LSI LOGIC®

4420 ArrowsWest Drive
Colorado Springs, CO 80907

October 30, 1998

To: T10 Technical Committee
From: John Lohmeyer, LSI Logic Principal Member of T10
Subj: Proposed Domain Validation Annex for SPI-3

The proposals received by T10 so far on domain validation have only addressed enhancements to the READ BUFFER and WRITE BUFFER commands in the SPC-2 draft standard. SPI-3 is completely silent on this important feature. I think this omission needs to be corrected.

I have drafted an informative annex on domain validation. Since it is informative, it contains no actual requirements. However, it does define several terms and it provides some minimal guidance on performing domain validation. Please review this proposed annex.

Domain Validation

(Informative)

1. Introduction

'Domain validation' is the act of verifying that the cable plant is sufficient to support the negotiated speed and width between the initiator and target. The cable plant includes any cables, printed circuit cards, and expanders that are present between the two SCSI devices. For example, two wide devices connected with a narrow cable will fail domain validation testing.

'Fall back' is the act of re-negotiating to a set of physical parameters that are less demanding of the cable plant. After falling back, domain validation is again performed to verify the new parameters. This cycle may be repeated until an acceptable set of physical parameters is found.

While the concepts of domain validation and fall back have always been supported in SCSI protocol, they have recently gained increased attention because higher data rates put increasing demands on the cable plant. It is unacceptable for two SCSI devices to negotiate a set of physical parameters that then fail to achieve reliable operation.

This annex defines some terminology related to domain validation and fall back. It also points to various tools available to support these concepts. Key tools are the recent additions to SPC-2 to enhance the READ BUFFER and WRITE BUFFER commands to facilitate domain validation testing.

This annex further points to facilities in the Desktop Management Interface to support domain validation and fall back.

2. Domain validation levels

2.1 Simple communication tests

Simple communication tests consist of sending a receiving known data patterns. These tests may detect a number of cable plant problems including:

1) Path width errors (i.e., narrow cable used with wide devices)
2) Down-revision expanders (e.g., expanders not capable of the negotiated data rate)
3) Gross cable errors (e.g., broken wire)
4) Incorrect termination (e.g., missing or bad terminator).

Since most legacy SCSI devices support the READ BUFFER and WRITE BUFFER commands, these commands are used for the simple communication tests. It is important that simple communication test routines insure that no other process is using the target device to avoid unwanted side effects. If there is any possibility that there may be multiple initiators present on the same SCSI bus, the RESERVE command should be used to lock out other initiators while doing the simple communication test.

Some data patterns are more stressful on the physical layer and should be used during the simple communication tests to assist in detecting certain failure modes. In particular, four data patterns have been identified that may be useful in detecting various failures:

1) Counting (0001h, 0203h, 0405h, …)
2) Alternating ones and zeros (0000h, FFFFh, 0000h, FFFFh, …)
3) Crosstalk (5555h, AAAAh, 5555h, AAAAh, …)
4) Walking ones and zeros

These data patterns are defined in greater detail in 98-228r0.

2.2 Signal margin tests

Margin testing verifies that the agreed upon physical parameters have some degree of margin. That is, known data patterns are transferred with slightly 'weakened' signals to verify that no errors occur on the transfers. The presumption is that if no errors occur with the 'weakened' signals, then transfers with normal signals should have sufficient margins to insure reliable operation. Should errors occur with the weakened signals, then the initiator should fall back to a lower transmission speed.

There are several techniques under investigation for weakening signals. The most obvious technique is to reduce the driver strength (slightly lower current or voltage level). Other techniques are also possible such as moving the timing window of the data signals with respect to the REQ or ACK signal. This would reduce the setup or hold time available to the receiver.

At the time of this writing, it has not been determined which, if any, of the driver margining techniques work best. Consequently, any driver margining technique is permitted.

While it is possible to achieve a similar effect by adjusting receiver sensitivity, there are some practical issues with allowing this option; only the driver or receiver should be adjusted at a given time. Adding the option to margin the receiver would necessitate still more controls to specify whether the driver or receiver is to be adjusted. To avoid this complexity, it has been agreed to margin the drivers and not the receivers.

