Eliminating Non-Determinism During Test of High-Speed Source Synchronous Differential Buses

Kartik Mohanram and Nur A. Touba

Computer Engineering Research Center Department of Electrical and Computer Engineering University of Texas at Austin, Austin, TX 78712-1084 E-mail: {kmram, touba}@ece.utexas.edu

Abstract

The at-speed functional testing of deep sub-micron devices equipped with high-speed I/O ports and the asynchronous nature of such I/O transactions poses significant challenges. In this paper, the problem of nondeterminism in the output response of the device-under-test (DUT) is described. This can arise due to limited automated test equipment (ATE) edge placement accuracy (EPA) in the source synchronous clock of the stimulus stream to the high-speed I/O port from the tester. A simple yet effective solution that uses a trigger signal to initiate a deterministic transfer of test inputs into the core clock domain of the DUT from the high-speed I/O port is presented. The solution allows the application of at-speed functional patterns to the DUT, while incurring a very small hardware overhead and trivial increase in test application time. An analysis of the probability of non-determinism as a function of clock speed and EPA is presented. It shows that as the frequency of operation of high-speed I/Os continues to rise, non-determinism will become a significant problem that can result in an unacceptable yield loss.

1. Introduction

The increasing demand for throughput has accelerated the move away from hierarchical shared bus architectures like PCI and PCI-X towards high performance, packet switched architectures like RapidIOTM [Bouvier 00], InfinibandTM [Infiniband 01], and HyperTransportTM [HyperTransport 01]. These protocols span a wide range of applications, both general purpose and otherwise.

Hierarchical shared bus topologies have failed to meet the data bandwidth requirements and increased system concurrency needs in today's demanding high-speed, high-performance processing environments. The fact that a device has to wait for the bus arbiter to complete an existing transaction before serving the next request poses a severe limitation on the available bandwidth on shared buses. Improvements like increasing the bus width and the bus frequency not only counter each other due to skew between signals but also increase the pin count on the device. Moreover, routing issues constrain the number of devices that can directly communicate with each other as the complexity of shared bus protocols increases. It is interesting to note that today's 133 MHz PCI-X standard allows only 2 devices per bus segment, and is for all practical purposes a pointto-point connection [Jaenicke 01].

These limitations have led to the emergence of high-speed, source synchronous, low power-low voltage differential signaling (LP-LVDS) parallel (or serial) buses such as the RapidIOTM interconnect architecture (optimized for intra-system interconnections). Such high-performance interconnect technologies employ packet switching to transfer data from source to destination thereby providing a high degree of scalability and are projected to attain performance levels scaling to 10 Gbps and beyond. Other examples of emerging high-speed interconnect architectures include InfinibandTM (optimized for system area networks) and HyperTransportTM (optimized for system level interconnections).

The development of packet switched interconnect architectures is not without its share of challenges for the test community. The challenges are manifold, requiring innovative ideas not only from an automated test equipment (ATE) hardware perspective but also from a design-for-test (DFT) and test generation perspective. There has been some work in the area of ATE hardware development to meet the requirements of supply and capture of high-speed differential stimuli to the device-under-test (DUT). In [Oshima 01], a pin electronics integrated circuit capable of testing high-speed differential buses up to 3 Gbps is presented. In [Keezer 01], a multiplexing / de-multiplexing approach to realize higher performance from existing ATE is presented.

Despite high levels of structured testability provided by scan and built-in self test, deep sub-micron designs, with multiple clock domains and internal asynchronous boundaries generally require a full complement of "at-speed" functional tests to be exercised in test mode. This is important not only for speed binning but also for detecting complex defects that are only uncovered when the full capabilities of the DUT are exercised [Maxwell 00].



