Abstract

This paper analyzes the requirements of at-speed functional testing of high-speed devices. We conclude that, although the requirement forecast by the International Technology Roadmap for Semiconductors is excessive, higher accuracy at-speed functional test systems are needed to qualify the high-performance devices anticipated in the next decade. We also conclude that, while challenging, the necessary higher speeds and accuracies can be realized.

1 Introduction

At-speed functional testing has been used to certify leading-edge digital integrated circuits for the last quarter century and more. During this period, many other techniques have also been used, and on occasion the high cost of at-speed functional testing has led to predictions of its demise[1]. Additionally, the ability to perform the required functional tests at anticipated device speeds has been challenged from the standpoint of basic physics.

The principal argument for at-speed functional testing is that if the part executes a full functional test at speed with margins in voltage, temperature, and timing, then we can be comfortably sure it will work properly in its final application.

At-speed functional test equipment has been costly because the electronics and mechanics of applying full functional test patterns at maximum device operating speed plus margin are complex, costly, and power-hungry. The design of such testers is also very demanding on device performance and packaging. Yet the approach survives because of the value added at the process performance limits[2].

At-speed functional testing is further challenged by device complexity. If internal access is not adequate to fully test a device in a reasonable time, then at-speed functional testing won’t help. Access to internal nodes gets more difficult as devices get more complex. Some means to establish and to evaluate all required internal states during execution of a functional test must be available. Structural additions to functional test (sometimes referred to as design-for-test, or DFT) are discussed further in a companion paper[3].

2 Background

Test rate and timing accuracy have long been recognized as the key limiters in at-speed functional testing. NTRS1997, the SIA 1997 National Technology Roadmap for Semiconductors[4] asserts that tester overall timing accuracy (OTA) should be +/-5% of test period at maximum test rate. These numbers are +/-15 to +/-5 ps at 3-10 GHz by 2012, depending somewhat on how one reads the material. If the tester OTA cannot keep up, then higher test rates are of little value. The 1998 update to the now renamed International Technology Roadmap for Semiconductors (ITRS - 1998 update)[5] forecasts that OTA will fall woefully behind.

Tester timing errors come from a number of places, including the characteristics of the components in the timing path, the interconnecting signal paths, and the mechanism by which the tester pin electronics connect to the device being tested (which is lovingly called the Device Under Test, or DUT). Some of these contributions can be systematically reduced by careful calibration, and techniques of automatically calibrating functional test equipment have a long and colorful history[6], [7].

Occasionally, basic physics has been raised as an objection to increased test rate or timing accuracy. Katz[8] points out that the round-trip between the tester electronics and the DUT limits test rate on IO pins on basic physics grounds, but then immediately fineses the problem with a fly-by connection dating back at least 20 years. Keezer and Zhou[9] and Mydill[10] also propose ways around this particular problem. So it is not a physics problem, merely an implementation issue.

Basic circuit noise also presents some difficulties. This physics problem is discussed by Keating[11]. As he pointed out, shot noise and flicker do not appear to be important constraints, and other circuit noise issues are once again due primarily to implementations.

Another limitation was also discussed by Keating (op. cit.). He pointed out that metastability in the capture electronics impacts the potential accuracy, with a reported limitation of 8 ps in the AMD 685 he used as a test case. However, metastability as such is not so limiting. This issue will be covered below.
Fixturing errors and path errors - particularly characteristic impedance mismatches - also limit the ability of a test system to apply and measure high-speed waveforms accurately. However, recent progress in fixtures and probe cards alleviates some of these worries. High-pin-count probe cards with bandwidths up to 9GHz have been demonstrated[12].

Deep-submicron technology has additional, more direct impact on device timing. As the interconnect gets denser and denser, the effects of electrical interaction in the dynamic behavior of interconnects get worse and worse[13]. Timing paths have been shown to be poorly correlated with overall device performance[14].

3 Observations on ITRS Accuracy Forecasts


The ITRS also calculates the yield loss (down-binning) due to the forecast overall timing error over time. The accuracy requirement is based on an assumption of 5% of device period not explicitly justified in the Roadmap document itself.

Using the yield model of Reference [2], we calculated the yield resulting from the ITRS equipment accuracy forecast. We also calculated the accuracy required for 95% yield. The results are exhibited in the following table.

