Proof of Reverse Engineering Barrier: SEM Image Analysis on Covert Gates

Tasnuva Farheen, Ulbert Botero, Nitin Varshney, Damon L. Woodard, Mark Tehranipoor, and Domenic Forte
Department of Electrical and Computer Engineering, University of Florida
Email: {tasnuvafarheen, jbot2016, nitinvarshney1}@ufl.edu, {dwoodard, tehranipoor, dforte}@ece.ufl.edu

Haoting Shen
Department of Computer Science and Engineering, University of Nevada,
Email: hshen@unr.edu

Abstract

IC camouflaging has been proposed as a promising countermeasure against malicious reverse engineering. Camouflaged gates contain multiple functional device structures, but appear as one single layout under microscope imaging, thereby hiding the real circuit functionality from adversaries. The recent covert gate camouflaging design comes with a significantly reduced overhead cost, allowing numerous camouflaged gates in circuits and thus being resilient against various invasive and semi-invasive attacks. Dummy inputs are used in the design, but SEM imaging analysis was only performed on simplified dummy contact structures in prior work. Whether the e-beam during SEM imaging will charge differently on different contacts and further reveal the different structures or not requires extended research. In this study, we fabricated real and dummy contacts in various structures and performed a systematic SEM imaging analysis to investigate the possible charging and the consequent passive voltage contrast on contacts. In addition, machine-learning based pattern recognition was also employed to examine the possibility of differentiating real and dummy contacts. Based on our experimental results, we found that the difference between real and dummy contacts is insignificant in SEM imaging, which effectively prevents adversarial SEM-based reverse engineering.

***Index Terms*—Reverse Engineering, IC Camouflaging, Scanning Electron Microscopy, Machine Learning, Countermeasure.**

Introduction

Security-sensitive assets, modules, and intellectual property (IP) are ubiquitous in modern system-on-chip circuits (SoCs). A key objective of security architectures, such as Intel SGX and ARM Trustzone, of these SoCs is to protect such assets from non-invasive attacks. However, they remain jeopardized by the evolution of front-side and back-side invasive and semi-invasive attacks [1], [2], [3], [4]. Many physical inspection methods such as focused ion beam (FIB) micro-probing, FIB circuit editing, and contact-less optical probing, for failure analysis and defect localization [5], [6] have been utilized to identify regions of interests and/or extract sensitive information.

SEM-based imaging, in combination with iterative integrated circuit (IC) deprocessing, can be used to recover the layout and eventually the gate-level netlist of a semiconductor IP [7], [8]. In some cases, the netlist alone is all that is needed by the attacker. However, in other cases, such as logic locked or obfuscated IPs, a recovered netlist is used as starting point for executing non-invasive attacks such as SAT [9], [10] or ATPG / key sensitization [11]. Possession of the layout or netlist can also aid in other invasive and semi-invasive attacks on assets. Thus, there is motivation for developing countermeasures to SEM-based RE. To disrupt the attacker's ability to identify logic gates from SEM images, camouflaging techniques such as manipulating the threshold voltage [12], [13], [14] of transistors and dummy contacts [11] have been proposed. However, in these early camouflaging processes, the attacker knows which cells are camouflaged in the design. Also, because of their significant area, power, and performance overheads, the number of camouflaged gates is limited, which permits an attacker to conduct various types of non-invasive attacks [9] and invasive attacks (e.g., probing) to recover the functionality.

To solve the issues mentioned above, in our previous work [15], we took a different IC camouflaging approach called "Covert Gates". This approach leverages doping concentration to form always-on transistors and dummy contacts to create always-off transistors. A combination of always-on and always-off transistors make up a covert cell. These covert logic gates are supposed to appear no different from regular logic gates under SEM imaging and should not be part of the original circuit functionality. As a result, the attackers first have to find out which gates are covert before proceeding with any attack, which leads to a considerable increase in complexity in contrast to the scenario where the attackers can quickly tell which ones are camouflaged and which are not. The attacker's ability to resolve the covert gates by FIB milling/probing or laser voltage probing (LVP) is also limited compared to other camouflaging techniques. As covert gates are identical to regular cells, these attacks have to be applied in every single gate, which is infeasible considering the required time, cost, and number of gates in modern designs. Although non-invasive SAT and ATPG based attacks are not as expensive and time-consuming as invasive ones, since every gate is suspect, their scalability also suffers as clearly demonstrated in our previous work [15]. Specifically, we have shown that the SAT attack must encode a vast number of gates to resolve the covert gates and the ATPG based attack cannot distinguish 98% of covert gates even under optimistic circumstances.

This paper expands on our previous one to validate some of our prior assumptions. In [15], it was claimed that doped regions of regular and always-on cases, real contacts, and dummy contacts of always-off transistors are indistinguishable under secondary electron (SE) and back scattered electron (BSE) imaging conditions. The discussion about the silicon doping for always-on transistors (essentially depletion mode devices) was very close to real scenarios. However, the dummy contact structure for always-off ones was simplified. In real devices, the contact connection could be different. For example, some contacts are connected to metal traces, some contacts are connected to silicon, while dummy contacts are simply embedded in isolating materials (e.g., silicon oxide). Consequently, the charge accumulation resulting from the SEM electron beam can be different based on the contact connections. Since the SEM imaging is sensitive to the charge accumulation, it is theoretically possible that the different accumulations will give clues to reverse engineering for the differentiation of real and dummy contacts [16]. Therefore, a further investigation on the charging effect on SEM imaging of dummy contacts is necessary.

Contributions. Our main contributions in this paper are summarized as follows:

- We fabricate real and dummy contacts in different structures to study the charging effect on SEM imaging.
- As a proof-of-concept, we perform a systematic study based on different dwelling time, voltage contrast, and intensity showing experimental results on SE and BSE images. Our main goal is to see whether dummy contacts are indistinguishable from real contacts or not considering real scenarios of variation of charging volume.
- We also verify the indistinguishability by performing both manual analysis and machine learning.

The rest of the paper is organized as follows. Section II reviews the SEM imaging for reverse engineering and IC camouflaging techniques. It also introduces the covert gate approach along with its operating principle. Section III describes the fabrication of dummy contact. Section IV describes our imaging setup and analyzes the images manually and using pattern recognition. Section V concludes the paper and discusses future directions.

