SAND2018-3511C

Semi-Supervised Learning and ASIC Path Verification

James Obert*, Tom J. Mannos
Sandia National Labs
Albuquerque, NM, USA
{jobert, tjmanno}@sandia.gov

Abstract— **To counter manufacturing irregularities and ensure ASIC design integrity, it is essential that robust design verification methods are employed. It is possible to ensure such integrity using ASIC static timing analysis (STA) and machine learning. In this research, uniquely devised machine and statistical learning methods which quantify anomalous variations in Register Transfer Level (RTL) or Graphic Design System II (GDSII) formats are discussed. To measure the variations in ASIC analysis data, the timing delays in relation to path electrical characteristics are explored. It is shown that semi-supervised learning techniques are powerful tools in characterizing variations within STA path data and has much potential for identifying anomalies in ASIC RTL and GDSII design data.**

Keywords—ASIC Verification; ASIC Design Assurance; ASIC Integrity; Trusted ASIC Design; Supply Chain Integrity.

James Obert and Tom Mannos are actively involved in ASIC design research at Sandia National Labs. The research described in this paper was funded by the Sandia National Lab's Lab Directed Reseach and Development program.

I. INTRODUCTION

A. Overview

Static timing analysis (STA) validates the timing performance of a design by checking all possible paths for timing violations. Because no actual functionality check is performed, STA does not require use case simulation nor vector generation. Conversely, Dynamic Timing Analysis (DTA) requires the generation of an exhaustive set of input vectors to check the design path timing and behaviour characteristics through simulation. The amount of analysis required for DTA versus STA is exponentially greater, and for this reason STA is the most often used ASIC design verification method. STA though more efficient, does not in itself have the capability to detect significant electrical path variations between STA instances (proxies) representing the same ASIC design. As an ASIC design is ported to a new technology library or a new analysis tool, verification using a baseline reference proxy is essential. The research in this paper illustrates how ASIC STA design verification is enhanced such that significant path variations are sensed within an ASIC design using semi-supervised machine learning [1]. Semi-supervised learning is utilized to identify anomalous ASIC design paths by comparing fully-labelled STA electrical path characteristics of a baseline STA proxy with other STA proxies of the same ASIC design.

As most design alternations are initiated via modifications to Register Transfer Level (RTL) [2] or Graphic Data System II (GDSII) format [3-4], this research investigates methods capable of sensing when an unintended change has been made to a design. The STA toolsets used in the research utilize variants of Critical Path Methodology (CPM) [5] and Program Evaluation and Review Techniques (PERT) [6] to find the worst-case delay of the circuits over all possible input combinations (See Section I.B). Using a devised algorithm for identifying proxy path differences called Semi-Supervised Anomalous Path Detection (SAPD), it shown that individual cross-proxy path variations are efficiently identified.

The content of this paper is arranged as follows. In Section II relevant STA equations are described along with the analysis approach and assumptions that were made. In Section III path group and individual path comparative analysis using our semi-supervised learning methodology is explained. Finally, in Section IV alternative statistical methods are described and contrasted with the semi-supervised algorithm. Finally, in Section V concluding comments are made along with future recommendations for related research.

B. Background and Assumptions

In this section, the STA analysis equations are briefly discussed. The equations below define the propagation characteristics of a signal along a path:

t_{req} : Worst case maximum time for a signal to propagate from its launch register (the start point) to its capture register (the endpoint).

T: Timing path clock period ($\frac{1}{f}$)

t_{ce}: Clock propagation time to the endpoint

t_s : Setup time of the endpoint which is a physical property of the memory element used for the register.

t_{arr}: Observed arrival time of signal

t_i: Gate delays on a path

t_{cs} : Clock propagation delay

t_{sl} : Setup margin or slack time

$$t_{req} = T + t_{ce} - t_s \quad (1)$$

$$t_{arr} = t_{cs} + \sum_i t_i \quad (2)$$

$$t_{sl} = t_{req} - t_{arr} \quad (3)$$

The slack time t_{sl} is an important metric in that it compares the worst case maximum signal propagation time t_{req} to the observed arrival time t_{arr}. A significant change in t_{sl} may indicate a path circuitry modification.

