Digital Synchronization for Reconfigurable ATE

Dr. Burnell G. West and Michael F. Jones
NPTest
150 Baytech Drive
San Jose, CA 95134
bwest@npitest.com, mjones@npitest.com

Abstract
This paper introduces a digital synchronization technique for a highly reconfigurable ATE platform that overcomes inherent scaling, multisite, and other limitations in currently used instrument synchronization methods.

1 Introduction
Automatic Test Equipment normally requires well synchronized instruments of a number of different types for consistency and repeatability. Examples include DPS triggers, analog waveform syncs, and scope pictures.

As more and more features are embodied in modern IC’s, the need for synchronization continues to grow. At the same time, continuing cost pressures drive the need for higher and higher site counts in order to keep the cost of test per IC manageable. Consider the following facts.
(1) Modern I/O strategies are being incorporated in IC’s. Some of these strategies create a new level of indeterminacy in the overall test flow. For multi-site testing, this creates a need for independent test flows per site.
(2) Very high performance SOCs will require very exact synchronization between analog and digital instruments.
(3) These same SOCs are being built with dozens or more time domains; testing them effectively has led to the need for test systems supporting many time domains.
(4) Multiple-instrument Platform ATE adds another complexity. In principle, instrument triggering requirements are unknown but must be supported, both in terms of the number of unrelated triggers and their timing accuracy.

All of these trends together create a need for a synchronization mechanism that supports many different sync triggers from many different instruments for many different purposes, without interfering with each other and without creating arbitrary limits on site count, site independence, or time domains per site.

This paper describes a new digital sync approach.

2 Background
In the past, test systems normally provided a single time domain for a single test site. Such test systems were expanded to multiple site capability by simply replicating the test (either by fanning out the signals in hardware or by duplicating vector patterns in memory) and driving all sites identically. Of course, DUT activity had to be captured individually; for high site count testing (as in DRAM’s) go/no-go testing was OK as long as it could be augmented by datalogging a single site at a time. This worked well enough as long as the tester timebase could be distributed across all the devices being tested without compromise. More recently, test systems incorporating multiple independent time domains have been introduced [1]. The use of multiple time domains in these systems gave more condensed vector sets and much easier test development, but did nothing to change the requirement for simultaneity in multi-site applications.

Synchronization in ATE traditionally has taken two forms. The first, and simplest, consists of running a high-bandwidth signal path from a trigger source to a trigger destination or set of destinations. This is easy to understand, but may not be as easy to implement if multiple sources or destinations are envisioned. The classic solution to this problem is a “star trigger” [2,3]. This trigger mechanism incorporates analog multiplexers and selectors, with a substantial number of individual cables for the various destinations or a “trigger bus” of some sort. Either approach adds crosstalk problems and limits the accuracy of the triggers provided to the overall bandwidth of the star trigger network. Either approach also limits the number of trigger sources and destinations to the lesser of the number of inputs, outputs, or switches in the trigger multiplexing network.

The second synchronization mechanism in ATE is a vector synchronization (“Enable Instrument” or EINST) that depends on a single sequencer and distribution throughout the tester of a sequence control vector containing the EINST token. This mechanism inherently has no better timing
resolution than the start time of the vector – which for most synchronization purposes in the past has been quite ample. For multiple time domains, it is obvious that the EINST mechanism has to be replicated for each time domain (as was done in the dual-time-domain system reported in [1]). This approach will not work in a multiple-time-domain system for two very obvious reasons: there are not enough wires available to distribute the tokens for all the time domains, and the trigger resolution (one vector) is simply not good enough for all present and near future requirements.

3 Digital Time Reference

In 1990 the Sequencer per Pin Test System was introduced [4]. A key contribution of this system architecture was its widespread use of digitally computed time values, incorporating all required terms from vector offset, edge placement, and cal before making the transformation to the analog time domain. The underlying philosophy of this approach was “keep it digital as long as possible”, a strategy that has served the industry well.

