Trojan Scanner: Detecting Hardware Trojans with Rapid SEM Imaging combined with Image Processing and Machine Learning

Nidish Vashistha, Hangwei Lu, Qihang Shi, M Tanjidur Rahman, Haoting Shen, Damon L Woodard, Navid Asadizanjani and Mark Tehranipoor
Florida Institute for Cybersecurity (FICS) Research
ECE Department, University of Florida
{nidish, tehranipoor}@ufl.edu

Abstract

Hardware Trojans are malicious changes to the design of integrated circuits (ICs) at different stages of the design and fabrication processes. Different approaches have been developed to detect Trojans namely non-destructive (electrical tests like run-time monitoring, functional and structural tests) and destructive (full chip reverse engineering). However, these methods cannot detect all types of Trojans and they suffer from a number of disadvantages such as slow speed of detection and lack of confidence in detecting all types of Trojans. Majority of hardware Trojans implemented in an IC will leave a footprint at the doping (active) layer. In this paper, we introduce a new version of our previously developed "Trojan Scanner" [1] framework for the untrusted foundry threat model, where a trusted GDSII layout (golden layout) is available. Advanced computer vision algorithms in combination with the supervised machine-learning model are used to classify different features of the golden layout and SEM images from an IC under authentication, as a unique descriptor for each type of gates. These descriptors are compared with each other to detect any subtle changes on the active region, which can raise the flag for the existence of a potential hardware Trojan. The descriptors can differentiate variation due to fabrication process, defects, and common SEM image distortions to rule out the possibility of false detection. Our results demonstrate that Trojan Scanner is more reliable than electrical testing and faster than full chip reverse engineering. Trojan Scanner does not rely on the functionality of the circuit rather focuses on the real physical structure to detect malicious changes inserted by the untrusted foundry.

1. Introduction

Outsourcing integrated circuit (IC) design, fabrication, and test facilities have reduced the costs and time to market. Building and maintaining an advanced technology node foundry can cost up to several billion of dollars [2]. Hence, most of the design companies have become fabless or they have migrated their fabrication team offshore [3], on which they have to rely for fabrication. Outsourcing semiconductors fabrication also brings in trust issues between design house and foundry, because the latter has full access to all the design details including GDSII layout and test vectors (test inputs and test responses). As a result, this trust issue has opened up an avenue to various types of threats in ICs including hardware Trojan insertion, overproduction, IP piracy, and out-of-specification/defective ICs appearing in the market [4].

Among all of these trust issues, hardware Trojans are the most dangerous threat as they can compromise the security and trustworthiness of a system and they are very difficult to detect due to their stealthy nature [5]. Hardware Trojan is a malicious modification to the circuit during any phase of design, integration or fabrication [6]. Using hardware Trojan an adversary can cause a denial of service, control (integrity violation) or leak (confidentiality violations) sensitive information from the system. The hardware Trojans can be a major threat to all electronic devices, civilian applications and most importantly military and space systems. There have been instances reported where a system's security was compromised because of suspected "back-doors" [7] [8].

Hardware Trojans can be inserted during any step of the IC design process due to the involvement of untrusted entities [9]. The classification of different kinds of Trojan insertion scenarios into attack models is essential to understand the origin of the hardware Trojans and hence to develop detection techniques and countermeasures based on the model. For the sake of simplicity, one could assume that IC design involves three main entities, namely (1) third-party intellectual property (3PIP) vendors who provide functional blocks i.e., IP blocks, (2) system-on-chip (SoC) developers who develop the architectural platform for a design and (3) the last entity is the foundry that fabricates the ICs. There can be different kinds of attack models based on the trust assumption with any of these entities [10]. Among them, the threat model of the untrusted foundry has been widely discussed in the hardware security community [10]. In this model (Fig. 1), the foundry is the only untrusted entity and perceived as a threat for malicious hardware insertion during fabrication. A Trojan can be inserted in unused spaces on the chip or by moving the cells in the layout to create space for inserting Trojan. Further, Trojan can replace a de-coupling MOS capacitor or existing filler cells. It can also be created by re-sizing the existing cells or by thinning, interconnects which can cause an early failure (i.e. denial of service attack). Such reliability hardware Trojans are the only one with no physical footprint on the doping layer, among all types of Trojans (categorized based on their physical, activation, and action characteristics). Fortunately, they can be easily detected by other means including accelerated wear out /aging tests [11] or transient/quiescent current test methods (IDDT/IDDQ) [12]. This paper is focusing only on the hardware Trojans with footprint changes at the active layer and only concerned with the untrusted foundry model, where a golden layout is assumed to be available.

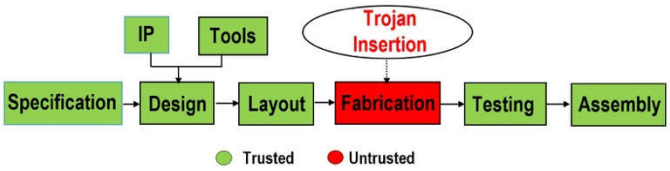

***Figure 1**. Untrusted foundry attack model.*

There has been extensive research on detecting hardware Trojans using run-time monitoring [13] and logic test approaches [4] [14] [15] [16] [17] [18]. However, such detection techniques have a serious number of limitations. For example, the run-time monitoring techniques increase resource utilization on IC by using on-chip sensors or structures to detect malicious activities. Such techniques consume extra CPU usage, power, memory and silicon area on the chip while the confidence in detection is quite low especially when Trojan is activated under rare condition. While test time methods like logic testing cannot easily detect large Trojans as it is difficult to generate test vectors for triggering them, side channel signal analyses approach needs a golden IC and it is vulnerable to circuit noise and process variations hence they cannot detect small size Trojans [19]. As a result, the confidence level in detecting Trojans using the above-mentioned techniques are quite low. Another approach discussed in the community is a destructive method, where the full-blown reverse engineering of the IC must be performed [20] [21]; in this case, a chip is fully reverse engineered [22] to reconstruct the circuit netlist. This approach is quite expensive, slow, and requires more than tens of ICs to successfully reverse engineer the chip. The sample preparation and delayering process are very sensitive and destructive; therefore, many samples are sacrificed before the right recipes for delayering are prepared. However, we believe that the reverse engineering of a chip for trust verification is the most effective one for the untrusted foundry threat model.