Considering equations 1-3, certain analysis approach assumptions are made when comparing BL1 (baseline) and VAR1 (test) proxies:

- Per Equation 3 as arrival time (t_{arr}) increases, the slack time (t_{sl}) decreases, and we expected to observe this when inspecting the slack and arrival times for each path group.
- It is assumed on average that the slack time for a path group during each STA session for BL1 and VAR1 should not vary greatly.
- It is assumed that the slack times and arrival times for each path are for all practical purposes statistically independent.
- To detect design changes between proxies globally, overall path group slack and arrival time relationships are inspected.
- When comparing individual proxy paths, it is expected that in many cases a variety of STA tools, libraries, and abstractions may be used. Therefore, it is plausible that slight variations between the statistical distributions of specific random variables such as pin capacitance, wire capacitance and delay will be present.

In Section II the principal methods used in performing *path group analysis* and quantifying the amount of variation present between proxies is described. Additionally, in Section II *individual path analysis* methods which quantify the amount of variation between proxies are defined.

II. Machine Learning and ASIC Path Analysis

A. Path Features

Each data field within the BL1 and VAR1 proxy datasets for our analysis is termed a feature as defined in Table 1 below.

Table 1 ASIC STA Features

Feature	Description
f_{s3}	Slack Time
f_{s4}	Arrival Time
f_{s5}	Pin Capacitance
f_{s6}	Wire Capacitance
f_{s7}	Transition Time
f_{f2}	Total Capacitance
f_{f4}	Delay Time

B. Path Analysis Flow

In Figure 1 below the same ASIC design is represented by the BL1 proxy and VAR1 proxy datasets. Individual and path group STA session data from these datasets are introduced to the SAPD system. In steps 1 and 2 baseline and test proxy

path group distributions and data clusters are compared to identify anomalous path groups. In step 3 those features having the largest influence on average on the delay time are identified within the proxies and this lowers the number of feature data distributions that need to be compared in step 4. Finally, in step 4 BL1 and VAR1 individual path distributions are compared and anomalous paths are identified.

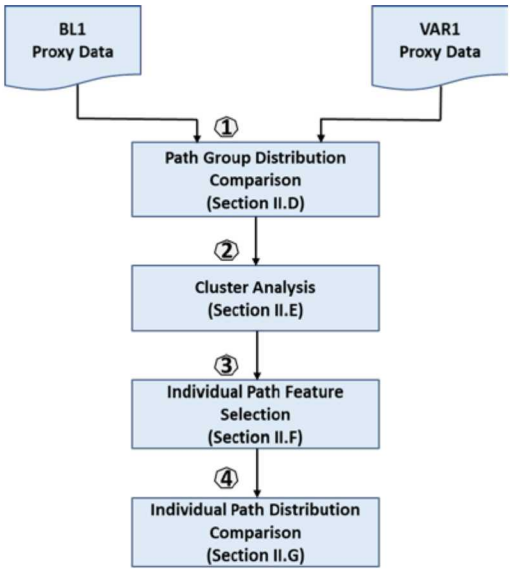


Figure 1: Semi-supervised Anomalous Path Detection.

C. Path Group Analysis

Proxy path files are organized such that groups of individual paths are associated with a path group designator. It should be noted that there is a one-to-one correspondence between paths and path groups within BL1 and VAR1 proxies. Specifically, the exact same number of distinct paths and path groups are present in each proxy file. Because BL1 and VAR1 proxies are assumed to be identical designs, it is assumed that only slight STA system induced variations between BL1 and VAR1 path group characteristics will exist. The methodologies used in comparing BL1 and VAR1 proxy path groups includes distribution and data clustering analysis which are explained in the sections that follow. Experimental results summarizing the performance of these methods are described in Section III.

D. Path Group Distribution Comparison

Quantitatively speaking distribution differences between BL1 and VAR1 proxies are measured using Kullback-Leibler Divergence (KL-Divergence) as defined below. Analyzing useful feature distributions, K-L Divergence measures the relative entropy increase between feature f_i and f'_i distributions. Where F_i is a baseline feature probability distribution and F_i' is a target or test probability distribution to be analyzed and i = 1, 2, ……,6.

$$D_{KL}(F_i'||F_i) = \sum_{f_i, f'_i} Prob(f'_i) \ln \frac{Prob(f'_i)}{Prob(f_i)} \quad (4)$$

D_{KL} >= 0 and is equal to zero when no change in entropy is detected between sampled sets.

