Thermal Characterization of BIST, Scan Design and Sequential Test Methodologies

Muzaffer O. Simsir and Niraj K. Jha
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544
{msimsir, jha}@princeton.edu

Abstract—It is a well known fact that during testing of a complex integrated circuit (IC), power consumption can far exceed the values reached during its normal operation. High power consumption, combined with limited cooling support, leads to overheating of ICs. This can cause permanent damage to the chip or can invalidate test results due to changes in the path delay. Therefore, even good chips can fail the test. To prevent this problem, a methodology to generate the thermal profile of chips during test is needed. If such profiles are provided beforehand, temperature-aware testing techniques can be devised. In this paper, we address this problem by presenting a methodology for thermally characterizing circuits under test. In our methodology, first, the test sequences for each targeted test strategy, namely, built-in self-test (BIST), scan design and sequential test generation, are generated automatically. Then, power profiles are extracted by using the switching activity information obtained from simulations. Finally, a very fast thermal profiling tool is used to produce the final thermal profiles. To the best of our knowledge, this is the first work on characterizing the thermal effects of different test methods. Such a thermal characterization can be leveraged for temperature-aware system-on-chip (SoC) test scheduling. Our experimental results present the maximum temperature values attained when using different testing techniques on several benchmarks. Results also demonstrate that low power testing techniques are not necessarily temperature-aware.

1. INTRODUCTION

Advances in silicon manufacturing technology enable designers to build complex ICs with higher transistor densities while keeping the chip size same. Furthermore, to improve the performance of ICs, designers often increase the clock frequency. The increase in transistor density and clock frequency results in an increase in power density. High power density, combined with limited cooling support, causes chips to heat excessively.

To overcome the above problem, temperature-aware chip design methodologies have been proposed [1], [2], [3]. Although the proposed methods provide effective ways for placement and routing by incorporating temperature information in the design flow, they target lowering the maximum temperature during normal operation. On the other hand, it is well known that power consumption during test is several times higher than during normal mode (the ratio can be as high as 30X [4]). Excessive power consumption, which can result in overheating, can cause good ICs to fail the test, leading to a decrease in overall yield and rise in production cost. Even if the

chips escape permanent damage, they can still fail the test. This may happen because the interconnection delay increases approximately 5% for every 10°C increase in temperature [3]. Increased interconnection delays can cause temporary path delay faults on a good chip and, hence, produce incorrect results. On the other hand, even if a chip passes the test, exposure to high temperature during test reduces the lifetime of the chip (aging effect). Another problem can stem from increased leakage current [5]. With increasing temperature, leakage current, and thereby leakage power, increases exponentially, and contributes to further temperature increase. This loop results in even higher temperatures, thus elevating the importance of the problem even more.

In order to reduce excessive power consumption during test, low power test methods have been proposed [6], [7]. In scan design, power reduction is achieved by scan chain modification [8], [9], test vector ordering [10], careful selection of don't care values in test patterns [11], [12], [13] and blocking the switching activity during scan-in cycles [14]. Similarly, for BIST, power-aware test methods use vector filtering [15], reseeding [16] and correlated test patterns [17]. Furthermore, modification of clock trees is also a common approach for both BIST and scan designs [18]. Since the main goal of the above-mentioned methods is to decrease the power consumption during test, their effects on the final thermal profiles are not investigated yet. However, in this paper, we analyze some of these power-aware test methods using our methodology and show how the thermal profiles of these designs change.

Besides the low power techniques, temperature-aware test methods have started to emerge for SoC and network-on-chip designs [4], [19], [20]. These methods aim at decreasing the maximum temperature by creating new test schedules. They assume that each core has a pre-determined test strategy and do not consider the effect of using different test strategies for the cores in the system.

In this paper, we analyze the effect of BIST, scan design and sequential test generation methods on the thermal profile of the circuit under test by presenting a methodology for thermal characterization. The contributions of our work can be summarized as follows:

- It integrates a very fast, yet accurate, thermal profiler that makes complete thermal characterization of test methods

possible.

- It presents a methodology for complete characterization of a circuit under test in terms of power and temperature for a given testing technique.
- It enables comparisons of the effects of different testing techniques on the final thermal profile.
- It provides an infrastructure for developing a temperature-aware SoC test scheduling methodology.
- It shows that power-aware test methods are not necessarily temperature-aware.

The rest of this paper is organized as follows. First, an overview of the test methods thermally characterized in this paper is presented in Section 2. Then, a motivational example is presented in Section 3. The methodology for thermal characterization is explained in Section 4. Implementation details and evaluation of the testing techniques are presented in Section 5. Conclusions are drawn in Section 6.

2. TEST METHODOLOGIES

In this section, we briefly explain the different test methodologies characterized in this paper.

2.1. Built-in self-test

Two BIST methods, BIST per clock and BIST per scan [21], are thermally characterized. In the BIST per clock architecture [Fig. 1(a)], depending on the bit-width of the input signals, there can be multiple pattern generators (PGs). In our experiments, the PGs are inserted at the primary inputs only. Similarly, the multiple response analyzers (RAs) are added to the primary outputs only. Moreover, after the initial seed is loaded, BIST is run till the end of test application time without any further adjustments. Although using more sophisticated BIST per clock methods, such as inserting PGs even at intermediate registers or BIST re-seeding or test vector filtering, can increase the fault coverage, their power profiles are similar to the ones obtained using the above setup. Hence, they lead to similar thermal profiles. On the other hand, in the BIST per scan architecture [Fig. 1(b)], there is a single PG that feeds the scan chains (SCs) during the scan-in cycles. Similarly, there is one RA, which creates the signature by capturing the scan-out signals.

