Methods to Assess Sensitivity Limits of Second Order Effects Measurements

Carl T. Boone, Sean C. Stuart, Dmitry Veksler,
& Brendan Foran
Electronics & Photonics Laboratory
The Aerospace Corporation
El Segundo, CA 90245
Email: carl.t.boone@aero.org

Garret Chan, Salam Zantout and Vikram Rao
Hardeware Security & Specialty
Engineering
The Aerospace Corporation
El Segundo, CA 90245
Email: vikram.rao@aero.org

***Abstract*— Sensitivity limits of a specific second order effects (2OE) testing method were studied with respect to the detection of a specific set of circuit modifications using flash-enabled FPGAs as a testbed. The 2OE method tested was an expanded variant of Sandia's Power Spectrum Analysis (PSA) and the circuit modifications to be studied were hardware Trojans of different size and with controlled changes to physical layout on the test FPGA. The focus of this work was data analysis comparing the utility of traditional Hotelling t^2 and corresponding p-values as separability metrics to confusion matrix analysis derived from machine learning classifiers based on logistic regression. The goal of the analyses was to optimize data collection for 2OE features that improve separability while clarifying the limitations of the method with respect to identifying the specific risk defined by the part/problem pair. The optimized 2OE methods easily identified physical differences in FPGAs (associated with manufacture changes related to lot dates) and certain hardware Trojan varieties were clearly identified as separable while others proved more difficult for the 2OE features analyzed. This work provides a path for 2OE methods optimization and quantification of utility to mitigate specific risks.**

Keywords—second order effects; power spectra analysis; Trojan detection; quantitative assurance

I. INTRODUCTION

Second order effects (2OE) testing methods have been proposed to provide cost-efficient screening for counterfeit microelectronic parts and could also be used to identify reliability escapes as a means to quantitative assurance. 2OE are measurable characteristics, separate from primary functionality, that are related to the physical implementation of devices. These characteristics include but are not limited to electromagnetic emission, power absorption, and other side channel signals. 2OE characteristics have been used to develop signature "fingerprints" allowing detection of differences between several types of microelectronic parts ranging from diodes and capacitors to integrated circuits such as field programmable gate arrays (FPGAs) [1,2,3]. Further, 2OE has been used to identify parts degradation by thermal cycling or radiation exposure. However, quantitative assessment of the goodness of a 2OE method to detect specific anomalies is lacking. The effectiveness of a specific 2OE method to identify differences in parts needs to be quantified by a separability metric [4]. The work we present here uses the Hotelling t^2 distance to quantify separability and corresponding p-value that measures confidence of data separation [5]. Further, we test the application of machine learning classifiers for determining separability.

The work presented here relied on 2OE data collected using a field-programmable gate array (FPGA) based "2OE-testbed" as described previously [6]. This platform allows for fast and easy collection of 2OE data for a significant number of circuits with controlled modifications to allow the quantification of 2OE method effectiveness to detect such differences. This experiment also allows us to study how modification of data collection and data analysis/reduction techniques relate to quantifiable separability.

While the work described here focused on a specific 2OE method (described in the following section), the purpose of this paper is rather to address the assessment of any particular 2OE method for the identification of specific part variability. The method is composed of both the data collection and analysis parts; choices of test equipment, cabling, sampling parameters all determine the relative value of data collected which may be simplified in terms of signal to noise, where signal is defined as elements in the data that supports the identification while noise is defined as elements in the data that do not support the identification. After data collection, data analysis methods are similarly critical in denoising and defining the effective signal and noise levels that may be optimized if detectability limits are to be understood. We believe that the assessment of a method requires optimized separation of signal from noise, which can be done during either or both the data collection and analysis steps.

The *Methods* section is separated into sections describing what was tested (*FPGA Test Bed and Circuit Configurations Tested*), how data was collected (*Measurements*) and how it was analyzed (*Data Reduction and Analysis*). The focus of this work is the assessment of data reduction and analysis. The purpose of data reduction is to identify which combinations of

measured features contribute value to identifiable separability defined with an appropriate figure of merit. The purpose of our analysis is to quantify separability and identify means to optimize a method and understand its limitations with respect to specific risks.

