William E. Ham

Recent Advances in Basic Physical Technology for Parallel SCSI: UltraSCSI, Expanders, Interconnect, and Hot Plugging

DIGITAL uses SCSI technology in most of its storage products and consequently has led major standards and industry bodies to improve the technology in the following areas: increased synchronous data phase speed beyond fast SCSI; longer, more complex electrical configurations by means of expander circuits; versatile and more manageable connectivity through a smaller, improved physical interconnect; and dynamic device insertion and removal. Data phase transmission rate extension is achieved through understanding and controlling silicon chip timing and transmission media parameters. Using expander devices to confine transmission line effects to shorter segments allows large increases in the maximum distance between devices and in the device population within the same SCSI domain. Expanders enable complex, hublike configurations to be created without changing existing SCSI devices or software. The use of 0.8-millimeter connector technology and consideration of cable losses has reduced the physical size of the external shielded interconnect by approximately two thirds, decreased the number of parts required to support complex configurations by a factor of 10, and increased the interconnect density to the same level used in serial SCSI. Finally, the mating and demating events that occur during device insertion and removal produce a spectrum of small, undetectable, electrical disturbances on the active bus that appear to be limited by the physics of the media and device capacitance.

Introduction

Parallel Small Computer System Interface (SCSI) is the workhorse technology for most of the storage applications in DIGITAL products today. This device and interconnect technology spans all system offerings from the simplest to the most complex. SCSI was introduced to the higher-end products in the early 1990s as the open systems follow-on to the DIGITAL proprietary Digital Storage System Interconnect (DSSI) and Computer Interconnect (CI) technologies.

As system demands have increased, SCSI has evolved to meet the needs. DIGITAL has made considerable contributions to the technology and led the effort to achieve industry standardization. This paper details the most significant developments in the physical features of parallel SCSI technology over the last several years that have allowed it to continue to serve DIGITAL customers in an effective, competitive way. The discussion targets the following four important areas:

1. Speed increases in the synchronous data phase, which resulted in the ANSI definition of UltraSCSI (Fast-20 SCSI) technology[1]
2. Development of software-invisible circuits, generally called expanders, that enable segmentation of SCSI domains into easily managed pieces
3. New connector and cable technology, namely the Very High Density Cabled Interconnect (VHDCI) device, that decreases the interconnect size and complexity by many fold[2]
4. Dynamic removal and replacement of devices on an active bus, which is referred to as hot plugging

DIGITAL made substantial contributions in the four areas. This work included creating the expander and interconnect standards projects; leading the working groups that defined the Fast-20, expander, and interconnect standards; providing data for the Fast-20 and hot-plugging projects; and proposing and gaining approval for the hot-plugging standard.

The author has taken a phenomenological approach throughout, because in most cases there are too many unknowns to achieve a rigorous analytical result. This

paper focuses on developments from SCSI-2 through UltraSCSI and specifically does not address the new Low Voltage Differential (LVD) technology being introduced for the highest-speed applications.

Pedigree

SCSI is defined in several ANSI standards[1,3,4] and in the material that was developed to create these standards.[5,6] The standards were generated over the last decade through a cooperative effort of approximately 60 major companies in the computer and computer support industry. As a result of this pedigree, the prime directive for SCSI technology is interoperability of devices designed and manufactured by different companies.

The details of the physical designs used to implement SCSI may not be visible to users and researchers; these details contain much of the marketing and technical differentiation between the products of the participating companies and are therefore hidden in the silicon design. The behavior at the device connector pervades the SCSI specifications. The basic assumption is that as long as the properties are compatible at these connectors, device substitution is possible. Thus, SCSI devices may be both interoperable and of different designs.

Basic Architecture

This section reviews the basic architecture of parallel SCSI. The SCSI bus is a parallel, multidrop, wired-OR configuration.

Signal Multiplexing and Phases The parallel signal construction of the bus allows multiplexing of some signals during different phases of communication so that the same signal lines may have very different functions in different phases. The physical behavior of signals is usually limited by the phase during which the shortest pulses are used and the demands for signal integrity are the highest. The limiting SCSI phase is the data phase (payload phase) that is executed with the highest synchronous rate. For UltraSCSI, this peak repetition rate is 20 megahertz (MHz). Table 1 contains the generally accepted terminology related to data phase speeds.

Because of the wired-OR property, each signal in the bus must be driven to a known state even if no SCSI device is actually driving the signal. SCSI uses the logical 0 state (negated state) as the undriven state and uses the bus terminators to drive the signal to this state in the absence of any driving devices. The device signal drivers must overcome this terminator-driven logic state of 0 in order to send a logical 1 (asserted state) onto the signal line.

SCSI signals must support all frequencies, from statically driven by the terminators only (DC) to the third harmonic of the fastest signal edge in the synchronous data phase. In many cases, the same wire must support all these frequencies at different times during the SCSI protocol.

The highest signal edge slew rates for UltraSCSI are approximately 500 millivolts per nanosecond (mV/ns). A 2-volt (V) transition requires approximately 4 ns/5.4 ns/meter (m) = 0.74 m for a signal edge (assuming 5.4 ns/m as the propagation velocity of the signal edge). Therefore, some relief exists because the connectors and cable assembly terminations are much smaller than the signal edge length; the connectors and terminations do not need to have carefully controlled characteristic impedance properties. This allows the use of the technology available in the connector and cable assembly industry to optimize the interconnect properties without the considerable design, manufacturing, and test burden imposed by controlled impedance requirements.

Transmission Modes The transmission mode of a SCSI bus is determined by the properties of the terminators that, by definition, constitute the ends of the bus. Terminators also supply most of the energy required to operate the single-ended transmission-mode devices and additionally provide the required matching

Table 1
Terminology for Data Phase Speeds

Data Phase Speed Name	Maximum Transfer Rate (Million transfers/second)[1]	Maximum Byte Rate (Narrow) (Megabytes/second)	Maximum Byte Rate (Wide) (Megabytes/second)
Asynchronous	Unspecified	Typically ~ 3	Typically ~ 6
Slow (synchronous)	5	5	10
Fast (synchronous)	10	10	20
Ultra (synchronous)[2]	20	20	40
Ultra2 (synchronous)[3]	40	40	80
Ultra3 (synchronous)[4]	80 to 100	80 to 100	160 to 200

[1]One transfer is 1 byte in narrow mode and 2 bytes in wide mode; 1 byte equals 8 data bits plus 1 parity bit.
[2]Ultra is synonymous with Ultra1 and Fast-20.
[3]Ultra2 is synonymous with Fast-40.
[4]Rates not yet finalized; Ultra3 is synonymous with Fast-80 or Fast-100.

to the characteristic impedance of the transmission line. In differential SCSI, the terminators provide a small portion of the overall energy required to operate the bus; the differential drivers supply the remainder of the energy.

Drivers that want to transmit an asserted state must overcome the biasing provided by the terminators. The drivers operate locally on the bus and alter the state in their immediate vicinity when they switch on and off. For single-ended SCSI, the 0 state is approximately 2.5 V and the 1 state is approximately 0.5 V. For high-voltage differential SCSI, the 0 state is approximately –1 V to –2 V, and the 1 state is approximately 2 V. (The difference between a state 1 and a state 0 is higher with differential—typically, approximately 4 V.)

For single-ended transmissions, the drivers operate on energy previously stored in the bus by the terminators. This energy is mostly electrostatic energy in the charge stored in the capacitance of the transmission line for negated states and electromagnetic energy in the current flowing through the inductance of the transmission line for asserted states. Ultimately, the terminators will set the state back to negated after the drivers cease to source or sink current; however, this only happens after the round-trip propagation delay from the driver to the farthest terminator if the bus does not have matched characteristic impedance properties.

Approximately the same energy transformations occur for differential SCSI, but significant current is supplied by the drivers for both the asserted and the negated states.

Multidrop Requirements The multidrop architecture requires a continuous low-resistance path called the bus path between the terminators and allows devices to be attached to this path. The number and properties of these attached devices vary widely because of many factors including the speed of operation, the overall length of the bus, and the transmission mode. Attached devices always disturb the transmission line properties of the bus path; the key to successful operation is in the management of the magnitude of these disturbances.

Generally, the more capacitance or electrical length the device has, the more disruptive it is. Placing devices too close together along the bus path can cause them to appear electrically as a single super disruptive device. Placing them too far apart can result in an overall bus length that is too long.

Wired-OR Glitches During the arbitration phase, when the SCSI devices decide which devices will be sending payload data to or from each other, multiple devices may assert the same control line (BSY) at the same time. Each device that wishes to communicate asserts both the BSY line and its respective device identification (ID) line. After examining the asserted ID lines to determine which device has the highest ID, all but the device with the highest ID release the BSY line. This leaves only one device, the winner, asserting the BSY line. While the current in the BSY line is readjusting itself from a multiple-driver asserted condition to a single-driver asserted condition, noise pulses (called wired-OR glitches) propagate throughout the length of the signal line and may be detected collectively as an erroneous phase. Therefore, one of the architectural limits for parallel SCSI is the time required for these wired-OR glitches to settle. This bus settle time is set by protocol at 400 ns and must be interpreted as a round-trip propagation time when using a simple SCSI bus. Allowing some time for propagation through driver and receiver chips yields a maximum physical length for a simple bus of 25 meters.

Areas of Improvement

Thus, the opportunities for improving SCSI derive from appropriately managing the transmission lines, taking advantage of the multidrop architecture offered by a parallel wired-OR structure, using state-of-the-art technology from the interconnect and silicon industry, and making innovative use of the time required for the wired-OR glitches to settle. These techniques are the basis of the development by DIGITAL in the four areas addressed in this paper.

Speed increases in the synchronous data phase are based primarily on increasing the timing precision in the silicon transceivers by using newer silicon technology. The interconnect properties remain largely unchanged from those used for fast SCSI.

Circuits that enable segmentation of SCSI domains into easily managed pieces are based on systematic isolation of transmission line properties and use of wired-OR noise pulse properties. No software, interconnect, or device changes needed to use these circuits.

New connector and cable technology is based on an innovative 0.8-millimeter (mm) ribbon-style connector technology that optimizes the total SCSI electrical requirements with the capabilities of cable and connector design.

Dynamic removal and replacement of devices on an active bus, i.e., hot plugging, is based on the multidrop architecture, which enables devices to be added or replaced without affecting continuity between other devices. Hot plugging depends on understanding and managing the electrical disturbances created during the insertion or removal.

The remainder of this paper provides details of these four areas of improvement. The end result of these extensions to the basic physical architecture of parallel SCSI is a major increase in its capabilities, accompanied by only a very minor disturbance to the installed base, especially the software.

Increasing the Synchronous Data Phase Speed

Beginning with the SCSI-2 standard, the synchronous transmission mode is available for transferring payload data between SCSI devices. The devices select this mode by mutual agreement before any synchronous data is passed. The agreement is achieved by using the asynchronous transmission mode, which is slow but usually reliable.

The synchronous data phase uses the DATA and PARITY bit lines for the data and either the REQ or the ACK control line as a signal that the receiver uses for capturing the data. The term synchronous derives from a specified timing relationship between the bit line signal edges and the REQ or ACK signal edges. (The falling edge of the ACK signal is used when the data phase transmission originates from the SCSI initiator, and the falling edge of the REQ signal is used when the transmission originates from the target.) There is no synchronous relationship between the internal timing references on different SCSI devices, so the receiver must buffer the received data before introducing the data into its internal data management structure. This buffering is usually accomplished by means of a first in first out (FIFO) circuit that uses the REQ or the ACK signal as the latching signal for the incoming data. For convenience, in this paper we only refer to the ACK signal, with the understanding that the same discussion applies to the REQ signal when it is used as the data-latching signal.

Since only the falling edge of the ACK signal is used in the presently specified SCSI versions and an ACK signal is required for every data transfer, it follows that the ACK signal cycles at least twice as fast as the data bits. When a continuous stream of transfers is transmitted, the ACK signal is a regularly repeating signal, nominally, a square wave. An alternating 1/0 pattern produces the highest fundamental frequency for the data bits at half the frequency of the ACK signal. Therefore, the ACK signal requires careful attention since it is the most demanding on the transmission process.

The focus of this section is to examine how the speed of the synchronous data phase was increased by a factor of two to achieve the Fast-20 (UltraSCSI) specification.

Status before UltraSCSI

In 1993, the SCSI-2 standard[3] had been in place for two years, and a follow-on standard called SCSI-3 Parallel Interface (SPI)[4] was technically stable. SPI had been created largely because the specifications in the SCSI-2 standard were not effective in implementing the single-ended version of the synchronous transmission (10 megatransfers per second). The differential version specified in SCSI-2 worked well but was much more expensive in cost, power, and space than the single-ended version. Therefore, most of the interest was in making the fast single-ended version work adequately.

Taking single-ended SCSI from asynchronous and slow synchronous (5 megatransfers per second) to the fast synchronous technology was difficult. The prevailing opinion was that the SPI standard represented the final improvement to parallel SCSI. This view set the stage for a number of alternate physical technologies based on the serial point-to-point transmission schemes used in communications technologies, e.g., Fiber Distributed Data Interface (FDDI) and Ethernet, to be used for higher-performance storage applications.

