# Detecting Trojans Though Leakage Current Analysis Using Multiple Supply Pad $I_{DDQ}$s

Jim Aarestad, Dhruva Acharyya*, Reza Rad+, and Jim Plusquellic
Department of Electrical and Computer Engineering, Univ. of New Mexico
*Verigy Inc.
University of Maryland, Baltimore County

## Abstract[1]

*Hardware Trojans have emerged as a new threat to the security and trust of computing systems. Hardware Trojans are deliberate and malicious modifications to the logic function implemented within digital and mixed signal chips. In contrast to software Trojans, it is not possible to simply 'scan the hard drive' to eradicate a hardware Trojan. Hardware Trojans can be designed to shutdown the chip at some pre-determined time and/or when some specific signal or data pattern is received. They may also be designed to remain hidden while leaking confidential information covertly to the adversary. Determining whether a hardware Trojan has been inserted into a chip is extremely difficult for a variety of reasons, e.g., nanometer feature sizes and chip design complexity combine to make optical inspection difficult or impossible. This paper presents experimental results demonstrating the effectiveness of a Trojan detection method that is based on the analysis of a chip's $I_{DDQ}$s (steady-state current), which are measured simultaneously from multiple places on the chip. The proposed method also incorporates a technique for virtually eliminating process and test environment (PE) variations effects which act to reduce detection sensitivity of traditional testing approaches. Used together, resolution enhancements of up to a 1000x are possible over conventional single power supply current measurement techniques. A regression-based statistical technique is applied to the data collected from a set of chips fabricated in a 65 nm process to illustrate the detection capabilities and limitations of this type of approach.*

## 1 Introduction

The security and trust of the chip design and fabrication processes are threatened by the horizontal dissemination of semiconductor companies and globalization of the industry to off-shore facilities [1][2]. In particular, integrated circuit (IC) "trust" relates to the degree of confidence one has that a fabricated instance of a chip implements only those functions described in the original specification -- nothing more and nothing less. A chip for which this does not hold true is said to have a hardware Trojan. A hardware Trojan, which is added to the chip by an adversary prior to fabrication, is designed to cause the chip to fail or leak confidential information while operating in the field. The adversary cleverly hides the Trojan to make it nearly impossible to detect it during the manufacturing test process for defects. Therefore, new testing approaches are needed to ensure IC trust.

There are several challenging aspects to the IC trust problem. The nanometer-sized physical dimensions of the wires and transistors that make up the chip, in combination with the complexity of current designs that integrate 100's of millions of such components precludes approaches that involve physically comparing micro-photographs of the chip with the original design. Secondly, the adversary can cleverly connect the Trojan circuit in the original design such that accidental or purposeful discovery using logic-based testing methods or parametric methods, such as those that measure power and delay, is highly improbable.

The main deficiency with parametric testing approaches as they have been defined for the purpose of detecting manufacturing defects is related to their sensitivity. As technology is scaled further into the nanometer domain, there is a corresponding increase in manufacturing process variations. Moreover, the increasing number of components that are integrated onto a single chip decreases the electrical signature of each component (the proverbial needle-in-a-haystack phenomenon). These trends combine to make it more difficult to identify the electrical signature of a Trojan circuit.

Even given these deficiencies, we believe that a parametric testing approach is the only universal solution for detecting the wide variety of possible Trojan implementations. This is true because, unlike logic-based testing methods or mission-mode watch-dog monitoring methods, parametric methods target the detection of "anomalies" in the chip's electrical behavior, and are not dependent on any specific function carried out by the Trojan. One can easily argue that any change to the chip's layout will introduce some type of electrical anomaly. The challenge of implementing an effective parametric Trojan-detection method is to design it with enough sensitivity to detect these small anomalies, while simultaneously building in a mechanism to filter out the natural electrical variations that occur because of manufacturing process variations.

In this paper, we propose a parametric approach that is based on the analysis of a chip's $I_{DDQ}$ (steady-state or quiescent current). Power supply analysis methods have been used for at least a decade in the manufacturing test community and more recently have been proposed for the Trojan detection problem [3]. A key contribution of our proposed approach is that we measure $I_{DDQ}$ at multiple places simultaneously across the 2-D surface of the chip. Our region-based $I_{DDQ}$ method directly addresses the adverse impact

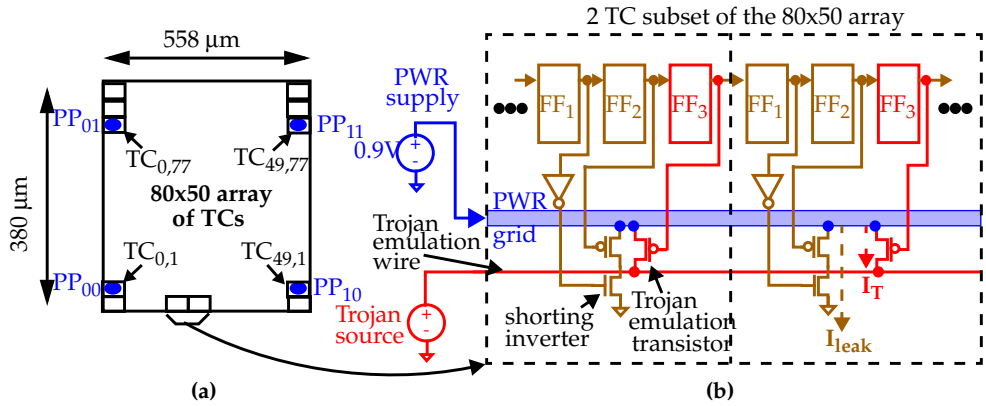



**Fig. 1. (a) Block diagram of the test structure and (b) details of the test circuits (TC).**

of increasing levels of process variations and leakage currents[1]. In previous work, we used simulation experiments to demonstrate that regional signal analysis significantly increases the resolution of power analysis methods to Trojans [4]. However, by itself, it is not sufficient for dealing with the adverse effects of process and test environment (PE) variations on detection resolution. To fully leverage the resolution enhancements available in a region-based approach, such methods must be combined with signal calibration techniques that are designed to attenuate and remove PE signal variation effects.

In this work, we apply linear regression analysis to the data collected from a special test structure designed to emulate hardware Trojans on a set of chips. The chips are fabricated in IBM's 65 nm, 10 metal layer SOI technology. The chips incorporate an array of cells that allow a Trojan to be emulated in one of 4,000 distinct locations on the chip. The design permits control over the position and magnitude of the Trojan current as well as the magnitude and distributional characteristics of the overall (chip-wide) leakage current. The results of our analysis demonstrate that detection sensitivity is strongly correlated with 1) the magnitude of the Trojan current, 2) the position on the power grid from which the Trojan sinks current and, 3) the pattern and variation in the chip-wide leakage current.

The rest of this paper is organized as follows. Previous work is described in Section 2. The test chip design is presented in Section 3. The procedure to emulate Trojans is described in Section 4 as well as our multiple supply port (MSP) and signal calibration processes. Section 5 provides an analysis of the impact of leakage current variations on the sensitivity of our method. We give our experimental results in Section 6 and conclude in Section 7.

## 2 Background

Security and trust is a major concern in the design and test of chips, particularly in situations that involve protecting secret keys and IPs. The malicious insertion of hardware Trojans in ICs is a new security and trust concern that must now be addressed in combination with conventional security risks. The following summarizes the published work on this topic.

The authors of [3] were the first to address the hardware Trojan issue. They propose the use of side-channel signals, e.g., transient power supply currents, to identify Trojans in chips. The authors of [5] propose a method that first determines a set of target 'hard-to-observe' sites for a Trojan with $q$ inputs and then uses automatic test pattern generation (ATPG) to generate patterns to activate the Trojan. A Trojan detection method that measures the combinational delay of a large number of register-to-register paths internal to the functional portion of the IC is proposed in [6]. In [7], the authors propose a region-based stimulation strategy and analyze the global power consumption to detect Trojans. In [8], the authors introduce special circuitry that enables the direct control of the least controllable nodes in the circuit as a means of triggering the activation of a Trojan. In [9], the authors build a path delay fingerprint of Trojan-free chips by running high coverage input patterns.