To clarify, a measured "feature" refers to an independent variable (e.g., a frequency in a power spectrum). A spectrum consisting of 1000 frequency points would have 1000 features, and a spectrum consisting of 1000 frequency points measured over two input voltages would have 2000 features.) However, because differences between parts may be localized in feature space, not all measured features may contribute equally to separability for a specific part/problem combination. Methods that we, and others, have explored in the past perform a dimensionality reduction on the data, for example using principal component analysis (PCA), to remove noise-dominated features, to improve overall signal to noise, and to potentially improve detectability. We consider noise to be any aspect of the data that does not contribute positively to separability. However, the final dimensionality to be retained is subject to user preference – well-defined cutoffs for the number of principal components retained may not exist, and subjectivity is used to determine the number. This adds an unknown to principal component analysis that, ideally, would be avoided. We address this by comparing separability metrics obtained with different numbers of principal components to assess if rigorous cut-offs can be established.

II. Methods

A. FPGA Test Bed and Circuit Configurations Tested

Our 2OE test-bed employs a ProASIC3 (A3P125) FPGA on a ProASIC3 development board that was modified by removal of filter capacitors. ProASIC3 FPGAs use non-volatile flash-based gates as configuration bits that are maintained when the FPGA is unpowered (unlike SRAM-configured FPGAs) and allows the same FPGA to be reconfigured many times. Circuit designs with small modifications can be sequentially programmed into an FPGA, enabling correlation of configurations with 2OE characteristics to optimize data collection and analyses and to identify detectability limitations of particular methods.

FPGA Serial No.	Lot Date Code ID
05	1427
07	1427
12	1731
15	1731
22	1731
28	1731

Table 1. Serial number and lot date code identifiers for separate FPGAs used in this manuscript.

Six individual FPGAs across two lot date codes are included in this work, as shown in Table 1. The two lot date codes represent

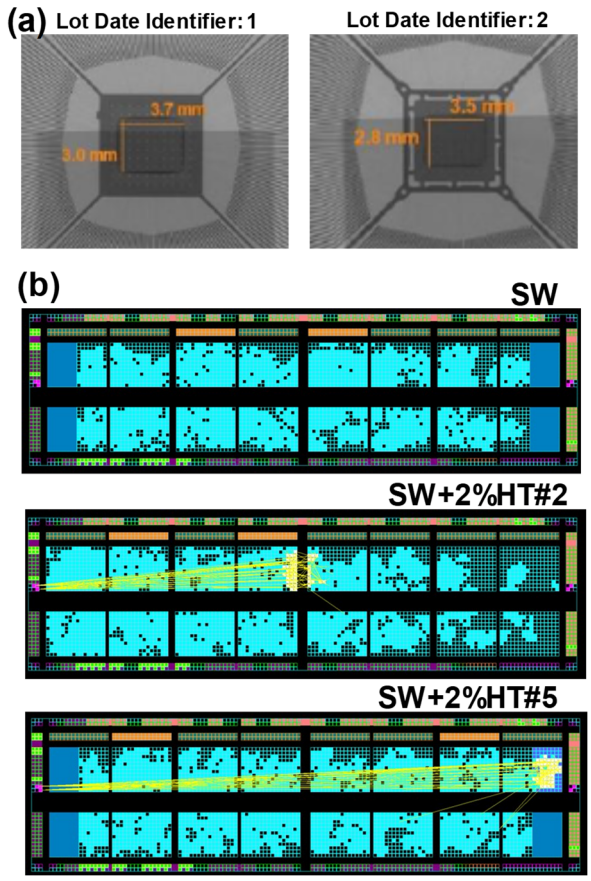

Figure 1. (a) X-ray images of FPGAs from each lot date code identifier, showing packaging and die size differences. (b) Gate usage layout for three example configurations generated from MicroSemi's Libero tool.

Configuration Identifier	Description
SW	Unmodified SpaceWire IP, no HT present
SW+2%HT#1	SW+2%HT, HT placed using default routing
SW+2%HT#2	SW+2%HT, HT placed in upper middle
SW+2%HT#5	SW+2%HT, HT located in upper right corner
SW+30%HT	SW IP plus 30%HT, HT placed by default routing

Table 2. Descriptions of FPGA configurations measured for this manuscript.

significantly different electrical lead design and die size, determined by x-ray imaging the parts. Fig. 1a shows x-ray images of two representative FPGAs, one from each lot code identifier.

