Noname manuscript No.
(will be inserted by the editor)

Detecting Hardware Trojans Inserted by Untrusted Foundry using Physical Inspection and Advanced Image Processing

Nidish Vashistha · M Tanjidur Rahman · Haoting Shen · Damon L Woodard · Navid Asadizanjani · Mark Tehranipoor

Received: date / Accepted: date

Abstract Hardware Trojans are malicious changes to the design of integrated circuits (ICs) at different stages of the design and fabrication process. Different approaches have been developed to detect Trojans namely: non-destructive (electrical tests like run-time monitoring, functional and structural tests) and destructive testing (full chip reverse engineering). However, none of the previously developed methods can be used to detect all types of Trojans and they all suffer from a number of disadvantages such as low speed of detection, low accuracy, low confidence level and poor coverage of Trojan types. Based on our literature survey of Trojan benchmarks, majority of the hardware Trojans implemented in an IC will leave a footprint at the active layer. In this paper, we propose a new technique based on rapid backside SEM imaging and advanced computer vision algorithms to detect any subtle changes on the active region of transistors that can show the existence of a hardware Trojan. Here, we are only concerned with untrusted foundry problem, where it is assumed the end-user has access to a golden layout/image of the IC. This is a common threat model for those organizations that fully design their IC but need access to untrusted foundry for fabrication. SEM image from a backside thinned golden IC is compared with a low-quality SEM image of an IC under authentication (IUA). We perform image processing to both golden IC & IUA images to remove noise. We develop a computer vision based algorithm to detect hardware Trojans based on their structural similarity. The results demonstrate that our technique is quite effective at detecting Trojans and significantly faster than full chip reverse engineering. One of the major advantages of our technique is that it does not rely on the functionality of the circuit, rather the real physical structure to detect malicious changes performed by the foundry.

Keywords Hardware Trust · Trojan Detection · Image Processing · Reverse Engineering · Scanning Electron Microscopy

1 Introduction

Outsourcing integrated circuit (IC) design, fabrication, validation and verification facilities have reduced the costs and time to market. Building and maintaining an advanced technology node foundry can cost up to several billions of dollars [4]. Hence, most of the design companies have become fabless or they have migrated their fabrication team offshore [5], for which they have to rely on 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, net-list, and test vectors. As a result, this trust issue has opened

Nidish Vashistha
nidish@ufl.edu

MT Rahman
mir.rahman@ufl.edu

Haoting Shen
htshen@ufl.edu

Damon L Woodard
dwoodard@ece.ufl.edu

Navid Asadizanjani
nasadi@ece.ufl.edu

Mark Tehranipoor
tehranipoor@ece.ufl.edu

Florida Institute for Cyber Security (FICS) Research
Electrical & Computer Engineering Department
University of Florida, Gainesville, FL, USA

up an avenue to various kinds of threats in ICs including hardware Trojan insertion, overproduction, IP piracy, and out-of-specification/defective ICs appearing in the market [27].

Among all of these trust issues, hardware Trojans are the most threating because they can compromise the security & trustworthiness of a system and they can be very difficult to be detected because of their stealthy nature [12]. A hardware Trojan is a malicious modification to the circuit during any phase of design, integration or fabrication [30]. Using hardware Trojan an adversary can cause a denial of service, control or leak sensitive information from the system. The hardware Trojans can be a major threat to all electronic devices (home automation devices, security cameras and locks etc.), civilian applications (aviation, law enforcement and health-care) and most importantly military and space systems. There have been instances reported where a system's security was compromised because of suspected "back-doors" [2, 8].

Hardware Trojans can be inserted during any step of the IC design process due to the involvement of untrusted entities [36]. 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. IC design involves three main entities. They are, third-party intellectual property (3PIP) vendor who provides functional blocks, system-on-chip (SoC) developers who develop the architectural platform for a design and the last entity is the foundry which fabricates the ICs. There can be different kinds of attack models based on the trust assumption with any of these entities [33]. Among them, the threat model of the untrusted foundry has been widely discussed in the hardware security community [33]. In this model (Figure 1), the foundry is the only untrusted entity and perceived as a threat for malicious hardware insertion during fabrication. These Trojans can be inserted into the chips during the wafer mask generation step of fabrication and various other ways. 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 of the interconnects which can cause an early failure (i.e. denial of service attack). This paper is only concerned with the untrusted foundry model, where a golden layout/IC is assumed to be available.

Previous studies and surveys on Trojan benchmarks [23, 25] shows the reliability Trojans are the only one with no physical footprint on doping layer among all types of Trojans (categorized based on their physical, activation, and action characteristics). These Trojans can cause failure by the acceleration of wear-out mechanisms in CMOS transistors, such as negative bias temperature instability (NBTI), hot carrier injection (HCI) or electromigration etc. [26]. These Trojans can be easily detected by aging test [9] or transient/quiescent current test methods (IDDT/IDDQ) [35].

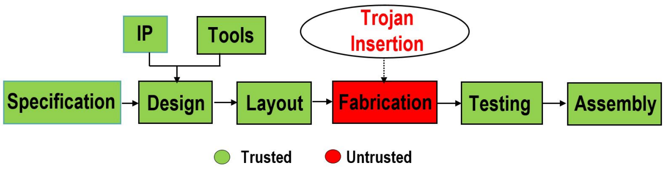


Fig. 1 Untrusted foundry attack model for Trojan insertion.

