Hardware Trojan Detection Using Backside Optical Imaging

Boyou Zhou, *Student Member, IEEE*, Aydan Aksoylar, Kyle Vigil, Ronen Adato, Jian Tan, Bennett Goldberg, M. Selim Ünlü, *Fellow, IEEE*, and Ajay Joshi, *Senior Member, IEEE*

***Abstract*— The high cost of integrated circuit chip production has driven more and more chip design companies to use overseas production services. Since the integrated circuit production cannot be closely monitored, the security of integrated circuit chips has become a major concern. Hardware Trojan (HT) insertion is one type of the hardware attack. HTs are extremely stealthy due to their small sizes and low triggering rates. HTs inserted during manufacturing can have minimum impact on the timing and power. In fact, this impact can be smaller than the timing and power variations caused by the process variations. Therefore, these HTs cannot be easily detected using traditional electrical methods. In this article, we propose a novel optical method, where we image the integrated circuit chip from the backside. Our method, can easily detect any replacements, modifications, or rearrangements of fill cells or functional cells for HT insertion. We use a noise-based detection method to achieve high HT detection rates in different testbenches. To further improve the robustness of our method, we strategically place high reflectance fill cells in the designs. Our approach provides high-resolution, nondestructive, and rapid means to detect HTs inserted during fabrication. We evaluate our approach using various hardware blocks where the HTs can occupy less than 0.1% of the total area or consist of fewer than three gates. In addition, we analyze our method with various magnitudes of noise, process variations, detection window sizes, and resolutions.**

***Index Terms*—Hardware security, hardware Trojan (HT) detection, near-infrared (IR) imaging, optical imaging.**

I. Introduction

INTEGRATED circuits (ICs) are the root of the trust in computing systems. However, today this trust can be broken. The large demands for IC chips has led to the globalization of the IC chip supply chain. Over the past two decades, IC design and manufacturing has become increasingly distributed across the world [1]. The standard IC chip production process consists of specification, design, fabrication, testing, and packaging phases [2]–[4]. Many U.S. companies still design their IC chips within U.S. However, given that overseas manufacturing can support a broad spectrum of technology nodes at a significantly lower cost, these IC companies prefer to fabricate their chips overseas.

Highly fragmented and distributed production brings efficiency and productivity to IC design and fabrication [5]. However, during the different phases of IC chip production, the IC chips face threats from hardware Trojans (HTs) [1], [4], [6]–[8], IP privacy [9], IC chip overbuilding [10], reverse engineering (RE) [11]–[16], side-channel analysis [17]–[19], [19]–[21], and IC counterfeiting [22]. HTs are malicious modifications or insertions of unwanted circuitry into the chip designs to sabotage the functionality or leak secret information. IP privacy violation discloses the intellectual properties of IC designs and causes economic damages to the IC company. Foundries can overbuild ICs and sell them to make illegal profits. Reverse engineering recovers designs of the chips to steal intellectual property. Side channel analysis focuses on using physical properties other than electrical signals on the chip to extract secret information from physical devices. IC counterfeiting refers to fabricating unqualified products but labeling them as products from other companies, which causes indirect loss to the companies that originally designed the chip. Given the pervasive use of the IC chips in consumer and industrial domains, it is mandatory to ensure that the security of these chips has not been compromised.

Among all the threats mentioned above, HT attacks are the worst attack among the above-mentioned threats. HTs can control, modify, disable or monitor the IC chips [22] in order to achieve information leakage, system compromise, and failure. Common HT insertion approaches are embedding the malicious IP cores, design modifications, and layout modifications during the fabrication phase [22]. We can detect the first two HT insertion approaches with functional verifications during the IC chip design process. In the third type of HT insertion, the attackers can reverse engineer the physical layout and modify the design during manufacturing. These HTs can be optimized to minimize the impact on timing and power of the extracted layout, and also they are extremely hard to detect. Besides, the attacker can design the HT trigger driven by leakage current [23], which can never been triggered during the post-silicon functional testing. Unless manufacturing

Manuscript received September 3, 2019; revised January 14, 2020; accepted April 7, 2020. Date of publication April 30, 2020; date of current version January 11, 2021. This work was supported in part by the Boston University, and in part by NSF under Grant ECCS-1641018. This article was recommended by Associate Editor Y. Jin. *(Corresponding author: Boyou Zhou.)*

Boyou Zhou is with Analog Garage, Analog Devices Inc., Santa Clara, CA 95054 USA. (e-mail: bobzhou@bu.edu).

Aydan Aksoylar is with the Autonomous Systems Group, Microsoft Corporation, San Francisco, CA 94103 USA.

Kyle Vigil is with Physics Department, Boston University, Boston, MA 02215 USA.

Ronen Adato is with Patent Agent at Choate, Hall & Stewart LLP, Boston, MA 02210 USA.

Jian Tan is with Teledyne LeCroy, Chestnut Ridge, NY 10977 USA.

Bennett Goldberg is with the Office of the Provost, Northwestern University, Evanston, IL 60208 USA.

M. Selim Ünlü and Ajay Joshi are with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA.

Digital Object Identifier 10.1109/TCAD.2020.2991680

is closely monitored, the IC design company needs to use expensive RE techniques like delayering the chip, SEM imaging, etc. to check for HTs. Hence, there is a pressing need for a new detection method, which can detect and locate HTs inserted during fabrication in a fast, accurate, and robust manner [24].

We propose a novel backside imaging method that can rapidly and accurately detect HTs inserted during the fabrication stage. We propose to use the optical responses from post-fabricated ICs, and compare the results to the responses from finite difference time domain (FDTD) simulations performed pretapeout on the chip. We can detect any compromises to the chip by identifying differences between the generated responses and the measured results. The contributions of this article are as follows.

1) *Backside Imaging Using Near-IR Light:* We propose a new approach to generate a backside image of an IC chip, which we call the "golden reference," using the detailed physical layout of the IC chip and FDTD simulations of individual standard cells. We compose the reference responses by mapping individual gate FDTD responses to the locations of standard cells and interpolate the response image to form the final result.
2) *Noise-Based Detection Method:* We implement a noise-based detection method that compares the imaged data and golden reference, which is improved from our correlation-based detection method The noise-based detection exhibits robustness against the additive white Gaussian noise (AWGN), and yields a higher detection rate compared to correlation-based detection method under the same SNR ratios.
3) *Engineering High Reflectance Fill Cells:* We engineer fill cells as the highest reflectance cells and strategically insert them before placing the functional cells of the digital designs to increase the contrast of imaging responses. We demonstrate that the insertion of extra fill cells improves the HT detection accuracy against noise.
4) *Evaluations Using Various Hardware Blocks:* We show the effectiveness of our approach through a variety of testbenches. We use standard *Cadence* tool flows and *Nangate* 45 nm technology to synthesize and place & route circuits from academic benchmarks. The area of inserted HT varies from 0.1% to 12% of the IC chip sizes. We use our method to detect the HTs in these testbenches under different signal to noise ratios. We also performed sensitivity analysis on the effects of process variations, the resolutions in imaging, and the detection window frame sizes.

II. Threat Model

We assume that the attacker can modify, shift or replace the fill cells in order to accommodate the malicious hardware blocks in the victim's design. The attacker can get access to the GDSII files used for fabrication, but cannot change the RTL, gate-level designs or any designs before the victim generates the GDSII files. We trust the IP blocks in the designs from any third party to be HT free.

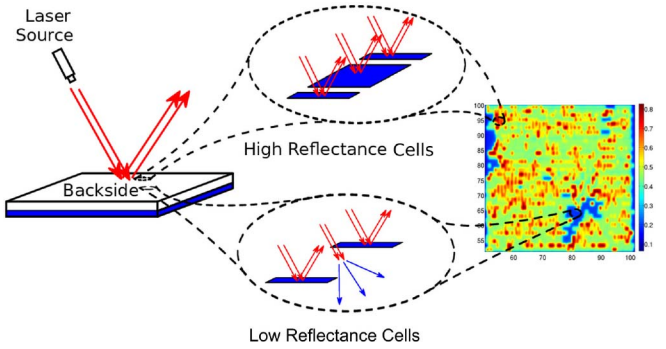

Fig. 1. Backside imaging of IC chips.

III. Backside Imaging of IC Chips

We propose a dramatically different approach to detect HTs inserted in IC chips during the fabrication phase. Metal in IC chips is strongly reflective to near-infrared (IR) light, while silicon is transparent to light at those wavelengths. Backside optical imaging of the fabricated chip enables us to extract the full standard cell layout of the chip with the watermarks, which in turn can be validated to detect any modifications to the IC layout (see Fig. 1). In addition to the original metal lines of standard cells, we also embed a maximal amount of metal in the M1 layer within the limit of the cell design specification in the fill cells to increase the total reflectance of the design. This strong reflected pattern forms the unique signature of chip design.

Our optical measurements are highly modular and independent of the multitude of connections in the full IC chip. In addition, the physical principles behind the implementation of our watermarking scheme are also highly distinct from previous approaches. Although, like the first PUF [25], it is an optical method that utilizes embedded scatterers, their intended design and functionality are fundamentally different. The scattering in [25] was explicitly designed to be random and impossible to predict and replicate. In contrast, determinism is essential to the functionality of our approach. We embed the watermark during the design phase and we can determine the optical response before fabrication. Our designed signatures can only be imaged within our analytical spectrum.

A. Optical Imaging Method

As the opaque metal layers prohibit front side imaging, backside imaging of integrated circuits is a well-established technique for failure analysis [26]–[28]. Bright-field images at near-IR wavelengths (e.g., $\lambda \sim 1 - 2\,\mu$m) can be used for passive measurements for fault detection and localization, such as for inspecting the fidelity of the metal wires [29]. In addition to passive measurements, the active functionality of the circuit can also be probed via techniques such as thermal imaging [26], which records power dissipation via heat generation or laser-voltage imaging (LVI) [27], which records the switching response of the transistors.

