ATA line termination issues on systems with high speed (Local Bus) interfaces

By

Intel, Maxtor, Quantum, Seagate Technology

3/17/95
Ver 3.0

Introduction

New PC motherboards and interface cards supporting local bus interfaces (both VLB and PCI) are running devices (disk drives under ATA, CD-ROMs and Tape Drives under ATAPI) up to the maximum transfer rate specified by ATA mode 3 (11.1 MB/s max) and mode 4 (16.67 MB/s max). The ATA Bus is currently specified as an unterminated bus. This lack of termination causes ringing on the signals traveling between the host and the device (see figure 1). Previous bus speeds were sufficiently slow that this ringing was not a problem, but higher bus speeds - ATA modes 3 and 4, are now being exercised by local bus interfaces. This could lead to a system failure if the signals are not properly terminated.

These transmission line effects are complicated by the support of dual port ATA in newer systems. Intended to increase the number of devices from 2 to 4 by utilizing the secondary address, cost pressures have lead designers to deviate from the straightforward technique of simply reproducing the system ATA logic used on the original port. Sharing of these system resources, particularly pins, can lead to further signal quality problems if proper precautions are not taken.

ATA Bus

The ATA interface specifies an 18-inch, 40-conductor ribbon cable as the means for connecting devices to the host interface. Up to two devices may be connected to the cable. The most common (and recommended by the ATA-2 standard) configuration is for the host to be at one end of the cable with a device located at the other end of the cable, up to 18 inches from the host. If a second device is configured, then it should be located within 6 inches from the device at the end of the cable.

The ATA cable is not specified as having any termination. With slow TTL edge rates this has not been an issue in the past. The advent of CMOS VLSI bridge

chips has increased the edge speeds on the bus, causing ringing and crosstalk problems. A long-time rule-of-thumb for analog designers has been that if the propagation delay of the cable exceeds one-quarter of the rise time then cable termination must be used. In the case of the ATA bus, the worst case propagation delay of the cable is approximately 4 ns, so rise times of less than 16 ns could cause problems. Many local bus to ATA bridge chips available today have edge rates of 1 to 2 ns, despite the ATA-2 standard requirement of 5 ns.

RF circuits solve the issues of signal reflections and ringing by driving the cable with a source impedance equal to the characteristic impedance of the cable, and terminating the far end of the cable with this same impedance. ECL circuit designers solve the problem by terminating only at the receiving end. Unfortunately, both of these cases assume a single source and a single termination. The SCSI interface designers understood this issue and specified that the physical devices at each end of the cable must be terminated. This has the disadvantage that the user must know the physical configuration and must install (or remove) the appropriate termination.

One of the innovations in ATA technology that some suppliers are offering system customers is dual port - using the primary and secondary ATA disk controller addresses to allow the connection of up to 4 devices in a system, two per cable. This expands the number of devices (e.g. 2 ATA disk drives, 1 ATAPI CD-ROM, 1 ATAPI tape drive), and in practice makes ATAPI CD-ROM usage possible.

Meanwhile, the electrical specifications for ATA are all based on the assumption of 1 or 2 devices per cable, and cables limited to 18 inches in length.

As conceived by the ATA and ATAPI standards, the new dual port feature works with these electrical limitations since two distinct cables are still used, each obeying the old electrical specifications.

