To: T10 Technical Committee
From: Rob Elliott, Compaq Computer Corporation (Robert.Elliott@compaq.com)
Date: 28 August 1999
Subject: QAS Arbitration Fairness

T10/99-160 standardized enabling/disabling arbitration fairness for normal devices. Currently, SPI-3 revision 8 just requires QAS devices to support arbitration fairness, providing some inconsistent rules about how to be fair to non-QAS devices. This proposal addresses QAS fairness to non-QAS devices.

Cases to consider
1. All devices are using normal arbitration. T10/99-160 defines how to enable/disable fairness.

2. All devices are using QAS. They could stay in QAS forever, where fairness is required. However, a hot-plugged device joins the bus with QAS disabled, so could be blocked from arbitrating or being selected if this assumption is made.

3. One set of initiators and targets is using normal arbitration, while another set is using QAS - each set is oblivious to the other set. The QAS set must drop to BUS FREE periodically to avoid starving the non-QAS set.

4. Some initiators and targets are using normal arbitration, while others are using QAS - they have all negotiated with each other. This case can be minimized, but not eliminated, by forcing devices to negotiate down to normal arbitration with this rule that was accepted by the SPI-3 Working Group (but rejected by the SCSI Working Group the next day) in May 1999:

> *Targets that receive a PPR message from any initiator with QAS disabled shall renegotiate to disable QAS with all initiators with which QAS had been enabled. Initiators that receive a PPR message from any target with QAS disabled shall renegotiate to disable QAS with all targets with which QAS had been enabled.*

There is no guarantee that the devices will all negotiate with each other, so the problem might still exist. Also, a hot plug would still break, since the new device joins with QAS disabled.

Conclusion: QAS devices must drop to BUS FREE periodically to avoid starving non-QAS devices.

Dropping back to normal arbitration
Several methods have been discussed for how often to drop back:
- After entering QAS, each target counts how often it wins. When a threshold is reached, the target forces a BUS FREE.
- After entering QAS, each target counts how often it wins. When a threshold is reached, the target stops arbitrating. Eventually all targets stop arbitrating and a BUS FREE occurs.
- After entering QAS, each target counts how often its fairness register clears (how often all the lower IDs are serviced). When a threshold is reached, the target forces a BUS FREE.
- After entering QAS, each target counts how often it sends QAS REQUEST messages. When a threshold is reached, the target forces a BUS FREE.

In a system with disconnect disabled, a target never wins; only the initiators arbitrate and win. Unless the initiators refused to arbitrate, non-QAS devices would be starved if targets were counting how often they won or how often the fairness register cleared. Counting how many QAS REQUEST messages are sent seems most robust.

Once in BUS FREE, all the QAS devices should set all the bits in their fairness registers to let non-QAS devices each have one shot at the bus. This way, during the first normal arbitration all the non-QAS devices will arbitrate and all the QAS devices will refrain from arbitrating. During subsequent arbitrations, the QAS devices should update fairness bits for both lower and higher priority devices. This ensures that a non-QAS device with a higher ID which did not win the first arbitration gets a chance.

(An alternative idea is to have an initiator arbitrate to select itself and force a BUS FREE periodically.)

Proposed arbitration fairness rules
Non-QAS device during normal arbitration (this is unchanged):

1) When not arbitrating, latch all lower priority IDs that lost arbitration into fairness register
2) If the device wants to arbitrate, wait for fairness register = 0. Latch all losers into fairness register for next round.

QAS device during QAS arbitration (modified with a counter):
1) When not arbitrating, latch all lower priority IDs that lost arbitration. These will only be QAS capable devices.
2) If the device wants to arbitrate, wait for fairness register = 0. Latch all losers into fairness register for next round.
3) If sending QAS_REQUEST the <n>th time since a BUS FREE, the target drops down to BUS FREE and sets all fairness bits.
4) All QAS devices set all their fairness bits upon going to BUS FREE after a QAS cycle.

QAS device during normal arbitration (new rules):
1) Set all fairness bits when leaving QAS into BUS FREE
2) When not wanting to arbitrate, latch *all* IDs that lost arbitration (higher or lower priority)
3) If the device wants to arbitrate, wait for fairness register = 0 and enter normal arbitration. The QAS capable target will send QAS REQUEST to disconnect, and QAS rounds will occur.

