Board Manufacturing Test Correlation to IC Manufacturing Test

C. Glenn Shirley[1], W. Robert Daasch[1], Phil Nigh[2] and Zoë Conroy[3]
[1]ECE, Portland State University, Portland OR; [2]IBM Inc., [3]Cisco Systems, Inc.

Abstract

A large end-to-end database was created by joining an IC test factory database with a board contract manufacturer (CM) database using the unique electrical identity fused into the IC and read at IC and CM test steps. A new rank correlation method was used to measure statistical dependencies of IC failure rates in the CM factory on failure rates and attributes measured in the IC factory. IC factory dependency, including radial wafer symmetry, was seen in the CM data providing opportunity to improve CM failure rate at the cost of IC yield. A special kind of receiver operating characteristic (ROC) was constructed to quantify the IC factory yield cost of CM failure rate improvements for hypothetical IC factory screens additional to the screens which generated the database. A screen based on RAM fuse repair count was found to be better than one based on blocking of edge dies.

1. Introduction

Ideally an IC's failure probability at a board assembly contract manufacturer's factory (CM) depends only on conditions such as particular manufacturing equipment, location on the board, etc. in the CM factory, and not on conditions such as location on the wafer in the upstream IC factory. In practice, imperfect test coverage in the manufacturing test screen at the IC manufacturer has been seen [1] to expose the CM factory to a population of ICs which still has some failure rate dependency on IC manufacturing conditions. This paper is a case study which finds and quantifies the most important failure rate dependencies on IC factory test attributes for ASICs used in a high-end communication board manufactured in a CM factory. The information can be used to set screening rules and limits in the IC factory to reduce the part of NTF (no trouble found at IC test for CM ASIC fails) attributable to the IC factory.

The strategy of the case study is as follows:

- Join IC factory data to CM factory data using a unique identifier (ID-tag) fused into each IC and recorded at each test socket in both factories.
- Use a group-ranking method, new to test applications, to isolate the strongest IC factory to CM factory dependencies and so identify IC factory attributes-of-interest which most strongly affect CM failure rates.
- Compute the receiver operating characteristic (ROC) for the IC yield/CM defect level tradeoff associated with IC factory attributes-of-interest identified in the previous step.

The IC factory and the CM factory are often different companies, so communication required to specify data requirements and change the IC factory test process based on CM data can be difficult. For example, parametric tests done on a sampling basis in the IC factory will not have been done on every die that reaches the CM factory yet the dependency of the CM failure rate on the particular parametric measurement is of interest. The group-ranking method is suited to naturally occurring data with no particular inter-factory data requirements. The group-ranking method also determines inter- and intra-factory dependency among attributes of very different types (spatial, parametric, intra-die test, pass/fail, finely or coarsely binned, etc.). The case study shows robustness of the group-ranking method against exigencies of real-world data.

The paper is organized as follows. Section 2 describes the way the database was built (2.1), and how the data were prepared using a grouping strategy (2.2) prior to measurement of rank dependency among attributes. Section 2.3 describes an old rank correlation method newly applied to test. Section 3.1 gives some descriptive rank statistics of the data and section 3.2 shows how to isolate the most important dependencies in the case study using the method described in 2.3. Section 3.3 quantifies the IC yield versus CM failure rate sensitivity using a special kind of receiver operating characteristic (ROC) analysis and compares effectiveness of screening strategies based on different ways of grouping dies in the group-ranking method. Sections 4, 5, and 6 wrap up the paper with discussion of limitations, possible improvements, and benefits.

2. Strategy

2.1 Building an End-to-End Database

The product used in this study is a large communication integrated circuit board assembled in the CM factory using 12 copies of an ASIC produced by the IC manufacturer. Each IC has a unique identifier (ID-tag) fused into it which is read at manufacturing test steps in the IC factory and in the CM factory. Each ID-tag encodes lot and wafer identity as well as die location on the wafer. Test steps in the CM factory also recorded the unique serial number of each board, and the location on the board of each of the 12 ICs.

978-1-4799-4722-5/14/$31.00 ©2014 IEEE

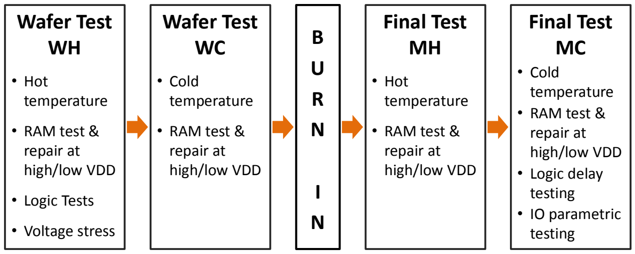

Figure 1. IC manufacturing test flow and test content. ICs were packaged after cold wafer test, WC. The ID-tag of each IC was recorded at each test socket.