In previous work, we proposed several region-based $I_{DDQ}$ and $I_{DDT}$ test methods for detecting manufacturing defects and showed that techniques for calibrating PE variations are critical to providing adequate detection resolution [10][11]. The same concern holds true for Trojan detection. In this paper, we emulate Trojans at multiple places in a set of chips and demonstrate the level of detection sensitivity we can achieve using calibration and multiple supply port (MSP) $I_{DDQ}$s.

## 3 Test Chip Design

A block diagram of the test chip design is shown in Figure 1(a). It consists of a 80x50 array of test circuits (TCs) that occupies an area of dimension 558 µm in width and 380 µm in height. The power grid is wired over 10 metal layers and is connected to an external power supply through four power ports (PP) labeled $PP_{00}$ through $PP_{11}$. Each TC consists of three FFs connected in a scan chain configuration, a *shorting inverter*, and a *Trojan emulation transistor* connected to a globally routed *Trojan emulation wire*. A schematic diagram of two adjacent TCs is shown in Figure 1(b). The shorting inverters and Trojan emulation

1. A region is defined as a portion of the layout that receives the majority of its power from a set of surrounding power ports or C4 bumps.

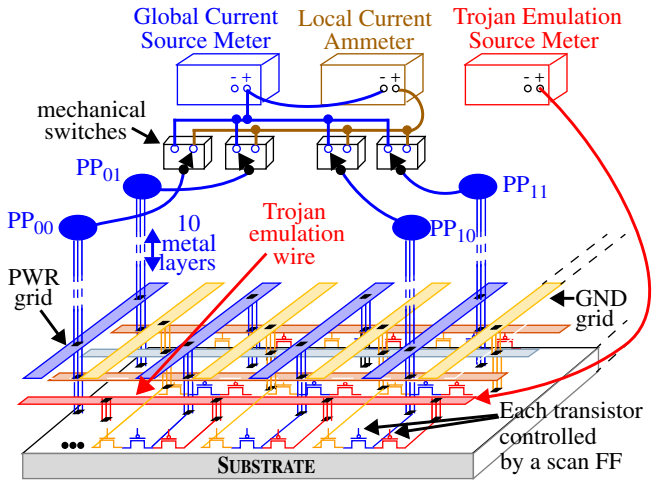



**Fig. 2. External Instrumentation Setup.**

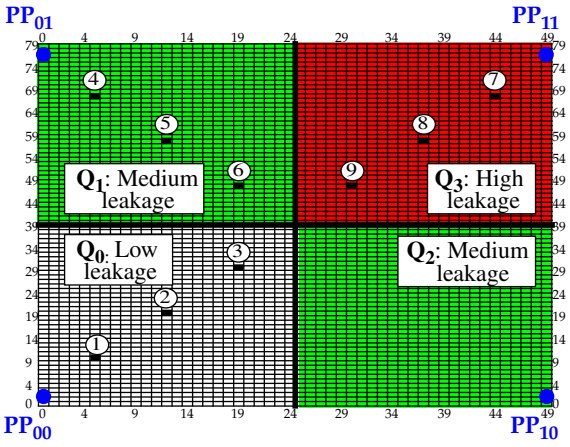



**Fig. 3. Positions of the nine TCs used for Emulating Trojans.**

transistors within each TC connect to the same point on the power grid.

The connection of the shorting inverters and the Trojan emulation transistors to point sources on the power grid enable two types of shorts to be introduced within any one (or more) of the 4,000 TCs. The first type shorts the power grid to ground through the inverter using $FF_1$ and $FF_2$ and the second type shorts the power grid to the Trojan emulation wire using $FF_3$. For the first type, the magnitude of the shorting current is pre-defined by the external power supply voltage, labeled *PWR supply* in Figure 1(b)[1]. We use the shorting inverters in the calibration process described below. For the second type, the magnitude of the shorting current is controlled through an external voltage source, labeled *Trojan source*. The scan chain enables a Trojan to be emulated at any point in the array by setting the Trojan source to a value less than the PWR supply voltage and scanning a bit pattern into the scan chain such that exactly one $FF_3$ contains a 0, which enables the Trojan emulation transistor, and the remaining 11,999 FFs contain 1's, i.e., everything else is disabled.

The scan chain also allows the ***off*** state of the shorting inverters to be configured into a high leakage (HL) and low leakage (LL) state. The leakage path is labeled as $I_{leak}$ in the right-most TC of Figure 1(b). For example, if the states of $FF_1$ and $FF_2$ are set with a '11', then both the n-channel and p-channel of the shorting inverter are off, resulting in a low leakage state. Alternatively, if the state is set to '01', then the n-channel is turned on while the p-channel remains off, resulting in a high leakage state. By configuring TCs in different regions into high- and low- leakage states, it is possible to create different leakage patterns across the array.

The external instrumentation setup is shown in Figure 2. As indicated above, the power ports are labeled $PP_{00}$ through $PP_{11}$ and wire out of the chip on separate pins in the package. The individual power pins are each wired to a low resistance mechanical switch as shown along the top portion in Figure 2. The switch can be configured in one of two positions, left or right. The left and right outputs of the four switches are connected to wires that route to a *Global Current Source Meter* (GCSM) and a *Local Current Ammeter* (LCA), respectively.

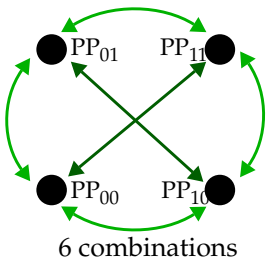
The GCSM is configured to provide 0.9 Volts to the PWR grid and is also able to measure the chip's global current with precision of approximately 1 µA. The LCA is wired in series with the GCSM and allows the individual power port (PP) currents to be measured at the same level of precision. As an example, the configuration of the switches as shown in Figure 2 enable the measurement of $PP_{00}$ current, $I_{00}$, through the LCA. The mechanical switches can be electronically controlled to enable the measurement of each of the other PP currents. The *Trojan Emulation Source Meter* (TESM) is used to set the voltage of the Trojan emulation wire, which routes out of the chip through a separate pin in the package, and to measure its current, $I_T$ (see Figure 1(b)).

## 4 Trojan Emulation Experiments

The hardware experiments are designed to investigate the capabilities of our methods for detecting emulated Trojans, hereafter referred to as Trojans, that connect to the power grid at various places. To accomplish this goal, we selected nine locations to emulate Trojans as shown in Figure 3. The grid of rectangles represent the 80x50 array of TCs. The Trojans are emulated by enabling the Trojan emulation transistors, one at a time, at each of nine labeled locations.

In order to investigate the impact of state-dependent leakage on the sensitivity of our methods, the states of the $FF_1$ and $FF_2$ in the array are set to implement the leakage pattern shown in Figure 3. The leakage pattern sets all 1,000 TCs in the upper right quadrant ($Q_3$) in the high leakage (HL) state. In addition, the leakage in quadrants $Q_1$ and $Q_2$ is set to medium leakage (ML) in which alternating TCs

1. The PWR supply is held constant in our experiments at 0.9 V.




**PP pairings**

$PP_{00}$-$PP_{01}$
$PP_{00}$-$PP_{10}$
$PP_{01}$-$PP_{11}$
$PP_{10}$-$PP_{11}$
$PP_{00}$-$PP_{11}$
$PP_{01}$-$PP_{10}$

**Fig. 4. Power port pairing combinations.**

are set in HL and LL states, while all TCs in $Q_0$ are set to the LL state. This pattern meets our goal of emulating an actual IC in which the leakage is state dependent, with regions of high and low leakage. We use this leakage pattern in the Trojan-free experiments and for each of the Trojan emulation experiments, described below.

