Defect detection under Realistic Leakage Models using Multiple I_{DDQ} Measurements

Chintan Patel, Abhishek Singh and Jim Plusquellic
cpatel2@cs.umbc.edu, abhishek@cs.umbc.edu, plusquel@cs.umbc.edu
Department of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Abstract

I_{DDQ} or steady state current testing has been extensively used in the industry as a mainstream defect detection and reliability screen. The background leakage current has increased significantly with the advent of ultra deep submicron technologies. The increased background leakage makes it difficult to use single threshold I_{DDQ} testing to differentiate defect-free chips from those with defects that draw small amount of currents. Several techniques that improve the resolution of I_{DDQ} testing have been proposed to replace the single threshold detection scheme. However, even these techniques are challenged to detect defects in the presence of leakage currents in excess of a few mA. All of these techniques use a single I_{DDQ} measurement per circuit configuration for detection and thus the scalability of these techniques is limited. Quiescent Signal Analysis (QSA) is a novel I_{DDQ} defect detection and diagnosis technique that uses I_{DDQ} measurements at multiple chip supply pads. Implicit in our methodology is a leakage calibration technique that scales the total leakage current over multiple simultaneous measurements. This helps in decreasing the background leakage component in individual measurements and thus increases the resolution of this technique to subtle defects. The effectiveness of this technique is demonstrated in this paper using simulation experiments on portion of a production power grid. Predicted chip size and leakage values from the International Technology Roadmap for semiconductors (ITRS) are used in these experiments. The performance of the proposed technique is evaluated using three different intra-die process variation distribution models.

1.0 Introduction

The advantages of analyzing power grid signals were recognized more than a decade ago with the introduction of I_{DDQ} testing. Here, an elevation in the steady-state current of a chip beyond a threshold was determined to be a reliable indication of the presence of a shorting defect in the circuit under test (CUT). Unfortunately, advances in silicon technology, in combination with increases in chip size and transistor density, have caused increases in the background steady-state current of defect-free chips, making it difficult to distinguish the defective chips using a single threshold technique [1]. Along with the increase in the magnitude of background leakage current, the variability in the current value from chip-to-chip (inter-die) as well as between different regions of a particular chip (intra-die) has increased significantly. However, the properties of the power grid continue to remain attractive from a testing perspective, and alternative multi-threshold I_{DDQ} methods and novel transient techniques are drawing considerable attention.

Several techniques that rely on a self-relative or differential analysis, in which the average I_{DDQ} of each chip is factored into the pass/fail threshold have been proposed. Although the application of these techniques to low power chips will continue, these methods are expected to become increasingly less effective for high performance ASICs with high background leakage currents. This is true because an increase in the thresholds employed by these techniques to account for increased leakage will reduce their resolution to defect currents.

An alternate calibration strategy that may have better scaling properties is to distribute the total leakage current across a set of measurements. This is accomplished by introducing probing hardware either on chip or off chip that allows access to individual power supply ports. The method proposed in this work called Quiescent Signal Analysis (QSA), is designed to exploit this type of leakage calibration as a means of increasing defect detection resolution. A secondary diagnostic benefit of such a technique is described in [2-5].

A linear regression analysis procedure is proposed for QSA that calibrates for high background leakage currents. This procedure is derived from our previous work on Transient Signal Analysis [6]. In TSA, multiple power supply transient signals are analyzed simultaneously as a means of both detecting the regional signal variations introduced by defects and diminishing the signal variations introduced by process variation effects. In QSA, this procedure performs a similar function of distinguishing globally distributed leakage current from the regional defect current.

In this work, an extensive set of spice simulations are used to demonstrate the defect detection capabilities of QSA in presence of significant background leakage noise and three different intra-device process variation models. The analysis is performed on a portion of a Production Power Grid (PPG) referred to as the Q9. The simulation models were derived using projected values for chip size, number of power supply ports and leakage currents obtained from the ITRS [7]. As it is infeasible to run spice simulations on the whole chip, the simulation models were derived by scaling the whole chip values to the size of the Q9. Defect-free chips were modeled using ITRS specified "whole chip" leakage values in the range of 1mA to

150mA. This range covers high and medium performance, medium power and low power chips. Three different intra-device process variation distributions were used in combination with the above mentioned base leakage values to generate 48 defect free models. Two of the local variation models were symmetric or regular in nature while the third model was random. 1800 defect models were generated using defect values in the range of 10µA to 100µA in combination with the above mentioned defect-free models to determine the detection resolution.

2.0 Related Work

Single-threshold I_{DDQ} technique relied on the fact that the steady state current distribution of defect-free chips is distinct from that of the defective ones. A chip that draws current that exceeds the defect-free current distribution by a fixed threshold is deemed as defective. With the advent of deep sub-micron technologies, the overlap in these distributions makes it difficult to set an absolute pass/fail threshold. The increase in sub-threshold and gate leakage currents in newer technologies can result in defect-free leakage currents that are significantly higher than the defect current. Thus, calibration methods are required to reduce the adverse effects of high leakage currents on defect current resolution. Several techniques based on a self-relative or differential analysis are proposed as a solution to this problem. A current signature method is proposed by Gattiker et. al. [8], that looks for discontinuities in the curve obtained by sorting I_{DDQ} measurements in ascending order. Delta I_{DDQ} is a differential I_{DDQ} method proposed by Thibeault [9] in which differences between successive I_{DDQ} measurements are compared to a threshold. Maxwell et. al. [10], proposed a current ratio method where chip specific thresholds are derived by using vectors that produce the minimum and maximum I_{DDQ} values. A clustering technique that groups good chips separately from bad chips is proposed by Jandhyala et. al. [11]. Daasch et. al. [12] describe a method that predicts device I_{DDQ} using the spatial proximity correlations among chips on a wafer. Variyam [13] proposes a linear prediction based technique in which each I_{DDQ} value among a set of values for a given chip is predicted from the remaining I_{DDQ} values in the set. Singh et. al [14] showed that I_{DDQ} readings of the neighboring die on a wafer can be used for variance reduction and to identify wafer-level spatial outliers. Sabade et. al [15][16] have also developed methods based on wafer-level spatial correlation analysis in which they derive a maximum defect-free I_{DDQ} threshold from the analysis of neighboring die.

