S-Parameter and Frequency Identification Method for ANN-Based Eye-Height/Width Prediction

Nikita Ambasana, Gowri Anand, Dipanjan Gope, *Senior Member, IEEE*, and Bhyrav Mutnury, *Senior Member, IEEE*

***Abstract*— Design and analysis of high-speed SerDes channels primarily deal with ensuring signal integrity (SI) for desired electrical performance. SI is predominantly judged by time domain (TD) metrics: bit error rate (BER), eye-height (EH), and eye-width (EW). With increasing bit rates and stringent BER criteria, TD simulations are becoming compute-time-intensive. A full-factorial, cost-effective design space exploration for SI is made possible by learning-based mapping of frequency-domain S-parameter data to EH/EW. A major challenge in this mapping procedure is the identification of relevant S-parameter data, such as return loss, insertion loss, crosstalks, and the frequency points at which they are sampled. This paper outlines a methodology to identify the critical S-parameters at specific frequency points using information theory-based definition of data relevance using a fast correlation-based filter solution for feature selection. This technique is applied for identifying relevant features for generating artificial neural network-based prediction models of EH/EW within 2.5% accuracy for channels with high data rates and complex topologies.**

***Index Terms*— Artificial neural network (ANN), eye-height (EH), feature selection, insertion loss, multilayer perceptron (MLP), PCIe, return loss, SATA, signal integrity (SI).**

I. Introduction

ENSURING signal integrity (SI) of data transfer between transmitter and receiver is a challenging task in designing high-speed SerDes (HSS) links. Traditionally, eye-diagrams have been used to assess the SI characteristics. With increasing signaling speeds, bit error rate (BER) requirements have become more stringent, resulting in longer random bit sequence simulations to capture low probability events in the channel. Large complex electrical systems such as blade and rack servers today support several high-speed serial data buses such as SATA, SAS, PCIe, and USB. Electrical designers need to ensure that the signal quality is maintained on all these high-speed links. For such systems, SI evaluation based on eye-height (EH) and eye-width (EW) at a given BER is intensive in compute-time requirements.

Manuscript received August 31, 2016; revised November 16, 2016; accepted January 14, 2017. Recommended for publication by Associate Editor D. Becker upon evaluation of reviewers' comments.

N. Ambasana and D. Gope are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru 560012, India (e-mail: ambasananikita@gmail.com; dipanjan.gope@gmail.com).

G. Anand is with DELL India, Bengaluru 560071, India (e-mail: gowri_anand@dell.com).

B. Mutnury is with DELL USA, Austin, TX 78754 USA (e-mail: bhyrav_mutnury@dell.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCPMT.2017.2661065

Time domain (TD) simulation of signal transmission can be performed in a SPICE circuit simulation environment using channel macromodels. Model order reduction techniques have been used in the past to reduce numerical complexity of simulation for channel models based on multiport-transmission lines [1] and frequency domain (FD) S-parameters [2]. Methodologies to enforce stability, passivity, and causality of such channel macromodels have also been investigated [3], [4]. Furthermore, to obtain accurate BER at higher operation frequencies, statistical methods such as LinkLab [5] and StatEye [6] were developed to speed up the simulation time by obtaining the eye-contour based on the calculation of cumulative distribution functions of received bit voltages using channel transfer functions. However, these methods assume that the underlying systems are linear, which ceases to hold true with present-day transmitters and receivers. Hybrid approaches were developed to capture the goodness of statistical estimates and parallel time-domain simulations of select bit sequences [7]–[9]. These constitute the present-day channel analysis techniques in TD.

For a selected channel topology, there are several variables impacting the signal quality, such as trace length, trace impedance, termination, and process/manufacturing tolerances. Several TD simulations are needed to cover the design space thoroughly enough to come up with a sufficient set of design rules. State-of-the-art methods for design space exploration involve techniques using design of experiments (DoE). DoE is a systematic approach to select the smallest set of designs that optimally capture the design space [10]. Response surface method [11] can be used to fit the DoE-based designs to EH/EW and create a fast and accurate prediction model for Monte Carlo analysis. Multilayer perceptron (MLP)-based artificial neural network (ANN) is used to predict EH and EW from channel topology variables in [12], using DoE for training the models. However, from [13], it is evident that for design spaces with large dimensions and highly nonlinear EH/EW behavior, a DoE-based approach can sometimes provide an inaccurate sensitivity analysis when compared to a full-factorial sweep of variables. For a more exhaustive coverage of the design space, a methodology for mapping S-parameters directly to EH/EW was proposed in [14] using learning-based modeling [15]–[17] and ANNs. A key element of this model generation is the identification of relevant FD metrics sampled at discrete frequency points. An arbitrary selection of frequency points and FD parameters may lead to the following:

1) undersampling of data, leading to loss of important information content and hence inaccurate models;
2) oversampling of data, leading to overtrained costly models with unnecessarily large size and computation cost.

Hence, selecting the right FD metrics and appropriate frequency step sizes is critical to building accurate and low-cost models.

In this paper, building on [18], a fast correlation-based filter (FCBF) [19] method is applied for feature selection (FS), i.e., to identify relevant S-parameters and frequency step sizes containing high amount of relevant information for each of the EH/EW values observed in the DoE set. In machine learning applications, FS is often used when large dimensional data is available, to increase prediction accuracy and make predictor models cost-effective by identifying input dimensions that are most relevant to the outputs [20], [21]. This process of input selection facilitates accurate modeling while reducing the input set size. In conjunction with ANN as the training module, the resultant FD to EH/EW mapping procedure can be trained using DoE-based FD/TD data and is capable of generating full-factorial sweep results without further expensive TD simulations. In the absence of such an information theoretic or data correlation-based method, the selection of FD metrics and frequency step size is dependent on designer intuition or interface expertise that would prevent a widespread usage or may result in unnecessarily large models with redundant or noisy data inputs.

The paper is organized as follows. In Section II, a brief overview of FD to EH/EW mapping is presented. In Section III, the FS process in the context of FD to EH/EW mapping is explained and the constituent steps are described in detail. In Section IV, the working of the FS method is explained for a simple channel with variants designed to have different dominant S-parameters. In Section V, numerical results are presented to demonstrate the efficacy of the proposed approach for SATA 3.0 and PCIe Gen 3 interfaces. Section VI concludes this paper.

II. ANN-Based FD to EH/EW Mapping

Neural networks are machines that are made to function in a manner similar to a brain. Their two main functions are to acquire knowledge from the environment through a learning process and store this knowledge in the "weights." The prediction function for a single hidden layered MLP ANN with N inputs, M hidden nodes, and one output is

$$y = f(\boldsymbol{x}) = \sum_{j=1}^{M} k_j * G\left(\sum_{i=1}^{N} w_{ij} x_i + b_j\right) \quad (1)$$

$$G(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (2)$$

where $\boldsymbol{x}$ is the input vector, w_{ij} is the weight connecting the ith input node to the jth hidden perceptron, and k_j is the weight connecting the jth hidden perceptron to the output node. An MLP is a two or more layer neural network with neuron units as perceptrons that have nonlinear activation functions in the hidden layer/s [22] as shown in Fig. 1. The intrinsic equation of an MLP makes it suitable for mapping input M-dimensional vector to continuous output functions.

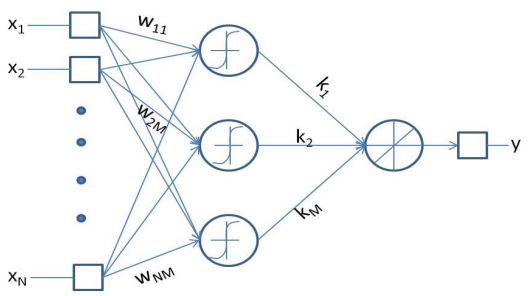


Fig. 1. Generic architecture of an ANN.

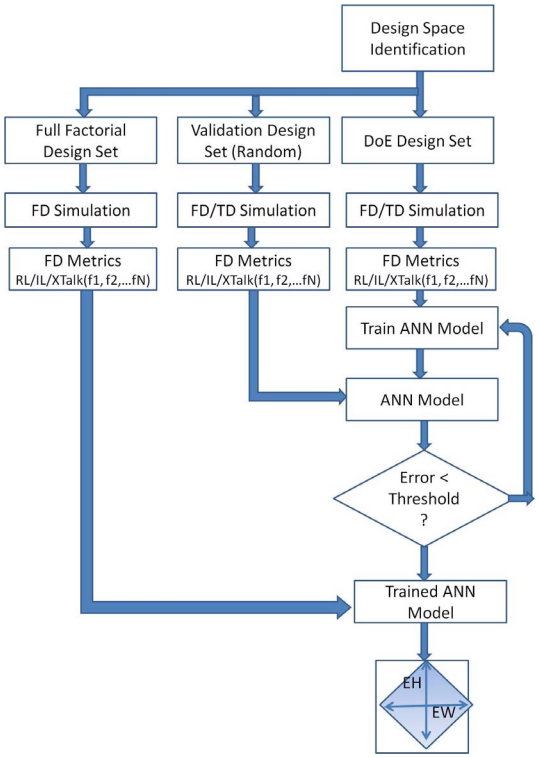


Fig. 2. ANN-based FD to EH/EW model generation and testing process.