There has been extensive research on detecting hardware Trojans using run-time monitoring [18] and logic test approaches [7,14,20,24,27,32]. However, such Trojan detection techniques have a number of serious limitations. For example, the run-time monitoring techniques increase resource utilization on IC by using on-chip sensors to detect malicious activities. Such techniques consumes extra CPU usage, power, memory and silicon area on the chip. 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 approaches are vulnerable to circuit noise hence they cannot detect small size Trojans [13]. As a result, the confidence level in detecting Trojans using above-mentioned techniques are quite low. Another approach discussed in the community is destructive method, where the full blown reverse engineering of the IC must be performed [21]; in this case, a chip is fully reverse engineered [29] to reconstruct the circuit net-list. 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 is very sensitive, and destructive, therefore many samples are sacrificed before the right recipes for delayering is prepared. However, reverse engineering of a chip for trust verification is the most effective one for the untrusted foundry threat model.

Not much attention was paid to Trojan detection using physical inspection techniques due to the high cost of advanced microscopy machines. However, over the past few years and with the advancements in the microscopy field these machines are more available to public and easier to rent by hours or purchase today. Courbon et.al [15, 16] proposed the basic concept of image processing to detect Trojans using SEM images. They have 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 only addition of logic gates or transistors as a Trojan insertion approach. Bao et. al [11] proposed a machine learning based technique to detect 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 et. al [37] 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 level.

Hence, we a need a new hardware Trojan detection technique that is reliable, fast, and covers all Trojans that leave footprint on the chip. We propose a new technique based on backside physical inspection, which can address shortcomings of existing Trojan detection techniques and is capable of detecting even smallest type of Trojans. In this paper, we make the following contributions:

- We develop a new physical inspection technique called ”Trojan Scanner” for hardware Trojan detection that is semi-invasive, where a chip's backside is thinned so that we can perform detailed imaging of the active layer. When 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 of an IC.
- As mentioned earlier, majority of the Trojans inserted by a foundry will have to make some modification (even minor) to the doping (active) layer. These changes can be easily detected from backside without the need of reconstruction of net-list or understanding the functionality of the IC under authentication (IUA). , this method does not need involvement of an engineer for imaging, image enhancement, comparison, and decision making. Hence, the entire process can be automated.
- In Trojan Scanner, fast SEM imaging has been performed by optimizing SEM parameters to detect differences between a golden IC and IUA SEM image with high confidence.

The remainder of this paper is organized as follows: Section 2 introduces the hardware Trojan taxonomy in details. Section 3 introduces our proposed technique called Trojan Scanner. Next, we discuss case studies of hardware Trojans, based on SEM imaging performed on a smart card circuit in Section 4. Section 5 presents the results of Trojan detection. Finally, we conclude this paper with our findings in Section 6 with a brief discussion about future work.

2 Trojan Taxonomy

Hardware Trojans can be classified based on their physical, activation and action characteristics [30].

- *Physical Characteristics:* Hardware Trojans which are classified on the basis of the type of geometrical modifications in the chip layout. It can be further sub categorized based on their *Type*: *Functional* category includes Trojans that are implemented by addition or deletion of transistors (logic gates) and *Parametric* category 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) [25]. Based on the *Size* of insertion or deletion it can be sub-categorized as *Big* or *Small.* Also based on their *Distribution* in the chip layout they can be classified as *Tight* (condensed) or *Loose* (scattered).
- *Activation Characteristics:* Some hardware Trojans are always on, taking actions such as leaking sensitive information; others remain silent until they get triggered by a particular event or stimulus (i.e. triggers). Based on triggering condition they can be classified as *internally triggered* (activated by an event inside the chip for example, temperature, voltage or frequency change etc.) and *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 external trigger signal [28].
- *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: *Modify - function* (changing the chip function through addition, removal or modification of a logic circuitry), *Modify - specification* (changing chip parameters like delay by modifying interconnects and transistor geometries) and *Transmit - info* (transmitting information to adversary) [30].

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

3 Trojan Scanner

Trojan Scanner is divided into three major phases: A) Sample preparation; B) Rapid SEM imagining and C) Image processing & computer vision algorithms to detect changes (insertions, deletions & modifications) between a Trojan-free golden layout & IUA SEM image. (Figure 2)

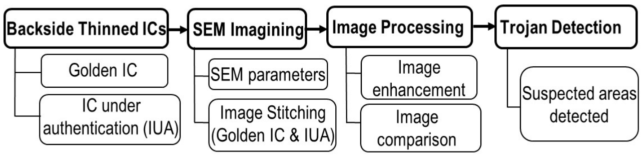


Fig. 2 Trojan Scanner : steps for Trojan Detection.

3.1 Sample Preparation

Packaged ICs are required to be de-capsulated to expose the silicon die. Mechanical polishing is one of the most commonly used techniques to remove the packaging material. However, even after exposing the bare die, it is still not ready for Scanning Electron Microscopy (SEM) imaging because the electrons cannot penetrate into a thick layer of the silicon substrate. The substrate needs to be further thinned by using precise polishing methods. Also, the bare die used in the technique is not flat and curvature keeps changing during the polishing; this change may cause uneven silicon substrate removal over the chip. To mitigate these issues, an advanced silicon die polishing technique called VarioMilll [1] can be used to perform backside thinning up to 1-2 μm. Using a 5-axis computer numerical control tool in combination with interferometer to adapt the polishing rate and the curvature shape is tracked to ensure uniform thinning across the die.