Over the past decade, not much attention was paid to Trojan detection using physical inspection techniques due to the high cost of advanced microscopy machines. However, with the advancements in the microscopy, these instruments are more prevalent in labs and available to the public, and easier to rent by hours or purchase today. Courbon et. al [23] [24] proposed the basic concept of image processing to detect Trojans using SEM images. They used front side SEM imaging and basic image processing functions like histogram equalization and image subtraction to detect Trojan insertion in the form of logic gates and transistors. They covered addition of logic gates or transistors as a Trojan insertion approach. Bao e.t. al [25] proposed a machine learning based technique to hardware Trojans. Their approach detects the changes in metal layers in IC but does not cover the detection of Trojans implemented by modifying doping regions. Using the backside approach, Zhou e.t. al [26] have used infrared based optical imaging to detect a Trojan implemented by replacement or re-routing of the standard cells. This approach would miss the small Trojans or minor changes at the active region because the resolution of infrared optical imaging is not sufficient to detect changes at the nanometer scale.

In our previous work [1], we have introduced a new hardware Trojan detection technique called Trojan Scanner, which uses SEM images of a golden chip as a reference to detect Trojan; however, a golden chip may not be always available for comparison. In this paper, we propose a new version of Trojan Scanner to detect hardware Trojans by comparing a golden layout with the SEM image taken from the backside of an IC under authentication (IUA). Here, we make the following contributions:

- The new Trojan Scanner framework is based on SEM imaging of a backside thinned IC. Compared to the front side approach, the sample preparation is easier for backside imaging, as it does not require complicated layer-by-layer wet/dry etching processes for removing heterogeneous layers i.e. metal, silicon oxide and polysilicon layers.
- Using supervised machine learning and image processing, Trojan Scanner identifies unique logic cells as a descriptor from the layout and SEM image of an IUA. As mentioned earlier, the majority of Trojans inserted by an untrusted foundry will have to make some modification (even minor) to the doping layer. These changes can be easily detected from backside without the need to reconstruct netlist or understand the functionality of IUA. During IC authentication, this method does not need involvement of an engineer for whole die imaging, image enhancement, comparison, and decision-making. Hence, the entire detection process can be automated.
- The descriptors consider doping area variations due to process variation, defects or SEM imaging noise to detect differences between a golden IC layout and IUA SEM image descriptor with high confidence.
- Since it is not expected for a user to have access to golden IC, the new Trojan Scanner technique addresses the shortcoming of the previous one, as it does rely on the availability of a golden chip.
- Imaging is an important step of the physical inspection and high dwelling time imaging usually takes a long time. As a result, Trojan Scanner takes low dwelling time and resolution images to reduce detection time significantly as supported by our results.

The remainder of this paper is organized as follows: Section 2 introduces the hardware Trojan taxonomy in detail. Section 3 introduces our new version of detection technique called Trojan Scanner 2.0. Next, we discuss the detection rate and results of Trojan detection in Section 4. Finally, we conclude this paper with our findings in Section 5 with a brief discussion about future work.

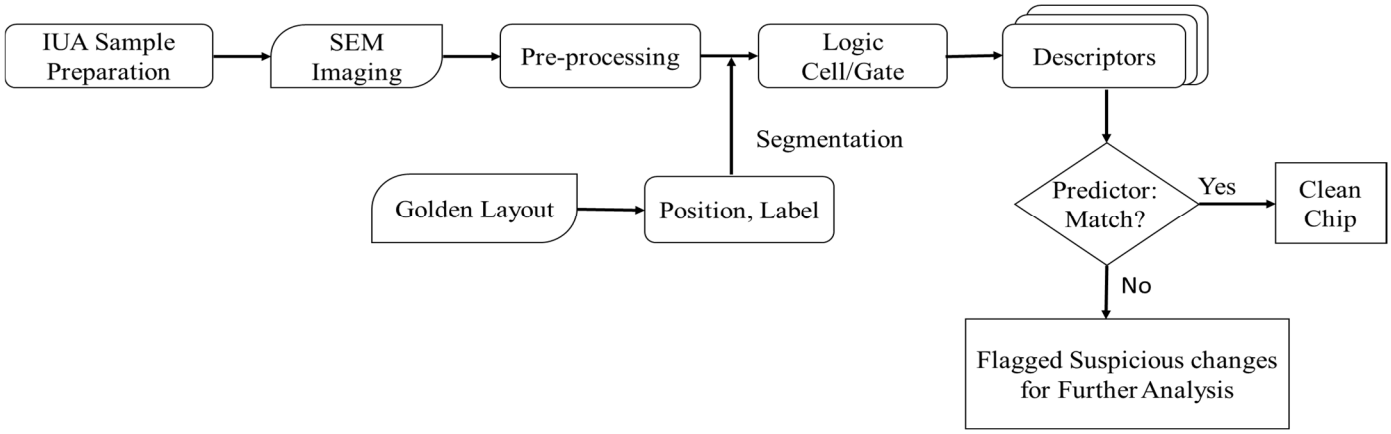

Figure 2. *Trojan Scanner 2.0 process flow.*

2. Trojan Taxonomy

Hardware Trojans can be classified based on their physical, activation and action characteristics [6].

1. *Physical Characteristics:* Hardware Trojans can be classified based on the type of geometrical modifications in the chip layout. It can be further subcategorized based on their (i) *Functional* category which includes Trojans that are implemented by addition or deletion of transistors (logic gates) and (ii) *Parametric* category which includes the modification of existing interconnects, via, or logic inputs. For example, thinning or widening of interconnects (critical path like power, ground line or a clock tree) [27].

2. *Activation Characteristics*: Some hardware Trojans are always on, taking actions such as leaking sensitive information or monitoring circuit activities; others remain silent until they are triggered by a particular event or stimulus i.e. triggers. Based on triggering condition they can be classified as (i) internally triggered (activated by an event inside the chip) for example, temperature, voltage or frequency change etc. and (ii) externally triggered (any user input in the form of a data stream or any other communicating signal). An externally triggered Trojan needs a sensing circuitry to receive the trigger signal [28].

3. *Action Characteristics:* These Trojans are classified on the basis of the malicious behavior they introduced in the chip or a system. Based on their action they are classified into three categories: (i) *Modify Function* (changing the chip function through addition, removal or modification of a logic circuitry), (ii) *Modify Specification* (changing chip parameters like delay by modifying interconnects and transistor geometries) and (iii) *Transmit Info* (transmitting information to adversary) [6].

Since Trojan Scanner is based on physical inspection of the backside of the chip, irrespective of the circuit functionality, in this paper, we keep our discussion on Trojans categorized by their physical characteristics only.

3. Trojan Scanner 2.0

Trojan Scanner 2.0 is divided into five major phases: A) Sample preparation; B) Rapid SEM imagining; C) Image pre-processing of images; D) Feature clustering and gates recognition using K-means and multi-class support vector machine (SVM); and E) Detection of gate-level changes between a golden layout and IUA SEM image (see Fig. 2).

