

QUALITY IMPROVEMENT AND COST REDUCTION USING STATISTICAL OUTLIER METHODS

Amit Nahar, Kenneth M. Butler, John M. Carulli Jr., Charles Weinberger
Texas Instruments, Dallas, TX

a-nahar2@ti.com, kenb@ti.com, jcarulli@ti.com, c-weinberger@ti.com

Abstract—Quality improvement and cost reduction in the overall IC manufacturing and test processes are being continuously sought. Outlier screening methods can address both of these needs. As technology scales, it has become increasingly difficult to screen outliers without excessive Type I or II errors. Hundreds of parameters are collected at wafer probe, but there lacks a systematic way of selecting outlier screens. In this paper we describe a statistical approach to both identify outliers and select beneficial screening parameters more effectively. Results on a 90nm design to reduce the burn-in fails are described.

I. INTRODUCTION

Quality requirements have become increasingly stringent for manufacturers in electronics. These requirements are specified in terms of initial or “time zero” quality (also known as “test escapes”) as well as product reliability, both typically using the metric of defective parts per million or DPPM. Targets for both time zero and reliability DPPMs are continually being driven downward. Table 1 shows reliability requirements for different product segments. This table is an adapted version as published in ITRS [1].

Table 1: Reliability requirements for different product segments[1]

Product Segment	Market Driver Advances	Market Driver Reliability	Market Driver Time-in-market	Technology Impact
PDA & Wireless	Function 2x every 2 yrs	50-2000 DPPM	Few Years	Low Power
Computer	Speed 2x every 2 yrs	50-2000 DPPM	Few Years	Performance
Internet	Bandwidth 4x every 3-4 yrs	99.999% up time	Long	Performance Power Reliability
Automotive	Function Ruggedness	Zero Defect	Long	Reliability

It can be seen that the DPPM requirements vary significantly across different product segments, with automotive products requiring very high reliability. There are different methods employed in the industry to ensure that the reliability requirements are met. For IC manufacturers, one traditional method of meeting reliability requirements is to employ burn-in. The burn-in step accelerates the failure time of unintended manufacturing defects in a circuit. This acceleration is often driven through elevated voltage and temperature conditions. Burn-in is typically comprised of a reduced pattern set. It can be either a proper subset of the manufacturing test or a set of patterns derived specifically

for burn-in. The application of burn-in also typically necessitates at least two post-packaging test insertions. The first test insertion is before burn-in to screen defects induced by the assembly process. The second test insertion is after the burn-in operation to verify known good units and to remove the units failing burn-in due to unintended manufacturing defects. The burn-in operation and post-burn-in retest can be very expensive. Burn-in oven costs often measured in six figures and high end automated test equipment in seven figures. IC manufacturers thus have a strong motivation to either reduce or eliminate burn-in whenever possible.

The IC industry is always striving to reduce the manufacturing and test cost but also maintaining the low DPPM numbers at the same time. Hence alternate methods to remove expensive tests are always being pursued. For 20 years or more, one method used to avoid burn-in has been IDDQ testing [2]. In this approach, the device is placed into a quiescent state, and then the power supply current is measured. The presence of a defect has been observed in many cases to elevate the quiescent current level by one or more orders of magnitude [3]. Devices that failed only IDDQ tests were considered reliability risks and often scrapped. As technology scales, the background currents are much higher [4]. Defects are therefore much harder to observe with respect to good units when measuring power supply currents.

Today, the class of devices to which simple single threshold IDDQ testing can be applied is considerably smaller than when it was first introduced. Instead, the techniques that must be applied to identify reliability risk die are increasingly complex, such as delta IDDQ or current signatures [5]. Besides IDDQ tests, other tests have been posited to be indicative, when failed, of possible reliability issues. As technology scaling continues, evidence suggests that these traditional methods of burn-in minimization are becoming less effective. The goal of the work reported here is to develop statistical methods to identify outlier die that correlate to reliability risks. The statistical methods can be used for burn-in reduction or elimination. They can also be used to improve quality by scrapping outlier die. These statistical methods can potentially be used to reduce the test cost and maintain the low DPPM numbers.

This paper is organized as follows. In section II, we describe the statistical method used for identifying outliers. In section III, we show how to set limits to distinguish outliers from the healthy population. Section IV describes the statistical procedure of selecting outlier screening parameters. Section V summarizes the results on a device.

