J. Electrical Systems 22-01 (2026): 157-182

[1]Vijay Sontakke

Improvements in Test Access Design and Scheduling for Checking Chip Manufacturing Quality

Abstract: - The use of 3-dimensional (3D) and network-on-chip (NOC) integrated circuits enables more functionality on chip, but they pose challenges for testing for manufacturing defects. Machine learning applications for artificial intelligence need to be highly reliable and so mandate in-system testing. All of these modern system-on-chip (SOC) devices manufactured using advanced process nodes integrate numerous heterogeneous embedded intellectual property (IP) cores, which require post-fabrication verification. To maintain practical testing procedures, a modular strategy is employed that enables each core to be exercised without revealing its internal structure, thereby facilitating straightforward test-pattern reuse. The primary challenge is the limited scan-in and scan-out bandwidth at the chip boundary, which is significantly lower than the aggregate channel demand of all cores. Further, enabling in-system and system-level tests, the use of functional pins instead of traditional test-only pins is required. Scan-compression techniques, such as Embedded Deterministic Test (EDT), reduce pattern volume but introduce additional complexity in bandwidth allocation. Consequently, an effective Test Access Mechanism (TAM) must address this disparity. Researchers have introduced various TAM architectures that balance wiring overhead, concurrency, compression efficiency, and scheduling complexity. This paper presents a survey of these TAMs, spanning the years and including recent ones. The goal of the paper is to compare several prominent approaches, examining strategies for dividing available TAM width, mapping cores to these divisions, jointly optimizing EDT parameters, and ultimately minimizing total test time while maintaining high fault coverage under realistic bandwidth constraints. It highlights challenges in their implementation and provides direction for further research on TAMs.

Keywords: 3D, TAM, Low Power, High Speed Input Output, test time.

I. INTRODUCTION

The electronics industry's rapid reduction of chip feature sizes below 10 nm, combined with the transition to three-dimensional integrated circuits, has significantly influenced chip design and testing. Modern system-on-chip (SOC) designs incorporate more than one billion transistors operating at gigahertz frequencies. As integrated circuits (ICs) expand to billions of transistors, designing, implementing, and testing them as monolithic entities has become impractical. A system-on-chip (SOC) is an IC composed of multiple components, known as cores. Each core is generally designed, implemented, and validated independently prior to integration. These designs encompass a range of digital, analog, mixed-signal, memory, optical, microelectromechanical, and radio-frequency cores or modules. Core-based SOC designs are prevalent in industry, with many contemporary SOCs containing hundreds of physical cores, also referred to as IPs, modules, blocks, or tiles. Within these, several classes of identical cores may exist, each comprising dozens of replicated cores. The increasing popularity of SOC designs is largely attributed to the ability to reuse independently designed and verified cores. However, to reuse core-level patterns and facilitate their access, test wrappers need to be developed [1].

The complexity of semiconductor designs has increased rapidly, and these components play a critical role in safety-related applications. There is a growing demand to collect comprehensive structural data from silicon throughout its lifecycle [2]. As process nodes advance, more sophisticated fault models and testing under varying voltage and temperature conditions are required [3][4]. Silicon Lifecycle Management (SLM) involves implementing structures during the design phase to monitor and analyze silicon data. This data can be leveraged to enhance performance and identify structural issues resulting from degradation or defects caused by functional workloads. Collecting such data on application platforms using scan-based infrastructure necessitates the use of active functional interfaces, such as PCIe, and researchers have proposed methods to address this need. Traditionally, SCAN and Built-In Self-Test (BIST) have utilized GPIO pins as the access mechanism [5]. While these pins are available during manufacturing tests on Automated Test Equipment (ATE), they are inaccessible in the system environment. The High-Speed Access and Test (HSAT) Solution [6][7] enables access to SCAN, JTAG, and BIST structures through functional interfaces such as USB and PCIe. A solution presented in [8] facilitates faster failure debugging, thereby accelerating project schedules. It also enables quicker debug turnaround for customer returns. Board-level testing (BLT) using this approach can provide additional coverage to detect defects at the board level and help isolate board-related issues. This capability is essential for achieving automotive ISO

---

[1] *Corresponding author: Vijay Sontakke, ARM Inc., Austin, TX, USA, Vijay.Sontakke@arm.com

specification compliance for Fault Tolerant Time Interval (FTTI) [9] and for improving the reliability of products in high-performance computing environments.

Testing such complex designs presents significant challenges. The concept of modular testing was proposed as early as 1999 [10]. The widespread adoption of SOC circuits has resulted in a substantial increase in testing costs. This escalation is primarily due to difficulties in accessing embedded cores during testing, extended test development and application times, and the large volume of test data[11].

II. BACKGROUND

Modern system-on-chip (SOC) designs require each core to have a defined number of input and output test channels. However, the limited number of chip-level test pins prevents simultaneous access to all core-level test channels. This constraint necessitates the adoption of hierarchical test strategies and efficient pattern retargeting to optimize test access. Hierarchical testing [12] has emerged as the most scalable and effective methodology for testing complex, core-based SOCs, as it addresses the challenges posed by device miniaturization, limited input/output resources, power constraints, and design-for-test (DFT) complexity. The hierarchical testing adds scan chains and compression logic in each core [13][14][15]. The hierarchical test approach remains the most promising and fully scalable solution for core-based SOC testing.

Scan test compression technologies, such as Embedded Deterministic Test (EDT) [16], are widely used to improve the logic testing of system-on-chip (SOC) designs. On-chip test compression has become a mainstream design-for-test methodology. In this approach, the tester supplies compressed test patterns, which on-chip decompressors expand into data for the scan chains. These techniques are extensively adopted to enhance testing efficiency in modular, core-based SOC architectures.

System-on-chip (SOC) test schemes utilize dedicated on-chip infrastructure, including test access mechanisms (TAMs), test wrappers, and various test pattern scheduling algorithms. The design of the test wrapper is presented initially in [17]. Later, the test wrapper design was standardized by IEEE Standard 1500 [18]. SOC designs typically integrate a diverse array of complex intellectual property (IP) cores, each operating at distinct clock rates, power requirements, and voltage levels. A persistent challenge in SOC testing is that the available test bandwidth, defined by the total number of input/output scan ports, is generally less than the aggregate scan channel requirements of all embedded cores. To address this limitation, test architectures employ TAMs [19] to transfer test data between the SOC's external scan ports and internal cores, while test wrappers standardize the interface between each core and the broader SOC environment. Use of TAMs has been proposed since 1998 [20][21]. Scan test mode configuration for Test Access Mechanisms (TAMs) or functional interfaces such as I2C, SPI, and UART [22] is most commonly performed using the IEEE 1149.1 standard. Paper [23] presents an alternative approach that utilizes the scan chain itself for configuration. This method requires a small number of additional flip-flops and latches to generate control signals.

TAMs that consider compression based on EDT have been proposed in [24][25][26][27][28]. Techniques that use on-chip comparators for identical modules to save pins for checking individual cores are also proposed [29]. These components collectively enable modular and scalable test methodologies. The approach presented in [30] uses high-speed chip pins typically used for serializer/deserializer to improve scan bandwidth. Testbus and testrail-based architectures are proposed in [20] and [21], respectively. Approaches to reduce test application time by reducing wrapper cells are also proposed [31][32]. Significant power dissipation occurs while applying test [33][34]. TAMs to minimize test power are presented in [35][36][37]. LFSR-driven encoded test data with time multiplexed test pins driven by ATE (automatic test equipment) is used in [38][39] to reduce test time. [40] presents TAM with XORs added on the output side to improve bandwidth. Using packet-switched networks on chip [41] for testing is also proposed, which uses an available network for data transfer and does not add any TAM, thereby eliminating routing overhead for them.

The rapid introduction of new electronic gadget models each year has led to shorter product usage cycles among consumers. Additionally, a segment of consumers is unable to afford high-priced electronic devices. These consumers prioritize low cost and a greater number of applications, often accepting minor compromises in quality or reliability of service. An increased number of applications typically requires more cores in a system-on-chip (SoC), which raises both testing costs and testing time. While comprehensive testing of all cores ensures product quality, it also increases overall costs. Therefore, to maintain a low product price while offering a large number of applications, some compromise in quality of service may be necessary. Partial testing of the SoC can help achieve lower product costs with minimal compromise in quality. Methods for generating test vector sets that result in minimal reductions in fault coverage are proposed, and corresponding test access mechanisms (TAMs) to support these methods are discussed [42].

Furthermore, a range of test scheduling techniques has been developed [43][44] to minimize overall test time by optimizing the allocation of TAM bandwidth and test resources. One of the most popular algorithms that uses bin-packing is presented in [17]. Test scheduling for SOCs with identical cores is proposed in [45][46]. In summary, TAM functions as the essential hardware link that facilitates communication between embedded core scan channels and the chip's top-level scan input/output ports. test data.

III. CLASSIFICATION

Designing an efficient test access mechanism and determining test scheduling to operate cores using chip pins is key to minimizing test cost. Researchers have proposed multiple techniques towards this. This survey attempts to review those techniques. The following are the major approaches based on the parameters they use to design TAM and/or plan a test schedule. The categories are based on the inherent principle used. A few of the techniques may fall in multiple categories; they have been classified based on their main characteristic.

A. Test Access Mechanisms for In-System and System Level Testing
B. Test Access Mechanisms for 3D Chips
C. Test Access Mechanisms for NOC Chips
D. Packetized Test Access Mechanisms
E. Efficient Test Scheduling
F. Test Access Mechanisms for Low Power
G. Options for Configuring Test Access Mechanisms
H. Combining Approaches for Bandwidth Management
I. Test Access Mechanism for Simultaneous Identical Cores Testing
J. Test Access Mechanisms for Pin-mux and EDT Based Architectures

IV. TEST ACCESS MECHANISMS FOR TESTING CHIPS

A. Test Access Mechanisms for In-System and System Level Testing

Traditionally, in-system test (IST) has relied on built-in self-test (BIST) methods such as X-tolerant Logic BIST (XLBIST) [5] for power-on/off or periodic testing. However, BIST patterns are generated locally on the device and typically provide only a pass-fail result. Using structural patterns driven by high-speed functional pins addresses this limitation and accelerates the debugging process. Additionally, reusing existing High-Speed I/O (HSIO) access mechanisms enables deterministic structural tests on cost-effective System Level Test (SLT) platforms. Previous techniques have utilized functional interfaces such as Serial Peripheral Interface (SPI) and Inter-Integrated Circuit (I2C) to drive the IEEE 1687 network [47][48]. However, these interfaces are unsuitable for delivering large volumes of scan data due to their low operating speeds and limited utility.

The Synopsys HSAT IP functions as a interface between the system-on-chip (SoC) High-Speed I/O (HSIO), such as a USB controller, and the Design-for-Test (DFT) infrastructure, transferring structural test patterns to internal scan chains and Test Access Port (TAP) controllers. Block diagram of the IP is shown in Fig. 1. The IP supports multiple configurations and loopback modes to facilitate debugging. Operating on the AXI clock, the HSAT IP processes data through its packet decoder and encoder via a standard AXI interface, which simplifies integration and verification. Incoming packets are directed to the Control Unit, SCAN Bridge, or TAP Bridge to manage IP operations, scan, and JTAG networks. Response data from these bridges is packetized and transmitted back through the AXI interface. The DFT-side interface drives scan-in, clocks, and controls into the Scan/TAP network in a manner consistent with Automated Test Equipment (ATE).

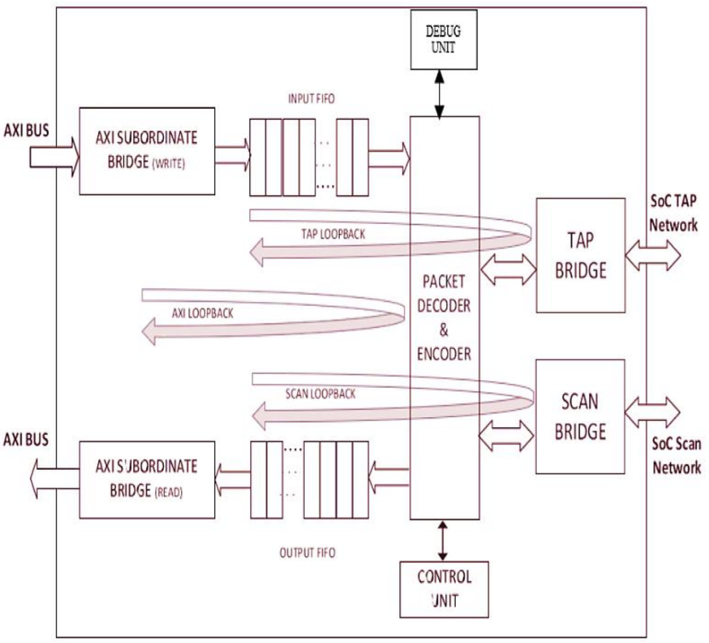

Fig. 1. HSAT Architecture [49]

Suri et al. presented a method [49] that utilizes high-speed chip pins for test access by integrating Synopsys's High Speed Access Test (HSAT) IP. This IP bridges the System-on-Chip's (SoC) High-Speed I/O (HSIO), such as USB, with its Design-for-Test (DFT) infrastructure to enable efficient pattern transfer. The approach repurposes the SoC's HSIO interface, including PCIe or USB, for structural test pattern application and capture. In contrast to IEEE 1149.10, this method requires minimal additional hardware, as it leverages existing mission-mode interfaces for test data transfer. Both Automated Test Equipment (ATE) and System Level Test (SLT) platforms can support this flexible access. The integration of HSAT IP enables the USB interface to serve as a bridge for high-speed test data.

Utilizing HSAT IP also enables diagnostic capabilities for both pre-silicon simulation and post-silicon System Level Test (SLT) failures. The TestMAX ALE solution is integral to this process, providing diagnostics that allow engineers to apply stimuli, inject errors, and analyze design responses. This approach facilitates early bug detection, cost-effective debugging, and enhanced observability during simulation. The author reported that testing of Chain, Stuck-At Fault (SAF), and Memory Built-In Self-Test (MBIST) patterns using this solution has resulted in a 100% pass rate.