For the ANN-based FD to EH/EW mapping scheme as described in [14], and shown in Fig. 2, the choice of frequency points and S-parameters used to generate the model, shown as insertion loss (IL), return loss (RL), near-end crosstalk (NEXT), far-end crosstalk (FEXT), and $f_1, f_2, \ldots f_{1N}$ are critical.

To illustrate, the frequency step size of inputs consisting of IL and total NEXT for a six differential port topology, is varied from 200 MHz to 2 GHz and the prediction accuracy of the ANN model is plotted in Fig. 3. The numbers marked for each point, indicate the total number of inputs to the resultant ANN model. Larger number of inputs results in ANNs with higher number of processing neurons in the hidden layer, increasing training time. On the other hand, lesser number of inputs result in loss of data for accurate mapping as can be seen at 2 GHz sampling. Hence, an intelligent technique to decide the S-parameters and their frequency points, for mapping

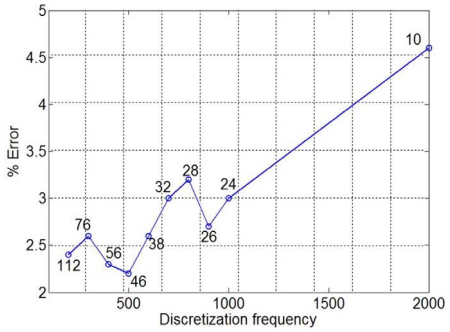


Fig. 3. Effect of frequency step size on prediction error.

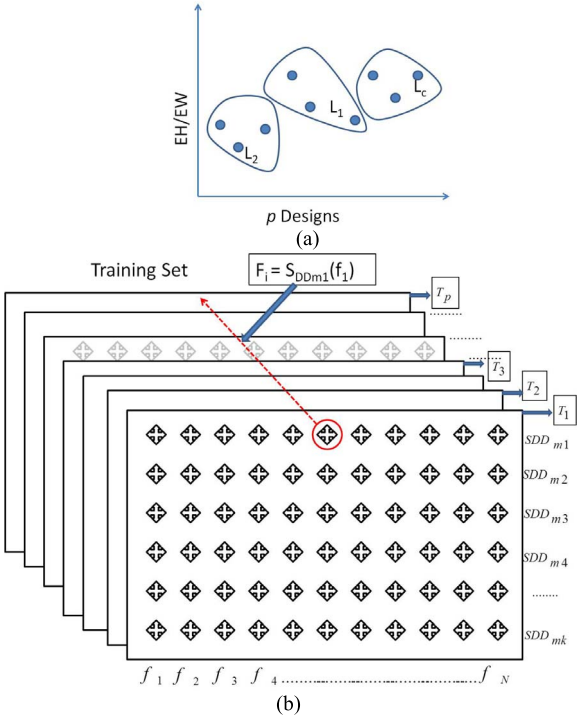


Fig. 4. (a) EH/EW grouped in classes with labels. (b) Training set and features.

is required. Such a technique has the benefit of preventing oversampling, removing redundancy in data, optimizing the number of inputs to the network so as to preclude user intuition to input selection.

III. Feature Selection For S-Parameters and Frequency Points

A "feature" refers to an aspect of the data that is specified before data collection. A "data point" refers to a single collection of feature values for all features [23]. In the context of FD to EH/EW mapping using ANN, each differential S-parameter for the output port "m" at which the eye is measured, at each sampled frequency is a "feature," shown as the diamond shapes in Fig. 4. The label of the EH/EW measured at the port m is the "class" to which that EH/EW belongs, shown as $T_1, T_2, \ldots T_p$. Each T_k takes a value from a fixed set of labels, assigned as shown in the inset of EH/EW versus design figure. The S-parameters at multiple frequencies and the target

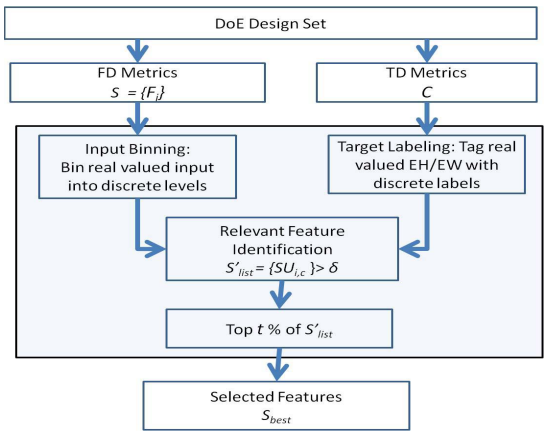


Fig. 5. Feature selection block diagram.

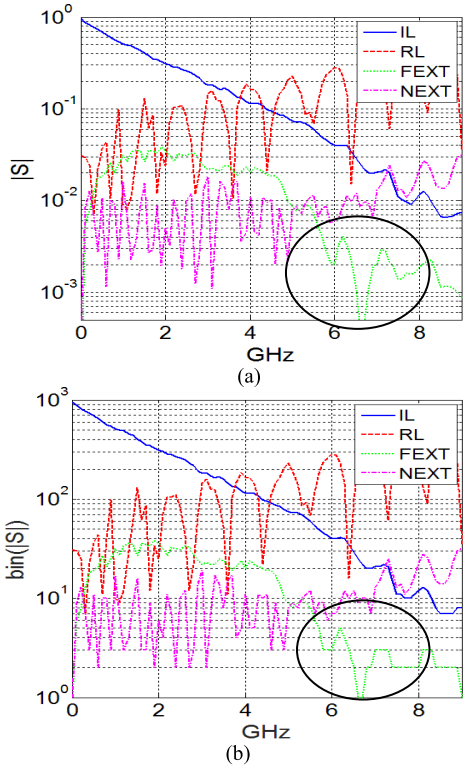


Fig. 6. Input binning. (a) Magnitude of S-parameters. (b) S-parameters binned into 1000 discrete levels.

class for a particular design form a "data point," indicated by each rectangle in Fig. 4. A DoE set of p such data points $D_i = \{F_1, F_2, \ldots F_{kN}; T_i\}, i \in \{1, 2, \ldots p\}$, makes a "training set."

For optimal feature selection in this paper, mutual information (MI)-based relevance definition as appearing in [24] and the technique termed as FCBF [19] is used. The detailed algorithm appears in [19], and a brief overview is presented here.

Consider the symbolic representation of certain terms for the description that follows: goodness measure SU is a normalized indicator of the MI content between two vectors, and the set S is a collection of kN features for p designs of the training

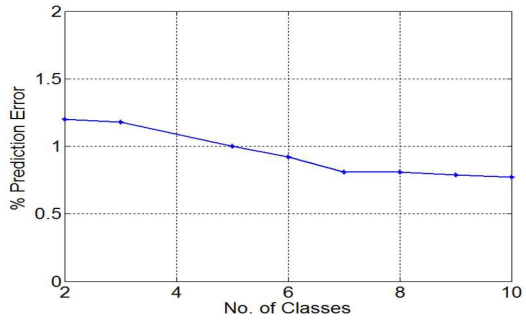


Fig. 7. Impact of the number of target labels on prediction accuracy.

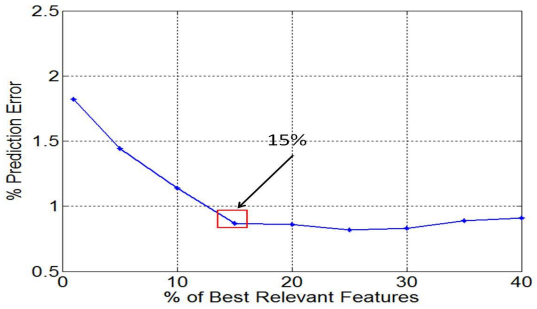


Fig. 8. Prediction accuracy versus percentage of features selected.

set

$$S = \{F_1, F_2, F_3 \ldots F_{kN}\} \tag{3}$$

where

$$F_i \in \left\{ \begin{matrix} \mathrm{SDD}_{m1}(f_1), \mathrm{SDD}_{m2}(f_1), \ldots, \mathrm{SDD}_{mk}(f_1), \\ \ldots\ldots\ldots, \mathrm{SDD}_{m1}(f_N), \ldots\ldots, \mathrm{SDD}_{mk}(f_N) \end{matrix} \right\} \tag{4}$$

and each F_i is a $p \times 1$ vector of discrete sampled FD values for p designs. The size of set S is thus $p \times kN$. The set of target labels C is given by

$$C = \{T_1, T_2, T_3, \ldots T_p\} \tag{5}$$

where

$$T_k \in \{L_1, L_2, \ldots L_C\}. \tag{6}$$

A fixed threshold δ is used for the goodness measure of correlation, $SU_{i,c}$, between the ith feature F_i and the target label vector C. S'_{list} is the subset of S that contains relevant features F_i that have a value of $SU_{i,c} > \delta$. Let t be the desired percentage of kN most relevant features, used to form the set S_{best} to generate the ANN model.