Background

A. SEM Imaging for Reverse Engineering

During the imaging step of IC reverse engineering (RE), SEM, a powerful magnification system, has become the tool of-choice with the shrinking of IC feature size. To capture information from a sample surface, SEM employs focused electron beams-(e-beams). It can efficiently image over a large surface and reveal various surface properties, such as chemical components, surface potential, topography, and conductivity. It also runs efficiently compared to other microscopic techniques such as Transmission Electron Microscope (TEM), Atomic Force Microscope (AFM), etc. Thus, SEM has become the most popular imaging technique for IC RE with high efficiency and resolution. As the SEM images are directly used for netlist extraction and annotation, any secure camouflaging technique must be resistant against this.

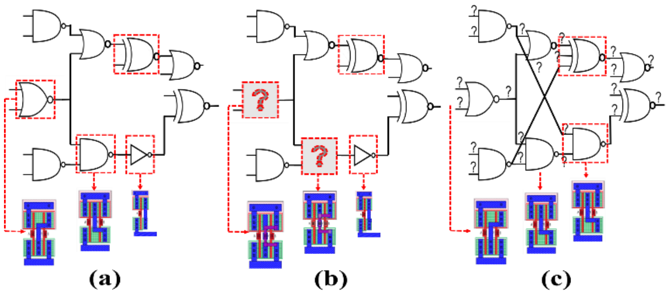

Fig. 1: (a) Original gate level netlist (b) Traditional camouflaging with dummy contacts where '?' indicates a gate that the attacker needs to resolve as NOR or NAND; (c) Covert gate camouflaging with potential dummy inputs ('?') that the attacker needs to resolve.

Recognizing gates and different doping regions of transistor layers in RE is the most challenging part due to reduced resolution and weak contrast resulting from similar materials [7], [12]. Surface materials and topography are two common SEM image contrast sources besides Passive Voltage Contrast (PVC). Surface potential varies with the doping type and concentration of materials. It also affects the contrast with the accumulation of surface charge. These phenomena can be captured by SEM [17], [18]. Penetration depth is another reason for e-beam energy being critical for SEM imaging. Higher energy provides more information going into deep layers but sometimes it causes loss of contrast because of the lower signal-to-noise ratio (SNR) in thin IC layers. For better images, in [16], [17], [18] low energy ($< 5keV$) is suggested. In our later experiments, we have taken higher to lower *kev* images and found the best resolution and contrast images to support our claims.

B. IC Camouflaging

IC camouflaging is the technique of disrupting the attacker's capability to identify logic gates from images obtained after delayering and imaging by SEM. This can be achieved by creating special standard cells that leverage doping concentration [19], dummy contacts [11], threshold voltage configuration [12] and interconnect obfuscation [20]. When an attacker delayers and images the IC, he/she should be incapable of revealing the camouflaged gates' true identity. As a result, some of the gates in the netlist obtained through the RE process become obscure as shown in Fig.1(a-b). All these techniques revolve around the same concept of creating a particular cell that can implement various types of functions depending on

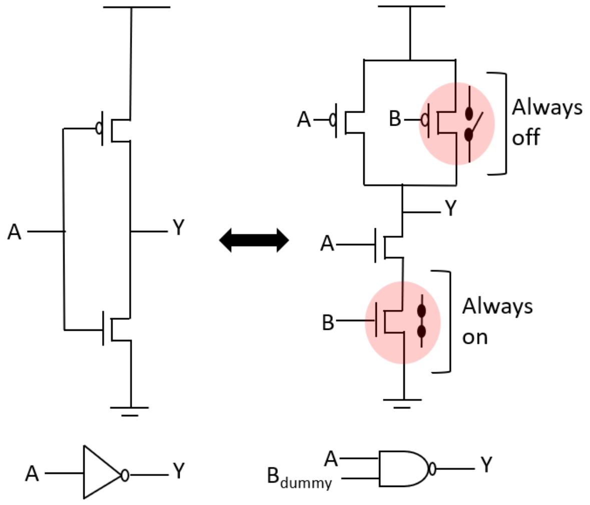

Fig. 2: Two input NAND gate behaving like an inverter with genuine input A and dummy input B. The attacker should not be able to determine that B is a dummy by SEM imaging. Thus, in the attacker's recovered netlist, it will appear as a NAND gate rather than an inverter.

how they have been configured or fabricated. For example, in Fig.1(b), the gates marked with '?' can exhibit NAND or NOR functionality depending on the contacts used.

Unfortunately, these cells can be compromised by exhaustive tests after obtaining a netlist by RE or by using advanced probing tools (e.g., LVP, FIB Microprobing) to make direct contact with camouflaged gates revealing their identity. It has also been shown that these camouflaged cells can be identified in minutes by SAT [9], [10] and ATPG [11] even if their precise functions are obscure during SEM imaging and RE steps.

C. Covert Gates

Considering the limitations and vulnerabilities mentioned above, we proposed a camouflaging technique in [15], reinstating the literal meaning of 'camouflaging' by creating 'Covert Gates' that leverage dummy contacts and doping concentration. Our primary focus was to create logic gates that resemble the regular gates in the design under SEM imaging but do not impact the circuit's original functionality because they contain one or more dummy inputs that have no effect on the logic gate's output. Nevertheless, since the SEM imaging cannot identify the dummy inputs, the netlist recovered by the attacker will not exhibit the correct functionality. The attacker has to first find which gates are covert gates before trying to execute an attack that identifies the dummy inputs. For example, in Fig.1(c), the gates marked with 'red' boxes have added dummy inputs compared to Fig.1(b-c), but since all gates appear as normal, all interconnects are suspect (marked by '?'). This approach's beauty is that the attacker cannot quickly tell which gates are camouflaged and which are not compared to other techniques. They have to perform the imaging and probing practically in every gate of a large design, which leads to an infeasible time and cost for the attacker. This significant increase in attack complexity is also applicable for SAT- and ATPG-based attacks. As to resolve the covert gates, the first one will need to encode a vast number of gates, and the second one will also be unable to generate and propagate the enormous number of test patterns to distinguish the gates.

Operating Principle: Variants of regular PMOS and NMOS are required to create standard gates like covert gates with dummy inputs. Regular PMOS and NMOS are only active when voltages with magnitudes greater than their threshold voltages are applied to their gates. For PMOS, the threshold voltage is negative and for NMOS it is positive. As the basis for covert gates, one must create NMOS and PMOS that are always off (open) or always on (closed), regardless of the gates' applied inputs. Using these variants, it is possible to create inputs that have no effect on a CMOS logic gate. For example, Fig.2 shows a two-input NAND with $Y = (AB)^0$ that actually behaves as an inverter ($Y = A^0$) because the input B is a dummy. An always-on transistor (shorted wire) in the pull-down network and always off transistor (open wire) in pull-up network means that the applied input on pin B shows no effect on the functionality of the NAND gate. An always-off transistor can be implemented by leveraging dummy contacts. An always-on transistor can be created with a heavily doped channel with the same dopants as the source and drain. This always-on transistor is out of the scope of this paper as it is already discussed thoroughly in our previous work [15]. Our primary focus will therefore be on always-off transistor where we will consider the structure closer to the real scenarios in different imaging conditions.