A SpaceWire IP (SW) variant that employed ~67% of the A3P125 FPGA's configuration resources was employed for this study. This circuit was modified by insertion of hardware Trojan functionality that required the use of an additional 30% of the FPGA's resources (SW+30%HT) or only an additional 2% of these resources (SW+2%HT). Further, for the SW+2%HT, several circuit variants were developed by modifying the physical position of FPGA resources [7] employed for the additional HT circuitry. Fig. 1b shows three examples identified as #1, #2, and #5. The five different configurations are included in the dataset analyzed for this manuscript, listed in Table 2.

B. Measurements

The 2OE measurements used in this work are adapted from Sandia's Power spectrum analysis (PSA) method [8 , 9] wherein a square wave of user-defined amplitude (voltage) and 10 kHz fundamental frequency was applied across the power pin (Vcc) and ground of the programmed A3P125 FPGA. The reflected square wave signal was measured with an oscilloscope, sampled every 5 ns. From earlier studies to optimize data collection, 16 oscilloscope measurements, with 50 square wave cycles each, were averaged (in software); this enabled 32-bit deep data in spite of our oscilloscope only delivering 8-bit data.

This PSA methodology has many positive features that make it valuable for this work. The square wave is inherently broad-band with a 1/f (slowly-decaying, where f is the frequency) profile, meaning that even a low-frequency square wave probes frequencies up to the sample rate, edge sharpness, and jitter limitations. Additionally, the pure square wave consists only of odd harmonics of the fundamental frequency [10]. If all components in the interrogated device had linear frequency responses, then only odd harmonics would appear in the measured spectrum. However, because components of interest (transistors, p-n junctions, etc) are nonlinear, even harmonics of the square wave fundamental frequency will appear due to modal coupling. Our prior work showed that the even harmonics did a better job of discriminating small differences between parts. Other spectroscopy methods for looking at nonlinear responses exist – such as side band analysis of a sine-wave input – but none are broadband over a single measurement.

The square wave amplitude is always chosen below the turn-on voltage for the transistors used in the FPGA, which is 1.5 V for these FPGAs. The probe signal avoids turning on and running the FPGA while interrogating the device. The device not being turned on means no interference from data streams, calculations, or hardware Trojan triggering. It also makes this technique part-agnostic: it can be adapted and optimized for any IC or component. A set of different voltage amplitudes from 0.4 V to 1.0 V in steps of 0.1 V are applied to the device during each measurement cycle.

The FPGA was then fully erased and reconfigured to a different circuit and the measurements repeated until characteristics of all the circuits of interest were obtained. This cycle was repeated for the same configuration set multiple times (referred in this manuscript as "samples") to determine systematic measurement noise and drift. An automated script to power the device, set the configuration, remove power from the device, and perform a PSA measurement was created in LabView.

To emphasize the scale of data collected, our overall method may be summarized as follows. For each of the six FPGAs (SNs shown in Table 1), we performed 2OE measurements for each of the five circuit configurations (shown in Table 2), with ten separate programming cycles made per configuration, for seven different square wave amplitudes (all without removing the FPGA).

This measurement scheme allows for direct comparison of the power spectra of different circuit configurations and individual FPGAs. The number of erase and reconfigure cycles was kept well below the 10000 cycles for which the Flash cells in the ProASIC3 are rated, and no systematic variation in the data related to the number of program cycles has been observed in this or our prior work.

C. Data Reduction and Analysis

Each measured waveform was normalized such that the measured square wave has an average of 0 during low amplitude segments and 1 during high amplitude segments. This allows measurements taken at different square wave amplitude voltages to be compared at the same level. Then each waveform, independently, was fast Fourier transformed to obtain the power spectrum. Given the 5 ns time step in the measurement over 50 square wave cycles at 10 kHz fundamental frequency, the frequency range for the power spectrum runs from 0 Hz to 100 MHz in steps of 200 Hz. The power spectrum was segregated into odd, even, and off harmonics of the square wave fundamental frequency. The off harmonics have been observed to contain only noise [11] and are discarded, while the odd and even harmonics are retained for further analysis. All of the harmonics of the power spectra are compiled into a dataset incorporating the measurement date and time, FPGA serial number identifier, configuration, and measurement voltage. In an extension of our prior work, we combine all voltage values from a single measurement cycle into one feature set. Different measurement cycles of the same FPGA and configuration are considered identical samples for training purposes – in this way measurement noise and drift are included in determining the limitations of the 2OE analysis.

