A Method for Skew-free Distribution of Digital Signals Using Matched Variable Delay Lines

Thomas Knight and Henry M. Wu

Artificial Intelligence Laboratory
and
Department of Electrical Engineering and Computer Science

Massachusetts Institute of Technology

A.I. Memo No. 1282 March, 1992

Abstract

The ability to distribute signals to all parts of a circuit with precisely controlled and known delays is essential in large, high-speed digital systems. We present a technique by which a signal driver can adjust the arrival time of the signal at the end of the wire using a pair of matched variable delay lines. We show how this idea can be implemented requiring no extra wiring, and how it can be extended to distribute signals skew-free to receivers along the signal run as well as the receiving end. We demonstrate how this scheme can be implemented as part of the pad and scan logic of a VLSI chip.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-89-J-3202 and by the National Science Foundation under grant number MIP-9001651.

Introduction

The ability to distribute digital signals to all parts of a digital system with known and adjustable delays has become essential in modern high-speed designs. We present a novel technique by which a signal driver can precisely control the signal's arrival time at the end of the wire. The approach involves measuring the round-trip delay of the signal and adjusting it with a pair of matched delay lines. This technique requires no instrumentation or special hardware at the end point, and can be implemented without extra wiring from source to destination. A variation on the basic implementation allows receivers along the signal run to compensate for shorter arrival times, so that all receivers on the same run can receive the signal with a single known delay and virtually free of skew. This method can readily be implemented using well-known circuit forms in various technologies, and is well-suited for incorporation into the boundary scan logic of a VLSI chip.

Skew-free distribution of signals is most essential in the fanout of clocks in synchronous designs. In large digital systems, such as supercomputers and multiprocessors, clock signals must be distributed to multiple parts of a PC board, which today can measure over two feet on the side. Often these clocks must be distributed among a number of boards, for example to facilitate synchronous accesses from the CPU to memory boards or communication among different processors. The propagation velocity of the clock wires cannot be accurately predicted due to variations in process and material. As the maximum distance through which these signals travel increases, the possible skew between two copies of the same clock goes up, causing potential setup or hold violations.

In general, the uncertainty or variation in the arrival time of a clock signal to all its destinations must be subtracted from its intended period to obtain the usable cycle time. For example, a CPU with a cycle time of 20 ns and 4 ns of skew in clock distribution has a usable cycle time of only 16 ns. In other words, 20% of the usable cycle is lost due to skew. The situation worsens as the cycle time of the system falls, making a strategy for distributing clocks without skew indispensable.

In synchronous transmission, data traveling on wires with delays longer than a clock period may cause metastability if arriving data violates the setup time required before the qualifying edge of the clock. [Rettberg] [6] describes one solution to this problem. A remedy is to precisely control the amount of time spent in transmission so that the clock always samples the data during the middle of a data cell. In high-speed designs it is often the case that multiple data cells are stored on the wire; the wire in effect acts as multiple pipeline latches. Here, the system needs to quantify the exact delay (in terms of numbers of pipestages) introduced by the wire.

Many approaches have been developed to deal with the problem of clock synchronization, although none are very effective at controlling the absolute phase of non-repetitive signals. The most common method uses a Phase-Locked Loop (PLL) at the clock receiver and the distribution of a slow master clock. The PLL multiplies the clock frequency and allows the phase of the regenerated clock to be adjusted. Although this method is effective for synchronizing the frequency of the local clock oscillator, the clock phase cannot be guaranteed because the phase of the reference - the copy of the master clock received - has already been varied by the delay of the distribution wire. In other words, a high edge rate, skew-free phase reference is still essential to properly compensate for phase error. If the receivers are physically far apart or if their positions are configuration dependent, the problem of skew will remain. Recently [Pratt/Nguyen][5] describes a method for using PLL's to synchronize a large number of local oscillators in both frequency and phase by averaging reference signals from the local oscillators' neighbors. However, their method requires precise placement of the phase-detector or analog error signals to be transmitted between oscillators. None of these methods work for non-repetitive signals.

