A Machine Learning Approach to Fab-of-Origin Attestation

Ali Ahmadi
Dept. of Electrical Engineering
The University of Texas at Dallas
Richardson, TX 75080

Mohammad-Mahdi Bidmeshki
Dept. of Electrical Engineering
The University of Texas at Dallas
Richardson, TX 75080

Amit Nahar
Texas Instruments Inc.
12500 TI Boulevard, MS 8741
Dallas, TX 75243

Bob Orr
Texas Instruments Inc.
12500 TI Boulevard, MS 8741
Dallas, TX 75243

Michael Pas
Texas Instruments Inc.
12500 TI Boulevard, MS 8741
Dallas, TX 75243

Yiorgos Makris
Dept. of Electrical Engineering
The University of Texas at Dallas
Richardson, TX 75080

ABSTRACT

We introduce a machine learning approach for distinguishing between integrated circuits fabricated in a ratified facility and circuits originating from an unknown or undesired source based on parametric measurements. Unlike earlier approaches, which seek to achieve the same objective in a general, design-independent manner, the proposed method leverages the interaction between the idiosyncrasies of the fabrication facility and a specific design, in order to create a customized fab-of-origin membership test for the circuit in question. Effectiveness of the proposed method is demonstrated using two large industrial datasets from a 65nm Texas Instruments RF transceiver manufactured in two different fabrication facilities.

1. INTRODUCTION

As the semiconductor industry has largely adopted the fab-less paradigm and as globalization has amplified concerns regarding integrity of the electronics supply chain, the ability to definitively identify the fabrication facility wherein an integrated circuit (IC) was manufactured has become imperative. Such a fab-of-origin attestation ability could constitute the cornerstone for numerous applications in the electronics industry, including intellectual property (IP) protection, licensing enforcement, quality and hardware integrity assurance, supply chain risk management, counterfeit IC detection and failure analysis, among others.

The importance of fab-of-origin attestation is highlighted by a recent US government research initiative whose objective is *to devise methodologies which use measurable electronic or physical characteristics for determining the specific fabrication facility of origin of a given electronic component* [4]. The various methods developed under this initiative seek to leverage the specifics of a manufacturing process, such as the use of particular materials or geometric rules during fabrication, in order to identify the fab-of-origin. Utilizing on-die laser markings during fabrication, atomic force microscopy (AFT), nanoscale structural, mechanical and electrical characterization based on transmission electron microscopy, device characterization, and using features of spectroscopic chemical signals from electronic components for identifying the source fab are among the explored directions [4]. All of these approaches, however, require additional complicated steps during manufacturing or specialized and expensive equipment during characterization in order to perform fab-of-origin attestation. An earlier approach to address this problem was introduced in [3], where the authors proposed a methodology to etch forensic information regarding the fabrication process on the die, or add memory elements to electrically store such data. While recovering this information can reveal the fabrication facility, this approach also incurs additional processing and characterization overhead and introduces the need for measures to ensure that the forensic information has not been tampered with.

Along a different direction, a methodology which leverages intrinsic variation of the semiconductor manufacturing process for foundry identification purposes was introduced in [16]. Process variation has been successfully exploited in various other tasks, including IC identification through physical unclonable functions (PUFs) [7, 11], hardware Trojan detection [2, 13], as well as counterfeit IC detection [5, 8]. The authors of [16], however, were the first to demonstrate its utility in this context. Specifically, they introduced a methodology for reverse engineering process parameters, such as threshold voltages and effective channel length of CMOS devices, using gate delay measurements obtained through an elegant path decomposition formulation. Statistical tests comparing the distribution of these parameters to the profiles of known foundries may, then, be used to identify which foundry fabricated the IC in question. While this method is design-independent, it requires access to the gate level implementation of the fabricated IC in order to reverse engineer these process parameters, which may pose an obstacle due to IP protection issues. Moreover, as explained in [16], reverse engineering of these parameters can become quite complicated in practice. Nevertheless, leveraging intrinsic process variation and simple electronic measurements for the purpose of fab-of-origin attestation is very appealing, as it does not require specialized process or characterization efforts and can be readily applied to existing ICs.

