SI/PI-Database of PCB-Based Interconnects for Machine Learning Applications

MORTEN SCHIERHOLZ[ID][1], (Graduate Student Member, IEEE), **ALLAN SÁNCHEZ-MASÍS**[2],
ALLAN CARMONA-CRUZ[2], (Student Member, IEEE), **XIAOMIN DUAN**[3], (Member, IEEE),
KALLOL ROY[ID][4], (Member, IEEE), **CHENG YANG**[ID][1], (Member, IEEE),
RENATO RIMOLO-DONADIO[ID][2], (Senior Member, IEEE),
AND CHRISTIAN SCHUSTER[ID][1], (Senior Member, IEEE)

[1]Institut für Theoretische Elektrotechnik, Hamburg University of Technology, 21073 Hamburg, Germany
[2]Department of Electronics Engineering, Instituto Tecnológico de Costa Rica, Cartago 30101, Costa Rica
[3]IBM Germany Research & Development GmbH, 71032 Böblingen, Germany
[4]Institute of Computer Science, University of Tartu, 51009 Tartu, Estonia

Corresponding author: Morten Schierholz (morten.schierholz@tuhh.de)

ABSTRACT A database is presented that allows the investigation of machine learning (ML) tools and techniques in the signal integrity (SI), power integrity (PI), and electromagnetic compatibility (EMC) domains. The database contains different types of printed circuit board (PCB)-based interconnects and corresponding frequency domain data from a physics-based (PB) tool and represent multiple electromagnetic (EM) aspects to SI and PI optimization. The interconnects have been used in the past by the authors to investigate ML techniques in SI and PI. However, many more tools and techniques can be developed and applied to these structures. The setup of the database, its data sets, and examples on how to apply ML techniques to the data will be discussed in detail. Overall 78 961 variations of interconnects are presented. By making this database available we invite other researchers to apply and customize their ML techniques using our results. This provides the possibility to accelerate ML research in EMC engineering without the need to generate expensive data.

INDEX TERMS Artificial neural network, database, electromagnetic compatibility, machine learning, power integrity, signal integrity.

I. INTRODUCTION

Numerical modeling and prediction of electromagnetic compatibility (EMC) related problems in engineering nowadays rely on a large variety of physics-based (PB) approaches and corresponding software tools [1], [2]. The approaches range from very efficient, like equivalent circuits, to more accurate approaches such as full-wave solvers based on e.g. the finite element method, the finite difference time domain method, and the method of moments [7]–[11], as shown in Fig. 1. SI and PI are two important aspects regarding the electromagnetic compatible design of printed circuit board (PCB)-based interconnects such as signal links between components, and the power delivery network (PDN). Specifically, SI engineering has to make sure that signals will be transmitted and

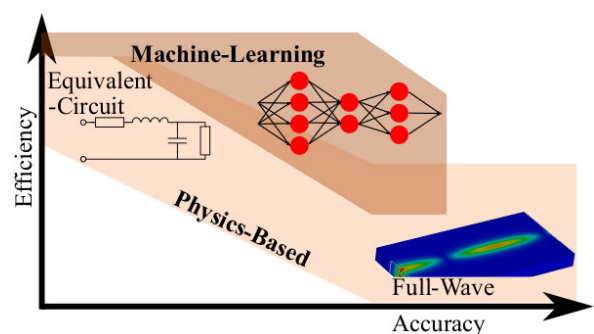The associate editor coordinating the review of this manuscript and approving it for publication was Joanna Kolodziej[ID].

FIGURE 1. Overview of efficiency vs. accuracy of physics-based (PB) and machine learning (ML) approaches.

detected adequately despite of losses and reflections, and PI engineering has to make sure that a constant voltage supply will be maintained despite ground bounce and simultaneous switching noise. The increasing complexity of problems
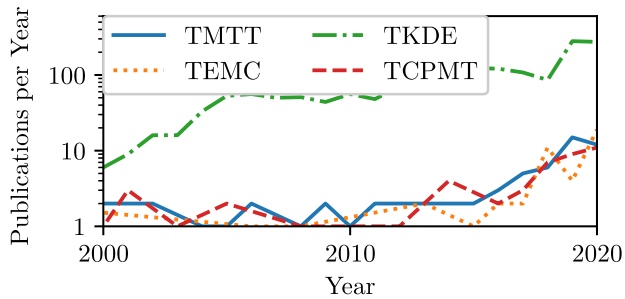
FIGURE 2. Publications found for search term "machine learning" in the Transactions on Microwave Theory and Techniques (TMTT), the Transactions on Electromagnetic Compatibility (TEMC), the Transactions on Components, Packaging and Manufacturing Technology (TCPMT) and the Transactions on Knowledge and Data Engineering (TKDE).

related to SI and PI of PCB-based interconnects demand constantly tools and techniques with capabilities at higher efficiency and higher accuracy [12]–[15].

Recently it can be observed that machine learning (ML) tools and techniques are increasingly used in the EMC domain either to improve PB approaches or to replace them [16]. The generally increasing interest in ML tools and techniques is also visible in the number of pertinent publications in the Transactions on Microwave Theory and Techniques (TMTT), the Transactions on Electromagnetic Compatibility (TEMC), and the Transactions on Components, Packaging and Manufacturing Technology (TCPMT) [17]–[24], as shown in Fig. 2. For comparison, the trend in Transactions on Knowledge and Data Engineering (TKDE) is shown, which focuses on computer science, artificial intelligence, and computer engineering topics. For the TKDE multiple times more publications are found but all Transactions share the same trend.

It is well known that in information technology ML tools and techniques have solved problems that previously were difficult to solve with standard algorithms [25]. Similarly, the hope is that ML tools and techniques will improve modeling for electrical engineering applications [16]. The existing state of the art in information technology and data science comprises a large spectrum of ML tools and techniques such as deep neural networks, long short-term memory with artificial neural networks (ANNs), recurrent neural networks, support vector machines, decision trees, nearest neighbor algorithms, Bayesian classifiers, etc.. However, these cannot be easily applied to EMC engineering problems and an adaption to the specific requirements of SI and PI is usually needed [26]–[29]. Mostly this is due to the complex electromagnetic behavior that SI, PI, and EMC problems show, and the difficulty to categorize and describe their three-dimensional nature consisting of a wide variety of different components and structures.

Besides tools and techniques, data is fundamentally important for ML investigations. In information technology data is commonly shared, e.g. large image sets are provided by the ImageNet or the Open Image Dataset [30], [31]. A database

in medical engineering is PhysioNet which provides medical data and software in an open access or restricted access format [32]. An overview of datasets for the visual object class challenge is given in [33]. In contrast, in the electrical engineering community publicly sharing of simulation data is not as common, which is understandable due to the cost of creating the data with PB approaches and intellectual property issues. Also, restrictions by licensing of simulation tools have to be taken into account possibly limiting their usage. Nevertheless, throughout different engineering domains the importance of sharing data is more and more recognized, e.g. the European Union is planning to implement the European Open Science Cloud which shall provide the distribution and collection of research data [34].