The PLL approach is often extended to allow correction for skew introduced by clock redistribution drivers. The clock driver is placed in the control loop of the PLL, adjusting the output of the amplified clock signal to be in phase with the reference clock input received. Examples of off-the-shelf chips designed to implement this idea include the Gazelle GA1110E and the Motorola MC88915. [Johnson] [3] uses PLL's to compensate for skew introduced by process variations in chips participating on a tri-state bus. A high-speed version of the Intel 486 chip uses PLL's to eliminate the delay between external and internal clocks caused by the on-chip clock driver.[9]

The approach that has so far been most effective in controlling skew is tight control of the length of signal runs. This works, but is extremely difficult to implement in densely populated card cages, backplanes, or PC boards. Autorouting is no longer possible, and the resulting long wires might lead to other problems, such as crosstalk. Furthermore, this approach will not work for distributed receivers on the same wire. Consequently, more signal lines are needed for point-to-point wiring, leading to further problems with skew and routing. Many high-performance supercomputers, such as those made by Cray, are designed using this discipline. [Greub][2] uses crystal-stabilized variable delay lines in place of matched-length wires, but the amount of delay is manually adjusted according to measurement at the receiver.

A Two-Wire Approach

Our basic idea is illustrated in Figure 1. We require that the signal sender (S) communicate with a single receiving point (R) with a forward and a reverse path that are of the same electrical length (or propagation delay). We assume that the delays in buffering the signal on output and input are the same at T_{pd}. If the time it takes for the signal to travel from S to R and back is $2T_{line}$, then it is guaranteed that the time it takes for the one-way trip is T_{line}. If we then insert delay lines of similar delay T_{delay} at the two endpoints of the round trip, the total round-trip delay is $(T_{pd} + T_{line} + T_{delay} + T_{delay} + T_{line} + T_{pd})$ or $2(T_{pd} + T_{line} + T_{delay})$, while the one-way delay will be $(T_{pd} + T_{line} + T_{delay})$. By adjusting both delay lines in tandem, we can guarantee that the arrival time of the signal at R is always exactly half of the total delay. We can therefore adjust both the total and one-way delay to be any value required[1] by phase locking the return signal to a reference delay at the sender: the arrival time at the receiver will always be one-half that of the reference delay.

A key feature of this technique is that it allows control of the arrival time of a signal with no adjustments or measurements necessary at the receiving end. The reference delay and phase adjustment are needed only at the sender, and thus can be limited to a small area where wire lengths are negligible.

If the propagation delay along the wire does not change, the delay line adjustment can be done once when the system is initialized. The signal used to

[1] The detection of phase may be ambiguous if the delay is longer than one signal period.

calibrate the line should be of short transition time and low repetition rate. This is because the phase detector can only unambiguously distinguish the return of the signal within one period, and the sharper the edge the more precise the detector can measure the time the edge returns.

An obvious drawback to this basic approach is that two wires are needed from the sender to each receiving site. Later we show how we can measure the round-trip delay with only a single wire. In practice even the two-wire approach is not hard to achieve. For example, in the case of PC boards, it is easy to modify an autorouter to always route the forward and reverse path next to each other, since this is similar to routing one thicker wire.

No extra wires are needed when there are multiple signals that have to travel from the source to the destination. Since we only need the return path when we calibrate the length of the wire, we can use two wires of the same length when the calibration occurs. Once we determine the length of this reference path, we can then adjust the delay of each of the other remaining wires using this measured arm as the reference. The arrival times for these other wires will not be one-half of the total delay, but it is possible to adjust for this knowing what the reference arm's delay is. Once the calibration is complete the reference arm may be reused as a regular signal wire.

Figure 1: The basic two-wire idea: the phase-detector locks the round-trip delay to the reference by controlling two matched variable delay lines in tandem. The arrival time at signal's destination is guaranteed to be half of the total delay, i.e. half of the reference delay.

Eliminating the Reverse Path

It is possible to eliminate the reverse path used to measure the round-trip delay by taking advantage of transmission line bounce. If we arrange for the receiver to have high impedance compared to the characteristic impedance of the line, i.e. we underterminate the line, a reflected wave of the same sign as the outgoing wave will appear at the driver after one round-trip delay. We can measure the arrival time of this reflected wave to properly adjust the matched delay lines. Series termination at the driver allows us to observe the reflection and prevents further bounces. If the series termination resistance is exactly the impedance of the wire, then the voltage at the wire end of the termination resistor doubles when the reflected wave returns. The new configuration is shown in Figure 2.

Figure 2: Using the reflected wave to measure return trip delay. The wire end of the series termination resistor sees a step in voltage. The second edge comes exactly one round trip delay after the first edge. The sender can detect the second edge and use that to feed the phase detector.