3.1 Sample Preparation

In this work, we use a smart card as our test sample. The smartcards are commonly used in financial payment systems like credit/debit cards, communication systems like cellphone SIM card or satellite television box and as an identification card by employers or as a national ID in some countries. Hence, an adversary can easily steal sensitive or confidential information, cause a data breach, and cause financial loss to a smart card using entities by implementing a Trojan in the circuitry.

A smart card die (see Fig. 3) is encapsulated into a thin epoxy resin, which is packaged into a plastic shield on one side and a metallic contact pad on another side. Smart card chip de-capsulation begins with removing the die by cutting the package with a sharp cutter. The die that is covered by epoxy resin can be further de-capsulated by using a few drops of fuming nitric acid followed by an acetone and iso-propyl alcohol wash. The silicon substrate is further thinned by using precise polishing equipment VarioMill [29].

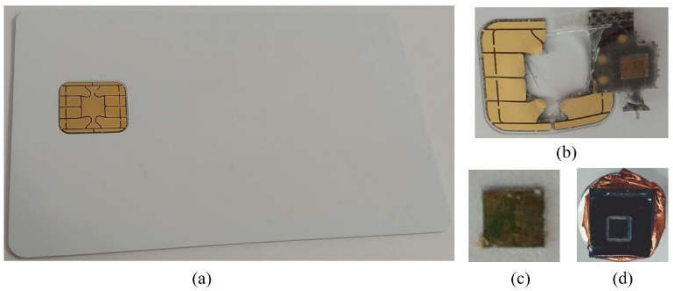

Figure 3*. Sample preparation: (a) Smart card, (b) Packaged die removal, (c) Bare die and (d) Backside thinned die mounted on SEM stub.*

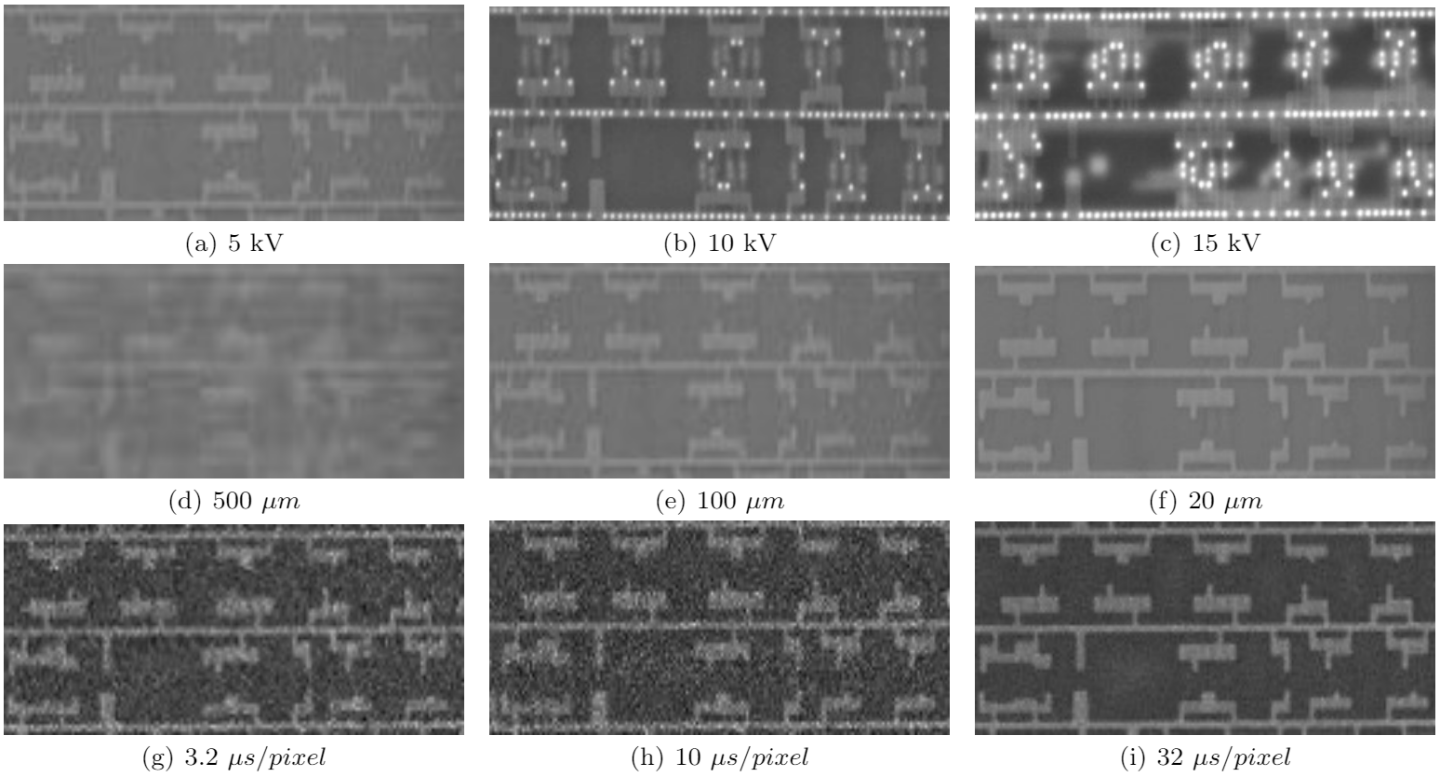

Figure 4. *SEM images variations with different [(a), (b) and (c)] Beam voltages, [(d),(e) and (f)] Field of views and [(g),(h) and (i)] Dwelling times.*

3.2 SEM Imaging

The objective here is to scan the whole die as fast as possible while capture sufficient feature details so to compare with the chip layout; this will significantly reduce the Trojan Scanner processing time. The timing and quality of the SEM images depend on the following SEM parameters. We have compared the effects of these parameters by varying one parameter at a time while keeping all other parameters constant (see Fig. 4).

i. *Beam voltage* - A 5kV beam can expose active regions while imaging from the backside, whereas 10kV, can further expose sub-surface features including the polysilicon and higher metal layers.
ii. *Field of view (FOV)* - A big field of view covers more features, but they are blurred because of the low magnification. Imaging time increases with the decrease in the field of view.
iii. *Dwelling time (speed)* - A higher dwelling time increases the signal-to-noise ratio of the image hence better quality of SEM images, but it increases the time to capture images. Meanwhile, it also affects the surface charging that can introduce artifacts in imagining.

After setting the above-mentioned parameters, the microscope can be programmed to scan the whole die in the form of small windows of images and these individual images are then stitched together to create a complete panorama image.