Dummy Contact Prototype Fabrication

In this section, we describe the construction of an always off transistor prototype to facilitate a more accurate analysis of covert gate indistinguishability under SEM. In [11], dummy contacts were initially proposed for camouflaging. Specifically, thin insulated films in contacts were used to interrupt the connection of transistor terminals to metal layers. This electrical disconnection of dummy contact-based camouflaging was used in 19 contacts within a single cell, allowing the cell to exhibit three functionalities (NAND, NOR, and XOR) depending on the real vs. dummy contacts. In our proposed covert gate approach, a thin insulating layer is required in between a metal connection and the contacts, not the full contacts. So, the advantage here is that these dummy contacts are quite similar to real contacts but only with a thin insulating layer to create the disconnection. Due to the insulating layer's use, the bias that is supposed to turn on the transistor cannot be effectively applied to the gate. Moreover, the insulating layer on the source contact cuts the transistor connection to VDD (for pullup networks) and GND (for pull-down networks). As a result, even with applied bias, the transistor always remains off by preventing the channel from forming a conducting channel. Another difference between [11] and the proposed approach is that we incorporate always-on transistors into the same cells as always-off transistors. This allows us to achieve dummy inputs rather than multiple logic functions. In our previous prototype [15], silicon doping manipulations and dummy contact structures were presented for SEM imaging analysis. However, in real devices, contacts can be connected with different material in various

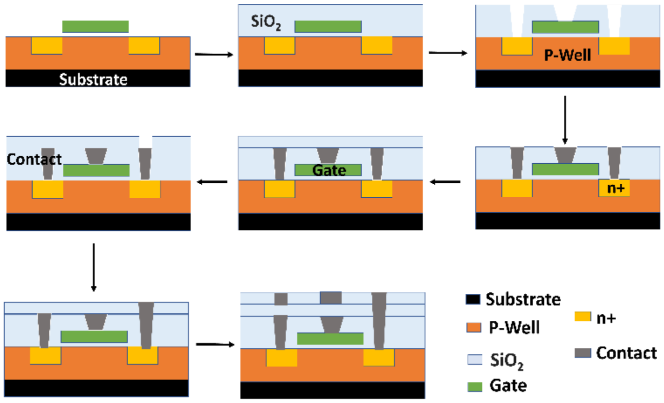

Fig. 3: Fabrication of contacts on always-off transistors by subsequent deposition, etching and filling.

structures, which may affect the charging from the e-beam in SEM imaging. Due to the PVC in SEM imaging, it is theoretically possible to differentiate different charging status, thus raising concerns about the robustness of the covert gate design against SEM based RE.

To further study this, we fabricated different contact structures, including both real contacts and dummy contacts, as later shown in Fig. 4. Different from the previous fabrication, now dummy contacts/vias are formed by the deposition and etching process, as shown in Fig. 3. At first, the dielectric material is deposited over the whole active region, and then a portion of it is etched to create trench contacts. After that, trench contacts are filled with metal, followed by further deposition, etching, and filling. The thin dielectric layer's sufficient thickness in between contacts serves as a stopper to prevent the switching of the transistor when it is supposed to work. The purpose is to create such dummy contacts that look like real contacts but are not functional. Thus, when an SEM image is taken from the top or bottom of an IC, the surface images of dummy contacts and real contacts remain indistinguishable.

A sample prototype is shown as in Fig. 4 to analyze the difference in imaging between dummy contacts and regular contacts. Due to the fab facility's limitation, we used gold as the metal material as it can be easily deposited and does not have oxidation like copper during fabrication. In Fig. 4, cases 1 and 2 represent regular contacts whereas 3 and 4 represent the respective counterpart dummy contact. Case 5 represents another regular contact with smaller charging volume to compare with other cases. The diameter and height are designed to be the same for both regular contacts and dummy contacts. The only difference is that there is a 17 nm insulated layer inserted in the dummy contacts that prevent forming a conducting channel. As a result, the transistors with dummy contacts always remain off despite an applied bias. The main goal of our experiment is to see whether the SEM images can detect this difference in dummy contacts and regular contacts through manual analysis as well as machine learning algorithm-based analysis.

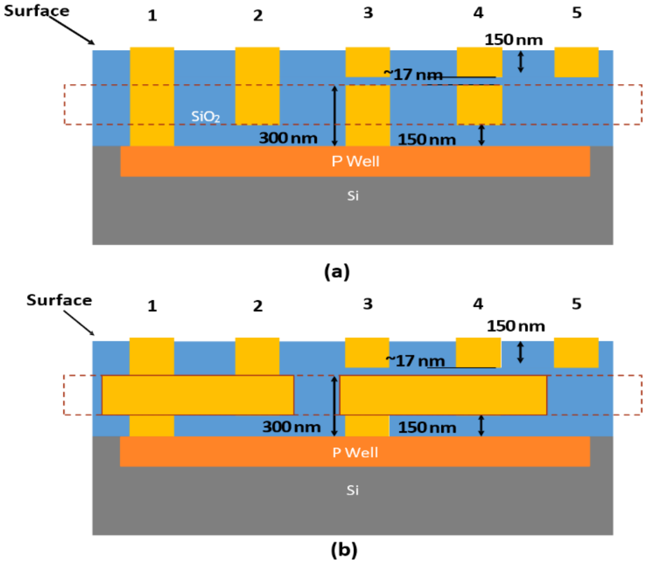

Fig. 4: Proposed Prototype of regular and dummy contacts being 1, 2 real contacts along with respective counterpart dummy contacts 3, 4. (a) represents regular and dummy contacts connected to silicon and well whereas (b) also adds metal layer connection with contacts.

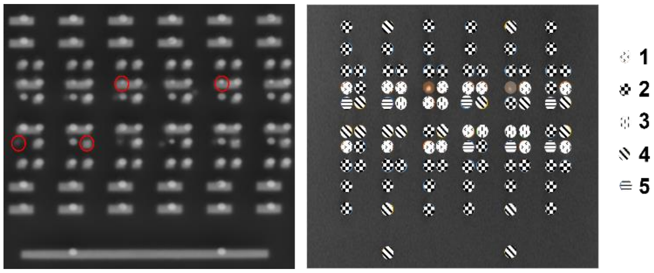

Fig. 5: Fabricated prototype along with color code (right) representing different types of contacts proposed in Fig.4. Red circles (left) show the failed contacts due to limited yield.

