NEIGHBOR SELECTION FOR VARIANCE REDUCTION IN I_{DDQ} and OTHER PARAMETRIC DATA

W. Robert Daasch†, Kevin Cota and James McNames
IC Design and Test Laboratory
Electrical and Computer Engineering, Portland State University
Portland, OR 97207

Robert Madge
LSI Logic Corporation
Gresham, OR 97030

Abstract

The subject of this paper is variance reduction and Nearest Neighbor Residual estimates for I_{DDQ} and other continuous-valued test measurements. The key, new concept introduced is data-driven neighborhood identification about a die to reduce the variance of good and faulty I_{DDQ} distributions. Using LSI Logic production data, neighborhood selection techniques are demonstrated. The main contribution of the paper is variance reduction by the systematic use of the die location and wafer- or lot-level patterns and improved identification of die outliers of continuous-valued test data such as I_{DDQ}.

1. Introduction

The key contribution of this paper is the introduction of new methods for identifying wafer patterns in continuous-valued test data. In the paper the methods will be demonstrated with I_{DDQ} data. The methods are however, applicable to any continuous-valued measurement. This contribution generalizes the idea of nearest neighbors and is robust to wafers with low yields, optimal in the sense of minimizing the residual and its variance and, can be applied to lots with as few as one measurement per die. The methods are data-driven and non-parametric, that is, they make no assumptions about wafer data pattern or the form of the original I_{DDQ} distribution.

The paper is organized into four main sections. The next section **Background** is a brief review of recent efforts in variance reduction techniques. This section also defines the terms used in the paper. The section **Neighbor Selection** introduces new techniques for identifying the data patterns in the intrinsic leakage and optimizing the final NNR formulation used in post-processing. It is worth repeating that the technique can be used to estimate any continuous-valued die measurement taken on the wafer. The fourth section **Results** demonstrates the techniques using production I_{DDQ} data for a $0.18\mu m$ process and a $0.25\mu m$ process. A short **Conclusion** ends the paper.

2. Background

In [1] it was shown that estimates of the I_{DDQ} residual can be made based on observed patterns at the wafer and lot level. Figures 1a and 1b are the familiar conceptual plots of the probability distribution functions (PDF) for non-faulty devices and faulty devices. A die defect is assumed to add a defect current to the current from intrinsic leakage and other normal processes. Die with the total current above some threshold are called outliers to the intrinsic leakage distribution. Traditionally a single threshold divides outliers from normal die and is a trade-off between the regions labeled returns and yield loss.

0-7803-7169-0/01 $10.00

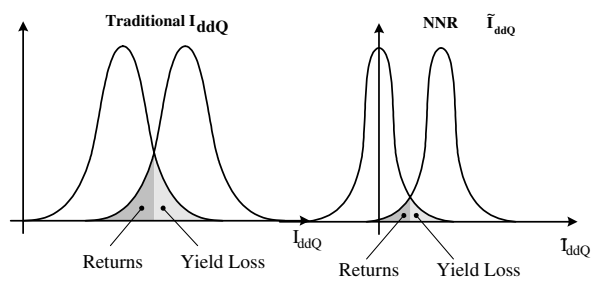

Figure 1: Idealized PDF (a) original I_{DDQ} (b) Residual ≡ $\tilde{I}_{DDQ}$

2.1. Statistical Post-Processing

Statistical Post-Processing (SPP) methods like Nearest Neighbor Residual (NNR) take a wider view of the test results. The goal of SPP techniques is to leverage the test measurement sample statistics to bin the die. A key step in Statistical Post-Processing is to reduce the variance of a distribution that will be used to identify outliers. In [1] it was shown that the variance of a well-defined residual is less than the variance of the original I_{DDQ} distribution. Mathematically, the I_{DDQ} residual, labeled $\tilde{I}_{DDQ}$, is defined in Equation (1)

$$\tilde{I}_{DDQ}(x, y, p) = I_{DDQ}(x, y, p) - \hat{I}_{DDQ}(x, y, p), \quad (1)$$

where $\hat{I}_{DDQ}(x, y, p)$ is an estimate of the original $I_{DDQ}(x, y, p)$. Each term is indexed by the x,y location and possibly other key parameters p. For NNR, the location of the die and the identification of its neighborhood is the core of the statistical post-processing. The central idea of NNR is the location x,y not parameters p. For readability the (x, y, p) is dropped from the notation for the rest of the paper. The residual in Equation (1) produces a distribution with reduced overlap of good and faulty die distributions. The benefit of this approach is that the variance of the residual is potentially much smaller while the difference between the $\tilde{I}_{DDQ}$ of a healthy device and that of a faulty device will remain the same.