The SEM imaging data in Table 1 summarizes the SEM image acquisition time to finish scanning of a 1.5mm x 1.5mm die, with different field of view versus dwelling time for 2048x2048 resolution. One can easily conclude based on the images in Fig. 4 and imaging time data in Table 1 that the images captured with a large field of view, small dwelling time takes less imaging time, but these images are unsuitable to detect changes. A small field of view with large dwelling time can capture the superior quality of images but it is collecting more data than required and makes the imaging process very long. Hence, there is a trade-off between the imaging time and suitable quality of images, to get better results for Trojan detection. To balance the time consumption and detection confidence, optimum SEM parameters are used (highlighted in green).

Table 1. *SEM imaging time variation over dwelling time and field of View.*

Dwellng Time (µs/pixel)	Field of View (µm×µm)			
	1500	500	100	20
1	6 s	54 s	22 min 30 s	9 hr 23 min
3.2	14 s	2 min 5 s	52 min 5 s	21 hr 42 min
10	1 min 25 s	6 min 25 s	5 hr 19 min	132 hr 49 min
32	2 min 52 s	24 min	10 hr 45 min	265 hr 30 min

3.3 Feature Identification

To simulate the presence of hardware Trojan in our IC, we have created similar looking doping region layout from the SEM image. Some of the logic gate footprints were modified in layout so that the areas of change in IUA SEM image can be emulated as a hardware Trojan. The feature identification includes four important parts as described below.

3.3.1 Image Pre-processing

In order to obtain a reliable shape descriptor for feature recognition, the gates need to be segmented from the blurry background, and the noise in SEM images needs to be removed [30]. Hence, it is necessary to apply image pre-processing algorithms to obtain clear features for characterizing them with a unique descriptor. The pre-processing of images for better feature detection involves the following steps (see Fig. 5):

- *Histogram equalization* - This method enhances image contrast by flattening pixel intensities [31] and the intensity of features can be better distributed on the histogram. This allows the SEM image features (logic gates at doping region) to gain better contrast as compared to the dark background. It is a well-developed and successfully used method in applications using medical and radar image processing [32].
- *Gaussian blur* - Gaussian white noise is considered as the predominant type of noise in SEM image [33]. A low pass filter called a Gaussian filter with the kernel size [34] of five is applied to address this problem.
- *Median filter* - The median filter is a smoothening technique that can effectively remove noise from the images and at the same time, it can preserve the edges information effectively [35][36]. In our work, edge detection with high accuracy is of critical importance to detect every unique footprint of a logic gate / digital circuit cell. Some of the features of our SEM images are very thin and have a very small number of pixels (i.e. two) to represent them, therefore, a small sized (i.e. size of 3) median filter is used.
- *Thresholding* – Binary Thresholding can segment the grayscale image into a binary image. It can remove the dark background and segment active region features for generating a shape descriptor.

3.3.2 2D Fourier Shape Descriptor

There are various well-developed 2D shape descriptors that generally fall into three categories based on: *Contour, Silhouette* and *Hybrid* which is the on fusion of contour and silhouette. To make our SEM features representation and classification faster, we use contour-based descriptor as it reduces the number of pixels for calculation. 2D Fourier feature descriptor (FD) is one of the most promising methods due its simplicity and robustness (image rotation or distortion noise) [37]. This feature descriptor is based on Fourier transform, it is easy to implement and requires less computational effort as compared to other contour-based descriptors such as the wavelet descriptor (WD) and the curvature scale space descriptor (CSS). These FDs can easily describe the closed shape of p-doped and n-doped active regions connected to the power and the ground rail of the circuit. FD for every logic gate (cell) can be obtained through extracting contour coordinates from a closed 2D shape, calculating shape signatures, and applying Fourier transform.

1) The contour coordinates are usually obtained from shape silhouette/mask (e.g. the threshold gate mentioned above).

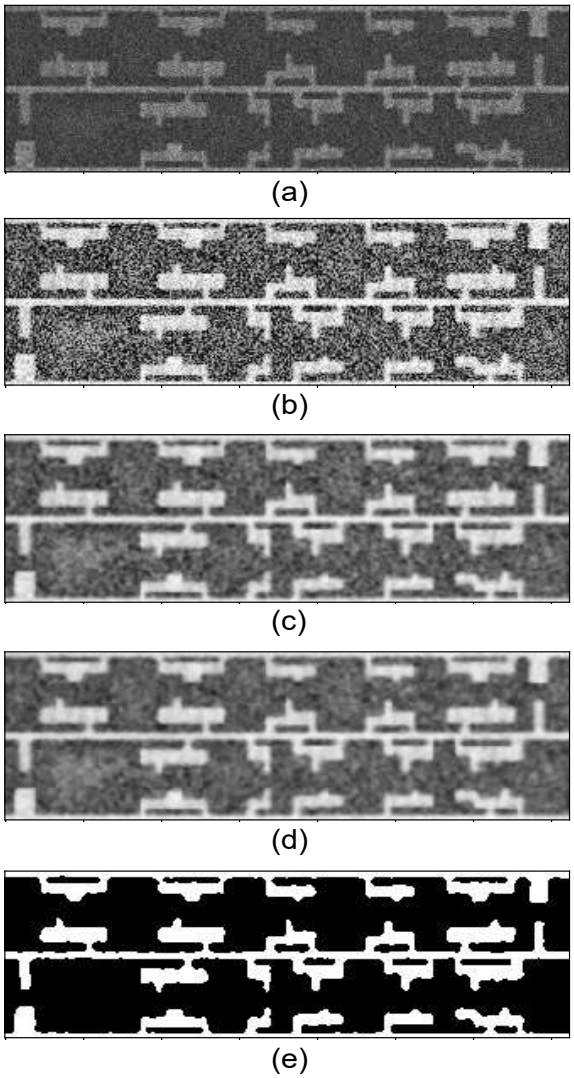

Figure 5*. Image pre-processing. (a) Original SEM image, (b) Histogram equalization, (c) Gaussian filtering, (d) Median filtering and (e) Thresholding.*

2) The shape signatures mostly are centroid distance, complex coordinates, triangular centroid area, and the length of the chord. According to [38], the centroid distance is sufficient to distinguish between different shapes of logic cells/gates.
3) The Fourier transform is applied to the shape signature as equation (1):

$$f[k] = DFT(C[n]) = \frac{1}{N} \sum_{n=0}^{N-1} C[n] e^{\left(\frac{-j2\pi kn}{N}\right)}, \quad k = 0,1,\ldots.N-1 \quad (1)$$

Where $f[k]$ is the Fourier transform of the coordinates at k^{th} shape signature denoted by $C[n]$