<table>
<thead>
<tr>
<th>Year</th>
<th>97</th>
<th>99</th>
<th>02</th>
<th>05</th>
<th>08</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period (ns)</td>
<td>1.3</td>
<td>1.1</td>
<td>.77</td>
<td>.59</td>
<td>.43</td>
<td>.33</td>
</tr>
<tr>
<td>ITRS Test Accuracy (+/-ps)</td>
<td>200</td>
<td>200</td>
<td>180</td>
<td>175</td>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td>ITRS Yield (%)</td>
<td>90</td>
<td>87</td>
<td>79</td>
<td>75</td>
<td>64</td>
<td>52</td>
</tr>
<tr>
<td>Miao-Dalal Yield[2]</td>
<td>92</td>
<td>88</td>
<td>81</td>
<td>67</td>
<td>33</td>
<td>20</td>
</tr>
<tr>
<td>Accuracy for 95% yield (+/-ps)</td>
<td>148</td>
<td>115</td>
<td>80</td>
<td>63</td>
<td>44</td>
<td>34</td>
</tr>
</tbody>
</table>

The ITRS predicts almost no change in delivered accuracy, anticipating that for the next decade test systems will be over 40% less accurate than equipment being delivered today.

Clearly, the ITRS forecasts have four problems.

• The yield loss at ITRS accuracy is understated
• The accuracy requirement is slightly overstated
• The statement of delivered accuracy is wrong
• The equipment accuracy forecast is pessimistic

The following figure exhibits the ITRS accuracy requirement forecast, the ITRS tester accuracy forecast, and the accuracy required to achieve 95% yield as calculated here. In the balance of this paper, we will analyze the reason for requirements of improved accuracy, and we will also consider the technical issues which must be addressed to produce ATE whose accuracy meets these requirements.

4 Functional Test Requirements

In the following subsections, we consider the requirements for:

• input timing accuracy
• input signal handling
• DUT connection
• I/O Switching
• output signal handling
• output differential signal handling
• output timing accuracy

that at-speed functional test must satisfy.

4.1 Input Timing Accuracy

With modern high-performance microprocessors, internal circuitry is no longer limited to the speeds attainable at the I/O boundary. Internal clocks are running three, four, five, or more times as fast as external data. This complicates the task of establishing the OTA requirement on the broad family of I/O pins.

There are two ways to look at this problem. The first way is to assume that input timing will be referenced to the external clock, with setup and hold times controlled to the device data sheet specifications. In this case, the ITRS I/O pin timing requirement will be 5% of the bus clock rate. Devices can be validated with this kind of timing requirement.

However, if the objective of the test is to establish detailed internal functionality, particularly for fault isolation to support subsequent failure analysis, the input timing accuracy must keep pace with the internal device clock in order to avoid phase ambiguity and assure that the intended test is actually applied.

This imposes a much more stringent requirement
errors. This level of OTA will be very difficult to reach in the near future.

However, if we pose a requirement of max 5% yield loss due to timing error budget, rather than 5% of period, the input timing accuracy requirement is relaxed to a manageable goal, as noted above.

4.2 Coping with Jitter

Much has been published recently about jitter measurements, as telecom systems drive serial data streams faster and faster[15][16]. Jitter, like bit error rate, is a good measure of the overall quality of a system. However, as a tool for detecting and eliminating marginal design or locating defects, it is very poor. Jitter consists of a combination of different causes of timing errors, most of which are related in one way or another to circuit dynamics and signal history - which is to say, most jitter is not due to noise, but is due to systematic errors.

It is straightforward to calculate jitter as a result of timing edge placement error, but it is difficult to impossible to discern specific edge placement errors from jitter measurements. Jitter measurement tools are of relatively little value in troubleshooting or optimizing device designs.

4.3 Input signal handling

Once the input pattern is generated and properly timed, it must be applied to the device input pins. An active buffer amplifier, commonly termed a PE (pin electronics) driver, is used for this purpose. The type of driver depends on the input signal character - for multi-gigahertz clocking differential signal drivers are required, while single-ended drivers are being used at data rates up to and beyond one GBPS. Single-ended drivers can be used in pairs for differential signals; however, this tightens the system edge alignment requirement. True differential drivers are inherently well aligned at the source. This alignment must be maintained by the fixture design.

The connection of the PE driver to the device input pin normally goes through a series of passive transmission lines and device fixturing. The high frequency response of this path is critical to the test quality[8][12][17].

This driver circuit is used not only to provide the required pattern logic, but to establish (on every significant cycle of the DUT input clock) that the device will respond properly to signals at the minimum high or maximum low level. DC accuracy in these level settings is important, and can be impaired by reflections and other signal mishandling issues.