Similar to our prior work, we performed PCA on the data and calculated Hotelling t^2 values between each sample cluster. From the t^2 values and the number of samples, the p-value, corresponding to the confidence that samples represent

statistically separate clusters, were calculated. The goal of clustering is to identify individuals that are separate from others. We note that, even if two clusters are statistically separate according to t^2 and p-values significant overlap may still occur in the clusters. This limit on the confidence that any single measurement will fall into a given cluster necessitates a way to classify individual samples within a set of known variants. To do this, we explored machine learning methods combined with different up-front data reduction methods, by training classifier functions [12] for lot date code, and circuit configuration.

A classifier function takes a data sample, consisting of values for a set of measured features as input, then outputs probabilities for membership in each of the classes trained and therefore assigns the likeliest class to that data sample. Creating the classifier function requires fitting ("training") to a set of data consisting of measured features and class labels for known samples. Classifier parameters and fitting routines are determined by the selected algorithm. Three common algorithms - logistic regression, support vector machines, and neural networks – were tested in the work leading to this paper using Mathematica's built-in "Classify" function [13].

Because the classifier is trained on data containing many features, some of which may be noise-dominated, different subsets of features may produce different classifier function outcomes. Part of this work is to assess the value of noise removal by dimensionality reduction for improving classifier accuracy. An additional benefit of data dimensionality reduction is the significantly reduced classifier training time and the resulting complexity of classifier functions; because dimensionality reduction through PCA is computationally efficient, the savings could be significant. To assess this aspect, we trained classifiers on different subsets of the data obtained in a variety of ways. In one set of methods, certain harmonics or regions of frequency were simply discarded. This was motivated by the fact that higher frequencies contained no discernable information leading to separability, and that certain harmonics may not be excited significantly within a given device under test.

We compare a variety of data subsets – both harmonics covering the complete frequency range, even or odd harmonics only covering the complete frequency range, and PCA-reduced subsets of both, odd, and even harmonics, retaining two to nine principal components.

III. Results

Figure 2 shows the PCA results in two plots for the first three principal components calculated using even harmonics. In these plots, the hierarchy of sample, serial number, and lot date is observed. The tightest clusters represent one circuit configuration within a single FPGA. The next larger groupings represent SW and SW+2%HT configurations within a single FPGA. The largest groupings account for lot date code, with separation to left and right sides of the black dashed line. Circuit configurations tend to separate along PC2 and higher PC axes. The 30%HT is well separated

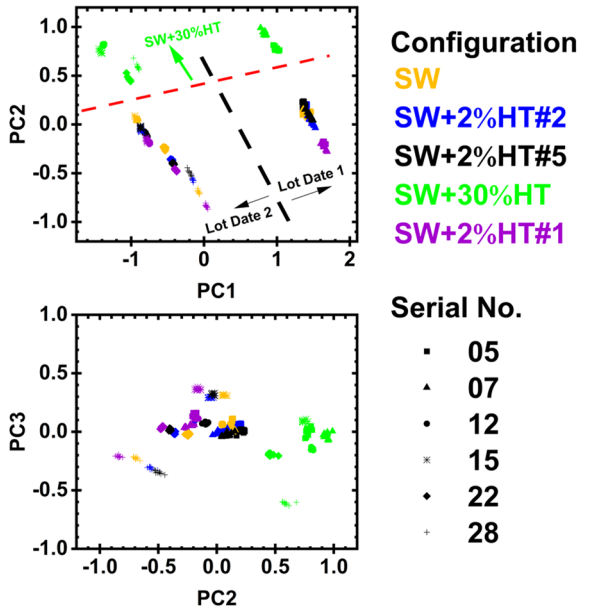

Figure 2. PCA results for even harmonics using covariance PCA, showing PC1 vs PC2 and PC2 vs PC3 for all configurations (denoted by color) and FPGA serial number identifiers (denoted by marker shape). Lot date separates along PC1 (black dashed line), while the SW+30%HT configuration separates along PC2 (red dashed line).

qualitatively from other circuit configurations, as denoted by the red dashed line in Fig. 2. Separation of other 2%HT configurations is not well-resolved in the plots of Fig. 2 but is resolved by the classifier analysis discussed later.

