Test Bus Sizing for System-on-a-Chip

Vikram Iyengar, *Student Member*, *IEEE*, and
Krishnendu Chakrabarty, *Senior Member*, *IEEE*

Abstract—System-on-a-chip (SOC) designs present a number of unique testability challenges to system integrators. Test access to embedded cores often requires dedicated test access mechanisms (TAMs). We present an improved approach for designing efficient TAMs and investigate the problems of improved deserialization of test data in the core wrapper, optimal test bus sizing, and optimal assignment of cores to test buses in the system. Place-and-route and power constraints are incorporated in the model. This work represents an important first step towards combining TAM design with efficient wrapper design for test data deserialization. Experimental results demonstrate that the proposed TAM optimization methodology provides efficient test bus designs for minimizing the testing time.

Index Terms—Core-based systems, embedded core testing, integer linear programming, linearization, test access mechanism (TAM), test bus, testing time.

1 Introduction

VLSI technology has now evolved to the point where million-gate integrated circuits (ICs) are common. As a result, large system-on-a-chip (SOC) ICs are being built from complex intellectual property (IP) blocks known as embedded cores. While embedded core-based design has become popular as a means for providing rich functionality with short design cycle times, testing core-based systems is difficult. Core vendors often conceal structural information about IP cores and provide only a set of precomputed test patterns with a specified fault coverage. Thus, system integrators cannot insert design-for-testability (DFT) hardware to ease test generation. Moreover, embedded cores are not always directly accessible from chip I/Os. Dedicated test access mechanisms (TAMs) are necessary for propagating test stimuli to core inputs and propagating test responses to chip outputs, as illustrated in Fig. 1. Designing efficient TAMs, especially under testing time, area, and power constraints is therefore recognized as a major problem facing core test integration [3], [4], [10].

A number of test access strategies and TAM designs have been proposed in the literature. These include multiplexed access [8], partial isolation rings [12], core transparency [6], dedicated test bus [13], reuse of the existing system bus [7], and a scalable bus-based architecture called TESTRAIL [10]. Bus-based TAMs, being flexible and scalable, appear to be the most promising. However, their design has largely been ad hoc and previous methods have seldom addressed the problem of minimizing testing time under place-and-route and power constraints. While [1] presents several novel TAM architectures (e.g., multiplexing, daisy chaining, and distribution), it does not directly address the problem of optimal sizing of test buses in the SOC. In particular, only internal scan chains are considered in [1], while wrappers and functional I/Os are ignored.

TAM design is of critical importance in SOC system integration since it directly impacts testing time and, thereby, affects test cost. The 1999 International Technology Roadmap for Semiconductors (ITRS) [9] clearly identifies test access for SOC cores as one of the key challenges for the near future. The TAM must also adhere to place-and-route constraints arising from functional interconnections between the cores. In addition, an effective TAM should incorporate system-level power constraints.

The relationship between testing time and test bus widths was recently examined in [3]. Several integer linear programming (ILP) models were developed to solve the design problems of assigning cores on the SOC to specific test buses and determining optimal widths of the test buses to minimize testing time. However, the problem of efficient deserialization of test data at core I/Os, which is required when the TAM width is less than the number of core I/Os, was not addressed in [3]. The deserialization model in [3] provided direct (parallel) access from the test bus to core I/Os (such as scan chains) that transport more test data, while grouping the remaining I/Os on a single test bus line. This led to an overestimation of the testing time and the width required for the test buses. While test bus design under the constraints of power and routing area was studied in [4], the deserialization model used for TAM design was based on [3]. Improved deserialization models are therefore necessary to generalize the designs in [3], [4] to other scenarios. These deserialization methods also require new ILP models for TAM optimization.

We present a new approach to the design of efficient bus-based TAMs for SOCs using improved test data deserialization at core I/Os. We develop improved ILP models for designing optimal TAMs that minimize the testing time. We extend the ILP models to incorporate the effects of routing and power constraints with the new deserialization strategy. The new approach delivers substantial reductions in testing time as compared to [3], [4] and, also, reduces the test bus width required to achieve mandated test

• *The authors are with the Department of Electrical and Computer Engineering, Duke University, 30 Hudson Hall, Box 90291, Durham, NC 27708. E-mail: {vik, krish}@ee.duke.edu.*

Manuscript received 28 Dec. 2000; revised 26 June 2001; accepted 18 Sept. 2001.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 113374.

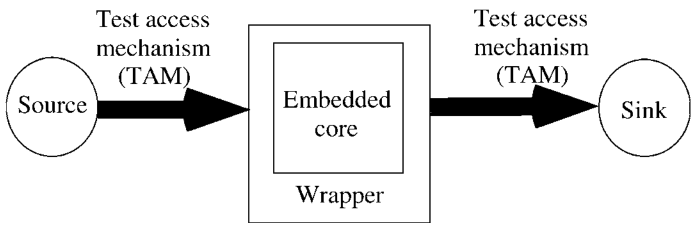

Fig. 1. Embedded core test architecture: test pattern source and sink, test access mechanism, and test wrapper.

application times. Even though our ILP models are aimed at obtaining optimal test bus widths, the search process may be halted to obtain intermediate, nearly optimal solutions within a few seconds of CPU time. This feature is available in most ILP solvers.

Our ILP models do not address the problem of testing the interconnect and wiring between cores. A complete solution for SOC testing that also addresses these issues requires additions to the basic formulation that is presented here for isolation testing of embedded cores. The proposed model also does not incorporate the optimal wrapper design proposed recently [11]. Several enhancements to the ILP model presented here are necessary for optimal TAM/wrapper codesign; nevertheless, this work represents an important first step towards that direction.

The proposed optimization approach is not limited to external testing using TAMs. The TAMs described in this paper can also be used to transport test patterns from on-chip sources to cores under test. If the same BIST engine is used in multiple cores, then the proposed techniques can be used to size these TAMs. The problems of TAM sizing is however of less relevance for BIST-ed cores that are tested entirely using BIST patterns. Nevertheless, BIST is seldom used as the only test mechanism for cores in an SOC. BIST tests are often augmented with external tests for hard-to-detect faults, which can be efficiently applied using optimized TAMs.

The organization of the paper is as follows: In Section 2, we describe the improved approach to test data serialization at core I/Os and the new ILP model for TAM design. In Section 3, we develop ILP models for minimizing the testing time by determining optimal test widths and optimal assignments of cores to test buses. Special linearization techniques are used to remove nonlinearity from the mathematical programming models. We solve the various ILP models using the *lpsolve* software package from Eindhoven University of Technology [2]. In Section 4, we examine the design of test buses that fork out into narrower parts, which can test several smaller cores in parallel to save testing time. Finally, in Section 5, we extend our basic ILP models for TAM design under power and routing constraints.