The logic gate/cell are not a closed shape feature (see Fig. 7). In order to apply Fourier descriptor to the whole gate, we segment each gate/cell as upper and lower part from the center, calculate the 20 Fourier coefficients for every part separately and concatenate them to generate a 40 dimensional FD vector for each gate by using equation (2)

$$FD_g = [f_{upper}[i], f_{lower}[i]], i = 0, 1, ..., 19 \quad (2)$$

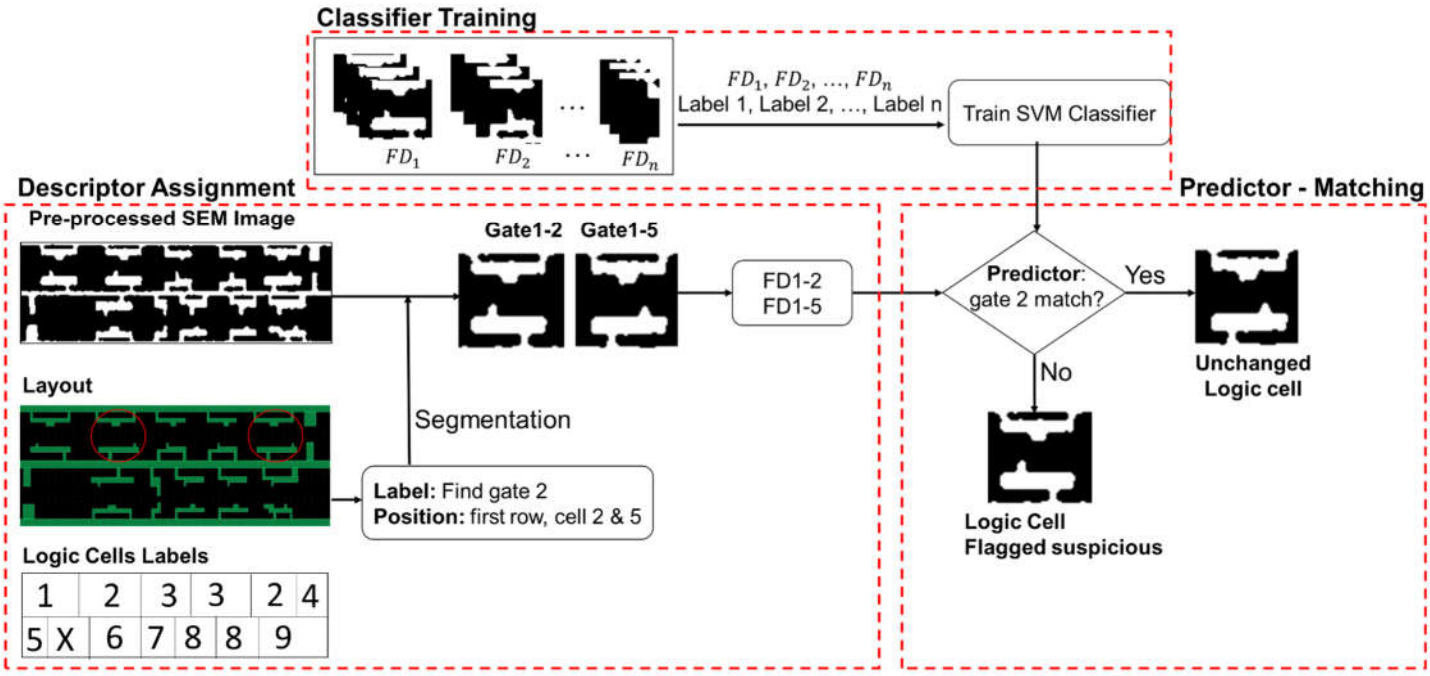

Figure 6. *Trojan Scanner matching engine.*

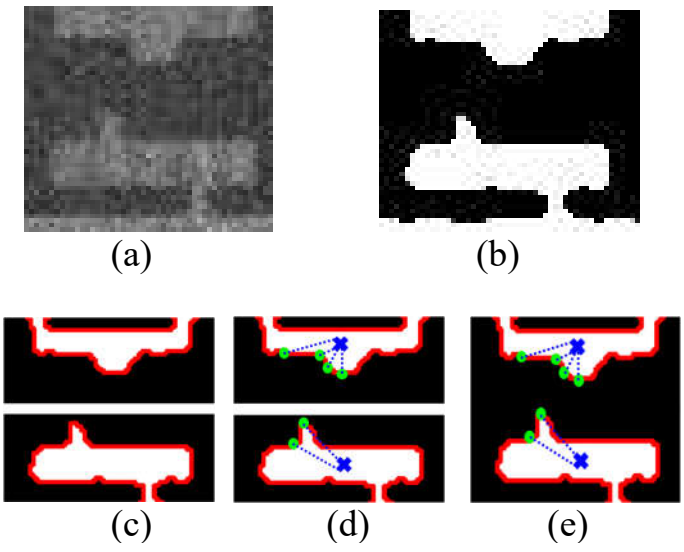

Figure 7. *Generation of Fourier descriptors. (a) Segmented SEM image of a logic cell, (b) Binarized image of logic cell, (c) Converting logic cell into closed shape, (d) Generating Fourier descriptor for upper and lower closed shape. (e) Combining upper and lower descriptors for whole logic cell.*

3.3.3 Multi-class SVM Classifier

In this study, the classifier is responsible for recognizing the gate types based on the SEM images, which is the key decision-making entity for of hardware Trojan detection. Previously Kulkarni et. al [40], Bao et. al [25] used one-class SVM to detect hardware Trojans. Another more efficient solution is to employ multi-class SVM to perform the same task in parallel. Once the multi-class predictor is well trained, it can be

efficiently applied on overall SEM image and validate different types of logic gates. In addition, the RBF kernel has used in multi-class SVM, which is more efficient in distinguishing

between non-linearly separable classes [41]. In this study, SEM images of different types of gates are labeled and fed to the classifier for training.

3.3.4 Matching Engine or Predictor

After training the predictor, the golden layout is used as a dictionary of logic cells to search target logic cells at the corresponding location in SEM image. Next, the targeted logic cells from SEM image are segmented and assigned a unique Fourier descriptor. Once the matching is made between the Fourier descriptor and predictor, a decision can be made on the authenticity of the logic cell whether the gate is genuine or changed maliciously.

For example (see Fig. 6), there are two instances of a logic gate (labeled Gate-2) in the first row of the layout at cell location 2 and 5. If we search these logic gates at the corresponding location of the SEM image, we find gate1-2 and gate1-5 with assigned a Fourier descriptor FD1-2 and FD1-5 respectively. Using predictor when both of these SEM and layout descriptors are compared one gate is found to be modified and another one as unchanged.