As ICs shrink in size, a key challenge in these imaging techniques is to obtain sufficient spatial resolution to resolve each individual structure in each gate. The diffraction limit imposes

a fundamental restriction on the maximum spatial frequency that can be imaged using conventional optical systems which limits the resolution. The resolution of an optical system can be determined by the size of the impulse response of the system which takes the form of an Airy function [30]

$$I(\rho) \propto \left[2\frac{J_1(2\pi\rho)}{2\pi\rho}\right]^2 \tag{1}$$

where $\rho = \mathrm{NA}r/(M\lambda)$ is the image space coordinate. The size of the impulse response is

$$\Delta x = 0.61\frac{\lambda}{\mathrm{NA}}. \tag{2}$$

Here, λ is the wavelength of the light in free space and NA is the numerical aperture ($\mathrm{NA} = n\sin\theta$) of the system, where n is the refractive index of the material in which the light propagates. J_1 is bessel function of the first kind.

Due to their high NA capability (∼3.4), complex solid immersion lenses provide a high resolution for fault analysis of integrated circuits [28]. In this article, we can eliminate the need for the high resolution and can rapidly and accurately detect malicious tampering and the presence of an HT at relatively low NAs (∼0.8).

Our approach is based on the fact that for low NAs in the near-IR region, we can achieve impulse response functions with widths on the order of the gate size in 45 nm or lower technology nodes [see (2)]. NAs of 0.14, 0.42, and 0.5 correspond to spot sizes of approximately 4.6 μm, 1.5 μm, and 1.3 μm, respectively, at $\lambda = 1.064$ μm. These correspond to common near-IR commercial objectives capable of imaging over 0.1–1 mm fields of view (several thousand to half a million gates simultaneously). An image collected in this manner (i.e., at low NA, without a solid immersion lens) would, rather than resolve the detailed substructure of individual gate, produce a slowly varying image that tracks the average reflectance of each gate over its area. Although the individual gates are comprised of unique layouts of metal lines, Fig. 2 shows that with both 15 nm and 45 nm technology, the responses of three different gates have distinctive signatures across the spectrum. Fig. 2 shows that fill cells that are engineered to contain maximum amount of metal, while meeting the metal density design rule constraints, achieve distinguishably stronger response compared to common functional cells. We can leverage these stronger responses to strengthen the contrast of the response images in order to achieve higher HT detection rates.

Therefore, backside imaging of IC chips can result in clear patterns depending on the standard cell layout. These patterns can serve as a robust, easily record-able optical watermark of IC chip. Any modifications through movement of cells or insertion of unwanted cells will result in a change in the watermark that can be measured with high fidelity. Imaging large fields of views provides the potential to perform these measurements on a large number of gates simultaneously. It also has the added benefit of considerably simplifying the required optical setup in comparison with commercial failure analysis tools. Our approach is a simple, rapid test to check for tampering of the IC chip at the fabrication stage, and it utilizes off-the-shelf tools to embed our engineered signature.

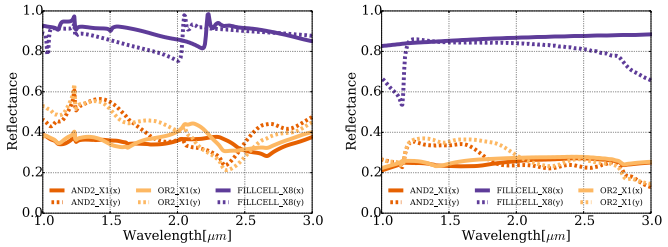

Fig. 2. (a) Reflectance spectrum of functional gates and fill cells designed in 45 nm technology (computed via FDTD simulations). The response is computed for both X and Y polarizations of the illuminating field (solid and dashed lines, respectively). For X polarization, the incident electric field is polarized along the VDD and VSS rails. For Y polarization, the polarization is perpendicular. (b) Reflectance spectrum of functional gates and fill cells designed using 15 nm technology. Fill cells still show much higher reflectance compared to functional cells.

B. Technology Scalability

While we demonstrate our approach at the 45-nm technology node level, we expect it to scale well to newer technology nodes because our technique *does not* require high resolution imaging. We can estimate the efficacy of our proposed technique as gate sizes are reduced by considering the size of the impulse response of our optical system as described by (2). The gate size-reduction directly affects our ability to detect a shift in the position or a change in the magnitude of the responses, and therefore the performance of our technique. In the older technologies, such as 90 nm or 130 nm, our imaging results have a more pronounced difference between different gates compared to smaller technology nodes. Given that the reflectance from each gate is proportional to the amount of metal inside of the gate, our methodology can detect the HTs with higher fidelity as the older technology nodes will have larger metal content in each gate.

We can show that the reflectance difference between different gates in optical imaging can be clearly distinguished. Case in point, in a 45-nm technology node [see Fig. 2(a)], the reflectance difference between AND gate and OR gate can be as much as 0.15 (with the wavelength of 1.3 μm). Replacement of AND gate with an OR gate results in 50% change in reflectance variation. In contrast to the AND gate and OR gate, the reflectance of fill cells is as high as 90%. A similar distinguishable difference between various gates is obscured in a 15-nm technology node [see Fig. 2(b)]. The gates in the 15-nm technology node have smaller sizes, and also the reflectance of all the gate is weaker compared to gates in the 45-nm technology node. However, the reflectance ratio between AND gate and OR gate still remains the same as that in the 45-nm technology node. At the same time, reflectance of fill cells is as high as 87%. By comparing the relative reflectance, we can detect the changes in the gate type. Moreover, in our method, the uniqueness of the image of a design is determined using the responses from both the individual gates and the neighbors. The pixel size of our optical imaging method is 0.1 $\mu\mathrm{m}^2$ while the smallest gate size of 45 nm technology node is 0.4 $\times 1$ $\mu\mathrm{m}^2$. As a result, each gate has multiple sampling points for the HT detection. In our

case, the minimum gate size of 45 nm technology is 0.26 μm^2, which can contain as much as 30 data points. Even with 15-nm technology, each gate can still have up to $\sim$ 15 data points.

By applying interpolations on sampled points, our technique also scales well in modeling areas much larger than single gate. This enables measurements over a large field of view and therefore a large number of gates can be simultaneously imaged. Considering an IC with $\sim$ 1 B transistors, an average of four transistors per gate corresponds to approximately 250 M gates. Our method is a significant improvement over the high NA imaging methods, where the sample is typically scanned with a scan size of 10–100 nm as opposed to 400 nm that we consider in this article. Compared to other equally accurate methods, such as SEM scanning of ICs, our method does not require any delayering of the chip that can take several days. So our method is several times faster. As part of the follow-up work, we are developing novel imaging techniques to reduce the imaging time down to minutes. The time for the model generation, data processing, and results evaluations takes a few minutes.

C. Optical Simulation Methodology

The FDTD method is widely used in modeling different microscopes, and it has been shown to work both analytically and experimentally [31], [32]. FDTD computation for one gate area requires roughly half an hour computation time. The computation time grows exponentially with area, i.e., a larger gate with area of 5-6 times of single gate takes 1-2 days to finish. Simulating the response for the whole layout of a typical 3-4 cm^2 IC chip is computationally infeasible. Thus, in this article, we simulate each individual gate to construct a library of responses from near-IR excitement. The constructed response library determines the response of a given layout through simulations. Our method does not require the need for in-field measurements of golden reference.

In the previous work, we considered the FDTD simulations for each gate with periodic boundary conditions by simulating illumination on infinite tiling of the same gate [33]. With these boundary conditions, we simulated the response of each gate separately and obtained their optical response by calculating the reflected power of each gate, normalized by the source power. Since the size of the impulse response is on the order of the size of the gates, the normalized power was then convolved with the impulse response of the optical system to simulate the image of a chip. The periodic boundary conditions allowed us to obtain the optical response at each wavelength simultaneously, at a reduced computation time. We considered only the six basic gate types (AND, OR, NAND, NOR, XOR, XNOR) in the evaluation of our optical watermarking technique. In this article, we consider the complete gate library, which includes gates that are larger than the impulse response. We extract the internal metal structures and contacts from back-end GDS files and rectilinearly decompose polygon structures into rectangular structures for FDTD calculations. In order to calculate the optical response of those gates, we develop a rigorous simulation method for gates whose sizes are larger than the illumination spots. For this purpose, instead of using plane-wave illumination with the same gate as the boundaries, we use other gates as boundaries of the illuminated gate and use a focused beam with a predetermined NA and perfectly matched layer (absorbing) boundary conditions. The optical response of each gate is obtained by normalizing the power of the reflected light confined within the predetermined NA by the source power. In addition, since we are using a focused light illumination, the focused spot needs to be scanned at multiple locations as opposed to the case with plane-wave illumination. Selecting the scan size to be the same as the size of the impulse response allows us to obtain all information within a gate, while ensuring a rapid acquisition time.

We considered the inclusion of active regions and polysilicon transistor gate layers in our FDTD simulations. In modern CMOS process, the polysilicon and diffusion layers are metalized through a silicide process. However, these metalized layers are 5–10 nm thick, and therefore their contributions toward reflection is negligible. Thus, we did not include those layers in our simulations. As some of the gates are larger than the size of the impulse response, the focused spot is scanned starting from the center of the gate and then moving to the sides with the specified step size until the majority (more than 50%) of the illuminated area does not overlap with the gate. Note that given a layout, placing the responses from the constructed library for each gate at its location in the layout results in a set of response values on a irregular grid. Since the experimental measurements are on a regular grid with predetermined scan sizes, we interpolate the simulated data to a regular grid using bi-cubic interpolation [34]. To construct the overall layout on a regular grid, we set the horizontal step equal to the size of the impulse response, and the vertical step equal to the height of the gates.

IV. HT Detection Process

In this section, we use the AES-T100 circuit from Trust-HUB [35] to explain our proposed method of HT Detection using near-IR imaging. To detect the HT in the example circuitry, we need to: 1) generate golden reference from simulations (see Section IV-A) and 2) cross-compare the measured response and the golden reference to detect HTs in the chip (see Section IV-B). We present the results in Section IV-C.