Functional tests for high-speed I/O ports use a stream of stimulus packets sent to the receive port of the DUT. The DUT processes these packets and the tester monitors the response packet stream from the DUT. However, non-determinism can be introduced by external inputs as well as the presence of asynchronous boundaries internal to the DUT such that the output from the DUT is not cycle deterministic. One such asynchronous boundary inherent to the DUT is between the receive logic clock domain of the high-speed I/O port and the core logic clock domain. The receive logic is driven by the source synchronous clock (henceforth the Rx clock) from the tester, that is asynchronous with respect to the core clock of the DUT. It is important that the tester not classify a defect-free device as a failure just because the received response did not match the expected one due to non-determinism. Non-determinism can be introduced due to jitter and limited ATE edge placement accuracy (EPA) in the source synchronous clock of the stimulus stream to the DUT from the tester. The frequency of operation of high-speed differential buses has increased at a rate much faster than the rate of improvement in ATE EPA, eroding large margins of comfort that were previously available. As this trend continues, the ambiguity of just when the ATE generated Rx clock is captured by the DUT's core clock domain is increasing and guaranteeing cycle-for-cycle deterministic behavior of the DUT is becoming a challenge.

In this paper, we present a DFT technique aimed at removing the non-deterministic behavior in the response stream from the DUT at high performance rates. We also provide an analysis of the probability of non-determinism as a function of clock speed and EPA and show that as the frequency of operation of high-speed I/Os continues to rise, non-determinism will become a significant problem that can result in an unacceptable yield loss.

The rest of the paper is organized as follows. In Sec. 2, we describe the test environment for a high-speed I/O port and discuss the problem of non-determinism in greater detail. In Sec. 3, we present the proposed solution to the problem. In Sec. 4, we present a model for the EPA of a tester, and analyze the problem of non-determinism using an analytical framework. Section 5 is a conclusion.

2. Problem of non-determinism

The DUT in the test environment is shown in Fig. 1. The DUT has a core clock, a Tx (transmit) clock, and an Rx (receive source synchronous) clock. The core clock is created by an on-DUT PLL from an ATE supplied source and is phase aligned with the incoming reference clock from the ATE. The core clock and the Tx clock may be integer multiples of each other (including 1-1) and may also include non-integer modes such as 3-2, 5-2, etc. It is usually the case that the core clock is a slow one while the Tx clock, used by the DUT to transmit source synchronous data, is a fast one.

The Tx clock going into the ATE from the DUT is used by the ATE to capture the Tx data from the DUT source synchronously. We assume throughout that the tester is able to receive the source synchronous clock (the Tx clock of the DUT) of the response stream, capture the Tx data, and that there is no problem of resolution here. Some recent ATE provide the ability to capture a source synchronous bus with its own clock at frequencies up to 2.5Gbps [Teradyne 00], [Agilent 01], [Schlumberger 02].



Figure 1. DUT in the test environment

Packets to the high-speed I/O port can be classified into two types – data and control. Control symbols are issued to initiate a request for and to acknowledge the receipt of packets in the system, as well as for system maintenance. Control symbols are also used for flow control, when retry and idle symbols are inserted to manage the flow of packets in the system. LP-LVDS buses are never tri-stated, but issue idle control symbols into the transmit stream when not processing any data.



Figure 2. Sample input stream from ATE to DUT

Most packet switched I/O architectures use an elasticity buffer in the receive circuitry to temporarily store packets before transferring them into the core clock domain. The source synchronous data (Rx data) that is embedded in the stimulus stream from the ATE is de-serialized as it is clocked into the elasticity buffer with the source synchronous clock (Rx clock). Several cycles later, once the data is stable, it is pulled out of the



elasticity buffer using the slow core clock. An asynchronous boundary is explicitly crossed when the packet is transferred into the slow core clock domain of the DUT. Figure 2 shows a sample input stream to the DUT from the tester. Note that the I/O port functions in double-data-rate (DDR) mode, since packets are clocked on the rising as well falling edges of the Rx clock.

The packet switched nature of the I/O protocol can complicate test response analysis in the presence of asynchronous boundaries internal to the DUT. In an ideal environment, as modeled in simulation, all events on the tester and the DUT occur at known absolute times. However, in the "real world", there is the issue of ATE EPA which causes variability in when exactly events occur. The EPA is a measure of the accuracy with which the ATE can place an input drive edge at exactly the same point in time relative to a common reference. While these problems have always existed, they were never a cause for concern because the EPA of the ATE was so small compared to cycle times such that a "sweet spot" was always found for timing edge sets where the DUT would respond in a deterministic manner. However, in [ITRS 01], it is estimated that off-chip I/O speeds have improved at a rate of 30% per year, while ATE EPA has improved at a rate of 12% per year. Thus, as cycle times of high-speed buses are rapidly decreasing and EPAs are not decreasing at the same rate, we are approaching a point where non-determinism becomes an issue. A detailed analysis of the probability of non-determinism with respect to clock frequencies and tester EPA is provided in Sec. 4.

