# High-Performance Bidirectional Signalling in VLSI Systems

Larry R. Dennison Whay S. Lee William J. Dally

Artificial Intelligence Laboratory
Laboratory for Computer Science
Massachusetts Institute of Technology
October 12, 1992

**Abstract**

Interchip I/O bandwidth is a critical bottleneck in VLSI systems. To make the best use of this resource the conventions and circuits used for inter-chip signaling must be optimized to achieve the maximum bit rate with minimum power dissipation. This paper describes a set of I/O pads that we have developed at MIT. They operate with small signal levels to reduce power dissipation. To achieve high bandwidth, they operate simultaneously in both directions over a single wire and are isolated from supply noise. Bidirectional signalling is achieved by superimposing transmitted current waveforms and can be used with either transmission lines or lumped loads. This paper describes the theory behind our signaling method, the circuit design of the I/O transceivers, and the results of our experiments with prototype transceiver test chips.

## 1 Introduction

As integrated circuit densities increase, the density and operating frequency of logic gates is growing at a much greater rate than that of chip I/Os. While the number of devices per chip is increasing quadratically as critical dimensions are reduced, I/Os on the other hand are usually confined to the chip perimeter and hence scale only linearly with linear dimensions. Even then, I/O pad pitch scales at a slower rate than on-chip wire pitch. Gate switching frequency scales with improvements in device dimensions while I/O bandwidth per pin is limited by external wiring. Also, device switching energy scales as the cube of linear dimensions [6], while with standard signalling levels, I/O switching energy remains constant.

The net effect of this uneven scaling of logic and I/O is that the increase in gate switching events per second on a chip is greatly outpacing the increase in I/O switching events per second. As a result, the bandwidth of a chip's pinout has become a major bottleneck to system performance unless locality

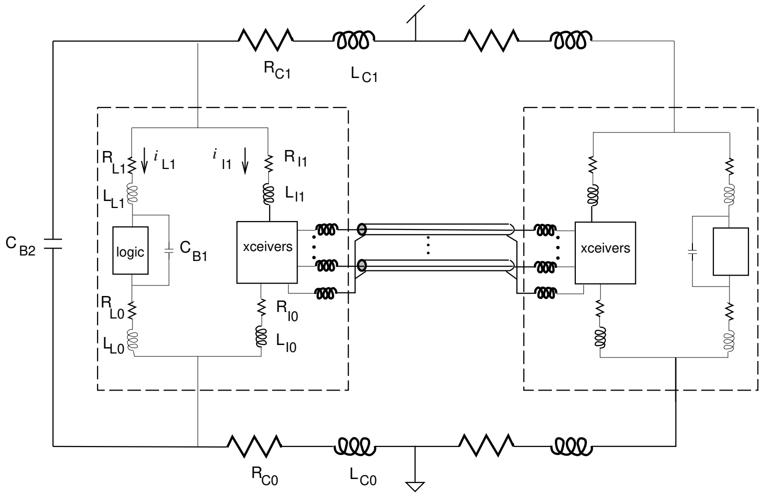



Figure 1: Transceivers on chips $A$ and $B$ exchange data over a limited number of signal pins connected by transmission lines. The signalling system must deal with noise due to power supply parasitics, reflections, crosstalk, and component tolerances.

can be exploited to keep communication on chip. Hence efficient signalling between chips in a VLSI system is critical.

In this paper we describe an approach to high-performance digital signalling currently under development at MIT. Section 2 introduces the problems involved in high-speed signalling and discusses the shortcomings of conventional signalling. Section 3 describes our approach to overcoming these problems: current-mode signalling and simultaneous bidirectional transmission. The circuits to realize this approach are the topic of Section 4. The limits of the performance of interconnections are discussed in Section 5. Finally, Section 6 discusses some early experiments with preliminary versions of these circuits.

## 2 The Problem

The digital signalling problem is illustrated in Figure 1. Two integrated circuit chips, $A$, and $B$ must exchange information at high-bandwidth over a limited number of signal pins. In general $A$ and $B$ may each also exchange signals with other chips, not shown. The problem is to transmit signals over each wire at maximum rate while dissipating minimum power.

The big obstacle to high-speed, low-power signalling is *noise*. In contrast to analog systems where noise is often due to physical phenomena (e.g., Johnson or shot noise), in digital systems almost all noise is self induced. The major sources of noise in digital systems affecting an I/O transceiver are:

1. Self-induced power supply noise.
2. External power supply noise.
3. Transmitter and receiver offsets.
4. Reflections.
5. Crosstalk.

Self-induced power supply noise, sometimes called "ground bounce", occurs when the I/O transceivers draw a current $i_{IX}$ from power supply $X$[1]. This current induces a self-induced power supply noise voltage on supply $X$ of:

$$v_{PX} = i_{IX} \sum_{Y} R_{YX} + \frac{di_{IX}}{dt} \sum_{Y} L_{YX}. \quad (1)$$

This voltage is measured between the *root* of the power supply tree for supply $X$ and the *leaf* of the tree where it connects to the transceivers. It is due to voltages developed across the parasitic inductances $L_{YX}$ and resistances $R_{YX}$ in each branch $Y$ of the $X$ power supply tree between the root and the transceivers.

Self-induced noise can be extreme in chips that attempt to drive fast edges with high-voltage swings. In addition to corrupting signals that are referenced to the power supplies, I/O induced supply noise can cause the system logic to malfunction because of the noise voltage developed across the parasitics on shared branches of the power supply tree.

