Abstract—Design for manufacturability (DFM) is essential because of the formidable challenges encountered in nano-scale integrated circuit (IC) fabrication. Unfortunately, it is difficult for designers to understand the cost-benefit tradeoff when tuning their design through DFM to achieve better manufacturability. This paper attempts to assist the designer in meeting this challenge by providing a methodology, called rule assessment of defect-affected regions (RADAR), which uses failing-IC diagnosis results to systematically evaluate the effectiveness of DFM rules. RADAR is applied to the fail data from a 90 nm Nvidia graphics processing unit to demonstrate its viability. Specifically, evaluation of various DFM rules revealed that via-enclosure rules play a more important role than the density-related rules. The yield impact of resolving violations is also quantified. In addition, comprehensive simulation experiments have shown RADAR to be accurate and effective for performing DFM evaluation.

Index Terms—Design for manufacturability (DFM), diagnosis, recommended design rules, yield modeling.

I. INTRODUCTION

INTEGRATED circuit (IC) manufacturing technology has progressed tremendously in the past several decades. It has now reached a point where fabrication challenges necessitate design intervention in order to achieve acceptable yield. In short, design for manufacturability (DFM)\(^1\) is now inevitable. DFM requires that design-feature fabrication challenges must be anticipated/discovered, and communicated to the designers so that the design can be remedied accordingly. Semiconductor foundries aid designers by providing DFM design rules [2], [3]. The DFM rules/guidelines are traditionally developed during process characterization or derived from experience learned from earlier technology nodes. Unfortunately the amount or level of DFM necessary to ensure a given design has sufficient yield cannot be predicted a priori. In other words, it is very difficult to know beforehand how much yield will be recovered through the partial or full deployment of a given DFM rule. Unfortunately, because of power, area, performance, and time-to-market constraints on the product, more often than not, the DFM rules are adhered to in an opportunistic manner [2]. Obviously, selection criterion in this manner may not be optimal as far as yield improvement is concerned. For example, it may turn out that DFM rules are followed only in noncongested layout areas but the yield of these areas may not be significantly affected by adhering to a specific DFM rule.

A. Motivation

Having DFM rules alone is insufficient for reaping the full benefits of DFM. The yield impact (YI) due to violations of a particular DFM rule must be understood and quantified. If this information is unavailable, then the designers must at the very least know the relative importance of a set of DFM rules in terms of their YI. For instance, rule A can be deemed more important than rule B, if violating rule A results in more yield loss. As far as we know, there is no existing method for answering such questions [4]. We believe this paper is a step in this direction, that is, an approach for prioritizing DFM rules in terms of YI.

The methodology proposed here is called rule assessment of defect-affected regions (RADAR). RADAR systematically evaluates the importance of a particular DFM rule. Specifically, RADAR diagnoses a large amount of test fail data and extracts the layout regions corresponding to the suspect failure locations. By collecting the DFM rule adherence/violation statistics for these regions, an analysis is performed to: 1) infer the rule violation most likely responsible for a given failure; 2) determine if the adherence of a particular DFM rule will have positive or negative YI; and 3) prioritize DFM rules in terms of their relative ability to prevent yield loss.

B. Prior Work

Studying diagnosis candidates in the context of DFM rules is not a new concept. In [5], a DFM rule check is performed on a design layout, and each area where the check fails is used as a starting point for identifying systematic defects (i.e., nonrandom defects caused by design-process interactions) in a manufacturing process. A similar work in [6] uses critical area analysis [7] to generate an expected net failure rate in nets where the DFM rule check fails. This is used to again identify systematic defects in nets that exhibit an observed failure rate that deviates from expectation. In [8], the DFM rules are “tightened” (i.e., the constraints imposed

---

\(^1\)There is no single agreed-upon definition for DFM. In this paper, DFM is taken to mean some level of adherence of a design to a set of DFM guidelines/rules for the purpose of improving yield.
by the DFM rules are increased) and applied to a design to identify areas that marginally pass the DFM rule check. An example of tightening would be to increase the via enclosure requirement by 10%. Specific tests are then used to target these susceptible areas to ensure quality. Conversely, the deployment of DFM rules can be prioritized based on detectability of systematic defects [9]. More specifically, locations with hard-to-detect faults (identified by fault simulation) are given a higher priority. Dongok et al. [10] mapped the layout locations where DFM rules are violated to a list of faults to improve diagnostic resolution and identify the rules responsible for failures. RADAR significantly differs from past work by analyzing the diagnosis results of actual failed ICs for statistically quantifying the YI of individual DFM rules. Instead of assuming that a given set of DFM rules is a correct starting point for subsequent analysis, we ask: is a given DFM rule truly relevant (i.e., does it impact yield significantly) for a given fabrication process? In addition, instead of assuming a loose hierarchy of importance for all the DFM rules, we ask: if a DFM rule is indeed relevant, how does it relatively rank when compared to other relevant rules? In other words, while prior work focuses on identifying or testing for systematic defects, our paper focuses on precisely understanding the economic tradeoffs of applying a particular set of DFM rules deployed within a given fabrication process.

C. Pros/Cons of RADAR

Since diagnosis candidates of failed product ICs are used, no additional test structure is required and therefore no extra silicon area is consumed by RADAR deployment. In addition, since product ICs typically occupy much more silicon area than test structures, using product ICs can substantially increase the amount of actionable information that can be collected. Moreover, in a manufacturing production line, tests must be applied to screen out failed ICs. Diagnosis of the test fail data consumes only CPU cycles and therefore does not incur a substantial cost to production. In other words, using diagnosis candidates to evaluate and rank DFM rules can be performed at a very low cost.

Another advantage of using product ICs is that the DFM rules are deployed in a realistic environment, i.e., the product itself. This is important because, even if the YI of a DFM rule can be accurately quantified using test structures, there is no guarantee that the YI remains the same when the rule is deployed in actual product ICs, which typically contain far more diverse layout features than conventional test structures.

Finally, a manufacturing process is not static. It changes over time. This means, of course, that the relevance of a DFM rule may change over time as well. It is therefore important to monitor these changes for effective DFM and quality control. Using on-going diagnosis to detect changes in DFM rule ranking is a cost-effective and nonintrusive way to monitor the process.

One limitation of RADAR is that its accuracy is affected by noisy data. Specifically, poor diagnosis results in terms of resolution and accuracy will adversely affect RADAR's accuracy. For instance, when more DFM rules are suspected due to poor diagnostic resolution, it is obviously unclear which one is the source of IC failure. This issue can be mitigated by better diagnosis and defect localization [11]–[13]. However, even when the defect location is precisely known, it is still possible that multiple rules are violated at this location, resulting in a degradation of accuracy. This noise is inherent and cannot be completely removed. The expectation-maximization (EM) approach used in RADAR addresses this problem however, and allows inference to be drawn in the face of these uncertainties.

D. Practical Considerations

RADAR is applicable in a variety of settings. First, it can be applied to the failed ICs currently in production to learn the important DFM rules and prioritize them for future designs in the same process. In addition, the ranking of the DFM rules can indicate the relative severity of the process-design interactions, which can also be used to prioritize process improvement effort for the current product. Second, RADAR can also be applied to test chips, which are now being deployed by fabless design houses to gauge design characteristics (cells, layout patterns, etc.) in a new process before fabricating the flagship product. These test chips are generally designed to reflect the actual product as much as possible. Therefore, using RADAR, the prioritized DFM rules learned from the analysis of failed test chips can be incorporated into the design flow to maximize yield of product ICs. Third, some products, such as microprocessors, typically undergo further design revisions (for timing and power improvement, for example) when the products are already in the market. RADAR can be applied to the failed ICs in the current revision in order to provide feedback for future revisions of the same design to maximize yield improvement at the same technology node.

It should be noted that although this paper only focuses on studying the DFM rules, other techniques for yield improvement (e.g., the use of phase shift mask [14]) can also be studied in a similar manner and is one aspect of our on-going work.