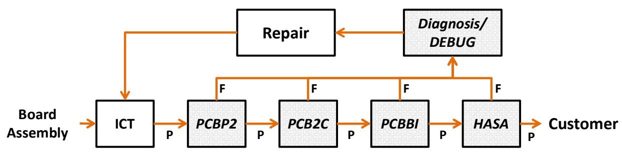

Figure 2. CM board manufacturing test flow. ID-tagged data for each IC in the board is recorded at each shaded/italic board test step. Failing boards are sent to a diagnosis/debug area where further board testing and logging of ID-tagged data is done.

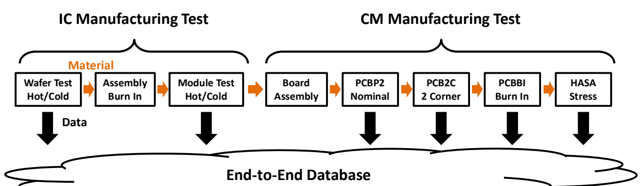

Figure 3. End-to-end database built by joining data from the IC manufacturing test flow and data from the CM test flow using the unique ID-tag of each of the 12 ASICs on each board.

The IC test material flow and test content is shown in Figure 1. ICs were packaged after the cold wafer test (WC) and sent through burn in and final test. Not all tests were performed nor was burn-in done on all ICs.

The CM factory test material flow is shown in Figure 2. After an initial test at nominal conditions (PCBP2) the board is tested at voltage and temperature corners (PCB2C) followed by a test after board-level burn in (PCBBI) and a non-destructive mechanical and thermal cycle stress monitor (HASA). Rework is a critical to the board level manufacturing flow, and data including ID-tag of all ASICs on the board is acquired during diagnosis/debug of any failing boards. Keep in mind that board-level testing is focused on the board, not the ASIC of interest in this study. The set of ASIC-related board failures is a subset of all CM board failures for which an ASIC replacement in diagnosis/debug/repair *changed* the board-level failure signature. The replaced ASIC was tagged as a failure in the CM. For example, an ASIC which when replaced changes the board signature to "pass" is a CM-failing ASIC. Failing ASICs are tagged with the test step (PCBP2, etc.) which routed the board to diagnosis/debug, and by the category of board failure it is associated with (traffic, memory, logic, etc.). About 240,000 ICs were shipped to the CM factory, so about 20,000 boards were sent through the test flow in Figure 2.

For proprietary reasons this paper cannot be specific about the number of ICs manufactured in the IC factory, or about failure rates at each IC factory or CM factory test step.

Data generated from the test steps in the IC factory and the CM factory were joined on the unique identity of each IC read at each test step in both factories to build the end-to-end database shown in Figure 3. A "perfect" board generates 12 (ASICs) x 4 (test steps) = 48 records. A board going through diagnosis/debug/repair generates many more records so it will be appreciated that the database was large. We mention in passing that many CM data integrity issues needed to be handled while building the database. Data integrity information acquired while constructing the database was very useful to the CM.

2.2 Preliminary Data Analysis – Grouping Strategy

The IC factory passes or fails each IC at the test steps WH, WC, MH, and MC shown in Figure 1 and parametric data were recorded on a sampling basis. Not every IC that reaches the CM has had parametric data recorded at the IC test factory, yet the relationship of parametric data to fail rates in the CM is of interest since a parametric measurement, combined with a test limit, may give an improved test screen. A way to avoid the usual requirement [2] that each die has all attributes measured is to group dies by a chosen "grouping attribute", and to compute all other attributes of each group from corresponding attributes of dies in each group. Table I gives the 59 group attributes, divided into 4 categories, used in this study. In the IC factory parametric values marked by * in Table I were acquired for every die. Other parametric attributes were sampled at rates of between 1% and 4%, except for VMAuto which was sampled at 55%.

Table I Group attributes used in this study. The color of text matches the color of bars and lines in Figures 6, 7, 8, and 9. Attributes marked with * were measured for every die.

Attribute	Description	Attribute Category
x	x coordinate	Spatial
y	y coordinate	
R	Distance from wafer center	
WHfr	Fail fraction, wafer-level hot	Fail fractions in IC test.
WCfr	Fail fraction, wafer-level cold	
MHfr	Fail fraction, package-level hot	
MCfr	Fail fraction, package-level cold	
WafHIDD*	IDDQ at wafer test (hot, core area of chip)	Median parametric value in IC test.
WafCIDD*	IDDQ at wafer test (cold, core area of chip)	
ModIDD*	IDDQ at module test (core area of chip)	
Repairfuses*	Number of RAM repairs	
ProcMon*	Average of process monitors (smaller ⇒ faster)	
WMinVDD1-4	Min VDD at Wafer Test - various patterns	
MMinVDD1-4	Min VDD at Final Test - various patterns	
VM1 - 10	Min VDD for various LOS Delay patterns	
VM11 - 26	Min VDD for various LOC Delay patterns	
VMLB	Min VDD for LBIST	
VMAUTO	Min VDD for Autotest	
CMfr	Total fail fraction in CM	Fail fractions in board factory (CM).
PCBP2fr	Fail fraction at PCBP2 in CM	
PCB2Cfr	Fail fraction at PCB2C in CM	
HASAfr	Fail fraction at HASA in CM	
PCBBIfr	Fail fraction at PCBBI in CM	
DEBUGfr	Fail fraction at Diagnose/debug in CM	
TRAFFICfr	Fail fraction in "traffic" category in CM	
MEMORYpf	Fail fraction in "memory" category in CM	
LOGICpf	Fail fraction in "logic" category in CM	
BOOTpf	Fail fraction in "boot" category in CM	
OTHERpf	Fail fraction in other categories in CM	