4 Results

4.1 Data samples

To demonstrate the detection of any kind of change, we recreate a layout of doping area based on the smart card die's SEM image (see Fig. 8) to mimic the presence of a hardware Trojan. In addition to this, we captured two data sets of IUA images with different imaging time i.e. 32 µs/pixel (Speed: 6) and 10 µs/pixel (Speed: 5) while other parameters are kept same (see Fig. 9). Based on the above-mentioned image processing and Fourier descriptor techniques a unique feature vector is assigned to golden layout and IUA images.

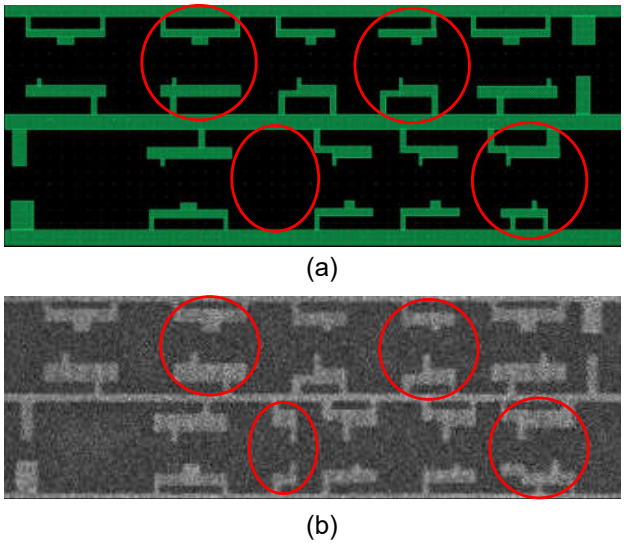

***Figure 8**. (a) Layout of the smart card chip (b) SEM image of the corresponding area. (Circled areas reflect modification or insertion of the logic cell.)*

The training data set can be generated by using the SEM images captured under different imaging conditions (see Fig. 9). These different SEM images have different levels of noise and pixel counts. As a proof of concept, we identified nine types of logic cells for training our model (see Fig. 10).

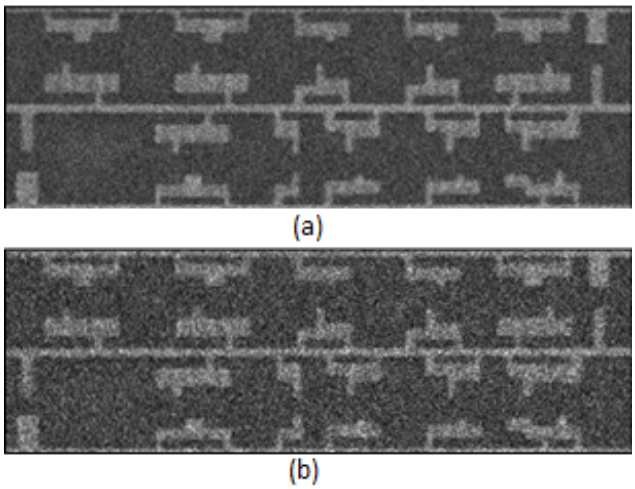

***Figure 9**. SEM images obtained by different imaging speed. (a) Speed 6: 32 µs/pixel and (b) Speed 5: 10 µs/pixel.*

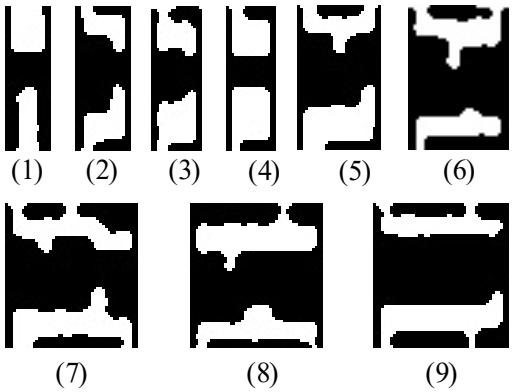

***Figure 10**. Different logic cells identified to train machine-learning model.*

4.2 Performance of Predictor

The multi-class SVM is trained with RBF kernel and verified by using five cross-validations. Fig. 11 shows the learning curve of the classifier training procedure for SEM images captured at two different imaging conditions i.e. SVM 6: Speed 6 & SVM 5: Speed 5.

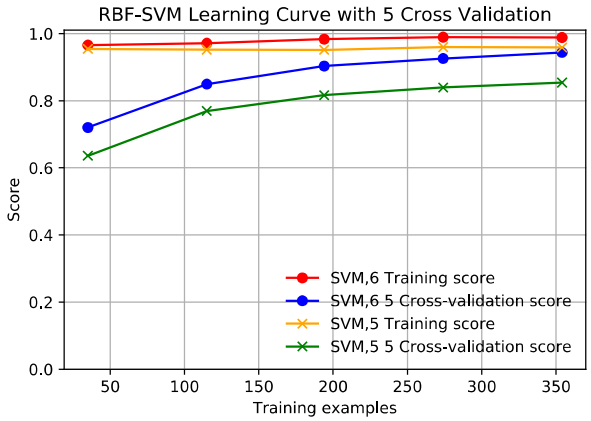

***Figure 11**. The learning curve of RBF-SVM classifier in SEM image obtained with speed 5 (SVM 5) and speed 6 (SVM 6).*

According to the classifier curve, the average recognition accuracy for 9 types of logic cells is 98±2% and 89±4% in SEM image with speed of six and speed of five respectively. The Trojan detection accuracy approaches nearly 100% when using a good quality (Speed 6) image as compared to the 93% while using somewhat noisy (Speed 5) image. These are the results when we have only used the basic image pre-processing methods (Gaussian and median filter), the performance is expected to be improved when other pre-processing steps (e.g. adaptive Wiener filter [33][42]) are applied.

The detection rate on the whole SEM image depends on the accuracy of the trained predictor that shown previously. Figure 12 shows three kinds of changes detection based on logic gate/cell modification, insertion/deletion, and substitution.

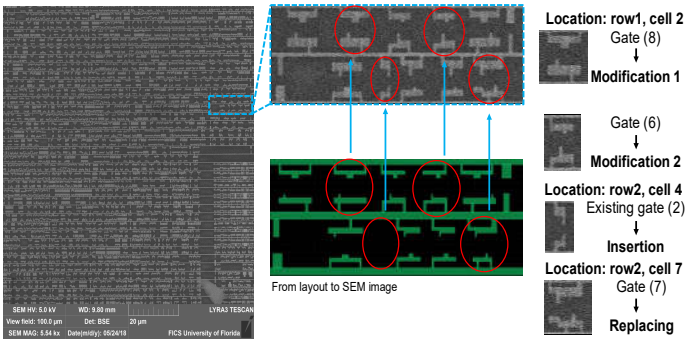

***Figure 12**. Layout vs. SEM image comparison results.*

It is evident from Table 2 that better quality images have better change detection capability.