ICCAD '16 November 07-10, 2016, Austin, TX, USA
© 2016 ACM. ISBN 978-1-4503-4466-1/16/11…$15.00
DOI: http://dx.doi.org/10.1145/2966986.2966992

In contrast to the aforementioned design-independent approaches to fab-of-origin attestation, in this paper we introduce a machine learning approach which leverages the *interaction* between the idiosyncrasies of a fabrication facility and a particular design. Specifically, we develop solutions for four variants of the fab-of-origin attestation problem. The first two variants assume availability of the test data profile from the ratified fabrication facility *only*, and seek to attest whether a single chip or a batch of chips, respectively, has been fabricated therein or not. The other two variants assume availability of the test data profile from *all* facilities which fabricate this chip and seek to identify whether a single chip or a batch of chips, respectively, were fabricated in the ratified fab or not. The proposed solutions rely only on the typical parametric test measurements[1] of a fabricated IC and require neither knowledge of the design, nor any additional provisions during manufacturing or any specialized measurement equipment.

Effectiveness of the proposed solutions is demonstrated using two large industrial datasets from a 65nm Texas Instruments RF transceiver produced in two geographically dispersed fabrication facilities. Considering that alternate fabrication facilities within the same company are highly tuned to resemble each other as much as possible, we point out that our evaluation is performed not only using realistic datasets but also ones that are very hard to tell apart.

The rest of this paper is organized as follows. In Section 2, we briefly define the four variants of the fab-of-origin attestation problem which we address herein. In Section 3, we describe the proposed solutions. In Section 4, we experimentally assess the accuracy of these solutions and, in Section 5, we draw conclusions.

2. PROBLEM DEFINITION

The methods proposed in this work seek to identify whether an IC was manufactured in a ratified fabrication facility based solely on the parametric measurements obtained during post-manufacturing testing. We note that these measurements have predefined acceptable ranges; any IC whose values fall outside these ranges is considered faulty and is discarded. Hence, what we are seeking is the ability to distinguish between the footprints of healthy chips from the ratified fab and the footprints of healthy chips from other fabs *within* the hyper-dimensional parametric space of acceptable performances. The conjecture here is that, for the same design and process, certain idiosyncrasies stemming from manufacturing tool installations, chemical sources, as well as altitude and geomagnetic location of the fabrication facility, lead to minor, yet systematic disparities in the resulting products of different fabs. These disparities may, therefore, be leveraged through machine learning methods in order to attest the source of origin of a given IC [4].

Four variants of the fab-of-origin attestation problem are considered herein:

- **AttestMe-I:** In this variant of the fab-of-origin attestation problem we assume that the only data to which we have access is the parametric test data profile from a statistically significant number of chips manufactured in the ratified fab. Given this profile and the parametric tests of a single IC, we seek to decide whether it was manufactured in the ratified fab or not.
- **AttestUs-I:** This variant assumes availability of the same information as above; instead of making a decision for a single IC, however, it considers the parametric tests of an entire batch of ICs and seeks to make a collective decision for the batch, assuming that they were all manufactured in the same fabrication facility.
- **AttestMe-II:** In this variant, we assume availability of the parametric test data profile from a statistically significant number of chips manufactured in each of the fabs wherein a given design could have been produced. Given these profiles and the parametric tests of a single IC, we seek to decide whether it was produced by the ratified fab or any other fab.
- **AttestUs-II:** Using the same information as above, this variant seeks to decide whether an entire batch of ICs, originating from the same facility, was manufactured in the ratified fab or any other fab.

We note that the *Attest(Me/Us)-I* variants require less training data, since they only rely on the profile of the ratified fab, as opposed to all fabs, yet are more difficult than their *Attest(Me/Us)-II* counterparts. Similarly, the *AttestMe-(I/II)* variants require less test data, since they make decisions for individual ICs, as opposed to batches of ICs, yet are more difficult than their *AttestUs-(I/II)* counterparts.

3. PROPOSED SOLUTIONS

In this section, we present the proposed solutions for the four variants of the fab-of-origin attestation problem, which we introduced in Section 2.