ATA Bus Termination

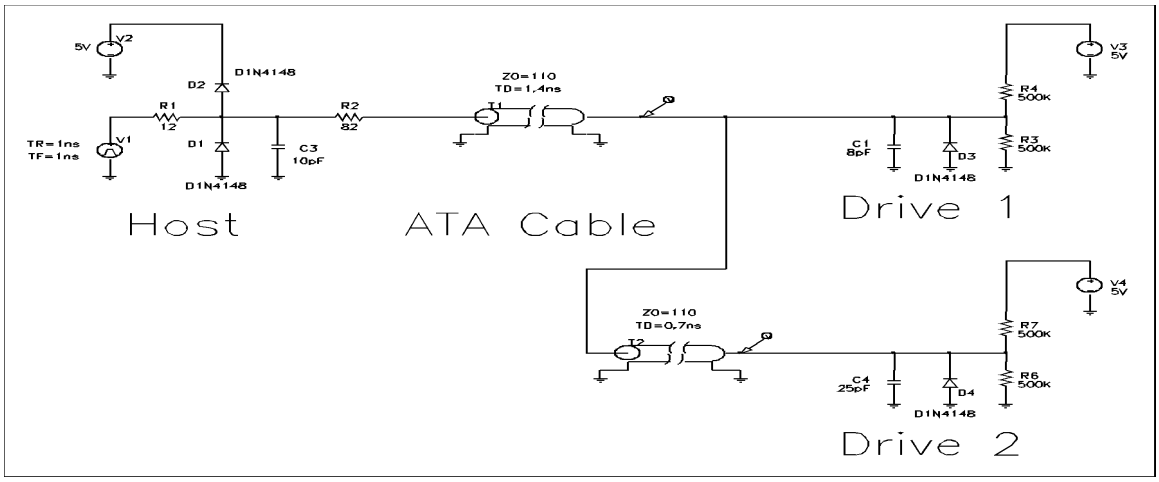

Figure 1 — Terminated host with two different devices

The ribbon cable commonly used for the ATA bus has a characteristic impedance of approximately 110 ohms. Theoretically, the optimum solution would be to drive the cable with a 110 ohm source and ensure that the device at the end of the cable was terminated in 110 ohms. One problem with this solutions is that when a 110 ohm cable is driven with a 110 ohm source, the voltage is divided by two. Another problem is how to terminate the end of the cable and not have a termination in the middle when the user decides to install a second device.

There is no technically “proper” solution to this problem. Redefining the ATA bus to use termination would lose backward-compatibility which is one of the strong features of the ATA bus. The only recourse available to system and device designers is partial termination at both the host and the device. No one solution at a single point will work under all circumstances.

