Power-Aware System-on-Chip Test Optimization through Frequency and Voltage Scaling

by

Vijay Sheshadri

A dissertation submitted to the Graduate Faculty of Auburn University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Auburn, Alabama
May 4, 2014

Keywords: System-on-Chip, Test Schedule Optimization, Voltage and Frequency Scaling, Session-Based Test Schedule, Sessionless Test Schedule

Copyright 2014 by Vijay Sheshadri

Approved by

Prathima Agrawal, Samuel Ginn Distinguished Professor of Electrical and Computer Engineering
Vishwani D. Agrawal, James J. Danaher Professor of Electrical and Computer Engineering
Adit Singh, James B. Davis Professor of Electrical and Computer Engineering
Abstract

A System-on-Chip (SoC) is a complete system that has been integrated onto a single chip. An SoC is often designed by embedding reusable blocks called cores. With shrinking device sizes, SoC cores are growing in number and complexity, which has led to high volumes of test data and resulted in long test times. Therefore, reducing test cost by minimizing the overall test time is one of the main goals of System-on-Chip (SoC) testing. Power dissipation during test mode is often much higher than that of functional mode and hence, test power management is also a major concern in SoC testing. To efficiently manage test resources and power dissipation, tests for the SoC cores are arranged into test schedules. Within these test schedules, the core tests may (as in the case of session-based test schedule) or may not (as in the case of sessionless test schedule) be grouped into test sessions. Traditional SoC test methods assume a constant test frequency and supply voltage (V_{DD}) for the entire test schedule. However, test time and test power can be regulated by V_{DD} and test clock frequency to optimize SoC test schedules for a given power budget.

The research presented in this dissertation focuses on power-aware optimization of SoC test schedules to minimize test time by scaling the supply voltage and test clock rate. This scaling can be session wise (in the case of a session-based test schedule) or dynamic (in case of sessionless test schedule). SoC testing can be sped up by increasing the test clock rate. However, test clock is constrained by the rated power limit (power constraint) and the critical path delay (structure constraint) of the SoC cores. These constraints can be manipulated using V_{DD}. Therefore, by scaling V_{DD} and clock rate, an optimal test time and schedule can be obtained for an SoC.

For the session-based test scheduling, the optimization problem is mathematically formulated and solved through Integer Linear Program (ILP) based methods to provide optimal
solutions. For SoCs with large number of cores, Integer Linear Programs are NP-hard and, in general, computationally expensive. To overcome this difficulty, a simulated annealing based heuristic method capable of providing near-optimal solutions is developed. Results show that the overall SoC test time can be considerably shortened by scaling the test clock and supply voltage. A similar heuristic method that is based on simulated annealing algorithm, is developed for the optimization of sessionless test schedules. The heuristic approach is capable of both preemptive (tests can be halted and resumed at will) and non-preemptive scheduling (tests cannot be interrupted at any time). Here also, the optimization results show a significant test time reduction over conventional reference test schedules where \( V_{DD} \) and clock are fixed at given nominal values.
Acknowledgments

There are many people to whom I would like to express my gratitude for their help during the pursuit of my doctoral dreams. Foremost among them are Professors Prathima Agrawal and Vishwani D. Agrawal, without whose constant support and guidance this dissertation would not have been possible.

I would like to thank Dr. Adit Singh, for serving as a member of my advisory committee, and Dr. Sanjeev Baskiyar, the external reader for this dissertation. I am also thankful to Dr. Victor Nelson, for teaching me the various VLSI-CAD tools through his course ‘CAD for VLSI’. Thanks are in order, as well, for Dr. Alice Smith and Dr. Chase Murray from the Dept. of Industrial Engineering. I have immensely benefited from their courses on Adaptive Optimization and Linear Programming. I would also like to acknowledge the support of Ms. Shelia Collis and the Wireless Engineering Research and Education Center (WEREC), Ms. Jo Ann Loden, Ms. Penny Christopher and Ms. Carol Lovvorn in helping me keep my school and immigration paper work in order.

My heartfelt thanks go out to my colleagues in Broun Hall, in particular, Dr. Pratap Prasad, Dr. Santosh Kulkarni, Dr. Suraj Sindia, Praveen Venkataramani, Gopalkrishna Iyer and Muralidharan. I owe much gratitude to all my friends, especially to the Auburn Kannada Koota gang.

Above all, I would like to express my deepest gratitude to my caring parents and my loving wife for their constant love compassion and support.

Finally, I must acknowledge that the research presented in this dissertation, is supported in part by the National Science Foundation Grants CCF-1116213, IIP-0738088 and IIP-1266036.
Table of Contents

Abstract ................................................................. ii
Acknowledgments ......................................................... iv
List of Figures ............................................................. vii
List of Tables .............................................................. ix
1 Introduction ............................................................ 1
   1.1 Problem Statement ............................................. 2
   1.2 Organization of Dissertation ................................ 3
2 Background and Prior Work on SoC Testing ......................... 4
   2.1 Test Infrastructure ............................................. 5
       2.1.1 Core Test Wrapper ..................................... 6
       2.1.2 Test Access Mechanism (TAM) ......................... 7
   2.2 Test Scheduling ............................................... 7
   2.3 Power Constrained Testing ................................. 9
   2.4 Frequency and Voltage Scaling ............................ 10
3 Optimization of Session-Based Test Schedules ..................... 12
   3.1 Background .................................................... 12
   3.2 Mixed-Integer Linear Program (MILP) Based Optimization .... 18
       3.2.1 Introduction ......................................... 18
       3.2.2 MILP Formulation .................................. 19
   3.3 Heuristic Based Optimization ................................ 21
       3.3.1 Simulated Annealing (SA) .......................... 21
   3.4 Results ........................................................ 25
       3.4.1 Experimental Setup ................................ 25
List of Figures

2.1 A simple test set-up showing the SoC under test, the test source and sink. The test data from the test source to different cores and from the cores to the test sink, is carried over the test bus. ......................... 4

2.2 SoC Test scheduling modeled as 3D optimization problem. .................. 5

2.3 Overview of IEEE1500 wrapper [40]. (WBR = wrapper boundary register; WBY = wrapper bypass; WP(I/O) = wrapper parallel (input/output); WS(I/O) = wrapper serial (input/output); WIR = wrapper instruction register.) .......... 6

2.4 Two test scheduling strategies, session-based (non-partitioned) and sessionless(partitioned) are illustrated. Sessionless testing can be non-preemptive (b) or preemptive (c) [33,35]. ...................................................... 8

2.5 Test time as a function of $V_{DD}$ [64]. The nominal and the optimal $V_{DD}$ are denoted by $V_{nom}$ and $V_{sync}$, respectively. ......................... 11

3.1 Overview of the SA heuristic algorithm. ................................. 22

3.2 Generating the initial solution for the SA algorithm. ...................... 23

3.3 Components of ASIC Z and their test time (in arbitrary units) and test power (in mW). ................................................................. 25

3.4 CPU time (in seconds) of the MILP optimization method for the ITC benchmarks. 30
3.5 CPU time (in seconds) of the heuristic optimization method for the SoC benchmarks. ......................................................... 32

4.1 ‘Merging’ test sessions to convert a session based test schedule into a sessionless test schedule. ......................................................... 37

4.2 The ‘Merge’ function erases the session boundaries in a session based test schedule and combines the tests together to form a sessionless test schedule. ......................................................... 38

4.3 Best-fit decreasing (BFD) algorithm for sessionless test scheduling. Test schedules obtained from this algorithm are used as reference cases in this paper where voltage and frequency are fixed at their nominal values for the entire schedule. ......................................................... 39

4.4 Convergence of the SA based optimization algorithm. The plot shows the heuristic algorithm converging towards the optimum test time as the temperature parameter \( T \) reduces. ......................................................... 40

4.5 Optimized sessionless test schedules for ASIC Z, under non-overlapping voltage range condition, for both (a) non-preemptive and (b) preemptive cases. (Note: Diagram not to scale.) ......................................................... 45

4.6 Comparing session-based and sessionless test schedules. ......................................................... 47
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Test Data set for ASIC Z at nominal supply voltage (1.0V)</td>
<td>26</td>
</tr>
<tr>
<td>3.2</td>
<td>Optimized Test Schedule for ASIC Z for nominal and custom clock rate (Cases 1 and 2). The supply voltage is at nominal value for both cases.</td>
<td>28</td>
</tr>
<tr>
<td>3.3</td>
<td>Test times (in arbitrary units) of ASIC Z system for custom $V_{DD}$ and clock rate (Case 3).</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Test times (in arbitrary units) for benchmark SoCs, obtained by MILP optimization method for the three optimization cases considered.</td>
<td>29</td>
</tr>
<tr>
<td>3.5</td>
<td>Test times (in arbitrary units) for MILP and heuristic test scheduling methods, with customized $V_{DD}$ and clock rates.</td>
<td>30</td>
</tr>
<tr>
<td>3.6</td>
<td>Test times (in arbitrary units) for benchmark SoCs, obtained by the heuristic optimization method for the three optimization cases considered.</td>
<td>31</td>
</tr>
<tr>
<td>3.7</td>
<td>Optimal test times (in arbitrary units) for nominal and custom clock rate (Cases 1 and 2) compared with the lower bound on test time at nominal $V_{DD}$.</td>
<td>33</td>
</tr>
<tr>
<td>3.8</td>
<td>Optimal test times (in arbitrary units) for custom $V_{DD}$ and clock frequency (Optimization case 3) compared with the lower bound on test time at $V_{min}$.</td>
<td>34</td>
</tr>
<tr>
<td>3.9</td>
<td>Optimized test times (in arbitrary units) for ASIC Z, for various power budget values.</td>
<td>34</td>
</tr>
<tr>
<td>3.10</td>
<td>Test times (in arbitrary units) and optimal voltages of ASIC Z system, for dual voltage ranges.</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>Test times (in arbitrary units) for sessionless test scheduling with voltage and frequency scaling.</td>
<td>41</td>
</tr>
<tr>
<td>4.2</td>
<td>CPU times* for all the benchmarks SoCs for the heuristic optimization algorithm.</td>
<td>42</td>
</tr>
<tr>
<td>4.3</td>
<td>Test times (in arbitrary units) for sessionless test scheduling with voltage and frequency scaling.</td>
<td>43</td>
</tr>
<tr>
<td>4.4</td>
<td>Sessionless test schedule optimization for ASIC Z, subject to various power budget values.</td>
<td>44</td>
</tr>
</tbody>
</table>
4.5 Comparing test time results for session-based and sessionless test schedule optimization. ................................................................. 46

4.6 Comparison of CPU times between session-based and sessionless testing, for the heuristic optimization algorithm. .................................................. 48

1 Overview of the benchmarks used in this dissertation. .............................. 60
Chapter 1
Introduction

Technological developments have made it possible to integrate an entire system onto a single chip. Termed as ‘System-On-Chip’ (or SoC for short), these devices have core based architecture where each core is often a reusable logic or memory block. Owing to their modularity, small area, high performance and low power consumption, SoC devices are becoming increasingly popular. In recent times, SoCs are extensively used in networking and communication applications. Emerging cellular and wireless technologies, such as WiMAX and LTE require high data rates, low latency at very low power budgets. SoCs are well suited for such requirements as they offer low power, programmable and cost-effective hardware solutions. SoCs are the driving force behind modern-day smart phones and tablets, and can also be found in other wireless applications such as radios, wireless access points, Bluetooth, etc.

The number of cores being embedded in SoC devices is increasing due to device size miniaturization. The resulting complexity and increase in the number of fault-sites has complicated testing of SoCs. Consequentially, the test data volume also grows in proportion to the number of cores in the SoC, since each core is associated with one or more tests, leading to longer test times. Thus, test time minimization has become a major challenge in the field of SoC testing. While testing multiple cores simultaneously can reduce the test time significantly, such concurrent execution is limited by excessive power dissipation due to increased switching activity. The power dissipation of a circuit during test mode is often higher than functional mode. Elevated power levels and heat dissipation by neighbouring cores can lead to the formation of thermal hotspots and undesirable power droops. Thermal hotspots may eventually cause irreversible damages to the chip whereas power droops
induce clock stretching which may lead to a good chip incorrectly failing a timing test [18]. Therefore, power-aware test strategies are needed for efficient test power management.

1.1 Problem Statement

The complexity of SoC testing is mitigated to an extent by adopting modular testing, which is equivalent to a “divide-and-conquer” approach. In modular testing, block-level tests can be applied to individual blocks (cores) of the SoC, such that these blocks can be tested almost independent of one another. As discussed earlier, concurrent testing of cores presents a trade-off between test time and test power. Hence, an optimal arrangement of core tests can be formed so as to yield a minimal test time while maintaining the test power under a safe limit. Such an arrangement is termed as SoC test schedule (discussed in detail in the next chapter).

The general SoC test scheduling problem can be stated as: Given an SoC with \(N\) cores, where each core may be associated with one or more tests, and a test power budget, find a test schedule (concurrency and sequence of test application) to:

a. Test all cores.

b. Reduce the overall test time.

c. Conform to the SoC test power budget.

The contribution of this work is a power-aware test scheme that optimizes the overall test time of an SoC by exploiting the influence of \(V_{DD}\) and clock over test power and test time of individual cores. In this work, both exact and heuristic approaches for test optimization are provided; while the exact method provides the most optimal result, the heuristic method achieves near-optimal results but addresses the problem of scalability.
1.2 Organization of Dissertation

The rest of this dissertation is organized as follows: Chapter 2 discusses the basics of SoC testing methodologies and summarizes the previous work in this field. Optimization techniques for SoC test schedules are presented in Chapters 3 and 4. Chapter 5 concludes the research work presented in this dissertation. The details of the SoC benchmarks, used in this research work, are provided in the appendix.
Chapter 2
Background and Prior Work on SoC Testing

As mentioned previously, in SoC testing, the modularity of an SoC is exploited by treating each core as a testable unit. A simple test set-up for an SoC is shown in Figure 2.1. The test source stores and provides the test stimuli for all the cores. The test bus relays the stimuli to the corresponding core and the test response of the cores to test sink. The test sink stores all responses which are then compared to the response of a fault-free version of the device to identify the faults.