Many of these process-tolerant I_{DDQ} methods use relative pass/fail thresholds instead of absolute thresholds. Also the other major similarity in all of the above techniques is that they use a single I_{DDQ} measurement per circuit configuration per die. As the variance in the I_{DDQ} values increases, it tends to increase the threshold bands in most of these techniques, thus decreasing their defect resolution. QSA differs from these methods by correlating individual supply I_{DDQ}s within each state vector. A regression analysis procedure in combination with outlier analysis is used to differentiate defect-free and defective devices. Therefore, the cross-correlation performed in QSA additionally calibrates for vector-to-vector variations. This is likely to further improve the process tolerance of the method. Also this method can be used in combination with all of the above mentioned vector-to-vector analysis techniques to further improve defect resolution.

Another advantage of a method that uses multiple supply port measurements is the natural scalability that this approach incorporates. The scalability features of QSA should make it possible for it to remain effective at detecting defects as chips get larger and incorporate larger numbers of more densely packed transistors. QSA is designed to exploit design trends that add additional supply ports (pads that interface to the external supply) as chip sizes and current requirements increase. However, it should be noted that this benefit of increased resolution comes with the cost of increased test time as multiple measurements need to be performed per vector.

Perhaps a greater benefit of using multiple power supply signals is that they offer information beyond defect detection. In our previous work, we have demonstrated the ability of QSA for application to defect diagnosis [2-5]. The procedure predicts the (x,y) coordinates at which a defect draws current from the power grid in the layout. To our knowledge, no other method that is based on the analysis of a chip's electrical signals is able to provide this type of information. Such information is extremely useful in failure analysis procedures, which are designed to determine the root cause of chip failures.

3.0 QSA Detection Procedure

QSA analyzes a set of I_{DDQ} measurements, each obtained from individual supply pads from the Chip-Under-Test (CUT). The resistive nature of the power grid causes the current drawn by the defect to be non-uniformly distributed to each of the supply pads. In particular, the defect draws the largest fraction of its current from supply pads topologically "nearby". The same is true of the leakage currents. However, only the leakage currents in the vicinity of the defect contribute to the measured current in these pads. The smaller background leakage component improves the accuracy of the defect current measurement.

The fraction of the defect current provided by each of the pads in the region of the defect is proportional to the equivalent resistance between the defect site and each of the pads. Consider the resistance model of a simple power supply grid as shown in Figure 1.

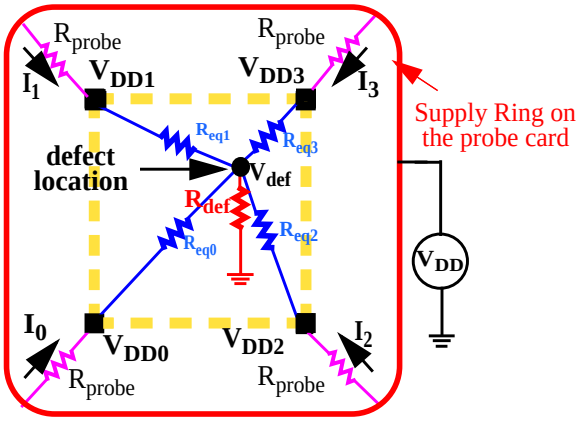

Figure 1. Equivalent resistance network with defect inside the circuit.

Here, R_{eq0} through R_{eq3} represent the equivalent resistances between each of the supply pads and the defect site shown in the center of the figure. The following set of equations describe the relationship between the power supply branch currents, I_0 through I_3 and V_{def}, the voltage at the defect site.

$$I_i \times (R_{eqi} + R_p) = V_{DD} - V_{def} \quad \text{for i = 0,1,2,3} \quad (1)$$

Consider the example shown in Figure 1. As the defect is topologically closer to V_{DD3}, it will have the lowest equivalent resistance to that pad and thus source the highest amount of current from that pad. Therefore, a defect causes regional variations where the current drawn from each pad is dependent on the equivalent resistance.

The defective device's I_{DDQ} consists of two components, the current drawn by the defect, and the leakage current. If the transistor density in the layout is regular, then the leakage current will be distributed evenly among all the supply ports. Each supply port in this case draws the same amount of leakage current as the other ports. A defect, in such a scenario, will cause more current to be sourced from a topologically closer pad and can be detected. However if the transistor density in the layout varies across the design, shown for example in Figure 2, the leakage current sourced by each supply pad will vary. This is due to the fact that the leakage current will be distributed by the power grid proportionally, as a function of resistance. This localized variation of the leakage currents will adversely affect a regional information based defect detection scheme.

The key observation concerning leakage current is that it is effected most significantly by the global variations introduced by changes in process and technology-related parameters. In other words, the current variations introduced by variations in these parameters will affect all transistors and junctions in a device in a similar manner. We are not claiming that intra-device variations do not exist, but rather, they are smaller in magnitude. The global nature

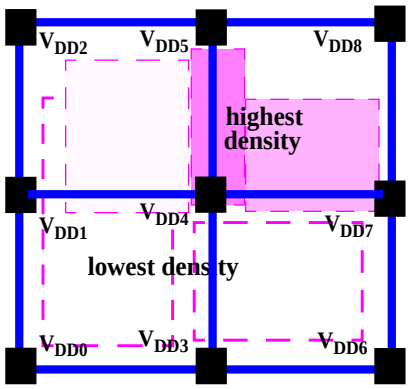

Figure 2. Unequal transistor densities in the layout.

of process variations scale the leakage currents to all supply ports, making it possible to track it using regression analysis.

Linear regression is used to track these global background leakage currents and provides a means of distinguishing them from the regional defect currents. The procedure is based on the analysis of scatter plots obtained by plotting the I_{DDQ} values at two supply ports. For example, Figure 3 represents a power grid with 16 V_{DD} supply ports. A set of defect-free spice simulations are run on the circuit where the leakage under each simulation is varied randomly across the grid.