Non-determinism due to limitations in tester EPA occurs because of ambiguity in just when the source synchronous clock from the tester arrives at the DUT. When crossing the asynchronous boundary internal to the DUT, the cycle in which the packet is read from the elasticity buffer by the core logic could be misaligned by one internal core clock cycle. This variability in when the packet is received can cause the subsequent output stream to be different than expected (without the presence of a defect).

Non-determinism is illustrated in Fig. 3. Packet a is clocked into the core clock domain in cycle 1 as expected. Now consider packet b from the tester, which is expected to be clocked in from the elasticity buffer in cycle 3 of the DUT's slow core clock. Limitations in the EPA of the tester clock cause it to arrive later such that it is clocked into the elasticity buffer in cycle 4 of the DUT's core clock (instead of cycle 3 as originally expected). This causes the test stimuli that are synchronized to be offset by one core clock cycle – something that simulation did not predict, and something that can cause a deviation from the expected response (as will be elaborated on

shortly). The ATE can end up classifying the part as defective simply because the packet was received a cycle later than intended. Variability in packet arrival times is not a problem in normal operation because the functional logic can handle packets in any order and will provide a correct output response for the order it receives the packets in. The problem exists in test because it is necessary that the output response match what is expected by the tester so that the part can be correctly classified.



Figure 3. Example of non-determinism where packet *b* is read in a different core clock cycle

2.1 Types of non-determinism

Two forms of the non-determinism in the output stream are possible. The first is that the response from the DUT occurs in the same order as the expected response, but is simply delayed by an extra idle in the Tx stream. This can be handled by "match mode" on testers today where they automatically compare and synchronize the actual and expected data streams. The second, and more difficult, case to handle is when the DUT issues all the test response packets and controls correctly, but in an order other than that predicted by simulation. This can occur because the order in which things are processed by the core logic is altered due to the variability in when the packet is received with respect to the other test stimuli synchronized to the core clock.

2.2 Possible solutions

A naïve solution to this problem would be to test the DUT at a slower speed so that the input stream is less likely to be affected by any jitter in the fast source synchronous clock from the tester and the slow core clock of the DUT. Although a viable option, a direct testing of the DUT's capabilities through "at-speed" functional tests is indispensable especially since such tests can uncover complex defects undetectable at lower speeds of operation.

A second option would be to include an explicit post-processing phase, where an algorithm is used to examine the contents of the capture RAM of the ATE to determine if the response stream can be matched to that from a non-defective part. The algorithm would try to match the actual response to one that could have been obtained from a non-defective part by examining the places where mismatch occurred, and try to fit that with a possible sequence of valid events in the DUT. Depending on the extent of non-determinism in the response, the test data volume (both input and output), and the complexity of the algorithm used, the costs associated with this approach could be prohibitive.

3. Proposed solution

The basic idea is to ensure that the DUT receives data from the tester on deterministic cycle boundaries of the slow core clock, i.e., to ensure that the DUT clocks in packets from the elasticity buffer into the slow core clock domain in a deterministic fashion. Since the DUT processes test packets in a deterministic way once they enter the slow core domain, the output response sent to the tester will also be deterministic.



Figure 4. Use of the trigger signal to synchronize the reading of packets into the core clock domain to eliminate non-determinism

We propose the use of a trigger signal between the DUT and the tester that is another primary input synchronized to the slow core clock. Whenever a test packet is to be clocked in across the asynchronous

boundary, the trigger signal is asserted a fixed number of cycles before the desired event. When the trigger is asserted, a counter is started on the DUT that prompts the reading of the elasticity buffer a fixed number of cycles later. The timing of the trigger signal is computed so that the packet will always be ready in the elasticity buffer even with the worst-case arrival time taking the EPA of the source synchronous clock from the tester to the DUT into consideration. If the packet arrives earlier, it is just held in the elasticity buffer until the predetermined core clock cycle. The trigger signal ensures that the packet crosses the asynchronous boundary in a deterministic fashion.