The methods that we propose are statistically based and make use of 'golden models' of the chip to establish statistical limits of Trojan-free chip behavior. The golden models are derived by extracting RC-transistor models from the layout of the (Trojan-free) chip using a set of process parameters that characterize the chips to be tested. Since we have control over the insertion of the Trojan in our chips, we instead derive the statistical limits from a Trojan-free configuration of the chips themselves in this work. The advantage of this strategy is that our characterization of Trojan-free chip behavior is determined from actual hardware. The drawback is that such an approach is difficult to implement in practice because it is not known which chips (if any) are Trojan-free. We plan to investigate the alternative 'simulation-defined' golden models approach in future work.

The *Trojan-free* configuration is implemented in our experiments by disabling all Trojan emulation transistors and setting the TESM to 0.9V. A Trojan-free data set is collected for each chip by measuring the global and local currents through each of the power ports. This process produces a set of four global and four local currents for each chip.

For each Trojan experiment, exactly one Trojan emulation transistor is enabled and the TESM is swept across a sequence of voltages, from 0.89 V to 0.80 V in 10 millivolt steps, for a total of 10 steps. At each TESM voltage, the local and global currents are again measured. The value of the TESM voltage determines the magnitude of the current sinked through the Trojan emulation transistor. Table 1 shows the mean and standard deviation of currents that are introduced under each of the TESM voltages across the set of chips used in the experiments. The currents scale

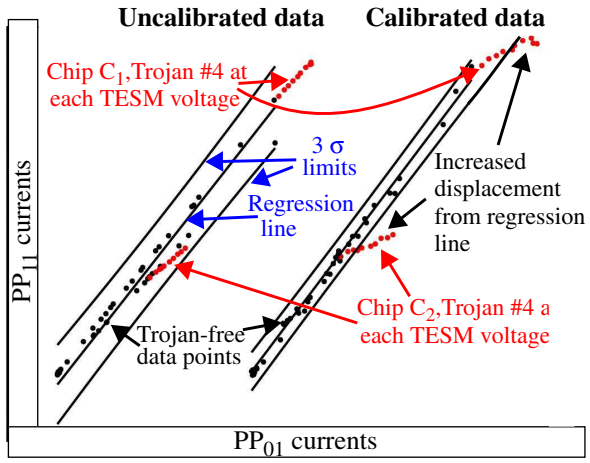



**Fig. 5. Scatter plot of raw currents for PP pairing $PP_{01}$ (x-axis) and $PP_{11}$ (y-axis) for Trojan-free experiments and emulated Trojan #4 using chips $C_1$ and $C_2$.**

approximately linearly by 6 µA from 0.89 V to 0.80 V.

| TESM V | 0.89 | 0.88 | 0.87 | 0.86 | 0.85 | 0.84 | 0.83 | 0.82 | 0.81 | 0.80 |
|---|---|---|---|---|---|---|---|---|---|---|
| µ (uA) | 8 | 15 | 21 | 27 | 34 | 40 | 45 | 51 | 57 | 62 |
| σ (uA) | 2.0 | 3.1 | 4.3 | 5.4 | 6.6 | 7.6 | 8.7 | 9.7 | 10.7 | 11.6 |

**Table 1: Trojan current statistics: mean µ and standard deviation σ in µA.**

In summary, for each chip, the data collection procedure produces 91 data sets, 1 Trojan-free data set and 90 Trojan data sets (9 Trojans * 10 TESM voltages). With 45 chips, there are a total of 45 Trojan-free data sets and 45x90 or 4,050 Trojan data sets.

### 4.1 Regression Analysis

We implement our statistical analysis method using scatter plots. The scatter plots are created by plotting the currents measured from one PP (power port) against the corresponding values measured at a second PP. Figure 4 illustrates the six PP pairings used in the construction of the scatter plots. As an example, Figure 5 plots the raw PP currents, $I_{01}$, along the *x*-axis against the corresponding $I_{11}$ on the *y*-axis for the 45 chips, measured without any Trojan emulation transistors enabled. These points are identified as *Trojan-free data points* in two separate data sets, labeled **Uncalibrated data** and **Calibrated data** (to be discussed). The following applies to either data set.

Since the individual ICs each have a unique leakage current associated with them, the Trojan-free data points are dispersed along the line labeled *regression line*. The regression line is actually derived from the Trojan-free data points and can be thought of as the 'best fit' line through them. The data points are not co-linear because of measurement noise and process variation effects, such as regional leakage current variations. Two parabolic curves, labeled *3*

σ *limits*, represent the prediction limits of the Trojan-free data points. The curves delimit a region in which 99.73% of the data points from Trojan-free chips are expected to fall. We use 3 σ for our limits because it is the most commonly used value in experiments carried out by industry, and is considered the industry standard.

The statistical limits are used to detect the Trojans. We consider a Trojan detected when its data point(s) falls outside these limit curves, i.e., above the top curve or below the bottom curve, **in at least one** of the scatter plot pairings (in our experiment, there are six pairings). For example, the data points labeled "Chip $C_1$, Trojan #4 at each TESM voltage" in Figure 5 are the ten data points produced under the Trojan emulation experiments for chip $C_1$. Each data point represents the currents measured at $PP_{01}$ and $PP_{11}$ under one of the applied TESM voltages. The same is true for the data points labeled "Chip $C_2$, Trojan #4 at each TESM voltage". In the *Uncalibrated* data set, none of the Trojan data points fall outside the limits and therefore these Trojans are considered undetected in this scatter plot pairing (if the same holds true in the other scatter plots then the Trojan escapes detection). The same is not true in the *Calibrated* data set however. The Trojans under larger voltage drops, e.g., TESM voltages 0.85 through 0.80, are detected in both chips.

### 4.2 Signal Calibration

The dispersion in the data points among the Trojan-free chips is caused primarily by chip-to-chip variations in the power grid resistance and the series resistance variations to the power ports from the power supply[1]. The differences in series resistances occur within the package and on-chip and as contact resistance variations in the clam-shell-style ZIF socket on the test board. In any case, these resistance variations adversely affect the sensitivity of our analysis to small Trojan current anomalies.

In previous work, we developed and demonstrated a process and environmental (PE) calibration technique [13][14] outlined here to deal with these chip-to-chip variations. The method makes use of the data collected from a special set of "calibration circuits" (CC), that are similar in design to the TCs shown in Figure 1(b) without the Trojan emulation transistor. Under the proposed signal calibration scheme, one CC is placed underneath each PP. Figure 1(a) shows the positions of the TCs, labeled $TC_{0,1}$, $TC_{0,77}$, etc., that are used as the CCs in our experiments. The calibration data is collected by enabling the shorting inverter in each of the TCs, one at a time, and measuring the PP and global currents. Leakage measurements are also made and subtracted from the shorting inverter currents. The shorting inverter currents are then normalized by dividing through by the global current.

The matrix of data collected under the calibration tests is used to calibrate the PP currents measured under subsequent Trojan tests. This is accomplished using the matrices obtained from a chip, $C_x$, and the data collected from calibration tests applied to a simulation model, *S*. The simulation model serves as the reference or 'golden chip' standard. For the chip and simulation model, the matrix is 4x4 in our experiments because the power grid has only 4 PPs. Equation (1) gives the expression for computing the transformation matrix, *X*. Once *X* is obtained, the four PP currents from $C_x$, measured using a test designed to detect Trojans, are calibrated using the linear transformation operator defined by Equation (2).

$$X = C_x^{-1} * S$$

$$\begin{bmatrix} x_{00} & x_{01} & x_{02} & x_{03} \\ x_{10} & x_{11} & x_{12} & x_{23} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \end{bmatrix} = inv\left(\begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{bmatrix}\right) \times \begin{bmatrix} r_{00} & r_{01} & r_{02} & r_{03} \\ r_{10} & r_{11} & r_{12} & r_{13} \\ r_{20} & r_{21} & r_{22} & r_{23} \\ r_{30} & r_{31} & r_{32} & r_{33} \end{bmatrix} \quad \textbf{Eq. 1.}$$

$$N_1 = I_1 * X$$

$$\begin{bmatrix} N_0 & N_1 & N_2 & N_3 \end{bmatrix} = \begin{bmatrix} I_0 & I_1 & I_2 & I_3 \end{bmatrix} \times \begin{bmatrix} x_{00} & x_{01} & x_{02} & x_{03} \\ x_{10} & x_{11} & x_{12} & x_{23} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \end{bmatrix} \quad \textbf{Eq. 2.}$$

The effect of the signal calibration operation is shown in Figure 5 in the **Calibrated** data set. Here, the Trojan-free data points are distributed more uniformly along the regression line, and the limits are 'tighter' when compared with the uncalibrated limits. More importantly, there is an **increase** in the displacement of the data points derived from the Trojan experiments away from the regression line in the calibrated data set, which illustrates that calibration *amplifies* Trojan current anomalies in the presence of PE variations.

