Clustering of Assertions using Machine Learning in Formal Verification

Viraj Y. Rawal S R Pavitra Vishwajith V. Rao *

*Synopsys India Private Limited, Bengaluru, Karnataka, India

Motivation

1. Verification Engineers verify a design by writing System Verilog Assertions.
2. Formal Verification Environment (FVE) typically has many System Verilog Assertions (SVA) for the Verification of a specific Design Under Test (DUT).
3. Initial phase of the Verification will have many failing System Verilog Assertions.
4. Problem: Identifying the System Verilog Assertions failing due to the same root cause.

Figure 1. Design Under Test

Failing System Verilog Assertions:

1. No Push when FIFO is Full.
2. No Pop when FIFO is Empty.
3. FIFO Full and FIFO Empty can not happen at the same time.
4. Write Pointer should roll back to zero if incremented beyond DEPTH.
5. Read Pointer should roll back to zero if incremented beyond DEPTH.

There is a possibility that multiple assertions are failing due to the same root cause.

Challenges:

1. Debugging each Assertion individually will consume a lot of the engineer's time and effort.
2. Wastage of compute resources.
3. Risk of exceeding the project deadline.

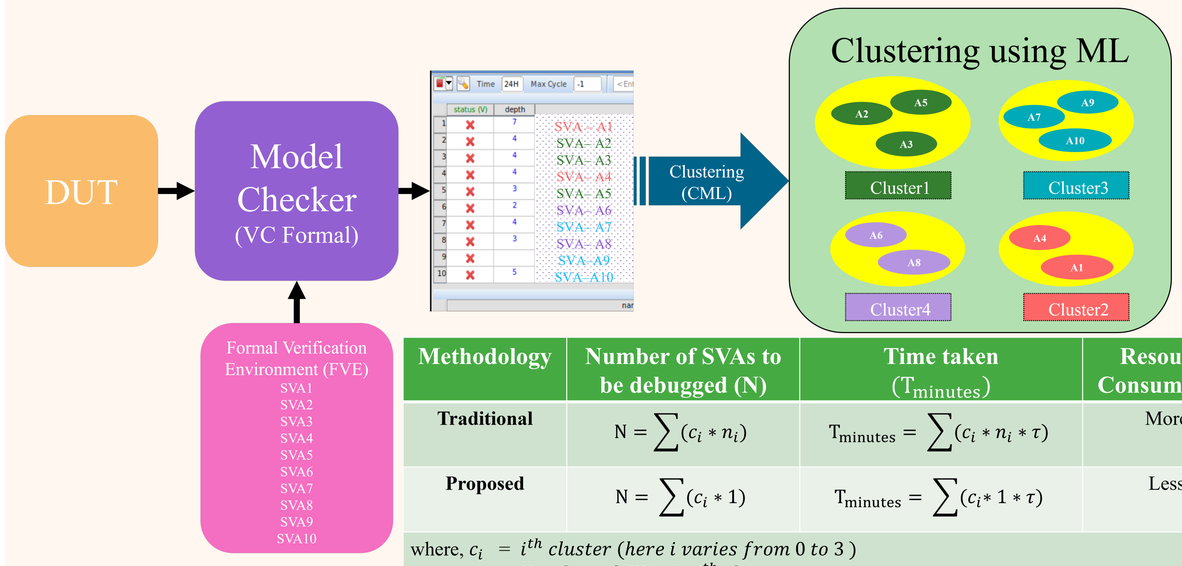
Main Idea (Overview)

Formal Verification Environment (FVE):

Failing System Verilog Assertions and its related data extracted from Model Checker (VC Formal).

Clustering using Machine learning (CML) :

Clustering the Failing Assertions with similar root cause identified based on the dataset given as the input to the Clustering Algorithm (Unsupervised Learning Algorithm).