![Figure 2.1: A simple test set-up showing the SoC under test, the test source and sink. The test data from the test source to different cores and from the cores to the test sink, is carried over the test bus.](image)

It is easily seen that as the number of cores increases, the overall test time of the SoC also increases. While concurrent testing of cores may cut down the overall test time, there may be other factors influencing it. For instance, there may be some test resources, such as test bus, external pins of the SoC, etc., that may be commonly shared among cores. This may lead to a conflict when such cores are tested simultaneously. Test power may also limit concurrent testing of cores.

The SoC testing problem can be modeled as a 3-dimensional optimization problem, where the SoCs power limit, test time and resources (such as pin count, etc.) form the three
axes. The power limit is fixed for the SoC and the resources have a limited availability. The objective of the 3-D optimization would be to minimize the test time by effective allocation of resources such that the power limit is not exceeded. This optimization problem has been modeled as a 3-D bin packing problem [24] as shown in Figure 2.2. Each core in the SoC can be modeled as a cuboid, where the core’s test power, test time and test resources, such as BIST resources, wrapper width, etc., constitute the three dimensions. The idea here is to place the cores in the cuboid representing the SoC in such a way that the test time is minimized while satisfying the power and resource constraints. This bin packing problem differs from the general bin packing problem in that if two cores are tested simultaneously, they overlap only on the time axis and not on the other two axes.

2.1 Test Infrastructure

The test infrastructure of SoC consists of a wrapper and a test access mechanism (TAM) [7].

Figure 2.2: SoC Test scheduling modeled as 3D optimization problem.
2.1.1 Core Test Wrapper

The test wrapper aides in the access and isolation of embedded cores. It acts as an interface between the core and the on-chip structure for test data transportation (TAM). The IEEE 1500 standard for embedded cores defines a standardized, scalable and configurable core wrapper for both logic and memory cores [1]. This wrapper consists of scan and control registers, data and control signals and instruction set. The IEEE 1500 wrapper architecture is shown in Figure 2.3. The boundary registers (WBR) form the wrapper chains which interface the TAM with the internal scan chains through the parallel pins (WPI/WPO). The instruction register (WIR) provides the necessary control information.

The wrapper may also perform serial-parallel or parallel-serial conversion to provide width adaptation in case of a mismatch between the available TAM width and the core input/output terminals. Wrapper configuration can be optimized for effective utilization of test bandwidth [20,21,31,35,39,47,48].

Figure 2.3: Overview of IEEE1500 wrapper [40]. (WBR = wrapper boundary register; WBY = wrapper bypass; WP(I/O) = wrapper parallel (input/output); WS(I/O) = wrapper serial (input/output); WIR = wrapper instruction register.)
2.1.2 Test Access Mechanism (TAM)

TAM is the infrastructure responsible for transporting the test data between SoC external pins and embedded cores. The TAM can be dedicated solely for test purposes or can be an existing on-chip bus structure. TAM design involves trade-offs between the transport capacity of the mechanism and the test application cost incurred in terms of test time, area overhead, etc.

Multiple TAM architectures have been proposed in the past. Aerts and Marinissen introduce [5]:

- A multiplexing architecture, where the entire TAM width is allocated to each core and the cores are tested sequentially,
- A distributed architecture, where a fixed TAM width is assigned to each core,
- A daisy chain architecture, where all the TAM width is assigned to every core but a bypass structure is added to shorten the access path for the cores.

More recently, a flexible TAM architecture has been proposed, where the TAM assignment to the cores is flexible; hence the TAM varies dynamically after each core test [26, 27, 70].

Previously published work formulates the test time as a function of TAM width and optimizes TAM allocation among cores, to achieve test time minimization. It has been shown [27] that the relation between the test time and TAM width is that of a ‘staircase’ function, meaning that the test time will only reduce after the TAM assignment to a core exceeds a certain core threshold value. Some of the published references on TAM optimization are [25–27, 34, 48, 70].

2.2 Test Scheduling

As mentioned earlier, a test schedule is an ordered arrangement of core tests often optimized for lowering test time and/or test power. A simple test schedule can be sequential
Figure 2.4: Two test scheduling strategies, session-based (non-partitioned) and sessionless (partitioned) are illustrated. Sessionless testing can be non-preemptive (b) or preemptive (c) [33,35].

or concurrent. In sequential scheduling, only one core is tested at a time. As a result, the overall test time is simply the sum of all individual core test times. While this is the simplest scheme to implement, the overall test time is longest. Concurrent scheduling, on the other hand, makes use of concurrency by simultaneously testing multiple cores. Existing concurrent scheduling strategies may be broadly categorized into:

- Session-based (non-partitioned) test scheduling, where no new test is allowed to start until all tests of a previous session are completed. A test session refers of a set of tests initiated simultaneously and run concurrently [13,14,35,55,56,72]. See the illustration in Figure 2.4(a).

- Sessionless (partitioned) test scheduling, where test session boundaries are ignored and a test may be scheduled to start as soon as possible [23,44,54,57,58,73]. The sessionless or partitioned test scheduling can be further divided into preemptive and non-preemptive scheduling. In the preemptive strategy, tests can be interrupted or restarted at
any time [25, 34]. The non preemptive strategy does not allow such interruptions, i.e., a test once initiated must be completed. Figures 2.4(b) and (c) illustrate non-preemptive and preemptive schedules, respectively.

2.3 Power Constrained Testing

To be used during test mode, the test vectors are so designed as to maximize switching activity in the circuit, in order to detect more faults per vector and hence minimize the test time. Therefore, test power can be up to four times the functional power [67]. To guarantee that this increased power dissipation will not cause heat induced failures in the device, a peak power budget for the entire SoC is defined. Power constrained test scheduling focuses on optimizing the overall test time of the SoC for a given power budget.

Many power-constrained testing strategies have been studied in the past [13, 14, 18, 25, 34, 53, 55, 56, 70]. The concept of fixing a single power budget for the SoC is known as the global peak power model and has been widely used. However, this model is regarded as a pessimistic approach since the single power limit value is based on the peak power consumption of the circuit and the circuit’s power consumption may not often reach peak power. Samii et al. proposed a cycle-accurate power model where there is a power value for every clock cycle [52]. While this model is more accurate than the global peak power model, it is more complex and computationally expensive. Alternatively, Larsson [32] proposed a power grid model aimed at countering local hot spots. This model allocates cores to a set of power grids. During test scheduling, cores are selected such that not only the global peak power limit, but also the grid’s power limit is not exceeded.

In the research presented in this dissertation, we adopt a global peak power model where the power consumption during simultaneous execution of multiple tests is given by the sum of their peak powers, and this value must not exceed the peak power budget of the SoC at any given time. The additive model for estimating power consumption was introduced by
Chou et al. [13,14]. In this model, the test power consumption of a block is approximated to a single value corresponding to the peak power consumption over the test time of the block.

### 2.4 Frequency and Voltage Scaling

The idea of scaling voltage and frequency has been prevalent in the field of microprocessors and SoCs. In [60], a locally placed configurable dynamic voltage and frequency scaling (DVFS) controller enables a large number of on-chip processors to switch $V_{DD}$ by selecting from two power grids and also independently controls their clock rates, in order to improve the energy efficiency of the multi-processor SoC (MPSoC). Voltage and frequency scaling techniques have also been employed in testing of SoCs.

Recently, multi-frequency SoC testing has been investigated. Sehgal et al. [53] have focused on the use of a multi-channel ATE capable of providing multiple frequencies. Zhao et al. [69] discuss test optimization through wrapper design in order to perform bandwidth matching between the ATE’s clock input and the core’s frequency. The PMScan system, introduced by Devanathan et al. [16], utilizes an adaptive supply voltage regulation scheme that lowers the $V_{DD}$ to balance the power dissipation and the frequency during the shift operation, while satisfying the timing requirements. Scheduling with multiple voltage islands and testing of cores at multiple voltages has also been considered [29]. These authors schedule core tests at multiple voltage levels and clock domains and reduce the clock frequency during low voltage testing to enable a time division multiplexing scheme for concurrent testing of cores.

Venkataramani et al. [62–65] discuss two aspects of testing, namely, power constrained testing where the test clock speed is limited by the circuit’s rated power and structure constrained testing where the test clock speed is limited by the critical path or other timing constraints of the circuit. The supply voltage is used for the purpose of balancing these two constraints to allow higher test clock rates in order to achieve test time reduction. Since test power is two to four times higher than the functional power, test clock is often power constrained, i.e., any increase in the clock would cause the power to exceed the device’s rated
maximum. The power consumption can be reduced by lowering the operating voltage. As a result, the clock rate can be increased without exceeding the power constraint of the core. However, reducing the voltage causes the delay of a circuit to increase, hence, elongating the critical path of the device. Thus, as we reduce $V_{DD}$, on one hand, the lowered power consumption allows higher clock rates thereby shrinking the test time but, on the other hand, the increased circuit delay requires slower clock rate and a longer test time. As Figure 2.5 [63,64] shows there exists an optimal point where the two constraints are satisfied and at the same time test time is significantly reduced. Experiments on ISCAS benchmark circuits by those authors show test time reductions of up to 62% at optimal values of $V_{DD}$ [62].
Chapter 3
Optimization of Session-Based Test Schedules

3.1 Background

Objectives of SoC testing, as outlined previously, are to test all cores of the SoC while managing the power dissipation so as to avoid thermal stress. Consider an SoC with \( n \) cores \( C_1, \ldots, C_n \), where each core \( C_i \) is associated with a test \( t_i \), and a peak power budget. The power budget for an SoC, \( P_{\text{max}} \), is defined as the maximum allowable power dissipation during the testing of SoC. The power budget is chosen so as to account for power droops and thermal hotspots that may occur due to peak activity during testing. Let there be \( n \) cores, \( C_1, \ldots, C_n \) in an SoC and let the test corresponding to a core \( C_i \) be \( t_i \), where \( i \in 1, 2, \ldots, n \). Each test is associated with test time and test power, which have been characterized at nominal operating conditions (nominal voltage and clock rate). Let \( T_{t_i} \) and \( P_{t_i} \) be the time and power of the test \( t_i \). Let tests, \( t_1, \ldots, t_n \), be distributed among \( k \) sessions, \( S_1, \ldots, S_k \) such that each session, \( S_j \) contains one or more tests. The test time of a session \( S_j \), given by \( T_{S_j} = \max(T_{t_i} | \forall t_i \in S_j) \) and the power dissipated during session, \( S_j \) is given by \( P_{S_j} = \sum(P_{t_i}), \forall t_i \in S_j \).

The general test scheduling problem can be expressed as an optimization problem:

Objective:

Minimize \( \sum_{j=1}^{k} T_{S_j} \cdot x_j \)

where \( x_j = \begin{cases} 1, & \text{if } S_j \text{ is scheduled} \\ 0, & \text{otherwise} \end{cases} \)
Subject to:

1) Power constraint: $P_{S_j}x_j \leq P_{\text{max}}$, $P_{\text{max}}$ being the power budget for the SoC.

2) Test completeness constraint: each test, $t_i, i \in \{1, 2, \cdots, n\}$ is executed at least once.

The test time and test power can, however, be influenced by the test clock. A faster clock reduces the test time but increases the power consumption. Lowering the clock frequency, on the other hand, reduces the test power but leads to longer test time. Thus, there exists a trade-off between the test time and test power. However, the energy spent during testing remains constant. Energy spent per core test, $E_{ti} = P_{ti} \times T_{ti}$. The total energy spent during the testing of the entire SoC can then be expressed as, $E_{\text{total}} = \sum_{i=1}^{n} P_{ti} \times T_{ti}$. In [63,64], the authors have proved that a lower bound on the total test time is given by the ratio of the total energy spent during the test and the power budget,

$$TT_{LB1} = \frac{E_{\text{total}}}{P_{\text{max}}} = \frac{\sum_{i=1}^{n} P_{ti} \times T_{ti}}{P_{\text{max}}} = \sum_{i=1}^{n} \frac{P_{ti}}{P_{\text{max}}} T_{ti}, \quad (3.1)$$

where $TT_{LB1}$ is the lower bound on SoC test time. Let $\frac{P_{\text{max}}}{P_{ti}} = F_{ti}$, where $F_{ti}$ is the scaling factor by which the clock frequency of a test $t_i$ is varied with respect to the nominal value. This scaling factor shall be referred to as frequency factor for the remainder of this work. Hence, the total test time of an SoC is lowest when each core test is scheduled sequentially at a clock rate equal to $\frac{P_{\text{max}}}{P_{ti}} f_{\text{nom}}$.

**Theorem 3.1** Concurrent scheduling of core tests at a test clock rate $\frac{P_{\text{max}}}{P_{ti}} f_{\text{nom}}$, cannot improve the lower bound on the total test time of the SoC, obtained by the sequential test schedule at the same clock rate.

**Proof:** Let there be $n$ tests, $t_1, \cdots, t_n$. Let $T_{ti}$ and $P_{ti}$ be the test, $t_i$’s test time and power dissipated, respectively. Let $P_{\text{max}}$ be the power budget.

**Case 1:** Sequential test scheduling (One test scheduled per session).
Let each session $S_i$ contain a single test, $t_i$. This implies that the session’s test time and power are the same as that of the test, i.e., $T_{S_i} = T_{t_i} = T_i$ and $P_{S_i} = P_{t_i} = P_i$. Now, if the clock frequency is altered, for speeding up the testing, the frequency factor, $F_i = P_{\text{max}} / P_i$. The modified session test time, now, is $T_i / F_i$ or, $T_i P_i / P_{\text{max}}$. The total test time (say $TT_1$), therefore, is $\sum_{i=1}^{n} T_i P_i / P_{\text{max}}$ or,

$$TT_1 = (T_1 P_1 + \cdots + T_n P_n) / P_{\text{max}} \quad (3.2)$$

Case 2: Concurrent test scheduling (Multiple tests scheduled in each session).