DIGITAL's Storage Bus Technical Office had seen many instances of difficult implementations that were the result of less-than-optimal understanding and management of the specification margins. No credible study had been presented on the margins available in SCSI, so the thrust was to create baseline characteristics of multidrop parallel SCSI to determine where unused margin might exist.

Little data was available on the precise reasons why specific implementations of fast synchronous SCSI did not work. The system would hang or report various error messages with almost no indication of the basic causes. A method that could report margin to failure and mechanism of failure was needed to unravel this situation. Therefore, the approach DIGITAL took was to step back from full SCSI implementations and to examine the pieces without the encumbrance of the SCSI protocol.

One of the most mysterious areas was the behavior of SCSI receivers. The SCSI-2 and SPI specifications used bipolar transistor-transistor logic (TTL) levels as the basic receiver input levels. Almost all SCSI devices were being designed with complementary metal-oxide semiconductor (CMOS) technology, so the differences between the receiver properties presented a key opportunity for hidden margin. Other unknown areas were jitter, cross talk, skew, ground offset, effects of stubs, and worst-case configurations.

DIGITAL built a special test environment to systematically examine each piece of parallel SCSI. The environment was named the PBDIT, an acronym for parallel bus data integrity tester. This test environment made it possible to systematically examine the real margins to failure for the key pieces and to develop the confidence that SCSI could be used at elevated speeds and be made highly robust at the slower speeds.

Special Test Environment

The test environment was built to allow known data patterns to be transmitted across a SCSI device, into SCSI transmission media, and then into another SCSI device. The same data pattern is loaded into both sides so the receiver knows exactly what data it is supposed

to receive. The transmitting side is called the exciter, and the receiving side is called the comparator. Received data is committed to the comparator by using one bit line as the latching ACK signal in a manner exactly like that specified in synchronous SCSI transmissions. The test environment allows the position of the ACK signal to be adjusted with respect to the data signal edges.

Since the comparator knows the data pattern that is transmitted, it is possible to isolate the precise data bit that caused the transmission error. This kind of error-directed methodology has found widespread use in the integrated circuit industry.

Other features of this test environment include detachable load boards that contain the SCSI drivers, terminators, receivers, connectors, or any other physical media-dependent components. The minimum requirements for a load board are that the exciter contain the SCSI driver and a connector and that the comparator contain the SCSI receiver and a connector. Other components may be placed between the load boards for different test conditions. The SCSI driver must have accessible points for the exciter logic, and similarly, the SCSI receiver must have output points to drive the comparator. These requirements eliminate drivers and receivers that are imbedded within chips with other functions. Fortunately, separate SCSI drivers are available for both single-ended and differential versions. (The differential versions normally use separate chips, but only a few choices are presently available for the separate single-ended versions.)

The test environment is useful for developing the understanding of operating mechanisms and for measuring the margins for specific hardware configurations. This environment is not useful for deriving specifications, since the performance at the specified interfaces, i.e., the device connectors, is not directly observable. Oscilloscope measurements provide the basis for setting compliance specifications, since these measurements can be performed at the connectors. The basic question that needed an answer was, Can parallel SCSI be operated at elevated speeds with reasonable margin to failure? DIGITAL optimized the special test environment to answer this question. Other specifications that would be necessary to ensure interoperable operation between UltraSCSI devices could be derived if it appeared possible to achieve the end result.

The data pattern loading and digital control of the exciter and the comparator were achieved through optically coupled means. This allowed the ground offset voltage to be adjusted between the driver and the receiver without compromising the operation of the logic.

The data flows only from the exciter to the comparator. If bidirectional information is desired, the physical connections between the exciter and comparator have to be reversed. This scheme leaves untested the cross-talk effects on the REQ signal that is traveling in the opposite direction to the ACK signal (if ACK is synchronized with the data as in a write operation). Separate measurements are necessary to examine this issue. Cross talk into other control lines is addressed by holding these lines constant in the data pattern transmitted.

The SCSI standard deals with the REQ cross-talk issue by requiring that the data lines be physically separated from the REQ and ACK lines in the transmission media. Measurements not reported in this paper have confirmed negligible speed-related cross talk into the REQ line.

Up to 27 pairs of 3-byte-wide lines (wide SCSI uses only 18 pairs for high-speed transmissions) can be tested with the special test environment. Figure 1 is a functional diagram of the test environment. The SCSI terminators are shown as separate from the load

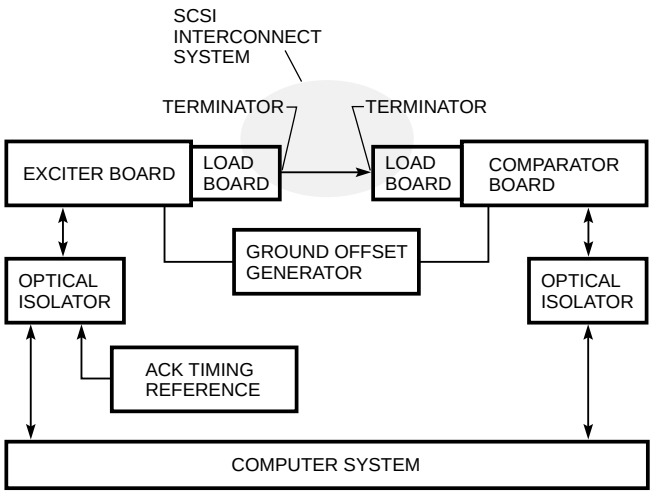


Figure 1
Special Test Environment

boards in this case. A key feature of this kind of testing is that the test does not necessarily stop when an error is detected. In fact, the environment may detect errors 100 percent of the time. This acceptable behavior allows mapping of the complete bit-error response of the system.

Sample Data from the Special Test Environment The test environment allows a multitude of tests to be performed. The test scheme described in this section is the one that was used to establish the basic timing margins available from normal SCSI silicon, cables, connectors, and terminators.

A random repeating data pattern with 16 thousand different bit combinations was used as the basic data pattern. This pattern was transmitted over a period of time, and the number of errors detected was recorded. In this test, an error is defined as one or more bits in the received data transfer that do not match the transmitted bit. To acquire a new error rate data point, the transmission test is repeated by using exactly the same number of transfers in the same time period with the same data pattern but with some test parameter changed.

Virtually any parameter can be varied for different tests. For a given physical configuration, the most useful parameter for determining the timing margin is the position of the ACK pulse with respect to the data edges. The basic data then becomes the number of errors detected and the position of the ACK pulse edge.

There are two basic random variables operating in this scheme: the data pattern and the jitter induced by non-data-dependent sources. It is easy to separate these two variables by using extremes in the data pattern: very few transitions and the maximum number of transitions (every data edge has a transition, i.e., alternating 1/0 pattern). Although this level of precision is available, we will see that we really do not need to bother for parallel SCSI at the maximum UltraSCSI rate.

Figure 2 shows a typical error rate plot from a simple single-ended configuration made from ordinary SCSI interconnect hardware and transceivers being tested at the maximum UltraSCSI rate. Each data point represents a 3-second sample (60 million transfers) at each ACK position. The ACK position is incremented in 0.1-ns steps for a total of 240 independent tests in the plot. To minimize the testing time, we tested only the time ranges from –3 to 9 ns and 44 to 56 ns. The individual data points are not distinguishable in this presentation, and there is very little scatter between neighboring points. In Figure 2, the error rate of 1 is used to indicate that no errors were detected, since the log of 0 is not easy to plot.

Examination of the raw data reveals that the plot is monotonic in detected error rate to the fourth decimal place. This indicates an extremely predictable situation as far as behavior of the same set of hardware is concerned. That is, there is virtually no Gaussian jitter present, and a SCSI system could be designed to be quite reliable and stable at the maximum UltraSCSI rate.

Extending the sample period to 5 minutes made no difference in the position of the key features. Using the 3-second sampling time, the entire data set could be acquired automatically in approximately 12 minutes.

The onset of errors is extremely sharp as the ACK position approaches the critical position. One hundred picoseconds changes the observation from 0 to 864 errors near the 8-ns position. On the other end, the 50.1-ns time produced 7 errors, and the 50.2-ns time produced 425 errors. No errors were detected at any of the times between 50.1 ns and 7.9 ns. This data shows that there are no strange effects that prevent SCSI from operating at the maximum UltraSCSI rate.

As the ACK position proceeds into the region of more errors, a condition is finally reached in which *all* the transfers have errors. On the one hand, the probability that one transfer has the same data content as its

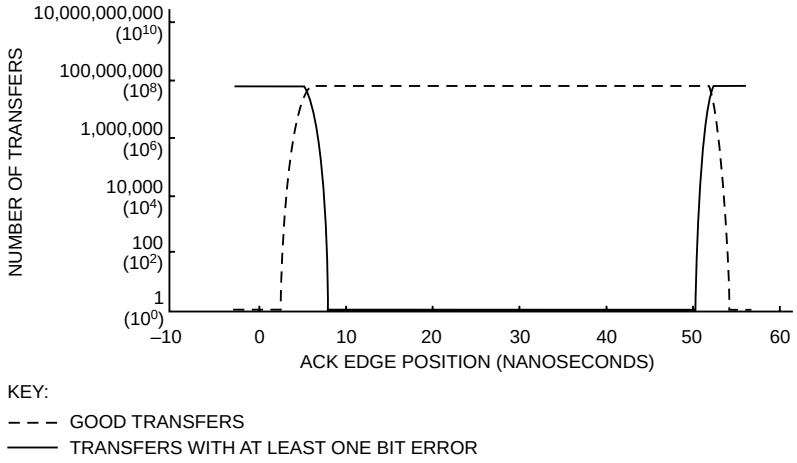


Figure 2
Typical UltraSCSI Error Rate Plot

neighbor's is very small with this random data pattern. On the other hand, since a random data pattern is being used, there is a reasonable chance that a bit will actually match that transmitted in one state but not in the other state. The random data pattern tends to spread out the time between the first error and the last good transfer. In the limit, for perfectly random data, this time is a measure of the total timing imprecision in the system.

This imprecision includes skew in the exciter and comparator boards, in the SCSI drivers and receivers, and in the cable transmission media (including loads, if any), and all forms of jitter. For the test conditions shown in Figure 2, the total difference is 3.6 ns near the 5-ns point and 5.4 ns near the 52-ns point. This shows that the skew specifications in the SCSI standard are overspecified as compared to actual hardware performance.

The data shown in Figure 2 is representative of a large variety of configurations up to approximately 3 meters long and loaded or up to much longer point-to-point lengths (20 meters or more [see Figure 6]). The error-free window can be made to collapse by adding too many loads or by using the wrong impedance cable, improper terminators, receivers with the wrong threshold voltages, or other bus component and configuration parameters. However, the details of the actual hardware and configuration do not affect the basic conclusion derived from Figure 2, namely, that a great deal of timing margin is available at the maximum UltraSCSI rate when ordinary SCSI hardware is used.

To put this into perspective, basic gigabit-per-second serial transmissions with approximately twice the basic bandwidth of UltraSCSI have bit times of about 1 ns and timing margins of a few hundred picoseconds. UltraSCSI has an effective margin window of a few tens of nanoseconds. This represents two orders of magnitude more margin for the parallel SCSI application.

The initial errors usually originate from the same bit. This bit is the one with the most unfavorable timing skew with respect to the ACK signal. The cliff is not perfectly sharp because there is a 50 percent chance that the data transmitted is the same as that expected even under the error case and, more importantly, because there is some level of jitter present. It is this jitter that softens the cliff. Thus, the first errors detected happen when the skew of the weakest bit adds to the tail of the jitter distribution. Only a few errors are present because only a small part of the jitter population extends far enough to trigger the error. SCSI systems will experience virtually no errors because of these mechanisms in service if one operates 1 ns or more away from an error cliff.

Note that these results from the special test environment almost always yield margins higher than those calculated from a set of interoperability specifications. This is because the interoperability specifications must allow margin for each piece, and the special test environment reports the integrated result from many pieces in the complete SCSI connection.

Higher Speeds The main effect of further increasing the transfer rate above the maximum UltraSCSI rate in the same set of hardware is to change the time position of the onset of nonzero error rates and to narrow the error-free region. Figure 3 shows an example of data from Fast-40 transmissions using separate high-voltage differential transceivers on each bit. (This data was acquired by DIGITAL's Storage Bus Technical Office in 1994.)

The error-free zone has narrowed to approximately 15 ns, and the time between first error and 100 percent errors has widened on both sides, but still no uncontrolled regions exist. This strongly suggests that at least Fast-40 transfer is possible with no major technology changes in the interconnect.