***Table 2**. False positive rates in different quality SEM images.*

False Positive	Speed 5 Image	Speed 6 Image
Gate1	0	0.03
Gate2	0.04	0
Gate3	0.33	0.01
Gate4	0	0
Gate5	0.01	0.05
Gate6	0.03	0
Gate7	0.15	0.08
Average	**0.08**	**0.02**

Tables 3 and 4 show the comparison results between our Trojans Scanner and full reverse engineering method as well as with other electrical test techniques to detect Trojans respectively.

***Table 3**. Full chip reverse engineering vs. Trojan Scanner.*

Metric	Full Reverse Engineering	Trojan Scanner
# of samples required	50-100	1
Detected Trojans	All types (except reliability Trojan)	All types (except Reliability & Parametric)
Processing time	Months	Hours
Functionality extraction	Required	Not required
Image processing (Polygon Extraction)	Required	Not required

***Table 4**. Trojan Scanner vs. Electrical tests.*

Hardware Trojans	Logic Test	Power SCA	Delay SCA	Run Time	Trojan Scanner
Functional	Maybe	Maybe	Maybe	Maybe	✓
Parametric	×	✓	✓	×	✓
Big	Maybe	✓	Maybe	✓	✓
Small	✓	×	✓	Maybe	✓
Tight	✓	✓	✓	Maybe	✓
Loose	✓	Maybe	✓	Maybe	✓

4.3 Confidence level and Sample size

To ensure ICs from a batch under authentication are Trojan-free, one can use a more efficient approach called acceptance sampling instead of testing all ICs, however, depending on the user's goals this method may or may not be chosen. An acceptance sampling approach uses Acceptable Quality Limit (AQL) ISO 2859 standard tables, which is a widely used method to measure if the production order has met the client's satisfaction or not. These quality limits are classified by critical defects - 0% (totally unacceptable, a user might get harmed or not meet regulatory conditions), major defects - 2.5% (unacceptable by end user) and minor defects (4% some departure from specification, end user won't mind using it) [43]. Based on these quality limits, the client has the inspection sample size to make an informed decision to accept or reject the lot [44]. As discussed earlier the Trojans have the capability to cause a security threat to a nation or a major scale financial and human life loss, so we can follow the critical defect level for our Trojan inspection. For example, suppose a government orders a lot 1000 ICs (government orders are limited in thousands) and based on AQL model inspection level II [44], we need to inspect 80 ICs for Trojan detection. Since we are using a critical defect limit, if we detect a Trojan in a single IC, the whole lot can be rejected. Using our technique, if we do not find any Trojan in 80 samples of ICs, then with a confidence level of 95%, we can claim the rest of 920 ICs are Trojan-free.

5 Conclusions

Current Trojan detection techniques available in the market and studied by researchers usually lack the coverage, speed and/or confidence of detection. Hardware Trojans, also leave a footprint either on active layer or a metal layer of an IC. In this paper, we have enhanced our previously developed technique "Trojan Scanner" by making it a golden IC free detection technique. We discussed possible scenarios of Trojan insertion and demonstrated their detection approach by comparing golden layout with IUA SEM image. We observed that during developing our technique, there is a trade-off between the accuracy of detection and SEM parameters (dwelling time, FoV). Faster SEM imaging will cause a performance drop. The challenge is to apply advanced image pre-processing method to reconstruct the noisy image. Some possible solutions are employing adaptive Wiener filter, noise filter under frequency domain, or applying an artificial neural network. Although we consider minimized time spending on image processing, we may use more complex contour-based shape descriptors to balance the overall processing time. For example, if we can apply complex shape descriptor that requires more time on the faster-scanned image, the overall processing time of Trojan detection will decrease dramatically. We finally discussed the confidence level of Trojan Scanner and the minimum number of ICs required establishing a trust in the supply chain.

References

[1] N. Vashistha, M. Tanjidur Rahman, H. Shen, D. L. Woodard, N. Asadizanjani, and M. Tehranipoor, "Detecting Hardware Trojans Inserted by Untrusted Foundry using Physical Inspection and Advanced Image Processing," Springer journal of Hardware and Systems Security, special issue on Hardware Reverse