To help accelerate the research in the domain of ML in combination with EMC engineering problems we present here the first publicly available database for training of custom made ML tools and techniques. The database contains different types of simulation results of structures which represent SI and PI problems. Details of the database are presented in Sec. II. The structures are presented in Sec. III. To give an impression what kind of ML techniques can be used on the data, use cases are given in Sec. IV. These examples mainly consist of ANNs which predict scattering parameters or are used for classification. In Sec. V conclusions and possible future directions are given.

II. SI/PI DATABASE

Key questions for the SI/PI-Database design were how to provide simple access to the data and how to ensure that data provided are valuable and can be validated by users. Other publicly available databases for the investigations of ML techniques were used as reference for our work, e.g. the Open Image dataset [31]. The Open Image dataset is one of the largest organized image datasets with more than 9 million images and more than 7 million images labeled [35]. It is possible to download images with the corresponding image labels from the Open Image Dataset either as a complete dataset or individual subsets. Inspired by this we decided to set up the SI/PI-Database in a way where data sets for specific interconnects can be downloaded individually and with a documentation that is comprehensive enough to allow reproduction of the data by the user using his or her own tool of preference.

A. ACCESS AND DATA FORMAT

The SI/PI-Database is accessible at the homepage of the Institut für Theoretische Elektrotechnik at the Hamburg University of Technology (TUHH) at

www.tet.tuhh.de/en/si − pi − database

To access the required data first one of the PCB interconnects shown in Tab. 1 is chosen on the website. Then by submitting an e-mail address the download link for the zip-archive is provided.

TABLE 1. Overview of PCB-based interconnects and parameter variations thereof available in the SI/PI-Database.

Sec.	Interconnect	Parameter Variations	No. of Cavities	No. of Ports	Frequency Range	Max. el. Length	No. of Simulations	Size of Data Set	Image
III-A	PWR/GND Plane PCB with 11x11 Via-Array [3], [4]	Nr. Capacitor Capacitance Cavity Height Permittivity	1	2	Start: 1 MHz Stop: 1 GHz Step: 3.33 MHz	2 λ	58 399	1.44 GB	
III-B	Link on 11 Cavity PCB with two 10x10 Via-Arrays [5]	Via Pitch Via Rad. Antipad Rad. Permittivity Loss tangent Cavity Height Trace Length	11	12	Start: 0.5 GHz Stop: 100 GHz Step: 0.5 GHz	79 λ	7031	2.56 GB	
III-B	Link on 8 Cavity PCB with two 10x10 Via-Arrays [5]	Via Pitch Via Rad. Antipad Rad. Permittivity Loss tangent Cavity Height Trace Length	8	12	Start: 0.5 GHz Stop: 100 GHz Step: 0.5 GHz	79 λ	7031	2.65 GB	
III-C	5x5 Via-Array on 10 Cavity PCB [6]	Antipad Rad. Pad Rad. Via Pitch Cavity Height Plate Thickness Permittivity Loss Tangent	10	34	Start: 1 GHz Stop: 40 GHz Step: 0.196 GHz	2.6 λ	5000	50.4 GB	
III-D	Link on 10 Cavity PCB with two 5x5 Via-Arrays [6]	Antipad Rad. Pad Rad. Via Pitch Cavity Height Permittivity Loss tangent Array Distance Trace Width	10	68	Start: 1 GHz Stop: 40 GHz Step: 0.196 GHz	59 λ	1500	57.2 GB	

The zip-archive has all necessary information for the chosen PCB structure, as shown in Fig. 3. Included are a *<description.pdf>* file, a *<parameter.csv>* file, and a *<variation>* folder. The *<variation>* folder contains all simulation results for the variations of the PCB structure. The simulation results are stored as single-ended (SE) scattering parameters in ASCII Touchstone® file format according to [36]. Many tools and applications are able to import the Touchstone® format. All Touchstone® files are named *<simu_[index].sNp>*. With the [index] it is possible to retrieve the parameter variations that have been used for the specific simulation from the *<parameter.csv>* file.

B. USAGE AND VALIDATION

Correct usage of the data is based on the *<parameter.csv>* file. This file has multiple rows and columns. The first row has the names of the columns e.g. *via_radius*, *via_pitch*. Starting with the second row, each row has the information for one specific variation of the PCB structure, e.g. *via_radius* = 5 mil, *via_pitch* = 80 mil. The last column is the [index] to connect the parameter variations with the scattering parameter in the *<variation>* folder, as shown in Fig. 3.

To validate the simulation results all required geometry and material parameters are provided in the file *<description.pdf>*. With the information it is possible to

recreate the interconnect with a modeling tool of own choice to model and re-simulate with a suitable full-wave solver. The results of the re-simulated interconnect can be compared against the simulation results provided in the SI/PI-Database. This makes it possible to check whether the [index] of the *<parameter.csv>* file is correctly assigned to the Touchstone®-file, and whether the simulations for creating the data for the SI/PI-Database in the fist place shows results of sufficient accuracy.

III. DESCRIPTION OF PCB-BASED INTERCONNECTS

Five different interconnects are currently part of the SI/PI-Database. These structures provide different functionalities and allow the investigation of different problems in SI and PI. All interconnects are based on PCBs and include vias, via-arrays and striplines. An overview is given in Tab. 1. All PCB-based structures have ports placed directly on the top, or bottom side of the stackup. Specifically, a port is defined by the area covered by the via barrel and the antipad (via hole) surrounding it in either the top most or bottom metallic plane of the PCB, as shown in Fig. 4 (b). This coaxial-like structure allows the fundamental mode to be used for excitation. The transmission, reflection and crosstalk between the ports are modeled with a PB approach [37]–[40].
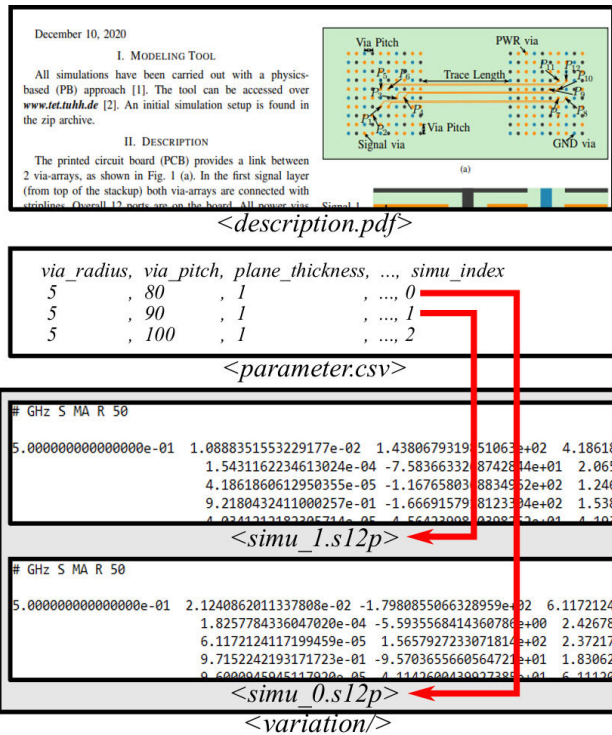
FIGURE 3. Overview of the files that are found for each PCB-based structure. *<description.pdf>* provides general information of the geometry and material setup, including the applied simulation tool. *<parameter.csv>* has one row for each variation of the structure. Column *simu_index* is the index to find the according Touchstone® file in the *<variation/>* folder e.g. Touchstone®-file *<simu_0.s12p>* was simulated with a *via_radius* of 5 mil, *via_pitch* of 80 mil, and *plane_thickness* of 1 mil.
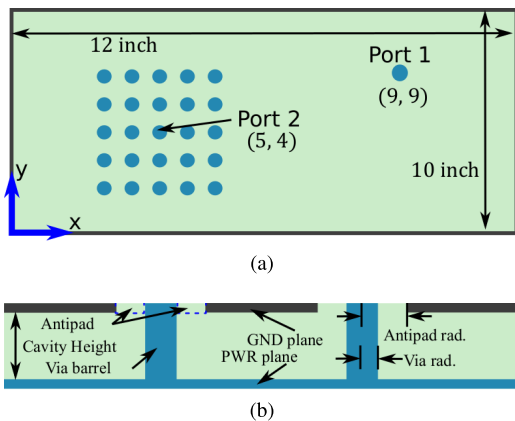