When the xy location is chosen as the "grouping attribute" there are 159 groups, each corresponding to a die location. The spatial attributes (x and y and radial distance from wafer center) of each group are computed from the x and y coordinates of the die. Fail fractions at IC factory test steps and at CM factory test steps become failure rate attributes (suffix "fr" in Table I) at the group-level by dividing the number of failures in the group by the number of passes and fails in the group, ignoring units which don't have a pass or a fail in the group. For attributes which are parametric values in the IC factory, the median of the attribute for all ICs which fall into the group defines the attribute value for the group.

Besides xy location, the dies may be grouped by any attribute by which a die is classified. For example, when "Repairfuses" is chosen as the grouping attribute there are 438 non-empty groups corresponding to the number of fuse repairs required for a die. Group attributes other than the grouping attribute are computed as for xy groups.

2.3 Surveying Dependency of Attributes

The traditional way of measuring correlation between pairs of attributes is to compute Pearson's rho which is the sample covariance of attribute values divided by the product of the standard deviations of attribute values. A problem with Pearson's rho is that it depends on both the "scatter" and the non-linearity of the relationship between attributes. For example, for $a = b^4$ (no scatter) with b randomly and uniformly distributed in the range $1 < b < 10$ Pearson's rho is 0.89. One would prefer a measure of correlation between attributes that is unity for a monotonic functional dependence (no scatter) between attributes, irrespective of the functional form of the dependence. Principal component analysis [3] uses the multidimensional analog of Pearson's rho, and so has the problem. The problem can be avoided by using statistics which depend on *ranks* of the attributes, not on the *values* of the attributes. One rank statistic is the correlation coefficient of the ranks of the attribute values, called Spearman's rho. Spearman's rho is computed in the same way as Pearson's rho except that ranks of attributes, rather than their values are used in the calculation. Another rank statistic called Kendall's tau, τ, measures the portion of pairs of sample instances that have matching ranks of the two attributes. In the example involving a and b both Spearman's rho and Kendall's tau are unity because the ranks of the two attributes for every pair of instances are the same. When the two attributes do not have a monotonic functional dependence like $a = f(b)$ Spearman's rho and Kendall's tau will differ from unity and lie in the range [-1,1]. The "degree of scatter" of *ranks* measured by Spearman's rho and Kendall's tau is called "dependency" which is to be distinguished from the term "correlation" which involves *values*.

Kendall's tau was chosen as a measure of dependency between pairs of attributes for three reasons:

- A rank-based measure (Spearman or Kendall) is well suited to quantifying the dependency between unlike attributes such as spatial measures (eg. xy location of a die on the wafer), failure rates, and parametric measurements. Rank-based dependency measures of dissimilar attributes do not depend on the shape of distributions of attributes taken individually (marginal distributions).
- Methods exist [4][5] to compute Kendall's tau between rankable attributes with different, even very different, numbers of ties. In test applications every attribute measured is binned, either coarsely (eg. pass/fail) or finely (eg. three 9's resolution) so many dies will have the same value (be tied) for a given attribute.
- A method exists [6] to compute the standard deviation of Kendall's tau, including the effect of ties, for the null hypothesis of independence. The significance of τ measured for a pair of attributes is gauged by comparing it to the standard deviation of τ expected if the attributes were independent.

A method given by Simon [5][6] for computing tau and the standard deviation for the null hypothesis considering ties for many pairs of attributes was implemented in Python code[1]. Computation time is roughly proportional to the square of the sample size, and the square of the number of attributes. In principle the calculation can be done when the sample consists of instances of individual dies. Since there are hundreds of thousands of dies, computation times are prohibitive when instances are individual dies. But the grouping strategy used to handle attributes for which many dies have missing data reduces the sample size from hundreds of thousands to a few hundred at most, if groups are regarded as instances. For typical group and attribute counts run time on a PC was a few minutes.

Simon's method of computing tau and the null hypothesis standard deviation considering ties is too technical to describe here, but the idea can be simply explained in the case of no ties. Suppose attributes a_1 and a_2, are known for each of n groups so that every pair of groups may ranked by a_1 and by a_2 without ties. The number of group-pairs, $n(n-1)/2$ comprises k "concordant" pairs and d "discordant" pairs. For a concordant pair, the relative ranks of a_1 for a group pair is the same as the relative ranks of a_2 of the group pair. For a discordant pair the relative ranks are opposite. Kendall's tau for the sample is $\tau = (k-d)/(k+d)$. If the two attributes were statistically independent and n groups were sampled many times from an infinite population, then the measured τ has an asymptotically normal distribution centered on zero with standard deviation [4][6] given by

$$\sigma_\tau = \sqrt{\frac{10+4n}{9n(n-1)}} \qquad (1)$$

[1] The code also implements Simon's generalization of the concept of Kendall's tau to any even number of attributes but the present paper uses only pairs of attributes.