External power supply noise is the voltage developed across shared branches of the power supply tree by other parts of the system:

$$v_{EX} = \sum_{S} \left( i_S R_{SI} + \frac{di_S}{dt} L_{SI} \right). \quad (2)$$

This voltage is developed across parasitics $R_{SI}$ and $L_{SI}$ through which both the transceivers and noise generator $S$ are connected to the root of the $X$ supply tree. For example, $R_{C0}$, is in a branch of the ground supply tree that is shared between the logic and transceivers on the left chip in Figure 1. Even if the I/O pads themselves are very quiet, they are connected to power supplies that may have substantial noise induced by on-chip logic and clock generators, and even other chips on the same board. While I/O drivers may have their own power supply pads to limit shared parasitics, board ground planes and board connector pins are shared.

Bypass capacitors can reduce supply noise when it is balanced, however they are not a panacea. Figure 1 shows bypass capacitors both on chip, $C_{B1}$, and off chip, $C_{B2}$. To zeroth order, these capacitors act as AC shorts between their

[1]In the figure, $X$ is 0 for the ground supply and 1 for the $V_{DD}$ supply.

terminal points on the power and ground supply trees. As an AC short, bypass capacitors are very effective when the power and ground currents are equal and opposite (as in the case of entirely on-chip logic or perfectly balanced differential signalling). In this balanced case, the AC current flows out of the bypass capacitor, rather than from one supply tree to the other. However, when driving single-ended lines (or differential lines whose termination points are far apart compared to a signal risetime), the AC power current is not equal and opposite that of the AC ground current. In this unbalanced case, all the bypass capacitor does is put the two power supply trees in parallel. At best, this will reduce the noise due to unbalanced currents by a factor of two. Also, even in the balanced case, power supply oscillations can result when the bypass capacitor forms a tank circuit with parasitic power supply inductances.

Transmitter and receiver offsets occur when device parameter variations cause the transmitted signal level and receiver threshold to differ from their nominal values. These offsets can be particularly large when a level is set by ratioing two devices whose parameters do not track one another – for example in a CMOS inverter receiver where the receive level is set by ratioing an N-FET against a P-FET.

Reflections, sometimes called “ringing”, are due to an impedance mismatch at any point of the transmission line. The ends of the line are the most common place for a mismatch to occur. If the line has a characteristic impedance of $Z_0$ (for typical PC boards, $50\Omega \leq Z_0 \leq 75\Omega$) and is terminated into an impedance of $R$ to the return plane, an incident waveform $v_i$ will cause a waveform

$$v_r = v_i \left( \frac{R - Z_0}{R + Z_0} \right) \tag{3}$$

to be reflected back down the line. An impedance mismatch in the middle of a line (e.g., at a connector) can be tolerated only if it is small compared to the rising edge of the signal. Usually an impedance mismatch of 10% can be tolerated resulting in a reflected waveform that has 5% of the amplitude (and .25% of the power) of the incident waveform.

Rise-time control is important to minimize reflections because: (1) the impedance of termination structures may be frequency dependent, and (2) small impedance discontinuities are only *seen* by high frequencies. Thus, to minimize the high-frequency component of the transmitted signal, it is important that the signal rise time be kept at the maximum allowed by the design bit rate. A rule of thumb is that the rise time should be about half of the bit cell time (e.g., a 1ns rise time for a 500Mbit/s signal)[2].

Crosstalk is the injection of one signal into another signal via parasitic capacitance and mutual inductance. If two signals $X$ and $Y$ run on adjacent

[2] This ratio gives a good compromise between maximum bit rate (cell size equal to rise time) and jitter tolerance.

copper traces on a PC board for a length $l$, a transient on $X$ will induce two types of crosstalk on $Y$: forward crosstalk which has the same waveform as $Y$ with amplitude scaled by the coupling coefficients and $l$, and backward crosstalk which has a waveform of $Y$ convolved with a pulse of length $l$. Forward crosstalk can be completely eliminated by adjusting the capacitive and inductive coupling coefficients to cancel each other while backward crosstalk is dealt with by having the source impedance matched to the line so that the crosstalk noise is absorbed at the source. Rise-time control is also important to minimize crosstalk. While the coupling coefficients of a stripline are independent of frequency, coupling due to package and device parasitics is frequency dependent and usually increases at high frequencies.

Not only do these five noise sources affect a signal by causing the level of a signal to be misinterpreted, they also introduce *jitter* into the system. Adding a noise source with magnitude $v_N$ to a signal that swings through $\Delta V$ with rise time $t_r$ introduces an uncertainty of

$$t_j = \frac{t_r v_N}{\Delta V} \tag{4}$$

in the timing of when the signal is detected crossing the receiver threshold. Power supply noise introduces additional jitter by modulating the delay of logic elements and drivers.

Conventional approaches to high-speed digital signalling in the presence of these noise sources have attempted to extend standard signalling techniques by heroic but brute-force methods involving very large numbers of power and ground pins (to minimize power supply noise) and very high current drivers (to achieve incident wave switching with standard voltage swings). The result is a system with high power dissipation.

The remainder of this paper discusses our solution to the problem of high speed signalling which (1) isolates the signal from noise using current-mode signalling, rather than suppressing noise using brute force techniques, (2) simultaneously sends signals in both directions over a single conductor to maximize the communication bandwidth per package pin, and (3) uses very low voltage levels to minimize I/O power dissipation.