In this work, we use a smart card's as our test sample. The smart cards 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, causing a data breach, big financial loss to these smart card using entities by implementing a Trojan in the circuitry.

A smart card die (Figure 3) is encapsulated into a thin epoxy resin, which is packaged into 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 which 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. Finally, the bare die can be back side thinned by using the Variomill.

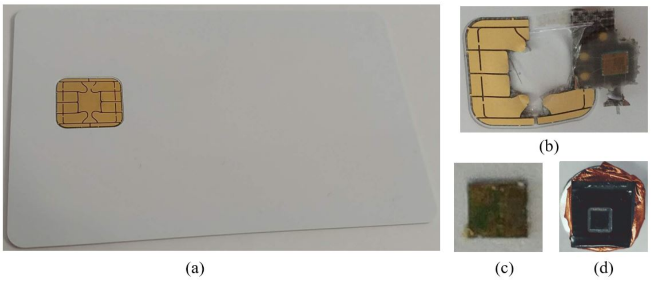


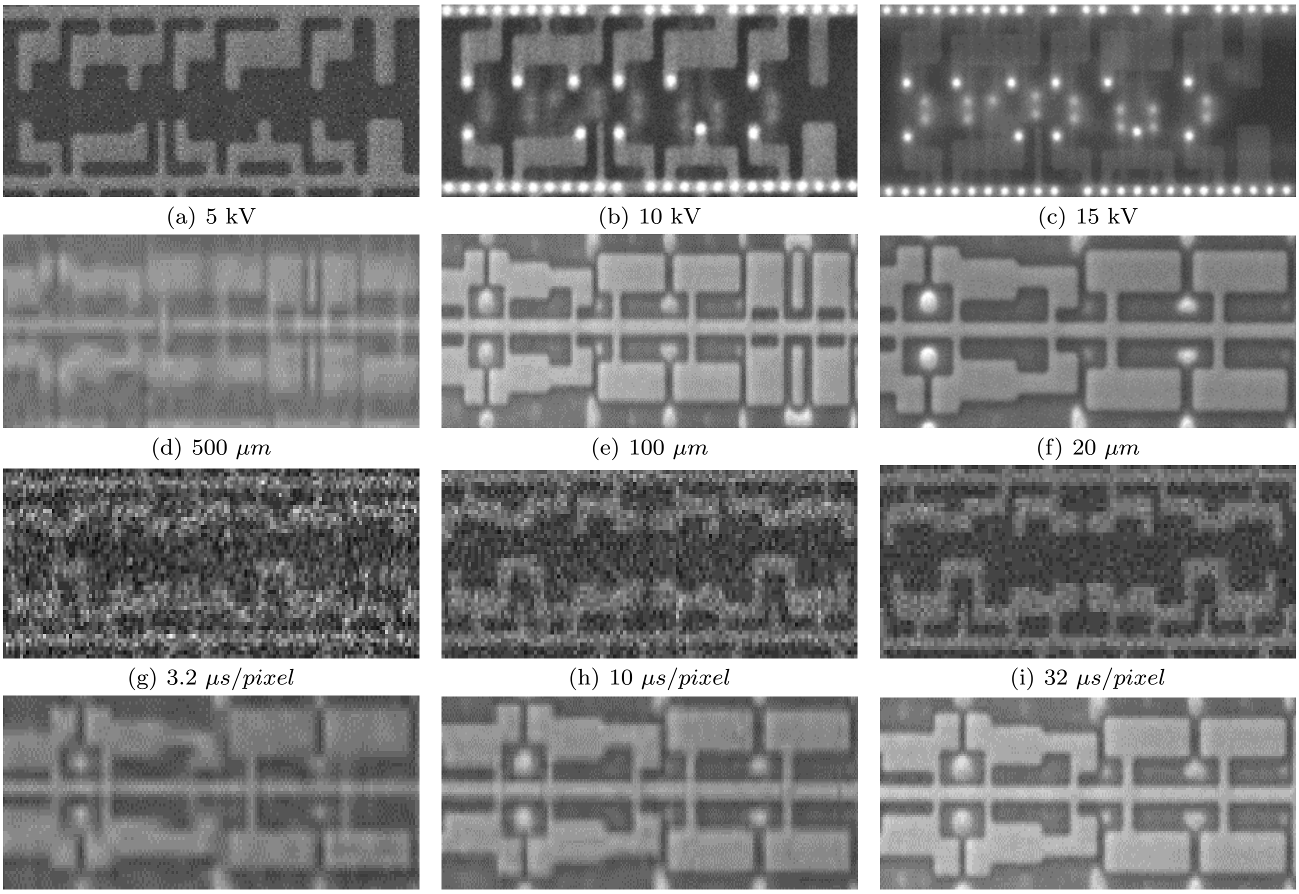
Fig. 3 Sample preparation: (a) Smart card, (b) die removal, (c) bare, die and (d) backside thinned die.