QAS device disconnect-reconnect mode page:
Interpret the 3-bit FAIR ARBITRATION field as the number of times that the target may send QAS_REQUEST before dropping to BUS FREE. 0 means be unfair in normal arbitration, fair within QAS, but not fair from QAS devices to non-QAS devices. 1 through 7 give are fair in normal arbitration, fair within QAS, and progressively less fair from QAS devices to non-QAS devices.

Example
Assume there is a QAS device at ID 4. Non-QAS devices are at IDs 6, 5, and 2. Data bus and fairness register bits are shown in this order: 7654_3210.

Enter BUS FREE from QAS (perhaps because the target reached <n>)
0000_0000 = fairness reg (device 6)
0000_0000 = fairness reg (device 5)
1110_1111 = fairness reg (device 4 QAS)
0000_0000 = fairness reg (device 2)
Start normal arbitration
All the QAS devices avoid arbitrating because fairness bits are set
All the non-QAS devices want to arbitrate
0110_0100 = arbitration IDs
Device 6 wins 0100_0000 = winner ID
Update fairness
0010_0100 = fairness reg (device 6)(won)
0000_0000 = fairness reg (device 5)(lost)
0010_0100 = fairness reg (device 4 QAS)(waiting)
0000_0000 = fairness reg (device 2)(lost)
The QAS devices fairness registers now have IDs of all non-QAS devices that wanted in.
Run cycles…
BUS FREE
Enter normal arbitration
QAS devices still waiting. Non-QAS devices still arbitrating
0010_0100 = arbitration IDs
Device 5 wins 0010_0000 = winner ID
Update fairness
0000_0100 = fairness reg (device 6)(wants to rearb)
0000_0100 = fairness reg (device 5)(won)
0000_0100 = fairness reg (device 4 QAS)(waiting)
0000_0000 = fairness reg (device 2)(lost)
Run cycles…
BUS FREE
Enter normal arbitration
QAS devices still waiting. Non-QAS devices still arbitrating
0000_0100 = arbitration IDs
Device 2 wins 0000_0100 = winner ID

Update fairness

0000_0000 = fairness reg (device 6)(wants to rearb)
0000_0000 = fairness reg (device 5)(done)
0000_0000 = fairness reg (device 4 QAS)(waiting)
0000_0000 = fairness reg (device 2)(won)

Now, the non-QAS devices are able to rearbitrate again. But, the QAS device can arbitrate also. Thus, the non-QAS devices will get to win for one to two rounds (if the QAS device had lowest priority) before QAS takes over.

Enter normal arbitration
Assume QAS device arbitrates, as do non-QAS devices 6 and 2.

0101_0100 = arbitration IDs
Device 6 wins 0100_0000 = winner ID
Update fairness

0001_0100 = fairness reg (device 6)(won)
0001_0100 = fairness reg (device 5)(done)
0000_0100 = fairness reg (device 4 QAS)(waiting)
0000_0000 = fairness reg (device 2)(wants to rearb)

Run cycles…
BUS FREE
Enter normal arbitration

0001_0100 = arbitration IDs
Device 4 wins 0001_0100 = winner ID
Update fairness

0000_0100 = fairness reg (device 6)(done)
0000_0100 = fairness reg (device 5)(done)
0000_0100 = fairness reg (device 4 QAS)(won)
0000_0000 = fairness reg (device 2)(lost)

Select a QAS target
Run cycles…
Target sends QAS_REQUEST message instead of BUS FREE
Enter QAS

When we eventually return to BUS FREE, all the non-QAS devices still have non-zero fairness registers indicating the device 2 should go first. Their fairness registers do not get reset during QAS arbitrations. The QAS device will reset its register to all 1's when going back to BUS FREE. Device 2 will go first, then every device's fairness register will clear and the QAS device will win in order. Only one normal arbitration round should occur whenever non-QAS devices were left hanging as QAS resumed.

Proposed changes to SPI-3 revision 8
Changes are needed in 18.1.1 (Disconnect-reconnect mode page), 10.2 and 10.2.1 (Arbitration bus phase), and Annex B (SCSI Bus Fairness). The bare requirements for the fairness algorithm have been copied here from annex B. The goal is to put the basic rules here and let annex B describe how the algorithm works. Italicized notes in brackets are just explanatory, not intended to go into the standard.

