Wafer to Lot-level S-parameter Prediction in Radio Frequency Testing Using Radial Basis Function Neural Network

Huimin Wang Yasuhiko Iguchi Chika Tanaka

Memory Division, Kioxia Corporation
2-5-1, Kasama, Sakae-ku, Yokohama 247-8585, Japan
e-mail : huimin1.wang@kioxia.com

Abstract – To reduce the production testing costs without sacrificing quality, a wafer to lot-level performance prediction has gained traction as a key enabler for production tests. Although many effective prediction methods for physical properties have been proposed, there are a few methods that can predict S-parameters accurately. In this study, we propose a novel S-parameter prediction method based Radial Basis Function neural network to make the accurate predictions using a minimal data set.

I. Introduction

Recently, the reduction of wafer testing costs without sacrificing quality has been required in the manufacturing of large-scale integrated circuits (LSIs). An increasing number of researchers has been focusing on the direct prediction of physical parameters, aiming to bypass the complex procedures of traditional numerical simulations and enhance modeling efficiency and predictive accuracy. To solve this problem, machine learning methods such as Gaussian process [1] have been proposed due to improvements in computational power and innovations in algorithms. Although existing methods for predicting wafer-level physical properties have achieved satisfactory results, the prediction of S-parameters obtained from radio frequency (RF) measurements remains challenging. Fig. 1 illustrates the definition of S-parameters for two-port network. Since the digit precision in the numerical difference of S-parameters affects the prediction results, ordinary machine learning models cannot accurately predict the characteristics of the datasets [2, 3, 4].

Moreover, when S-parameters are used to analyze circuit characteristics, we need to handle large-scale samples because S-parameters are typically obtained through frequency scanning. As a result, the time and computational costs associated with processing large-scale data increase, making it challenging to obtain the predicted S-parameters.

Considering these issues, we propose a novel RF testing technique using a Radial Basis Function (RBF) neural network [5] to predict the S-parameters.

The RBF neural network model demonstrates exceptional capabilities in localized feature extraction and nonlinear modeling. It is currently widely used in various fields, such as financial data prediction and physical property prediction. It effectively mitigates the issue of traditional models that are difficult to learn characteristics accurately due to small differences in numerical accuracy and the complexity of learning patterns in large datasets. By applying the RBF neural network model, it is expected that the number of measurement devices under test (DUT) on a wafer will be reduced, resulting in cost savings.

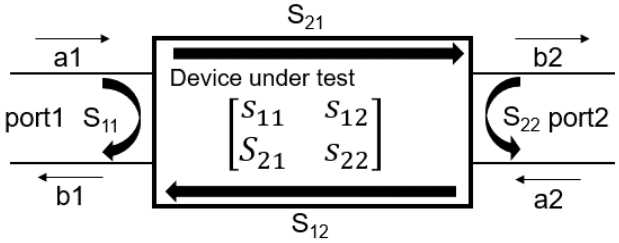

Fig.1. S-parameters for two-port network

This paper is organized as follows: Section II introduces the specific details of the RBF neural network model within our testing strategy, as well as the experimental setup and evaluation methods. Section III presents the results of the experiments, while Section IV provides a summary and conclusions.

II. Experimental Methodology

A. *Data preparation*

The dataset used in this work was obtained through on-wafer RF measurements of fabricated devices. Each measurement consists of the following parameters: bias voltages, chip position, and measurement frequency, along with the corresponding S-parameters. The frequency range spans from 200 MHz to 20 GHz, sampled at 100 points per voltage condition.

The dataset is structured into the input features and targets depending on the prediction task.

The input vector X includes:

$$X = [\text{wafer\_number}, x, y, V_d, V_g, \text{frequency}]^T \in \mathbb{R}^d \quad (1)$$

where x and y represent the chip coordinates on the wafer, and V_d, V_g are the drain and gate voltages, respectively.