3.2 SEM Imaging

The objective here is to take SEM images from the whole die in short time while capturing enough details for the following comparison between the golden IC and IUA. Hence, capturing SEM images with a proper resolution is an important step in Trojan Scanner. At the time of this research, we do not possess the golden layout of the chip under authentication, therefore high-quality images are used as our golden IC data set, as it has a minimum amount of noise, and clearly captures all features at doping (active) layer. It takes a significantly longer time to obtain these high-resolution images, but these images are as close to the layout of the chip. In summary, high resolution image is used as a golden layout.

Low-quality image of IUA is captured through a very fast scanning process; We manipulate these images and inject Trojans of different types and sizes to serve as our IUA image. 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 (Figure 4).

1. **Beam voltage** - The accelerating voltage (in kV) of electron beam decides the penetration depth of electrons inside the object. For example, 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.
2. **Field of view (FOV)** - It is the area covered by SEM in a single raster. The field of view is inversely related to the magnification of the image. 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.



(a) 5 kV (b) 10 kV (c) 15 kV

(d) 500 μm (e) 100 μm (f) 20 μm

(g) 3.2 $\mu s/pixel$ (h) 10 $\mu s/pixel$ (i) 32 $\mu s/pixel$

(j) 512 x 512 (k) 1024 x 1024 (l) 2048 x 2048

Fig. 4 SEM images variations with different Beam Voltages [(a),(b) and (c)], Field of Views [(d),(e) and (f)], Dwelling Times [(g),(h) and (i)] and Resolutions [(j),(k) and (l).]

3. **Dwelling time (speed)** - It is the time taken by the SEM detector to integrate signal for one pixel. 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.

4. **Resolution** - It denotes the pixel counts in the image. A higher resolution image is more clear and sharp to detect features but it takes much more time to capture higher resolution images.

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 stitched 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, giving different field of view vs. dwelling time for 2048x2048 resolution. One can easily conclude based on the images in Figure 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, large dwelling time and high-resolution 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

Table 1 SEM imaging Time variation over Dwelling time and Field of View.

Dwelling Time (μs/pixel)	**Field of View**			
	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 40 min
32	2 min 52 s	24 min	10 hr 45 min	265 hr 30 min

in green). However, they can be changed to use even faster imaging parameters in future as more advanced and intelligent computer vision techniques developed for Trojan detection.

3.3 Image Processing & Computer Vision

To simulate the presence of hardware Trojan in our IC we have carefully manipulated the IUA SEM image by performing changes at the doping level (figure 5) to represent different types of Trojans.

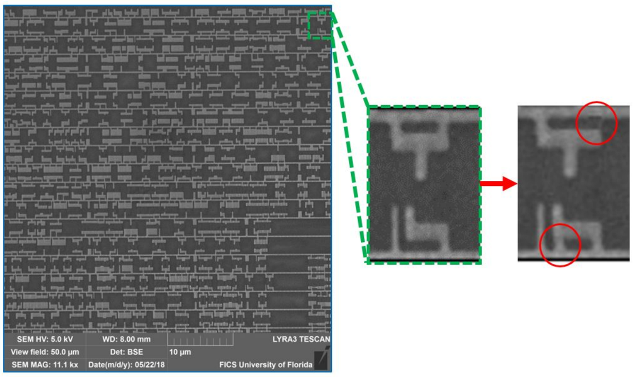


Fig. 5 IUA image to emulate the presence of Trojan.

Before applying an image processing algorithm to detect changes, the raw golden IC and IUA SEM images need to be aligned by using image registration [38] and enhanced for better feature detection. After registration, the images are filtered using FFT [17] bandpass filter to remove high-frequency noise components. Then, images are segmented to separate the doping region (foreground) features from (background) dark area. An adaptive thresholding method has used to segment images because of the variation in contrast and brightness during the whole die imaging. This step returns a binary image as an output and we can detect the sharp edges of features [19]. These features are further smoothened by using a Gaussian filter to smoothen the edges. Since we cannot remove all noise during this enhancement process, the remaining noise may create small holes. These holes can possibly create false positives during image comparison step, so these holes need to be filled using flood fill operation [17].

After the above-mentioned image enhancement process on both golden and IUA images, these images can be compared by Structural SIMilarity (SSIM) [31] to generate a comparison index for every pixel to search the areas of difference. SSIM index is a metric used for measuring the similarities between two images based on their luminosity, contrast and structural difference. It can be calculated between two windows x and y of same size $N \times N$ (equation 1).

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

The SSIM formula is based on the three measurement comparisons between the two samples x and y: luminosity, contrast & structure (equation 2, 3 and 4)

$$\mathrm{luminosity}(x,y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} \tag{2}$$

$$\mathrm{contrast}(x,y) = \frac{(2\sigma_{xy} + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

$$\mathrm{structure}(x,y) = \frac{(\sigma_{xy} + C_3)}{(\sigma_x\sigma_y + C_3)} \tag{4}$$

where

- μ_x and μ_y are the average of x and y;
- σ_x^2 and σ_y^2 are the variance of x and y;
- $\sigma_x y$ the covariance of x and y;
- C$_1$=$(k_1 L)^2$
- L is the dynamic range of the pixel values ($2^{\#bits}$ *per pixel* - 1)
- Using default values in MATLAB of $k_1 = 0.01$ and $k_2 = 0.03$