Figure 4 shows how the proposed solution will work to resolve the problem presented in Fig. 3. In this case, the trigger is asserted such that packets are clocked in cycles 2 and 4 of the slow core clock of the DUT. If the packet arrives at the expected time, it is held in the elasticity buffer for one core clock cycle before it is read (as shown in the upper part of the diagram). However, if jitter and limited tester EPA cause the packet arrival time to be delayed, it is still read in during the same core clock cycle as the previous case (as shown in Fig. 4(b)). Hence the behavior is deterministic irregardless of the packet arrival time.

3.1 Overhead of the proposed solution

The DFT hardware for implementing the trigger is shown in Fig. 5. The trigger signal initiates the counter, and the counter gates the packet-ready signal when in test mode. The delay in clocking packets across the asynchronous boundary leads to the introduction of extra idle symbols in the response stream from the DUT. However, this is something than can be deterministically predicted. Note that there is a minor latency penalty to be paid for using the trigger scheme (one extra core clock cycle for each packet read). An additional drawback is that the DUT operation is not completely a normal



Figure 5. Proposed solution



functional mode, but the difference in the logic used in normal and test mode is minimal.

This scheme also comes at the expense of increased hardware complexity on the DUT. The receive circuitry of the high-speed I/O port of the DUT has to be modified to realize the proposed solution in test-mode. A counter that initiates the transfer of test packets a fixed number of cycles after the trigger is asserted needs to be implemented. Note that an extra pin is not required on the DUT to support the trigger signal since one of the primary inputs of the DUT can be multiplexed internally to function as the trigger signal in test mode.

4. Analysis of probability of non-determinism

In this section, we analyze the probability of nondeterminism in the output response of the DUT as a function of the Rx clock frequency and the EPA of the ATE. We address the important issue of when the problem of non-determinism becomes significant enough to warrant the implementation of the proposed solution.

4.1 EPA model

Since the EPA of the tester is a measure of the accuracy with which the tester can position any driven edge with respect to an absolute reference in time, it is possible to model the distribution of the placement of the driven edge by a Gaussian distribution about the reference point [Dalal 99]. The EPA of the tester is equal to three times the variance (σ) of this Gaussian distribution. Note that this translates to a very high probability (greater than 0.99) that the driven edge is within $\pm 3\sigma$ of the reference point. This is illustrated in Fig. 6.



Figure 6. Gaussian distribution for driven edge and EPA

Thus, it is apparent that a high variance for this distribution corresponds to a poor tester EPA and vice versa. The Gaussian distribution for both the Rx clock (supplied by the ATE) and the core clock (of the DUT) can be used to tabulate the probability of occurrence of non-determinism as a function of both tester EPA and the speed of operation of the high-speed I/O port. Note that the core clock is derived from the reference clock which comes directly from the tester and hence is affected by the EPA of the tester.

4.2 Probability of non-determinism

Consider the final packet f of an input stream that is

to be clocked in with the falling edge A_1 in Fig. 7. The corresponding core clock cycle in which this packet will be read from the elasticity buffer will depend on the relative position of the falling edges A_1 and A_2 . If, say, A_1 occurs before A_2 , the packet will be read into the core clock domain on a core clock cycle *i*. However, if A_1 occurs after A_2 , the packet will be read into the core clock domain on core clock cycle *i*+1. Thus there is a one cycle variability in the core cycle on which the packet enters the core clock domain which can lead to non-determinism. As the absolute time difference between the occurrence of the falling edges A_1 and A_2 shrinks (and becomes comparable to the EPA of the tester), the probability that the order of occurrence of the falling edges A_1 and A_2 differs from run to run increases.