The rest of this paper is organized as follows. Section II briefly discusses the key components and the flow of RADAR. Section III describes the details of the statistical analysis used to evaluate the effectiveness of DFM rules. This is followed by experiments that utilized real fail data from an N-vidia graphics processing unit (GPU), and simulated fail data from a number of benchmark circuits [15], [16] in Section IV. Finally, the paper is summarized in Section V.

II. DFM RULE EVALUATION

Before evaluating fail data, a rule-violation database for every DFM rule of interest must be created for all designs under consideration. This database is queried during rule evaluation. The rule-violation database is readily available from the design sign-off step, but can also be easily regenerated by performing the DFM rule check and saving the layout locations of the resulting violations. While this task is nontrivial, it is just a one-time effort for each design.
heuristic, such as [13], can also be used to further localize the defects not targeted by DFM or random contaminations.

The last category represents failures caused by systematic defects not targeted by DFM or random contaminations. At this point, although not strictly required, a layout analysis heuristic, such as [13], can also be used to further localize the affected area to improve accuracy. The region that corresponds to the affected area of the layout is queried in the rule-violation database to increment the adherence count (i.e., $X_k$). The corresponding area in the layout is called the failing layout region and can be categorized in one of three ways.

1) **Violated-Rule Failure**: Geometry targeted by a rule exists in the region and the rule is violated.
2) **Adhered-Rule Failure**: Geometry targeted by a rule exists in the region and the rule is not violated.
3) **Nonadhered Failure**: Geometry targeted by a rule does not exist in the region.

The last category represents failures caused by systematic defects not targeted by DFM or random contaminations.

At this point, although not strictly required, a layout analysis heuristic, such as [13], can also be used to further localize the affected area to improve accuracy. The region that corresponds to the affected area of the layout is queried in the rule-violation database to increment the adherence count (i.e., $X_{11}$) or the violation count (i.e., $X_{01}$) of the corresponding DFM rule if the rule is applicable to that region. These two counts are collected for every DFM rule of interest.

Similarly, the passing layout region corresponds to the portions of the circuit that have passing signal lines. This is typically a very large portion of the design and hence contains many signal lines. Therefore, this area is sampled to keep the analysis time tractable. Specifically, we randomly sample from the set of passing signal lines. This sampling procedure should have a negligible effect on accuracy, as long as a sufficient number is obtained. Again, for every DFM rule of interest, the same procedure is performed on the rule-violation database to obtain its violation count (i.e., $X_{00}$) and adherence counts (i.e., $X_{01}$).

Based on the information collected, statistical analysis is then employed to determine: 1) if there is strong evidence to indicate that the application of a DFM rule is independent of the observed failures (i.e., has no impact on yield one way or the other) and 2) the relative importance of the DFM rules in terms of their YI.

### III. Statistical Analysis

This section describes the statistical methods used to study the effectiveness of DFM rules in terms of their impact on yield. Fig. 2 shows the statistical analysis flow, which consists of two independent sub-flows. The right sub-flow starts by classifying the observed failures into different rule categories (violated, adhered, or nonadhered). When the rule violation responsible for causing failure is known, the yield loss of the product can be attributed to each rule violation. For example, suppose the yield loss of a product is 15% and suppose the classification results show that 10% of the failures are caused by rule-$k$ violations, then the yield loss due to rule-$k$ violation is 1.5% and the yield loss attribution is 10%. In other words, yield loss attribution simply refers to the fraction of failures in each rule category after classification for the purpose of decomposing the total yield loss into rule categories. This is desired because, when the yield loss due to each rule violation is known, the failure rate of each violated rule can also be analytically calculated afterwards. Although the right sub-flow is represented as a three-step process, our actual implementation uses the EM algorithm that analyzes the entire set of rules concurrently and yields the result for all three steps simultaneously.

In contrast, the left sub-flow examines each rule individually and it begins by analyzing the formal notion of statistical association between a DFM rule and yield loss—association is the antonym of independence in statistics. Subsequently, confirmation of a causal relationship between the DFM rule and yield loss, using one or more of the three approaches, is performed. The relative risk (RR), which is defined as the ratio of the probability of failure given rule violation to the probability of failure given rule adherence, is calculated for each DFM rule and is used to rank all rules.

It should be also noted that the left sub-flow can also be applied to multiple rules simultaneously. This is necessary if the rules are highly correlated and cannot be distinguished based on diagnostic resolution, orthogonal misbehavior, or severity of violation. Such cases can be easily handled, however, by combining such rules into a single-rule type.

With the RR estimated from the left sub-flow and the violated-rule failure rate estimated from the right sub-flow, the
YI of addressing a particular rule violation can be statistically quantified and used to guide design optimization. It should be noted that, despite similarity in name and in concept, yield loss attribution and YI estimation are different. Specifically, yield loss attribution is the fraction of failures in each rule category after classification while YI refers to estimated change in yield after a certain fraction of the rule violations are corrected. Therefore, yield loss attribution is specific to the current in-production design while YI estimation is not. YI estimation can be applied to the current in-production design (in this case, it should correlate with yield loss attribution) or future designs to predict how much yield loss will be recovered through partial or full deployment of the DFM rules.

It should also be noted that the classification results obtained from the first step of the right sub-flow can also be used as the starting data source to the left sub-flow to improve the accuracy of the follow-on analyses. However, this step is optional because there may not be sufficient data in each rule category after classification for the subsequent analysis. In addition, there is no dependency between the causation confirmation step and the follow-on steps. If it is computationally expensive to confirm causation for every rule, then the RR or the YI can be estimated first so that only the top-ranking rules are validated by causation confirmation. The flow in Fig. 2 is shown in this way since we believe it is more intuitive. The order can be easily altered, however, for a more flexible implementation of the RADAR methodology. In the following subsections, the left sub-flow is examined first since the right sub-flow is tightly coupled to the YI calculation.

A. Association Confirmation

The association confirmation described here is based on [17]. To determine the YI of a given DFM rule, for a given layout region that corresponds to a signal line, the random variables $R$ and $L$ are defined as follows:

$$R = \begin{cases} 0 & \text{if rule is applicable and is not enforced in the layout region} \\ 1 & \text{if rule is applicable and is enforced in the layout region} \end{cases}$$

$$L = \begin{cases} 0 & \text{if layout region is within the passing layout region} \\ 1 & \text{if layout region is within the failing layout region}. \end{cases}$$

The data collected using RADAR described in Fig. 1 can be summarized in the format shown in Table I.

<table>
<thead>
<tr>
<th>$R$</th>
<th>$L = 0$</th>
<th>$L = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$X_0$</td>
<td>$X_0$</td>
</tr>
<tr>
<td>1</td>
<td>$X_o$</td>
<td>$X_{10}$</td>
</tr>
</tbody>
</table>

Each $X_{ij}$ represents the number of observations for which $R = i$ and $L = j$. The dotted subscripts represent sums, $X_i = \Sigma X_{ij}$ and $X_j = \Sigma X_{ij}$. If $X$ is modeled as a multinomial distribution, i.e., $X = (X_{00}, X_{01}, X_{10}, X_{11})^T \sim \text{multinomial}(n, p)$, where $n$ is the number of trials and $p = (p_{00}, p_{01}, p_{10}, p_{11})^T$ is the event probability vector, then the maximum likelihood estimate [18] for the individual event probability $p_{ij}$ is given by $X_{ij}/n$ [17].

Intuitively, before ranking a DFM rule, one must know if the DFM rule of interest affects yield at all (i.e., if $R$ is independent of $L$). In other words, the following hypothesis should be tested:

$$H_0: p_{ij} = p_i \times p_j \quad \text{versus} \quad H_1: p_{ij} \neq p_i \times p_j.$$

We can use the Pearson’s $\chi^2$ test and define the statistic $T$ to be

$$T = \sum_{i=0}^{1} \sum_{j=0}^{1} \left( \frac{X_{ij} - E_{ij}}{E_{ij}} \right)^2,$$

where $E_{ij} = \frac{X_i X_j}{n}$.