Based on the SSIM Index, a Trojan plot map has generated to label the area(s) of suspicion (figure 6)

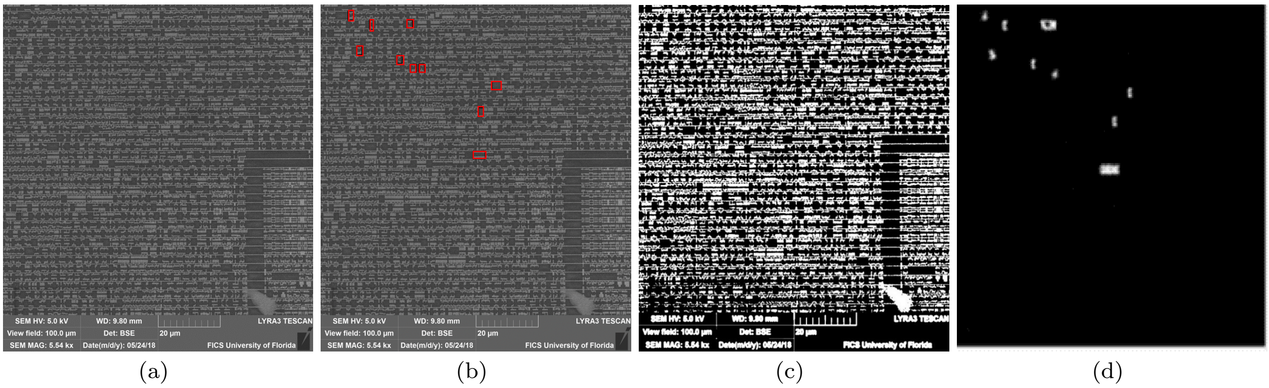
4 Case Studies

Hardware Trojan's modification or insertion-deletion cases can be created by performing malicious changes at active / metal layer. We have generated some of the possible scenarios from the smart card chip's SEM images (Figure 7).

4.1 Modification based Trojans

A hardware Trojan created by modifying a standard digital logic cell or a custom logic cell at the doping level. An adversary can replace a logic gate by another logic gate or a custom logic, change the number of inputs, resize a logic gate, split a P well into a P well and N well [12].

- **NAND to NOR** A hardware Trojan implemented by replacing a NAND gate with a NOR gate or vice versa to implement a logic of their own interest.



(a) (b) (c) (d)

Fig. 6 Trojan Scanner: FFT Filtered Images (a) Golden IC Image, (b) IUA Image, with changes from golden IC highlighted, (c) Binarized & Gaussian filter image and (d) Highly suspected areas detected by optimized SSIM algorithm.

- **NAND to A.B+C (or any custom logic)** Similarly an adversary can replace a logic gate with their own custom designed logic cell, which upon triggering can change the functionality of the circuit.
- **Narrowing Power /Ground lines** Narrowing important interconnects is also another example of Trojan (EthernetMAC10GE-T400/T500) [23, 25]

4.2 Insertion - Deletion based Trojans

Another way to create hardware Trojan is to insert or delete a logic gate or any active layer component. Although it is very hard to find an empty space inside a chip, so a smart adversary can replace filler cells or MOS de-coupling capacitors to insert Trojan.

- **Inverter** Insertion of chain of inverters has proposed as a Trojan (RS232-T1800) [23, 25]
- **Capacitor** Yang et. al has [34] designed a circuit that uses capacitors to siphon charge from nearby wires as they transition between digital values. Then by using the fully charged capacitor to deploy an attack that forces a victim flip-flop to the desired value. A capacitor can be implemented by using a MOSFET that utilizes a polysilicon & silicon doping region as a parallel plate and thin oxide layer as a dielectric of the capacitor.

All other possible cases of hardware Trojans which can change either active layer or metal later are summarized in Table 2.

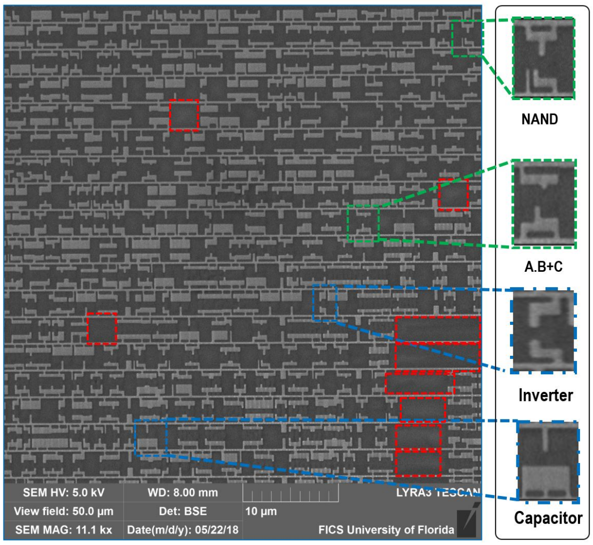


Fig. 7 Smart card doping level Footprints (Red contour areas show empty spaces in the chip, green contour depicts the Trojans implemented by modification of a logic gate and blue area shows Trojans added by insertion.)