Prototype Indistinguishability Analysis

A. SEM Imaging Setup and Parameters

A fabricated sample prototype is shown in Fig. 5 and SEM images are taken with TESCAN FERA3 FIB-SEM. The contacts are the bright circles shown in Fig. 5 (left), whose associated structures are indicated with different colored circles in Fig. 5 (right). The colors are shown in the legend on the far right and correspond to the different cases in Fig. 4. The gray trace shown in Fig. 5 (left) are the metal traces shown in Fig. 4b, increasing the charging volume of the contacts that go through the traces. Due to the yield loss (e.g., residue photoresist), some contact structures were not successfully fabricated as expected, which are circled in red in Fig. 5 (left). To systematically study the possible effects of different structures on SEM imaging, we manipulated the e-beam energy from 1 keV to 25 keV, varied the dwell time from 3μs per pixel to 320μs per pixel, and employed both SE and BSE detectors.

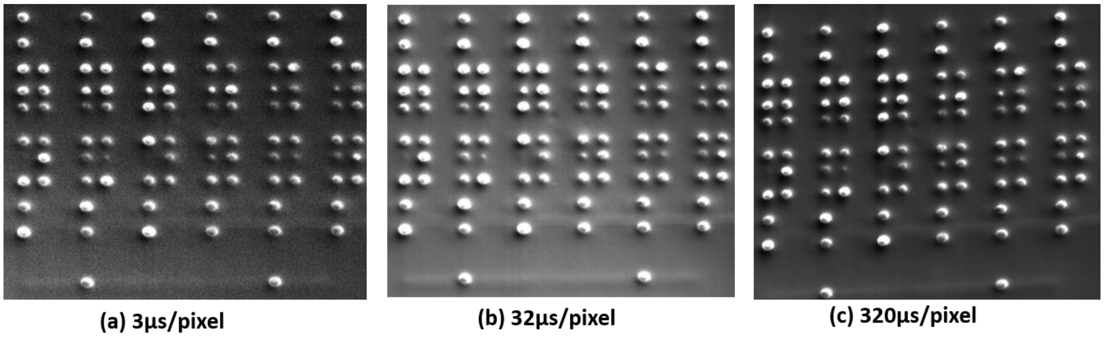

Fig. 6: Comparison of dwell times in SE detector at 1 KeV.

B. Manual Analysis

Based on the images taken with the detectors at different operating conditions of e-beam energies and dwell time, a manual analysis is performed. We first took the image at a low e-beam energy (1 keV) to focus on the shallow surface of the sample. Because 1 keV is lower than the threshold of BSE detector, the BSE imaging is not taken under this condition. As shown in Fig. 6, we can see that the 1 keV SE imaging is only sensitive to the sample surface condition. The contacts showing the highest intensity (white) are not related to the structure (comparing Fig. 6 with Fig. 5). We believe the variation in intensity can be contributed to the surface residues from sample cleaning process, which is not helpful for the differentiation of contact structures.

Distinguishability Under High keV Imaging. We then increased the e-beam energy to 5 keV, 15 keV, and 25 keV for SE and BSE with different dwell time. The imaging results from 5 keV are close to 15 keV showing slightly better image qualities. While with 25 keV, the samples under observation were quickly charged due to the large SiO_2 surface and thus result in poor image quality. The SE and BSE images taken with 15 keV and different dwell time are shown in Fig. 7. Overall, we can see that the longer dwelling time provides slightly better signal-to-noise-ratio (SNR), but nothing is essentially different on the images taken with shorter dwelling time. On both SE and BSE images, the metal layers beneath the sample surface are shown, indicating the electrons of 15 keV penetrated deeply enough to convey the structure information from the corresponding regions. As expected, the SE imaging is more sensitive to the surface while BSE is more sensitive to the material elements; therefore, the contacts exposed on surface are emphasized more (or brighter) in the SE images than they are in the BSE images.

In either SE or BSE case, it seems difficult to categorize the contacts into the corresponding structures based on the intensity by simply looking at the images. Therefore, we performed a qualitative analysis, as shown in Fig. 8.

Distinguishability in SE Images. According to the analysis of SE images in Fig. 8 (a, b), the intensity distributions of structures 1, 2, 3, and 4 are highly overlapping with each other, making it infeasible to differentiate the structure. With respect to structure 5, the mean intensity is significantly lower than other structures and the distribution of intensity is barely overlapping with other structures. This can be explained by the considerably smaller charging volume of contacts in Structure

5. Charges resulted from e-beam will be more easily accumulated on contacts in structure 5, lowering the intensity. While on structures 3 and 4, although the upper part of the contacts is not directly connected to the silicon, the shallow gap (17 nm) allows sneaky charge releasing and reduces the charge accumulation. Therefore, on SEM images, we can point out the contact in structure 5 from other four types of structures but cannot differentiate the remaining four structures which are the ones that are most important to an attacker.

Distinguishability in BSE Images. Compared to SE images, BSE images are less sensitive to PVC. To observe the difference among different contact structures based on PVC will be more difficult. As shown in Fig. 8 (c, d), although the mean intensity of structure 5 is still lower than the mean intensity other structures, the distribution is overlapping with the distribution of others. Given BSE observation result of a single contact, it is no longer possible to tell if this is structure 5 or not.

Takeaway. In summary, structures 3 and 4 are used as dummy contacts for always-off transistors while structures 1 and 2 are used as real contacts for regular transistors. Our analysis shows that although PVC results in different intensity for different contact structures, the differences are not enough to differentiate the real contacts (structure 1 or 2) from the dummy contacts (structure 3 or 4), as the intensity variations of each structure are larger than the difference of the mean intensities. In other words, the intensity distributions of structures 1, 2, 3, and 4 are largely overlapped. *Based on this manual and preliminary quantitative analysis, our claim that always-off transistors in covert gates (corresponding to dummy inputs) are indistinguishable from normal transistors in standard logic gates still holds.*

B. Machine Learning Based Classification

1) Preliminaries: A proper analysis of the camouflaging ability of these covert gates requires inspection through pattern recognition to evaluate their indistinguishability in an automated fashion. A big advantage of our approach is the impractical time and resources required to simply identify the camouflaged contacts in a design prior to any further reverse engineering. Therefore, it is important to ensure their camouflage even if an attacker were to use an automated computational analysis that would drastically reduce cost. To