A. Methodology to Generate Golden Reference

To generate the golden reference, we synthesize, floor plan and place&route a digital block, using *Cadence* RC and Encounter tools. From the final fabrication-ready Geometrical database for information exchange (GDSII) file, we extract the locations and orientations of each gate and export the geometric information of the gates to a design exchange format (DEF) file. We simulate the near-IR response of each gate in the standard cell library and create a collection of gate responses. According to the geometric information of the design, we map the FDTD calculated responses of individual gates into the locations and orientations of the corresponding gates. The mapping of the optical responses generates the optical response from the original design file in Fig. 3(a) of

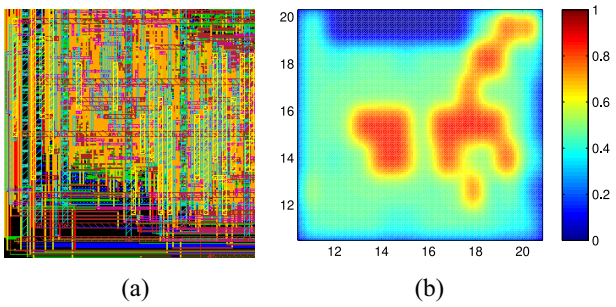

Fig. 3. (a) Physical layout of a 10 μm × 10 μm region of the AES-T100 hardware block. (b) Backside image (reflectance value) of the 10 μm × 10 μm region. The fill cells have the highest reflectance (red areas).

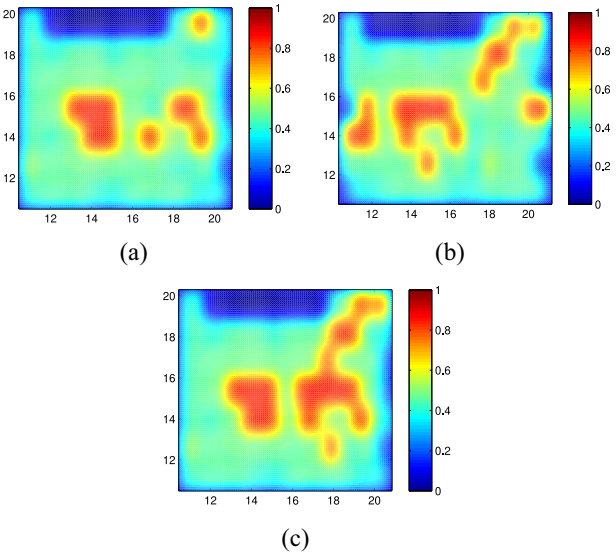

Fig. 4. (a) Backside image (reflectance value) of the 10 μm × 10 μm region where the fill cells are replaced with functional gates that constitute the HTs. (b) Backside image (reflectance value) of the 10 μm × 10 μm region where the bottom 3 rows are shifted by 5 μm to the left to make room for cells that constitute the HT. (c) Backside image (reflectance value) of the 10 μm × 10 μm region where the functional cells are replaced by a different set of functional cells that constitute the HT.

the AES-T100 circuitry. To form the image in Fig. 3(b), we convolve the reflectance map of the AES-T100 circuit with an impulse response from near-IR imaging.

In our attack model, an attacker needs to replace, remove, or shift fill cells to make room for HTs [25]. Any replacements or shift will result in a different optical image compared to the one we generated from the original design, because fill cells are significantly different from functional cells. The stark difference between the fill cells and other cells forms the signatures of the design. The response from the fill cells also increase the qualities of the neighboring cell responses. Convolved with the fill cell responses, the overall response of these neighboring cells will change if any of these fill cells are replaced with other functional cells. We applied these properties of our engineered fill cells to increase our HT detection rates.

Fig. 4(a) and(b) shows the optical response image when an attacker replaces or shifts fill cells, respectively. The optical response image is different compared to the optical response

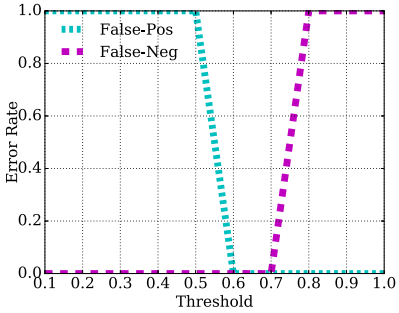

Fig. 5. Monte-Carlo simulation results of detection error rate against detection threshold. The optimized detection threshold should be 0.65 for an SNR of 10.

image of the design without HTs in Fig. 3(b). Any modifications or shifts of these fill cells (red areas in Fig. 4) are more prominently observed in the near-IR image than any modification/shifts of the nonengineered gates (blue and green areas in Fig. 4). Fig. 4(c) shows an example of replacing functional cells with functional cells of HTs. The responses from engineered fill cells significantly increase the contrast of the overall image. Although the modifications in Fig. 4(c) are not observable by comparing the responses, we can compute the differences between measured results and the golden reference using signal processing techniques to identify the HT inserted ICs from nontampered ICs.

B. Correlation Method

To reliably and accurately detect HTs in large systems requires an automation of the comparison between golden reference and measured results. In our previous work [33], we use 2-D-correlation coefficients as the threshold metric for the HT detection.

In image processing, the correlation coefficient describes the similarity between two images. Here, we use the correlation coefficient to compute the differences between images. If the correlation coefficient is higher than the threshold, we consider that the circuitry has enough similarity with the original design. If the correlation coefficient is lower than the threshold, we consider that the circuitry is different from the original design, which is tampered by HTs. We describe our 2-D correlation metric

$$\rho_{M_X M_Y} = \frac{\text{cov}(M_X, M_Y)}{\sigma_{M_X} \sigma_{M_Y}} \tag{3}$$

where cov is the co-variance of the two matrices M_x&M_y and σ is the standard deviation. Here, we use M_X to denote the near-IR imaged matrix and M_Y to represent the imaged matrix with noise.

In our previous work [33], in order to minimize false positives and false negatives in HT detection, we predetermined the threshold for the correlation value between the measured results and golden reference based on the Monte-Carlo simulations of AES-T100 circuit (AES with HTs from [35] in Fig. 5). Our simulations showed that given any HTs and the IC design, we can optimize our detection threshold at an SNR of 10 to maximize our detection rate. We minimized the false positives and false negatives and achieved 100% detection rates in

TABLE I
AREA OVERHEAD OF ALL THE TESTBENCHES USED IN THIS ARTICLE (IN μm^2)

Testbench	Area without Trojans [μm^2]	Trojan Area [μm^2]	Trojan Area Percentage (%)	Testbench	Area without Trojans [μm^2]	Trojan Area [μm^2]	Trojan Area Percentage (%)
c1355	266	4.522	1.7	PIC400	2969.9	292.1	9.83
c1908	251.6	0.798	0.317	AES100	274177.6	253.2	0.0923
c2670	333	0.798	0.240	AES200	274177.6	169.5	0.0618
c499	257	4.522	1.760	AES700	322705	297.4	0.0922
c880a	197.6	0.798	0.403	AES900	318359	267.3	0.0840
PIC100	4215.0	351	8.33	AES1000	274177.6	251.1	0.0915
PIC200	4215.0	89.6	2.13	AES1200	323348	450.3	0.1392
PIC300	4215.0	253.2	6.01	AES1700	320670	1388.3	0.4329

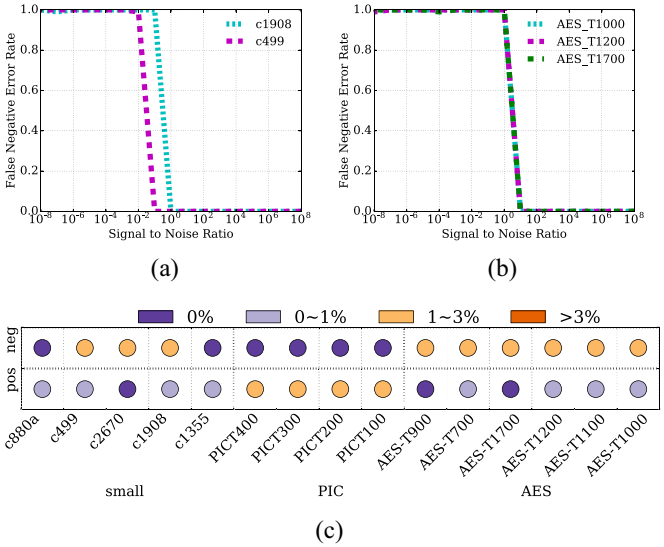

Fig. 6. (a) False-negative error rates versus SNR of c1908 and c499 testbenches from [36], and (b) AES-T1000, AES-T1200, and AES-T1700 circuits from [35] using correlation comparison method. (c) Summary of various testbenches from both [36] and [35]. We use colors in the legend to denote the error rates at SNR from 10^{-8} to 10^{-1} belonging to the corresponding range.

the given condition. In real-life, we can calibrate our optical responses to optimal noise levels, and optimize our threshold with the corresponding noise level to increase the detection rates. Our simulations were limited under all these conditions.

C. Results

We used testbenches from Trust-HUB [35] and [36] to evaluate our proposed approach. Using 45 nm Nangate library, we generated the GDS files for testbenches. The area of all testbenches is listed in Table I. Fig. 6(a) and (b) shows the error rates versus SNR for the testbenches from [36] and [35], respectively. The sizes of the HTs range from 0.06% to 9.83% of the total area in the testbenches. Fig. 6(c) presents the evaluations of our method using all the testbenches. Each dot in the Fig. 6(c) represents the error rate analysis of SNR varying from 10^{-8} to 10^{-1}. If all error rates in this SNR spectrum fall within the error rate range, we use the corresponding color dot to represent it. For example, all the false-negative rates of c880a testbench are 0% in the SNR ranging from 10^{-8} to 10^{-1}. We use a blue dot to denote it in the table. We use three other types of color dots to represent if all error rates are in $0 \sim 1\%$, $1 \sim 3\%$, and $> 3\%$, respectively. From this we can see that our correlation detection method presents higher detection rates in c880a and c1355 compared to other benchmarks. Our method has lower detection rates in PIC testbenches because in our method the threshold is optimized for designs where the area occupied by the HTs is less than 2%. These observations show that the correlation method requires predetermined HT sizes to achieve high detection rates.