The version of Eq. (1) taking into account ties given in [6] was used in the analysis.

Simon's estimate [6] of the standard deviation of tau for the null hypothesis of independence is crucial to gauging the significance of any measured value of tau. In this paper a pair of attributes will be taken to have significant dependence if the absolute value of a "standardized tau", $z = |\tau/\sigma_\tau|$, exceeds 3. $z = 3$ corresponds to a 99.7% confidence level.

3. Results

3.1 Descriptive Statistics

Wafer maps of the ranks of failure rates for each xy group (die location) are shown in Figure 4 as a function of IC factory test step, and in the CM factory.

It is apparent in Figure 4 that the radial dependence is reduced by successive test steps. Ideally, ICs would be shipped to the CM with no trace of the radial pattern, but Figure 4 shows a residual radial dependence observed in the CM (board) factory.

The data in Figure 4 are also shown in Figure 5 as a sequence of rank correlation plots, with Kendall's tau measuring the dependence between failure rate ranks of successive test steps.

3.2 Dependency Analysis

Extending the idea of measuring taus of failure rates of xy groups shown in Figure 5 to measuring tau between *every* pair of xy group attributes in both factories, absolute values of standardized taus for all 1711 pairs of attributes described in were computed. Of the 1711 pairs 627 pairs linking the IC factory and the CM factory are of primary interest. Of the 627 pairs, the standardized tau of 57 pairs involving the rank of the CM overall failure rate (CMfr) with all other attributes is shown as a pareto plot in Figure 6. Strong dependency ($z > 3$) between CMfr and the top 4 attributes is expected since TRAFFIC, PCBP2, etc. are subsets of CMfr. Also expected is the strong dependency between R, WHfr, WCfr, MHfr, and MCfr already observed in Figure 4 and Figure 5. Newly identified with strong dependency are: WMinVDD3, VM14, VM19, Repairfuses, WMinVDD4, VM4, and ProcMon. Of these, Repairfuses and ProcMon were measured on every die.

xy grouping was used because the radial pattern "escaping to" the CM (Figure 4) suggested that significant IC to CM dependency would be observed with xy grouping. However, grouping dies by other attributes is also possible and groupings more (and less) sensitive to the IC to CM dependency may exist.

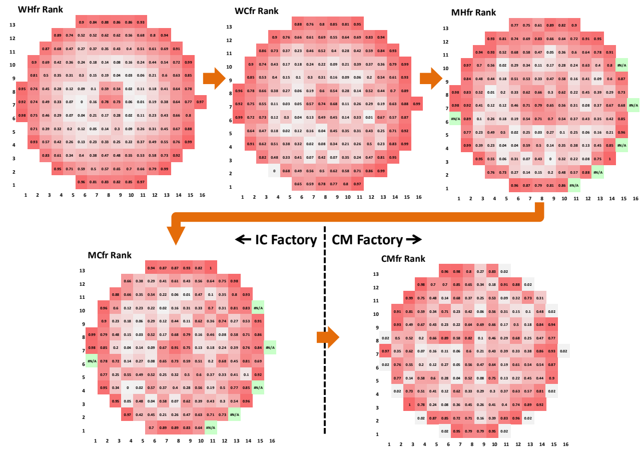

Figure 4. Wafer maps of failure rate rank at each test step in the IC factory and at the CM. A pronounced radial dependence becomes more random at each test step but does not disappear in the CM.

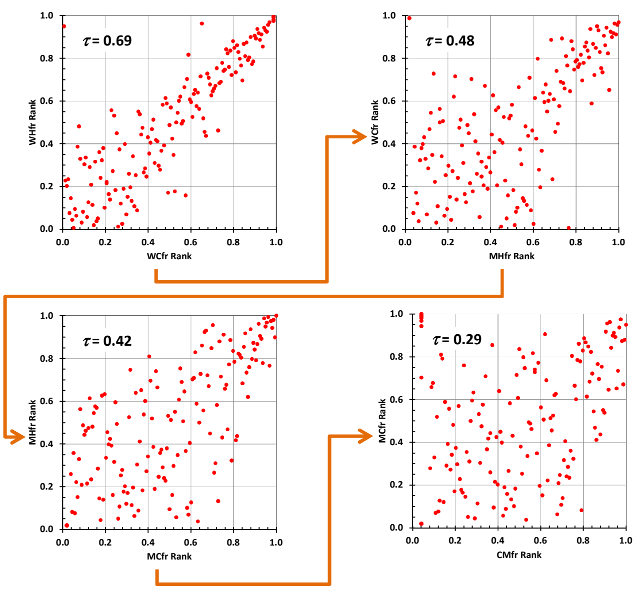

Figure 5. Kendall's tau measures the degree of rank correlation of failure rates between successive IC test steps and overall fail rates in the CM factory. Each point is an xy group (die location).

