Missouri University of Science and Technology

Scholars' Mine

Electrical and Computer Engineering Faculty Research & Creative Works

Electrical and Computer Engineering

01 Jun 2022

Training Set Optimization in an Artificial Neural Network Constructed for High Bandwidth Interconnects Design

Bo Pu

Heegon Kim

Xiao Ding Cai

Bidyut Sen

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/ele_comeng_facwork/4645

Follow this and additional works at: https://scholarsmine.mst.edu/ele_comeng_facwork

 Part of the Electrical and Computer Engineering Commons

Recommended Citation

B. Pu et al., "Training Set Optimization in an Artificial Neural Network Constructed for High Bandwidth Interconnects Design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 6, pp. 2955 - 2964, Institute of Electrical and Electronics Engineers, Jun 2022.
The definitive version is available at https://doi.org/10.1109/TMTT.2022.3162209

This Article - Journal is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

Training Set Optimization in an Artificial Neural Network Constructed for High Bandwidth Interconnects Design

Bo Pu, *Senior Member, IEEE*, Heegon Kim, *Member, IEEE*, Xiao-Ding Cai, *Member, IEEE*, Bidyut Sen, Chunchun Sui, *Member, IEEE*, and Jun Fan, *Fellow, IEEE*

***Abstract*— In this article, a novel training set optimization method in an artificial neural network (ANN) constructed for high bandwidth interconnects design is proposed based on rigorous probability analysis. In general, the accuracy of an ANN is enhanced by increasing training set size. However, generating large training sets is inevitably time-consuming and resource-demanding, and sometimes even impossible due to limited prototypes or measurement scenarios. Especially, when the number of channels in required design are huge such as graphics double data rate (GDDR) memory and high bandwidth memory (HBM). Therefore, optimizing the training set selection process is crucial to minimizing the training datasets for developing an efficient ANN. According to rigorous mathematical analysis of the uniformity of the training data by probability distribution function, optimization flow of the range selection is proposed to improve accuracy and efficiency. The optimal number of training data samples is further determined by studying the prediction error rates. The performance of the proposed method in terms of accuracy is validated by comparing the scattering parameters of arbitrarily chosen strip and microstrip type GDDR interconnects obtained from EM simulations with those predicted by ANNs using default and the proposed training-set selection methods.**

***Index Terms*— Artificial neural network (ANN), design-of-experiment (DoE), high bandwidth interconnect, microwave modeling, optimization, probability analysis, uniform distribution.**

I. Introduction

ELECTROMAGNETIC (EM) modeling is regarded as an effective way to characterize the microwave components. Several methods based on computational EM have been developed for an accurate characterization, such as finite-difference time-domain (FDTD) [1], finite element method (FEM) [2], method of moment (MoM) [3], and partial element equivalent circuit (PEEC) [4], etc. Those methods have been widely implemented in modeling microstrip lines, stripline, and coplanar structures [5]–[10]. The above-mentioned methods are all grid-based or cell-based techniques. A meshless modeling approach was proposed to avoid the very complicated and even time-consuming demerit caused by the practical structure with complex geometries [11]. In general, conventional numerical methods ensure accurate modeling by analyzing the inherent characteristics of microwave structures. Comprehensive understanding of complicated microwave theory and mathematics is the prerequisite to make use of those methods. To bring a convenience to a practical engineering level, complicated numerical methods have been realized by commercial simulation tools, and users do not need to face the original numerical challenges in their realistic usage. However, simulator-based approaches embody the following disadvantages: time-consuming, costly license fees, and expensive hardware resources [12].

In recent years, artificial neural networks (ANNs) have attracted attentions for modeling and design optimization of microwave components since the mechanism of ANNs can handle new designs by establishing a training network on the primitively known structures, comprehended in the early designs. Previous works have enabled fast and accurate modeling of microwave components with the neural network and machine learning approaches. A method, which can build neural models automatically by adding samples and neurons as needed, was proposed in [13]. An advanced algorithm for automated parametric modeling of microwave components using combined neural networks and interpolation techniques was introduced in [14]. Tremendous speedup over conventional methods was realized by ANNs for the modeling of electronic device and circuits [15]. Electromagnetically trained ANN (EM-ANN) was first introduced in [16] for accurate and efficient modeling of microstrip vias and interconnects in monolithic microwave/millimeter-wave integrated circuit (MMIC). An approach to map the complex relationship between the physical and electrical parameters of interconnect structures in an efficient manner by ANNs was presented in [17]. To develop libraries of neural models for passive and active components, a hierarchical neural network approach was proposed in [18]. With the development of the technology, the functions of neural networks have got enhanced and been widely used to handle the challenges in modeling of 3-D substructures [19], complex high-speed channel with equalization [20], and updated with multidimensional extrapolation techniques [21]. Neural networks have also been implemented

Manuscript received October 26, 2021; revised December 21, 2021 and February 11, 2022; accepted March 3, 2022. Date of publication April 11, 2022; date of current version June 3, 2022. *(Corresponding author: Jun Fan.)*

Bo Pu is with DeTooLIC Technology, Ningbo, Zhejiang 315899, China (e-mail: bobpu@ieee.org).

Heegon Kim and Jun Fan are with EMC Laboratory, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: jfan@mst.edu).

Xiao-Ding Cai and Bidyut Sen are with Cisco Systems, San Jose, CA 95035 USA.

Chunchun Sui is with Tesla, Inc., Palo Alto, CA 94304 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMTT.2022.3162209.

Digital Object Identifier 10.1109/TMTT.2022.3162209

0018-9480 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

into the crosstalk prediction, sensitivity analysis, performance assessment and signal/power integrity design for high speed and microwave circuits [22]–[25].

Above mentioned frontier researches profoundly promote the development of modeling technology toward an efficient direction. However, in previous publications on ANN-related applications in microwave and high-speed areas, only the implementation and enhancement of ANN itself has been discussed. The accuracy of the ANN method is significantly affected by training data size, but so far the selection of training data in a rigorous probability analysis has yet to be investigated. Using a large number of training data samples can enhance the general accuracy of an ANN. However, this approach also increases the overall effort because more training data must be generated. Especially, in graphics double data rate (GDDR) memory and high bandwidth memory (HBM) which solve the ever-increasing bandwidth issue by enlarging the number of channels [26], [27], maintaining a limited number of training data samples for all channels is desirable to save time while still achieving high accuracy. Hence, a better method for the selection of training datasets is of utmost importance.

The design-of-experiment (DoE) method can be used to efficiently select training data [28]. Any DoE method generates a distribution of plausible collections of parameter values from a multidimensional distribution. It enables rapid and precise ANN model construction by using a small number of training data, because the method spreads the sample points more evenly across all possible values, on the basis of the original multidimensional distribution. Many studies have constructed ANN models of transmission line structures based on training data generated by a DoE method. However, as described earlier, the training data samples generated by a DoE method can significantly affect the accuracy of the constructed ANN model, and the DoE method setup necessary to optimize efficiency has not been carefully examined in prior studies. In our previous study [29], the inputs used in the DoE method were physical dimensions, such as metal width, spacing, and dielectric height. The outputs, which are resistance, inductance, conductance, and capacitance (*RLGC*) matrices, can be first extracted from EM simulations on the basis of the model established by physical dimensions and material information, and used as the criteria in the training process. Moreover, *RLGC* matrices are able be converted to corresponding scattering parameter (S-parameter) by circuit simulators for a quick check of the signal quality.

Although the metal width and spacing parameters in the training datasets were uniformly selected by the DoE method, the ratio of the metal width to spacing, an important factor for mutual *RLGC* components, was not uniformly distributed. The consequence of this issue was a larger prediction error at the boundary of the ranges (i.e., for an extreme channel structure, such as a channel with a minimum width and maximum spacing), which would result from the lack of training data.