Pandey et al. [6] also utilized HSAT, uses the PCIe interface to drive scan patterns at high speed. Their work demonstrated successful application of HSAT with PCIe on a System-on-Chip (SoC), achieving bandwidth comparable to 318 General Purpose Input/Outputs (GPIOs) by using 64 GPIOs supplemented with several PCIe Gen2 I/Os. The authors further explored the use of HSIO for in-field Silicon Lifecycle Management (SLM) [50], leveraging the HSAT architecture to drive scan, JTAG, sensors, and on-chip monitors. As semiconductor complexity increases and expectations for SoC performance and longevity rise, continuous silicon monitoring throughout its lifecycle becomes essential to maximize performance and detect defects before they affect system operations. SCAN vectors are reused for SLM alongside embedded sensors, enabling proactive failure detection. The substantial volume of data generated necessitates a robust network and access mechanism. The industry's shift toward High-Speed IO for in-field test access presents an opportunity to use the same mechanism for SLM, as these interfaces' native protocols provide significant bandwidth. The authors also proposed employing HSAT with streaming scan and DFTMax-based designs [51].

Yilmaz et al. presented the mechanism to access test-data over High-Speed Link (MATHS) [8]. It is a high-throughput system based on Peripheral Component Interconnect Express (PCIe) for structural testing of SoCs at wafer, package, and system levels. While the architecture is PCIe-based, it can be extended to other high-speed functional interfaces. The system eliminates the need for costly test equipment by removing the I/O and memory per I/O requirements typically associated with conventional test setups. Because the mechanism adheres to PCIe standards [52], it is portable across Automated Test Equipment (ATE), System Level Test (SLT), board, and in-field testing platforms. MATHS serves as a tool for enabling ATE-to-system correlation, which directly impacts turn around time (TAT). MATHS operates as a separate chip mode and requires both an on-chip hardware component and unified MATHS software, which is portable to any supported platform.

B. Test Access Mechanisms for 3D Chips

Following the assembly of a three-dimensional (3D) package containing multiple dies, testing is conducted to verify that each die within the package operates within specified parameters. This process typically requires each die to have a direct connection to the package input/output (IO). However, in packages with a large number of dies, providing every die with a direct IO connection is difficult. Consequently, such packages are designed so that one die serves as the IO access point (IO-die), while the remaining dies (core-dice) connect to the IO-die via a die-to-die interconnect, as illustrated in Fig. 2. In this architecture, the die-to-die interconnect can become a bottleneck for test data transfer. When the interconnect functions as a high-speed serial interface with a limited number of lanes, parallel test data transfer is constrained. To address this challenge, serializing the test data is one solution. Such serial/de-serial paths are shown in Fi. 3. This serialization and deserialization process must be transparent to automatic test pattern generation (ATPG) tools. Further, the approach should allow for the reuse of ATPG patterns in both wafer sort and final package testing. Paper [53] introduced a scan serializer/deserializer (SerDes) implementation that reduces testing costs for multi-die packages with shorter die-to-die link width. This solution requires minimal programming overhead, and involves only a few additional test data register (TDR) configuration writes. It also resolves indeterminism associated with SerDes links. The author applied this solution across multiple products which enabled ATPG tools to utilize the SerDes for data transfer. ATPG patterns were generated by increasing the total pipeline stage count by an amount corresponding to the latency introduced by the scan SerDes interface.

Prior to packaging a 3D integrated circuit (IC), pre-bond testing of the cores in each layer need to be conducted. Following packaging, post-bond testing of the entire 3D IC is required. The combined need for pre-bond and post-bond testing increases the number of test pads required for 3D ICs compared to traditional non-stacked ICs. This increase in test pads leads to larger circuit area and increased routing congestion. Hsu et al. presented an algorithm [54] to reduce test time by minimizing the number of test pads. Their approach utilizes an integer linear programming (ILP) formulation to optimize test scheduling based on these constraints.

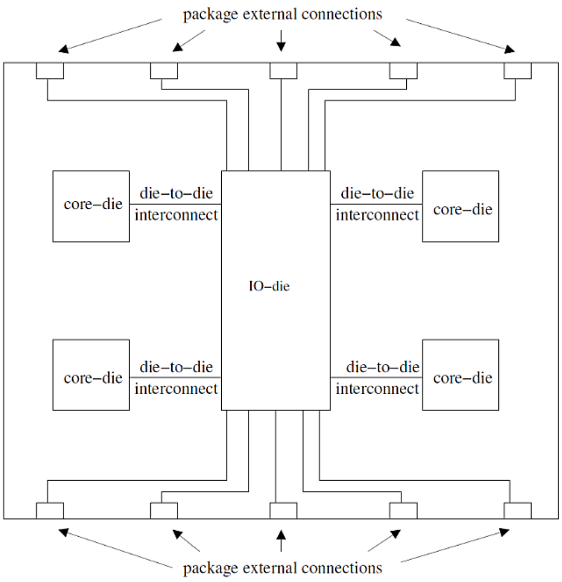

Fig. 2 Multi-die Package [53]

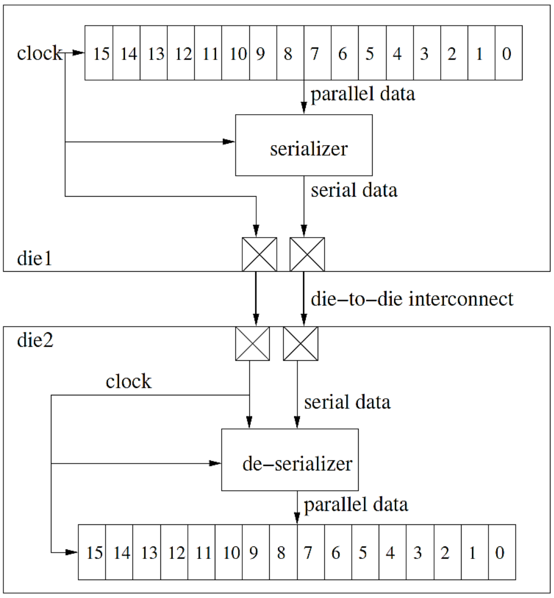

Fig. 3. SerDes Concept [53]

Choi et al. presented a reconfigurable Test Access Mechanism (RTAM) for three-dimensional integrated circuits (3D ICs) [55]. The RTAM is based on the IEEE 1687 scan network, and provides flexibility for reconfiguring the scan path. This standard allows the scan path to be altered without interrupting ongoing tests while test data are being transmitted. The states of the Scan Interface Block (SIB) chains simultaneously modify the test widths and the test order of the cores under test (CUTs). Fig. 4 illustrates an example of the test access structure for a core with three scan chains, where the test width ranges from 1 to 3. SIBs 1, 3, and 5 determine whether the corresponding scan chains are included in the test path. The RTAM extends the architecture of reconfigurable core test access. One set of SIBs controls the transmission of test data to the scan chains, while another set manages reconfiguration. The authors presented experimental results for their reconfigurable TAM in pre-bond, mid-bond, and post-bond tests. These results were compared with those of a non-configurable TAM presented in [56], and they showed improvement in test time.

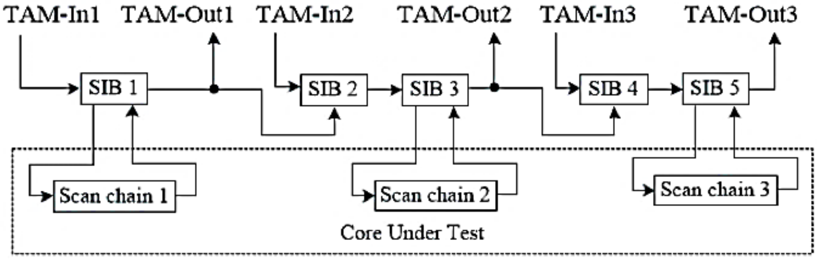

Fig. 4. Reconfigurable Core Scan Test Access [55]

Karmakar et al. presented two approaches for testing three-dimensional integrated circuits (3D ICs) [57]. The first, a step-by-step approach, develops power-restricted test schedules for each die and subsequently determines test concurrency between the dies. It checks number of test pins, through-silicon vias (TSVs), and power limits of the stack while deciding schedule. A Particle Swarm Optimization (PSO)-based technique is employed to select resource allocation, power distribution to individual dies, and their internal test schedules. Using PSO in two stages of optimization results in a significant reduction in the overall test time of the IC. The second, an integrated approach, utilizes PSO to generate a power-constrained test schedule for the entire IC in a single optimization step. Experimental results demonstrated that the integrated approach yields a schedule with shorter test time than the step-by-step approach due to its greater flexibility and fewer restrictions.

A basic approach to optimized test scheduling is to initiate testing of all cores simultaneously. However, concurrent testing of all cores is typically infeasible due to several system and die limitations. They include constraints on Test Access Mechanisms (TAMs) and power consumption. For systems with a relatively small number of cores, optimal scheduling can be generated automatically, but heuristic algorithms are required for larger systems. The traditional method organizes core tests into sessions, where all tests in a session begin simultaneously and the session concludes with the longest test. Paper [58] presented a session-less scheduling algorithm, that allows the test procedure of any core to commence at any time, provided that TAMs are available and power constraints are met. The session-less schedule showed an average of 44% improvement in test time during experiments compared to session-based schedule.

C. *Test Access Mechanisms for NOC Chips*

A shared BUS-type data interconnect structure has been widely used for multi-site testing to apply patterns to multiple devices under test (DUTs) simultaneously. However, increasing the number of DUTs and the testing frequency introduces significant challenges related to signal and power integrity within the shared bus architecture. Additionally, the number of DUTs that can be connected to the bus is limited. Complexity of bus controller operations, such as bus prioritization, increases proportionally with the number of connected DUTs. To address these limitations and improve parallel testing efficiency, new models and templates tailored for multi-site testing are required. One promising approach is the Network-On-Chip (NOC)-based test access architectures. An NOC is an interconnection model implemented on a chip as a micro-network. Fig. 5 illustrates several well-known NOC topologies. In this figure, rectangles labeled 'R' represent routers, while unlabeled rectangles denote DUTs. With NOC-based platforms, the number of DUTs that can be connected to a single interface board is theoretically unlimited when NOC is used as the interconnect structure.

Typically, NOC is implemented using field-programmable gate arrays (FPGAs) or dedicated application-specific integrated circuits (ASICs). Automated test equipment (ATE) serves as a test sink, and a test result analyzer can be installed on the interface board to facilitate the analysis process as shown in Fig. 6. Because the NOC can buffer data to compensate for differences in operating speeds between connected modules, the speed settings among the ATE, NOC interconnect, and DUTs are flexible. Therefore, at-speed testing using low-speed ATE can be achieved without the need for precise clock control or synchronization circuits. Previous multisite test methods have been described in [59][60][61]. In [62], Hong et al. presented a test architecture based on NOC interconnect that employs packet-type test patterns to enhance test parallelism and enable simultaneous testing of many DUTs.

Table I presents results reported by the authors when this architecture was implemented using ISCAS CS298 circuits, which shows up to a 28.47% reduction in test time.

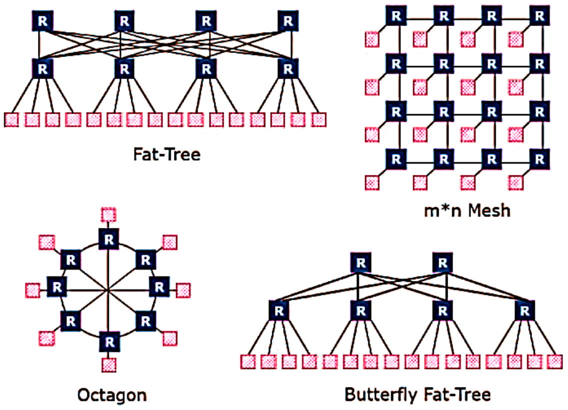

Fig. 5. NOC Topologies [62]

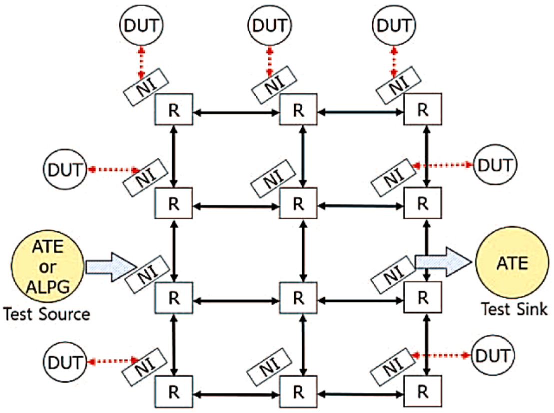

Fig. 6. Multi-die Package [62]

TABLE I. EXPERIMENTAL RESULTS [62]

# of DUTs	Test Time (us)	Reduction Ratio (%)
1	3.261	0
16(1)	41.113	26.91
16(2)	41.481	25.78
32	81.865	27.47
64	163.369	27.75
128	324.137	28.78
230	583.809	28.47

Many test schemes have been published that propose reusing on-chip bus-based interconnects as test access mechanisms (TAMs). However, these approaches are not effective for TAMs implemented with networks-on-chip (NoCs), because packet-based NoCs do not allow partitioning and separate control of network links without modifying the network routers for test mode. So, specialized test schemes for NoC-reused TAMs have been proposed [63]. Time-division-multiplexing (TDM) enables the delivery of test data to different devices under test (DUTs) during distinct time slots (TS) through assigned input access points (APs). The TDM technique has been incorporated into many test schemes [64], resulting in effective solutions. In NoC-based architectures, the core transmits data, and the packetizer generates packets. These packets are injected into the network and traverse the network to reach their destinations. A flit is the smallest unit of flow control, and a packet consists of multiple flits. The flit size is equal to the network channel width, except for header flits.