To separate the desired information about defects from normal leakage other researchers have correlated I_{DDQ} to process parameters [2]. Intra-die comparisons such as ΔI_{DDQ} [3] and current ratios [4] compare one I_{DDQ} vector to another. Ideally, any technique should be without significant tooling, increased test time or what has proven a difficult challenge without a dependence on test vector selection or sequence.

2.2. Nearest Neighbor Residual

For a given test measurement, NNR post-processing estimates the test measurement of each die based on a robust selection of a wafer or wafer-lot pattern around the die of interest. The relationship between die location and yield has been observed many times [5,6]. Outliers (NNR fails) are detected by large residuals (differences) between the estimate and the measurement. In the case of I_{DDQ}, the Nearest Neighbor Residual replaces the original I_{DDQ} measurement in Equation (1) with

$$I_{DDQ} \rightarrow \bar{I}_{DDQ} = \frac{1}{N_v} \sum_{i=1}^{N_v} I_{DDQ.i}$$

$$\tilde{I}_{DDQ} = \bar{I}_{DDQ} - \hat{I}_{DDQ}, \quad (2)$$

where $\bar{I}_{DDQ}$ is the average of N_v test vectors used in the I_{DDQ} test. The $\hat{I}_{DDQ}$ is the estimate of a die's average value based on its neighbors' $\bar{I}_{DDQ}$. The die I_{DDQ} average will be used throughout the paper to demonstrate the technique. In practice this has been very successful parameter to bin the die. However, NNR is not limited to this choice. NNR could equally be applied to data from a single I_{DDQ} vector or any continuous-parameter measured at each die.

Identifying a neighborhood is one of the key contributions of this paper. In [1] the NNR $\hat{I}_{DDQ}(x, y)$ was defined to be the median of at least eight die closest to the (x, y) location. This paper extends the idea of neighborhood selection to include data-driven techniques. The new, data-driven, selection highlights the relationship between the neighborhood, the robustness of the estimate, manufacturing variability and, the final variance reduction that can be obtained.

3. Neighbor Selection, Location Averaging

A fixed neighborhood selection such as the eight die closest to the position x, y works well in practice when the data pattern is smoothly varying. Smooth contours have been observed many times. However, stepper patterns have also been observed and they are not smooth but systematic. They impose a checkerboard effect across the wafer. In particular, I_{DDQ} data show a mixture of the two patterns, smooth combined with systematic or vice versa [1]. As shown in this paper, a mixture is further complicated by the fact that it varies from lot to lot. The lot to lot variation is difficult to specify in advance with a neighborhood pattern and any fixed pattern is difficult to use when yield is low.

A key attribute of NNR is its use of data-driven techniques to define the neighborhood. Within some general limitations the NNR neighborhood used for a lot is determined from the data for the lot. By defining the neighborhood from the data itself NNR can easily respond to a changing mixture of systematic and smooth patterns from one lot to the next. As NNR is implemented off-tester the required computation of identifying the neighborhood adds very little if any, to test time.

The neighborhood selection scheme called Location Averaging is demonstrated with a simple example shown in Figure 2. The demonstration wafer consists of a single row of twenty die and a lot is one wafer. The upper wafer shows a smooth transition from left to right. The lower wafer shows a systematic pattern representative of a stepper effect.

A location average neighborhood is based on a candidate template denoted by $T = \{\}$. In the example, the candidate template is the set of die within ±2 of a location,

$$T = \{X-2, X-1, X+1, X+2\}.$$

The X denotes the location of the die to be substituted into the template and is called the target location or target die. If location 6 was the current target location, the candidate die locations are positions 4, 5, 7 and 8.