4.4 DUT connection

Accuracy at the inputs and outputs of the pin electronics circuitry is of little value unless it can be carried directly to the DUT input and output pins[18]. The input and output pins on the DUT are subject to significant process variation - the load variation that results presents a challenge to the fixturing as well as to the basic design of the pin electronics circuit itself.

For input pins, the PE driver is source-terminated. This compensates for a fairly wide range of device input characteristics. Focused calibration can also help to compensate, by adjusting the edge timing to allow for device input capacitance. The adjustment is easy to make, but the measurement required to make it correctly is not.

For output pins, the connection problem is very different. Modern high-speed devices are designed to drive fairly specific loads, and in some cases these loads are inconsistent with test system designs. At the extremely high speeds forecast by the SIA roadmap, device current outputs must be adequate to generate the necessary signals. A 50-ohm line directly from the device pin to the test system comparator input is preferred. This line should be terminated at the comparator. Any mismatches in the line can show up through reflections from the mismatch to the DUT output pin, which can only absorb the reflection cleanly if it exactly matches the transmission line impedance.

For generality, tester input and output channels are frequently combined into single driver/comparator circuits since there is substantial commonality in the timing and pattern requirements of both inputs and outputs, and device pins are not both at the same time. As long as input and output do not switch, this combination does not present major problems (although the combined loading makes circuit design more exacting).

4.5 I/O Switching

Very high speed device pins that can be either inputs or outputs present the additional complexity of turnaround time. This difficulty can be resolved by locating the measurement comparator close to the device pin[10] or using a dual transmission line hooked up in fly-by (assuming the device output can handle the load)[8]. In either case, ghosts of activity from one end impair timing accuracy on the other end if the signal path itself isn’t treated with sufficient care.

4.6 Output signal handling

The DUT outputs are typically measured by two comparators, one set at the minimum acceptable HIGH output and the other at the maximum acceptable LOW output. The dynamic range and compare-side bandwidth are the key parameters associated with response signal handling. Dynamic range is established by the operating characteristics of the comparator itself. Compare side bandwidth is established by the comparator and the parasitics and quality of the signal path leading to it.

Here, again, DC level accuracy is important but not as critical as timing accuracy. Comparators with
high dispersion (\(T_{PD}\) variation over edge speed, threshold setting overdrive, or slope) add some timing error, but in digital systems this can be minimized because the waveform to be captured does not exhibit these inconsistencies to a great extent.

4.7 Output differential signal handling

Little has been published about the special needs of differential signal measurement. Figures of merit in single-ended signals bear little relationship with those in differential signals.

The absolute value of the high and low voltages generated by the two outputs of a differential pair are not as important as the voltage difference generated at the two nominal logic states. Similarly, the time at which either of the signal pairs cross any particular voltage threshold is not as significant as the time at which the difference between the two signals is at or close to zero volts.

These characteristic differences have not impacted test system design requirements in the past, because the actual circuit behavior could be effectively measured by single-ended circuit techniques and real devices were generally sufficiently well-behaved that the translation of these characteristics into single signal requirements was reasonably straightforward.

As device outputs reach and exceed 1 GBPS, however, these techniques may produce some yield loss. Time delay to differential crosspoint may be more precise than time delay to specific level crossing if the output fall time differs from the output rise time in a differential pair. In this case, measuring the propagation delay of the two signals independently will produce erroneous results. The good news is that no false passes will result.

4.8 Output timing accuracy

Current CMOS devices exhibit significant output propagation delay variation with process variation and with temperature. As a consequence, device designers have begun to provide source clock outputs that are more precisely aligned with device outputs. This presents an output timing measurement challenge, as the ATE is normally not designed to adjust the synchronization of the strobes with the process variation or the temperature of the device being tested.

4.9 Source-Synchronous Clocks

A serious output timing challenge is the source-synchronous clock discussed above. Assuring that specifications are met requires measuring the output timing with respect to an output clock. Since there are at least two different sources of synchronous output timing variation, it is not surprising that one might consider two different solutions.

The output timing variation that depends on the process parameters can be established by making specific measurements on specific pins (generally the source clock pin), and then readjusting the output strobe event sequences to match. This technique has been used effectively in the past.

The temperature dependent output timing is not consistent over the test pattern if the pattern activation causes significant differences in power consumption - which it generally will, and probably should for thorough device testing. In order to assure that the required accuracy can be achieved, equipment has been developed that can hold the device junction temperatures acceptably constant over the pattern duration[19].