Paper [65] checks stimulus and response flits with varying automatic test equipment (ATE) channel widths for the d695 benchmark system-on-chip (SoC). The analysis utilizes the rectangle packing test algorithm [66] for three distinct cases. Experimental results for the d695 benchmark SoC are presented in Fig. 7 and Fig. 8, with ATE channel widths set at 32, 64, and 128 bits. Each stacked bar in the figures represents the total number of injected test data flits for all modules in the d695 benchmark SoC. As shown in Fig. 7, increasing the flit size, or ATE TAM width, leads to a decrease in the number of test stimulus flits across all three cases. This reduction occurs because (1) a wider flit enables allocation of a wider TAM width to a DUT, which is effective in case 1, and (2) a wider flit allows embedding more test vectors within a single flit, which is effective in cases 2 and 3. For cases 1 and 2, if the TAM width of a DUT is narrower than the flit size, only a single test response set is embedded in each flit, resulting in the transport of a significant number of idle bits. In contrast, for case 3, this number is significantly reduced due to the accumulation of multiple test-response sets within each flit.

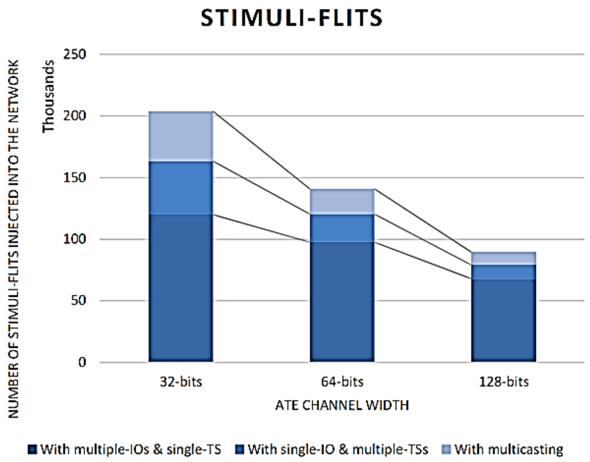

Fig. 7. Test Stimuli Flits [65]

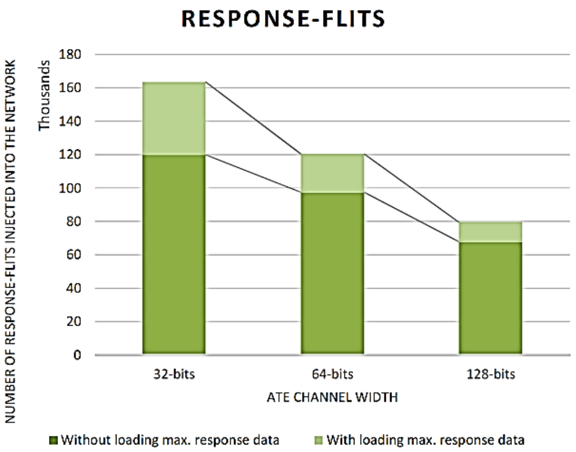

Fig. 8. Test Response Flits [65]

D. *Packetized Test Access Mechanisms*

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

1) Streaming Scan Network: Pattern count and chain length differences of cores result in creating inefficient pattern retargeting when those cores are tested in parallel. Jean-Francois presented an architecture and scheduling algorithm [67] called Streaming Scan Network (SSN) that reduces these efficiencies greatly. With this method, any number of cores could be tested with limited chip pins. In hierarchical testing, groups of cores to be tested need to be identified early in the design cycle. Due to limited pattern count information at the stage, the planning often does not result in optimal scan channel bandwidth usage. Broadcasting scan data to input pins is popularly used to save input channels per core. However, with this method, outputs need to be independently observed and so need at least one scan channel per core. The packetized bus-based data transfer architecture presented in [29] limits the minimum number of channels and constrains the number of channels for a core.

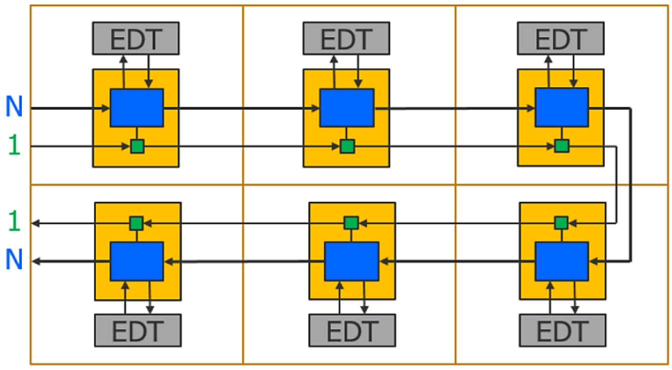

Fig. 9. SSN Architecture [67]

Authors of this architecture presented a mux-based scan pin switching network in [68] and showed dynamic bandwidth management enablement using it in [27]. One drawback of these architectures is the high cost of routing, which adds constraints for tile-based SOC designs. Structural test fabric has overhead for the data word in a parallel bus, in general. Capture clock is applied simultaneously to all cores with this method. But due to differences in pipeline stage based on the physical location of the core, simultaneous application of capture clock won't become possible as scan loading operation for them won't finish simultaneously for all cores.

The author proposed an architecture in which each core contains a host controller (streaming scan host - SSH) that controls the scan operation of the core. As shown in Fig. 9, the cores are connected to the scan data bus, which is called here the SSN bus. The host controller takes data from the bus and loads it into the scan chains of the core. It puts unloaded data from chains as output on the bus. Though the diagram shows EDT as scan compression logic, SSN could be used with uncompressed chains or a combination of compressed and uncompressed chains in a block. The SSH has two interfaces as shown in Fig. 10 – one for configuring various parameters and another for payload scan data. IEEE 1687-based [15] IJTAG network is used for configuration and performs setting up multiple configurations, including the bus width of SSN getting used, the location of the block on the network, and shift cycles required for the scan pattern for the block. After performing the configuration, test data is loaded in SSH as packets of data. No opcode or address information of the block is sent in the packet, and only scan data is transferred in the packet. Each SSH identifies which data bits to use from the bus. Control signals for scan and EDT operation are generated using local logic in the SSH controller. The SSH contains bypass muxes, which enable passing data to the next block without modification when the block is not under the current test. Further, muxes could be inserted on the data path, which enables including or excluding a portion of the network from the current operation. The SSN bus width is determined based on the availability of chip pins and does not depend on the number or sizes of cores.

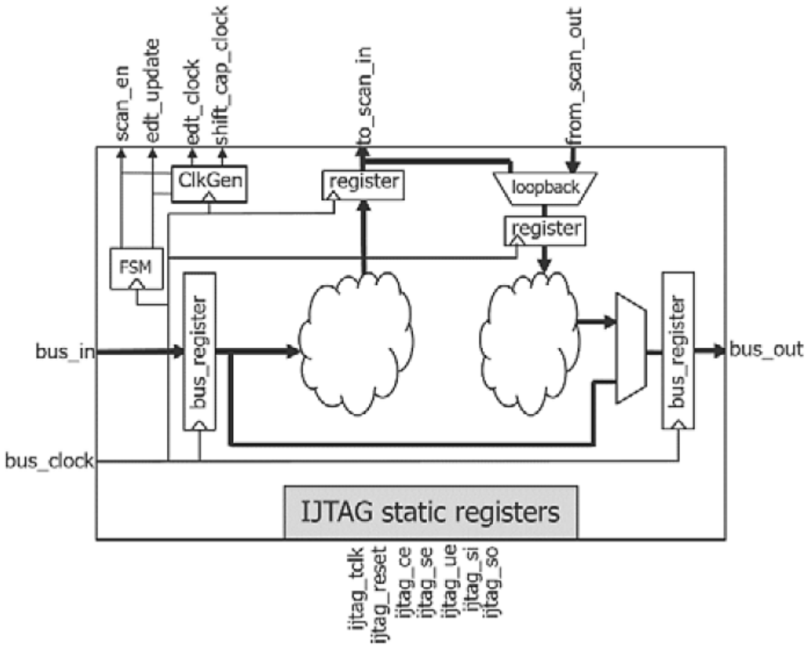

Fig. 10. Streaming Scan Host (SSH) node [67]

Scan data can be routed to any core on the bus, and cores can be selected dynamically while data is streamed through the bus. This enables the capability of choosing any combination of cores to be tested in parallel. Pin-mux-based architectures like [14][69] also offer the choice of combinations of cores for testing, but usually they need to deal with congestion for routing wires. Architectures presented in [12] and [13] use a serializer/deserializer. It enables using high frequencies for chip pins, thereby improving scan bandwidth utilization. However, they also have routing challenges for tile-based SOCs. Architecture presented in [14] also suffers from this limitation. SSN-based architecture offers this capability without any possibility of routing congestion. SSN is also compatible with the packet distribution architecture described in the IEEE 1149.10 standard [70] for a high-speed test access port.

a) *Packetized Data Transfer:* A packet consists of a bit stream required to perform a shift operation for all blocks under the current test. The ATE continuously drives these packets to be delivered to various cores. This technique enables blocks with a varying number of scan channels to be used and makes their design choices independent of the SSN bus width. Contrary to pin mux architectures [14][69][71], 8-bit data could be driven from ATE using SSN, while they would have required a 9-bit bus for testing together two blocks with 4 and 5 channels. The SSH host controls the operation of choosing bits from the bus and applying them over EDT channels during load operation, and placing data from EDT on appropriate bus bits during unload operation. Time slots in a packet are usually the same for scan-in data and scan-out data for a block.

b) *Efficiency in Packing Data:* Differences in shift length across blocks result in padding in the pattern in pin-mux architectures. A block with a shorter chain needs to match the length of the longest chain among cores and add padding in patterns as shown in Fig. 11. This padding results in wastage of bandwidth and inefficiency in retargeting patterns. Further generation of control signals and clocking locally enables retargeting of blocks with varying pattern count. Blocks with smaller scan data could finish early, and they could be replaced with other blocks subsequently. Also traditional pin mux architectures needed to pad the scan data of such smaller blocks, and this padding further reduces efficiency. Since control signals for scan operation are generated locally within each block with SSN based architectures, the scan operation for it does not depend on other blocks. Capture can be performed as soon as loading is finished, and it need not wait for the finishing loading of other blocks participating in the current test. This offers a great advantage over pin-mux architectures and results in significantly improved efficiency in packing data during pattern retargeting.

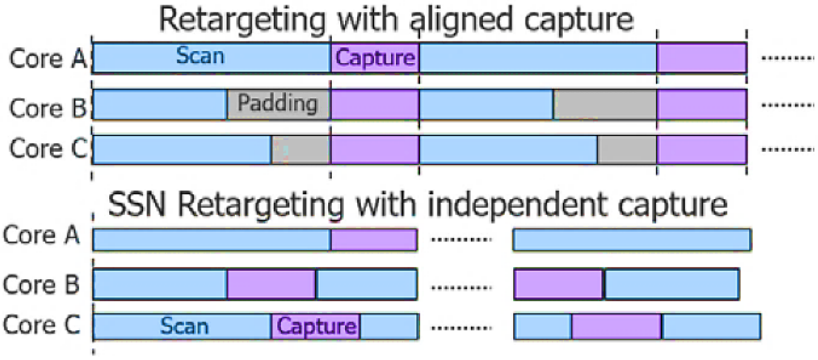

Fig. 11. Retargeting with aligned vs. independent capture [67]

2) SSN with High Speed IOs: Paper [72] implemented SSN with high-speed input/outputs (IOs) and utilized PCIe for extended structural testing [8]. The study presented a comparative analysis of scan architectures, including point-to-point, broadcast, and streaming scan approaches, when integrated with high-speed interfaces such as PCIe. In one configuration, all general-purpose input/outputs (GPIOs) were used, while other configuration combined a reduced number of GPIOs with PCIe for scan operations. The latter configuration showed up to a 7X improvement in test time. Most functional logic was tested using the MATHS infrastructure, which uses an attached personal computer (PC) for pattern storage. The requirement for automatic test equipment (ATE) vector memory for stuck-at and transition delay faults was reduced by nearly 20X. This testing strategy eliminates the need for costly ATE equipment by removing the necessity for memory allocation per I/O. As process technology nodes advance and design complexity increases, the demand for memory to store automatic test pattern generation (ATPG) patterns also rises. The proposed solution provides a cost-effective storage method for ATPG patterns. The authors also reported that using SSN improved the minimum voltage required for testing (Vmin). With the latest process technology nodes, testing at the lowest possible voltage is must for defect detection. The new scan architecture significantly improved Vmin with the PCIe-based streaming scan solution, achieving values close to timing signoff voltages compared to traditional scan designs. There was no adverse impact on Vmin due to shift noise, as the shift frequency was reduced and shift clock edges were staggered across cores during concurrent testing. Vmin thus accurately reflects functional capture paths and aligns with timing closure targets.

3) Methods Of Securing SSN: SSN architecture poses security risks in the absence of robust access control mechanisms. Malicious actors may use the SSN to extract sensitive information from the system-on-chip (SoC), thereby threatening overall system security. Although SSN facilitates efficient SoC testing, it introduces vulnerabilities. For example, hackers can shift in crafted test data and extract test responses to obtain confidential information. Such vulnerabilities may result in severe consequences, including reverse engineering, counterfeiting, and integrated circuit (IC) overproduction [73][74]. Additionally, SoC cores from untrusted third-party intellectual property (IP) vendors increases the risk of compromised cores, which can enable data alteration or sniffing attacks [75][76]. Therefore, implementing stringent access controls and restrictions within the SSN is essential to prevent unauthorized access and protect sensitive data. To address these security challenges, many countermeasures have been proposed.