engineering and Obfuscation 2018. (publication pending)
[2] "Semi industry fab costs limit industry growth | EE Times." [Online]. Available: https://www.eetimes.com/document.asp?doc_id=1264577. [Accessed: 25-Jul-2018].
[3] "Samsung Breaks Ground on $14 Billion Fab | EE Times." [Online]. Available: https://www.eetimes.com/document.asp?doc_id=1326565. [Accessed: 25-Jul-2018].
[4] M. Tehranipoor and F. Koushanfar, "A Survey of Hardware Trojan Taxonomy and Detection," *IEEE Des. Test Comput.*, vol. 27, no. 1, pp. 10–25, Jan. 2010.
[5] G. T. Becker, F. Regazzoni, C. Paar, and W. P. Burleson, "Stealthy Dopant-Level Hardware Trojans," Springer, Berlin, Heidelberg, 2013, pp. 197–214.
[6] X. Wang, M. Tehranipoor, and J. Plusquellic, "Detecting Malicious Inclusions in Secure Hardware: Challenges and Solutions."
[7] "The Hunt for the Kill Switch - IEEE Spectrum." [Online]. Available: https://spectrum.ieee.org/semiconductors/design/the-hunt-for-the-kill-switch. [Accessed: 25-Jul-2018].
[8] "The Navy Bought Fake Chinese Microchips That Could Have Disarmed U.S. Missiles - Business Insider." [Online]. Available: https://www.businessinsider.com/navy-chinese-microchips-weapons-could-have-been-shut-off-2011-6. [Accessed: 25-Jul-2018].
[9] X. Zhang and M. Tehranipoor, "Case study: Detecting hardware Trojans in third-party digital IP cores," in *2011 IEEE International Symposium on Hardware-Oriented Security and Trust*, 2011, pp. 67–70.
[10] K. Xiao, D. Forte, Y. Jin, R. Karri, S. Bhunia, and M. Tehranipoor, "Hardware Trojans: Lessons Learned after One Decade of Research," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 22, no. 1, pp. 1–23, May 2016.
[11] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, "Circuit Failure Prediction and Its Application to Transistor Aging," in *25th IEEE VLSI Test Symmposium (VTS'07)*, 2007, pp. 277–286.
[12] G. Zhang, D. Das, R. Xu, and M. Pecht, "IDDQ trending as a precursor to semiconductor failure," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–7.
[13] Yier Jin and Y. Makris, "Hardware Trojan detection using path delay fingerprint," in *2008 IEEE International Workshop on Hardware-Oriented Security and Trust*, 2008, pp. 51–57.
[14] F. Wolff, C. Papachristou, S. Bhunia, and R. S. Chakraborty, "Towards Trojan-free trusted ICs: Problem analysis and detection scheme," in *Proceedings of the conference on Design, automation and test in Europe*, 2008, pp. 1362–1365.
[15] R. S. Chakraborty, F. Wolff, S. Paul, C. Papachristou, and S. Bhunia, "MERO: A statistical approach for hardware Trojan detection," in *Cryptographic Hardware and Embedded Systems-CHES 2009*, Springer, 2009, pp. 396–410.
[16] S. Narasimhan, X. Wang, D. Du, R. S. Chakraborty, and S. Bhunia, "TeSR: A robust temporal self-referencing approach for hardware Trojan detection," in *Hardware-Oriented Security and Trust (HOST), 2011 IEEE International Symposium on*, 2011, pp. 71–74.
[17] H. Salmani, M. Tehranipoor, and J. Plusquellic, "A novel technique for improving hardware trojan detection and reducing trojan activation time," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 20, no. 1, pp. 112–125, 2012.
[18] J. Aarestad, D. Acharyya, R. Rad, and J. Plusquellic, "Detecting Trojans Through Leakage Current Analysis Using Multiple Supply Pad I_{DDQ}s," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 4, pp. 893–904, Dec. 2010.
[19] S. Bhunia, M. S. Hsiao, M. Banga, and S. Narasimhan, "Hardware Trojan Attacks: Threat Analysis and Countermeasures," *Proc. IEEE*, vol. 102, no. 8, pp. 1229–1247, Aug. 2014.
[20] S. E. Quadir, J. Chen, D. Forte, N. Asadizanjani, S. Shahbazmohamadi, L. Wang, J. Chandy, and Mark Tehranipoor, "A survey on chip to system reverse engineering," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 1, p. 6, 2016.
[21] E. L. Principe, N. Asadizanjani, D. Forte, M. Tehranipoor, R. Chivas, M. DiBattista, S. Silverman, M. Marsh, N. Piche, and J. Mastovich, "Steps Toward Automated Deprocessing of Integrated Circuits." ASM, 07-Nov-2017.
[22] R. Torrance and D. James, "The State-of-the-Art in IC Reverse Engineering," in *Proceedings of the 11th International Workshop on Cryptographic Hardware and Embedded Systems*, Springer-Verlag, 2009, pp. 363–381.
[23] F. Courbon, P. Loubet-Moundi, J. J. A. Fournier, and A. Tria, "A High Efficiency Hardware Trojan Detection Technique Based on Fast SEM Imaging," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015*, 2015, pp. 788–793.
[24] F. Courbon, P. Loubet-Moundi, J. J. A. Fournier, and A. Tria, "SEMBA: A SEM based acquisition technique for fast invasive hardware trojan detection," in *Circuit Theory and Design (ECCTD), 2015 European Conference on*, 2015, pp. 1–4.
[25] C. Bao, D. Forte, and A. Srivastava, "On application of one-class SVM to reverse engineering-based hardware Trojan detection," in *Fifteenth International Symposium on Quality Electronic Design*, 2014, pp. 47–54.
[26] B. Zhou, R. Adato, M. Zangeneh, T. Yang, A. Uyar, B. Goldberg, S. Unlu, and A. Joshi., "Detecting hardware trojans using backside optical imaging of embedded watermarks," in *Proceedings of the 52nd Annual Design Automation Conference on - DAC '15*, 2015, pp. 1–6.
[27] B. Shakya, T. He, H. Salmani, D. Forte, S. Bhunia, and M. Tehranipoor, "Benchmarking of Hardware Trojans and Maliciously Affected Circuits," *J. Hardw. Syst. Secur.*, vol. 1, no. 1, pp. 85–102, Mar. 2017.
[28] M. Tehranipoor and C. Wang, *Introduction to*

Hardware Security and Trust. Springer Science & Business Media, 2011.

[29] "varioscale | varioMill." [Online]. Available: https://www.varioscale.com/variomill. [Accessed: 29-Mar-2018].

[30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, 1992.

[31] R. C. Gonzalez and R. E.Woods, "Histogram Equalization," no. June, pp. 1–3, 2008.

[32] Y. T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, 1997.

[33] K. S. Sim, V. Teh, and M. E. Nia, "Adaptive noise Wiener filter for scanning electron microscope imaging system," *Scanning: The Journal of Scanning Microscopies*, vol. 38, no. 2, pp. 148–163, 2016.

[34] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 1, pp. 26–33, 1986.

[35] S. J. Ko and Y. H. Lee, "Center Weighted Median Filters and Their Applications to Image Enhancement," *IEEE Trans. Circuits Syst.*, vol. 38, no. 9, pp. 984–993, 1991.

[36] P. L. A. T. Loupas, W.N.Mcdicken, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images," vol. 510, no. 1, p. 11, 1989.

[37] D. Zhang and G. Lu, "Generic Fourier descriptor for shape-based image retrieval," *Proc. - 2002 IEEE Int. Conf. Multimed. Expo, ICME 2002*, vol. 1, no. July, pp. 425–428, 2002.

[38] G. Zhang, Z. Ma, L. Niu, and C. Zhang, "Modified Fourier descriptor for shape feature extraction," *J. Cent. South Univ.*, vol. 19, no. 2, pp. 488–495, Feb. 2012.

[40] A. Kulkarni, Y. Pino, and T. Mohsenin, "SVM-based real-time hardware Trojan detection for many-core platform," *Proc. - Int. Symp. Qual. Electron. Des. ISQED*, vol. 2016–May, pp. 362–367, 2016.

[41] B. C. Kuo, H. H. Ho, C. H. Li, C. C. Hung, and J. S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 1, pp. 317–326, 2014.

[42] K. S. SIM, K. K. LAW, and C. P. TSO, "Mixed Lagrange Time Delay Estimation Autoregressive Wiener Filter Application for Real-Time SEM Image Enhancement," *Infect. Immun.*, vol. 70, pp. 919–927, 2007.

[43] "What is the 'AQL' (Acceptance Quality Limit) in simple terms?" Available: https://qualityinspection.org/what-is-the-aql/ Nov-2011.

[44] "Acceptable Quality Limit (AQL)." Available: http://www.asiainspection.com/aql-acceptable-quality-limit [Accessed: 29-Mar-2018].