In order to analyze the thermal profile of a low power BIST method, we implemented the technique presented in [17]. In this method, power reduction is achieved by increasing correlation between consecutive test patterns. To be more specific, in between the original test patterns, four intermediate test patterns are inserted. The new patterns are derived from the two original patterns in such a way that the number of bit changes in the new sequence is minimal.
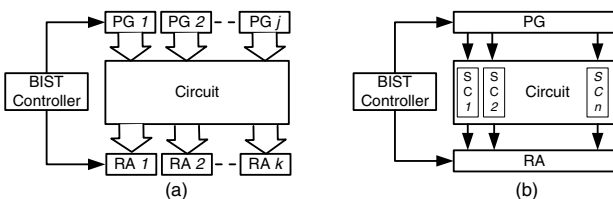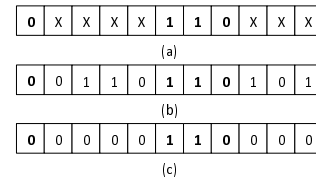

Fig. 2. A scan test pattern after using (a) no modifications, (b) random fill, (c) adjacent fill.

2.2. Scan design

After incorporating full scan in the original circuits, test patterns are derived using a combinational test generator. A large fraction of such test patterns, however, are don't care bits. Selection of values to these bits plays an important role in determining the power consumption and duration of scan test. We implemented the two most commonly used don't care bit filling methods, namely, random fill and adjacent fill [13]. In random fill, the value of a don't care bit is assigned randomly, whereas in adjacent fill, the value of the last defined bit is used. Fig. 2 demonstrates an application of these methods to a test pattern example. The adjacent fill method aims to decrease the number of bit flips in the test pattern, which in turn reduces the switching activity in the combinational circuit during the scan-in cycles. On the other hand, the random fill method makes this decision randomly. However, this method creates a positive side effect, which is the detection of untargeted faults. For this reason, although the test patterns generated by the random fill method cause high power consumption, the test application time is shorter than that of adjacent fill.

Similar to the BIST case, we also implemented a low power scan method [14] for comparison purposes. This method uses an AND gate to block the switching activity during scan-in cycles for each scan cell. Figs. 3(a) and 3(b) demonstrate the original scan chain and the modifications required to implement the low power version, respectively. When scan is enabled in the low power scan architecture, one of the inputs of the AND gate, which is the complement of the scan enable signal, is set to zero. Therefore, the combinational circuit beneath the scan cells receives a constant zero signal even though the values in the scan chain toggle. Since this architecture does not

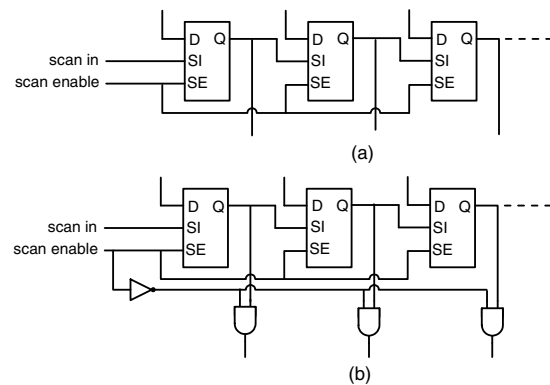Fig. 1. Implementation of (a) BIST per clock, (b) BIST per scan.

Fig. 3. A section of a scan chain that belongs to the (a) original design, and (b) low power design that blocks switching activity with AND gates during scan cycles.
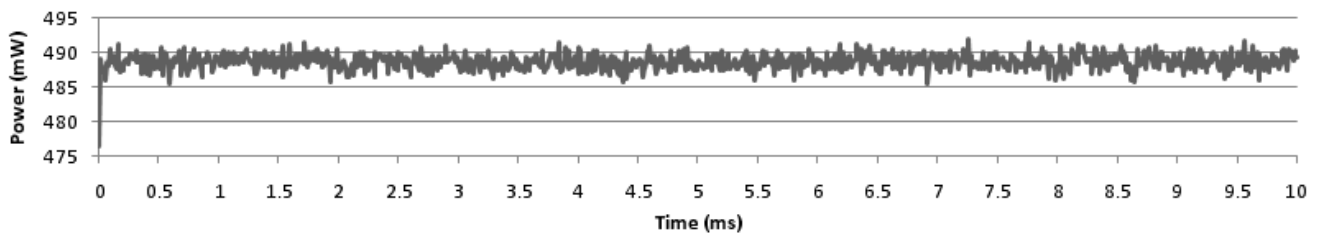
Fig. 4. BIST power profile.

modify the scan chain or the combinational logic, the same test patterns generated for the original design can be used without affecting the fault coverage or the test application time.

Finally, we used a test pattern compaction method to analyze its effect on the final thermal profile. In order to use a compaction method, initially, all test patterns are generated. Then, compaction is achieved by overlapping the set bits of a test pattern with the don't care bits of other patterns. After merging is complete, the remaining don't care bits are filled using either the random or adjacent fill methods described above. Since don't care bit filling is a post-processing step, unlike the case without compaction, application of different fill methods does not change the total number of test patterns, hence test application time.

2.3. Sequential test generation

For sequential test generation, we employed MIX [22]. MIX is a very efficient test generator that can achieve high fault coverages by mixing deterministic, state-driven and genetic optimization based test generation techniques. Moreover, the tool produces results very fast with the aid of partial state transition graphs and computation of flip-flop dependencies. Any other sequential test generator can also be thermally characterized in our framework. However, even though such test generators yield test sequences that have a high power density, because of the resultant small test application times, their impact on temperature is minimal, as we will see later.

