Random Forest Model for Silicon-to-SPICE Gap and FinFET Design Attributes Identification

Hyosig Won[1] and Katsuhiro Shimazu[2]
[1,2] Infrastructure Design Technology, System LSI, Samsung Electronics
1-1 Samsungjeonja-ro, Hwaseong-si, Gyeonggi-do 445-330, Korea
E-mail: [1]hs.won@samsung.com, [2]kats.shimazu@samsung.com

Abstract: Random forest model was applied to analyze design attributes influence on Silicon-to-SPICE(S2S) gap. In order to have enough model accuracy to discuss S2S gap, scaled learning data was used with random design attributes count to build each tree in forest model. From the improved model, indices so-called '*importance*' and newly defined '*impact*' can be extracted to identify significant design attributes which determine S2S gap. The identified design attributes classify S2S gap well and show clear trend of it. Finally the key FinFET structures can be identified as the representative layout structure to cause large S2S gap.

Keywords—**Random Forest, Attribute, Importance, Impact, SPICE, FinFET**

1. Introduction

To accelerate product yield ramp up, it is important to characterize the silicon device accurately by measuring device-under-test (DUT) designed exactly as same as that in real production chips. We have implemented cell-level model-hardware correlation vehicles (CMV) and product-level model-hardware correlation vehicles (PMV) in 10nm FinFET technologies. These vehicles consist of transistors tapped to test terminals by metal routing with real design layout same as cell library and product chip respectively.

Such vehicles include variety of devices with many different layout parameters also known as design attributes. Design attributes are layout features such as fin number, stack number, gate length, active cut types, area, density, and all possible local dimensions around transistor layout.

S2S gap discussed here is the difference in device characteristics between silicon DUT and SPICE simulation. We define it by the simple equation as (1).

$$\text{S2S gap} \equiv \frac{Idsat(measured) - Id \quad (SPICE)}{Idsat(SPICE)} \qquad (1)$$

where *Idsat*(*measured*) is the measured saturated drain current, *Idsat*(SPICE) is that of SPICE simulation at measured threshold voltage. For example, S2S gap = −0.1 means 10% smaller drain current than SPICE simulation. S2S gap may come from incorrect modeling for particular design layouts, high layout sensitivity to process fluctuation or defects, etc. Finding design attributes that result in large S2S gap and fixing the systematic causes behind them are crucial for timely yield ramp up.

But the number of design attributes is increasing significantly in the recent technology node, and the impact of design attributes is sometimes interdependent to each other. So it becomes more and more difficult to analyze the impact of individual design attributes accurately from the test vehicle such as CMV and PMV.

We applied random forest model [1] to analyze such test vehicle for the following reasons. First, the model can handle both of numerical data and class data (or may called categorical data) as input to predict numerical data as output. In our model, the input is design attributes which have both data types and S2S gap is output of numerical data. Secondly the model can handle many descriptive variables such as design attributes. And depending on the contribution to S2S gap determination, the important design attributes are selected in the model. The third reason is effectiveness to extract design attributes contribution to S2S gap by analyzing the model structure.

Random forest model is the ensemble of decision trees [2], the tree consists of continuous nodes, and the node is the most basic element to split data into next nodes. To build the model, a relational data set of design attributes and observed (measured) S2S gap is used as learning data.

In modeling process for each tree, data is randomly sampled from the learning data by so-called bootstrap sampling by which the data is sampled with the same size of learning data but allowing redundant sampling, in other words, data is returned to sampling source every time. And design attributes are also randomly selected for each tree modeling so that various types of trees are built. Figure 1 shows a node structure as data split branch. At each node in each tee, the best design attribute is selected as a data split criterion to split S2S gap data in order to maximize variance reduction in the path from one node into the next two nodes, i.e. left node and right node. The varance reduction dS is defined by below eqation (2) with the expression weithted by data count of each node.

$$dS \equiv Sm \times nm - Sl \times nl - Sr \times nr \qquad (2)$$

where Sm , Sl , Sr are variance of S2S gap data in the node to split, left node, right node, and nm, nl, nr are data count in the node to split, left node, right node respectively.

The data split criteria is defined by the best design attributes as below expression (3) or (4).