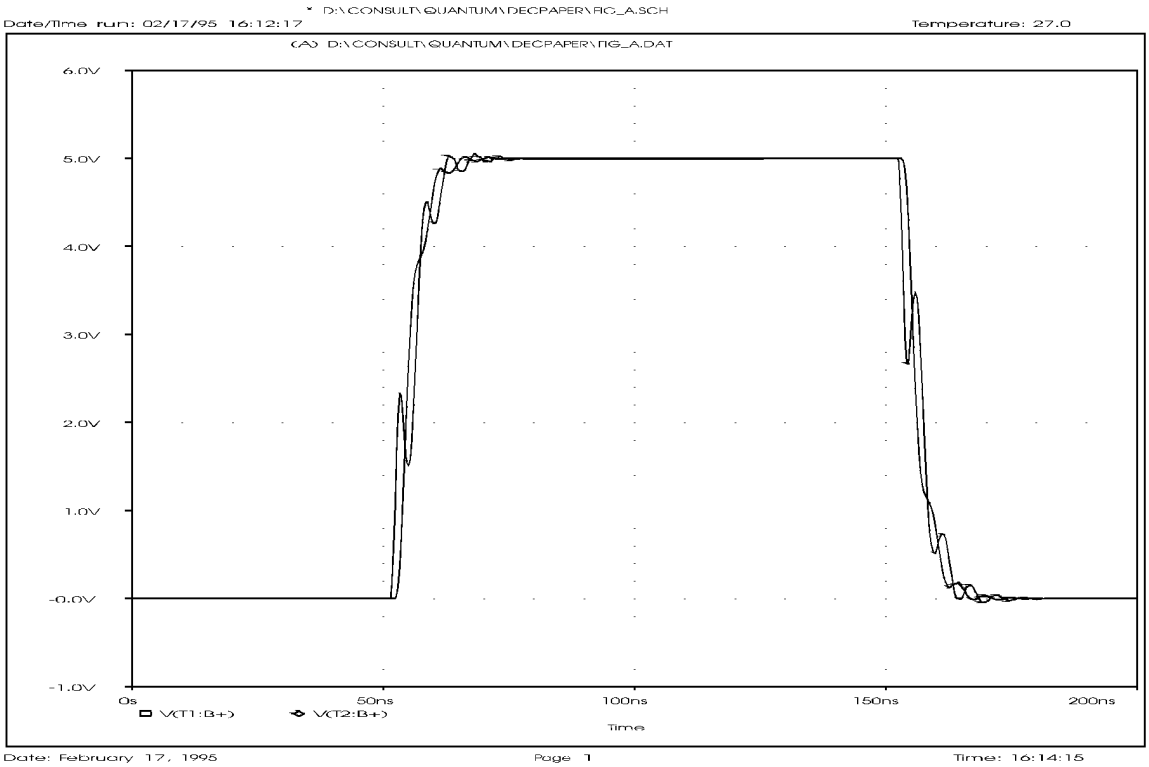

Figure 2 — Waveform at device inputs

One of the common misconceptions is that limiting the slew rate of the source to 5 ns will fix the problem. The rule-of-thumb used by analog designers tells us that we need 16 ns edges to truly eliminate the ringing problem. The controlled slew rate helps, but does not solve the problem by itself.

Another misconception is that source termination is sufficient to eliminate the ringing problem. This is not the case. Simple simulations with a host and a single device may give this false impression, but a more complete simulation with two devices reveals the problem. Figure 1 shows a SPICE model with a source, 12 inches of cable, a device, 6 inches of cable, and another device. The host is terminated with 82 ohms, but the devices are unterminated. Figure 2 shows the waveforms at the inputs to each of the two devices resulting from the model of Figure 1. The signal at device 2 is fine, but the signal at device 1 shows a significant glitch at half of the source voltage.

This is easily explained by transmission line theory. The incident wave begins by traveling down the cable. The magnitude of the incident wave is the source voltage reduced by the voltage divider of the source impedance and the cable impedance (110/(82 + 110)). For now we will assume that this is about half of the applied voltage. The incident wave reaches the first device, and then passes on down to the second device. The second device is at the end of the cable, so the incident wave reflects at this point. At the point of reflection the incident and reflected waves are added with no time delay, so the full signal amplitude

appears. But the reflected wave still has to travel back to the source. As it does, it passes the first device. Now the first device sees the sum of the incident and reflected waves, and the voltage steps up to the applied voltage value.

The problem occurs because the incident and reflected waves occur at different times at the first device. The greater the delay between the incident and reflected waves, the greater the problem. The reflected wave eventually arrives at the host where it is mostly dissipated in the source termination. Note that if the devices are more than 6 inches apart, the problem gets worse.

ATA Bus Timing

Terminating the ATA bus does not come for free. The impact of partial termination at the host and the device is to increase propagation delays throughout the system. The ATA document specifies that timing is referenced to the input pins of the device. This means that most of the timing issues must be addressed by systems manufacturers and bridge chip designers.

The most significant timing issue is the propagation delay of the cable. This needs to be accounted for in the host-side timing. The SPICE model in Figure 6 shows a host with very fast rise times driving a cable with worst case loads. Two devices are assumed with the maximum allowed capacitance loading of 25 pF (ATA-2, Revision 3, section 4.5, table 3). The simulation results are shown in Figure 7. It can be proven that the period of the ringing is four times the propagation delay of the cable. This simulation shows that the worst-case cable propagation delay is 4.2 ns. Note that this is twice the value that would be calculated by assuming an 18-inch cable with a propagation velocity of 70%. The additional delay is due to the presence of the capacitance loads on the cable. This is important to system designers who must take into account worst-case cable delay when specifying the bridge chip timing.

From the host point of view all of the ATA timings must be corrected by the propagation delay of the cable to insure that the timing is correct at the input pins of the furthest device. Figure 8 shows a typical result using the read cycle data setup time as an example. The ATA standard says that the setup time at the device must be T5 (e.g. 20 ns for PIO Mode 3). This means that the host must allow for a setup time of (T5 - transit time) (e.g. 20 - 2x4.2 = 11.6 ns for PIO Mode 3).

Note that this time is referenced to the pins of the ATA interface at the host. If the host has a series partial termination resistor then the bridge chip must include additional timing margin to account for the RC delay of that resistor. The additional cable delay added by a series termination resistor at the host is dependent on the value of the resister and the lump capacitance of the system. Some simulations have shown a range of 1 to 11 ns for series resister values

ranging from 22 to 82 ohms. A higher resister value increases the delay, reducing the timing margin. This implies that a lower resister value, such as 33 ohms, might be the best.