Rajski et al. introduced a Root of Trust mechanism that integrates a scrambler, descrambler, ring generator, secure server, hash core, random number generator, and secret memory [77]. This mechanism uses a challenge-response protocol, which requires users to generate a correct response to a given challenge to obtain authorization. If the response is incorrect, an obfuscated output is provided. Subsequently, the authors proposed a stream cipher-based technique to secure test data in SSN architecture [78]. This scheme utilizes an input stream cipher to decrypt test data before it enters the SSN and an output stream cipher to encrypt test responses before they are shifted out. In [79], Shukla et al. proposed a secure SSN architecture designed to prevent data sniffing attacks. This architecture restricts the offset of each SSH node, ensuring that no node can access data belonging to another node, thereby effectively mitigating data sniffing threats.

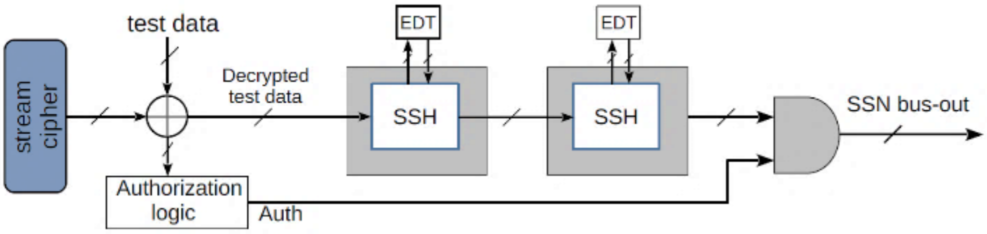

Fig. 12. Secure SSN Architecture [80]

Riaz et al. presented a secure SSN architecture [80] that uses test data encryption and test response masking. This architecture is shown in Fig. 12. It ensures that only users possessing the correct authentication key can access the test response, thereby enhancing the security of the testing process. Test response masking is implemented through user authorization. By eliminating the stream cipher on the output side, the architecture achieves a significant reduction in area overhead while maintaining the same security level as [78].

4) Bandwidth Improvement by Eliminating Redundancies: Automatic Test Pattern Generation (ATPG) patterns contain randomly distributed don’t care data, which are redundant and do not contribute to improving the Quality of Results (QoR). These redundant data increase the cost of testing by increasing the cycle count. Fig. 13 illustrates ATPG patterns for cores A and B with both care and don’t care cycles. In this figure, white bars indicate don’t care data cycles while colored bars represent care cycles. Interleaving patterns by exploiting these redundancies, rather than serializing them over a shared channel, can significantly reduce test application time. By using the structural distinction between care cycles and don’t care cycles in a sequential compressor, Gnawali et al. proposed a protocol-aware Streaming Fabric-based solution [81] to improve bandwidth utilization. This solution consists of two components. The first is a programmable streaming fabric that enables cores to detect and ignore dataless cycles. It maintains clock signals for autonomous operation. The second component is software that schedules pattern data for different cores into the identified dataless cycles. Interleaving care data into each other's underutilized bandwidth improves overall bandwidth utilization and reduces turnaround time (TAT). The authors reported a 43% reduction in test cycles across four industrial designs using this solution compared to the baseline streaming fabric.

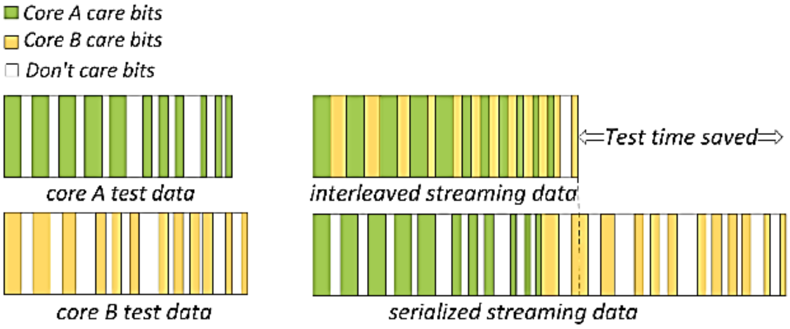

Fig. 13. Data Streaming [81]

E. *Efficient Test Scheduling*

The Various approaches have been developed to address the test scheduling problem. Some studies have modeled the problem as an integer linear programming (ILP) problem, formulating the connection information between modules and test access mechanisms (TAM) as well as the wrapper design for each module. The optimal solution is then obtained using an ILP solver. However, since the ILP problem is classified as NP-hard, solving the scheduling problem is highly time-consuming. To mitigate the issue of extended computational time, alternative strategies utilizing metaheuristic algorithms, such as the BAT algorithm, have been proposed [82]. These methods can generate solutions in a relatively short time through efficient search strategies. Repeated searches still require considerable time, and the solutions obtained are not guaranteed to be optimal. Given the computational complexity of NP-hard problems, recent research has increasingly focused on heuristic algorithms, which offer significantly faster solutions compared to traditional optimization methods [83], [84].

Kim et al. presented a dynamic pairing algorithm [85] that accounts for the mutual influence of tests during the scheduling process. This method identifies available test resources for specific pairs and minimizes the test time of paired modules under defined constraints. The algorithm uses a heuristic-based approach to reduce CPU time required for scheduling. It sequentially schedules target modules and determines the optimal test schedule through dynamic pairing. Scheduling is conducted by traversing the modules under test following a preprocessing phase. The process consists of three phases: greedy placement, dynamic pairing, and optimization. Experimental results showed substantial improvement compared to the scheme presented in [83].

F. Test Access Mechanisms for Low Power

In three-dimensional (3D) integrated circuits (ICs), the upper dies are accessible only through the bottom-most die, which significantly complicates testing compared to conventional two-dimensional (2D) ICs. Therefore, the test access mechanism (TAM) must be extended to all dies in the 3D IC via through-silicon vias (TSVs). Test infrastructures such as test pins, TSVs, and routing area in 3D ICs are highly constrained. So designing TAM requires to use all these information. Karmakar et al. [86] previously introduced metaheuristic-based, thermal-aware test scheduling algorithms for 3D ICs. However, evaluating power during scheduling increases test time. Metaheuristic optimization techniques, including particle swarm optimization and simulated annealing, are computationally slower than purely heuristic approaches. To address this issue, Chatterjee et al. presented a purely heuristic approach [83] to balance test time and peak power in the generated schedule using a cost function. This method proposed a heuristic test scheduling algorithm for 3D ICs that reduces total test time while adhering to TSV constraints and optimizing peak power consumption. The algorithm employs a dimensionless parameter, α, referred to as the weight factor.

When α equals one, the algorithm prioritizes the test time, resulting in increased parallelism and reduced test time without considering power dissipation. Conversely, setting α to zero emphasizes power reduction, thereby minimizing the peak power dissipation of the schedule. Fig. 14 shows the test time achieved using this approach as the TAM width varies.

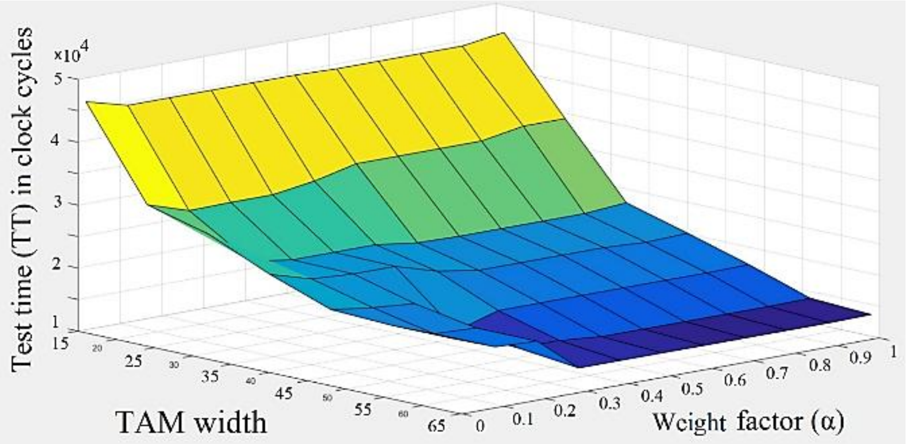

Fig. 14. Variation in Test Time with TAM Width and Weight Factor [83]

Paper [42] proposed a test methodology to reduce Test Power (TP) and Test Data Volume (TDV) while maintaining acceptable testing quality. This approach utilizes approximate computing to decrease both test data volume and power consumption. In [87], an Approximate radix-2 hybrid redundant multiply-and-accumulate (ApproxMAC) unit was presented which also uses approximate computing. The core concept of partial testing involves selecting output bits of System-on-Chip (SoC) cores that can tolerate errors, thereby excluding faults in sub-circuits within the cone of influence of these outputs from testing. Faults excluded in this manner are termed insignificant faults. In most SoC core cases, these insignificant faults are typically absent from the circuit, ensuring that quality of service is not compromised despite reductions in TDV and TP. To maximize the number of bit positions that can be omitted from each test vector with minimal reduction in fault coverage, a Particle Swarm Optimization (PSO)-based detector is used. Test vectors are applied to the core via the wrapper chain, which consists of serially connected wrapper cells. Total power consumption is directly related to the number of transitions in these wrapper cells. Based on this principle, the author proposed a power-aware Test Access Mechanism (TAM), which also reduces the time required to load vectors and utilizes a modified set of vectors. The author reported an average power reduction of 16.65% with this architecture and the modified test vectors.

G. *Options for Configuring Test Access Mechanisms*

Since the test application employs several configurations of channel allocation, the amount of control data is significant, and so it also impacts test time. Paper [28] proposed 3 options to load control data and described how their architecture supports them.

1) IJTAG Interface: The Internal Joint Test Action Group (IJTAG) standard-based test control manages on-chip test and debug instruments. It provides a way to describe operations using IEEE1149.1-based Test Data Registers (TDRs). Fig. 15 shows a typical chip with a TAP controller and two cores. A path is selected using segment insertion bits (SIBs). A SIB enables adding an instrument to the current test. Data is loaded in TDRs of core 1 or core 2 by writing to the corresponding SIB. The TDR in the switching register works based on control data, which determines the core to be selected for the current configuration and to which chip pins they need to be connected. Since the IJTAG network is used to control various features of the core, using it for control-data delivery of the switching network could be done easily. IJTAG works on a slow clock, typically 10 to 20 times slower than the scan shift-clock. So, using this method for delivering a huge number of network configurations incurs significant test setup overhead. Though there could be ways to improve the serial operation, like using parallel paths for accessing multiple cores TDRs or loading control data during the last scan pattern, they add design complexities.

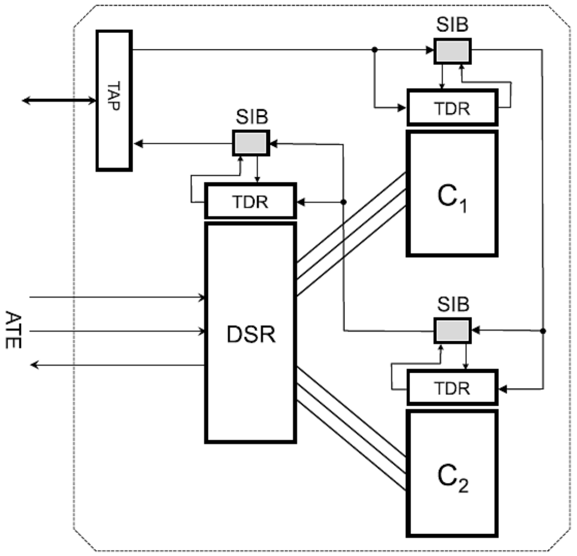

Fig. 15. Using IJTAG network to transfer control data [28]

2) Dedicated Control Chain: Separate chains could be built to load control data bits into registers. A pattern could be used to load this chain, which configures demux and mux in the input and output channels paths. Since this chain needs to be checked, the pattern could have only loaded data. Most ATEs mandate that the length of all chains be matched, so the loading operation cycles of this chain have to match the regular scan chain cycle count. During the test application, the first pattern needs to consist of control data bits. After applying this pattern, regular scan patterns for selected cores can be applied. Since the control chain also works on the shift-clock, the speed of operation of the configuration network is high, and so applying many test configurations does not incur much test setup overhead that occurs when the IJTAG interface is used.

3) Pipeline Registers: Control data bits registers could be added as shadow elements of regular scan chains. The author proposes a way to operate them to configure a switching network to connect selected ATE channels with cores using testing. The shadow registers are updated to load the required value only at the end of the load operation of the regular chain. The data values are loaded as part of the last scan pattern so that the network is ready to be used for the next test configuration. Also, since values to control bits are updated at the end of each scan pattern, they also need to be loaded as part of every scan pattern, even when the same test configuration is used for them. Since the number of control data bits is usually small, adding extra bits in the regular scan chain adds a tiny fraction of overhead to the test setup.

H. Combining Approaches for Bandwidth Management

An industrial multicore-based system on a chip (SOC) employing embedded deterministic test faces significant challenges in optimizing test time due to inefficient bandwidth allocation to cores. Paper [28] presents multiple architectural features that enable the test schedule to optimize bandwidth allocation. The architectural features also handle physical constraints efficiently. The test scheduling algorithm is adjusted for defining effective test configurations, partitioning chip pins, allocating pins to cores using test data volume, and flexible ATE channels usage for cores.

SOCs contain wrapped cores with individual EDT controllers. The EDT controllers interface with chip pins and are driven by ATE. Test schedule and TAM assign a subset of ATE channels to each core. This allocation results in varying compression ratios, test data volume, test application time, and interface design complexity. The scheme presented by the authors in [26] is capable of allocating varying pins to cores. Since chip pins to be used for scan testing are limited, EDT channels of all cores cannot be driven simultaneously, and achieving the best compression is quite a difficult task. Since a core is used in multiple SOCs, channel count for the core in one SOC may not be optimal for another SOC. Chip designers typically generate patterns for each core, analyze them, and create a test schedule to test them in groups. Since some blocks may have too large pattern count, the allocation of channels may result in increased test time and increased test data volume.