Digital time reference enables precise, digitally controlled relationships between a system clock and a vector time, even if the vector time is neither constant nor a multiple of the clock. The advantage of this process is detailed in [1] and [4]; it can be summarized as follows.

- The first vector in a test starts on a clock cycle boundary.
- Every subsequent vector starts at a later time that is exactly the sum of all previous vector times.
- The start time of every subsequent vector is defined by a specific clock cycle and a fraction of the clock cycle.

Clearly, therefore, there is no possible ambiguity in the relationship between a vector start time and the system clock cycles. Without digital timing, it is impossible to control clocks well enough for this kind of guarantee and still provide adequate DUT vector period flexibility.

In fact, analog instruments frequently require time references with much higher precision than digital tests require. As long as such analog clocks are locked with synthesizer techniques to the digital system timebase, a similar time relationship of the analog clock with the digital timebase can be assured.

The digital time reference strategy is adapted to the triggering task. For a new test system platform introduced last year, NPTest has developed a digital triggering strategy based on the “keep it digital” philosophy. Rather than a wideband network of limited trigger source and destination paths, the approach uses a trigger distribution process based on messages sent through a low-cost network between instruments. An initialization process develops a universal time plane for all instruments in the system, and a sync message to all instruments identifies a trigger and sets its time. This universal time plane is called the Isochronous Fabric Interface™ (IFI)[5]. The accuracy of this trigger is not constrained by the message network, but by the resolution of the trigger time word and the accuracy of the system-wide clock. Both of these can be controlled to much higher precision at a much lower cost than is possible in an analog trigger distribution network. In the current system, the clock distribution is precise enough to hold clock jitter below 1 ps RMS; the resolution of the trigger word as established in the current protocol is 1.22 ps. These numbers are accurate enough for current and near future needs; both can be improved as needed without replacing the platform itself.

The balance of this paper describes in detail the implementation of the system time plane and the synchronization strategy employed.

3.1 System Synchronization Messages

The basic communication process of the system is shown in the diagram of Figure 1. Isochronous Fabric Interface

The figure indicates twelve instruments allocated to four sites. The system clock is a 400 MHz clock. Instrument time is tracked in a 10-bit word that contains the Global Clock Number (GCN). This number is initialized at the time the system is powered up. It provides the master time plane reference for all synchronization.

The shortest possible message in the IFI is a single 4-byte word. The IFI ring itself is a 43-pair parallel source-synchronous bus running at 200 MTPS (800 megabytes per second plus flags). Sync messages are instances of the shortest
possible IFI message – a sync message occupies one word time. So sync messages can be sent every five ns.

The current protocol defines three sync types, as shown in Figure 2.

<table>
<thead>
<tr>
<th>Type</th>
<th>Clock</th>
<th>ResCount</th>
<th>Fine</th>
<th>SubType</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1</td>
<td>37 36</td>
<td>30 29</td>
<td>20 19</td>
<td>12 11</td>
</tr>
<tr>
<td>high-resolution trigger</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Clock</th>
<th>ResCount</th>
<th>Fine</th>
<th>SubType</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1</td>
<td>37 36</td>
<td>30 29</td>
<td>20 19</td>
<td>12 11</td>
</tr>
<tr>
<td>low-resolution trigger</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>ResCount</th>
<th>Fine</th>
<th>SubType</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1</td>
<td>32 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>partial sync</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. **Sync Types**

The first of these sync types is a high-resolution sync type, the second uses part of the time field to expand the number of sync types. The third sync type, “partial sync”, is a special case used where a sync event must be recognized by coincidental events on multiple instruments.

Note that confirmed sync words contain two time fields. The first is the Global Clock Number (GCN); the second is a binary fraction of a clock cycle. When an instrument generates a sync message, it uses the current GCN and inserts the proper fraction in the offset field. This defines the time of the sync to the precision of the fraction. A high-resolution trigger has a precision of 1.22 ps, while that of a low-resolution trigger is 78 ps. The GCN resolution is 5 ns.