3. Motivational Example

In this section, we use a discrete cosine transform (DCT) benchmark [23] to demonstrate the effect of several test methods on the final thermal profile of an IC. In order to mimic the behavior of complex ICs, the DCT core is replicated as many times as necessary to reach a $1mm \times 1mm$ core size. Furthermore, it is assumed that the chip under test contains 100 cores, i.e., has dimensions of $1cm \times 1cm$. During testing, it is assumed that there are five time epochs. During each epoch, 20% of the cores are randomly selected and tested. Of course, a core and chip of another size with a different number of test epochs are also thermally characterizable using the same approach. The power and thermal analysis methodologies used for characterizing different test methods for this core are discussed later in Section 4.

BIST, scan design, and sequential test generation methods, which are explained in Section 2, are characterized using the setup defined above. For scan design, one, two, four and eight scan chains are appended to the DCT core. Then, scan test
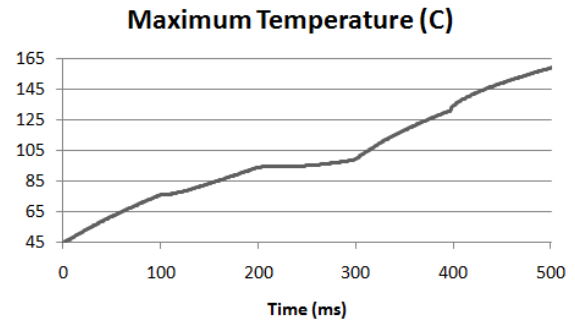

Fig. 5. Temperature vs. time for a BIST-per-clock version of the DCT core tested in the last time epoch.

patterns are obtained by using both random fill and adjacent fill methods with and without compaction.

In order to run thermal analysis, power profiles for the targeted test strategy are needed. Although cycle-by-cycle power characterization would provide the most accurate results, this process suffers from extreme storage space and run-time requirements. In order to overcome this problem, we analyzed the distribution of BIST power consumption over time (see Fig. 4). In the figure, the power curve does not fluctuate much over time. This implies that instead of using cycle-by-cycle power consumption values, using the average power consumption after test proceeds beyond the initial phase is a reasonable approach.

Before comparing the thermal effects of the above-mentioned test methods, it is instructive to look at the thermal characterization of cores during test application. For example, the thermal profile of a single DCT core, which is tested in the last time epoch, is given in Fig. 5. In this experiment, BIST per clock test method is used, and the test application time for a single core is set to $100ms$. The ambient temperature is assumed to be $45°C$. The figure shows that, although the core is not tested in the first four time epochs, i.e., the first $400ms$, its temperature still increases. This is due to the fact that, in prior epochs, other cores were tested and the temperature rise of these cores has a lateral thermal impact on the core in question. Furthermore, depending on the proximity of the other cores being tested, the temperature increase can be significant. On the other hand, in the last time epoch, during which the core is tested, the core temperature shows a steep increase at the beginning. Later on, this increase approaches the rate of increase observed in the previous time epochs.

Although analysis of the thermal profile is useful for understanding the thermal behavior of the circuit under a
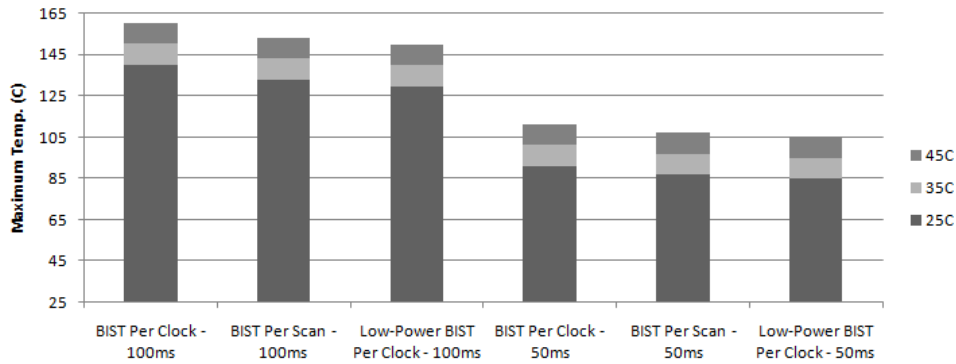
Fig. 6. Comparison of maximum temperature values attained using different BIST methods for DCT.
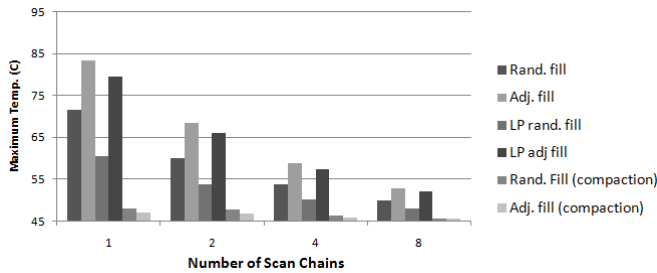

Fig. 7. Effect of varying the number of scan chains in the DCT core on the maximum temperature values.
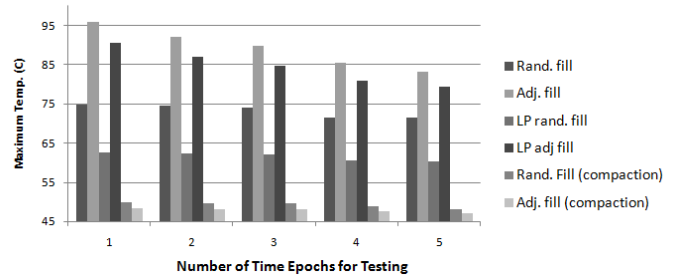

Fig. 8. Number of time epochs vs. maximum temperature values for the scan design of the DCT core.

specific test method, a more important criterion is the maximum temperature reached. For example, Fig. 6 shows the maximum temperature values for various BIST methods for $25°C$, $35°C$ and $45°C$ ambient temperature values. The graph illustrates that longer the BIST methods are applied, higher the maximum temperature rises. This observation suggests that if BIST methods are not used with caution, the chip can be damaged during test. Although using low power BIST helps reduce the maximum temperature, the reduction is not significant. Another observation can be made on the effect of ambient temperature on the maximum temperature. The figure shows that ambient temperature does not affect the rise in temperature. The difference between the maximum temperature values for the same test method under $45°C$ and $35°C$, and $35°C$ and $25°C$, is $10°C$, which is equal to the difference between the ambient temperatures. From this point on, we assume that the ambient temperature is $45°C$.

