To: T10 working group 00-199r2
From: Bill Ham / Robert Elliott, Compaq
Date: May 15, 2000
Subject: Proposal for expander requirements, domain validation basic integrity check, and related target requirements

Table of contents

1. Introduction

This document contains a proposal for technical content to be added to SPI-4 for the purposes of ensuring that simple expanders (per the definitions in EPI) have adequate information to operate effectively in SCSI domains containing SPI-4 segments. Much of the material is directly copied from EPI, some is by reference to other sections of SPI-4 and some is new original material.

In order to define the requirements for expanders the definitions for segments and other related terms and concepts are proposed for adoption in SPI-4.

2. New content for SPI-4 relating to expanders

Extended configurations are described that incorporate the ability to separate the SCSI domain into electrically isolated parts by using active circuits. The requirements for these active circuits are described in clauses 3 through 6. These circuits are specified to produce almost no additional requirements beyond that for use in ordinary configurations for initiators and targets. Initiators and targets shall not produce any behavior that causes extended

configurations defined in this section to fail provided that the extended configuration meets the requirements in Clauses 3 through 6.

Initiators shall implement the basic integrity check for domain validation presently described in Annex L in SPI-3.

There are no known additional requirements for targets in the extended configurations.

3. Definitions and concepts:

3.1 Definitions:

bus-path: The electrical path directly between the bus terminators.

byte: Indicates an 8-bit construct.

contact: The electrically-conductive portion of a connector associated with a single conductor in a cable.

differential: A signaling alternative that employs differential drivers and receivers.

initiator: An SCSI device containing application clients that originate device service and task management requests to be processed by a target SCSI device. See the SCSI-3 Architecture Model standard for a detailed definition of an initiator.

logical unit: An externally addressable entity within a target that implements an SCSI device model. See the SCSI-3 Architecture Model standard for a detailed definition of a logical unit.

path: The cable, printed circuit board or other means for providing the conductors and insulators that connect two or more points.

SCSI bus: consists of all the conductors and connectors required to attain signal line continuity between every driver, receiver, and terminator for each signal.

stub: Any electrical path connected to the bus that is not part of the bus-path.

target: A SCSI device that receives SCSI commands and directs such commands to one or more logical units.

3.2 SCSI domain related terminology and concepts

SCSI domain:

A SCSI domain is a logical bus with at least one bus segment, at least one initiator, and at least one target. Domains with multiple bus segments are enabled through the use of bus

expanders. Domains consist of the set of SCSI devices that are addressable from an initiator or target.

Simple expander:

Devices that couple bus segments together without using SCSI ID's in the device are called simple bus expanders.

Bridging expander:

Devices that couple a bus segment to another SCSI segment or another kind of port by using addressable SCSI ports in the device.

4. Bus segments in a SCSI domain

Previous SCSI standards define parameters for SCSI busses based on the assumption that there is a single electrically conducting path between bus terminators for each signal and that a SCSI domain contains all the devices between these two terminators. This electrical path is assumed to pass signals in both directions without delay other than that caused by the propagation delay of the transmission line associated with the path. It is assumed that there are no intervening active components in the path between the bus terminators.

A more general concept recognizes that it is possible to build SCSI domains that use more complex physical implementations where there may be active electrical components between SCSI devices.

A building block for these more complex implementations is the bus segment which is defined as two bus terminators and the associated single electrically conducting path between these terminators (for each signal) that satisfies the assumptions in the first paragraph of this clause. Multiple bus segments may be functionally connected together by special coupling circuits described in clause 5.

Each bus segment has TERMPWR sources and TERMPWR distribution parameters.

Bus segments always use the same transmission type (LVD or SE) within the segment. A domain may contain segments that use different transmission types.

Bus segments follow the same rules individually that are described in the existing standards (with important exceptions). Using multiple bus segments with coupling circuits in the same domain allows much more of the full properties supported by the SCSI bus protocol to be realized than when using single segment domains.

Some of the salient properties impacted are device count limits, physical length limits, ground voltage shifts, dynamic removal and replacement of portions of domains, and mixing of device types (SE, LVD).