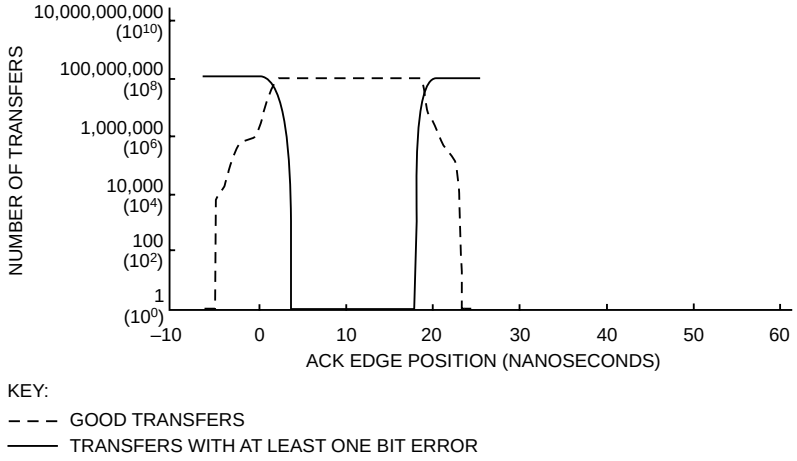


Figure 3
Fast-40 Error Rate Plot

Additional Tests Other tests that are useful with the special test environment are ground offset effects, terminator power effects, correlation of time domain plots on the signals with error rate distributions, hot-plugging testing (which results in good error detection), and comparison of the impact of different cables and transceivers. Test results of this nature are not included in this paper because the impact of these variations depends on many parameters and the results may not be generally applicable.

Timing Specification Methodology

With the increased emphasis on timing precision for UltraSCSI technology, it was necessary to introduce better specifications for the measurement of timing parameters than those in the SCSI-2 and SPI standards. Figure 4 shows the precise measurement points and features used for the specification of single-ended UltraSCSI signals.

The effects of the finite slew rate on the signal edges are accounted for largely by specifying the voltage levels that coincide with the receiver input levels. Thus, the setup time ends when the receiver is able to detect an asserted state at 1.3 V, and the asserted period begins when the asserted state has been detected. On the negation side, the signal must rise to at least 1.6 V before the receiver can detect a negated state, and a negated state must be detected if the input signal reaches 1.9 V. In the SCSI-2 and SPI standards, any point between 0.8 V and 2.0 V could be used as the timing measurement.

Sample UltraSCSI Signals

Numerous variations on the details of the signals can be produced in UltraSCSI configurations. This section shows two types of signals as representative examples that validate UltraSCSI as viable under certain conditions. The first case explores a configuration that actually exceeds the recommended specifications. This is a complex cabled environment with a cluster of loads on one end and some distributed loads on the other end. The second case shows the signals over a 25-meter, single-ended point-to-point bus.

Complex Loads Figure 5 specifies a complex configuration and the single-ended SCSI signals that result at various positions along the bus. The logic

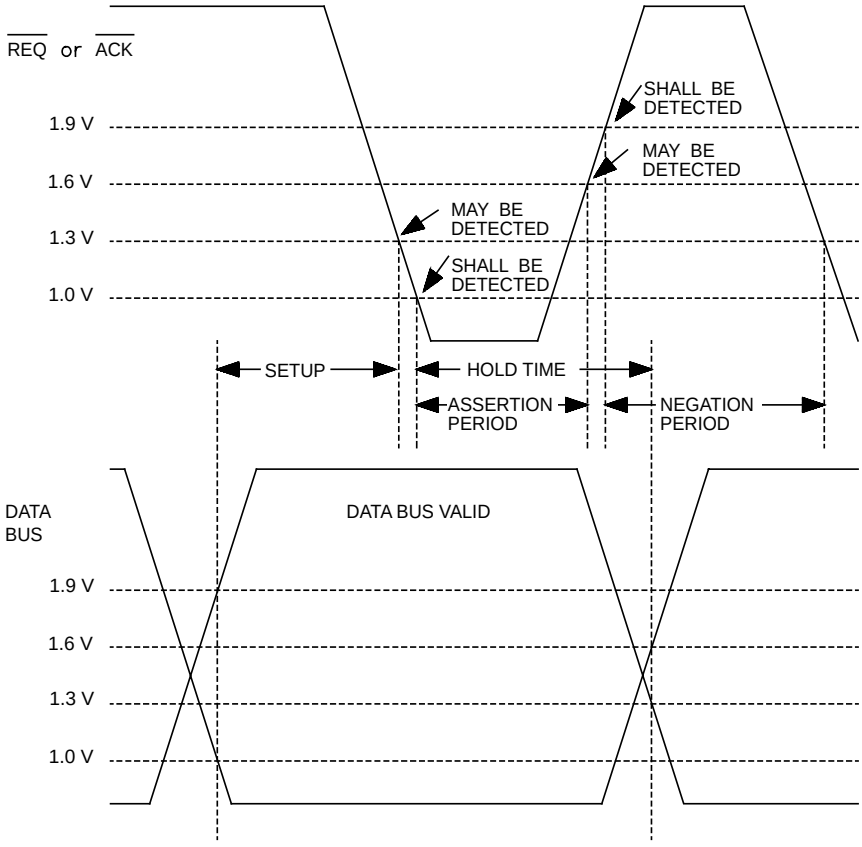


Figure 4
Single-ended UltraSCSI Timing Measurements

signal that is driving the SCSI driver chip is the first trace at the top; it provides a common timing reference for all the signals. The weakest signal is at device position 4, just after a relatively long run with no loads. This signal is below the 1-V level but has a very slow assertion slew rate that causes considerable loss of asserted state pulse width. This complex configuration works with the receivers used but does not have the timing margin required by the Fast-20 standard.

By varying the position of the loads so that there are no loads between the driver and the first load (not shown), the signal at the first load device is degraded even more than at position 4 in Figure 5. This is one reason that the overall length of single-ended UltraSCSI with many loads is restricted to 1.5 meters and that the total number of loads is limited to 8.[1] UltraSCSI devices connected to backplanes may be especially sensitive to attached cables that extend the total bus length more than 6 to 8 centimeters (cm) beyond the backplane. This reduced bus length is rather severe when compared to that allowed at the maximum fast SCSI transfer rate (a total of 3 meters).[4] In the section Small, Improved Interconnect, we show how to overcome this 1.5-meter, 8-device limit by using an active SCSI interconnect.

Applying the timing measurement methods shown in Figure 4 to the waveforms in Figure 5 illustrates that more careful timing specification methods do indeed help significantly to keep the timing margin high enough to use.

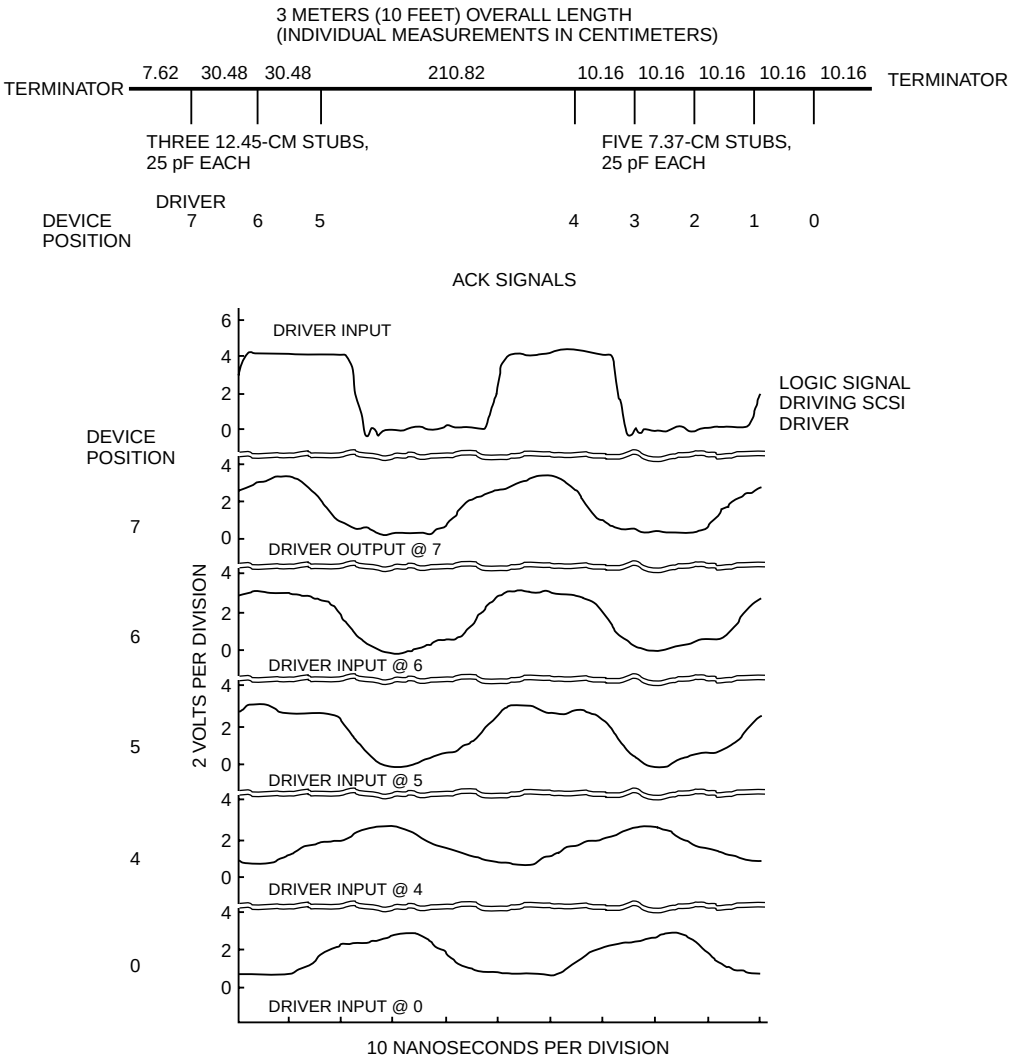


Figure 5
UltraSCSI Signals in a Complex Bus

Point-to-point Configuration If loads are present only at the ends of the bus, the transmission line between SCSI devices improves electrically. This occurs simply because the loads significantly disrupt the characteristic impedance and cause reflections and attenuation. The point-to-point signal at 25 meters has better amplitude and timing margins than signals in much shorter buses with closely spaced loads. Figure 6 shows a typical example of a point-to-point UltraSCSI signal. The format used in Figure 6 is the same format used in Figure 5.

Differential UltraSCSI

Differential UltraSCSI uses the same configuration rules as fast SCSI (25-meter total length, 20-cm [8-inch] stubs, 16-device load)[1] and uses the same timing values as single-ended UltraSCSI. The larger signal amplitudes and the common mode rejection property of differential transmissions help overcome the transmission line weaknesses in heavily loaded and long buses. As with any high-voltage differential system the costs—in terms of money, power, and space—are higher.

Other Requirements for UltraSCSI

The Fast-20 standard[1] contains a number of detailed requirements on the components used in UltraSCSI configurations. Included are slight modifications to the cable impedance, active negation requirements for drivers, special length limits for certain loading conditions, restrictions regarding the kinds of single-ended terminators to use, and timing budgets.

Summary of Developments in the Area of Increased Synchronous Data Phase Speed

The UltraSCSI (Fast-20) speed increase can be attributed to a systematic examination of the margins present in actual SCSI hardware and to the elimination of the excess margins. Advances in the integrated circuit industry enabled silicon designs to be specified with tighter controls on the driver and receiver timing and threshold properties than were possible when the SCSI-2 or SPI standards were developed. All the important changes needed for SCSI devices are contained in the silicon designs for the drivers and receivers. As a result, the user sees no difference between the appearance of UltraSCSI and that of ordinary SCSI.

The system integrator must use a more restrictive set of configuration rules than required for fast and slow SCSI. Also, the only impact on software is the addition of a new speed agreement code for the rates uniquely supported by UltraSCSI. This negotiation is done precisely the same way for UltraSCSI as for any other form of SCSI. Finally, UltraSCSI devices are 100 percent backward compatible with fast and slow devices. Although a device may be capable of the maximum UltraSCSI rate, it may be needed in a configuration that does not support UltraSCSI. In such a case, the UltraSCSI device would be used in the fast or slow mode and would have more margin at those slower speeds than it would if it were not UltraSCSI capable.