In this article, a training set optimization is proposed to overcome the accuracy issue caused by nonuniform mutual parameters. The proposed optimization approach for training dataset enhances the efficiency of the ANN with the least computation resources, easing the knowledge requirement for users and designers, compared to traditional numerical solutions [1]–[4], and achieves higher accuracy than that of the traditional DoE setup by using the same number of training data samples [29]. The rigorous analysis process introduced herein is a general dataset optimization approach to provide an efficient training for ANN. Thus, the merit requiring less training data samples can be implemented into modeling of any arbitrary, complex, multiconductor transmission line systems, and potentially capable to deal with inhomogeneous, non-isotropic, multilayered structures with less difficulty to obtain sufficient simulated or measured data.

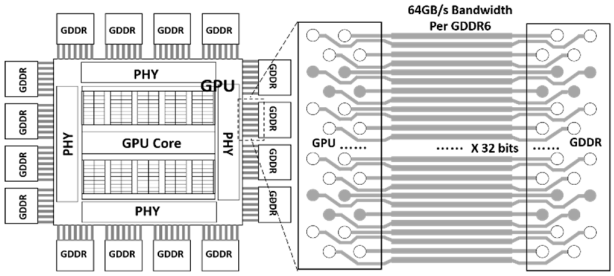


Fig. 1. Structure of GPU-GDDR6 system with high bandwidth (32 bits)/speed (16 Gbps/channel) interconnects.

Section II introduces an ANN construction procedure to predict the *RLGC* matrices of a typical target interconnects structure in GDDR memory. In Section III, various DoE method setups for improved training dataset generation are discussed, and an optimal DoE setup is proposed based on uniformity probability investigation and error analysis. The performance of the proposed method is validated by S-parameter simulations performed with arbitrarily selected parameters of strip lines and microstrip channels in Sections IV and V, respectively.

II. ANN Construction Procedure for Interconnects Modeling in GDDR6 Memory

32 bits high-speed channels between GDDR6 and graphics processing unit (GPU) enables the data communication in a fast speed and high bandwidth as illustrated in Fig. 1. An electrical characterization of channels is usually demanded for the evaluation of performance. However, large number of lines for high bandwidth make the *RLGC* extraction based on full wave simulation not efficient. Since the structure of lines are designed with common characteristics, ANN is used here to provide a fast but accurate prediction of *RLGC* for channels on the basis of geometrical parameters, especially in the early design stage when the layout and model for simulation have not prepared yet. Those channels are designed as stripline and microstrip line types depending on specific functions as shown in Fig. 2. Geometrical parameters of one differential pair for clock and other single-ended for data are linewidth (W), intrapair spacing (S), and interpair spacing (S_P). The heights of the prepreg and core layers are H_P and H_C, respectively. Dielectric material is often predefined, geometrical parameters here as W, S, S_P, H_P, and H_C are

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RANGES OF DESIGN PARAMETER FOR TRAINING SET OPTIMIZATION (STRIP LINE)

Parameter	Min	Max	Parameter	Min	Max
W	4 mil	8 mil	H_P	3.5 mil	15 mil
S	5 mil	10 mil	H_C	3.5 mil	15 mil
W/S	0.2	1.0	S_P	5 mil	50 mil

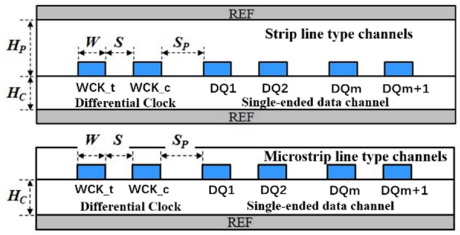


Fig. 2. Target channel for modeling: Strip lines and microstrip lines between GPU and GDDR6. Design parameters are W, S, S_P, H_P, and H_C.

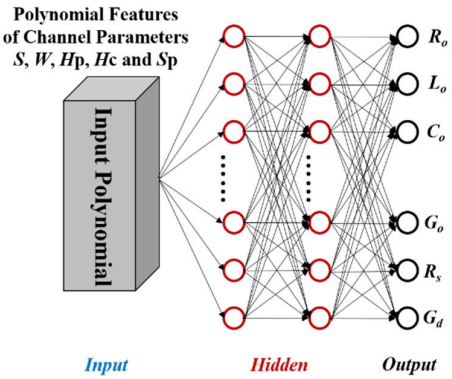


Fig. 3. ANN model for channel RLGC components. The polynomial features of design parameters and the corresponding channel *RLGC* matrices are the inputs and outputs of the established ANN model.

the design parameters of the target channel in the training process. Table I lists ranges of the design parameters for optimization, which is used in [29] and defined as default case here. The conductor thickness is fixed at 0.6 mil.

The ANN model constructed for high speed and bandwidth channel design is illustrated in Fig. 3. The design parameters of physical dimensions are the input parameters of the ANN model, for example, S, W, S_p, H_p, and H_c. The total number of the output nodes is 30 as mentioned in our previous research, which is the reference [29]. The corresponding per-unit-length *RLGC* matrices of the W-element model, which consists of six frequency-independent *RLGC* (Lo, Co, Go, Ro, Gd, and Rs) matrices, are the ANN outputs. Each element, such as Ro, includes five terms (11, 12, 13, 23, and 14). Because of symmetry, reciprocity, and negligible crosstalk between the leftmost and right-most pairs, the 36 terms in each matrix can be replaced by only five terms mentioned, which were also explained in detail in [29]. Two hidden layers containing 40 hidden neurons are used for the ANN model. The degree of input polynomial and regularization factor are 2 and 0.003, respectively. Five channel parameters were transformed into their polynomial features with degree of 2 as the input of ANN, which is a more suitable format for efficient training [30]. We would like to mention here that the ANN itself is not the main content of this manuscript while the optimization of the training datasets is the most significant contribution.

The mutual terms in the extracted *RLGC* components vary exponentially with S and S_P. To achieve a nearly linear relationship between the inputs and outputs of the training datasets, the data values of the training set outputs are converted by logarithmic functions. This post-processing step makes the ANN model sufficiently precise despite the use of a small number of training data samples.

Although the polynomial chaos expansion [31] and support vector machine methods have been widely used for optimization, their merits relate to variability analysis and process regression. An ANN computes in an interpolative manner. Uniformly scattered training samples, therefore, provide the network with better information to make comprehensive and efficient predictions [32]. As explained in reference [33], the uniformity helps the gradient descent converge more quickly, because it makes more uniform steps through the feasibility space of the error function. In addition, in some cases, it helps zero-center the data, thus preventing zigzag behavior in gradient descent algorithm optimization.

Approaches to achieved uniform samplings have been studies in many publications. Among them, adaptive rejection samplings, low-discrepancy sequences (LDS), and Latin-hypercube sampling (LHS) method [34] have very good performance. Adaptive rejection sampling seems to be the best candidate with superior performance, while it requires a prerequisite that the probability density function (PDF) should satisfy log-concave, and this limits its application range. LDS such as Sobol's or Halton's sampling methods can also generate uniform distribution as well. However, LDS performs well for solving high dimensional problem and the LHS has proven to be very effective in application to metamodeling where a not huge number of sample points are required, and a sampling design of such points can be efficiently optimized [35]. Therefore, to ultimately realize an efficient optimization for reduced training data size, LHS is used herein to generate the initial level uniform data for each geometrical parameter in its optimal range determined by proposed rigorous probability analysis. The given lower and upper bounds for the range are shown in Table I, and the training dataset inputs were selected. The corresponding *RLGC* values, which are the training dataset outputs, were extracted by EM simulations. Because of symmetry, reciprocity, and negligible crosstalk between the leftmost and right-most pairs, the 36 components in each matrix can be replaced by only five components. If *WCK_t* to DQ_{m+1} correspond to strips 1–6, then *RLGC* components 11, 12, 13, 23, and 14 are sufficient to represent the full 6×6 matrix as mentioned in [29].