E. Cluster Analysis

In cluster analysis BL1 and VAR1 proxy data characteristics are compared by using agglomerative hierarchical clustering. If no modifications have been made to the ASIC design, it is expected that BL1 and VAR1 clustering patterns will be similar. Agglomerative hierarchical clustering builds a hierarchy from the bottom-up, and doesn't require the that the number of clusters be specified beforehand as other methods do (ie. K-means). The algorithm starts by putting each data point into its own cluster, and then identifies the closest two clusters. The closest two clusters are then combined, and the process is repeated until all data samples are formed into a single cluster. The clustering process is represented as dendrogram (see Figures 4 & 5). There are several methods available for determining the distance between clusters throughout the agglomeration process. The two most common methods are complete linkage (CL) and mean linkage (ML) methods which either finds the maximum possible distance between points belonging to two different clusters or finds all possible pairwise distances for such points and then calculates the average respectively. Ensuring good clustering requires the measurement of the group sum of squares (GSS) and the within group sum of squares (WGSS) between cluster data samples while employing CL and ML.

Sections II.F and II.G describe methods capable of identifying significant differences between individual corresponding BL1 and VAR1 proxy paths.

F. Individual Path Feature Selection

In this section a Bayesian Variable Selection (BVS) method that calculates the activation probability for each data feature/predictor is explained. Those predictors exhibiting the highest activation probability on average are most likely to affect the value of the target variable (delay time). The BVS method is used to find the activation probabilities for wire capacitance (f_4), transition time (f_5), total capacitance (f_6) and pin capacitance (f_3). The regression expression as shown in Equation 5 is utilized within BVS to find the activation probabilities for each of the features. In Equation 5, β_3, β_4 ……. β_6 are the corresponding regression coefficients, α = regression constant and the response variable O = the delay time.

$$O = \alpha + \beta_3 f_3 + \beta_4 f_4 + \beta_5 f_5 + \beta_6 f_6 \quad (5)$$

BVS requires that selection of a prior distribution to predict a posterior distribution mean. If we assume:

$p(\mu)$: Prior distribution with μ representing the mean of the prior

$p(f_i|\mu)$: Likelihood of a sample f_i being observed given assumed mean μ

The posterior distribution is found using Equation 6:

$$p(\mu \mid f_i) = \frac{p(f_i \mid \mu)p(\mu)}{p(f_i)} \quad (6)$$

The normalizing constant $p(f_i)$ is found using Equation 7:

$$p(f_i) = \int p(f_i \mid \mu)p(\mu)d\mu \quad (7)$$

The normalizing constant $p(f_i)$ which is unconditional, ensures the posterior integrates to one. $p(f_i)$ is a multidimensional integral over all the model parameters and is approximated using Markov chain Monte Carlo (MCMC) algorithms [10].

G. Individual Path Distribution Comparison

Once the activation probabilities for the features are identified and those features exhibiting the highest significance are identified, individual path distribution analysis is performed using Equation 4. Those paths showing the greatest KL-Divergence between BL1 and VAR1 distributions are considered anomalous.

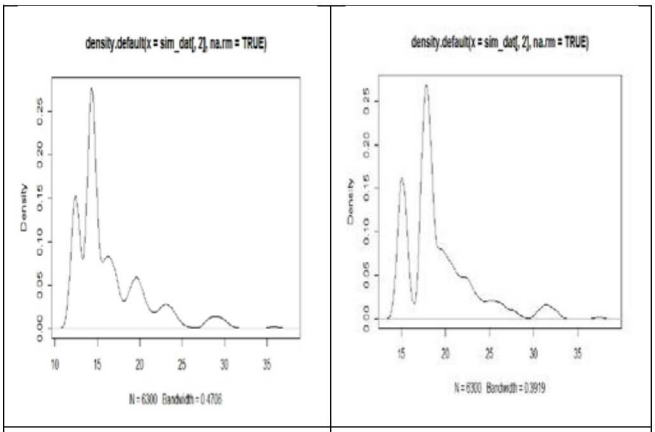
III. Experimental Results

A. Experimental Data

To illustrate the detection capabilities of the SAPD algorithm, two path group databases representing the same ASIC design were created. The baseline database designated as BL1, consisted of 6425 paths where the following defects were removed: data-entry errors, missing values, outliers, unusual (e.g. asymmetric) distributions, changes in variability, clustering, non-linear bivariate relationships and unexpected patterns. A second database was derived from BL1 designated VAR1 also consisted of 6425 paths, but random path variations were made in order to simulate unintended design modifications.