'$att = x$ or others' (if x is class/categorical data) (3)
'$att \leq x$ or others' (if x is numerical data) (4)

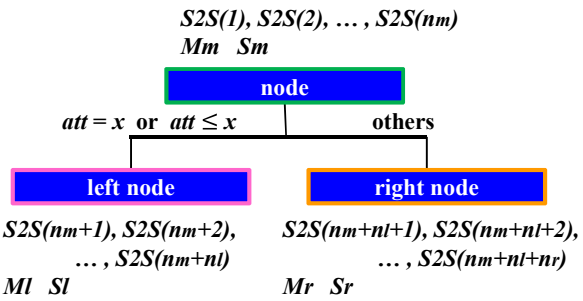

Figure 1. Node structure as data split branch. *S2S(i)* is *i*-th S2S gap data. *nm*, *nl*, *nr*, *Mm*, *Ml*, *Mr*, *Sm*, *Sl*, *Sr*, are data count, mean, variance in each node. *att*, *x* are design atribute and its value as a data spit criterion.

where *att* and *x* are the selected best design attribute and its value. The node split ends at a terminal node where there is no variance reduction any more by data splitting.

In prediction, node path of each tree is traced according to the input design attributes to reach the terminal node where the tree's prediction is given by the mean value of S2S gap data. The prediction by the random forest model is given by the mean value of all trees' predictions.

2. Model

To analyze design attributes influence on S2S gap precisely, the model acuracy was improved by two method, randam design attributes count and learning data scaling.

2. 1 Random design attributes count

In conventional random forest model, the design attributes count is fixed at $n/3$ or $\sqrt{n}$ for all the trees modeling, where n is total design attributes count to be taken into the model, i.e. design attributes count in the leaning data. Model fitness to learning data can be improved by increasing this count, but it may cause over-fit. To improve model accuracy considering over-fit, design attributes count is randomized, i.e. randomly selected from $\sqrt{n} \sim n$ for each tree modeling. Then with the selected count, design attributes are selected randomly to build each tree model.

To estimate the over-fit, over-fit rate of the random forest model was defined by below equation (5).

$$\text{over-fit rate} \equiv \frac{\Sigma(residue)^2 of\ testing\ data}{\Sigma(residue)^2 of\ learning\ data} \quad (5)$$

where $\Sigma(residue)^2$ means prediction error which is the sum of squared residue between observation data and prediction result by random forest model (not each tree's prediction), '*learning data*' is randomly selected 50% of S2S gap data and '*testing data*' is remaining 50% data which is not used to build model. If the design attributes count for tree modeling is simply increased, when it exceeds ~$n/2$, the over-fit rate starts to increase abruptly and reaches up to ~10 at the count equal to n. By applying random design attribute count for tree modeling, the over-fit rate is maintained at 1.0~1.2 and effective to improve the model fitness with avoiding or minimizing over-fit risk.

2. 2 Learning data scaling

Another way to improve the model accuracy is to scale the learning data. At first, pre-model is built by using the original learning data. Then scaling coefficients are extracted as linear regression coefficients between pre-model prediction result and the original learning data. Then the original learning data is scaled by the obtained scaling coefficients. If the liner regression model is given by equation (6), the original learning data is scaled by the same coefficients a and b as equation (7) in order to pre-magnify the learning data to compensate the gap between original learning data and pre-model prediction.

(original learnig data)
$$= a \times (\text{ pre-model prediction }) + b \quad (6)$$
(scaled learning data)
$$= a \times (\text{ original learning data }) + b \quad (7)$$

These scaling coefficients should be extracted and applied to scale the data for NMOS and PMOS separately because the best coefficients are different between device types. Then the final model is built by the obtained scaled learning data to give more precise prediction which fits to the original leaning data.

Learning data scaling has the effect to extend and adjust the distribution of tree's prediction so that the mean value of it locates closer to the observation value. Figure 2 shows the example of tree's prediction histogram. In case of the model by original learning data, wrong predictions exist in one side and pull the mean value away from the observation value. In case of the model by scaled learning data, the distribution of tree's prediction is magnified so that the mean value of the distribution matches to the observation value. This is the case observed in the large plus side of S2S gap prediction range. The same phenomenon is observed in the large minus side of prediction range.

2. 3 Model accuracy improvement

By applying these two methods, the model fitness to the learning data is improved as shown in Figure 3. The model improvement is summarized in table 1. The slope between observed S2S gap (=learning data) and predicted S2S gap is improved to ~1 at the final model. The prediction error $\Sigma(residue)^2$, which is sum of squared gap between observation and prediction, is reduced to ~1/4 times by random design attributes count, another ~1/6 times by learning data scaling, totally ~1/25 times compared with conventional model.