Figure 3 shows the separability metric t^2 calculated for nine principal components between all samples. In this plot the t^2 value between each sample set is represented by the color intensity of the square intersecting the respective row and column labeled for each sample. The diagonal is always calculated to be 0, as each sample set has zero separation from itself. Larger separation by lot date code is observed, and lines indicating the greater separability for the SW+30%HT circuit configuration can be seen.

Figure 4 shows that the value of t^2 tends to increase rapidly with the first few principal components, then slowly flattens. This is because the first few principal components capture almost all the sample to sample variation, as described previously. Small sample to sample variation in higher principal components contributes less to the t^2 value. However, even if t^2 continually increases, the p-value will have a minimum as more features tends to reduce confidence and increase the p-value [14]. This is illustrated in Fig. 4b-c, which shows t^2 and p-values relative to SW configuration for serial number 05 (showing only similar configurations within a single FPGA), versus number of principal components retained in the calculation for close clusters. We note that very well-separated clusters produce p-values below machine

precision for all principal components and, for that reason, are not be included in Fig 4b-c.

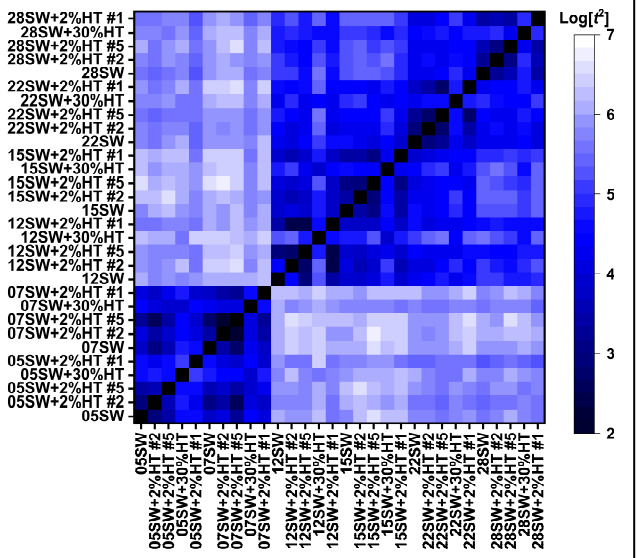

Figure 3: Hotelling t^2 distance between sets for all configurations and serial number identifiers, considering

A good cut-off for the number of principal components can be determined as the number that minimizes the p-value (thereby maximizing confidence in separation). This number of principal components will depend on the number of configurations included in the dataset as well as the measurement features, and will not, in general, be obtained in a nonempirical manner. For example, in this dataset based on Figure 4c, six principal components is a natural cut-off.

As mentioned in the "Methods" sections, classifier functions were trained on the data, with training data consisting of a random subset of seven measurements of each sample. Two classifiers were trained: one whose class set was the two lot date codes, and one whose class set was the configuration. Lot date codes easily classify accurately no matter the data reduction method, and the results are not discussed further here. To compare the value of different data reduction methods, we trained classifiers on different data subsets and reduction methods: even and odd harmonics; PCA reduction of the data retaining different numbers of principal components; the entire frequency range using complex FFT (that is, keeping phase information) and just the FFT amplitude (removing phase information). Figure 5 shows a subset of the confusion matrices generated during analysis for a logistic regression classifier algorithm [12], comparing different levels of data inclusion and dimensionality reduction. We tested other classifier algorithms including support vector machines and a neural network; however, logistic regression tended to perform well and require the shortest training time. The effectiveness of the classifier algorithm for determining separability from the available 2OE data is qualitatively seen in the extent to which the confusion matrices maximize the diagonalized values with minimized