5. Simple bus expanders

Bus expanders are elements used for connecting segments together. There are two basic types: simple and bridging. Simple expanders do not occupy a SCSI ID and are intended to be "invisible" to initiators and targets. Bridging expanders have SCSI IDs on all ports, participate in SCSI arbitration and messaging, and are “devices” in the SCSI sense.

The following features shall be the required properties of simple bus expanders:

- No SCSI Ids used on simple expanders
- No arbitrations initiated by the simple expander
- No messages originating with the simple expander shall be sent by the simple expander (messages sent from initiators and targets could be read if desired – for example the simple expander could need to know the negotiated data phase speed or some other variable property of a transaction of an initiator / target connection)
- Retransmitted signals from the simple expander transmitter shall meet the same requirements as any other SCSI device at the expander boundary (which may or may not be at a separable connector)
- Simple expander receivers shall operate error free with the most degraded signal allowed for any SCSI device receiver at the expander boundary (which may or may not be at a separable connector)
- Simple expanders shall not interfere with the REQ/ACK offset count in any initiator or target in the domain (other than that caused by the propagation time through the expander)
- Expanders that are powered on shall retransmit RESET assertions from one segment to the other regardless of the state of any other SCSI signals on either side (This allows the domain to be reset due to catastrophic events on one side that could lock up the expander.)
- Simple expanders shall operate with any arbitrary placement of the initiators and targets with respect to the simple expander (for example all targets and initiators could be on the same side of the expander or there could be initiators and targets on both sides of the expander)
- TERMPWR shall not be connected in the expander between the segments being coupled
- DIFFSENS shall not be electrically or logically connected between segments being coupled
- Transmission mode (SE/LVD, etc.) changes on one segment shall cause the simple expander to issue a SCSI bus RESET on the other side

It is very desirable that simple expanders consume minimal propagation time during arbitration so that the end to end domain propagation time budget may be used primarily for longer physical length connections. See clause 5.4.3.

Figure 1 shows a single simple expander between two bus segments.

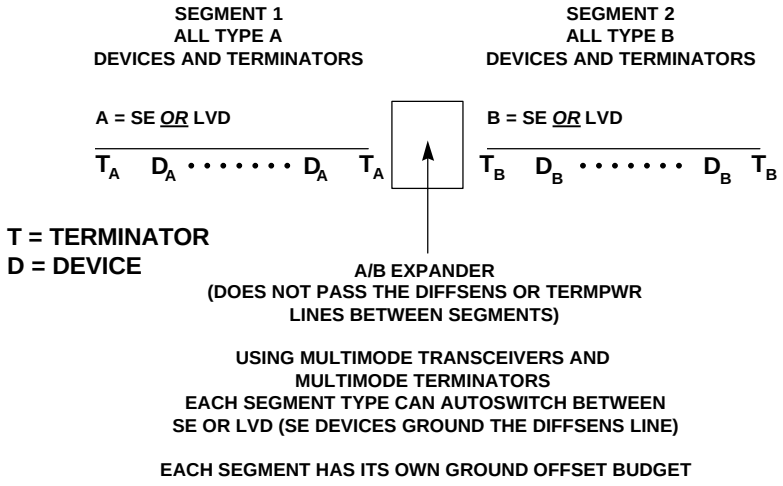

Figure 1 - A two segment domain using a single expander circuit

Figure 2 shows three ways that simple expanders may be used to connect bus segments.