Section 3.1 Definitions
3.1.x **QAS device**: a SCSI device which has negotiated to enable QAS with any other SCSI device.
3.1.xx **non-QAS device**: a SCSI device which has not negotiated to enable QAS with any other SCSI device.
[these make some text in section 10 easier to read]

Section 18.1.1:
(in the text after table 67)

> "If the FAIR ARBITRATION field is set to 000b, the target shall not use arbitration fairness during normal arbitration. If this field is set to a nonzero value, the target shall use arbitration fairness during normal arbitration.
>
> Regardless of the value in the FAIR ARBITRATION filed the target shall use arbitration fairness during QAS.

A target using QAS shall interpret the FAIR ARBITRATION field as a QAS LIMIT value, from 0 to 7. If QAS LIMIT is zero, the target shall always send QAS REQUEST messages to start a new arbitration phase. If QAS LIMIT is nonzero, a target using QAS shall drop to BUS FREE instead of sending a QAS REQUEST message if it has sent a QAS_REQUEST message QAS LIMIT times since the last BUS FREE. See 10.2.3 for more details.

10.2 Arbitration

Arbitration allows one SCSI device to gain control of the SCSI bus to allow that SCSI device to initiate or resume a task.

There are two methods that a SCSI device may use to arbitrate for the SCSI bus~~:~~ normal arbitration and QAS. Normal arbitration is mandatory and requires the detection of a BUS FREE phase on the SCSI bus before starting. QAS is optional and, when enabled, requires the detection of a QAS REQUEST message (see 16.2.10) before starting.

SCSI devices with arbitration fairness enabled shall maintain a fairness register which records the SCSI IDs of devices that need a chance to arbitrate (see Annex B). Fairness in normal arbitration is enabled in targets by the Disconnect-Reconnect mode page (see 18.1.1). Fairness is always enabled in quick arbitration.

10.2.1 NORMAL ARBITRATION phase

The procedure for a SCSI device to obtain control of the SCSI bus is as follows:

a) The SCSI device shall first wait for the BUS FREE phase to occur. The BUS FREE phase is detected whenever both the BSY and SEL signals are simultaneously and continuously false for a minimum of a bus settle delay.

NOTE 21 - This bus settle delay is necessary because a transmission line phenomenon known as a wired-OR glitch may cause the BSY signal to briefly appear false, even though it is being driven true.

a.2) After detecting BUS FREE after a QAS phase, each QAS device with arbitration fairness enabled in normal arbitration (e.g. targets with QAS LIMIT set to a non-zero value - see 18.1.1) shall set all bits in its fairness register except its own.

b) The SCSI device shall wait a minimum of a bus free delay after detection of the BUS FREE phase (i.e. after the BSY and SEL signals are both false for a bus settle delay) before driving any signal.

c) Following the bus free delay in step (b), the SCSI device may arbitrate for the SCSI bus by asserting both the BSY signal and its own SCSI ID~~, h.~~ However the SCSI device shall not arbitrate (i.e. assert the BSY signal and its SCSI ID) if more than a bus set delay has passed since the BUS FREE phase was last observed. If arbitration fairness is enabled ~~in the Disconnect-Reconnect mode page (see 18.1.1)~~, the SCSI device shall not arbitrate until its fairness register is cleared ~~(see Annex B)~~.
[initiators might be fair, but don't have a mode page to enable it]

NOTE 22 - There is no maximum delay before asserting the BSY signal and the SCSI ID following the bus free delay in step (b) as long as the bus remains in the BUS FREE phase. However, SCSI devices that delay longer than a bus settle delay plus a bus set delay from the time when the BSY and SEL signals first become false may fail to participate in arbitration when competing with faster SCSI devices, and may not be be ensured fair arbitration by the arbitration fairness algorithm.

d) After waiting at least an arbitration delay (measured from its assertion of the BSY signal) the SCSI device shall examine the DATA BUS.

A) If no higher priority SCSI ID bit is true on the DATA BUS, then the SCSI device has won the arbitration and it shall assert the SEL signal.

B) If a higher priority SCSI ID bit is true on the DATA BUS (see table 27 for the SCSI ID arbitration priorities), then the SCSI device has lost the arbitration and the SCSI device shall release the BSY signal and the SCSI ID after the SEL signal becomes true ~~and~~ , within a bus clear delay after the SEL signal becomes true. Any losing SCSI devices may return to step (a).