Let the $n$ tests be scheduled in $k$ sessions such that every test is covered by exactly one test session. The test time and power of a session, $S_j$ are given by $T_{S_j} = \max\{T_i\}$ and $P_{S_j} = \sum(P_i), \forall t_i \in S_j$, respectively. If the clock frequency per session is varied, the frequency factor per session, $F_j = P_{\text{max}} / P_{S_j} = P_{\text{max}} / \sum P_i, \forall t_i \in S_j$. The modified session test time for session, $S_j$ is given by, $(\max\{T_i\} \cdot \sum P_i) / P_{\text{max}}, \forall t_i \in S_j$. The total test time (say $TT_2$), therefore, is $[\sum_{j=1}^{k} T_{S_j} (\sum P_i)] / P_{\text{max}}, \forall t_i \in S_j$ or,

$$TT_2 = [T_{S_1} (P_1 + \cdots + P_x) + \cdots + T_{S_k} (P_y + \cdots + P_n)] / P_{\text{max}}, \text{where } x, y \in \{1, \cdots, n\} \quad (3.3)$$

For any session, $S_j$, $T_{S_j} \geq T_i, \forall t_i \in S_j$, i.e, the session’s test time is always greater than or equal to the test times of the tests in that session. This implies that, if tests $t_x, t_y, t_z \in S_j$, then $T_{S_j} P_x + T_{S_j} P_y + T_{S_j} P_z \geq T_x P_x + T_y P_y + T_z P_z$. The LHS of this inequality resembles 3.3 and the RHS resembles that of 3.2. Hence, from this and 3.2 and 3.3, we can say that $TT_2 \geq TT_1$ and therefore, the total test time of concurrent scheduling is at most as small as the lower bound when the test clock rate is $P_{\text{max}} / P_{t_i} f_{\text{nom}}$.

The test power of a core test, characterized at nominal supply voltage ($V_{\text{nom}}$), is dependent on voltage. As $E \propto P \propto V^2$, the energy per core test and hence the total test energy also varies with the supply voltage. Therefore, the lower bound on test time given in
Equation 3.1 is applicable at the nominal supply voltage. The energy per core test is given by, $E_{ti} = P_{ti} \left( \frac{V_{min}}{V_{nom}} \right)^2 T_{ti}$, where $V_{min}$ is the lowest value to which the supply voltage can be reduced, without disrupting the circuit’s functionality. Hence, the new lower bound for the total test time of the SoC is,

$$TT_{LB2} = \frac{\sum_{i=1}^{n} P_{ti} \left( \frac{V_{min}}{V_{nom}} \right)^2 T_{ti}}{P_{max}},$$

where $TT_{LB2}$ is the new lower bound on SoC test time.

Note that here a constant $V_{min}$ is assumed for all SoC cores; however, the value of $V_{min}$ may vary among the cores.

The theorem showed that scheduling core tests at a clock frequency such that the power consumed per test is the same as the power budget, yields the lower bound and that concurrent test scheduling cannot improve this lower bound. This is under the assumption that the power dissipation of a core can be raised to equal the power budget without any physical limitation on the individual core power limit or the clock. In reality, the clock rate of individual cores is often limited by their structural constraints (for e.g., critical path delay) and power constraints (rated maximum power). Consequently, the maximum clock frequency of a session is decided by the maximum clock frequency of the slowest core in that session, i.e., $f(S_j) \leq \min\{f_{max}(t_i) | \forall t_i \in S_j\}$, where $f(S_j)$ is the clock rate of session $S_j$ and $f_{max}(t_i)$ is the maximum clock frequency of a test $t_i$. Since all cores of the SoC are tested at the nominal clock frequency, $f_{nom}$, it is valid to assume that $f_{nom}$ is the clock rate of the slowest core in the SoC (say $f_0$). Then, frequency factor of a session, $F_j = \frac{f(S_j)}{f_0}$. Note that the test session containing the slowest core of the SoC will possess a unity frequency factor. The maximum frequency factor is given by:

$$\max\{F_j\} = \min\left[ \frac{\min\{f_{max}(t_i) | \forall t_i \in S_j\}}{f_0}, \frac{P_{max}}{P_{S_j}} \right]$$

(3.5)
The structural and power constraints that limit a core’s maximum frequency are also influenced by the supply voltage. The power consumed \((P)\) varies in direct relation with supply voltage \((V_{DD})\) and clock frequency \((f)\),

\[
P \propto V_{DD}^2 \cdot f \quad (3.6)
\]

This implies that power consumption can be reduced by lowering the operating voltage. As a result, the clock rate can be increased without exceeding the power constraint of the core. However, the delay of a circuit also varies with the voltage, as given by the alpha power law \([46,50,51]\):

\[
delay \propto \frac{V_{DD}}{(V_{DD} - V_{TH})^\alpha}, \quad (3.7)
\]

where \(\alpha\) is the velocity saturation index.

The value of \(\alpha\) lies between 1 (for complete velocity saturation) and 2 (no velocity saturation) \([46,50,51]\). As seen from the above expression, reducing the voltage causes the delay to increase, which in turn, slows down the execution speed and hence, results in longer a test time. Thus, as we reduce \(V_{DD}\), on one hand, the lowered power consumption allows higher clock rates thereby shrinking the total test time and on the other hand, the increased circuit delay results in slower clock rate and a longer test time. Therefore, it is required to find an optimal \(V_{DD}\) that will allow us to balance the two trade-offs and at the same time achieve a test time reduction.

Let \(f_p\) and \(f_s\) be the frequency limits corresponding to the power and structural constraints of a core, respectively. The relationship given in Equation 3.6 can now be written as:

\[
P_{core} \propto V_{DD}^2 \cdot f_p
\]

where \(P_{core}\) and \(f_p\) are the power rating and the power constrained frequency limit for a core, respectively. Since the power rating for a core is a constant, the \(f_p - V_{DD}\) relation can be
rewritten as:

\[ f_p \propto 1/V_{DD}^2 \]  \hspace{1cm} (3.8)

The \( f_s - V_{DD} \) relationship can be expressed, in accordance with the alpha power law (Equation 3.7), as:

\[ f_s \propto \frac{(V_{DD} - V_{TH})^\alpha}{V_{DD}} \]  \hspace{1cm} (3.9)

From these expressions it can be noted that as \( V_{DD} \) is decreased, \( f_p \) increases allowing higher clock rates. At the same time, \( f_s \) decreases with decreasing \( V_{DD} \), thus restricting the clock rate. Both these constraint limits are independent of each other, i.e., the power constraint limit \( f_p \) is only decided by the rated power of the core, with no regards to the critical path of the core and similarly, the critical path of the core dictates the structure constraint limit, ignoring the rated power limit of the core. The maximum clock rate of a core, therefore, is the minimum of the two frequency limits. Now, the clock frequency of a session, which is the same as the slowest core in that session, is given by

\[ f(S_j) \leq \min \left\{ f_p(t_i), f_s(t_i) \right\} \forall t_i \in S_j \]

and since the frequency factor of a session, \( F_j = \frac{f(S_j)}{f_0} \), its maximum value is given by,

\[ \max \{F_j\} = \min \left\{ \frac{\min \left\{ f_p(t_i), f_s(t_i) \right\} \forall t_i \in S_j }{f_0}, \frac{P_{\text{max}}}{P_{S_j}} \right\} \]  \hspace{1cm} (3.10)

The lower bound for the SoC test time, defined in Equation 3.4, does not take into account, the structure constraint of the clock rate. As a result, the equation predicts that the test time continually reduces as \( V_{DD} \) is lowered. However, from Figure 2.5, we know that beyond an optimal \( V_{DD} \) point, the test time increases with decreasing \( V_{DD} \). Therefore, Equation 3.4 is revised to include the optimal voltage, \( V_{\text{opt}} \), instead of \( V_{\text{min}} \).

\[ TT_{LB2} = \sum_{i=1}^{n} P_{t_i} \left( \frac{V_{\text{opt}}}{V_{nom}} \right)^2 T_{t_i} / P_{\text{max}}, \]  \hspace{1cm} (3.11)

It can be noted that Equation 3.4 would be the same as Equation 3.11, when \( V_{\text{opt}} = V_{\text{min}} \).
Let us assume that \( f_s = k \cdot f_p \) at \( V_{nom} \). As \( V_{DD} \) is lowered, both \( f_s \) and \( f_p \) vary accordingly. At \( V_{opt} \), \( f_s(V_{opt}) = f_p(V_{opt}) \), i.e., \( f_s \cdot \frac{V_{nom}}{V_{opt}} \cdot \left( \frac{V_{opt} - V_{th}}{V_{nom} - V_{th}} \right)^\alpha = f_p \cdot \left( \frac{V_{nom}}{V_{opt}} \right)^2 \). Since \( f_s/f_p = k \),
\[
k = \frac{V_{nom}}{V_{opt}} \cdot \left( \frac{V_{nom} - V_{th}}{V_{opt} - V_{th}} \right)^\alpha \tag{3.12}
\]
The value of \( \alpha \) for recent short-channel MOSFETs is 1.3 [50]. For the sake of simplicity, let us assume the value of \( \alpha \) as 1. Now, Equation 3.12 can be written as, \( k(V_{opt})^2 - kV_{th}V_{opt} - V_{nom}(V_{nom} - V_{th}) = 0 \), which is of the form \( ax^2 + bx + c = 0 \). Solving for \( V_{opt} \),
\[
V_{opt} = \frac{kV_{th} \pm \sqrt{(kV_{th})^2 + 4kV_{nom}(V_{nom} - V_{th})}}{2k} \tag{3.13}
\]

### 3.2 Mixed-Integer Linear Program (MILP) Based Optimization

#### 3.2.1 Introduction

Linear Programming (LP) [28] is an optimization tool designed to achieve the best outcome in a mathematical model where the relationship among the factors involved in the model can be formulated as linear equalities or inequalities. Linear programming consists of a linear objective function that has to be optimized under restrictive conditions that are expressed as linear equalities or inequalities. The canonical form of LP problems is:

\[
\text{maximize } c^T x \\
\text{Subject to } Ax \leq b \text{ and } x \geq 0
\]

where \( c \) and \( b \) are vectors of constant coefficients and \( A \) is a matrix of pre-determined coefficients whereas \( x \) is a vector of variables (known as decision variables) whose values are to be determined.

Integer linear programming (ILP) is a special case of linear programming wherein all the variables are restricted to integers. ILP problems are NP-complete and hence, large optimization problems are intractable through this method. Similar to ILP, MILP (mixed integer linear programming) is also a special case of linear programming since it contains
a combination of integral and real-valued decision variable. MILP problems are also NP-complete and solving them can be cumbersome and time consuming. In the past, MILP based optimization techniques have been used for SoC test scheduling [8–10, 27, 29]. The MILP model presented in this section takes into account the influence of $V_{DD}$ and clock on the test time and test power and optimizes the overall test time of the SoC for a pre-defined peak power limit.

3.2.2 MILP Formulation

The mixed-ILP model for optimizing $V_{DD}$ and clock rate per test session is formulated in this section. The voltage range is divided into multiple steps of voltages and for each step, the test power and frequency limits (structure and power constraint) of each test session is pre-computed. Let $P_{ij}$ be the test power of $j^{th}$ session at $i^{th}$ voltage. Similarly, let $F_{s_{ij}}$ and $F_{p_{ij}}$ be the frequency limit imposed by the structure and power constraints, respectively, for the $j^{th}$ session at $i^{th}$ voltage. For each session, the $V_{DD}$ is chosen by a binary variable whereas the clock rate of the session is a real-valued variable. $T_j$ and $F_j$ are the test time and frequency factor of a test session. $x_{ij}$ is a binary variable that selects a test session and its optimal $V_{DD}$ among all possible test sessions and voltage steps. The test schedule optimization can then be described as follows:

**Objective:**

$$\text{Minimize } \sum_{i,j} \left( \frac{T_j}{F_j} \right) x_{ij},$$

where $x_{ij} = \begin{cases} 
1, & \text{if } j^{th} \text{ session is scheduled at } i^{th} \text{ voltage} \\
0, & \text{otherwise}
\end{cases}$

**Subject to:**

1. $F_j \cdot \sum_i (P_{ij} \cdot x_{ij}) \leq P_{max}$
2. \( \sum_i x_{ij} = 1 \)

3. a. \( F_j.x_{ij} \leq F_{s_{ij}} \) b. \( F_j.x_{ij} \leq F_{p_{ij}} \)

4. each test, \( t_i, i \in \{1,..,n\} \) is executed at least once.

The first constraint of the ILP ensures that the power consumption of the test session does not exceed the power budget. The second constraint specifies that each session should be associated with exactly one voltage value. The clock frequency is bounded by the structure constraint \((F_{s_{ij}})\) and the power constraint \((F_{p_{ij}})\) in the ILP’s third constraint. The fourth one is a test completeness constraint which ensures that all the core tests are scheduled.

The above non-linear model is linearized using simple substitutions. Let \( 1/F_j = u_j \) and \( u_j.x_{ij} = q_j, \forall i, j \). These substitutions necessitate the inclusion of two more constraints: 1) \( q_j \geq u_j - M(1 - x_{ij}), \forall i, j \), where \( M \) is a large number such that \( M >> u_j \), 2) \( q_j \geq 0 \). The new ILP formulation is as follows:

**Objective:**

\[
\text{Minimize } \sum_j T_j.q_j,
\]

**Subject to:**

1. \( \sum_i (P_{ij}.x_{ij}) \leq P_{max}.u_j \)

2. \( \sum_i x_{ij} = 1 \)

3. a. \( x_{ij} \leq F_{s_{ij}}.u_j \) b. \( x_{ij} \leq F_{p_{ij}}.u_j \)

4. each test, \( t_i, i \in \{1,..,n\} \) is executed at least once.