The paper [28] presents a bandwidth management scheme allowing users to do a tradeoff in channel allocations along with improving physical constraints to minimize routing congestion due to the TAM network. It also presents several techniques to deliver data to control channel configurations. The SOC test mechanism proposed by the author consists of two switching networks. Data to different cores is passed using paths determined by control data bits generated by the test scheduler. An identical module receives the same data when in broadcast mode. The switching network consists of demuxes, which connect a channel driven by a chip pin with one of many cores. The selection of the core is determined by the address register contents. The address registers also decide which core to observe on the chip pins. Flexibility in allocating ATE channels increases encoding efficiency and compression. The assignment of allocating channels dynamically to cores based on their needs results in shorter test application time.

I. Test Access Mechanism for Simultaneous Identical Cores Testing

In systems-on-chip (SOCs) with multiple identical cores, efficient parallel testing and reliable verification of complex interconnections are essential [88]. Kang et al. introduced the Majority-based Test Access Mechanism (TAM) [89], which finds a majority value by analyzing response data from all cores. The majority value, defined as more than half of the test responses, is computed by the majority analyzer (MA) module. The Majority-based TAM compares this value with the response data from each core to identify faults. The automatic test equipment (ATE) receives the majority value and verifies it with the expected result. If the majority value differs from the expected value, it indicates that more than half of the cores are faulty. This TAM is particularly suited for systems with low yield, where a majority of cores may be defective. To facilitate this process, a register is added to each XOR gate. When the majority value matches the expected value, the system operates as a standard majority-based TAM. If the majority value differs, the register functions as a shift register to extract response data, enabling the ATE to identify faulty cores. The TAM then restarts the testing process using only the non-faulty cores.

Later, the authors presented an enhanced majority-based TAM [90] that optimizes hardware resources and test scheduling. The proposed architecture incorporates multiple majority analyzers (MAs) and partitions the cores into sections, as illustrated in Fig. 16. The MA architecture allows selection between individual core response data and the majority value. Each MA processes four response data inputs, and the outputs from four MAs are combined to form a higher-level MA. This enables scalable extension of the architecture. This design computes the majority value across all cores and compares it with the response data from the remaining cores, similar to the original majority-based TAM. Additionally, the architecture can select specific core response data and test cores within designated sections to manage test power consumption using multiplexers within the MA. Another variant, the NoC-reused TAM [91] showed that optimizing test schedules with awareness of pin count can reduce test time for a given number of pins. Kang et al. also proposed a majority-based approach for network-on-chip (NoC) systems [63], which reduces hardware overhead compared to [89] by using existing NoC infrastructure.

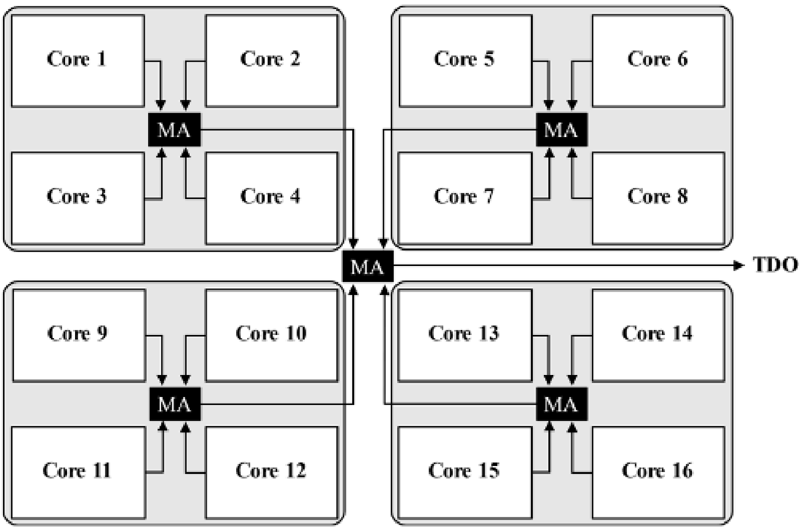

Fig. 16. Architecture of Proposed TAM in [90]

Paper [92] presented an algorithm to optimize test time. The algorithm is based on the majority analyzer and comparator presented in [89] and [93]. The author also sees the scope of improving test time using other methods. Paper [92] considers design information like the total width of the TAM bus and the number of cores on chip. It then decides the number of TAM partitions required and finds a possible combination of TAM widths(say X1) as per the TAM partition. Then it finds a possible assignment of cores to the TAM based on width and stores that information(say X2). The algorithm then selects a width from the X1 array and calculates the test time for all combinations of assignments in X2. It performs more analysis of test time with combinations of cores, TAM widths, and assignments. This way, it evaluates every possible combination and chooses the combination that offers the least test time.

J. Test Access Mechanisms for Pin-mux and EDT Based Architectures

1) Impact of EDT Architecture on Scan Bandwidth: For manufacturing testing, compression technologies like Embedded Deterministic Test (EDT) are used widely. Each core is designed to work in isolation and contains an EDT controller. Patterns generated using EDT have a high density of test cubes for initial patterns, and this fact is used by methods, such as in [94], to allocate channels to cores based on this. Paper [95] presented an analysis of different configurations of EDT for reducing the total test time required for an SOC. It presents data from industrial experiments for different EDT architectures. EDT decompressor provides continuous test stimuli to internal chains, and data is applied to them, while data coming out of chains is compressed using EDT compactors on the output side of the core. Compactors like Xpress Compactor use data from the scan chain to mask certain chains and observe others for a pattern. Keeping power in limits during scan shifting and capture is desired for modern low-power devices. Data from the scan chain is also used for controlling power during scan shifting.

Based on whether separate scan chains are used for controlling parameters of EDT, such as power and X-masking, EDT architectures are divided into Separated Control Data (SCD) and Non-Separated Control Data (NSCD). For NSCD architectures, control bits for power and x-masking are distributed across multiple chains, and the channels are used for both control and data bits for application purposes. While for SCD architectures [96][97], separate channels are used to apply control bits and data bits. The number of shift cycles for the NSCD architecture [98] consists of initialization cycles, the length of the longest chain, and pipeline stages on chains meant for control data purposes. SCD architectures save shift cycles required for pipeline stages as control data is driven by dedicated control channels, and the number of shift cycles does not change per pattern for them. The SCD architecture has a disadvantage in the form of the requirement of additional scan channels for achieving the same performance in terms of pattern count and test coverage(TC), similar to the NSCD-based architecture.

The author presented test coverage and pattern count data for 36 cores by using NSCD and SCD EDT architectures. It collected baseline data using the NSCD architecture initially. They added extra channels for the control data delivery purpose. Numbers for extra input channels are decided by dividing the number of control bits required by the length of the longest chain in the core. Fig. 17 shows comparative data for these cores. Green bars indicate pattern difference for SCD architectures compared to NSCD architectures, while red lines show differences for test coverage. The chart indicates that test coverage is maintained as much, while pattern count is reduced for most of the cores using SCD architectures. SCD gives 69% to 93% test time reduction at the expense of one extra channel.

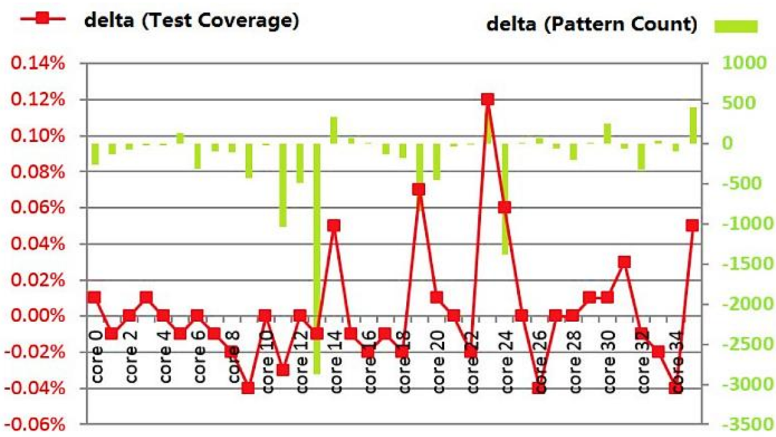

Fig. 17. Comparison of SCD versus NSCD [95]

a) Test Scheduling Using SCD and NSCD EDT Architectures: In previous studies [96][97][98] done by the author, an SCD-based architecture was used where dedicated control channels and shared data channels for non-identical cores are used. SCD was also used for identical cores by authors in other studies. As shown in Fig. 4, SCD saves both test time and improves test coverage at the expense of an extra channel. However, when data channels are shared across many identical cores, this overhead gets eliminated. The author performed a study using the proposed architectures for 8 SOCs. In the initial stage of design, it uses a compression-analysis utility to find the best possible scan channel to chain ratio for EDT. These EDT configurations form baseline architecture. Then it finds dedicated channels required for control purposes and uses them for SCD architectures. Table II shows number of configurations for these two schemes. For a few SOCs, an increase in scan channels for SCD architectures increases the number of scan groups while scheduling tests, which results in increased test time. For these SOCs, an NSCD-based architecture is selected. Table III shows the test time comparison for stuck-at fault testing for the 8 chips while using the two EDT architectures. For most chips, SCD saves 15% to 24% test time. The author reported similar savings when transition fault testing is employed.

TABLE II. SCAN GROUPS FOR NSCD VERSUS SCD [95]

SOC Name	#Scan Groups for NSCD	#Scan Groups for SCD
A	9	9
B	6	7
C	3	3
D	3	3
E	3	3
F	2	3
G	2	3
H	2	3

TABLE III. TEST TIME REDUCTION IN DC TEST [95]

Scan Group	A	C	D	E
0	23.42%	21.22%	21.28%	16.05%

Scan Group	**A**	**C**	**D**	**E**
1	23.08%	21.04%	20.19%	15.42%
2	21.32%	15.29%	2.82%	21.48%
3	20.45%	–	–	–
4	18.53%	–	–	–
5	21.32%	–	–	–
6	19.67%	–	–	–
7	18.97%	–	–	–
8	16.95%	–	–	–

2) Impact of EDT Architecture on Scan Bandwidth: Huang et al presented a step-by-step flow [99] to improve bandwidth management in SOCs using Embedded Deterministic Test (EDT). Bandwidth management is important for reducing test time. A dynamic bandwidth management may include a solver that assigns input and output channels dynamically, TAM schemes, and a scheduling algorithm as presented in [26]. Various solvers that assign input and output channels to cores, test scheduling algorithms, and test access mechanisms have been proposed for a design that uses EDT. These features enable assigning only a few of the ATE channels to cores during pattern application. The Dynamic Bandwidth Management flow proposed by the author does not require any upfront information about core patterns. It creates hardware called the bandwidth manager and inserts it into the core. This information is described in a Tessent Core Description (TCD) file automatically, which is used during the test scheduling stage.

Fig. 18 shows a design with DBM with 3 inputs and 4 output scan channels. The muxes and demuxes are used to connect individual cores with these chip pins. The red dots are registers that control selection through these muxes and demuxes. The proposed flow sets up the shift path by writing to configuration select registers(CSR). Then, a subsequent pattern loads data in Control Registers (CRs) using a dedicated control channel selected by the previous pattern. This creates a new connection from top-level pins to a subset of cores' scan channels. Connection of control chain to the muxes and demuxes is shown in Fig. 19. Such a pattern is used for each group of scan patterns. Retargeted patterns for the cores can be applied from top-level pins once these connections are made.

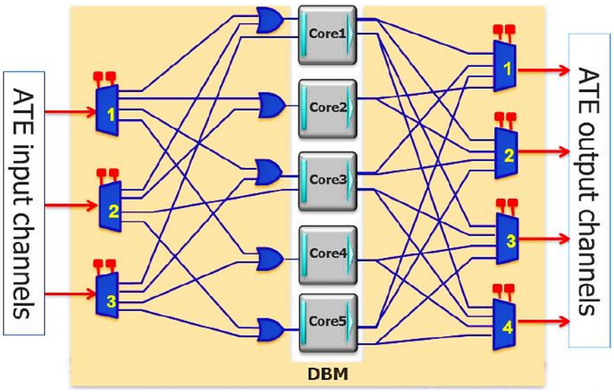

Fig. 18. Conceptual View of DBM TAM [99]

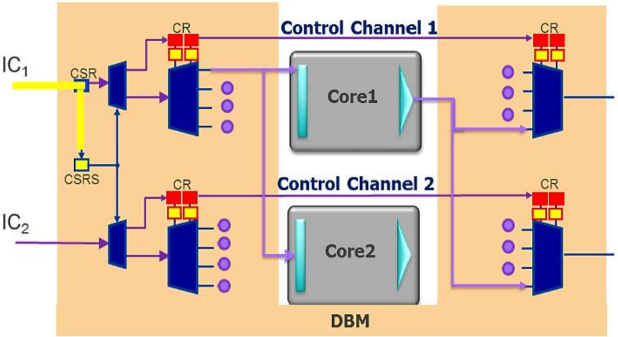

Fig. 19. Control Registers and Control Channels [99]

For the generation of DBM IP, information about cores present on the SOC, which are identical cores, how many EDT blocks are in each core, the number of input and output channels for each EDT, and the number of uncompressed chains, is provided. This information is provided to the tool as a DFT specification. Since test responses of each core need to be checked separately, the number of identical cores tested in parallel is limited based on the availability of chip pins.

Once the DBM IP is added to each core and patterns are generated for them, the test scheduling process can be started. The test scheduler extracts information about top-level pins and DBM IPs. It determines the number of configurations required, which cores to connect with top-level pins in each configuration, and programming values of various CRs. The test scheduling algorithm is based on two-dimensional bin packing [28]. All tasks in test scheduling are done automatically by the tool and produce a single pattern file in the form of STIL/WGL. All configurations are written in the pattern file irrespective of the number of configurations.

