IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 October 2023, accepted 13 November 2023, date of publication 20 November 2023,
date of current version 28 November 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3335306*

RESEARCH ARTICLE

# Fast Multi-ANN Composite Models for Repeater Optimization in Presence of Parametric Uncertainty for On-Chip Hybrid Copper-Graphene Interconnects

**SUYASH KUSHWAHA[1], (Graduate Student Member, IEEE), AVIRUP DASGUPTA[2], (Senior Member, IEEE), SOURAJEET ROY[2], (Senior Member, IEEE), AND ROHIT SHARMA[1], (Senior Member, IEEE)**

[1]Department of Electrical Engineering, Indian Institute of Technology Ropar, Rupnagar, Punjab 140001, India
[2]Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India

Corresponding author: Rohit Sharma (rohit@iitrpr.ac.in)

This work was supported in part by Semiconductor Research Corporation (SRC) under Grant 2022-IR-3168.

**ABSTRACT** In this paper, composite models are developed to predict the statistics of the optimal number and size of repeaters required to minimize the power delay product (PDP) of on-chip hybrid copper-graphene interconnects when subject to parametric uncertainty. Specifically, two distinct artificial neural network (ANN) based composite models are developed in this paper. Each composite model is made up of three individual ANNs that are interconnected. Depending on the way in which the ANNs are interconnected, the total number of full-wave electromagnetic (EM) simulations and SPICE circuit simulations required for training the composite models are reduced. Overall, the composite models enable the use of analytic expressions instead of expensive and repeated full-wave EM and SPICE simulations to solve the repeater optimization problem within a Monte Carlo framework for efficient statistical analysis.

**INDEX TERMS** Artificial neural networks (ANNs), copper-graphene interconnects, fin field effect transistors (FinFETs), parametric uncertainty, repeaters, statistical analysis.

## I. INTRODUCTION

With the scaling of very large scale integration (VLSI) technologies below the sub-22 nm levels, conventional on-chip copper interconnects are reaching the limits of their performance due to aggravated scattering [1]. Moreover, the diffusion of copper ions from the interconnects into the dielectric layer further increases the dielectric conductivity and leakage losses [2]. As a remedy for both the above problems, recently ultra-thin graphene barrier layers have been placed around the copper traces [3], [4]. These ultra-thin graphene barrier layers, by virtue of their high mean free path of electrons compared to copper, not only serve as low resistance paths parallel to the copper conductor for enhanced current conduction but also prevent the diffusion of copper ions into the dielectric, all without significantly decreasing the cross-sectional area of the interconnects. Therefore, hybrid copper-graphene interconnects have the potential to address the limitations of on-chip copper interconnects in ultra-scaled technology nodes [3], [4].

An important aspect of on-chip interconnect design is to place repeaters along the conductors to optimize some performance metric of the network. For example, repeaters have often been introduced into the interconnects to minimize the power, delay, power delay product (PDP), bandwidth, crosstalk delay, and chip area of the network [5], [6], [7], [8], [9], [10], [11], [12], [13]. The main challenge in such scenarios is to accurately calculate the optimal number and size of these repeaters. To that end, various works have reported approximate closed-form expressions of the optimal

The associate editor coordinating the review of this manuscript and approving it for publication was Davide Patti.

number and size of the repeaters albeit for minimizing the power consumed, delay, or crosstalk delay separately for on-chip interconnects [5], [6], [7], [8], [9], [10]. Recently, in [11], a particle swarm optimization (PSO) algorithm has been developed to infer the optimal number and size of the repeaters for minimizing the more holistic PDP of both conventional copper and multi-walled carbon nanotube interconnects. In all these works, the optimal number and size of the repeaters are expressed as functions of the per-unit-length (p. u. l.) parameters of the interconnects.

Within the context of hybrid copper-graphene interconnects, full-wave electromagnetic (EM) solvers are required to accurately calculate the p. u. l. parameters at massive computational time costs [14], [15]. This issue of massive computational time cost is further exacerbated if the statistical information (e.g., mean, standard deviation, and probability density function) of the optimal number and size of the repeaters have to be quantified using repeated Monte Carlo evaluations due to parametric uncertainty in the interconnects [12]. In order to address this computational burden, recently analytic models have been developed for predicting the p. u. l. parameters of hybrid copper-graphene interconnects [16]. However, while the models of [16] are extremely fast, they lack the requisite level of accuracy compared to EM solvers [14].

One effective approach to mitigate the massive computational costs behind repeated evaluations of the p. u. l. parameters, and consequently, the optimal number and size of repeaters of general interconnects is by using machine learning [14], [15], [17]. This is because machine learning (ML) surrogate models or metamodels such as artificial neural networks (ANNs) can emulate the p. u. l. parameters of interconnects, and thereby, the optimal number and size of the repeaters as analytic functions of the geometrical, physical, and material parameters of the network [14], [15], [17]. These analytic functions can be used in lieu of computationally slow full-wave EM simulations within a Monte Carlo framework to efficiently determine the statistics of the optimal number and size of the repeaters required in presence of parametric uncertainty. Despite this advantage of ML metamodels, their one major limitation is that they are very data hungry. This means that a massive amount of training data is required to ensure that these metamodels are able to accurately predict the p. u. l. parameters of interconnects [17]. Since, generating the massive amount of training data requires repeated full-wave EM simulations, ML metamodels end up incurring prohibitively large training time costs. Therefore, performing repeater optimization for hybrid copper-graphene interconnects even using conventional ML techniques are often computationally intractable.