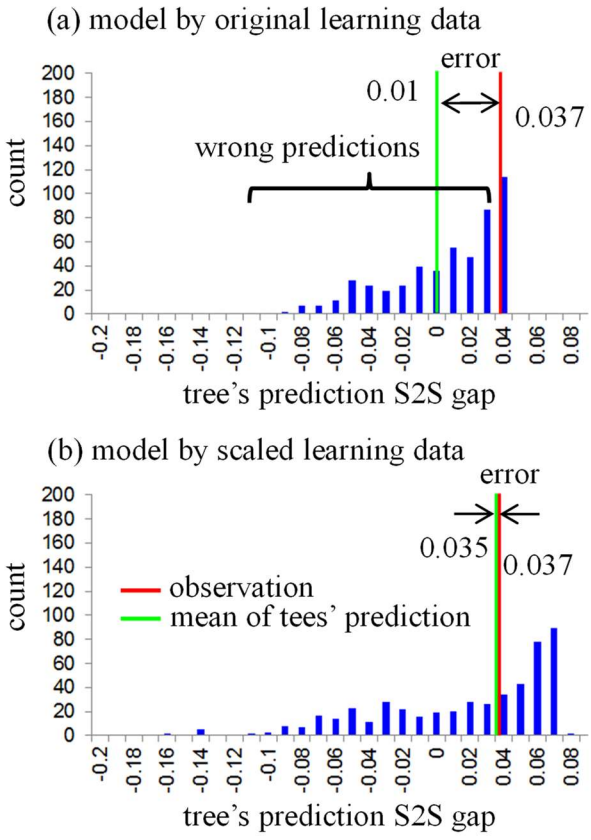

Figure 2. Histogmam of tree's prediction. Example of 1,000 trees to predict in the large plus side of prediction range.

The tree count in model also affects the model accuracy. From modeling exercise for various cases, the prediction error $\Sigma(residue)^2$ reduces rapidly with increasing tree count and saturates at and over ~200 tree count. In this paper, 1,000 trees were built to model S2S gap.

3. Design Attribute Identification

The above explained method was applied to model S2S gap to find the key layout structures which lead to large S2S gap. Our primary interest is the S2S gap in large minus direction, which means smaller drain current than expected by SPICE simulation. Indices '*importance*' and '*impact*' are extracted from the obtained precise model. And significant design attributes can be identified.

3. 1 Significant design attributes

Index '*importance*' is extracted for each design attribute as the sum of variance reduction at each node split, which is defined by the normalized formula (8).

$$\text{'}importance\text{'} \equiv \frac{\sum_{att} dS(att)}{\sum_{All} dS(All)} \quad (8)$$

where dS is variance reduction given by equation (2), Σatt means sum by nodes at which a particular design attribute att is used to split S2S gap data into next nodes, ΣAll means sum by all nodes in the model except for terminal nodes.

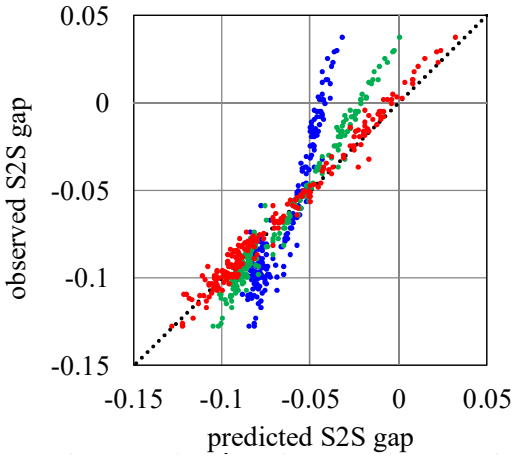

Figure 3. Observation vs. prediction.
● bule: conventional model,
● green: premodel by random design atttributes count,
● red: final model by random design atttributes count and scaled learning data, ••• dotted line: ideal line of slope=1

Table 1. Model acccuracy improvement summary. Total design atributes used to build model n=14, generated tree count =1,000 , prediction data count =211

	conventional model	pre-model	final model
dsign attributes count	fixed at $\sqrt{n}$	random $\sqrt{n} \sim n$	random $\sqrt{n} \sim n$
learning data	original data	orioginal data	scaled data
$\Sigma(residue)^2$	0.1364	0.0351	0.0057
slope	2.35	1.41	1.02

The larger value of '*importance*' means the larger contribution to determine S2S gap in the model.