5. \( q_j \geq u_j - M(1 - x_{ij}), \forall i, j \)

6. \( q_j \geq 0 \)

Note that the voltage step size determines the precision of the solution. However, reducing the step size to enhance the precision would increase the number of variables in the formulation and hence render the problem intractable through ILP.
3.3 Heuristic Based Optimization

Integer Linear Programs are NP-hard, in general and are computationally expensive for large SoCs. The CPU time required to obtain an optimal solution increases exponentially as the number of cores and the complexity of the SoC increases. The proposed MILP method also shares the same issues with scalability in terms of the problem size. Hence, a simulated annealing (SA) based optimization technique is presented that is much faster than ILP for larger SoCs and also capable of producing results similar to that of the ILP method. Heuristic algorithms, often employing greedy approaches, perform much better in terms of CPU time as compared to exact methods such as ILP. While a heuristic method does not guarantee an optimal solution, a good algorithm can produce near-optimal values consistently. Many heuristic optimization approaches in the field of SoC testing have been published in the past [15, 19, 23, 35, 44, 57, 73].

3.3.1 Simulated Annealing (SA)

Simulated annealing is a directed search algorithm inspired from the annealing process in metallurgy, first proposed by Kirkpatrick et al. [30] has been used in the past for SoC test scheduling [23, 35, 73]. The algorithm accepts a non-improving solution with a finite probability so as to avoid getting stuck at a local optimum. The probability of accepting worse solutions is controlled by the temperature parameter \( T \). As the temperature of the process cools down, it becomes less and less likely for the algorithm to accept non-improving solutions. Theoretical studies on simulated annealing have shown that simulated annealing based algorithms converge to a global optimum with a probability of 1 under certain specified conditions on the updating and iteration of temperature values [61]. The overview of our SA based optimization algorithm is as shown in Figure 3.1.
Simulated Annealing Heuristic

\( T = \) temperature
\( K = \) cooling parameter
\( T_f = \) final temperature
\( X_B = \) best solution obtained so far
\( X_C = \) current solution

generate initial solution, \( X_0 \) (test schedule and corresponding test time)
\( X_B = X_0, X_C = X_0 \)

while \( T \geq T_f \) do
  perform SA move operation (swapping of tests) on \( X_C \)
  scale clock rate and voltage to optimize the new test schedule
  compute test time of the optimized test schedule, \( X_{new} \)
  \( Diff = X_{new} - X_C \)
  if \( Diff \leq 0 \) or \( \exp\left(\frac{-Diff}{KT}\right) \geq \) random \((0,1)\) then
    \( X_B = X_{new}, X_C = X_{new} \)
  else
    discard \( X_{new} \) and retain \( X_C \)
  end if

  \( T = K \times T \)
end while

Figure 3.1: Overview of the SA heuristic algorithm.

Initial solution

The initial solution is developed by inserting a randomly selected test into a session until the session’s power consumption (\( P_{ses} \)) is close to the peak power budget (\( P_{max} \)). This step is repeated until all the tests are grouped into sessions such that no two sessions contain the same test.

The test schedule, thus generated, serves as the starting point for the simulated annealing heuristic. Frequency and voltage scaling (described in Cost Calculation) are also applied to optimize the test time obtained through Figure 3.2.

SA move operator

The move operator in our simulated annealing algorithm is a swapping of tests between two sessions. Among the many test sessions of the test schedule, two sessions \( s_1 \) and \( s_2 \) are
**Initial Solution**

list1 = list of core tests to be scheduled {initially contains all core tests}
list2 = list of core tests currently executed {initially empty}
\( t_{sch} = 0 \) {overall test time of the test schedule}

while list1 is not empty do
    list2 = empty list
    while \( P_{ses} < P_{max} \) do
        insert random test \( i \) into list2
        delete test \( i \) from list1
        \( P_{ses} = \sum P_i \), \( \forall i \in list2 \)
    end while
    if \( P_{ses} > P_{max} \) then
        remove recently added test from list2
    end if
    \( t_{sch} = t_{sch} + \max (t_i, \forall i \in list2) \)
end while

Figure 3.2: Generating the initial solution for the SA algorithm.

selected at random, such that \( s_1 \neq s_2 \). A randomly chosen test in each session is swapped with each other, thus forming a new test schedule. The cost of the resultant solution, which is the test time of the new test schedule, is computed. The new test schedule is accepted if the new solution is better than the best solution obtained so far or if their difference (\( d \)) is such that \( \exp\left(\frac{d}{K T}\right) \geq \text{random}(0, 1) \), where \( K \) is the cooling parameter and \( T \) is the annealing temperature (described in Annealing Schedule), else the swap is discarded.

Simulated annealing is a neighborhood evaluation based class of algorithms where neighboring solutions are examined and accepted or discarded. The neighboring solutions, in this case, are obtained by swapping of the tests in the sessions. In the worst case, the number of sessions may be the same as the number of tests implying that each session will contain one unique test. Hence, the number of neighboring solutions for an SoC with ‘\( n \)’ core tests would be \( \frac{n(n-1)}{2} \).

**Cost calculation**

The cost in this optimization problem refers to the test time of the test schedule. The overall test time for the session-based test schedule is the sum of the test time of the longest
test in the session. The test clock frequency and the supply voltage are scaled to optimize the test time. The scaling factor for the clock, referred to as the frequency factor (F), is updated after addition of every test during the initial solution phase and after every swap, in the SA move operation phase. The frequency factor is limited by both the peak power budget and the clock rate constraints of individual core of the SoC, as given in Equation 3.5.

Voltage scaling is done for each session as given below:

- reduce $V_{\text{DD}}$ by one step.
- Calculate the power and structure constraint limits of the tests using Equations 3.8 and 3.9 respectively.
- Update the frequency factor using Equation 3.10.
- Repeat the steps if the resulting session test time is lower than before, else quit the voltage scaling procedure.

**Annealing schedule**

Annealing schedule refers to the temperature ($T$), the cooling parameter ($K$) and their effects on the optimization procedure. True to the metallurgical annealing process, the initial value of the temperature is high. Each iteration of the heuristic, which produces a new solution, corresponds to a value of the temperature. After each iteration, the temperature is reduced according to $T_{\text{new}} = K \times T$, where $K \leq 1$. Hence, the number of iterations is dependent on both, the temperature and the cooling parameter. The stopping criteria for the procedure would be the temperature value reducing below a certain specified limit.
3.4 Results

3.4.1 Experimental Setup

The exact and the heuristic optimization methods were experimented on several ITC'02 benchmarks [3] and ASIC Z. The ASIC Z was introduced by Y. Zorian in [72] and consists of RAM, ROM and other blocks. These blocks, along with their test time and power are shown in Figure 3.3. The peak power budget for the ASIC Z is given as 900 mW.

The test time and test power data for the ITC benchmark SoCs have been provided by Millican and Saluja [41]. The peak power budget for these SoCs were assigned based on the test power information of individual cores in the SoCs. To account for power and structure constrained limits on the frequency of individual cores, maximum clock rates are allocated for each core. The values for the power constrained limit ($f_p$) are computed based on the test length and test power of the blocks/cores. The block with the highest test power and longest time is regarded as the slowest and the rest of the cores in the SoC are
normalized with respect to the slowest core. Hence, the test with a low test power value can be clocked at a faster rate. For assigning the structure constraint limit \((f_s)\), the fact that the test power can be as high as four times the functional power is taken into account, i.e., \(P_{\text{func}} \leq P_{\text{test}} \leq 4P_{\text{func}}\). \(P_{\text{test}} \propto f_p\) and since the structure constraint limit decides the functional clock rate, \(P_{\text{func}} \propto f_s\). Hence, the structure constraint limit \((f_s)\) for each core is set to \(k \times f_p\), where \(k\) is a uniform random number generated in the range(1,4).

For illustration, the complete data of ASIC Z, including the frequency limits, is given in Table 3.1. This test data set for ASIC Z is specified at a nominal supply voltage. The test time, test power and the power budget were provided by Y. Zorian in [72]. The frequency constraints for each block were derived by the steps described previously. Similarly, the test data for the remaining benchmarks is given in the Appendix section.

Further, the range of operating voltage, in this work, is assumed to be between 1.0V (nominal) and 0.6V (minimum). The other parameters for the alpha-power law, namely,

<table>
<thead>
<tr>
<th>Block</th>
<th>Test time</th>
<th>Test power</th>
<th>(f_p^+)</th>
<th>(f_s^{++})</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAM1</td>
<td>69</td>
<td>282</td>
<td>1.75</td>
<td>6.65</td>
</tr>
<tr>
<td>RAM2</td>
<td>61</td>
<td>241</td>
<td>2</td>
<td>7.55</td>
</tr>
<tr>
<td>RAM3</td>
<td>38</td>
<td>213</td>
<td>3</td>
<td>5.6</td>
</tr>
<tr>
<td>RAM4</td>
<td>23</td>
<td>96</td>
<td>5</td>
<td>8.8</td>
</tr>
<tr>
<td>ROM1</td>
<td>102</td>
<td>279</td>
<td>1.5</td>
<td>4.6</td>
</tr>
<tr>
<td>ROM2</td>
<td>102</td>
<td>279</td>
<td>1.5</td>
<td>3.83</td>
</tr>
<tr>
<td>*RL1</td>
<td>134</td>
<td>295</td>
<td>1.2</td>
<td>2.74</td>
</tr>
<tr>
<td>*RL2</td>
<td>160</td>
<td>352</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>**RF</td>
<td>95</td>
<td>10</td>
<td>8</td>
<td>17.6</td>
</tr>
</tbody>
</table>

\(^+\)power constraint \(^{++}\)structure constraint
*Random Logic **Register File
$V_{th}$ and $\alpha$ are assumed to be 0.5V and 1.0. These values are in tune with the 45-nm technology [71]. In [59], Tran and Baas show the operation of a 32-bit adder, designed in 45-nm technology node, for a range of $V_{DD}$, starting from 1.0V all the way down to 0.1V. The authors note that the operation of the circuit enters sub-threshold region below 0.5V. Other related work have reported the functioning of memory and logic circuits, for sub-70nm technology, at voltages as low as 0.6V (non sub-threshold operation) [6,49,66]. Keeping this in mind, let us revisit the lower bound on SoC test time. As mentioned earlier, the structure constraint of the SoC cores’ clock rate limits the scaling of $V_{DD}$. Assuming the least restrictive condition on the structure constraint, we have $f_s = 4f_p$. Substituting $V_{nom} = 1.0V, V_{th} = 0.5V$ and $k = 4$ in Equation 3.13 yields a $V_{opt} \approx 0.69V$. This value of $V_{opt}$ can be used in Equation 3.11 to derive the lower bound on the test time of the SoC benchmarks considered in this work.

The experiments were performed on a Dell workstation with a 3.4GHz Intel Pentium processor and 2GB memory. The MILP models were solved using IBM CPLEX Optimization Solver (student edition) whereas the SA based heuristic algorithm was developed using the Python scripting language [4].

### 3.4.2 MILP Results

The results for the proposed MILP method are presented in this section. In order to evaluate the optimization results, three optimization cases are considered. The first one, referred to as Case 1, is the nominal case where the $V_{DD}$ and the test clock are fixed at a nominal value. In the second case (Case 2), the $V_{DD}$ is fixed at a nominal value but the clock frequency is optimized per test session [55]. Finally, in Case 3, both $V_{DD}$ and the clock are optimized [56].

Let us consider the ASIC Z system. Previously published optimal test times for the ASIC Z include 392 by Zorian [72], 330 by Chou et al. [13, 14] and 300 by Larsson and Peng [35]. For the nominal case (Case 1), the test schedule and test time (300 units) are similar to the one obtained by Larsson and Peng [35]. Customizing the test clock per session
Table 3.2: Optimized Test Schedule for ASIC Z for nominal and custom clock rate (Cases 1 and 2). The supply voltage is at nominal value for both cases.

<table>
<thead>
<tr>
<th>Session</th>
<th>Block</th>
<th>Test time</th>
<th>Session</th>
<th>Block</th>
<th>Freq. factor</th>
<th>Test time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RL1, RL2, RAM2</td>
<td>160</td>
<td>1</td>
<td>RAM1, ROM2</td>
<td>1.5</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>RAM1, ROM1, ROM2</td>
<td>102</td>
<td>2</td>
<td>RAM2, RAM3</td>
<td>1.98</td>
<td>30.77</td>
</tr>
<tr>
<td>3</td>
<td>RAM1, RAM4, RF</td>
<td>38</td>
<td>3</td>
<td>RAM4, RF</td>
<td>4.71</td>
<td>4.88</td>
</tr>
<tr>
<td></td>
<td>ROM1, RL1, RL2</td>
<td></td>
<td>4</td>
<td>ROM1, RL1, RL2</td>
<td>0.97</td>
<td>164.624</td>
</tr>
</tbody>
</table>

Total test time = 300

Total test time = 268.274

(Case 2) reduces the test time by 10.5% (Table 3.2). The frequency factor in the table indicates the speed-up of the clock, done to reduce the test time. A frequency factor of 1.5 implies that the test clock frequency of that session was increased to 1.5 times the nominal value. The lower bound on the overall test time for ASIC Z at nominal $V_{DD}$, calculated using Equation 3.1, is 220.2 units. The difference between the lower bound and the test time at nominal clock rate and voltage (case 1) is 26.6% and the difference between the lower bound and the test time for optimization case 2 (customized test clock frequency) is 17.9%. One can observe that by customizing the clock rate, the test scheduling result moves closer to the lower bound but is constrained by the maximum clock rate of individual cores.