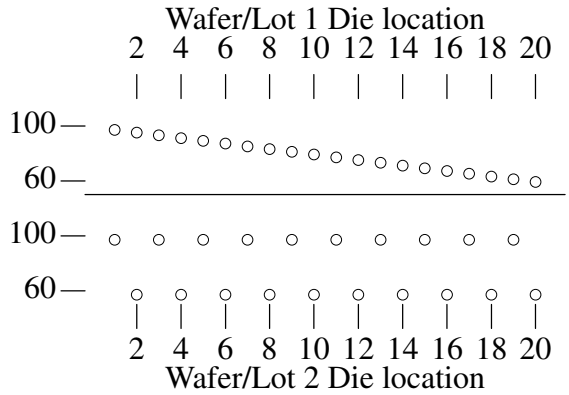

Figure 2: Simple example of Neighborhood Selection

Using the algorithm in Figure 3, Location Averaging loops over the template entries one by one. In the example, the first iteration template entry $TE = X-2$ would be used, $X-1$ next and so on. The die estimates ($\hat{I}$) are assigned using the current template entry. For $TE = X-2$ and location $D = 6$, the die in position 4 is used to estimate the value of location 6. For all die in the lot ($D = 1..20$) a nearest neighbor residual ($\tilde{I}_{TE,D}$) is calculated using the estimates from the current template entry. In the example, NNR returns $\tilde{I}_{X-2,D}$ as a set of 20 values, one for each die location. In line 4, the median of this residual set is saved for template entry $TE = X-2$. The algorithm loop then repeats for the next template entry, i.e. $TE = X-1$.

Line 1: Loop over Template Entry (TE)
Line 2: Assign $\hat{I}_D = \bar{I}_{TE}$
Line 3: Run NNR for $\hat{I}_D$
Line 4: Assign $R_{TE} = \left| Median(\tilde{I}_{TE,D}) \right|$
Line 7: End Template Loop
Line 8: Sort (R_{TE})

Figure 3: Location Averaging Algorithm

The algorithm processes each template entry and assigns each entry a single index value, the median of the die residuals for the lot.

Sorting these values ranks the template entries by their estimate of the expected value of the parameter of interest, in this case $\bar{I}$. The template entry with the smallest median residual for the lot identifies the location for the best single die estimate of a target die measurement. In the example, the location $X-1$ has the smallest residual. The template ranking for the first lot is

$$T_1 = X-1, X+1, X-2, X+2$$

The second lot template is dominated by the stepper and the template ranking is

$$T_2 = X+2, X-2, X+1, X-1.$$

The example highlights the robust nature of location averaging and its ability to use data-driven techniques to address the difficulty of predicting neighborhoods in advance. As will be shown in the next section, location averaging can readily identify a template ranking for data with a mixture of smooth and systematic effects.

4. Results

Several NNR results are reported in this paper in an effort to understand the ability of NNR to reduce variance and identify outliers. The methods for neighborhood selection and NNR dependence on yield were tested using LSI Logic production data. The data reported has been renormalized to avoid releasing any proprietary information. The data is for a $0.18\mu m$ process and includes approximately 75 wafers from three different lots. A of total 25,000 die are used in the experiments.

Results for neighborhood selection, die yield and single vector I_{DDQ} are presented in this section. These studies demonstrate the stability of the NNR $\tilde{I}$ residual and NNR's utility at identifying outliers.

4.1. Neighbor Selection

The first study examined data-driven neighborhood selection schemes. For neighborhood selection two lots of twenty-five wafers are analyzed. Each wafer has about 300 die per wafer. A lot has approximately 7,500 die. The candidate template is the 120 die set radiating out from the target. Each 120 die template is a die borough from which a smaller neighborhoods can be selected. The Location Averaging algorithm described earlier ranked the template entries. All I_{DDQ} data collected for the two lots was used. Twenty I_{DDQ} vectors were measured on the die that passed functional and gross I_{DD} tests. The NNR neighborhood used for the estimate was the twelve, lowest template entries.