II. STATISTICAL METHODS FOR IDENTIFYING OUTLIERS

Over the years, several papers have been published to identify outliers which are at a risk to either fail at a future test point or in an end application. These outlier die are considered to be a risk for reliability failure. Interquartile (IQR) range methods have traditionally been a simple way of detecting outliers [6]. The Automotive Electronics Council has guidelines for using Part Average Testing (PAT) to identify outliers on semiconductor devices [7]. Techniques based on median statistics such as median of absolute deviations have also been used for outlier identification[8].

With newer semiconductor technologies it has been more and more difficult to identify clear and effective outlier populations. Thus advanced statistical techniques are needed to identify outliers more efficiently. A number of papers have been published on neighborhood-based techniques such as Statistical Post ProcessingTM or SPPTM. The concept behind neighborhood-based techniques is that there are two distributions in the parametric data; one for the healthy die and the other for the “bad” die which will be the outliers and will deviate from the healthy die distribution. Traditionally with IDDQ testing there is significant overlap between the healthy die and bad die distributions. Ordinarily, to distinguish outliers efficiently, this overlap should be minimal. Neighborhood-based techniques try to reduce this overlap between the distributions by using the concept of residuals [10]-[14]. The term “residual” refers to the difference between a measured parameter such as IDDQ or minVDD and an estimate of the expected value for that measurement. The use of residual transformation reduces the variance of the two distributions. There has been other work in the outlier identification area[15], [16].

To date, all of the work on SPP has been published using 180nm and 130nm production data. As published, SPP uses a variety of parametric measurements like IDDQ, subthreshold IDDQ (STIDDQ) and minimum operating voltage (minVDD) to identify outlier die. The authors of this paper have also published on a neighborhood-based technique to identify outliers which has been used for burn-in reduction on 90nm and 65nm process driver devices[12].

We have developed customized statistical methods for outlier detection using various transformations such as Nearest Neighbor Residuals (NNR), Location Averaging (LA), and Principle Component Analysis (PCA). In this paper, the Location Averaging method will be explained. We first explain NNR, which is the foundation for LA. As the name implies, NNR is a neighborhood-based technique. For each parameter under study, the healthy die estimate is constructed from the median of the same measurement for the eight surrounding die on the wafer. The median is used instead of the mean because it is less sensitive to outlier responses. If one or more of the measurements is missing in the neighborhood, it is allowed to expand slightly so that sufficient measurements can be used to construct the estimate. When the estimate is constructed, a residual is formed by subtracting it from the measurement from the die of interest. The use of residuals reduces data variance and also can rotate the distribution as it centers it on zero. This

permits the use of tighter limits, thus minimizing overkill. LA differs from NNR in that the definition of the “neighborhood” is intentionally varied. A method is employed in which a window of die, typically a 7x7 or 9x9 array, is scanned across the wafer, and for each measurement, the X and Y (row and column, respectively) locations around the die of interest are ranked based on how close are the die electrically to the die of interest.

This difference is illustrated graphically in Figure 1, which shows an example of a calculated LA template. The 7x7 matrix is swept across the wafer surface and each location within the matrix is ranked according to its predictive capability. The highest ranking die (shown in white) forms the “neighborhood”, and again, this method is employed parameter-by-parameter and wafer-by-wafer. The rest of the analysis is much the same as for the case of NNR. At the end, we will obtain an estimate and residual value for every good die on the parameter.

41	33	29	17	26	32	43
48	27	5	15	13	24	46
39	19	3	7	11	22	38
35	9	1	2	10	36	
37	21	12	8	4	20	40
45	23	14	16	6	28	47
44	31	25	18	30	34	42

Fig. 1: Example Location Averaging template showing die ranked which are closest to the die of interest (blue die)

III. LIMIT SETTING

A very important aspect of both NNR and LA is limit setting. That is deciding where to draw the line between the intrinsic distribution and the outliers. The limits should be set in such a way that the type I (false positive) and II (false negative) errors should be minimized. The false positives are the outliers and the false negatives are the non-outliers. We will describe a limit setting method in this section. We approach the limit setting challenge in this method by defining the intrinsic (healthy) die population first. Once the intrinsic population is defined, all die outside that population are outliers. This concept of limit setting is exactly opposite to some of the earlier work published under SPP, which used the bad die distribution to define the limits.