V. NOISE-BASED DETECTION METHOD

In Section IV-B, we presented the correlation coefficient as the metric for HT detection. Correlation enables a quantitative, straight forward comparison between the reference and the imaged data. However, it suffers from two major drawbacks. First, there is no prior knowledge of the HT before performing the detection in most of the detection environments. A tester can only achieve high HT detection rate with given experimental conditions, including the HT sizes and SNR values, if and only if the testing process has this information, and the tester can optimize the HT detection threshold for further computation. A tester can estimate the noise level of the optical responses. However, the sizes of the HTs are still unknown in most cases. In practice, the limitations on the choosing the threshold hinder the flexibility and applicability of this approach. Second, in correlation-based detection method, we average out all our extracted information of the response from the ICs to one metric for comparison. Using averaged information limits our capabilities of differentiating details between the colored maps. One metric is suitable in concluding the overall image; yet it is not ideal for comparing minor differences between two images. In our case, we need to identify the individual gates in the layout. In the correlation method, the information that we gleaned from data sampling only contains means and variations of the overall results (the imaged matrix). To improve the detection method, we must extract the information to compare between the imaged results as much as possible. Hence, we propose a noise-based detection method to enhance robustness of our method. Noise-based

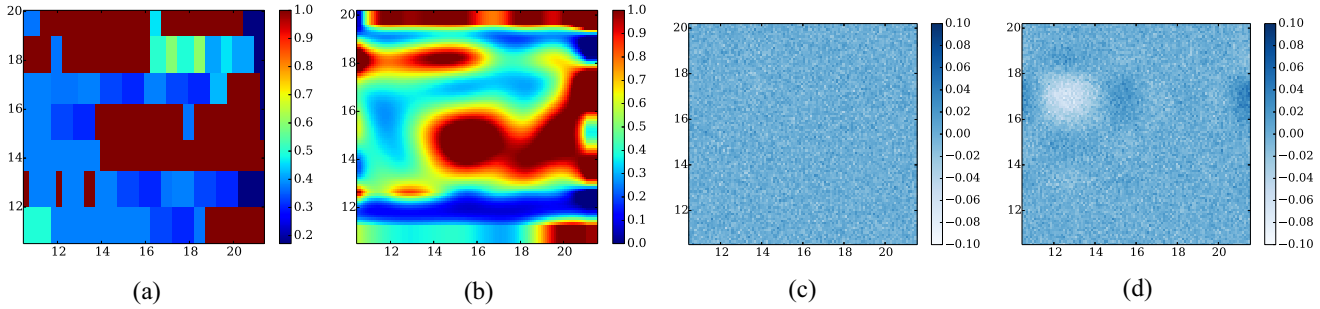

Fig. 7. (a) Part of reflectance layout in testbench c1355. We use single gate FDTD results of the optical responses to represent every pixel in the gate location. (b) Cubic interpolated results from (a) as the reference of nontampered circuit, which we denote using M_Y. (c) AWGN with a variance of 0.01 with the same area. (d) The tampered circuit example, in which we replace one NAND gate with an AND gate. We use the measure response, M_Y, subtracted by the golden reference, M_X, to get the image in (d).

detection evaluates the fitness of the noise introduced to the response image in our predetermined model. Different noise models require corresponding modeling of the noise in the detection method. We assume that the measurement noise is AWGN [37], [38] as a typical example to demonstrate the effectiveness of applying noise-based detection. The discussion of modeling of other noise sources is beyond the scope of this article.

A. Methodology Explanation

To explain our use of the noise-based detection method, we use an example of replacing one NAND gate with an AND gate in the c1355 testbench.[1] In Fig. 7, the four figures represent four numerical representations of the noise-based detection method. After we extract geometric information of the gates from the GDSII files, we map the results from the FDTD simulation of each individual gate to the corresponding location and orientation. We assign values from the FDTD simulation to each pixel in each gate [Fig. 7(a)]. As we discussed in Section III, we interpolate the image in Fig. 7(a) with the bicubic function, and populate the images as shown in Fig. 7(a). We represent the post interpolated results with M_X, shown in Fig. 7(b). M_X matrix is the golden reference. We use M_Y to denote the response image with noise. In our assumption, the noise is additive to the reference, therefore,

$$M_R = M_Y - M_X \tag{4}$$

where M_R is noise. If M_R follows the AWGN [Fig. 7(c)], we consider that the imaged circuit is not tampered, because the difference between the golden reference and measured response consists of noise only. If M_R is not AWGN, then the measured matrix may have other components in its mean or variance which can indicate an HT. Fig. 7(d) presents the imaged figure after the subtraction from the golden reference. By comparing Fig. 7(c) and (d), we can see the difference in the top left corner, which indicates an HT. Thus, we test the matrix M_R to see whether it has a mean of zero and follows Gaussian Distribution to detect HTs in the designs. We determine both false positive and false negative tests of gathered data with different resolutions. We use *D'Agostino's X-squared test* to identify the shape of the expected probability density function (PDF). *D'Agostino's X-squared test* is a metric to evaluate how well a statistic set of data fits Gaussian distribution's PDF. *D'Agostino's X-squared test* utilizes a suggested p-value to evaluate the skewness of a data set compared to Gaussian distribution, which is independent from the mean or variance of the data set. In experimental tests here, we use 0.05% as the p-value threshold as it is commonly adopted in estimation of Gaussian distributions. This p-value is independent from all the testbenches or even any HTs. In this article, we used the p-value in testing Gaussian distribution of the noise matrix M_R. Our method is still applicable for other noise models, given that the other testing methods exist for the other noise models. Compared to the correlation-based detection method, the misalignment between the measured results and the golden reference can impact the quality of the detection system. To ensure the alignment of the system, we can preplace markers in the ICs to align the measured results and the golden reference.

Most side channel-based HT detection techniques consist of two parts: 1) data gathering and 2) data analysis. In the first part, under a given excitation on the side channel, both a benign chip and a malicious chip will respond to inputs based on their physical properties. For example, when using thermal imaging [39] in HT detection, the thermal map is the physical property of a chip under test. The physical excitements are the input electrical signals of the chip, and the output response is the IR image of the chip. Our proposed technique utilizes near-IR illumination as the physical property, near-IR laser signals as excitement, and near-IR images as responses. In the second part, by comparing the simulated results to backside image of DUT image, changes due to HTs can be detected. Here, our proposed noise-based comparison method does not require any prior assumptions about the HTs before testing and is independent from any statistical information from any data gathering method. Therefore, a noise-based detection method can also be a generalized method that can be applied in all kinds of data analysis in HT side-channel detection.

B. Results

We use the noise-based detection method for data analysis. We use false positive and false-negative rates versus SNRs

[1]The example shown here is NAND gate replaced with AND gate. Other gates, such as AND gate replaced with OR gate would show similar results. We do not show other cases here due to space constraints.

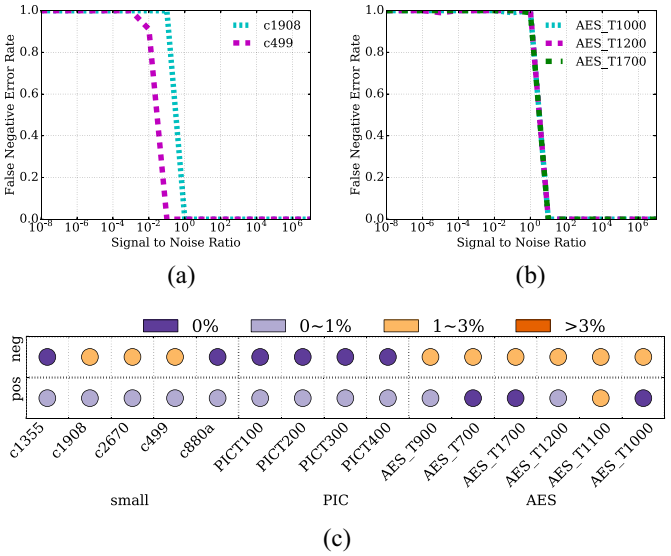

Fig. 8. False-negative error rates versus SNR of (a) two testbenches from [36] using noise detection method and (b) three testbenches from [35] using noise detection method. (c) Summary of various testbenches from both [36] and [35]. We use colors in the legend to denote the error rates at SNR from 10^{-8} to 0.1 belonging to the corresponding range.

to evaluate the performance of our proposed noise-based HT detection method. We use small circuits from [36], PIC circuits and AES large circuits from [35] in our simulations. In the small circuits from [36], we have different designs for each HT testbench. In PIC and AES circuits from [35], we have two basic designs, AES and PIC, with different HTs in each testbench.

Similar to the correlation method, for each test case, we synthesize, floor-plan and place&route each circuit with *Nangate* 45 nm technology using *Cadence* RC and Encounter tool flow to generate GDSII files. We generate and use those files along with DEF files, with both location and orientation information of each gate, to generate the optical response data. We interpolate the reflective response image to generate reference image for later comparisons. Depending on the resolution used for the experiments, we adjust the image interpolation density.

We then use interpolated data in the noise-based detection method to determine whether the circuit has an HT or not. For a given circuit, we divide the imaged matrices into batches of smaller areas in the layouts. Each batch has the same area and represents the reflectance data from a certain area of the image. We denote an area as detection window frame. If we detect HTs in one of the window frames, we state that the batch fails the test. Considering that the simulation time for area of 1 mm^2 is on the order of weeks, we choose a window frame with the size of 10 μm $\times$ 10 μm for small test benches, and 250 μm $\times$ 250 μm for large testbenches. In Section VI-B, we analyze the impact of the window frame sizes on the detection rates.

Fig. 8(a) shows false positive and false-negative rates for c499 and c1908, two testbenches from [36], at SNR from 10^{-8} to 10^{8}. Since triggers in these testbenches are designed for low probability triggering rates, they are nearly impossible to detect using functional testing [36]. The area of HT payload circuits is much larger than the areas of the HT triggers. In this article, we illustrate the sensitivity of our HT detection by only detecting the HT triggers instead of the HT payload, because the triggers are much harder to be detected. Fig. 8(b) shows the results of evaluation of the AES testbenches from the Trusthub website using noise detection method. From Fig. 8(a) and (b), we can see that in c1908, c499 from [36] and AES-T1000, AES-T1200, AES-T1700 from [35], error rates drop to zero with SNR starting from 10^1 in the false-negative error rates.