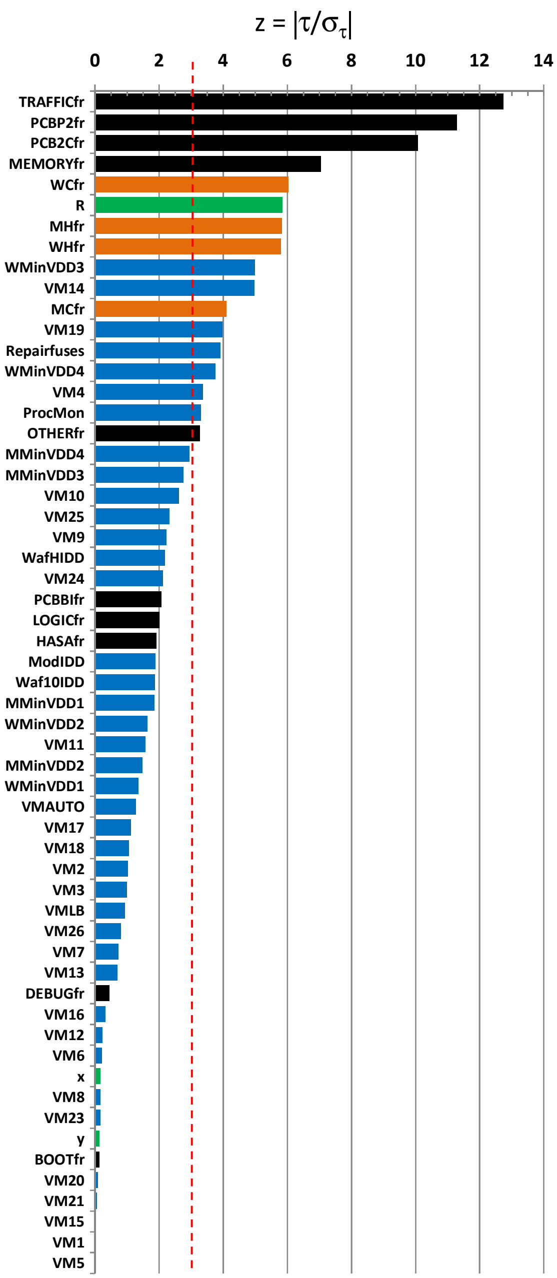

Figure 6 xy group standardized tau, z, for overall CM failure rate CMfr vs all other attributes listed in **Table I**. Bar color matches categories in the table. The $z = 3$ significance limit is shown.

3.3 IC Yield vs CM Failure Rate Sensitivity Analysis

Attributes involved in the strongest dependencies (measured by z) are candidates to be used for IC factory test screen improvements. The effectiveness of a candidate attribute can be gauged by constructing a receiver operating characteristic (ROC) for an hypothetical IC factory screen supplementing the test process that generated the data. The ROC shows the IC factory yield loss and the CM factory defect fraction improvement for IC factory screens available by exploiting the candidate attribute.

The ROC is constructed as follows:

- Divide dies into groups labeled by a grouping attribute g. Grouping attributes are attributes by which individual dies may be classified. For example, g may be xy where x and y are coordinates of the die on the wafer.
- Compute all other group attributes besides the grouping attribute for each group by the method described in section 2.2. For example, a group parametric attribute is the median value of the parametric attribute of dies in the group.
- Assign a rank, i, to each group g by ranking groups according to the candidate attribute. The candidate attribute is not necessarily the same as the grouping attribute. For example, xy groups may be ranked by R, the distance from the wafer center or by any other group attribute. More than one group may have the same rank.
- For each group, count dies passing and failing for each test (WH, WC, MH, MC) and for the entire CM factory. For example, numbers of units tested, passing, and failing at WH are $c_{\mathrm{WH}} = p_{\mathrm{WH}} + f_{\mathrm{WH}}$, p_{WH}, f_{WH}. Note that not all passing units are necessarily sent to the next test step so, for example[2], $c_{\mathrm{MH}} \le p_{\mathrm{WC}}$.
- Compute IC factory yield. Suppose that the hypothetical supplementary screen kills dies at the first test step (WH) in the IC factory according to whether the rank of the group they belong to exceeds a reference rank, r. In this case the yield of test WH is

$$Y_{\mathrm{WH}}(r) = \sum_{\{g(i):i \le r\}} p_{\mathrm{WH}}(g) \Big/ \sum_{\{\forall g\}} c_{\mathrm{WH}}(g) \tag{2}$$

and the yields of subsequent test steps are

$$\begin{aligned} Y_{\mathrm{WC}}(r) &= \sum_{\{g(i):i \le r\}} p_{\mathrm{WC}}(g) \Big/ \sum_{\{g(i):i \le r\}} c_{\mathrm{WC}}(g) \\ Y_{\mathrm{MH}}(r) &= \sum_{\{g(i):i \le r\}} p_{\mathrm{MH}}(g) \Big/ \sum_{\{g(i):i \le r\}} c_{\mathrm{MH}}(g) \\ Y_{\mathrm{MC}}(r) &= \sum_{\{g(i):i \le r\}} p_{\mathrm{MC}}(g) \Big/ \sum_{\{g(i):i \le r\}} c_{\mathrm{MC}}(g) \end{aligned} \tag{3}$$

so the overall IC factory yield is

$$Y(r) = Y_{\mathrm{WC}}(r) \cdot Y_{\mathrm{WH}}(r) \cdot Y_{\mathrm{MH}}(r) \cdot Y_{\mathrm{MC}}(r)\,. \tag{4}$$