Under $H_0$, $T$ converges in distribution [19] to $\chi_1^2$, which means the distribution of $T$ will increasingly resemble that of $\chi_1^2$ as more data is collected. An appropriate $\alpha$-level test is to reject $H_0$ when $T > \chi_1^{2,\alpha}$. A common choice of $\alpha$ is 0.05 which means that the probability of incorrectly rejecting $H_0$ is at most 5%. In addition, the $p$-value (which is the smallest $\alpha$ value that will result in $H_0$ being rejected with the current data) is usually calculated along with $T$ to indicate the probability of observing the same or more extreme data.

Alternatively, we can use the likelihood ratio test and define

$$U = 2 \sum_{i=0}^{1} \sum_{j=0}^{1} X_{ij} \log \left( \frac{X_{ij}X_iX_j}{X_{i}X_j} \right).$$

Under $H_0$, $U$ also converges in distribution to $\chi_1^2$, and an appropriate $\alpha$-level test is to reject $H_0$ when $U > \chi_1^{2,\alpha}$. In this paper, both tests are used with $\alpha = 0.05$ and the $p$-values for both tests are also calculated.

B. Causation Confirmation

As noted in [20], an association between random variables $R$ and $L$ does not imply a causal relationship between $R$ and $L$. In other words, even if a DFM rule is associated with yield loss, it does not imply that violating the DFM rule will cause yield loss. This can occur due to the existence of confounding variables [20] that distort the relationship between $R$ and $L$. For example, suppose a particular DFM rule is not enforced in the congested regions of a design because of area constraints. These congested regions typically have relatively high yield loss due to critical area [7]. Hypothesis testing for independence of $R$ and $L$ will likely reject $H_0$ since the data will indicate that the DFM rule violation (i.e., $R = 0$) is associated with yield loss (i.e., $L = 1$). For this example, it is therefore tempting to conclude that adherence of the DFM rule leads to yield enhancement. However, it is clear that the conclusion may not be valid since these congested regions can suffer from relatively high yield loss regardless of whether the DFM rule is enforced. In other words, critical area can be a confounding factor that causes random variables $R$ and $L$ for a given rule to appear to have a causal relationship.

Fig. 3(a) illustrates the effect of confounding variables using a hypothetical plot. The upper and lower horizontal lines correspond to the average yield loss for congested and noncongested regions, respectively. Black dots represent observed data points
(i.e., the DFM rule $X_{ij}$ counts) while gray dots represent unobserved data points. The dashed line is the best-fit line for the observed data. It is clear from Fig. 3(a) that yield loss is independent of DFM adherence for both the congested and noncongested regions. The best-fit line, however, indicates that there is an association between low yield loss and the adherence of a particular DFM rule. This occurs because the observed data is dependent on critical area, which is the confounding factor in this hypothetical example. This example also shows that an association does not imply a causal relationship.

Since enforcing a DFM rule can be costly, the rule should be continually applied (e.g., to future designs in the same process) only when violation of the DFM rule is confirmed to cause yield loss. To combat the problem of confounding variables, three approaches are examined: 1) active conditioning [21]; 2) confounding adjustment [20]; and 3) hypothesis confirmation via simulation.

1) **Active Conditioning**: Active conditioning [21] refers to the process of actively changing the random variable $R$ (in this case, enforcing a particular DFM rule) and then checking if it impacts $L$ (in this case, yield loss). This means that instead of passively observing $L$ from a given set of data $(R, L)$, $R$ is actively changed to derive new data. Using the same example, this would mean intentionally enforcing a particular DFM rule in some of the congested areas and doing the opposite for some noncongested areas. Clearly, this method is very expensive since it involves design changes. Specifically, modifying the layout for identifying yield-loss mechanisms is likely not justifiable. Nonetheless, this method can be viable for test chip design to gather as much feedback as possible before using the process to manufacture the actual product.

Fig. 3(b) illustrates the effect of using active conditioning to study the causal relationship. Comparison with Fig. 3(a) shows that the new set of data collected is more complete and is now independent of the confounding variable (critical area). When a random variable can be actively conditioned, the issues of confounding are avoided and an associative relationship is equivalent to a causal relationship. Again using the example of Fig. 3, each data point $(R, L)$ is classified into groups defined by level of critical area. It must be emphasized that for the causal inference to be valid, all confounding variables must be identified and isolated. Although there is no manufacturing cost to adjust for confounding (unlike active conditioning), the measurement of the confounding variable(s) can still be expensive. (These are also challenges for active conditioning.) In this case, critical area must be extracted for all regions of interest in the design (not just the failing layout regions) to separate the data (i.e., the DFM rule $X_{ij}$ counts) into groups that have similar levels of critical area.

Fig. 3(c) illustrates the groups formed for confounding adjustment. The lower circle denotes the low critical-area group while the upper circle denotes data with high critical area. The lower and upper horizontal lines are the best-fit lines within each group. The two lines are averaged to give the average best-fit line (dashed), which now indicates the correct relationship between $R$ and $L$.

3) **Simulation Validation**: The last approach explored for combating confounding uses simulation. Simulating portions of the layout that correspond to failing layout regions can suggest a likely root-cause of failure and therefore confirm the causal relationship between a DFM rule and yield loss. For example, lithographic simulation [22] and chemical-mechanical polishing (CMP) simulation [23] can be performed. Sub-wavelength lithography inevitably introduces distortion in the printed pattern, which is a significant source of yield loss in nano-scale manufacturing. Performing lithography simulation enables the identification of hotspots due to the specific spatial arrangement of the layout features. Another commonly occurring yield loss mechanism is due to CMP. Defects, which include metal stringers [24] and delamination [25], can be induced by the CMP process. CMP simulation can therefore be performed on the failing layout region to determine if the observed yield loss is likely caused by a side effect of CMP. Although only lithography and CMP simulation are suggested here, any simulator that accurately identifies potential yield issues can be used to confirm the causal relationship.

In this paper, the confounding-adjustment approach is used since we have no control on the failure data collection for the industrial data and therefore cannot perform active conditioning. In addition, we also do not have access to process simulation models to perform simulation-based hypothesis confirmation.
C. RR Estimation

After a causal relationship has been established, the odds ratio ($\psi$) can be estimated to measure strength of association as follows:

\[
\text{Odds Ratio} = \psi = \frac{P(R'|L)}{P(R'|L')} \approx \frac{X_{01}X_{10}}{X_{11}X_{00}} = \psi
\]

where $R'$ ($R$) represents $R = 0$ ($R = 1$) and $L'$ ($L$) is defined similarly. Since the failure rate of a layout feature $P(L = 1)$ is typically very small, under the “rare disease assumption,” it can be shown that the odds ratio approximates the RR [17]

\[
RR = \frac{P(L|R')}{P(L|R)} \approx \frac{P(R'|L)}{P(R'|L')} \approx \frac{X_{01}X_{10}}{X_{11}X_{00}} = \frac{(X_{01} + 0.5)(X_{10} + 0.5)}{(X_{11} + 0.5)(X_{00} + 0.5)}.
\]

In the last expression, 0.5 is added to each $X_{ij}$ term to reduce the variance when the sample size is small [17]. This method is used in this paper by default.

One interesting property of the odds ratio is that it does not depend on the amount of passing data collected (i.e., the violation/adherence count collected in the passing layout regions), assuming that a sufficient amount of data has already been sampled. For example, if sampling for passing data increases (decreases), then $X_{10}$ and $X_{00}$ will increase (decrease) by approximately the same factor. Since $X_{10}$ and $X_{00}$ appears in the numerator and denominator of the odds ratio expression, respectively, their effect will be cancelled out. It should be noted that a different metric was used in the earlier version of RADAR [1]. When performing validation of RADAR using simulation experiments (to be discussed in detail in Section IV-B), it is found that, while the earlier metric gave the correct rule ranking, it cannot be reliably interpreted because it is dependent on the amount of passing data collected and therefore is not invoked here.