NOTE 23 - Step (d) above requires any device that begins NORMAL ARBITRATION phase to complete the NORMAL ARBITRATION phase to the point of SEL being asserted if it begins the NORMAL ARBITRATION phase as stated in step (c). This precludes the possibility of the bus being hung.

d.5) After the bus free delay in step (b), SCSI devices with arbitration fairness enabled which are not arbitrating shall wait a bus set delay and start sampling the DATA BUS to determine which SCSI devices attempted arbitration, which SCSI device won, and which SCSI devices lost. This sampling shall continue for an arbitration delay after the bus free delay in step (b). A QAS device shall update its fairness register with all device IDs that lost arbitration. A non-QAS device shall update its fairness register with all lower-priority device IDs that lost arbitration.

NOTE xx: For ease of implementation, this sampling may begin when BSY is true following BUS FREE and end when SEL is true.

e) The SCSI device that wins arbitration shall wait at least a bus clear delay plus a bus settle delay after asserting the SEL signal before changing any signals.

The SCSI ID bit is a single bit on the DATA BUS that corresponds to the SCSI device's unique SCSI address. All other DATA BUS bits shall be released by the SCSI device. During the NORMAL ARBITRATION phase, DB(P_CRCA), and DB(P1) (if present) may be released or asserted, but shall not be actively driven false.

10.2.2 QAS protocol

QAS allows a target that has information unit transfers enabled and QAS enabled that is currently connected to an initiator that has information unit transfers enabled and QAS enabled to transfer control of the bus to another SCSI device that has information unit transfers enabled and QAS enabled without an intervening BUS FREE phase. SCSI devices that support QAS shall report that capability in the INQUIRY command.

An initiator that supports QAS ~~shall~~ may negotiate the use of the QAS phase with each target that has indicated support of QAS any time the data transfer agreement is in an indeterminate state, using the PPR message~~, in order to enable QAS~~. SCSI devices that support QAS shall implement the fairness algorithm (see Annex B) during all QAS ~~and normal~~ arbitrations. SCSI devices shall negotiate the use of QAS with a particular SCSI device before using QAS to select or reselect that SCSI device. Also, targets shall have negotiated the use of QAS with a particular initiator before using QAS REQUEST message to do a physical disconnect from that initiator, and initiators shall have negotiated the use of QAS with a particular target before accepting a QAS REQUEST message from that target. If an initiator receives a QAS REQUEST message from a target that has not negotiated the use of QAS, then the initiator shall create an attention condition for the QAS REQUEST message, and shall report MESSAGE REJECT on the following MESSAGE OUT phase.
[don't require QAS if endpoints support it – software may choose not to use it to keep the system all non-QAS, or an expander in the domain might not support it.]

~~In an environment where some SCSI devices have QAS enabled and other SCSI devices do not, it is possible for the SCSI devices that have QAS enabled to prevent SCSI devices that do not have QAS enabled from arbitrating for the bus. This occurs when SCSI devices that have QAS enabled never go to a BUS FREE phase.~~

~~To prevent this from occurring targets with QAS enabled, when doing a physical disconnect following an initial connection, should do a physical disconnect using a BUS FREE phase if their fairness register is empty.~~

~~If a target has QAS enabled and the fairness register is not empty or if the connection is not the initial connection (i.e., a physical reconnection) the target should use QAS to arbitrate for the bus.~~

~~In a mixed environment SCSI devices that do not have QAS enabled should disable their fairness algorithm, and those SCSI devices should be assigned a higher priority SCSI ID than SCSI devices with QAS enabled.~~
[old fairness scheme removed]

10.2.2.1 QAS phase

The procedure for a target to indicate it wants to release the bus is as follows:

a) The target shall change to a MESSAGE IN phase and issue a single QAS REQUEST (55h)

message.

b) After the initiator negates the ACK signal for the QAS REQUEST message ~~and~~, if the initiator did not create an attention condition ~~then~~, the initiator shall release all SCSI signals within two system deskew delays after detecting MSG, C/D, and I/O signals false.

c) After detection of the last ACK signal being true ~~and~~, if there is no attention condition~~,~~ the target shall negate REQ. After detection of the last ACK signal being false, the target shall release all SCSI signals except the BSY, MSG, C/D, I/O and REQ signals. Then the target shall negate the MSG, C/D, and I/O signals within two system deskew delays. The target shall wait two system deskew delays after negating the C/D, I/O, and MSG signals before releasing the REQ signal.

d) If the target detects the SEL signal being true, the target shall release the BSY, MSG, C/D, and I/O signals within one QAS release delay.

e) After waiting at least a QAS arbitration delay from releasing the SCSI signals in step (c), if there are no SCSI ID bits true the target shall transition to the BUS FREE phase.

f) After waiting at least a QAS arbitration delay from releasing ~~of~~ the SCSI signals in step (c), if there are any SCSI ID bits true the target shall wait at least a second QAS arbitration delay. If the SEL signal is not true by the end of the second QAS arbitration delay the target shall transition to the BUS FREE phase.