Figure 3: Plot of voltages vs. time for the circuit in Figure 2. When the delay lines are correctly adjusted, the reflected wave arrives at the phase detector after one reference delay.

This is a particular application of a more general technique: a signal sender can compensate for the characteristics of the line it is driving by figuring out the parameters of the line through measurement of the reflected wave. In this case we measure and compensate for the propagation delay. It is also possible to measure the line impedance [4] or frequency response characteristics, for example.

Also interesting to note is that we can view this setup simply as an example of full-duplex communication on a single wire. The crucial functionality required is that the sender be able to feed a signal to the receiver, which in turn must send the signal back so that the former can measure the round trip delay. Since the sender knows what it is putting on the line, by superposition the signal returned is the content on the line with the sender's own signal subtracted. Similarly the receiver can cancel out the return signal it sends and obtain the original signal. So instead of making use of transmission line bounce, the receiver can properly terminate and buffer the incident signal before sending it back with its own driver; the sender decodes the return signal by subtracting the outgoing signal from the line. Figure 4 illustrates this idea.

Figure 4: Full-duplex communication between sender and receiver. Each device sends and receives a signal. The received signal is simply the content on the wire with the signal being sent subtracted.

The hybrid coil in a telephone, in use for decades, is an example of a circuit using this idea to effect full-duplex transmission. [Dally] [1] uses a similar approach for full-duplex communication between nodes in a multiprocessor.

Distributed Receivers

Once we can control the arrival time of the signal at the end point, it is also possible to allow receivers distributed along the line to receive the signal at the same time. The method calls for the distributed receivers to detect the arrival of the signal on both the forward and reverse trips. The arrival time at the end is guaranteed to be the midpoint of these two instances. If we take the forward arrival and delay it through two matched variable delay lines and phase lock the delayed signal to the reverse arrival, then a tap in between the two delay lines will present a signal whose timing matches that of the signal at the end of the line. The scheme is shown in Figure 5.

Figure 5: Compensation at distributed receivers. A distributed receiver detects both the incident and returned signals. A pair of matched variable delay lines slow the former until it coincides with the latter. A tap in the middle of the two delay lines has a signal similar in timing to that of the signal at the end of the wire.

The elegance of this approach is that its implementation requires a circuit very similar to the one used in the original matched delay line technique to control skew. No new circuits components are needed.

Limitations

There are some practical limitations and drawbacks in practical implementations of the above techniques. The most important of which is perhaps the failure of the method to account for the variations in the speed of the signal buffers between input and output due to differences in circuit topology and loading conditions. This can be solved by: 1) artificially slowing down the input buffer, 2) compensating in the adjustment of the variable delay lines, or 3) connecting the return signal through an output buffer circuit seeing a similar load as the real output driver. None of these remedies are optimal, but each will probably be sufficient. Since the buffer are part of the control loop, part-to-part variations are in fact compensated for in our schemes.

Another major drawback is the difficulty to build delay lines with small minimum delays. Since distributed receivers require that the incident and returned signals be spaced at least two delay-line delays apart, a large minimum delay means that these receivers cannot be too close to the end of the wire.

Measuring transmission line bounces is easier to describe on paper than it is to implement. Real-life transmission lines with distributed capacitive and inductive loads have hard-to-predict behaviour. Often multiple bounces due to these loads occur. A simple threshold detection scheme may not be adequate to pick out the reflected wave from the end of the line. Even in point-to-point transmission, the dissipation and limited high-frequency response of the transmission line generate effects that must be accounted for. Also, using a single wire for full-duplex transmission will lower the noise margins.

Fundamentally, these methods rely on the ability to perform precise triggering on voltage levels present on the transmission lines. Any source of noise in voltage level will introduce errors in timing and lessen the effectiveness of our schemes. More sophisticated means for measuring and compensating the parameters of the transmission line are needed for better results.

VLSI Implementation

The techniques described above requires three basic circuit components: a pair of matched variable delay lines, either a threshold detector/comparator or a differential amplifier, and a phase detector. Optimal implementations of these circuit elements will depend largely on the technology used.

Ideal delay lines for this application have a small minimum delay (optimally under 1 ns), a large range, have fine adjustment levels, and are easy to match. In technologies with very fast gate delays, such as ECL or very fast CMOS, the matched variable delay lines can be implemented as a brigade of inverter/buffers feeding a large multiplexor. The advantages of this implementation is that only digital control signals are necessary, and that the parts are readily available in semi-custom processes such as gate-array or standard cells. The drawback is that the granularity of adjustment is coarse (equal to one inverter delay), the minimum delay is large (one multiplexor delay), and the delay is subject to temperature and voltage variations.