In order to illustrate the proposed optimization methods and to demonstrate their effectiveness, we use core-based SOC $\mathcal{S}$ (described as system $\mathcal{S}_2$ in [3]) as a running example throughout the text—see Fig. 2. Companies are reluctant to disclose details of real circuits for publication in the open literature [5]; hence, we are using $\mathcal{S}$ as a representative example. It contains cores of various sizes and I/O widths. This hypothetical but nontrivial SOC consists of two combinational ISCAS 85 and eight sequential ISCAS 89 benchmark circuits. We assume that the ISCAS 89 circuits contain internal scan chains. The proposed test bus optimization methodology is also applicable to SOCs containing nonscanned sequential cores since these cores can be treated as combinational (having no scan test data) for the purpose of testing time calculation. The complexity of the ILP models developed here depends more on the number of cores in the SOC than on the sizes of the cores. While only two test buses are shown in Fig. 2, our ILP models can be easily used for any number of test buses.

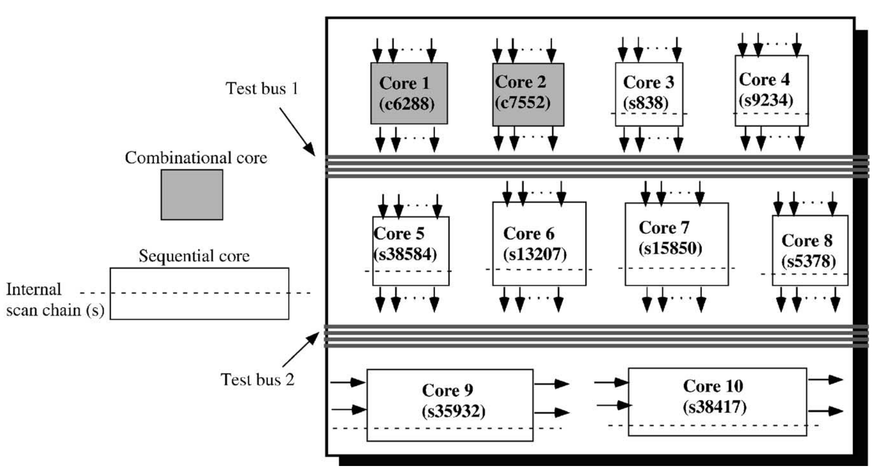

Fig. 2. Example core-based SOC $\mathcal{S}$ with two test buses, containing two combinational cores and eight sequential cores with internal scan.

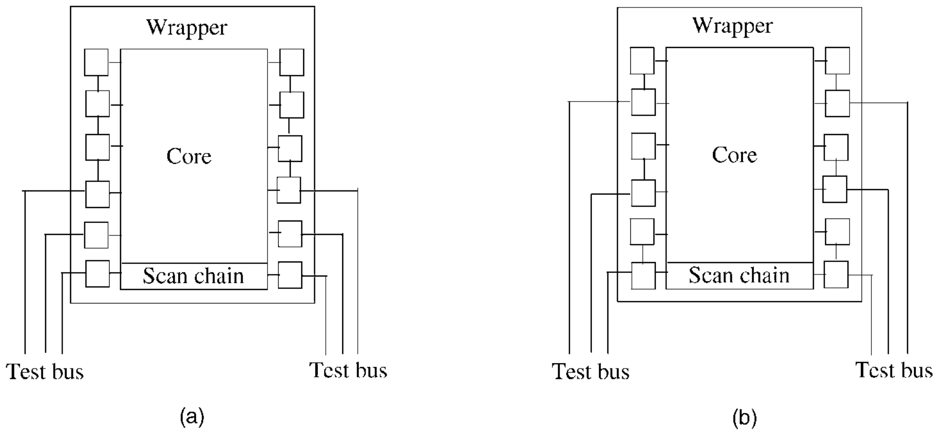

Fig. 3. Test data deserialization: (a) "worst case" approach from [3] and (b) the proposed uniform distribution.

2 Improved Test Data Deserialization

In this section, we present the improved deserialization model and the ILP formulation based on this model. First, we briefly describe ILP. The goal of ILP is to minimize a linear objective function on a set of integer variables, while satisfying a set of linear constraints. A typical ILP model can described as follows:

$$\begin{aligned} \text{minimize}: &\quad \mathbf{Ax} \\ \text{subject to}: &\quad \mathbf{Bx} \leq \mathbf{C}, \mathbf{x} \geq 0, \end{aligned}$$

where $\mathbf{A}$ is a cost vector, $\mathbf{B}$ is a constraint matrix, $\mathbf{C}$ is a column vector of constants, and $\mathbf{x}$ is a vector of integer variables. Efficient ILP solvers are now readily available, both commercially and in the public domain.

Test data deserialization is required at core I/Os when the width of the TAM is less than the number of core terminals [3]. This happens often because the number of core terminals is determined by the functionality of the core, while the TAM width is determined by the test pattern source, the SOC routing and area constraints, and in some cases, the width of the existing system bus. Therefore, test data deserialization must be performed in the test wrapper if the number of core terminals is larger than the TAM width.

Consider an SOC consisting of N cores and let Core i, $1 \leq i \leq N$, have n_i inputs and m_i outputs (including data and scan I/Os). Let the SOC have B test buses with widths $w_1, w_2, \ldots, w_B$. If Core i is assigned to Bus j, the amount of test data serialization at the I/Os of Core i is related to the difference between Core i's test width ϕ_i and the width w_j of Bus j, where $\phi_i = \max\{n_i, m_i\}$. Let t_i be the testing time in cycles required by Core i when no deserialization is necessary. For combinational cores, t_i is equal to the number of test patterns p_i. However, for cores with internal scan, $t_i = (p_i + 1)\lceil f_i/s_i \rceil + p_i$, where Core i contains f_i flip-flops and s_i internal scan chains. Here, we assume that the core scan chains are balanced, which is an important scan chain design consideration to reduce the time taken to scan-in test patterns to the core. In [3], the testing time (with test data deserialization) for Core i assigned to Bus j was given as

$$T_{ij} = \begin{cases} t_i, & \text{if } \phi_i \leq w_j \\ (\phi_i - w_j + 1)t_i, & \text{if } \phi_i > w_j. \end{cases}$$

However, this assumes a "worst case" deserialization of test data, in which the first $(w_j - 1)$ test bus lines are connected to $(w_j - 1)$ core I/Os in parallel and the last test bus line is serially connected to the remaining $(\phi_i - w_j + 1)$ core I/Os; see Fig. 3a. This model is motivated by the need to provide parallel access to core terminals such as scan chain inputs that transport more test data. However, substantial reduction in testing time can be achieved in general if there is a more uniform distribution of test bus lines among the core I/Os as in Fig. 3b. In the improved test data deserialization scenario,

$$T_{ij} = \left\lceil \frac{\phi_i}{w_j} \right\rceil t_i.$$

As a special case, if $\phi_i \leq w_j$, we have $T_{ij} = t_i$.