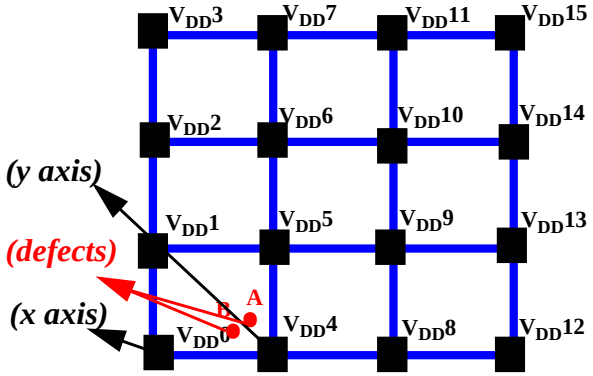

Figure 3. Power Grid with defects inserted at points A and B.

Figure 4 shows the scatter plot obtained by plotting the I_{DDQ} values at V_{DD1} against the I_{DDQ} values at V_{DD5} obtained under each of these simulations. A least squares estimate of the regression line is drawn through these defect-free data points. Two curves representing 99.95% prediction limits are shown around the regression line, delimiting a region referred to as the Process Variation Zone (PVZ). Here, the prediction limits are sensitive to both the number of simulations or samples and the amount of dispersion of the data points around the regression line (Mean Square Error or MSE).

The PVZ represents the defect-free chip *space* and accounts for intra-device process variations and measurement noise. Two more spice simulations are run, one with

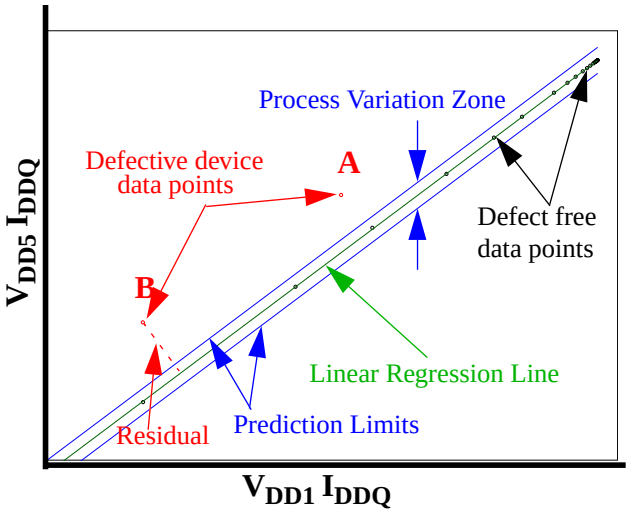

Figure 4. Scatter Plot, Regression Line and prediction limits (PVZ) for V_{DD1} and V_{DD5}.

defect A and another with defect B inserted in the circuit as shown in Figure 3. The I_{DDQ} values measured at V_{DD1} and V_{DD5} under these two simulation models are also plotted in Figure 4, labeled as A and B respectively. The regional variation caused by these defects in V_{DD5} is not well correlated with the variation measured at V_{DD1} on the same chip. The large I_{DDQ} at V_{DD5} in combination with the small I_{DDQ} at V_{DD1} generates data points outside the PVZ. For this pairing of V_{DD}s, the position of the data points outside the PVZ suggests that the last two circuit models are defective.

The standard statistical method of analyzing variance in scatter plots is through residuals. A residual is defined to be the shortest distance from a data point to the regression line, as shown in Figure 4. Residual Analysis, used in combination with the 99.95% prediction limits, make it straightforward to decide the pass/fail status of a chip. If more than one scatter plots are analyzed, a test chip fails if it produces at least one data point outside the corresponding PVZs.

One metric to evaluate the effectiveness of the technique would be to count the number of pairings for which the defective device data points fall outside the PVZ. However, in addition to this metric, it is also meaningful to examine the magnitudes of the residuals. In order to make this value meaningful for comparisons with other values, the magnitude of the residuals are normalized or standardized using Equation 2.

$$ZRES = \frac{residual}{\sqrt{MSE}} \qquad (2)$$

Here, MSE is the variance of the defect-free simulation residuals. For the experiments in this paper, the prediction bands are used as the pass/fail threshold for identifying the defective devices and the ZRES values are used to evaluate the confidence of the prediction. In other words, a device fails if at least one data point falls outside of a predetermined prediction band for any V_{DD} pairing. Moreover, the confidence that a test device is defective is higher for larger values of ZRES.

4.0 Production Power Grid

Figure 5 shows the 80,000 by 80,000 unit layout of the PPG. The PPG interfaces to a set of external power supplies through an area array of V_{DD} and GND C4 pads. A C4 pad is a solder bump for an area array I/O scheme. The PPG has 64 V_{DD} C4s and 210 GND C4s (not shown in Figure 5). The 64 V_{DD} C4s divide the PPG into 49 different regions called Quads. Due to space and time constraints, it was not possible to run spice simulations on the entire PPG. Rather, a portion of the PPG consisting of 9 quads was simulated using spice. This portion consists of the lower left 9 Quads as shown in Figure 5, and is subsequently referred to as the Q9. The Q9 occupies a 30,000 by 30,000 unit area and has 16 V_{DD} C4's.

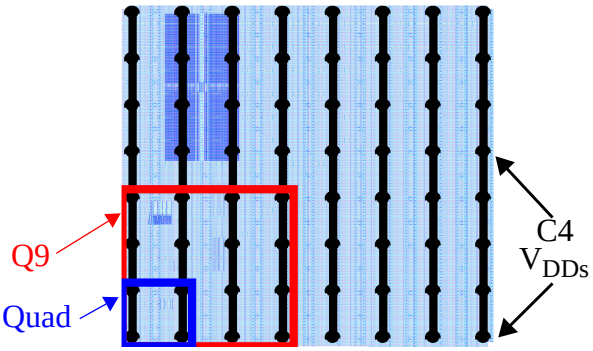

Figure 5. Layout of the PPG.

Figure 6(a) expands on the lower left corner of the PPG by showing a more detailed diagram of the 10,000 by 10,000 unit region called the Quad. This is again expanded in Figure 6(b). At this level, it can be seen that the grid is constructed over 4 layers of metal with metal 1 and 3 running vertically and metal 2 and 4 running horizontally. The C4s are connected to wide runners of vertical metal 5, indicated as m5 in Figure 6(a), that are in turn connected to the m1-m4 grid. In each layer of metal, the V_{DD} and GND rails alternate. In the vertical direction, each metal 1 rail is separated by a distance of 432 units. The alternating vertical V_{DD} and GND rails are connected together using alternating horizontal metal runners. Stacked contacts are placed at the appropriate crossings of the horizontal and vertical rails.