From its definition, it is clear that the RR based on the odds ratio is simply a ratio of the probability of failure given rule violation [i.e., $P(L|R)$] to the probability of failure given rule adherence [i.e., $P(L|R')$]. Therefore, the RR can be interpreted as follows: 1) when $RR \approx 1$, rule violation and yield loss are independent; 2) when $RR \gg 1$, rule violation is RR times more likely to result in yield loss than rule adherence; and 3) when $RR \ll 1$, rule adherence is 1/RR times more likely to result in yield loss than rule violation.

The standard error (SE) of the odds ratio (and therefore the RR) can be calculated analytically using [17]

\[
se(\psi) = \psi \sqrt{\frac{1}{X_{00}} + \frac{1}{X_{01}} + \frac{1}{X_{10}} + \frac{1}{X_{11}}}
\]

Thus, the normal-based (1 − α) confidence interval of the estimated odds ratio can be calculated as $\psi \pm z_{\alpha/2}se(\psi)$, where $z_{\alpha/2}$ is the value of the inverse cumulative distribution function (CDF) of a standard normal distribution when the probability is $\alpha/2$.

In RADAR, RRs are calculated for each rule and are used to rank every DFM rule. Rules are sorted in decreasing order of the RR. Rules with RR substantially larger than unity (i.e., > 10) should be enforced. Rules with RR larger than unity should be further investigated. In the rare and paradoxical scenarios that a rule is found to have RR substantially less than unity (i.e., < 0.1), then an in-depth investigation should be performed to understand why rule adherence is seemingly causing failures. Particular attention should be paid to the top-ranked rules to prioritize DFM adherence to improve yield.

D. YI Estimation

The RR metric gives the relative importance of a given set of rules and therefore is useful as a general guideline for prioritizing DFM rule adherence. However, it does not take into account the specific layout features in the design. Therefore, it does not quantify the yield impact (YI) of adhering to a particular rule. For example, rule A may be more important than rule B in terms of RR, but if rule B has much more violations than rule A in a particular design, then enforcing rule B may have a larger YI than enforcing rule A. To quantify the YI of a DFM rule $k$, let $v_k$ be the total number of violations and $\beta_k$ be the fraction of violations corrected by DFM rule enforcement, where $0 \leq \beta_k \leq 1$. YI of rule $k$ can be calculated in the following way:

\[
YIk = \left[1 - \left(1 - P_k(L|R)\right)^{v_k}\right] - \left[1 - \left(1 - P_k(L|R')\right)^{(1-\beta_k)v_k}\right].
\]

The first term is the estimated yield loss due to violations of rule $k$ before any correction. The second term is the estimated yield loss after $\beta_k \times v_k$ violations have been corrected. The YI calculation measures the difference between these two scenarios. This calculation, however, requires parameters $P_k(L|R')$, $P_k(L|R)$, and the total number of violations $v_k$ for each rule. $v_k$ is available from the rule-violation database. For parameters $P_k(L|R')$ and $P_k(L|R)$, it is tempting to use the data collected from the RADAR flow to estimate them directly. However, this is not possible since the sampling scheme of RADAR is a case-control sampling [17], i.e., the passing and failing data are collected independently and therefore the data collected is not representative of the overall population. Nonetheless, since the RR, which is simply a ratio of $P_k(L|R')$ to $P_k(L|R)$, is estimable, the YI can be estimated if an estimate for either $P_k(L|R')$ or $P_k(L|R)$ is available. In the following, two approaches for estimating $P_k(L|R')$ and $P_k(L|R)$ are discussed.

1) Estimation of $P_k(L|R')$: In the first approach, $P_k(L|R')$ is estimated. By definition, $P_k(L|R')$ is the failure rate of the layout pattern described by rule $k$ given that rule $k$ is violated. This failure rate can be estimated if the yield loss due to violation of this particular rule is known. Unfortunately, the yield loss contributed by violations of a rule is not known in general and can be considered as an unobserved class label for each failed IC. To cope with this uncertainty, EM [26] is used for maximum likelihood estimation of the probabilities we define next. Specifically, let $K$ be the number of rules under evaluation, then each failed IC can be assigned one of the $(K + 1)$ different class labels, that corresponds to each of the $K$ rules and an unknown category, where the unknown category defined here is the union of adhered- and nonadhered failure categories. Let $y$ represent the assignment of a failed IC to a rule category, i.e., $y$ can take values from 1 to
Let $\pi_k = P(y = k)$ denote the yield loss attribution for rule $k$ (also known as the class prior probability). Let $N$ be total number of failed ICs and let $D_1, D_2, \ldots, D_N$ denote the data collected for each failed IC, where $D_i$ is a binary vector of length of $(K + 1)$. The $k$th component of $D_i$ (represented as $D_{i,k}$) is one if and only if rule $k$ is violated in the failing layout region of the $i$th failed IC. If none of the $K$ rules are violated, then it is assigned to the unknown category, i.e., $D_{i,(K+1)} = 1$. Using this data representation, the EM algorithm consists of the following steps.

i) Initialization: Initialize current estimates of $\pi_k$. A reasonable initialization for $\pi_k$ is the uniform distribution, i.e., $\pi_1 = \pi_2 = \cdots = \pi_{K+1} = 1/(K + 1)$. Once $\pi_k$ is estimated, $P_k(L|R')$ can be estimated as well because the estimated yield loss due to violations of rule $k$ is now known, which is the total observed yield loss multiplied by $\pi_k$. Specifically, the initial value of $P_k(L|R')$ is set to $(\pi_k \times YL|YI)$, where $YL$ is the observed yield loss.

ii) Expectation: Calculate the posterior probabilities using Bayes’ theorem

$$
\gamma_{ik} = P(y = k|D_i) = \frac{P(D_i|y = k)\pi_k}{\sum_j P(D_i|y = j)\pi_j}
$$

where $P(D_i|y = k)$ is the probability of failure for the $i$th failed IC given that the violation of rule $k$ is the cause of the failure. Thus

$$
P(D_i|y = k) = \begin{cases} 0 & \text{if } D_{i,k} = 0 \\ P_k(L|R') & \text{if } D_{i,k} = 1. \end{cases}
$$

In other words, rule $k$ is assumed not to be the cause of failure if it is not violated. If it is violated, the probability of it causing failure is simply $P_k(L|R')$.

iii) Maximization: Update the current estimates of $\pi_k$ and $P_k(L|R')$ using the posterior probabilities as follows:

$$
\pi_k = \frac{\sum \gamma_{ik}}{N}
$$

$$
P_k(L|R') = \frac{\sum_{i} yD_{i,k}}{\sum \gamma_{ik}}
$$

iv) Iteration: Repeat steps 2 and 3 until convergence.

The EM approach described here corresponds to the implementation of the right sub-flow in Fig. 2 and, as mentioned before, the results of the three-step flow are generated simultaneously. Specifically, the posterior probabilities $\gamma_{ik}$ provides the results for step one (the class label for the $i$th IC failure is simply arg max $k \{ \gamma_{ik} \}$), the yield loss attribution $\pi_k$ directly provides the results for step two, and the estimated $P_k(L|R')$ directly produces the results for step three.

2) Estimation of $P_k(L|R')$: The second approach is to estimate $P_k(L|R')$ instead of $P_k(L|R')$. Again, by definition, $P_k(L|R')$ is the failure rate of the layout pattern described by rule $k$ given rule adherence. If we assume that this failure rate is attributable to the random yield loss of the pattern only, then $P_k(L|R')$ is estimable using critical-area analysis [7]. This approach, however, is not always applicable because the pattern described by a rule $k$ may not be well defined (a density check, for example). In such cases, critical-area analysis of pattern is not possible. Therefore, in this paper, the EM approach of estimating $P_k(L|R')$ is used for YI estimation.

### IV. Experiment

In this section, experiment results involving both real silicon and simulation data are presented. Two sets of experiments are performed to evaluate the effectiveness and accuracy of RADAR. The first experiment applies RADAR to tester fail data measured from an industrial design (an N-vidia GPU in 90 nm technology) to evaluate some of DFM rules utilized in that chip. The second experiment evaluates RADAR using eight representative DFM rules over a set of seven placed-and-routed benchmark circuits [15], [16]. This latter experiment is more comprehensive in nature since we know a priori the actual source of failure. In addition, we can and do explore the impact of other influences on accuracy such as the resolution and correctness of failure diagnosis.