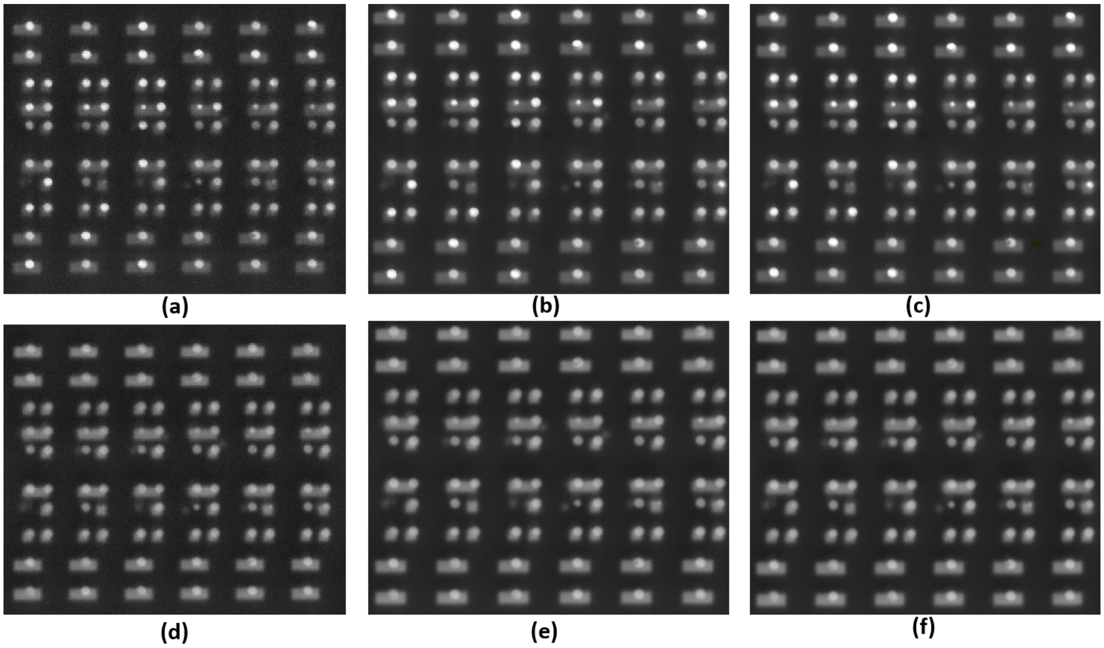

Fig. 7: Comparison of SE (a, b, c) and BSE (d, e, f) detector images at 15 KeV energy with dwell times 3μs per pixel, 32μs per pixel, and 320μs per pixel, respectively.

do so, we utilized a series of practical pattern recognition techniques for classification in combination with a statistical analysis for explainability of our result.

Pre-processing Steps. As mentioned earlier, the data of covert and genuine contacts has been set up into 4 different classes depending on the detector type and under a variety of imaging conditions. These imaging conditions have a very close relationship between the amount of noise one can expect in the scanned contact's image and the amount of time necessary to scan. An effective automated analysis and detection method ideally should be able to identify these contacts in the presence of noise to leverage the minimal scan time these conditions provide. Thus, some pre-processing is required before any further pattern recognition and classification analysis. This pre-processing consists of simply utilizing a *median filter* to remove the characteristic salt and pepper noise seen in lower resolution images, followed by *thresholding to segment* the contact from its surroundings and finally *registration* to center all detections in the same Euclidean space.

Feature Representations. Now with the data properly preprocessed, we begin our efforts of effectively discriminating genuine contacts from covert camouflaged contacts in these images. The first step in any machine learning or pattern recognition methodology is to effectively identify the features to use in one's classifier. Since we are focusing on contacts from SEM images, the features that we predominantly will utilize, especially in an unsupervised fashion, are the intensity profiles of the contacts in question and their respective shape features. The shape features we utilize are the area, the perimeter, the circularity, and the average intensity of the extracted contact. These shape features are particularly useful since they give us insight into the characterization of the contacts we are looking to classify. Ideally, genuine contacts should have a similar profile as far as intensities across their structure. Not to mention their area and perimeter measures should be relatively consistent. Lastly, circularity, which measures how circular an object is, is a key characteristic of expected genuine contacts as that is their prototypical shape.

Classification Algorithms and Evaluation. With all this in mind, we evaluate our classifiers based on using solely the intensity information versus using solely the shape feature information. Specifically, we focus on unsupervised techniques (no user intervention or a prior knowledge incorporated) that can be completed with minimal data. We do not utilize the current trend of deep/supervised learning for classification due to their necessity of large amounts in training data (tens to hundreds of thousands of samples). In a practical use case, the reverse engineer or attacker will not have a database of these covert contacts in this design available for supervised techniques that require vast volumes of training data. While there would be a vast improvement in performance if deep learning was applied there currently is not enough data of camouflaged contacts to effectively train an artificial neural network (ANN) or convolutional neural network (CNN) for classification between covert and genuine gates.

For our experiments we also experience the very real likelihood of class imbalance/skew. Specifically, there are instances where the contacts imaged from one class outnumber its counterpart. This is the case for our experiments where often