5 Meeting Functional Test Requirements

We have now established and quantified the architectural and timing requirements that an at-speed functional tester must satisfy.

As these requirements go significantly beyond current practice, it is reasonable to ask whether they can be met with practical test systems.

Systems being delivered today generate patterns on up to 1000 pins at data rates up to 1.3 GTPS. Associated with these systems are high-speed clocks capable of generating fixed clock trains well over 1 GHz. However, the overall timing accuracy goals remain challenging.

Key contributions to the timing error budget depend to some extent on the system architecture. Most modern ATE uses precise system clocks together with some kind of single-cycle interpolation, in order to place timed edges where they are wanted. In the following figure, we indicate a possible allocation of a timing edge placement error budget as it can realistically be expected to evolve over the next decade.

As shown in the figure, the contributors are clock phase noise, interpolation linearity, pattern effects, calibration resolution, and measurement error. Pattern effects occur in the timing circuitry, the pin electronics circuitry, and the interconnect between the various devices. The extrapolations shown are net of calibration corrections. The numbers in parentheses are net of digital resolution.

Edge placement applies to both input and output
side; in general OTA will be either twice input edge placement error (in the case of setup/hold measurements) or input plus output plus I/O alignment error (in the case of propagation delay or access time measurements). PE and timing path pattern errors include both active circuits and interconnect.

5.1 Stability of Critical Components

Timing accuracies of a few picoseconds, in paths consisting of several consecutive devices, requires very stable components with extremely small pattern sensitivity. Pattern error contributions get exacerbated as data rate goes up and minimum pulse width goes down inversely. In general, this requires keeping the paths as short as possible. It also requires emphasis on technologies that are inherently stable against pattern variation.

In the early days of ATE, mainframe computer technology drove the ECL technology on which ATE timing paths depended, and the required components were pretty much off the shelf. The higher speeds reached by CMOS have shouldered ECL aside in many applications, and GaAs speeds have attacked the top end. However, the pattern effects of GaAs continue to render them unsuitable for high-performance timing circuits. Meanwhile, silicon ECL technology has been advanced to exhibit transistor $F_t$ in the range of 30 GHz, and more recently, SiGe has been reported to reach 75 GHz. SiGe circuits have been demonstrated at data rates of 50 GBPS[20].

The stability of these devices remains their key advantage; the cost their key disadvantage. Modern telecommunications systems utilizing ECL-type devices in increasing volumes give us significant basis to expect these costs to decrease over time.

5.2 The Metastability Issue

Metastability is well known[11] - when a D-type flipflop is clocked close to the time the data input changes, the propagation delay is extended. Fail data in a tester are typically captured by such a circuit, as shown here.

Metastability, as expressed by this equation, is normally discussed in the context of the probability of failure of a capture circuit, rather than the accuracy of the time value of a capture.

However, if we design our circuitry in such a way that we can wait a little while after the measure flipflop is strobed, then this equation gives us a very comfortable time resolution. For example, let's suppose $\Delta t/\tau = 5$, which is to say $T_D - T_P = 5*\tau$. For a Motorola 10E131, the value of $\tau$ is 200 ps, so the delay extension is one ns. In this case, the ambiguity region is $10^{-5}$ times the normal propagation delay of 700 ps, or seven femtoseconds.

We can therefore conclude that, when strobing a flipflop for event time capture, metastability is not a significant contributor to timing error. We recently used this fact as a means to improve our critical device characterization measurement accuracy tools and techniques, as discussed in the next section.

5.3 Characterization Measurement Accuracy

Edge placement accuracy of a few picoseconds, in paths consisting of several consecutive devices, requires very good component measurement and characterization capability. Traditionally, bench setups using pulse or pattern generators and high-performance oscilloscopes are used.

The limitations of these pieces of equipment were not well understood, so we investigated the timing accuracy of two typical oscilloscopes. The two figures shown here present the measured timing error as a function of edge position across the oscilloscope face. These timing errors represent the difference between an observed level crossing and that determined by the physical setup.

Measurements on Tektronix 11801B:

The figure shows the data and an approximate moving average. Errors of as much as 10 ps occurred fairly frequently, but smoothing over the collected data reduces the error observed to 2-3 ps. This smoothing exhibits a distinct periodicity.