FIGURE 4. PWR/GND plane PCB with via array from Sec. III-A. For clarity the size of the via array is increased and the number of vias is reduced. The center of the via array is the port under investigation. All dimensions are in inch. (a) Top view, (b) cross section.

A. PWR/GND PLANE PCB WITH 11 × 11 VIA-ARRAY

This structure consists of a PDN on a PCB with one cavity, a power plane on the bottom and a ground plane on top, as shown in Fig. 4. A more detailed description is given in [3], [4]. Port 2 is in the center of a via-array and is the port that is decoupled by placing decoupling capacitors (decaps) on the surrounding via-array. At this port the impedance of the PDN is observed and the dependence on the configuration of
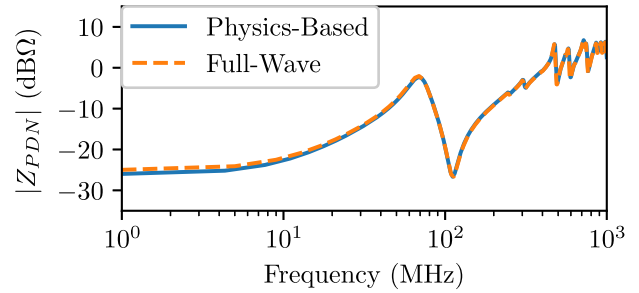

FIGURE 5. Comparison of a physics-based approach vs. a full-wave solver (finite element method) for the structure in Fig. 4 without any decaps. Port 1 is terminated with a 50 mΩ resistance.
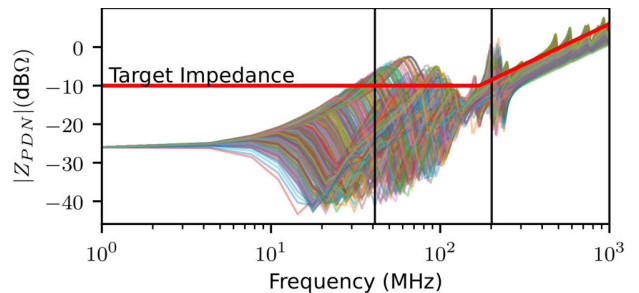

FIGURE 6. Impedance profiles for 600 simulations of the power delivery network in Fig. 4 with different decap distributions and one cavity.

decaps is analyzed. The variations for this structure include changing the distance of the planes, the material properties, and the decaps. For the decaps the location, capacitance, and number was varied. The capacitors were chosen from a library containing overall 15 values including equivalent series resistance and equivalent series inductance. A comparison of the setup without any decaps was performed with a full-wave solver with good agreement of both modeling tools [41], as shown in Fig. 5. The variety of the impact of decap distributions on the PDN impedance in the frequency range from 1 MHz to 1 GHz is illustrated in Fig. 6 for 150 different decap distributions of the 58 399 total variations.

B. LINKS ON PCB WITH TWO 10 × 10 VIA-ARRAYS

Via-arrays on PCBs are commonly used to provide an electrical connection between components placed on top or bottom of the PCB and structures on inner layers, such as striplines and power or ground planes. In the following we refer to the term "link" when there is full connectivity between two distant ports on a PCB consisting of at least two vias and one stripline thus enabling signal transmission. In [5] two links were concatenated to produce results for a setup that consists of two PCB-based interconnects. The setup is shown in Fig. 7 (a), with the backplane and daughtercard connected. The connector in Fig. 7 (a) is not included in the simulation setup, but shown here for illustrative purposes. For the modeling the daughtercard and backplane were treated separately, as shown in Figs. 8 (a) and (b). Both stackups contain two via-arrays with the connection of planes, vias, and striplines, as shown in Fig. 7 (b). The links between
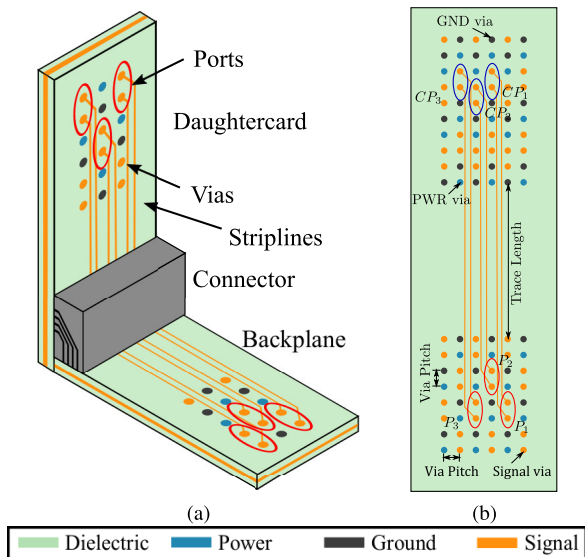
FIGURE 7. Setup of the connected daughtercard and backplane as described in Sec. III-B. (a) 3D view including the connector, (b) view of the PCB showing the ports for the differential signaling, (c) legend for the color coding. Images adapted from [5].
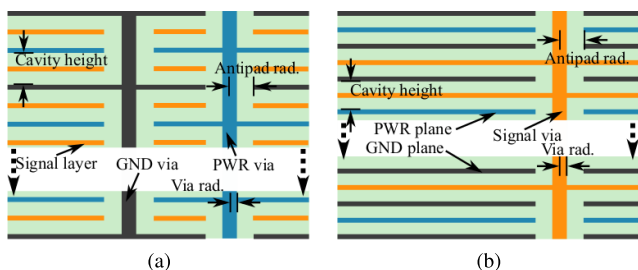

