Physically-Aware Diagnostic Resolution

John A. Porche  
Dept. of ECE  
Carnegie Mellon University  
Pittsburgh, PA 15213

R. D. (Shawn) Blanton  
Dept. of ECE  
Carnegie Mellon University  
Pittsburgh, PA 15213

Abstract—Physical failure analysis (PFA) is crucial to improving yield during the production life of an integrated circuit (IC). PFA begins with software-based diagnosis, which reports a set of likely failing candidates. A large set, which can result from test-set equivalent (TSE) faults, makes PFA success unlikely because each candidate may potentially have to be examined. Reducing the cardinality of TSE fault classes can reduce the time required for PFA since it likely decreases the faults that must be examined. Additional (diagnostic) test vectors can be generated to reduce class size. In this work, several layout-based physical metrics are defined and investigated for assessing the layout characteristics of a TSE fault class, which serves as an estimate of the time/difficulty required to perform PFA. In short, these metrics are used to define a physically-aware diagnostic resolution (PAR) for a fault class. A TSE fault class with a bad PAR likely requires more PFA effort than one with good PAR. To demonstrate the utility of PAR, a test-selection algorithm that chooses tests from an N-detect test pool based on the improvement in PAR is compared to techniques that use conventional definitions of diagnostic resolution. Comprehensive experiments reveal that fewer additional tests are needed when using PAR.

I. INTRODUCTION

Physical failure analysis (PFA) typically examines a representative integrated circuit (IC) that has failed post-fabrication test. The purpose of PFA is to find the root-cause of the failure. Once the root-cause is determined, the manufacturing process and/or the design can be modified to reduce re-occurrence for yield improvement, or the test-vector set can be modified to ensure a future failed IC of the same type does not escape. Repeating this process to improve yield during product lifetime is one form of yield-learning [1]. Before PFA can be performed however, possible candidates (i.e., failure locations) must be identified. Data gathered during test is analyzed by software-based diagnosis, along with the IC’s logic netlist, to determine possible defect locations in the form of a candidate list. More often than not, the area occupied by the candidates can be quite large, leading to both expensive and unsuccessful PFA being performed on the failing IC [2] [3].

A failing IC with a large number of candidates may not even be chosen for PFA, thus possibly removing a critical sample point for improving either yield or quality. A large number of candidates can be due to the fact that the stuck-at faults analyzed during diagnosis fall into several test-set equivalent (TSE) classes [4]. A TSE fault class is a set of faults (in this case, stuck-at) that have the same fault-simulation response for a given test-vector set.

The process of determining the cause of failure (i.e., test, diagnosis, and PFA) is typically referred to as root-cause failure analysis. The time required for and success rate of root-cause analysis impacts the overall success of yield-learning. Since PFA is a significant portion of root-cause analysis, reducing the time and improving the accuracy of PFA will improve its overall success [5]. The conventional objective of diagnosis is to analyze the tester data to provide candidates for potential examination during PFA. In most cases, multiple candidates will be reported by diagnosis. When multiple candidates are examined, root-cause determination may be unsuccessful due to contamination or destructive examination performed on other candidate sites [6]. Therefore, both the timeliness and likelihood of success of PFA can be improved if the number of candidates can be decreased (i.e., resolution improved). The methods described in [7] and [8] deal with improving physical resolution within a failing cell, i.e., intra-cell resolution. Physically-aware diagnostic tools can report not only candidates, but physical location, layer, shape, and size [9] [10]. Tools such as these can be used to create a new notion of resolution.

Diagnosis quality for a given chip can be quantified using notions of accuracy and resolution. The latter, resolution, is typically defined as the number of candidates reported by diagnosis. Accuracy, on the other hand, is more ill-defined since it requires actually knowing the location of the failure. But in general, a diagnosis result is considered accurate if the actual failing location(s) are among the candidates reported by diagnosis.

Diagnosis can be improved however if the quality of test data is increased using additional tests aimed at improving resolution. (This assumes, of course, that the entire test response of the failing IC is generated and kept, that is, testing continues after failure occurs.) Diagnostic automatic test pattern generation (DATPG) [4] produces tests that improve resolution by distinguishing faults, and are applied only in cases where the diagnostic quality of the test set is of utmost importance. Diagnostic test vectors aim to produce unique responses from each targeted fault, rather than just achieving fault coverage. While faults can be distinguished to a certain extent by the tests created by DATPG, further candidate isolation and localization is accomplished using PFA.