Measurements on an HP54120 show very different behavior, and the periodicity is more obvious than
in the other case. The amplitude of the periodic error appears to be about +/- 5 ps.

Note that the edge position errors had very different character in the two different pieces of equipment, while the overall measurement setup was identical except for the oscilloscopes chosen. The linearly increasing error in the second figure is believed to be a minor calibration misadjustment. The deviation from this linear error is our concern.

The measured time is the DC level crossing of the waveform. The timing path of high-performance test equipment is usually implemented with ECL differential signals, so a DC level crossing time is a crude representation of the actual differential signal timing. For this reason, measuring a level crossing time is not as useful as measuring the instant at which two signals cross each other.

5.4 Differential Signal Path Characterization

The ambiguity range of a flipflop can be made very small. This is discussed in detail in Section 5.2. Differential timing components can be characterized very precisely by taking advantage of this fact.

As an example, we wanted to learn how much noise was added to a signal path by a newly announced device, the EDGE E118 (a 1-18 differential fanout clock buffer). So we connected a pair of adjustable air-dielectric rigid delay lines ("trombones") to a Motorola 10EL52 flipflop and an EDGE 118 component as shown in the figure below.

The trombone included a vernier screw adjustment mechanism which provided adjustment resolution of 0.02 mm. As the signals are differential with approximately equal rise and fall times, adjustment of one trombone in the pair and not the other gives a signal crosspoint setting resolution of 0.01 mm. This results in a timing adjustment of 30 femtoseconds.

The signal path lengths were adjusted so that the clock and data arrived at the same time, with the consequence that we could adjust one trombone slightly to cause the data arrival time to cross the ambiguity region of the flipflop.

Following is the curve showing the probability of clocking "1" against the adjusted trombone setting.

From these data, we concluded that the noise sigma of the measurement was approximately 86 fs. This includes the flipflop aperture ambiguity, flipflop noise, edge propagation noise, and the target noise - that due to the EDGE 118.

Since the circuits used in this evaluation can be considered typical circuits for development of high-performance ATE, we can also conclude that timing accuracy is not threatened by basic circuit noise, if the signal timing path is well designed.

More importantly, it supports this technique as a means to qualify and characterize differential timing components, and arrangements of such components, with subpicosecond resolution and accuracy.

5.5 System Implementation Requirements

Knowing very precisely the behavior of individual components is not really enough to build a very accurate system. Combining those components is also key. A fundamental rule is to minimize the length of the timing path. This rule translates into the very simple rule "keep the timing path digital as long as possible"[22]. It also provides the opportunity to cost-optimize by implementing the bulk of an ATE system in CMOS with the critical timing and pin electronics implemented in more stable circuit technologies, such as high-speed bipolar silicon or silicon-germanium.
Deskew and calibration should be done digitally, and wherever feasible systematic errors should be translated into digital offsets so that the total combination can be held to round-off error. Minimizing roundoff error then clearly implies carrying digital resolution well below the timing accuracy requirement - which in not very much time from now will mean subpicosecond resolution in all timing calculations.

5.6 Achievable Equipment OTA

The Overall Timing Accuracy (OTA) of equipment which can be implemented over the next decade if it is designed according to the preceding discussion can be compared to the 95% yield objective stated above. Following is the comparison of the achievable equipment performance against this requirement forecast.

6 Conclusion

In this paper, we have pointed out that the timing demands of the very fast, very complex devices anticipated over the next decade are overstated by the ITRS document, and that that same document grossly underestimates the potential deliverable equipment.

Nevertheless, we have shown that the requirement will be very demanding, and we have discussed several reasons for this increasing demand. These include increased operating speeds, substantial crosstalk resulting from finer geometries, delay fault diagnosis, design validation, and source-synchronous output timing issues.

We also showed that these demanding timing measurements can be addressed by properly designed and configured at-speed functional test systems, using well-characterized, very stable components and effective system implementation rules. The most important of these rules is to minimize the analog portion of the system.

We have also pointed out that differential edge placement measurements with resolution well under 1 ps - a characterization requirement, not for the complex digital devices to be tested, but for the components that are used to implement these test systems - can be made.

Concerns over round-trip delay, noise, metastability, and other allegedly fundamental limits to timing accuracy have been refuted.

7 Acknowledgements

Device measurements discussed above were made by Masashi Shimanouchi. Yield model calculations were made by Song Miao. Both of these contributions are greatly appreciated, as were the contributions of the several reviewers of this paper in terms of its content, structure, and organization.

References