The testbenches in Fig. 8(c) are AES encryption engines with HTs, and PIC testbenches are HT tampered PIC16 $\times$ 64 circuit. In the AES testbenches, we apply a window frame of 250 μm $\times$ 250 μm in the simulation setup. Fig. 8(c), similar to Fig. 6(c), summarizes all the testbenches used in this evaluation. Each dot in the Fig. 8(c) represents the error rate analysis of SNR varying from 10^{-8} to 10^{-1}. If all error rates in this SNR spectrum fall into the error rate range, we use corresponding color dot to represent it. As we can see from Fig. 8(c), most of the false positive and false-negative results from various testbenches have low error rates in our HT detection method.

VI. Optimizations of HT Detection

To analyze the robustness and applicability of our approach, we consider the impact of process variations (Section VI-A), imaging resolutions and window frame sizes (Section VI-B), and fill cell pattern insertions (Section VI-C) on our proposed method. Process variations have become the major concern for all HT side-channel detection methods. The attacker can design the HTs with smaller power and area overhead than the ones caused by the process variations. The detection method must overcome the impact of process variations, and robustly identify the HTs with fidelity. In our data analysis method, the detection window frame size determines the number of *D'Agostino's X-squared tests* in one chip test and the resolution of sampled data decides the quantity of sampling data in single test. Both factors affect the detection rate; therefore, we evaluate the impacts of both factors on the detection rates in the Monte-Carlo simulations. In addition to engineering fill cells to increase the detection rates, we propose to strategically place the fill cells in the design in order to improve our method.

A. Process Variation

In order to analyze the impact of process variations on the near-IR reflectance, we model the process variations on the functional cells by stretching or shrinking the metal structures in the horizontal or vertical direction.[2] We model the process variation of metal in this form, because this form of process variations has the most impact on our results. The stretching or shrinking of the structures in metal layers can directly change the optical reflectance responses, while the structures in other layers have orders of magnitude lower responses. The process variations in the dopant or the polysilicon have minimum impact on our results (less than 1% [33]). Other process

[2]We would like to note that our methodology is not affected by voltage and temperature variations and hence we do not consider those process variations in our analysis.

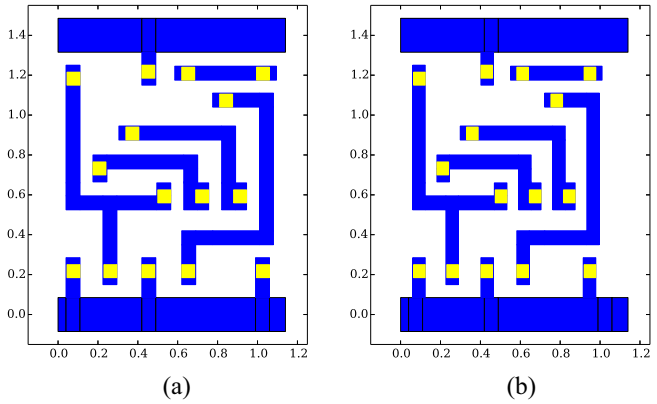

Fig. 9. (a) Process variation free XOR2_X1 gate. (b) XOR2_X1 with 10% process variation in the *X* dimension. Here, all the metal structures inside of the gate have been compressed by 10% in the *X* dimension to model process variations.

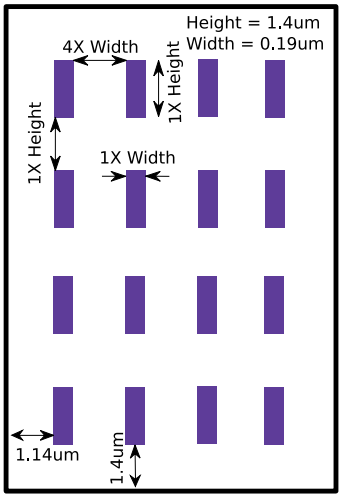

Fig. 10. Fill cell pattern that we inserted before place&route of the design. We put this array of fill cells to secure regions from shifts or replacements of fill cells and replacement of functional cells with other set of functional cells.

variations models such as the irregular shapes on the edges of the metal structures have significantly lower impact.

In Fig. 9, we show the XOR2_X1 as an example gate to explain our methodology of modeling the process variations in *Nangate* 45 nm technology. We apply 10% shrink in the horizontal dimension, except the power rails. We then perform FDTD analysis on the modified layouts to analyze the response changes after the process variations. In Fig. 9, we show the examples of 2_X1 cell with no process variation and with $\pm 10\%$ in the horizontal dimension for the M1 metal layer and M1 contacts. In this modeling, we only consider the effects of the process variations on the inner structures of M1 and contacts, since the all the optical responses are from the inner structures We simulate the responses of eight kinds of basic cells with/without process variations on the spectrum from 1 μm to 3 μm on *X* and *Y* polarizations. Our analyses (see Section VI-D) show that the determinism of gate response does not change due to process variations, which means that the difference in optical response caused by process variations is smaller than the difference between two different gates. In order to minimize the impact of process variations, we choose the wavelength and polarization that is the least affected by process variations (1.2 μm and *Y* polarization in our case) for our evaluation of the noised-based detection method in Section VI-D.

B. Resolution and Window Size

In noise-based data analysis, resolution determines the quantity of data and the detection window size decides the number of individual *D'Agostino's X-squared tests* in HT detection of one chip. Intuitively, higher resolution imaging can better differentiate the details of the designs. Yet the physical property of the side channel in data gathering constrains the resolution of imaged results. In order to quantitatively analyze impacts of resolution on detection rates, we use the down-sampled near-IR imaging matrices to conduct comparisons between different resolutions, as we do not have the data from other side-channel techniques. In near-IR imaging, we use 0.1 μm as the resolution for interpolation. Here, we apply 0.2 μm and 0.4 μm as interpolation resolutions in simulations to evaluate imaging techniques with lower resolutions. These resolution comparisons show that the decisive factor in side-channel HT detection is resolution.

In Section V, for a given testbench, we divide the imaged matrices into batches of the detection window frames, with the same sizes. If we detect HTs in one of the window frames from the tested circuitry, we assume that all the batches fail the test. In this way, we utilize much more data than one value for evaluation. Since our test is based on the distribution of the noise, if one window frame covers a smaller area, the data set of white noise is more likely to fail the noise-based test. If one window frame covers much larger area, the HTs might not be detected since the noise will average out the small modifications in the data set. Further investigations in Section VI-D explain the optimal sizes on the choices of detection window sizes.

C. Pattern Insertion

To further improve the detection rates, we propose to strategically place fill cells into the floor plan before the *place&route* step. We engineer the fill cells to have higher reflectance than other cells. Compared to other cells, which have reflectance around 50%, engineered fill cells have reflectance of almost 100%. As we discussed in the Section IV, HTs often replace fill cells with functional cells, or move fill cells to make room for inserting malicious functional cells. The fill cells have the highest optical responses in backside imaging compared to other gates. Hence, any changes to these preplaced fill cells can be easily identified. In the case when HTs replace functional cells with other set of functional cells, the preplaced fill cells contribute to improvements of the detection rates. The reflectance of a cell is not only determined by its own reflectance but also the reflectance of neighboring gates.

In designing the pattern, we ensure the engineered fill cells can cover as much neighboring area as possible. In our

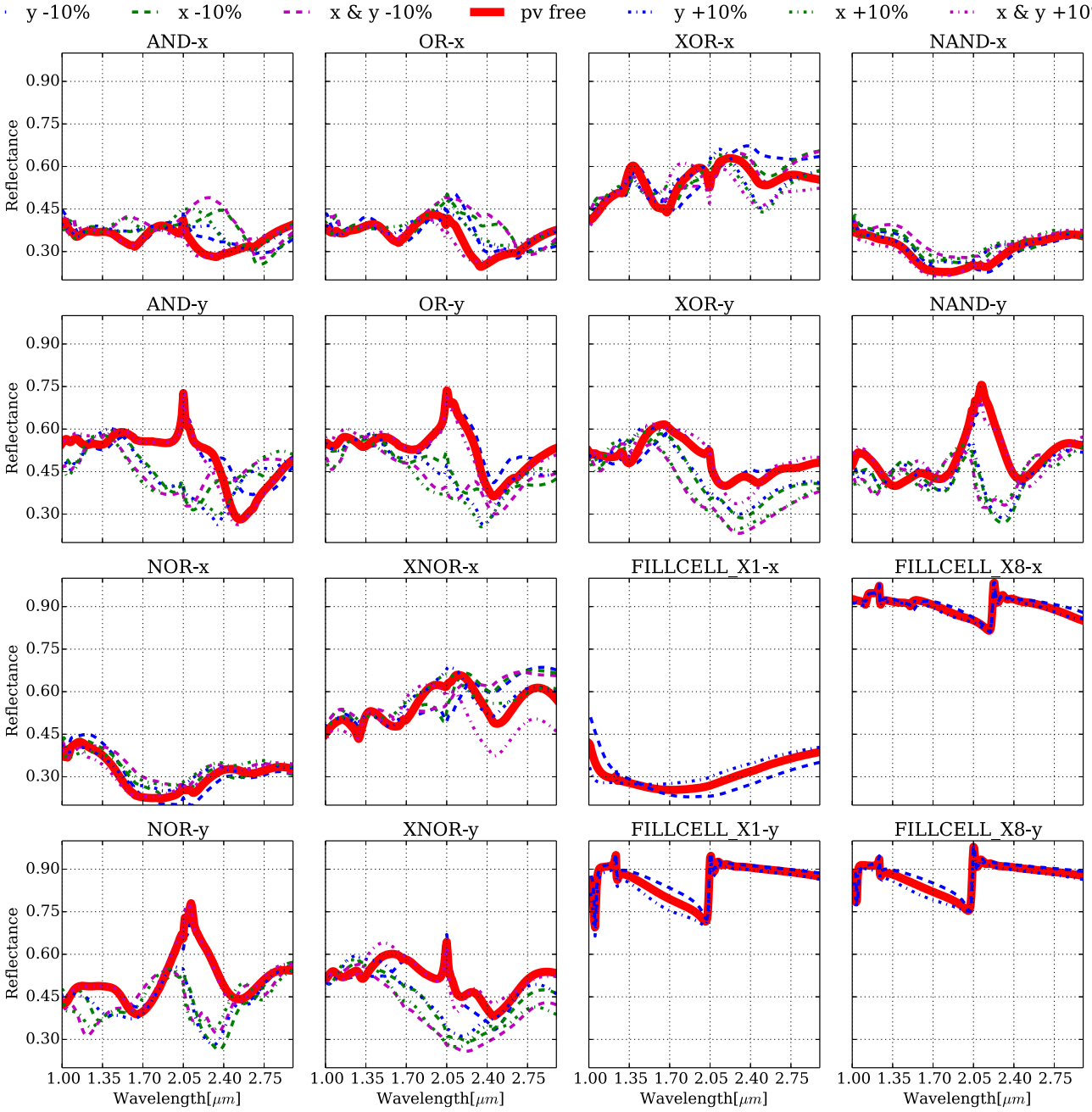

Fig. 11. Reflectance under ±10% variation in dimensions of metal1 structures through near-IR laser of wavelength from 1 μm to 3 μm. *X* polarization laser reflectance is on the left and *Y* polarization laser reflectance is on the right. The thick red line represents the process variation free reflectance of a cell. The other lines are the reflectance with process variation.