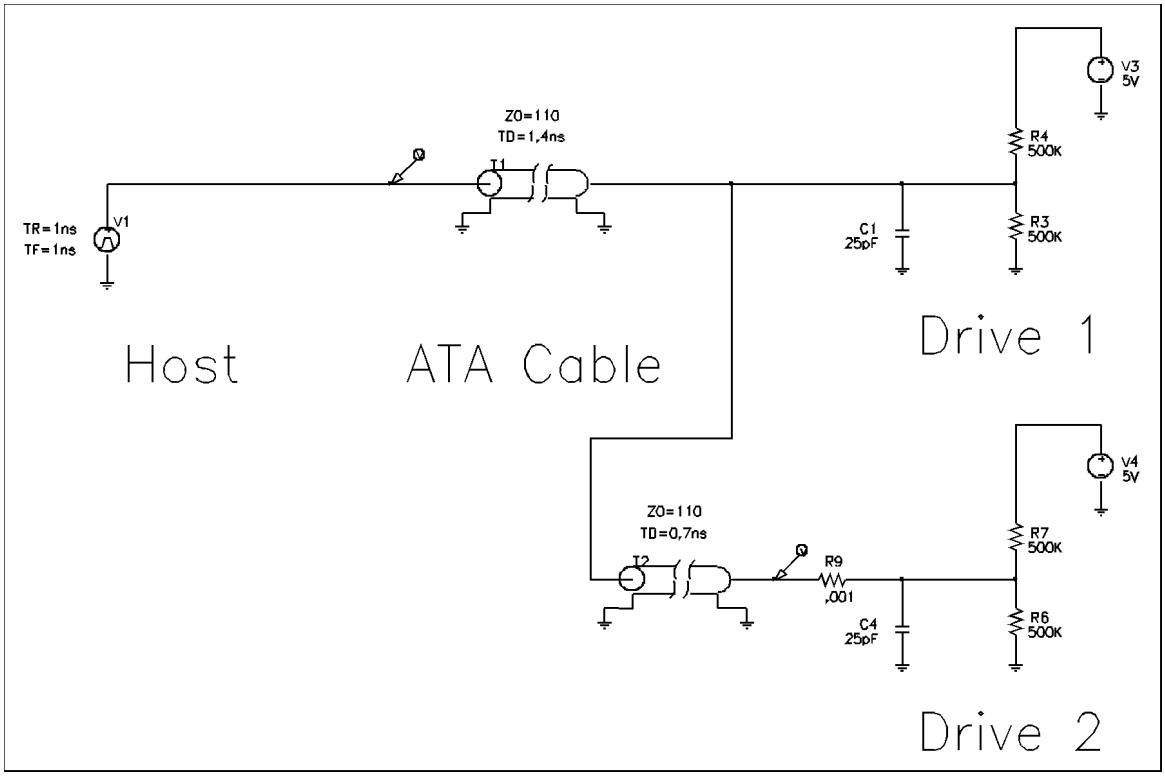

Figure 6 — SPICE model of ATA cable with worst case loads

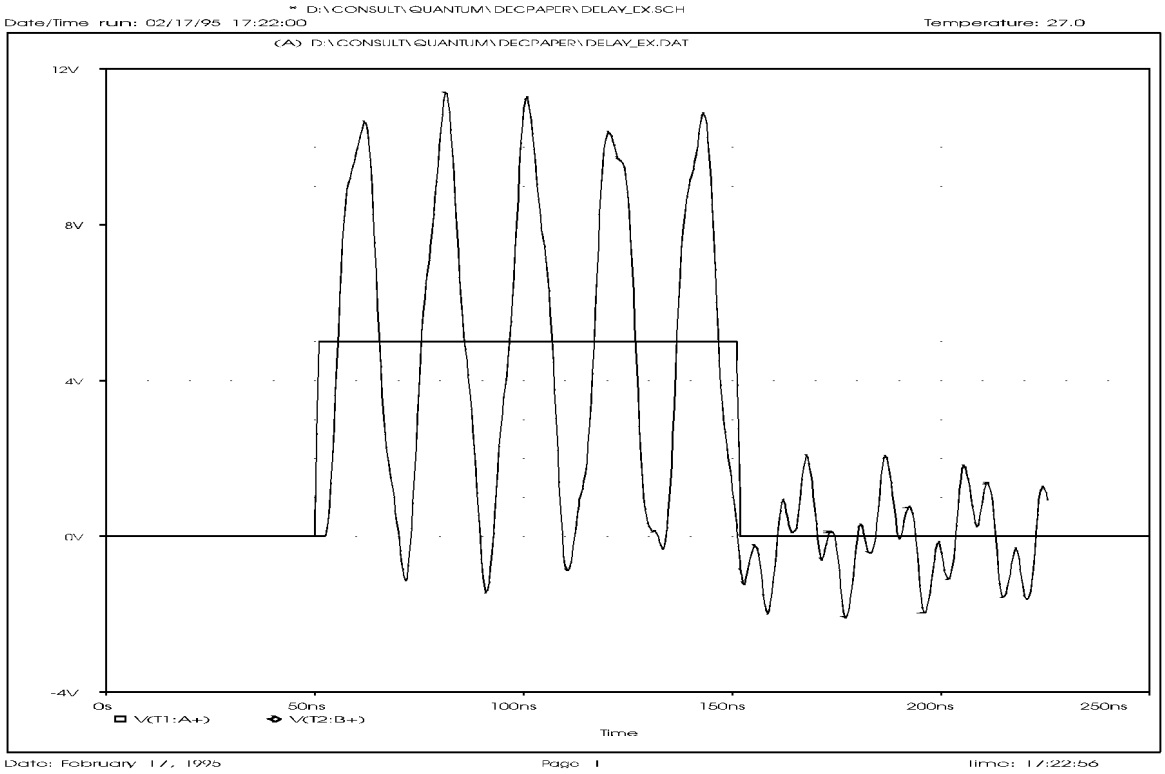

Figure 7 — Simulation of unterminated ATA cable with worst case loads

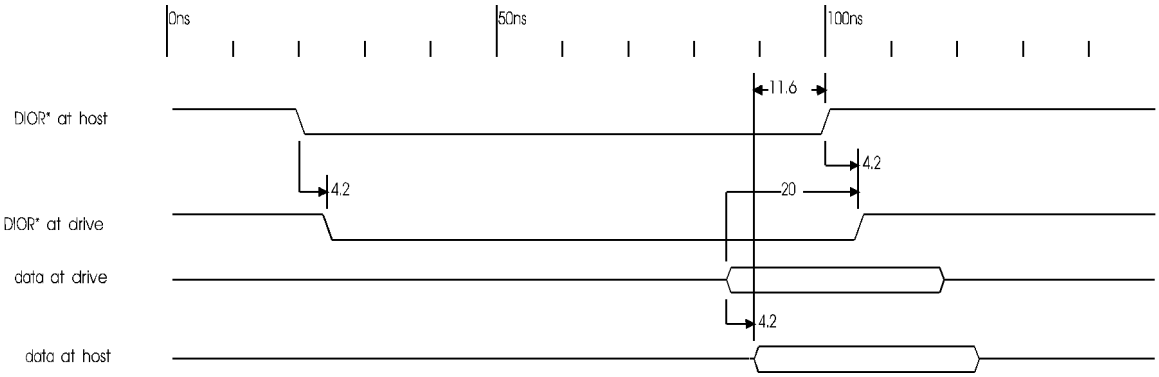

Figure 8 — Host data setup time during a read cycle

One of the common mistakes made by chip designers is to use a lumped capacitance model for simulating the delay of the output cell. For ATA simulations this sometimes consists of adding the maximum capacitance allowed for the host and the devices (3 x 25 pF) to an estimated capacitance value for the cable (25 pF). The simulation is then performed with 100 pF capacitance on the output. This is not an accurate measurement of the timing. The proper

simulation is to use an output capacitance for the motherboard, a host end termination resistor, and transmission lines to the devices (Figure 9).

To illustrate how these are different, suppose that the output cell turns out to be 2 ns too slow. The designer (using a 100 pF model) might be inclined to increase the drive of the output devices. With enough drive into a purely capacitance load another 2 ns can be removed, bringing the output cell back into spec. But look at what happens with the transmission line model. Increasing the drive of the output cell does not decrease the length of the cable nor increase the speed of signal propagation in the cable. The 2 ns required cannot be achieved by increasing the output drive. The time must come from somewhere else in the internal circuitry. Increasing the output drive only increases the edge speed, making the ringing worse at the device end. (In reality some time is gained with the faster edge speed, but not nearly as much as would be predicted with a simple capacitance load model.)

