The Cadence® Modus DFT Software Solution introduces a ground-breaking new physically aware 2D Elastic Compression architecture which reduces manufacturing test time by up to 3X, saving test cost and making chips more profitable. This innovative patented technology can also reduce the overhead of compression logic on chip routing resource by up to 2.6X, improving die size and accelerating time to market. The Cadence Modus DFT Software Solution is natively integrated with Cadence’s full-flow digital solution, which provides faster design closure, better predictability, and best-in-class power, performance, and area (PPA).

Introduction

Every year, the chip industry spends roughly $4B on automated test equipment (ATE) for manufacturing test, according to industry experts. This cost must be absorbed into the overall production cost of each working chip coming off production lines. At the same time, chips at advanced nodes are becoming increasingly complex to enable ever richer user experiences. This results in more and more logic gates to be tested on ATEs, further driving up the $4B market (see Figure 1).

Time on ATEs breaks down into several components – testing of memories, chip I/O interfaces, analog logic, and digital logic. Testing of digital logic typically ranges from 10-50% of total test time, which equates to a significant portion of the overall $4B annual ATE industry market size.

The Cadence Modus DFT Software Solution is a new design-for-test (DFT) solution that reduces test time for digital logic by up to 3X compared to current industry solutions, with no impact on chip size or yield. Alternatively, for the same test time as current industry solutions, the Cadence Modus DFT Software Solution can reduce the overhead of DFT logic on chip routing resources by up to 2.6X. Its patented physically aware 2D Elastic Compression architecture is the foundation behind these unique benefits.

XOR Compression

XOR-based test compression is the approach most widely used in the chip industry today to minimize digital logic ATE test cost. XOR compression reduces test time by partitioning registers in a design into more scan chains than there are scan pins on the chip to connect to the ATE.
As the ratio between the number of scan chains and the number of scan pins increases, the length of each scan chain decreases, which means fewer clock cycles to shift in each test pattern. For a constant pattern count, fewer shift clock cycles per pattern means less total test time on the ATE.

Compression Ratio Impact on Coverage and Test Time

Fewer clock cycles per pattern also means fewer bits of information in each pattern to control register values and detect faults. At some point, if the compression ratio becomes too high, the achievable fault coverage drops since some faults will require more register values to be controlled than there are bits in a test pattern. Also, as the compression ratio increases, even if fault coverage can be maintained the number of patterns required to maintain this coverage rises rapidly since it becomes more difficult to pack the detection of multiple faults into a single pattern (Figure 2, above). As a result, there is a diminishing benefit on total test time from increasing the compression ratio (Figure 3). The asymptote of this curve is typically around 50–100X the compression ratio.

Compression Ratio Impact on Chip Size

In addition to fault coverage drop and pattern count growth, increasing the compression ratio also has a significant impact on the physical implementation of a chip. That’s because every scan chain must be connected to and from the XOR compression logic (Figure 4).

At a typical industry compression ratio of 100X, the average impact of XOR compression logic across a range of common digital components is 3–5% of total chip routing resource. If the compression ratio is increased to 400X, then the impact on chip routing resource increases to a staggering 10%, which would require a significant increase in die size to accommodate and ultimately outweigh any potential cost savings from the reduced test time. See Table 1.

<table>
<thead>
<tr>
<th>Design</th>
<th>#Instances (M)</th>
<th>100X</th>
<th>400X</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>1.3</td>
<td>5%</td>
<td>15%</td>
</tr>
<tr>
<td>GPU</td>
<td>2.6</td>
<td>3%</td>
<td>8%</td>
</tr>
<tr>
<td>Networking</td>
<td>1.6</td>
<td>3%</td>
<td>9%</td>
</tr>
<tr>
<td>DSP</td>
<td>2.3</td>
<td>5%</td>
<td>11%</td>
</tr>
<tr>
<td>Automotive</td>
<td>2.5</td>
<td>3%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 1: Impact of increased compression ratio on chip routing.
2D Elastic Compression

The Cadence Modus DFT Software Solution introduces two major innovations backed by an extensive portfolio of over 18 US patents granted to address the fundamental challenges of rising compression ratio: 2D Compression and Elastic Compression. When combined, the Modus 2D Elastic Compression architecture enables compression ratios beyond 400X with up to a 3X reduction in test time compared to traditional XOR compression at a 100X compression ratio. There is no impact on fault coverage or chip routing resource.

Alternatively, if 2D compression is used on a traditional non-elastic XOR compression structure at a 100X compression ratio, then the impact of the compression logic on chip routing resource is only 2%, an up to 2.6X improvement over current industry solutions.