A. Silicon-Data Experiment

In this experiment, RADAR is applied to an N-vidia GPU that is manufactured using a TSMC 90 nm process. The chip is nearly $9 \times 9$ mm, contains over 12 million gates and 700 K flip-flops, and is tested with over 4000 test patterns. Synopsys Tetramax [27] is used to design nearly 9000 failing GPUs to generate diagnosis candidates for each failure. Layout-based pruning [13] is not applied to the failing layout regions. The Mentor Graphics tool Calibre [28] is used for both rule-violation database generation and subsequent queries. Rule-check analyses are not performed for rules involving metal-1 and poly layers since we have incomplete layout information for these layers. Detailed descriptions of the DFM rules are intentionally omitted since they are foundry-confidential. Table II summarizes the rule categories under evaluation (column 1) and their descriptions (column 2). Column 3 gives the rule count for each category. For example, MX.END.1R is applicable to metal layers 2–8, and therefore has seven rules. After rule checking, it is found that certain rules have no violations in the design. These rules cannot be fully evaluated since their RRs cannot be estimated (their failure rates can still be estimated and can be used as a measure of the effectiveness of rule adherence, however). Therefore these rules are removed from further consideration. Column 4 shows the number of remaining rules for each category after removing those rules that do not have any violations. In addition, after classification using EM, insufficient data existed for certain rule categories (e.g., MX.END.3R). Therefore, the EM-based classification is not applied to this data and the statistical analysis is instead performed directly on the rule violation/adherence count collected using the left sub-flow in Fig. 1. In the following discussion, the prefix “MX” is replaced with the actual layer(s) when referring to a specific rule. For example, M2.END.1R refers to a rule in
the category MX.ENC.1R that has been applied to metal-2, while M23.WD.4R refers to a rule in category MXY.WD.4R applicable to metal layers 2 and 3.

As already discussed, critical area is likely a confounding variable in DFM rule evaluation. Thus, critical area for both open and bridge defects are extracted for both passing and failing layout regions using Calibre [28]. To keep the analysis time manageable, only the critical area with a radius equal to half the pitch of the corresponding layer is used since this quantity is proportional to the weighted critical area for most defect sizes of interest. Particularly, radii of 0.08 μm are used for metal layers 2–6, while radii of 0.15 μm are used for metal layers 7 and 8. The critical area extracted for both bridge and open defects are averaged. The mean is used to separate the data (i.e., the DFM rule $X_{ij}$ counts) into two groups, namely the high critical-area group and the low critical-area group. Because critical area analyses for both opens and bridges are performed, the data is partitioned into four groups for confounding adjustment.

Table III summarizes the experiment results. Column 1 gives the rule name. Columns 2 and 3 show the hypothesis testing results using the likelihood ratio test and the corresponding $p$-values, respectively. Similarly, columns 4 and 5 show the hypothesis testing results using the Pearson’s $\chi^2$ test and the corresponding $p$-values, respectively. It is clear that the likelihood ratio test and the Pearson’s $\chi^2$ test give very similar results.

The only rule in which the likelihood ratio test and Pearson’s $\chi^2$ test differs is M8.ENC.2R. This happens because one of the $X_{ij}$ counts is zero. The likelihood ratio test cannot evaluate $H_0$ in this scenario, but the Pearson’s $\chi^2$ test does not require this missing data. A majority of the DFM rules is associated with yield loss (all except M67.WD.4R and M78.WD.4R). The associated $p$-values are extremely small, indicating a very low probability of Type-I error (i.e., associating a DFM rule with yield loss when they are actually independent).

Columns 6–9 show hypothesis testing results for each of the four groups after adjusting for confounding (critical area) using the Pearson’s $\chi^2$ tests only. It is clear that results with confounding adjustment are different from those without confounding adjustment. The subscripts H and L in the column labels indicate whether the critical area is high or low, respectively. For example, the label “Open CAL” means that the data in this group has low critical area for open defects. An entry “N/A” indicates that either $X_{iH}$ or $X_{L}$ is zero, which means hypothesis testing cannot be performed even with Pearson’s $\chi^2$ test. Only 11 rules (highlighted rows in the table) out of 20 rules considered are completely confirmed to have a causal relationship after confounding adjustment. Column 10 of Table III shows the RR for each rule. As mentioned in Section III-C, RR is simply a ratio of the probability of failure given rule violation to the probability of failure given rule adherence. Thus, RR quantifies the changes in the likelihood of failure when the rule is violated. Therefore, to show the relative importance of each rule, the rows of Table III have been sorted according to the RR in decreasing order. Using the interpretation suggested in Section III-C, it is clear that all the via-enclosure rules (MX.ENC.1R and MX.ENC.2R) exhibit strong association with yield loss and therefore should be enforced. The first four rules in the MXY.WD.4R category exhibit nonnegligible association with yield loss but they seem to be not as important after confounding adjustment. This situation suggests that their nonnegligible association with yield loss is contributed by confounding factors. Therefore, they should be further investigated by collecting more data and/or more conclusive analysis such as physical failure analysis. The rest of the rules in the MXY.WD.4R category have negligible association with yield loss and therefore can likely be safely ignored. Surprisingly, adhering to M2.DEN.3R and M3.DEN.3R seems to have an adverse effect on yield. This is probably due to diagnosis inaccuracy. It is interesting to note that all the rules, whose causal associations

<table>
<thead>
<tr>
<th>Rule name</th>
<th>Likelihood ratio test</th>
<th>Pearson’s $\chi^2$ test</th>
<th>Pearson’s $\chi^2$ test (confounding adjustment)</th>
<th>Relative yield loss</th>
<th>Yield impact ($\beta = 0.5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reject $H_0$</td>
<td>$p$-value</td>
<td>Reject $H_0$</td>
<td>$p$-value</td>
<td>Open CAL</td>
</tr>
<tr>
<td>M8.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M3.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M7.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M4.ENC.1R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M2.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M4.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M2.ENC.1R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M5.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M6.ENC.2R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M6.ENC.1R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M3.ENC.1R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M5.ENC.1R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M23.WD.4R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M45.WD.4R</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
<td>0.0000</td>
<td>Yes</td>
</tr>
<tr>
<td>M34.WD.4R</td>
<td>Yes</td>
<td>0.0019</td>
<td>Yes</td>
<td>0.0023</td>
<td>Yes</td>
</tr>
<tr>
<td>M56.WD.4R</td>
<td>Yes</td>
<td>0.0072</td>
<td>Yes</td>
<td>0.0111</td>
<td>Yes</td>
</tr>
<tr>
<td>M67.WD.4R</td>
<td>No</td>
<td>0.9741</td>
<td>No</td>
<td>0.9740</td>
<td>No</td>
</tr>
<tr>
<td>M78.WD.4R</td>
<td>No</td>
<td>0.6719</td>
<td>No</td>
<td>0.6542</td>
<td>No</td>
</tr>
<tr>
<td>M2.DEN.3R</td>
<td>Yes</td>
<td>0.0008</td>
<td>Yes</td>
<td>0.0029</td>
<td>Yes</td>
</tr>
<tr>
<td>M3.DEN.3R</td>
<td>Yes</td>
<td>0.0008</td>
<td>Yes</td>
<td>0.0027</td>
<td>Yes</td>
</tr>
</tbody>
</table>
relationship with yield loss has been confirmed by confounding adjustment, have a high RR (>10). This means that confounding adjustment and RR estimation agree with each other. The resulting rule order in column 10 reveals that the enclosure rules are, in general, more effective than the density rules which suggests that via issues may cause more yield loss than planarity/density problems for this particular process/design.