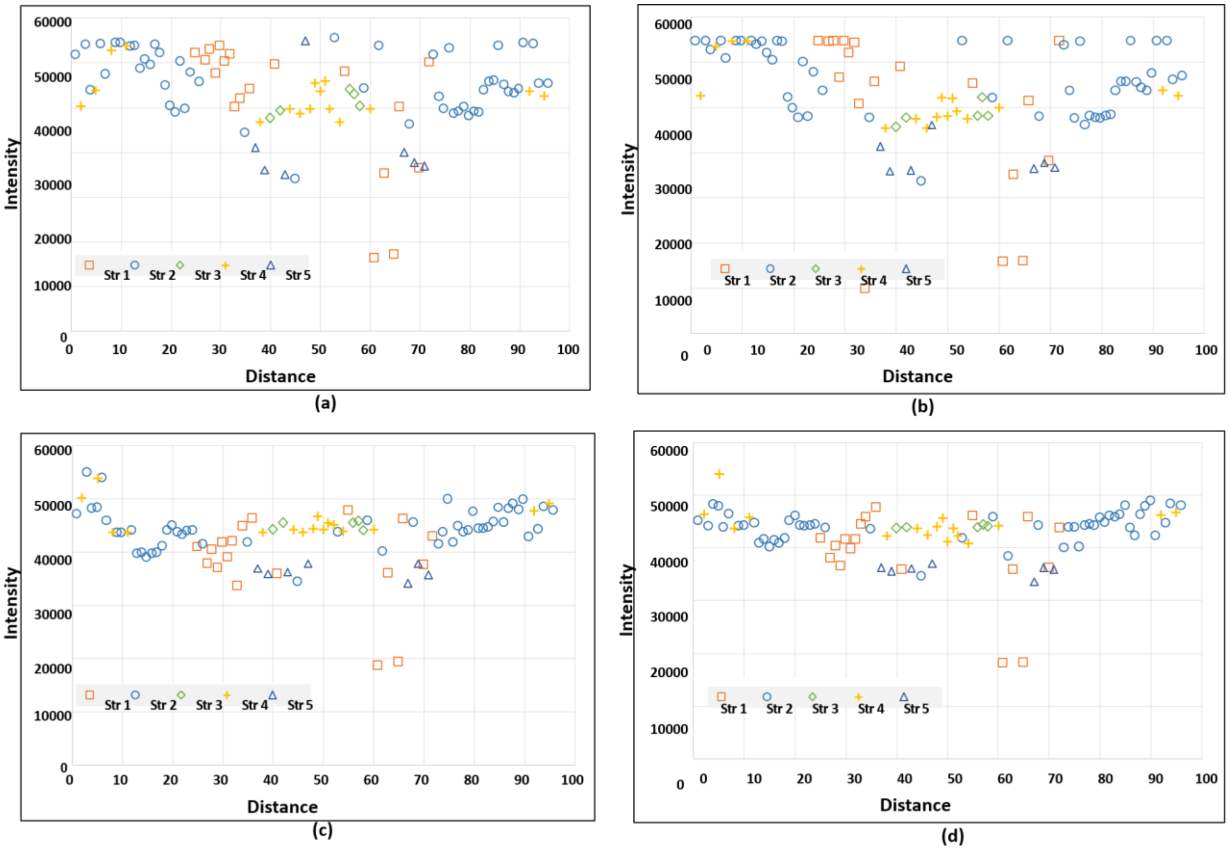

Fig. 8: Intensity Profile of SE (a, b at 15 KeV with dwell times 3µs per pixel, 320µs per pixel correspondingly) and BSE (c, d at 15 KeV with dwell times 3µs per pixel, 320µs per pixel correspondingly) images at different operating conditions. Note data points below 20,000 intensities are from failed contact structures.

the number of genuine contacts is twice as large, if not more, as the covert contact class itself. This is another reason that supervised learning techniques are avoided – to minimize the likelihood of overfitting due to class imbalance. Instead, we propose a technique that consists of assuming the attacker has access to a small set of data that can be used to create a characteristic reference for each class. This characteristic reference, referred to as a centroid of the features used in classification, is simply the averages of all the intensities for each pixel location for the intensity-based approach or the average of every sample area, perimeter, circularity, and average intensity for the shape feature approach. Afterwards, we evaluate the likelihood of a sample classifying to one class versus the other by measuring the distance between the sample's feature vector and the characteristic centroid of the covert contact versus the genuine contact.

Additionally, we performed classification experiments using the popular K-means clustering algorithm since this is another popular unsupervised learning algorithm, albeit often used for gaining intuition about one's data as opposed to classification. Nonetheless, it is reasonable to assume if these genuine contacts and covert contacts were easily discriminated, they should cluster together effectively.

For our centroid based classification methodology, we evaluate intensity and feature focused approaches using Euclidean distance and cosine similarity as the objective functions. We use these metrics since they are the two predominant distance metrics for data like ours. For each, a score of 0 means a perfect match and a score of 1 is a complete mismatch. We evaluate this methodology using a sweeping threshold from 0 to 1 for each threshold to evaluate the sensitivity of each class and the respective thresholds production of true and false positive classifications. Furthermore, to address the class imbalance and provide a fair comparison between covert and genuine contact characterization/classification, we randomly select the same number of samples of the class with a greater number of samples so that it matches the number of samples of the smaller class. Then, we compute the centroids of each class with these samples and compute the false positives and true positives based on the sweeping thresholds as mentioned earlier. Using k-fold validation, we perform this action 10 times and average the results to acquire a good estimation of the performance of the covert gates versus the genuine gates. Additionally, we use the same partitions of data and perform k-means clustering on the data to evaluate the average classification performance with this algorithm. Lastly, we are focused on two instances of comparing genuine gates versus covert gates. We refer to each

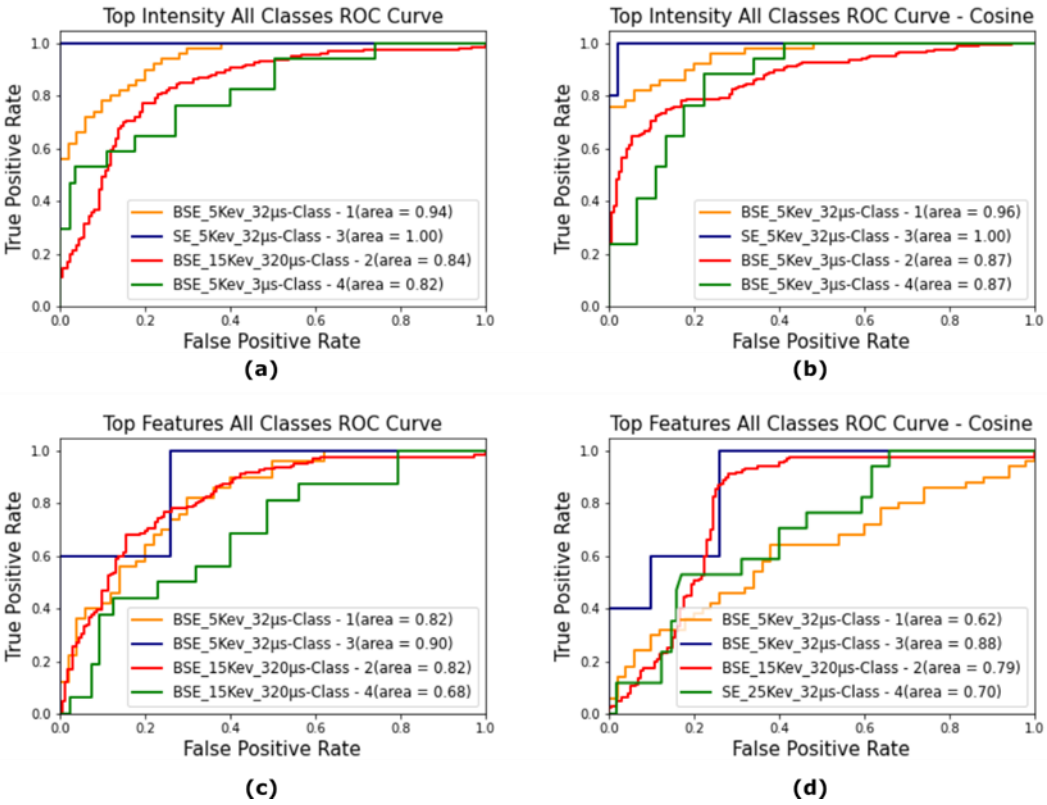

Fig. 9: Best ROC curve results for centroid based classification utilizing (a) intensities with Euclidean distance (b) intensities with cosine distance (c) shape features with Euclidean distance (d) shape features with cosine distance.

class 1 vs. 3 and class 2 vs. 4, which correspond to the class in Fig. 4.