We assume in our deserialization model that the test sets for the SOC cores are available in scan format, in which the functional input values remain unchanged (i.e., they are don't cares) during successive scan cycles. This implies that the scan input values are specified in multiple cycles, instead of being specified as independent values in the same cycle. A similar assumption was implicitly made in [3], [4]. If the test sets are obtained before the translation to scan format, then alternative deserialization models involving the lengths of the scan chains can be used to reduce the testing time even further. Such a model was presented in [11] and work is currently underway for wrapper/TAM codesign that incorporates such wrapper designs.

We next introduce binary variables x_{ij} (where $1 \leq i \leq N$ and $1 \leq j \leq B$), which are used to determine the assignment of cores to test buses in the SOC. Let x_{ij} be a 0-1 variable defined as follows:

$$x_{ij} = \begin{cases} 1, & \text{if Core } i \text{ is assigned to Bus } j \\ 0, & \text{otherwise} \end{cases}.$$

The time needed to test all cores on Bus j is given by

$$\sum_{i=1}^{N} T_{ij} x_{ij}.$$

Since all the test buses can be used simultaneously for testing, the system testing time equals

TABLE 1
Test Data for the Cores in $\mathcal{S}$ [3]

Circuit (core)	Number of test inputs n_i	Number of test outputs m	$\phi_i = \max\{n_i, m_i\}$	Number of flip-flops f_i	Number of scan chains	Number of test cycles t_i	Number of test patterns p_i
c6288	32	32	32	–	–	12	12
c7552	207	108	207			73	73
s838	36	3	36	32	1	2507	75
s9234	40	43	43	211	4	5723	105
s38584	70	336	336	1426	32	5105	110
s13207	78	168	168	638	16	9634	234
s15850	93	166	166	534	16	3359	95
s5378	39	53	53	179	4	4507	97
s35932	67	352	352	1728	32	714	12
s38417	60	138	138	1636	32	3656	68

$$\max_j \sum_{i=1}^{N} T_{ij} x_{ij}.$$

The problem of minimizing the system testing time when test bus widths are known can be formally stated as follows:

- $\mathcal{P}_{assign}$: Given N cores, B test buses of test widths $w_1, w_2, \ldots, w_B$, determine an assignment of cores to test buses such that the total testing time is minimized.

A mathematical programming model for this problem can be formulated as:

Objective: Minimize cost function

$$\mathcal{T} = \max_j \sum_{i=1}^{N} T_{ij} x_{ij}$$

subject to

1. $\sum_{j=1}^{B} x_{ij} = 1$, $1 \le i \le N$.
2. $x_{ij} = 0$ or 1.

The above minmax cost function can be easily linearized to obtain the following ILP model.

Objective: Minimize $\mathcal{T}$, such that:

1. $\mathcal{T} \ge \sum_{i=1}^{N} T_{ij} x_{ij}, 1 \le j \le B$.
2. $\sum_{j=1}^{B} x_{ij} = 1$, $1 \le i \le N$.

We solved this simple ILP model to determine optimal assignments of cores to test buses for the SOC introduced in Section 1. The number of variables and constraints for this model (a measure of the complexity of the problem) is given by NB and $N + B = O(N)$, respectively. Table 1 presents the test data for each core in the system. The ILP model was solved using the *lpsolve* package [2] on a Sun Ultra 10 workstation with a 333 MHz processor and 256 MB of memory. The user time was only 0.06 seconds. We solved the ILP model for several bus width values; the value of B was 2 in all cases. The optimal assignment of cores to the two test buses is given in Table 2. Note that the testing times shown are not the minimum testing times that can actually be achieved with the total test widths given. For example, for $w_1 + w_2 = 48$, a testing time of 11,1140 cycles can be achieved using $w_1 = 26$, $w_2 = 22$, and the test bus assignment vector (2,2,2,1,1,2,1,2,2,1), where a 1 (2) in position i of the vector indicates that Core i is assigned to Bus 1 (2). In the next section, we will examine the problem of determining an optimal distribution of the total test data width among the individual test buses.

The following theorem presents a lower bound on the total testing time when the widths of the test buses are known. This lower bound can indeed be achieved in practice—we illustrate this below using the system $\mathcal{S}$ as an example. We also make use of this theorem in Section 3 to derive a lower bound on the testing time when only the total test data width is known and the optimal widths of the test buses have to be determined.

Theorem 1. *For an SOC with N cores and B test buses with widths $w_1, w_2, \ldots w_B$, respectively, a lower bound on the total testing time $\mathcal{T}$ is given by*

$$\mathcal{T} \ge \max_i \left\{ \left\lceil \frac{\phi_i}{\max_j\{w_j\}} \right\rceil t_i \right\}.$$

Proof. The testing time for Core i depends on the width of the test bus to which it is assigned. Clearly, the testing time for Core i is at least

$$\min_j \left\{ \left\lceil \frac{\phi_i}{w_j} \right\rceil t_i \right\}.$$

TABLE 2
Optimal Assignment of Cores to Test Buses of Predetermined Widths

Total test width $w_1 + w_2$	Distribution (w_1, w_2)	Optimal test bus assignment	Optimal testing time
24	(16,8)	(2,2,2,2,1,1,2,2,1,2)	219253
32	(16,16)	(2,2,2,2,2,1,1,2,2,1)	166604
48	(32,16)	(2,1,2,2,1,1,2,2,1,2)	112690
64	(32,32)	(2,1,2,1,1,2,1,2,2,2)	88344

Since the overall system testing time is constrained by the core that has the longest test time,

$$\mathcal{T} \geq \max_i \left\{ \left\lceil \frac{\phi_i}{\max_j\{w_j\}} \right\rceil t_i \right\}.$$

Intuitively, this value is the time needed to test the core that has the longest testing time when assigned to the widest bus. □

For the example SOC with two test buses of 32 bits and 16 bits, respectively, Theorem 1 provides a lower bound on the testing time of 56155 cycles. This corresponds to a test bus assignment in which only Core 5 is assigned to test Bus 1. Such an assignment is indeed optimal and the lower bound of Theorem 1 is achieved if either the width of the second test bus is increased or if a third test bus is used.

3 Optimal Test Bus Sizing

In this section, we address the problems of determining optimal widths of test buses and optimal assignments of cores to test buses. This is a generalization of the core assignment problem in Section 2. We assume that the total system test bus width can be at most W and that the width of a test bus does not exceed a designer-specified value w_{max}. It is desirable to specify such a value in practice, since it is not feasible to have arbitrarily large test buses on chip. TAMs that use existing system buses are limited by the width of the system bus, which is often 32 or 64. Moreover, limiting the size of W makes the ILP model tractable. For our experiments, we have chosen $w_{max} = 32$. The problem of minimizing testing time by optimal allocation of width among the test buses can be stated as follows:

- $\mathcal{P}_{mintime}$: Given N cores and B test buses of total width W, determine the optimal width of the test buses, and an assignment of cores to test buses such that the total testing time is minimized.