Diagnostic resolution is a function of IC structure, and the test-vector set utilized. Faults that have identical responses for all tests are said to be in the same diagnostic equivalence.
fault class [11] or functionally equivalent [12]. As already mentioned, faults that are equivalent under a test-vector set are said to be in a test-set equivalent (TSE) fault class. The number of TSE fault classes and the cardinality of each is a measure of the overall diagnostic resolution of a circuit and test-vector set pair [13]. The resolution can be improved using additional test vectors, or by modifying the circuit with test points or other on-chip instrumentation. Resolution can also be improved using advanced, physically-aware diagnosis techniques [14] [15].

Improving conventional notions of diagnosis can be costly, either in terms of increased test and computation time, or in terms of silicon area for circuit modification. Optimally, efforts to improve the resolution of a TSE fault class should be undertaken only when the impact on the PFA process is most significant. Utilizing various metrics, the physical properties of each TSE fault class can be examined to determine if improving its resolution (reducing its size) will improve PFA. The conventional approach for characterizing the resolution of a TSE fault class uses cardinality, that is, how many faults are in the class. In this work, various physically-aware metrics that are based on the layout characteristics of the members of the TSE fault class are proposed. We believe these metrics or a combination of them will allow test resources, such as DATPG and PFA, to be better directed for improving root-cause failure analysis and many volume - diagnosis applications (e.g., [10]).

In the remaining parts of this paper, we discuss background concepts (section II), propose physical metrics for diagnostic resolution (section III) and show how the physical metrics of several ISCAS-85, and ISCAS-89 benchmark circuits are used to improve DATPG (section IV).

II. BACKGROUND

A. Diagnostic Resolution

Diagnostic resolution is a measure of how well the faults are distinguished [13] [16]. In this work, diagnostic resolution for a given circuit and test-vector set is defined as the number of faults that reside in each TSE fault class. A TSE fault class with high-cardinality conventionally implies inferior diagnostic resolution, whereas a low-cardinality fault class implies superior diagnostic resolution. The average diagnostic resolution is the average number of faults per fault-set for all of the TSE fault classes that a circuit and its corresponding test-vector set produce. Work in [17] and [18] define diagnostic resolution in terms of distinguished fault-pairs to measure the effectiveness of a DATPG algorithm. In [13], diagnostic resolution is equated to the number of TSE fault classes and their corresponding cardinality. In [11], diagnostic coverage is defined as the number of TSE fault classes divided by the number of faults in the circuit, and is simply the reciprocal of average diagnostic resolution. It is worth noting that the set of faults used in [11] are locally-equivalent, collapsed stuck-at faults. In this work, the faults are not collapsed since such an abstraction obfuscates physically-aware resolution which is defined later.

<table>
<thead>
<tr>
<th>Set size</th>
<th>C432</th>
<th>C499</th>
<th>C880</th>
<th>C1355</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>380</td>
<td>6.69</td>
<td>514</td>
<td>236</td>
</tr>
<tr>
<td>2</td>
<td>347</td>
<td>30</td>
<td>88</td>
<td>192</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>4.2</td>
<td>193</td>
<td>146</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>6</td>
<td>30</td>
<td>238</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>7</td>
<td>29</td>
<td>103</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>2</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>FC%</td>
<td>99.3</td>
<td>99.0</td>
<td>100.0</td>
<td>99.5</td>
</tr>
</tbody>
</table>

Table I: Each table entry is the number of test-set equivalent (TSE) fault classes of a given cardinality which indicated by the row label. The test-vector sets used achieve the stuck-at fault coverage listed in the bottom row.

B. Test-Set Equivalence Fault Classes

Faults that produce the same fault-simulation response under a test-vector set are TSE faults. Previously referred to as fault groups [11] or potentially distinguished faults [19], these TSE faults are found using fault simulation and analysis of fault-output responses. Because the simulation responses of TSE faults are identical for a given test-vector set, a conventional diagnosis approach typically reports all faults of the set as candidates. (The work in [14] is one exception since it uses layout information to distinguish faults with identical responses.) Distinguishing tests can be used to break a TSE fault class into two or more subclasses, thus improving resolution. Performing this step on all TSE fault classes can improve average diagnostic resolution of a circuit under test.