III. PROPOSED TRAINING SET OPTIMIZATION FOR EFFICIENT ANN TRAINING

A. Probability Analysis for a 2-D Uniform Distribution

As described earlier, to construct the ANN shown in Fig. 3, W and S were uniformly selected by the DoE method to form the training datasets [29], but W/S was not uniformly

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

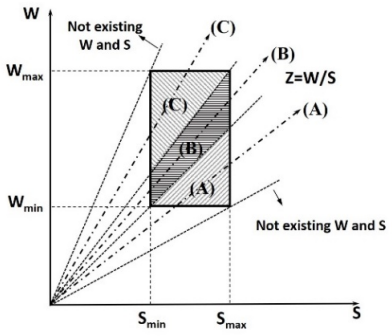


Fig. 4. Uniformity analysis of W/S for the given ranges of W and S.

distributed. Because W/S can significantly affect the accuracy of the mutual *RLGC* components, large prediction errors occurred for the cases with the boundary W/S values, which were insufficiently covered by the training data. As shown in Table I and Fig. 3, five of the six design parameters except W/S were the inputs of the ANN. The uniformity of these five parameters was automatically enforced by the DoE method if they were directly used as the input parameters of the DoE. The remaining design parameter, W/S, however, was not independent. Its uniformity can be investigated rigorously by performing probability analysis.

Let us assume that W and S are uniformly distributed in the ranges of $[W_{\min},\ W_{\max}]$ and $[S_{\min},\ S_{\max}]$, respectively. As shown in Fig. 4, all possible W and S combinations form a rectangular area, which can be further divided into three parts: A to C. Of note, for an arbitrary point $(W,\ S)$ in the figure, W/S is the slope of the line passing through the point and the origin of the coordinate system. Then regions A, B, and C correspond to the W/S ranges of $[W_{\min}/S_{\max},\ W_{\min}/S_{\min}]$, $[W_{\min}/S_{\min},\ W_{\max}/S_{\max}]$, and $[W_{\max}/S_{\max},\ W_{\max}/S_{\min}]$, respectively (assuming $W_{\max}/S_{\max} > W_{\min}/S_{\min}$; other cases can be similarly derived). Furthermore, the PDFof W/S determines its uniformity. If it is constant, then W/S is uniformly distributed. To obtain the PDF, the cumulative distribution function (CDF) is first calculated in each part. Then the PDF is obtained as the derivative of the CDF.

The PDF of the 2-D uniform distribution W and S with ranges of $W_{\min}$–$W_{\max}$ and $S_{\min}$–$S_{\max}$ is described in the following:

$$f_{w,s}(W,S) = \begin{cases} \dfrac{1}{W_{\max}-W_{\min}} \times \dfrac{1}{S_{\max}-S_{\min}}, & \text{Inside rectangle} \\ 0, & \text{elsewhere.} \end{cases} \tag{1}$$

In part A, the CDF of W/S can be obtained as shown in (2), where Z is an arbitrary value in the range of $[W_{\min}/S_{\max},\ W_{\min}/S_{\min}]$

$$\begin{aligned} &\mathrm{CDF}_1(W/S \le Z) \\ &= \mathrm{CDF}_1(W \le ZS) \\ &= \int_{W_{\min}/Z}^{S_{\max}} \int_{W_{\min}}^{ZS} \frac{1}{W_{\max}-W_{\min}} \times \frac{1}{S_{\max}-S_{\min}} dW dS \\ &= \frac{ZS_{\max}^2 - 2W_{\min}S_{\max} + \left(W_{\min}^2/Z\right)}{2(W_{\max}-W_{\min})(S_{\max}-S_{\min})}. \end{aligned} \tag{2}$$

Only when the CDF is a linear first-order function of Z does the derivative, which is the PDF, become constant. Clearly, the PDF of W/S in part A is not constant, and thus W/S is not uniformly distributed in this part

$$\begin{aligned} &\mathrm{CDF}_2(W/S \le Z) \\ &= \mathrm{CDF}_2(W \le ZS) \\ &= \int_{S_{\min}}^{S_{\max}} \int_{W_{\min}}^{W_{\min}S} \frac{1}{W_{\max}-W_{\min}} \times \frac{1}{S_{\max}-S_{\min}} dW dS \\ &\quad + \int_{S_{\min}}^{S_{\max}} \int_{S_{\min}S}^{ZS} \frac{1}{W_{\max}-W_{\min}} \times \frac{1}{S_{\max}-S_{\min}} dW dS \\ &= \frac{(Z - W_{\min}/S_{\min})(S_{\max}+S_{\min}) + W_{\min}(S_{\max}/S_{\min}-1)}{2(W_{\max}-W_{\min})(S_{\max}-S_{\min})}. \end{aligned} \tag{3}$$

In part B, the corresponding CDF can be derived similarly, as shown in (3). The CDF_2 is clearly a linear first-order function of Z, which is an arbitrary value in the range of $[W_{\min}/S_{\min},\ W_{\max}/S_{\max}]$. Therefore, the PDF_2 becomes constant, and W/S is uniformly distributed in this part.

In part C, the CDF of W/S is similarly derived, as in (4), where Z is an arbitrary value in the range of $[W_{\max}/S_{\max},\ W_{\max}/S_{\min}]$. The CDF is clearly not a linear function of Z, and thus W/S is not uniformly distributed in this part

$$\begin{aligned} &\mathrm{CDF}_3(W/S \le Z) \\ &= \mathrm{CDF}_3(W \le ZS) \\ &= \int_{S_{\min}}^{S_{\max}} \int_{W_{\min}}^{S_{\min}S} \frac{1}{W_{\max}-W_{\min}} \times \frac{1}{S_{\max}-S_{\min}} dW dS \\ &\quad + \int_{W_{\max}/Z}^{S_{\max}} \int_{S_{\min}S}^{W_{\max}} \frac{1}{W_{\max}-W_{\min}} \times \frac{1}{S_{\max}-S_{\min}} dW dS \\ &\quad + \int_{S_{\min}}^{W_{\max}/Z} \int_{S_{\min}S}^{ZS} \frac{1}{W_{\max}-W_{\min}} \times \frac{1}{S_{\max}-S_{\min}} dW dS \\ &= 1 - \frac{ZS_{\max}^2 + W_{\max}^2/Z - 2W_{\max}S_{\min}}{2(W_{\max}-W_{\min})(S_{\max}-S_{\min})}. \end{aligned} \tag{4}$$

In summary, when both W and S are uniformly distributed, W/S is uniformly distributed in part B only. Of note, in our practical problem, W, S, and W/S are all design parameters with specified ranges. If the W/S range is larger than $[W_{\min}/S_{\min},\ W_{\max}/S_{\max}]$, as is the case in this article, W/S is clearly not uniformly sampled when W and S are uniformly sampled in their ranges by using the DoE method. This is the underlying reason for the issue with the training set selection in reference [29] discussed earlier.