This problem can be shown to be NP-hard using the techniques presented in [3]. However, it can be solved exactly for nontrivial, realistic SOCs using an ILP formulation. A mathematical programming model for $\mathcal{P}_{mintime}$ is shown below.

Objective: Minimize cost function

$$\mathcal{T} = \max_j \left\{ \sum_{i=1}^{N} \left\lceil \frac{\phi_i}{w_j} \right\rceil t_i x_{ij} \right\}$$

subject to:

1. $\sum_{j=1}^{B} x_{ij} = 1,\ 1 \leq i \leq N$ /* Core i is connected to exactly one test bus */.
2. $\sum_{j=1}^{B} w_j = W,\ 1 \leq j \leq B$ /* The sum of the widths of all test buses is W */.
3. $w_j \leq w_{max},\ 1 \leq j \leq B$ /* The maximum of any test bus is w_{max} */.
4. $x_{ij} = 0$ or 1, $1 \leq i \leq N$, $1 \leq j \leq B$ /* Core i is either assigned to Bus j or is not assigned to Bus j */.

The nonlinear term $\frac{\phi_i}{w_j}$ in the cost function can be replaced by $\phi_i v_j$, where $v_j = \frac{1}{w_j}$, by adding new binary indicator variables δ_{jk} (where $1 \leq j \leq B, 1 \leq k \leq w_{max}$) to the mathematical programming model, such that:

$$\delta_{jk} = \begin{cases} 1, & \text{if Bus } j \text{ is } k \text{ bits wide} \\ 0, & \text{otherwise.} \end{cases}$$

In addition, the following constraints are included in the model:

1. $w_j = \sum_{k=1}^{w_{max}} k\delta_{jk}, \quad 1 \leq j \leq B.$
2. $v_j = \sum_{k=1}^{w_{max}} \frac{1}{k}\delta_{jk}, \quad 1 \leq j \leq B.$
3. $\sum_{k=1}^{w_{max}} \delta_{jk} = 1, \quad 1 \leq j \leq B.$

Intuitively, for every Bus j there is exactly one value of k for which δ_{jk} equals 1; therefore, the new indicator variable values determine the width w_j of each test bus.

The ceiling operator in the cost function is another source of nonlinearity. However, it can also be easily linearized. Since $\phi_i v_j \leq \lceil \phi_i v_j \rceil < \phi_i v_j + 1$, the nonlinear term $\lceil \phi_i v_j \rceil$ can be replaced by the new integer variable $\psi_{ij} = \lceil \phi_i v_j \rceil$, with the added constraint: $\psi_{ij} \geq v_j \phi_i, 1 \leq i \leq N, 1 \leq j \leq B$. Finally, the nonlinear term $\psi_{ij} x_{ij}$ in the cost function can be replaced by $y_{ij} = \psi_{ij} x_{ij}$, with three additional constraints as given in [3]. The following theorem demonstrates the equivalence between the nonlinear term $\psi_{ij} x_{ij}$ and the new variable y_{ij}, when the three new constraints from [3] are added to the mathematical programming model.

Theorem 2. *The nonlinear expression $\psi_{ij} x_{ij}$ can be linearized by replacing it with the variable y_{ij} and the following three constraints:*

1. $y_{ij} - \max_i\{\phi_i\} x_{ij} \leq 0, \quad 1 \leq i \leq N, \quad 1 \leq j \leq B,$
2. $y_{ij} - \psi_{ij} \leq 0, \quad 1 \leq i \leq N, \quad 1 \leq j \leq B,$
3. $\psi_{ij} + \max_i\{\phi_i\} x_{ij} - y_{ij} \leq \max_i\{\phi_i\}, 1 \leq i \leq N, 1 \leq j \leq B,$

where $\psi_{ij} \geq 0$ and $y_{ij} \geq 0$ are integer variables and x_{ij} is a binary variable.

Proof. To establish equivalence between y_{ij} and $\psi_{ij} x_{ij}$, we need to show that:

$$y_{ij} = \begin{cases} 0, & \text{if } x_{ij} = 0 \\ \psi_{ij}, & \text{if } x_{ij} = 1 \end{cases}$$

Consider first the case when $x_{ij} = 0$. From constraint 1, we have $y_{ij} \leq 0$; therefore, y_{ij} must equal 0. When $x_{ij} = 1$, from constraints 2 and 3, we have $y_{ij} \leq \psi_{ij}$ and $\psi_{ij} \leq y_{ij}$; therefore, $y_{ij} = \psi_{ij}$, which completes the proof. □

The new variables and constraints yield the following ILP formulation:

Objective: Minimize cost function $\mathcal{T} = \max_j \sum_{i=1}^{N} y_{ij} t_i$, subject to:

1. $y_{ij} - \max_i\{\phi_i\} x_{ij} \leq 0, 1 \leq i \leq N, 1 \leq j \leq B$.
2. $y_{ij} - \psi_{ij} \leq 0,\ 1 \leq i \leq N, 1 \leq j \leq B$.
3. $\psi_{ij} + \max_i\{\phi_i\} x_{ij} - y_{ij} \leq \max_i\{\phi_i\}, 1 \leq i \leq N, 1 \leq j \leq B$.
4. $\psi_{ij} \geq v_j \phi_i, \quad 1 \leq i \leq N, 1 \leq j \leq B$.
5. $w_j = \sum_{k=1}^{w_{max}} k\delta_{jk}, \quad 1 \leq j \leq B$.
6. $v_j = \sum_{k=1}^{w_{max}} (1/k)\delta_{jk}, \quad 1 \leq j \leq B$.
7. $\sum_{k=1}^{w_{max}} \delta_{jk} = 1, \quad 1 \leq j \leq B$.
8. $\sum_{j=1}^{B} x_{ij} = 1, 1 \leq i \leq N$.
9. $\sum_{j=1}^{B} w_j = W, \quad 1 \leq j \leq B$.
10. $w_j \leq w_{max}, \quad 1 \leq i \leq N, \quad 1 \leq j \leq B$.

The ILP model for $\mathcal{P}_{mintime}$ can be used to determine the effect of increased total test bus width on the testing time. However, there is a limit to which the testing time can be decreased by simply increasing bus widths. This limit is presented in the following theorem and can be used to determine the maximum test width beyond which testing time cannot be decreased.