Methodology	Number of SVAs to be debugged (N)	Time taken ($T_{minutes}$)	Resource Consumption
Traditional	$N = \sum(c_i * n_i)$	$T_{minutes} = \sum(c_i * n_i * \tau)$	More
Proposed	$N = \sum(c_i * 1)$	$T_{minutes} = \sum(c_i * 1 * \tau)$	Less

where, c_i = i^{th} *cluster (here i varies from* 0 *to* 3)
n_i = *Number of SVAs in* i^{th} *cluster*
τ = *Approximate time taken to debug* 1 *SVA*

Figure 2. Overview of Clustering of Assertions with the similar root cause

What are the Main Contributions?

1. **Clustering using ML**: Proposed the Machine learning-based methodology to cluster the System Verilog Assertions failing due to similar root cause.
2. **Data set**: Clustering (Unsupervised Model Learning) is applied on readily available data to the user corresponding to each failing System Verilog Assertion (SVA) from Synopsys VC Formal tool.

Results and Discussion

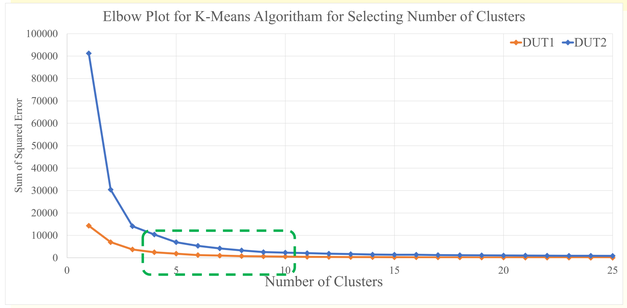


Figure 3. Elbow Plot: Sum of Squared Error v/s Number of Clusters

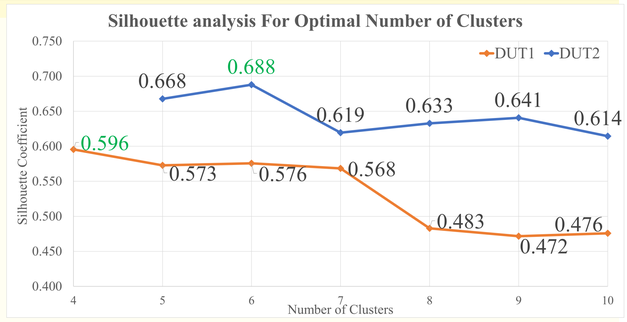


Figure 4. Silhouette Coefficient v/s Number of Clusters

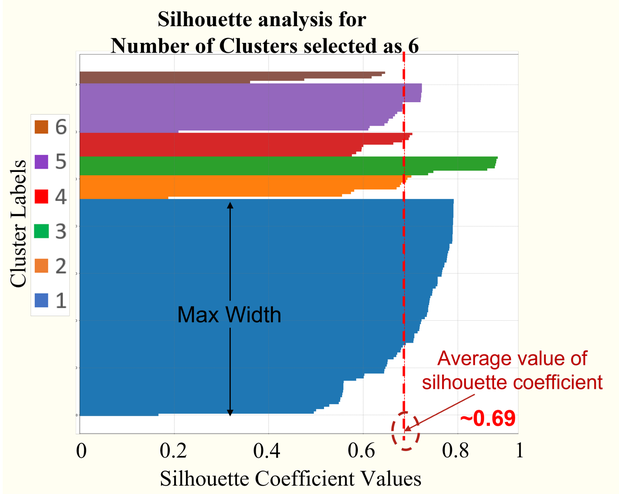


Figure 5. Analysis for Selecting Optimum Number of Clusters

Table 1. Results and Validation

Attributes	DUT 1	DUT 2
Registers	285	1364
Primary Inputs	44	71
Primary Outputs	29	38
Total SVAs	128	475
Falsified SVAs	43	146

Results: Clustering Using ML (CML)		
Silhouette Coefficient	0.6	0.69
Optimal Number of Clusters	4	6
Purity	74.42%	65.07%

1. Sum of squared error obtained by the K-Means algorithm results in an Elbow Plot, which helps to decide a suitable range of cluster numbers as shown in Fig 3.
2. Further, Silhouette Analysis is used to get the optimal number of clusters from the range given by the Elbow Plot as shown in Fig 4.
3. Silhouette Analysis results in a Silhouette Coefficient for each cluster number.
4. Highest value of the Silhouette Coefficient helps to identify the optimal value of the cluster number.
5. Purity is used as a measure to validate the results given by ML Algorithm as shown in Table 1.