Once the Location Averaging analysis is complete on a parameter, an estimate and residual value is obtained for every good die on the wafer. Figure 2 shows the residual and the estimate data for one parameter of a wafer. A robust linear regression of the residual and estimate data is performed. Outliers in the dataset have an impact on the slope of the regression line. To negate their impact, the robust regression model is used to define the regression line. An iteratively re-weighted least squares model is used in the robust regression model. After defining the regression line, limits are set to identify outliers from the intrinsic population. Prediction limits are defined about the regression line to identify outliers. Prediction limits are

calculated based on the confidence level provided by the user. The confidence levels can vary anywhere between 95% and 99.99%. The prediction limits will become wider as the confidence level is increased. This concept is similar to the fact that 6Sigma limits are wider than 3Sigma limits. These prediction limits are calculated for every parameter and wafer. The limits will automatically adjust as the distribution of the estimate and residual data changes for every wafer. All die outside the prediction limits are outliers. The blue lines in Figure 2 are the prediction limits and red die are outliers.

In the Location Averaging method, we use a neighborhood model to obtain residuals, and outliers are identified in the residual space. The idea behind using the neighborhood model is to identify die which are different from their neighbors. If there a large cluster of defects on a wafer (e.g.: a large region on a wafer with high IDDQ), then there is a possibility that a few die in that region will have large estimate values, but small residual values because the die of interest will be similar to its neighbors. Since the limits are defined only in the residual space, in such cases some defective die will not be identified as outliers. This is a limitation of the current Location Averaging method.

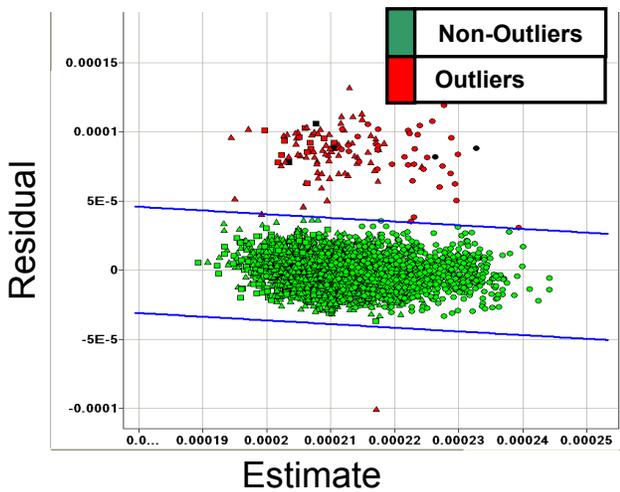


Fig. 2: Robust regression of Residual Vs Estimate. Die outside the prediction limits are outliers.

The question is why the residual data transformation is required? Isn't the raw test data enough to identify outliers? We will illustrate with an example how the residual data is more effective in reducing the intrinsic data variance and also in amplifying the outlier signature. Figure 3 shows the raw data distribution for a parameter and one wafer. The raw data for this parameter has been normalized. Interquartile range (IQR) limits are defined for this distribution to identify outliers. A burn-in fail for this device is in the middle of this raw data distribution and inside the IQR limits. Figure 3 also shows the residual data distribution obtained from the Location Averaging method for the same parameter. The intrinsic distribution for the residual data is much tighter than the raw data distribution. Table 2 compares the variance and the standard deviation of the intrinsic distribution for the raw and residual data. The

variance and standard deviation of the residual data is smaller than the corresponding metrics for the raw data. The IQR limits for residual data are also tighter compared to the IQR limits for raw data. The burn-in fail becomes an outlier in the residual data and is outside the IQR limits. This shows that the outlier signature is amplified with the residual data.