5 Results

The high-quality SEM image with dwelling time 3.2 *μs/pixel* has selected as a golden IC image. We captured another three data sets of the IUA images with different imaging parameters i.e. 32μs/pixel, 10μs/pixel and 3.2 μs/pixel while other parameters are kept same. These IUA images are edited to implement the above-mentioned case of hardware Trojans (modification, insertion and deletion). Based on above-mentioned image processing techniques these golden and IUA images are enhanced to remove noise and segmented for feature detection. After image enhancement, they are compared with a golden IC image using SSIM algorithm. Figure 9(b), 8(b) and 8(c) shows the results of comparison with the golden IC images, the white spots mark the suspected areas of change and possible places of hardware Trojans. We also notice false positive detections in Figure 8(b) & 8(c) due to noise in IUA images are typically

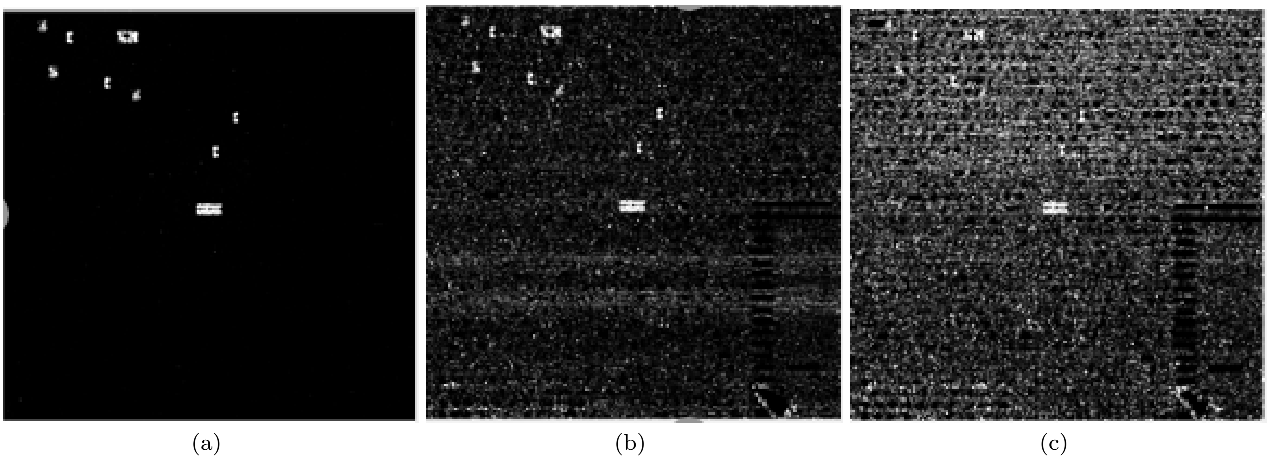


Fig. 8 Results of IUA image comparison with a golden image captured at different dwelling time (a) 32 μs/pixel, (b) 10 μs/pixel and (c) 3.2μs/pixel respectively.

Table 2 Trojan Footprints on IC

Size of Change	Change Type	Example	Footprint
Smallest	Modification	NAND ↔ NOR	Active Region
↓		NAND → A.B+C (or any custom logic)	Active Region
		Splitting active P well → P + N well	Active Region
		Changing number of inputs	Active Region
		Resizing 1x → 2x	Active Region
		Interconnects / Power / GND - Thinning	Metal Layer 1
	Camouflage Cells	NOR ↔ NAND	Metal Layer 1
	Insertion/ Deletion	Invertor NOT	Active Region
		NAND / NOR	Active Region
Biggest		Capacitor	Active Region

small spot consisting of only a few pixels. It is clear that the quality of change detection (hence suspected areas of Trojans detected) is correlated with the quality (noise) of the IUA image used. These false positives can be further removed (Figure 9) by proper eroding [17].

5.1 Comparison: Trojan Scanner vs. Other techniques

Tables 1 and 2 show the comparison results between our Trojans Scanner and full reverse engineering method (Table 3) as well as with other electrical test techniques to detect Trojans respectively (Table 4).

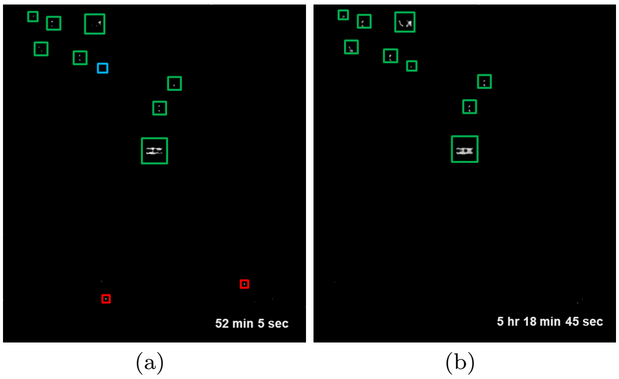


Fig. 9 Results after optimization of false positives (a) Golden 32μs /pixel: IUA 3.2μs /pixel and (b) Golden 10μs /pixel: IUA 3.2μs /pixel image comparison. (Contours: Green - correct detection, Red - False positives, Blue - false negative).