The implementation of the filter-based FS consists of three main processes, input binning, target class labeling, and relevant feature identification, explained in detail as follows. Feature selection toolbox (FEAST) [25] for MATLAB and C, that provides implementation of common MI-based filter feature selection algorithms was used to perform FS for the results presented in this paper. The block diagram involving the subprocesses for FS is shown in Fig. 5.

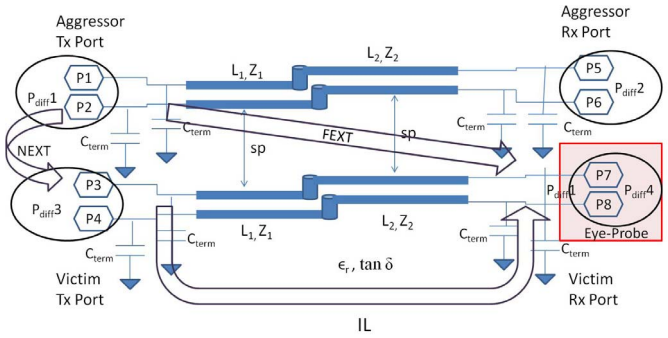


Fig. 9. 5-Gbps generic channel topology with six design variables and four differential ports.

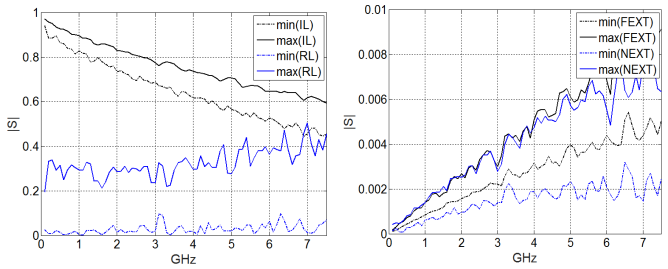


Fig. 10. Minimum–maximum profiles over different training sets for (a) IL, RL and (b) FEXT, NEXT for RL dominant channel variant.

A. Input Binning

The input to the feature selector is a training set that consists of FD and TD simulations of DoE specific designs for a given design space. These simulations are used to form sets of kN features F_i. As the phase response of the channels considered for this paper is linear, for all further discussions, only the magnitude of the differential S-parameters is taken into consideration. The magnitude can be represented in dB, i.e., the log scale or the real value. The method for FS selected for this paper is highly sensitive to the underlying probability distributions of the input. The real value of magnitude was selected as it is the accurate representation of the actual power spread of the losses in the channel, unlike the value in dB that is a logarithmic function of the magnitude. The FS technique used here requires calculation of probabilities of occurrence of specific inputs. To make the probability calculations, the real valued input is divided into discrete levels, by a simple process of thresholding. The input space is first studied to determine the number of levels required to retain the information in as few levels as possible. Fig. 6(a) and (b) shows the differential S-parameters for the victim receiver port (IL, RL, FEXT, and NEXT) for a simple two differential pair topology and the binned values. As shown here, for this paper, the input was binned into 1000 distinct levels. The circles on the figure indicate the effect of binning at very low levels of magnitude variations.

B. Target Labeling

FD to EH/EW mapping is a prediction problem and FS as described here is a classification problem. The most common

TABLE I
FEATURE DISTRIBUTION FOR RL DOMINANT CASE

Features Selected (45 of 300)	Number of points in IL	Number of points in RL	Number of points in FEXT	Number of points in NEXT
EH	16	29	0	0
EW	13	32	0	0

way to convert a prediction problem into a classification one is to tag the outputs with distinct labels or assigned classes. The EH/EW can be distributed into classes in different ways. A primary study was performed on a representative channel to identify the impact of selecting number of classes for a given DoE, on the prediction accuracy of the ANN model, as shown in Fig. 7. As can be seen, the prediction error across the different number of target labels or classes remains consistently within a 2% margin. Based on several such studies, it was observed that the EH/EW spread is captured well within four or five different target labels or classes to achieve accurate prediction. Furthermore, to obtain a tradeoff between the number of classes and the number of DoE sets for any given class, each of the channels' EH/EW were tagged in four or five different classes.

C. Relevant Feature Identification

The FS method presented here falls under the "filter" technique of the broad categorization of "filter" and "wrapper" techniques. This is because features are filtered out due to the lack of their relevance to the target labels or classes based on a given goodness measure. If entropy of a random variable X is defined as (7), where $P(x_i)$ is the probability of occurrence of x_i

$$H(X) = -\sum_i P(x_i)\log_2 P(x_i) \tag{7}$$

$$H(X|Y) = -\sum_j P(y_j)\sum_i P(x_i/y_j)\log_2 P(x_i/y_j). \tag{8}$$

Then, the goodness measure, SU is given by (5)

$$\mathrm{SU}(C, Fi) = \frac{H(C) - H(C/Fi)}{H(C) + H(Fi)} \tag{9}$$

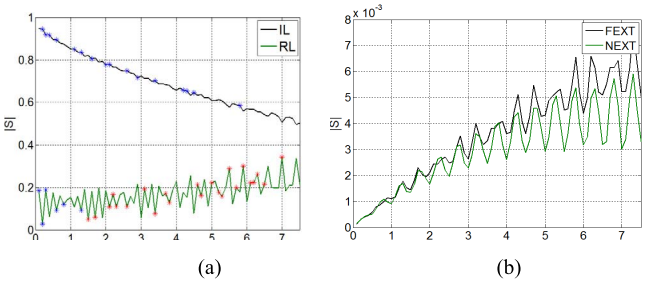
The threshold δ is set to 0, to enable listing all the features that have any correlation with the class distribution, with the most relevant features appearing at the top of the list. The percentage t determines the size of the ANN model and hence a study was performed to identify the smallest number of features that results in saturation of the prediction accuracy. Fig. 8 shows the trend of prediction accuracy versus percentage of features selected. Based on this study, 15% of the total features were set to be selected for all following cases.

IV. PROOF OF CONCEPT: INFORMATION THEORY AND SERDES CHANNELS

A four differential pair topology for a 5-Gbps channel, as shown in Fig. 9 was considered to prove the efficacy of the FS algorithm presented in Section III. P1 and P2 form

Algorithm 1 Finding S_{best} for Class Set C

Input: $S = \{F_1, F_2, \ldots F_{kN}\}$, $C = \{T_1, T_2, \ldots T_P\}$, t, δ
Output: S_{best}
1:begin
2: $S'_{list} = \{\}$; $SU_{val} = \{\}$;
3: for $i = 1{:}kN$ do begin
4: Calculate $SU_{i,c}$ using (9)
5: If $SU_{i,c} > \delta$ do begin
6: $S'_{list} = \{S'_{list}; F_i\}$;
7: $SU_{val} = \{SU_{val}; SU_{i,c}\}$;
8: end;
9: end;
10: order S'_{list} for descending values in SU_{val};
11: for $j = 1{:}t{*}0.01{*}length(S'_{list})$ do begin
12: $S_{best} = \{S_{best}; F_j\}$;
13: end;
14:end;



(a) (b)

Fig. 11. Distribution of selected features over differential S-parameters. (a) IL and RL. (b) FEXT and NEXT. Blue points indicate top 50% of the 45 selected features and red points the next 50%.

differential port 1, P5 and P6 form differential port 2, P3 and P4 form differential port 3, and P7 and P8 form the differential port 4. The aggressor is the top differential pair shown in Fig. 9 with $\mathrm{P_{diff}}1$ as the transmitter and $\mathrm{P_{diff}}2$ as the receiver. The victim is the bottom differential pair shown in Fig. 9, with $\mathrm{P_{diff}}3$ as the transmitter and $\mathrm{P_{diff}}4$ as the receiver. The eye simulations are performed for $\mathrm{P_{diff}}4$. IL is the loss from $\mathrm{P_{diff}}3$ to $\mathrm{P_{diff}}4$, NEXT is the energy coupled from victim to aggressor transmitter ports and FEXT is the energy coupled from victim transmitter port to aggressor receiver port. Even though NEXT is at the transmitter, at high frequencies due to the impedance mismatch between the transmission line and driver (transmitter) impedance, there will be reflections at Tx and NEXT will show up at the receiver due to the impedance mismatch at the source. The design variables, as shown, are the termination capacitances C_{term} that take the values 0.1, 0.2, and 0.3 pF; lengths of traces L_1 vary as 4, 5, and 6 in and L_2 as 6, 7, and 8 in; and the characteristic impedances as $Z_1, Z2$.