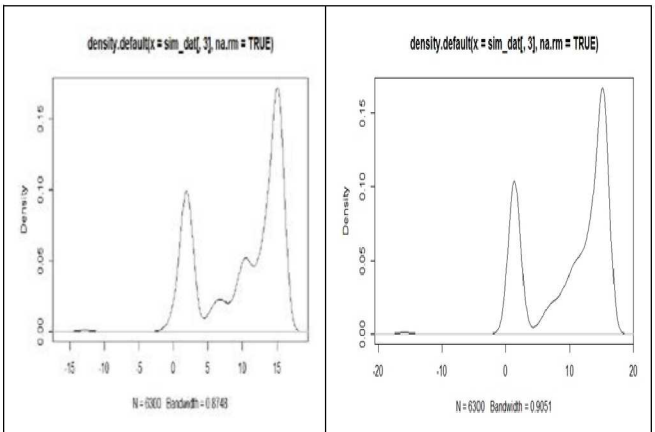
B. Path Group Distribution Analysis

In Figures 2 and 3 show BL1 and VAR1 path group distributions for arrival and slack times. A comparison of BL1 and VAR1 distributions showed significant differences in arrival time and slack time shapes indicating divergence and the possibility of path group modifications.



(a) BL1 arrival times (b) Var1 arrival time

Figure 2: BL1 & VAR1 arrival time distribution comparison.



(a) VAR1 slack time (b) VAR1 slack time

Figure 3: BL1 & VAR1 slack time distribution comparison.

C. Cluster Analysis

Cluster analysis was performed on BL1 and VAR1 path groups as shown in dendrograms in Figures 4 & 5 and the cluster plots in Figure 6 & 7 below.

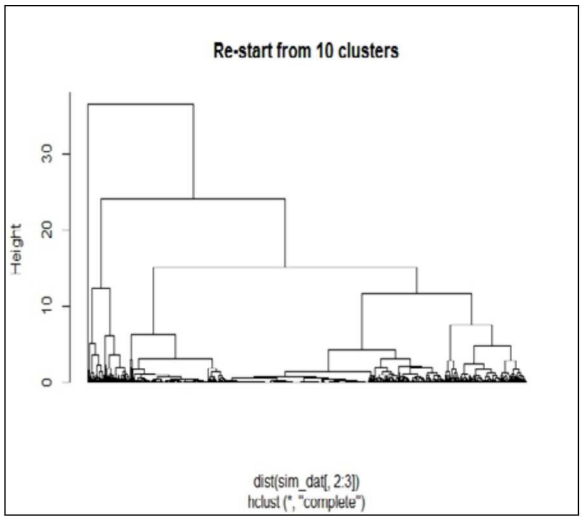


Figure 4: Dendrogram for BL1.

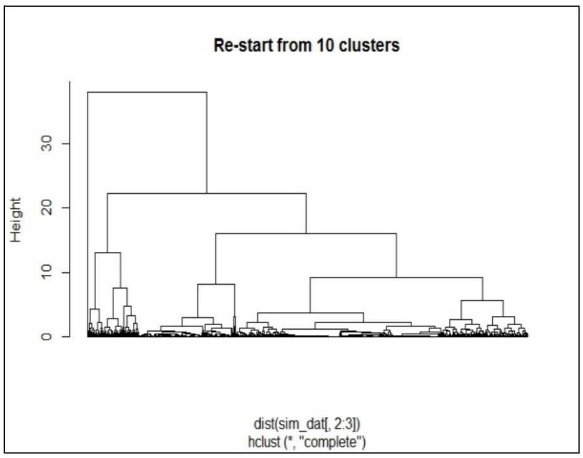


Figure 5: Dendrogram for VAR1.