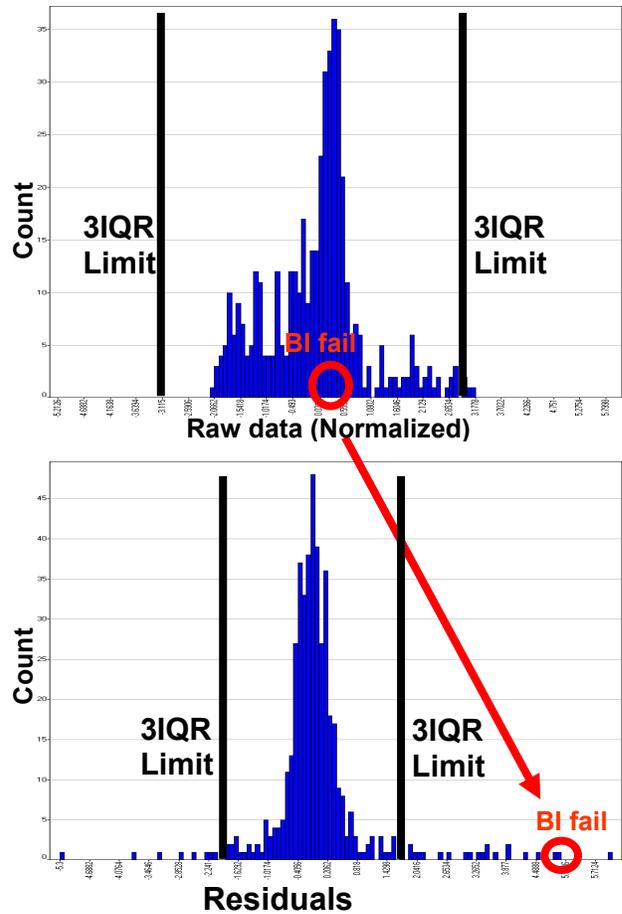


Fig. 3: Raw data and the residual data for the same parameter. Variance of the healthy population reduced in the residual space and outliers are amplified.

Table 2: Comparison of variance and standard deviation for raw data and residuals

	Raw data	Residuals
Standard deviation	0.99	0.56
Variance	0.99	0.31

IV. SELECTION OF PARAMETERS FOR OUTLIER SCREENING

In a production environment, hundreds of tests (parameters) are executed and recorded at wafer probe and/or package test. It is often not practical or justified to perform outlier screening on all parameters collected in production. Outlier screening is typically implemented only on a subset of parameters so that the overkill can be minimized. The challenge is how to select the parameters for outlier production implementation. We have developed a

flow to identify outlier screening parameters by correlating outliers at wafer probe with burn-in fails. This flow allows the identification of tests which can catch these fails and also are statistically significant. The idea is that outlier screening will be implemented in production on these identified parameters and potential future burn-in will be screened at wafer probe. There are three steps in the flow.

STEP 1: Identifying dataset for analysis

This is a simple step. Typically a sample of wafers containing at least one fail of interest are selected for the analysis.

STEP2: Outlier Identification

Outlier identification has been discussed in a number of papers over the years. Once the wafers are selected for analysis, outliers are identified for every parameter and wafer using any statistical outlier method. In this paper, the Location Averaging method is used for outlier identification. At the end of this step, a list of outliers is obtained for every parameter and wafer.

STEP3: Statistical Selection of parameters

Once the list of outliers is obtained for every parameter and wafer, the next step is to correlate the fails to the outliers at wafer probe. If at least one fail is an outlier for a parameter, a 2x2 contingency table is created for the parameter as shown in Figure 4. Chi-square and the corresponding p-value metrics are computed for every test using the 2x2 contingency table.

All parameters with a p-value of < 0.05 are considered to be statistically significant [17]. An iterative approach is used for the selection of statistically significant parameters. The iterative approach is inspired by static test compaction methods used in automatic test pattern generation (ATPG) and proceeds as follows:

1. Compute p-value and Chi-square for every parameter catching the fail
2. From the remaining parameters, select the parameter with the lowest p-value
3. Remove all the outliers from the dataset for the selected parameter
4. Recalculate Chi-square and p-value for every parameter catching the fail in the reduced dataset.
5. Repeat steps from 2 to 5 till any parameter has a p-value of < 0.05 or till all the fails are caught

Contingency Table:

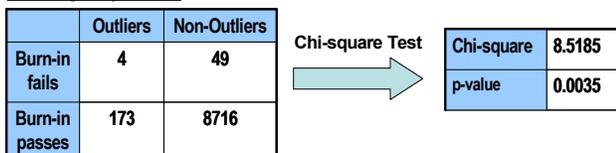


Fig. 4: 2x2 contingency table used to calculate the Chi-square and p-value statistic

At the end of this iterative approach, a list of statistically significant parameters is identified which can catch

maximum fails and also have least overkill. Typically we begin the analysis with anywhere between 50 and 1500 parameters. After following the iterative approach as described above, we end up with a list of anywhere between 1 and 20 statistically significant parameters catching the fails. The analysis focuses on parameters related to chip functionality and performance. These may include tests such as IDDQ, minVDD, Fmax, pin leakages and tests using on-chip parametric structures. These are the types of tests used for the selection of outlier screens. Even though all the identified parameters are statistically significant outliers, it is important to perform a sanity check and confirm that there is a relation between the identified parameter(s) and the failure modes observed on the fails of interest.

