Unified Compression and LBIST in a Physically Aware Environment

By Christos Papameletis, Principal Software Engineer; Vivek Chickermane, Sr. Group Director R&D, Cadence

Unified compression is a new approach that unifies scan compression and logic built-in self-test (LBIST). It leverages recent innovations from Cadence in physically-aware design for test (DFT) to solve routing congestion and area issues from traditional discrete approaches and delivers a confident path to high-quality test. 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 could instead be reduced by half to reduce the overall test time with the same fault coverage.

Contents

Introduction

Scan compression is a critical technology for addressing the rapid rise of test cost without sacrificing coverage requirements. It has become widely adopted throughout the semiconductor industry but is facing challenges. The rise of safety-critical semiconductors demands not just high coverage, but also the ability to verify that the design is working in the field. Traditional approaches have used discrete scan compression and LBIST, as shown by Figure 1. Issues with additional area overhead and routing congestion limit the effectiveness of this architecture.

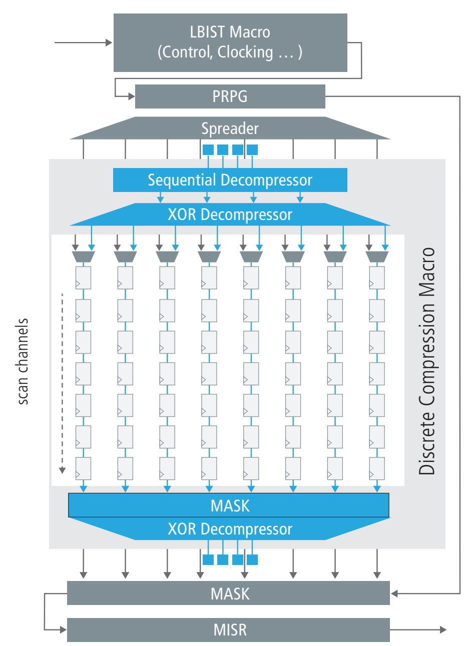

Figure 1: Traditional discrete LBIST and scan compression

Cadence 2D Elastic Compression is a new technology that uses a two-dimensional physically aware sequential compressor-decompressor (CoDec) design to address the severe wiring congestion as well as the test coverage droop and pattern spike at high compression ratios [1]. Traditional scan compression clusters the CoDec in the center and communicates directly with each scan channel, causing a routing nightmare as the number of scan channels increases, Figure 2. 2D Elastic Compression enables high compression ratios by organizing the CoDec and the compression channels into a 2D grid that reduces wiring congestion, Figure 3. The multiple input shift-register (MISR) is located inside the boundary compressors that reside along the periphery of the design, with each bit observing the XOR of a whole row or column of channels. This results in significant area and wiring congestion savings as the MISR is kept small and localized. The X-masking logic is included in the 2D grid, allowing it to be placed locally, near the scan channel that is masked.

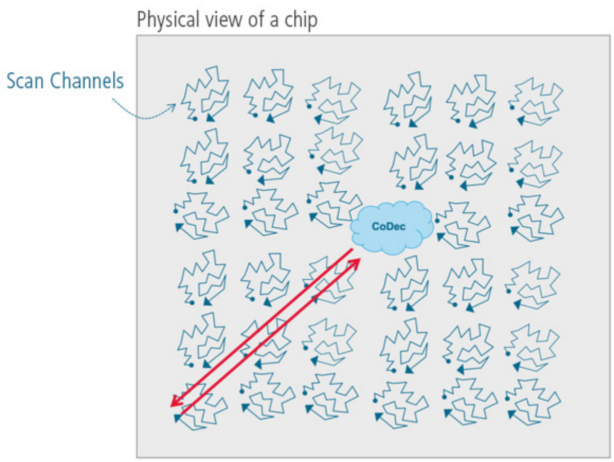

Figure 2: Traditional scan compression architecture

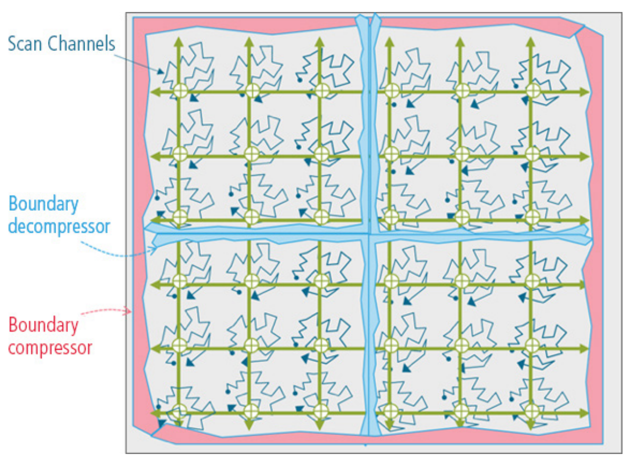

Figure 3: Elastic 2D Compression architecture

Unified Compression: Architecture

Figure 4 illustrates the concept of Unified Compression. An Elastic decompressor enables the target test coverage to be maintained at the higher compression ratios where channel lengths are reduced. The sequential nature of the Elastic decompressor enables the fewer bits of information that are available in each (shorter) test pattern to be used more efficiently to control register values and detect faults.