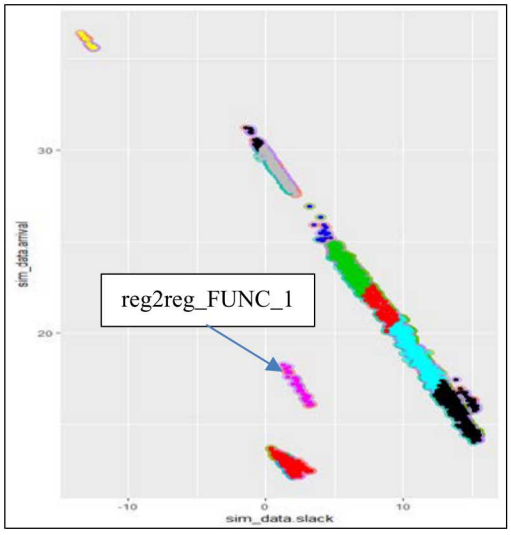


Figure 6: Clustering of BL1**.**

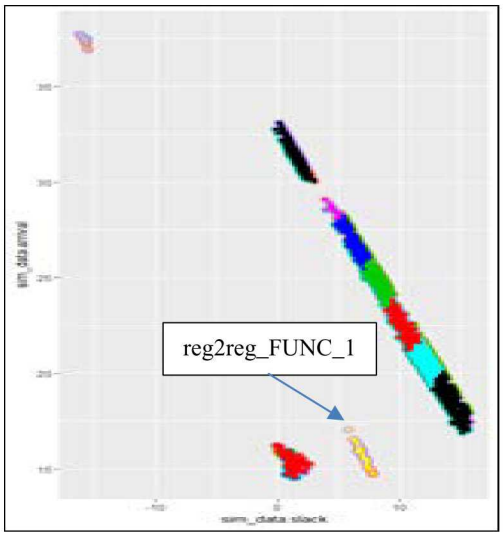


Figure 7: Clustering of VAR1**.**

Measurement of the group sum of squares (GSS) and the within group sum of squares (WGSS) between cluster data samples while employing CL and ML, and found that the highest value for the ratio of BGSS/WGSS occurred on average more often when using the CL method.

Comparing the clustering patterns between BL1 (Figure 6) with VAR1 patterns (Figure 7) it is evident the data vector components in BL1 and VAR1 proxies do not have identical values. Specifically, it is seen that a cluster shift between path group reg2reg_FUNC_1 occurs. As shown in Figures 8 and 9, the cluster composition between proxies was inspected to detect individual sample shifting between clusters.

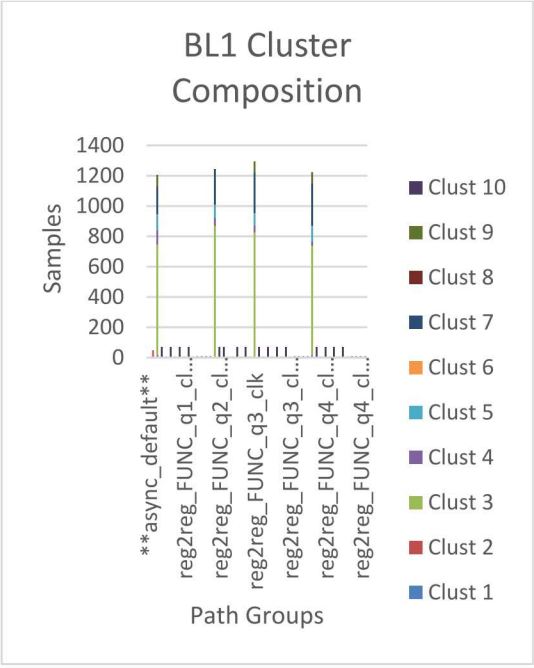


Figure 8: BL1 Cluster composition**.**

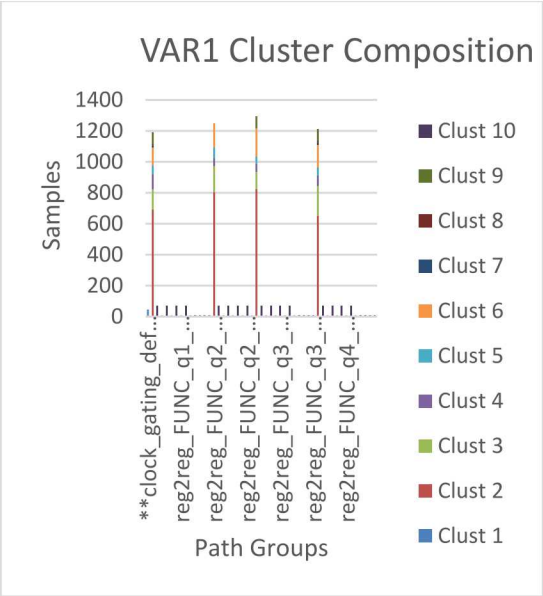


Figure 9: VAR1 Cluster composition**.**

Figures 10 BL1 and 11 VAR1 cluster compositions show small changes in individual cluster sample composition as well.

D. Individual Path Feature Selection Analysis

The BVS analysis results are shown in Figure 10 below.