## 5 The Impact of Leakage Current Patterns and Variations

Calibration is very effective in reducing the adverse impact of resistance variations in the power distribution system. However, regional (within-die) leakage current variations are not calibrated for using this process. If leakage variations are large, they can produce **false positive** Trojan detections. Since our limits are derived from Trojan-free chips (instead of simulation models), within-die leakage variations tend to add to the dispersion in the Trojan-free data points, which widens the limits and makes the regression test less sensitive to Trojan signal anomalies.

To determine the level of leakage current variations in our test chips, we ran a separate set of experiments in which we placed individual regions into the HL state and made global current measurements. Figure 6(a) shows a sequence of block level diagrams of the TC array. The diagram labeled 'LL', is a 'low-leakage-everywhere' configuration of the array that we use as a base pattern. The diagrams labeled '$HL_x$', where $x$ = 1 to 64, are configura-

1. A second, less significant source for the dispersion is leakage current variations, to be discussed.

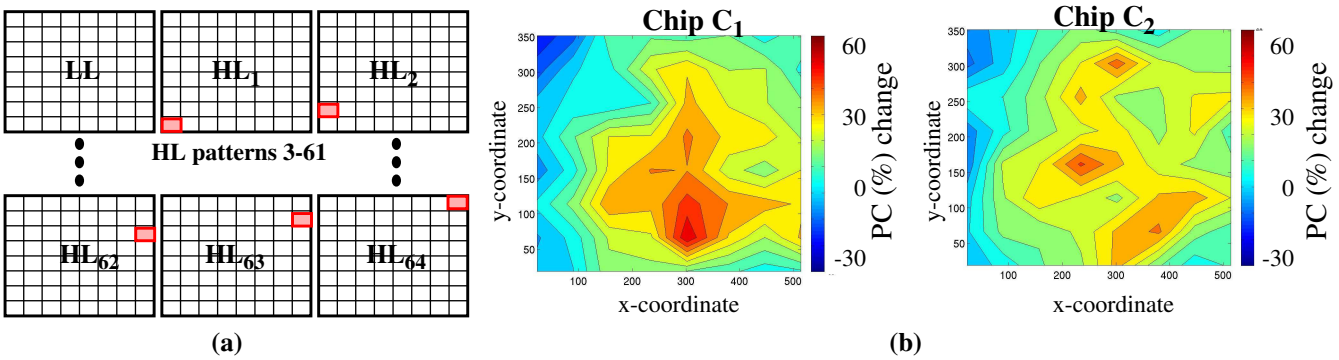



**Fig. 6. Analysis of within-die leakage current variation, (a) block diagrams (b) contour plots.**

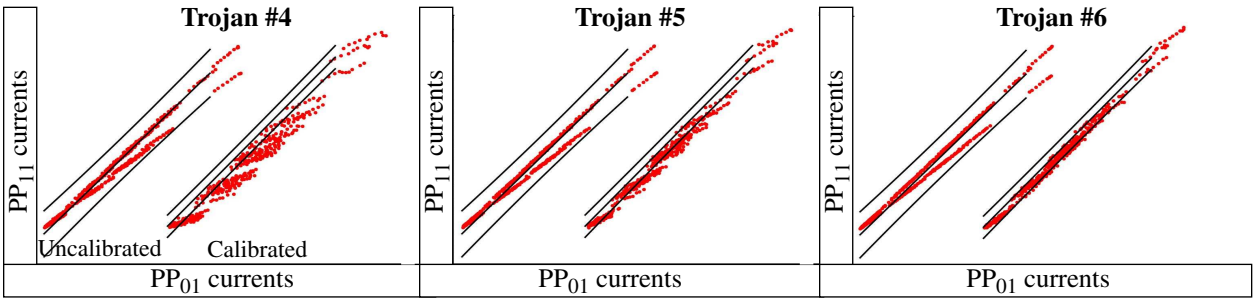



**Fig. 7. Regression analysis of Trojans #4 (left) through #6 (right) for all chips at all TESM voltages for $PP_{01}$-$PP_{11}$.**

tions that set the TCs within the shaded region in a HL state (all other TCs are configured in a LL state). By subtracting the global current measured under the LL configuration from each of the $HL_x$ configurations, we obtain the leakage characteristics of the P-channel devices within each of these 64 regions. Variations in their magnitude across the array correspond to within-die leakage current variations.

Contour plots depicting the variations in leakage for chips $C_1$ and $C_2$ are shown in Figure 6(b). The center (x,y) coordinates of each of the 64 regions are plotted in the (x,y) plane of the contour plots. The contours depict the regional leakage variation of each chip as a percentage change (PC), computed using Eq. 3 with respect to the global current measured under the $HL_1$ configuration. The darker shaded

$$PC_x = \frac{(HL_x - HL_1)}{HL_1} \times 100 \qquad \textbf{Eq. 3.}$$

contours identify regions with smaller leakage current than the $HL_1$ reference region while the lighter contours show regions with higher currents. The $PC_x$ values range from +60% to -30%.

Variations in leakage current add to the measurement noise and the corresponding dispersion in the Trojan-free data points. The calibration process described earlier subtracts leakage current from the shorting inverter current and therefore is not able to correct for these variations. We define the calibration process this way to avoid needing to carry out calibration for every Trojan test that is applied (each Trojan test defines a different core logic state and corresponding leakage variation pattern). However, we give an analysis in Section 6.2 that demonstrates a moderate improvement in Trojan sensitivity when leakage is not subtracted, which shows that calibration can, in fact, correct for leakage variations.

### 5.1 Sensitivity Impact of Leakage Patterns and Trojan Position in the Layout

Beyond process and environmental noise, there are two other factors that influence the sensitivity of our method to Trojan current anomalies, (1) the location of the Trojan in the layout with respect to the surrounding PPs and (2) the leakage state of the gates in the core logic. The first is illustrated in Figure 7 which depicts a sequence of scatter plots. The raw currents measured from power ports $PP_{01}$ and $PP_{11}$ are plotted along the x- and y-axis respectively in each of the plots. The regression limits are derived using only the data points from the first 20 chips (of the 45). The data points depicted in each of the scatter plots are derived from the Trojan experiments for Trojan #4 (left) through Trojan #6 (right) (the Trojan-free data points used to derive the limits are not shown). Each plot includes 450 data points which corresponds the product of 45 chips and 10 TESM voltages. Both the uncalibrated and calibrated data sets are shown. The analysis below focuses on the calibrated data sets.