The R model of the PPG was obtained from an extraction script using parameters characterizing TSMC's 0.25mm process. A well characterized probe card model described in [17] was used to model the tester power supply and probe card contact resistance to the chip. The combined resistance network contains approximately 27,000

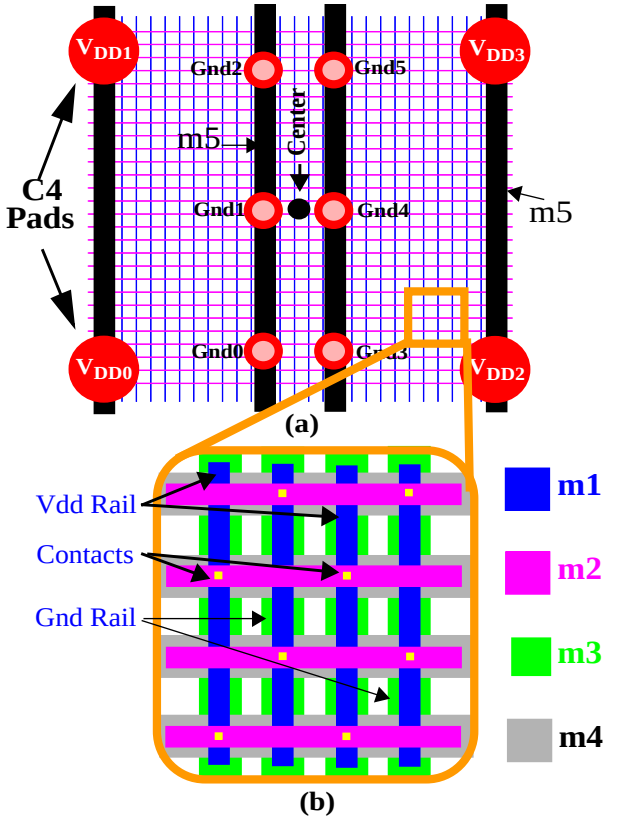

Figure 6. Detail of the "Quad": Portion of the PPG.

resistors per quad.

5.0 Simulation Models

The simulation models were derived according to the current technology node, the expected chip size and nominal I_{DDQ} for different categories of chips as described in the ITRS. The maximum I_{DDQ} for high performance ASICs is predicted to be anywhere from 70mA to 150mA. I_{DDQ} for low power, low speed chips will be significantly lower than these values and can be anywhere from 1mA to a few tens of mA. The area of the chip, once is production, is predicted to stay relatively constant around 140 mm^2. The total number of V_{DD}/GND pads would be around 1700 for high performance ASICs out of which we expect $1/3^{rd}$ will be V_{DD}s (400 - 500 pads). As mentioned earlier, due to memory and time constraints it is infeasible to run simulations using the power grid for the whole chip. Therefore a portion of the chip namely the Q9 is used for running simulations to validate the proposed technique. The I_{DDQ} and chip area values shown above are scaled to derive the background leakage values for the Q9. The area of Q9, if fabricated in the 0.13µm technology node, would be 4.85 mm^2. Therefore, if the I_{DDQ} for the whole chip is about 150 mA the I_{DDQ} for the Q9 will be around 5.2 mA. To ensure that the model is not overly optimistic the number of V_{DD} pads can be compared. There are 16 V_{DD} pads in the Q9 which would translate to about 340 V_{DD} pads in the whole chip. This number is lower than the actual number of V_{DD} pads predicted for the whole chip indicating that the model is not overly optimistic, as it uses less measurement points than available.

As the background leakage current has a wide range depending on the type of chip being tested a range of 1mA to 150mA was used to model the leakage current. 19 values were selected in this range as the leakage values for defect-free chips and the corresponding leakage values for the Q9 were derived. These 19 defect-free models are referred to as the uniform leakage models. 8 of these values were in the range of 70mA to 150mA to model high performance ASICs. The other 11 were from 1mA to 70mA that model medium and low power chips. The values in each of these subsets model chip-to-chip or inter-device process variations in the base leakage. The leakage current is modeled by placing about 31,500 current sources on the metal1 rails in the Q9. The metal1 rails in the layout represent the transistor density in a particular region. Regions with higher transistor densities have more metal1 rails than regions with lower transistor densities. Placing the leakage sources regularly along the metal1 rails emulates the effect of having irregular transistor densities in the layout.

With decreasing device dimensions, one of the other major problems facing most parametric testing techniques is that of intra-device or region-to-region process variations. Although these local variation effects are significantly lower than the global inter-device leakage variations they cannot be ignored for current and future technology nodes. These variations can be caused during any of the several complex processing steps and are thus hard to model. They could either be completely random over the whole chip or could vary in different regions of the chip in a regular fashion. For deriving our simulation models, we consider three different intra-device process variation distributions, one random and two regular in nature.

The random distribution is modeled by first creating 4 random boxes with known minimum dimensions over the Q9 as illustrated in Figure 7. The values of the uniform leakage model sources that fall within each of these 4 regions were varied by +/-2.5% and +/-5%. This model is referred to as *Random-Boxes*.

The first regular distribution model is called *Edge-to-Edge* and is illustrated in Figure 8. Here the uniform leakage model current sources are varied from +5% to -5% from one edge to the other. 20 rectangular vertical slices are generated where the variation in each slice is 0.5%. Although the overall leakage current is not affected significantly, this type of variation changes the local leakage distribution in different regions of the Q9.