Theorem 3. *For a core-based system with N cores, a lower bound on the total testing time $\mathcal{T}$ is given by*

$$\mathcal{T} \geq \max_i \left\{ \left\lceil \frac{\phi_i}{w_{max}} \right\rceil t_i \right\}.$$

Proof. Let the system consist of B test buses with (undetermined) test widths $w_1, w_2, \ldots w_B$ such that $\max_j\{w_j\} \leq w_{max}$. We know from Theorem 1 that,

$$\mathcal{T} \geq \max_i \left\{ \left\lceil \frac{\phi_i}{\max_j\{w_j\}} \right\rceil t_i \right\}.$$

Since $\max_j\{w_j\} \leq w_{max}$, we have

$$\mathcal{T} \geq \max_i \left\{ \left\lceil \frac{\phi_i}{\max_j\{w_j\}} \right\rceil t_i \right\}.$$

This completes the proof of the theorem. □

Closely related to $\mathcal{P}_{mintime}$ is the problem $\mathcal{P}_{minwidth}$ of determining the minimum test width W needed to satisfy testing time constraints:

- $\mathcal{P}_{minwidth}$: Given N cores, B test buses, and a maximum testing time specification $\mathcal{T}$, determine the optimal width of the test buses and an assignment of cores to test buses such that the testing time is less than $\mathcal{T}$.

This problem has been shown to be NP-hard as well [3]. However, an ILP formulation for $\mathcal{P}_{minwidth}$ can be derived from the ILP model for $\mathcal{P}_{mintime}$ simply by replacing the earlier cost function with:
Objective: Minimize cost function $W = \sum_{j=1}^{B} w_j$ with the additional constraint:

$$\sum_{i=1}^{N} y_{ij} t_i \leq \mathcal{T}, 1 \leq i \leq N, 1 \leq j \leq B.$$

We solved the ILP models for problems $\mathcal{P}_{mintime}$ and $\mathcal{P}_{minwidth}$ for several values of W and $\mathcal{T}$, keeping $B = 2$. The number of variables and constraints for these ILP models is given by $B(3N + w_{max} + 2) = O(NB)$ (since w_{max} is a constant) and $B(5N + 5) + N = O(NB)$, respectively. The user time was approximately 24 seconds in all cases. In Table 3, we present the values of test bus width distribution and testing time obtained using the new deserialization strategy and improved ILP model for $\mathcal{P}_{mintime}$. The testing time using uniform deserialization is at least an order of magnitude less than the time required in [3] for all cases. This is expected since an inefficient deserialization model was used in [3]. Table 4 shows the optimal test bus width and width distribution for a given maximum testing time. The test width required to achieve a specified testing time increases gradually at first, as the testing time decreases. However, when the individual bus widths approach $w_{max} = 32$, there is a sharper rise in test width, since the w_{max} constraint limits the number of available choices for bus width optimization.

4 Test Bus Subdivision

In this section, we investigate the effects of subdividing individual test buses into several parts that can test a set of cores in parallel. This is especially useful in reducing testing time when several small cores with relatively smaller test widths are assigned to a wide test bus. In this case, the test bus can fork out into a set of narrower test buses that transport test data to the smaller cores in parallel. Other large cores with a large test width may be assigned to the undivided part of the test bus as before, thus, experiencing no change in their testing time.

We first examine a special case of the test bus subdivion problem, in which the individual undivided test bus widths are known, before considering the more general case of the problem, in which the test bus width w_j for each Bus j must be calculated. This special case of the problem is stated as follows:

- $\mathcal{P}_{fork}$: Given N cores and B test buses of predetermined widths $w_1, w_2, \ldots, w_B$, respectively, and an upper limit j_{max} on the number of subdivisions allowed for test Bus j, determine 1) an optimal subdivision of the test buses and 2) an assignment of cores to test buses such that the total testing time is minimized.

Problem $\mathcal{P}_{fork}$ may easily be shown as NP-hard by restriction to a special case (Problem $\mathcal{P}_{assign}$).

Let x_{ij} be the 0-1 core-to-bus assignment variable as before. We now introduce a new 0-1 assignment variable z_{ijl}

TABLE 3
Optimum Testing Time and Optimal Width Distribution with Two Best Buses and a Given System Test Width for the Example SOC

Total test width W	Optimal distribution (w_1, w_2)	Optimal testing time	Testing time in [3]
32	(16,16)	166604	2202286
36	(20,16)	147627	2174501
40	(24,16)	134060	2149720
44	(28,16)	120542	2123437
48	(28,20)	11140	2099390
52	(24,28)	103458	2086542
56	(28,28)	96544	2069738
60	(28,32)	91287	2044346
64	(32,32)	88332	2029753

that represents core assignments to subdivided test buses. The 0-1 variable z_{ijl} is defined as follows:

$$z_{ijl} = \begin{cases} 1, & \text{if Core } i \text{ is assigned to the } l\text{th subdivision of test Bus } j \\ 0, & \text{otherwise} \end{cases}$$

Let the j_{max} possible subdivisions of test Bus j have widths $w_{j1}, w_{j2}, \ldots, w_{jj_{max}}$, respectively. Then,

$$\sum_{l=1}^{j_{max}} w_{jl} = w_j, 1 \leq j \leq B.$$

Note that if Core i is assigned to subdivision l of test Bus j, such that $z_{ijl} = 1$, then x_{ij} cannot equal 1, i.e., Core i cannot be assigned to the entire undivided test Bus j. The following constraint is therefore added to our formulation:

$$\sum_{j=1}^{B} \sum_{l=1}^{j_{max}} z_{ijl} + \sum_{j=1}^{B} x_{ij} = 1, 1 \leq i \leq N.$$

The testing time for Core i assigned to the undivided test Bus j is given by

$$\left\lceil \frac{\phi_i}{w_j} \right\rceil t_i,$$

while the testing time changes to

$$\left\lceil \frac{\phi_i}{w_{jl}} \right\rceil t_i,$$

if Core i is assigned to the lth subdivision of test Bus j. The overall testing time for the SOC may therefore be written as:

$$\mathcal{T} = \max_j \{ \max_{l \in \{1,2,\ldots,j_{max}\}} \sum_{i=1}^{N} \left\lceil \frac{\phi_i}{w_{jl}} \right\rceil t_i z_{ijl} + \sum_{i=1}^{N} \left\lceil \frac{\phi_i}{w_j} \right\rceil t_i x_{ij} \}.$$

The nonlinear term $\frac{\phi_i}{w_{jl}}$ in the cost function can be replaced by $\phi_i v_{jl}$, as was done in Section 3, where $v_{jl} = \frac{1}{w_{jl}}$, by adding new binary indicator variables δ_{jkl} (where,

$$1 \leq j \leq B, 1 \leq k \leq w_{max}, 1 \leq l \leq j_{max})$$

to the ILP model, such that:

TABLE 4
Optimal Width and Width Distribution for $\mathcal{S}$ with Two Test Buses and a Given Maximum Testing Time

Maximum testing time $\mathcal{T}$	Optimal test width W	Optimal width distribution (w_1, w_2)
200000	27	(4,23)
190000	28	(5,23)
180000	30	(2,28)
170000	32	(6,26)
160000	34	(6,28)
150000	36	(8,28)
140000	39	(22,17)
130000	41	(18,23)
120000	46	(18,28)
110000	49	(28,21)
100000	56	(29,28)

$$\delta_{jkl} = \begin{cases} 1, & \text{if the } l\text{th subdivision of Bus } j \text{ is } k \text{ bits wide} \\ 0, & \text{otherwise.} \end{cases}$$

Additional constraints must be added to the model to 1) ensure that, for every subdivision of Bus j, there is only one value of δ_{jkl} that equals 1, i.e., subdivision l of Bus j can have only one value of width k; and 2) linearize the ceiling operator in the cost function. Since these additional constraints are similar in form to the constraints added to perform similar functions in the ILP model for Problem $\mathcal{P}_{mintime}$ in Section 3, we do not elaborate upon them here. The number of variables and constraints in the ILP model for $\mathcal{P}_{mintime}$ is each $O(NBj_{max})$, which can be written as $O(NB)$, if j_{max} is limited to a constant value. Limiting j_{max} to a constant value is desirable since a large number of test bus subdivisions may not be viable for routing and layout.