3.1 AttestMe-I

The *AttestMe-I* variant is, essentially, a one-class classification problem, for which numerous solutions exist in the literature [10]. Specifically, given a statistically significant set of parametric test data from the ratified fab, we need to learn a boundary that encloses this population in the multidimensional space of these measurements. The trained one-class classifier then compares the footprint of a new IC to this boundary, in order to decide whether it came from the ratified fabrication facility or not.

The key challenge in the context of *AttestMe-I*, however, is the high dimensionality of our data, which is typically in the few hundreds (i.e., number of probe-tests). Indeed, due to the curse of dimensionality, it is practically impossible to capture the underlying interaction between the design and the idiosyncrasies of a specific fab and to establish any meaningful boundary in the raw data space. Instead, the proposed method employs the following steps:

Dimensionality Reduction: In order to reduce the dimensionality of the test data, we employ the t-Distributed Stochastic Neighbor Embedding (t-SNE) [15] technique. t-SNE is a non-linear transformation of the parametric test data into a lower-dimensional feature space, wherein enough discriminative power exists for learning the boundary that encloses the population.

Clustering: Once the data is projected in the transformed space, we use the GAP statistic method [14] to estimate the number of clusters that the data consists of, fol-

[1] While in this work we only consider probe-test data, should on-die process control measurement (PCM) data be available, they can be seamlessly integrated into our solutions.

lowed by k-means clustering to separate the data into the corresponding number of clusters.

Boundary Identification: A simple one-class classifier (i.e., a convex hull) is then trained to enclose the data of each cluster. Collectively, the acceptance region of the trained one-class classifiers for all the clusters, define the space where ICs from the ratified fab are expected to reside.

Decision Making: Given the test data of a new IC, its footprint in the transformed space is computed and compared to the acceptance region. The IC is considered as originating from the ratified fab *if and only if* this footprint falls within any of the learned clusters.

We note that we considered the simpler and very popular Principal Component Analysis (PCA) [9] method for dimensionality reduction. However, as we will demonstrate in Section 4.1, the variance of our data appears to be highly non-linear; therefore, PCA, which linearly transforms the original data to a lower dimensional subspace, while retaining most of its variance, performs poorly. We also note that we considered training an advanced one-class classifier (i.e., SVM) to directly learn a single boundary in the reduced feature space. However, the data in this space is highly discontinuous, with the vast majority of the points congregating in small clusters. Therefore, as we will demonstrate in Section 4.1, learning a single boundary to successfully include all these discontinuous regions while excluding the rest of the space is of limited effectiveness.

3.2 AttestUs-I

Our solution to the *AttestUs-I* variant seeks to take advantage of the fact that process variations are expected to affect ICs produced within the same fab in a correlated way. Accordingly, this correlation can be leveraged to improve fab-of-origin attestation effectiveness for a batch of ICs, all of which originate from the same fab. To achieve this, we assess the underlying distribution of performance parameters for this batch against the profile of the ratified fab using a non-parametric statistical test. In particular, we employ the Anderson-Darling (AD) test [1], which is a well-known procedure for determining whether a sample of k observations comes from a given distribution or not. Its main advantages include its sensitivity to the distribution shape and its applicability to small sample sizes. In order to utilize the AD test in the fab-of-origin attestation context, we apply the following procedure:

Density Estimation: For every performance parameter t of the device under attestation, we use the parametric measurements in the statistically significant training set from the ratified fab, along with Kernel Density Estimation (KDE) [12] using Epanechnikov kernel, in order to compute its probability density function, PDF_t.

Membership Test: Consider m_t as the measurement vector of performance parameter t from all ICs in the batch under attestation. We apply the AD test to the estimated density and the measurements from the chips in the batch, i.e., $AD(PDF_t, m_t)$, where the null hypothesis is that the measurement vector, m_t, comes from the estimated density of the ratified fab. The output of the AD test is an asymptotic p-value in the range 0 to 1. For a p-value less than a chosen threshold (usually 0.05), the null hypothesis is rejected and we deduce that the distribution of the measured data, m_t, is dissimilar to the estimated density (i.e., this batch of chips does not originate from the ratified fab).