To solve this problem, the ranges of W and S can be increased such that part B in Fig. 4 covers the entire range of W/S. Thus, on the basis of probability analysis, W/S in the specified range is ensured to be uniformly sampled when W and S are uniformly sampled in their ranges. In the case under study, if the S range remains the same as that in the original, and the W range is increased to [1, 10] mil, the specified range of W/S [0.2, 1] becomes the same as that in part B. That is, when W is uniformly sampled in [1, 10] mil, and S is uniformly sampled in [5, 10] mil, W/S is sampled in the range

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

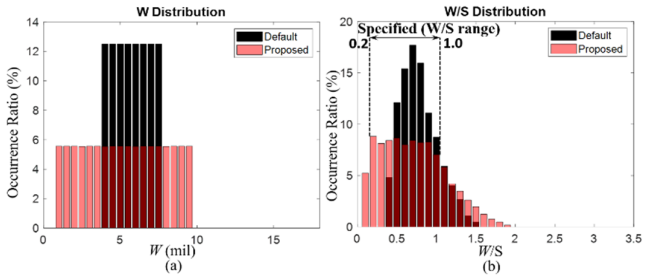


Fig. 5. (a) W and (b) W/S distributions of training data when the DoE method uses an increased W range while the S range is maintained (proposed is case 2 in Table II).

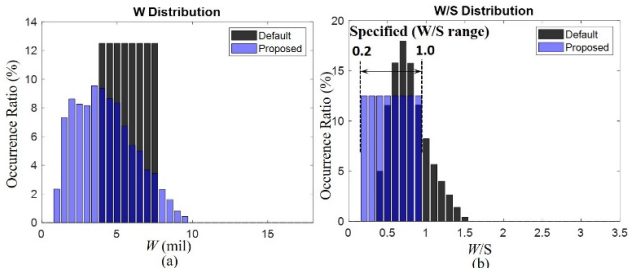


Fig. 6. (a) W and (b) W/S distributions of training data when the DoE method uses a uniform W/S (proposed is case 7 in Table II).

TABLE II
PARAMETER SETUP TO GENERATE TRAINING DATASETS BASED ON PROPOSED DOE METHODS

Case #	Uniform part	W_{MIN} (mil)	W_{MIN} (mil)	$(W/S)_{MIN}$	$(W/S)_{MAX}$
1: (Default) [29]	***W***	**4**	**8**	0.4	1.6
2: W=1~10 mil (Optimal)		**1**	**10**	0.2	1.0
3: W=2~16 mil		**3**	**12**	0.3	2.4
4: W=1~32 mil		**2**	**16**	0.2	3.2
5: W=3~12 mil		**1**	**32**	0.1	6.4
6: W/S=0.4~1.6	***W/S***	2	16	**0.4**	**1.6**
7: W/S=0.2~1.0 (Alternative)		2	10	**0.2**	**1.0**

of [0.1, 2], within which in [0.2, 1], W/S is also uniformly sampled. Using this method, we can ensure the uniformity of W/S in its specified range; however, this process results in sampling points outside of the specified ranges of W and W/S.

The range of [1, 10] mil for W ensures the generation of uniformly distributed W/S in its specified range. Further increasing the range of W can still provide a uniform distribution of W/S in its specified range; however, more sample points will fall outside of the specified ranges of W and W/S, thus decreasing the efficiency and accuracy.

Alternatively, instead of using W and S as input parameters, we can use W/S and S. If W/S is uniformly sampled in [0.2, 1], and S is uniformly sampled in [5, 10] mil, W is then sampled in the range of [1, 10] mil, but clearly the sampling of W is no longer uniform, as can be easily seen in Fig. 4: only part B is sampled. This nonuniformity can clearly affect accuracy when the number of training data samples is small. With enough training samples, the effect on accuracy is negligible, because the entire range of W is covered.

B. DoE Setups to Enhance ANN Model Accuracy

To validate the proposed solutions through probability analysis, we tested different DoE setups in the case under study to compare the performance of the different training set selections. The black bars in Fig. 5(a) and (b) show the W and W/S distributions in the training datasets generated by the initial default setup as case 1 in Table II. The default setup ensured a uniform W distribution because the input training datasets of W were generated with the LHS method; however, the uniformity of the W/S distribution was not achieved by the uniformly distributed W and S, as shown in Fig. 5(b). Although the occurrence ratio at $W/S = 0.7$ is more than 15%, those at the boundary ($W/S < 0.5$) are less than 5%. This low uniformity of W/S reduces the accuracy of the mutual *RLGC* elements, particularly when the geometry of channel has a W/S value near the boundary.

The first solution proposed in Section III-A increases the range for input data W to [1, 10] mil while maintaining the S range. This is case 2 listed in Table II. Consequently, the uniformity of the W/S distribution can be relatively enhanced. The red bars in Fig. 5 show the W and W/S distributions when the training datasets are generated with an increased range of W, from 1 to 10 mil.

Although the occurrence ratio of training datasets within the specified W and W/S ranges is relatively reduced, the uniformity of W/S within the specified W/S range is significantly enhanced. As long as the number of training data samples included in the specified W and W/S ranges is sufficient, the proposed modification can improve the accuracy of the ANN model compared with that of the original DoE set up under the same number of training data samples.

The alternative solution proposed in Section III-A uses W/S and S as the input parameters, and samples W/S uniformly in [0.2, 1] and S uniformly in [5, 10] mil, as in case 7 listed in Table II. The resulting distributions of the training datasets are compared with the default case in Fig. 6. Although this new DoE setup sacrifices the uniformity of the W distribution, which dominates the self *RLGC*, the accuracy of the mutual *RLGC* models can be enhanced because of the completely uniform W/S distribution.

To estimate the error rates of the ANN models trained on the training datasets generated in the different cases listed in Table II, we randomly selected 140 additional channel structures within the parameter ranges in Table I, which were simulated numerically as the test data for error analysis. The learning curves of the average error rates of all cases as functions of the number of training data samples used in the ANN training are compared in Fig. 7. The average error rates of all cases clearly decrease as the number of training data samples increases, but the decreasing rate of the original DoE setup is the smallest because more evenly distributed W/S is achieved in all other cases, thus further resulting in better model accuracy, particularly at the boundary. The performance of the proposed DoE modifications in Section III-A is clearly

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

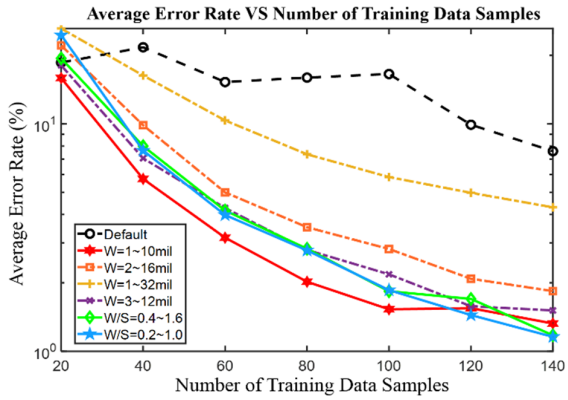


Fig. 7. Learning curves of different DoE modifications for training dataset generation. The lowest average error rates are achieved by the proposed modification with uniform W sampling with an increased range from 1 to 10 (solid red line), or the proposed alternative modification with uniform W/S sampling from 0.2 to 1.0 (solid blue line).

the best when the number of training data samples is more than 60. The decreasing rate of the original DoE setup is the smallest because more evenly distributed W/S is achieved in all other cases, thus further resulting in better model accuracy, particularly at the boundary. As derived in the probability analysis, case 2 has the minimum required range of W that ensures the uniformity of W/S in its specified range. Other DoE cases, such as cases 3 and 4, have larger W ranges than the original case and are thus better, but they cannot ensure the uniformity of W/S in its specified range. Case 5 can ensure the uniformity of W/S in its specified range; however, the larger W range results in fewer useful sampling points in the required ranges, thus lowering accuracy. Case 6 again cannot ensure the uniformity of W/S in its specified range. The learning curves shown in Fig. 7 validate the proposed modifications based on the probability analysis.