# 3 Our Solution

Our approach to digital signalling is to isolate our transmitter signal levels and receiver threshold from power supply noise rather than to try to reduce the noise through brute force. We also extract the maximum bandwidth per pin by operating single-ended and in both directions simultaneously[3]. The key elements of our approach are:

[3] Related previous efforts include [4] and [5]

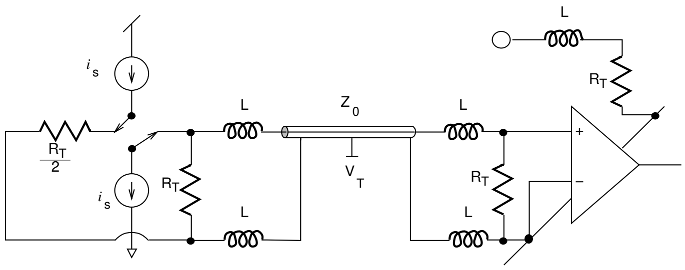



Figure 2: A high-performance current-mode signalling system. A current mode driver isolates signal levels from transmitter power. The receiver senses the voltage across the termination.

1. Connections are made point-to-point by connecting a controlled impedance transmission line between a pad on one IC, $A$, and a pad on another IC, $B$.

2. Digital signals are propagated in both directions simultaneously between $A$ and $B$, doubling the available I/O bandwidth of the technology.

3. Signals are transmitted using a current-source driver with voltage sensed across the on-chip termination resistor in the receiving chip. This isolates the signals from power supply noise.

4. Signals are parallel terminated on-chip in each transceiver.

5. The signal return plane is independent of the IC power supplies. This isolates the signals from power supply noise in the transmission plane or through the terminators.

Figure 2 shows a simplified circuit of our driver and receiver. The driver operates in current-mode. That is it supplies a current of $+I_S$ or $-I_S$ to the line to signal a 1 or a 0 respectively. The figure shows the driver sinking $I_S$ to send a zero. To transmit a 1, the two switches reverse and the driver sources $I_S$. This switching is performed at a controlled rate (as slow as possible given the design bit rate) to minimize $\frac{dI}{dt}$ into the transmission line. With an ideal current source, the signal level is completely isolated from transmitter power supply noise. Also the impedance of the current source is infinite, so the source impedance is entirely due to the on-chip termination resistor, $R_T$. With a 50Ω line (a 25Ω load), a value of $I_S$ of 10mA gives a 250mV swing and a termination power dissipation of 2.5mW (however, without on-chip transformers, transmitter power is at least $V_{DD}I_S$, 30mW for a 3V supply). Compare this with the 360mW required to drive a 3V swing signal into the same 25Ω load.

The receiver is an ideal differential amplifier that senses the voltage across the on-chip termination resistor, $\pm\frac{1}{2}I_S R_T$. The ideal amplifier rejects all power supply noise and supplies are not used to generate the threshold. The push-pull current signalling results in a threshold equal to the termination voltage which is particularly easy to generate without offsets.

The lead inductances of the signal and $V_T$ leads are shown in the figure. The signal lead is an issue if its distributed capacitance is such that it appears as an impedance discontinuity just before the termination. In that case, signal rise times will be limited to be at least some multiple of the lead length. Also transmitter transients result in $L\frac{di}{dt}$ drops across the signal inductance that are seen by the receiver. The edge rate must be selected to make these drops small compared to the signal level. The $V_T$ inductance is more of a concern because it is shared by several signals. A signal transition on one lead will induce a voltage, $L\frac{di}{dt}$, that will be seen by the other receivers and injected as a backward traveling signal onto their attached transmission lines. For a given signal rise time $t_r$ the number of signal leads $N$ that can share a $V_T$ lead will be limited to keep the voltage across the inductor less than some small fraction $f$ (typically $f \leq 0.25$) of a full signal:

$$V_L = L\frac{di}{dt} = \frac{LNI_S}{t_r} < fI_S R_T, \tag{5}$$

so

$$N < \frac{ft_r R_t}{L}. \tag{6}$$

This design produces very little self-induced power supply noise. The current driver draws constant current from the power supply and injects zero net current into $V_T$. The upper current source always sources $I_S$ into $V_T$ while the lower current source always sinks $I_S$ out of $V_T$. One of the two sources passes all of its current through the local $R_T/2$ resistor while the other splits its current between the line and the local terminating resistor. The low voltage swings result in current levels that are an order of magnitude lower than in a conventional system. Moreover the current mode operation completely isolates the transmitted voltage levels and the receiver threshold from external power supply noise.

The on-chip terminations, which can be made self-adjusting, minimize noise due to signal reflections and absorb crosstalk noise when it reaches its first termination. Careful layout to minimize coupling coefficients, to match coefficients to eliminate forward crosstalk, and to enforce an upper limit on parallel wire length is still required to keep crosstalk from being a problem.

Figure 3 shows how simultaneous differential signalling is achieved. Current-mode transmitters are attached to both ends of the line. Signals generated by $A$ propagate from left-to-right while signals generated by $B$ propagate from right-to-left. The voltages and currents in the line are the superposition of the two waveforms. At $A$ we receive $B$'s signal by subtracting out the voltage due to $A$'s signal on the transmission line. Figure 4 shows how the two signals

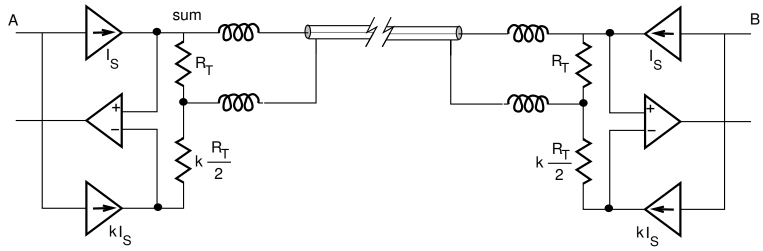



Figure 3: Simultaneous bidirectional signalling is accomplished by transmitting signals in both directions across a transmission line. At each end, the received signal is recovered by subtracting out the effects of the transmitted signal. A current source scaled by factor $k$ is used to subtract out the transmitted signal to reduce power dissipation.