2D Compression

Modus 2D Compression targets the impact of high compression ratios on chip routing resource. The key insight behind Modus 2D Compression is to leverage the two-dimensional nature of a physical chip layout to build an XOR network which can unfold into a grid structure across the chip (Figure 5).

In Figure 6, the wirelength of the 2D grid structure scales sub-linearly with the compression ratio, and at a 400X compression ratio is still no worse for routing resource than traditional “one-dimensional” XOR compression at a 100X compression ratio.

Elastic Compression

Modus Elastic Compression pairs with Modus 2D Compression to mitigate the fault coverage and pattern count impact of high compression ratios. It differs from traditional XOR compression by leveraging registers and sequential feedback loops alongside traditional XOR logic in the decompressor circuit.

The sequential nature of an elastic decompressor enables the automatic test pattern generation (ATPG) algorithm to leverage multiple shift cycles to control register values in a single fault capture cycle (Figure 7). Conceptually, this is

<table>
<thead>
<tr>
<th>Design</th>
<th>#Instances (M)</th>
<th>1D-100X</th>
<th>1D-400X</th>
<th>2D-100X</th>
<th>2D-400X</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>1.3</td>
<td>5%</td>
<td>14.6%</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>GPU</td>
<td>2.6</td>
<td>2.9%</td>
<td>7.7%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>Networking</td>
<td>1.6</td>
<td>3.3%</td>
<td>9.4%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>DSP</td>
<td>2.3</td>
<td>4.6%</td>
<td>11.5%</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>Automotive</td>
<td>2.5</td>
<td>2.6%</td>
<td>5.6%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4%</td>
<td>10%</td>
<td>2%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Figure 6: 2D Compression via the Cadence Modus DFT Software Solution yields a 400X compression ratio.

Figure 5: In 2D Compression, an XOR network unfolds into a grid structure across the chip.

Figure 7: With the sequential elastic decompressor, the ATPG algorithm can control register values in a single fault capture cycle.

Scan chain

Elastic Decompressor

Scan_in pins

Sequential circuit with XOR gates and registers

Scan_out pins

XOR Compressor

Scan chain

Figure 7: With the sequential elastic decompressor, the ATPG algorithm can control register values in a single fault capture cycle.
like being able to “borrow” scan bits from previous clock cycles to help detect more challenging faults in the current clock cycle. An elastic decompressor can also adaptively increase the number of shift cycles in a test pattern to be larger than the scan chain length to provide yet more controllability to detect tough faults.

Unified Compression and LBIST in a Physically Aware Environment

Building on the 2D Elastic architecture, Unified Compression is a new approach that unifies scan compression and logic built-in self-test (LBIST). This new physically aware approach to LBIST allows designs to target the high coverage needed for safety-critical applications without impacting the design convergence. On a sample design, area savings of 35-47% and scan wirelength savings of 63-77% for the same channel length can be demonstrated. Also, with the same area and scan wirelength budget, the channel length can instead be reduced by half to reduce the overall test time with the same fault coverage. Automotive and safety-critical designs stand to benefit from this unique solution that is effective for both in-system test and 0-DPPM manufacturing test.

For more information, read the “Unified Compression and LBIST in a Physically Aware Environment” white paper.

Integration with Implementation Flow

Bringing the two dimensionality of the physical world into the creation of XOR compression logic requires a seamless integration of compression logic insertion, logic synthesis, and gate placement. This is achieved through a native code level unification of the Cadence Modus DFT Software Solution with RTL physical synthesis using Cadence’s Genus™ Synthesis Solution. The output of the Genus Synthesis Solution with Modus 2D Elastic Compression is a fully placed design, including a placed 2D XOR grid structure.

Also, the complete suite of Cadence® tools used for digital implementation—including test in the Cadence Modus DFT Software Solution, the Genus Synthesis Solution, place and route in the Innovus™ Implementation System, and timing signoff in the Tempus™ Timing Signoff Solution—shares a common unified user interface for TCL scripting and reporting (Figure 8). This streamlines flow development, simplifies user training, and improves productivity of multi-tool users.

Comprehensive Functionality

In addition to Modus 2D Elastic Compression, the Cadence Modus DFT Software Solution includes comprehensive support for all other industry-standard DFT structures, such as fullscan, XOR and MISR compression, X-masking, low-pin-count test, programmable memory BIST, logic BIST, JTAG controllers, IEEE 1500 wrappers, iJTAG embedded instrument access, and timing-driven automatic testpoint insertion.

Modus ATPG supports hierarchical test, low-power ATPG with scan and capture toggle count limits, and distributed test pattern generation with near-linear runtime scalability. Modus Diagnostics includes single- and multi-die volume diagnostics, both with physical defect location callout and root-cause analysis.

Further Information

For more information, please check out our data sheet and product page at www.cadence.com/modus.