Index '*impact*' is the newly defined index. It is extracted for each design attribute and its value as the sum of mean value shift in the path from one node to the next left node. It is defined by the normalized formula (9).

$$\text{'}impact\text{'} \equiv \frac{\sum_{att=x} dM(att=x)}{\sum_{All} |\, dM(All)\,|} \quad (9)$$

$$dM \equiv Ml - Mm \quad (10)$$

where is dM, Ml, Mm are mean value shift, mean value of the next left node, mean value of the node before split respectively, and $\Sigma att=x$ means sum by nodes at which a particular design attribute att and its value x is used to split S2S gap data into next nodes, ΣAll means same as the definition in equation (8). Larger minus or larger plus value of '*impact*' should mean the more power to drive S2S gap to minus or plus direction respectively.

Figure 4 shows bar chart of '*importance*' extracted from improved random forest model. The result is summarized for NMOS transistors. Top 4 design attributes, vtType, fin_edge, fin_LR, and nfin with large *'importance'* were selected to show their '*impact*' in Figure 5. The design attributes showing large minus '*impact*' are identified as;
(A) vtType (threshold voltage type) = LVT
(B) fin_edge (gate to active edge in an arbitrary unit) ≤ 32
(C) fin_LR (fin count ratio at adjacent gate) = 1:3, 1:2, 2:3
(D) nfin (fin count) = 3

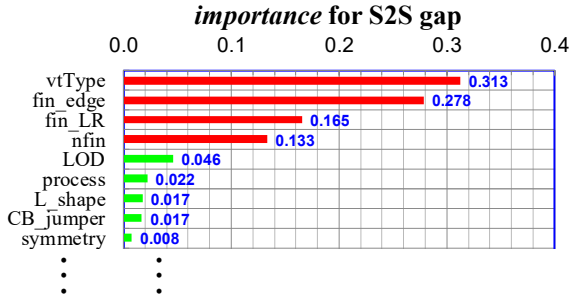

Figure 4. Design attributes in '*importance*' order.

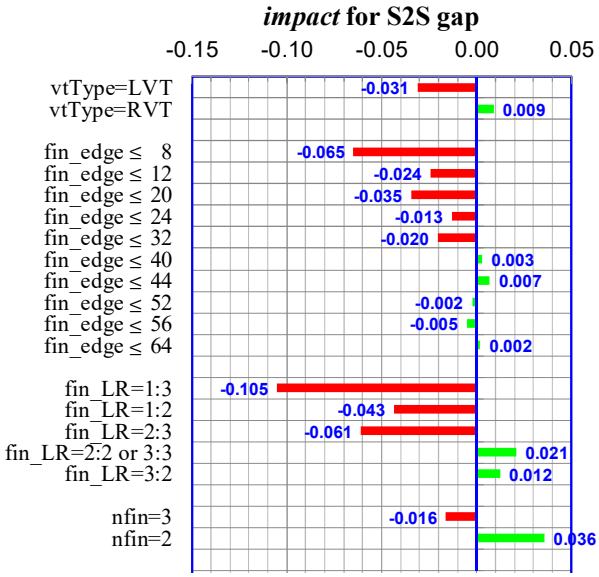

Figure 5. Design attributes showing minus '*impact*' (red bar) are supposed to drive S2S gap to minus direction.

These are supposed to be the significant design attributes to cause large minus S2S gap, i.e. smaller drain current than that of SPICE model.

3. 2 Design attributes dependency

Figure 6 and Figure 7 show S2S gap trend classified by the identified design attributes with plotting the observation data and prediction result. The model has enough accuracy and resolution to reproduce observed dependency. And S2S gap is clearly classified by the identified design attributes. As '*impact*' indicates, S2S gap was verified to show clear trend of larger minus S2S gap at following conditions;
at vtType = LVT rather than at vtType = RVT
at nfin = 3 rather than at nfin = 2
at fin_LR = 1:3, 1:2, 2:3 rather than at 2:2, 3:3, 3:2
at fin_edge ≤ 32 rather than at fin_edge=40 ~ 64.

3. 3 Representative structure

From the significant design attributes (A), (B), (C) and (D), two representative structures were identified as the layout feature to satisfy these design attribute conditions as shown in Table 2. 125 transistors among 10,080 were identified to have these structures.