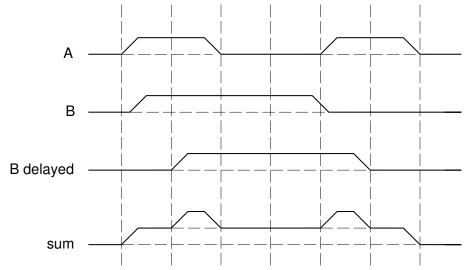



Figure 4: Waveforms for bidirectional signalling. The superposition of signal A with a delayed version of signal B gives the waveform on the left end of the transmission line.

are summed together on the transmission line.

We have avoided using differential signalling because while it provides twice the signal swing per unit pad current and gives completely balanced loads, it is less bandwidth efficient than single-ended signalling. To achieve the same bandwidth per pin with differential signalling would require operating at almost twice the signal rate as with single-ended signalling[4]. While this requires the same $\frac{di}{dt}$ to get the same voltage swing, the faster rise-times make smaller transmission line discontinuities a factor and the shorter bit-cell periods make jitter of timing signals more of a concern. Finally, the gain-bandwidth product of the receiving amplifier would have to double to achieve this rate. Hence, it is usually not possible to go twice as fast with differential signals.

We have not used series terminating drivers [4] because these drivers inherently send a voltage signal referenced to a local power supply voltage. Thus it is impossible to isolate them from power supply noise. We have attempted to duplicate the power advantages of these drivers by using pulsed signalling as described in Section 6.

# 4 Circuit Design

In this section we describe the circuit designs that approximate the ideal elements used to describe the concept in Section 3. The driver design described draws zero AC power and injects zero DC current into the termination plane, $V_T$. The receiver separates the two bidirectional waveforms by subtracting a copy of the transmitted waveform from the waveform on the line.

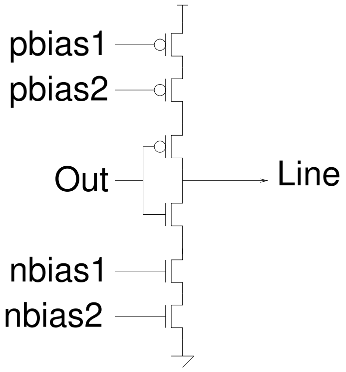
## 4.1 Driver

The driver may be constructed by using standard CMOS current sources, as shown in Figure 5. These are implemented using transistors operating in the saturation region. If the output driver is designed using cascoded current sources, output impedances greater than 100K ohm are easily obtained. The power supply rejection in this case is bounded below by -100dB.

Power supply noise also enters the line from parasitic capacitances, such as gate-source/gate-drain overlaps, back-gate, diffusion-well, and diffusion-bulk capacitances. Since all plates are connected via DC paths to some portion of the circuit, these capacitors couple high-frequency noise into the line. The effects of some of these parasitics may be reduced by careful layout techniques and the use of source-coupled wells.

Figure 6 shows an HSPICE AC analysis of the power supply rejection. The cascode current sources lead to 60dB of power supply rejection for frequencies

[4] Because fewer $V_T$ pins are required for differential signalling the same bandwidth per pin can be achieved at slightly less than twice the single-ended signalling rate.



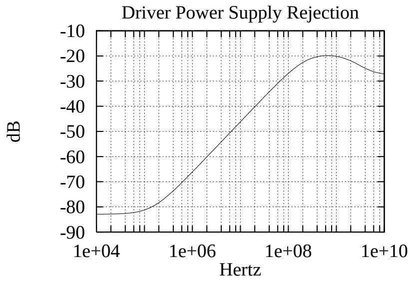




Figure 5: Cascoded Current-Source Driver

Figure 6: Driver Power Supply Rejection

below 1 MHz. Power supply noise in this range usually results from switching power supplies. Between 1MHz and 100MHz, we still achieve 20dB of rejection. Noise in this range of frequencies is typically the result of other IC's on the same board. Coupling becomes most noticeable at frequencies over 100MHz, which is typically self-induced. We thus want to limit the amount of power supply noise $V_{PX}$ in the higher frequencies.

## 4.2 Current Steering

One of the problems with the cascoded current sources is that the internal nodes tend to have large parasitic capacitances. When decoupled from the line, the current sources charge these capacitances to the rails. Later, when the output switches, the drive transistor sees the parasitic capacitor, not the current source. This results in undesirable initial current overshoot.

A solution is to *steer* the current from the cascoded current sources into either the line or a dummy load. The circuit show in figure 7 is similar to that used for a differential driver except that the complementary output never leaves the chip. In the absence of a received signal, this circuit gives a constant voltage on all nodes in the current source, eliminating the initial overshoot.

The current steering driver has two other beneficial properties. First, it draws constant current from both the $V_{DD}$ and $GND$ supplies, irrespective of the output state. This reduces $V_{PX}$ which in turn reduces the required number of supply pins and the noise effects on the system logic. Second, the total current supplied from an output driver into $V_T$ is close to zero[5]. This suggests that the $V_T$ supply could be as simple as a capacitor and a resistor divider.