Table 1 shows the number and corresponding cardinality of the TSE fault classes for several ISCAS-85 benchmark circuits using stuck-at test-vector sets. It was constructed via diagnostic fault simulation as described in [11].

C. Physical Failure Analysis

One purpose for developing a diagnostic metric based on physical characteristics derived from the layout centers on improving the success of PFA. PFA involves physical examination, de-processing, and testing of the failing IC to determine the root-cause of failure. Diagnosis uses tester data and a model of the design to identify a set of candidates, where each candidate is defined as a circuit location and sometimes a defect type. Various electrical examinations and forms of inspection are then performed on the candidate to determine the cause of test failure. Inspections during PFA typically include de-processing, emissivity analysis, electrical probing, and optical examination [20].

The number of candidates reported by diagnosis can be reduced by performing some preliminary PFA. This analysis could involve e-beam [21] or PICA [22] probing for localizing electrical anomalies. Further techniques for analysis include emissivity and thermography which can observe photon and
infrared emission to aid in candidate reduction [23].

Figure 1: Illustrative examples of physically-aware diagnostic metrics for three different TSE fault classes: (a) rectangular area (RA), (b) supplemental area (SA), and (c) area union (AU).

III. PHYSICALLY-AWARE METRICS

As mentioned, average diagnostic resolution for a given circuit and associated test-vector set is a function of the number and cardinality (resolution) of each TSE fault class [13]. Instead of characterizing the resolution of a fault class using its cardinality, the portion of the design layout associated with the fault class can be used instead. The idea is to use features of the layout that correspond with the fault class for various purposes that include, for example, gauging the effort required for PFA, the physical resolution achieved by diagnosis, and a measure for diagnostic resolution that is based on the physical characteristics of the fault class. Similar to [24], this work examines physical features related to diagnosis. But unlike [24], the physical metrics are based on a TSE fault class rather than on the physical characteristics of an individual fault site.

Physically-aware diagnostic resolution (PAR) is a measure of a physical placement and size of a TSE fault class in context of the design layout and the fault model assumed. PAR can be based on various metrics that correspond to physical characteristics of a fault class. In this work we employ the stuck-at fault model to the interconnect wires between standard cells. The physical characteristics of the interconnect includes total wire-length (WL), rectangular-area (RA), area union (AU) and supplemental area (SA).

Wire Length. The wire-length (WL) metric is the summation of the length of the polygon segments associated with all members of a TSE fault class. Specifically, the WL metric sums the metal and polysilicon segment length of the node associated with each member of the class, and thus measures the length of wire that may be inspected for PFA. A large value for WL thus implies more effort for performing PFA.

Rectangular Area. The rectangular-area (RA) metric approximates the die area that a TSE fault class occupies. It is meant to capture the portion of die that potentially has to be de-processed or examined during PFA. Specifically, RA is the area of the rectangle that contains all the polygon segments that correspond to the nodes of a TSE fault class. Figure 1a illustrates how the RA is measured for a TSE fault class. RA can indicate the breadth of area the nodes occupy on the die. Because some PFA examination methods have limited observation scope, high RA could indicate that multiple examinations are required to observe all the nodes, thus leading to increased PFA effort.

Supplemental Area. The supplemental-area (SA) metric measures the area contained in bounding box of the nodes in a TSE fault class minus the area of the actual nodes. In other words, SA measures how much the members of the class are spread out. Figure 1b shows a TSE fault class with a large RA which is shown as the black-white patterned area. While the RA for the class of Figure 1b is large, the actual area to examine or observe is relatively small. A TSE fault class with high SA may require additional observations to examine each node in the TSE fault class. Small SA however indicates a class where the nodes are tightly spaced or overlapping, a property that is likely advantageous for PFA.

Area Union. Area union (AU) measures the area of the geometric union of polygon segments of each node in the TSE fault class. For this metric, the union measures the area that covers the poly, metal, contacts, cells that make up the nodes, while subsuming overlapping areas. Figure 1c illustrates how AU is determined. Area union is an approximation of the the total IC area occupied by all the nodes of a TSE fault class.