Summary

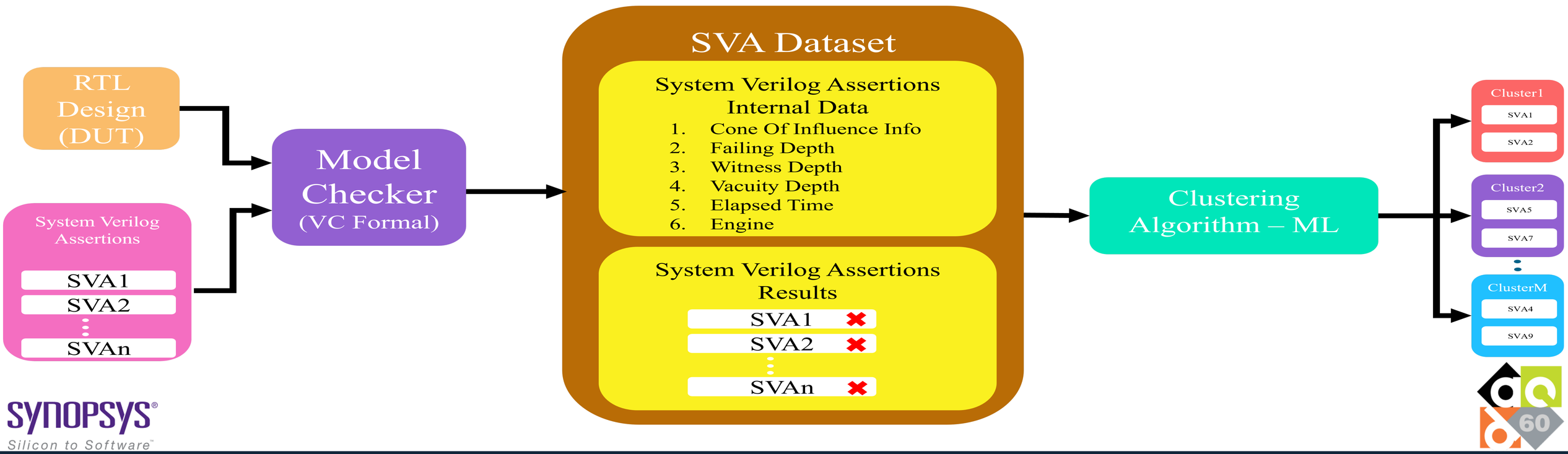
1. It is challenging for Formal Verification engineers to debug a large number of falsified assertions having a similar root cause – resulting in time consumption and wastage of resources.
2. The approach of clustering assertions using Machine Learning helps to ease assertion debugging and improve the overall effort estimation.
3. Machine Learning Algorithm is fed with data from Model Checker (Synopsys VC Formal Tool), which results in the formation of Clusters of Assertions having a similar Root Cause.

What are the Advantages?

1. It reduces verification engineers' efforts and time spent in analyzing and debugging multiple failures by correctly identifying the clusters given by the proposed methodology.
2. It gives the Verification Managers/Leads the key to efficient Verification task planning and execution by correctly estimating the efforts and resources required for the Verification task.

Practical Implications

1. **What is the Run time of the Machine Learning Algorithm in Formal Verification Environment if it has 1000+ System Verilog Assertions?:** Machine learning is normally quite fast- The typical time to generate the clusters is within a few minutes.
2. **Is there any prior step to using the proposed methodology?:** Yes, the user needs to run the Formal Tool on the coded System Verilog Assertions first.



SYNOPSYS®
Silicon to Software™