FIGURE 8. Setup of the daughtercard and backplane as described in Sec. III-B. (a) and (b) cross section view of the backplane and daughtercard respectively. The daughtercard has 8 cavities, the backplane has 11 cavities. Images adapted from [5].
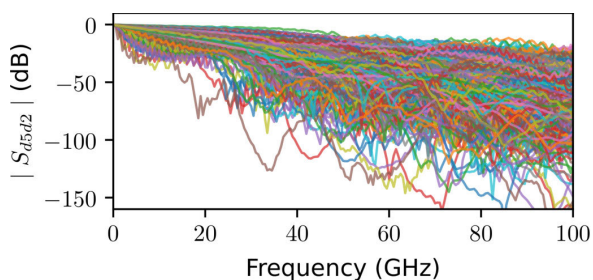

FIGURE 9. Examples for transmission of the DIFF link for the concatenation of the structures described in Sec. III-B for 1310 geometry variations.

the via-arrays are designed for differential (DIFF) signaling. Since the simulation is based on SE ports the DIFF scattering parameters were obtained in a postprocessing step. Backplane and daughtercard each have 12 SE ports. Overall 7031 design variations were simulated from $0.5\,\text{GHz}$ to $100\,\text{GHz}$. The impact on the transmission of DIFF ports of the concatenated backplane and daughtercard is shown in Fig. 9 for 1031 variations.
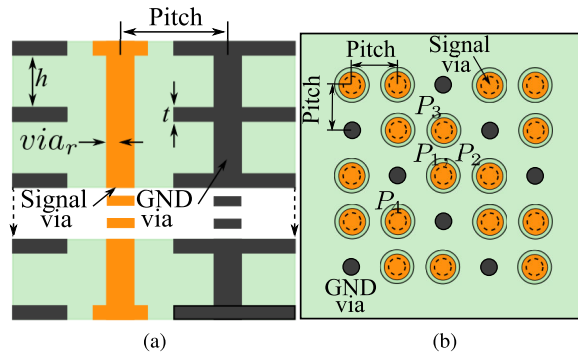

FIGURE 10. Structure of the setup discussed in Sec. III-C. The cross section and top view are shown in (a) and (b) respectively. P_j is the port j at the end of the via. For example P_1 is at the top, P_2 is at the same via at the bottom, P_3 is on another via at top.
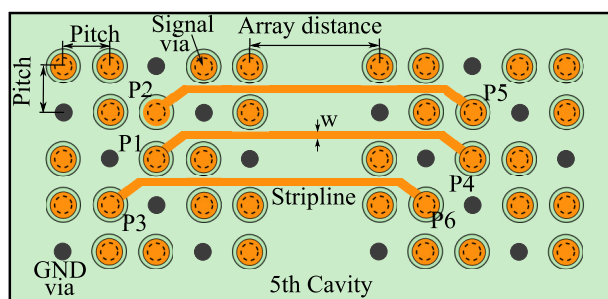

FIGURE 11. Top view of the two 5 × 5 via-arrays in the fifth cavity as outlined in Sec. III-D. The ports P_j are on top of the stackup. No ports are at the bottom.

C. 5 × 5 VIA-ARRAY ON 10 CAVITY PCB

The 5 × 5 via array is located on a 10 cavity PCB, as shown in Fig 10. The connectivity of GND and signal vias is shown in Fig 10 (a), with a via ratio of 2 : 1 for signal and GND, respectively. The ports P_1, P_3 and P_4 are placed on the top side of the PCB, P_2 at the bottom side of the vias, as shown in Fig. 10 (b). A more thoroughly description is given in [6]. All planes of the stackup are connected by GND vias. The radius of the via barrel is smaller than the radius of the via pads on top and bottom of the stackup, as shown in Fig. 10 (a). Further variations of the geometry are stated in Tab. 1. The discontinuity of the vias and variations of the geometry have an impact on the scattering parameters. Overall 5000 variations were simulated.

D. LINK ON 10 CAVITY PCB WITH TWO 5 × 5 VIA-ARRAYS

The interconnects consist of a link between two 5 × 5 via-arrays. Both arrays are located on the same multilayer stackup of 10 dielectric cavities. In the fifth cavity signal vias of the first array are connected by striplines to the corresponding signal vias of the second array, as shown in Fig. 11. GND vias connect all planes. The ports of interest for this case are P_1, P_2, P_3, P_4, P_5 and P_6. Geometry variations such as via pitch, distance between the arrays, and stripline distance influence the transmission and crosstalk between striplines and vias. Overall 1500 simulations were made. For

the simulation signal pads of the striplines and the coupling between striplines were not considered.

IV. ML INVESTIGATIONS USING THE SI/PI-Database

Here, for some of the structures past investigations with respect to ML as well as new results are presented. These investigations are based on ANNs where typically a genetic algorithm (GA) was used to identify the best topology and set of hyperparameters. Scattering parameters and classification against a threshold are predicted. Further, frequency domain values, e.g. weighted power sum of transmission (WPT) and weighted signal to crosstalk ratio (WSXTR), are calculated from predicted scattering parameters in a postprocessing step.

A. PREDICTION OF POWER SUPPLY IMPEDANCE VIOLATIONS OF A PCB

In Fig. 6 the impact of decap placement on the PDN impedance is shown. The large variety of possible positions in combination with capacitance values and number of decaps creates a challenge to chose an optimum distribution. Hence, an ANN was trained to predict, depending on the decap positions and values whether a specific PDN impedance (target impedance, TI) is violated. This can be understood as a classification problem with two labels (the target impedance (TI) being violated or not).

For the ML investigations the materials and geometries of the structure were not varied at first. Each of the 121 vias in the via array provides a possible decap position except for the center via (port 2) which is the port under investigation. To predict whether the PDN impedance fulfills the TI with the applied decap distribution, the values of the capacitance values and position of the placed decaps are important, as indicated in Fig. 6.

Investigations of the ANN topology and training parameters were reported in [3]. An ANN with two hidden layers with 15 and 5 neurons respectively showed good performance. The first approach to consider both, capacitance value and position, was implemented by using an individual input feature to the ANN for each via. This resulted in 121 input features, one per neuron, independent whether a decap was placed or not. For this approach the capacitance value was used for the input feature and a prediction accuracy of about 80 % was achieved. As stated in [3] the accuracy of the prediction could be increased when the data was preprocessed differently. An accuracy of more than 90 % was obtained by grouping of vias which either are close to each other or have similar distances to port 2. This highlights the importance of a domain dependent preprocessing qualified for a specific problem.

B. PREDICTION OF CUMULATIVE VALUES IN THE FREQUENCY DOMAIN FOR LINKS BETWEEN TWO VIA-ARRAYS

In [5] the setup described in Sec. III-B was investigated using ANNs. Besides scattering parameters, cumulative values

TABLE 2. Calculated and Predicted WPT of Concatenated Link on PCB with two Via-Arrays, Calculated from PB S-Parameter (Sim), calculated from ML predicted S-Parameter (Calc), ML prediction from Link Geometry (Pred).