- Compute the overall CM factory failure rate as follows

$$FR(r) = \sum_{\{g(i):i \le r\}} f_{\mathrm{CM}}(g) \Big/ \sum_{\{g(i):i \le r\}} c_{\mathrm{CM}}(g)\,. \tag{5}$$

- The ROC is a parametric plot of IC factory yield, Y, versus CM factory fail fraction, FR, using r as the parameter.

---

[2] The inequality is a useful data integrity check.

The obvious grouping of dies is xy grouping. Figure 7 shows ROCs for xy groups ranked by several choices of candidate attribute. The ROCs are normalized by the IC factory yield and the CM fraction failing without application of any screening of xy groups. That is, the point (1,1) in the upper right corner represents non-application of the hypothetical screen, and as xy groups are screened, one moves away from that corner down the ROC curve. The heavy green curve represents screening xy groups with R exceeding smaller and smaller values of radial distance from the wafer center as one moves down the ROC from the (1, 1) point. For example, when all dies with R > 6.2 are screened by the hypothetical screen, the cost is 9% of the IC yield (without the hypothetical screen) for a benefit of 10% reduction of CM fail fraction (without the hypothetical screen).

Two limiting ROC curves corresponding to special candidate attributes are of interest. The first limiting ROC curve is when groups are shuffled by making the candidate attribute a random number (uniform on [0,1]). The random ROC curve ("Rand") shows the case when there is no information to rank groups so that any screen will reduce yield with no failure rate benefit. Generation of several random ROC curves shows the effect of sample size. The second limiting ROC curve is given by ranking groups in ascending[3] order of each group's fraction of ICs failing in the CM

$$CMfr = f_{\mathrm{CM}}(g)/c_{\mathrm{CM}}(g)\,. \tag{6}$$

The ROC with the CMfr candidate attribute is a limiting case because the highest failure rate groups are eliminated first from the sum in Eq. (5) as r is reduced from the maximum rank. The CMfr-based ROC cannot be used to improve quality delivered to the CM because it is based on post-hoc knowledge of events in the CM, but it is useful because it bounds all ROC curves for the given data and the particular grouping attribute.

xy-grouped ROCs for the other candidate attributes given in Figure 7 were chosen by selecting attributes measured on every die as follows

- R, WHfr, MCfr, and Repairfuses all showed significant dependency (large z) in Figure 6.
- WafHIDD showed weaker dependency in Figure 6.

Unlike many ROCs, the curves in Figures 7, 8, and 9 are not monotonically decreasing. This is because both denominator and numerator in Eq. (5) are subject to group-to-group sampling error. In the limit of very large samples all curves would be smooth and monotonically decreasing. In this limit no curve could exceed unity on the normalized CMfr (vertical) axis. The sampling error can be gauged by the variation of the random ROC curve ("Rand" in Figures 7, 8, and 9) which in the very large sample limit would be a horizontal line at unity on the vertical axis.

[3] Worst case ROCs can be constructed by ranking groups in descending order of CMfr, but these are of no interest.

Once a ROC is established, it may be used to screen further dies. An accept/reject limit on the candidate attribute value is set to accept or reject groups. Two ways of using the ROC may be identified depending on how die membership in a group is established:

<u>On-tester screen.</u> When individual dies have the same attribute as the candidate attribute then die membership in an accepted or rejected group is apparent from an individual die attribute so that a die-by-die on-tester accept/reject decision can be made. For example dies in xy groups may be screened on the tester by the value of R of the xy groups. And dies in Repairfuses groups may be screened on the tester by their value of Repairfuses. In general, a ROC derived when the candidate (ranking) attribute is the same as the grouping attribute may be used for on-tester die-by-die screening decisions. ROCs useful for on-tester screens are shown as heavier lines in Figures 7, 8, and 9.

<u>Off-tester screen.</u> When statistics from multiple dies must be accumulated to calculate an attribute before a group (and the dies therein) can be classified as accepted or rejected, "off-tester" screening decisions in the manner of Madge et. al [7] will be required. For example, the WHfr ROC in Figure 7 based on the fail fraction observed at the hot wafer insertion requires data to be accumulated over an entire lot (strictly the entire dataset for the example shown) before a enough data have been acquired to classify the xy group by its value of WHfr and thereby accept or reject dies in the xy group.