There are many DFT resources that the above compression architecture has in common with an LBIST architecture and that can be shared between the two test configurations. Intercepting the scan inputs of the Elastic decompressor and forcing them to constants enables it to operate as a pseudo-random pattern generator (PRPG) for LBIST. Creating a serial load/unload interface enables the PRPG, the MISR, and the mask register to be included in the LBIST test data register (TDR) to program initial values for the PRPG and mask and observe the final MISR signature at the end of an LBIST run.

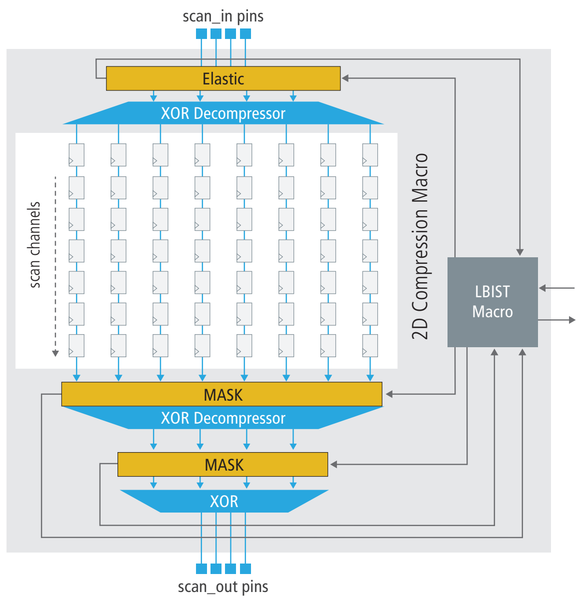

Figure 4: Unified Compression architecture

Sharing these resources between compression and LBIST comes with multiple benefits. Reusing the Elastic decompressor as the LBIST PRPG and spreader, the MISR compressor as the LBIST MISR, and the cycle-by-cycle compression masking logic as the LBIST mask (fixed for a whole LBIST run) conserves area. It also enables all the benefits of physically-aware 2D Elastic Compression to be leveraged by LBIST without any additional overhead. The LBIST macro is simply reduced to a finite state machine (FSM) along with a set of counters, comparators, and configuration registers. Interfaces are built in both the compression macro and the LBIST macro that allow the shared resources that reside inside the compression macro to be accessed by LBIST. For example, in LBIST mode, the LBIST FSM takes over control of the Elastic register and the MISR. It may reset them at the beginning of the run as needed and controls their clocks appropriately throughout the run to ensure that the MISR does not get corrupted while the channels are flushed.

The 2D grid that enables reducing wiring congestion and building shorter compression channels is also leveraged by LBIST. Unlike compression, which may experience test coverage droop and pattern spike as channel lengths decrease, LBIST relies on pseudo-random patterns and does not suffer from these issues. So, the benefit of being able to reduce the channel length by 50%, for example, will usually translate directly to a 50% reduction in LBIST runtime to reach approximately the same fault coverage. The efficiency stemming from the sequential nature of the Elastic decompressor helps to also maintain fault coverage in compression mode in the presence of these shorter channels.

Unified Compression: Results

Table 1 compares the area and scan wirelength of unified 2D Elastic Compression and LBIST compared to discrete compression and LBIST for two designs, A and B, at two compression ratios, 200X and 400X. In the latter cases, the channel length was halved compared to the former cases. In all cases examined, significant savings were achieved for the same channel length. Using 2D Elastic Compression and sharing the same physically-aware DFT resources for both compression and LBIST led to area savings of 35-47%, and scan wirelength savings of 63-77% for the same channel length. Also, with the same area and scan wirelength budget, the channel length could instead be reduced to half to reduce the overall test time with the same fault coverage.

Design (Compression ratio)	Area (um^2)		Scan Wirelength (um)	
	Discrete Compression + LBIST	Unified 2D Elastic Compression + LBIST	Discrete Compression + LBIST	Unified 2D Elastic Compression + LBIST
A (200X)	31287	16391 (–47%)	1441715	526040 (–63%)
B (200X)	41131	26402 (–35%)	11235374	2542247 (–77%)
A (400X)	59191	34752 (–41%)	2141133	582565 (–72%)
B (400X)	81586	50838 (–37%)	15556871	3777880 (–75%)

Table 1: Area and scan wirelength savings from unifying compression and LIBST

For the same designs and configurations, Table 2 shows that the LBIST fault coverage was maintained when the same number of pseudo-random patterns was applied to channels of half the original length. In other words, when the channel length was reduced to half, the same target coverage was reached in half of the original LBIST runtime.

Design (Compression ratio)	LBIST Fault Coverage (%)		LBIST Runtime (cycles)	
	Discrete Compression + LBIST	Unified 2D Elastic Compression + LBIST	Discrete Compression + LBIST	Unified 2D Elastic Compression + LBIST
A (200X)	91.61	91.61	14675831	14674350
B (200X)	99.14	99.1	1010504	1007732
A (400X)	91.76	91.76	7476248	7481471
B (400X)	99.12	99.05	522656	517029

Table 2: LBIST fault coverage and runtime savings from unifying compression and LBIST

Conclusion

This new approach of unifying scan compression and LBIST in a physically-aware insertion allows designs to target the high coverage needed for safety-critical applications without impacting the design convergence. Sample data has shown the ability for designs to implement a unified compression architecture which achieves equivalent test coverage in half the time as contemporary discrete solutions.

References

1. K. Chakravadhanula, et al., “Advancing test compression to the physical dimension,” in IEEE International Test Conference (ITC), Fort Worth, TX, 2017.