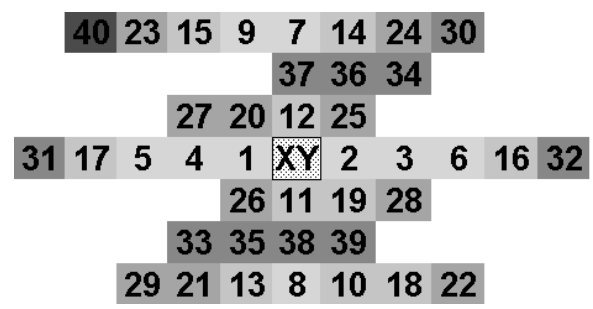

Figure 3: Forty best die based on Location Averaging Neighborhood Selection

Two views of this result are shown in the Figures 3 and 4. Figure 3 shows the rank and position relative to a target die of the top forty candidate die based on the $|\tilde{I}|_{median}$. The lightest areas are near the target location but not immediately adjacent to it. The die in the same row provide the best six estimates. The other six die in the eight die annulus around the target die are ranked well below other die in the borough. This is rather different from the annulus about a die that was used in [1]. The second view shown in Figure 4a is a top-down and Figure 4b is a 3D perspective of a die template. The rows and stepper pattern are clearly evident as well as the rapid drop off of estimates greater than 40.

The pattern of best candidate die varies from lot to lot. The computation to identify the pattern is simple to complete off-tester without increased test time or assembly time. To demonstrate the variation in neighborhood patterns the experiment was repeated using a

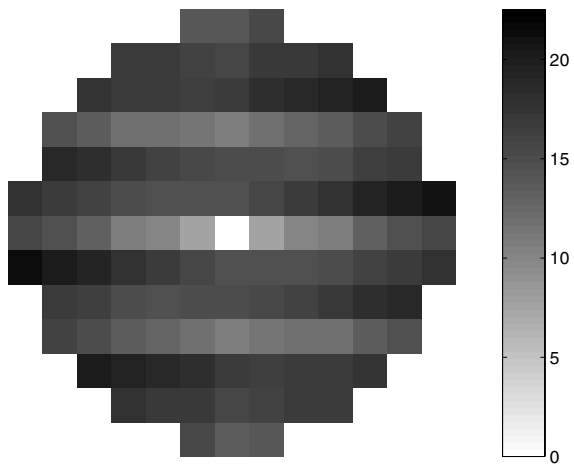

Figure 4a: Location Averaging Template (First lot). Target Location is white, light regions are best neighbors and dark regions are poor neighbors

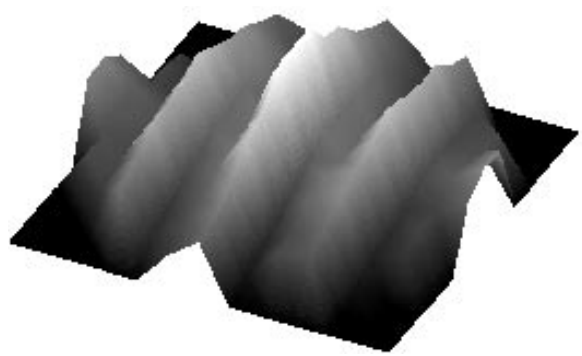

Figure 4b: Location Averaging Template (First lot). Target Location is white, light regions are best neighbors and dark regions are poor neighbors

different lot. In Figure 5, the strong row and 3X3 stepper pattern in Figure 4 has been replaced by a conical shape (i.e. concentric contour rings) very similar to the eight nearest neighbors described in [1].

Location averaging is an efficient and robust technique to meet the key challenge of neighborhood selection for NNR and other statistical post-processing techniques. To track manufacturing variations the method requires lot by lot updates of the template ranking. This calculation is easy to implement off-tester. In a manufacturing setting the main requirement is that the template ranking and subsequent die binning be completed prior to assembly.

Figure 5: Location Averaging Template (Second lot) Target Location is white, light regions are best neighbors and dark regions are poor neighbors

4.2. Yield Dependence

The second study looks into the relationship between yield and the estimate $\hat{I}$ and the residual $\tilde{I}$. Using high yielding wafers, the NNR method is implemented on single wafer I_{DDQ} data and the point estimate for the residual distribution is calculated. The neighborhood is defined as the twenty best die from the location averaging neighborhood selection algorithm.