Four different variants of the above topology were created to make the channel: 1) return loss (RL) dominant; 2) insertion loss (IL) and RL dominant; 3) crosstalk dominant with dominant NEXT or FEXT with use of microstrip traces and stripline traces; and 4) as having high mode conversion. This was achieved by changing the material properties (ϵ_r, $\tan\delta$),

TABLE II
PREDICTION ACCURACY FOR RL DOMINANT CASE

Metric	Prediction Error	Maximum Error
EH	1.02%	23.8mV
EW	0.26%	1.1ps

spacing between differential pairs (*sp*), and characteristic impedances Z_1, Z_2. The changes made to create the different channel properties are mentioned with each case. All the four cases were trained with a DoE set of 45 designs and a test set of 100 random designs. The EH were classified into five different classes, selected to have equal representation in the DoE. All the differential S-parameters coupling power in to the victim port, SDD43 (IL), SDD44 (RL), SDD41 (FEXT), and SDD42(NEXT), sampled at 100 MHz up to the third harmonic of 7.5 GHz, were supplied to the FS algorithm. This amounts to $k = 4$ features per frequency point and a total of $N = 75$ sampled frequency points, resulting in $kN = 300$ features each with a training set size of $p = 45$. Of the 300 training sets, the FS algorithm was tuned to select the top ranking $t = 15\%$ or 45 features that showed the highest relevance to the class distribution, as defined in (5). The 45 selected features were fed into ANN of the MLP type, with 45 input nodes, 5 neurons in the hidden layer, and 1 output node [14], with hidden layer having sigmoidal activation function, as explained in (2) and output node with a linear function. Levenberg–Marquardt (LM) [26] technique with gradient descent and back-propagation was used for weight and bias training. The selected features and the prediction results are presented for each case.

A. RL Dominant Channel

To generate a channel with high amount of reflections so that the EH is largely affected by RL, the dielectric was made to be less lossy by using material with ϵ_r as 3.2, tanδ as 0.001, large Z_1, Z_2 variations with 80, 100, and 120 Ω with a trace-to-ground height (h) 8 mil and *sp* set to 6 h or 48 mil. The profiles of the differential IL, RL, FEXT, and NEXT for the DoE set are shown in Fig. 10(a) and (b), respectively.

As can be seen from the red continuous and dotted lines, RL has the highest variation across the 45 DoE designs. The distribution of the selected 45 features across the four S-parameters is shown in Table I and the frequency points picked, for EH as targets, are represented on the S-parameter curves in Fig. 11.

As expected, the majority of the features are from RL and the concentration of the features below the Nyquist frequency of 2.5 GHz indicates that band to have maximum relevance to the EH. The prediction accuracy for the ANN model is shown in Table II. The histogram of error in prediction of EH and EW is shown in Fig. 12(a) and (b), and correlation plot for simulated and predicted values is shown in Fig. 12(c) and (d).

B. IL and RL Dominant Channel

To increase the impact of IL by making the channel more lossy, the material properties were changed to ϵ_r as 4.2, tan δ as 0.022, with a reduced trace to ground height (h) 4 mil and *sp* set to 6 h or 24 mil. Z_1 and Z_2 variations were retained as in Section IV-A. The profiles of the differential IL, RL, FEXT, and NEXT for the DoE set are shown in Fig. 13(a) and (b), respectively. The selected feature distribution across S-parameters is shown in Table III and Fig. 14.

As expected, the more relevant features are from IL and the concentration of the features below the Nyquist frequency

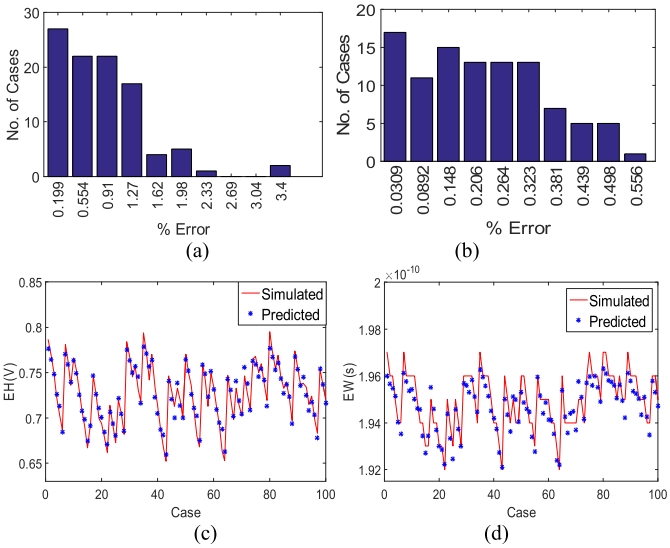


Fig. 12. Histogram of percent error distribution of (a) EH and (b) EW. Correlation between (c) EH and (d) EW.

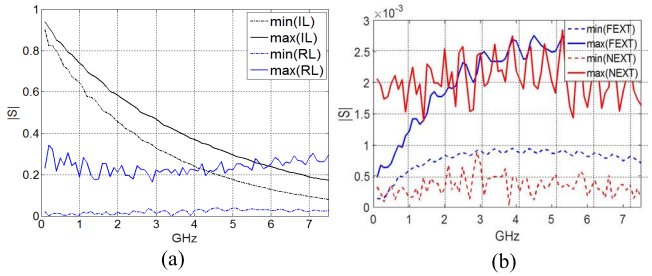


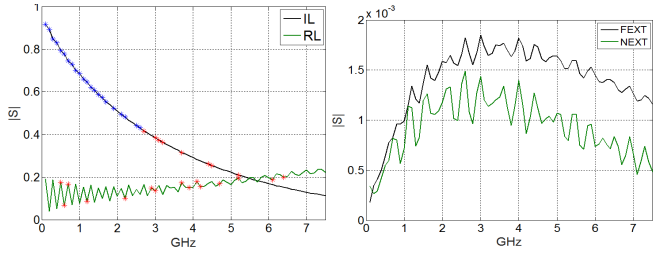
Fig. 13. Minimum–maximum profiles over different training sets for (a) IL, RL and (b) FEXT, for IL and RL dominant channel variant.

TABLE III
FEATURE DISTRIBUTION FOR IL DOMINANT CASE

Features Selected (45 of 300)	Number of points in IL	Number of points in RL	Number of points in FEXT	Number of points in NEXT
EH	30	15	0	0
EW	34	11	0	0

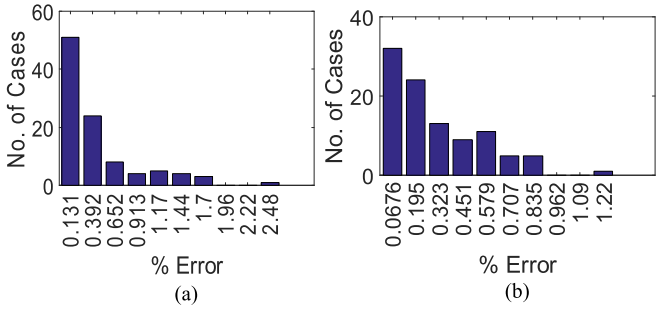
TABLE IV
PREDICTION ACCURACY FOR IL DOMINANT CASE

Metric	Prediction Error	Maximum Error
EH	0.63%	8.7mV
EW	0.39%	1.8ps



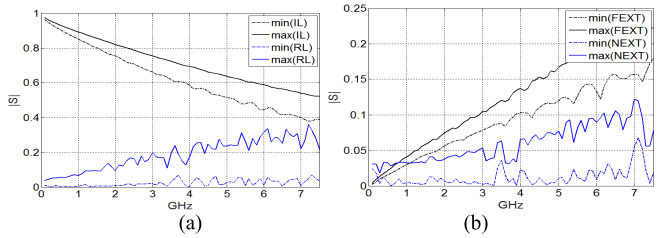
(a) (b)

Fig. 14. Distribution of selected features over differential S-parameters. (a) IL and RL. (b) FEXT and NEXT. Blue points indicate top 50 % of the 45 selected features and red points the next 50%.



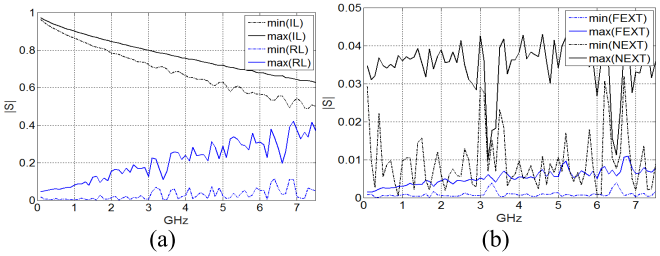
(a) (b)

Fig. 15. Histogram of percent error distribution of (a) EH and (b) EW.