benchmarks, the preplaced fill cells do not increase the overall area of the design, since there are extra spaces in the original design for fill cells. In general, we need to design the pattern of the preplaced fill cells such that we use the minimum amount of cells to cover the entire chip. We experimented using multiple designs with different sizes of fill cells and different distances between cells. The optimal solution without increasing the area of the design, which yielded highest detection rates among various SNR, gave the pattern in Fig. 10. We present the sensitivity analysis results of the inserted patterns in Section VI-D.

D. Results

1) Process Variations: As discussed in Section VI, process variations can be minimized by choosing the wavelength of the optical source that is most resistant to process variations. We simulate the responses of eight different standard cells with/without process variations on the spectrum ranging from 1 μm to 3 μm on *X* and *Y* polarizations (Fig. 11). In Fig. 11, the thick red line represents the reflectance of functional cells that are free of process variation. The other lines represent the reflectance with process variations. At the wavelength of 1.2 μm, the changes to the reflectance for cell is less than ±5%

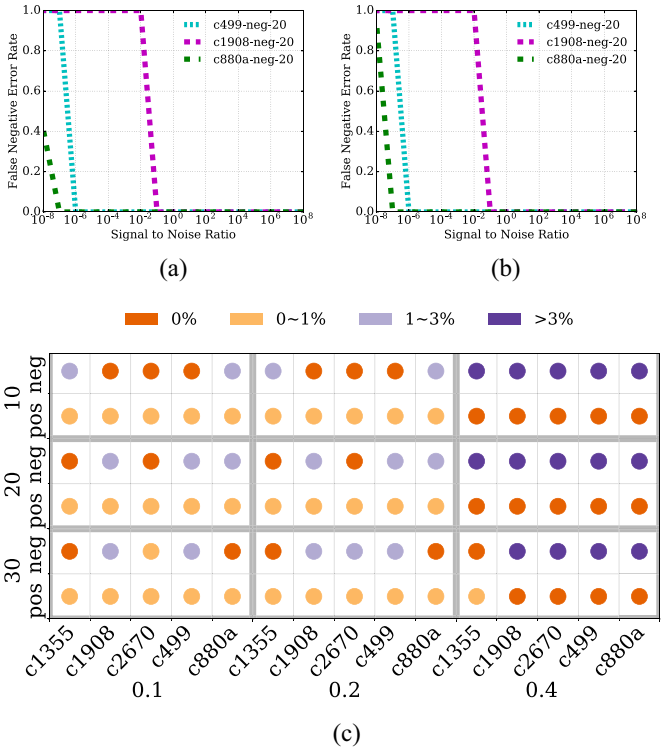

Fig. 12. False-negative rates versus SNR of three testbenches from [35] (a) with an imaging resolution of 0.1 μm and (b) with an imaging resolution of 0.2 μm using noise detection method. (c) Summary of various testbenches from both [35] and [36]. We use colors in the legend to denote the error rates at SNR from 10^{-8} to 10^{-1} belonging to the corresponding range. Here, 10, 20, and 30 refers to detection window sizes 10 μm, 20 μm, and 30 μm. On the X-axis, 0.1, 0.2, and 0.4 refers to resolutions of the image as 0.1 μm, 0.2 μm, and 0.4 μm, respectively.

in X polarization imaging, and at most $\pm$10% in Y polarization imaging, except FILLCELL_X1. Only in FILLCELL_X1, the reflectance of the fill cell in X polarization is unusually low, due to narrow, thin cell shape with much less metal fillings in the horizontal direction.

2) Resolution and Window Size: Here, we use five various testbenches to evaluate the effects of the resolution and the window frame sizes on the detection rate. Fig. 12(a) and (b) shows false-negative rates versus SNRs for imaging in three testbenches from [36]. Fig. 12(a) uses 0.1 μm interpolation resolution for imaging, while Fig. 12(b) uses 0.2 μm. Fig. 12(a) and (b) clearly shows that higher resolution provides performance improvements at lower SNR. Here, we conclude that in smaller testbenches, such as c1355, improving resolution from 0.1 μm to 0.2 μm does not affect detection rate much.

Fig. 12(c) summarizes the relationship between the resolutions and the window frame sizes for different testbenches. If all false-negative rates in this SNR spectrum fall into the error rate range, we use corresponding color dot to represent it. Since we are using 45 nm *Nangate* technology, if we decrease the resolution to 0.4 μm, many of the metal structures with minimum designed sizes become blurry in imaging. In Fig. 12(c), false-negative rates for a resolution of 0.4 μm are much lower than that for resolutions of 0.1 μm and 0.2 μm. In the detection window size analysis, false-negative rates

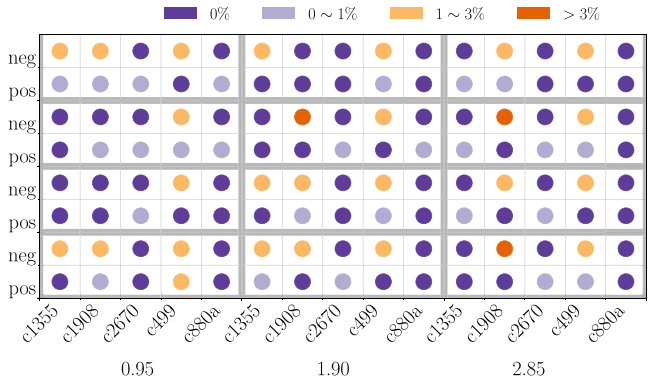

Fig. 13. Detection rates with different patterns. In all the patterns, we use the same number of rows and columns of fill cell arrays. Y-axis represents number of rows and columns in fill cell array. X-axis represents the size of fill cells in the array.

TABLE II
AREA (μm^2) OF STRATEGICALLY PLACED INSERTED FILL CELLS, 0.19, 0.38, 0.57 REPRESENTS THE WIDTH OF EACH FILL CELL IN THE ARRAY IN [μm], 1, 2, 3, 4 REPRESENTS NUMBER OF ROWS AND COLUMNS OF THE ARRAY

	1	2	3	4
0.19	0.266	1.064	2.394	4.256
0.38	0.532	2.128	4.788	8.512
0.57	0.798	3.192	7.182	12.768

increase as the detection window size increases because the imaged response of the HT gets averaged out in large detection windows. The colored dots represent the binned results instead of the absolute values. The results show a trend of increase in false-negative rates as the window sizes increase. Fig. 12(c) shows that high resolution imaging can increase the HT detection rates.

3) Pattern Insertions: By strategically placing the engineered fill cells before placing the designs, we can further improve detection rates. As shown in Fig. 10, our fill cells start from bottom left corner. The distance from the left and bottom edges are 1.14 μm and 1.4 μm, respectively. We fixed the distance between two fill cells to be 4$\times$ the width of the fill cell in the horizontal direction and 1$\times$ the height of the fill cell in the vertical direction. We vary the size of fill cells and number of the fill cells inserted in the design to find out the optimal amount of fill cells that we need to place in the floor plan. In Fig. 13, we show that more fill cells placed in the design does not necessarily improve HT detection rates. The results may be affected by the *place&route* automation tool. A detailed investigation of the impact of amount of metal pattern, placement algorithm and routing algorithm on the HT detection rates will be completed as part of our future work. Table II shows all the area increase for placing fill cells. The extra area cost of all the designs is less than 5% of the design area.

VII. RELATED WORK

HTs inserted during the fabrication are hard to detect. These HTs are small in terms of power and area, and have extremely

low triggering probability. Conventional verification methods may not be effective since the timing and power variations due to the HTs are lower than those due to the process variation. Low triggering rates make it much harder to identify HTs during chip verification. Various approaches have been proposed to either prevent or detect HTs during fabrications. Three categories of protection methods have been proposed: prefabrication protector designs, post-fabrication detection, and side-channel-based detection method.

A number of prefabrication protector designs have been explored at both circuit and architecture levels. At the circuit level, split manufacturing, gate level obfuscation, and limitations of gate usage have been proposed to protector designs. Split manufacturing [40]–[42] proposes that the lowest layer of IC is fabricated by one trusted vendor, while other layers are fabricated by untrusted vendors at a lower cost. This approach is too expensive since the trusted vendor has to use high cost through silicon vias (TSVs) to assemble different layers of logic. HARPOON [10] provides netlist-level obfuscation, which can be integrated into the synthesis of SoC designs. Other techniques, [9], [43]–[45] also provide protection by replacing the fill cells with functional cells. Such obfuscation designs do not require extra die area but they do not provide strong protection against SAT solver analysis [46]. Bao *et al.* [47] proposed to both prevent and detect HTs by limiting the usable gate types in the standard library. This limits the performance of the resulting designs. All these abovementioned techniques are either expensive or sacrifice area in order to prevent HT insertion. At the architecture level, logic state scrambling, power signatures in potential tampered logic, and circuitry encoding have been proposed. ODETTE [48] protects the circuitry from HTs by increasing the number and variations of reachable states. VITAMIN [49] utilizes an inverted voltage scheme that aims to activate some targeted Trojans with a higher triggering rate. The Trojan's power consumption will become significant in the power analysis against HTs in this way. Linear complement dual codes [50] are designed to encode the instructions inside the circuitry and to be resistant to error injection and side channel analysis. These methods potentially increase the complexity of logic designs, and they also increase the power consumption of the overall systems.