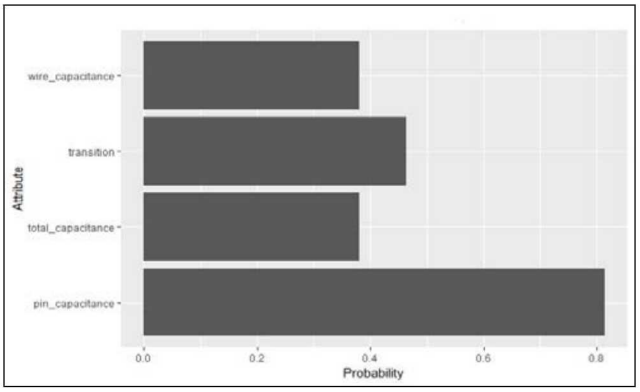


Figure 10: Feature activation probabilities

BVS analysis performed on the BL1 and VAR1 datasets consistently showed the pin capacitance feature (f_3) to have the highest activation probability ($f_3 > 0.8$) while all other features were ($f_4, f_5, f_6 < 0.5$).

E. Individual Path Distribution Comparison

With the pin capacitance feature (f_3) having an average activation probability ($f_3 > 0.8$) and all others ($f_4, f_5, f_6 < 0.5$), we confined the path distribution analysis to the differences between individual BL1 and VAR1 path pin capacitance distributions. Figure 11 below shows the results of the distribution analysis where the vertical axis is the KL-Divergence (Equation 4) between corresponding individual BL1 and VAR1 pin capacitance distributions, and the horizontal axis is the respective path numbers.

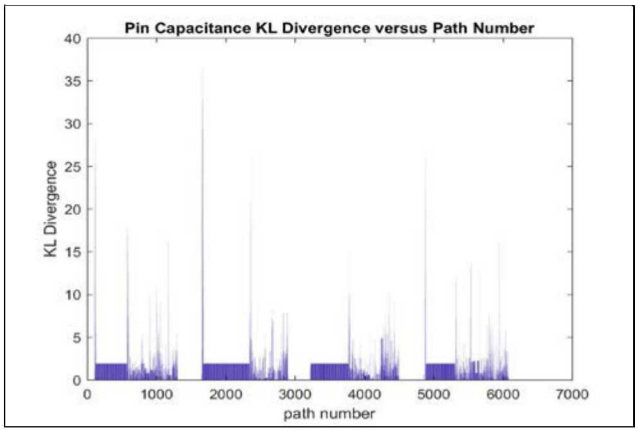


Figure 11: KL-Divergence of pin capacitance per path.

As Figure 11 shows, most feature f_3 distribution KL-Divergence (f_3 KLD) values fall approximately into the range of ($5 \geq f_3 \geq 0$). Those f_3 path distributions outside of the ($5 \geq f_3 \geq 0$) range required further inspection as these paths are anomalous. The degree to which individual f_3 KLD values were found to be anomalous was measured by establishing a set of thresholds (T_n, n = 1, 2, ….., 16) and observing which individual f_3 KLD values exceeded each T_n. T_n was calculated by first taking the mean (m) and standard deviation (std) of all f_3 KLD values, and then multiplying std by a factor (F_n). Each T_n value is equal to std times a distinct F_n added to m as illustrated in Equation 8.

$$T_n = m + std * F_n \quad (8)$$

Figure 12 below shows the results of the thresholding analysis. As illustrated in Figure 12 as F_n increases T_n increases, which results in the number of identified anomalous paths (P_n) to decrease exponentially. Table 2 shows a sample of identified anomalous paths and the respective f_3 KLD values when n = 13 and F_{13}.

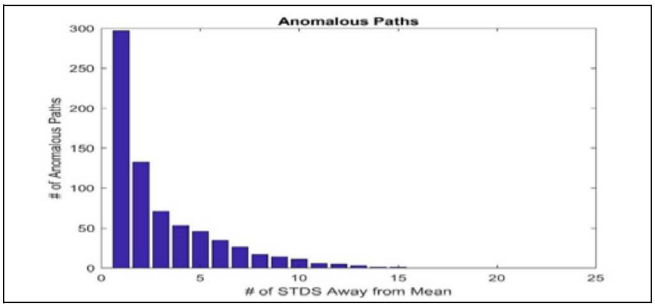


Figure 12: Identification of anomalous paths.