Composite Score. The four aforementioned metrics can be combined to form a single composite score (CS). It consists of a normalized value of for each metric:

$$\sum ||wl|| + ||ra|| + ||sa|| + ||au||$$

The values are normalized by utilizing the maximum and minimum physical measurements for each metric. Combining these metrics into a composite score provides a comprehensive characterization of a TSE fault class that is not achieved by any one metric. A fault class with a high composite score indicates PFA will be more challenging than one with a lower score.

Figure 2: Rectangular area (RA) of 11 TSE faults from the ISCAS-85 benchmark C1355.

IV. PAR APPLICATIONS

A number of ISCAS benchmark circuits are placed and routed using a 180nm, six-metal CMOS technology that uses a basic standard-cell library. Layout characteristics for the PAR

1For a single fault class, $ra = sa + au$, however, $sa$ and $au$ are included in the equation because they are normalized. Thus, $sa + au \neq ||sa|| + ||au||$.

2We use 180nm in this work, however, no change in results or methods is expected with smaller, modern processes.
metrics are extracted from the resulting LEF/DEF (Library Exchange Format/Design Exchange Format) design files.

A. PAR of Benchmark Circuits

In order to measure the physically-aware diagnostic resolution (PAR) of a given circuit and test-vector pair, the TSE fault classes for the circuit must be first determined. To accomplish this first step, a diagnostic fault simulator has been implemented that locates identical fault-simulation responses efficiently. We applied the simulator to various benchmark circuits using test-vector sets that achieve approximately 100% stuck-at fault coverage. Faults that produce the same simulation response reside in the same TSE class. Now given the faults in each class, it is straightforward to correlate the node associated with each fault to the physical features in the LEF/DEF. For example, Figure 2 shows the RA for an 11-member class for the ISCAS-85 benchmark circuit c1355.

One of the advantages of a physically-aware notion of resolution can be observed by comparing a PAR metric to conventional logic resolution. A plot of various PAR metrics against class cardinality (Figure 3) reveals no correlation between the physical metric and class size. This result shows that reducing cardinality of a TSE fault class via diagnostic test generation, for example, would not necessarily reduce the die area implicated by the smaller class. The results in Figure 3 imply: (1) the layout location and features associated with a fault is entirely dependent on place-and-route, and (2) the number of members of a given TSE fault class is a strong function of the circuit logic structure and the test-vector set employed.

The results of Figure 3 also imply that the traditional notion of resolution reported by a software-based diagnosis tool that uses cardinality can be quite misleading. To further illustrate this point, we compare logic resolution (i.e., cardinality) and various PAR metrics combined into a composite score. Specifically, we applied a commercial diagnosis tool to the failure data described in [25].

The commercial tool computed the cardinality of each injected defect. The possible fault locations were analyzed for their physical characteristics and combined into a composite PAR. Similar to Figure 3, there is no correlation between cardinality and PAR.

An ideal candidate for PFA would be one with a small area to examine, and a small number of nets. A TSE class with cardinality equal to one would conventionally be the first target for PFA. Further use of the composite score for size-1 sets means nodes can be chosen for PFA that minimize the PFA effort. The TSE fault sets that have a cardinality of one are particularly interesting since these failures are the primary targets for undergoing PFA. The use of the composite score to further identify a fault set that will minimize PFA effort will improve PFA success. It may be possible that a set with cardinality of one, could face more PFA effort than another set with cardinality greater than one, when layout is considered. Figure 4 shows classes of 2-cardinality that have significantly lower composite PAR score than some of the 1-cardinality, implying that there may a few multi-cardinality classes that may have less PFA effort than single-cardinality ones.

B. Test-Pattern Selection

The results shown in Figures 3 and 4 motivates an analysis of diagnostic automatic test pattern generation (DATPG). Conventional DATPG aims to improve diagnostic resolution by creating additional test vectors for distinguishing faults [11], i.e., reducing TSE fault class cardinality. Focusing on cardinality alone will not always improve the PAR. To explore this issue, we implemented a greedy physically-aware diagnostic test-selection procedure that is based on PAR (as described in Figure 5).