(a) (b)

Fig. 16. Minimum-Maximum profiles over different training sets for MS for (a) IL, RL and (b) FEXT, NEXT for crosstalk dominant channel variant.



(a) (b)

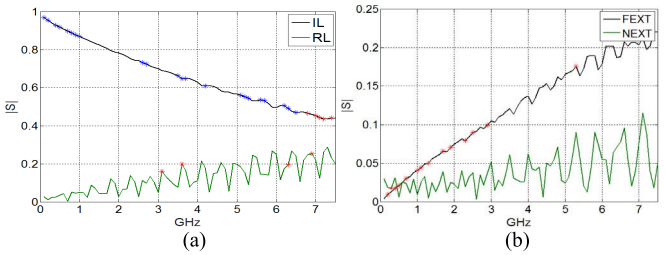
Fig. 17. Minimum-Maximum profiles over different training sets for SL for (a) IL, RL and (b) FEXT, NEXT for cross-talk dominant channel variant.

of 2.5GHz indicates that band to have maximum relevance to the EH. The prediction accuracy for the ANN model is shown in Table IV. The histogram of error in prediction of EH and EW is shown in Fig. 15(a) and (b).

C. Crosstalk Dominant Channel

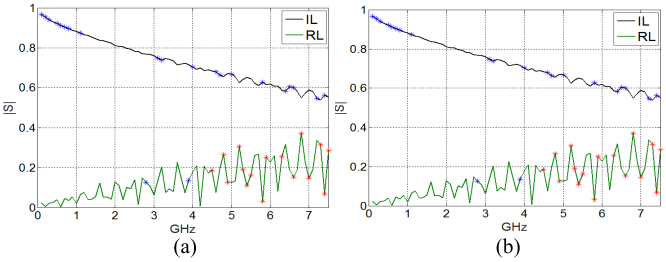
To create a channel with enhanced crosstalk and minimal IL and RL, the above channel topology and material were varied to reduce loss by setting ϵ_r as 3.2, tanδ as 0.001 and reduce reflections by setting Z_1, Z_2 variations between 98, 100, and 102 Ω. Both microstrip and stripline were used to show the comparative reduction in far end coupling. For microstrip trace, trace to ground height (h) was set to 5 mil and sp was reduced from the previous two cases, to 10 mil. For stripline trace, trace to ground height (h) was set to 9 mil and sp to 10.8 mil. The FS algorithm was applied over channels with both microstrip (MS) and stripline (SL) traces to study the impact of FEXT and NEXT. The profiles of the differential IL, RL, FEXT, and NEXT for the DoE set are shown in Figs. 16 and 17(a) and (b) for MS and SL, respectively. It can be seen that FEXT is much higher in micro-strip case. The distribution of selected features is given in Table V and Figs. 18 and 19.

It can be clearly seen from Figs. 18 and 19 that for micro-strip traces, the EH have a high correlation to FEXT and NEXT whereas for stripline the EH have almost no correlation



(a) (b)

Fig. 18. Distribution of selected features for MS over differential S-parameters. (a) IL and RL. (b) FEXT and NEXT. Blue points indicate top 50% of the 45 selected features and green points the next 50%.



(a) (b)

Fig. 19. Distribution of selected features for SL over differential S-parameters. (a) IL and RL. (b) FEXT and NEXT. Blue points indicate top 50% of the 45 selected features and blue points the next 50%.

TABLE V

FEATURE DISTRIBUTION FOR CROSSTALK DOMINANT CASE

Trace Type	Features Selected (45 of 300)	Number of points in IL	Number of points in RL	Number of points in FEXT	Number of points in NEXT
MS	EH	27	4	13	1
	EW	25	0	16	4
SL	EH	20	18	0	7
	EW	18	0	0	27

TABLE VI

PREDICTION ACCURACY FOR CROSSTALK DOMINANT CASE

Trace Type	TD Metric	Prediction Error	Maximum Error
Micro-strip	EH	0.54%	10.0mV
	EW	0.21%	1.3ps
Strip-line	EH	0.34%	5.8mV
	EW	0.21%	1.2ps

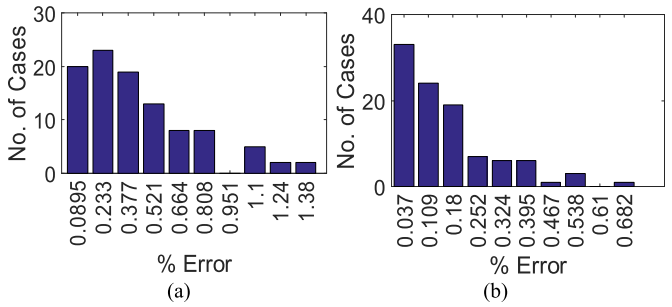


Fig. 20. Histogram of percent error distribution of (a) EH and (b) EW for microstrip.

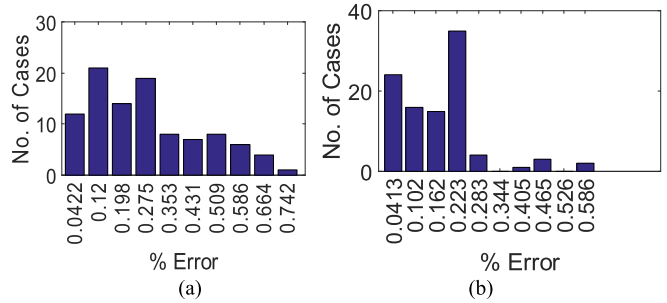


Fig. 21. Histogram of % error distribution of (a) EH and (b) EW for stripline.

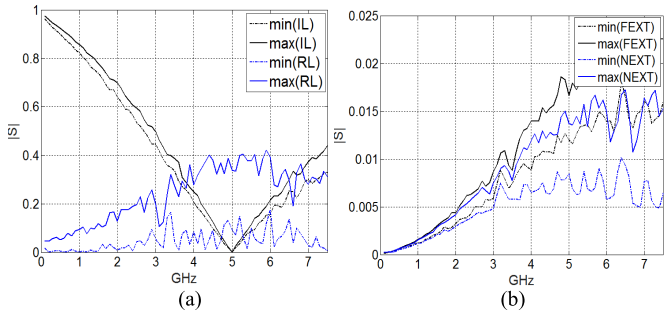


Fig. 22. Minimum-Maximum profiles over different training sets for (a) IL, RL and (b) FEXT, NEXT for CD dominant channel variant.

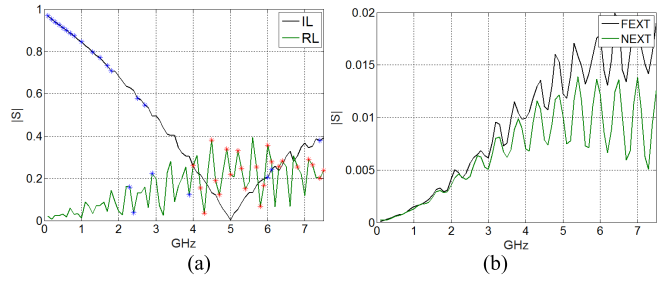


Fig. 23. Distribution of selected features over differential S-parameters. (a) IL and RL. (b) FEXT and NEXT. Blue points indicate top 50% of the 45 selected features and red points the next 50%.

to FEXT. The prediction accuracy for the ANN model is shown in Table VI. The histogram of error in prediction of EH and EW is shown in Figs. 20 and 21(a) and (b).

D. Common-Mode Dominant Channel

The above topology with stripline traces was modified from that in *C* so that *sp* is 6 h or 54 mil with material and impedance set to create minimal loss and reflections.

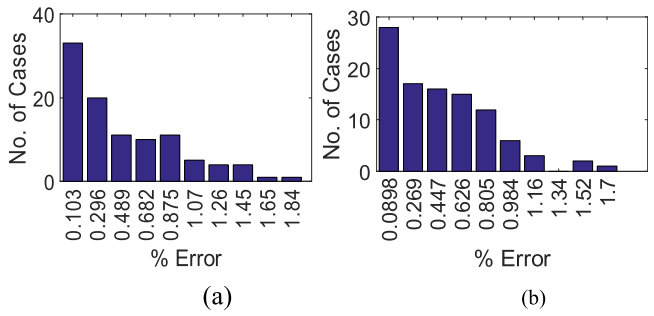


Fig. 24. Histogram of % error distribution of (a) EH and (b) EW.

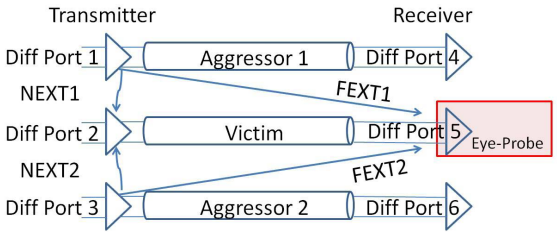


Fig. 25. Aggressor victim topology used for FD to EH/EW mapping in all numerical examples.