To simulate reduced yield on the wafer, ten die are randomly eliminated from the I_{DDQ} data-set and the process is repeated. As a simulation the possible increased variability of the I_{DDQ} data itself is not modeled. As data is found in the database the connection between yield, increased data variance and NNR will be studied. As the yield for the wafer approaches zero, a series of point estimates $|\tilde{I}|_{median}$ are obtained. Large changes in $|\tilde{I}|_{median}$ indicate instability in the estimate and the decreasing ability of NNR to consistently detect the outliers.

Figure 6 plots the point estimate versus yield for four typical wafers. The estimate is relatively constant even at substantially reduced yields. At very low yields, few of the twenty

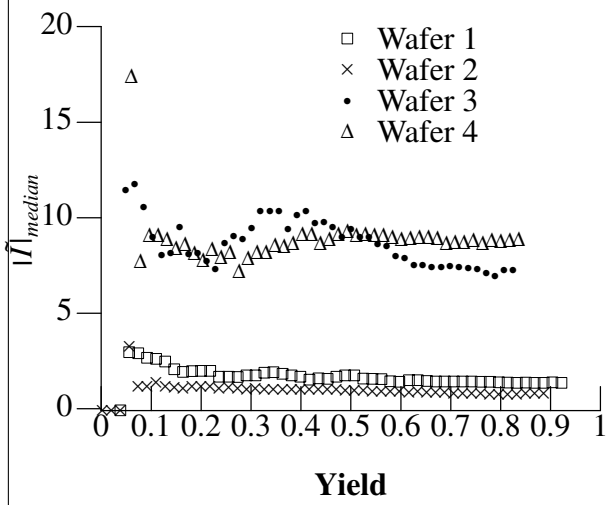

Figure 6: NNR stability as a function of wafer yield

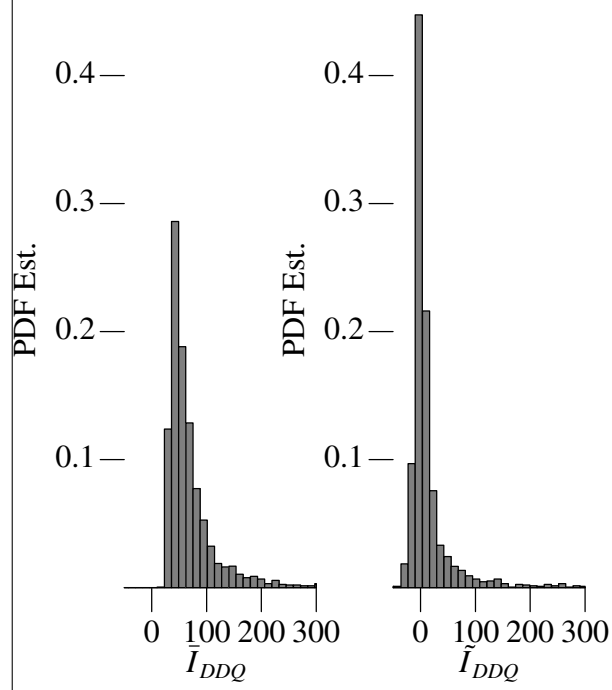

Figure 7: Improved Variance from nearest neighbor estimation

neighbors are present and the stability of the estimate suffers. At the very lowest simulated yields, estimates vary by 100% or more and many die would be erroneously binned. The preliminary conclusion from this study is that NNR data-driven neighborhood selection is robust to diminished wafer yield.

4.3. Variance Reduction

An example of variance improvement is shown in Figure 7. The left-hand histogram is a plot of the mean $\bar{I}_{DDQ}$ for a lot containing I_{DDQ} measurements on 5,300 die. Twenty I_{DDQ} vectors were tested and averaged. Location averaging was completed on the lot and ranked 120 neighbors. The twelve best neighbors were used to compute the nearest neighbor estimate. The right-hand histogram is a plot of the residual $\tilde{I}_{DDQ}$ following the NNR statistical post-processing. As can be readily seen the variance is reduced by about a factor of two. Histogram bin sizes and scales are identical for the two graphs.