Table 3.3 shows, however, that customizing both $V_{DD}$ and the test clock (Case 3) lowers the test time by as much as 50%. It can also be noted from the table that the sessions in the schedule not only have different clock rates but also different $V_{DD}$ (which is the optimum $V_{DD}$ for that session). The lower bound in Equation 3.4, calculated at $V_{min} = 0.6V$ is 79.27 units. The difference between this lower bound and the optimal test time as seen in Table 3.3 is 46.5%. The test time from optimization case 3 deviates from the lower bound as the optimal $V_{DD}$ for each session in the test schedule is higher than $V_{min}$. 

28
Table 3.3: Test times (in arbitrary units) of ASIC Z system for custom $V_{DD}$ and clock rate (Case 3).

<table>
<thead>
<tr>
<th>Session</th>
<th>Block</th>
<th>Freq. factor</th>
<th>$V_{DD}$</th>
<th>Test time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RF</td>
<td>12.5</td>
<td>0.8V</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>RAM1, RAM2, RAM3, RAM4</td>
<td>2.56</td>
<td>0.65V</td>
<td>26.95</td>
</tr>
<tr>
<td>3</td>
<td>ROM1, ROM2, RL1, RL2</td>
<td>1.33</td>
<td>0.75V</td>
<td>120.5</td>
</tr>
</tbody>
</table>

Total Test time = 148.25

Table 3.4: Test times (in arbitrary units) for benchmark SoCs, obtained by MILP optimization method for the three optimization cases considered.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>No. of cores</th>
<th>$P_{max}$ (mW)</th>
<th>Case 1 Test time</th>
<th>Case 2 Test time</th>
<th>Case 3 Test time</th>
<th>% Reduction over Case 1</th>
<th>% Reduction over Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>a586710</td>
<td>7</td>
<td>800</td>
<td>14271856</td>
<td>13011130.61</td>
<td>6799115.12</td>
<td>52.36</td>
<td>47.74</td>
</tr>
<tr>
<td>h953</td>
<td>8</td>
<td>800</td>
<td>122636</td>
<td>121715.34</td>
<td>79318.76</td>
<td>35.32</td>
<td>34.84</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>9</td>
<td>900</td>
<td>300</td>
<td>268.274</td>
<td>148.25</td>
<td>50.58</td>
<td>44.74</td>
</tr>
<tr>
<td>d695</td>
<td>10</td>
<td>400</td>
<td>15188</td>
<td>12733.2</td>
<td>7173</td>
<td>52.77</td>
<td>43.67</td>
</tr>
</tbody>
</table>

Case 1 = $V_{DD}$ and test clock fixed at nominal value; Case 2 = $V_{DD}$ fixed at nominal value, clock scaled per test session; Case 3 = $V_{DD}$ and clock scaled per test session.

Similarly, the optimized test times for the benchmarks for the various optimization cases considered, is tabulated in Table 3.4. The percent reduction specified in the last two columns of the table refer to the reduction in test time achieved by case 3 ($V_{DD}$ and clock scaling) with respect to the other two optimization cases. For instance, in case of ASIC Z, the test time for the optimization case 3 is about 50% lower than that of case 1 (fixed $V_{DD}$ and clock) and 45% lower than case 2 (only frequency scaling). From the table it can be noted that by customizing both voltage and frequency can reduce the test time in half.

The plot in Figure 3.4 shows the CPU time of the MILP optimization method. As seen from the plot, optimization through frequency and voltage scaling consumes most CPU time and also the run time grows very quickly with the SoC size.
Figure 3.4: CPU time (in seconds) of the MILP optimization method for the ITC benchmarks.

Table 3.5: Test times (in arbitrary units) for MILP and heuristic test scheduling methods, with customized $V_{DD}$ and clock rates.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>SA based heuristic method</th>
<th>MILP method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test time</td>
<td>CPU time</td>
</tr>
<tr>
<td>a586710</td>
<td>6799118.34</td>
<td>0.12 sec</td>
</tr>
<tr>
<td>h953</td>
<td>79319.12</td>
<td>0.09 sec</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>150.26</td>
<td>0.11 sec</td>
</tr>
<tr>
<td>d695</td>
<td>7177.53</td>
<td>0.17 sec</td>
</tr>
</tbody>
</table>

3.4.3 Heuristic Algorithm Results

A comparison of optimized test times obtained from the MILP and the SA based test scheduling algorithm is provided in Table 3.5. Since the heuristic can be dependent on the initial solution, the algorithm is repeated for hundred starting points and the best solution among them is selected. The CPU time of the algorithm is averaged over the hundred simulations. As seen from the table, the difference between the heuristic solution and the exact solution is marginal. The table also shows that the CPU time for the heuristic does not vary much with respect to the SoC size.
Table 3.6: Test times (in arbitrary units) for benchmark SoCs, obtained by the heuristic optimization method for the three optimization cases considered.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>No. of cores</th>
<th>$P_{\text{max}}$ (mW)</th>
<th>Case 1 Test time</th>
<th>Case 2 Test time</th>
<th>Case 3 Test time</th>
<th>% Reduction over Case 1</th>
<th>% Reduction over Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1023</td>
<td>14</td>
<td>400</td>
<td>21245</td>
<td>19888.7</td>
<td>12193.05</td>
<td>42.6</td>
<td>38.7</td>
</tr>
<tr>
<td>p34392</td>
<td>19</td>
<td>400</td>
<td>952199</td>
<td>758199.76</td>
<td>369921.1</td>
<td>61.17</td>
<td>51.24</td>
</tr>
<tr>
<td>t512505</td>
<td>31</td>
<td>400</td>
<td>5589002</td>
<td>5414047.16</td>
<td>3038172.5</td>
<td>45.64</td>
<td>43.88</td>
</tr>
<tr>
<td>p93791</td>
<td>32</td>
<td>400</td>
<td>178568</td>
<td>160618.71</td>
<td>90391.8</td>
<td>49.38</td>
<td>43.72</td>
</tr>
<tr>
<td>R100</td>
<td>100</td>
<td>900</td>
<td>1347</td>
<td>1213.56</td>
<td>730.4</td>
<td>45.77</td>
<td>39.81</td>
</tr>
<tr>
<td>R200</td>
<td>200</td>
<td>900</td>
<td>2837</td>
<td>2502.29</td>
<td>1536.35</td>
<td>45.84</td>
<td>38.60</td>
</tr>
<tr>
<td>R500</td>
<td>500</td>
<td>900</td>
<td>7706</td>
<td>6653.01</td>
<td>4212.27</td>
<td>45.34</td>
<td>36.68</td>
</tr>
</tbody>
</table>

To emphasize this point, the heuristic methods was employed to solve the test scheduling problem for larger ITC benchmarks, for which the MILP solver would struggle to provide a solution. In order to further evaluate the performance of the heuristic, SoCs with 100, 200 and 500 cores (referred to from now on as R100, R200 and R500, respectively) were created. The test time and test power data for the R100 SoC was generated using a uniform random number generator, in the range (10, 100) and (50, 500) respectively. The R200 and the R500 are multiple copies of the R100 SoC. The peak power budget for these SoCs was set to 900mW. Table 3.6 summarizes the optimized test times obtained through the heuristic method for these SoCs. From the table, it can be noted that the heuristic method also achieves a test time reduction of up to 60%.

The CPU time for the heuristic optimization is plotted in Figure 3.5. As seen from the table, the heuristic algorithm is able provide an optimized test schedule for the 500 core SoC in just over 6 seconds.

### 3.4.4 Lower Bounds on SoC Test Time

Section 3.1 introduced two lower bounds on the SoC test time; one applicable at nominal value and the other at the optimum point of the supply voltage. In Table 3.7, the SoC test
time for optimization cases 1 and 2 (nominal and custom clock rate at nominal V<sub>DD</sub>) is compared with the lower bound on test time at nominal V<sub>DD</sub> (Equation 3.1). From the table one can observe that, as the test clock rate is scaled, the optimal test time moves closer to the lower bound but this progression is hindered by limits on the individual clock rates of the SoC cores. It can be noted from Table 3.7 that, for benchmarks h953 and t512505, the difference between the lower bound and the optimal test time is much larger than the rest of the benchmarks. This because, from the theorem, we know that the lower bound on test time is reached by scaling the test clock at the rate P<sub>max</sub>/P<sub>test</sub>. For some of the cores in these two benchmarks, this ratio is as large as 2000. The individual clock constraints, however, are not as high as the ratio, P<sub>max</sub>/P<sub>test</sub>. As a result, there is a marked difference between the lower bound and the optimal test times for these two benchmarks.

The lower bound on SoC test time defined by Equation 3.1 does not apply for optimization case 3, since the supply voltage is also scaled along with the clock rate and the lower bound on the scaling of V<sub>DD</sub> would be V<sub>opt</sub>. The results for optimization case 3 (both V<sub>DD</sub> and clock scaled per test session) are compared with the lower bound computed at
Table 3.7: Optimal test times (in arbitrary units) for nominal and custom clock rate (Cases 1 and 2) compared with the lower bound on test time at nominal $V_{DD}$.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>No. of cores</th>
<th>$P_{max}$ (mW)</th>
<th>Lower Bound (LB)*</th>
<th>Heuristic optimization test times for</th>
<th>% Difference from LB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case 1</td>
<td>Case 2</td>
</tr>
<tr>
<td>a586710</td>
<td>7</td>
<td>800</td>
<td>11476950.1</td>
<td>14271856</td>
<td>13011130.61</td>
</tr>
<tr>
<td>h953</td>
<td>8</td>
<td>800</td>
<td>41511.06</td>
<td>122636</td>
<td>121715.34</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>9</td>
<td>900</td>
<td>220.2</td>
<td>300</td>
<td>268.274</td>
</tr>
<tr>
<td>d695</td>
<td>10</td>
<td>400</td>
<td>9193.4</td>
<td>15188</td>
<td>12733.2</td>
</tr>
<tr>
<td>g1023</td>
<td>14</td>
<td>400</td>
<td>11400.31</td>
<td>21245</td>
<td>19888.7</td>
</tr>
<tr>
<td>p34392</td>
<td>19</td>
<td>400</td>
<td>516245.20</td>
<td>952199</td>
<td>758199.76</td>
</tr>
<tr>
<td>t512505</td>
<td>31</td>
<td>400</td>
<td>1587297.02</td>
<td>5589002</td>
<td>5414047.158</td>
</tr>
<tr>
<td>p93791</td>
<td>32</td>
<td>400</td>
<td>121480.20</td>
<td>178568</td>
<td>160618.71</td>
</tr>
<tr>
<td>R100</td>
<td>100</td>
<td>900</td>
<td>1132.26</td>
<td>1347</td>
<td>1213.56</td>
</tr>
<tr>
<td>R200</td>
<td>200</td>
<td>900</td>
<td>2264.52</td>
<td>2837</td>
<td>2502.29</td>
</tr>
<tr>
<td>R500</td>
<td>500</td>
<td>900</td>
<td>5661.3</td>
<td>7706</td>
<td>6653.01</td>
</tr>
</tbody>
</table>

*Lower Bound calculated at nominal $V_{DD}$, by Equation 3.1; Case 1 = $V_{DD}$ and test clock fixed at nominal value; Case 2 = $V_{DD}$ fixed at nominal value, clock scaled per test session.

$V_{opt} = 0.69V$ (Equation 3.11) in Table 3.8. The difference between the lower bound and the optimal test time can be attributed to the fact that while calculating the optimal $V_{DD}$ point, a least restrictive condition was assumed for the structure constraint. This, however, is not the case for all cores of the SoC and therefore, $V_{opt}$ for such cores will be higher than the calculate value of 0.69V.

Once again, one can notice that there is a large gap between the lower bound and the optimal test times for benchmarks h953 and t512505, which could not be bridged by voltage scaling.

### 3.4.5 SoC Power Budget

The optimization techniques proposed in this work increase the test power consumption close to the power budget. While this strategy may not come across as a low-power testing
Table 3.8: Optimal test times (in arbitrary units) for custom $V_{\text{DD}}$ and clock frequency (Optimization case 3) compared with the lower bound on test time at $V_{\text{min}}$.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>No. of cores</th>
<th>$P_{\text{max}}$ (mW)</th>
<th>Lower Bound(LB)*</th>
<th>Optimal test time</th>
<th>% Difference from LB</th>
</tr>
</thead>
<tbody>
<tr>
<td>a586710</td>
<td>7</td>
<td>800</td>
<td>5464175.96</td>
<td>6799115.12</td>
<td>19.63</td>
</tr>
<tr>
<td>h953</td>
<td>8</td>
<td>800</td>
<td>19763.41</td>
<td>79318.76</td>
<td>75.08</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>9</td>
<td>900</td>
<td>104.83</td>
<td>148.25</td>
<td>30.23</td>
</tr>
<tr>
<td>d695</td>
<td>10</td>
<td>400</td>
<td>4376.98</td>
<td>7173</td>
<td>39.02</td>
</tr>
<tr>
<td>g1023</td>
<td>14</td>
<td>400</td>
<td>5427.69</td>
<td>12193.05</td>
<td>55.48</td>
</tr>
<tr>
<td>p34392</td>
<td>19</td>
<td>400</td>
<td>245784.34</td>
<td>369692.1</td>
<td>33.5</td>
</tr>
<tr>
<td>t512505</td>
<td>31</td>
<td>400</td>
<td>755712.11</td>
<td>3038172.5</td>
<td>75.12</td>
</tr>
<tr>
<td>p93791</td>
<td>32</td>
<td>400</td>
<td>57836.72</td>
<td>90391.77</td>
<td>36.01</td>
</tr>
<tr>
<td>R100</td>
<td>100</td>
<td>900</td>
<td>539.07</td>
<td>730.40</td>
<td>26.2</td>
</tr>
<tr>
<td>R200</td>
<td>200</td>
<td>900</td>
<td>1078.14</td>
<td>1536.34</td>
<td>29.82</td>
</tr>
<tr>
<td>R500</td>
<td>500</td>
<td>900</td>
<td>2695.35</td>
<td>4212.27</td>
<td>36.01</td>
</tr>
</tbody>
</table>

*Lower Bound calculated at $V_{\text{opt}} = 0.69V$, by Equation 3.11.

method, it can be noted that by controlling the power budget, one can choose to make savings in the test power. There will, however, always be a trade-off between the test time and the test power. This phenomenon is evident in Table 3.9, which gives the optimum test time for ASIC Z for power budgets. As seen from the table, lower value of $P_{\text{max}}$ increases the test time whereas a higher value reduces the test time. However, the percent reduction in test time for the different power budgets is similar.