For scan design, the maximum temperature values at the end of scan tests with random and adjacent fill methods for one, two, four and eight scan chains are presented in Fig. 7. The figure demonstrates that increasing the number of scan chains decreases the maximum temperature. This is because increasing the number of scan chains decreases the test application time, but does not change the power density of the chip much. Since a single scan chain causes the highest temperature increase, all subsequent results for scan design are given for a single scan chain. The figure also shows the surprising result that the adjacent fill method without compaction causes a higher temperature increase than that of random fill without

compaction. This is because although the adjacent fill method decreases switching activity during scan-in cycles, longer test application times, required to reach the same fault coverage values as the random fill method, result in a higher increase in temperature. This example shows that, although reducing power consumption during test is important, test application time should also be considered for temperature-awareness, and power-aware test methods are not necessarily temperature-aware. As expected, the low-power scan method reduces the maximum temperature. When compaction is used, the maximum temperature falls significantly because of the reduced test application time even though the power consumption values are comparable.

One of the assumptions for the experiments described above is that there are five time epochs during which 20% of the cores are tested. Fig. 8 presents the maximum temperature values for a single-chain scan design if this assumption is changed to have one, two, three and four time epochs. For example, if there is only one time epoch, all cores are tested at the same time. From the figure, it can be concluded that, although testing more cores in parallel decreases the overall test application time, the temperature of the chip increases. This shows that decreasing the test application time is also not enough for generating temperature-aware tests – it is important to take into account both the power consumption and test application time simultaneously.

The last test method investigated in this paper, i.e., sequential test generation, causes minimal temperature increase for the DCT core because of the extremely short, and yet efficient,
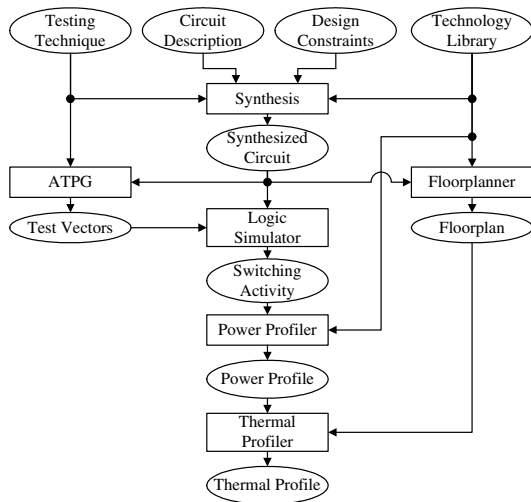
Fig. 9. Overall flow for thermal characterization.

test sequence it produces.

The conclusions drawn from the comparisons presented above provide the motivation for this work. If the thermal effects of different test strategies are known beforehand, then it will be possible to protect the chip by devising temperature-aware test methods. Since power-awareness does not imply temperature-awareness, there is a need to develop temperature-aware test methods directly.

4. THERMAL CHARACTERIZATION

In this section, we explain the overall methodology of thermal characterization, and discuss the thermal profiling tool.

4.1. Methodology

The overall flow of our methodology is shown in Fig. 9. The inputs to the flow are choice of the testing technique, circuit description, design constraints and technology library. The choice of testing technique determines which automatic test pattern generator (ATPG) will be used. Moreover, it can also affect the synthesis step. For example, full scan design requires usage of scan flip-flops and extra controller circuitry. The other inputs, circuit description, design constraints and technology library are used to map the design onto physical gates or register-transfer level modules.

The first step, which is synthesis, ensures that design rules, which are defined in the technology library, and user-defined design constraints are not violated. Next, a floorplanner takes the synthesized circuit and creates a floorplan, which is used by the thermal profiler later on. At the same time, a logic simulator extracts the switching activity information for each element in the design by applying generated test vectors to the circuit. Then, the power profiler estimates the power consumption of each element, i.e., generates the power profile, by processing the switching activity information. Finally, the thermal profiler uses the power profile and floorplan information to generate the estimated temperature value for each element in the design.

Thermal profiling is the key step in our methodology and is explained in detail in the next subsection.

4.2. Thermal modeling

In general, thermal profilers can be used to perform both static and dynamic analysis. The aim of static analysis is to generate the thermal profile of a circuit when the power profile does not change for a very long time. It computes the temperature distribution when the thermal elements in the design reach a steady-state condition. Hence, static analysis can be used if fluctuations in the power profile are small or infrequent. However, if drastic changes in the power profile are observed during operation, then using static analysis can result in neglecting short-term peak temperature values.

Unlike static analysis, dynamic analysis is used to determine the thermal profile of the circuit for a given time period. To be able to run dynamic analysis, the power profile must be characterized correctly for each time period. Dynamic analysis enables observation of thermal fluctuations within the desired time granularity. Furthermore, since the duration of test application is limited, using dynamic analysis is necessary for thermal characterization of test methods. However, this method is much more computationally expensive than the static analysis method.