4.4. Outlier Detection

This study highlights the improved outlier detection that the reduced variance of the residual can provide. The data used in this study was a LSI test chip. The die area was much larger than the other experiments with about 100 die per wafer. There are approximately 2,500 die represented in this study with I_{DDQ} measurements for die that passed functional and gross I_{DD} tests. There are twenty I_{DDQ} measurements for each die. Location averaging was completed in the usual way to define a template of 40 die. The twelve die lowest ranked die with I_{DDQ} data were used to compute estimates and residuals.

Two key graphs summarize this portion of the study. Plotted in Figure 8 is the die estimate from its nearest neighbors versus the average of the I_{DDQ} for the die. Several features are evident. First, note that the majority of the estimates lie in a large mass in the lower left. The widening at approximately 100 units is evidence that the estimate variance increases slightly with the value of $\bar{I}_{DDQ}$. Secondly note the clear outliers to the right. These large $\bar{I}_{DDQ}$ are not at all

predicted by the nearest neighbors in the ranked, forty die template. LSI's ΔI_{DDQ} post-processing binned approximately 95% of the outliers in the data set. All of the ΔI_{DDQ} escapes, primarily in the right-hand part of the plot above 150, are detected as outliers by NNR.

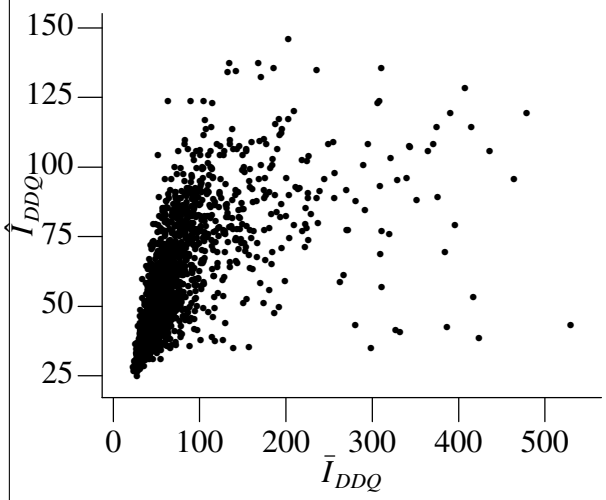

Figure 8: Scatter plot of $\hat{I}_{DDQ}$ vs $\bar{I}_{DDQ}$

In Figure 9 the die residual is plotted versus the mean of die. As a reminder the residual is of the difference of an estimate and observed value

$$\tilde{I}_{DDQ} = \bar{I}_{DDQ} - \hat{I}_{DDQ}.$$

In effect this plot is the difference of the x-axis and y-axis in Figure 8. Note again the mass of residuals is evident in the lower left of the plot. The smaller residual values represent die where the estimate predicted the mean value. Second, there are negative values. Negative residuals were also shown in the NNR histograms in Figure 7. The negative values are an over prediction of the estimate compared to the mean. Die with the negative residuals were examined with respect to other test measurements (such as device speed) and have not shown any evidence of being outliers in the sense of test escapes. Third, note the much improved linear correlation of residual versus mean compared to the estimate versus mean. The die represented by the large residuals, away from the mass, are the ones that would be binned as outliers.