Column 11 shows the yield loss attribution \(\pi_k\) (i.e., the fraction of failures in each rule category after classification) using EM analysis. The total yield loss for the GPU is assumed to be 15%. Multiplying the total yield loss with yield loss attribution provides an estimate of the yield loss caused by the violation of each rule, which in turn allows the failure rate given violation and YI to be calculated. The YI for each rule, assuming half of all the violations are corrected, is shown in column 12. The sum of the entries of column 12 is 0.072, which is about half of the total yield loss, as expected. Comparing columns 10 and 12, it is clear that the RR and the YI do not rank the rules in the same order. As mentioned in Section III-D, RR is useful as a general guideline; the specific geometries present in a particular design must be taken into account to accurately quantify the YI of the rule violations. RR and YI are related but they do not necessarily correlate. RR measures the likelihood of failure if a rule \(k\) is violated. YI measures the reduction in yield loss if some proportion (which is captured by the parameter \(\beta_k\)) of the existing rule-\(k\) design violations are fixed. So a rule \(k\) with a high RR may not have significant YI if the rule is not significantly violated in the design. On the other hand, a rule that is not very “risky” can have significant YI if there is a significant number of violations in the design that are then fixed (i.e., \(\beta_k\) is high).

Lastly, as mentioned in the introduction, diagnosis inaccuracies/ambiguities can degrade the accuracy of RADAR. Therefore, it is of interest to quantify the accuracy of RADAR to provide some reassurance that this experiment result is valid. This issue is examined next.

B. Simulation-Data Experiment

In this section, two sets of simulation experiments are performed to examine the accuracy and effectiveness of RADAR. The first experiment examines the effectiveness of RADAR to detect association relationships between yield loss and DFM rule violations, while the second experiment evaluates the accuracy of RADAR for DFM rule ranking.

1) Experiment Setup: Both experiments use the in-house defect simulation framework SLIDER (simulation of layout-injected defects for electrical responses) [29], [30] to generate populations of virtual IC failures under various DFM adherence/violation scenarios. SLIDER achieves both efficiency and accuracy using mixed-signal simulation. Specifically, SLIDER injects defects at the layout-level and extracts a circuit-level netlist of the defect-affected region(s). This ensures that the defects are accurately represented. Circuit-level simulation is then performed using the extracted netlist of the defect-affected region(s) to ensure accuracy while logic-level simulation is performed for the rest of the circuit to keep the runtime tractable. By using a mixed-signal simulator (such as Cadence AMS Designer [31]), appropriate signal conversions are performed automatically at the interface that connects the digital and analog domains. SLIDER can generate defects that share a common layout pattern (i.e., systematic defects). This is achieved by using the Mentor Graphics Calibre Pattern Matching [32] tool, which identifies all the layout locations that share a similar layout pattern.

Eight virtual DFM rules are created for use in both experiments. Each rule describes a layout pattern that contains hard-to-manufacture layout features (violation pattern). An additional pattern (adherence pattern) is defined to describe the scenario when the rule is adhered to. Both patterns are needed for creating the virtual failure data. Specifically, both patterns are inputted into the Calibre Pattern Matching tool, which identifies the matching locations for defect injection and simulation in SLIDER. Rows 2 and 3 of Table IV show both the adherence and violation patterns for all eight DFM rules, respectively. Rule 1 (in column 2), for example, describes a pattern whereby three vias in close proximity interact and increase the likelihood for a bridge defect whose location is indicated by the black circle. To address this violation, the via at the bottom is spaced out from the other two. Thus, the bottom via is no longer visible in the adherence pattern (row 3). A short description for each rule is provided in row 4 of Table IV.

It should be emphasized that the pattern matching does not need to be exact. The Calibre Pattern Matching tool can be and has been used to identify locations with similar layout patterns, where similarity is defined by a set of constraints. An example of a constraint would be: the via-to-via distance for the adherence pattern in rule 1 must be between 0.3 and 0.5 \(\mu m\). (The interested reader is referred to [32] for more details on how to use constraints to define pattern similarity). In addition, although the patterns in Table IV are only shown in one orientation, the experiments use all the different orientations of the patterns. From Table IV, it is clear that violation and adherence patterns are generally very similar. Their difference represents the correction that designers use to address the violation. In each pattern, the black circle indicates the location of defect injection. The corresponding defect type (open or bridge) is shown in row 5 of Table IV.

Table V summarizes the characteristics of the designs, tests, and rule violation/adherence statistics for each of the eight rules in each design. In both experiments, seven benchmark circuits [15], [16] are used, as shown column 1 of Table V. These circuits are placed and routed using Cadence First Encounter [33]. The technology used is the TSMC 180 nm CMOS process from the publicly-accessible MOSIS webpage [34]. The tests for these circuits are generated using Synopsys Tetramax [27] with each achieving a 100% fault coverage. The number of gates, layout area, and number of tests for each circuit are shown in columns 2–4, respectively. Columns 5–20 of Table V show the number of adherences and violations for each of the eight DFM rules. The layout locations corresponding to these adherences and violations are randomly sampled, and defects are injected to create two failure populations, each of size 100, of violated- and adhered-rule failures, respectively. These two failure populations (henceforth referred to as the violation and adherence populations) are created for each rule in each design. Unfortunately, some circuits have less than 100 rule violations and/or adherences for some of the rules.
TABLE IV
CHARACTERISTICS OF DEFECTS INJECTED FOR THE VIOLATED AND ADHERED PATTERNS OF EIGHT DFM RULES

<table>
<thead>
<tr>
<th>Rule</th>
<th>Rule 2</th>
<th>Rule 3</th>
<th>Rule 4</th>
<th>Rule 5</th>
<th>Rule 6</th>
<th>Rule 7</th>
<th>Rule 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violation pattern</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 vias interaction</td>
<td>2 vias and a long wire interaction</td>
<td>Via corner-to-corner interaction</td>
<td>Short wire corner-to-corner interaction</td>
<td>3 vias interaction</td>
<td>2 vias and a long wire interaction</td>
<td>Via corner-to-corner interaction</td>
<td>2 vias and a long wire interaction</td>
</tr>
<tr>
<td>Adherence pattern</td>
<td>bridge</td>
<td>bridge</td>
<td>open</td>
<td>open</td>
<td>bridge</td>
<td>bridge</td>
<td>open</td>
</tr>
</tbody>
</table>

In both experiments, confounding adjustment is not validated, however, because of the following reasons.

1) SLIDER does not have the capability to inject defects at signal lines with a particular critical area for a particular defect size.
2) Confounding adjustment is simply performing the same analysis with careful data partitioning and therefore validation without confounding adjustment should suffice.