First, all of the Trojan-free control samples fall within the limits (this also holds true for the remaining 5 PP pairings not shown). Therefore, there are no **false positive Trojan detections** in the analysis[1]. Second, the number of

1. We analyze false positives in Section 6.4.

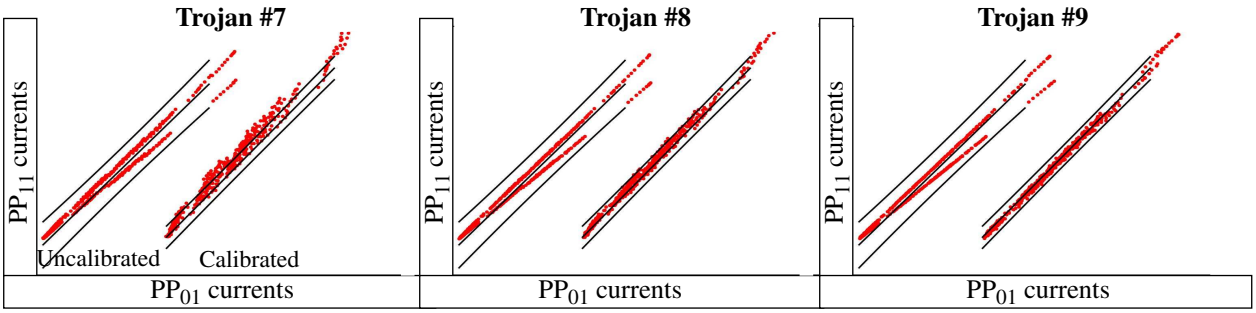



**Fig. 8. Regression analysis of Trojans #7 (left) through #9 (right) for all chips and TESM voltages for $PP_{01}$-$PP_{11}$.**

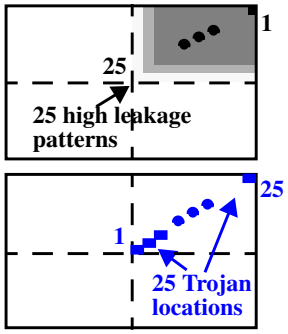



**Fig. 9. High leakage patterns (top), Trojan locations (bottom).**

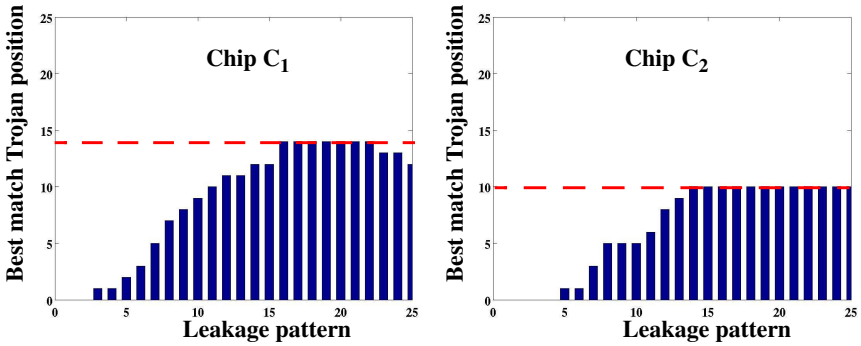



**Fig. 10. Leakage-current-center analysis for two chips. Bar height identifies high leakage pattern that represents best match to Trojan location.**

detected Trojans is largest for Trojan #4 (left-most scatter plot) when compared with the scatter plots for Trojans #5 and #6. This is evident from the larger number of data points that fall outside the limits. The reason this is true is related to the layout position of Trojan #4 as shown in Figure 3. Since Trojan #4 is closer to $PP_{01}$ than Trojans #5 and #6, the fraction of the Trojan current sourced from $PP_{01}$ is larger, at the expense of the fraction sourced from $PP_{11}$. The relative increase in Trojan current in $PP_{01}$ is reflected in the 'lower-right-pointing' direction in the dispersion of the data points. The larger relative current in $PP_{01}$ (which is plotted along the x-axis) 'draws' the data points away from the regression line because the Trojan current sourced from $PP_{11}$ is only slightly larger than it is for the Trojan-free case. The larger the Trojan current anomaly, the larger the excursion of the Trojan data points from the regression line.

In contrast, Trojan #6 distributes current almost uniformly to $PP_{01}$ and $PP_{11}$ because its layout position is closer to the midpoint between the PPs. This nearly equal distribution makes it more difficult to detect the anomaly. An opposite, but similar trend occurs in the scatter plots for Trojans #7 through #9 as shown in Figure 8. Given that these Trojans are closer to $PP_{11}$, the direction of dispersion in the Trojan data points is to the upper left to reflect the relative increase of the Trojan current in $PP_{11}$ with respect to $PP_{01}$.

A subtle difference exists across the corresponding scatter plots of Figures 7 and 8 however, that is most clearly observed for Trojans #6 and #9. Given that the layout positions of these Trojans are mirror images of each other along the y axis, we would expect the scatter plots for these two Trojans to be mirror images as well, but this is not the case. The number of instances of Trojan data points that fall outside the limits is larger for Trojan #6 than it is for Trojan #9. The reason this is true is related to the asymmetry in the underlying leakage pattern introduced into the array as shown in Figure 3, and to a smaller degree, the difference in the magnitude of the leakage current in each of these regions. In reference to the asymmetry, the layout position of Trojan #9 is closer to $PP_{11}$ and therefore, sinks more of its current from $PP_{11}$. But this distribution pattern is very similar to the distribution of current introduced by the leakage pattern, and therefore the Trojan current anomaly is indistinguishable from the Trojan-free chips under this leakage pattern. We refer to this position as the 'leakage-current-center' for this leakage pattern. With regard to the second condition, the larger magnitude of the leakage current in the region surrounding Trojan #9 decreases the signal-to-noise ratio of the Trojan anomaly and makes it more difficult to detect.

### 5.2 Leakage-Current-Center Analysis

To better understand the relationship between a leakage pattern and its leakage-current-center, we ran a sequence of experiments on two chips. For each chip, we configured a sequence of high leakage regions in the upper left quadrant of the test chip array, labeled 1 through 25 in the top block level diagram in Figure 9. $HL_{25}$ configures all 1,000 TCs in this quad into the HL state. Each of the remaining HL patterns sets incrementally fewer rows and

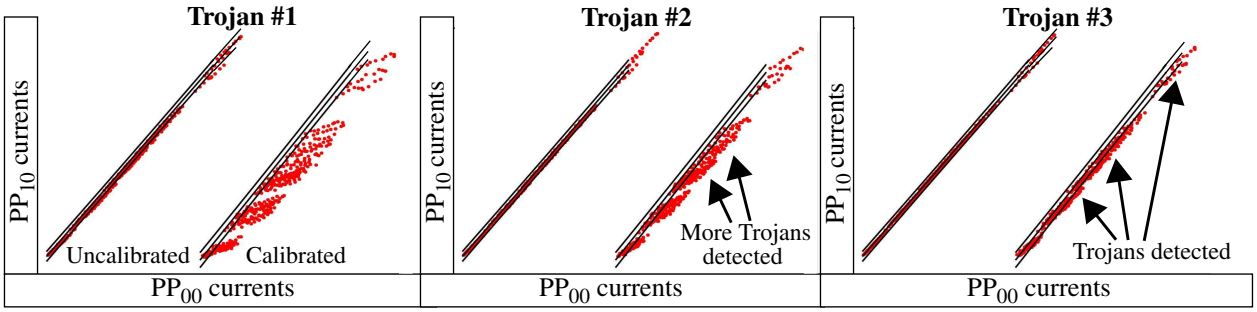



**Fig. 11. Regression analysis of Trojans #1 (left) through #3 (right) for all chips at all TESM voltages for $PP_{00}$-$PP_{10}$.**

columns in the HL state, eliminating rows and columns on the left and bottom of the previously numbered pattern. $HL_1$ sets only a single TC in the HL state, the TC under $PP_{11}$, and therefore is very similar to the LL pattern.

In a separate set of 25 experiments, we configured the entire array into a LL state and enabled the Trojan emulation transistors, one at a time, at a series of locations along the diagonal from the center of the TC array to the TC under $PP_{11}$, as shown in the bottom block level diagram of Figure 9. The four PP currents were measured in each HL and Trojan emulation experiment.

The objective of this analysis is to determine, for each leakage configuration, the Trojan position that produces a set of PP currents that are the **closest match** to the PP currents measured under the leakage configuration. The position of the Trojan in the layout producing the best match represents the leakage-current-center for that leakage pattern. The closest match is determined by treating the 4 PP currents as a point in 4-D space and finding the smallest Euclidean distance between the set of Trojan positions and a given leakage configuration. This process is repeated for each of the leakage patterns.

The bar graphs shown in Figure 10 give the results for chips $C_1$ and $C_2$. The best matching emulated Trojan position is given on the y-axis for each of the leakage patterns on the x-axis. From the bar graphs, it is clear that the best match for $HL_1$, which enables only a single TC in HL, is Trojan 1, which is centrally located in the grid. This is the intuitive result because this Trojan position distributes current uniformly to the PPs and is similar to the $HL_1$ (and LL) pattern, which has a leakage-current-center in the geometric center of the TC array.