Bitrate (Gbps)	Sim (dB)	Calc (dB)	Pred (dB)
5	68.8	68.7	68.5
15	71.9	71.6	71.5
25	72.9	72.6	72.5
35	74.9	74.6	74.4
45	73.1	72.8	72.6
55	73.9	73.6	73.5

describing the performance in frequency domain, e.g. the weighted power sum of transmission (WPT), were predicted [42]. For the ML investigations the WPT was predicted with two different approaches. The first approach used predicted scattering parameters from an ANN to calculate the WPT (Calc). The ANN had 4 hidden layers and 2000 neurons per layer. The second approach predicted the WPT directly from the link geometry without the scattering parameters (Pred). The comparison showed a good correlation of the two methods with the analytical approach (Sim), as shown in Tab. 2. Deviations of less than 1 dB were observed. Similar observations could be made for the weighted signal to crosstalk ratio (WSXTR) in [5]. This showed that it is possible to predict frequency dependent values as the WPT directly from geometry related parameters.

C. PREDICTION OF S-PARAMETERS FOR A 5 × 5 VIA-ARRAY

To identify the performance of ANNs to predict the magnitude and phase of electrically short PCB-based interconnects a regression was performed on the data set described in Sec. III-C. The prediction of an ANN is compared against a PB modeling approach in Fig. 12. As input features the parameters in combination with the frequency are used. Good agreement of simulation and prediction could be achieved with an ANN with 4 hidden layers and 24, 32, 24 and 24 neurons in the hidden layers, and 12 neurons in the output layer. For the hidden layers a hyperbolic tangent and for the output layer a linear activation function was used. For the training process the data was split in three sets, training, validation and testing. For the three data sets a mean-squared error (MSE) of less than 3×10^{-4} was achieved. This indicates a good generalization of the data and the absence of an overfitting problem. The prediction of scattering parameter shows important features such as the resonance at about 18 GHz, as shown in Fig. 12 (a). Above 28 GHz a deviation in the magnitude of about 4 dB is observed. The prediction of the phase shows most of the characteristic variations up to 32 GHz, as shown in Fig. 12 (b). The maximum observed deviation in the phase is 18 degrees. This shows that the prediction of scattering parameters of electrically short PCB-based structures is possible using ANNs.
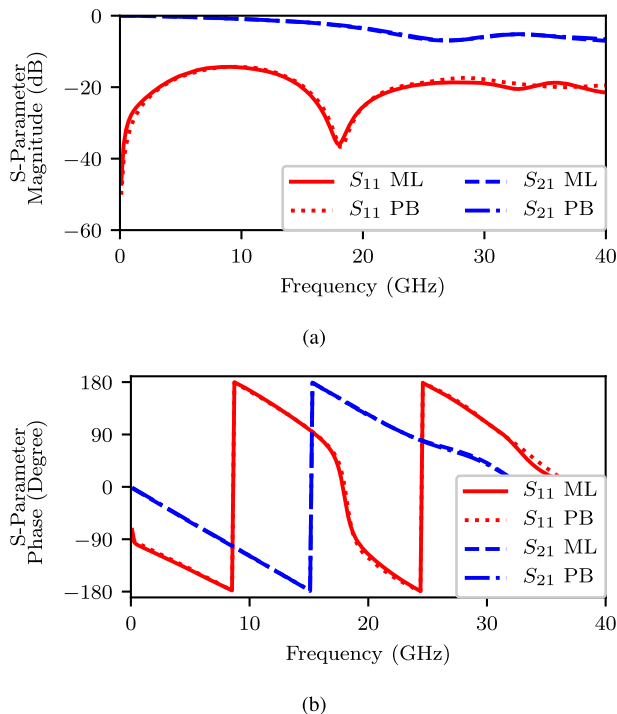
(a)

(b)

FIGURE 12. Prediction of scattering parameters based on an ANN for the interconnect of a 5 × 5 via-arrays on 10 cavity stackup as described in Sec. III-C. (a) Magnitude and (b) Phase with physics-based (PB) modeling and machine learning (ML) prediction.

D. PREDICTION OF CLASSIFICATION PARAMETER FOR A 5 × 5 VIA-ARRAY

For the data set with the 5 × 5 via-array described in Sec. III-C a multilabel classification was performed in order to investigate whether the magnitudes of scattering parameters S_{31}, S_{32} and S_{11} are below −21 dB, −20 dB and −11 dB, respectively [44]. The investigations were performed up to 20 GHz. 10 % of the 5000 data samples complied with these restrictions, resulting in an unbalanced data set. To predict whether a data sample fulfills the classification parameters an ANN with 16 and 32 neurons in the hidden layers and 3 neurons in the output layer was trained. Training ANNs with unbalanced data sets can cause problems as reported in [45]. To overcome the challenges presented by unbalanced data one solution is to use a cost sensitive training method, which penalizes wrongly predicted cases of the under represented class. Here, the score F_1 was used to consider the underrepresented class more [43].

In Fig. 13 the accuracy of the prediction of S_{31} with two ANNs is compared. In Fig. 13 (a) the ANN is trained with the mean-squared error as loss function. In Fig. 13 (b) the F_1 score was used during the training of the ANN. An improvement of the prediction accuracy considering the F_1 score is observed. This improvement is most dominant for the underrepresented class. This shows that the MSE is not a good indicator for the performance of unbalanced classification problems. However, techniques as using the F_1 score help to improve the performance of the classifier.
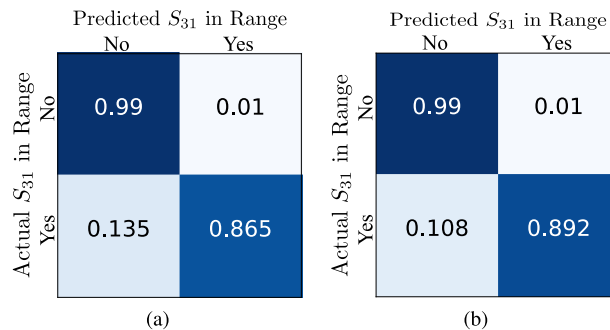

(a)

(b)

FIGURE 13. Confusion matrix for the prediction accuracy that S_{31} is below −21 dB for the interconnect of 5 × 5 via-arrays on 10 cavity stackup as described in Sec. III-C: (a) Training process depending on the mean-squared error, (b) training process depending on the F_1 score to consider the unbalanced data [43].
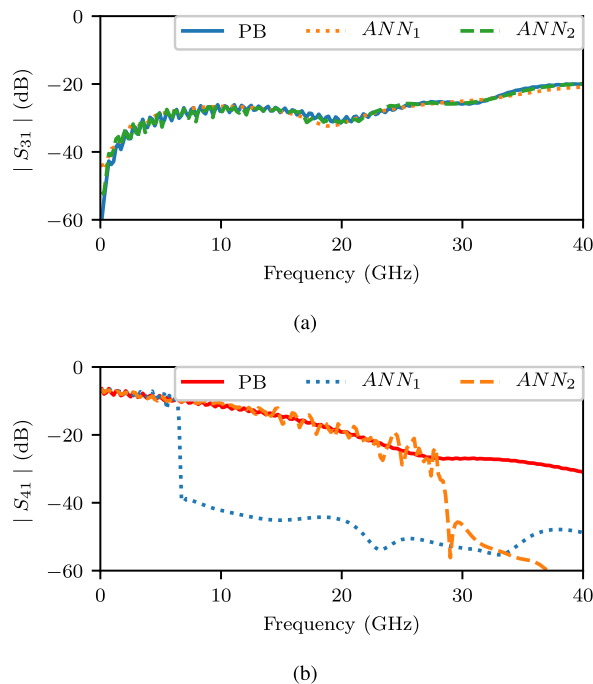