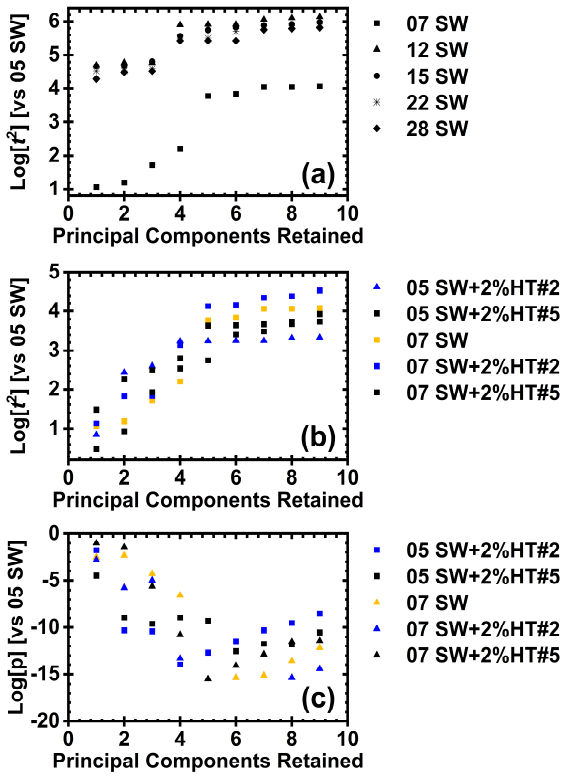

Figure 4. a) t^2 vs number of principal components retained for one circuit configuration (SW) measured on five different physical FPGAs. b) t^2 vs number of principal components for different, low- t^2 circuit configurations (recorded on two different FPGAs of the same lot date code). c) p values for the t^2values from (b).

values in the off-diagonal elements; for example, in all of the matrices presented in Figure 5, the SW+30%HT was classified correctly most often.

Retaining six principal components, and considering only even harmonics, resulted in the best sensitivity/specificity values across all the circuit configurations tested. Including all measured 2OE features decreased classifier accuracy for this data indicating that regions in the data consisting mainly of noise "confuse" the classifier and/or prevent the classifier from efficiently finding a solution. While the SW+30%HT was found most clearly separable, the smaller circuit modifications represented by the set of 2%HT circuits are much less separable across FPGAs. With six to nine principal components, and only even harmonics considered, one of the SW+2%HT varieties was fully distinguishable while others were not well separable from each other; correlation of the physical differences between the different SW+2%HT variants with separability (based on our optimized PSA methods) is the focus of ongoing work.

