Machine Learning in Physical Design

Bowen Li
Electrical and Computer Department
North Carolina State University, Raleigh, NC
bli11@ncsu.edu
Tel: 919-561-3601

Paul D. Franzon
Electrical and Computer Department
North Carolina State University, Raleigh, NC
paulf@ncsu.edu
Tel: 919-515-7351

***Abstract*—Machine learning, a powerful technique for building models, can rapidly provide accurate predictions. Since Integrated Circuit (IC) design and manufacturing have tremendously high complexity and enormous data, there is a surge in adapting machine learning approach in IC Design stages, as machine learning can provide fast predictions. Recently, machine learning has been used in some IC Design stages (e.g. Physical Verification), but not in Physical Design. In this research, machine learning is adapted to Physical Design. Surrogate Modeling is implemented to predict results after GR in Physical Design. Machine learning models for predicting Detailed Route (DR) results using Global Route (GR) results are also discussed. With surrogate models and machine learning methods, circuit performances after Physical Design (e.g. hold violation check and area) would be predicted quickly.**

Keywords—Surrogate Modeling; Physical Design; Global Route; Detailed Route; machine learning; neural network; decision tree; relationship; prediction

I. INTRODUCTION

As the size of semiconductor process technology nodes further scales down, the industry is greatly challenged in terms of IC design and manufacturing. During IC design, Physical Design can convert circuit components into an integrated circuit layout. Only after Physical Design, can the circuit layout be checked as to whether it satisfies Design Rules. However, physical Design is a very time-consuming process. How to quickly predict circuit performances after Physical Design is one of the key challenges in IC design. Machine learning is one way to predict results for such complex processes. Nowadays, machine learning adapted in Physical Verification has been proposed [1]. As far as we know, machine learning adapted to Physical Design hasn't been explored. If such an approach is found, performances after Physical Design (e.g., hold slack, area and power) will be predicted rapidly.

One way to predict the results for complex processes is to use machine learning (i.e., neural networks and decision trees). Thousands of randomly selected inputs and outputs should be used in machine learning processes to train a model. However, Physical Design is very complex. Inputs should be chosen consciously, aiming to build an accurate model. Surrogate Modeling (SUMO) has such sampling methods and many machine learning models to use [2], for instance kriging and radial basis function. Surrogate modeling is a math method, which is constructed using a data-driven approach. Within a surrogate model, attention is given to estimating relations between inputs and outputs, while ignoring the physical aspects of the phenomenon.

In this paper, surrogate models for predicting results after the GR step is proposed. It will save lots of time modeling GR results instead of DR results in Physical Design. SUMO can also sample inputs, aiming to build better models. Models for predicting DR results using GR results have been found using a machine learning approach; GR can be related to DR. Random sampling inputs are suitable for this process. Finally, people can use this framework to predict the performances of the circuit design after Physical Design within seconds, instead of running the Physical Design simulation for 40 minutes.

II. FRAMEWORK

The framework of Physical Design using machine learning used in this paper is shown in Fig. 1. In Stage 1, Surrogate Modeling would run Physical Design thousands of times to obtain enough inputs and outputs after GR. During this process, SUMO will generate models for each outputs to predict GR results in the future. In Stage 2, thousands of GR results and DR results are set as inputs and outputs respectively in machine learning models. After training, these machine learning models can precisely predict results after DR using GR results.

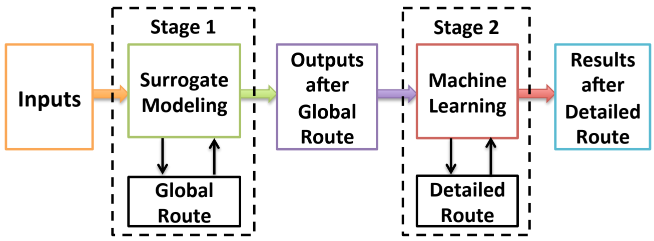

Fig. 1. Framework of Physical Design Simulation

There are two reasons to predict outputs after GR instead of DR using Surrogate Modeling in Stage 1. First, it consumes lots of time to build models for outputs after DR directly. Second, building surrogate models needs particular samples for different model builders. SUMO has such sampling methods. In Stage 2, GR can provide some instructions to DR, which means that they have direct relations. Random sampling and machine learning models can predict results after DR using GR results.