Table 3.9: Optimized test times (in arbitrary units) for ASIC Z, for various power budget values.

<table>
<thead>
<tr>
<th>$P_{\text{max}}$</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>% Reduction Case 1</th>
<th>% Reduction Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>600mW</td>
<td>434</td>
<td>347.21</td>
<td>194.23</td>
<td>55.25</td>
<td>44.06</td>
</tr>
<tr>
<td>900mW</td>
<td>300</td>
<td>268.27</td>
<td>148.25</td>
<td>50.583</td>
<td>44.74</td>
</tr>
<tr>
<td>1200mW</td>
<td>262</td>
<td>207.6</td>
<td>131.1</td>
<td>49.96</td>
<td>36.85</td>
</tr>
</tbody>
</table>
3.4.6 Multiple Supply Voltages

Modern SoCs are typically heterogeneous and may consist of mixed-signal circuits, logic and memory blocks, each of which may function at separate voltages and clock frequencies. For instance, the analog and mixed-signal cores usually belong to much older semiconductor technologies and operate at higher voltages compared to the memory blocks, which often operate at voltages much less than 1V and are of the latest semiconductor technology. This would mean that the SoC cannot be tested at a single \( V_{DD} \) point. However, the optimization model presented in this work is able to take the various voltage ranges of the cores into account and find the optimum in each case. To demonstrate this, two voltage ranges are considered; 1. \([1.5V, 1.2V]\) with nominal \( V_{DD} = 1.5V \) and 2. \([1.0V, 0.6V]\) with nominal \( V_{DD} = 1.0V \). Each core of the ASIC Z benchmark is assigned to one of the two ranges. The non-overlapping voltage ranges place an additional restriction that cores with different \( V_{DD} \) range cannot be tested concurrently. The test schedule for ASIC Z, along with the optimal voltages, is given in Table 3.10. As seen from the table, while the tests are grouped into sessions, the test sessions are grouped according to their voltage ranges.

Table 3.10: Test times (in arbitrary units) and optimal voltages of ASIC Z system, for dual voltage ranges.

<table>
<thead>
<tr>
<th>Voltage range</th>
<th>Test session</th>
<th>Freq. factor</th>
<th>Optimal ( V_{DD} )</th>
<th>Test time</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.5V, 1.2V)</td>
<td>RF</td>
<td>12.5</td>
<td>1.2V</td>
<td>0.8</td>
</tr>
<tr>
<td>nominal</td>
<td>RAM2, ROM1,RL2</td>
<td>1.33</td>
<td>1.3V</td>
<td>120.17</td>
</tr>
<tr>
<td>= 1.5V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.0V, 0.6V)</td>
<td>RAM3, RAM4</td>
<td>5.19</td>
<td>0.7V</td>
<td>7.31</td>
</tr>
<tr>
<td>nominal</td>
<td>RAM1, ROM2, RL1</td>
<td>1.72</td>
<td>0.75V</td>
<td>77.83</td>
</tr>
<tr>
<td>= 1.0V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Total test time} = 206.12 \]
Chapter 4
Optimization of Sessionless Test Schedules

As discussed earlier, in sessionless testing, tests are scheduled, not in sessions, but simultaneously one test after another. As a result, sessionless test scheduling often has test time that is at least equal to, but often better than, that of session-based scheduling. In this section, a test optimization algorithm for sessionless test scheduling is proposed, which is a heuristic approach very similar to that of the session-based test scheduling, in that, it also based on a simulated annealing algorithm. The optimization algorithm employs voltage and frequency scaling, and can provide solution to both preemptive and non-preemptive scheduling schemes. Since only a single clock and $V_{DD}$ input is assumed, tests that are scheduled together have the same clock rate and $V_{DD}$. As a result, now the frequency factor corresponds to a clock scaling factor for sets of test scheduled concurrently. The lower bound on test time, provided by Equation 3.4, is valid for sessionless test schedules as well.

4.1 Heuristic Approach to Optimization

The initial solution and the SA move operator of this method remains the same as that of the heuristic for session-based testing. However, after the swap move, session boundaries in the new test schedule are erased and consecutive test sessions are merged together to form a sessionless test schedule. The cost of the resultant solution is determined; this solution is accepted if the new solution is better than the best solution obtained so far, or if their difference ($d$) is such that $\exp(-\frac{d}{kT}) \geq \text{random}(0, 1)$.

To compute the test time of the sessionless test schedule, firstly, consecutive test sessions in the test schedule resulting from the swap move are merged together by scheduling tests from the next session as soon as a test in the current session completes, as illustrated in
Figure 4.1: ‘Merging’ test sessions to convert a session based test schedule into a sessionless test schedule.

Figure 4.1. This process of ‘merging’ sessions is repeated until all tests are scheduled. The function \textit{Merge} is described in Figure 4.2. The test session that will be merged with its predecessor is passed as an argument to the \textit{Merge} function. The tests in this test session are added to the sessionless test schedule as and when the tests in the previous test session complete. In case of non-preemptive strategy, tests currently scheduled are run to completion and new tests are added from the next session as the tests that are currently scheduled, end. In case of preemptive scheduling strategy, on the completion of a test, the remaining tests that are yet to complete are preempted. A preemption implies that the tests are suspended and the remainder of the tests are treated as new tests to be scheduled later. The ‘new’ tests are included in the next session along with the tests that are already scheduled in that session.

The test clock frequency and the supply voltage are scaled for every set of concurrently scheduled tests to optimize the test time. However, the clock rate and the voltage for concurrently scheduled tests remain constant until the completion of a test; the frequency and voltage scaling is performed after the completion of every test, unlike the session-based test optimization method where the frequency and voltage are scaled after every test session. The annealing schedule remains the same as that for session-based test optimization algorithm.
**Merge(session)**

\[ \text{slsch} = \text{sessionless test schedule \{} \text{initially empty} \} \]

if \( \text{slsch} \) is empty then
   add all tests in the session to \( \text{slsch} \)
else
   while \( \text{session} \) not empty do
      if test in \( \text{slsch} \) completes then
         select a test from session and add to \( \text{slsch} \)
         \[ P = \sum P_i, \forall i \in \text{slsch} \]
         if \( P > P_{\text{max}} \) then
            remove the added test from \( \text{slsch} \)
         end if
      end if
   end while
end if

Figure 4.2: The ‘Merge’ function erases the session boundaries in a session based test schedule and combines the tests together to form a sessionless test schedule.

### 4.2 Optimization Results

The experimental setup including the benchmarks for sessionless test optimization remains the same as that of the session-based test optimization. For comparison with voltage and frequency scaled schedules, an algorithm to generate reference sessionless schedules with voltage and frequency fixed at nominal values is provided. The test scheduling process is modeled as a bin packing problem. An individual core test is treated as a block with test power as height and test time as width. A best-fit decreasing (BFD) heuristic then solves the bin packing problem. The tests are sorted in decreasing order of their power consumption and stacked together in such a way that at any given time in the test schedule the total power does not exceed a specified \( P_{\text{max}} \). The algorithm is provided in Figure 4.3.

The procedure in Figure 4.3 first sorts the list of unscheduled core tests in the decreasing order of their test power. Next, each test from this list is ‘scheduled’ by relocating it to a new list. This new list contains tests that are currently running. This step is repeated until the total test power is as close to the power limit as possible. After the completion of a test, a new test is added to the schedule from the sorted list. This whole process is repeated until
BFD Heuristic
list1 = list of core tests to be scheduled {initially contains all core tests}
list2 = list of core tests currently executed {initially empty}
t_{sch} = 0 {overall test time of the test schedule}
list1.sort(key = power, reverse = True)
while list1 is not empty or list2 is not empty do
    for each test i in list1 do
        if $P < P_{max}$ then
            insert test i into list2
            delete test i from list1
            $P = \sum P_i, \forall i \in list2$
        else
            remove recently added test from list2
        end if
    end for
    $t_{sch} = t_{sch} + \min(t_i, \forall i \in list2)$
    delete the test with smallest test length from list2
    for all remaining tests in list2 do
        update test length
    end for
end while

Figure 4.3: Best-fit decreasing (BFD) algorithm for sessionless test scheduling. Test schedules obtained from this algorithm are used as reference cases in this paper where voltage and frequency are fixed at their nominal values for the entire schedule.

all core tests are scheduled. The end time of the final test is the total test time of the test schedule.

As scaling voltage and frequency alters the test time and power of a core test, clock and supply scaled test schedules cannot be modeled as a bin packing problem. Hence, the SA based heuristic algorithm is adopted. Test times obtained for the benchmarks, for both preemptive and non-preemptive scheduling, are given in Table 4.1. Reference cases (column 4) are the fixed nominal voltage and clock frequency schedules obtained from the algorithm of Figure 4.3. This algorithm has no randomization elements in it and hence requires only one iteration. The heuristic, however, has some randomization and can be dependent on the initial solution. Therefore, the algorithm is repeated for hundred starting points and the best solution among them is selected. The CPU time of the algorithm is averaged over the hundred simulations. In each iteration, the starting point is a test
Figure 4.4: Convergence of the SA based optimization algorithm. The plot shows the heuristic algorithm converging towards the optimum test time as the temperature parameter \( T \) reduces.

scheduled with test sessions which are then merged to yield a sessionless test schedule. As the temperature reduces the algorithm moves from one feasible solution to another, with every new solution being better than the previous one. Occasionally, based on a finite probability, a ‘worse-than-previous’ solution is accepted to avoid saturation at the local optima. This probability is much lower at lower temperatures. The results from one of the iterations are plotted in Figure 4.4. The plot shows the convergence of the algorithm for both the preemptive and non-preemptive test optimization of the ITC benchmark \( p93791 \).

As seen from the plot, the initial solution for both preemptive and non-preemptive schedules is greater than 90000. However, as the temperature reduces, the quality of the solution improves and the test time moves closer to its optimal value. It must be noted that a single iteration was randomly chosen and plotted in the Figure 4.4 and hence, the final test time seen in the plot is not the best solution obtained for that benchmark.

As in the case of session-based test scheduling, the heuristic method of optimization is tested on the R100, R200 and R500 SoCs in order to further evaluate the performance of the heuristic algorithm. Table 4.1 summarizes the optimized test times obtained through
Table 4.1: Test times (in arbitrary units) for sessionless test scheduling with voltage and frequency scaling.

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>Test time without scaling&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Non-preemptive testing</th>
<th>Preemptive testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test time</td>
<td>% Reduction&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Test time</td>
</tr>
<tr>
<td>a586710</td>
<td>14090716</td>
<td>5797578.6</td>
<td>58.85</td>
</tr>
<tr>
<td>h953</td>
<td>122636</td>
<td>60805.62</td>
<td>50.42</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>262</td>
<td>137.85</td>
<td>47.38</td>
</tr>
<tr>
<td>d695</td>
<td>13301</td>
<td>5210.05</td>
<td>60.83</td>
</tr>
<tr>
<td>g1023</td>
<td>18084</td>
<td>8898.82</td>
<td>50.79</td>
</tr>
<tr>
<td>p34392</td>
<td>701684</td>
<td>279570.6</td>
<td>60.15</td>
</tr>
<tr>
<td>t512505</td>
<td>5344747</td>
<td>2940986.25</td>
<td>44.97</td>
</tr>
<tr>
<td>p93791</td>
<td>139008</td>
<td>68638.25</td>
<td>50.62</td>
</tr>
<tr>
<td>R100</td>
<td>1208</td>
<td>625.83</td>
<td>48.2</td>
</tr>
<tr>
<td>R200</td>
<td>2366</td>
<td>1337.4</td>
<td>43.47</td>
</tr>
<tr>
<td>R500</td>
<td>5807</td>
<td>3497.6</td>
<td>39.8</td>
</tr>
</tbody>
</table>

<sup>1</sup> Test time at fixed voltage and frequency, obtained from Best-Fit Decreasing algorithm (Figure 4.3).

<sup>2</sup> Percent reductions are with respect to the reference case.

the heuristic method for these SoCs. From Table 4.1, one can notice that by scaling the voltage and frequency dynamically the test time can be shortened by 45-60%. One can also observe that the preemptive and non-preemptive strategies yield almost identical solutions. This because, even though the preemptive scheme may enhance concurrency by partitioning tests, the partitioned tests have the same clock scaling factor and the same limits on the clock rate as the original test. This means that the behavior of the test time with respect scaling of voltage and frequency, in the preemptive scheme will be very similar to that of the non-preemptive case. Therefore, the difference between the test time for the cases is marginal. Also, from the perspective of the optimization method, as preemption progressively increases the number of tests to be scheduled, this causes the solution space to widen immensely, at the same time increasing chances of local optima saturation. This phenomenon is more pronounced in larger SoCs. One can observe, in Table 4.1, that for larger benchmarks the non-preemptive scheme marginally outperforms the preemptive scheme.
Table 4.2: CPU times* for all the benchmarks SoCs for the heuristic optimization algorithm.

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>CPU time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-preemptive</td>
</tr>
<tr>
<td>a586710</td>
<td>0.27</td>
</tr>
<tr>
<td>h953</td>
<td>0.265</td>
</tr>
<tr>
<td>d695</td>
<td>0.46</td>
</tr>
<tr>
<td>g1023</td>
<td>0.63</td>
</tr>
<tr>
<td>p34392</td>
<td>1.00</td>
</tr>
<tr>
<td>t512505</td>
<td>2.00</td>
</tr>
<tr>
<td>p93791</td>
<td>2.02</td>
</tr>
<tr>
<td>R100</td>
<td>4.68</td>
</tr>
<tr>
<td>R500</td>
<td>23.22</td>
</tr>
</tbody>
</table>

*CPU time averaged over 100 iterations of the heuristic.