Furthermore, the error rate distributions of the self and mutual *RLGC* components for all cases are plotted in Fig. 8, with the number of training data samples fixed at 120. The default case for both the self and mutual elements is the poorest in terms of the accuracy of the predicted ANN elements. The proposed alternative modification (case 7) is the best, and the other proposed modification (case 2) is the second best, as shown in Fig. 8(b). Case 5, which has the largest W range, is clearly worse than all other modifications.

As shown in Fig. 8(a), most of the self *RLGC* components are precisely predicted, with an error rate of less than 5%. In contrast, the error rates of the predicted mutual *RLGC* components increase overall, to values greater than 10% in some cases, as depicted in Fig. 8(b). This change in error rate occurs because the mutual *RLGC* values are dominated by the ratio of W/S. Although the relative standard deviation of the self *RLGC* components in the test set is 0.161, those of the mutual *RLGC* components 12, 23, 13, and 14 are 0.404, 0.865, 1.202, and 1.739, respectively, where WCK_t to DQ_{m+1} in Fig. 2 correspond to strips 1–6. Because the mutual *RLGC* variations due to different channel structures are significantly larger than the self *RLGC* components, the error rates of the

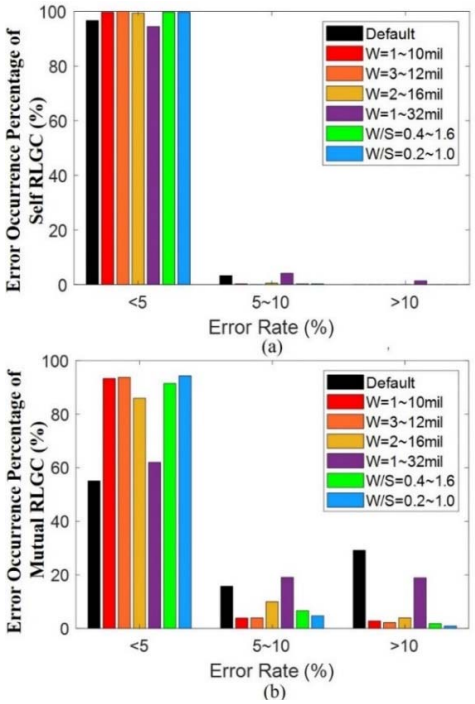


Fig. 8. Error rate distributions of different DoE modifications. (a) Self *RLGC* components. (b) Mutual *RLGC* components. The number of training data samples is fixed to 120.

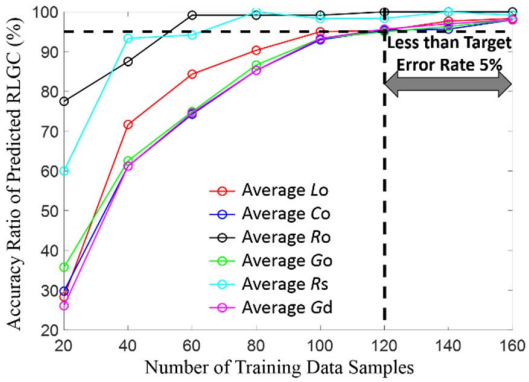


Fig. 9. Ratios of predicted RLGC values with an error rate less than 5%, as functions of the number of training data samples. Training data samples are generated by the proposed DoE modification with a uniform W from 1 to 10 mil. The optimal number of training data needed to achieve the target ratio of 95% is 120.

predicted mutual *RLGC* components are relatively higher than those of the self *RLGC* components, as shown in Fig. 8.

Notably, that the self *RLGC* components are accurately predicted regardless of the DoE modifications used, as shown in Fig. 8(a), whereas the accuracy of the predicted mutual *RLGC* components is significantly dependent on these factors, as shown in Fig. 8(b) because the mutual *RLGC* models are more dominated by the W/S distribution in the specified range.

C. Optimal Number of Training Data Samples

Although using a greater number of training data samples can lead to higher accuracy of an ANN model, too many

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

TABLE III
COMPARISON OF THE COMPUTATIONAL TIME AND ACCURACY

Number of Training Data Samples	Average Accuracy Ratio (%)		Prediction Time for *RLGC* parameters (Minutes)	
	Default Method	Optimal Method 1 Case 2 W=1~10 mil	Proposed Method	Computing Time by EM Solver
20	52.89	40.99	10.34	62.10
40	53.45	73.25	15.28	168.32
60	58.70	84.25	20.32	222.12
80	65.43	89.33	26.18	328.24
100	67.14	93.75	30.30	370.31
120	74.40	96.22	35.19	468.17
140	70.85	98.46	41.42	490.23
160	78.37	98.19	46.81	544.36

training data samples can negatively affect simulation time and resources. A learning curve that indicates the prediction error rates of the test datasets in accordance with the number of training data samples can be used to investigate the tradeoff between performance and cost. In this study, we defined a 5% target error rate for each element, including *L*o, *C*o, *R*o, *G*o, *R*s, and *G*d. The percentages of the predicted elements with an error rate less than 5% are plotted in Fig. 9 as functions of the number of training data samples used in the ANN training. The training datasets were generated by using the proposed alternative modification method (case 2) with a uniform *W* distribution from 1 to 10 mil, which has the best learning curve for the constructed ANN, as shown in Fig. 7. As the number of training data samples increases, the accuracy of the element models is enhanced. When the number of training data samples is more than 120, all the element values achieve the target specification. Therefore, 120 is the optimal number of training data samples because, in this example, it enables the most efficient channel model to be developed based on the proposed DoE modification.

Table III lists the computational times in constructing an ANN with the number of the training samples and the corresponding accuracy already shown in Fig. 7. These costs are necessary to construct an ANN; however, after an ANN is constructed as a generic model of the target channel, predicting the electrical performance of a channel with arbitrary geometrical parameters within the predefined ranges becomes very fast and does not require any EM simulation skills. Less than 30 s is required for the ANN to predict the *RLGC* elements for the channel studied herein, compared with at least 180 s when a 2-D analysis tool is used under the same hardware condition. As a result, the time and resource cost for the parameter prediction of a large number of channels such as GDDR and HBM by our proposed method is much less than those based on 2-D or 3-D solvers, and the more channels, the more efficiency is achieved.

IV. VALIDATION OF THE PROPOSED METHOD

To validate the performance of the proposed optimal DoE approach for ANN, the commercial EM simulator is considered the "gold standard" for generating S-parameters for various structures of channels. The ANN results with the default and proposed alternative setup of DoE were compared with those from simulations to validate the performance of our method. The channel structures used for validation were randomly selected within the parameter ranges in Table I. The chosen channel structures are listed in Table IV. The number of training data samples for the ANN was fixed to 120.

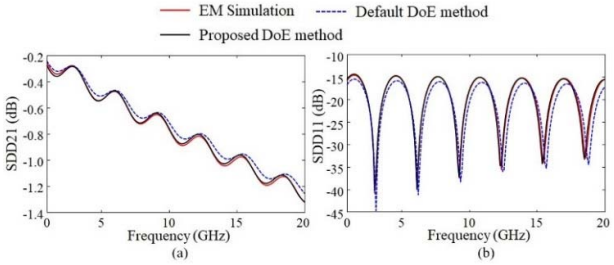


Fig. 10. Comparisons of loss of strip lines obtained from EM simulations, predicted by ANN models trained by using the default DoE method and predicted by ANN models trained by using the proposed optimal DoE setup. (a) $|S_{DD21}|$. (b) $|S_{DD11}|$.