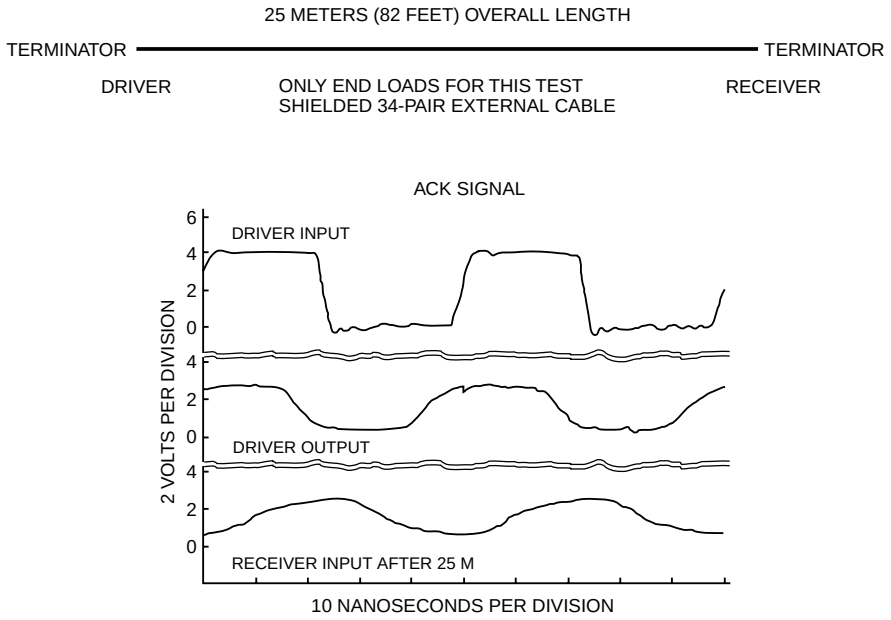


Figure 6
Point-to-point UltraSCSI Signals

Bus Expanders

As noted previously in the discussion of complex loads, there are rather severe limits on the configurations that can be achieved with single-ended UltraSCSI when implemented in a single bus. The extension to parallel SCSI architecture that overcomes this constraint involves using active circuits that connect SCSI buses electrically but isolating them from each other in a transmission line sense. These circuits have the general name expanders, since they expand the configuration capabilities of parallel SCSI.

Each individual bus has two terminators and its own transmission mode (single ended or differential) and obeys transmission line–based configuration rules as if it were the only bus in the system. When used with expanders, these individual buses are called bus segments. The collection of SCSI devices in all the bus segments that are electrically connected together is called the SCSI domain. One example of a SCSI domain using expanders is shown in Figure 7. Note that when using expanders, it is possible to have bus segments that do not have any SCSI initiators or targets but only serve to form an electrical interconnect between other bus segments.

Expander Properties

Expanders are available in two basic types: simple and bridging. Bridging expanders behave as a SCSI initiator or target, whereas simple expanders have a set of properties that make them look like a piece of wire with delay to the protocol. Simple expanders

- Cannot initiate SCSI IDs and arbitrations and cannot originate messages, although the expanders can read messages sent from initiators and targets
- Allow minimal arbitration propagation delay
- Yield a retransmitted signal timing skew (both delay and high/low) no worse than from valid SCSI initiators or targets
- Do not interfere with the REQ/ACK offset count
- Allow min/max pulse widths to be maintained
- Require the filtering of the SCSI RESET line
- Allow arbitrary placement of the initiator and the targets
- Require that terminator power not be connected between the segments being coupled
- Do not need to know the negotiated data phase speed or any other variable property of a transaction
- Require that there be no electrical or logical connection of the DIFFSENS line (a single-ended signal that indicates the transmission mode being used on the bus segment) between segments being coupled
- Issue a SCSI bus RESET signal on one segment on detecting transmission mode (single-ended/LVD, etc.) changes on the other segment

Simple expanders are becoming available from several sources in the industry for use with UltraSCSI.

Domain Rules Using Simple Expanders

When using only simple expanders in a domain, six rules must be observed:

1. All bus segments in the domain must comply with their individual bus segment length limits and other segment-related requirements.
2. Any segment between two other segments must support the highest performance level that can be negotiated between the two other segments. For example, two wide UltraSCSI segments must not be separated by a segment that does not support both wide SCSI and UltraSCSI.
3. The maximum propagation delay between any two devices in the domain cannot exceed 400 ns. A special case exists for devices that use extremely long times for responding to BUS FREE (the so-called BUS SET DELAY)—the one-way propagation limit is 300 ns instead of 400 ns.
4. The number of addressable devices cannot exceed 16 unless the domain contains bridging expanders.
5. A branch/leaf architecture must be observed; loops are not allowed.
6. The REQ/ACK offset negotiated between any two devices must be large enough to ensure that adequate offset and buffering is available to accommodate the round-trip time between the devices. For the maximum UltraSCSI rate with a 400-ns maximum one-way domain propagation time, the

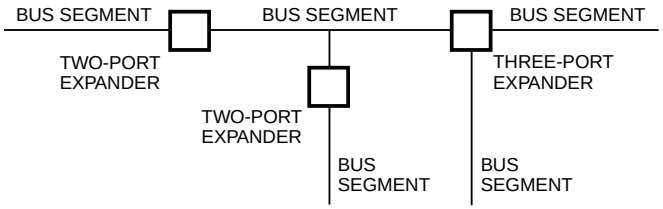


Figure 7
SCSI Domain Built Using Expanders

minimum offset is 18. (This offset level is derived by considering a maximum round-trip time of 800 ns at 50 ns per transfer [800/50 = 16] and somewhat arbitrarily adding two transfers to account for some additional delay due to the processing time in the silicon.)

Achieving the 400-ns one-way domain delay requires expanders that will not pass the wired-OR glitch (noted earlier in the introduction) between bus segments. This filtering of the glitch allows the bus segments to settle individually.

The propagation delay through an expander directly subtracts from the physical distance between devices. It is therefore desirable to use expanders with small delays. For a single-ended–to–single-ended application, the delay can be as low as 10 ns. For a single-ended–to–differential application, the delay is typically around 100 ns, which is another significant penalty to using differential bus segments.

More detail concerning these rules and other properties is available in the draft ANSI document: *SCSI Enhanced Parallel Interface,*[5] which was edited by the author of this paper.

Summary of Improvements Related to Bus Expanders

The use of simple expanders dramatically extends the utility of single-ended UltraSCSI. The most obvious example is the ability to introduce point-to-point segments where additional length is needed. A less obvious example is the ability to create star or hub configurations by clustering simple expanders into a local physical area. An example of a three-port SCSI hub is shown in Figure 7. Note the three simple expander circuits internally connected within the hub. Simple expanders also make it possible to mix single-ended and differential SCSI devices in the same domain, to achieve the full 16-device count, to add and remove bus segments without shutting down the entire domain, and to achieve differential performance without incurring the extra cost of differential. Bridging expanders offer the same transmission isolation as simple expanders and may allow increasing the number of devices in the domain to as high as 946,[5] but bridging expanders are not as well developed as simple expanders and will not be explored in depth in this paper.

Note that the improvement in signal integrity is dramatic when using expanders with backplane applications. Therefore, it is good practice to use an expander whenever connecting a SCSI cable to a backplane that contains SCSI devices.

Smaller, Improved Interconnect

Another recent development in parallel SCSI technology is the introduction of much smaller external physical interconnects and more capable internal device interconnects. The SCSI connectors and shielded cables have historically been large, bulky, and generally difficult to manage.

Spearheaded by activities that began in 1995 in the SFF (formerly Small Form Factor) industry group, standardization is under way of two new connector families that offer unprecedented levels of functionality and true multisourcing of complete connectors for parallel SCSI. These families are the Very High Density Cabled Interconnect (VHDCI)[2] shielded connectors that reduce the overall size of an external connector by two thirds and the Single Connector Attachment–2 (SCA-2)[7] unshielded connectors that integrate into a single connector all the functions needed to run a peripheral. The VHDCI family revolutionizes the external SCSI interconnect and the controller parts of the internal SCSI interconnect; the SCA-2 family does the same for the internal device interface.

For the first time, complete connectors—not just the mating interface—are being standardized. This feature is essential to achieving interchangeability and second sourcing for connectors with the same style of termination-side contact. The VHDCI family is specified in 26 different forms, all with exactly the same mating interface, so that virtually any kind of device or cable assembly design can be accommodated. Interestingly, this array of choices for the connectors does not increase the complexity of the interconnect but rather opens up new ways for product developers to design products while maintaining a simple and physically interoperable separable connector interface. In fact, this ability to accommodate a variety of product design requirements without changing the separable interface is one reason that SCSI is becoming *less* complicated.

Similarly, the family of SCA-2 connectors for SCSI internal devices and cables is following the VHDCI standardization model, with a significant number of intermatable forms being standardized. These connectors offer the ability to bring all the SCSI signals, all the power and ground connections, and all the optional signals, such as IDs, spindle sync, and power fail, out of the device through a single unshielded connector. This feature dramatically shrinks the cost and complexity of interconnecting an array of SCSI devices.

Using an SCA-2 connector, the device may be inserted into a backplane without using cables. If the SCA-2 and backplane combination is not used, a SCSI cable (50-pin or 68-pin conductor), a four-lead power cable for ground and power (5-V and 12-V), and one or more smaller cables for the IDs etc., are required for *every* device in the system. Each of these cables is routed differently, has different current carrying and other electrical requirements, and has very different connectors. Although this cabled option is flexible and offers significant advantages in some systems, it is usually not the best solution in the device array and modular packaging applications that are required for the

higher-end applications. Therefore, the SCA-2 is a significant factor in the dramatic reduction in complexity of higher-end SCSI device applications.

VHDCI Connectors

The physical size of the VHDCI connectors is much smaller than the earlier versions, as seen in Figure 8. Because of its low profile, the VHDCI 68-pin family is approximately half the height and twice the width of the latest Fibre Channel external connector, the High-Speed Serial Data Connector (HSSDC). Figure 9 shows a comparison of the VHDCI and HSSDC connectors. The same panel space is required for either technology.

The VHDCI connectors shown in Figure 9 are closely spaced, but the orientation of the polarizing shield connection is 180 degrees different between the upper and lower connectors. This arrangement allows an offset cable assembly to be used where one side is flat. This same cable assembly may be used on both the upper and lower connectors without interference. The specifications of the VHDCI interface ensure that neighboring PC option slots will not have interference even if all the SCSI ports have cable assemblies attached.

The VHDCI connector is useful for multiport applications such as RAID (redundant array of inexpensive disks) controllers. Figure 10 shows examples in which the wide version of the connector family has allowed at least a doubling of the number of ports possible in a single controller form factor. As illustrated in Figure 10, the device design enables up to four wide SCSI ports on a single PC option card cutout.

The VHDCI retention scheme is also significantly simplified by introducing a three-way retention post for the bulkhead connector. This post accepts (1) the conventional jackscrews, (2) a squeeze-to-release clip for positive retention with rapid release, or (3) a detent ring retention that requires a stronger pull than that required with no retention but no action other than

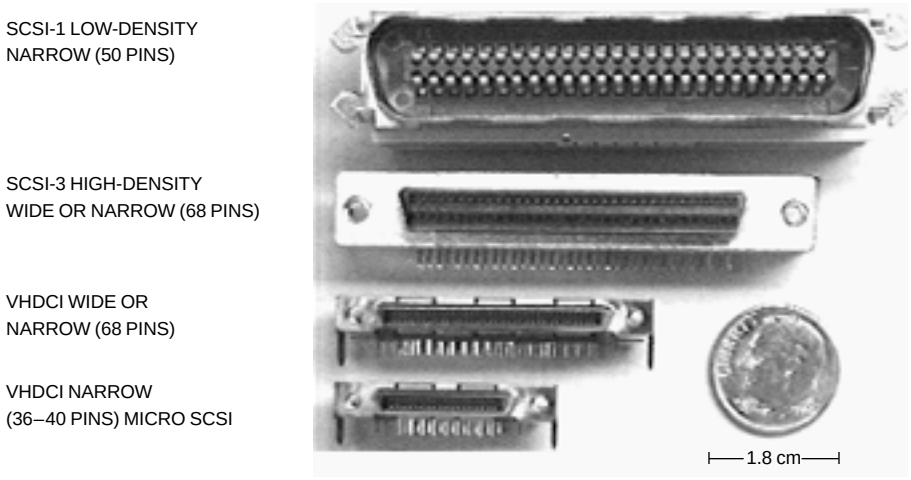


Figure 8
External SCSI Connectors

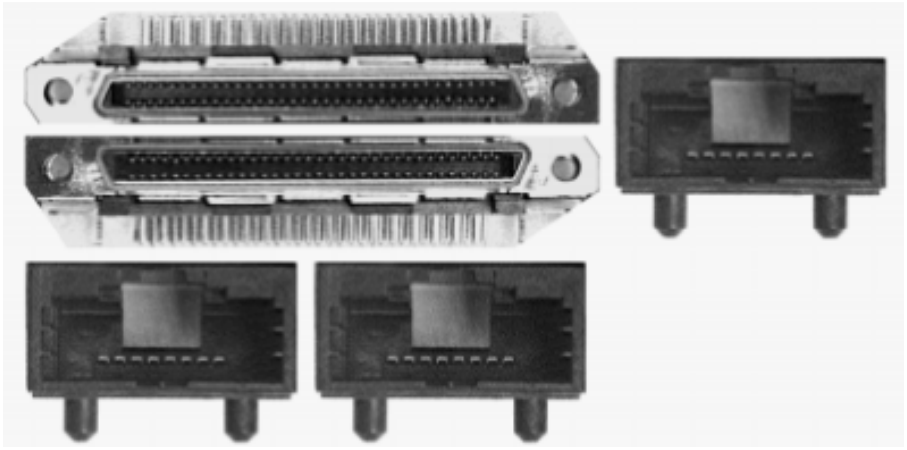


Figure 9
Comparison of the 68-Pin VHDCI and Fibre Channel HSSDC Connectors

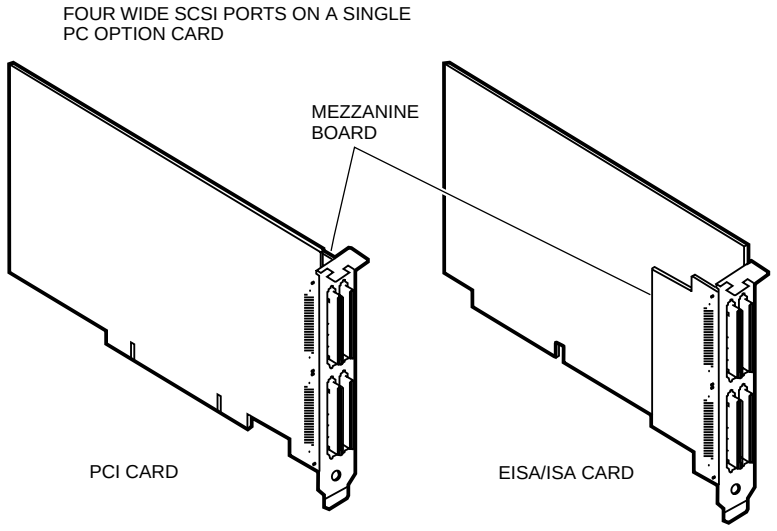


Figure 10
Four Wide SCSI Ports on a Single PC Option Card

pulling or pushing. The choice of retention type is made in the cable assembly. All 68-pin VHDCI cable assemblies that comply with the SFF specifications work on all 68-pin VHDCI mating connectors.