Decision Making: This procedure is repeated individually for each performance parameter. A majority vote is, then, employed to provide the final decision for the batch.

3.3 AttestMe-II

The *AttestMe-II* variant of attesting an individual chip, when parametric measurements from a statistically significant number of chips from both the ratified and all other (i.e., undesired) fabs are available, boils down to a two-class classification problem. Availability of populations from both classes simplifies the problem drastically and eliminates the need for clustering. Instead, our solution to this variant involves the following steps:

Classifier Training: We use the available training data to train Deep Neural Networks (DNNs), which have recently achieved state-of-the-art performance in a wide range of classification tasks of high dimensionality [6]. For our two-class classification problem, a five-layer DNN with two output neurons is trained using measurement data from both the ratified and the undesired fabs. To train the entire network, a generative pre-training step is applied to train one layer at a time. Then, the whole network is fine-tunned using the back-propagation learning algorithm. Note that dimensionality reduction is intrinsic to the DNN, and a separate step is not required in this approach.

Decision Making: Given a new IC whose source of origin needs to be attested, its performance parameters are measured and provided to the trained DNN, which determines which of the two classes the IC belongs to, i.e., whether it was produced in the ratified fab or in an undesired fab.

3.4 AttestUs-II

Our solution to the *AttestUs-II* variant follows the general principles of what we described in Section 3.2 and consists of the following steps:

Density Estimation: For every performance parameter of the IC batch under attestation, we compute its probability density function (PDF) in both the ratified fab and the undesired fab(s) by applying KDE on the corresponding training sets.

Membership Test: For every performance parameter, we apply the AD test using the measurement vector from all ICs in the batch under attestation and the PDFs of the ratified fab and the undesired fab(s). The combination of the two p-values determines whether, with respect to this performance parameter, the ICs in the batch were manufactured in the ratified fab or in an undesired fab.

Decision Making: This procedure is repeated individually for each performance parameter. A majority vote is, then, employed to provide the final decision for the batch.

4. EXPERIMENTAL RESULTS

In this section, we experimentally evaluate the effectiveness of the proposed solutions using actual production test data from a 65nm RF transceiver currently in high volume manufacturing (HVM) by Texas Instruments.

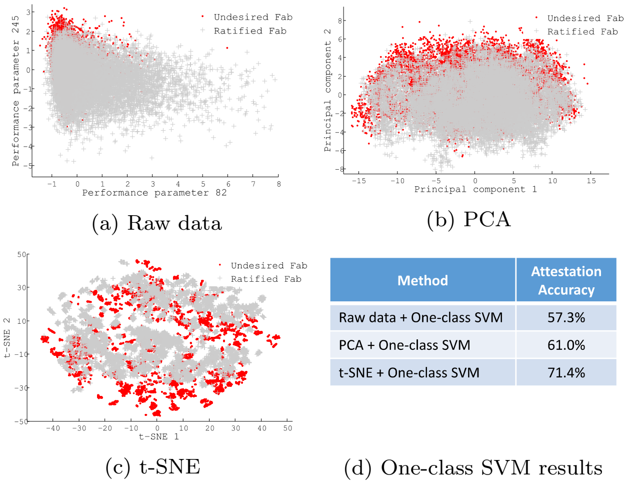
Our dataset comprises devices from two geographically dispersed fabs wherein this RF transceiver is fabricated. For the purpose of this study, we consider one of these facilities as the ratified fab and the other one as the unknown or undesired fab. The dataset for the ratified fab includes 600 wafers from 20 lots, with approximately 1500 die per wafer. For each die, 276 probe-test measurements are provided. These tests are the typical measurements performed at wafer probe to ensure compliance of the performances of an RF transceiver design to its specifications (i.e., production tests). They include both structural tests (open/short circuit, power consumption, I_{DDQ}, input voltage threshold, output voltage level, etc.) and functional tests (BER, EVM, CMMR, receiver sensitivity, output power, phase noise, etc.) and indirectly cover a broad range of process parameters. We note that we chose not to use in-line tests (e-tests) which directly reflect process parameters, since they are on the wafer scribe-lines rather than the die, hence they are not available at the final chip level. The dataset of the undesired fab includes the same 276 probe-test measurements from 500 wafers in 20 lots. These two datasets were obtained from the two fabs at approximately the same period. Using this dataset, we seek to:

- Visualize the overlap of the two populations in the raw data space and in the linearly transformed PCA space, as well as the effectiveness of the non-linear t-SNE transformation in increasing discrimination, and demonstrate the limited effectiveness of training a one-class classifier (i.e., SVM) to separate the populations through a single boundary, due to data discontinuity.
- Quantify the effectiveness of *AttestMe-I* and *AttestUs-I*, which use data solely from the ratified fab for learning the underlying model, in distinguishing between ICs produced in the ratified and in an unknown fab.
- Assess the attestation accuracy improvement achieved by *AttestMe-II* and *AttestUs-II*, which are trained with datasets from both the ratified and the undesired fabs.
- Demonstrate the effectiveness of our solutions in handling process variations by assessing attestation accuracy on ICs from future production.

4.1 Population overlap

To demonstrate population overlap, we randomly select 5 wafers from each of the 20 lots in the ratified fab and we use all probe-test data of all die on these 100 wafers as our training set. We, then, train a one-class SVM to learn the boundary that encloses the population originating from the ratified fab in three different spaces: (i) in the raw data space which includes all 276 dimensions, (ii) in a PCA transformed space where the data is linearly projected on the first 30 principal components, and (iii) in the t-SNE transformed space where the retained data is non-linearly projected on 3 dimensions. As our validation set, we use all die from a randomly selected wafer from each of the 20 lots of the ratified fab (excluding the wafers used for training) and from each of the 20 lots of the undesired fab. The trained SVMs are, then, used to individually decide whether each die in the validation set originated from the ratified fab or not.

Figures 1a-c visualize the training and validation data on the space of the two most discriminative raw measurements, on the two main components of the linearly transformed PCA space, and on the two components of the non-linearly transformed t-SNE space, respectively. As may be observed, there is an almost complete population overlap in the first case, which is only slightly reduced after linear transformation in the second case, because the variability of the data is non-linear. The non-linear transformation of the third case, however, performs significantly better in separating the two populations. While this is visualized only in a two-dimensional space, our extensive experimentation with multiple dimensions has confirmed this observation, justifying the use of t-SNE as the method of choice for enhancing discrimination via dimensionality reduction in this context.



(a) Raw data

(b) PCA

(c) t-SNE

Method	Attestation Accuracy
Raw data + One-class SVM	57.3%
PCA + One-class SVM	61.0%
t-SNE + One-class SVM	71.4%

(d) One-class SVM results

Figure 1: Population overlap and single boundary classification accuracy in raw and transformed measurement spaces

The results reported in the table of Figure 1d, which quantify the effectiveness of a single boundary established by training a one-class SVM in each of the three spaces mentioned earlier, are also consistent with this observation. Indeed, attestation accuracy of a single IC in the raw data space is only 57.3%, barely higher than a coin-toss. Learning the boundary in the 30-dimensional PCA space only slightly improves accuracy to 61%, while doing so in the 3-dimensional t-SNE space boosts accuracy to 71.4%. This rather low accuracy is attributed to the highly discontinuous nature of the data in the projected space, which calls for a clustering-based classification approach, as we show next.

4.2 Learning only from ratified fab

In order to assess the effectiveness of *AttestMe-I*, we perform clustering and boundary identification on the t-SNE transformed space of the training data, as detailed in Section 3.1. Then, for each IC in the validation set, we apply the decision making step which examines whether its footprint in this space lies within the boundary of any of the clusters assigned to the ratified fab. Table 1a reports the attestation accuracy for *AttestMe-I*, noting that we consider as *positive*, (P), a chip originating from the ratified fab and as *negative*, (N), a chip originating from an undesired source. In this confusion matrix, True Positive Rate (TPR) denotes the percentage of ICs that are correctly identified as originating from the ratified fab, while True Negative Rate (TNR) refers to the percentage of ICs that are correctly labeled