This framework can help people build models quickly, while SUMO for DR will consume so much time. GR results during Physical Design would also be saved. Thanks to surrogate models and machine learning models in these two stages, the framework can be used to predict results after Physical Design very quickly.

III. SURROGATE MODELING METHODOLOGY

A. Stage 1 in the Framework of Physical Design

In Fig. 1, Surrogate Modeling at Stage 1 will sample input features, and send those samples to Physical Design flow. After GR, the most concerned outputs will be sent back to Surrogate Modeling. Then surrogate models for each output would be generated. After that, the errors for each current model will be measured. If the error is not met with the target, new samples will be selected and sent to Physical Design. Surrogate Modeling will run the process mentioned thousands of times above to generate the best models for results after GR in Physical Design.

B. Selection of inputs and outputs

There are hundreds of features in the Physical Design flow. It's important to choose the most influenced features and most concerned outputs. At first, 4 target results have been chosen after DR, namely the number of DRC violations, hold slack, power and area, which are very concerning issues in Physical Design flow and the manufacturing process. There are six features, which influence the targets most, namely target clock period (clock_target), number of layers (num_layer), init density (init_density), aspect ratio (ratio), max transaction time at sink (SinkMaxTran) and max transaction time at buffer (BufMaxTran). Hence, SUMO has 6 inputs.

Outputs after GR should also be selected for SUMO. Worst negative slack (WNS), total negative slack (TNS), the number of violating paths (violating_path) and hold slack (hold_slack_trial) after GR would be used to predict final hold slack. In the congestion distribution report, the sum of the proportion of remaining tracks from -5 to -8, -1 to -4, 0 to 10, and 11 to 20 are set as four groups, labeled as x_neg_5_8, x_neg_1_4, x_pos_0_10, x_pos_11_20 respectively. These four groups can predict the number of DRC violations. As for power, power results after GR (power_trial) may predict the final power. Considering area, area after GR (area_trial) would predict the final area.

C. Model Builder Performances

After inputs and outputs are selected, the surrogate model builder would be chosen. The common used models are kriging [3] and artificial neural networks (ANN) [4]. Besides those, there are also some adaptive model builders. Each adaptive model builder is a combination of a model type and an optimization algorithm to choose the model parameters. For example, a model builder named kriginggenetic builds a kriging model using genetic optimization.

In this paper, the root relative square error (RRSE) is used to test model accuracy. When the RRSE is nearly 0, the model is a perfect fit. When the RRSE is nearly 1, the model is not as good as simply predicting the average of the actual values. In this research, an $RRSE < 0.5$ is the target.

$$RRSE = \sqrt{RSE} = \sqrt{\sum \left(f\left(\vec{x}_i\right) - y_i \right)^2 / \sum \left(\overline{y}_i - y_i\right)^2} \qquad (1)$$

Where $f\left(\vec{x}_i\right)$ is the value predicted by the Surrogate Model for one sample case; y_i is the target value from physical design flow result; $\overline{y}_i$ is the mean of total target values.

During Surrogate Modeling, 16 models have been used, including kriging, radial basis function, Gaussian process, rational, ANN, neighbor interpolation, extreme learning machine, and support vector machine. Each model builder generates 10 surrogate models for each outputs after GR. After training, 8 outputs, namely area_trial, WNS, TNS, violating_path, and four groups of remaining tracks, can be modeled properly with particular models. While, hold_slack_trial and power_trial can't be modeled by any model builders, because the RRSE of these 2 outputs are nearly 1. After comparison, the anngenetic model builder provides the best performance. Ann, annfixed and kriginggenetic model are also better than other models.

The reason why hold_slack_trial and power_trial cannot be modeled is that all these results have very complex relationships with inputs. In addition, the presence of noise and lack of representative samples can also be constraints on building a good model.