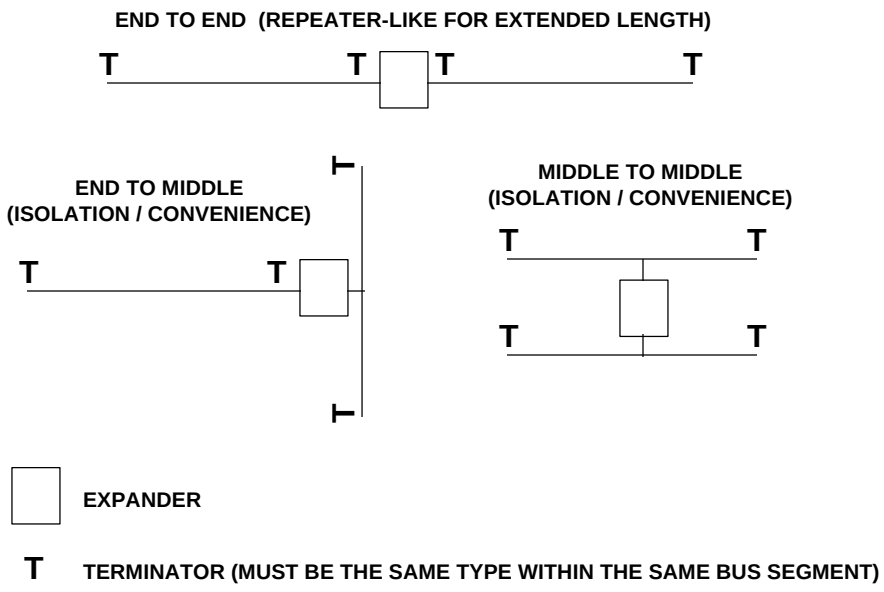

Figure 2 - Three ways to couple bus segments together with expanders

5.1 Homogeneous type

If an expander has the same type of segment on both sides it is termed a homogeneous expander. The homogeneous expander does not do type conversion (e.g. SE to LVD).

This kind of expander may be useful in existing systems where domain length increases may be achieved by inserting a single homogeneous expander in the right place. Such a condition exists, for example, in a domain where several SE devices are connected to a backplane and subsequently to a host adapter by a shielded external cable. By placing a homogeneous expander near the backplane one creates a short, heavily loaded backplane segment and a point to point segment to the host adapter. Increases in overall domain physical length may be achieved because the point to point segment length limit is longer than the multi-drop segment length.

5.2 Heterogeneous types

Expanders that have different bus transmission types on each side are heterogeneous expanders.

5.3 Domain examples using simple expanders

Figure 3 shows two examples of domains built using only simple expanders.

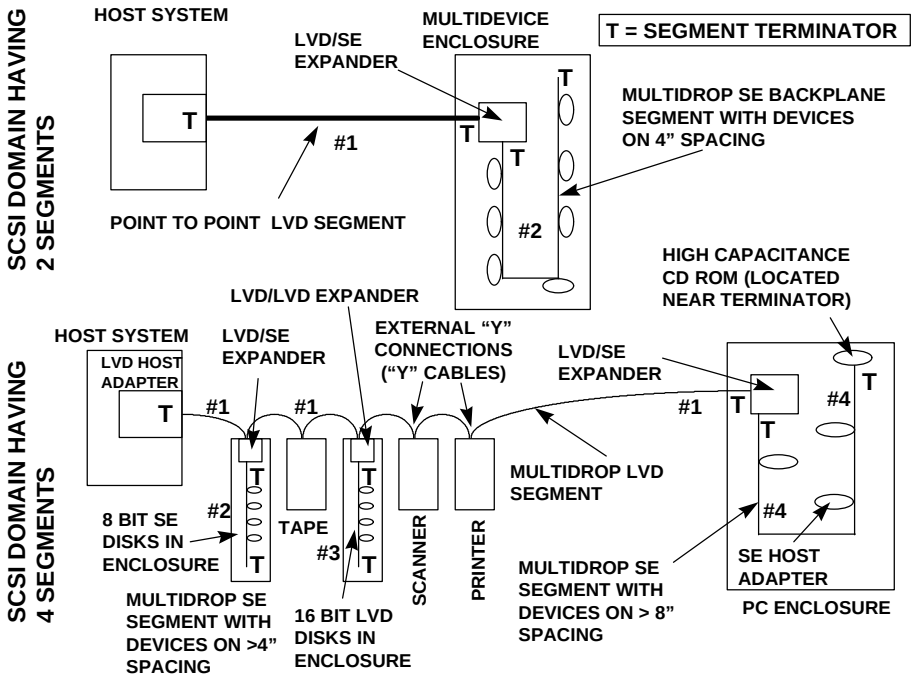

Figure 3 - Examples of domains using simple expanders

5.4 General rules for SCSI domains using simple expanders

The rules are summarized in 5.4.1 followed by detailed discussion for each in the subsequent clauses.