Post-fabrication detection techniques includes RE, timing and power analysis, and data monitoring during runtime. The generic post-fabrication detection method is to reverse engineer the fabricated ICs. SEM and FIB have been proposed to reverse engineer ICs for HT Detection. However, these approaches are expensive and slow. FANCI [51] and FIGHT [52] flag possible HT wires, which can reduce the complexity in RE. Unfortunately, the process to RE the fabricated chip is only affordable to large semiconductor companies [1], [4]. Standard testing methods have been proposed in HT detection, which do not require full RE of the fabricated chip. Timing-based analysis [53]–[55] analyzes the delay changes in the circuitry to detect HTs. Dynamic power detection, such as the sustained vector technique, measures the dynamic power consumption difference in the circuit and identifies the gates that may contain HTs [56]–[58]. Gate level characterization (GLC) [59]–[62] describes the gates in terms of leakage power, switching power, and delays. It leverages linear programming to solve a system of equations to describe the entire circuit. Such time and power-based analysis, including [59], [61], [63], [64] are less preferable on higher technology nodes, since the impact of process variations on power and delays become even larger than the impact caused by the HTs, i.e., analog capacitance HTs [23]. Control monitoring includes *TrustNet* and *Datawatch* to detect malicious attacks on processors [65], [66] during runtime. This introduces overhead in the performance in order to detect HTs.

Side-channel detection utilizes physical properties of ICs other than electrical properties to detect HTs hiding from electrical tests. Emission measurements [67], [68] and thermal analysis [39], [69], [70] are good examples of side channel analysis, but all these techniques are either not as fast as our technique, or as faster but do not have as high a resolution as our technique.

VIII. Discussion

In this article, we have only focused on simulation-based detection. We have taped out ICs with and without HTs using our modified logic gates. We will report the measurement results as part of a future publication.

We can apply our detection method to the detection of changes in the design. If the adulteration of the circuits is caused by replacing or shifting the logic gates, which changes the structure in metal 1, our method should be able to detect it. If the adulteration of the circuits is caused by the modifications such as changing the dopant concentration in the Si layers or changing the transistor dimensions, our method cannot detect such changes.

In our previous work, we have modeled the reflectance based on infinite tiling as the boundary. In this article, we model the optical response with the boundary conditions of random gates. It is true that ideally we should consider all the combinations of the surrounding gates. However, it is not practical to do this due to the excessively long FDTD simulation time for each combination. (around 8 h per gate combination). Thus, we have modeled our gate boundary condition using random neighboring gates.

IX. Conclusion

In this article, we developed a new technique for HT detection. Our technique uses near-IR light to image the metal structures from the backside of the IC. We use the FDTD simulations to generate the imaged results from the metal structures in the gate library. After the IC design, we can extract the locations of each gate from our design. Combined with the FDTD results of the gates, we can generate the reflection results of the design as golden reference. During the testing process, we modeled the image of the manufactured ICs, compared against the golden reference, and identified the differences to detect HTs.

In order to achieve the high contrast in backside imaging to detect HTs, we modified the fill cells in the library to fill with metal. Any modifications, shift or replacements of these fill

cells can be detected by our technique. We improved our detection rate by implementing a noised-based detection method. We evaluated our techniques with different noise levels and different observation windows. Our analysis shows that we are able to detect HTs that occupy less than 0.1% of the total chip area.

ACKNOWLEDGMENT

The authors would like to thank Sheng Wei for providing source codes from his paper [36] for our testbenches. The authors also thank Mahmoud Zangeneh and Tianyu Yang for their initial work on this project. Thanks to the design team of Michigan A2 chip, who provided them with their chip examples.

REFERENCES

[1] M. Tehranipoor and F. Koushanfar, "A survey of hardware Trojan taxonomy and detection," *IEEE Design Test Comput.*, vol. 27, no. 1, pp. 10–25, Jan./Feb. 2010.

[2] *Trust in Integrated Circuits (TIC)—Proposer Information Pamphlet*, DARPA, Arlington, VA, USA, 2007. [Online]. Available: https://bit.ly/2Uz37jy

[3] R. S. Chakraborty, S. Narasimhan, and S. Bhunia, "Hardware trojan: Threats and emerging solutions," in *Proc. Int. High Level Design Validation Test Workshop (HLDVT)*, San Francisco, CA, USA, 2009, pp. 166–171.

[4] R. Karri, J. Rajendran, K. Rosenfeld, and M. Tehranipoor, "Trustworthy hardware: Identifying and classifying hardware Trojans," *Computer*, vol. 43, no. 10, pp. 39–46, Oct. 2010.

[5] (2019). *The Semiconductor Industry and the Power of Globalisation*. Accessed: Jan. 3, 2019. [Online]. Available: https://econ.st/2KRcwiB

[6] E. Love, Y. Jin, and Y. Makris, "Proof-carrying hardware intellectual property: A pathway to trusted module acquisition," *IEEE Trans. Inf. Forensics Security (TIFS)*, vol. 7, no. 1, pp. 25–40, Feb. 2012.

[7] A. Waksman and S. Sethumadhavan, "Silencing hardware backdoors," in *Proc. IEEE Symp. Security Privacy (SP)*, Berkeley, CA, USA, 2011, pp. 49–63.

[8] M. Hicks, M. Finnicum, S. T. King, M. M. Martin, and J. M. Smith, "Overcoming an untrusted computing base: Detecting and removing malicious hardware automatically," in *Proc. IEEE Symp. Security Privacy (SP)*, Oakland, CA, USA, 2010, pp. 159–172.

[9] J. A. Roy, F. Koushanfar, and I. L. Markov, "EPIC: Ending piracy of integrated circuits," in *Proc. Conf. Design Autom. Test Eur. (DATE)*, 2008, pp. 1069–1074.

[10] R. S. Chakraborty and S. Bhunia, "HARPOON: An obfuscation-based SoC design methodology for hardware protection," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. (TCAD)*, vol. 28, no. 10, pp. 1493–1502, Oct. 2009.

[11] R. Torrance and D. James, "The state-of-the-art in semiconductor reverse engineering," in *Proc. 48th ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, New York, NY, USA, 2011, pp. 333–338.

[12] (2019). *Intel 22-nm Trigate Transistors Exposed*. Accessed: Apr. 3, 2019. [Online]. Available: https://bit.ly/2EW4Cll

[13] *Integrity and Reliability of Integrated Circuits (IRIS)*, DARPA, Arlington, VA, USA, 2019. Accessed: Nov. 3, 2019. [Online]. Available: https://bit.ly/2NZtigF

[14] (2019). *Iphone 5 A6 SoC Reverse Engineered, Reveals Rare Hand-Made Custom CPU, and Tri-Core GPU —Extremetech*. Accessed: Apr. 3, 2019. [Online]. Available: https://bit.ly/2O16AVO

[15] (2019). *How to Reverse Engineer Software (Windows) in a Right Way*. Accessed: Apr. 3. 2019. [Online]. Available: https://bit.ly/2u0GRU7

[16] (2019). *Reverse Engineering Integrated Circuits With Degate—Home*. Accessed: Apr. 3, 2019. [Online]. Available: https://bit.ly/2VSRZym

[17] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Proc. Annu. Int. Cryptol. Conf. (CRYPTO)*, 1999, pp. 388–397.

[18] A. Schlösser, D. Nedospasov, J. Krämer, S. Orlic, and J.-P. Seifert, "Simple photonic emission analysis of AES," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst. (CHES)*, 2012, pp. 41–57.

[19] F. Koeune and F.-X. Standaert, "A tutorial on physical security and side-channel attacks," in *Foundations of Security Analysis and Design III*. Heidelberg, Germany: Springer, 2005, pp. 78–108.

[20] D. Genkin, A. Shamir, and E. Tromer, "RSA key extraction via low-bandwidth acoustic cryptanalysis," in *Proc. Int. Annu. Cryptol. Conf. (CRYPTO)*, 2014, pp. 444–461.

[21] P. Rohatgi, "Electromagnetic attacks and countermeasures," in *Cryptographic Engineering*. Boston, MA, USA: Springer, 2009, pp. 407–430.

[22] M. Rostami, F. Koushanfar, and R. Karri, "A primer on hardware security: Models, methods, and metrics," *Proc. IEEE*, vol. 102, no. 8, pp. 1283–1295, Aug. 2014.

[23] K. Yang, M. Hicks, Q. Dong, T. Austin, and D. Sylvester, "A2: Analog malicious hardware," in *Proc. IEEE Symp. Security Privacy (SP)* San Jose, CA, USA, 2016, pp. 18–37.

[24] N. G. Tsoutsos and M. Maniatakos, "Fabrication attacks: Zero-overhead malicious modifications enabling modern microprocessor privilege escalation," *IEEE Trans. Emerg. Topics Comput. (TETC)*, vol. 2, no. 1, pp. 81–93, Mar. 2014.

[25] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, 2002.

[26] S. B. Ippolito, S. A. Thorne, M. G. Eraslan, B. B. Goldberg, and M. S. Ünlü, "High spatial resolution subsurface thermal emission microscopy," *Appl. Phys. Lett.*, vol. 84, no. 22, p. 4529, 2004.

[27] U. Kindereit, G. Woods, J. Tian, U. Kerst, R. Leihkauf, and C. Boit, "Quantitative investigation of laser beam modulation in electrically active devices as used in laser voltage probing," *IEEE Trans. Device Mater. Rel. (TDMR)*, vol. 7, no. 1, pp. 19–30, Mar. 2007.

[28] F. H. Köklü and M. S. Ünlü, "Subsurface microscopy of interconnect layers of an integrated circuit," *Opt. Lett.*, vol. 35, no. 2, pp. 184–186, Jan. 2010.