Selection of the thermal profiling tool is very important. Using a very accurate, but slow, profiler limits the usage of our methodology. Since our methodology may be used in a feedback loop of a temperature-aware test scheduler, long run times will prevent it from being embedded in the loop. On the other hand, using fast thermal profilers, which trade off speed for accuracy, is not preferred since it will lead to incorrect thermal profiles. Then the temperature-aware test scheduler using our methodology will not produce optimal results. We use ISAC [24] for thermal profiling as it is both fast and accurate. It is described in Section 5.

5. EXPERIMENTAL RESULTS

In this section, we provide the implementation details of the proposed methodology, and evaluate the testing techniques discussed in Section 2.

5.1. Implementation details

As mentioned earlier, the inputs to our methodology are the choice of the testing technique, circuit description, design constraints and the technology library. The choice of the technology library is very important for producing results which are representative of the physical implementation since it is used to generate the floorplan and characterize the power profile of the circuit. We use the technology library for the $65nm$ technology node provided by Taiwan Semiconductor Manufacturing Company (TSMC) [25] as the properties of the gates defined in this library are validated in silicon.

The first tool required to implement the proposed methodology is the synthesis tool. For this purpose, we use the commercially available Synopsys Design Compiler [26]. In addition, with the integration of Synopsys DFT Compiler [27], incorporation of scan test circuitry, if necessary, is done automatically. However, integration of BIST circuitry was done using our own programs.

TABLE 1
BASIC PROPERTIES OF BENCHMARKS

Circuit	#inputs	#outputs	#gates	Max. freq. (MHz)
dct	1025	1024	16806	500
des3	239	64	4377	500
dhrc	65	8	4271	250
diffeq	160	96	24536	200
multiplier3x3	320	58	30579	500
s13207	62	152	8589	278
s35932	35	320	17793	500
s38584	38	304	20697	250

TABLE 2
COMPARISON OF BIST METHODS FOR 50ms TEST APPLICATION TIME

Circuit	BIST per clock			BIST per scan			Low power BIST per clock		
	Power den. (mW/mm^2)	Max. temp ($^\circ C$)	Fault cov. (%)	Power den. (mW/mm^2)	Max. temp ($^\circ C$)	Fault cov. (%)	Power den. (mW/mm^2)	Max. temp ($^\circ C$)	Fault cov. (%)
dct	489	111.0	49.7	459	106.9	59.0	445	105.0	49.5
des3	322	88.4	95.5	397	98.6	53.0	340	90.9	97.9
dhrc	74	55.0	87.1	293	84.5	52.3	72	54.7	86.8
diffeq	285	83.5	2.4	341	91.0	50.8	161	66.7	2.2
multiplier3x3	532	116.8	96.8	320	88.2	51.7	442	104.6	96.5
s13207	97	58.1	34.5	257	79.7	66.4	83	56.2	34.2
s35932	512	114.1	60.5	436	103.8	51.8	457	106.7	60.0
s38584	119	61.1	69.8	214	73.9	59.0	107	59.4	69.9

TABLE 3
COMPARISON OF BIST METHODS FOR 100ms TEST APPLICATION TIME

Circuit	BIST per clock			BIST per scan			Low power BIST per clock		
	Power den. (mW/mm^2)	Max. temp ($^\circ C$)	Fault cov. (%)	Power den. (mW/mm^2)	Max. temp ($^\circ C$)	Fault cov. (%)	Power den. (mW/mm^2)	Max. temp ($^\circ C$)	Fault cov. (%)
dct	487	160.1	49.8	455	153.1	59.0	445	149.8	49.7
des3	325	120.8	95.5	399	138.5	53.0	344	125.1	98.1
dhrc	75	62.4	87.2	295	114.0	52.4	73	61.9	86.9
diffeq	282	112.1	2.7	342	125.3	50.9	164	82.9	2.6
multiplier3x3	530	170.3	96.9	320	120.3	51.7	439	149.1	96.5
s13207	98	67.8	34.5	255	105.5	68.4	84	64.5	34.4
s35932	512	165.6	60.5	438	147.7	51.8	460	152.6	60.1
s38584	117	73.0	69.8	216	95.4	59.9	106	70.2	70.0

The tools for floorplanning, logic simulation and power profiling need to be compatible with the TSMC technology library to produce correct results. For this reason, JupiterXT [28], VCS [29] and Power Compiler [30] from Synopsys are used as floorplanner, logic simulator, and power profiler, respectively.

In order to generate test patterns for scan designs, another tool from Synopsys, TetraMAX [31], is used. This tool supports both random fill and adjacent fill methods. On the other hand, as BIST generates the test patterns on-chip, it does not need an ATPG tool. As mentioned before, the sequential test sequences are generated by MIX [22].

Finally, ISAC [24] is used for thermal profiling. ISAC is a fast and accurate thermal analysis software that supports static and dynamic analysis. It performs operation-by-operation dynamic adaptation of temporal and spatial resolution in order to dramatically reduce computational overhead without sacrificing accuracy. It has been validated against reliable commercial thermal analysis tools using industrial and academic synthesis test cases and chip designs. The spatial- and temporal-adaptation techniques proposed for ISAC improve CPU time by 21.64–690.00X and 122.81–337.23X, respectively.

ISAC requires information on the floorplan, power profile

and cooling system setup as input. We implemented a plastic packaging cooling setup, which is a common cooling environment for the type of benchmarks analyzed in this paper. To be more specific, we assumed the thickness of the plastic packaging to be $2mm$. The thermal conductivity and heat capacitivity of plastic are set to $0.53W/mK$ and $0.13 * 10^6 J/m^3K$, respectively. Finally, no heat sink is assumed and the resistivity of the air flow is set to $10^\circ C/cm^2 W$ [32], [33].

5.2. Simulation results

We thermally characterized the test methods explained in Section 2 by using our methodology. In this section, we present the statistics of the simulation results.