2) Hypothesis Testing Validation: To investigate the effectiveness of RADAR for detecting association relationships, failure populations with different proportions of adhered- and violated-rule failures are generated by random sampling (with replacement) of the adherence and violation populations described in the experiment setup and combining the samples into a single population. Specifically, in the first iteration, the population has 90 violated-rule failures and 10 adhered-rule failures; in the second iteration, the population has 80 violated-rule failures and 20 adhered-rule failures, and so on. The goal of this experiment is to evaluate how the hypothesis-testing statistic, the associated p-value, and the RR change, as the proportion of violated-rule failures in the population varies. The procedure is repeated for each rule for each circuit. It should be emphasized that, by changing the proportions of adhered- and violated-rule failures, the adherence and violation counts are varied in both the failing and passing layout regions. This is true since the total number of violations/adherences is fixed for a given design. Similar to the silicon experiment, Synopsys Tetramax [27] is used to diagnose all virtual failures to generate diagnosis candidates for each failure. The Mentor Graphics tool Calibre [28] is again used for both rule-violation database generation and subsequent queries. Since seven benchmarks are used in the evaluation, the results for each rule are averaged across designs. To examine how the effectiveness of RADAR changes as the proportion of violated-rule failures varies, the statistic $T$ (for Pearson $\chi^2$ test), the logarithm of its p-value, and the RR are plotted against the percentage of violated-rule failures in the failure population, as shown in Fig. 4. Since eight DFM rules are evaluated, there are eight corresponding curves in each plot, whose legend is shared and shown in Fig. 4(a). In addition, the percentage of violated-rule failures is the independent variable here and therefore forms the common x-axis among the three plots. The dependent variables, namely, the statistic $T$, the logarithm of its p-value, and the RR form the y-axes of each plot in Fig. 4(a)–(c), respectively. Fig. 4(a) clearly shows that the statistic $T$ correlates with the level of failure corresponding to rule violation as expected. The associated p-value (Fig. 4(b)) also decreases, indicating the increasing confidence that the association is valid. To interpret Fig. 4(a) and (b) more precisely, it is important to recall that $T$ can be approximated by $\chi^2_1$, which means that rule violation and yield loss are deemed to be associated when $T > 3.84$ (or equivalently p-value $< 0.05$) at 95% confidence under $H_0$. Using this threshold, Fig. 4(a) and (b) shows that hypothesis testing in RADAR is able to detect
association even when there are only 20% of violated-rule failures. It is interesting to note that rule 3 has a U-shape in the statistic-\(T\) plot in Fig. 4(a). This implies that yield loss becomes associated with rule adherence when the proportion of violated-rule failures is small. This is expected since both the adherence and violation patterns of rule 3 are very common in all the designs (as shown in columns 9 and 10 of Table V). As a result, when the population has a significant portion of adhered-rule failures, rule adherence becomes associated with yield loss. The other rules do not exhibit this behavior because their adherence patterns are generally more common than their violation patterns and therefore require a much greater amount of adhered-rule failures for this phenomenon to occur. In addition, Fig. 4(a) and (b) shows that the changes in statistics \(T\) and \(p\)-value are not monotonous. This is expected since random sampling was used to mix different adhered- and violated-rule failures into a single population.

Fig. 4(c) shows that the RR is increasing as the amount of violated-rule failures increases which indicates that the failure rate given violation is increasing higher than the failure rate given adherence, as expected. Looking again at the RR of rule 3, it is found that it has an average value of 0.31 when the amount of violated-rule failures is 0.1 (i.e., the amount of adhered-rule failures is 0.9), this verifies that RADAR is able to correctly identify the rare scenario that rule adherence can cause higher yield loss.

<table>
<thead>
<tr>
<th>Table VI</th>
<th>DIFFERENT SCENARIOS CONSIDERED FOR DFM RULE RANKING EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule no.</td>
<td>Unknown</td>
</tr>
<tr>
<td>Linear</td>
<td>90  80  70  60  50  40  30  20  80</td>
</tr>
<tr>
<td>Equal</td>
<td>50  50  50  50  50  50  50  50  50</td>
</tr>
<tr>
<td>Dominant-1 (Dom1)</td>
<td>90  90  90  90  90  90  90  90  90</td>
</tr>
<tr>
<td>Dominant-2 (Dom2)</td>
<td>90  90  90  90  90  90  90  90  90</td>
</tr>
<tr>
<td>Dominant-3 (Dom3)</td>
<td>90  90  90  90  90  90  90  90  90</td>
</tr>
</tbody>
</table>

3) Rule Ranking and EM Validation: To evaluate the accuracy of the rule ranking produced by RADAR, several different sets of virtual failure populations are created and summarized in Table VI. Each row corresponds to a scenario while each column specifies the number of injected rule-violation failures. For example, in the “Linear” scenario, 90 violated-rule failures are selected for rule 1, 80 failures are selected for rule 2, etc. In addition, there are 80 adhered-rule failures (ten for each of the eight rules) for each scenario. The adhered-rule failures are labeled as unknown because in the estimation of \(P(I|\tilde{R})\), only rule-violations found in failing layout regions are considered. The adhered-rule failures instead of random failures are included to ensure a nonzero entry in the adherence count, which is necessary for RR estimation. In this experiment, five scenarios are considered, namely: 1) “Linear,” in which the number of violations is linearly decreasing for each rule category; 2) “Equal,” in which the number of violations for each
rule is equal; 3) “Dominant-1 (Dom1),” in which one rule has substantially more failures than the rest; 4) “Dominant-2 (Dom2),” in which two rules have substantially more failures than the rest; and 5) “Dominant-3 (Dom3),” in which three rules have substantially more failures than the rest.

Since failures due to different rule violations exist in the same population, diagnosis accuracy and resolution is very important. (A misdiagnosed candidate can lead to a rule violation being incorrectly attributed as the cause, for example.) To isolate the “noise” that can happen due to inaccuracies/ambiguities in diagnosis, this experiment is performed under two controls: ideal diagnosis and “real” diagnosis. Ideal diagnosis here means that the outcome of diagnosis correctly pinpoints the faulty signal line(s) corresponding to the injected defect with no ambiguity, while real diagnosis can be inaccurate (i.e., the diagnosis outcome does not contain the injected-defect signal line(s)) or suffer from nonideal resolution (i.e., additional signal lines are reported along with the actual injected-defect signal line(s)). Since virtual failure data is used, ideal diagnosis data can be easily obtained by just finding the signal line(s) whose geometries overlap with that of the injected defect. Real diagnosis is obtained by applying commercial diagnosis again to the virtual failure data. In addition, since the exact numbers of violated-rule and adhered-rule failures are known, the parameters of interest (namely, the yield loss attribution $\pi_k$, the RR, and the YI) can also be evaluated to provide the “golden” reference point for comparison. Again, the results are averaged over the seven benchmark circuits used in this experiment.

Fig. 5 shows the results in the form of a normalized bar chart for the yield loss attribution $\pi_k$, the RR, and the YI. Table VII summarizes the average percentage errors (PEs) for these three parameters for both ideal and real diagnosis, where PE is calculated as follows:

$$PE = \frac{|\text{Observed Value} - \text{Correct Value}|}{\text{Correct Value}}.$$ 

Each bar in Fig. 5 has a label in the y-axis that consists of two parts, namely the scenario and the data source. The scenario refers to one of the failure scenarios defined in Table VI while the data source can be “golden,” “ideal diagnosis,” and “real diagnosis.” Fig. 5(a) and row 2 of Table VII shows that the yield loss attribution is accurate in all scenarios even when real diagnosis is used. Fig. 5(b) and row 3 of Table VII shows that RR estimation is not as accurate since it uses both the violation and adherence information. The accuracy is affected if either the adherence or violation count significantly deviates from the actual count due to diagnosis noise or EM classification error (e.g., a rule is wrongly attributed to be the cause of failure during the EM analysis). Further investigation reveals that RR estimation is particularly sensitive to EM classification errors. To give better insight, recall that RR is calculated using

$$RR = \frac{P(L|R')}{P(L|R)} \approx \frac{X_{01}X_{10}}{X_{11}X_{00}} \approx \frac{(X_{01} + 0.5)(X_{10} + 0.5)}{(X_{11} + 0.5)(X_{00} + 0.5)}$$

where $X_{01}$ and $X_{11}$ are the violation and adherence counts, respectively, collected in the failing layout region and are therefore affected by the EM classification error. Consider the scenario that an adhered-rule failure is misclassified to a violated-rule failure. In this case, $X_{01}$ will increase by one while $X_{11}$ will decrease by one. In other words, misclassification errors can result in an overestimation of the numerator and an underestimation of the denominator at the same time. This situation amplifies the error in RR estimation and is found to be the major cause of loss of accuracy in RR estimation.

Despite the mediocre relative-risk accuracy, Fig. 5(c) and row 4 of Table VII shows that the YI, which is estimated using both RR and $P_{LKL}(L|R')$, still agrees very well with the actual data assuming $\beta = 0.5$ (i.e., half of the violations are corrected). The accuracy of YI is expectedly worse than that of yield loss attribution because of the additional error introduced by the RR estimation. It is also clear from Table VII that all three parameters have a worse accuracy in the real-diagnosis scenario than in the ideal-diagnosis scenario, as expected.