Figure 7. Analysis of non-determinism

Let the absolute time displacement between the occurrence of the two events A_1 and A_2 (in the simulation environment) be Δ . σ_{Rx} and σ_{core} are the variance of the Rx and the core clock probability distributions (Gaussian), respectively. Let the order of events in the simulation environment be A_1 precedes A_2 ($A_1 \rightarrow A_2$). Then, the probability distribution (Gaussian) of the placement of the edge A_1 (assumed to occur at an instant y), relative to the placement of the edge A_2 (which is used as the reference, with mean $\mu = 0$), is given by:

$$P(A_1 \text{ occurs at } y) = \exp\left(\frac{-(y - (-\Delta))^2}{2 \cdot \sigma_{Rx}^2}\right)$$

If the event A_2 occurs at an instant *x*, the probability that A_1 occurs after *x* (and hence, in an order other than the simulation predicted one) is given by the integral:

$$P(A_1 \text{ occurs after } x) = \int_{x}^{\infty} \exp\left(\frac{-(y+\Delta)^2}{2 \cdot \sigma_{Rx}^2}\right) dy$$

The above integral is then evaluated over the interval $[-\infty, +\infty]$ to compute the total probability of interchange of the events A₁ and A₂. Thus:

$$P(A_2 \to A_1) = \int_{-\infty}^{\infty} \int_{x}^{\infty} \exp\left(\frac{-(y+\Delta)^2}{2 \cdot \sigma_{Rx}^2}\right) dy \exp\left(\frac{-x^2}{2 \cdot \sigma_{core}^2}\right) dx$$

When Δ is 0, i.e., when the time difference between the events A_1 and A_2 in the simulation environment is 0, $P(A_2 \rightarrow A_1)$ evaluates to 0.5. We show the value of



 $P(A_2 \rightarrow A_1)$ as a function of the time difference Δ in the semi-log plot in Fig. 8.



Figure 8. Semi-log plot of $P(A_2 \rightarrow A_1)$ as a function of displacement Δ . (Rx clock EPA, core clock EPA) varies from (25ps, 25ps) to (200ps, 200ps)

To plot the graph in Fig. 8, the EPA of both the Rx clock and the core clock was varied from 25ps to 200ps. On the X-axis, the displacement Δ was varied from 0ps to 600ps in increments of 1ps. For each pair of values of the EPA of the Rx clock and the core clock, the log (base 10) of the probability of interchange of events A₁ and A₂ as a function of increasing displacement Δ is graphed. Note that the probability of non-determinism (i.e., P(A₂ \rightarrow A₁)) diminishes rapidly as the displacement Δ between the events A₁ and A₂ increases. In addition, for a fixed value of Δ , there is a significant increase in the probability of non-determinism as the EPA becomes less precise.



Figure 9. Semi-log plot of $P(A_2 \rightarrow A_1)$ as a function of Δ , when Rx clock EPA is 50ps and core clock EPA varies from 50ps to 200ps

Often the ATE has only a limited number of pins that have very precise EPA, while the other pins have a less precise EPA. In some cases, it may not be possible to use the very precise EPA pins to drive both the Rx clock and the core clock. In that case, the core clock is driven with a less precise EPA. We show a semi-log plot in Fig. 9 where the Rx clock EPA is 50ps while the core clock EPA varies from 50ps to 200ps. In Fig. 10, the Rx clock EPA is 25ps while the core clock EPA varies from 25ps to 200ps. Note that the probability of non-determinism is significantly increased as the core clock EPA becomes less precise.



Figure 10. Semi-log plot of $P(A_2 \rightarrow A_1)$ as a function of Δ , when Rx clock EPA is 25ps and core clock EPA varies from 25ps to 200ps

4.3 Probability of non-determinism for the test session

Packets are latched on both edges of the Rx clock (since it is DDR). There are several edges on the Rx clock during each core clock period. The Rx clock edge that is nearest the falling edge of the core clock will have the smallest displacement Δ (such Rx clock edges will be referred to as critical edges). All the other edges during the core clock period will have a much larger displacement and hence a negligible probability of non-determinism. The best-case value for the displacement Δ for the critical edges is equal to half the period of the Rx clock. Let P_{nd} be the probability that a packet on a critical edge causes non-determinism. Then the probability of non-determinism for the entire test session is is equal to $[1 - (1 - P_{nd})^m]$ where m is the total number of packets that are received on critical edges.



Figure 11. Probability of non-determinism for test session as a function of EPA for different I/O frequencies



In Figs. 11 and 12, we show the probability of non-determinism for the entire test session if *m* is 100 and the displacement Δ for the critical edges is equal to half the period of the Rx clock. Figure 11 shows how the probability of non-determinism for the test session varies with EPA for a given I/O frequency. As the I/O frequency continues to increase, the EPA must rapidly improve to keep the probability of non-determinism low.