IV. MACHINE LEARNING METHODOLOGY

In Stage 2, machine learning models (e.g., neural networks and decision tree) are implemented to predict results after DR using GR results. 1027 new samples created from Physical Design are used to build machine learning models.

First, linear relationships are analyzed among GR results and DR results. The correlation coefficient can measure the strength and the direction of a linear relationship. The value of the correlation coefficient is in the range of [-1, 1]. If two data sets have a strong positive or negative linear correlation, this value is close to +1 or -1 respectively. If there is no linear correlation or a weak linear correlation, it's close to 0.

The correlation coefficients between 4 target outputs of DR and 10 outputs from GR are analyzed. Power and power_trial, area and area_trial have strong linear correlations, which can be used to construct models with linear regression. Meanwhile, hold slack has a linear relation with hold_slack_trial and power_trial. The number of DRC violations has a linear relation with area_trial.

A. Power and Area

As discussed above, linear regression would be used to find relationships between power_trial and power as well as between area_trial and area. The power and area model function are given as:

$$power = 0.95755 \times power\_trial + 8.3282e-05 \qquad (2)$$

$$area = area\_trial \qquad (3)$$

Using these functions, we can correctly predict the final power and area from GR result.

B. Hold slack

From I.Ahold slack has a linear relation with hold_slack_trial and power_trial. However, linear regression cannot give an accurate prediction. In this case, neural network and decision tree are common methods.

1) Neural network

The 10 results after GR are set as inputs and the final hold slack values are set as output in neural network. 80% of 1027 data points are set as training samples, 10% are set as validation samples and 10% as testing samples. The number of hidden neurons is 30. The Mean Squared Error (MSE) is used to measure the difference between outputs and targets. The training MSE is 6.41743e-3 and testing MSE is 7.40701e-2. Predicted values have a close relationship with target values; in this case, neural networks model for hold slack can correctly predict results after DR using GR results.

2) Decision tree

A decision tree is a classification technique (or classifier), which is a systematic approach to building classification models from input data sets. A decision tree can easily handle redundant or irrelevant attributes.

Before a decision tree for hold slack is built, one assumption should be made: if hold slack is extremely close to 0 (e.g., -0.035), the circuit can pass hold violation check under careful design. Based on this assumption, a decision tree using two attributes, hold_slack_trial and power_trial, is shown in Fig. 2. In this structure, hold slack will be predicted within a range. The training error rate for this decision tree is 0.36%; the testing error rate is 0.98%. Hence, a decision tree can also provide an accurate prediction for hold slack.

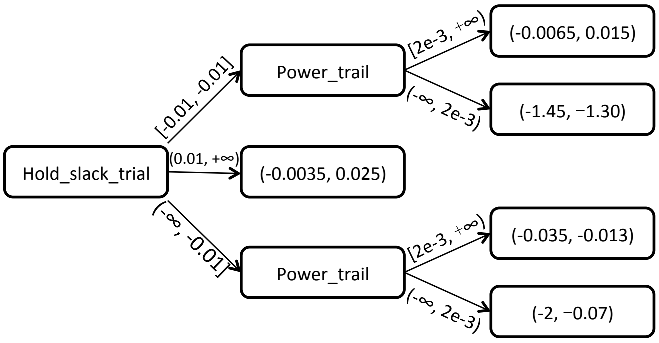

Fig. 2. Decision tree for hold slack

C. The number of DRC violations

As discussed in section II, the proportion of remaining tracks may predict the number of DRC violations. Four groups of remaining tracks are set as 4 inputs and the number of DRC violations is set as output in neural network and decision tree. area_trial is also set as an input in the decision tree, because it has a linear relationship with the number of DRC violation.

1) Neural network

80% of 1027 data points are set as training samples, 10% are set as validation samples and 10% as testing samples. The number of hidden neurons is 50. After training, although predicted values are closely related to target values, the training MSE and testing MSE are very large. The overall performance of the neural networks model for the number of DRC violations is not good. In this case, the neural network model cannot correctly predict the number of DRC violations after DR.