Table 3 Full Chip Reverse Engineering vs. Trojan Scanner

Metric	Reverse Engineering	Trojan Scanner
# of samples required	50 - 100	1
Trojans detected	All (except reliability)	All (except reliability & parametric)
Processing time	Months	Hours
Image Processing (Functionality extraction)	Required	Not required
Polygon Extraction	Required	Not required

5.2 Confidence level & 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 users 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) [3]. Based on these quality limits, the client has the

Table 4 Trojan Scanner vs. Electrical Tests

Trojan Type	Logic Test	Power SCA	Delay SCA	Run Time	Trojan Scanner
Functional	Maybe	Maybe	Maybe	Maybe	Yes
Parametric	No	Yes	Yes	No	In Future
Big	Maybe	Yes	Maybe	Yes	Yes
Small	Yes	No	Yes	Maybe	Yes
Tight	Yes	Yes	Yes	Maybe	Yes
Loose	Yes	Maybe	Yes	Maybe	Yes

inspection sample size to make an informed decision to accept or reject the lot [6]. 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 [6], 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 don't 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.

5.3 Challenges

Although Trojan Scanner is efficient enough to detect changes at the doping layer, still there are some challenges that need to be addressed:

- **Camouflage cells.** A camouflage cell has been used by designers as a countermeasure to reverse engineering, as it mimics the original logic cell [22]. For example, a NOR and NAND camouflage cells look similar on the active region and metal-1 layer. Only the detection of via contact between a metal layer-1 and active layer can differentiate between a NOR and NAND gate. So, if an adversary can replace a NAND with a NOR camouflage cell, we need to modify our technique by incorporating high kV SEM imagining or possibly using a Focused Ion Beam (FIB) to delayer the die to detect a via contact.

- **Design for Manufacturability (DFM).** Besides the Design rule check (DRC) performed by the design house, the foundry can optimize the circuit layout to increase the yield. To the best of our knowledge [10], the foundry optimizes the interconnects routing, create/remove new contacts to avoid open / shorts in design. In some cases, they may move the logic or resize logic cell. In this case, it will be a good assumption that they will convey these changes to the design house so that we can account for DFM in Trojan Scanner.

6 Conclusion and Future work

Current hardware 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 proposed and demonstrated a backside imaging method combined with advanced computer vision techniques to physically inspect chips and detect hardware Trojans by using ”Trojan Scanner”. We demoed possible scenarios of Trojan insertion and the detection approach by comparing SEM images of a golden IC with an IC under authentication. We observed during developing our technique, there is a trade-off between the accuracy of detection and SEM parameters (dwelling time, FoV). We finally discussed the confidence level of Trojan Scanner and the minimum number of ICs required to establish a trust in the supply chain.

Our future work will be focused on using more advanced computer vision techniques in combination with machine learning to detect Trojans even using lower quality SEM images which is faster to acquire. Also, for many real case scenarios the golden IC is not available, but designers have access to their design layout. (Trojans inserted by third-party IP is out of the scope of this study). So we are extending this work to detect hardware Trojan by comparing GDSII layout with SEM image as well.

References

1. Our technologies bridge the gap in the micron scale. URL https://www.varioscale.com/variomill
2. The navy bought fake chinese microchips that could have disarmed u.s. missiles (2011). URL http://www.businessinsider.com/navy-chinese-microchips-weapons-could-have-been-shut-off-2011-6
3. What is the aql (acceptance quality limit) in simple terms? (2011). URL https://qualityinspection.org/what-is-the-aql/
4. Semi industry fab costs limit industry growth (2012). URL https://www.eetimes.com/document.asp?doc_id=1264577
5. Samsung breaks ground on $14 billion fab (2015). URL https://www.eetimes.com/document.asp?doc_id=1326565
6. Acceptable quality limit (aql) (2018). URL https://www.asiainspection.com/aql-acceptable-quality-limit