As we will explain further in section 5, we need to limit the edge rate of the driver. This is done by splitting the driver into several small parallel drivers. The turn-on of each driver is then staggered via a buffer chain. This technique allows the construction of not only a slow straight edge, but also a piecewise

[5]The DC current into $V_T$ is equal to the mismatch between the two current sources.

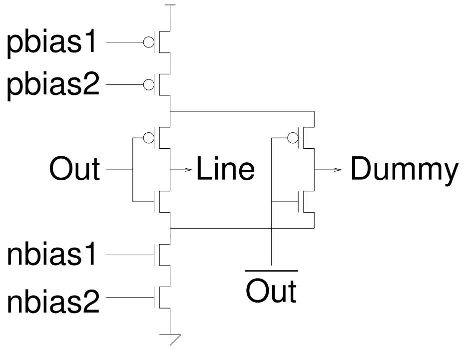



Figure 7: Cascoded Current Sources with Steering

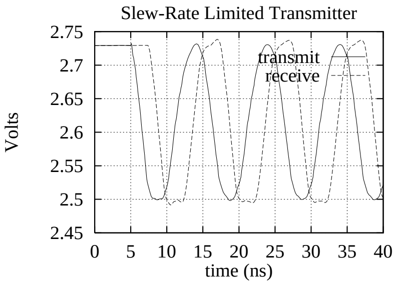



Figure 8: Transmitter Transient Response, 200Mbit

approximation to more elaborate curves. These could more closely match the desired low-pass characteristic. Figure 8 shows the transient response of a slew-rate limited driver.

### 4.3 Receiver

As we decreased the signalling voltage level, we kept the output voltage swing of the receiver constant (full CMOS levels). In doing so, we increased the required gain-bandwidth product of the receiver. The input voltage swing of the differential amplifier is 250mV, ignoring noise. For 3V CMOS levels, a gain of 12 (22dB) is required. To achieve a 200Mbit signalling rate, we need a receiver with a gain-bandwidth product of 1.2GHz.

The gain-bandwidth product must be measured large-signal, as the high-frequency output swing must match the low-frequency output swing to avoid inter-symbol effects. Using a good CMOS process, it is fairly easy to construct a standard transconductance amplifiers with a unity-gain frequency near 2GHz. For a standard transconductance amplifier, these gains are only for a limited output range which is typically $\pm 1V$. Outside of that range, the transistors in the output stage come out of saturation and enter the triode region. Once in triode, they supply less current and the output slew rate falls off.

The transconductance amplifier can be adjusted to give large output swings by allowing one of the two output drive transistors to enter cut-off. This produces the rail-to-rail swings shown in Figure 9. The input waveform (labeled "input") is 200mV peak-to-peak, with 1ns edges. The waveform labeled "amp output" is the output of the differential amplifier. This output is then cleaned up using an inverter whose output is labeled "output".

Figure 10 shows the result of a 500Mbit/sec input waveform. The input waveform is 200mV peak-to-peak, with .25ns edges. Notice that the amplifier output no longer goes rail-to-rail and is now slew-rate limited. The inverter following the amplifier restores full rail signalling.

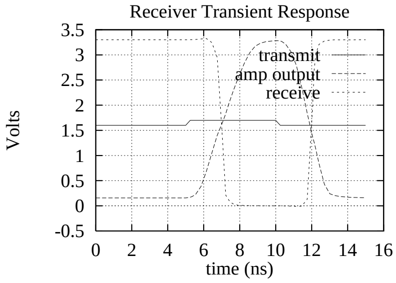



Figure 9: Receiver Transient Response, 200Mbit

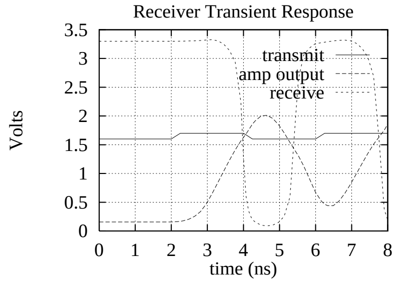



Figure 10: Receiver Transient Response, 500Mbit

# 5 Achievable Performance

The performance of an inter-chip signalling system is limited by the following factors:

- The bandwidth and frequency characteristics of the transmission media and its terminations.
- The gain-bandwidth product of the transmitter.
- The gain-bandwidth product of the receiver.
- The voltage noise and timing jitter present in the system.

The eye diagram 11 illustrates the effect of these factors on system performance. The transmitter produces a waveform with amplitude $v_{tx}$, bit cell period $t_c$, and rise time $t_r$. The rise time $t_r$ is limited both by the bandwidth of the transmitter and by the bandwidth of the media. If the transmitter input amplitude is $v_i$, the transmitter gain-bandwidth product, $f_{tx}$, must be at least

$$f_{tx} \geq \frac{v_{tx}}{2 v_i t_r} \tag{7}$$

to generate this waveform. Transmitter gain-bandwidth is unusually not the factor limiting performance because it is easy to build transmitters in $0.8\mu$ CMOS with gain-bandwidth products in the GHz range and the transmitter is attenuating ($v_{tx} < v_i$). The transmission media is the limiting factor on the achievable transmitted waveform. Generating a waveform with significant frequency components above the flat response of the transmission system will result in excessive attenuation, excessive reflections, or both.