Figure 11 shows the details of the 68-pin VHDCI system. The lip in the jack post provides the securing point for squeeze-to-release clips and for split-ring detent retention. The center of the jack post is threaded for use with jackscrews.

Although smaller than the high-density connector, the VHDCI connector is durable. It has no pins that can bend; its retention scheme uses the same-size jackscrew thread as the high-density wide connector; and its contacts are imbedded in the housing where they cannot move or become distorted.

SCA-2 Connectors

The SCA family uses an 80-position, leaf-style contact to interface all active SCSI lines, three power voltages, and device control signals. This connector is considerably smaller than the collection of the three different connectors used for power, options, and SCSI bus in a cabled system. There are two basic versions of SCA connectors: SCA-1 and SCA-2. Both versions are unshielded and useful only within shielded enclosures. The SCA-1 has 80 positions with all contacts designed

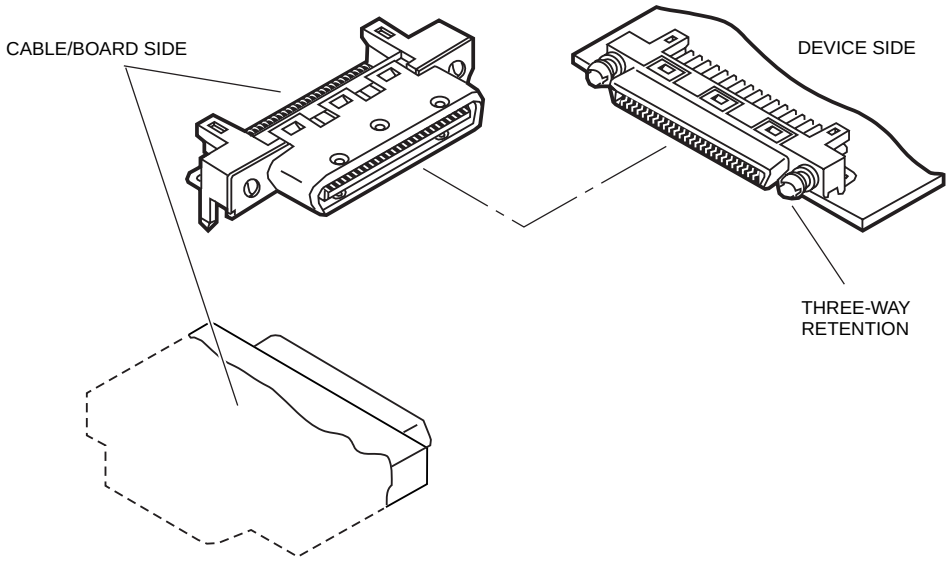


Figure 11
Overall View of the 68-pin VHDCI System

to be the same length. The SCA-2 can be mated to the SCA-1 but has advanced grounding contacts and sequenced signal and power contacts for supporting hot plugging and blind mating (no visual feedback during mating). Both versions are available in many styles, which differ by the termination-side structure and overall orientations.

The SCA-1 is not a documented standard and is being replaced by the SCA-2. The SCA-2 connector was introduced to SFF in 1995[7] as the first step toward formal standardization.

Two special features exist in the SCA-2 connector. First, two contacts, one on each side of the connector, provide the first make/last break for the ground connection. This design ensures that a common electrical ground is established between the device and the system before any power or signal connections are made on device insertion. Upon removal, these contacts ensure that the ground stays intact throughout the disengagement of the signal and power pins.

The second feature allows the special long power contacts to precharge bypass capacitors before the main power contacts make. This reduces the disturbance to the power distribution system and eliminates any arcing on the service power pins. Two pins at the extreme ends of the connector indicate that the connector is fully mated. The overall view of the SCA-2 system is shown in Figure 12.

The size of the connectors in the SCA family has not decreased dramatically. The connectors need to maintain enough size to achieve blind mating alignment, and, for backplane applications, there is little advantage in having a connector that is smaller than the device. With 89-mm (3.5-inch [in]) or the newly proposed 76-mm (3-in) form factor devices, the SCA connector comfortably fits within the device boundaries.

The use of backplanes for direct device attachment is possible because all the electrical connections for the device are available in one connector on the device. This design eliminates the cables used to attach the device and the space required for the connectors, thus significantly shrinking the size required to package multiple devices.

External SCSI Cable

The external cable for SCSI is shrinking also, through the use of smaller-gauge wire, better dielectrics, and less jacketing material, as illustrated in Figure 13. Formerly, wide SCSI required a cable of approximately 12.70 mm (0.50 in) in diameter (a 126.677-mm^2 [0.196-in^2] cross section) with 28-gauge wire. Today, wide SCSI cables with 30-gauge wire are shipping with diameters of 9.40 mm (0.37 in) (69.398-mm^2 [0.107-in^2] cross sections). Cables with 7.62-mm (0.30-in) diameters (45.61-mm^2 [0.07-in^2] cross sections) are possible with 32-gauge wire and inexpensive dielectrics for wide SCSI. Cables with 6.35-mm (0.25-in) diameters (4.987-mm^2 [0.049-in^2] cross sections) for narrow SCSI (45.61-mm^2 [0.07-in^2] cross sections) are flexible and manageable—similar in size and flexibility to a desktop computer power cord and smaller than many serial cables. When used with active single-ended, LVD, or HVD terminators, the 32-gauge wire is adequate for distributing terminator power and SCSI signals in most applications. Long cables should not be used for terminator power distribution.

Further reductions in the connector and cable sizes need to be weighed against the ease of handling, the need for sufficient strength to survive normal service stresses, and the cost increases at very small sizes. The combination of the VHDCI connector and 30/32-gauge wire sizes is a good optimization.

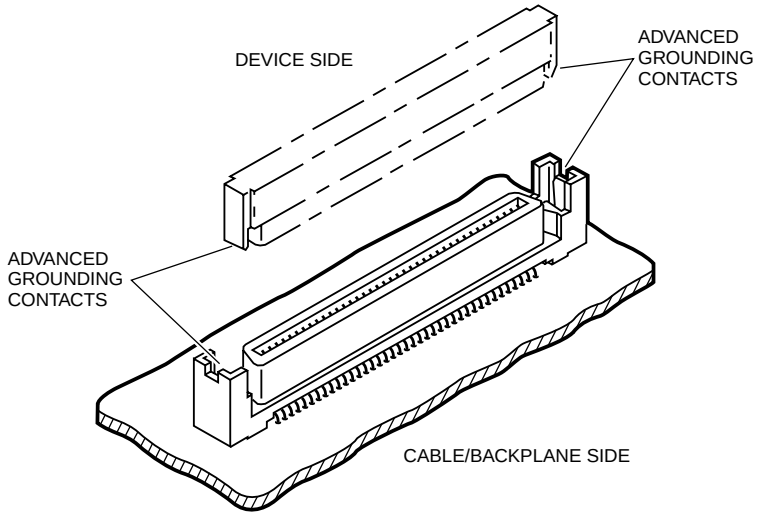


Figure 12
Overall View of the SCA-2 Connector System

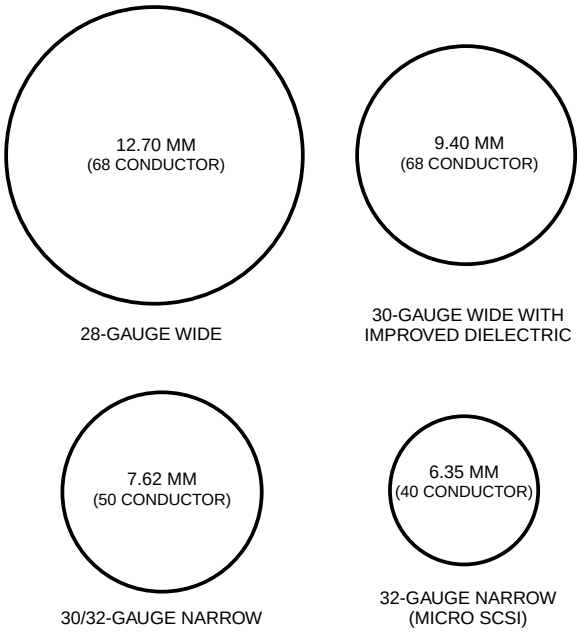


Figure 13
External SCSI Cable Diameters

Summary of the Benefits Derived from a Smaller, Improved Interconnect

The VHDCI connector and smaller cables combine to offer a robust yet user-friendly revolution in SCSI interconnect. The leaf-style contact of the SCA connector eliminates problems with bent pins that frequently bedevil the older wide SCSI connector. The ability to use up to four wide UltraSCSI ports in a single PCI option slot increases the SCSI connectivity per PCI slot to 60 devices (from 15 devices). By using multiple PCI slots, hundreds of SCSI devices can be connected to a single PC or workstation. In addition, the SCA-2 connector implements the essential contact sequencing required to perform SCSI device hot plugging.

Device Insertion and Removal Bus Transients

The multidrop feature of the SCSI bus allows device removal and replacement without disturbing the communications between other SCSI devices, if the electrical disturbances caused by the device being added or removed are not detected by any other SCSI devices. Thus, it is architecturally possible to dynamically reconfigure the device population without interrupting existing data transmission processes between operational devices.

The transients involved with device insertion and removal include mechanical vibrations, power distribution instabilities, SCSI terminator power noise, electrostatic discharge (radiation and induced current), and SCSI signal line noise. All except the SCSI signal line noise and the terminator power noise are handled by the storage system design and therefore are not directly part of the advancements in parallel SCSI. The SCSI terminator power noise is determined by the size of the decoupling used on the SCSI terminators and the size of the capacitance on the device being inserted. This noise is easily controlled by ensuring that these sizes meet the values specified in the SPI standard.[4]

The delicate case is when the SCSI signal lines are involved, which is the subject of this section. To determine the nature and magnitude of these signal line disturbances, one must understand the following three mechanisms: (1) the overall sequence of events, (2) the electrical dynamics of connector contacts when used in the SCSI application, and (3) the electrical consequences on the bus when the device makes/breaks contact with the SCSI signal line.

There are two sequences of interest: insertion and removal. The removal process is easy to grasp after the insertion process is understood.

Single-ended Device Insertion

The initial conditions considered for SCSI device insertion assume a SCSI device with its ground solidly and continuously connected to the ground of the SCSI bus. This connection is easily accomplished, for example, by using sequenced contacts where the device ground makes connection well before any signal connection. In this state, the SCSI device pins present a maximum fully discharged capacitance of approximately 25 picofarads (pF). After the device signal pin contacts the bus, this capacitance becomes charged (by extracting charge from the bus) to the voltage on the signal line at the time of the insertion.

These values range from approximately 3 V for negated lines to nearly 0 V for asserted lines.

Since the SCSI device being inserted is logically off (i.e., there is no driver current), the only current that needs to flow is that required to charge the 25-pF capacitance. This is sharply different from many connections in electronics in which current flows through the contact after an electrical contact has been established.

In the case where no bus voltage changes occur except as a result of the device insertion, the insertion transient begins with the initial contact and ends when there is no further bus voltage change with time (the steady state voltage). Once the device pin voltage reaches the steady state bus voltage, no further current flows through the contact.

Therefore, once the device capacitance becomes charged to the steady state signal line voltage, no further disturbances to the signal line voltage will occur even if the contact opens momentarily during a chattering event. The voltage on the device capacitance changes during the transient from a discharged state (zero voltage) to the steady state signal line voltage, with the current always flowing into the device capacitance.