It is thus not possible to decrease the overhead by increasing the drive current beyond the 12 mA specified by ATA-2.

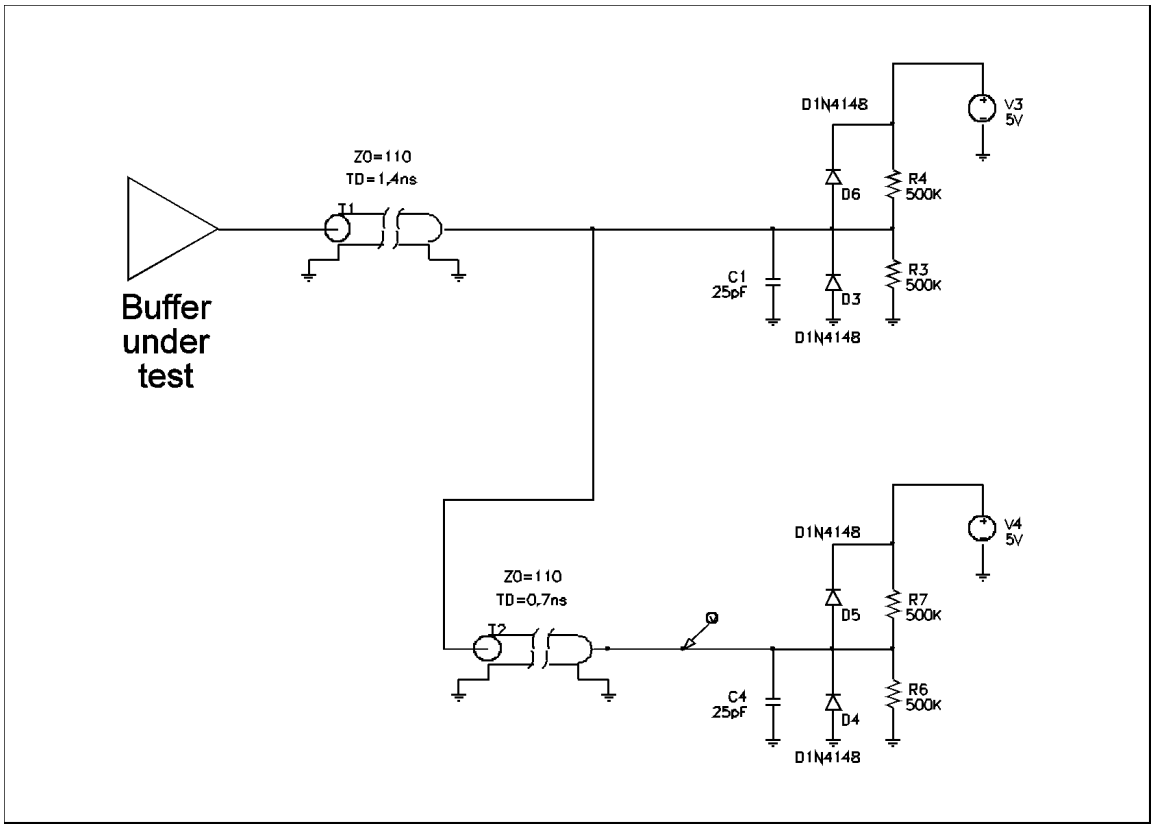

Figure 9 — Standard cable load for measuring ATA signal timing

Propagation delay times at the host (and at the device) must be measured to the ATA standard of 0.8 V for high to low transitions and 2.0 V for low to high transitions. Many IC manufacturers measure to the “typical” switch point for TTL of 1.4 V. But since both the hysteresis window and hysteresis offset of a given receiver will move with process, voltage, and temperature, the only guaranteed switch points are the TTL high and low values.

Device manufacturers need to insure that any partial termination circuits they implement present an effective capacitance of 25 pF or less. What is an effective capacitance? From a practical point of view, any circuit is valid provided it does not increase the propagation delay of a worst-case cable. This is because systems manufacturers are counting on a certain cable delay in their design. The easiest way to answer the question of acceptability is to do a simulation with the proposed termination and measure the delay. The simulation needs to include all possible device combinations including generic devices with ordinary 25 pF loading. The simulations should also include read cycle timing with the worst-case load on the data lines. The difference here is that the device manufacturer gets to make their measurements at the output of the device, not at the far end of the cable.