IV. Conclusions

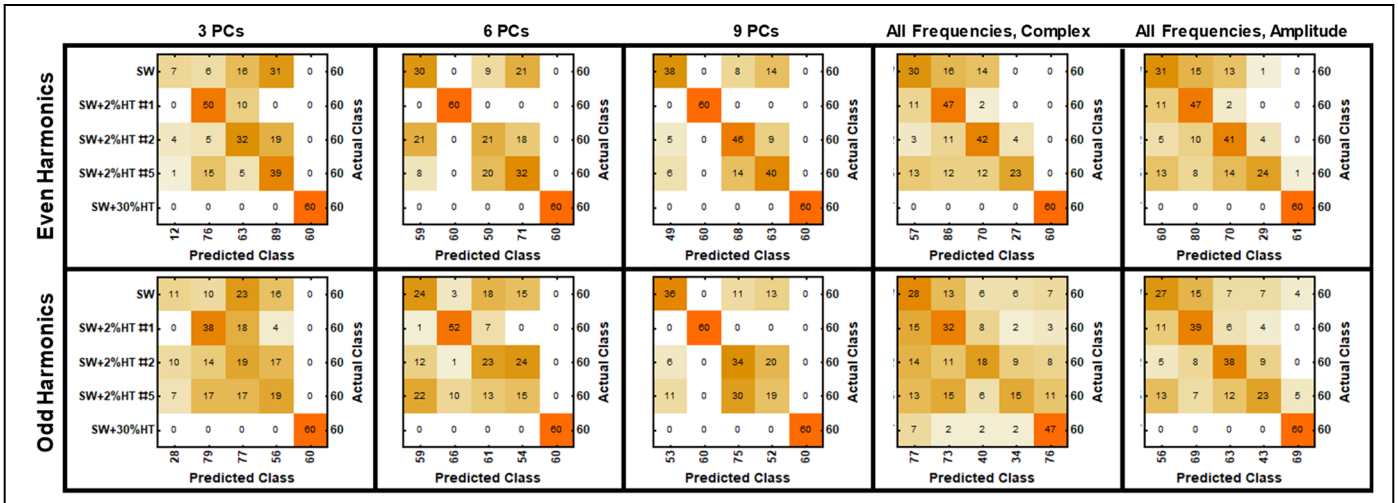

Figure 5. Example confusion matrices for classifiers for different subsets of the data: even and odd harmonics, retaining 3, 6, and 9 principal components, retaining the entire frequency range and complex FFT, and retaining the entire frequency range with just the amplitude. These classifiers use a logistic regression algorithm.

Using reprogrammable FPGAs as a testbed with a modified PSA-type 2OE method, variations in FPGA structure related to lot date code, and circuitry modifications represented by HT insertion, were readily quantified for separability across a variety of data analysis techniques including t^2 tests and machine learning classifiers. Dimensionality reduction by PCA appears to be a valuable pre-classification method for removing noisy features, focusing on important regions, and speeding up training processes. However, small circuit modifications showed limited separability despite improvements made by dimensionality reduction and employing machine learning classifier algorithms. The technique may need to be further optimized or combined with other methods to separate very similar parts and very small modifications. Identification of the limitations of specific 2OE methods, such as presented in this work, is critical for evaluating the value of potential risk mitigation for specific parts/problem pairings.

While clearer experiments may focus on optimizing the separability of specific circuit modifications (for example, we have also developed and analyzed data sets focused on identifying separability of small 2%HT modified circuits using a single FPGA), the complicated data set presented in this work - covering by multiple individual FPGAs, lot date related changes in ICs and circuit configuration changes - provides a broader basis for testing the value derived from application of data dimensionality reduction and machine learning methods and understanding the breadth of the problem for identifying optimization and understanding limitations of a specific 2OE method for identifying a specific problem.

ACKNOWLEDGMENT

This work was funded in part by Aerospace Corporation's Independent Research and Development (IRAD) Program.

REFERENCES

[1] Loubriel, Guillermo M.and Tangyunyong, Paiboon, *Power Spectrum Analysis (PSA) for Detecting Counterfeit Electronics.*. United States: N. p., 2017, ,https://www.osti.gov/servlets/purl/1476785

[2] Katie Liszewski and Thomas Bergman, "Battelle Barricade: Microelectronic Device Authentication and Counterfeit Detection using Power Analysis," Proceedings from the 43rd International Symposium for Testing and Failure Analysis, pp 79-86 (2017)

[3] Walter J. Keller , "ADEC Depot Testing and the Shifting Counterfeit Electronic Threat", Center for Advanced Life Cycle Engineering (CALCE) Counterfeit Symposium, June 28, 2016

[4] Richard C. Ott, et. al. "A Separability Metric for Second Order Effects Data," Proccedings of GOMACTech Conference, San Diego, CA, March 16-20 (2020).

[5] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd Ed., Wiley, 2003

[6] Carl T. Boone, et al., "Detectability of Small Circuit Modifications by Second Order Effects Testing," Proccedings of GOMACTech Conference, San Diego, CA, March 16-20 (2020).

[7] Libero SOC User Guide, MicroSemi Corporation (2015)

[8] Paiboon Tangyunyong, et al., "Power Spectrum Analysis (PSA)." United States, https://www.osti.giov/servlets/purl/146244

[9] Paiboon Tangyunyong, et al., US patent **9,188,622.**

[10] Mary Boas, *Mathematical Methods in the Physical Sciences*, 2nd ed., John Wiley and Sons (1983)

[11] The determination of off harmonic signals as noise was also made by calculating PCA-based t-tests for the off-harmonic data only and finding no separability with good confidence.

[12] *Encyclopedia of Machine Learning*, ed. Claude Sammut and Geoffrey I. Webb, Springer (2010) https://doi.org/10.1007/978-0-387-30164-8

[13] Descriptions of Mathematica functions can be found at Wolfram Research https://resources.wolframcloud.com/FunctionRepository

[14] This assumes the t^2 values are increasing much more slowly with higher principal components. If t^2 continues to increase rapidly, p-values will also decrease rapidly