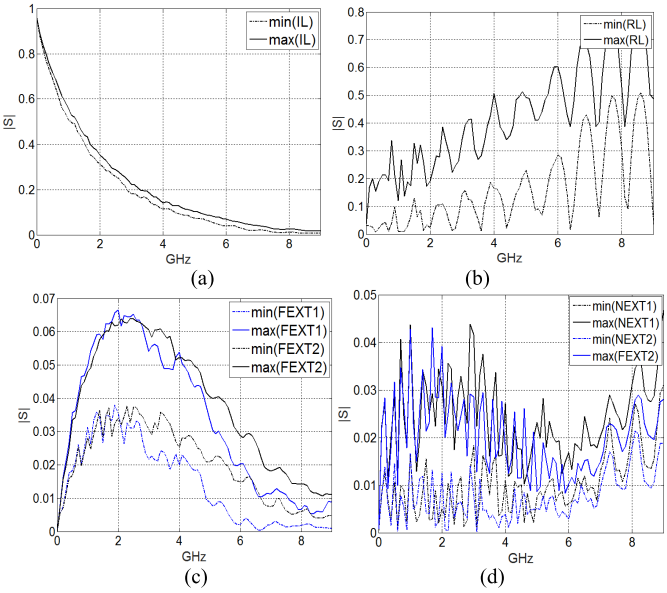


Fig. 26. Minimum–maximum profiles over different training sets for (a) IL, (b) RL, (c) FEXT1, FEXT2, and (d) NEXT1, NEXT2 for SATA 3.0.

To enhance mode conversion effect, the second trace was made longer by 0.66 in. This change is visible in the resonance in the IL, as shown in Fig. 22(a), as compared to Figs. 10, 16, and 17(a). The profiles of the differential IL, RL, FEXT, and NEXT for the DoE set are shown in Fig. 22(a) and (b), respectively. The selected feature distribution across S-parameters is shown in Table VII and Fig. 23.

The prediction accuracy for the ANN model is shown in Table VIII. The histogram of error in prediction of EH and EW is shown in Fig. 24(a) and (b).

The prediction accuracy for all the channel variants is within 1.5% and the maximum error voltage and error EW is less than 30% of the maximum variation across the test cases.

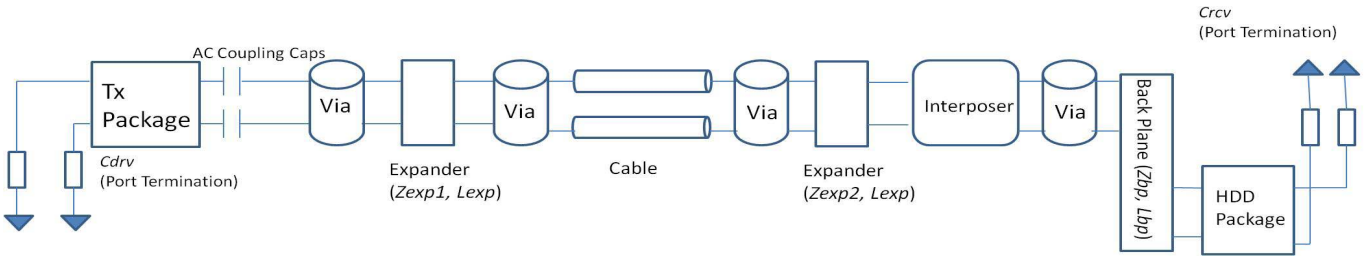


Fig. 27. SATA 3.0 topology from controller to hard disk drive (HDD) for a single differential pair.

TABLE VII
FEATURE DISTRIBUTION FOR DOMINANT COMMON-MODE CASE

Features Selected (45 of 300)	Number of points in IL	Number of points in RL	Number of points in FEXT	Number of points in NEXT
EH	18	27	0	0
EW	7	38	0	0

V. NUMERICAL RESULTS

In this section, numerical experiments are performed using the proposed FS algorithm to select relevant features that are used to generate ANN models to accurately predict EH/EW values from S-parameters. Numerical experiments are performed on two HSS interfaces, SATA 3.0 and PCIe Gen3, to quantify the performance of ANN to map S-parameters to EH/EW. The metrics for error measurement was chosen as relative L2-Norm error (ε) and maximum error (δ) as defined as follows:

$$\varepsilon(\vec{a},\vec{b}) = \sqrt{\frac{\sum_{i=1}^{N}|a_i-b_i|^2}{\sum_{i=1}^{N}|a_i|^2}} \tag{10}$$

$$\delta(\vec{a},\vec{b}) = \underset{i\in[1,N]}{\text{MAX}}\,|a_i-b_i| \tag{11}$$

where a_i are the predicted EH/EW values for the ith channel topology using the proposed mapping methodology, and b_i are the values obtained after simulating the channel in TD using a SPICE simulator. N is the size of the test set.

A. SATA 3.0

The topology shown in Fig. 27 is used in three serial-link differential pairs shown in Fig. 25. The victim port used for eye-measurements is differential port 4.The design space is formed of five variables: the driver port capacitance (C_{drv}), expander board impedances (Z_{exp1}, Z_{exp2}), back-plane impedance (Z_{bp}), and receiver port capacitance (C_{rcv}). The length of expander board (L_{exp}) is fixed to 3" and length of back-plane (L_{bp}) to 11". The differential IL, RL, FEXT1, FEXT2, NEXT1, and NEXT2 for the DoE set are shown in Fig. 26 and constitute $k = 6$ features per frequency point. The S-parameters sampled at 100 MHz from 0–9 GHz, resulting in $N = 91$ frequency samples, are fed into the FS algorithm. This leads to $kN = 546$ total number of features. As the DoE set size has 27 designs, the EH/EW are binned into four different classes to have maximum number of possible classes with six or more samples. The real valued S-parameter inputs are binned into 1000 levels. 82 (or 15% of 546) of the selected features are used to build the ANN model for FD to EH/EW mapping. Table IX and Fig. 28 show the distribution of selected features across the S-parameters

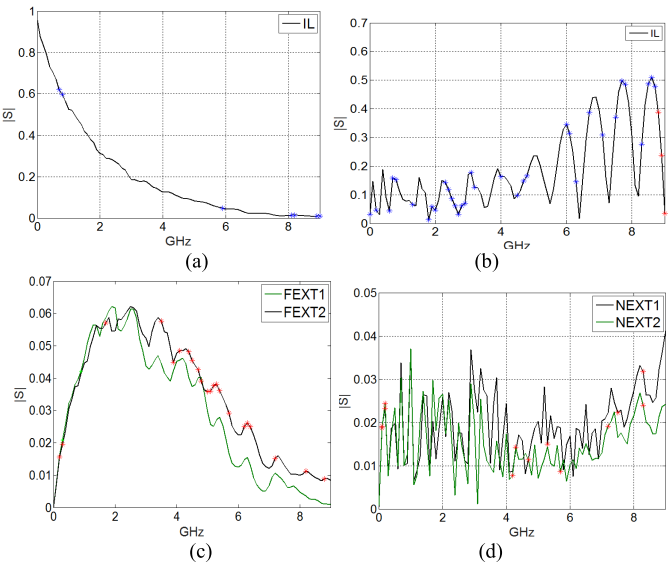


Fig. 28. Distribution of selected features over differential S-parameters. (a) IL. (b) RL. (c) FEXT1, FEXT2. (d) NEXT1, NEXT2. Blue points indicate top 50% of the 82 selected features and red points the next 50%.

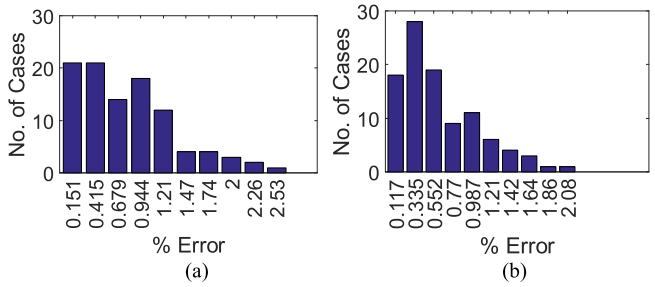


Fig. 29. Histogram of percent error distribution of (a) EH and (b) EW for SATA 3.0.

TABLE VIII
PREDICTION ACCURACY FOR DOMINANT COMMON-MODE CASE

Metric	Prediction Error	Maximum Error
EH	0.66%	15.3mV
EW	0.59%	3.2ps

The prediction accuracy for the ANN model, with 82 input nodes, 7 neurons in the hidden layer, and 1 output node, is shown in Table X. The histogram of error in prediction of EH and EW for 100 random test cases is shown in Fig. 29(a) and (b).