The last model is also a regular distribution as shown in Figure 9 and is referred to as *Center-Out*. Here 20 squares are generated and the uniform leakage model current

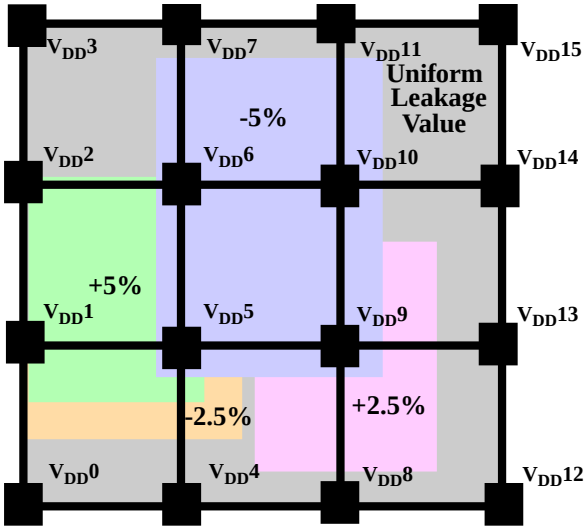

Figure 7. Random-Boxes (random distribution) model for local variations.

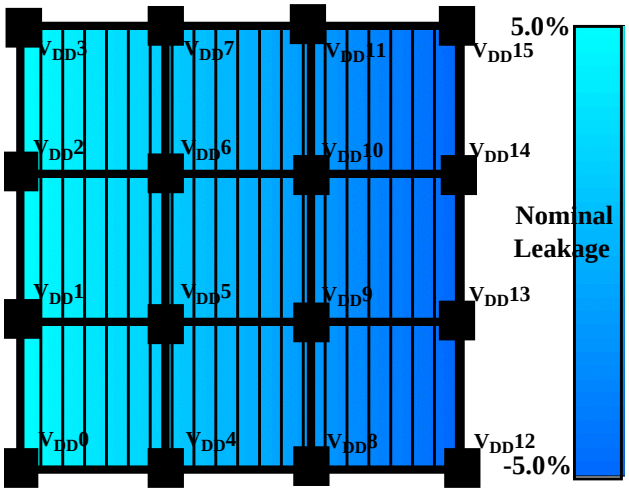

Figure 8. Edge-to-Edge (regular distribution) model for local variations.

sources are varied by -5% from the center to +5% at the outermost square, with a 0.5% variation per square. This model will not only change the local leakage distribution but also affect the overall leakage current as the size of the squares gradually increase as we move away from the center.

Using the 19 uniform leakage values mentioned above and the 3 local variation distributions a total of 76 defect free simulation models were derived. 19 models incorporated no local variations and were just the uniform leakage models. The others were combinations of each of these 19 uniform models with (1) Random-Boxes with +/-2.5% and +/-5% variation regions, (2) Edge-to-Edge variation of +/-5% and (3) Center-Out variation of +/-5%. A defect is modeled by inserting one extra current source among the 31,500 leakage sources. Defects were placed in the quad located in the center of Q9 as shown in Figure 10 and

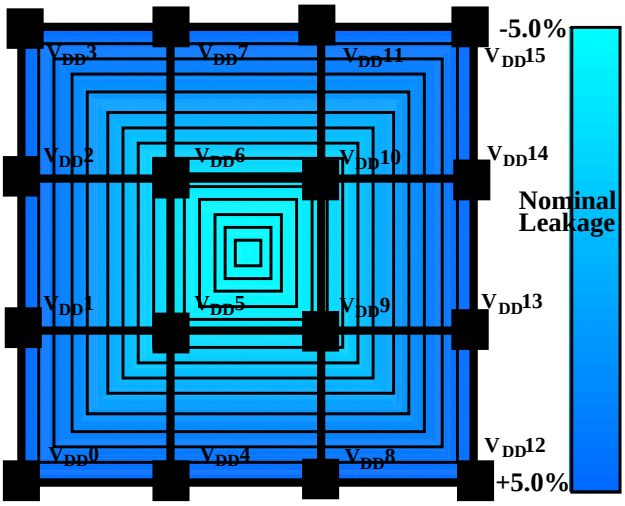

Figure 9. Center-Out (regular distribution) model for local variations.

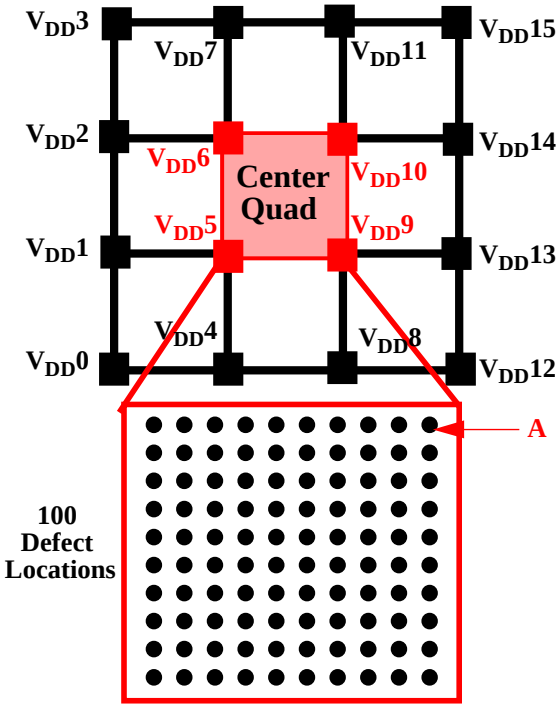

Figure 10. Center Quad and defect locations.

referred to as the Center Quad. 100 defect locations were selected in the Center Quad such that they are regularly distributed in a two-dimensional mesh like structure as shown in the figure. Different defect current values in combination with a leakage current model from above and the 100 defect locations were used to generate 1800 defective device simulation models.

6.0 Results and Discussion

Defect simulations were run using six different defect current and leakage current combinations (DLCs). The uniform leakage boundary values for high performance devices were used for the defect simulations and as

described earlier, they were scaled by the dimensions of Q9. Each of these combinations, shown in Table 1, were used in conjunction with the 100 defect locations and 3 different local variation models to derive the 1800 defective device models.