Figure 7 shows that Repairfuses has a better ROC than the ROC based on R because it is closer to the ideal CMfr ROC, but it can't be used for an on-tester screen for xy groups. However, Repairfuses *can* be used for an on-tester screen for dies grouped by the Repairfuses attribute. Accordingly, ROCs shown in Figure 8 were constructed by grouping dies by the Repairfuse attribute measured on each die. Notice that the CMfr ROC grouped by Repairfuses (Figure 8) gives a potentially better screen than grouping by xy (Figure 7). The heavy dashed blue curve gives the ROC for the on-tester screen based on ranking Repairfuses groups by the Repairfuses attribute. When all dies with Repairfuses > 50 are screened, the cost is 4% of the IC yield without the screen for a benefit of 10% reduction of CM fail fraction without the screen.

Figure 9 shows ROC curves for dies grouped by Idd measured at the hot wafer insertion. It is apparent that there is no opportunity for a screen based on this attribute.

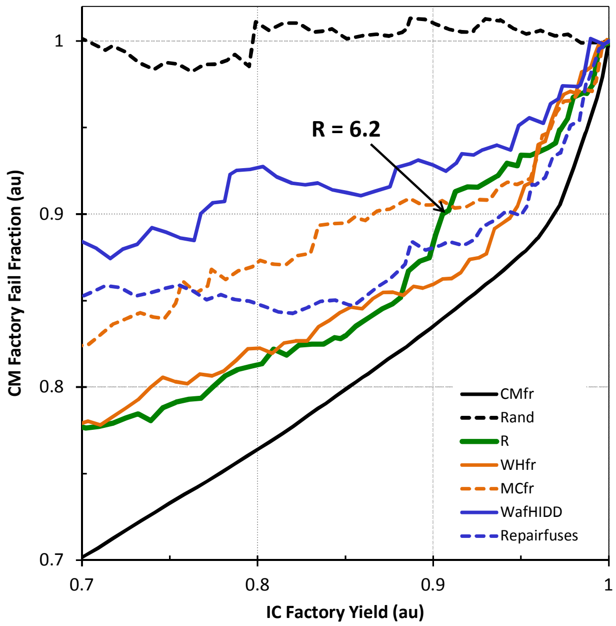

Figure 7. ROCs for xy-groups of dies ranked by attributes in the legend. All ROCs are bounded by CMfr and Rand. The heavy line (R) is the ROC for an on-tester die-by-die R-limit screen (edge die block) in the IC factory.

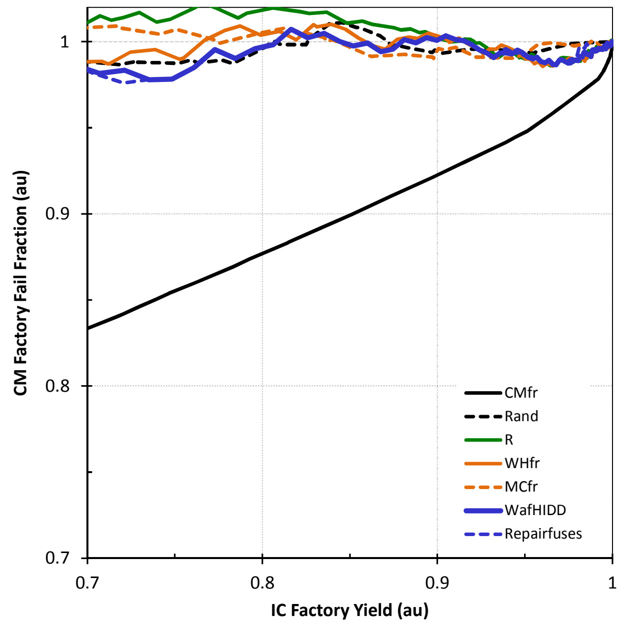

Figure 9 ROCs for WafHIDD-groups of dies ranked by attributes in the legend. All ROCs except CMfr are indistinguishable from Rand. The heavy line (WafHIDD) is the ROC for an on-tester die-by-die WafHIDD-screen in the IC factory.

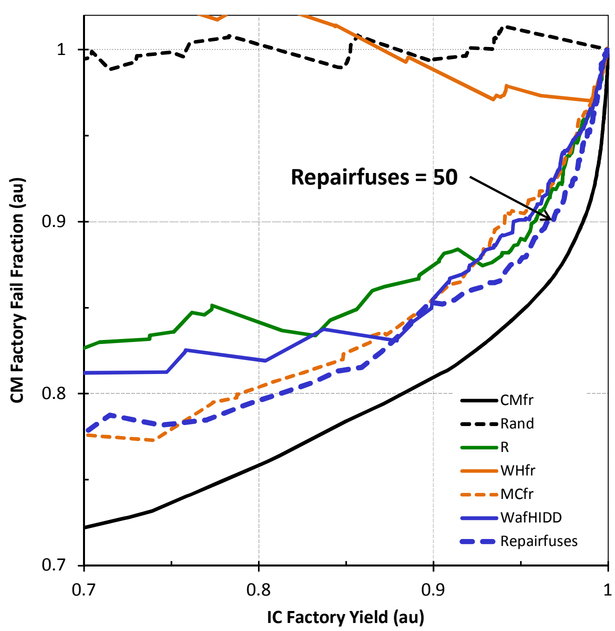

Figure 8 ROCs for Repairfuses-groups of dies ranked by attributes in the legend. All ROCs are bounded by CMfr and Rand (apart from sampling error). The heavy line (Repairfuses) is the ROC for an on-tester die-by-die Repairfuses limit screen in the IC factory.