5.4.1 Rule summary

Valid SCSI domains shall follow these six rules:

1. All bus segments in the domain shall comply with their individual bus segment length limits and other segment related requirements.

2. Any segment between two other segments shall support the highest performance level that can be negotiated between the two other segments. For example, two wide LVD Fast-40 segments shall not be separated by a segment that does not support both wide and Fast-40. See Figure 4 for examples.

3. The expander between two segments shall support the maximum performance levels supported on each SCSI interface of the expander.

4. The maximum propagation time between any two devices in the domain shall not exceed 400 ns.

5. The number of addressable devices in the domain shall not exceed the addressability of the devices in the domain.

6. Loops topologies are not allowed.

In addition to these six rules which affect the basic operation of the domain there is a special performance consideration for the REQ/ACK offset level supported when using simple expanders.

Since the round trip domain propagation time can be as large as 800 ns when using simple expanders, the REQ/ACK offset negotiated between any two devices should be larger than used for domains that do not use expanders. If the offset and buffering is not sufficient to accommodate the round trip time between the devices the domain will experience a performance degradation. This minimum offset level increases with increasing data phase rate. The minimum offset levels to avoid performance degradation for a variety of conditions are shown in Table 2.

5.4.2 Rule 1

Requirements for domains consisting of a single SE segment or a single LVD segment are specified in detail in other clauses of this document. Every segment in a multi-segment domain shall conform to the requirements for single segment domains of the same transmission mode type.

5.4.3 Rule 2

Rule 2 relates to intermediate segments (which only exist in domains of at least three segments). The segment between the two other segments is the intermediate segment. The intermediate segment shall be wide if both other segments are wide. The intermediate segment shall support the lowest common speed between the other segments.

An example of a rule 2 configuration violation is shown in Figure 4.