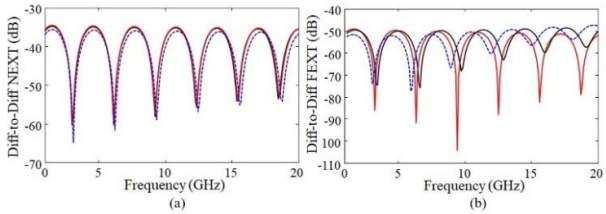


Fig. 11. Comparisons of crosstalk of strip lines obtained from EM simulations, predicted by ANN models trained by using the default DoE method and predicted by ANN models trained by using the proposed optimal DoE setup (of note: proposed one takes less training data samples and time to achieve the same accuracy here comparing with default one). (a) Differential-to-differential NEXT. (b) Differential-to-differential FEXT between adjacent lines of the different pairs.

Both Differential pairs and single-ended transmission lines act as the high speed and huge bandwidth channels. The type of differential pairs has more complicated characteristics and was used as the case for validation to verify our proposed method. Differential insertion loss $|S_{DD21}|$, differential return loss $|S_{DD11}|$, differential-to-differential near-end crosstalk (NEXT), and far-end crosstalk (FEXT) between adjacent differential lines are shown in Figs. 10 and 11, respectively, for case 1 listed in Table IV. The *W* and *W*/*S* values in this case are 4 and 0.4 mil, respectively. The proposed alternative DoE setup clearly achieves significantly better predictions than the default DoE setup when the same number of training data samples is used. Although the magnitudes of the differential-to-differential FEXTs are predicted well, the resonance frequencies of the FEXTs simulated by the ANN models and the EM simulation results correlate less well, as shown in Fig. 11(b). The crosstalk between the differential lines is a function of S_P, which has a very wide parameter range, from 5 to 50 mil. In addition, the FEXT value between strip lines is usually very small, owing to the approximate homogeneous property. Thus, the effects of the inaccurate resonance frequency are negligible.

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
GDM COMPARISON RESULTS FOR COMPLICATED DIFFERENTIAL CHANNEL STRUCTURES

#	S (mil)	W (mil)	H_P (mil)	H_C (mil)	S_P (mil)	SDD21		SDD11		NEXT		FEXT	
						Default	Proposed	Default	Proposed	Default	Proposed	Default	Proposed
1	10	4	6	10	12	G (0.26)	E (0.09)	F (0.43)	E (0.09)	G (0.37)	E (0.08)	VP (1.78)	P (1.02)
2	9.3	4.1	14.4	5.5	14.6	G (0.27)	VG (0.18)	F (0.42)	G (0.23)	F (0.44)	VG (0.14)	F (0.58)	VP (2.02)
3	7	6.9	4.4	8.3	17.8	E (0.07)	E (0.09)	F (0.46)	E (0.09)	P (1.42)	VG (0.19)	P (1.08)	F (0.64)
4	9.7	7.7	4.6	7	30.9	F (0.62)	G (0.25)	P (0.84)	G (0.31)	P (1.59)	G (0.35)	VP (2.45)	F (0.65)
5	9.4	5.7	6	10	18.2	E (0.09)	E (0.04)	F (0.48)	VG (0.11)	VG (0.15)	E (0.07)	P (1.59)	P (1.59)
6	5.8	4.6	8.6	9	9.9	E (0.04)	E (0.04)	VG (0.14)	VG (0.1)	VG (0.17)	E (0.05)	VP (4.34)	P (1.53)
7	7.9	7.6	11	11.7	21.2	VG (0.14)	E (0.04)	G (0.21)	F (0.54)	VG (0.15)	E (0.06)	VP (2.62)	VP (2.26)
8	9.2	6.1	12.4	7.6	28.8	VG (0.18)	E (0.06)	G (0.21)	G (0.21)	G (0.24)	VG (0.11)	P (1.58)	VP (1.94)

TABLE V
CORRELATION LEVELS AND CORRESPONDING GDM VALUES

Correlation Levels	E	VG	G	F	P	VP
Meaning	Excellent	Very Good	Good	Fair	Poor	Very Poor
GDM Range	< 0.1	0.1–0.2	0.2~ 0.4	0.4~ 0.8	0.8– 1.6	>1.6

Eye diagram is a good index to describe the channel quality. However, it is data rate and eye mask dependent and not a figure of merit for an obvious judgment. To numerically compare the correlation of the S-parameter curves, the global difference measure (GDM) in the feature selective validation (FSV) method is used alternatively [36]. The GDM is an overall single figure of merit between the two datasets being compared. It allows a simple decision to be made about the quality of a comparison. It is obtained from the overall values for the two components, the amplitude difference measure (ADM), and the feature difference measure (FDM). The relationship between the correlation levels and the corresponding ranges of GDM values are listed in Table V. According to the calculated GDM values, the overall accuracy of the trained ANN models by using the proposed DoE modification is significantly better than that of the ANN models based on the original DoE setup.

V. IMPLEMENTATION IN MICROSTRIP LINE MODELING

In this section, generic models of microstrip channels are obtained similarly with ANNs constructed by using the training datasets selected according to the default and the proposed alternative DoE method. With the same method for error sets analysis and an error margin of 5%, the number of training data used for the ANN model was selected to be 120. The structure of the microstrip channel under study is shown in Fig. 2. Each of the three differential pairs has an identical conductor width (W), intrapair spacing (S), and interpair spacing (S_P). The height of the dielectric layer is H_P. W, S, S_P, H_P, and W/S are the design parameters of the target channel. The upper and lower bounds of each design parameter are listed in Table VI. Although this application has only five design parameters, the W and S ranges are increased. The allowable W/S ratio is limited to between 0.2 and 1.0. Both the conductor and solder-mask thicknesses are fixed to 0.6 mil.

A boundary case with randomly selected geometries was selected to validate the improved performance of the proposed DoE method. The selected W, S, S_P, and H_P of the target

TABLE VI
RANGES OF DESIGN PARAMETER FOR TRAINING SET OPTIMIZATION (MICROSTRIP)

Parameter	Min	Max	Parameter	Min	Max
W	5 mil	12 mil	H_C	3.5 mil	15 mil
S	5 mil	12 mil	S_P	5 mil	50 mil
W/S	0.2	1.0			

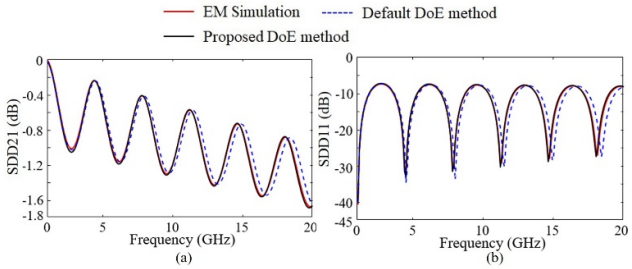


Fig. 12. Comparisons of loss of microstrip lines obtained from EM simulations, predicted by ANN models trained by using the default DoE method and proposed optimal DoE setup. (a) $|S_{DD21}|$. (b) $|S_{DD11}|$.

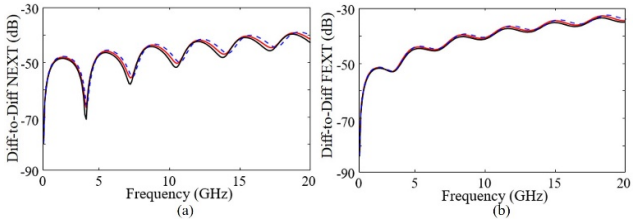


Fig. 13. Comparisons of crosstalk of microstrip lines obtained from EM simulations, predicted by ANN models trained by using the default DoE method and proposed optimal DoE setup (of note: proposed one takes less training data samples and time to achieve the same accuracy here comparing with default one). (a) Differential-to-differential NEXT. (b) Differential-to-differential FEXT between adjacent lines of the different pairs.