As the HL region grows larger around $PP_{11}$ for leakage patterns greater than 1, the index of the best matching Trojan also grows larger. The leakage-current-center continues to move for patterns up through 16 for chip $C_1$ and pattern 14 for chip $C_2$, at which point it stops and remains constant (and even recedes for $C_1$).

Although the movement of the leakage-current-center in the two bar graphs is similar, there are differences that are beneficial to the detection of Trojans in these regions. First, the actual leakage-current-center for any given leakage pattern is different for each chip in most cases. For example, the largest Trojan position is 14 for chip $C_1$ while it is 10 for chip $C_2$. This suggests that the probability of detecting Trojans in these 'hidden' areas is larger for more diverse chip populations.

The high degree of control over leakage current distribution in the TC array makes it possible to use different leakage patterns to eliminate 'hidden' regions. For example, although we do not show it, the poor detectability of Trojan #9 can be resolved by using the complimentary leakage pattern, with $Q_0$ configured in HL and $Q_3$ configured in LL. Given the symmetry of the power grid, this new configuration would generate results similar to those shown in Figure 11 for Trojans #1 through #3 using power port pairing $PP_{00}$-$PP_{10}$. Trojan #9 is not detected under any TESM voltage in Figure 8, where, in contrast, more than 25% of the Trojan data points that fall outside the limits for Trojan #3 in Figure 11. The difference between Trojan #8 in Figure 8 and Trojan #2 in Figure 11 is even more pronounced.

Unfortunately, the leakage characteristics in the core logic of commercial designs is not as easily controlled as it is in our TC array. This is true because the multi-level nature of logic gates in commercial designs creates situations in which configuring a HL state for one gate using ATPG can produce a LL state in the successor gate. Also, it is computationally difficult to determine patterns that create leakage distributions with widely varying characteristics. However, the fact that leakage current distribution characteristics are state dependent can be leveraged, particularly for detecting Trojans in the leakage-current-center as described above.

Beyond manipulating leakage, larger power grids provide an alternative means of dealing with the leakage-current-center problem. Larger power grids will have more power ports, which are inserted by designers to maintain a stable voltage across the 2-D plane of the chip. The additional power ports can be used to eliminate the leakage-current-center of a region altogether. This is accomplished by using power port pairings that include one of the power ports in the region surrounding the Trojan, e.g., $PP_{11}$ and one that belongs to an adjacent region, e.g., $PP_{21}$ ($PP_{21}$ does not exist in our power grid but would be to the right of $PP_{11}$ in a larger grid). However, the most sensitive power port pairings are those that are closest to the Trojan, i.e., both located in the region surrounding the Trojan, and therefore manipulating state leakage through different test

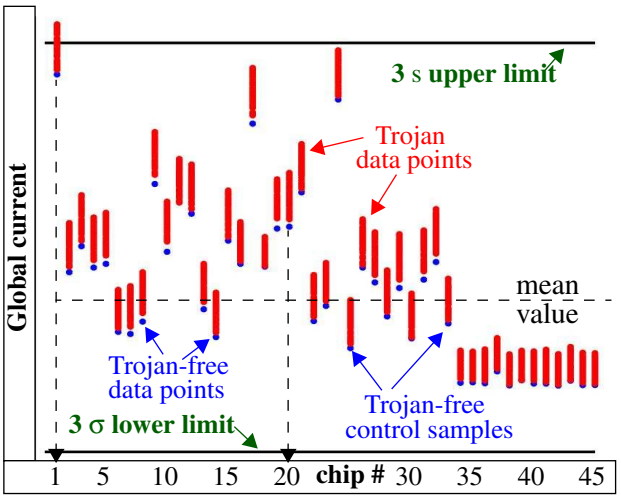



**Fig. 12. Global current 1-D statistical analysis with chips on the x-axis and global currents on the y-axis.**

vectors is likely to be the most effective way to deal with the leakage-current-center problem.

## 6 Experimental Results

### 6.1 Global Current Analysis

In order to determine the improvement of our strategy over conventional power supply methods, we first carry out a global current analysis. Figure 12 plots chip number (x-axis) against the measured global current (y-axis) from the chips under both the Trojan-free and Trojan experiments. The 3 σ upper and lower limits are derived using the data point from the first 20 Trojan-free data points while chip numbers 21 and above serve as **control samples**. The purpose of the control samples is to determine the potential for false positives. By excluding these Trojan-free samples from the data used to compute the limits, they are evaluated against the limits in the same way as the Trojan data points.

The Trojan data points are shown above the single Trojan-free data point for each chip. With 9 Trojan locations and 10 TESM voltages, there are 90 Trojan data points per chip, many of which are superimposed above the Trojan-free data point of each chip. The upward displacement of these data points occurs because the Trojan always adds to the existing Trojan-free leakage current. From the figure, the Trojan data points fall within the limits for 44 of the 45 chips, i.e., only chip $C_1$ has Trojan data points above the 3 σ limit. The large Trojan-free 'base' leakage current associated with this chip allows Trojans with larger currents (lower TESM voltages) to be detected.

### 6.2 Regression Analysis

As indicated earlier, a Trojan is counted as 'detected' in our regression analysis procedure if **at least one** of its data points falls outside the limits in the 6 scatter plots constructed from the 4 power port pairings (see Figure 4). We use the term **false positive** for Trojan-free data points that fall outside the limits. The term *limit-setting* is used in reference to the Trojan-free chips used to establish the statistical limits, while *control* samples is used for the remaining Trojan-free chips.

The results of applying regression analysis to the Trojan-free and Trojan data collected from our 45 chips is shown in Table 2. The first row indicates that the number of chips producing false positives is 0 under each of the three analyses, namely, the 1-D global current analysis (column 2) and regression analyses using uncalibrated (column 3) and calibrated data (column 4).

| Total # of detections possible is 4,050 | Global Current | Uncalibrated Regression | Calibrated Regression |
|---|---|---|---|
| False Positives | 0 | 0 | 0 |
| Trojans detected (w/o leakage calibration) | 45 (1.1%) | 291 (7.2%) | 2185 (54.0%) |
| Trojans detected (with leakage calibration) | | | 2306 (56.9%) |

**Table 2: Total number of Trojan detections under global current and regression analysis.**

The second row of Table 2 gives the number and percentage of the Trojans that are detected. The total number of Trojans is 4,050, which is derived from the product of the number of chips (45), the number of Trojans (9) and the number of TESM voltages (10). From the data in columns 2 and 3, Trojan detection sensitivity increases by a factor of **6.5**, computed as (7.2%)/(1.1%)*100, just by going to a MSP-based technique. Bear in mind that the number of global current detections is likely to decrease rapidly toward 0 as the chip size increases because the method does not scale. On the other hand, the results obtained for MSP are likely to be similar for larger chips because the additional PPs available allows MSP to scale with chip size.

When calibration is used, another factor of **7.5** is obtained in the number of detections over the uncalibrated case. Overall, MSP and calibration improve sensitivity by a factor greater than **49** over global current analysis. With larger chips, this factor is likely to increase to between 100 and 1000 times. The value in the column 4 on the last row of the table gives the results using the 'leakage' calibration method described in Section 5. As we indicated earlier, the improvement in Trojan detection sensitivity is modest and is probably not worth the extra time and effort to carry out calibration on a vector-by-vector basis.

As a visual aid for understanding the impact of calibration, Figure 13 shows regression analysis for all 6 of the PP pairings and all Trojans, with uncalibrated data on the left and calibrated data on the right. The clusters of Trojan data associated with each PP pairing have been offset in the two dimension plane to a position that corresponds to the physical location of the PP pairing as given in Figure 4. The Trojan-free data points are not shown to help with the visualization. The larger dispersion around the regression limits for the calibrated data illustrates the level by which calibration amplifies the Trojan current anomalies.