Table 2 Path KLD	
Path #	KLD Value
1665	32.92
1666	36.46
1667	29.11
1668	32.10

IV. Discussion

A. Statistical Methods Explored

Comparisons were made between standard statistical methods (cross-correlation and regression) and SAPD. Correlation between slack and arrival times for both the BL1 and VAR1 proxy datasets were measured. The correlation coefficients and thus the degree of correlation between slack and arrival time using both the Pearson and Spearman methods were found. The Pearson correlation coefficient is based on the degree to which slack and arrival times change linearly relative to one another. If non-linearities exist between the variables, the Pearson method lacks the accuracy of the nonlinear Spearman method [7].

In Figures 13(a) BL1 path group arrival and slack time vectors are correlated and cross-correlated. In Figures 13(b), VAR1 path group arrival and slack time vectors are correlated and cross-correlated as well. As shown, significant differences were observed between the correlation values Pearson: {-0.1920 versus -0.5277}, Spearman: {- 0.0258 versus 0.0534} between BL1 and VAR1 proxies. Because an identical order and one-to-one correspondence between BL1 and VAR1 path groups exists, the cross-correlation [8] values of arrival and slack times were compared as well. As shown we also observed significant differences in these values as well between plots 13(a) and 13(b).

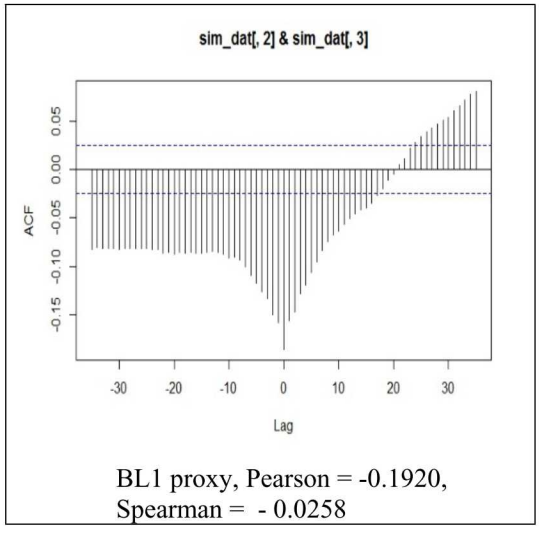


Figure 13(a)**:** Correlation analysis of all path groups BL1

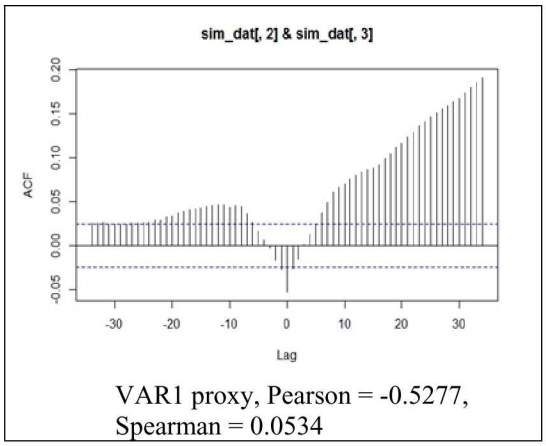


Figure 14(b)**:** Correlation analysis of all path groups VAR1

Correlation values using both Pearson and Spearman were found to be relatively low negative correlations with Pearson performing better with the inherently linear relationship between slack and arrival time. Because there is a relatively large amount of value change between correlation values between the two proxies, correlation as an ASIC modification indicator was not found to be a reliable metric when many outlier points are present.

Linear regression modeling was performed to estimate the relationship between slack and arrival times with slack time as the dependent/response variable and arrival time as the independent/predictor variable. Figure 14 below shows the results of the analysis.

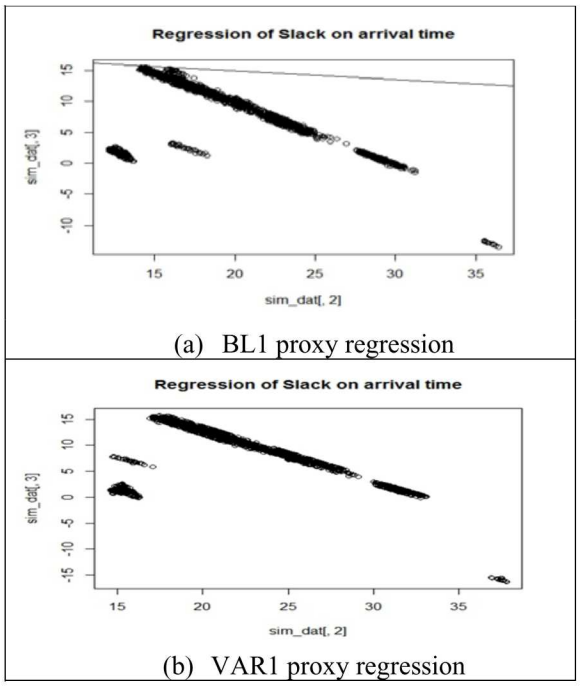


Figure 15: Plot of Slack time versus Arrival time. (slack time vertical axis, arrival time horizontal)

