e., there is no direct routing or table routing match and there is no subtractive phy)(see 7.12.4.3 and 7.12.4.4);
- 2) Arb Reject (Bad Connection Rate) if the ~~connection~~Request Path request does not map to any expander phy that supports the connection rate (i.e., none of the prospective physical links support the requested connection rate); or
- 3) Arb Reject (Pathway Blocked) if the ~~connection~~Request Path request maps to expander phys that all contain blocked partial pathways (i.e., are all returning Phy Status (Blocked Partial Pathway)) and pathway recovery rules require this ~~connection~~Request Path request to release path resources (see 7.12.4.6).