The basic properties of the benchmarks used for thermal characterization are presented in Table 1. The first column shows the name of the benchmark. "dct" is a benchmark that performs discrete cosine transform, "des3" performs encryption, "dhrc" performs differential heat release computation, "diffeq" solves a differential equation, "multiplier3x3" performs a 3×3 matrix multiplication, and "s13207", "s35932", and "s38584" are taken from the ISCAS'89 benchmark suite. Columns 2, 3, 4 and 5 represent the number of inputs, outputs, total number of gates and the maximum frequency for each benchmark, respectively.

TABLE 4
COMPARISON OF SCAN TEST METHODS FOR ORIGINAL DESIGN

Circuit	Random fill					Adjacent fill				
	Power den. (mW/mm^2)	Test appl. time (ms)	Max. temp $(^\circ C)$	#Test patterns	Fault cov. (%)	Power den. (mW/mm^2)	Test appl. time (ms)	Max. temp $(^\circ C)$	#Test patterns	Fault cov. (%)
dct	144	72.9	71.4	369	100	77	278.8	83.3	1414	100
des3	91	26.0	52.2	197	100	56	28.1	49.7	214	100
dhrc	133	18.5	53.1	196	98.6	66	20.6	49.4	217	98.4
diffeq	133	4.5	47.3	183	99.9	58	4.6	46.0	185	99.9
multiplier3x3	119	15.6	51.3	252	99.7	61	42.5	52.2	687	99.6
s13207	137	55.3	65.1	623	100	60	71.7	55.8	809	100
s35932	143	7.9	49.2	81	100	78	33.9	52.6	352	100
s38584	119	70.8	66.3	916	100	64	129.8	63.4	1680	100

TABLE 5
COMPARISON OF SCAN TEST METHODS FOR LOW POWER SCAN

Circuit	Random fill					Adjacent fill				
	Power den. (mW/mm^2)	Test appl. time (ms)	Max. temp $(^\circ C)$	#Test patterns	Fault cov. (%)	Power den. (mW/mm^2)	Test appl. time (ms)	Max. temp $(^\circ C)$	#Test patterns	Fault cov. (%)
dct	84	72.9	60.4	369	100	69	278.8	79.3	1414	100
des3	62	26.0	49.9	197	100	48	28.1	49.0	214	100
dhrc	51	18.5	48.1	196	98.6	33	20.6	47.2	217	98.4
diffeq	14	4.5	45.2	183	99.9	9	4.6	45.2	185	99.9
multiplier3x3	36	15.6	46.9	252	99.7	30	42.5	48.5	687	99.6
s13207	45	55.3	51.6	623	100	32	71.7	50.8	809	100
s35932	48	7.9	46.4	81	100	39	33.9	48.8	352	100
s38584	41	70.8	52.3	916	100	32	129.8	54.2	1680	100

The statistics for the BIST methods with $50ms$ and $100ms$ test application times are presented in Tables 2 and 3, respectively. In these tables, the power densities, maximum temperature values and stuck-at fault coverages at the end of testing are shown in Columns 2, 3 and 4, respectively, for BIST per clock. The following columns present the same statistics for BIST per scan and low power BIST per clock test methods, respectively. A comparison of power densities between these BIST methods demonstrates that the low power method achieves, on an average, 13% power reduction compared to the original BIST per clock method. On the other hand, it is not possible to find a direct correlation between the power densities of BIST per clock and per scan methods. However, whenever the per clock methods yield very low power densities, the per scan method gives much higher values. This stems from the fact that, in these cases, the switching activity at the primary inputs does not significantly penetrate the sequential depth of the circuit. On the other hand, in the BIST per scan method, the switching activity is not blocked by any elements in the design, thereby leading to higher power density. If the power densities of the same BIST method for $50ms$ and $100ms$ test application times are compared, it can be seen that the power densities are almost same. This result is expected as the example in Section 3 shows that BIST power consumption does not fluctuate much over time.

The other metric presented in Tables 2 and 3 is the maximum temperature values. When the duration of test is fixed, i.e., $50ms$ or $100ms$, the maximum temperature values depend on power densities only. The results demonstrate that higher the power density, higher the maximum temperature. If the test application time is increased from $50ms$ to $100ms$, the maximum temperature generally rises to much higher levels.

The final statistic presented in Tables 2 and 3 is the stuck-at fault coverage. The tables show that the fault coverage stayed

almost at the same level although the test application time was increased from $50ms$ to $100ms$. Although the fault coverage seems to be low for some cases, it can be improved by using one of the methods mentioned in Section 2.1.

Thermal characterization statistics of scan methods are presented in Tables 4 and 5. For the experiments, the scan frequency is set to $150MHz$. In Table 4, power density, test application time, maximum temperature values, the number of test patterns and stuck-at fault coverages are presented in Columns 2, 3, 4, 5 and 6, respectively, for random fill. The following columns present the same statistics for adjacent fill. Similarly, Table 5 presents the same statistics for random fill and adjacent fill methods when the low power scan architecture, which is presented in Section 2.2, is used. The power density and test application time values for the random and adjacent fill methods are as expected: the random fill method has higher power densities but smaller test application times. The random fill method achieves the same stuck-at fault coverage as the adjacent fill method using fewer test patterns. Both methods provide almost 100% stuck-at fault coverage for all benchmarks. Although the power densities are as expected, the maximum temperature values demonstrate interesting results. For some benchmarks, such as "dct", "multiplier3x3" and "s35932", the adjacent fill method leads to a higher temperature than the random fill method. This shows that temperature-aware testing and power-aware testing are different concepts and low power methods may not necessarily be low temperature ones.