(a)

(b)

FIGURE 14. Scattering parameter prediction for the link on a 10 cavity stackup with two via-arrays as discussed in Sec. III.(a) magnitude S_{31} by physics-based (PB) model and prediction by ANNs (b) magnitude of S_{41} by physics-based (PB) model and prediction by artificial neural networks (ANNs). ANN_1 has 24, 40, 40 and 24 neurons in the hidden layers respectively, ANN_2 has 512 neurons in each of the hidden layers.

E. PREDICTION OF S-PARAMETERS FOR ELECTRICALLY LONG LINKS BETWEEN VIA-ARRAYS

To investigate the accuracy of ANNs to predict the magnitude of scattering parameters of electrically long PCB-based interconnects the data set described in Sec. III-D was used. The transmission S_{41} between two arrays and the near-end crosstalk (NEXT) S_{31} was investigated. Looking at the phase of transmission and NEXT shows strong variations for the transmission, and much less variations for the NEXT. The hyperparameters as well as the topology of an ANN were manually tuned in an iterative process along multiple generations of ANNs to achieve a good prediction of the NEXT.

The comparison of the magnitude of S_{31} modeled with a PB approach and predicted with an ANN (ANN_1) with 24, 40, 40 and 24 neurons in the hidden layers is shown in Fig. 14 (a). An ANN with the same topology was trained to predict the magnitude of the transmission, a comparison with the PB approach is shown in Fig. 14 (b). Beyond a frequency of 7 GHz a deviation of more than 20 dB is observed. To increase the frequency range with an adequate prediction accuracy another ANN (ANN_2) with 512 neurons in each of the 4 hidden layers was trained. The increased size of the ANN results in a similar drop in accuracy but beyond 28 GHz. This shows that parameters which show strong phase variations over frequency require a increased size of the ANN to represent the complexity of the link.

V. CONCLUSION AND OUTLOOK

The presented results based on data available from the SI/PI-Database show that predictions of scattering parameters, frequency dependent values as the WPT, and classifications are possible with moderate to high accuracy using ANNs. Among other things it was found that scattering parameter prediction for electrically large structures are challenging for ANNs. Predicting classification labels for the transmission and reflection of the 5 × 5 via array showed good performance. Also it was observed that the preprocessing of the geometry and material parameter may have a large impact on the accuracy. Finally, it was found that when handling unbalanced data sets for classification problems it is important to take into account the underrepresented class.

With the introduction of the SI/PI-Database we hope to enable other researchers to build further ML based tools for SI and PI applications, thus taking a step towards the generation of better ML techniques for EMC related problems in general. This might bring new and different ideas from different points of views, e.g. ideas that might have already proven to be effective in other domains and can be adapted.
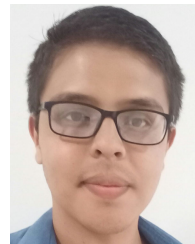
REFERENCES

[1] H.-D. Bruns, C. Schuster, and H. Singer, "Numerical electromagnetic field analysis for EMC problems," *IEEE Trans. Electromagn. Compat.*, vol. 49, no. 2, pp. 253–262, May 2007, doi: 10.1109/temc.2007.897152.

[2] H.-D. Bruns, A. Vogt, C. Findeklee, A. Schroder, M. Magdowski, M. Robinson, F. Heidler, C. Schuster, and K. Y. See, "Modeling challenging EMC problems," *IEEE Electromagn. Compat. Mag.*, vol. 6, no. 3, pp. 45–54, 2017, doi: 10.1109/memc.0.8093837.

[3] C. M. Schierholz, K. Scharff, and C. Schuster, "Evaluation of neural networks to predict target impedance violations of power delivery networks," in *Proc. IEEE 28th Conf. Electr. Perform. Electron. Packag. Syst. (EPEPS)*, Montreal, CA, USA, Oct. 2019, pp. 1–3.

[4] M. Schierholz, C. Yang, K. Roy, M. Swaminathan, and C. Schuster, "Comparison of collaborative versus extended artificial neural networks for PDN design," in *Proc. IEEE 24th Workshop Signal Power Integrity (SPI)*, May 2020, pp. 1–4, doi: 10.1109/spi48784.2020.9218157.

[5] K. Scharff, C. M. Schierholz, C. Yang, and C. Schuster, "ANN performance for the prediction of high-speed digital interconnects over multiple PCBs," in *Proc. IEEE 29th Conf. Electr. Perform. Electron. Packag. Syst. (EPEPS)*, San Jose, CA, USA, Oct. 2020, pp. 1–3.

[6] A. Sánches-Masís, A. Carmona-Cruz, M. Schierholz, X. Duan, K. Roys, C. Yang, R. Rimolo-Donadio, and C. Schuster, "ANN hyperparameter optimization by genetic algorithm for via interconnect classification," *Proc. IEEE 25th Workshop Signal Power Integrity (SPI)*, Jul. 2021, pp. 1–4.

[7] R. Laroussi and G. I. Costache, "Finite-element method applied to EMC problems (PCB environment)," *IEEE Trans. Electromagn. Compat.*, vol. 35, no. 2, pp. 178–184, May 1993, doi: 10.1109/15.229423.

[8] P. P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, vol. 3. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[9] T. Weiland, "Time domain electronic field computation with finite difference methods," *Int. J. Numer. Model.*, vol. 9, no. 4, pp. 319–395, Jul. 1996.

[10] R. F. Harrington, *Field Computation by Moment Methods*, D. K. Ray, Ed. Piscataway, NJ, USA: IEEE Press, 1993.

[11] B. Archambeault, C. Brench, and O. M. Ramahi, *EMI/EMC Computational Modeling Handbook*. Cham, Switzerland: Springer, 2001, doi: 10.1007/978-1-4615-1557-9.

[12] 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.

[13] Y.-S. Li, H. Yu, H. Jin, T. E. Sarvey, H. Oh, M. S. Bakir, M. Swaminathan, and E.-P. Li, "Dynamic thermal management for 3-D ICs with time-dependent power map using microchannel cooling and machine learning," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 9, no. 7, pp. 1244–1252, Jul. 2019.

[14] S.-Y. Hung, C.-Y. Lee, and Y.-L. Lin, "Data science for delamination prognosis and online batch learning in semiconductor assembly process," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 2, pp. 314–324, Feb. 2020.