4. Limitations

The analysis here was done on a complete set of naturally occurring data for an entire production cycle, so conclusions about potential test improvements were made long after they could have been applied. More useful would be an adaptive version of the method which can be applied during the CM manufacturing ramp.

The method depends on sufficiently large volumes, and sufficiently coarse groups, such that groups will be populated with CM passes and fails as well as ICs that have had a parametric measurement in the IC test factory.

The method of choosing grouping attributes described in this paper is somewhat ad hoc, starting from xy groups. An automatic method to find the best way to group dies would be very useful.

5. Discussion

A large part of the effort for the case study was constructing the database from naturally occurring data with no specific inter-factory data requirements. Rework in the CM factory made the data complicated to understand. A key data requirement is time-stamping of all data records so that causality can be invoked. Resolution of data integrity issues that arose in the course of the work provided much useful feedback to the CM.

The obvious attribute to exploit to improve screens in the IC factory is the radial dependence of failure rates revealed in the wafer maps of Figure 4. But comparison of Figures 8 and 9 with Figure 7 shows that more and less effective groupings than xy groupings are possible. The effectiveness of the method depends on 1) the choice of die "grouping attribute" by which to group dies, and 2) the choice of "candidate attribute", derived from die attributes, by which to rank the groups. Further development will seek automatic methods to make these choices and to reveal less obvious dependencies by finding "meta" attributes which are functions of other attributes and grouping and ranking by the meta-attributes. For example, R is a meta-attribute depending (by Pythagoras) on x and y. If on-tester die-by-die screens are required, the range of choices is narrowed to cases where the grouping attributes and candidate (ranking) attributes are the same.

6. Conclusions

Adaptive application [8] of the group-ranking method during the CM production ramp would be more useful than application of the method long after all the data were acquired as in the case study here. Adaptive application would require pre-production definition of data requirements and analysis programs in preparation for real-time data analysis and test program changes on a few thousand ICs over the first weeks of production. Pre-production coordination and communication of results at the level of detail required for the method may be difficult if the IC vendor and CM customer are different companies. If the CM factory is a different company from the IC supplier, the method may more easily be applied if a copy of CM-like system-level test [9] is used as an outgoing monitor in the IC factory.

The receiver operating characteristic (ROC) quantifies the cost/benefit tradeoff for the attributes identified by the group-ranking method. The ROC is the basis for negotiations between the IC factory and the CM factory. On-tester die-by-die ROCs for both xy groups and Repairfuses groups asymptote to a value of about 0.75 for CM fail fraction as the ROCs are extended beyond the left edges of Figures 7 and 8. So no degree of aggressive screening by blocking edge dies or by Repairfuses count can improve the CM fail fraction by better than 25%. This fact provided by the ROC is a useful boundary condition for negotiations. Improvements of this order may well be less expensively available elsewhere in the CM factory.

7. Acknowledgements

Support of the PSU authors by a grant from the Cisco Foundation is gratefully acknowledged.

8. References

[1] Zoë Conroy, Geoff Richmond, Xinli Gu, Bill Eklow, "A Practical Perspective on Reducing ASIC NTFs", *International Test Conference,* 2005, paper 14.2.

[2] Harry H. Chen, Roger Hsu, PaulYoung Yang, J. J. Shyr, "Predicting System-Level Test and In-Field Customer Failures Using Data-Mining", *International Test Conference,* 2012, paper 2.3.

[3] P. O'Neill, "Production Multivariate Outlier Detection Using Principal Components", *International Test Conference*, 2008, paper 1.2.

[4] M. G. Kendall and Jean Dickinson Gibbons, "Rank Correlation Methods", 1990, Oxford University Press.

[5] Gary Simon, "Multivariate Generalization of Kendall's Tau with Application to Data Reduction", *Journal of the American Statistical Association*, 1977, Vol. 72, No. 358, pp. 367-376.

[6] Gary Simon, "A nonparametric test of total independence based on Kendall's tau", *Biometrika* 1977, Vol. 64, No. 2, pp. 277-282.

[7] Robert Madge, Manu Rehani, Kevin Cota, W. Robert Daasch, "Statistical Post-Processing at Wafersort - an Alternative to Burn-in and a Manufacturable Solution to Test Limit Setting for Sub-micron Technologies," *Proceedings 20th IEEE VLSI Test Symposium* 2002, pp69-74.

[8] "Adaptive Test" *in International Technology Roadmap for Semiconductors (ITRS)*, Vol.: Test and Test Equipment, p21 (2011).

[9] Sounil Biswas, Bruce Cory, "An Industrial Study of System-Level Test", *IEEE Design and Test of Computers*, Vol. 29, No. 1, pp19-27 (2012).