A TAM architecture was designed for the example SOC $\mathcal{S}$, consisting of two test buses, the first of which was allowed to be subdivided. The total width W was fixed at 40, with $w_1 = 24$ and $W_2 = 16$. The testing time obtained with test bus subdivision was 132,895 cycles, which is 0.8 percent lower than the time obtained without test bus subdivision (134060 cycles in Table 3). The optimal subdivisions obtained for Bus 1 were $w_{1a} = 4$ and $w_{1b} = 20$ and the test bus assignment vector was (2,2,2,2,2,1b,1b,1a,1a,1b).

Next, we present the general case of the test bus subdivision problem, in which the optimal individual undivided test bus widths must be determined as well. The formal statement of this problem is as follows:

- $\mathcal{P}_{general}$: Given N cores and B test buses of total width W and an upper limit j_{max} on the number of subdivisions allowed for test Bus j, determine 1) optimal widths for each test bus, 2) an optimal subdivision of the test buses, and 3) an assignment of cores to test buses such that the total testing time is minimized.

Problem $\mathcal{P}_{general}$ is clearly NP-hard as well since it can be restricted to Problem $\mathcal{P}_{mintime}$. The ILP model for $\mathcal{P}_{general}$ was obtained by combining the ILP models of $\mathcal{P}_{mintime}$ and

TABLE 5
Optimal Test Bus Width Distribution and Testing Time for $\mathcal{S}$ Under Place-and-Route Constraints ($W = 48$)

Constraint level $R-P$	Reward R	Penalty P	Width distribution	Testing time	Core assignment
-12	0	12	(28,20)	111140	(2,2,2,1,1,2,1,2,2,1)
-9	3	12	(28,20)	111140	(2,2,2,1,1,2,1,2,2,1)
-6	3	9	(28,20)	111140	(2,2,2,1,1,2,1,2,2,1)
-3	3	6	(28,20)	112934	(2,2,2,2,1,2,1,1,2,1)
0	0	0			infeasible
0	6	6	(28,20)	112934	(2,2,2,2,1,2,1,1,2,1)
3	6	3	(26,22)	116828	(2,2,1,2,1,2,1,2,2,1)
6	9	3	(26,22)	116828	(2,2,1,2,1,2,1,2,2,1)
12	12	0	–	–	infeasible

$\mathcal{P}_{fork}$, and adding a few variables for the purpose of linearizing the various constraints. The number of variables and constraints for the ILP model of $\mathcal{P}_{general}$ is each $O(NB)$. Again, we do not present the entire ILP model here for purposes of brevity.

5 Test Bus Sizing Under Routing and Power Constraints

In this section, we investigate the problem of TAM design under the additional constraints of power dissipation and routing costs. While testing time is an important criterion in test bus design, viable solutions to the TAM design problem must also take into account the issues of routing costs and power dissipation in order to make hardware implementation feasible. We first present a novel *reward-penalty* model for test bus sizing that incorporates designer preferences arising from place-and-route constraints. This extends the reward-model based test bus design presented in [4].

5.1 Satisfying Place-and-Route Constraints

Designer preferences related to test bus routing can be expressed as 0-1 constants r_{ij} and p_{ij}, where $r_{ij} = 1$ if the designer desires Cores i and j to be assigned to the same test bus, and $p_{ij} = 1$ if assigning Cores i and j to the same bus will have an adverse affect on routing overhead. The *reward* R for a specific TAM design is measured by

$$R = \sum_{k=1}^{B} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_{ik} x_{jk} r_{ij}$$

and the penalty is given by

$$P = \sum_{k=1}^{B} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_{ik} x_{jk} p_{ij}.$$

This provides a more detailed model than in [4], where only rewards were considered. The problem of TAM design under place-and-route constraints can now be formalized as:

- $\mathcal{P}_{route}$: Given N cores and B test buses of total width W, a minimum reward R and maximum penalty P, and a preferred assignment of cores to test buses, determine the optimal widths of the test buses and an assignment of cores to test buses such that the reward is at least R and the penalty does not exceed P.

A mathematical programming model for $\mathcal{P}_{route}$ can be derived from that for $\mathcal{P}_{mintime}$ by including the two constraints:

1. $\sum_{k=1}^{B} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_{ik} x_{jk} r_{ij} \geq R$.
2. $\sum_{k=1}^{B} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_{ik} x_{jk} p_{ij} \leq P$.

Once again, the nonlinear term $x_{ik}x_{jk}$ in the constraints above can be replaced with a new binary variable u_{ijk} by adding two new constraints:

1. $x_{ik} + x_{jk} \leq u_{ijk} + 1$
2. $x_{ik} + x_{jk} \geq 2u_{ijk}$.

The total number of variables and constraints is now given by $B(N^2 + 2N + w_{max} + 2) = O(N^2 B)$ and

$$B(2N^2 + 5N + 5) + N + 2 = O(N^2 B),$$

respectively. The reward-penalty model was added to the ILP formulation to determine the best testing times for $\mathcal{S}$ that can be achieved under place-and-route constraints. We detemined the testing times for test bus width $W = 48$ for several combinations of R and P, and compared these times with the time obtained for $W = 48$ in $\mathcal{P}_{assign}$. The designer preferences chosen were related to the optimal core assignment A = (2,2,2,1,1,2,1,2,2,1) obtained for $W = 48$ using $\mathcal{P}_{assign}$, i.e., Bus 2 is used for Cores 1, 2, 3, 6, 10, and Bus 1 is used for 4, 5, 7, 8, 9. We used a first set of rewards to guide the assignments towards A (to minimize testing time) by encoding $r_{13}, r_{36}, r_{49}, r_{78} = 1$ and a second set of rewards $r_{14}, r_{17}, r_{25}, r_{29}, r_{38}, r_{39}, r_{64}, r_{810} = 1$ (based on hypothetical routing preferences) to lead the assignment away from A. Similarly, a first set of penalties $p_{15}, p_{24}, p_{67}, p_{910} = 1$ was chosen to lead the assignment towards A and a second set of penalties $p_{12}, p_{16}, p_{23}, p_{610}, p_{45}, p_{47}, p_{58}, p_{59} = 1$ was chosen to lessen the chances of choosing A.