We have used this selection approach on a number of products. In practice the sanity check has exposed a remarkably small number of retained but unrealistic parameters. When they are detected the unrealistic parameters are removed from the production screen deployment.

V. RESULTS

We have performed the Location Averaging outlier analysis and used the methodology of selecting statistically significant parameters for outlier screening on a number of products. In this paper, we will discuss a case study from a 90nm wireless SOC. Burn-in testing was performed on this device and a number of post burn-in fails were obtained over a large sample of production material. The goal of this analysis was to identify outlier screens which could predict the burn-in fails at wafer probe to avoid future costs associated with burn-in.

Wafers containing the burn-in fails were identified for the outlier analysis. The Location Averaging and robust linear regression model was used to identify outliers on all the parameters collected in the test program. A 99% confidence level was used to set the prediction limits about the regression line. After identifying outliers for every parameter, Chi-square and a p-value statistics were calculated for every parameter. The approach of selecting statistically significant parameters as described earlier was followed to select the outlier screens.

We started the analysis started with approximately 400 parameters. After following the iterative approach of selecting statistically significant parameters, 15 parameters were retained. Table 3 shows the cumulative percentage of burn-in fails and outliers caught by the parameters. Using all the parameters in the Table 3, 55.3% burn-in fails were detected with 2.5% outliers. A sanity check was performed on these results to make sure that the identified parameters had a relation to the failure mode observed on the burn-in fails. In all cases it was confirmed that there was a direct relation. By screening 55.3% of burn-in fails at probe, we met the customer DPPM requirements for this device. These outlier screens were deployed in production and burn-in testing was eliminated on the non-outlier population for this device. In general, we have seen IDDQ, minVDD and Fmax tests useful to screen burn-in fails and customer returns.

Table 3: List of statistically significant outlier parameters useful to screen the burn-in fails. 55.3% of burn-in fails could be caught with 2.5% outliers.

Parameter	Cumulative Burn-in fails caught (%)	Cumulative Outliers (%)
Parameter 1	7.9	0.1
Parameter 2	10.5	0.2
Parameter 3	13.2	0.2
Parameter 4	15.8	0.3
Parameter 5	21.1	0.4
Parameter 6	23.7	0.5
Parameter 7	28.9	0.7
Parameter 8	31.6	0.8
Parameter 9	36.8	1
Parameter 10	42.1	1.6
Parameter 11	44.7	1.8
Parameter 12	47.4	2
Parameter 13	50	2.1
Parameter 14	52.6	2.3
Parameter 15	55.3	2.5

We have compared the results of Location Averaging with other outlier methods such as Part Averaging Testing. We have observed in a number of cases that the Location Averaging method has been more effective than Part Average Testing in screening burn-in fails or customer returns. We will show a couple of examples comparing Part Average Testing and Location Averaging. Figure 5 shows the raw data distribution for one parameter and one wafer. It shows that the burn-in fail on this wafer is in the middle of the distribution. This burn-in fail cannot be screened using the Part Average Testing on this parameter. Figure 6 shows the residual and estimate data obtained from Location Averaging for the same parameter and wafer. In the residual data plot, the same burn-in fail becomes an outlier and it is beyond the 99% Prediction Limit (blue lines). Another comparison is shown in Figure 7 and Figure 8. In this example, a burn-in fail is on the tip of the raw data distribution. This burn-in fail is not an outlier. The same burn-in fail becomes an outlier in the residual plot. Both these examples illustrate that Location Averaging is very useful in amplifying the outlier signature.