Unfortunately the delay of a given termination circuit at the drive end also depends on the termination used at the host end. There is an interaction between these two termination. Simulations show that the incremental cable delay due to the 82 ohm termination at the device circuit is 1.4 ns for a 33 ohm host termination and 5.0 ns for an 82 ohm host termination. This delay is in addition to the fundamental RC delay introduced by a series termination resistor at the host.

Devices internally need to account for any delay that their termination circuitry inserts. This delay must be added to the timing specifications of the interface controller chip such that the ATA timing specs are still met at the input pins to the device.

Dual Port ATA Systems

Supporting ATA-2 on multiple ports with a single bridge chip can introduce new complications. Unless proper attention is paid to these issues, a new class of system integration problems will be created.

In the interest of reducing cost through saving pins on these system bridge chips, the chip makers have often multiplexed the pins between the ports. Some control lines obviously must be driven separately (e.g. IRQ). But clever design can allow some control lines to be shared (directly connected but with isolation) (e.g. IOR and IOW), and the data lines can be shared. Unfortunately, they cannot be physically directly connected without proper care.

The restriction mentioned earlier on number of devices on a cable and length of a cable should really refer to "bus" not "cable." In the old, simple ATA world the bus was the cable (with some stubbing). But with 2 ports it is possible to create a system with 2 cables but one bus by sharing signals between the ports.

The figure below illustrates this case. With IOR, IOW, and data directly connected, the physical bus for these lines begins at the point marked "bus starts" and ends at the point marked "bus ends." It contains the two ATA cables and the traces on the system board connecting the cables to one another. Stubs are present at each device (4 stubs) and one in the middle of the bus to connect to the system chip. If the individual cables are 18 inches, then the total bus is 36-40 inches long (depending on the length of the traces on the system board) with 5 device loads. Under these circumstances the transmission line effects would generate errors for any device operating at Fast ATA speeds, and perhaps even at slower speeds. To avoid these problems proper termination must be used and sufficient timing margin for the additional propogation delay must be provided.