A circuit is characterized using two methods based on PAR: 1) an overall PAR is computed by summing the PAR for all TSE fault sets. 2) an average PAR is computed using median values of each metric for all TSE fault set. The overall PAR is computed by summing the PAR computed for all of the TSE fault sets. The average PAR is computed by averaging the computed PAR for all of the TSE fault sets.

The greedy-test selection process (Figure 5) begins with generating an n-detect pool of test vectors. Each test in a N-detect pool is fault simulated to determine all the new, smaller TSE fault classes established by each vector. The effectiveness of a test vector is defined to be the amount of change in the PAR (ΔPAR). Fore each loop iteration in Figure 5, the test vector that maximizes ΔPAR and meets a user-defined minimum is removed from the pool and entered into the final test set. The TSE fault classes are subsequently updated and the process is repeated as long as the ΔPAR meets the threshold, overall PAR is not met, and the updated vector pool is non-empty. For comparison purposes, we also implemented a version of the greedy-test selection procedure that uses diagnostic coverage (DC), which is based solely on reducing the cardinality of TSE fault sets. Using various benchmarks, the test-selection procedure is used to select tests that optimize for aggregate PAR, average PAR, and DC. The starting point for each benchmark circuit is the 100% stuck-at fault test set, and the resulting TSE fault classes that result from applying the diagnostic fault simulator to the circuit and test-set pair. This approach is followed because it mimics the conventional approach of appending diagnostic vectors to production test vectors. Using N-detect test pools that are at least five times larger than the 100% stuck-at test-vector set, tests are selected until the ΔPAR saturates. For all the benchmark circuits considered, the tests selected using average PAR as a guide, improve faster than using the diagnostic coverage which is based on cardinality. In other words, conventional DATPG does not optimize PAR serendipitously.

Figure 6 shows the results for the ISCAS-89 benchmark circuit B08, which is representative of all benchmarks investigated. Diagnostic test vectors added to the initial stuck-at test set using three different test selection goals are depicted. The

The work in [25] describes an environment for constructing and simulating various types of defects that are modeled and injected at the layout level.
cumulative change in the aggregated-PAR (labeled as “PAR”), average PAR, and conventional cardinality (labeled as “DC”) is shown. Test selection using the average PAR improves the composite metric 42% over the conventional use of TSE fault set cardinality reduction as the selection criteria. Test selection using aggregated-PAR as the target does not even achieve even the gains of the conventional method.

Average PAR improves when a new test is added and splits a TSE fault set into two or more new TSE fault sets. Aggregate PAR does not take into account the number of fault sets (it merely sums the physical values of the members of all the TSE fault sets). Because of this, a metric like WL will stay constant, regardless of the number of tests added. This problem is somewhat addressed with the area-based metrics of AU, SA,
and RA. The aggregate of these metrics can potentially get smaller (or bigger) with additional tests. Therefore a reduction (or increase) of the aggregate PAR could occur. This non-monotonic behavior can influence the greedy test selection stopping criteria when using the aggregate PAR. Thus, a local minima can cause the process to terminate early (Figure 6).

For a majority of the selected vectors, the average PAR-based selection improves the score for B08 more than the same number of test vectors selected based on the cardinality-based DC. Specifically, for the first seven vectors selected, selection based on average PAR outperforms not only the DC-based selection (by as much as 50%) but also selection based on aggregate PAR.

![Fig. 6: Cumulative change in aggregate PAR (“PAR”), average PAR and DC for test selection from a 30-detect test pool for ISCAS-89 benchmark B08.](image)

V. CONCLUSION

New metrics are introduced for quantifying the physical diagnostic resolution of a set of faults. These metrics measure the physical characteristics of a fault class by examining the layout geometry associate with a fault. The metrics can be used to compare the relative PFA effort required to examine candidates reported by diagnosis. It was also demonstrated how PAR can be used to improve diagnostic test-vector generation. For estimating PFA effort, PAR measurements for diagnosis candidates are shown not to correlate with cardinality, which is the conventional measure of diagnostic resolution. Similarly, for diagnostic test vector generation, cardinality is shown to perform poorly when compared to directly targeting a PAR metric.

Our current work is focused on further developing various PAR metrics that better capture the difficulty of PFA. For example, PAR metrics that account for interconnect shape, number of vias, layer depth, etc. are currently under development.

REFERENCES