### 3.2 Confirmed and Partial Syncs

Most triggers are generated by specific events at specific instruments. The generation process, as described above, is quite straightforward. When we realized the need for partial syncs, we elected to call these simpler sync events confirmed syncs. Some devices require pattern matches. In some of these devices, the pattern match involves more than one device pin. It would be an unacceptable constraint to require that all pins involved in pattern matches for any given site be contained in a single instrument.

Partial syncs are used to avoid this constraint. A partial sync event is one in which a portion of a pattern match must be recognized by one of the instruments, and the remainder by one or more other instruments. The Partial Sync word consists of type and State History (which is aligned with the low-order 5 bits of the GCN).

State History is a bit field of the in-sync status for the previous 32 clocks. A bit value of 1 indicates in-sync condition for that clock cycle.

The State History word indicates which of 32 time periods the participating instrument is in sync as determined by its local match condition. A specific instrument generates partial sync words for a given sync type, based on its internal in-sync condition on a cycle-by-cycle basis.

Suppose three instruments widely separated in the system all detect the pattern required for their portion of a match. These three instruments will compute a partial sync representing that fact and insert a “1” bit in the State History word at the time field corresponding to the clock cycle in which the matched vector originates. One of these three instruments is assigned by software to be the Partial Sync Master (PSM) for the sync type assigned to that match condition. As soon as the 32 bits of its State History word are filled, the now non-zero word owned by the PSM is transmitted to the ring.

All instruments participating in the same sync type monitor the IFI ring for that specific State History sync word. Each instrument ANDs its incoming Partial Sync word of that sync type with its internally generated state history word for the same time frame. The resulting word is passed on only if non-zero.

As soon as a non-zero partial sync word reaches the originator, a sync (pattern match, in this case) has been found and the originator converts the word to a confirmed sync, calculates the GCN that is associated with the match, and places this word on the ring. This confirmed sync then propagates to all participating instruments.

### 3.3 Trigger Latency Considerations

As with all triggering, there is a period of time required between when the trigger requirement is determined and when the triggered instrument can respond. This time is the trigger latency.

Since any instrument can trigger any instrument, a system-level latency has to be defined for a given test program and loadboard configuration (if DUT test considerations require multiple loadboards, the latency for the worst-case loadboard is used). The test program sets this latency value based on two key considerations. The first is the minimum latency of the specific system configuration, which is set by the relative positions of trigger source and destination instruments on the ring, established by the particular system configuration.
and the loadboard hookup. The second is the maximum latency acceptable by the instruments being synchronized.

The overall latency of the IFI ring in a 40-slot head is just over one microsecond. For many applications, this is satisfactory. If smaller latency is required, then both instrument configuration and loadboard design can be used to reduce it. The least latency achievable (using the IFI ring) depends on local pipelines and analog signal considerations but has the ring insertion/removal overhead attached.

Instruments generate and propagate syncs in five bus cycles, so the minimum latency if the source is immediately ahead of the destination in the ring is local processing plus eleven bus clock cycles (55 ns).

Minimum latency for a partial sync includes a full ring cycle. The latency is ring time plus the path between the PSM and the last participating instrument. If two adjacent instruments generate the sync, then this minimum latency will be ring delay plus 55 ns plus local processing time.

We make one final observation. The 10-bit GCN word length puts an upper limit on latency: 5 us.

### 3.4 Periodic Syncs and Keepalive

Another application of the sync mechanism is a Test Head Interface (THIF) capability to generate periodic syncs. A periodic sync is a confirmed sync message designed to provide a trigger every \( n \) vectors. Since vector times and the system clock are not commensurate, periodic syncs are another specific advantage of creating syncs with digital precision.

Keepalive is a mechanism that keeps a DUT PLL active during unconstrained CPU activity (such as reloading focused calibration tables for a different VDD setting). Its value is reduction in test time as the time to restore a PLL is often quite long.

If a device test requires multiple time domains (as with fractional clocks [1]) keepalive is challenged by the need to keep the time domains properly in sync. Since the CPU activity required is neither constant nor predictable, exit from keepalive has to align with those vectors that coincide in each time domain. This is enabled by the periodic sync mechanism. The period of the periodic sync is the least common multiple of the periods of each of the time domains.