The received waveform is also shown in Figure 11. Here the reliable received amplitude has been reduced by a combination of voltage noise $v_N$ and attenuation $A$ to $v_{rx} = Av_{tx} - v_N$. The rise-time and timing jitter limit the

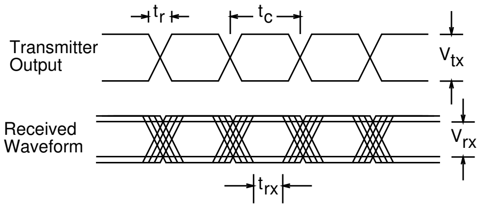



Figure 11: An eye diagram showing transmitted and received waveforms. Attenuation, noise, and jitter reduce the height and width of the *eye* that the receiver must sense.

*window* or *eye* within each bit cell during which $v_{rx}$ can be sensed by the receiving amplifier to $t_{rx}$. If the receiver output has amplitude $v_o$, the receiver gain-bandwidth product, $f_{rx}$, must be at least

$$f_{rx} \geq \frac{v_o}{v_{rx} t_{rx}}. \tag{8}$$

At present, receiver gain-bandwidth product is a limiting factor on system performance. For very small signal swings, two stage amplifiers are required to achieve the required gain without sacrificing bandwidth. Also, the required $f_{rx}$ can be traded off against the required $f_{tx}$ by increasing signal level $v_{tx}$, although this is at the expense of increased power dissipation. However, in $0.8\mu$ CMOS, one can build a receiver that operates at bit rates that push the capabilities of conventional package and PC-board transmission media. As circuit bandwidth improves with smaller devices, transmission media will become the single limiting factor on performance.

**Interconnect Model:** As described above, even with perfect circuits, performance is limited by the transmission media: bond-pad parasitics, lead inductance, pin capacitance, and transmission line imperfections all affect signal quality and bandwidth. Consider the model for the interconnection between two integrated circuit chips shown in Figure 12. The circuits are in separate packages attached to a common PC board. The PC trace is modeled as an ideal transmission line with a $50\Omega$ characteristic impedance. Both ends of the transmission line are terminated on-chip in the characteristic impedance. We add in three parasitic elements to each end: pad capacitance, lead inductance, and pin capacitance. For these, we assign the nominal values of $5p$F, $3n$H, and $5p$F. These values are representative of high-performance ASIC packages available today. This model is representative for a short connection across a PC board where package parasitics dominate. For a longer run or one that included connectors, line attenuation and discontinuities would be included as well.

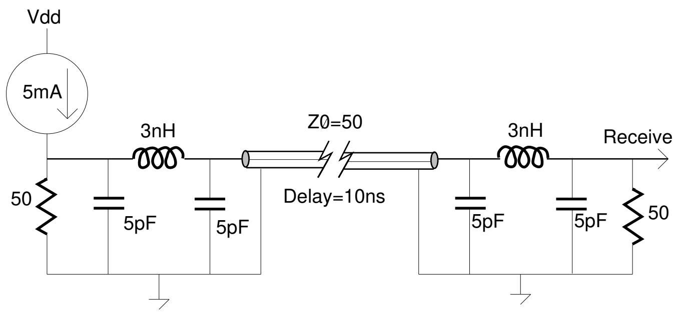



Figure 12: A circuit model of a package and PC board transmission media. Packaging parasitics at either end of the line affect the frequency response of the termination.

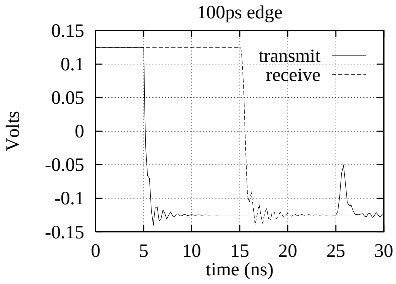



Figure 13: Line response to 100ps edge. Ringing is caused by the parasitic tank circuit and a pulse is reflected back to the transmitter because the parasitics make the termination frequency dependent.

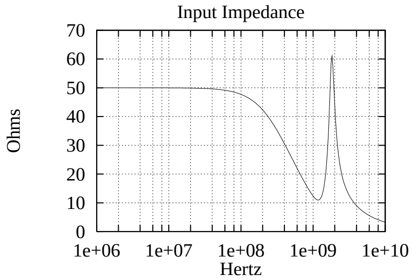



Figure 14: Effective input impedance for the model circuit. The $Pi$ low-pass filter causes the impedance to roll off above 100MHz and then peak at 2GHz, the resonant frequency of the parasitic tank circuit.

**Step Response:** Using SPICE, we look at the step response of the interconnect model (Figure 13). We see that the initial edge at the driver pad looks clean (with only a small amount of ringing due to the parasitic tank circuit), as does the initial edge at the receiver. However, a very large spike is reflected back at the driver. In the actual circuit, we are recovering data at both ends of the interconnect. A spike of this magnitude could easily cause an error in data recovery.

This spike is mainly the result of the frequency-dependent termination. The topology of pad capacitor, lead inductor, and pin capacitor is that of a $\Pi$ (low-pass) filter. The effective termination impedance is shown in Figure 14. High frequency components are reflected back toward the driver. Very little energy at those frequencies is actually transferred to the pad, explaining the clean waveform seen at the pad.

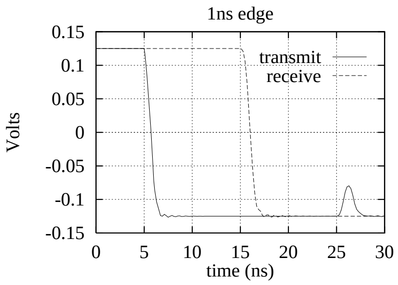



Figure 15: Line response to 1ns edge. Slowing the edge rate reduces the ringing and halves the amplitude of the reflected pulse.