If the signal line voltage changes after the insertion transient is completed (because of events such as being driven by other devices, by noise, or by the inserting device beginning to use its own driver), then current will again begin to flow through the contact. This is a normal SCSI condition for contacts in service. If the signal line voltage changes during the insertion transient because of events other than the connector contact effects (e.g., signals changing because of being driven by other devices, other noise), then it is more difficult to determine exactly where the insertion transient ends. The beginning of the insertion transient will still be marked by a charging of the device capacitance. Examples presented later in this paper show both insertion events and driving events from other devices occurring at the same time.

The time required for complete contact mating on all SCSI signals in the bus is up to six orders of magnitude greater than the time required for a SCSI signal to change state. Therefore, signal level changes are likely during the insertion process. The electrical behavior of the contact as it continues wiping (sliding after initial contact is made) from its initial contact point to its final resting position becomes a critical part of the process. The following subsections explore this behavior in detail.

Connector Insertion Dynamics The data presented in this section were derived from a DIGITAL DSSI bus in 1990. The DSSI bus is nearly identical to the SCSI bus, and many of the results apply without modification to SCSI. Similar data have been observed on the SCSI bus, but the complete set of data presented in this paper is not presently available from actual SCSI hardware. The disturbances in the DSSI bus are larger than those seen in the SCSI bus, because the DSSI voltages are slightly higher (3.5 V for DSSI compared to 2.8 V for single-ended SCSI), and the instrumentation capacitance (~10 pF) adds significantly to the device capacitance because of the state of the art for scope probing in 1990. Numerous tests with modern scope probes (0.6 pF or less) of SCSI hardware have shown that the SCSI disturbances are indeed qualitatively the same but significantly less in size than those shown here from the DSSI hardware.

The mechanisms described apply to any system in which the insertion transient is caused by the charging of a small capacitance. Figure 14 shows the basic test setup. A device is inserted into a connector with scope probes attached on either side of the mating interface and with an additional probe attached to the bus some distance from the connector. The voltage on the device side of the connector is used as the trigger signal into a digital storage scope so that the events before, during, and after the mating event can be examined. This is clearly a single-event type of measurement, so a high sampling rate (1 billion samples per second) and significant scope memory is required to capture the waveforms. The scope probes used have a 1-megohm input resistance.

The connector used for the tests in this section has multiple parallel pins that all mate and demate in the same general time period. There is no intentional difference in the pin lengths. The time relationship between the mating events on two neighboring pins was explored first. By choosing neighboring pins, the differences between the pins is kept to a minimum so the time differences observed should represent the best pin-to-pin synchronization in a mating event.

For this test, a probe was attached to each of two pins, and the connector alone (not part of a device) was mated to the bus segment connector. Figure 15 shows the results.

Both pins appear to show instantaneous transitions between the charged and discharged states on the time scale that was required to capture both events on the same plot. The mating events are separated by approximately 19 milliseconds, and there is no evidence of any discharging after the initial charging has occurred. Since the scope probes have a 1-megohm input resistance, any lack of contact during the wipe portion of the mating will allow the capacitance to discharge through the probe with a time constant of approximately RC, where R is the scope probe resistance and C is the sum of the connector pin and probe capacitances. Assuming a total of 10 pF, this gives a decay time constant of 10 microseconds.

Figure 16 shows another mating event on pin 1 at a 500 times more sensitive time scale. In this case, some evidence of momentary opens is seen with the

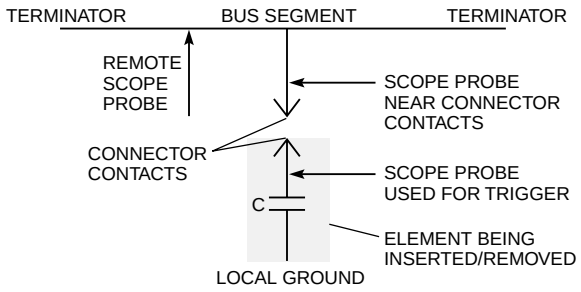


Figure 14
Test Setup for Insertion/Removal Transients

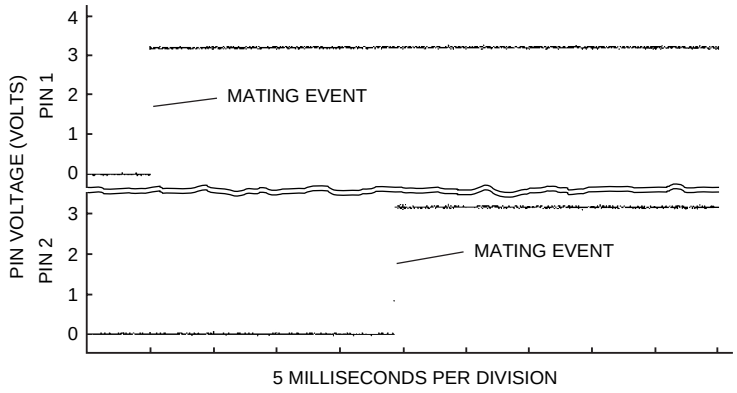


Figure 15
Time Relationship between Mating Events for Two Pins in the Same Connector

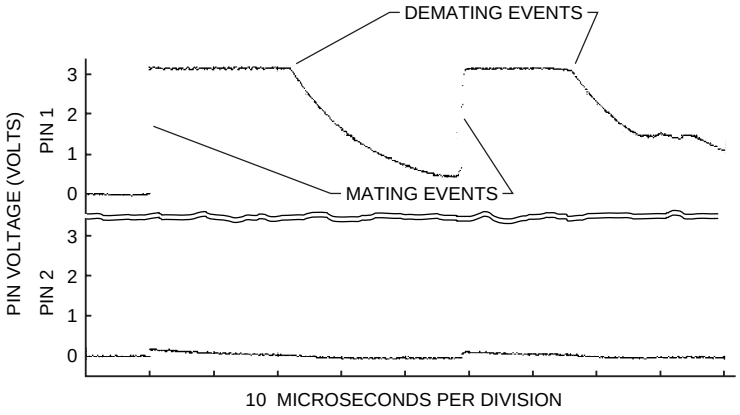


Figure 16
Contact Bounce Events

expected decay dynamics. The actual time constant is a bit longer than 10 microseconds because of some capacitance in the connector pin. This bounce behavior may or may not be present during the initial stages of the event shown in Figure 15, but clearly the behavior is not visible in the figure. To observe the suite of transients that exist in the mating process, one must examine the transients at several different time scales. In general, this requires repeating the mating events, since the dynamic range of the scopes used was insufficient to capture all the detail in a single event.

The initial mating event on pin 1 still appears to be instantaneous on the time scale used in Figure 16, but some slope is visible in the second bounce event. Also, during the second decay period, a shelf in the decay indicates that a partial, high-resistance contact was briefly experienced. Pin 2 is not close to making a contact at the time range shown in Figure 16. The figure shows a small amount of cross talk in the pin 2 voltage waveform caused by the pin 1 transients.

This data clearly shows that the details of the mating process are highly complicated and intrinsically

unpredictable. Therefore, the best we can hope for is to establish some limiting cases for the important parameters. The limiting features shown in Figure 16 are the extremely rapid initial mating event and the decay times. We examine these rapid transients in detail later in this section. The decay times are determined by the actual contact resistance and the resistance of the leakage path to local ground. For normal SCSI devices, there is very little leakage to ground on the device pin so the opens produced by the bounce have no effect.

Some cases observed indicate much more complex bounce structures. Figure 17 shows a case in which the mating connection is not established until more than 700 microseconds have passed.

The data in Figures 15 through 17 were all acquired from the same connector contact during separate mating processes. Typically, the details of the mating event are very different even under nominally identical conditions.

Another type of mating event is shown in Figure 18. This event requires approximately 10 microseconds to make the transition from uncharged to charged, and there is no bounce. This particular event produces almost no cross talk into pin 2. Events with these characteristics are somewhat rare and are called gradual transients in this paper.

Figure 19 shows a closer look at the rapid transient type of mating event. In this figure, we have added a device capacitance of approximately 20 pF to the scope probe for a total of approximately 30 pF. Notice that the transient requires 2 to 3 ns to substantially complete its charging. There is a ratio of nearly 10^7 between the mating events on different pins in the same connector and the rapid transient of a single contact.

Limiting Parameters for the Rapid Transient The question of whether the rapid transient shown in Figure 19 is the worst case needs to be explored because the duration of the transient affects the disturbance on the bus. Some bounding features and some implications of the observed behavior of the rapid transient are noted in this subsection.

Assuming that the transient event occurs in 2 ns and that the velocity of impingement just prior to the first mating event is 1 meter per second, then the distance traveled by the contact would be 2 nanometers (nm).

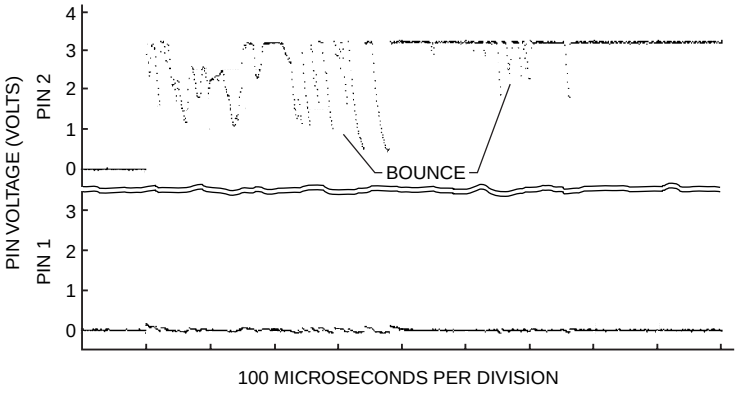


Figure 17
Extended Mating Bounce Events

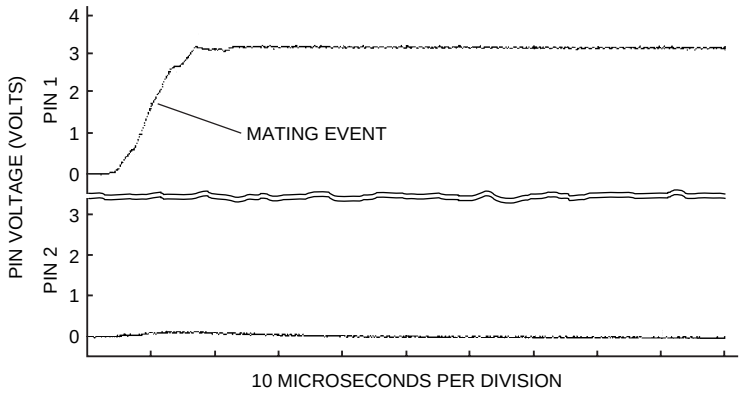


Figure 18
Gradual Mating Event, No Bounce

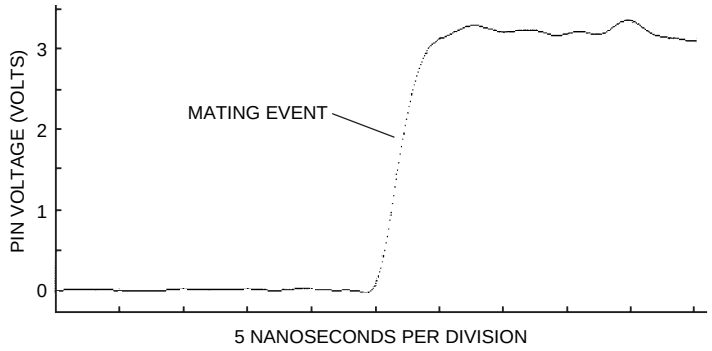


Figure 19
Detailed Structure of the Rapid Transient

This distance is equivalent to a few atomic distances. The distance traveled during the gradual transient shown in Figure 18 would be approximately 10 microns, and during the extended bouncy case shown in Figure 17, approximately 1 mm. The velocity for the latter two cases would likely be somewhat reduced because of the mechanical interference between the pins, and the actual distance traveled is probably significantly less. There is little opportunity, however, for the velocity to be reduced for the rapid transient, and this distance of 2 nm is probably at least the correct order of magnitude.

The following calculation shows the total current levels required to charge the capacitance in 2 ns.

$$Q = CV = 30 \times 10^{-12} \text{ pF} \times 3.5 \text{ V}$$
$$= 10.5 \times 10^{-11} \text{ coulombs},$$

where Q represents the total charge, C is the capacitance, and V is the voltage. Since this charge is transferred in a time t of 2 ns, the average current is

$$Q/t = 10.5 \times 10^{-11} \text{ coulombs}/(2 \times 10^{-9} \text{ ns})$$
$$= 52.5 \text{ milliamperes (mA)}.$$

For a gradual transient that takes 10 microseconds, the average current is approximately 10 microamps. These calculations show that the most severe amplitude disruption to the signal on the bus occurs with the rapid transients, since relatively large current must be supplied in a short time to charge the capacitor.

The next item to be examined is the current density that must exist during the transient. Since the contacts move only 2 nm and the surface finish of actual contacts is not nearly this smooth, it is reasonable to assume a square 2-nm contact. Clearly, this assumption is not rigorously defensible and could be the subject of an entire study area in its own right; however, there is no basis for assuming that the lateral contact region would be any different than the contact area in the mating direction. The basic conclusions would not be affected even if we assumed a hundredfold lateral increase in contact area. Attempts to use scanning electron microscopy to examine the actual contact area were not fruitful in establishing the actual initial physical contact area because of the severe physical disruption that occurs on the microscopic level and because of the small sizes involved.