### 3.5 Trigger Calibration and Sync Accuracy

Precise instrument triggers are of no value unless they can be referred accurately to the relevant time domain. Final time reference is a calibration responsibility.

We have mentioned the system time plane that the IFI defines. However, we have not considered its relationship to the DUT (or DUTs).

An instrument maintains a calibration offset as well as its Global Clock Number. This calibration offset, for a DUT input event, is the time between the clock edge used to establish a trigger time and the arrival time of an analog event generated at that trigger time at the DUT (a different offset value is associated with each DUT resource on an instrument). Here, again, the calibration offset is added to the sync time offset to establish a delay program for an analog delay in the event path.

For a DUT output event, the calibration offset is the difference between the path from the DUT pin and the path from the trigger clock edge.

The overall accuracy of a digitally defined sync is therefore limited by:

- Clock fraction resolution
- Calibration accuracy
- Clock distribution quality

Currently, by far the largest of these error sources is the calibration accuracy. Today, the best tools available produce edge placement accuracies in the range of \( \pm 20 \) ps. In future, we can expect the accuracy to improve, as measurement techniques and DUT path modeling get better.

### 3.6 Digital Synchronization Scope and Limits

From the foregoing, it is clear that the bandwidth of the IFI ring will impose a constraint on the sync trigger mechanisms. This limit is affected by bus traffic from other sources.

Multiple independent syncs cannot “collide”. The GCN asserts the sync clock cycle time, so the ring slot the trigger word occupies is not critical. Bus protocol requires that instruments that can create triggers must have enough IFI FIFO space so bus read/write activity can be delayed a cycle without penalty. Instruments cannot delay incoming sync messages; they must wait for a non-critical slot.

During the activity of a specific functional test, the bus traffic might be expected to be rather light. If we assume that 80% of the bus is available for sync messages, then the number of triggers that can be active simultaneously on the IFI ring in a 40-slot head is 80% of five cycles times 40 instruments: 160 triggers. Each of these triggers can be unrelated to all the others, in time, source, or destination.
This is not the limit on the number of triggers that can be generated; only the limit that can be active on the IFI ring simultaneously.

The sync word definition allows 16 different types of periodic sync, 63 different high-resolution sync types, and over 4000 low-resolution sync types. Of course, synchronization activity wholly within a specific instrument does not require the use of any of these sync types.

What are the practical implications of these limits? First, the limit of simultaneous triggers on the bus must be considered. This limit is only theoretical in nature; since sync messages can stall each other, the issue becomes more one of latency program in a test than of maximum simultaneous activity on the IFI ring. With a maximum latency of 5 us, as many as a thousand sync messages could be generated referring to exactly the same time. Scenarios requiring more are not known.

Second, the number of types must be considered. If partial syncs (as match mode) are used, there are only 16 sync types available. This limits the number of sites that can require match using more than one instrument for the match word to 16. This limit is not likely to occur in a practical case, but it is not theoretically inconceivable.

Confirmed sync triggers are much more common in practice, for such actions as scope triggers and DPS Vbump. For Vbump, a trigger is generated at a specific vector and the device power is indexed to a new value. The accuracy required of this kind of trigger is limited by DPS settling time, much larger than the low-resolution accuracy that is available. Consequently, the test programmer can define hundreds of different triggers for DPS settings if desired, or DPS Vbump triggers could be defined for hundreds to a few thousand test sites.

The digital synchronization mechanism described here is sufficient for test requirements for some time to come.

4 Conclusion

We have described a new strategy for integrated circuit test synchronization across several active test and measurement instruments. This strategy is much more flexible than currently available star or bus trigger arrangements, and is more accurate as well. We have shown how this digital sync can be used for multiple independent test site match operation, for wide-ranging DPS triggers, and for multisite tests with multiple time domains per test enabling synchronized keepalive capability. This synchronization strategy has been shown to enable extensive growth in both instrumentation complexity and test and measurement accuracy.

References
[2,3] VXI and PXI star trigger definitions (Year)