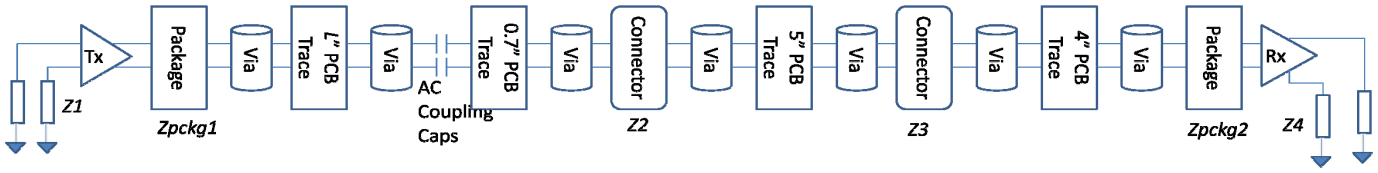


Fig. 30. PCIe Gen3 topology for a single differential pair.

TABLE IX
FEATURE DISTRIBUTION FOR SATA 3.0

Features Selected (82 of 546)	No. IL	No. RL	No. FEXT1	No. FEXT2	No. NEXT1	No. NEXT2
EH	7	37	2	22	5	9
EW	6	55	5	6	4	6

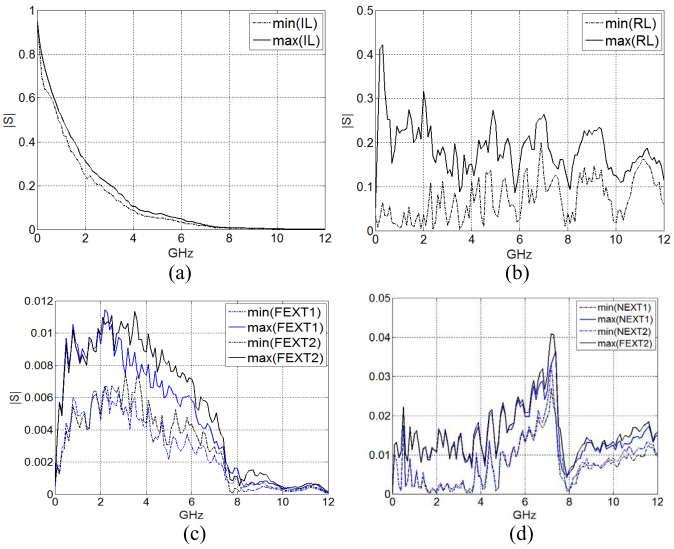


Fig. 31. Minimum-Maximum profiles over different training sets for (a) IL, (b) RL, (b) FEXT1, FEXT2, and (c) NEXT1, NEXT2 for PCIe Gen 3.

B. PCIe Gen 3

A PCIe Gen 3 channel with the topology shown in Fig. 30 is considered for the second set of experiments. The design space consists of six variables *Zpckg1*, *Zpckg2*, being the package impedances, *Z1*, *Z4* being the termination impedances and *Z2* and *Z3* being the connector impedances, as marked in the topology with *L* set to 8.9". PCIe topology involves receiver equalization that is not needed for SATA 3.0. For the experiments here, a three-tap feed-forward equalizer with a DC bias were used at the transmitter and a single tap decision feedback equalizer (DFE) was used at the receiver end. The equalizer coefficients were allowed to adapt to the channel response. The DoE training set consists of $p = 45$ designs. The test set consists of 100 randomly selected designs. The differential IL, RL, FEXT1, FEXT2, NEXT1, and NEXT2 for the DoE set are shown in Fig. 31 at 100-MHz step size from 0 to 12 GHz. This constitutes $k = 6$ features per frequency point, $N = 121$ sampled frequency points and $kN = 726$ total number of features. The S-parameters are binned in 1000 values to retain maximum information. The EH is distributed into five different classes.

The FS algorithm selects 109 (15%) of the 726 input features which act as input to the ANN. Because of adaptive

TABLE X
PREDICTION ACCURACY FOR SATA 3.0

Metric	Prediction Error	Maximum Error
EH	0.97%	4.1mV
EW	0.74%	1.69ps

TABLE XI
FEATURE DISTRIBUTION FOR PCIe GEN 3 CONSTANT EQUALIZATION

Features Selected (109 of 726)	No .IL	No. RL	No. FEXT1	No. FEXT2	No. NEXT1	No. NEXT2
EH	26	65	0	1	9	8
EW	14	73	0	0	8	14

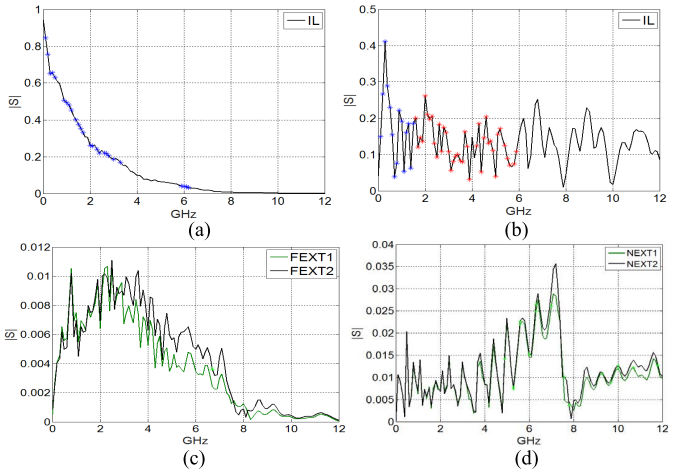


Fig. 32. Distribution of selected features over differential S-parameters. (a) IL. (b) RL. (c) FEXT1, FEXT2. (d) NEXT1, NEXT2. Blue points indicate top 50% of the 109 selected features and green points the next 50%.

TABLE XII
PREDICTION ACCURACY FOR PCIe GEN 3 WITH CONSTANT EQUALIZATION

Metric	Prediction Error	Maximum Error
EH	2.12%	4.02mV
EW	1.74%	2.74ps

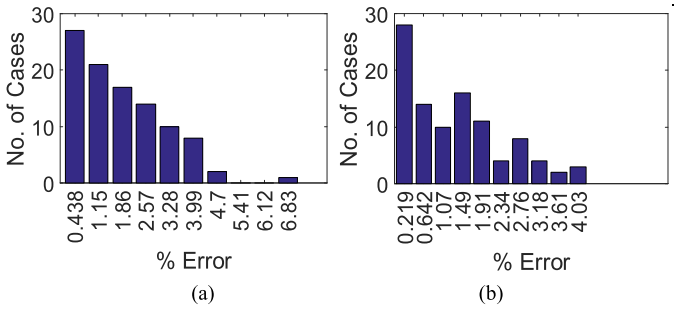


Fig. 33. Histogram of percent error distribution of (a) EH and (b) EW for PCIe Gen 3 with constant equalization.

equalization process, the ANN is also supplied with the equalization data for each case. This consists of the DFE tap coefficients, the transmitter equalization coefficients, and the dc bias value. The ANN used is of MLP type with 114 input nodes, one hidden layer formed of eight neurons having sigmoid activation function and one output neuron with linear activation function. The distribution of the selected features is shown in Table XI and Fig. 32. The prediction accuracy is tabulated in Table XII. The histogram of error in prediction of EH and EW is shown in Fig. 33(a) and (b).

VI. Conclusion

With the growing complexity of channel topologies and increasing bit rates for communication a full-factorial sweep over design variables using explicit time-domain simulations for channel characterization become prohibitive. On the other hand, DoE-based methods can sometimes provide inaccurate sensitivity analysis. In this paper, a feature selection method is proposed based on an information theoretic framework to identify the relevant FD S-parameters metrics that have considerable impact on the EH/EW of the channel. This is achieved by correlating underlying probability densities of the sampled S-parameters at discrete frequencies to the probability densities of the EH/EWs. The selected features are fed into an FD to EH/EW mapping process using ANN models. Therefore, the FS algorithm provides a way to prevent undersampling or oversampling of data, resulting in compact accurate and cost-effective ANN models with only relevant inputs, without any requirement of channel interface expertise. The state-of-the-art interfaces, SATA 3.0 and PCIe Gen 3, have been studied in Section V to demonstrate that the method works for complex high-speed topologies, and it helps achieve prediction accuracies within 2%.

References

[1] R. Achar and M. S. Nakhla, "Simulation of high-speed interconnects," *Proc. IEEE*, vol. 89, no. 5, pp. 693–728, May 2001.

[2] W. T. Beyene and J. Schutt-Aine, "Accurate frequency-domain modeling and efficient circuit simulation of high-speed packaging interconnects," *IEEE Trans. Microw. Theory Techn.*, vol. 45, no. 10, pp. 1941–1947, Oct. 1997.