Figure 12. Probability of non-determinism for test session as a function of I/O frequency for different EPAs

Figure 12 shows how the probability of non-determinism for the test session varies with I/O frequency for a given EPA. If a tester with a particular EPA is to be used to test successive generations of chips, the probability of non-determinism will rapidly increase as the I/O frequency improves.

Table 1. Probability of non-determinism for test session if m = 100

Rx Clk Freq.	Δ ps	EPA (Rx Clock, Core Clock) (ps, ps)				
		(25, 25)	(50, 50)	(100,100)	(150,150)	(200,200)
1 GHz	500	0	0	0	0	$5.6 \cdot 10^{-6}$
2 GHz	250	0	0	$5.6\cdot10^{\text{-}6}$	0.0201	0.3306
5 GHz	100	0	0.0011	0.8190	0.9997	0.9999
7 GHz	71	0	0.1216	0.9989	0.9999	0.9999
10 GHz	50	0.0011	0.8190	0.9999	0.9999	0.9999
15 GHz	33	0.2256	0.9997	0.9999	0.9999	0.9999

Table 1 gives the probability of non-determinism for a test session (m = 100) for a few example cases. As can be seen from the Table, if a tester with a 50ps EPA is used to test a chip with a 7 GHz I/O port, the probability of non-determinism is 0.1216. This could result in a significant yield loss. From this data, it is apparent that as the rapid increase in I/O frequencies continues to outstrip improvements in tester EPA, non-determinism in the output response of the DUT will be a significant problem. The DFT technique presented in this paper can be used to address this problem.

5. Conclusion

As the I/O frequencies for high-speed source synchronous differential buses continue to rise and ATE EPA becomes relatively less precise, techniques for limiting sources of non-deterministic behavior in the DUT will be needed. In this paper, we have proposed a DFT technique for the elimination of non-determinism that arises due to limited ATE EPA in the source synchronous clock of the stimulus stream to a high-speed I/O port from the tester. It allows the application of at-speed functional patterns to the DUT while avoiding yield loss due to nondeterminism. The proposed DFT scheme requires a very small hardware overhead and trivial increase in test application time.

Acknowledgements

The authors would like to thank Carol Pyron at Motorola for suggesting this topic and for her help in preparing this paper. This material is based on work supported in part by the National Science Foundation under Grant No. MIP-9702236.

References

- [Agilent 01] Technical Specifications, *Agilent 93000 SOC* Series NP-models, Agilent Technologies, 2001.
- [Bouvier 00] Bouvier, D., "RapidIOTM: An Embedded System Component Network Architecture," RapidIOTM Trade Association, 2001.
- [Dalal 99] Dalal, W., and Miao, S., "The Value of Tester Accuracy," *Proc. International Test Conference*, pp. 518-523, 1999.
- [HyperTransport 01] "HyperTransportTM Technology I/O Link: A High Bandwidth I/O Architecture," Advanced Micro Devices, Inc., 2001.
- [Infiniband 01] InfinibandTM Architecture Specification, Vols. 1 and 2, InfinibandTM Trade Association, 2001.
- [ITRS 01] The International Technology Roadmap for Semiconductors, Test and Test Equipment, 2001.
- [Jaenicke 01] Jaenicke, R., "RapidIO raises PCI stature in the box," *EETIMES*, April 2001.
- [Keezer 01] Keezer, D. C., Zhou, Q., Bair, C., Kuan, J., and Poole, B., "Terabit-per-second Automated Digital Testing," *Proc. International Test Conference*, pp. 1143-1151, 2001.
- [Maxwell 00] Maxwell, P., Hartanto, I., and Bentz, L., "Comparing functional and structural tests," *Proc. International Test Conference*, pp. 400-407, 2000.
- [Oshima 01] Oshima, A., Poniatowski, J., and Nomura, T., "Pin Electronics IC for High Speed Differential Devices," *Proc. International Test Conference*, pp 1128-1133, 2001.
- [Schlumberger 02] Technical Specifications, *ITS9000ZX Test System*, Schlumberger Semiconductor Solutions, 2002.
- [Teradyne 00] Technical Specifications, J973-400 VLSI Test System, Teradyne, Inc., 2000.