The defect draws the maximum amount of current from pads topologically closer to the defect site. Thus most of the defect current sourced by a defect in any quad, is supplied by the four V_{DD} pads that constitute the defective quad. In other words, the defect causes minimal change in the current sourced by pads outside the defective quad as compared to the defective quad pads. This helps in reducing the number of V_{DD} pairings analyzed for defect detection. The probability of detection is higher in each of the scatter plots that include one pad from the defective quad in combination with a pad from a neighboring quad. For example, if the defect is located in the Center Quad in Figure 10, most of the current drawn by the defect is supplied by V_{DD} pads $V_{DD}5$, $V_{DD}6$, $V_{DD}9$ and $V_{DD}10$. The V_{DD} pairings with the highest detection probability in this case will be, $V_{DD}1$-$V_{DD}5$, $V_{DD}4$-$V_{DD}5$, $V_{DD}5$-$V_{DD}9$, $V_{DD}5$-$V_{DD}6$, $V_{DD}2$-$V_{DD}6$, $V_{DD}6$-$V_{DD}7$, $V_{DD}6$-$V_{DD}10$, $V_{DD}10$-$V_{DD}11$, $V_{DD}10$-$V_{DD}14$, $V_{DD}10$-$V_{DD}9$, $V_{DD}9$-$V_{DD}13$ and $V_{DD}9$-$V_{DD}8$. Thus for any defect in the Center Quad we need to analyze the scatter plots obtained using the above 12 V_{DD} pairings. A similar procedure can be used to construct the scatter plot combinations for defects that occur in other quads.

This reduced set of scatter plots can be analyzed only if the defective quad can be identified. In most cases, it is simple to identify the defective quad by sorting the I_{DDQ}s drawn from each pad in descending order. If the first three pads are non-colinear and constitute a quad then that quad is the defective quad. However, if the defect is very close to the boundary of two quads this condition might not hold. Consider the defect marked A in Figure 10. This defect will draw maximum current from $V_{DD}10$. The second and third highest in the list can be $V_{DD}9$, $V_{DD}11$, $V_{DD}6$ or $V_{DD}14$, depending on the resistance profile of the grid in that region. In such cases, either all possible scatter plots for each of these quads can to be considered or a technique similar to the one proposed in our previous work on defect diagnosis using QSA can be used to identify the defective quad [5]. The second solution requires a small DFT structure to be inserted under each V_{DD} C4 (see [5] for details).

DLC #	Chip Uniform Leakage	Scaled Q9 Uniform Leakage	Defect Current
1	150mA	5.192mA	100µA
2	150mA	5.192mA	50µA
3	150mA	5.192mA	25µA
4	70mA	2.422mA	50µA
5	70mA	2.422mA	25µA
6	70mA	2.422mA	10µA

Table 1: Defect and uniform leakage combinations used for defect simulations.

6.1 Edge-to-Edge Local Variation Model

A total of 600 defect simulation models incorporated this type of local variation. The defect free scatter plots were generated using 38 defect free models namely, 19 uniform leakage models and 19 Edge-to-Edge defect-free models. The data analysis for these set of simulations is presented in Table 2. As shown in the first two rows of the table all the 600 defects were detected in this case. Also shown in row three is the average number of detections for all the defects over all the 12 scatter plots. A higher number suggests that each defect was detected multiple number of times in different scatter plots.

The higher probability of detections in this case would be for scatter-plots between pads that are well correlated in presence of this type of intra-device process variations. Closely studying Figure 8 reveals that all scatter plots between V_{DD} pads that are located vertically adjacent to each other should provide the best results. This is confirmed by looking at the number of detections per scatter plot (not shown in table), where all such scatter plots consistently have higher number of detections than the ones that analyze horizontally adjacent V_{DD} pads.

Figure 11 shows the detection sensitivity for all the 100 defect locations over the 12 scatter plot pairings for DLC #3. This combination has the minimum defect current in the presence of 150mA of uniform leakage current. The x and the y axis give the location of the defect in the center quad and the z dimension reports the maximum *difference,*

	DLC # 1	DLC # 2	DLC # 3	DLC # 4	DLC # 5	DLC # 6
Total number of defects	100	100	100	100	100	100
Defects detected	100	100	100	100	100	100
Average number of detections over 12 scatter plots	9.03	7.18	6.09	6.91	5.88	5.01

Table 2: Edge-to-Edge Local Variation Detection Data

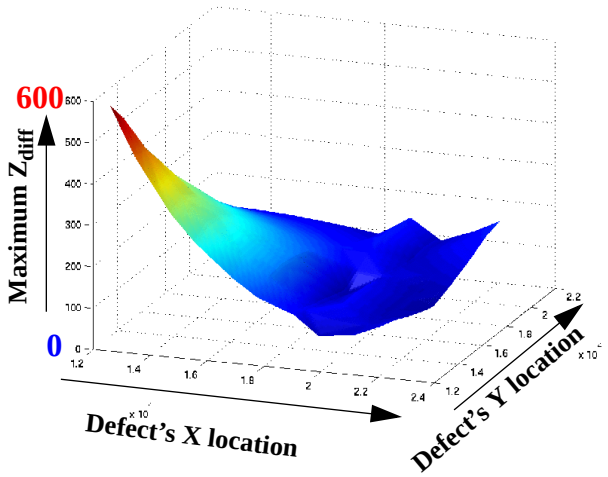

Figure 11. Edge-to-Edge model: Maximum Z_{diff} value distribution for DLC # 3.

Z_{diff}, between the standardized residuals (ZRES) of a defective device data point and the prediction band. The maximum Z_{diff} value gives the measure of confidence with which a device can be deemed as defective. In cases where the device data point falls outside the prediction bands of more than one scatter plot the probability of false detection is reduced. However, if the device data point is an outlier in only one or very few scatter plots then a safety threshold can be used for the minimum value of Z_{diff} required in at least one scatter plot to deem the device defective. If the maximum Z_{diff} value reported here is greater than the threshold the device can be identified as defective. As described earlier in Section 3, the standardized residuals are computed as the ratio of the defective device residual and the square root of the MSE. The MSE of a particular scatter plot is determined by variance of the defect free residuals. Thus scatter plots with highly correlated defect free device data points will have lower MSE values and thus higher detection sensitivity. *Similar analysis was performed for all defect models and also other local variation models. Due to space limitations only the detection results are presented here for these models.*

6.2 Center-Out Local Variation Model

A total of 600 defect simulation models incorporated this type of local variation model. The defect free scatter plots were generated using 38 defect free models namely, 19 uniform defect-free models and 19 Center-Out defect-free models. The data analysis for these set of simulations is presented in Table 3. As shown in the first two rows of the table all the 600 defects were detected in this case. Also shown in row three is the average number of detections for all the defects over all the 12 scatter plots. Compared to the previous model, the affect on detection sensitivity for this model is higher with decreasing defect currents. Also the absolute values suggest that devices with this type of variations will be harder to screen than former regular type of variation. Close inspection of Figure 9 would suggest that in this case scatter plots between V_{DD} pads that fall inside the same local variation band should provide better results. In our case, that translates to scatter plots between the four V_{DD} pads that surround the Center Quad and this trend was observed in the average number of detections per scatter plot.