It is also worth explicitly noting the error in the relative ranking of the DFM rules. For the five scenarios considered in Table VI, namely “Linear,” “Equal,” “Dom-1,” “Dom-2,” and “Dom-3,” the possible rule rankings include 8, 1, 2, 2, and 2, respectively. For Fig. 5(a) and (c), it can be observed that the derived rank of each rule very much matches the rank given in the golden result for both ideal and real diagnosis. For example, for “Dom-3,” rules 1, 2, and 3, are always in the larger class while the remaining rules (4–8) all belong to the smaller class, demonstrating no error in the ranking. For Fig. 5(b) (RR), the number of possible ranks depends on the number of sites in the design where the rule is applicable and thus is different than the number of ranks for Fig. 5(a) and (c) which are directly based on the number of failures given in Table VI. It is still true, however, that ranking error (or lack thereof) can be obtained from Fig. 5(b). For example, for “Dom2” it can be observed that the lowest-ranked rule is rule 8 for the golden, ideal, and real diagnosis scenarios. While it is somewhat straightforward to derive the rule rankings from Fig. 5, it is more easily observed in tabular form. Consequently, Tables VIII and IX provide the rankings for the results given in Fig. 5(a) and (b), respectively. There is no ranking error in Fig. 5(c). Highlighted table entries indicate a mismatch between the golden (actual) ranking and the derived ranking.

Since EM classification errors directly affect the accuracy of subsequent analyses, EM classification accuracy is carefully examined for both the ideal- and real-diagnosis scenarios. Since the class label for each failure is known in the virtual data, EM classification accuracy can be evaluated by simply checking the proportion of failures whose class labels have been correctly identified. The results are summarized in Table X. The leftmost column provides the seven benchmark circuits used in this experiment. Columns 2–6 show the EM classification accuracy using ideal diagnosis for the five scenarios defined in Table V, respectively, while columns 7–11...
TABLE VIII
RULE RANKS IN TERMS FOR YIELD LOSS ATTRIBUTION FOR FIVE DIFFERENT FAILURE SCENARIOS, NAMELY LINEAR, EQUAL, DOM1, DOM2, AND DOM3

<table>
<thead>
<tr>
<th></th>
<th>Linear</th>
<th>Equal</th>
<th>Dom1</th>
<th>Dom2</th>
<th>Dom3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gold</td>
<td>Ideal</td>
<td>Real</td>
<td>Gold</td>
<td>Ideal</td>
</tr>
<tr>
<td>R1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

TABLE IX
RULE RANKS IN TERMS FOR RELATIVE RISK FOR FIVE DIFFERENT FAILURE SCENARIOS, NAMELY LINEAR, EQUAL, DOM1, DOM2, AND DOM3

<table>
<thead>
<tr>
<th></th>
<th>Linear</th>
<th>Equal</th>
<th>Dom1</th>
<th>Dom2</th>
<th>Dom3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gold</td>
<td>Ideal</td>
<td>Real</td>
<td>Gold</td>
<td>Ideal</td>
</tr>
<tr>
<td>R1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>R2</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>R3</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>R4</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>R5</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>R6</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>R7</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>R8</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

TABLE X
EM CLASSIFICATION ACCURACY FOR IDEAL AND REAL DIAGNOSES

<table>
<thead>
<tr>
<th></th>
<th>Ideal diagnosis</th>
<th>Real diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Equal</td>
</tr>
<tr>
<td>c5351</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>c7552</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>s13207</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>s15850</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>s35932</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>s38584</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>s9234</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.89</td>
<td>0.88</td>
</tr>
</tbody>
</table>

show the results with real diagnosis. The last row of Table X shows the average accuracy. As expected, classification accuracy is worse with real diagnosis, reflecting the ambiguities inherent in logic-level diagnoses. The average classification accuracies for ideal and real diagnoses are 88.5% and 76.3%, respectively.

C. Result Summary

In this section, experiment results are presented using both silicon and simulation data. The silicon experiment using the Nvidia GPU fail data demonstrates how RADAR can be applied in an industrial context to identify important DFM rules and quantify their YI. This is useful for improving yield in future designs. In addition, rule-ranking results may indicate the relative importance of the processing steps (e.g., via issues seem to cause more yield loss than density issues), which suggests that process improvement effort can also be prioritized for the current product. Comprehensive simulation experiments using eight virtual DFM rules on seven benchmark circuits have demonstrated the accuracy and effectiveness of RADAR. Hypothesis testing is first examined. The proportion of violated-rule failures is actively changed and the resulting changes in the statistic $T$, $p$-value, and RR are examined and interpreted. The results show that RADAR is able to detect association effectively. Rule ranking parameters, namely yield loss attribution, RR, and YI, are then evaluated under different scenarios. Additionally, two sources of inaccuracy are examined, namely EM classification errors and diagnosis ambiguities. The results show that, despite nonideal diagnosis and imperfect EM classification, YI estimation is still reasonably accurate.

For both the silicon and simulation experiments, only certain types of rules are considered. For the silicon experiment involving an Nvidia GPU design, this means that there can be other types of rules that are more or less important than the rules analyzed. On the other hand, for the simulation experiment, the rules chosen for analysis, while limited in nature, have no consequence on the efficacy of RADAR since “golden results” are available for checking accuracy. This implies that other rules that target other layout geometries will not in our opinion impact the practicality of RADAR.

V. Conclusion

In this paper, a novel methodology called RADAR has been developed to study the effectiveness of DFM rules using the results of IC failure diagnosis. Since the cost-benefit tradeoff of DFM is still not well understood, this paper serves as a first step in systematically making this assessment. Application of RADAR to a fabricated 90 nm GPU from Nvidia produced rule rankings that showed that some rules are effective in preventing yield loss while others are not. This provides a quantitative basis to prioritize rule adherence in future designs. At the same time, process improvement effort can also be prioritized to improve the yield of the current product. Additionally,
comprehensive simulation experiments have demonstrated the capability of RADAR to effectively determine the relevance of a DFM rule using statistical analysis and accurately quantify its YI. Deployment of RADAR will therefore allow designers to make informed decisions concerning DFM-rule adoption. We believe the accuracy of RADAR can be improved however. Specifically, improved diagnostic resolution, the elimination of superfluous rule violations, and an improved EM formulation are the subject of future enhancements, all of which can be simultaneously incorporated into RADAR. Improved resolution means fewer rule violations are implicated for each chip failure and thus reduces noise from the EM formulation. Elimination of superfluous rules can be performed based on the defect type that the rule is meant to prevent. For example, if diagnosis strongly indicates that the failure is caused by a two-way bridge, then the rule violation that is intended to prevent an open defect can be eliminated as a possible source of failure. Finally, the EM formulation can be improved using more insightful domain knowledge. For example, instead of simply attributing to a failed IC a “yes = 1” or “no = 0” for each type of rule violation, it is more accurate to include a weighted value based on the number and severity of violations. All of these areas present significant opportunity for improving RADAR.

REFERENCES

[34] [Online]. Available: http://www.mosis.com

Wing Chiu Tam (S’03) received the B.S. degree from the National University of Singapore, Singapore; and the Ph.D. degree from Carnegie Mellon University, Pittsburgh, PA, USA, in 2003 and 2012, respectively. He was with Advanced Micro Devices, Sunnyvale, CA, USA, and Agilent Technology from 2003 to 2006, Singapore. He is currently with Nvidia, Santa Clara, CA, USA, working on in-house DFT tools and methodology development. His current research interests include CAD for DFM, IC test, defect modeling, and yield modeling.

Shawn Blanton (S’93–M’95–SM’03–F’08) received the B.S. degree in engineering from Calvin College, Grand Rapids, MI, USA, the M.S. degree in electrical engineering from the University of Arizona, Tucson, AZ, USA, and the Ph.D. degree in computer science and engineering from the University of Michigan, Ann Arbor, MI, USA, in 1987, 1989, and 1995, respectively. He is currently a Professor of Electrical and Computer Engineering with Carnegie Mellon University, Pittsburgh, PA, USA, where he also serves as a Director at the Center for Silicon System Implementation, and as the Founder and Leader of the Advanced Chip Testing Laboratory. His current research interests include test and diagnosis of integrated heterogeneous systems.