The target vector Y corresponds to the complex S-parameters at a given frequency, represented as:

$$Y = [Re(s11), Im(s11), Re(s12), Im(s12), Re(s21), Im(s21), Re(s22), Im(s22)] \quad (2)$$

Min-Max normalization was applied to the data to improve the model's ability to learn relationships between vectors.

B. *Experimental setup*

In this work, we describe the experimental setup used to

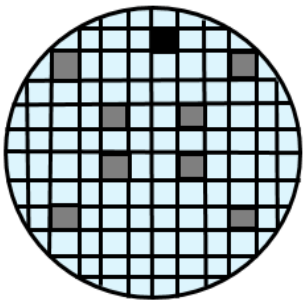

Fig. 2. Examples of the 8 selected chips

evaluate the model performance across wafer-level and lot-level.

For wafer-level prediction, 8 chips were selected from a single wafer as training data. According to related work [1], the chosen chips can have a slight impact on the prediction results; therefore, we try to choose the points that are evenly distributed on the wafer surface. Examples of the 8 selected chips are shown in fig. 2, where the gray chips indicate those with the training data. The light blue chips represent the prediction targets, and the black chip is presented as an example of a prediction result.

For lot-level prediction using multiple wafers, the same approach was taken to select 8 chips on a single wafer, as shown in fig. 2. Subsequently, the same 8 chips were consistently chosen from 3 wafers as the training data. This approach allows us to determine whether the neural network can accurately capture the relationship between the wafers. We selected wafer numbered 2, 5 and 10 (Wafer02, Wafer05, Wafer10) to predict wafer number 8 (Wafer08).

C. *Structure of RBF neural network*

The standard RBF neural network has a three-layer forward structure: the input layer, the hidden layer and the output layer. The RBF neural network employs a Gaussian kernel activation as the first hidden layer, followed by two fully connected layers with Rectified Linear Unit (ReLU) activations. The final layer produces the predicted S-parameters via a linear transformation.

The proposed model consists of an input layer with 6 neurons corresponding to spatial, electrical, and frequency features as shown in fig.3. We consider $(X_{\text{train}}, Y_{\text{train}}) = \{(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), \ldots, (X_n, Y_n)\}$ as the training datasets and $X_{\text{test}} = (X_1^*, X_2^*, X_3^*, \ldots, X_m^*)$ as the testing datasets, where $m >> n$.

The RBF layer is as follow:

$$\emptyset_i(X) = exp\left(-\frac{\|X-c_i\|^2}{2\sigma_i^2}\right),\ i = 1, 2, 3, \ldots, n \qquad (3)$$

$$h^{(1)} = [\emptyset_1(X), \emptyset_2(X), \emptyset_3(X), \ldots, \emptyset_n(X)]^T \epsilon \mathbb{R}^n \qquad (4)$$

where $c_i \in \mathbb{R}^d$, which represents the center of the i-th kernel. σ_i is defined as the average Euclidean distance between the i-th center and its neighboring centers. The output of RBF layer is given as follows in Eq. (4).

The proposed model, the centers of the radial basis functions are defined as trainable parameters. Specifically, the centers

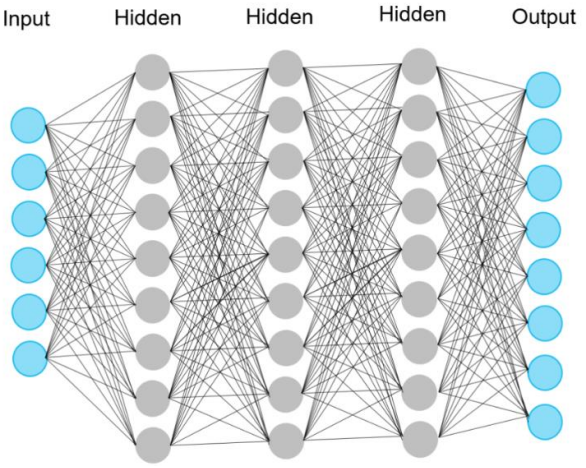

Fig. 3. RBF neural network in this study

are initialized randomly and updated through backpropagation during training, enable the network to adaptively learn optimal kernel locations based on the input distribution. And the proposed model was trained using the mean squared error (MSE) loss function to minimize the squared difference between predicted and true values. Parameter optimizer was performed using the Adam optimizer, an adaptive learning rate method that computes individual learning rates for each parameter based on estimates of first and second moments of the gradients. This approach facilitates fast and stable convergence during training.