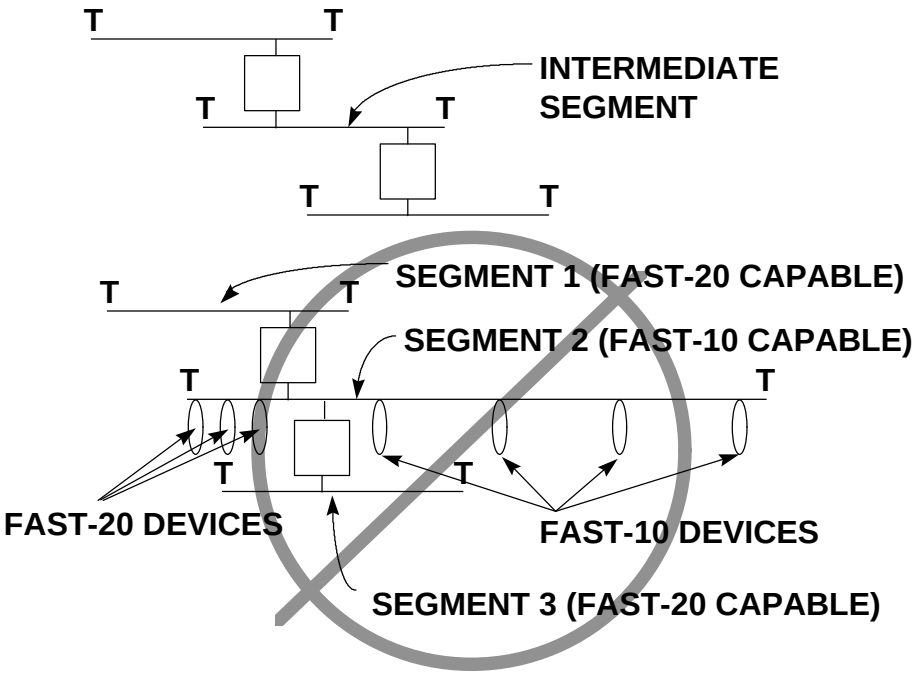

Figure 4 - Intermediate segments and performance ranking

The configuration in Figure 4 is valid only if the data phase rate is limited to Fast-10 for any data phase transactions between segment 1 and segment 2, segment 2 and segment 3 or segment 1 and segment 3. Even though the Fast-20 devices in segment 2 are located close to the expanders and the distance between the expanders is small, the segment length is defined by the distance between the terminators - not by the distance to the expander connection or to the devices. The intermediate segment is not Fast-20 capable and may not be used for Fast-20 transactions between segment 1 and segment 3. Segment 2 is also not to be used for Fast-20 transactions within segment 2. Fast-20 transactions are allowed between devices in segment 1 or between devices in segment 3.

The intermediate segment in this example will see signals at the higher data rates on the DATA and parity lines but since the devices in the intermediate segment are not participating in the higher data rate transmission and are waiting for the next BUS FREE, RESET or other general SCSI phase they are unaffected by the higher speeds.

For multimode segments, any dynamic change of transmission mode (LVD to SE etc.) is treated as a fault and the expander shall assert the RESET line on the segment opposite the one that experienced the transmission mode change. The expander shall detect this state change by sensing the DIFFSENS line. This scheme ensures that the initiators on the other segments are aware of the change in transmission mode and can reassess whether this mode change is consistent with the performance requirements for the segments and the overall parameters for the domain before allowing traffic to resume. Once RESET is asserted initiators shall renegotiate with all targets in the domain.

5.4.4 Rule 3

Homogeneous expanders between two segments shall support the maximum performance levels supported on each SCSI interface of the expander. If one SCSI interface of the expander is wide both SCSI interfaces of the expander shall be wide. Both SCSI interfaces of the expander shall support the same maximum data phase rate. This rule means that the expander itself shall not be the weak link in the domain.

5.4.5 Rule 4

5.4.5.1 Effects of wired-or glitches

Wired-or glitches occur when two on more drivers are asserting the same signal line and one subsequently ceases to drive the line. This condition happens frequently during arbitration on the BSY signal and may happen on other wired-or signals. This change in the number of asserted drivers causes a redistribution of current in the segment (with resulting voltage glitches) and may cause false detection of BUS FREE and other errors. The worst case condition is when two devices near a segment terminator are involved. In this case it requires a full segment length round trip time before the line is again stable (after the device stopped asserting the line). If this condition applies, the round trip time allowed is 400 ns. The one way time is 200 ns.

Waiting the entire domain round trip time may be avoided by ensuring that wired-or glitches do not pass through the expander. This standard does not describe how expanders implement this capability but it is within the presently available technical art. If wired-or glitches are not propagated through the expanders, then the maximum round trip domain signal propagation time is 800 ns and the one way domain propagation time is 400 ns.

If a simple expander does not implement the wired-or glitch filter it shall be labeled indicating that it allows propagation of wired-or glitches. This standard assumes that wired-or glitch blocking expanders are used and that the maximum domain round trip time of 800 ns is available.

5.4.5.2 Expander propagation delay effects

The expander is said to be in series with initiators and/or targets when the path between the initiators and/or targets goes through an expander. In this case the propagation delay through the expander shall be counted as part of the 400 ns budget between those devices.

The delay varies depending on the implementations. Care shall be exercised when considering expanders to understand the capabilities of the expanders being used. When two expanders are in series the delay across the pair may be much less than twice the individual delays. This is because the “direction” change that consumes much of the propagation delay during arbitration will only apply to one of the expanders at a

time. The single expander delay, Tds and the expander series pair delay, Tdp should be specified.

If the expander is attached to a segment (as in case of the device enclosures in the bottom part of Figure 3) it is only in series between the devices in the enclosure and other devices in the domain. The expander in the enclosure would not be in series between the two host ports for example.

The propagation time through the differential transceivers of initiators and targets does not need to be separately accounted for if the wired-or glitches cannot propagate through the expander. The differential transceiver delay effects are confined to the differential segments. Using expanders that do not pass the wired-or glitch prevents one segment's delays from being passed on to the next.