In MOS technology a common adjustable delay element is an RC delay line. The capacitor is implemented by a large area of diffusion, and the resistor is a pass transistor with its gate tied to the control voltage.[7][3] This implementation has the advantage that very fine adjustments are possible. The drawbacks are that the chip area required to implement the capacitor is large, that it is more difficult to match two delay lines, and that an analog control voltage needs to be generated. Variable delays can also be implemented by interpolating the delay between gates. A bipolar implementation of this idea is described in [Walker] [8].

In the one-wire approach where the reflected wave is measured, we need either two kinds of logic gates with different trip points or a differential amplifier for subtracting the sender's signal. In ECL it is possible to adjust the threshold by varying the reference voltage fed to one side of the emitter-coupled pair. Generating the two appropriate threshold voltages may not be easy. In CMOS proper scaling of the transistors serves to vary the threshold. The CMOS DMC differential comparator is a fast circuit that can be used to implement a range of trip points. It can also be used to implement an effective differential amplifier.

The requirements for the phase detector is simple in this application. Since the phase detector need not lock to a dynamically changing signal, it does not need to provide output to indicate the magnitude of the error in phase; only the direction of the error is required. An edge-triggered register, for example, can provide simple detection. If an analog delay voltage is needed an XOR-type phase detector can be used, with the penalty that phase detection range is reduced from 360° to 180°. Alternatively the digital control signal can be converted to analog form via a cheap, mostly digitally implemented D/A conversion technique, such as Pulse-Width Modulation (PWM).

Scan Logic Implementation of Compensation

If the wire delays do not vary once the system is configured, the compensation process (delay line adjustments) can be performed once at setup. In this case the adjustments can be controlled by the boundary scan logic of the chip. Not only does this simplify the control logic of the PLL, it makes the amount of compensation applied available to the rest of the system. The system can then gain knowledge of the absolute delay in time or pipestages inherent in the critical paths of its communication wires. Control logic can then manage the internal pipeline delays and bypasses to compensate for variable wire delays. The technique is another example of compensating for manufacturing and design variation in interconnect by sophistication in on-chip circuitry.

Conclusions

We have presented a novel technique that allows skew-free distribution of digital signals with known and controllable arrival times. The technique requires adjustments and measurements only at the sender. Our method is based on measuring the round trip delay of a signal and then adjusting it with a pair of matched delay lines. This technique can be modified to work without extra wires, and is effective for receivers in the middle as well as the end point of a wire. This method can be readily implemented using well-known circuit forms in ECL and CMOS technologies, and incorporates well into the boundary scan logic of a custom VLSI.

References

[1] Kevin Lam, Larry R. Dennison and William J. Dally. Simultaneous Bidirectional Signalling for IC Systems. In *IEEE International Conference on Computer Design: VLSI in Computers & Processors*, 1990.

[2] Hans J. Greub. *Apparatus for Skew Compensating Signals.* United States Patent 4,833,695. 1989.

[3] Mark G. Johnson. A Variable Delay Line Phase Locked Loop for CPU-Coprocessor Synchronization. In *IEEE International Solid-State Circuits Conference*, 1988.

[4] Thomas Knight and Alex Krymm. Self Terminating Low Voltage Swing CMOS Output Driver. In *IEEE Custom Integrated Circuits Conference*, 1987.

[5] Gill A. Pratt and John Nguyen. Synchronizing Oscillators in Mesh-Connected Networks. *In preparation.* MIT Laboratory for Computer Science.

[6] Randall D. Rettberg and Lance A. Glasser. *Digital Phase Adjustment.* United States Patent 4,700,347. 1987.

[7] Jonathon Taft. *A Calibrated Digital Delay Line Implemented in VLSI.*, SB Thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA (1983).

[8] Richard Walker, Jieh-Tsorng Wu, Cheryl Stout, Benny Lai, Chu-Sun Yen, Tom Hornak, and Pat Petruno. In *IEEE International Solid-State Circuits Conference*, 1992.

[9] Ian A. Young, Jeff K. Greason, Jeff E. Smith, and Keng L. Wong. A PLL Clock Generator with 5 to 110MHz Lock Range for Microprocesors. In *IEEE International Solid-State Circuits Conference*, 1992.