2) Results & Discussion: Following the experiments outlined above, we present our results showing the discriminatory ability of pattern recognition approaches between covert and genuine gates. A good representation of our results using the centroid-based classification technique is the receiver operating characteristic (ROC) curves. These curves show the performance of a classification model at all classification thresholds and plots two parameters: True Positive Rate (TPR) and False Positive Rate (FPR). An ideal model would yield a point in the upper left most corner signifying a TPR of 100% and FPR of 0% for all thresholds. Additionally, we can compute the area under the ROC curve to provide a scalar estimation of the model's performance.

In Fig. 9, we provide the ROC curves of top performing models for classes 1 through 4 across all SEM HV and dwell time imaging conditions for both intensity and feature based approaches using Euclidean distance and cosine similarity. In Tables I and II, we also provide the area under the ROC curve (AUC) for intensity-based classifiers and feature-based classifiers, respectively, for all imaging conditions. For the most part, all models perform sub-par aside from the images of class 3 taken by SE detector at 5 KeV energy with dwell time 32µs per pixel. In that case, the ROC curves display a score of 100% and 99% for intensity-based approach using Euclidean and cosine distance, respectively. However, this was definitely an outlier. In fact, class 3 was the top performer in almost every instance when looking at Tables I and II. Nevertheless, as one can see from the other curves in Fig. 9 and scores in Tables I and II, there remains a lot to be desired in terms of effective classification performance.

These results are reinforced when observing the accuracy scores from using K-means clustering instead of our centroid based approach. Again, K-means is traditionally more for data comprehension than classification but if our data was easily discriminated in an unsupervised fashion the covert and genuine contacts should cluster well. We see in Table III that is not the case with scores barely above chance regardless of using shape or intensity profiles as the feature vector for sample clustering. The best-case scenario is 74% accuracy.

Takeaway. Based on these results, we can effectively say that any of our methods utilizing automation will have great difficulty for classification of covert versus genuine contacts. The vast majority of scores regardless of distance metric, utilizing shape-based features or intensities, or imaging conditions have sub-par results. Furthermore, the best performing instances of utilizing intensity-based classification with cosine similarity still leave room for improvement. Especially considering the poor results utilizing the same methodology for classes 2 and 4. Classes 1 and 3 are a substantially smaller set comparatively to 2 and 4 so it is likely that the better results comparatively are due to working with this smaller subset of data and when extrapolating out to larger sets of data with more variance we see results more like classes 3 and 4.

TABLE I: ROC area under curve (AUC) after k-fold cross validation for all imaging conditions using intensity-based classifiers using Euclidean distance (I.E.) and cosine similarity (I.C.). The best result for each class and similarity measure are highlighted in bold.

Detector	**Imaging Conditions**		**Class 1**		**Class 3**		**Class 2**		**Class 4**	
	Energy (KeV)	Dwell Time (μs/pixel)	I.E.	I.C.	I.E.	I.C.	I.E.	I.C.	I.E.	I.C.
BSE	5	3	.87	.88	.92	.95	.53	.80	**.80**	.82
	5	32	**.89**	.83	.86	.95	.59	.81	**.80**	**.84**
	5	320	.70	**.90**	.98	.96	.62	.83	.72	.80
	15	3	.79	.80	.64	.82	.78	.83	.59	.82
	15	32	.70	.77	.52	.82	.76	.78	.52	.75
	15	320	.65	.84	.61	.94	**.82**	**.84**	.69	.80
	25	3	.74	.93	.74	.91	.68	.72	.51	.71
SE	5	3	.35	.84	.98	.95	.60	.75	.68	.76
	5	32	.43	.88	**1.0**	**.99**	.52	.69	.64	.70
	15	3	.43	.79	.94	.78	.48	.72	.68	.75
	15	32	.49	.83	.92	.84	.47	.69	.65	.72
	15	320	.41	.78	.94	.90	.53	.69	.63	.71
	25	3	.47	.83	.71	.80	.45	.67	.65	.72
	25	32	.34	.85	.96	.97	.36	.72	.79	.76
	25	320	.43	.89	.96	.98	.47	.79	.73	.80

TABLE II: ROC area under curve (AUC) after k-fold cross validation for all imaging conditions using feature-based classifiers using Euclidean distance (F.E.) and cosine similarity (F.C.). The best result for each class and similarity measure are highlighted in bold.

Detector	**Imaging Conditions**		**Class 1**		**Class 3**		**Class 2**		**Class 4**	
	Energy (KeV)	Dwell Time (μs/pixel)	F.E.	F.C.	F.E.	F.C.	F.E.	F.C.	F.E.	F.C.
BSE	5	3	.74	.53	.81	.72	.53	.57	.49	.50
	5	32	.75	.64	.85	.84	.47	.46	.62	.55
	5	320	.56	.43	.64	.78	.59	.60	.48	.42
	15	3	.73	.55	.32	.55	.71	.70	.56	.45
	15	32	.59	.47	.42	.46	.70	.71	.54	.50
	15	320	.54	.53	.55	.68	**.80**	.77	**.68**	.61
	25	3	.61	.57	.34	.52	.71	.72	.43	.37
SE	5	3	.47	.58	.76	**.88**	.72	.65	.36	.36
	5	32	.69	**.66**	**.90**	.84	.51	.56	.54	.48
	15	3	.46	.48	.56	.71	.50	.61	.56	.45
	15	32	.65	.43	.60	.75	.57	.71	.50	.33
	15	320	.32	.30	.79	**.88**	.58	.65	.47	.38
	25	3	.40	.45	.63	.50	.54	.52	.49	.52
	25	32	.41	.35	.83	.87	.44	.36	.59	**.70**
	25	320	.50	.50	.61	.67	.58	.50	.41	.52

TABLE III: K-Means accuracy after k-fold cross validation for all imaging conditions using intensity and feature based approaches. The best result for each classification scenario and approach are highlighted in bold.