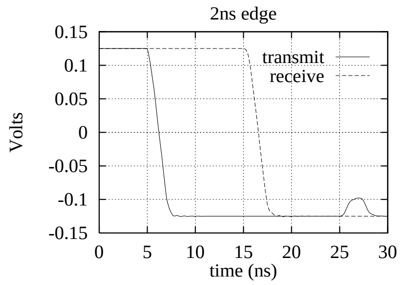



Figure 16: Line response to 2ns edge. The reflected pulse is now reduced to about 10% of signal amplitude.

**Rise-Time Control:** The results of the step response analysis suggest that the source waveform not have any frequency components which are incorrectly terminated. Such components may be attenuated by low-pass filtering or by shaping the drive current transition. We adopt the second approach by slowing down the transition. The SPICE runs in Figure 15 and Figure 16 show the result of edge times of 1ns and 2ns respectively. These figures show that slowing the edge rate is effective in reducing the reflection.

**Performance Limits:** Consider the circuits described in Section 4. The driver is able to supply very high speed edges, however, as described above, edge shaping is needed to reduce reflections. To limit the reflections to less than 10%, we restrict $t_r \geq 2$ns. Thus, based on packaging constraints, we are restricted to less than 500Mbit/sec per direction per pin.

As the line rate increased or edge rate decreased, the amplifier in the receiver became slew-rate limited. This results in a kind of inter-symbol interference. A long transmission of zeros results in difficulty in recovering an isolated one. A similar effect holds true for a long transmission of ones followed by a zero. Even at 500Mbit/sec, the isolated pulse width at 90% of final value is slightly over 1ns.

Thus, for an $0.8\mu$ CMOS process, using the circuits of Section 4, the performance-limiting factor is the package and the interconnect, not the circuits or devices.

# 6 Experiments

The design presented in Section 4 has been influenced by our experience in fabricating and testing three prototype chips. Our first chip demonstrated the basic concept of bidirectional signalling. However its performance was limited by slow on-chip test circuitry. This chip was described in [5] and will not be discussed further here. The second chip, described below, overcame

PHOTO

Figure 17: Transmitted Data and Transmission Line Signal Waveforms for Static Test Chip.

the problems with on-chip test circuitry. Its performance was limited by noise and the gain-bandwidth of the $2\mu$ CMOS receiver. The third chip tested a pulsed version of the bidirectional pads. Unfortunately fabrication difficulties prevented this chip from being fully operational.

For economy and fast turnaround, all three chips were fabricated in $2\mu$ CMOS. The first chip was fabricated through MOSIS while the second and third chips were fabricated through the Massachusetts Microelectronics Center.

## 6.1 Static Pad Test Chip

The system shown in Figure 3 was fabricated and tested. The driver was realized as a simple two-transistor CMOS inverter, sized to source and sink 10mA. A differential sense amp was used as the receiver. A scaled down (5:1 ratio) driver was used to produce the reference copy of the transmitted signal for subtraction, to help reduce power dissipation. A static differential sense amp was used as the receiver . The complete design was fabricated in a $2\mu$ N-well process. Total simulated power consumption was 41.5mW per transceiver.

A pseudo-random data stream generator and a compare circuit were included on the prototype chip for the purpose of testing. Operating bidirectionally at 100Mbits/s through a $75\Omega$, 2m long coaxial cable, our transceiver was monitored for 24 hours without a compare mismatch. Functionality of the system is tolerant of fluctuations of about $\pm 200$mV in $V_T$.

Figure 17 shows the two transmitted data waveforms along with the superimposed signal seen on the transmission line. The noise seen on the superimposed signal is rejected by the differential receiver amplifier.

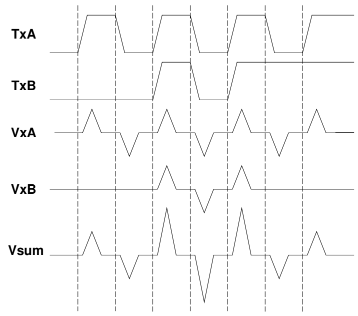



Figure 18: Timing Diagram for Pulsed Transceiver System

## 6.2 Pulsed Bidirectional Test Chip

Our third test chip included a pulsed bidirectional transceiver based on a technique reported by Gabara & Thompson [2]. This design pulses the current drive high or low in response to a rising or falling transition as shown in Figure 18. Thus the transmitter dissipates no static power. A combination of fabrication problems (transistor betas off by a factor of three) and design problems with the pulse shaping circuitry prevented these chips from operating as designed.

## 6.3 Self-Adjusting Terminations

Both of the designs described above include self-adjusting termination resistors. A finite state machine (FSM) switches a series of exponentially sized complementary MOS passgate resistors. During the reset sequence, the FSM switches passgates in and out until the termination resistance matches an external reference resistor. Using an on-chip terminator reduces board parts count, and avoids the lead inductance between the transceiver and terminator inherent with external termination.

## 6.4 Discussion of Experimental Results

Both of the prototypes described above suffered from a large discrepancy between the SPICE model parameters used for simulation[6] and the actual device parameters. Actual transistor transconductance was measured to be less than half the simulation value.

Despite modeling inaccuracies, the digitally-adjusted termination mechanism operated properly. By operating digitally, it eliminated many of the stability and noise problems associated with the analog-controlled termination scheme used in [5].