D. *Evaluation methods*

In this paper, the Mean Absolute Percentage Error (MAPE) was chosen as the evaluation method for errors. MAPE is a measure of prediction accuracy used in statistical forecasting. It typically expresses accuracy as a ratio defined by the following formula:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{Y_{true}-Y_{pre}}{Y_{true}}\right\| \qquad (5)$$

where Y_{true} is the actual value, and Y_{pre} is the prediction value. The difference between the actual value and the predicted value is divided by the actual value. Then, the absolute value of this ratio is summed for every predicted point and then divided by the number of fitted points n.

III. Results and Discussion

In this chapter, we focus on summarizing our experimental results. Under the test conditions of 85 °C, at the same device name with V_d = 1.2V and V_g = 1.2V in the frequency range of 200M ~ 2G Hz, two chips were selected for the presentation of results: one is $(x, y) = (0, 0)$ and the other is $(x, y) = (1, 3)$.

The discussion is organized into two sections: one is wafer-level prediction results, and the other is lot-level prediction results. For the wafer-level case, we additionally compared our method with previous approaches.

A. *Wafer-level prediction results*

Fig.4. shows a comparison of the S-parameters prediction

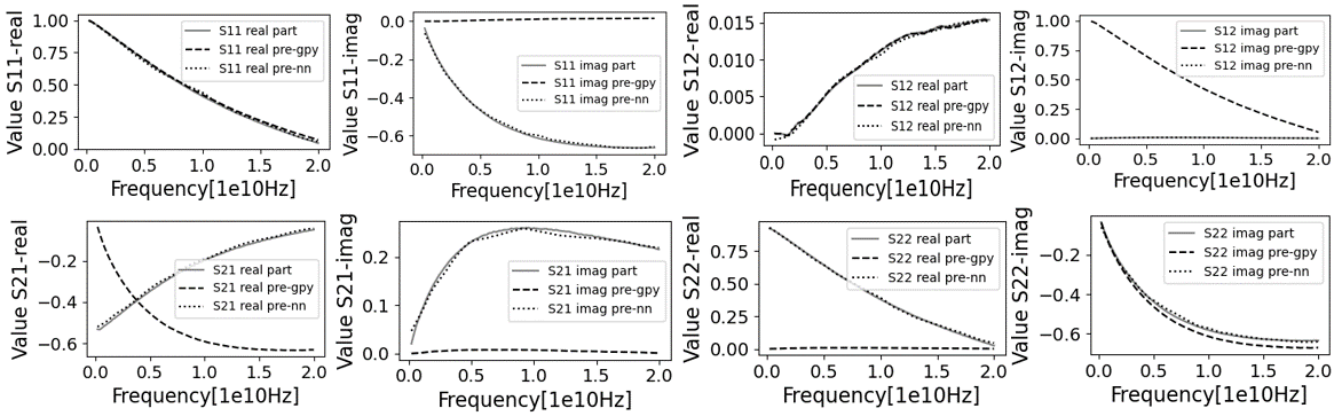

Fig. 4. S-parameters prediction results for wafer-level test on Wafer02 between the proposed RBF neural network model and the GPy model. The dot-dashed line corresponds to the RBF model's prediction, the dotted line indicates the prediction result from the GPy model, and the gray solid line represents the actual measured values.