Although the preemptive and non-preemptive strategies yield identical test times, they differ, slightly, in their run time (CPU time). This is because after the completion of each test, the preemptive strategy introduces extra complexity in the scheduling process by adding the preempted tests as new tests to the list of unscheduled core tests. With more tests being added to the scheduling list due to preemption, the number of while loops executed in the heuristic increases as do the calls to the voltage scaling function. The combined effect leads to a longer CPU time for the preemptive algorithm. As seen from Table 4.2, the heuristic algorithm is able provide an optimized test schedule for the 500 core SoC in approximately 35 seconds for the preemptive strategy and in approximately 23 seconds for the non-preemptive strategy.

4.2.1 Lower Bound on SoC Test Time

Table 4.3 compares the SoC test times for the preemptive and non-preemptive sessionless test schedules, to the lower bound given by Equation 3.11. It can be noted that the difference between the optimal test time and the lower bound of an SoC is as high as 74% in some cases. The optimal test time of sessionless testing is evidently closer to the lower bound than compared to the session-based testing. Here again, the difference between the optimal test
Table 4.3: Test times (in arbitrary units) for sessionless test scheduling with voltage and frequency scaling.

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>Lower bound (LB)</th>
<th>Non-preemptive testing</th>
<th>Preemptive testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test time</td>
<td>% Difference</td>
<td>Test time</td>
</tr>
<tr>
<td>a586710</td>
<td>5464175.96</td>
<td>5.75</td>
<td>5803598.28</td>
</tr>
<tr>
<td>h953</td>
<td>19763.41</td>
<td>67.48</td>
<td>60771.52</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>104.83</td>
<td>23.95</td>
<td>129.98</td>
</tr>
<tr>
<td>d95</td>
<td>4376.98</td>
<td>15.99</td>
<td>5205.9</td>
</tr>
<tr>
<td>g1023</td>
<td>5427.69</td>
<td>39.0</td>
<td>8898.82</td>
</tr>
<tr>
<td>p34392</td>
<td>245784.34</td>
<td>12.08</td>
<td>281358.1</td>
</tr>
<tr>
<td>R100</td>
<td>539.07</td>
<td>17.37</td>
<td>652.42</td>
</tr>
<tr>
<td>R200</td>
<td>1078.14</td>
<td>25.95</td>
<td>1455.97</td>
</tr>
<tr>
<td>R500</td>
<td>2695.35</td>
<td>27.99</td>
<td>3743.39</td>
</tr>
</tbody>
</table>

1Lower Bound calculated at $V_{opt} = 0.69V$, by Equation 3.11. 2Percent difference is computed with respect to the lower bound.

time and the lower bound is caused by the fact that the $V_{opt}$ for the various cores of the SoC may be higher than the value for which the lower bound is calculated. As noted previously in Table 3.7 and Table 3.8, the optimal test times for benchmarks h953 and t512505 are much higher than the lower bound, indicating that erasing session boundaries and adopting sessionless test schedules has still not overcome the significant difference between the optimal and lower bound for the test times of these benchmark SoCs.

4.2.2 SoC Power Budget

In this section, the influence of the power budget over sessionless testing is examined. Similar to the experiment with the session-based testing, the benchmark ASIC Z is subjected to optimization under three different power budgets, 600mW, 900mW and 1200mW. The results, tabulated in Table 4.4, show a similar trade-off between test time and test power.
Table 4.4: Sessionless test schedule optimization for ASIC Z, subject to various power budget values.

<table>
<thead>
<tr>
<th>$P_{\text{max}}$ (mW)</th>
<th>Test time without scaling$^1$</th>
<th>Test time Non-preemptive testing</th>
<th>Test time Preemptive testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test time</td>
<td>% Reduction$^2$</td>
<td>Test time</td>
</tr>
<tr>
<td>600mW</td>
<td>364</td>
<td>183.71</td>
<td>49.53</td>
</tr>
<tr>
<td>900mW</td>
<td>262</td>
<td>137.85</td>
<td>47.38</td>
</tr>
<tr>
<td>1200mW</td>
<td>204</td>
<td>114.35</td>
<td>43.99</td>
</tr>
</tbody>
</table>

$^1$ Test time at fixed voltage and frequency, obtained from Best-Fit Decreasing algorithm (Figure 4.3).  
$^2$ Percent difference is computed with respect to the reference case.

4.2.3 Multiple Supply Voltages

As mentioned earlier, modern SoCs combine cores with varying technologies and requirements. As a result, the SC maybe divided into voltage islands. Cores in one island may have a different operating voltage range than compared to cores in another voltage island. The heuristic algorithm for sessionless testing optimization is also capable of handling cores with different voltage requirements. The condition that cores belonging to the same voltage range can be tested concurrently still applies to sessionless testing. As a consequence of this restriction, the sessionless test schedule gets divided into as many sessions as the number of voltage islands in the SoC, implying that the test schedule will be a hybrid of both sessionless and session-based scheduling. The scheduling algorithm treats this exclusivity requirement as a resource constraint and checks the voltage compatibility of cores while scheduling them concurrently. The multi-voltage experiment in Section 3.4.6 is repeated for the sessionless testing. Figure 4.5 shows the result for the test. The overall test time for ASIC Z is 179.34 units for the non-preemptive scheme, and 181.82 units for the preemptive.

The above experiment featured non-overlapping voltage ranges for the cores. However, the scheduling becomes slightly complicated in case of overlapping ranges since cores from different islands may have a common operating voltage range. This implies that the scheduling algorithm needs to keep track of common voltage levels among cores while scheduling them together.
Figure 4.5: Optimized sessionless test schedules for ASIC Z, under non-overlapping voltage range condition, for both (a) non-preemptive and (b) preemptive cases. (Note: Diagram not to scale.)

4.3 Comparison With Session-Based Testing

As stated earlier, the SoC test time obtained through sessionless test scheduling is always better than or same as that of session-based test scheduling. This section compares the various aspects of sessionless testing with that of session-based testing, with the objective of pointing out the strengths and weaknesses of this strategy. For comparison with session-based test scheduling, the non-preemptive testing scheme is chosen since it fits the description of classic sessionless test schedule and also, the difference in test time for preemptive and non-preemptive schemes is not significant.

The test times for session-based and sessionless test schedules are compared in Table 4.5, at both, nominal and scaled voltage and test clock frequency. It can be inferred from the table that the difference in test time between the two test scheduling strategies is less than 30% for the SoC benchmarks considered. Session-based testing introduces some idle time gaps in the test schedule by waiting for the longest test in the session to complete and hence
Table 4.5: Comparing test time results for session-based and sessionless test schedule optimization.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Optimal test time for</th>
<th>Nominal V_{DD} and clock</th>
<th>Scaled V_{DD} and clock</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>session-based</td>
<td>sessionless</td>
<td>%*</td>
</tr>
<tr>
<td>a586710</td>
<td>14271856</td>
<td>14090716</td>
<td>1.27</td>
</tr>
<tr>
<td>h953</td>
<td>122636</td>
<td>119357</td>
<td>2.67</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>300</td>
<td>262</td>
<td>12.66</td>
</tr>
<tr>
<td>d695</td>
<td>15188</td>
<td>13301</td>
<td>12.42</td>
</tr>
<tr>
<td>g1023</td>
<td>21245</td>
<td>18084</td>
<td>14.88</td>
</tr>
<tr>
<td>p34392</td>
<td>952199</td>
<td>701684</td>
<td>26.31</td>
</tr>
<tr>
<td>t512505</td>
<td>5589002</td>
<td>5344747</td>
<td>4.37</td>
</tr>
<tr>
<td>p93791</td>
<td>178568</td>
<td>139008</td>
<td>22.15</td>
</tr>
<tr>
<td>R100</td>
<td>1347</td>
<td>1208</td>
<td>10.32</td>
</tr>
<tr>
<td>R200</td>
<td>2837</td>
<td>2366</td>
<td>16.6</td>
</tr>
<tr>
<td>R500</td>
<td>7706</td>
<td>5807</td>
<td>24.64</td>
</tr>
</tbody>
</table>

*Percent difference between test times of session-based and sessionless test schedules.

leads to longer test times than sessionless testing. In comparison, sessionless testing does not allow idle time gaps since the test scheduling occurs immediately after completion of older tests. This, however, is most effective when the test times of the SoC cores are very different from each other, as demonstrated through the block diagram in Figure 4.6.

The left-half of the figure shows a case where the SoC has a combination of lengthy (‘T1’) and short (‘T2’, ‘T3’) tests whereas the right-half depicts a case where the core test times are of similar length. The test time reduction achieved by the sessionless testing method over the session-based method, in the latter case is marginal compared to the former case. It can be concluded from the experiment that the advantage of sessionless testing over session-based testing is dependant on the test times of individual cores of the SoC.

The CPU times for the optimization of session-based and sessionless test schedules through the heuristic method, is tabulated in Table 4.6. It can be noted that the time spent on optimizing a sessionless test schedule is much higher than that for the session-based case.
While this may be the artefact of the optimization algorithm, it can be stated that, in general, it is simpler, and hence, easier to distribute tests into sessions.

**Pros and Cons of Sessionless Testing**

This section discusses some of the advantageous and disadvantageous aspects of sessionless testing.

**Pros:** Sessionless test scheduling provides the lowest test time compared to sequential or session-based test scheduling (the same as session-based, in the worst case). It is most beneficial when test times of cores vary from each other by a great margin, as can be seen in Figure 4.6.

**Cons:** Sessionless testing method complicates the test control infrastructure. In general, parallel or concurrent testing assumes that multiple TAM or test buses are available and can be distributed among the SoC cores. The sessionless scheme necessitates that the multiple
Table 4.6: Comparison of CPU times between session-based and sessionless testing, for the heuristic optimization algorithm.

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>CPU time* (in seconds)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session-based testing</td>
<td>Sessionless testing</td>
</tr>
<tr>
<td>a586710</td>
<td>0.12</td>
<td>0.27</td>
</tr>
<tr>
<td>h953</td>
<td>0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>ASIC Z</td>
<td>0.11</td>
<td>0.34</td>
</tr>
<tr>
<td>d695</td>
<td>0.17</td>
<td>0.46</td>
</tr>
<tr>
<td>g1023</td>
<td>0.16</td>
<td>0.63</td>
</tr>
<tr>
<td>p34392</td>
<td>0.19</td>
<td>1.00</td>
</tr>
<tr>
<td>t512505</td>
<td>0.19</td>
<td>2.00</td>
</tr>
<tr>
<td>p93791</td>
<td>0.18</td>
<td>2.02</td>
</tr>
<tr>
<td>R100</td>
<td>1.36</td>
<td>4.68</td>
</tr>
<tr>
<td>R200</td>
<td>2.6</td>
<td>9.4</td>
</tr>
<tr>
<td>R500</td>
<td>6.28</td>
<td>23.22</td>
</tr>
</tbody>
</table>

*CPU time averaged over 100 iterations of the heuristic.

TAM operate independently so that core tests on the TAM can be scheduled independent of each other and thereby erase session boundaries. This implies that each of these test buses or TAM would require its own test control resource, such as scan-enable signal, shift and capture clocks, wrapper instruction register (WIR) etc., incurring a significantly complex control overhead. The feasibility of this approach would then depend on the ability of the ATE and/or the on-chip interface (for e.g., JTAG) to provide such a test control infrastructure. On the other hand, in case of session based test scheduling, the test control is much simplified since all the tests in a session can be provided the same control signals [38], [45]. When the test times of cores are very similar, the test time of a sessionless test schedule may be fractionally lesser than a session-based schedule (Figure 4.6). In such a case, the control overhead costs may not be offset by the test cost reduction achieved by the sessionless testing scheme.

Another point to note is that, resource constraints (such as availability of pins, buses, BIST engines etc.), precedence among core tests (in case of some hierarchical SoCs) or
voltage/power islands can cause sessionless test schedules to be pseudo-sessionless (Figure 4.5(a)) or in the extreme case, session-based test schedule. In recent work, Millican and Saluja propose a MILP model to optimize session-based and sessionless test schedules with voltage and frequency scaling [43]. In their model, concurrently scheduled tests are not restricted to having the same $V_{DD}$ and test clock rate. The results presented in the work, for a set of 16 benchmarks, do not show a significant difference between the test times of session-based and sessionless test schedules. Interestingly, the CPU time reported in the paper for the MILP model for optimizing session-based schedule is much higher than that for the sessionless schedule.
SoCs continue to grow in size and complexity due to continued advancement in IC technology. The test data required to test such large, complex SoCs is voluminous, leading to longer test times. Also, since power consumption during test mode is much higher than during functional mode, the SoC test power may cause local hotspots and supply voltage droops. In this work, a power-aware optimization method of SoC test schedules through voltage and frequency scaling is proposed. The test clock frequency can be scaled by a factor (referred to as frequency factor in this work) to speed-up the SoC test time. This factor, however, is limited by SoC power budget and also by the maximum clock rates of individual cores of the SoC. Restrictions on the core-level clock rate may be imposed by a power constraint (maximum power dissipation limit of the core) or a structural constraint (critical path delay). Voltage can be reduced to lower the power dissipation and increase the clock frequency without exceeding the power limit of the core, thereby resulting in test time minimization. However, in accordance to the alpha power law, further reduction in the voltage causes the critical path delay to increase which, in turn, leads to the increase of test time. Hence, a proper choice of both $V_{DD}$ and test clock rate is required to optimize the SoC test time. The voltage and frequency scaling method of optimization is applicable to both session-based and sessionless SoC test schedules and has been demonstrated on several SoC benchmarks.