The ILP model required approximately 28 seconds to be solved in all cases. Table 5 presents the test width distribution and testing times obtained under routing constraints. The first line in Table 5 shows the testing time 111,140 obtained in Section 3 for $W = 48$, $R = 0$, and $P = 12$, i.e., when no routing constraints are in effect. The remaining rows of the table present test width distribution and testing times for increasingly constrained assignments. We measure the routing constraint level by $R - P$, where

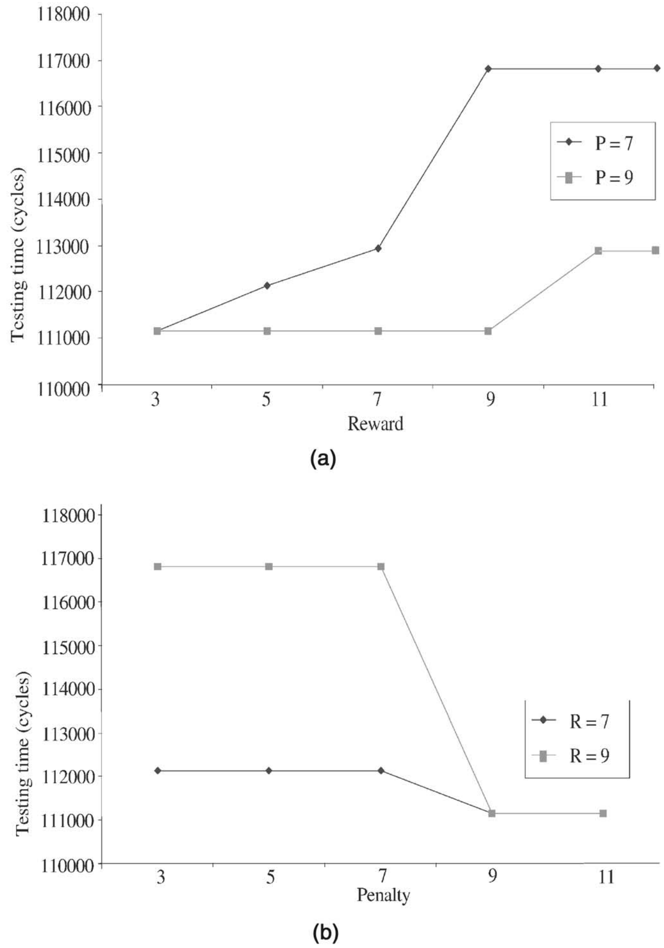

Fig. 4. Variation of testing time under place-and-route constraints: (a) increase in testing time with increasing R and (b) decrease in testing time with increasing P.

level $0 - 12 = -12$ is the least constrained, and level $12 - 0 = 12$ is the most constrained. Note that the testing time does not increase monotonically with increasing constraint level. This is because the testing time is not linearly dependent on $P - R$, but varies with both P and R. For instance, the problem is found to be infeasible whenever $P = 0$ (the most stringent penalty level), while solutions for $R = 12$ (the most stringent reward level) are possible.

Fig. 4 illustrates how testing time varies with increasing R (keeping P constant) and increasing P (keeping R constant) for $W = 48$. In Fig. 4a, for $P = 9$, testing time increases from the optimum value of 111,140 cycles at $R = 3$ to the maximum value of 116,828 cycles at $R = 9$. Testing time levels off at this higher value (116,828), since a maximally constrained assignment has been reached and further increases in R have no effect on testing time. Similarly, in Fig. 4b, for $R = 9$, testing time is at its maximum when $P \leq 7$ since the assignment is maximally constrained for $P \leq 7$. Testing time decreases as P is increased and reaches the optimum value of 111,140 cycles at $P = 9$.

5.2 Satisfying Power Constraints

Power dissipation is an important consideration in SOC testing. A high degree of test parallelism can cause the power dissipation to exceed the maximum power rating of the IC during test mode, resulting in permanent damage to the system. If the peak power exceeds a threshold value, it can cause structural damage to the silicon or to the package. Likewise, elevated average power can also cause structural damage to the silicon, bonding wires, or the package. Power dissipation also adds to the thermal load that must be transported away from the device under test.

Here, we investigate the problem of TAM optimization, while incorporating system-level power constraints. The power dissipated during core testing is related to the speed at which the tests are applied to the core. Let Bus j (width w_j) be used to apply tests to Core i in the SOC. Using the test data serialization model in Section 2, the average number of cycles required to apply a test pattern to Core i is given by

$$c_{ij} = \left\lceil \frac{\phi_i}{w_j} \right\rceil \frac{t_i}{p_i} = \left\lceil \frac{\phi_i}{w_j} \right\rceil \frac{(p_i + 1)\lceil f_i/N_i \rceil + p_i}{p_i}.$$

Therefore, test patterns are applied to Core i at the rate of $1/c_{ij}$ per cycle. Let the highest amount of energy dissipated by Core i, on application of any one test pattern be G_i. Note that G_i is the maximum energy dissipated by any test pattern, irrespective of the previous test patterns applied. Then, the peak power dissipated during the test is given by $\frac{G_i}{c_{ij}}$. Therefore, while an increase in w_j reduces testing time, it

TABLE 6
TAM Design Under Power Dissipation Constraints

Total test width W	$P_1 = P_2 = 325$ Optimal distribution (w_1, w_2)	$P_1 = P_2 = 325$ Optimum testing time	$P_1 = P_2 = 400$ Optimal distribution (w_1, w_2)	$P_1 = P_2 = 400$ Optimum testing time
36	(9,27)[1]	150877	(7,29)	154428
40	(8,32)	158330	(12,28)	135058
44	(12,32)	141462	(20,24)	122834
48	(24,24)[1]	113605	(24,24)	112887
52	(20,32)	121308	(23,29)	107496
56	(28,28)	96624	(28,28)	96624
60	(32,28)	94458	(29,31)	94380
64	(32,32)	91303	(32,32)	88986

Intermediate result, not provably optimal.

also causes c_{ij} to decrease and, thus, increases the peak power dissipation during the test. While our problem formulation addresses test bus optimization under peak power dissipation constraints, it may be easily adapted to address the problem of average power dissipation as well.

A tradeoff between testing time and power dissipation must therefore be made in order to keep testing time practical, while ensuring that the maximum power rating of the SOC is not exceeded. We assume, without loss of generality, that the maximum energy dissipated by a test pattern, denoted by G_i, is proportional to the gate count of the circuit. Furthermore, the maximum power rating P_{max} is divided equally among the B test buses in the SOC. Alternative strategies to divide the power rating among the test buses, e.g., based on test bus width, can be easily incorporated in this model.