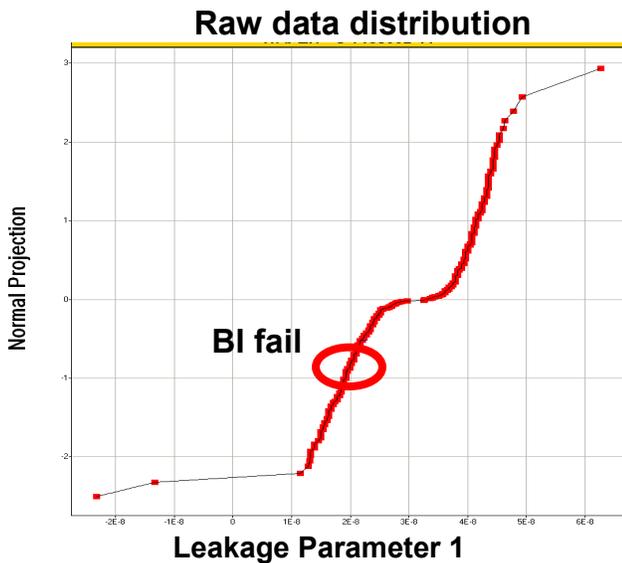


Fig. 5: Raw distribution for a parameter. Burn-in fail in the middle of the distribution.

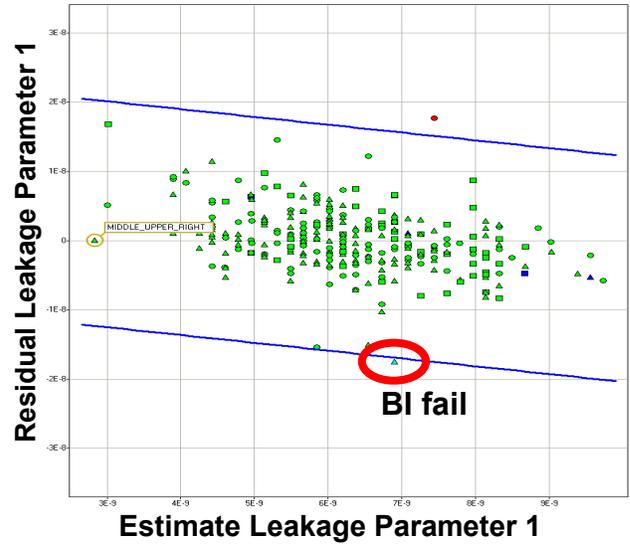


Fig. 6: Residual and Estimate regression for the same parameter as shown in Fig 5. Burn-in fail is an outlier.

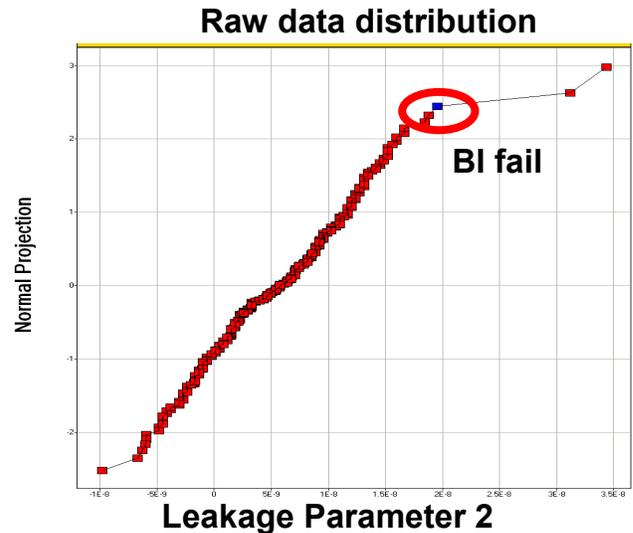


Fig. 7: Raw distribution for a parameter. Burn-in fail in the middle of the distribution.

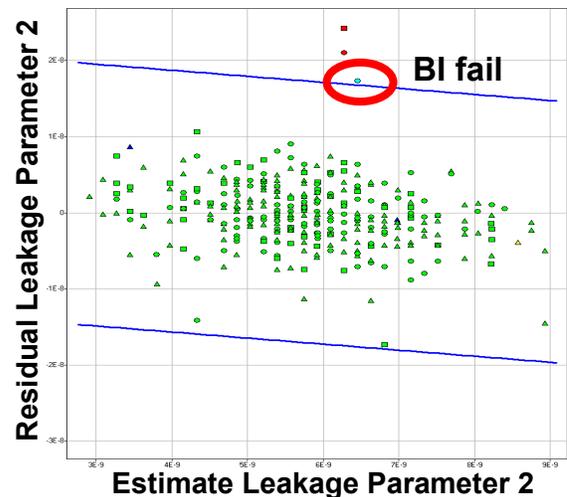


Fig. 8: Residual and Estimate regression for the same parameter as shown in Fig 7. Burn-in fail is an outlier.