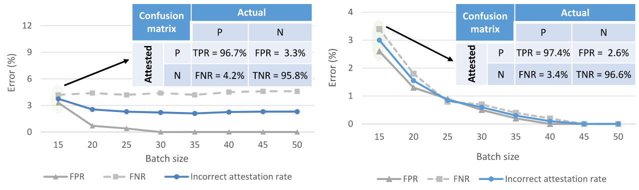
Table 1: Single IC attestation results

(a) *AttestMe-I*

Confusion matrix		Actual	
		P	N
Attested	P	TPR = 85.5%	FPR = 14.5%
	N	FNR = 15.5%	TNR = 84.5%

(b) *AttestMe-II*

Confusion matrix		Actual	
		P	N
Attested	P	TPR = 97%	FPR = 3%
	N	FNR = 4%	TNR = 96%



(a) *AttestUs-I* (b) *AttestUs-II*

Figure 2: Attestation results for various batch sizes

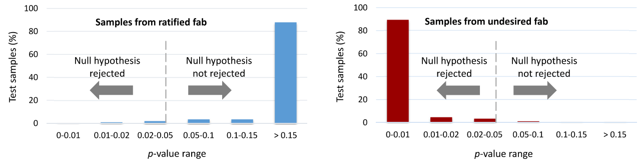


Figure 3: Histogram of p-values for AD test against the ratified fab distribution for batches of 15 chips (*AttestUs-I*)

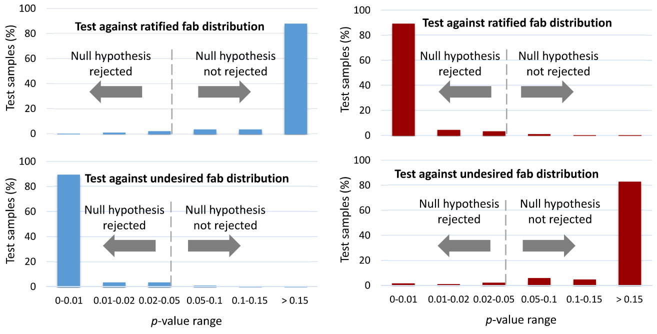
as originating from an undesired fab. False Positive Rate (FPR) and False Negative Rate (FNR) are defined similarly. As may be observed, the overall attestation accuracy is 85%, clearly outperforming the one-class SVM reported in Figure 1d. This is expected due to the manifold nature of the t-SNE transformed data, which makes it difficult to separate via a single boundary, as the SVM tries to do.

Effectiveness of *AttestUs-I* is assessed by first estimating the performance parameter densities of the ratified fab through the training set. Then, for a batch of ICs originating from the same fab, we measure the performance parameters from all ICs in the batch and we perform the AD membership test for each of the parameters, as detailed in Section 3.2. In our experiment, we randomly draw batches of sizes in the range [15, 50] from the validation sets of the ratified and the undesired fab; this procedure is repeated 1000 times for each batch size. Figure 2a reports the *AttestUs-I* results. The horizontal axis denotes the batch size, while the vertical axis is the attestation error rate. As may be observed, this method is very successful in attesting the fab-of-origin of a batch, with accuracy exceeding 96% for batch sizes of as small as 15 ICs. The confusion matrix for this batch size is also provided in the figure.

To gain further insight, in Figure 3 we show the distribution of p-values for a batch size of 15 ICs, where the x-axis represents the range of p-value and the y-axis shows the percentage of 500 randomly selected batches which have a p-value within a given range. In the Anderson-Darling distribution test, the null hypothesis is that the 15-dimensional measurement vector of the 15 ICs in the batch comes from a specific population, which in our case is the distribution of the ratified fab. As shown in the left histogram, for the vast majority of the 500 samples from the ratified fab, the p-value is larger than 0.05, hence the null hypothesis is not rejected, i.e., these batches are correctly assumed to have originated from the ratified fab. Conversely, as shown in the right histogram, for the vast majority of the 500 samples from the undesired fab, the p-value is smaller than 0.05 and the null hypothesis is rejected, i.e., these batches are correctly assumed to have originated from the undesired fab.