7. Aarestad, J., Acharyya, D., Rad, R., Plusquellic, J.: Detecting trojans through leakage current analysis using multiple supply pad. IEEE Transactions on information forensics and security **5**(4), 893–904 (2010)
8. Adee, S.: The hunt for the kill switch–are chip makers building electronic trapdoors in key military hardware? the pentagon is making its biggest effort yet to find out. IEEE Spectrum (2008)
9. Agarwal, M., Paul, B.C., Zhang, M., Mitra, S.: Circuit failure prediction and its application to transistor aging. In: VLSI Test Symposium, 2007. 25th IEEE, pp. 277–286. IEEE (2007)
10. Aitken, R.: Dfm metrics for standard cells. In: Proceedings of the 7th International Symposium on Quality Electronic Design, pp. 491–496. IEEE Computer Society (2006)
11. Bao, C., Forte, D., Srivastava, A.: On application of one-class svm to reverse engineering-based hardware trojan detection. In: Quality Electronic Design (ISQED), 2014 15th International Symposium on, pp. 47–54. IEEE (2014)
12. Becker, G.T., Regazzoni, F., Paar, C., Burleson, W.P.: Stealthy dopant-level hardware trojans. In: International Workshop on Cryptographic Hardware and Embedded Systems, pp. 197–214. Springer (2013)
13. Bhunia, S., Hsiao, M.S., Banga, M., Narasimhan, S.: Hardware trojan attacks: threat analysis and countermeasures. Proceedings of the IEEE **102**(8), 1229–1247 (2014)
14. Chakraborty, R.S., Wolff, F., Paul, S., Papachristou, C., Bhunia, S.: Mero: A statistical approach for hardware trojan detection. In: Cryptographic Hardware and Embedded Systems-CHES 2009, pp. 396–410. Springer (2009)
15. Courbon, F., Loubet-Moundi, P., Fournier, J.J., Tria, A.: A high efficiency hardware trojan detection technique based on fast sem imaging. In: Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, pp. 788–793. EDA Consortium (2015)
16. Courbon, F., Loubet-Moundi, P., Fournier, J.J., Tria, A.: Semba: A sem based acquisition technique for fast invasive hardware trojan detection. In: Circuit Theory and Design (ECCTD), 2015 European Conference on, pp. 1–4. IEEE (2015)
17. Gonzalez, R.C., Woods, R.E.: Digital image processing. Pearson (2018)
18. Jin, Y., Makris, Y.: Hardware trojan detection using path delay fingerprint. In: Hardware-Oriented Security and Trust, 2008. HOST 2008. IEEE International Workshop on, pp. 51–57. IEEE (2008)
19. Krig, S.: Computer vision metrics. Springer (2016)
20. Narasimhan, S., Wang, X., Du, D., Chakraborty, R.S., Bhunia, S.: Tesr: A robust temporal self-referencing approach for hardware trojan detection. In: Hardware-Oriented Security and Trust (HOST), 2011 IEEE International Symposium on, pp. 71–74. IEEE (2011)
21. Quadir, S.E., Chen, J., Forte, D., Asadizanjani, N., Shahbazmohamadi, S., Wang, L., Chandy, J., Tehranipoor, M.: A survey on chip to system reverse engineering. ACM journal on emerging technologies in computing systems (JETC) **13**(1), 6 (2016)
22. Rajendran, J., Sam, M., Sinanoglu, O., Karri, R.: Security analysis of integrated circuit camouflaging. In: Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, pp. 709–720. ACM (2013)
23. Salmani, H., Tehranipoor, M., Karri, R.: On design vulnerability analysis and trust benchmarks development. In: Computer Design (ICCD), 2013 IEEE 31st International Conference on, pp. 471–474. IEEE (2013)
24. Salmani, H., Tehranipoor, M., Plusquellic, J.: A novel technique for improving hardware trojan detection and reducing trojan activation time. IEEE Transactions on Very Large Scale Integration (VLSI) Systems **20**(1), 112–125 (2012)
25. Shakya, B., He, T., Salmani, H., Forte, D., Bhunia, S., Tehranipoor, M.: Benchmarking of hardware trojans and maliciously affected circuits. Journal of Hardware and Systems Security **1**(1), 85–102 (2017)
26. Shiyanovskii, Y., Wolff, F., Rajendran, A., Papachristou, C., Weyer, D., Clay, W.: Process reliability based trojans through nbti and hci effects. In: Adaptive Hardware and Systems (AHS), 2010 NASA/ESA Conference on, pp. 215–222. IEEE (2010)
27. Tehranipoor, M., Koushanfar, F.: A survey of hardware trojan taxonomy and detection. IEEE design & test of computers **27**(1) (2010)
28. Tehranipoor, M., Wang, C.: Introduction to hardware security and trust. Springer Science & Business Media (2011)
29. Torrance, R., James, D.: The state-of-the-art in ic reverse engineering. In: Cryptographic Hardware and Embedded Systems-CHES 2009, pp. 363–381. Springer (2009)
30. Wang, X., Tehranipoor, M., Plusquellic, J.: Detecting malicious inclusions in secure hardware: Challenges and solutions. In: Hardware-Oriented Security and Trust, 2008. HOST 2008. IEEE International Workshop on, pp. 15–19. IEEE (2008)
31. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
32. Wolff, F., Papachristou, C., Bhunia, S., Chakraborty, R.S.: Towards trojan-free trusted ics: Problem analysis and detection scheme. In: Proceedings of the conference on Design, automation and test in Europe, pp. 1362–1365. ACM (2008)
33. Xiao, K., Forte, D., Jin, Y., Karri, R., Bhunia, S., Tehranipoor, M.: Hardware trojans: Lessons learned after one decade of research. ACM Transactions on Design Automation of Electronic Systems (TODAES) **22**(1), 6 (2016)
34. Yang, K., Hicks, M., Dong, Q., Austin, T., Sylvester, D.: A2: Analog malicious hardware. In: Security and Privacy (SP), 2016 IEEE Symposium on, pp. 18–37. IEEE (2016)
35. Zhang, G., Das, D., Xu, R., Pecht, M.: Iddq trending as a precursor to semiconductor failure. In: Prognostics and Health Management, 2008. PHM 2008. International Conference on, pp. 1–7. IEEE (2008)
36. Zhang, X., Tehranipoor, M.: Case study: Detecting hardware trojans in third-party digital ip cores. In: Hardware-Oriented Security and Trust (HOST), 2011 IEEE International Symposium on, pp. 67–70. IEEE (2011)
37. Zhou, B., Adato, R., Zangeneh, M., Yang, T., Uyar, A., Goldberg, B., Unlu, S., Joshi, A.: Detecting hardware trojans using backside optical imaging of embedded watermarks. In: Design Automation Conference (DAC), 2015 52nd ACM/EDAC/IEEE, pp. 1–6. IEEE (2015)
38. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and vision computing **21**(11), 977–1000 (2003)