Margin testing is normally done using the WRITE BUFFER and READ BUFFER commands (see SPC-2). For each target, the initiator sends a series of WRITE BUFFER commands interleaved with READ BUFFER commands. After each READ BUFFER command the received data is compared to the original data to verify successful transfer.

The margin testing is divided into two phases. During the first phase, the initiator's signals are weakened when performing the WRITE BUFFER command, while the target's signals are sent normally. During the second phase, the target's signals are weakened when performing the READ BUFFER command, while the initiator's signals are sent normally.

The WRITE BUFFER and READ BUFFER commands include a margin adjustment field to specify the vendor-specific margin setting used for each transfer. Legacy devices do not support this field, so it is not possible to do complete signal-margining tests with legacy devices. Nonetheless, if either device supports this field, it is possible to do partial testing (i.e., inbound or outbound testing).

2.3 Signal calibration

Signal calibration adjusts signal line drivers (and possibly receivers) to improve signal transmission characteristics during normal data transfers. Parameters that may be adjusted include driver strength, receiver sensitivity, and timing parameters such as setup time and hold time.

This annex does not address signal calibration beyond defining the term.

3. Fall back

Fall back is the act of re-negotiating to a lower (or less-demanding) set of physical parameters (e.g., Fast-40 instead of Fast-80). It is accomplished either by a new PPR negotiation (assuming both devices support the PPR message) or by a new WDTR/SDTR negotiation for older devices.

The fall back procedure can introduce some complexities since message-based negotiations are often handled at a lower level of the operating system than the portion that deals with command delivery. Thus the routine that performs domain validation testing may need a service from a lower layer that was previously not provided. Extensions to several operating system interfaces are being proposed to support fall back. These include the Desktop Management Interface (DMI), Advanced SCSI Programming Interface (ASPI), and others.

4. System considerations

The READ BUFFER and WRITE BUFFER commands were first defined in SCSI-2. Devices implemented prior to SCSI-2 should report CHECK CONDITION status and ILLEGAL REQUEST sense key in response to attempts to issue these commands. It may be impractical to perform communications tests with these devices. This is not a significant issue since very few of the devices are still in service and these devices only support lower transfer rates that should not stress the cable plant.

4.1 Buffer Corruption

SCSI-2 defined the READ BUFFER and WRITE BUFFER commands as accessing physical buffers in the target. Many implementations do not protect the buffer contents if there is an intervening command from any other process. Therefore, domain validation software should ensure that no processes are active while performing tests.

The RESERVE command may be useful in blocking commands from other initiators. However, using the RESERVE command is not sufficient to prevent commands from the same initiator

(possibly issued by other processes) from corrupting the buffer contents. Also, targets with multiple logical units may corrupt the buffer if commands are processed on other logical units.

A proposal for SPC-2 (98-184r2) defines an enhancement to the READ BUFFER and WRITE BUFFER commands that helps with this problem. The proposal defines a 'logical buffer' that is protected from corruption by other commands. This logical buffer may be especially valuable when performing tests during normal operation of the system.

4.2 System Hangs

Domain validation testing may cause several kinds of error conditions on cable plants that are inadequate for the negotiated set of physical parameters:

a) Parity or CRC errors
b) Data comparison mismatches
c) Bus hangs.

These error conditions usually result in falling back to a less demanding set of physical parameters. The first error condition is a detected error. The second error condition is an undetected error. The third error condition may require special handling as described in the next paragraph.

Bus hangs occur when the target fails to detect an ACK pulse from the initiator. This is a frequent failure mode on marginal cable plants. It is recommended that initiators include provisions to avoid extended bus hangs. Two recovery actions are possible when an initiator detects an extended delay for the target to switch phases at the end of a data phase while performing domain validation testing:

a) Assert the RST signal

b) Send additional ACK pulses.

The exact time of an 'extended delay' is vendor specific and is not defined by this standard.