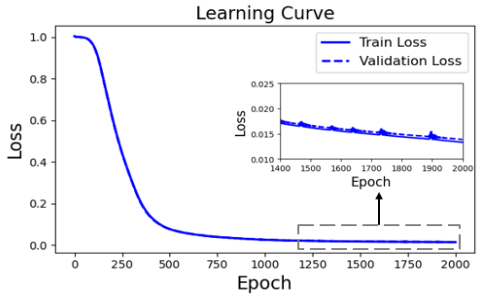

Fig. 5. Training and validation loss curves

results for Wafer02 between our proposed model and the Gaussian process method (GPy) described earlier. As shown in the fig. 4, the predicted results from the proposed method for S-parameters S11, S12, S21 and S22 fit well with the corresponding measured S-parameters at chip (0, 0).

From the results, GPy performs well on certain outputs but shows inconsistent performance across the entire output set. This indicates that GPy struggles to effectively capture the relationship between inputs and outputs. A contributing factor is the large variability or disparity within the input data, which poses challenges for the GPy model. Even after normalization, GPy model is still unable to learn the underlying correlations effectively, likely because such heterogeneity in the data distribution hinders the Gaussian process ability to generalize.

In contrast, compared with GPy model, the RBF neural network model achieved better results under the same input conditions.

While, to verify the training stability and generalization performance of the proposed model, we monitored the loss values on both the training and validation datasets during the training process.

As shown in fig.5, the solid line represents the loss on the training set, while the dashed line represents the loss on the validation set. The RBF neural network model was optimized using the Adam algorithm with the MSE loss function. The smooth and consistent convergence of both curves indicates stable learning and good generalization performance.

Therefore, our proposed method demonstrates a good

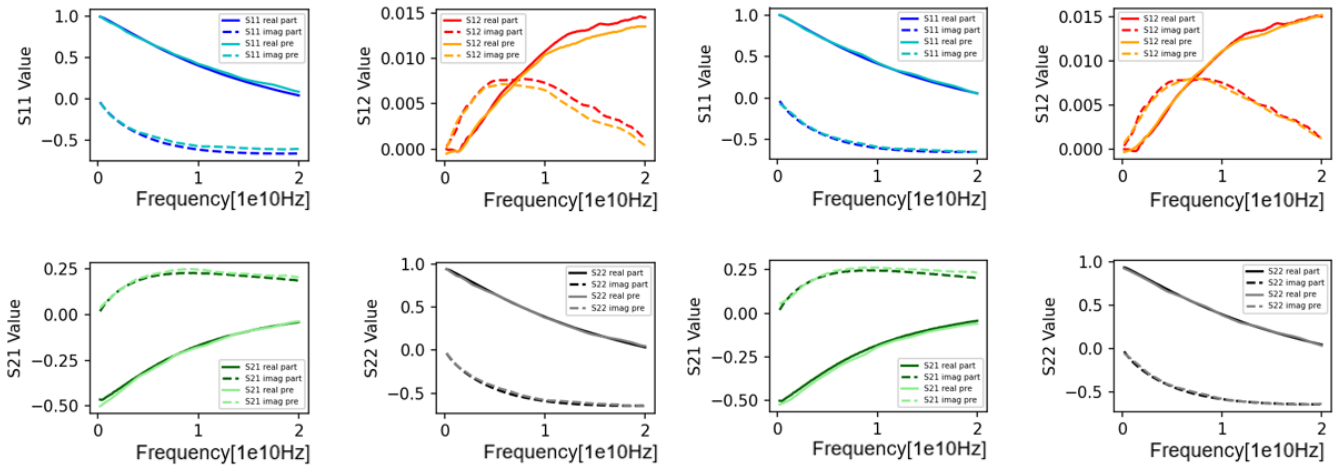

Fig. 6. S-parameters prediction results for lot-level test on Wafer08 of (a) $(x, y) = (0, 0)$ and (b) $(x, y) = (1, 3)$. The solid lines represent the real parts of S-parameters, and the dashed lines represent the imaginary parts of S-parameters. For both (a) and (b), darker colors and lighter colors represent the measured data and the predicted results, respectively.