As shown in Figure 14 and per Equation 3 slack time decreases linearly as arrival time increases. It was observed in Figure 6 that this linear relationship existed for disjointed clusters of data with a principal linear cluster of samples at the top of the plot. Table 3 below shows the regression statistics for BL1 and VAR1 proxies.

Metric	BL1	VAR1
Adjusted r^2	0.0367	0.00263
p-value	2.2e-16	2.78e-5
Residual standard Error (RSE)	4.03	3.93

Table 1: Regression Statistics

The adjusted r^2 measures the amount of variation witnessed in slack time when the arrival time value changes. Table 3 indicates that only a slight change in slack value occurs when arrival times change: approximately 3.6% for BL1 and 0.26% for VAR1 proxies. A low p-value (p-value < 0.05) for a predictor (arrival time) is significant in the regression model. Both BL1 (p-value = 2.2e-16) and VAR1 (p-value = 2.78e-5) are much lower than 0.05 and thus arrival time value changes significantly affect the variation in the value of slack time. The residual standard error (RSE) values for both BL1 (4.03) and VAR1 (3.93) indicate the regression models as slack time prediction tools will exhibit a nominal amount of error when many outlier points are present.

There were notable differences observed between the data plots of Figure 14. This difference is both a horizontal and vertical shift in the notated data cluster in 14(a) and 14(b). From the Table 3, it was apparent that significant differences between BL1 and VAR1 proxy adjusted r^2 and p-values existed.

B. Method Comparisons

Cross-correlation and regression techniques could identify path group variations between BL1 and VAR1 proxies, but unlike SAPD could not be used to identify distinct path anomalies. Although distinct path anomalies could not be verified using either method alone, the regression linearity properties (r^2 and p-value) provided a wealth of predictive information.

V. CONCLUSIONS AND FUTURE WORK

SAPD provides both a coarse and find grained approach to ASIC verification during manufacturing. Using lightweight unsupervised learning (clustering) it can effectively sense path variations between proxies. Additionally, SAPD is capable of efficiently identifying individual path anomalies by first using BVS to reduce the dimensionality of the detection model and then sensing which paths have been altered using KL-Divergence.

Armed with SAPD as a lightweight path variation detection algorithm, future research should include unique methods for self-correction of anomalous paths.

REFERENCES

[1] Chapelle, Olivier, Bernhard; Zien, Alexander (2006). *Semi-supervised learning*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-03358-9

[2] Vahid, F. (2010). "Digital Design with RTL Design, Verilog and VHDL (2nd ed.)." John Wiley and Sons. p. 247. ISBN 978-0-470-53108-2.

[3] Buchanan, J. R., "The GDSII Stream Format", June 11, 1996.

[4] GDSII™ Stream Format Manual (B97E060), Release 6.0, Calma, February 1987.

[5] Kelley, J., M. Walker, "Critical-Path Planning and Scheduling.", 1959 Proceedings of the Eastern Joint Computer Conference.

[6] Brennan, M., "PERT and CPM: a selected bibliography", Monticello, Ill., Council of Planning Librarians, 1968. p. 1.

[7] Croxton, F. E., Cowden; D. J., Klein, S. (1968) "Applied General Statistics", Pitman. ISBN 9780273403159 (page 625).

[8] Bracewell, R., "Pentagram Notation for Cross Correlation." The Fourier Transform and Its Applications. New York: McGraw-Hill, pp. 46 and 243, 1965.

[9] Freedman, D. A., *Statistical Models: Theory and Practice*, Cambridge University Press (2005).

[10] Gelman, A., Carlin, John B.; Stern, Hal S.; Dunson, David B.;Vehtari, Aki; Rubin, Donald B. (2013). *Bayesian Data Analysis*, Third Edition. Chapman and Hall/CRC. ISBN 978-1-4398-4095-5.