It is difficult to determine the best configuration of allocating SOC pins to each core. Using the compression_analyzer utility gives the best channel-to-chain ratio for the EDT controller; it won't help in determining SOC pins allocation to each core when a group of cores is selected for pattern retargeting. DBM test scheduling optimizes the allocation of SOC pins to cores, enabling testing some cores in parallel in order to maximize bandwidth usage. The DBM-based scheduling proposed by the author reduces pattern count by 2 to 3 times compared to non-DBM scheduling, such as in [28].

3) Independent Scan Mode Operation Cores: Differences in scan chain length result in inefficient bandwidth utilization, as smaller chains keep using the scan pins and add padding bits for the unused portion. EDT comes with additional features like x-tolerance low power controls for shifting. Based on the number of clocks and timing exceptions like multi-cycle paths, different designs need varying x-tolerance capability, which results in extra shift cycles for their programmability. For a large SOC, even if shift lengths are perfectly balanced, effective shift length varies due to these features. The author reported a variation of up to 30 percent in shift length for a large SOC. Traditionally, common controls have been used for scan operations like shift and capture. However, such a design necessitates all tiles to have the same number of over-shift cycles to match the length of the longest chain in the group. Due to these constraints, SOC architectures [14][16][100] employing EDT inherently face inefficiencies in bandwidth utilization. The author proposed the idea of an independent scan operation for each tile. He proposed to generate scan control signals locally in each tile. With such local control, each tile can perform a scan operation irrespective of operations happening in other tiles in the group. With such a scheme, the need for a global scan load/unload operation is eliminated. Each tile contains control registers to program various features like scan shift chain length, number of capture cycles, and number of patterns in the tile. The pattern retargeting algorithm accounts for each tile's unique sequences and maps the data to chip pins. The author presented the idea of SRF in [101]. Later, he presented more results of using the SRF on 4 quadrants of the chip in [68]. Scan bandwidth utilization improvements for them were up to 96%. With baseline scheme of fixed pin-muxing, utilization for quadrants were 53.79%(BL), 61.00%(BR), 58.74%(TL), 70.71%(TR). With SRF usage, utilization was improved to 92.16%(BL), 93.21%(BR), 96.13%(TL), 89.82%(TR) respectively.

4) Scan Routing Fabric: Dong et al presented an architecture [68] based on networks and switches called Scan Routing Fabric (SRF). Available chip pins can be connected to any tiles using the network and box packing algorithms can be used for test scheduling for them. During testing, most of the tile patterns in a retarget group finish, and just one tile pattern keeps applying data for a long time after them. A lot of pins do not apply data for the majority of the time [28]. This wastes bandwidth allocated to the tiles whose patterns have finished, as pins allocated to them remain idle.

Since the pattern count of tiles is unknown at the initial stage of design, determining the scan pins allocation to them is problematic. The author proposed a switching network to ease this problem. With the box packing algorithm, test scheduling to reduce white spaces in bandwidth usage can be achieved. The switching network enables connectivity of every tile with any scan pin on the chip boundary. Fig. 20 shows various components of the architecture proposed by the author.

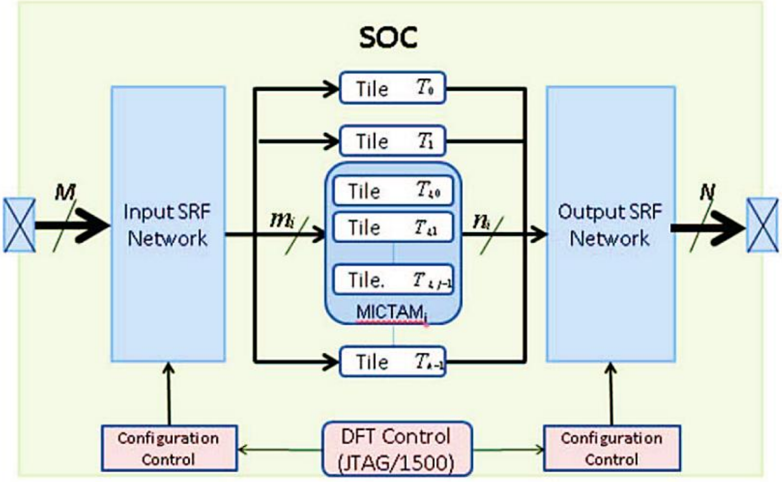

Fig. 20. Test Access Mechanism with SRF [68]

5) *Coverage-ramp Based Channel Allocation:* Abraham et al presented a channel allocation scheme [94] to optimize test time. When patterns are generated using EDT, the test cube density percent reduces drastically after the initial few hundred patterns. This observation suggests that fixed allocation of channels to cores does not yield optimum test time. The pattern is split when a test pattern is not allocated enough channels to use the number of care bits. The method that focuses on test cubes [102][103] produces improved test data volume and test time. However, they do not produce an optimal test schedule. Method in [27] tries to improve it. As shown in Fig. 21, coverage ramp changes when different channels are allocated. When optimal channels are allocated, the pattern ramp achieves maximum coverage faster.

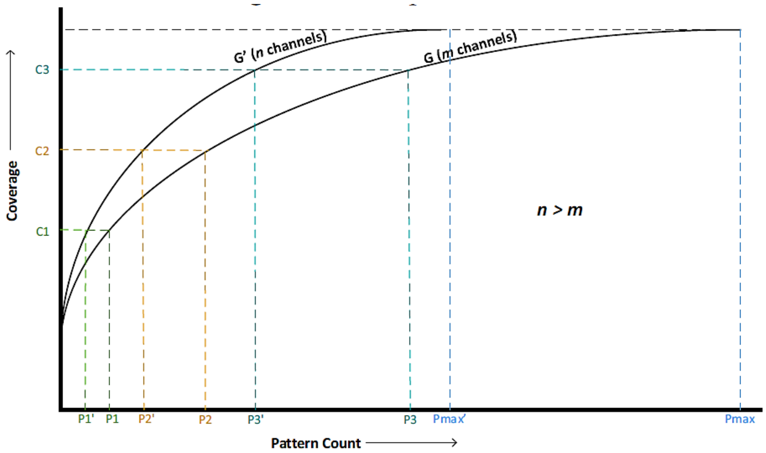

Fig. 21. Coverage ramp vs pattern count of an IP for two different scan channel configurations [94]

The coverage ramp indicates that an IP benefits if more channels are allocated to it during the later pattern generation stage than the initial stage. This means that optimal assignment starts with less channel allocation at the early part of pattern generation and gradually increases channels in later phases. The authors analyzed 8 designs for test time and test volume by allocating two sets of test channels, and the results of this analysis are presented in Fig. 22. In the first scheme, the maximum number of channels is allocated to the IP at the start of pattern generation, and they are reduced gradually at regular intervals of coverage gains. In second scheme, a lower number of channels is allocated at the beginning of pattern generation, and they are increased at regular coverage gains. The results show that Scheme II always produces a lower test time compared to Scheme I.

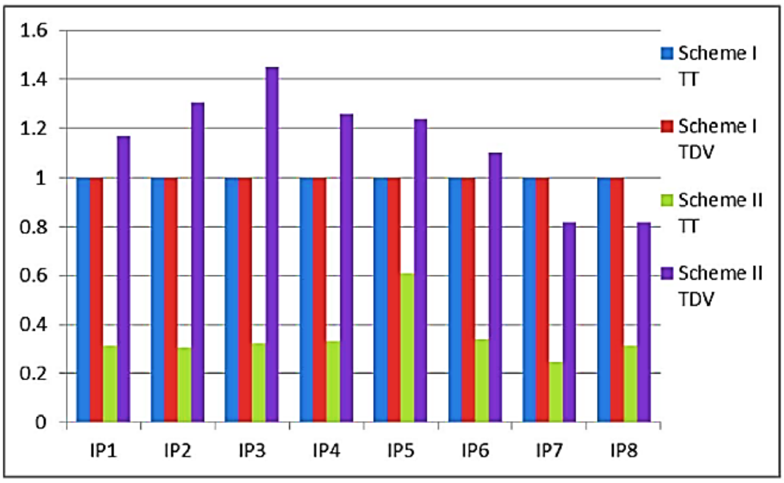

Fig. 22. Normalized TDV & TT for Scheme I, II [94]

The author presented a heuristic that allocates available channels to IPs dynamically. It assigns extra channels to IPs that result in the least pattern count. The IPs share channels during the early stage of coverage ramp. The heuristic allocates a larger number of channels during the later stage of the test schedule to those IPs that produce the biggest reduction in total pattern count. The author used the heuristics to check their effectiveness and applied them to 3 SOCs. Table IV shows the results achieved using the heuristic. It also presents the results of using the other two scheduling methods. In the first schedule, all channels are allocated to each IP, while the second scheme allocates one channel to each IP. Schedule 3 in the table is based on a heuristic proposed by the authors. The table shows that a reduction of up to 93% could be achieved using the proposed heuristic compared to Scheme I.

TABLE IV. COMPARISON OF TEST COST METRICS [94]

	SOC1	**SOC2**	**SOC3**
No. of IPs	8	11	12
No. of Channels	6	6	3
Schedule I - #Patterns	51616	83226	82902
Schedule II - #Patterns	73344	10207	32049
Schedule III - #Patterns	34560	6962	16800

V. FINDINGS AND DISCUSSION

After reviewing the recent advances in the related literature, this section summarizes approaches discussed in previous section. It also discusses the benefits of various TAM design approaches and test scheduling techniques. Then, it highlights the main challenges that remain in key aspects and discusses a direction for future work.

- *Test Access Mechanisms for In-System and System Level Testing*: For safety-critical and high-performance applications, continuous silicon monitoring is required. Companies developed their own solution to perform testing during system operation. Using functional pins is key for these tests, and various pin interfaces, such as PCIe, SerDes, and DDR, have been used. While they enable in-system and system-level tests, they also reduce test time as they operate at very high speed compared to traditional slow-speed GPIOs. Synopsys' HSAT IP is a readily available solution that chip designers can use to implement high-speed chip pins as a test access mechanism.
- *Test Access Mechanisms for 3D and NOC chips*: 3D chips need to be tested at various stages – individual dies and post-packaging. For post-packaging testing, test access needs to consider TSVs and any test pads added to the package. These factors, along with power consumption to operate dies, significantly affect test time and test cost, which may be prohibitive. So, efficient test access mechanisms and scheduling algorithms are needed. Although researchers presented multiple approaches for leveraging additional infrastructure on 3D chips and NOCs, there is still no unified approach or standard, and no EDA companies have provided solutions.

- *Packetized Test Access Mechanisms*: One of the major challenges in designing TAM is physical routing, specifically for tile-based SOCs. The pattern-count differences in the core also lead to inefficiencies in test scheduling. The streaming scan network methodology eliminates these efficiencies by using a packet-based data delivery network. It also generates all control signals locally for the core's scan operation, thereby making them independent of one another. Independence greatly improves data processing, as the core no longer needs to wait for others to finish their data applications. The method uses an IEEE 1687-based IJTAG network to control the entire operation. The control circuits could be generated and integrated in the core using the vendor's EDA tool. At the same time, pattern generation performs tasks such as selecting patterns for each core and assigning them to chip pins. The method and solution are highly efficient, as they involve fewer physical implementation challenges, use an IEEE-standard-based interface, support the generation and integration of control units, and support pattern generation using their own tool. SSN brings the challenge of exposing chip information and poses a security risk. So, chip designers also need to employ a mechanism for stopping attackers from extracting sensitive information from the system-on-chip.
- *Combining Approaches for Bandwidth Management*: Different aspects of test scheduling, such as scan configurations, chip pin assignments to cores considering their multiple instances, physical locations, and control data delivery mechanisms, affect test application time. TAM design also significantly impacts physical routing. The paper analyzes these factors and proposes an approach to selecting them. The author explores a few design decisions to understand their impact on bandwidth. It described the impact of the EDT chain-to-channel ratio, core selection to avoid control data delivery, using the input or output path of a bidirectional pin at all times, and assigning more pins to cores with high data demand. These decisions are quite helpful to reduce bandwidth further.
- *Test Access Mechanisms for Low Power*: One technique to reduce test time is to test multiple cores simultaneously. However, simultaneous toggle activity increases power consumption and may cross the package thermal limit or decrease design performance. Algorithms have been presented that calculate power during test scheduling, accounting for varying chip pins, the number and location of cores, and TSVs for 3D chips. The analysis provides a mapping of test time and power under these configurations. Chip designers can use such methods to optimize the test schedule.
- *Options for Configuring Test Access Mechanisms*: Selecting which chip pin to connect to which core channel requires a programmable mechanism. The author presented several methods for loading data into these programs. It provided details on 3 methods: using the IEEE 1687-based IJTAG protocol, building separate chains for control data, and using pipeline registers. The user could make use of one of these methods based on physical implementation constraints and the test optimization goal.
- *Test Access Mechanisms for Pin-mux and EDT Based Architectures*: While packet-based test access mechanisms are gaining industry acceptance, a majority of SOCs still use pin-mux-based architecture. Over the years, many architectures have been proposed for routing pins to various cores. Several test scheduling algorithms leverage these architectures to minimize test time and test data volume. Compression architectures have been used since their inception, and their features significantly affect test time and test data volume. Many variants of TAMs are proposed, considering various features of the compression logic block, such as EDT. SOC designers need to consider chip pins, the total number of cores, and the number of identical cores for designing TAM. They also need to analyze TAM wiring for physical implementation and various features of the compression solution. They also need to consider power consumption while deciding the test schedule. SOC designers need to perform exhaustive analysis to design an optimized TAM, and this should be done as early in the SOC design cycle as possible.