VI. CONCLUSION

We have described a methodology to utilize production wafer probe data to identify at risk material early in the production process. Variance reduction techniques are employed to amplify defect signatures and compress natural variability. Statistical techniques identify parameter correlations to fails of interest and an ATPG-inspired method is used to compact the number of screening parameters to minimize overkill. A case study illustrates the power of the technique which has proven effective across a wide range of products, technologies, and manufacturing situations.

REFERENCES

- [1] "International Technology Roadmap for Semiconductors", <http://public.itrs.net/>
- [2] C. Crapuchettes, "Testing CMOS IDD on Large Devices," *Proc. 1987 IEEE Int. Test Conf.*, pp. 310-315, Sep. 1987.
- [3] J. M. Soden, C. F. Hawkins, R. K. Gulati, and W. Mao, "IDDQ Testing: A Review," *J. Electronic Testing: Theory and Applic. (JETTA)*, vol. 3, no. 4, pp. 291-303, Dec. 1992.
- [4] T. W. Williams, R. H. Dennard, R. Kapur, M. R. Mercer, W. Maly, "IDDQ Test: Sensitivity Analysis of Scaling," *Proc. 1996 IEEE Int. Test Conf.*, pp. 786-792, Oct. 1996.
- [5] A. E. Gattiker, W. Maly, "Current Signatures: Application," *Proc. 1998 IEEE Int. Test Conf.*, pp. 1168-1177, Oct. 1998.
- [6] P. Buxton, P. Tabor "Outlier Detection for DPPM Reduction," *Proceedings IEEE Int. Test Conference, 2003*, pp. 818-827, Sep-Oct 2003.
- [7] Automotive Electronics Council, "Guidelines for Part Average Testing (PAT)," AEC Q001, rev. C, 18 July 2003, <http://www.aecouncil.com/AECDocuments.html>.
- [8] S.S. Sabade, D. M. Walker, "Evaluation of Effectiveness of Median of Absolute Deviations Outlier Rejection-based IDDQ Testing Burn-in Reduction" *Proceedings 2002 VLSI Test Symp.*, pp. 81-86, Apr-May 2002.
- [9] W. R. Daasch, J. McNames, D. Bockelman, K. Cota, R. Madge, "Variance Reduction Using Wafer Patterns in IDDQ Data" *Proceedings IEEE Int. Test Conference, 2000*, pp. 189-198, Oct 2000.
- [10] R. Daasch, K. Cota, J. McNames, R. Madge, "Neighbor Selection for Variance Reduction in IDDQ and Other Parametric Data" *Proceedings IEEE Int. Test Conference, 2001*, pp. 92-100, Oct 2001.
- [11] C. Schuermyer, B. Benware, K. Cota, R. Madge, R. Daasch, L. Ning, "Screening VDSM Outliers Using Nominal and Subthreshold Supply Voltage IDDQ," *Proceedings IEEE Int. Test Conference, 2003*, pp. 565-573, Sep 2003.
- [12] K. M. Butler, S. Subramaniam, A. Nahar, J. M. Carulli Jr., T. J. Anderson, W. R. Daasch, "Successful Development and Implementation of Statistical Outlier Techniques on 90nm and 65nm Process Driver Devices," *Proc. 2006 Int. Reliability Physics Symp.*, pp. 552-559, Oct. 2006.
- [13] D. I. Bergman, H. Engler, "Improved IDDQ Testing with Empirical Linear Prediction", *Proc. 2002 IEEE Int. Test Conf.*, page 954, Nov. 2002.
- [14] S.S. Sabade, D.M.H. Walker, "Improved wafer-level spatial analysis for I_{DDQ} limit setting," *Proc. 2001 IEEE Int. Test Conf.*, pp 82-91, Nov. 2001.
- [15] S.S. Sabade, D. M. H. Walker, "CROWNE: Current Ratio Outliers with Neighbor Estimate", DFT 2003, pp 132-139.
- [16] C. Thibeault, "Diagnosis Method Using Δ IDDQ Probabilistic Signatures: Theory and Results", *Journal of Electronic Testing: Theory and Applications 2000*, Vol 4., pp 339-353.
- [17] D. S. Yates, D. S. Moore, D. S. Starnes, "The practice of statistics"