[15] C. K. Ku, C. H. Goay, N. S. Ahmad, and P. Goh, "Jitter decomposition of high-speed data signals from jitter histograms with a pole–residue representation using multilayer perceptron neural networks," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 2227–2237, Oct. 2020.

[16] 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, doi: 10.1109/tcpmt.2020.3011910.

[17] R. R. Alavi, R. Mirzavand, A. Kiaee, R. Patton, and P. Mousavi, "Detection of the defective vias in SIW circuits from single/array probe(s) data using source reconstruction method and machine learning," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 9, pp. 3757–3770, Sep. 2019, doi: 10.1109/tmtt.2019.2931298.

[18] Y.-F. Shu, X.-C. Wei, J. Fan, R. Yang, and Y.-B. Yang, "An equivalent dipole model hybrid with artificial neural network for electromagnetic interference prediction," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 5, pp. 1790–1797, May 2019, doi: 10.1109/tmtt.2019.2905238.

[19] H. M. Torun and M. Swaminathan, "High-dimensional global optimization method for high-frequency electronic design," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 6, pp. 2128–2142, Jun. 2019, doi: 10.1109/tmtt.2019.2915298.

[20] 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, doi: 10.1109/tmtt.2020.3031204.

[21] X. Ye, Y. Bai, R. Song, K. Xu, and J. An, "An inhomogeneous background imaging method based on generative adversarial network," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4684–4693, Nov. 2020, doi: 10.1109/tmtt.2020.3015495.

[22] H. M. Torun, A. C. Durgun, K. Aygun, and M. Swaminathan, "Causal and passive parameterization of S-Parameters using neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 10, pp. 4290–4304, Oct. 2020, doi: 10.1109/tmtt.2020.3011449.

[23] F. Cilici, M. J. Barragan, E. Lauga-Larroze, S. Bourdel, G. Leger, L. Vincent, and S. Mir, "A nonintrusive machine learning-based test methodology for millimeter-wave integrated circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 8, pp. 3565–3579, Aug. 2020, doi: 10.1109/tmtt.2020.2991412.

[24] P. Zhao and K. Wu, "Homotopy optimization of microwave and millimeter-wave filters based on neural network model," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 4, pp. 1390–1400, Apr. 2020, doi: 10.1109/tmtt.2019.2963639.

[25] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012, doi: 10.1109/tpami.2011.155.

[26] C. H. Goay, A. Abd Aziz, N. S. Ahmad, and P. Goh, "Eye diagram contour modeling using multilayer perceptron neural networks with adaptive sampling and feature selection," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 9, no. 12, pp. 2427–2441, Dec. 2019, doi: 10.1109/tcpmt.2019.2938583.

[27] B. Li, B. Jiao, C.-H. Chou, R. Mayder, and P. Franzon, "Self-evolution cascade deep learning model for high-speed receiver adaptation," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 6, pp. 1043–1053, Jun. 2020.

[28] M. Larbi, R. Trinchero, F. G. Canavero, P. Besnier, and M. Swaminathan, "Analysis of parameter variability in an integrated wireless power transfer system via partial least-squares regression," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 11, pp. 1795–1802, Nov. 2020.

[29] X. Wang, T. Nguyen, and J. E. Schutt-Aine, "Laguerre–volterra feedforward neural network for modeling PAM-4 high-speed links," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 12, pp. 2061–2071, Dec. 2020.

[30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[31] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, A. Veit, and S. Belongie. (2017). *OpenImages: A Public Dataset for Large-Scale Multi-Label and Multi-Class Image Classification*. [Online]. Available: https://storage.googleapis.com/openimages/web/index.html

[32] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.

[33] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009, doi: 10.1007/s11263-009-0275-4.

[34] P. Ayris, J.-Y. Berthou, R. Bruce, S. Lindstaedt, A. Monreale, B. Mons, Y. Murayama, C. Södergord, K. Tochtermann, and R. Wilkinson. (2016). *Realising the European Open Science Cloud. Publications Office of the European Union*. Accessed: Dec. 10, 2020. [Online]. Available: https://data.europa.eu/doi/10.2777/940154

[35] G. Ai. (2020). *Open Image Data Set*. Accessed: Dec. 10, 2020. [Online]. Available: https://storage.googleapis.com/openimages/web/index.html

[36] *Touchstone File Format Specification*, IBIS Open Forum Official Subcomitee Under SAE-ITC, Warrendale, PA, USA, 2009.

[37] R. Rimolo-Donadio, X. Gu, Y. H. Kwark, M. B. Ritter, B. Archambeault, F. de Paulis, Y. Zhang, J. Fan, H.-D. Bruns, and C. Schuster, "Physics-based via and trace models for efficient link simulation on multilayer structures up to 40 GHz," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 8, pp. 2072–2083, Aug. 2009.

[38] X. Duan, R. Rimolo-Donadio, H.-D. Brüns, and C. Schuster, "A combined method for fast analysis of signal propagation, ground noise, and radiated emission of multilayer printed circuit boards," *IEEE Trans. Electromagn. Compat.*, vol. 52, no. 2, pp. 487–495, May 2010, doi: 10.1109/temc.2010.2041238.

[39] S. Müller, X. Duan, M. Kotzev, Y.-J. Zhang, J. Fan, X. Gu, Y. H. Kwark, R. Rimolo-Donadio, H.-D. Brüns, and C. Schuster, "Accuracy of physics-based via models for simulation of dense via arrays," *IEEE Trans. Electromagn. Compat.*, vol. 54, no. 5, pp. 1125–1136, Oct. 2012, doi: 10.1109/temc.2012.2192123.

[40] S. Muller, F. Happ, X. Duan, R. Rimolo-Donadio, H.-D. Bruns, and C. Schuster, "Complete modeling of large via constellations in multilayer printed circuit boards," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 3, no. 3, pp. 489–499, Mar. 2013, doi: 10.1109/tcpmt.2012.2234211.

[41] *ANSYS Electromagnetics Suite, Release 16.2*, ANSYS, Canonsburg, PA, USA, 2015. [Online]. Available: http://www.ansys.com

[42] K. Scharff, H.-D. Bruns, and C. Schuster, "Performance metrics for crosstalk on printed circuit boards in frequency domain," in *Proc. IEEE 23rd Workshop Signal Power Integrity (SPI)*, Jun. 2019, pp. 1–4, doi: 10.1109/sapiw.2019.8781662.

[43] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," in *Machine Learning and Knowledge Discovery in Databases*, vol. 8725. Berlin, Germany: Springer, Sep. 2014, pp. 225–239.

[44] G. Tsoumakas and I. Katakis, "Multi-label classification," *Int. J. Data Warehousing Mining*, vol. 3, no. 3, pp. 1–13, Jul. 2007, doi: 10.4018/jdwm.2007070101.

[45] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/tkde.2008.239.

MORTEN SCHIERHOLZ (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Hamburg University of Technology, Hamburg, Germany, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Institut fü Theoretische Elektrotechnik. He has authored and co-authored articles with respect to machine learning (ML) investigations for signal integrity (SI) and power integrity (PI) applications in conferences as EPEPS and SPI. His research interest are the analysis of SI and PI of printed circuit board-based interconnects with ML tools and techniques.