Under these assumptions, the physical contact area is assumed to be $(2 \text{ nm})^2$ or 4×10^{-14} cm^2 in the following calculations. The current density to support the 50-mA rapid transient current is therefore approximately 10^{12} A/cm^2. Typical current densities in copper and other metals are less than 10^6 A/cm^2. The electromigration onset current is of this same order. The current density in the rapid SCSI transient is a million times greater than that which metal can support.

To support the massive current density, the actual contact area must be much larger than the initial physical contact area assumed in the above calculations. The author believes that this can be explained by a micromolten metal-to-metal joint that is formed upon initial contact and that the front of the melt propagates (probably through phonon interaction) at approximately the speed of sound in the metal. This process would create crudely a thousandfold increase in the effective insertion velocity and would result in a millionfold increase in contact area, since the melt would propagate in all directions.

This mechanism would produce reasonable current densities and would form an intimate metal-to-metal interface with both contacts that would aid in reducing the contact resistance. The micromelt size becomes rapidly self-limiting, with the expanding contact area causing decreased current density, which in turn, causes decreased melt temperature.

As discussed in the next section, the actual contact resistance during the rapid transient cannot be large. If this resistance is large, as in the case of the gradual transient, the mating event is much less disruptive.

Many variations on the mating transients can be observed, but we do not attempt to show all of them

in this paper. One special variation, however, is worth noting—the combination of rapid and gradual transients in the same mating process. Sometimes the mating process starts with a gradual transient and then shifts to a rapid transient. Figure 20 shows a complex mating process in which (1) a gradual transient initiates, (2) a rapid transient starts but does not complete, (3) the rapid transient ends, (4) another gradual transient process starts, and (5) another rapid transient finishes the charging process.

This observation is consistent with several possible microprocesses during which the initial rapid transient extinguishes before completion.

- The micromelt becomes physically torn apart by the advancing motion of the contacts. (This process is unlikely because of the excessively slow physical motion.)
- The micromelt explodes. (This process is likely.)
- The micromelt becomes resistive through the contamination of the melt with insulating material.
- The micromelt front reaches a thin region and opens because of the lack of material.
- The micromelt front reaches an insulating region.

On further movement of the contacts, a new rapid transient condition is encountered between different metallic peaks of the contacts, and a new rapid transient begins. Figure 21 shows a conceptual representation of this process.

Gradual transients appear to be associated with normal current densities (i.e., 10^6 A/cm^2) and much higher contact resistance than rapid transients. In cases where a micromelt is not initiated, the low contact resistance associated with the liquid metal–to–solid metal interface and the expanded contact area are not present. Therefore, one way to eliminate the mating disturbance caused by the rapid transients is to ensure that a micromelting process is not possible.

In the process shown in Figure 20, it is probable that a gradual-type contact is being maintained somewhere else in the contact, since no voltage decay is evident when the rapid transient ends. Indeed, it is to be expected that the rapid transient mechanism would not operate after the capacitance is charged to a certain level, since there would not be enough energy difference to initiate and sustain a rapid transient. Therefore, the gradual transient is the behavior derived from an extrapolation of the normal mechanisms that produce contact resistance. This detailed discussion is pursued because we must understand the basic physical mechanisms to gain confidence that we are considering the worst-case disturbances.

Single-ended Device Removal

During the process of removal, the device pin separates from the bus. Since both the bus and the device are at the same voltage just before the separation, no current is flowing unless the bus voltage changes when the contact is in the process of separating. Therefore, in most cases the separation process causes no disturbance.

Bounce can occasionally be observed during the demating process when there is a leakage-to-ground path present on the device side. Of course, if a voltage decay occurs and the contacts re-connect, the mechanisms are essentially the same as for the insertion transient. The key point is that no additional mechanisms have been noted for device removal that could be more disruptive than those operating during the insertion process. In the limit, the removal process could produce as much disruption as the insertion process.

Figure 22 shows two examples of demating. The demating events shown in Figure 22a have only approximately a 60-microsecond separation. This separation is exceptionally small, and it is theoretically possible to have coincidental contact-to-contact events (within the precision of the instrumentation). The demating event with bounce shown in Figure 22b was acquired on exactly the

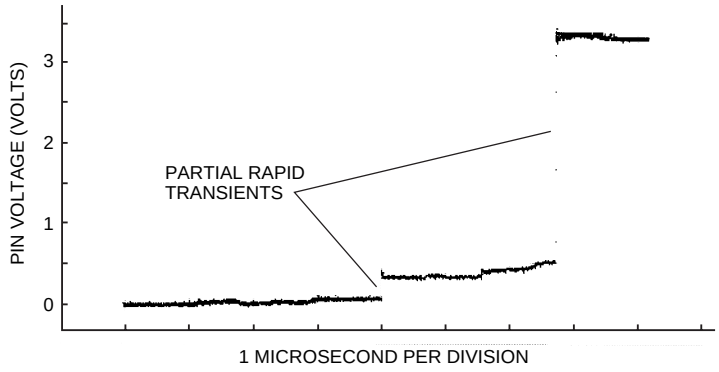


Figure 20
Gradual and Rapid Transients in the Same Mating Process

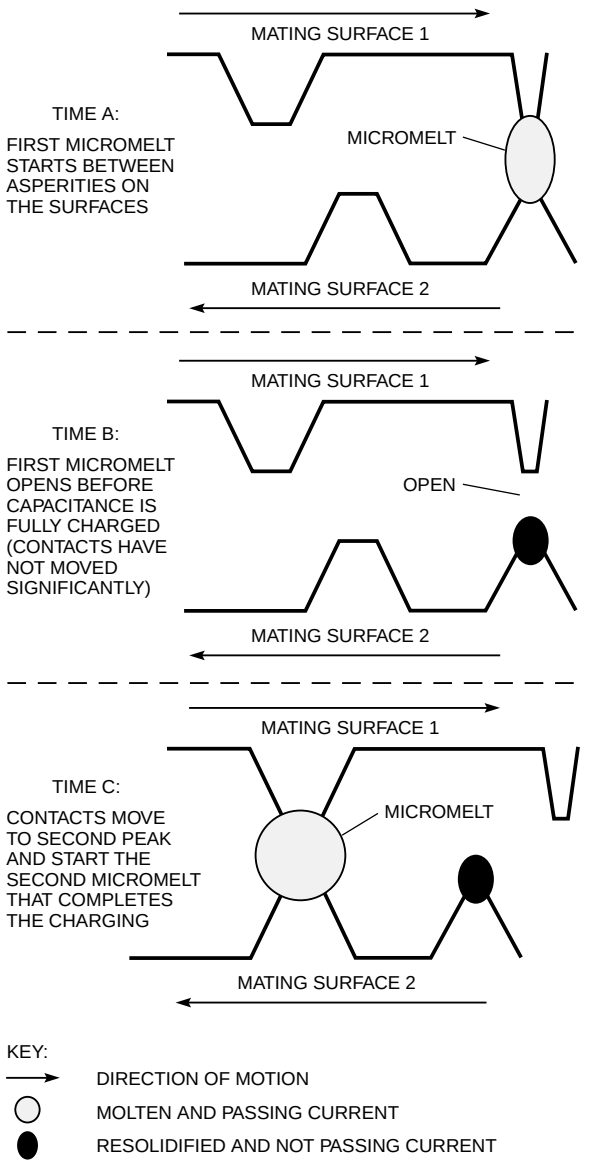


Figure 21
Architecture of Combination Gradual/Rapid Mating Event

same pins in exactly the same connector used for the events in the top of the figure, and there is no evidence of any activity on pin 2. Pin 2 demated long before any activity was seen on pin 1. Again, this underscores the unpredictability of the details of any given event.

Impact of Device Insertion and Removal on Bus Signals

This section contains several examples of the noise produced on the bus side of the connector. Actual devices with approximately 25 pF of capacitance were used to obtain the data. This capacitance value is increased by the probe capacitance. On the bus side, there is also some increased capacitance caused by the probe used to acquire the bus side signal. Figure 23 shows the basic impact of a rapid transient on the bus side of the connector and the time relationship of the bus disturbance to the voltage on the device side. The bus voltage is reduced while it supplies the necessary charge to the device pin. After the device capacitance is charged, the bus resumes its voltage level before the insertion transient (more or less).

In this test, the bus pulse is approximately 3-ns wide at its midpoint; its peak amplitude is approximately 1.25 V. This pulse is significantly larger in amplitude than that produced from a device alone.

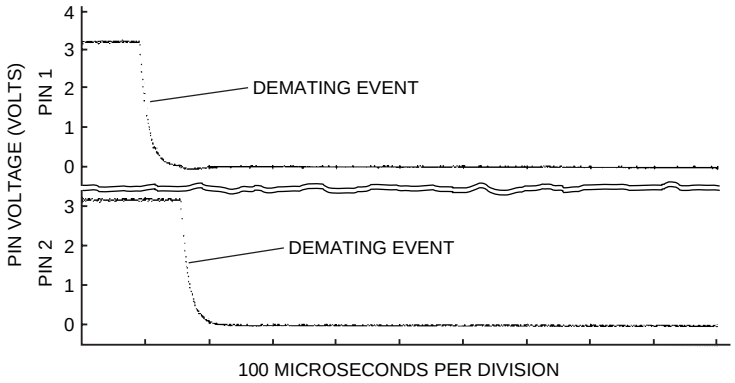
One of the more interesting features of the signals in Figure 23 is the lack of commonality or tracking in the signals after the rapid transient has passed. In the simplest interpretation, one would expect both sides of the connector to have nearly the same voltage (at the least to be within the accuracy of the 0.1-ns propagation time between the probes). The following discussion addresses the author's current thinking on the reasons for this lack of tracking.

Instrumentation effects, such as resonance or differences in probe properties, were ruled out by using both probes on the same signal and noting that there was little difference in the signals reported from each channel. Later, typically after a few microseconds, the voltages do become effectively the same.

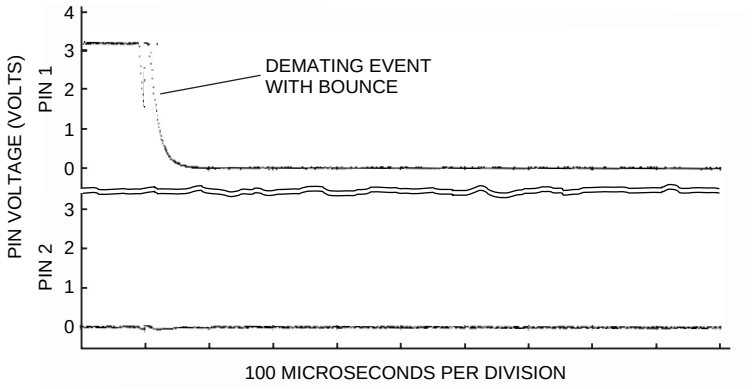
Because a significant voltage difference is present for relatively long times, there must be a significant voltage source between the contacts to support this observed difference. In the initial stages, the difference between the pin voltages is approximately 3 V. If the current is the one calculated in the section Limiting Parameters for the Rapid Transient, that is, approximately 50 mA, then the current-limiting impedance must be at least $3/0.05 = 60$ ohms. This impedance, coupled with the parasitic capacitances and inductances, serves to blunt the instantaneous electrical energy transfers that would be implied by a very low source impedance. If the source impedance were very low, then both sides would have to track shortly after the initial contact.

Part of this limiting impedance is the loaded or local transmission line impedance of the bus. The characteristic impedance is nominally approximately 100 ohms for an unloaded bus. Since the bus connector is attached to the middle of the line, both sides are available to supply charge and the effective charging impedance would be approximately 50 ohms. A 30-pF capacitance would have a charging time constant of 1.5 ns. This time constant fits the observations well during the rapid transient itself but does not fit the timing parameters of the voltage differences observed well after the rapid transient.

Elevated local temperatures are almost certainly present during the rapid transient (near the melting point of the metals!), so it seems plausible that the mystery voltage source is basically thermal electromotive force (EMF) between the pins. Allowing a few microseconds to achieve thermal equilibrium and subsequent loss of the thermal EMF also seems quite plausible. These details are inviting further detailed investigation but



(a) Demating Event with 60-microsecond Separation



(b) Demating Event with Bounce

Figure 22
Demating Events

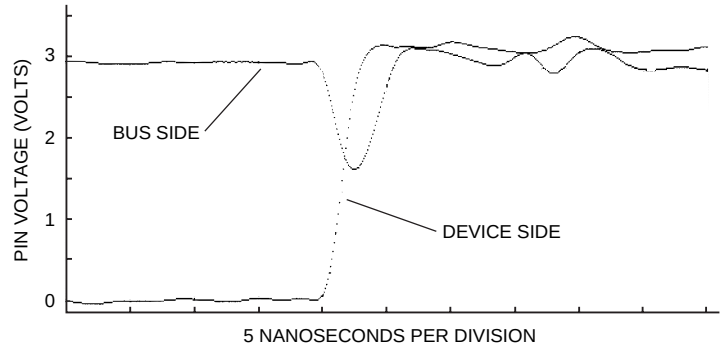


Figure 23
Most Severe Noise Pulse Observed

do not affect the practical conclusions as applied to parallel SCSI.