VI. CONCLUSION

Structural testing using scan plays an important role in today's system-on-chip (SoC) manufacturing testing. Modern packaging techniques, such as 3D and network-on-chip (NOC), are not test-friendly, so newer test access mechanisms need to be developed for them. The rapid development of semiconductor manufacturing technology has enabled the use of many cores, facilitating the addition of multiple functions on a chip. Testing these cores incurs significant test time, so designing an efficient test access mechanism to apply test data and operate these cores is crucial. Furthermore, testing the chips while operating in the system requires using functional pins instead of test pins, so test access needs to be designed with functional interfaces in mind. In-system testing is necessary for many applications, including machine learning, automotive, space, and medical applications. This paper presents a survey of various TAM design and test-scheduling techniques that address these requirements. Their design and physical implementation challenges are discussed. Various test scheduling approaches are also discussed that help to reduce test application time. As semiconductor techniques keep being invented, more and more cores will get added to chips, and newer manufacturing defects keep showing. To overcome these challenges, efficient scan techniques, including innovative test access mechanisms, will be needed. Using a combination of techniques to reduce overall test time is a promising approach that leverages the shortcomings of individual techniques. This paper is closed with a summary of the benefits and challenges of each approach.

REFERENCES

[1] E. J. Marinissen, S. K. Goel and M. Lousberg, "Wrapper design for embedded core test," Proceedings International Test Conference 2000 (IEEE Cat. No.00CH37159), Atlantic City, NJ, USA, 2000, pp. 911-920, doi: 10.1109/TEST.2000.894302.

[2] R. Kashyap, "Silicon lifecycle management (SLM) with in-chip monitoring," *2021 IEEE International Reliability Physics Symposium (IRPS)*, Monterey, CA, USA, 2021, pp. 1-4

[3] P. H. Hochschild, P. Turner, J. C. Mogul, R. Govindaraju, P. Ranganathan, D. E. Culler, and A. Vahdat. "Cores that don't count". In: Proceedings of the Workshop on Hot Topics in Operating Systems. 2021, pp. 9-16

[4] A. Singh, S. Chakravarty, G. Papadimitriou and D. Gizopoulos, "Silent Data Errors: Sources, Detection, and Modeling," 2023 IEEE 41st VLSI Test Symposium (VTS), San Diego, CA, USA, 2023, pp. 1-12

[5] P. Wohl, J. A. Waicukauski, G. A. Maston and J. E. Colburn, "XLBIST: X-Tolerant Logic BIST," 2018 IEEE International Test Conference (ITC), Phoenix, AZ, USA, 2018, pp. 1-9

[6] A. Pandey, B. Tully, A. Samudra, A. Nagarandal, K. Natarajan and R. Singhal, "Novel Technique for Manufacturing & In-system Testing of Large Scale SoC using Functional Protocol Based High-Speed I/O," 2022 IEEE 40th VLSI Test Symposium (VTS), San Diego, CA, USA, 2022, pp. 1-7

[7] K. Kandula, A. Kapatkar and H. B. Reddy, "High Speed IO Access for Test," 2022 IEEE Women in Technology Conference (WINTECHCON), Bangalore, India, 2022, pp. 1-6

[8] M. Yilmaz, P. K. D. Jagannadha, K. Narayanun, S. Sarangi, F. da Silva and J. Sarmiento, "NVIDIA MATHS: Mechanism to Access Test-Data Over High-Speed Links," in IEEE Design & Test, vol. 40, no. 4, pp. 25-33, Aug. 2023

[9] P. K. Datla Jagannadha et al., "Special Session: In-System-Test (IST) Architecture for NVIDIA Drive-AGX Platforms," 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, USA, 2019, pp. 1-8

[10] Y. Zorian, E. J. Marinissen and S. Dey, "Testing embedded-core based system chips," Proceedings International Test Conference 1998 (IEEE Cat. No.98CH36270), Washington, DC, USA, 1998, pp. 130-143

[11] Q. Zhou and K. J. Balakrishnan, "Test Cost Reduction for SoC Using a Combined Approach to Test Data Compression and Test Scheduling," 2007 Design, Automation & Test in Europe Conference & Exhibition, Nice, France, 2007, pp. 1-6.

[12] A. Sanghani, B. Yang, K. Natarajan and C. Liu, "Design and implementation of a time-division multiplexing scan architecture using serializer and deserializer in GPU chips," 29th VLSI Test Symposium, Dana Point, CA, USA, 2011, pp. 219-224

[13] M. Sonawane et al., "Flexible scan interface architecture for complex SoCs," 2016 IEEE 34th VLSI Test Symposium (VTS), Las Vegas, NV, USA, 2016, pp. 1-6

[14] P. Wohl, J. A. Waicukauski, J. E. Colburn and M. Sonawane, "Achieving extreme scan compression for SoC Designs," 2014 International Test Conference, Seattle, WA, USA, 2014, pp. 1-8

[15] "IEEE Standard for Access and Control of Instrumentation Embedded within a Semiconductor Device," in IEEE Std 1687-2014 , vol., no., pp.1-283, 5 Dec. 2014, doi: 10.1109/IEEESTD.2014.6974961

[16] J. Rajski, "Embedded deterministic test," 2008 International Conference on Signals and Electronic Systems, Krakow, Poland, 2008, pp. 4-4, doi: 10.1109/ICSES.2008.4673339.

[17] Yu Huang et al., "Resource allocation and test scheduling for concurrent test of core-based SOC design," Proceedings 10th Asian Test Symposium, Kyoto, Japan, 2001, pp. 265-270, doi: 10.1109/ATS.2001.990293.

[18] "IEEE Standard Testability Method for Embedded Core-based Integrated Circuits," in IEEE Std 1500-2005 , vol., no., pp.1-136, 29 Aug. 2005, doi: 10.1109/IEEESTD.2005.96465.

[19] E. J. Marinissen, Y. Zorian, R. Kapur, T. Taylor and L. Whetsel, "Towards a standard for embedded core test: an example," International Test Conference 1999. Proceedings (IEEE Cat. No.99CH37034), Atlantic City, NJ, USA, 1999, pp. 616-627, doi: 10.1109/TEST.1999.805786.

[20] E. J. Marinissen, R. Arendsen, G. Bos, H. Dingemanse, M. Lousberg and C. Wouters, "A structured and scalable mechanism for test access to embedded reusable cores," Proceedings International Test Conference 1998 (IEEE Cat. No.98CH36270), Washington, DC, USA, 1998, pp. 284-293, doi: 10.1109/TEST.1998.743166.

[21] P. Varma and B. Bhatia, "A structured test re-use methodology for core-based system chips," Proceedings International Test Conference 1998 (IEEE Cat. No.98CH36270), Washington, DC, USA, 1998, pp. 294-302, doi: 10.1109/TEST.1998.743167.

[22] M. Sonawane *et al*., "Flexible scan interface architecture for complex SoCs," *2016 IEEE 34th VLSI Test Symposium (VTS)*, Las Vegas, NV, USA, 2016, pp. 1-6

[23] A. Floridia and M. Grosso, "Test mode selection and data I/O by means of a new Scan-based interface," 2024 39th Conference on Design of Circuits and Integrated Systems (DCIS), Catania, Italy, 2024, pp. 1-6

[24] J. Janicki et al., "EDT channel bandwidth management in SoC designs with pattern-independent test access mechanism," 2011 IEEE International Test Conference, Anaheim, CA, USA, 2011, pp. 1-9, doi: 10.1109/TEST.2011.6139170.

[25] J. Janicki, J. Tyszer, G. Mrugalski and J. Rajski, "Bandwidth-aware test compression logic for SoC designs," 2012 17th IEEE European Test Symposium (ETS), Annecy, France, 2012, pp. 1-6, doi: 10.1109/ETS.2012.6233003.

[26] J. Janicki, M. Kassab, G. Mrugalski, N. Mukherjee, J. Rajski and J. Tyszer, "EDT Bandwidth Management in SoC Designs," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 31, no. 12, pp. 1894-1907, Dec. 2012, doi: 10.1109/TCAD.2012.2205385.

[27] J. Janicki et al., "EDT bandwidth management - Practical scenarios for large SoC designs," 2013 IEEE International Test Conference (ITC), Anaheim, CA, USA, 2013, pp. 1-10, doi: 10.1109/TEST.2013.6651898.

[28] W. -T. Cheng et al., "Scan Test Bandwidth Management for Ultralarge-Scale System-on-Chip Architectures," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 6, pp. 1050-1062, June 2015, doi: 10.1109/TVLSI.2014.2332469.

[29] G. Giles, J. Wang, A. Sehgal, K. J. Balakrishnan and J. Wingfield, "Test Access Mechanism for Multiple Identical Cores," 2008 IEEE International Test Conference, Santa Clara, CA, USA, 2008, pp. 1-10, doi: 10.1109/TEST.2008.4700553.

[30] Qiang Xu and N. Nicolici, "Multi-frequency test access mechanism design for modular SOC testing," 13th Asian Test Symposium, Kenting, Taiwan, 2004, pp. 2-7, doi: 10.1109/ATS.2004.60.

[31] Qiang Xu and N. Nicolici, "Modular SOC testing with reduced wrapper count," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 24, no. 12, pp. 1894-1908, Dec. 2005, doi: 10.1109/TCAD.2005.852447.

[32] Qiang Xu and N. Nicolici, "On reducing wrapper boundary register cells in modular soc testing," International Test Conference, 2003. Proceedings. ITC 2003., Charlotte, NC, USA, 2003, pp. 622-631, doi: 10.1109/TEST.2003.1270889.

[33] Vijay Sontakke, John Dickhoff, “A survey of scan-capture power reduction techniques”, International Journal of Electrical and Computer Engineering (IJECE), 2023.

[34] Vijay Sontakke, John Dickhoff, “Developments in scan shift power reduction: a survey”, Bulletin of Electrical Engineering and Informatics (BEEI), 2023.

[35] V. Iyengar and K. Chakrabarty, "System-on-a-chip test scheduling with precedence relationships, preemption, and power constraints," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 21, no. 9, pp. 1088-1094, Sept. 2002, doi: 10.1109/TCAD.2002.801102.

[36] X. Kavousianos, K. Chakrabarty, A. Jain and R. Parekhji, "Test Schedule Optimization for Multicore SoCs: Handling Dynamic Voltage Scaling and Multiple Voltage Islands," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 31, no. 11, pp. 1754-1766, Nov. 2012, doi: 10.1109/TCAD.2012.2203600.

[37] D. Zhao and S. Upadhyaya, "Dynamically partitioned test scheduling with adaptive TAM configuration for power-constrained SoC testing," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 24, no. 6, pp. 956-965, June 2005, doi: 10.1109/TCAD.2005.847893.

[38] A. B. Kinsman and N. Nicolici, "Time-multiplexed test data decompression architecture for core-based SOCs with improved utilization of tester channels," European Test Symposium (ETS'05), Tallinn, Estonia, 2005, pp. 196-201, doi: 10.1109/ETS.2005.43.

[39] Wei Zou, S. M. Reddy, I. Pomeranz and Yu Huang, "SOC test scheduling using simulated annealing," Proceedings. 21st VLSI Test Symposium, 2003., Napa, CA, USA, 2003, pp. 325-330, doi: 10.1109/VTEST.2003.1197670.

[40] P. T. Gonciari and B. M. Al-Hashimi, "A compression-driven test access mechanism design approach," Proceedings. Ninth IEEE European Test Symposium, 2004. ETS 2004., Corsica, France, 2004, pp. 100-105, doi: 10.1109/ETSYM.2004.1347617.

[41] M. Richter and K. Chakrabarty, "Test pin count reduction for NoC-based Test delivery in multicore SOCs," 2012 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, Germany, 2012, pp. 787-792, doi: 10.1109/DATE.2012.6176601.

[42] K. Mrityunjay, S. Biswas and J. K. Deka, "ATPG for Incomplete Testing of SoC and Power Aware TAM Architecture," 2018 15th IEEE India Council International Conference (INDICON), Coimbatore, India, 2018, pp. 1-6

[43] Vikram lyengar, Krishnendu Chakrabarty and E. J. Marinissen, "Test access mechanism optimization, test scheduling, and tester data volume reduction for system-on-chip," in IEEE Transactions on Computers, vol. 52, no. 12, pp. 1619-1632, Dec. 2003, doi: 10.1109/TC.2003.1252857.

[44] E. Larsson and H. Fujiwara, "System-on-chip test scheduling with reconfigurable core wrappers," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 14, no. 3, pp. 305-309, March 2006, doi: 10.1109/TVLSI.2006.871757.

[45] Yu Huang, S. M. Reddy and Wu-Tung Cheng, "Core-clustering based SoC test scheduling optimization," Proceedings of the 11th Asian Test Symposium, 2002. (ATS '02)., Guam, USA, 2002, pp. 405-410, doi: 10.1109/ATS.2002.1181745.

[46] M. Sharma, A. Dutta, W. -T. Cheng, B. Benware and M. Kassab, "A novel Test Access Mechanism for failure diagnosis of multiple isolated identical cores," 2011 IEEE International Test Conference, Anaheim, CA, USA, 2011, pp. 1-9, doi: 10.1109/TEST.2011.6139171.

[47] H. -M. von Staudt, B. Van Treuren, J. Rearick, M. Portolan and M. Keim, "Exploring and Comparing IEEE P1687.1 and IEEE 1687 Modeling of Non-TAP Interfaces," 2021 IEEE European Test Symposium (ETS), Bruges, Belgium, 2021, pp. 1-10.

[48] E. Larsson, S. K. Gangaraju and P. Murali, "System-Level Access to On-Chip Instruments," 2021 IEEE European Test Symposium (ETS), Bruges, Belgium, 2021, pp. 1-6.

[49] J. Suri, R. Kinger, S. Nimmagadda and H. Fei, "Structural Testing on SLT Platform with HSAT IP & High-Speed I/O Access," 2025 IEEE International Test Conference (ITC), San Diego, CA, USA, 2025, pp. 542-545.