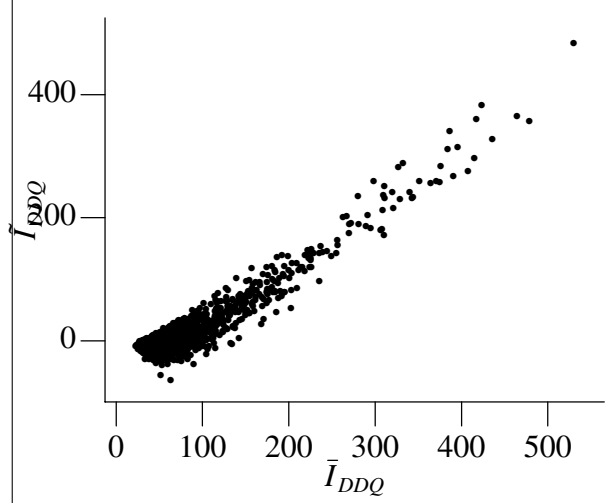

Figure 9: Scatter plot of $\tilde{I}_{DDQ}$ vs $\bar{I}_{DDQ}$

As a final example of outlier detection are the die represented by the histogram in Figure 10 are field fails, i.e. test escapes. The die passed all tests at production and final test. A single I_{DDQ} measurement was available for the escape die. From the histogram, approximately 40% of the die are above a $I_{DDQ} = 15\mu A$ threshold. The original test limit was $I_{DDQ} = 100\mu A$.

Two possible alternatives to catch the I_{DDQ} escapes are to reduce the I_{DDQ} test limit to a much lower level (e.g. $15\mu A$) or add test vectors to increase functional coverage. The first alternative is an example of the classic trade-off of yield for improved early failure rate and reduced infant mortality. For second alternative of improved test coverage, the longer test time has to be weighed against reduced EFR. A third alternative is look for the outliers of the existing test suite using NNR.

The NNR study of this product used a single lot and a single I_{DDQ} vector from the wafer

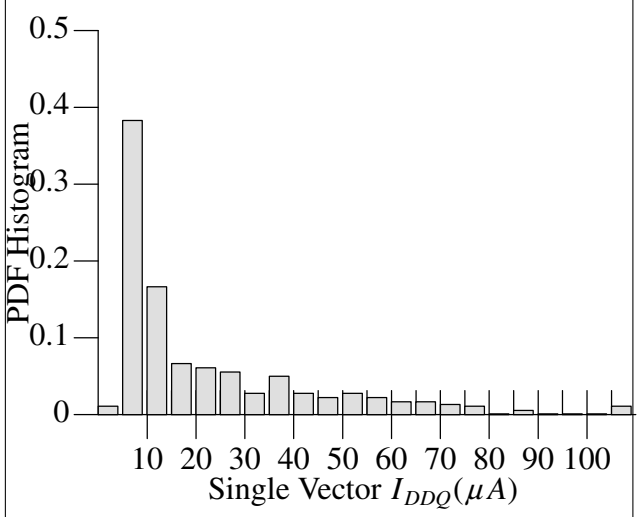

Figure 10: Histogram of single vector I_{DDQ} for field fails

test data, that is in Equation (2) the average is replaced with the one I_{DDQ} measurement for each die.

$$\bar{I}_{DDQ} \rightarrow I_{DDQ.single}.$$

The neighborhood was selected in the usual way using a lot of 25 wafers. Using a single test vector models the available field fail data. In this case, the nearest neighbor residual is the difference of the estimate from the template and the original measurement.

The results are summarized in the scatter plot in Figure 11. The vertical axis is the single vector I_{DDQ} and the horizontal is a die speed measurement denoted KP. The legend summarizes three test results. The open circle is an NNR fail, functional pass and I_{DDQ} pass. The solid dot is an NNR fail and functional fail. Recall that the I_{DDQ} limit is $100\mu A$. Finally a triangle is a functional or I_{DDQ} hard fail and the cross is all pass. From the histogram, the die of interest are $10\mu A < I_{DDQ} < 100\mu A$ because they are the source of 40% of the escapes. In this range 33 of the 39 die are NNR fails. Eleven die (open circles) in this region (28%) were identified only by NNR.

Two key ideas are demonstrated by this experiment. First, NNR is flexible and can be used with any parametric data such as a single I_{DDQ} vector. Second, NNR demonstrates a data-driven selectivity for outliers that could provide a dramatic improvement of Early Failure Rates.

5. Conclusion

This paper expands the understanding of the technique called Nearest Neighbor Residual. This paper presented new, data-driven alternatives for neighborhood selection. By expanding the selection process to include the best die estimates NNR was shown to be robust to yield fluctuations.

The location averaging has been applied to each lot of data analyzed to capture the key patterns in that specific lot. Location averaging was shown to identify pattern variation in a single product on a lot to lot basis.

After statistical post-processing variance reduction was observed in all cases and shown to improve the ability of I_{DDQ} testing to identify outliers.