The problem of optimizing test bus widths and testing time under power constraints can now be formalized as:

- $\mathcal{P}_{power}$: Given N cores, B test buses of total width W, and a maximum power rating $\frac{P_j = P_{max}}{B}$ for each test bus, determine the optimal widths of the test buses and an optimal assignment of cores to test buses.

A mathematical programming model for $\mathcal{P}_{power}$ can be derived from $\mathcal{P}_{mintime}$ by adding the constraint $x_{ij}\frac{G_i}{c_{ij}} \leq P_j$. This constraint can be linearized as $x_{ij}G_i \leq c_{ij}P_j$ or $x_{ij}G_i \leq \psi_{ij}P_j\frac{t_i}{p_i}$, using the variable ψ_{ij} to denote the expression

$$\left\lceil \frac{\phi_i}{w_j} \right\rceil,$$

as in Section 3. This model has a total of $B(5N + 5) + N = O(NB)$ variables and $B(6N + 5) + N = O(NB)$ constraints.

The ILP model for $\mathcal{P}_{power}$ was solved to obtain bus width distributions and testing times for several values of rated power P_{max}. The time needed to reach a solution was approximately 26 seconds on average. Table 6 presents results on power-constrained TAM design for SOC $\mathcal{S}$. The testing time under power constraints is clearly greater than the unconstrained testing time for the same total test bus width (Table 3). However, the increase in testing time decreases progressively with increasing test width. Therefore, while power constraint $P_j = 400$ units increases the testing time by 3.66 percent at $W = 36$, the corresponding increase at $W = 64$ is only 0.72 percent, thus demonstrating the scalability of the proposed technique.

6 Conclusions

We have presented a formal methodology for optimal test bus sizing for system-on-a-chip test. The techniques developed include an improved test data deserialization strategy that greatly reduces the system testing time compared to that reported earlier in the literature. Several important issues in TAM design have been explored. These include optimal test bus width distribution, assignment of cores to test buses, and determination of the minimum test bus width required to satisfy testing time specifications. A novel reward-penalty model incorporates designer place-and-route preferences in the TAM design, making it possible for the designer to trade testing time with ease of placement and routing. System-level power dissipation constraints have also been incorporated in the model to ensure that the maximum power rating of the system is not exceeded during test. The test bus optimization models developed have been solved for a realistic example SOC; the experimental results presented demonstrate the feasibility of the proposed methodology.

Acknowledgments

The authors thank Erik Jan Marinissen and Wim Verhaegh of Philips Research Laboratories for valuable comments on a draft version of this manuscript. This research was supported in part by the US National Science Foundation under grant number CCR-9875324.

References

[1] J. Aerts and E.J. Marinissen, "Scan Chain Design for Test Time Reduction in Core-Based ICs," *Proc. Int'l Test Conf.*, pp. 448-457, 1998.

[2] M. Berkelaar, *lpsolve*, version 3.0. Eindhoven Univ. of Technology, Design Automation Section, Eindhoven, The Netherlands, ftp://ftp.ics.ele.tue.nl/pub/lp_solve. 2000.

[3] K. Chakrabarty, "Design of System-on-a-Chip Test Access Architectures Using Integer Linear Programming," *Proc. VLSI Test Symp.*, pp. 127-134, 2000.

[4] K. Chakrabarty, "Design of System-on-a-Chip Test Access Architectures Under Place-and-Route and Power Constraints," *Proc. Design Automation Conf.*, pp. 432-437, 2000.
[5] S. Davidson, P1500 Benchmarking Task Force Chair, private correspondence, Dec. 1999.
[6] I. Ghosh, S. Dey, and N.K. Jha, "A Fast and Low Cost Testing Technique for Core-Based System-on-Chip," *Proc. Design Automation Conf.*, pp. 542-547, 1998.
[7] P. Harrod, "Testing Re-Usable IP: A Case Study," *Proc. Int'l Test Conf.*, pp. 493-498, 1999.
[8] V. Immaneni and S. Raman, "Direct Access Test Scheme—Design of Block and Core Cells for Embedded ASICs," *Proc. Int'l Test Conf.*, pp. 488-492, 1990.
[9] *International Technology Roadmap for Semiconductors.* Semiconductor Industry Assoc., http://public.itrs.net/files/1999_SIA_Roadmap/Home.htm. 1999.
[10] 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," *Proc. Int'l Test Conf.*, pp. 284-293, 1998.
[11] E.J. Marinissen, S.K. Goel, and M. Lousberg, "Wrapper Design for Embedded Core Test," *Proc. Int'l Test Conf.*, pp. 911-920, 2000.
[12] N.A. Touba and B. Pouya, "Using Partial Isolation Rings to Test Core-Based Designs," *IEEE Design and Test of Computers*, vol. 14, pp. 52-59, Oct./Dec. 1997.
[13] P. Varma and S. Bhatia, "A Structured Test Re-Use Methodology for Core-Based System Chips, *Proc. Int'l Test Conf.*, pp. 294-302, 1998.

Vikram Iyengar received the BE degree in electrical and electronics engineering from the Birla Institute of Technology, India, in 1996, and the MS degree in computer engineering from Boston University in 1998. He is presently a PhD candidate and research assistant in computer engineering at Duke University. Mr. Iyengar has published more than 20 papers in archival journals, refereed conference proceedings, and workshops. He is currently working on research projects related to test scheduling and test delivery architectures for system-on-chip designs. He received an IBM Graduate Fellowship in 2001. Beginning July 2002, he will be an advisory engineer at IBM, Burlington, VT. He is a student member of the IEEE.

Krishnendu Chakrabarty received the BTech degree from the Indian Institute of Technology, Kharagpur, in 1990, the MSE and PhD degrees from the University of Michigan, Ann Arbor, in 1992 and 1995, respectively, all in computer science and engineering. He is now an assistant professor of electrical and computer engineering at Duke University. Dr. Chakrabarty is a recipient of the US National Science Foundation Early Faculty (CAREER) award, the US Office of Naval Research Young Investigator award, and the Mercator Professor award from Deutsche Forschungsgemeinschaft, Germany. He received a best paper award at the 2001 Design, Automation and Test in Europe (DATE) Conference. His current research projects (supported by the US National Science Foundation, the Defense Advabced Research Projects Agencey, the US Office of Naval Research, and industrial sponsors) are in system-on-chip test, real-time operating systems, distributed sensor networks, and architectural optimization of microelectrofluidic systems. He has published more than 80 papers in archival journals and refereed conference proceedings, and he holds a US patent in built-in self-test. He is a senior member of the IEEE, a member of ACM, and ACM SIGDA, and a member of Sigma Xi. He serves as an associate editor for *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, and as an editor for *JETTA*. He is guest editing a special issue of *JETTA* on system-on-a-chip test, scheduled for publication in 2002. He also serves as vice chair of technical activities in IEEE's Test Technology Technical Council, and is a member of the program committees of several IEEE/ACM conferences and workshops.