As added evidence for thermal effects, experiments with early LVD SCSI devices that use a 1.2-V bus level instead of the 3.5-V bus level shown in Figure 23 transfer much less energy and have a much shorter settling time before both sides of the contact track. These LVD results will be reported separately.

The point extracted from these charging-impedance and settling-time observations is simply that the overall energy transfer rate is limited by the microphysics of the process. This means that Figure 23 almost certainly illustrates the worst-case disturbances.

It has been noted that the bus pulse is similar to that produced by a stub on the bus and a signal with a fast rise/fall time. In a sense, we really are charging a stub in either case,

and in both cases the loaded or local characteristic impedance of the bus limits the extent of the disturbance.

To more accurately measure the noise pulse produced when a device is added to the bus, measurements were performed without a scope probe attached to the device pin. To do so required triggering the scope from the noise pulse on the bus side. Consequently, it was not possible to see the device-side charging dynamics. Figure 24 shows the measured pulse near the device connector and at a point 2 meters away.

The pulse measured in Figure 24 has approximately half the amplitude of the pulse in Figure 23. This is more reduction in amplitude than one would expect from the removal of 10 pF from the effective device capacitance, and this difference, while not completely explained, is in the favorable direction. The noise pulse that reached the next device (where it could be detected as an error) would be even smaller, because of the dispersion and attenuation in the bus and because the neighboring device would need to have its 25-pF capacitance charged also. The signal at the measurement point 2 meters away in Figure 24 indicates the intensity of the attenuation and dispersion to be expected in the rapid transient bus pulses. The details of the attenuation and dispersion depend somewhat on the bus media used.

The rapid transient bus pulses are shown on actual data pulses in Figure 25. The top trace in the figure shows a rapid transient pulse on a negated part of a single-ended SCSI signal. There is a scope probe on this device, but the device capacitance is only approximately 15 pF so the total with the probe is approximately 25 pF. Note that the noise pulse is approximately 0.8 V and does not take the signal into the receiver detection range below 2 V. This negated state is a bit higher than usually found, so the bus pulse is starting from a higher point. If the pulse had started from a lower point, for example about 2.5 V, the pulse amplitude would not have been as large. Further discussion of the receiver detection range appears later in this section.

The signals in Figure 25 were purposely chosen to have broad falling edges of approximately 15 ns. Normal SCSI signals are 5 ns or faster. The broad edges maximize the chance that the bus pulse will produce a signal slope reversal of the type that can produce multiple edges. The bottom trace in Figure 25 shows a bus pulse in the most sensitive part of the falling edge. This pulse produces almost no slope reversal because by the

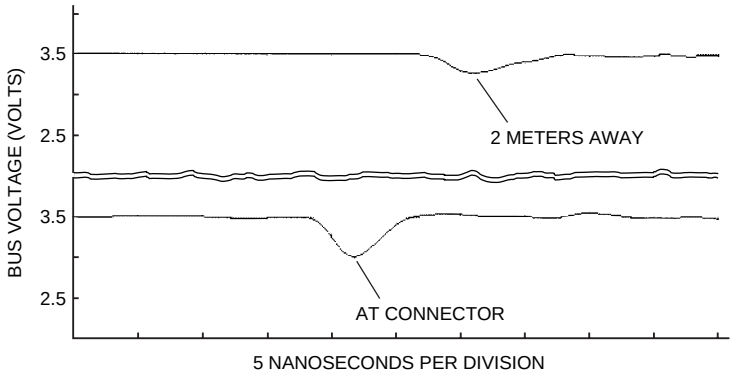


Figure 24
Bus Pulses with No Scope Probe Attached to the Device

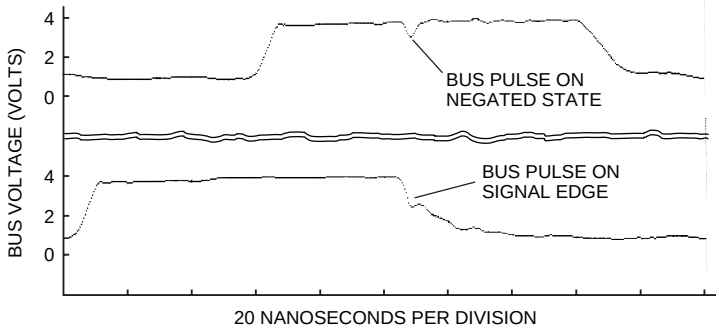


Figure 25
Bus Pulses on Actual Signals

time it is ready to become positive-going, the data signal has fallen so much that there is no voltage source to drive the signal more positive. At the beginning of the falling edge, the slew rate is increased by the bus pulse; in the middle, the edge is extended and consequently the overall time required for the falling edge is almost exactly the same as for the falling edge that has no bus pulse (see the top trace).

Therefore, the main effect of rapid transient pulses occurs when they intersect the signal edges (where a state change is expected anyway), and the effect is movement of the position of the edge by no more than 2 ns from the normal position. This movement is already accounted for in the SCSI standard as *pulse distortion skew,* so there is no important effect.

If the mating event happens while the bus signal is in the asserted state, there is little effect since little charge is transferred. If the event happens in the rising edge, there may not be enough voltage difference to start a rapid transient—again, there is little effect. If a rapid transient is initiated on a rising edge, the impact is still a small shift in the position of the edge. In any arbitrary combination of signal level and type of transient, the bus disturbance will not be greater than those shown in Figure 24 and Figure 25.

Differential

For differential SCSI systems, essentially the same behavior occurs as for the single-ended case except that the relationship between two contacts instead of just one must be considered. If insertion transients on the positive signal differential line are occurring at the same time as transients on the negative signal line, we must examine the difference between these transients to see what impact they have on the differential signals. Based on the time required between mating events on neighboring connector pins presented in the section Connector Insertion Dynamics and in Figure 15, it is evident that the differential case is almost always two independent and isolated single-ended cases. This is because the difference in the time required for different pins in the same connector to begin the mating process vastly exceeds the actual transient time on either signal.

In SCSI differential systems, both the positive and the negative signals are normally positive with respect to the local grounds. This means that the transients will be the same polarity on both signals.

In the very rare cases in which some overlap exists between the transient times on both signals, the rapid transient disturbances would usually be seen as common mode events that reduce the effective differential transient signal. These events are not seen if common mode noise exists where the signals have opposite polarity with respect to local grounds during the transients. In this case, it is theoretically possible to produce anticommon mode differential transients. However, the anticommon mode case will always have the positive and negative signal lines within a differential logical voltage level of ground, and the transients will therefore be small. Even in the anticommon mode case, the effect is at most a slight shift in the time when the differential state change is observed, since the transient disturbances are so small.

In the pathological differential case, large common mode levels exist on both the positive and the negative signals. The insertion transient will be larger because the bus voltage is larger. This case is even more rare since it requires both coincidental pin mating and coincidental large common mode.

The other case considered that can have a unique effect on differential systems is that of extended bounce. This case extends the effective mating time to the point when some overlap between the transient activity on the pins is more likely. Recall that the extended bounce case was only visible when a leakage mechanism was available to discharge the incoming device capacitance. In actual devices, no significant leakage occurs so a bounce event does not produce disturbances after this first contact.

The differential signal seen by the incoming device may be seriously affected by extended bounce if there is bus activity during this bounce. Consider, for example, a case in which the positive signal contact opened because of a bounce event after achieving a full charge. While it is still open, the negative signal changes state. Now both the positive and negative signals are at the same nominal potential, which is an indeterminate differential condition. Fortunately, this condition is not a problem because the only device that sees this condition is the device being inserted or removed and it is not in an operational state.

Summary of the Handling of Device Insertion and Removal Transients

After a complex, yet self-consistent, set of experimental data and interpretations, the concluding results are that the worst-case SCSI bus transients resulting from proper insertion and removal processes should not cause errors in the SCSI bus as presently specified in the SPI and Fast-20 (UltraSCSI) standards. The proper processes include pregrounding prior to insertion, avoiding excessive device capacitance, and using SCSI drivers and receivers that meet all of the SCSI requirements.[4]

As of this writing, all reports of device insertion/removal errors have been traced back to failure to use proper procedures or designs. The most common errors are lack of pregrounding, devices that do not maintain the high-impedance input state during power cycling, and power distribution or mechanical transient effects unrelated to SCSI proper.

The mechanisms that operate span a time spectrum from picoseconds in rapid transients to seconds in contact wipe and other macro connector operations.

The worst-case differential transients occur when one treats the differential system as two independent single-ended SCSI buses—one for the positive signal and one for the negative signal.

The rapid transient becomes more and more detectable as bus speeds increase and the receivers and timing margins become more sensitive. Schemes to encourage the gradual transient are the best protection against the ultimate problems caused by rapid transients. The best-known method for producing reliable gradual transients is to avoid a metal-to-metal contact during the initial contact and until the device capacitance is charged. At this time, no such connector system exists for SCSI applications.

Overall Summary

Evolution in four significant hardware technologies in the recent past has enabled parallel SCSI to break through the barriers that were preventing it from delivering excellent value, flexibility, and growth to the computer data storage industry. Application of more scientific methods, use of the latest silicon technology, and developments in the interconnect technology provided the foundation for these improvements. DIGITAL provided most of the basic data and led important standards and industry bodies to accomplish this.

Acknowledgments

Fee Lee, Keith Childs, Chuck Bagg, Pak Seto, and Jonathan Salles were instrumental in developing the special test environment and for acquiring much of the data presented in this paper. The author gratefully acknowledges the management and technical support from Pete Korce, Mike Chamberlain, Bob Passmore, Richie Lary, Ken Chester, Bob Rennick, Laura Woodburn, and Ellen Lary.

References

1. *ANSI X3.277-1996: Information Technology X3T9.2/375R SCSI-3 Fast-20, X3T10/1071D* (New York: American National Standards Institute, 1996). Available from Global Engineering, 15 Inverness Way East, Englewood, CO 80112-5704, tel. (800) 854-7159 or (303) 792-2181, fax (303) 792-2192.
2. Very High Density Cabled Interconnect (VHDCI) Specification, SFF-8441. Available from the SFF Committee, 14426 Black Walnut Court, Saratoga, CA 95070, voice fax-back service (408) 741-1600, tel. (408) 867-6630 ext. 303, fax (408) 867-2115.
3. *ANSI X.3.131-1994: Information Systems—Small Computer System Interface–2* (*SCSI-2*), X3T9.2/375R (New York: American National Standards Institute, 1994). Available from Global Engineering, 15 Inverness Way East, Englewood, CO 80112-5704, tel. (800) 854-7159 or (303) 792-2181, fax (303) 792-2192.
4. *ANSI X3.253-1995: Information Technology—SCSI-3 Parallel Interface* (*SPI*), *X3T10/855D* (New York: American National Standards Institute, 1995). Available from Global Engineering, 15 Inverness Way East, Englewood, CO 80112-5704, tel. (800) 854-7159 or (303) 792-2181, fax (303) 792-2192.
5. *SCSI Enhanced Parallel Interface* (*EPI*), *X3T10/1143D* (New York: American National Standards Institute, 1997). Available from Global Engineering, 15 Inverness Way East, Englewood, CO 80112-5704, tel. (800) 854-7159 or (303) 792-2181, fax (303) 792-2192.
6. Information about I/O interfaces is available at the T10 home page at http://www.symbios.com/t10.
7. Single Connector Attachment–2 (SCA-2) Specifications: SFF-8015, SFF-8046, SSF-8048, SSF-8066, and SSF-8451. Available from the SFF Committee, 14426 Black Walnut Court, Saratoga, CA 95070, voice fax-back service (408) 741-1600, tel. (408) 867-6630 ext. 303, fax (408) 867-2115.

Biography

William E. Ham
Bill Ham joined DIGITAL in 1983 and has been working in storage technology since 1990 as the manager of the Storage Bus Technical Office (SBTO). The SBTO group represents DIGITAL at most of the important standards and industry bodies that are involved with the transmission of storage data. Presently, these groups are SCSI, STA, Fibre Channel, and SFF. The SBTO group also creates information from actual laboratory testing and was responsible for introducing Fast-20 technology to ANSI in 1993 and hot-plugging technology to SCSI in 1992. Bill has a Ph.D. in electrical engineering from Southern Methodist University in Dallas, Texas, and has been active in the electronics industry in a variety of technologies since 1970. He has held significant positions in silicon chip, printed circuit board, multichip module technology, and storage bus technology. In addition, he is the past technical editor of the SSA physical standards, past editor of the new SPI-2 SCSI standard, editor of the entire family of SFF connector documents, editor of the new Enhanced Parallel Interface ANSI project for SCSI, and secretary for the Fibre Channel physical standards group (T11.2).