5.4.5.3 Sample calculations

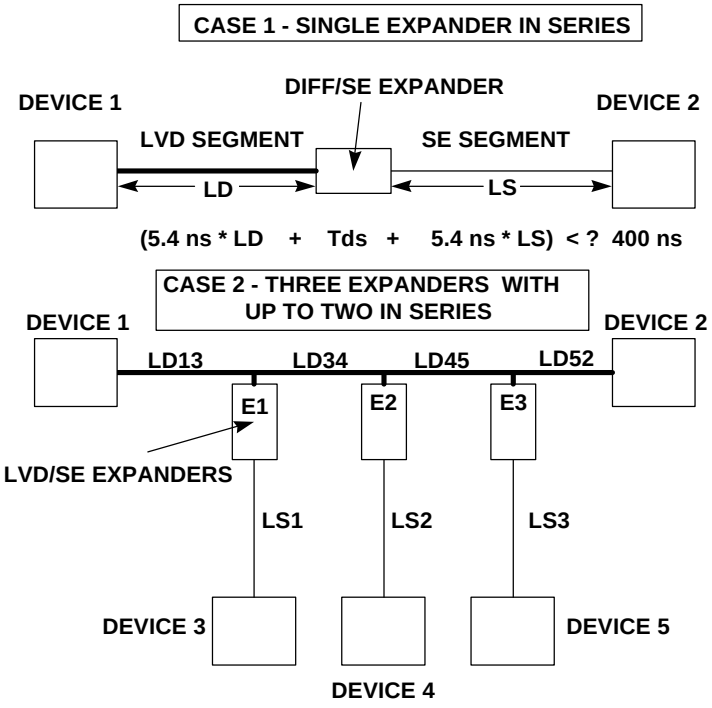

Figure 5 - Two configurations for domain delay calculations

Figure 5 shows two sample SCSI domain configurations. In Figure 5 parameters whose first letter is "L" are physical lengths, "D" refers to differential segments and "S" refers to single ended segments. In case 1 the delay calculations are shown in the figure. For the more complex case 2 one shall consider all the possible combinations between any two devices. These calculations are shown in Table 1. The device pair that has the largest combination of expander propagation time and interconnect propagation time determines if this configuration meets the 400 ns device to device maximum propagation time requirement.

While this may appear complex, the limiting cases may be obvious without the rigorous analysis.

Table 1 - Domain delay calculations

DEVICE PAIR	PATH BETWEEN DEVICES	EXPANDERS DELAY (ns)	INTERCONNECT DELAY (ns)
1-2	LD13,LD34,LD45,LD52	0	5.4*(LD13+LD34+LD45+LD52)
1-3	LD13,E1,LS1	Tds	5.4*(LD13+LS1)
1-4	LD13,LD34,E2,LS2	Tds	5.4*(LD13+LD34+LS2)
1-5	LD13,LD34,LD45,E3,LS3	Tds	5.4*(LD13+LD34+LD45+LS3)
2-3	LD52,LD45,LD34,E1,LS1	Tds	5.4*(LD52+LD45+LD34+LS1)
2-4	LD52,LD45,E2,LS2	Tds	5.4*(LD52+LD45+LS2)
2-5	LD52,E3,LS3	Tds	5.4*(LD52+LS3)
3-4	LS1,E1,LD34,E2,LS2	Tdp	5.4*(LS1+LD34+LS2)
3-5	LS1,E1,LD34,LD45,E3,LS3	Tdp	5.4*(LS1+LD34+LD45+LS3)
4-5	LS2,E2,LD45,E3,LS3	Tdp	5.4*(LS2+LD45+LS3)

5.4.6 Rule 5

Since simple expanders have no SCSI ID’s the maximum number of addressable devices in the domain is not increased or decreased by the use of simple expanders.

5.4.7 Rule 6

Loop topologies in any form are not allowed within a domain. Using expanders connected in a loop it is possible to create conditions where both an expander and a target or initiator are asserting the same line. Under these conditions the line will not return to the negated state when the initiator or target releases the line since it will continue to be driven by the expander. The logic state of the line will not change and a lock up condition exists.