Session-Based Testing

For the session-based test scheduling, a power-aware SoC test optimization technique by session-wise optimal selection of $V_{DD}$ and clock has been proposed. A mixed-integer
linear program (MILP) model is formulated to obtain the optimized solution. Results show more than 50% test time reduction over conventional reference test schedules where $V_{DD}$ and clock are fixed at given nominal values. While the MILP method was able to provide optimal solutions for smaller benchmarks, typical SoCs may contain hundreds of cores. Customizing $V_{DD}$ and clock for such SoCs by use of ILP is not practical since the CPU time required to obtain an optimal solution increases exponentially as the number of cores and the complexity of the SoC increases. Hence, a simulated annealing based heuristic optimization method that can provide near-optimal solutions with much less runtime than the MILP method for large SoCs is presented. The effectiveness of the algorithm was demonstrated through experiments on SoCs as large as 500 cores. From the results it can be concluded that the size of the SoC did not have a large impact on the performance of the heuristic approach, unlike that of the MILP method.

Sessionless Testing

The optimization technique through frequency and voltage scaling was also applied to sessionless test scheduling. The proposed heuristic method, which is based on simulated annealing, is capable of providing optimized solutions to both preemptive and non-preemptive type of testing. In preemptive testing, it is assumed that a test can be suspended and resumed at will whereas in the non-preemptive strategy, the tests run uninterrupted until completion. Results show up to 60% test time reduction over conventional reference test schedules where $V_{DD}$ and clock are fixed at given nominal values. Both test scheduling methods, preemptive and non-preemptive, yielded almost identical results for the SoCs considered. However, preemptive test scheduling introduces extra complexity of suspending and restarting tests at will. While the objective of this work is to provide optimization techniques for both, session-based and sessionless test scheduling methods, a fair comparison between the two methods was provided in Section 4.3. It can be inferred from the comparison that
sessionless test scheduling often yields the lowest test time but at an additional cost to the test control architecture.

**Frequency and Voltage Scaling**

In this work, the test clock frequency is varied within the bounds of structural constraint (such as critical path) and power constraint (rated power limit) of the cores. For timing tests where frequency is critical during the capture cycle, varying the clock frequency may be limited to the shift cycles where in the test data is shifted in/out. The shift cycle involves multiple clock cycles to shift the test data in/out as opposed to the capture cycle which is a single, often at-speed, clock cycle. Hence, shifting data faster by varying shift cycle clock frequency can lead to a considerable reduction in the overall test time of the SoC. The constraints on the shift clock rate would be the critical delay of the scan path (structural constraint) and the shift power limit (power constraint) of the scan tested core. At such an event, the proposed method of finding the optimal $V_{DD}$ may also be confined to just the shift cycles during the SoC testing, wherein the voltage can be reduced to regulate the shift power such that the shift cycle clock frequency can be increased (as done by PMScan [16]).

The voltage and frequency scaling schemes presented in this work are intended only for the reduction of test time and should not interfere with the fault coverage of the test. It has been shown that while $V_{DD}$ does not affect stuck-open defects, it may affect the behavior of resistive opens [17,37]. Chang and McCluskey conclude from their experiments that low voltage testing captures defects that can cause early-life and intermittent failures and that these defects are undetected at nominal voltage [11,12]. However, Engelke et al. showed that testing at very low voltages may contribute to coverage loss [17]. This does not, however, invalidate the proposed method but only restricts the available voltage range for the voltage scaling scheme. Hence, the contribution of this work is a method with enough flexibility that user can select the range of voltages based on the defect coverage requirement. Most of the previously reported work is on “very low” voltage testing [11,12,17,22,37].
5.1 Future Work

This section discusses possible extensions and applications of the work presented in this dissertation.

1. IEEE P1687: It is a newly proposed methodology to standardize access to embedded instruments (cores) for test and debug [2]. This standard offers flexibility in configuring the scan path through elements called segment insertion bits (SIB). As a result of this, the test time of a core, in a P1687 environment, depends on other tests that are scheduled concurrently. Due to this, existing test strategies may not be applicable to this environment since these strategies assume the test time to be a constant [68].

2. 3D-SICs: Modern day ICs are not only growing horizontally but also vertically thanks to 3D-Stacked IC (SIC) technology, where ICs can be stacked on top of each other. The connectivity between stacks is provided by special structures known as through-silicon via (TSV). Test scheduling for 3D-SICs poses new challenges such as exacerbated thermal stress, limited number of TSV, etc. [36].

3. Simultaneous frequency and voltage scaling: In this work, it was assumed that $V_{DD}$ and clock had a single input each, which imposes the restriction that concurrently scheduled tests have the same voltage and frequency. However, some of the modern day testers are capable of providing more than one clock inputs to cater to multi-clock domains in the SoC. They are also capable of driving sets of SoC pins at multiple voltages simultaneously. This provides more flexibility in terms of voltage and frequency scaling scheme, since each core in the concurrent set, can now have its own optimal $V_{DD}$ and scaled clock rate. This may, however, lead to a complicated control mechanism and higher overhead.
Bibliography


Appendix: SoC Test Benchmarks

The test data for the various benchmarks is given here. The data is similar to the one provided in Table 3.1. The first section contains a tabulated overview of all the benchmarks. The next section details the format in which the entire test data of the benchmarks are documented whereas the final section contains the actual test data. It must be noted that the data provided is for a nominal voltage of 1.0V.

Overview

The benchmarks used in this research work are listed in Table 1. The first seven benchmarks are a part of the ITC’02 SoC Benchmark initiative [3]. The test time and test power data for the ITC benchmark SoCs have been provided by Millican and Saluja [41,42]. The final three SoCs have been described in Section 3.4.3. It is assumed that each core in a benchmark SoC has a single individual test. Hence, the number of tests for an SoC is the same as the number of cores in it.

Test Data Format

The data for a benchmark contains five records in which core tests are identified as \{T_i\}. An ordered list of these records is given below:

1: SoC Benchmark name and the overall power budget \( P_{\text{max}} \) in mW.

2: Test power in mW for all core tests.

3: Test time in arbitrary time units at nominal voltage and frequency for all core tests.

4: Power constraint limit, \( f_p \), for each core (see Chapter 3).
Table 1: Overview of the benchmarks used in this dissertation.

<table>
<thead>
<tr>
<th>Benchmark name</th>
<th>Number of cores</th>
<th>$P_{\text{max}}$ (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a586710</td>
<td>7</td>
<td>800</td>
</tr>
<tr>
<td>h953</td>
<td>8</td>
<td>800</td>
</tr>
<tr>
<td>d695</td>
<td>10</td>
<td>400</td>
</tr>
<tr>
<td>g1023</td>
<td>14</td>
<td>400</td>
</tr>
<tr>
<td>p34392</td>
<td>19</td>
<td>400</td>
</tr>
<tr>
<td>t512505</td>
<td>31</td>
<td>400</td>
</tr>
<tr>
<td>p93791</td>
<td>32</td>
<td>400</td>
</tr>
<tr>
<td>R100</td>
<td>100</td>
<td>900</td>
</tr>
<tr>
<td>R200</td>
<td>200</td>
<td>900</td>
</tr>
<tr>
<td>R500</td>
<td>500</td>
<td>900</td>
</tr>
</tbody>
</table>

5: Structure constraint limit, $f_s$, for each core (see Chapter 3).

Let us consider the benchmark **a586710** as an example to explain this data format:

```
a586710 P_{\text{max}} = 800

\{
  'T6': 42.2, 'T7': 210.4, 'T4': 107.3, 'T5': 433.9, 'T2': 105.3, 'T3': 138.05, 'T1': 674.65
\}

\{
  'T6': 40431, 'T7': 1914433, 'T4': 181140, 'T5': 7739141, 'T2': 2679692, 'T3': 6029308, 'T1': 6351575
\}

\{
  'T6': 50.0, 'T7': 3.25, 'T4': 15.0, 'T5': 1.28, 'T2': 4.0, 'T3': 2.5, 'T1': 1.0
\}

\{
  'T6': 154.96, 'T7': 3.43, 'T4': 42.28, 'T5': 3.96, 'T2': 15.83, 'T3': 3.34, 'T1': 2.36
\}
```

Line 1 provides the benchmark name and the power budget in mW. Four subsequent records are enclosed within *braces*. The second record (Line 2) gives the peak power (mW) for the seven cores. The third record (Line 3) lists the test time of each core test in the SoC. For instance, the test time for core test ‘T6’ is 42.2 units and the test time for core test ‘T7’ is 210.4 units. Similarly, the succeeding lines provide information regarding the frequency limits (corresponding to power constraint and structure constraint) for each of the core tests of the SoC. Note that for larger SoCs with many more core tests, a record may contain several ‘lines’.
Test Data

a586710 Pmax = 800
\{ 'T6':42.2, 'T7':210.4, 'T4':107.3, 'T5':433.9, 'T2':105.3, 'T3':138.05, 'T1':674.65 \}
\{ 'T6':40431, 'T7':1914433, 'T4':181140, 'T5':7739141, 'T2':2679692, 'T3':6029308, 'T1':6351575 \}
\{ 'T6':50.0, 'T7':3.25, 'T4':15.0, 'T5':1.28, 'T2':4.0, 'T3':2.5, 'T1':1.0 \}
\{ 'T6':154.96, 'T7':3.43, 'T4':42.28, 'T5':3.96, 'T2':15.83, 'T3':3.34, 'T1':2.36 \}

d695 Pmax = 400
\{ T10:3863.00, T8:4605.00, T9:714.00, T6:9869.00, T7:3359.00, T4:5829.00, T5:5105.00, T2:73.00, T3:2507.00, T1:12.00 \}
\{ T10:4.25, T8:6.00, T9:8.00, T6:1.00, T7:2.75, T4:5.00, T5:1.25, T2:10.00, T3:25.00, T1:35.00 \}

h953 Pmax = 800
\{ 'T8':302.52, 'T6':204.25, 'T7':14.21, 'T4':0.33, 'T5':1.79, 'T2':575.38, 'T3':0.59, 'T1':56.6 \}
\{ 'T8':58139, 'T6':34037, 'T7':65, 'T4':1099, 'T5':13541, 'T2':3279, 'T3':1319, 'T1':119357 \}
\{ 'T8':1.0, 'T6':1.59, 'T7':40.0, 'T4':100.0, 'T5':25.0, 'T2':3.05, 'T3':50.0, 'T1':1.61 \}
\{ 'T8':1.82, 'T6':1.93, 'T7':130.4, 'T4':139.83, 'T5':73.45, 'T2':9.96, 'T3':156.09, 'T1':6.09 \}

g1023 Pmax = 400
\{ T8:4484.00, T9:419.00, T6:1679.00, T7:1695.00, T4:1715.00, T5:1775.00, T2:3131.00, T3:14794.00, T0:5939.00, T1:109.00, T10:1237.00, T12:512.00, T13:1024.00 \}

p34392 Pmax = 400
\{ T8:228.00, T9:236599.00, T6:512.00, T7:9930.00, T4:12336.00, T5:1965.00, T2:3108.00, T3:6180.00, T0:170276.00, T1:294064.00, T14:4440.00, T15:128.00, T16:28.20, T17:333.60, T18:27.95 \}
R500 $P_{\text{max}} = 900$

\{0:302.00, 1:134.00, 2:142.00, 3:140.00, 4:86.00, 5:82.00, 6:216.00, 
7:127.00, 8:326.00, 9:173.00, 10:79.00, 11:342.00, 12:231.00, 13:127.00, 
14:150.00, 15:170.00, 16:325.00, 17:167.00, 18:58.00, 19:181.00, 20:181.00, 
21:108.00, 22:176.00, 23:114.00, 24:118.00, 25:131.00, 26:274.00, 27:186.00, 
28:287.00, 29:213.00, 30:247.00, 31:335.00, 32:194.00, 33:279.00, 34:120.00, 
35:199.00, 36:327.00, 37:263.00, 38:271.00, 39:82.00, 40:252.00, 41:266.00, 
42:123.00, 43:269.00, 44:108.00, 45:307.00, 46:254.00, 47:65.00, 48:161.00, 
49:166.00, 50:169.00, 51:175.00, 52:253.00, 53:183.00, 54:272.00, 55:68.00, 
56:348.00, 57:268.00, 58:97.00, 59:148.00, 60:346.00, 61:108.00, 62:219.00, 
63:340.00, 64:217.00, 65:292.00, 66:335.00, 67:110.00, 68:327.00, 69:313.00, 
70:203.00, 71:230.00, 72:311.00, 73:280.00, 74:54.00, 75:266.00, 76:126.00, 
77:230.00, 78:274.00, 79:115.00, 80:137.00, 81:173.00, 82:92.00, 83:212.00, 
84:325.00, 85:126.00, 86:106.00, 87:210.00, 88:233.00, 89:293.00, 90:124.00, 
91:170.00, 92:139.00, 93:303.00, 94:310.00, 95:271.00, 96:79.00, 97:343.00, 
98:266.00, 99:86.00, 100:302.00, 101:134.00, 102:142.00, 103:140.00, 104:86.00, 
105:82.00, 106:216.00, 107:127.00, 108:326.00, 109:173.00, 110:79.00, 111:342.00, 
112:231.00, 113:127.00, 114:150.00, 115:170.00, 116:325.00, 117:167.00, 118:58.00, 
119:181.00, 120:181.00, 121:108.00, 122:176.00, 123:114.00, 124:118.00, 125:131.00, 
126:274.00, 127:186.00, 128:287.00, 129:213.00, 130:247.00, 131:335.00, 132:194.00, 
133:279.00, 134:120.00, 135:199.00, 136:327.00, 137:263.00, 138:271.00, 139:82.00, 
140:252.00, 141:266.00, 142:123.00, 143:269.00, 144:108.00, 145:307.00, 146:254.00, 
147:65.00, 148:161.00, 149:166.00, 150:169.00, 151:175.00, 152:253.00, 153:183.00, 
154:272.00, 155:68.00, 156:348.00, 157:268.00, 158:97.00, 159:148.00, 160:346.00,