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

microstrip channel are 5, 12, 40, and 11 mil, respectively. The *W*/*S* ratio is 0.416, which falls into the lower boundary case.

The predicted $|S_{\mathrm{DD21}}|$ together with $|S_{\mathrm{DD11}}|$ and the differential-to-differential NEXT together with differential-to-differential FEXT between adjacent lines are shown in Figs. 12 and 13, respectively, for the target channel. The predicted S-parameter curves when using the proposed DoE method nearly overlap with the EM simulation results. These findings demonstrate that using 120 training data samples with the proposed DoE method is sufficient to accurately model the target microstrip channel.

VI. Conclusion

In this article, training set optimization in an ANN constructed for ultraspeed and high bandwidth interconnects design is proposed. The enhanced ANN can precisely predict the RLGC components for any arbitrary structure for the channels between GPU and GDDR6 memory when the geometrical parameters are within predefined ranges. The uniformity of *W/S* sampling in its specified range in the training datasets is a dominant factor affecting the accuracy of the mutual RLGC elements in the modeling of multiconductor transmission lines. A probability analysis was performed to understand the reason underlying the *W/S* non-uniformity causing the accuracy issues, and two methods are proposed to address the issue in this article. When an ANN model of the target channel is trained on 120 training data samples generated by the proposed alternative optimized DoE modification, 95% of the predicted RLGC components for the test datasets achieve an error rate less than 5%, while the accuracy is only 74.40% by traditional DoE on the same condition. Validation is ultimately performed by comparing S-parameters obtained from proposed methods, traditional DoE, and gold standard EM simulations. Furthermore, around 10 times speedup than EM simulation is achieved by proposed approach as listed in Table III.

The proposed method should enable rapid and precise modeling in contrast with traditional numerical solutions, and allow for ANN predictions with much fewer training data samples than the traditional DoE setup. The rigorous analysis process introduced herein should be desired for implementation in modeling any arbitrary, complex, multiconductor transmission line systems, and be potentially capable to deal with inhomogeneous, non-isotropic, multilayered structures with less difficulty for costly training data samples preparation by simulation or measurement.

References

[1] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA, USA: Artech House, 2000.

[2] J.-M. Jin, *The Finite Element Method in Electromagnetics*. New York, NY, USA: Wiley, 2002.

[3] R. F. Harrington and J. L. Harrington, *Field Computation by Moment Methods*. Oxford, U.K.: Oxford Univ. Press, 1996.

[4] A. E. Ruehli and P. A. Brennan, "Efficient capacitance calculations for three-dimensional multiconductor systems," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-21, no. 2, pp. 76–82, Feb. 1973.

[5] M. A. Sanchez-Soriano *et al.*, "Peak and average power handling capability of microstrip filters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 8, pp. 3436–3448, Aug. 2019.

[6] V. Nayyeri, M. Soleimani, and O. M. Ramahi, "A method to model thin conductive layers in the finite-difference time-domain method," *IEEE Trans. Electromagn. Compat.*, vol. 56, no. 2, pp. 385–392, Apr. 2014.

[7] G. Macchiarella *et al.*, "Accurate modeling of stubs used as resonant coupling elements in SIW filters," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 12, pp. 1125–1128, Dec. 2020.

[8] D. Y. Heh and E. L. Tan, "Multiple LOD-FDTD method for multiconductor coupled transmission lines," *IEEE J. Multiscale Multiphys. Comput. Techn.*, vol. 5, pp. 201–208, 2020.

[9] B. Pu, K. H. Kim, S. Kim, and W. Nah, "Modeling and parameter extraction of coplanar symmetrical meander lines," *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 3, pp. 375–383, Jun. 2015.

[10] J. Zhang, F. Feng, J. Jin, and Q.-J. Zhang, "Efficient yield estimation of microwave structures using mesh deformation-incorporated space mapping surrogates," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 10, pp. 937–940, Oct. 2020.

[11] Y. Yu and Z. Chen, "A 3-D radial point interpolation method for meshless time-domain modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 8, pp. 2015–2020, Aug. 2009.

[12] W. Zhang, F. Feng, J. Zhang, Z. Zhao, J. Ma, and Q.-J. Zhang, "Parallel decomposition approach to wide-range parametric modeling with applications to microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 12, pp. 5288–5306, Dec. 2020.

[13] V. K. Devabhaktuni, M. C. E. Yagoub, and Q.-J. Zhang, "A robust algorithm for automatic development of neural-network models for microwave applications," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 12, pp. 2282–2291, Dec. 2001.

[14] W. Na and Q. Zhang, "Automated parametric modeling of microwave components using combined neural network and interpolation techniques," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2013, pp. 1–3.

[15] M. S. Nakhla and Q. J. Zhang, *Modeling and Simulation of High-Speed VLSI Interconnects*. Boston, MA, USA: Kluwer, 1994.

[16] P. M. Watson and K. C. Gupta, "EM-ANN models for microstrip vias and interconnects in dataset circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 44, no. 12, pp. 2495–2503, Dec. 1996.

[17] A. Veluswami, M. S. Nakhla, and Q.-J. Zhang, "The application of neural networks to EM-based simulation and optimization of interconnects in high-speed VLSI circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 45, no. 5, pp. 712–723, May 1997.

[18] F. Wang, V. K. Devabhaktuni, and Q.-J. Zhang, "A hierarchical neural network approach to the development of a library of neural models for microwave design," *IEEE Trans. Microw. Theory Techn.*, vol. 46, no. 12, pp. 2391–2403, Dec. 1998.

[19] S. Liao, H. Kabir, Y. Cao, J. Xu, Q.-J. Zhang, and J. Ma, "Neural-network modeling for 3-D substructures based on spatial EM-field coupling in finite-element method," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 1, pp. 21–38, Jan. 2011.

[20] T. Lu, J. Sun, K. Wu, and Z. Yang, "High-speed channel modeling with machine learning methods for signal integrity analysis," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 1957–1964, Dec. 2018.

[21] W. Na, W. Liu, L. Zhu, F. Feng, J. Ma, and Q.-J. Zhang, "Advanced extrapolation technique for neural-based microwave modeling and design," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4397–4418, Oct. 2018.

[22] A. A. Ilumoka, "Efficient and accurate crosstalk prediction via neural net-based topological decomposition of 3-D interconnect," *IEEE Trans. Adv. Packag.*, vol. 24, no. 3, pp. 268–276, Aug. 2001.

[23] R. J. Pratap, D. Staiculescu, S. Pinel, J. Laskar, and G. S. May, "Modeling and sensitivity analysis of circuit parameters for flip-chip interconnects using neural networks," *IEEE Trans. Adv. Packag.*, vol. 28, no. 1, pp. 71–78, Feb. 2005.

[24] R. Trinchero, P. Manfredi, I. S. Stievano, and F. G. Canavero, "Machine learning for the performance assessment of high-speed links," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 1627–1634, Dec. 2018.

[25] M. Swaminathan, H. M. Torun, H. Yu, J. A. Hejase, and W. D. Becker, "Demystifying machine learning for signal and power integrity problems in packaging," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 8, pp. 1276–1295, Aug. 2020.

[26] T. Hollis *et al.*, "16Gb/s and beyond with single ended I/O in high-performance graphics memory," in *Proc. DesignCon*, Santa Clara, CA, USA, Feb. 2018.

[27] B. Pu, J. S. Pak, C. Jo, and S. Moon, "Design of 2.5 D interposer in high bandwidth memory and through silicon via for high speed signal," in *Proc. DesignCon*, Santa Clara, CA, USA, Jan. 2019.

[28] D. C. Montgomery, *Design and Analysis of Experiments*. New York, NY, USA: Wiley, 1991.