[29] S. B. Ippolito, B. B. Goldberg, and M. S. Ünlü, "Theoretical analysis of numerical aperture increasing lens microscopy," *J. Appl. Phys.*, vol. 97, no. 5, 2005, Art. no. 053105.

[30] L. Novotny and B. Hecht, *Principles of Nano-Optics*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[31] P. Török, P. Munro, and E. E. Kriezis, "High numerical aperture vectorial imaging in coherent optical microscopes," *Opt. Exp.*, vol. 16, no. 2, pp. 507–523, 2008.

[32] İ. R. Çapoğlu, C. A. White, J. D. Rogers, H. Subramanian, A. Taflove, and V. Backman, "Numerical simulation of partially coherent broadband optical imaging using the finite-difference time-domain method," *Opt. Lett.*, vol. 36, no. 9, pp. 1596–1598, 2011.

[33] B. Zhou *et al.*, "Detecting hardware Trojans using backside optical imaging of embedded watermarks," in *Proc. 52nd ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, San Francisco, CA, USA, 2015, pp. 1–6.

[34] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process. (TASSP)*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[35] (2014). *Trust-Hub Website*. Accessed: Nov. 30, 2014. [Online]. Available: https://bit.ly/2TCHRwd

[36] S. Wei, K. Li, F. Koushanfar, and M. Potkonjak, "Hardware Trojan horse benchmark via optimal creation and placement of malicious circuitry," in *Proc. 49th Design Autom. Conf. (DAC)*, San Francisco, CA, USA, 2012, pp. 90–95.

[37] M. Popović and A. Taflove, "Two-dimensional FDTD inverse-scattering scheme for determination of near-surface material properties at microwave frequencies," *IEEE Trans. Antennas Propag. (TAP)*, vol. 52, no. 9, pp. 2366–2373, Sep. 2004.

[38] A. Zjajo, "Random process variation in deep-submicron CMOS," in *Stochastic Process Variation in Deep-Submicron CMOS*. Dordrecht, The Netherlands: Springer, 2014, pp. 17–54.

[39] K. Hu, A. N. Nowroz, S. Reda, and F. Koushanfar, "High-sensitivity hardware Trojan detection using multimodal characterization," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, Grenoble, France, 2013, pp. 1271–1276.

[40] K. Vaidyanathan, B. P. Das, and L. Pileggi, "Detecting reliability attacks during split fabrication using test-only BEOL stack," in *Proc. 51st ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, San Francisco, CA, USA, 2014, pp. 1–6.

[41] K. Vaidyanathan, R. Liu, E. Sumbul, Q. Zhu, F. Franchetti, and L. Pileggi, "Efficient and secure intellectual property (IP) design with split fabrication," in *Proc. IEEE Int. Symp. Hardw. Orient. Security Trust (HOST)*, Arlington, VA, USA, 2014, pp. 13–18.

[42] J. Valamehr *et al.*, "A 3-D split manufacturing approach to trustworthy system development," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. (TCAD)*, vol. 32, no. 4, pp. 611–615, Apr. 2013.

[43] A. Baumgarten, A. Tyagi, and J. Zambreno, "Preventing IC piracy using reconfigurable logic barriers," *IEEE Design Test Comput. (DTC)*, vol. 27, no. 1, pp. 66–75, Jan./Feb. 2010.

[44] K. Xiao, D. Forte, and M. Tehranipoor, "A novel built-in self-authentication technique to prevent inserting hardware Trojans," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. (TCAD)*, vol. 33, no. 12, pp. 1778–1791, Dec. 2014.

[45] K. Xiao, D. Forte, and M. M. Tehranipoor, "Efficient and secure split manufacturing via obfuscated built-in self-authentication," in *Proc. IEEE Int. Symp. Hardw. Orient. Security Trust (HOST)*, Washington, DC, USA, 2015, pp. 14–19.

[46] M. El Massad, S. Garg, and M. V. Tripunitara, "Integrated circuit (IC) decamouflaging: Reverse engineering camouflaged ICs within minutes," in *Proc. Netw. Distrib. Syst. Security Symp. (NDSS)*, 2015, pp. 1–14.

[47] C. Bao, Y. Xie, and A. Srivastava, "A security-aware design scheme for better hardware Trojan detection sensitivity," in *Proc. IEEE Symp. Hardw. Orient. Security Trust (HOST)*, Washington, DC, USA, 2015, pp. 52–55.

[48] M. Banga and M. S. Hsiao, "ODETTE: A non-scan design-for-test methodology for trojan detection in ICs," in *Proc. IEEE Symp. Hardw. Orient. Security Trust (HOST)*, San Diego CA, USA, 2011, pp. 18–23.

[49] M. Banga and M. S. Hsiao, "VITAMIN: Voltage inversion technique to ascertain malicious insertions in ICs," in *Proc. IEEE Symp. Hardw. Orient. Security Trust (HOST)*, San Francisco, CA, USA, 2009, pp. 104–107.

[50] X. T. Ngo, S. Bhasin, J.-L. Danger, S. Guilley, and Z. Najm, "Linear complementary dual code improvement to strengthen encoded circuit against hardware Trojan horses," in *Proc. IEEE Symp. Hardw. Orient. Security Trust (HOST)*, Washington, DC, USA, 2015, pp. 82–87.

[51] A. Waksman, M. Suozzo, and S. Sethumadhavan, "FANCI: Identification of stealthy malicious logic using boolean functional analysis," in *Proc. ACM SIGSAC Conf. Comput. Commun. Security (CCS)*, 2013, pp. 697–708.

[52] D. Sullivan, J. Biggers, G. Zhu, S. Zhang, and Y. Jin, "FIGHT-metric: Functional identification of gate-level hardware trustworthiness," in *Proc. 51st ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, San Francisco, CA, USA, 2014, pp. 1–4.

[53] J. Li and J. Lach, "At-speed delay characterization for IC authentication and trojan horse detection," in *Proc. IEEE Int. Workshop Hardw. Orient. Security Trust (HOST)*, Anaheim, CA, USA, 2008, pp. 8–14.

[54] Y. Jin and Y. Makris, "Hardware Trojan detection using path delay fingerprint," in *Proc. IEEE Int. Workshop Hardw. Orient. Security Trust (HOST)*, Anaheim, CA, USA, Jun. 2008, pp. 51–57.

[55] I. Exurville, L. Zussa, J.-B. Rigaud, and B. Robisson, "Resilient hardware Trojans detection based on path delay measurements," in *Proc. IEEE Int. Symp. Hardw. Orient. Security Trust (HOST)*, Washington, DC, USA, 2015, pp. 151–156.

[56] M. Banga and M. S. Hsiao, "A novel sustained vector technique for the detection of hardware Trojans," in *Proc. IEEE 22nd Int. Conf. VLSI Design (VLSID)*, New Delhi, India, 2009, pp. 327–332.

[57] 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. (VLSI) Syst.*, vol. 20, no. 1, pp. 112–125, Jan. 2012.

[58] I. Wilcox, F. Saqib, and J. Plusquellic, "GDS-II trojan detection using multiple supply pad VDD and GND IDDQs in ASIC functional units," in *Proc. IEEE Int. Symp. Hardw. Orient. Security Trust (HOST)*, Washington, DC, USA, 2015, pp. 144–150.

[59] M. Potkonjak, A. Nahapetian, M. Nelson, and T. Massey, "Hardware Trojan horse detection using gate-level characterization," in *Proc. 46th ACM/IEEE Design Autom. Conf. (DAC)*, San Francisco, CA, USA, 2009, pp. 688–693.

[60] S. Wei, S. Meguerdichian, and M. Potkonjak, "Gate-level characterization: Foundations and hardware security applications," in *Proc. IEEE Design Autom. Conf. (DAC)*, Anaheim, CA, USA, 2010, pp. 222–227.

[61] S. Wei and M. Potkonjak, "Scalable hardware Trojan diagnosis," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 6, pp. 1049–1057, Jun. 2012.

[62] S. Wei, A. Nahapetian, M. Nelson, F. Koushanfar, and M. Potkonjak, "Gate characterization using singular value decomposition: Foundations and applications," *IEEE Trans. Inf. Forensics Security (TIFS)*, vol. 7, no. 2, pp. 765–773, Apr. 2012.

[63] R. Rad, X. Wang, M. Tehranipoor, and J. Plusquellic, "Power supply signal calibration techniques for improving detection resolution to hardware Trojans," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, San Jose, CA, USA, Nov. 2008, pp. 632–639.

[64] Y. Alkabani and F. Koushanfar, "Consistency-based characterization for IC trojan detection," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, San Jose, CA, USA, Nov. 2009, pp. 123–127.

[65] A. Waksman and S. Sethumadhavan, "Tamper evident microprocessors," in *Proc. IEEE Symp. Security Privacy (SP)*, Oakland, CA, USA, 2010, pp. 173–188.

[66] J. Backer, D. Hely, and R. Karri, "Reusing the IEEE 1500 design for test infrastructure for security monitoring of systems-on-chip," in *Proc. IEEE Int. Symp. Defect Fault Tolerance VLSI Nanotechnol. Syst. (DFT)*, Amsterdam, The Netherlands, 2014, pp. 52–56.

[67] P. Song *et al.*, "Marvel—Malicious alteration recognition and verification by emission of light," in *Proc. IEEE Int. Symp. Hardw. Orient. Security Trust (HOST)*, San Diego CA, USA, 2011, pp. 117–121.

[68] F. Stellari, P. Song, and H. A. Ainspan, "Functional block extraction for hardware security detection using time-integrated and time-resolved emission measurements," in *Proc. IEEE 32nd VLSI Test Symp. (VTS)*, Napa, CA, USA, 2014, pp. 1–6.

[69] A. N. Nowroz, K. Hu, F. Koushanfar, and S. Reda, "Novel techniques for high-sensitivity hardware Trojan detection using thermal and power maps," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. (TCAD)*, vol. 33, no. 12, pp. 1792–1805, Dec. 2014.

[70] D. Forte, C. Bao, and A. Srivastava, "Temperature tracking: An innovative run-time approach for hardware Trojan detection," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, San Jose, CA, USA, 2013, pp. 532–539.