Another set of comparisons can be made between the original scan design and its low power version. Table 5 shows that the power density values of the low power scan architecture are drastically lower than that of the original scan architecture. As the low power scan architecture does not change the design, the test application times, number of test patterns and fault

TABLE 6
EFFECT OF COMPACTION ON SCAN TEST PATTERNS

Circuit	Test appl. time (ms)	#Test patterns	Fault cov. (%)	Random fill		Adjacent fill	
				Power den. (mW/mm²)	Max. temp (°C)	Power den. (mW/mm²)	Max. temp (°C)
dct	9.7	48	100	132	48.0	89	47.0
des3	19.5	148	100	90	49.0	64	47.9
dhrc	17.6	186	98.6	133	52.8	75	49.4
diffeq	3.9	154	99.9	131	47.0	72	46.1
multiplier3x3	5.6	91	99.7	101	47.2	62	46.3
s13207	25.5	290	100	137	55.7	70	50.5
s35932	3.4	35	100	137	46.8	112	46.5
s38584	12.9	166	100	119	50.4	78	48.5

TABLE 7
STATISTICS FOR SEQUENTIAL TEST GENERATION

Circuit	Power den. (mW/mm²)	Test appl. time (μs)	Max. temp (°C)	#Test vectors	Fault cov. (%)
dct	396	34.0	45.1	1695	97.9
des3	372	1.0	45.0	488	99.9
dhrc	51	0.1	45.0	811	88.5
diffeq	161	4.7	45.0	784	99.5
multiplier3x3	361	10.0	45.0	5304	99.5
s13207	80	1.2	45.0	325	26.1
s35932	429	1.6	45.0	773	89.8
s38584	107	8.1	45.0	2035	61.1

coverages are the same as for the random fill and adjacent fill methods used for the original designs. Decreasing the power density, while keeping the test application time the same, helps reduce the maximum temperature values.

Table 6 shows similar statistics for scan test when compaction is used on test patterns. Columns 2, 3 and 4 present the test application time, number of test patterns and stuck-at fault coverages, respectively. The power density and maximum temperature values for the random fill method are given in Columns 5 and 6, respectively. The last two columns present the same statistics for the adjacent fill method. The major difference between the scan tests with and without compaction is the number of test patterns, hence the test application time. When compaction is used, the number of test patterns is reduced, therefore test application time decreases. On the other hand, unlike the case without compaction, the number of test patterns for random fill and adjacent fill methods are the same because don't care bit filling is performed as a post-processing step after compaction. However, compacted test patterns still reach near 100% stuck-at fault coverage for all benchmarks. Another comparison can be made based on power densities. The power density of the random fill method with compaction is higher than that of adjacent fill with compaction, as expected. However, the power density of the random fill method is slightly lower than the case without compaction. On the contrary, the adjacent fill method with compaction causes slightly higher power densities. This stems from the fact that the test patterns after compaction contain fewer don't care bits. Therefore, bit filling methods have less effect on the power density. Finally, when the maximum temperatures of the random fill and adjacent fill methods are compared, it can be seen that the random fill method causes higher temperature increase. However, these values are much lower than that of the cases without compaction.

The statistics for the last test method, which is sequential

test generation, MIX, are presented in Table 7. This table presents the power density, test application time, maximum temperature value, the number of test vectors and stuck-at fault coverage for each benchmark. Although the statistics presented in this table are the same as in previous tables, the unit of test application time is different (μs vs. ms). This shows the efficiency and compactness of test sequences generated by the sequential test generation tool. The stuck-at fault coverages are very high for most benchmarks. The first metric, which is the power density, shows similarities with the ones reported for the low power BIST per clock method. However, the test application times are three to five orders of magnitude lower. This fact can also be seen by comparing the number of test vectors. The very small test application times result in almost no increase in the final temperatures. This implies that, although MIX was used for sequential test generation, using any other sequential test generation method would still yield very low temperature increases. At this point, it must be noted that, since the benchmarks are replicated as many times as necessary to reach a $1mm \times 1mm$ core size, the sequential test generator is used for the original benchmarks. For the replications, the same test sequences are used. For larger benchmarks, sequential test generation may produce longer test sequences, or may not even finish. Therefore, although sequential test generation leads to the lowest temperature, it may not be a practical option in many cases.

6. CONCLUSION

As transistor densities, consequently power densities, continue to increase with technological advances, thermal problems start to become more important due to the limited cooling support. Power consumption during test is much higher than during normal operation of an IC, which makes it important to evaluate the thermal profile of a chip under test. If chips get excessively heated during test, they can be damaged. On

the other hand, even if they are not damaged, chips can still produce incorrect results due to changes in path delays, consequently failing the test.

In this paper, we proposed a methodology for thermal characterization of different test methods. It is used to estimate their impact on the final thermal profile of the circuit. The implementation of the proposed methodology for several test methods based on BIST, scan design and sequential test generation was also presented. In addition to the original test methods, their low power versions were also analyzed. However, our methodology is not limited to the test methods presented in the paper and can be easily applied to any other test method.

Our experimental results showed that low power test methods are not necessarily temperature-aware. Although decreasing the power consumption during test is important, the maximum temperature can still reach higher values when the test application time is longer than that of the original test method. We also concluded that BIST methods should be used with care as they can lead to excessive overheating. Although scan test methods did not cause extreme temperature increases for our experiments, this needs further investigation for other benchmarks and cooling setups.

REFERENCES

[1] K. Skadron, M. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan, "Temperature-aware microarchitecture: Modeling and implementation," *ACM Trans. Archit. Code Optim.*, vol. 1, pp. 94–125, Mar. 2004.

[2] H. Yan, Q. Zhou, and X. Hong, "Efficient thermal aware placement approach integrated with 3D DCT placement algorithm," in *Proc. Int. Symp. Quality Electronic Design*, Mar. 2008, pp. 289–292.