[6] We based our design on the data supplied in [3], as recommended by the foundry.

Our experiments with the static test chip revealed several noise problems with our initial circuits. Our CMOS inverter drivers did not deliver zero total current into $V_T$ and hence induced a substantial amount of $L\frac{di}{dt}$ noise across the shared $V_T$ inductor that corrupted the received signals. This driver had inferior power supply rejection and passed substantial amounts of power supply noise into the transmitted signal. Without the use of current steering, the driver also generated substantial power supply noise. Finally, this driver did not employ rise-time control.

Our experiments, particularly since they were conducted using a slow $2\mu$ CMOS process also highlighted the importance of receiver gain-bandwidth product. In particular careful layout is required to minimize capacitance on the critical nodes of the differential amplifier.

The DIP package we used for the prototypes, contributed a large amount of lead inductance ( up to 19nH [7] ), which was the another major cause of noise and signal distortion. A high-performance package is essential to successful high-speed signalling.

## 7 Conclusion

The number of logic switching events per second per chip is growing much faster than the rate of I/O switching events per chip making I/O bandwidth a critical resource. Conventional I/O signalling conventions operate with large voltage swings, reference signal levels and receive thresholds to power supplies, and send signals in only one direction at a time. The net result is that conventional signalling techniques squander the scarce I/O pads.

We have described a high-performance bidirectional signalling method and its implementation with CMOS circuits. Our method uses small voltage swings (250mV) to minimize power dissipation and self-induced noise. We use current mode signalling to completely isolate the transmitted signal levels and receiver threshold from the noisy power supplies. Finally, we send signals simultaneously in both directions over a single point-to-point link between two chips – in effect doubling the effective I/O bandwidth.

Our circuit implementation uses a cascode current source driver to minimize power supply coupling. Current steering is used to eliminate the initial overshoot problem, reduce self-induced power supply noise, and reduce noise coupled through the $V_T$ lead inductance. The driver rise-time is controlled to minimize reflections, crosstalk, and self-induced noise. Our receiver is implemented as a differential CMOS amplifier. In $0.8\mu$ CMOS the amplifier achieves a gain-bandwidth product of 2GHz enabling operation at 500Mbits/s.

To gain experience with high-performance bidirectional signalling, we have implemented three test chips. The results from two of these experimental chips are reported here. Using a $2\mu$ CMOS process, we have successfully demonstrated bidirectional signalling at rates in excess of 100Mbits/s. However, experience with these prototype chips has led us to modify our driver design to reduce its noise generation and increase its noise immunity.

The current-mode, low-voltage swing, bidirectional signalling method we describe here is an attractive alternative both to (1) the conventional full-voltage swing CMOS signalling in use in many VLSI systems and (2) the low-voltage swing power-supply referenced signalling common in ECL systems. In the first case it dramatically reduces the power required to achieve a given data rate. In both cases it greatly increases noise immunity hence permitting operation at smaller signal levels.

This signalling method is suitable for use in all types of digital systems that use point-to-point signalling. It is particularly appropriate for connections between similar VLSI components, such as between the routers in a multicomputer interconnection network [1]. In these cases, there is no need to interface to any existing signalling standard and only one chip type needs to be designed to take advantage of the benefits of bidirectional current-mode signalling.

We envision industry adopting a signalling convention similar to the one described here as clock frequencies increase and straightforward extensions of TTL compatible signalling become impractical because of power dissipation and the number of supply leads required to keep power supply noise under control. Adapter chips will be required at first to allow chips using high-speed signalling to interface to chips using existing standards.

With today's $0.8\mu$ CMOS chips, data rates are limited both by receiver gain-bandwidth product and by the bandwidth of the transmission media. With high-performance packages and specially designed substrates, one can make the receiver circuit the rate limiting factor. In the long run, however, as circuit speeds continue to increase with shrinking device dimensions and as we gain the benefit of the high $f_T$ bipolar devices found in many modern BiCMOS processes, the transmission media will be the rate limiting factor. Within a very few years, the package design and the characteristics of the substrate used to connect the chips will largely determine the available inter-chip bandwidth.

## Acknowledgment

The research described in this paper was supported in part by the Defense Advanced Research Projects Agency under contracts N00014-88K-0738 and N00014-91J-1698, in part by a National Science Foundation Presidential Young Investigator Award, grant MIP-8657531, with matching funds from General Electric Corporation, IBM Corporation and AT&T Corporation.

## References

[1] William J. Dally. Network and processor architecture for message-driven comput-

ers. In Suaya and Birtwhistle, editors, *VLSI and Parallel Computation*. Morgan Kaufmann, 1990.

[2] Thaddeus J. Gabara and David W. Thompson. High speed, low power cmos transmitter-receiver system. *IEEE Transactions on Computers*, pages 344–347, May 1988.

[3] Lance. A. Glasser and Daniel. W. Dobberpuhl. *The Design and Analysis of VLSI Circuits*. Addison-Wesley, 1985.

[4] Thomas Knight and A. Krymm. Self terminating low voltage swing cmos output driver. *IEEE Custom Integrated Circuits Conference*, pages 289–292, 1987.

[5] Kevin Lam, Larry. R. Dennison, and William. J. Dally. Simultaneous bidirectional signalling for ic systems. *ICCD : VLSI in Computers and Processors*, pages 430–433, 1990.

[6] Carver Mead and Lynn Conway. *Introduction to VLSI systems*. Addison-Wesley, 1980.

[7] Signetics. *Signetics ABT Handbook*. Signetics, 1992.