[3] P. Triverio, S. Grivet-Talocia, M. S. Nakhla, F. G. Canavero, and R. Achar, "Stability, causality, and passivity in electrical interconnect models," *IEEE Trans. Adv. Packag.*, vol. 30, no. 4, pp. 795–808, Nov. 2007.

[4] S. Grivet-Talocia and B. Gustavsen, *Passive Macromodeling: Theory and Applications*, vol. 239. Hoboken, NJ, USA: Wiley, 2015.

[5] B. K. Casper, M. Haycock, and R. Mooney, "An accurate and efficient analysis method for multi-Gb/s chip-to-chip signaling schemes," in *Symp. VLSI Circuits Dig. Tech. Papers*, Honolulu, HI, USA, Jun. 2002, pp. 54–57.

[6] V. Stojanovic and M. Horowitz, "Modeling and analysis of high-speed links," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2003, pp. 589–594.

[7] K. Xiao, B. Lee, and X. Ye, "A flexible and efficient bit error rate simulation method for high-speed differential link analysis using time-domain interpolation and superposition," in *Proc. IEEE Int. Symp. Electromagn. Compat.*, Detroit, MI, USA, Aug. 2008, pp. 1–6.

[8] D. Oh, J. Ren, and S. Chang, "Hybrid statistical link simulation technique," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 1, no. 5, pp. 772–783, May 2011.

[9] Z. Chen, W. D. Becker, and G. Katopis, "A new approach to deriving packaging system statistical eye diagram based on parallel non-linear transient simulations using multiple short signal bit patterns," in *Proc. 62nd IEEE Electron. Compon. Technol. Conf. (ECTC)*, San Diego, CA, USA, May/Jun. 2012, pp. 160–167.

[10] K. Rekab and M. Shaikh, *Statistical Design of Experiments With Engineering Applications*. Boca Raton, FL, USA: CRC Press, 2005.

[11] B. Mutnury, F. Paglia, M. Wang, G. K. Singh, and A. C. Scogna, "Modeling and characterization of high speed interfaces in blade and rack servers using response surface model," in *Proc. IEEE Electron. Compon. Technol. Conf.*, May/Jun. 2011, pp. 183–190.

[12] W. T. Beyene, "Application of artificial neural networks to statistical analysis and nonlinear modeling of high-speed interconnect systems," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 26, no. 1, pp. 166–176, Jan. 2007.

[13] R. Mellitz, V. Ragavassamy, and M. Brownell, "Fast, accurate, 'simulation-lite' approach for multi-GHz board design," in *Proc. DesignCon*, 2011.

[14] N. Ambasana, G. Anand, B. Mutnury, and D. Gope, "Eye height/width prediction from *S*-parameters using learning-based models," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 6, no. 6, pp. 873–885, Jun. 2016.

[15] J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers, "Space mapping technique for electromagnetic optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 42, no. 12, pp. 2536–2544, Dec. 1994.

[16] M. H. Bakr, J. W. Bandler, K. Madsen, and J. Søndergaard, "An introduction to the space mapping technique," *Optim. Eng.*, vol. 2, no. 4, pp. 369–384, Dec. 2001.

[17] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq, "A surrogate modeling and adaptive sampling toolbox for computer based design," *J. Mach. Learn. Res.*, vol. 11, pp. 2051–2055, Jul. 2010.

[18] N. Ambasana, D. Gope, B. Mutnury, and G. Anand, "Automated frequency selection for machine-learning based EH/EW prediction from S-parameters," in *Proc. 24th IEEE Elect. Perform. Electron. Packag. Syst. (EPEPS)*, San Jose, CA, USA, Oct. 2015, pp. 53–56.

[19] L. Yu and H. Liu, "Feature selection for high-dimensional data: Fast correlation-based filter solution," in *Proc. Int. Conf. Mach. Learn.*, Washington, DC, USA, 2003, pp. 1–8.

[20] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.

[21] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[22] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[23] H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: An enabling technique," *Data Mining Knowl. Discovery*, vol. 6, no. 4, pp. 393–423, Oct. 2002.

[24] K. S. Balagani and V. V. Phoha, "On the feature selection criterion based on an approximation of multidimensional mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1342–1343, Jul. 2010.

[25] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 27–66, Jan. 2012.

[26] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.

Nikita Ambasana received the B.E. degree in electronics and communication engineering from the Dharmsinh Desai Institute of Technology, Dharamsinh Desai University, Nadiad, India, in 2009, and the M.Tech. degree in information and communication technology from the Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India, in 2012. She is currently pursuing the Ph.D. degree with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India.

Her current research interests include methods of using neural networks for faster design space analysis for signal integrity of high-speed channels and sensitivity analysis for power integrity using 2.5D solvers.

Gowri Anand received the M.Sc. degree in electrical engineering from the Arizona State University, Tempe, AZ, USA, in 2002.

She was the Storage and Server Signal Integrity Lead for SATA technologies with Dell Inc., Bengaluru, India, where she is currently a Senior Electrical Engineer with the Enterprise PG Group, with a focus on analyzing SATA channels in rack and blade server systems. She has over six years of experience in electrical modeling and simulation of package designs. She has four IEEE conference papers and holds two patents at Dell Inc.

Dipanjan Gope (M'05–SM'12) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology (IIT) Kharagpur, Kharagpur, India, in 2000, and the M.S. and Ph.D. degrees in electrical engineering from the University of Washington, Seattle, WA, USA, in 2003 and 2005, respectively.

He was a Senior CAD Engineer with Intel, Santa Clara, CA, USA, from 2005 to 2007. He was a Founding Member of Nimbic, Inc., Mountain View, CA, USA, a start-up pioneering in cloud-based 3-D full-wave electromagnetic simulation and acquired by Mentor Graphics, Inc., Wilsonville, OR, USA, in 2014, where he served as the Vice President of Research and Development from 2007 to 2011 and from 2013 to 2014. As part of the Ph.D. dissertation, he co-developed the PILOT technology for fast circuit-electromagnetic simulation, now licensed by the University of Washington. He has been an Assistant Professor with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India, since 2011. He has authored over 70 journal and conference publications. His current research interests include computational electromagnetics with applications in signal integrity, power integrity, electromagnetic interference for high-speed chip-package-systems, fast solver algorithms, radio-frequency sensing, parallel programming for many-core, and cloud computing.

Dr. Gope received the Divisional Award from Intel for contributions toward providing signal integrity solution to Intel architecture platforms. He has served on the Technical Program Committee of many international conferences, including the IEEE Design Automation Conference, the IEEE International Conference on Computer-Aided Design, the IEEE Design, Automation and Test in Europe Conference and Exhibition, and the IEEE Asia and South Pacific Design Automation Conference. He was the General Co-Chair of the 2014 IEEE Electrical Design of Advanced Packaging and Systems.

Bhyrav Mutnury (S'03–M'05–SM'11) received the M.Sc. degree in electrical engineering and the Doctor of Philosophy degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002 and 2005, respectively.

From 2000 to 2005, he was a Research Assistant with the Department of Electrical and Computer Engineering, Georgia Institute of Technology. His work was focused on modeling and analyzing high-speed power distribution networks and macromodeling of active circuits to cut down simulation time by 10–1000×. He was with the IBM USA, Austin, TX, USA, from 2005 to 2010, where he created an electrical design, modeling, and analysis algorithms that can reduce the modeling, simulation, and analysis of electrical interfaces anywhere from 100–1000×. He is currently a Distinguished Engineer and the Global Team Lead with the Enterprise Signal Integrity Group, Dell Inc., Austin, where he is responsible for storage, network, rack, and blade server designs. He is also driving the next-generation high-speed interfaces and modeling methodologies at Dell Inc. His innovative design of experiments-based methodologies and integrated tool suite to model high-speed SerDes for electrical and physical designs have provided huge improvement in productivity. He is the Electrical Interface Expert on GbE, XAUI, FC, USB, PCIe, QPI, DDR, and FBD. He has over 15 years of progressive experience in system design with strong focus on electrical modeling, analysis, and optimization of complex high-speed servers. He was involved in numerous research projects with the Georgia Institute of Technology, Atlanta, GA, USA, the Missouri Institute of Science and Technology, Rolla, MO, USA, and Pennsylvania State University, State College, PA, USA. He has authored or coauthored over 55 refereed publications in various IEEE and non-IEEE conferences. These publications covered various disciplines, including package and interconnect modeling, analysis, and optimization, active circuit and transistor-level circuit macromodeling, high-speed serial and multidrop interface design and modeling, and services science and management engineering. He has filed over 105 invention disclosures in the fields of electrical cable design, package and printed circuit board design, and electrical design space exploration.

Dr. Mutnury was a Secretary of the IEEE Components, Packaging, and Manufacturing Technology Austin Chapter from 2007 to 2012.