[3] W. Hung, C. Addo-Quaye, T. Theocharides, Y. Xie, N. Vijakrishnan, and M. Irwin, "Thermal-aware IP virtualization and placement for networks-on-chip architecture," in *Proc. Int. Conf. Computer Design*, Oct. 2004, pp. 430–437.

[4] P. Rosinger, B. Al-Hashimi, and K. Chakrabarty, "Rapid generation of thermal-safe test schedules," in *Proc. Design, Automation & Test in Europe Conf.*, Mar. 2005, pp. 840–845.

[5] Y. Liu, R. P. Dick, L. Shang, and H. Yang, "Accurate temperature-dependent integrated circuit leakage power estimation is easy," in *Proc. Design, Automation & Test in Europe Conf.*, Apr. 2007, pp. 1–6.

[6] P. Girard, "Survey of low-power testing of VLSI circuits," *IEEE Design & Test of Computers*, vol. 19, pp. 82–92, May-June 2002.

[7] N. Nicolici and X. Wen, "Embedded tutorial on low power test," May 2007, pp. 202–210.

[8] C. Giri, P. Choudhary, and S. Chattopadhyay, "Scan architecture modification with test vector reordering for test power reduction," in *Proc. Int. Symp. Integrated Circuits*, Sept. 2007, pp. 449–452.

[9] O. Sinanoglu, I. Bayraktaroglu, and A. Orailoglu, "Reducing average and peak test power through scan chain modification," *J. Electronic Testing: Theory and Applications*, vol. 19, pp. 457–467, Aug. 2003.

[10] P. Girard, C. Landrault, S. Pravossoudovitch, and D. Severac, "Reducing power consumption during test application by test vector ordering," in *Proc. Int. Symp. Circuits and Systems*, May 1998, pp. 296–299.

[11] K. Cho, S. Mitra, and E. McCluskey, "California scan architecture for high quality and low power testing," in *Proc. Int. Test Conf.*, Oct. 2007, pp. 1–10.

[12] W. Li, S. Reddy, and I. Pomeranz, "On reducing peak current and power during test," in *Proc. Annual Symp. VLSI*, May 2005, pp. 156–161.

[13] D. Czysz, M. Kassab, X. Lin, G. Mrugalski, J. Rajski, and J. Tyszer, "Low power scan shift and capture in the EDT environment," in *Proc. Int. Test Conf.*, Oct. 2008, pp. 1–10.

[14] S. Bhunia, H. Mahmoodi, D. Ghosh, S. Mukhopadhyay, and K. Roy, "Low-power scan design using first-level supply gating," *IEEE Trans. VLSI Systems*, vol. 13, pp. 384–395, Mar. 2005.

[15] S. Wang, "A BIST TPG for low power dissipation and high fault coverage," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 15, pp. 777–789, July 2007.

[16] A. A. Al-Yamani and E. J. McCluskey, "Built-in reseeding for serial BIST," in *Proc. VLSI Test Symp.*, Apr. 2003, pp. 63–68.

[17] M. Nourani, M. Tehranipoor, and N. Ahmed, "Low-transition test pattern generation for BIST-based applications," *IEEE Trans. Comput.*, vol. 57, no. 3, pp. 303–315, Mar. 2008.

[18] M.-H. Chiu and J. C.-M. Li, "Jump scan: A DFT technique for low power testing," in *Proc. VLSI Test Symp.*, May 2005, pp. 277–282.

[19] C. Liu, V. Iyengar, and D. Pradhan, "Thermal-aware testing of network-on-chip using multiple-frequency clocking," in *Proc. VLSI Test Symp.*, Apr. 2006, pp. 46–51.

[20] D. R. Bild, S. Misra, T. Chantem, P. Kumar, R. P. Dick, X. S. Hu, L. Shang, and A. Choudhary, "Temperature-aware test scheduling for multiprocessor systems-on-chip," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 2008, pp. 59–66.

[21] N. Jha and S. Gupta, *Testing of Digital Systems*. New York, NY, USA: Cambridge University Press, 2003.

[22] X. Lin, I. Pomeranz, and S. Reddy, "MIX: A test generation system for synchronous sequential circuits," in *Proc. Int. Conf. VLSI Design*, Jan. 1998, pp. 456–463.

[23] A. Oppenheim and R. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ, USA: Prentice Hall, 1975.

[24] Y. Yang, Z. Gu, C. Zhu, R. P. Dick, and L. Shang, "ISAC: Integrated space-and-time-adaptive chip-package thermal analysis," *IEEE Trans. Computer-Aided Design*, vol. 26, pp. 86–99, Jan. 2007, can be downloaded from: http://www.robertdick.org/isac/.

[25] "Taiwan Semiconductor Manufacturing Company," http://www.tsmc.com.

[26] "Design Compiler User Guide, 2007.03," http://www.synopsys.com.

[27] "DFT Compiler User Guide, 2007.03," http://www.synopsys.com.

[28] "JupiterXT Virtual Flat Flow User Guide, 2007.03," http://www.synopsys.com.

[29] "VCS User Guide, 2006.06," http://www.synopsys.com.

[30] "Power Compiler User Guide, 2007.03," http://www.synopsys.com.

[31] "TetraMAX ATPG User Guide, 2007.03," http://www.synopsys.com.

[32] L. Shang, L.-S. Peh, A. Kumar, and N. K. Jha, "Thermal modeling, characterization and management of on-chip networks," in *Proc. Int. Symp. Microarchitecture*, Dec. 2004, pp. 67–78.

[33] "Electronic Development Labs Inc., Plastic reference data," http://www.edl-inc.com/Plastic%20data.htm.