Figure 6 shows some examples of loops. Even if it appears that no deadlock condition is possible (in some symmetrical configurations for example) loops are still not allowed because the propagation time variability between components guarantees asymmetry and non-zero deadlock risk.

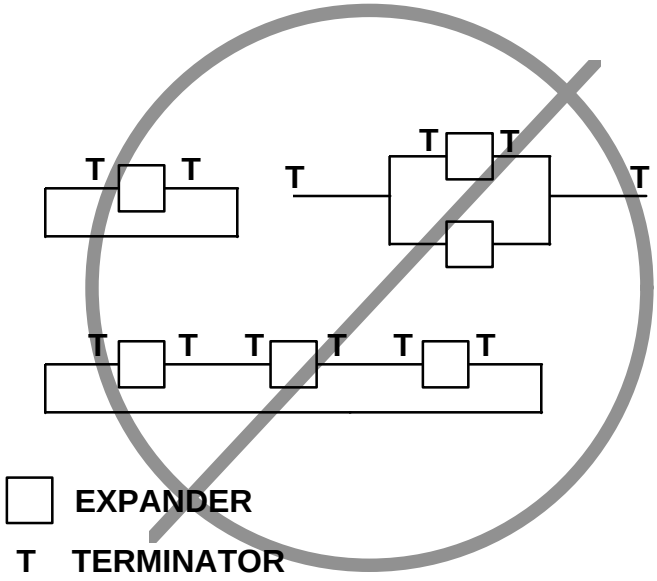

Figure 6 - Examples of illegal loops

5.4.8 Special performance considerations for domains with simple expanders

The REQ/ACK offset is the difference between the number of REQ pulses sent (received) and the number of ACK pulses received (sent) in a synchronous data phase transmission. This offset allows multiple transfers to be on the domain media at the same time.

The device REQ/ACK offset counter is set to zero before the data phase begins. When a REQ is sent or received the offset counter is incremented. When an ACK is sent or received the counter is decremented. After the data phase is completed the offset counter should again be at zero since the number of REQs issued and the number of ACKs received should be the same.

When the target sends the first REQ pulse there is a minimum of one round trip time before the first ACK pulse can be received from the initiator. This round trip time includes the data processing time at the initiator. The target may continue to issue REQ pulses until the offset counter in the target reaches the maximum REQ/ACK offset level that was negotiated.

If the maximum offset level in the target is reached, the target waits until it receives a decrementing ACK pulse before issuing another REQ pulse. The time spent waiting for a decrementing ACK pulse is lost.

The receiving device is required to accept up to at least the negotiated REQ/ACK offset level of data phase transfers in its buffers.

Negotiated REQ/ACK offsets do not affect the operation of simple expanders.

The minimum desirable offset value is given by:

[{2x one way domain propagation time} / {ACK (REQ) period}] + processing overhead

Table 2 gives some representative values from the above expression for round trip domain propagation times greater than 400 ns assuming the processing overhead to be 2 ACK(REQ) periods in all cases.

Table 2 - Minimum REQ/ACK offset for maximum performance

Domain round trip propagation time (ns)	Data phase speed	ACK(REQ) period (nominal min)	Minimum REQ/ACK offset to avoid performance degradation (assuming 2 overhead periods in all cases)
500	Fast-10	100	7
600	Fast-10	100	8
700	Fast-10	100	9
800	Fast-10	100	10
500	Fast-20	50	12
600	Fast-20	50	14
700	Fast-20	50	16
800	Fast-20	50	18
500	Fast-40	25	22
600	Fast-40	25	26
700	Fast-40	25	30
800	Fast-40	25	34
500	Fast-80	12.5	42
600	Fast-80	12.5	50
700	Fast-80	12.5	58
800	Fast-80	12.5	66
500	Fast-160	6.25	(82) 84*
600	Fast-160	6.25	(98) 100*
700	Fast-160	6.25	(114) 116*
800	Fast-160	6.25	(130) 132*
* rounded up to the next multiple of 4 - see xxxxx			