4.3 Learning from all fabs

In order to quantify the accuracy of the proposed fab-of-origin attestation solutions when test data from both the ratified and the undesired fab is available, we enhance the training set so that it contains data from both fabs. Specifically, in addition to all die from 5 randomly selected wafers in each of the 20 lots from the ratified fab, the new training set also includes all die from 5 randomly selected wafers in each of the 20 lots from the undesired fab. The validation set remains unchanged, i.e., it contains all die from a randomly selected wafer from each of the 20 lots of the ratified fab and from each of the 20 lots of the undesired fab (excluding the wafers used for training).



(a) Samples from ratified fab (b) Samples from undesired fab

Figure 4: Histogram of p-values for AD test against the ratified and the undesired fab distributions for batches of 15 chips (*AttestUs-II*)

Evaluation of the *AttestMe-II* solution starts with training a two-class classifier, in this case a five-layer DNN with two output neurons, using the training set, as explained in Section 3.3. The trained DNN is then applied to individually classify each IC in the validation set as originating from the ratified or the undesired fab. *AttestMe-II* results are summarized in Table 1b. As may be observed, the trained DNN is able to accurately attest 96.5% of all chips and is significantly better than the *AttestMe-I* approach. This is expected, as we now have access to data from both fabs, which simplifies the process of learning the boundary that separates them, as compared to the case where training data is available only from the ratified fab.

Effectiveness of *AttestUs-II* requires estimation of the performance parameter densities for both the ratified and the undesired fab using the enhanced training set. Then, for a batch of devices from the same fab, we measure the performance parameters from all ICs in the batch. For each performance parameter, we perform the AD membership test against the densities of both fabs to compute the corresponding p-values, and we decide which fab the batch originated from, as explained in Section 3.4. Once again, in our experiment we randomly draw batches of sizes in the range [15, 50] from the validation sets of the ratified and the undesired fab, and repeat this procedure 1000 times for every batch size. Figure 2b reports the *AttestUs-II* results, with the x-axis denoting the batch size and the y-axis showing the attestation error. As may be observed, for a batch size of as few as 25 ICs, the accuracy of this solution exceeds 99%, while for a batch size of 40 ICs, it achieves error-free attestation. A comparison to the curves in Figure 2a reveals that availability of the additional training information from the undesired fab enhances the accuracy of the membership test and reduces the error. As a point of reference, the confusion matrix for the batch of size 15 is also provided.

Lastly, Figure 4 presents the histogram of p-values for batches of 15 ICs. Figure 4a shows the p-values for 500 batches originating from the ratified fab, wherein the top and bottom graphs compare these samples against the ratified and the undesired fab distributions, respectively. Evidently, for the vast majority of samples the null hypothesis is not rejected for the ratified fab but is rejected for the undesired fab, hence these batches are correctly attested as originating from the ratified fab. Conversely, Figure 4b demonstrates the same results for 500 batches originating from the undesired fab, in which case the results are reversed.

4.4 Future production attestation accuracy

As a final experiment, we evaluate the robustness of the proposed solutions against fabrication process shifts. To do so, we use probe-test data from a new set of wafers from 10 lots, which were fabricated in each of the two fabs a few months after the wafers of our original dataset. We refer to these new wafers as "future wafers". Our training set remains the same, but our new validation set now comprises all die from 20 randomly selected future wafers, equally distributed across the 10 new lots from each of the two fabs. Table 2 compares the effectiveness of the *AttestMe-II* solution on the original validation set, which comprises current wafers (i.e., contemporary to those used for training) to the effectiveness on the new validation set, which comprises future wafers. As may be observed, the trained DNN network continues to attest ICs from future production with only a very slight reduction in accuracy. Similarly, Figure 5 compares the effectiveness of the *AttestUs-II* solution for ICs on current and future wafers, where the x-axis is the batch size and the y-axis denotes the attestation error. Once again, the difference in the two scenarios is small and reduces as the batch size increases. As an ancillary measure for maintaining robustness, the underlying trained models can be periodically updated.