Detector	**Imaging Conditions**		**Intensity**		**Features**	
	Energy (KeV)	Dwell Time (μs/pixel)	Class - 1 V. 3	Class - 2 V. 4	Class - 1 V. 3	Class - 2 V. 4
BSE	5	3	.66	.44	**.74**	.45
	5	32	**.67**	.52	.67	.44
	5	320	.55	.47	.54	.45
	15	3	.65	.47	.61	.57
	15	32	.46	.51	.55	.49
	15	320	.60	.53	.56	.72
	25	3	.55	.49	.53	.49
SE	5	3	.49	.57	.46	.50
	5	32	.44	.48	.50	.50
	15	3	.54	.51	.60	.52
	15	32	.60	.57	.58	.53
	15	320	.42	.55	.39	.53
	25	3	.57	.46	.58	.48
	25	32	.54	.49	.52	.51
	25	320	.63	.50	.53	.52

Conclusion and Future Work

In this study, we studied a modification to the fabrication processing of dummy contacts in covert gate camouflaging design. Then we fabricated real and dummy contacts in different structures, where the contacts are connecting silicon and/or metal with different charging volumes. We tried various combinations of e-beam energy and dwell time during SEM imaging to observe the possible PVC from different contact structures. According to our SEM images, it is difficult to correctly point out whether a contact is a real one or a dummy one. Machine-learning-based pattern recognition was also performed, further confirming the insignificant difference between real and dummy contact structures in SEM imaging. Therefore, the claim still holds that SEM imaging cannot correctly find out the dummy contacts at this stage, and the covert gate design is believed to be robust against SEM imaging-based reverse engineering. In the near future, we plan on imaging our samples under helium-neon ion beam (He-Ne) microscope, which typically offer better PVC. In addition, our research on the covert gate will move to performance verification. We will develop device models to evaluate the essential properties and build model cards of different camouflaged gates for circuit-level simulations. We will also explore tape-out opportunities to fabricate chips for more comprehensive and practical covert gate camouflaging design tests. Lastly, future work can see the acquisition of larger volumes of data to experiment and verify whether the indistinguishability of covert gates versus genuine maintains when utilizing deep learning or other supervised learning methods that are able to learn more non-linear discriminatory functions.

References

[1] C. Helfmeier, D. Nedospasov, C. Tarnovsky, J. S. Krissler, C. Boit, and J.-P. Seifert, “Breaking and entering through the silicon,” in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 733–744.

[2] H. Lohrke, S. Tajik, C. Boit, and J.-P. Seifert, “No place to hide: Contactless probing of secret data on fpgas,” in *International Conference on Cryptographic Hardware and Embedded Systems*. Springer, 2016, pp. 147–167.

[3] H. Wang, D. Forte, M. M. Tehranipoor, and Q. Shi, “Probing attacks on integrated circuits: Challenges and research opportunities,” *IEEE Design & Test*, vol. 34, no. 5, pp. 63–71, 2017.

[4] T. Krachenfels, F. Ganji, A. Moradi, S. Tajik, and J.-P. Seifert, “Realworld snapshots vs. theory: Questioning the t-probing security model,” *arXiv preprint arXiv:2009.04263*, 2020.

[5] C. Boit, C. Helfmeier, and U. Kerst, “Security risks posed by modern ic debug and diagnosis tools,” in *2013 Workshop on Fault Diagnosis and Tolerance in Cryptography*. IEEE, 2013, pp. 3–11.

[6] C. Boit, T. Kiyan, T. Krachenfels, and J.-P. Seifert, “Logic state imaging from fa techniques for special applications to one of the most powerful hardware security side-channel

threats," in *2020 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*. IEEE, 2020, pp. 1–7.

[7] R. Torrance and D. James, "Reverse engineering in the semiconductor industry," in *2007 IEEE Custom Integrated Circuits Conference*. IEEE, 2007, pp. 429–436.

[8] E. L. Principe, N. Asadizanjani, D. Forte, M. Tehranipoor, R. Chivas, M. DiBattista, S. Silverman, M. Marsh, N. Piche, and J. Mastovich, "Steps toward automated deprocessing of integrated circuits," in *ISTFA 2017: Proceedings from the 43rd International Symposium for Testing and Failure Analysis*. ASM International, 2017, p. 285.

[9] M. El Massad, S. Garg, and M. V. Tripunitara, "Integrated circuit (ic) decamouflaging: Reverse engineering camouflaged ics within minutes." in *NDSS*, 2015, pp. 1–14.

[10] P. Subramanyan, S. Ray, and S. Malik, "Evaluating the security of logic encryption algorithms," in *2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. IEEE, 2015, pp. 137–143.

[11] J. Rajendran, M. Sam, O. Sinanoglu, and R. Karri, "Security analysis of integrated circuit camouflaging," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 709–720.

[12] M. I. M. Collantes, M. El Massad, and S. Garg, "Threshold-dependent camouflaged cells to secure circuits against reverse engineering attacks," in *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2016, pp. 443–448.

[13] A. S. Iyengar, D. Vontela, I. Reddy, S. Ghosh, S. Motaman, and J.w. Jang, "Threshold defined camouflaged gates in 65nm technology for reverse engineering protection," in *Proceedings of the International Symposium on Low Power Electronics and Design*, 2018, pp. 1–6.

[14] N. E. C. Akkaya, B. Erbagci, and K. Mai, "A secure camouflaged logic family using post-manufacturing programming with a 3.6 ghz adder prototype in 65nm cmos at 1v nominal v dd," in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 2018, pp. 128–130.

[15] B. Shakya, H. Shen, M. Tehranipoor, and D. Forte, "Covert gates: Protecting integrated circuits with undetectable camouflaging," *IACR Transactions on Cryptographic Hardware and Embedded Systems*, pp. 86–118, 2019.

[16] T. Sugawara, D. Suzuki, R. Fujii, S. Tawa, R. Hori, M. Shiozaki, and T. Fujino, "Reversing stealthy dopant-level circuits," in *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 2014, pp. 112–126.

[17] A. K. Chee, "Quantitative dopant profiling by energy filtering in the scanning electron microscope," *IEEE Transactions on Device and Materials Reliability*, vol. 16, no. 2, pp. 138–148, 2016.

[18] S. Elliott, R. Broom, and C. Humphreys, "Dopant profiling with the scanning electron microscope—a study of si," *Journal of Applied Physics*, vol. 91, no. 11, pp. 9116–9122, 2002.

[19] S. Malik, G. T. Becker, C. Paar, and W. P. Burleson, "Development of a layout-level hardware obfuscation tool," in *2015 IEEE computer society annual symposium on VLSI*. IEEE, 2015, pp. 204–209.

[20] S. Chen, J. Chen, D. Forte, J. Di, M. Tehranipoor, and L. Wang,"Chip-level anti-reverse engineering using transformable interconnects," in *2015 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFTS)*. IEEE, 2015, pp. 109–114.