### 6.3 Trojan "Hit" Analysis

Unlike the manufacturing test whose objective is to

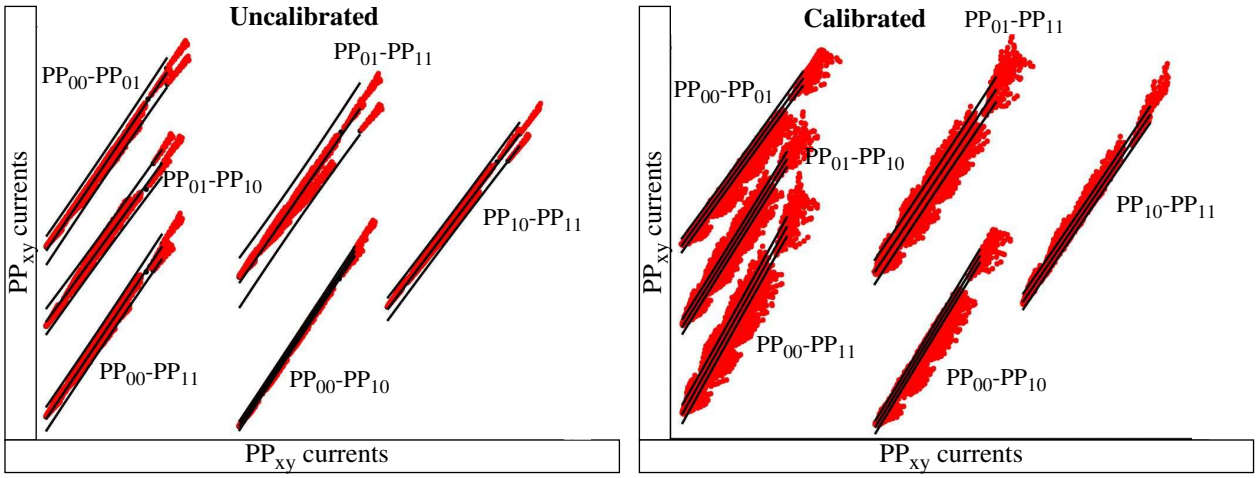



**Fig. 13. Regression analysis of uncalibrated (left) and calibrated data (right) of all Trojans, PP pairings and TESM voltages using 45 chips.**

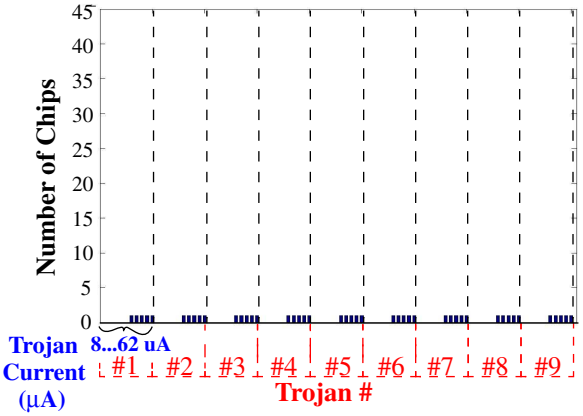



**Fig. 14. Number of chips in which each Trojan is detected using global current analysis.**

identify *every* defective chip, the objective of Trojan detection is to find at least one chip that contains a Trojan. Although finding one is sufficient, given the non-zero probability that any given chip may have a manufacturing defect (defects also produce current anomalies) suggests that the appropriate metric is to determine if a suspect anomaly is present in a larger group of chips. The chance that an entire group of chips is defective in the same way is very small because most defects are random in nature[1]. Therefore, the level of confidence that a measured anomaly is caused by a Trojan increases if the same pattern exists in more than one chip.

The histogram in Figure 14 shows the number of chips in which a Trojan is detected using global current analysis. The x-axis consists of 9 clusters of bars that correspond to each of the Trojans. Within each cluster are 10 bars that correspond to the TESM voltages used in the analysis, labeled with the range of mean currents sourced by the Trojan (see Table 1). For example, the smallest Trojan current, 8 μA, corresponds to the 0.89 TESM voltage while the largest Trojan current, 62 μA corresponds to the 0.80 TESM V. The y-axis plots the number of chips in which each Trojan is detected.

The 1-D statistical analysis shown in Figure 12 indicates that it is possible to detect about half of the Trojans in chip 1. Figure 14 reflects this by showing that about half of the bars have a height of 1. Overall, global current analysis performs poorly with regard to our metric which relates confidence in a positive Trojan detection to the height of the bars.

In contrast, the histograms shown in Figure 15 for regression portray higher levels of confidence in positive Trojan detections. The histogram results using **uncalibrated** data are shown in Figure 15(a), while the **calibrated** versions are shown in Figure 15(b). For example, in the uncalibrated regression results, Trojan #1 is detected in approx. 35 chips at the higher TESM voltages, while Trojans #7 is not detected in any chip under about half of the TESM voltages.

The calibrated data results shown in Figure 15(b) provide much higher levels of confidence that the chips have Trojans. The larger height of the bars indicate that most Trojans are detected more often in comparison with the uncalibrated data, yielding higher confidence, and the large number of *non-zero* bars indicates that most Trojans are detected in at least some chips (except for Trojan #9 which is missed in all chips at all TESM voltages). Moreover, except for Trojan #9, the other instances of Trojans that are missed in every chip (zero height bars) occur only at TESM voltage 0.89, which produces the smallest Trojan current. Overall, only 14 Trojans of the 4,050 Trojans are missed in every chip. As discussed in Section 5.2, the height of the bars are smaller for Trojan #8 and zero for Trojan #9 because these Trojans are near or at the *leakage-current-*

1. There are also systematic types of defects that can produce similar anomalies to the pattern expected for Trojans. The simultaneous presence of systematic defects may increase the level of false alarms.

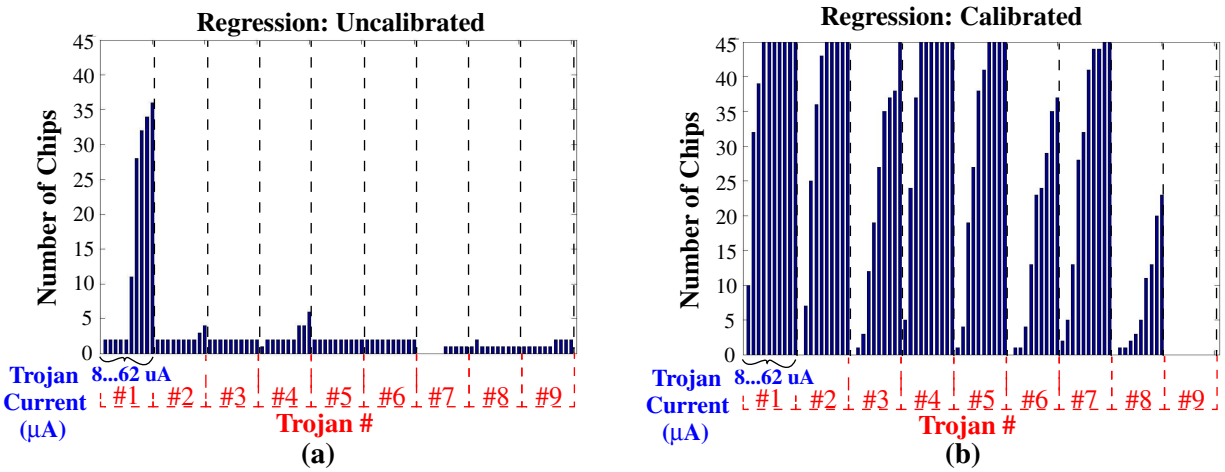



**Fig. 15. Number of chips in which each Trojan is detected using regression analysis on uncalibrated data (a) and calibrated data (b).**

*center* of the leakage pattern, respectively.

### 6.4 False Positive Analysis

Any type of statistical technique is subject to error based on the selection of samples used to construct the limits. In the analysis presented, we built the limits from the leakage currents of the first 20 chips, and, as shown in Table 2, this particular selection of chips established limits such that none of the data points in any of the 6 scatter plots from the entire set of 45 chips, i.e., the 20 limit-setting chips + the 25 control chips, fell outside of the limits. However, this need not be the case for other sets of limit-setting chips.