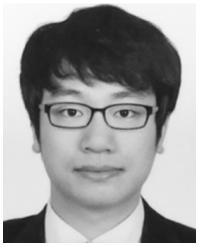
Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.

[29] H. Kim, C. Sui, K. Cai, B. Sen, and J. Fan, "Fast and precise high-speed channel modeling and optimization technique based on machine learning," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 2049–2052, Dec. 2018.
[30] J. Brownlee. (May 2020). *How to Use Polynomial Feature Transforms for Machine Learning*. [Online]. Available: https://www.mahinelearning mastery.com
[31] E. Frick, J. B. Preibisch, C. Seifert, M. Lindner, and C. Schuster, "Variability analysis of via crosstalk using polynomial chaos expansion," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2017, pp. 317–319.
[32] F. Tong and X. Liu, "Samples selection for artificial neural network training in preliminary structural design," *Tsinghua Sci. Technol.*, vol. 10, no. 2, pp. 233–239, Apr. 2005.
[33] Y. LeCun, L. Bottou, G. B. Orr, and K. R. M'uller, "Efficient backprop," in *Neural Networks, Tricks of the Trade* (Lecture Notes in Computer Science), vol. 1524. Berlin, Germany: Springer, 1998, pp. 9–48.
[34] M. Stein, "Large sample properties of simulations using Latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.
[35] S. Kucherenko, D. Albrecht, and A. Saltelli, "Exploring multi-dimensional spaces: A comparison of Latin hypercube and quasi Monte Carlo sampling techniques," 2015, *arXiv:1505.02350*.
[36] A. P. Duffy, A. J. M. Martin, A. Orlandi, G. Antonini, T. M. Benson, and M. S. Woolfson, "Feature selective validation (FSV) for validation of computational electromagnetics (CEM). Part I—The FSV method," *IEEE Trans. Electromagn. Compat.*, vol. 48, no. 3, pp. 449–459, Aug. 2006.

Bo Pu (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2009, and the Ph.D. degree in electronic and electrical engineering from Sungkyunkwan University, Seoul, South Korea, in 2015.

From 2015 to 2020, he was a Staff Engineer with Foundry Business, Semiconductor Research and Development headquarter of Samsung Electronics, Hwaseong, Korea. From 2020 to 2021, he was a Visiting Assistant Research Professor with National Science Foundation (NSF) Industry/University Cooperative Research Center for Electromagnetic Compatibility, Missouri University of Science and Technology, Rolla, MO, USA. In 2021, he joined the DeTooLIC Technology Company Ltd., Ningbo, Zhejiang, China, where he is currently a Vice President. His current research interest includes the design methodology of the electronic design automation (EDA) for chip-package-printed circuits board (PCB) systems. He also focuses on the researches of high-speed integrated circuits system up to 224 Gbps, 2.5-D Si-interposer for high bandwidth memory (HBM), and through silicon via (TSV) for 3-D ICs. He holds ten patents about high speed links and 2.5-D/3-D ICs.

Dr. Pu is a TPC Member of the Joint IEEE EMCS and APEMC 2018, APEMC 2022. He was the recipient of Best Student Paper Award (2011 IEEE APEMC) and Best SIPI Symposium Paper Award (Joint 2021 IEEE EMC+SIPI and EMC Europe) as first author, a Young Scientists Award from the International Union of Radio Science (URSI) in 2014, and the 2020, 2021 Distinguish reviewer of IEEE Transactions on EMC. He also obtained Ph.D. Fellowship Award in 2013, Best Innovation Award, Excellent Performance Award, and Excellent Project Award as the first awardee in 2015–2019 from Samsung Electronics for his outstanding contributions. He was as a Session Chair in IEEE APEMC 2017, IEEE EMC+SIPI 2020, 2021. He is currently an Associate Editor for IEEE ACCESS.



Heegon Kim (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2009, 2011, and 2015, respectively.

He worked at KAIST as a Post-Doctoral Researcher until 2015. In 2016, he joined Missouri University of Science and Technology, Rolla, MO, USA, as a Visiting Assistant Research Professor. His research interests include signal integrity (SI)/power integrity (PI) in chip-package-printed circuits board (PCB) hierarchical system, SI/PI co-simulation method for high-speed channel, 3-D IC based on through-silicon-via, and passive equalizer design.

Xiao-Ding Cai (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 1995.

He has worked with Nortel Networks, Ottawa, Sun Microsystems, Menlo Park, CA, USA, Juniper Networks, Sunnyvale, CA, USA, Rambus, Sunnyvale, and Cisco Systems, San Jose, CA, USA, as Senior Principal Engineer and Technical Leader in systems groups, specializing in high-speed signal integrity and power integrity. He has been actively promoting SI/PI methodologies, emphasizing fundamentals and applications with a number of technical publications and patents.

Bidyut Sen received the Ph.D. degree in physics from the State University of New York, Stony Brook, NY, USA, in 1986.

He is a Principal Engineer with the Unified Compute Server Group, Cisco Systems, Inc., San Jose, CA, USA. He has many years of experience in the computer industry and has worked in several companies, including Sun Microsystems, Sunnyvale, CA, USA, Fujitsu, San Jose, and LSI Logic, San Jose. He has several publications, presentations, and patents.

Chunchun Sui (Member, IEEE) received the B.S. degree in mechanical and electronics engineering from the Beijing Institute of Technology, Beijing, China, in 2005, the M.S. degree in software engineering from Peking University, Beijing, in 2009, and the Ph.D. degree in computer engineering from the Electromagnetic Compatibility Laboratory, Missouri University of Science and Technology, Rolla, MO, USA, in 2015.

Since 2015, he has been with Cisco Systems, San Jose, CA, USA, and Alibaba Group (us) Inc., Palo Alto, CA, USA, as a Hardware Engineer. Since 2020, he has been a Staff Hardware Engineer with Tesla, Inc., Palo Alto. His research interests include high-speed signal integrity and power integrity solution in chip/package/system design.

Jun Fan (Fellow, IEEE) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1994 and 1997, respectively, and the Ph.D. degree in electrical engineering from the Missouri University of Science and Technology, Rolla, MO, USA, in 2000.

From 2000 to 2007, he worked for NCR Corporation, San Diego, CA, USA, as a Consultant Engineer. In July 2007, he joined the Missouri University of Science and Technology, and became Tenured Professor in 2016. From October 2018, he was the Cynthia Tang Missouri Distinguished Professor of computer engineering. He served as the Director of the Missouri S&T EMC Laboratory, Rolla, and the Director of the National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) for Electromagnetic Compatibility (EMC) from 2013. He was a Senior Investigator of Missouri S&T Material Research Center, Rolla, as well. His research focuses on hardware design and fundamental research for electromagnetic compatibility (including signal and power integrity) at the levels of integrated circuit, package, printed circuits board (PCB) and systems, and development of specialized design tools and innovative measurement technologies.

Dr. Fan served as a member of the Board of Directors. He received an IEEE EMC Society Technical Achievement Award in August 2009. He was the Chair of the Technical Advisory Committee and the TC-9 Computational Electromagnetics Committee, and a Distinguished Lecturer, in the IEEE EMC Society. He was Technical Paper Chair and Technical Program Chair for a few IEEE International Symposia on EMC, General Chair for IEEE International Conference on Signal and Power Integrity, founding Chair for the SC-4 EMC for Emerging Wireless Technologies Special Committee, and so on. He currently is the inaugural Editor-in-Chief for the IEEE TRANSACTIONS ON SIGNAL AND POWER INTEGRITY, and an Associate Editor for the IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY.

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on February 21,2023 at 19:16:13 UTC from IEEE Xplore. Restrictions apply.