2) Decision tree

Because the number of DRC violations determines whether the physical layout of a particular chip satisfies design rules, it doesn't need to be predicted precisely. The assumption is made that if the number of DRC violations is fewer than 10, there will be no DRC violations under careful design. If the number of DRC violations is greater than or equal to 10, there would be many congestion in the design. Hence, a data set of the number of DRC violations can be divided into two groups: (i) the group that has fewer than 10 DRC violations, labeled as "Yes"; (ii) the group that has at least 10 DRC violations, labeled as "No". 80% of 1027 data points are set as training samples, and 20% as testing samples. The pruned decision tree for DRC violations using these two groups with 5 attributes is shown in Fig. 3.

In this decision tree structure, the conditions are shown in each nodes. Black arrows or red arrows mean the data meet or didn't meet the conditions respectively. The original tree is pruned down to 6 layers. Training error is 3%, and testing error is 17.18%. Therefore, the decision tree can be used to predict the number of DRC violation using GR results.

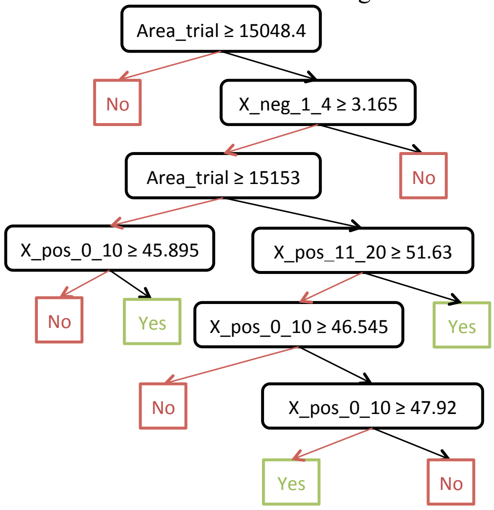

Fig. 3. Decision tree for DRC violation number

V. Conclusion And Future Work

An innovation approach is proposed in this paper, using SUMO to predict GR results and finding machine learning models for predicting DR results in Physical Design. There are two achievements in this paper: first, surrogate models are built to predict results after GR using Surrogate Modeling; second, machine learning models are found to predict DR results using GR results in Physical Design. These are important preparation works for implementing the framework in Section II. After trained, surrogate models for area, WNS, TNS, the number of violating paths and the proportion of remaining tracks can be built properly. The anngenetic model is the best model builder. In addition, linear regression models for the final power and area, neural network and decision tree model for final hold slack and the decision tree model for the number of DRC violations after DR are explored.

Future experimental work will be focused on finding proper surrogate models for hold slack and power after GR. Naïve Bayes Classifier can be adapted into the decision tree model for quicker convergence.

Acknowledgments

I would like to thank my advisor, Dr. Paul D. Franzon, for offering me this great research opportunity. I would also like to thank Dr. W. Rhett Davis, for teaching Physical Design.

References

[1] Bei Yu, David Z. Pan I, Tetsuaki Matsunawa, Xuan Zeng: "Machine Learning and Pattern Matching in Physical Design", *The 20th Asia and South Pacific Design Automation Conference*, 286 – 293, Jan. 2015

[2] Ilan Eskinazi, Benjamin J. Fregly: "An Open-Source Toolbox for Surrogate Modeling of Joint Contact Mechanics", *IEEE Transactions on Biomedical Engineering Volume:63*, 269 – 277, July 2015

[3] Forrester, A.I.J., Keane, A.J.: "Recent advances in surrogate-based optimization", *Prog. Aerospace Sciences 45*, 50–79, 2009

[4] Haykin, S.: "Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice-Hall", *Englewood Cliffs*, 1998

[5] S. Koziel, D. Echeverría Ciaurri, and L. Leifsson: "Surrogate-Based Methods", 2011

[6] Junyan Peng, P. P. K. Chan: "Revised Naive Bayes classifier for combating the focus attack in spam filtering, Machine Learning and Cybernetics (ICMLC)", *2013 International Conference*, 610 – 614, 2013