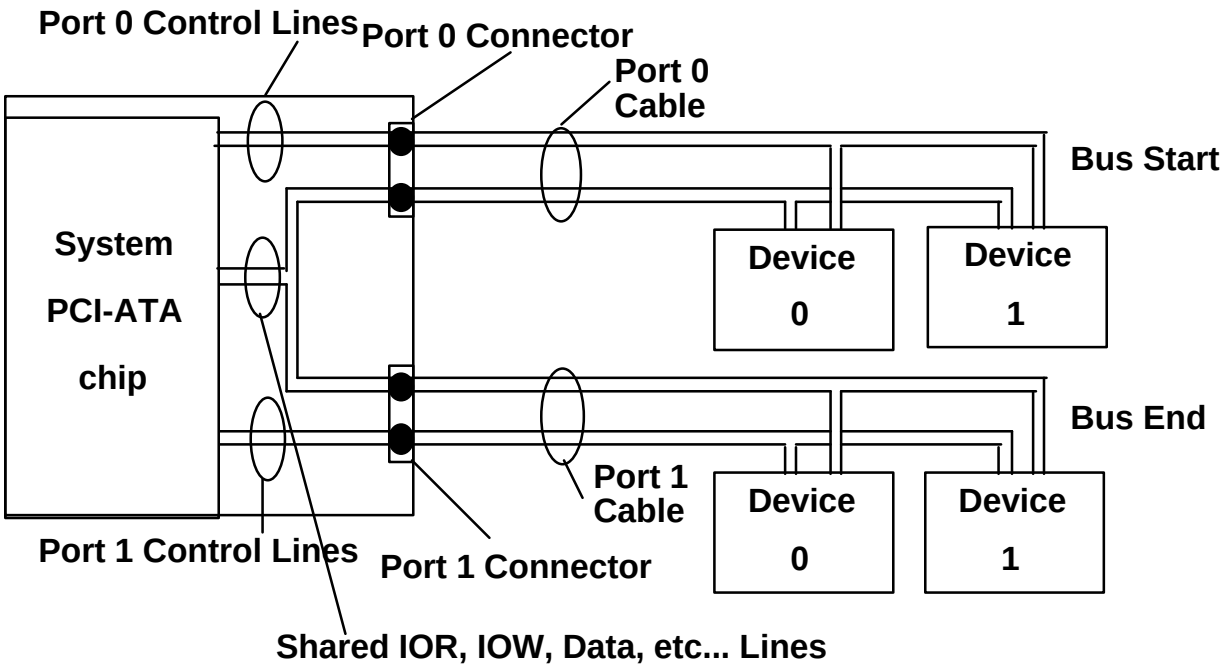

Figure 10 — Typical Dual Port Configuration

This is one of the classic issues that is a system design problem impacting devices on the ATA cable. System integrators should find a clean solution to this potential problem. Any solution should consider some method of isolation or controlling the signals on the motherboard. This will add a prop delay which should be considered in the overall system timing.

Another major problem is the implementation of DASP line on some motherboards. We have seen some cases where the primary and secondary port DASP- lines are wire-ORed together. This in itself is a problem. With these lines connected, if a Device 0 (historically known as Drive 0 or the Master) and a Device 1 (historically known as Drive 1 or the Slave) are connected to the primary port, while a Device 0 is connected to the secondary port, the secondary Device 0 will always incorrectly see a Device 1 present on its port during a bootup. This could cause the Device 0 on the secondary port not to boot in the time required. This problem would be for all drive manufacturers who follow the ATA CAM specifications.

During bootup with a Device 0 and a Device 1, the Device 1 asserts the DASP- line to inform Device 0 that it is present. The Device 1 needs to assert DASP- within 400 msecs after the RESET- line is released. Device 1 also has 30 seconds after the RESET- line is released to assert PDIAG- to indicate the Device 1 has finished its diagnostics. Device 0 will look for the DASP- line being asserted within the 400 msec window after the RESET- line is released. If Device 0 sees the DASP- line asserted it assumes that a Device 1 is present and waits for up to 30 seconds for the Device 1 to assert PDIAG- indicating the Device 1 has completed its diagnostics.

Some CD ROM devices hold the DASP- line asserted for 86 msec after the RESET- line is released. Since the DASP- lines for the secondary port and the primary port are connected together, any Device 0 following the ATA specifications that is connected to the other port will incorrectly determine that a Device 1 is present on its port.

Guidelines for system designers

There are not many guidelines in the ATA standard addressing the transmission line effects on the ATA cable. The dual port feature of newer systems further complicates system design, especially since the obvious techniques (using separate drivers and receivers for each signal on each port) may add cost to the system.

With this in mind, the following are strongly recommended for all system implementations:

- Do not use any value less than 1K ohm for pull up resistors on ATA bus open collector signals such as IORDY and IOCS16 (as per the ATA -2 standard).

- The chip driving the ATA signals should be located as close as possible to the ATA connector.

- Keep the capacitance of each pin of the IDE connector on the host below 25 pF when the cables are disconnected from the host, as per ATA-2.

- If series resistor termination is used, measure and analyze the other parameters on the same line in the system (load capacitance, distance from the source on the cable, propagation delay, rise and fall times of the buffer and the resistor value at the other end). Make sure that signal timing of the system is not compromised.

- Series resistors of 22 to 47 ohm should be used on all of the lines as close as possible to the ATA connector on the motherboard, except for the open collector signals (e.g. IORDY, IOCS16, DASP).

- For dual port implementation, one series resistor for each port and each shared line on the cable should be used. The chip should drive two series resistors, one for each port. These series resistors should be placed as close as possible to the ATA connector for each port.

- For dual port implementations, the following signals should not be shared between the two ports: INTR, DACK, DREQ, DASP, and PDIAG.

- For dual port implementations, the pair of signals CS0 and CS1 may be shared between the two ports, or the pair of signals IOR and IOW may be shared between the two ports, but not both. The pair of signals not shared should be treated the same as INTR, DACK, DREQ, DASP, and PDIAG.

- For dual port implementations, special considerations should be given when connecting IORDY between the two ports since it is an open collector line.

- For dual port implementations, increase the width of DIOR and DIOW signals to accommodate the additional delays.