In order to determine the sensitivity of our method to false positives, we repeated the analysis described above 100 times using different sets of 20 limit-setting chips. In order to maintain some diversity in the population of chips, we first partitioned the 45 chips into 2 nearly equal subsets based on their overall leakage characteristics. Chips with leakages above the mean value (see Figure 12) were placed in one group, while those below the mean were placed into the second group. A pseudo-random number generator was then used to construct 100 sets of 20 limit-setting chips, 10 from each group. The remaining 25 chips that were not selected were used as control samples. Regression analysis was carried out using each of these limit-setting chip sets.

The number of false positives produced in these experiments is given in Table 3. The first column reports 47 as the number of experiments that produced at least one false positive across the 100 analyses. Therefore, over half of the analyses (53) produced no false positives. Of the 47 that do, the second column reports the number of chips that produce a false positive in at least one of the scatter plots. Although false positives are theoretically possible among the limit-setting chips, all of the chips producing false positives (70 of them) come from the control group. The value of 2,500 in this column is derived from the product of 25 control chips * 100 analyses. The third column gives the number of actual data points that fall outside the limits as 104. Therefore, from the numbers in columns 2 and 3, it can be deduced that only one scatter plot (of the 6) has a false positive data point in most of the experiments. The value 15,000 is derived from 2,500 (defined above) * 6 scatter plots. Although not shown, all of the false positive data points are close to the regression limits and therefore, they can be easily eliminated by ensuring sufficient diversity exists among the chips used to set the limits. This can be accomplished in a straightforward way by selecting the limit-setting chip set such that the full range of overall chip leakage characteristics is represented.

| # of analyses with at least one chip producing false positives | # of chips that produce false positives | # of false positive data points |
|---|---|---|
| 47 of 100 | 70 of 2,500 | 104 of 15,000 |

**Table 3: False Positive Analysis Results.**

## 7 Conclusions

In this paper, we carried out hardware experiments in which Trojans are emulated in a set of 45 chips fabricated in a 65 nm technology. A multiple supply port technique, in combination with a power signal calibration technique, are shown to increase detection sensitivity dramatically (by a factor of at least 49) over a global power signal analysis method. Given that large commercial grids incorporate hundreds (sometimes thousands) of power ports, we expect that enhancements in sensitivity to Trojans could exceed three orders of magnitude when such techniques are applied in practice.

We emulated Trojans that sink as little as 8 μA of current. Detecting such small current anomalies is not possible using conventional global power signal analysis methods. By using additional test patterns that control background leakage currents in different ways, we believe that our methods are capable of detecting all emulated Trojans investigated in our experiments.

## 8 Acknowledgements

This work was supported in part by NSF grant CNS-0716559.

## References


[1] http://www.acq.osd.mil/dsb/reports/2005-02-HPMS_Report_Final.pdf

[2] http://www.darpa.mil/mto/solicitations/baa07-24/index.html

[3] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, B. Sunar, "Trojan Detection using IC Fingerprinting", Symposium on Security and Privacy, 2007, pp. 296 - 310.

[4] R. Rad, J. Plusquellic, M. Tehranipoor, "Sensitivity Analysis to Hardware Trojans using Power Supply Transient Signals", International Workshop on Hardware-Oriented Security and Trust, 2008, pp. 3-7.

[5] F. Wolff, C. Papachristou, S. Bhunia, and R. Chakraborty, "Towards Trojan-Free Trusted ICs: Problem Analysis and Detection Scheme", Design, Automation and Test in Europe, 2008, pp. 1362-1365.

[6] Jie Li and John Lach, "At-Speed Delay Characterization for IC Authentication and Trojan Horse Detection", International Workshop on Hardware-Oriented Security and Trust, 2008, pp. 8-14.

[7] M. Banga and M. S. Hsiao, "A Region Based Approach for the Identification of Hardware Trojans", International Workshop on Hardware-Oriented Security and Trust, 2008, pp. 40-47.

[8] R. S. Chakraborty, S. Paul and S. Bhunia, "On-Demand Transparency for Improving Hardware Trojan Detectability", International Workshop on Hardware-Oriented Security and Trust, 2008, pp. 48-50.

[9] Y. Jin and Y. Makris, "Hardware Trojan Detection Using Path Delay Fingerprints", International Workshop on Hardware-Oriented Security and Trust, 2008, pp. 51-57.

[10] D. Acharyya and J. Plusquellic, "Hardware Results Demonstrating Defect Detection Using Power Supply Signal Measurements", VLSI Test Symposium, 2005, pp. 433-438.

[11] J. Plusquellic, D. Acharyya, A. Singh, M. Tehranipoor and C. Patel, "Quiescent Signal Analysis: a Multiple Supply Pad IDDQ Method," IEEE Design and Test of Computers, vol. 23, no. 4, pp. 278-293, 2006.

[12] X. Wang, M. Tehranipoor and J. Plusquellic, "Detecting Malicious Inclusions in Secure Hardware: Challenges and Solutions", International Workshop on Hardware-Oriented Security and Trust, 2008, pp. 15-19.

[13] D. Acharyya, J. Plusquellic, "Calibrating Power Supply Signal Measurements for Process and Probe Card Variations", IEEE International Workshop on Current and Defect Based Testing, 2004, pp. 23 - 30.

[14] R. M. Rad, X. Wang, M. Tehranipoor, J. Plusquellic, "Power supply signal calibration techniques for improving detection resolution to hardware Trojans", IEEE/ACM International Conference on Computer-Aided Design, 2008, pp. 632 - 639.


## Author Biographies

Jim Aarestad is conducting VLSI and hardware security research as a Master's degree graduate student at the University of New Mexico and holds a Bachelor's Degree in Computer Engineering from UNM as well. He has been a student member of the IEEE for five years, and served for one year as the Treasurer of the UNM Chapter of IEEE. His previous work includes ten years with Intel Corporation in semiconductor manufacturing operations. Prior to joining Intel, Jim designed instrumentation systems for research projects at the University of North Dakota. He also holds an Associate's Degree in Applied Science from the North Dakota State College of Science.

Dr. Dhruva Acharyya holds a B.S. degree in Physics, Astronomy and Mathematics from Drake University, Des Moines, IA and a M.S and Ph.D degree in Computer Engineering from University of Maryland, Baltimore County, MD. His research interests include VLSI Testing, Hardware Security, Design for Manufacturability and Test Structure Design. He is currently employed at Verigy Inc., a premier semiconductor test solutions provider and automated test equipment manufacturer. Prior to joining Verigy, Dr. Acharyya was a researcher at IBM Austin Research.

Reza Rad received his Ph.D. degree in computer engineering from University of Maryland Baltimore County in 2008 and M.S. and bachelor degrees from University of Tehran in 2001 and 1998. From 2008 to 2010 he served as a faculty member of University of Maryland Baltimore County. Since 2010 he has joined Maxlinear Inc. as an ASIC design and test engineer. His research interests are in hardware authentication and security and defect based testing. He is also interested in design and implementation of novel hardware accelerator architectures for signal and image processing applications on reconfigurable devices. Dr. Rad has served as member of technical program committee of IEEE Symposium of Hardware Oriented Security and Trust (HOST) in 2010 and is a member of IEEE.

Professor Plusquellic received both his M.S. and Ph.D. degrees in Computer Science from the University of Pittsburgh in 1995 and 1997, respectively. He is currently an Associate Professor in Electrical and Computer Engineering at the University of New Mexico. His research interests are in the area of nano-scale VLSI and include authentication of IC hardware, silicon validation, design for manufacturability and delay test methods. Professor Plusquellic is on the program committee of the International Test Conference (ITC). He has held multiple positions on the organizing committee for the Defect-Based Testing Workshop including the General Chair position in 2006. He is the General Chair for the International Symposium on Hardware-Oriented Security and Trust (HOST-2010). Professor Plusquellic received the "5 Years of Continuous Service Award" from ITC, a Best Paper Award from VTS, an ACM Distinguished Service Award from SIGDA and two Austin CAS Fellow Awards from IBM. He is a member of the IEEE.