6.3 Random-Boxes Local Variation Model

Again, a total of 600 defect simulation models incorporated this type of local variation model. The defect free scatter plots were generated using 38 defect free models namely, 19 uniform defect-free models and 19 Random-Boxes defect-free models. The data analysis for these set of simulations is presented in Table 4. As shown in the first two rows of the table all the defects except some in DLC #6 were detected in this case. Also it should be noted that 2 defect-free devices fall outside the prediction bands by a very small margin, when 99.95% confidence limits are used. Chips that incorporate these type of intra-device pro-

	DLC # 1	DLC # 2	DLC # 3	DLC # 4	DLC # 5	DLC # 6
Total number of defects	100	100	100	100	100	100
Defects detected	100	100	100	100	100	100
Average number of detections over 12 scatter plots	7.21	4.75	3.3	5.41	3.87	3.3

Table 3: Center-Out Local Variation Detection Data

	DLC # 1	DLC # 2	DLC # 3	DLC # 4	DLC # 5	DLC # 6
Total number of defects	100	100	100	100	100	100
Defects detected	100	100	100	100	100	81
Average number of detections over 12 scatter plots	9.78	6.99	3.58	8.44	5.26	1.41

Table 4: Random-Boxes Local Variation Detection Data

cess variations are the hardest to screen as the change in leakage distribution over different regions of the chip is random in nature. More significant variations of random nature can reduce the defect detection sensitivity of this technique. This model was incorporated as it is expected to be present in a real processing environment. +/-2.5% and +/-%5 variations in the uniform leakage value might be too high or too low depending on the maturity and the control of the process. Also we have ensured in the model that the boxes affected by the variations are small enough to affect the leakage characteristics of the Quad and the Q9. If these type of variations are present over larger regions, such that they encompass regions bigger than that bounded by the 4 surrounding V_{DD} pads, their adverse effect on the detection sensitivity will be reduced. If these variations are completely random over very small regions or even at a single transistor level, we expect that they might be averaged out thus again aiding the detection sensitivity of our technique.

As this model is random in nature the correlation coefficients of many scatter plots are comparable and so the maximum Z_{diff} values are spread over all these scatter plots. In majority of cases, scatter plots that use V_{DD} pads in vicinity of the defect location are better at detecting the defect. Also it should be noted that the variance in the defect free data points is the highest for this model and therefore the maximum Z_{diff} are significantly lower than the other two models.

6.4 Discussion

As presented in the previous three subsections, in most cases the proposed QSA technique is able to detect defects drawing as low as 10μA and 25μA current in the presence of 70mA and 150mA of leakage current. The major advantage of this technique is that it is scalable with increasing chip size as it distribute the leakage over a set of measurements. Many defects are detected in more than one scatter plot in most cases. This suggests that a set of experimental test chips can be used to predetermine the number of scatter plots to be analyzed, thus decreasing the number of current measurements required by this technique. Although the test time is expected to increase at most linearly, it might not be an exact multiple of the number of measurements made. This is due to the fact that for steady state measurements the setup time for the test is common over all these measurements. These measurements can be made either using specialized hardware on the ATE, on chip monitors or off chip monitors mounted on the probe card. Some ATE today have more that one power supply unit and have current measurement capabilities on each of this units. Several low cost desktop DFT testers have been proposed that will be able to make multiple I_{DDQ} measurements. Along with the defect detection capabilities, QSA data can provide extra information that can be leveraged for (1) a more balanced power grid design, (2) solving over heating and power dissipation problems associated with scan-based testing, (3) to study variability in the fabrication process and (4) as described earlier to physically determine the location of the defect in the device. Like all other I_{DDQ} techniques, this technique will also be affected by the resolution of the measurement instruments. Although it is desirable to have highly accurate current measurement capabilities to optimally exploit the advantages of this technique, the loss of resolution due to less accurate measurements is of the same order as all other techniques.

A lot of I_{DDQ} techniques have been proposed in the last decade to address the challenges posed by high background leakage currents and process variations. All these techniques are based on a single I_{DDQ} measurement per circuit configuration. I_{DDQ} thus measured corresponds to the current drawn by the sensitized defect and the leakage current distributed over the whole chip. To overcome this diluting of defect current contribution, I_{DDQ} measured under different test vectors is analyzed for detection. It would be difficult for these techniques to detect defects with very low defect current in the presence of very high leakage currents. Also these techniques are susceptible to inter-device, state dependent and vector-to-vector variations. For example, if all the devices are affected by these variation effects and have a 1% variance in the uniform leakage value of 150mA, that translates to 1.5mA of variance between different devices over one vector. It would be very difficult for any vector-to-vector analysis technique to detect defects that draw a few tens of μA of defect current. As an alternative, the proposed technique uses multiple measurements for a single vector and analyzes them to reduce the adverse effects of these type of variations. However, the resolution of QSA will be affected by the magnitude and the distribution of intra-device process variations. In this paper, we used three intra-device process variation models with variations in the range of +/-5% to demonstrate the detection capabilities of QSA. The resolution of this technique is likely to reduce, than reported in this work, with higher values of these type of variations. Although it is not possible to fairly compare existing techniques that are based on single measurement and vector-to-vector analysis with QSA that uses multiple measurement and per vector analysis, it is clear that the resolution of QSA will be higher than most of the existing techniques. It must be noted that the increase in resolution is obtained at the expense of making multiple measurements, which in turn translates to increase in test time. However, the significant increase in resolution can enable I_{DDQ} testing in present and future technology generations and can compensate for the increase in the test cost. One other major advantage is that this technique can be used in combination with any existing vector-to-vector analysis technique to further improve the defect resolution of the entire I_{DDQ} test suite as conceptually represented in

Figure 12.