5. CONCLUSION

Parametric measurements, such as the ones taken during manufacturing testing, comprise valuable information which reflects the interaction between the design of an IC and the fabrication process through which it was produced. In conjunction with machine learning methods, this information may be harnessed to provide effective solutions to numerous variants of the fab-of-origin attestation problem, without requiring design modifications, custom processing steps, or specialized characterization equipment. Four such solutions were developed and evaluated using actual test data from a large number of ICs implementing an RF transceiver design, which were fabricated in two geographically dispersed foundries. Results indicate that the accuracy of these fab-of-origin attestation solutions reaches 96.5% when deciding whether a single IC originated from a ratified fab or an unknown/undesired facility and 100% when collectively making the same decision for a batch of as few as 40 ICs.

It is worth noting that while precise cloning of an IC could evade our methods, our study was performed on two fabs of the same manufacturer so it resembles the best cloned devices one can build. Thus, we expect even higher attestation accuracy when the fabrication facilities are independent. Also, it is possible that changes in fabrication process, such as machine part replacements, software updates or new material suppliers, may shift the process parameters and affect the accuracy of our models over time. Nevertheless, our methods were able to attest future productions with only minor accuracy reduction, demonstrating robustness of the models to such changes. To reinforce robustness, our future research focuses on attesting not only a specific fab but also a specific machine or material used in fabricating an IC.

Table 2: Comparison of *AttestMe-II* results for chips from current and future production

	AttestMe-II				
Validation set data	Accuracy	TPR	FPR	TNR	FNR
From current production	**96.5%**	97%	3%	96%	4%
From future production	**94.0%**	96%	4%	92%	8%

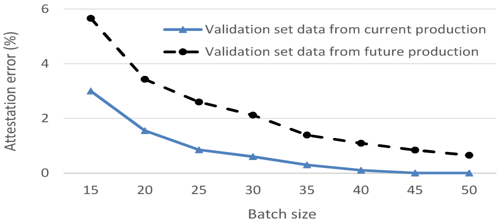


Figure 5: Comparison of *AttestUs-II* results for chips from current and future production

6. REFERENCES

[1] T. W. Anderson *et al.*, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *The annals of mathematical statistics*, pp. 193–212, 1952.

[2] S. Bhunia *et al.*, "Hardware trojan attacks: Threat analysis and countermeasures," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1229–1247, 2014.

[3] G. Bloom *et al.*, "Fab forensics: Increasing trust in IC fabrication," in *Proc. IEEE HST*, 2010, pp. 99–105.

[4] DARPA. SB133-003: Electronic component fingerprinting to determine manufacturing origin. [Online]. Available: http://www.acq.osd.mil/osbp/sbir/solicitations/sbir20133/darpa133.htm

[5] U. Guin *et al.*, "Counterfeit integrated circuits: A rising threat in the global semiconductor supply chain," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1207–1228, Aug 2014.

[6] G. E. Hinton *et al.*, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[7] D. E. Holcomb *et al.*, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *Trans. Computers*, vol. 58, no. 9, pp. 1198–1210, 2009.

[8] K. Huang *et al.*, "Recycled IC detection based on statistical methods," *Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 6, pp. 947–960, June 2015.

[9] I. Jolliffe, *Principal component analysis.* Wiley Online Library, 2002.

[10] L. M. Manevitz *et al.*, "One-class SVMs for document classification," *the Journal of machine Learning research*, vol. 2, pp. 139–154, 2002.

[11] M. Potkonjak *et al.*, "Differential public physically unclonable functions: architecture and applications," in *Proc. ACM DAC*, 2011, pp. 242–247.

[12] B. W. Silverman, *Density estimation for statistics and data analysis.* CRC press, 1986, vol. 26.

[13] M. Tehranipoor *et al.*, "A survey of hardware trojan taxonomy and detection," *IEEE Design & Test of Computers*, vol. 27, no. 1, pp. 10–25, Jan 2010.

[14] R. Tibshirani *et al.*, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[15] L. Van der Maaten *et al.*, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 2579-2605, p. 85, 2008.

[16] J. B. Wendt *et al.*, "Techniques for foundry identification," in *Proc. ACM DAC*, 2014, pp. 1–6.