TABLE I
MAPE between Predicted and Measurement Values

	Wafer-level		Lot-level	
Wafer number	Wafer02		Wafer08	
Chip	(0,0)	(1,3)	(0,0)	(1,3)
MAPE_s11_real	0.03%	0.04%	0.15%	0.05%
MAPE_s11_imag	0.02%	0.05%	0.07%	0.04%
MAPE_s12_real	0.45%	0.42%	0.59%	0.39%
MAPE_s12_imag	0.04%	0.10%	0.18%	0.07%
MAPE_s21_real	0.05%	0.09%	0.04%	0.11%
MAPE_s21_imag	0.04%	0.08%	0.07%	0.10%
MAPE_s22_real	0.05%	0.05%	0.05%	0.04%
MAPE_s22_imag	0.02%	0.05%	0.03%	0.03%

performance at the wafer-level predictions.

B. Lot-level prediction results

The RBF neural network demonstrated effective prediction of lot-level S-parameters as shown in fig.6.

In the lot-level experiment, only a limited number of data points from three wafers (Wafer02, Wafer05 and Wafer10) were used to train the model. Despite this limitation, the model accurately predicted the characteristics of Wafer08, which was excluded from the training set. The close agreement between the predicted and measured results suggests that the RBF neural network exhibits robust performance and generalization ability across different wafers within the same lot.

C. Evaluation and discussion

In this section, we summarize the experimental results and compute the MAPE between the predicted and measurement values. The calculation results are presented in Table I. The experimental evaluation results show that the MAPE of the proposed RBF neural network is less than 1% in both wafer-level and lot-level prediction tasks. The sub-1% error means the RBF neural network effectively captures the complex relationship between input characteristics such as voltage settings and spatial coordinates, and the real and imaginary components of the S-parameters.

However, this study was limited to a single manufacturing lot and four wafers. While preliminary results demonstrate promising accuracy, the model's robustness remains uncertain. To fully validate the model, future work should focus on the following areas:

1. Dataset Expansion: Future studies should evaluate model's performance under different process variations and environmental factors. This approach will help reduce the risk of overfitting.
2. Algorithm Innovation: We will also explore alternative approaches that have the potential to significantly improve prediction accuracy.

IV. Conclusions

In this paper, we proposed a novel Radial Basis Function neural network model for predicting S-parameters, achieving a Mean Absolute Percentage Error of less than 1% in both wafer-level and lot-level predictions. The Radial Basis Function neural network has been validated to demonstrate its capability in predicting S-parameters. Our proposed method effectively addresses the limitations of conventional approaches that fail to capture high-dimensional characteristics.

In future studies, we plan to further optimize and refine the model architecture. Additionally, we will investigate the potential application of this methodology in industrial production processes, with the aim of reducing testing costs while maintaining measurement accuracy.

References

[1] Shintani, Michihiro, *et al*., "Wader-level Variation Modeling for Multi-site RF IC Testing via Hierarchical Gaussian Process." *2021 IEEE International Test Conference (ITC)*. IEEE, pp. 103–112, 2021.

[2] Yuqing, Sun, Qiao Junfei, and Han Honggui. "Structure design for RBF neural network based on improved K-means algorithm." *2016 Chinese Control and Decision Conference (CCDC)*. IEEE, pp.7035-7040, 2016.

[3] Torun, Hakki Mert, *et al.*, "Causal and passive parameterization of S-parameters using neural networks." *IEEE Transactions on Microwave Theory and Techniques* 68.10: 4290-4304, 2020.

[4] Bansal, Vinayak, *et al*., "Parametric S-Parameter Prediction Using Deep Learning." *2024 IEEE 33rd Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, pp.1-3,2024.

[5] He, Jia, and Hongyang Liu. "The application of dynamic K-means clustering algorithm in the center selection of RBF neural networks." *2009 Third International Conference on Genetic and Evolutionary Computing.* IEEE, pp.488-491, 2009.