S2S gap of the representative structure surely distributes in the large minus area of all DUT data as shown in Figure 8. The mean S2S gap of it is −0.106 while that of all DUT data is −0.0495. The difference is tested to be significant as the p-value to observe such mean value (−0.106) is evaluated to be 0 from probability density curve for mean

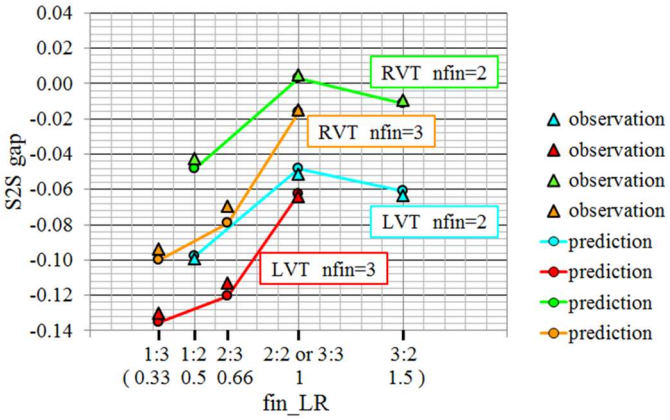

Figure 6. S2S gap trend by the identified design attributes.

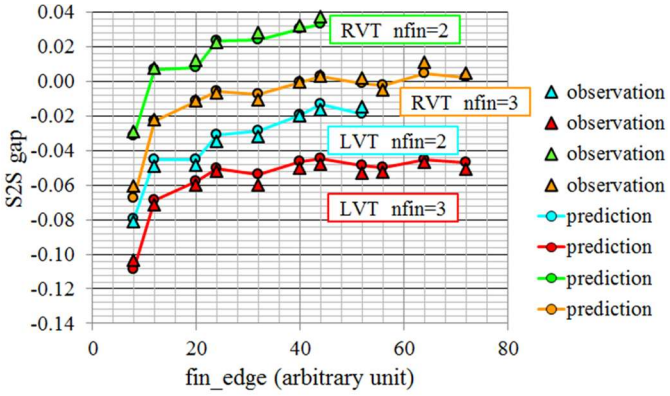

Figure 7. S2S gap trend by the identified design attributes.

S2S gap shown in Figure 8. The probability density was calculated by accumulating mean S2S gap of ramdomly sampled data from all DUT data with the sampling size same to the representative sturucture data count.

4. Conclusion

Random forest model was applied to design attributes analysis on S2S gap.

(1) The model accuracy was improved by random design attributes count in tree building and learning data scaling.
(2) The indices such as '*importance*' and '*impact*' extracted from the improved model are effective to identity significant design attributes for S2S gap.
(3) The key FinFET structures to cause large minus S2S gap were identified by the model analysis.

References

[1] Breiman, L., "Random Forests", *Machine Learning*, 45, pp.5-32, 2001
[2] Breiman, Friedman, Olshen and Stone, *Classification and Regression Trees*, Wadsworth, 1984.

Table 2. Representative structure for large minus S2S gap.

	vtType	fin_edge	Fin_LR	nfin
structure1	LVT	8	1:3	3
structure2	LVT	8	2:3	3

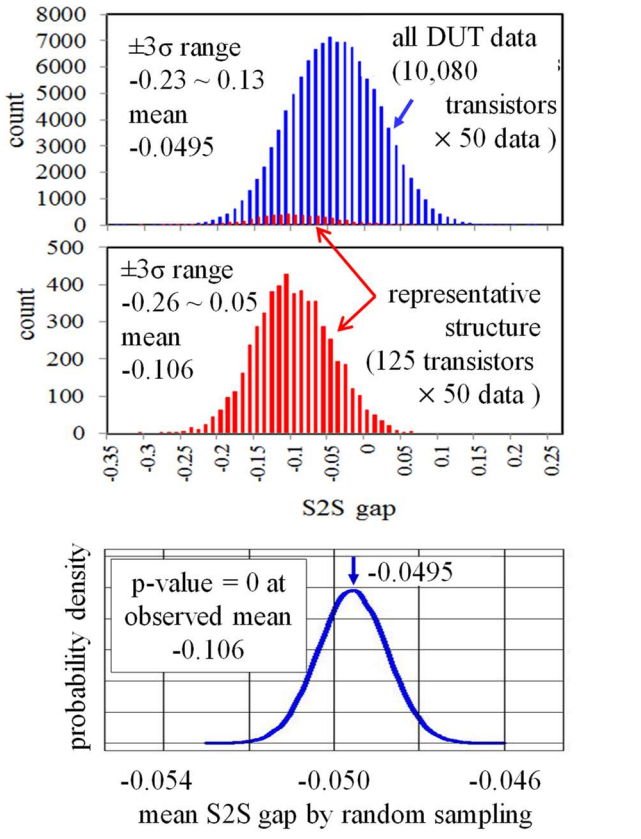

Figure 8. S2S gap distribution of the representative sutructure in all DUT data. Probability density was calculated by 100,000 random sampling from all DUT data.