[50] A. Pandey, B. Tully and K. Natarajan, "High Speed IO Access for Test forms the foundation for Silicon Lifecycle Management," 2022 IEEE International Test Conference (ITC), Anaheim, CA, USA, 2022, pp. 656-660

[51] P. Wohl, J. A. Waicukauski, J. E. Colburn and M. Sonawane, "Achieving extreme scan compression for SoC Designs," 2014 International Test Conference, Seattle, WA, USA, 2014, pp. 1-8

[52] PCI Express Base Specification Revision 5.0, 2021

[53] S. Upadhyay and A. Tokuz, "Scan SerDes* for Multi-die Packages," 2024 IEEE International Test Conference (ITC), San Diego, CA, USA, 2024, pp. 333-338

[54] M. -H. Hsu, C. -H. Cheng and S. -H. Huang, "3D IC test scheduling with test pads considered," 2016 5th International Symposium on Next-Generation Electronics (ISNE), Hsinchu, Taiwan, 2016, pp. 1-2

[55] I. Choi, H. Oh and S. Kang, "Test access mechaism for stack test time reduction of 3-dimensional integrated circuit," 2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Jeju, Korea (South), 2016, pp. 522-525

[56] B. Noia, K. Chakrabarty, S. K. Goel, E. J. Marinissen and J. Verbree, "Test-Architecture Optimization and Test Scheduling for TSV-Based 3-D Stacked ICs," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 30, no. 11, pp. 1705-1718, Nov. 2011

[57] R. Karmakar, A. Agarwal and S. Chattopadhyay, "Test Infrastructure Development and Test Scheduling of 3D-Stacked ICs under Resource and Power Constraints," 2015 IEEE 24th Asian Test Symposium (ATS), Mumbai, India, 2015, pp. 73-78

[58] Flottes, Marie-Lise & Azevedo, Joao & Natale, Giorgio Di & Rouzeyre, Bruno. (2015). Session-less based thermal-aware 3D-SIC test scheduling. 1-2. 10.1109/ETS.2015.7138732.

[59] Seo, Sungyoul & Lim, Hyeonchan & Kang, Soyeon. (2017). Off-chip test architecture for improving multi-site testing efficiency using tri-state decoder and 3V-level encoder. 191-195

[60] Kim, Haksong et al. “A Novel Massively Parallel Testing Method Using Multi-Root for High Reliability.” IEEE Transactions on Reliability 64 (2015): 486-496.

[61] Keezer, D.C. & Chen, T.H. & Moon, Thomas & Stonecypher, D.T. & Chatterjee, Abhijit & Choi, H.W. & Kim, S.Y. & Yoo, H.. (2015). An FPGA-based ATE extension module for low-cost multi-GHz memory test. 10.1109/ETS.2015.7138756.

[62] Hong, Chaneui et al. “Network-On-Chip Based Test Access Mechanism for Multi-Site Test.” 2018 International Conference on Computational Science and Computational Intelligence (CSCI) (2018): 1431-1433.

[63] Han, Taewoo & Choi, Inhyuk & Oh, Hyunggyo. (2015). Parallelized Network-on-Chip-Reused Test Access Mechanism for Multiple Identical Cores. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 35. 1-1. 10.1109/TCAD.2015.2481872.

[64] Vartziotis, Fotios & Kavousianos, Xrysovalantis & Chakrabarty, Krishnendu & Jain, Arvind & Parekhji, Rubin. (2015). Time-Division Multiplexing for Testing DVFS-Based SoCs. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on. 34. 668-681. 10.1109/TCAD.2015.2394462.

[65] Ansari, Muhammad Adil. (2018). Network Load Analysis During Test Mode for the Network-on-Chip Reused Test Access Mechanisms. 323-327. 10.1109/iCCECOME.2018.8659014.

[66] Hwang, Shian-Miin & Kao, Cheng-Yan & Horng, Jorng-Tzong. (1994). On solving rectangle bin packing problems using genetic algorithms. 2. 1583;1583 - 1590 vol.2;1590 vol.2.

[67] J. -F. Côté et al., "Streaming Scan Network (SSN): An Efficient Packetized Data Network for Testing of Complex SoCs," 2020 IEEE International Test Conference (ITC), Washington, DC, USA, 2020, pp. 1-10, doi: 10.1109/ITC44778.2020.9325233.

[68] Y. Dong et al., "Maximizing scan pin and bandwidth utilization with a scan routing fabric," 2017 IEEE International Test Conference (ITC), Fort Worth, TX, USA, 2017, pp. 1-10, doi: 10.1109/TEST.2017.8242053.

[69] D. Trock and R. Fisette, "Recursive hierarchical DFT methodology with multi-level clock control and scan pattern retargeting," 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, Germany, 2016, pp. 1128-1131.

[70] "IEEE Standard for High-Speed Test Access Port and On-Chip Distribution Architecture," in IEEE Std 1149.10-2017 , vol., no., pp.1-96, 28 July 2017, doi: 10.1109/IEEESTD.2017.7995164.

[71] J. Remmers, M. Villalba and R. Fisette, "Hierarchical DFT methodology - a case study," 2004 International Conferce on Test, Charlotte, NC, USA, 2004, pp. 847-856, doi: 10.1109/TEST.2004.1387348.

[72] Mangilal, Kunal & Yilmaz, Mahmut & Agarwal, Vishal & Sarangi, Shantanu & Narayanun, Kaushik. (2024). A Scalable & Cost Efficient Next-Gen Scan Architecture: Streaming Scan Test via NVIDIA MATHS. 400-406. 10.1109/ITC51657.2024.00062.

[73] Mardani Kamali, Hadi & Zamiri Azar, Kimia & Homayoun, Houman & Sasan, Avesta. (2020). On Designing Secure and Robust Scan Chain for Protecting Obfuscated Logic. 217-222. 10.1145/3386263.3407655.

[74] Limaye, Nimisha & Kalligeros, Emmanouil & Karousos, Nikolaos & Karybali, Irene & Sinanoglu, Ozgur. (2020). Thwarting All Logic Locking Attacks: Dishonest Oracle With Truly Random Logic Locking. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. PP. 1-1. 10.1109/TCAD.2020.3029133.

[75] Riaz, Anjum & Kumar, Gaurav & Tudu, Jaynarayan & Ahlawat, Satyadev. (2022). On Protecting IJTAG from Data Sniffing and Alteration Attacks. 10.1109/ISVLSI54635.2022.00038.

[76] Elnaggar, Rana & Karri, Ramesh & Chakrabarty, Krishnendu. (2020). Security Against Data-Sniffing and Alteration Attacks in IJTAG. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. PP. 1-1. 10.1109/TCAD.2020.3019167.

[77] Rajski, Janusz & Trawka, Maciej & Tyszer, Jerzy & Wlodarczak, Bartosz. (2022). Hardware Root of Trust for SSN-basedDFT Ecosystems. 479-483. 10.1109/ITC50671.2022.00056.

[78] Rajski, Janusz & Trawka, Maciej & Tyszer, Jerzy & Włodarczak, Bartosz. (2024). A Nonlinear Stream Cipher for Encryption of Test Patterns in Streaming Scan Networks. IEEE Transactions on Circuits and Systems I: Regular Papers. PP. 1-13. 10.1109/TCSI.2024.3447080.

[79] Shukla, Sonali & Kumar, Bhavika & Singh, Virendra. (2023). SSSN: Secured Streaming Scan Network. 1-6. 10.1109/LATS58125.2023.10154483.

[80] Riaz, Anjum & Kumar, Pardeep & Prasad, Yamuna & Ahlawat, Satyadev. (2025). On Securing SSN Architecture using Test Vector Encryption. 10.1109/ISCAS56072.2025.11043163.

[81] Gnawali, Krishna Prasad, Andrea Costa, Nathalie Etono, Denis Martin, Bala Tarun Nelapatla and Amit Purohit. “Test Pattern Aware Streaming Fabric-based Scan Test Methodology.” *2025 IEEE International Test Conference (ITC)* (2025): 438-441.

[82] Chandrasekaran, Gokul, Neelam Sanjeev Kumar, P. Karthikeyan, Kumarasamy Vanchinathan, Neeraj Priyadarshi and Bhekisipho Twala. “Test Scheduling and Test Time Minimization of System-on-Chip Using Modified BAT Algorithm.” *IEEE Access* 10 (2022): 126199-126216.

[83] Chatterjee, Subhajit, Surajit Kumar Roy, Chandan Giri and Hafizur Rahaman. “An Efficient Test Scheduling to Co-optimize Test Time and Peak Power for 3D ICs.” *2020 International Symposium on Devices, Circuits and Systems (ISDCS)* (2020): 1-6.

[84] Vartziotis, Fotis. “TDMS Test Scheduler: An Integrated Framework for Test Scheduling of DVFS-based SoCs with Multiple Voltage Islands.” *2021 IEEE European Test Symposium (ETS)* (2021): 1-2.

[85] Kim, Heetae, Hyojoon Yun, Doohyun Yoon and Sungho Kang. “An Efficient Test Scheduling Method Based on Dynamic Pairing.” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 44 (2025): 3606-3616.

[86] Karmakar, Rajit & Chattopadhyay, Santanu. (2016). Thermal-Safe Schedule Generation for System-on-Chip Testing. 475-480. 10.1109/VLSID.2016.47.

[87] Dutt, Sunil & Chauhan, Anshu & Bhadoriya, Rahul & Nandi, Sukumar & Trivedi, Gaurav. (2015). A High-Performance Energy-Efficient Hybrid Redundant MAC for Error-Resilient Applications. Proceedings of the IEEE International Conference on VLSI Design. 2015. 10.1109/VLSID.2015.65.

[88] Sadi, Mehdi & Guin, Ujjwal. (2021). Test and Yield Loss Reduction of AI and Deep Learning Accelerators. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. PP. 1-1. 10.1109/TCAD.2021.3051841.

[89] Han, Taewoo & Choi, Inhyuk. (2015). Majority-Based Test Access Mechanism for Parallel Testing of Multiple Identical Cores. Very Large Scale Integration (VLSI) Systems, IEEE Transactions on. 23. 1439-1447. 10.1109/TVLSI.2014.2341674.

[90] Lee, Sangjun & Park, Jongho & Lee, Inhwan & Lee, Kwonhyoung. (2021). Hybrid Test Access Mechanism for Multiple Identical Cores. 365-366. 10.1109/ISOCC53507.2021.9613908.

[91] Richter, Michael & Chakrabarty, Krishnendu. (2014). Optimization of Test Pin-Count, Test Scheduling, and Test Access for NoC-Based Multicore SoCs. Computers, IEEE Transactions on. 63. 691-702. 10.1109/TC.2013.82.

[92] H. Parmar and U. Mehta, "An improved algorithm for TAM optimization to reduce test application time in core based SoC," 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Dhaka, Bangladesh, 2015, pp. 443-446, doi: 10.1109/WIECON-ECE.2015.7443963.

[93] H. Ke, D. Zhongliang and H. Jianming, "Boundary scan with parallel test access mechanism," 2009 9th International Conference on Electronic Measurement & Instruments, Beijing, China, 2009, pp. 4-70-4-73, doi: 10.1109/ICEMI.2009.5274683.

[94] J. Abraham, S. Umapathi and S. Krishnamurthi, "Test Time Minimisation in Scan Compression Designs Using Dynamic Channel Allocation," 2016 29th International Conference on VLSI Design and 2016 15th International Conference on Embedded Systems (VLSID), Kolkata, India, 2016, pp. 593-594, doi: 10.1109/VLSID.2016.99.

[95] G. Li, H. Zhao, Q. Yang, J. Qian and Y. Huang, "Industrial Case Studies of SoC Test Scheduling Optimization by Selecting Appropriate EDT Architectures," 2018 IEEE International Test Conference in Asia (ITC-Asia), Harbin, China, 2018, pp. 109-114, doi: 10.1109/ITC-Asia.2018.00029

[96] Y. Huang et al., "Test Compression Improvement with EDT Channel Sharing in SoC Designs," 2014 IEEE 23rd North Atlantic Test Workshop, Johnson City, NY, USA, 2014, pp. 22-31, doi: 10.1109/NATW.2014.14.

[97] G. Li et al., "Hybrid Hierarchical and Modular Tests for SoC Designs," 2015 IEEE 24th North Atlantic Test Workshop, Johnson City, NY, USA, 2015, pp. 11-16, doi: 10.1109/NATW.2015.9

[98] X. Liu, C. Yu, Y. Qi, Y. Huang and J. Fu, "Case Study of Testing a SoC Design with Mixed EDT Channel Sharing and Channel Broadcasting," 2016 IEEE 25th North Atlantic Test Workshop (NATW), Providence, RI, USA, 2016, pp. 12-17, doi: 10.1109/NATW.2016.10

[99] Y. Huang, "EDT dynamic bandwidth management (DBM) in SoC testing," 2016 29th IEEE International System-on-Chip Conference (SOCC), Seattle, WA, USA, 2016, pp. 58-63, doi: 10.1109/SOCC.2016.7905435

[100] C. Barnhart, V. Brunkhorst, F. Distler, O. Farnsworth, B. Keller and B. Koenemann, "OPMISR: the foundation for compressed ATPG vectors," Proceedings International Test Conference 2001 (Cat. No.01CH37260), Baltimore, MD, USA, 2001, pp. 748-757, doi: 10.1109/TEST.2001.966696.

[101] Y. Dong et al., "Toward more efficient scan data bandwidth utilization on modern SOCs," 2016 29th IEEE International System-on-Chip Conference (SOCC), Seattle, WA, USA, 2016, pp. 64-68, doi: 10.1109/SOCC.2016.7905436.

[102] J. Janicki, M. Kassab, G. Mrugalski, N. Mukherjee, J. Rajski and J. Tyszer, "Test Time Reduction in EDT Bandwidth Management for SoC Designs," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 32, no. 11, pp. 1776-1786, Nov. 2013, doi: 10.1109/TCAD.2013.2263038.

[103] J. Janicki, M. Kassab, G. Mrugalski, N. Mukherjee, J. Rajski and J. Tyszer, "Erratum to “Test Time Reduction in EDT Bandwidth Management for SoC Designs” [Nov 13 1776-1786]," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 33, no. 1, pp. 167-167, Jan. 2014, doi: 10.1109/TCAD.2013.2292631.