NOTE 25 - The release of MSG, C/D, and I/O may cause release glitches; Step (f) above ensures these glitches occur at a time when no connection is established on the bus so that they do not interfere with proper operation.

The procedure for a ~~SCSI~~ QAS device to obtain control of the SCSI bus is as follows:

a) The ~~SCSI~~ QAS device shall first wait for MESSAGE IN phase to occur with a single QAS REQUEST (55h) message. When the ~~SCSI~~ QAS device detects the ACK signal being false for the QAS REQUEST message and the attention condition ~~is~~ being cleared it shall begin the QAS phase ~~if allowed under the fairness algorithm~~.

b) The ~~SCSI~~ QAS device shall wait a minimum of ~~a~~ two deskew delays after detection of the MSG, C/D, and I/O signals being false before driving any signal.

c) Following the delay in step (b), the ~~SCSI~~ QAS device may arbitrate for the SCSI bus by asserting its own SCSI ID within a QAS assertion delay from detection of the MSG, C/D, and I/O signals being false. If arbitration fairness is enabled, the QAS device shall not arbitrate until its fairness register is cleared.

d) After waiting at least a QAS arbitration delay (measured from the detection of the MSG, C/D, and I/O signals being negated) the ~~SCSI~~ QAS device shall examine the DATA BUS.

A) If no higher priority SCSI ID bit is true on the DATA BUS and the fairness algorithm allowed the ~~SCSI~~ QAS device to participate, then the ~~SCSI~~ QAS device has won the arbitration and it shall assert the SEL signal.

B) If a higher priority SCSI ID bit is true on the DATA BUS (see table 27 for the SCSI ID arbitration priorities) or the fairness algorithm prevented the ~~SCSI~~ QAS device from participating in QAS arbitration, then the ~~SCSI~~ QAS device has lost the arbitration.

C) Any ~~SCSI~~ QAS device other than the winner has lost the arbitration and shall release its SCSI ID bit after two deskew delays and within one QAS release delay after detection of the SEL signal being true. A ~~SCSI~~ QAS device that loses arbitration may return to step (a).

e) The ~~SCSI~~ QAS device that wins arbitration shall wait at least a QAS arbitration delay after asserting the SEL signal before changing any signals.

f) After the QAS arbitration delay in step (d), QAS devices with arbitration fairness enabled which are not arbitrating shall start sampling the DATA BUS to determine which QAS devices are attempting arbitration, which QAS device won, and which QAS devices lost. This sampling shall continue for a bus settle delay plus two system deskew delays. The devices shall update their fairness register with all device IDs that lost in arbitration.

The SCSI ID bit is a single bit on the DATA BUS that corresponds to the SCSI device's unique SCSI address. All other DATA BUS bits shall be released by the SCSI device. The DB(P_CRCA) and DB(P1) are not valid during the QAS phase. During the QAS phase, DB(P_CRCA), and DB(P1) may be released or asserted, but shall not be actively driven false.

10.2.3 QAS device arbitration fairness to non-QAS devices

To ensure QAS devices do not starve non-QAS devices, a target using quick arbitration shall drop to BUS FREE instead of sending a QAS REQUEST message if it has sent a QAS_REQUEST message QAS LIMIT times (see 18.1.1) since the last BUS FREE. Upon entering BUS FREE, all QAS targets shall set all the bits in their fairness registers (except their own). This restricts the first normal arbitration to non-QAS devices. While in normal arbitration, the QAS targets shall set fairness register bits for both lower and higher-priority devices which lose arbitration until their fairness registers are cleared. This ensures that each non-QAS device, even if higher priority than the QAS devices but lower priority than another non-QAS device, gets a chance to win the bus. Once their fairness registers clear, the QAS devices can participate in normal arbitration.

Annex B: SCSI Bus Fairness

Since Annex B is planned to be rewritten, no specific changes are proposed here. Text from section 10.2.3 could be copied to B.2.3, maybe with additional explanatory material.