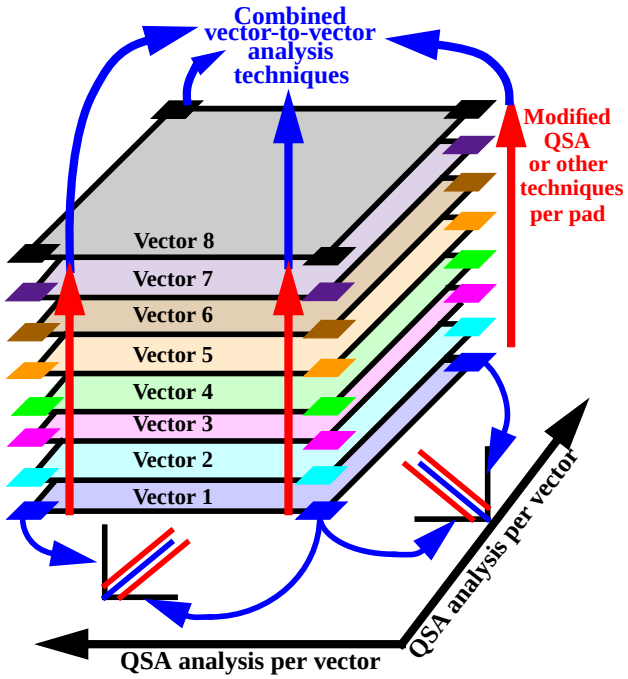

Figure 12. Combination of QSA and other vector-to-vector analysis techniques in a test suite.

The QSA analysis presented in this work can be used to perform a per vector analysis for each vector. Then either an enhanced version of QSA or any other pre-existing technique can be used to perform the vector-to-vector analysis. The vector-to-vector analysis can be performed either by adding the currents from all the measurements or individually at each supply pad. All the date reported in literature uses only one I_{DDQ} measurement and therefore not directly applicable to this technique. We are currently designing hardware experiments to validate most of the work presented here in hardware, however to make it statistically meaningful a large population would be required, which could only be provided by an industrial partner.

7.0 Conclusions

A novel defect detection technique based on leakage calibration using multiple I_{DDQ} measurements per vector called Quiescent Signal Analysis is proposed in this paper. The detection procedure is based on regression analysis in combination with outlier analysis. The defect detection capabilities of this technique are demonstrated using an extensive set of spice simulations. The robustness of this technique to very high background leakage currents and significant inter-device as well as intra-device process variations is presented. The detection sensitivity is analyzed in presence of three different type of intra-device leakage distribution models. Analysis has been provided to show that the scalability and sensitivity of this technique is expected to be better than existing I_{DDQ} techniques. The increased resolution provided by this method can enable I_{DDQ} testing in high performance ASICs and can compensate the increase in cost due to multiple measurements. We are currently developing a test chip to study the effectiveness of this method in silicon. This will also enable us to enhance the technique and propose a vector-to-vector analysis extension to this work.

8.0 References

[1] T.W.Williams, R.H.Dennard, R.Kapur, M.R.Mercer & W.Maly, "I_{DDQ} test: Sensitivity Analysis of Scaling", ITC, 1996, pp.786-792.

[2] Y. Ouyang and J. Plusquellic, "IC Diagnosis Using Multiple Supply Pad I_{DDQ}s" Design and Test, Volume 18, Number 1, pp. 50-61, Jan-Feb 2001.

[3] C. Patel and J. Plusquellic, "A Process and Technology-Tolerant I_{DDQ} Method for IC Diagnosis", VTS, 2001, pp. 145-150.

[4] C. Patel, E. Staroswiecki, D. Acharyya, S. Pawar and J. Plusquellic," A Current Ratio Model for Defect Diagnosis using Quiescent Signal Analysis", International Workshop on Defect Based Testing, 2002.

[5] C. Patel, E. Staroswiecki, S. Pawar, D. Acharyya and J. Plusquellic. "Defect Diagnosis using a Current Ratio based Quiescent Signal Analysis Model for Commercial Power Grids", JETTA, Volume 19, Issue 6, pp. 611-623, Dec 2003.

[6] A. Germida, Zheng Yan, J.F. Plusquellic and F.l Muradali, "Defect Detection using Power Supply Transient Signal Analysis", ITC, 1999, pp. 67-76.

[7] http://public.itrs.net.

[8] A.E.Gattiker and W.Maly, "Current Signatures", VTS, 1996, pp.112-117.

[9] C. Thibeault, "On the Comparison of Delta I_{DDQ} and I_{DDQ} test", VTS, 1999, pp. 143-150.

[10] P. Maxwell, P. O'Neill, R. Aitken, R. Dudley, N. Jaarsma, M. Quach, D. Wiseman, "Current Ratios: A self-Scaling Technique for Production I_{DDQ} Testing", ITC, 1999, pp.738-746.

[11] S. Jandhyala, H. Balachandran, A. P. Jayasumana, "Clustering Based Techniques for I_{DDQ} Testing", ITC, 1999, pp. 730-737.

[12] W. R. Daasch, J. McNames, D. Bockelman, K. Cota, "Variance Reduction Using Wafer Patterns in I_{DDQ} Data", ITC, 2000, pp. 189-198.

[13] P. N. Variyam, "Increasing the I_{DDQ} Test Resolution Using Current Prediction", ITC, 2000, pp. 217-224.

[14] A. Singh, "A Comprehensive Wafer Oriented Test Evaluation (WOTE) Scheme for the I_{DDQ} Testing of Deep Sub-Micron Technologies", Workshop on I_{DDQ} Testing, 1997.

[15] S. Sabade and D.M.H. Walker, "Improved Wafer-level Spatial Information for I_{DDQ} Limit Setting" ITC, 2001, pp. 82-91.

[16] S. Sabade and D.M.H. Walker, "Neighbor Current Ratios (NCR): A New Metric for I_{DDQ} Data Analysis", Defect and Fault Tolerance in VLSI Systems, 2002, pp. 381-389.

[17] D. Acharyya and J. Plusquellic, "Impedance Profile of a Commercial Power Grid and Test System", ITC, 2003, pp. 709-718.