ALLAN SÁNCHEZ-MASÍS received the Lic. Degree (Hons.) in electronics engineering from the Instituto Tecnológico de Costa Rica (ITCR), Cartago, Costa Rica, in 2020. He is currently a Student Research Assistant with the Department of Electronics, ITCR, where he is also a Student Mentor in electromagnetic fields and electrical communication lectures. His current research interests are the analysis of signal integrity for multilayer interconnects through machine learning techniques.

ALLAN CARMONA-CRUZ (Student Member, IEEE) received the B.S. and Lic. degrees in electrical engineering from the Instituto Tecnológico de Costa Rica (ITCR), Cartago, Costa Rica, in 2018, and the master's degree in microelectronics in the optimization of time expensive black-box systems on SI problems from the Tecnológico de Costa Rica. He is currently with Intel Corporation, Costa Rica.

From 2018 to 2020, he was a Research Assistant with the Integrated Circuit Design Laboratory, ITCR, in the area of signal and power integrity. He was an Academic Interchange with the Institute of Electromagnetic Theory, Hamburg University of Technology. He has authored or coauthored various articles with both institutions and participated in related conferences, such as SPI 2019 and LASCAS 2020. His areas of professional interest include signal and power integrity, artificial intelligence, and digital and analog circuit design and testing.

XIAOMIN DUAN (Member, IEEE) received the M.S. degree in microelectronics and microsystems and the Ph.D. degree *(summa cum laude)* in electrical engineering from the Hamburg University of Technology, Hamburg, Germany, in 2007 and 2012, respectively.

He was a Post-Doctoral Researcher with the Institute of Electromagnetic Theory, Technical University of Hamburg, from 2012 to 2013, and the RF and High Speed Group, Fraunhofer Institute for Reliability and Microintegration, Berlin, Germany, from 2013 to 2015. Since 2016, he has been with IBM Germany Research & Development GmbH, Böblingen, Germany. His research interests include signal, power integrity, and EMC for high-speed and RF systems.

Dr. Duan was a recipient of the IEEE EMC Best Student Symposium Paper Award in 2009, the IBM Ph.D. Fellowship Award in 2010, and the IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY Best Paper Award in 2012 and 2016.

KALLOL ROY (Member, IEEE) received the bachelor's degree in electrical engineering from IIT Kanpur, Kanpur, and the Ph.D. degree in electrical communication engineering from Indian Institute of Science (IISc) Bangalore. He was a Post-Doctoral Researcher with the Packaging Research Center, Georgia Institute of Technology, Atlanta, USA, and the Statistical Artificial Intelligence Laboratory, Ulsan National Institute of Technology, South Korea, and the Department of Mathematics, IISc Bangalore. He is currently an Assistant Professor with the Institute of Computer Science, University of Tartu, Estonia. His research interest is deep learning, computer vision, AI accelerator, and electromagnetics. He was a recipient of the APS-IUSSTF Physics Student Visitation Award, the Jawaharlal Nehru Scholarship Steel Authority of India Limited, in 2000, the IIT Guwahati MHRD Scholarship, Government of India 2007, the Sterlite Best Paper Award at Photonics 2010, and the 2012 Microsoft Travel Award.

RENATO RIMOLO-DONADIO (Senior Member, IEEE) received the B.S. and Lic. degrees in electrical engineering from the Instituto Tecnológico de Costa Rica (ITCR), Cartago, Costa Rica, in 1999 and 2004, respectively, and the M.Sc. degree in microelectronics and microsystems and the Ph.D. degree in electrical engineering from the Technische Universität Hamburg (TUHH), Hamburg, Germany, in 2006 and 2010, respectively.

From 2006 to 2012, he was with the Institute of Electromagnetic Theory, TUHH. From 2012 to 2014, he was with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. He is currently a Full Professor with the Department of Electronics Engineering, ITCR. His current research interests include the system-level modeling and optimization of interconnects, the analysis of signal and power integrity problems, and high-speed integrated circuit design and remote sensing techniques.

Dr. Rimolo-Donadio has been a part of several technical program committees for IEEE conferences. He was a recipient of the IEEE DATE Best IP Paper Award in 2008, the IEC DesignCon Best Paper Award in 2010, the Best IEEE EMC German Ph.D. Thesis of the Year in 2011, the CPMT Best Paper Award in 2012, the IEEE EMC Symposium Best Paper Award 2013, and the IEEE JLT Best Paper Award in 2014. He has served as the Chair for the Costa Rica IEEE Circuits and Systems Chapter 2015–2016. He has been a reviewer for several IEEE journals related to the fields of integrated circuits and signal and power integrity.

CHRISTIAN SCHUSTER (Senior Member, IEEE) received the Diploma degree in physics from the University of Konstanz, Germany, in 1996, and the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, in 2000. Since 2006, he has been a Full Professor and the Head of the Institute of Electromagnetic Theory, Hamburg University of Technology (TUHH), Germany. He was with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, where he was involved in high-speed optoelectronic package and backplane interconnect modeling and signal integrity design for new server generations. His currents interests include signal and power integrity of digital systems, multiport measurement and calibration techniques, and development of electromagnetic simulation methods for communication electronics.

He is an Adjunct Associate Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology. He was serving as a Distinguished Lecturer for the IEEE EMC Society from 2012 to 2013. He is a member of the German Physical Society and several technical program committees of international conferences on signal and power integrity and electromagnetic compatibility. He is a member of the Board of Directors of the EMC Society. He received the IEEE Transactions on EMC Best Paper Awards in 2001 and 2015, the IEEE Transactions on CPMT Best Paper Awards in 2012 and 2016, the IEC DesignCon Paper Awards in 2005, 2006, 2010, 2017, and 2018, three IBM Research Division awards from 2003 to 2005, and the IBM Faculty Awards in 2009 and 2010. He received the Sustained Service to the EMC Society Award in 2019. He was serving as the Chair for the German IEEE EMC Chapter from 2016 to 2019. He is currently an Associate Editor for the IEEE Transactions on EMC.

CHENG YANG (Member, IEEE) received the B.S. degree in electronic science and technology from Wuhan University, Wuhan, China, in 2009, the M.S. and Ph.D. degrees in electromagnetic field and microwave technology from the National University of Defense Technology, Changsha, China, in 2012 and 2016, respectively.

From 2013 to 2015, he was supported by the Chinese Scholarship Council as a joint-Ph.D. Student with the Technische Universität Hamburg (TUHH), Hamburg, Germany. From 2016 to 2019, he was a Faculty Member with the State Key Laboratory of Millimeter Wave, Southeast University, Nanjing, China. Since 2019, he has been a Senior Engineer with the Institute of Electromagnetic Theory, TUHH. His current research interests include computational electromagnetics, microwave measurement techniques, and design and analysis of nonlinearly loaded electromagnetic structures.

- - -