Finally, NNR was shown to work with a single I_{DDQ} measurement and predict outliers that previously were test escapes. The final example shows how NNR could be applied to detect outliers of other continuous-valued test results.

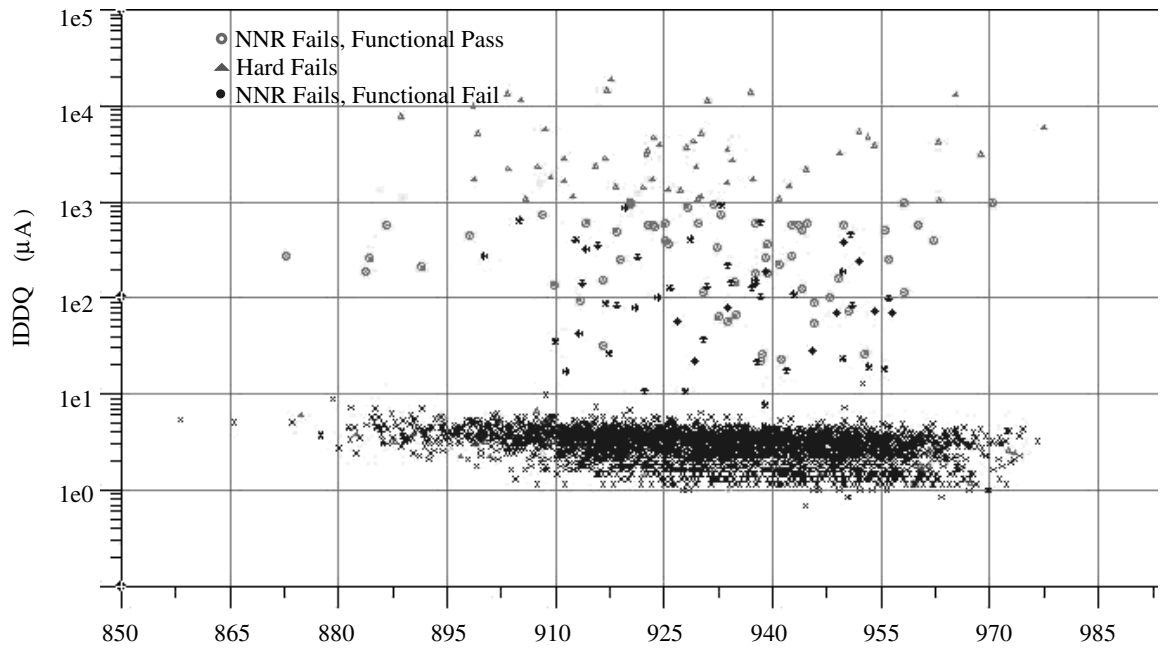

Figure 11: Scatter plot single vector I_{DDQ} vs die speed (KP)

References

1. W.R. Daasch, J. McNames, D. Bockelman, K. Cota, and B. Madge, "Variance Reduction Using Wafer Patterns in I_{DDQ} Data," *Proceedings of International Test Conference,* pp. 189-198, October 2000.
2. A, Keshavarzi, K. Roy, and C.F. Hawkins, "Intrinsic Leakage in Deep Submicron CMOS IC-measurement-based Test Solutions," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* 8, 6, pp. 717 -723, December 2000.
3. C. Thibeault, "A Histogram Based Procedure for Current Testing of Active Defects," *Proceedings International Test Conference,* pp. 719-723, October, 1999.
4. P. Maxwell, P. O'Neill, R. Aitken, R. Dudley, N. Jaarsma, M. Quach, and D. Wiseman, "Current Ratios: A Self-Scaling Technique for Production IDDQ Testing," *Proceedings International Test Conference,* pp. 738-746, October 1999.
5. W.C. Riordan, R. Miller, J.M. Sherman, and J. Hicks, "Microprocessor Reliability Performance as a Function of Die Location," *IEEE 37th International Reliability Physics Symposium,* pp. 1-11, San Diego, CA, 1999.
6. R.A. Richmond, "Successful Implementation of Structured Testing," *Proceedings of International Test Conference,* pp. 344-348, October 2000.