In this paper, two novel composite ML metamodels are developed to address the above limitation of hybrid copper-graphene interconnects. Each composite metamodel consists of three individual ANNs that are interconnected. The main difference between the two composite metamodels lies in the way in which the ANNs are interconnected. In one metamodel, the ANNs are interconnected in such a manner so that the prior knowledge of the p. u. l. parameters of the interconnects extracted from the crude but analytic models of [16] can be directly leveraged to accelerate its training. In the second metamodel, the ANNs are interconnected in such a manner so that the analytic models of [16] can be indirectly leveraged to accelerate its training. In both composite metamodels, the combination of the three ANNs enable the construction of a set of analytic expressions that can be used in lieu of the slow full-wave EM and SPICE simulations to efficiently solve the repeater optimization problem for hybrid copper-graphene interconnects for any given value of the geometrical, physical, and material parameters of the passive interconnect structures and the active fin field effect transistors (FinFETs) making up the repeaters.

This work is an expansion of the author's conference paper of [12]. In this paper, the following new ideas and techniques not seen in the conference paper is presented.

(1) In this paper, two multi-ANN composite metamodels have been intelligently developed with very low training time costs. In the first composite metamodel, the inaccurate yet fast analytic models of [16] are used as prior knowledge of the interconnects within the known source difference (SD) formulation [14], [15] and the new prior knowledge with source difference (PKID) formulation [18] to expedite its training. In the second composite metamodel, the inaccurate yet fast analytic models of [16] are included in a novel alternative formulation referred to as prior knowledge input with sparse postprocessing (PKISP) to expedite the training process. All these formulations represent key novelties of this work over that of [12] where conventional ANNs were used without any attempt to expedite their training. The outcome of these novelties is that the proposed composite metamodels can be trained 2-4 times faster than the conventional metamodel of [12] as demonstrated in Section III.

(2) In the proposed composite metamodels, a dedicated ANN is used to map the impact of parametric uncertainty in the FinFETs used to construct the repeaters on the optimal number and size of the repeaters. This inclusion of the parametric uncertainty in the active FET device besides the passive interconnects has not been considered in other existing works including the conference paper of [12]. Consequently, the proposed composite metamodels are the only ones that can truly account for a comprehensive set of uncertain parameters spanning both the active and passive parts of the interconnect network. In fact, the numerical examples of Section III of this paper clearly demonstrate that neglecting the parametric variability of the FinFETs can introduce 10-20% errors in the statistical solutions of the repeater optimization problem.

(3) An additional benefit of the dedicated ANN used to map the parametric uncertainty of the FinFETs is that it emulates the switching resistance, gate capacitance, and the parasitic capacitance of the repeaters as analytic functions

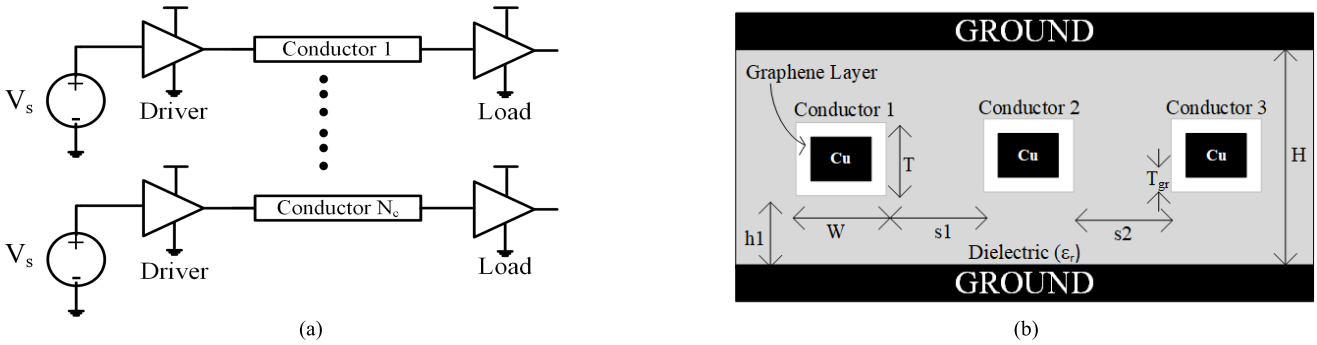



**FIGURE 1.** Multiconductor stripline on-chip hybrid copper-graphene interconnect network. (a) Circuit schematic of a general $N_c$-conductor hybrid copper-graphene interconnect network showing driver/load circuits. (b) Cross-sectional view of the hybrid copper-graphene interconnects showing the heterogeneous nature of the traces.

of the FinFET device parameters. These analytic functions can be used instead of multiple SPICE simulations of the repeaters to efficiently characterize the impact of the uncertainty present in the FinFET parameters on the optimal number and size of the repeaters. Thus, this ANN will enhance the computational efficiency of the proposed composite metamodels over standard non-ML approaches when performing repeater optimization. Importantly, this benefit is not explored in the work of [12] because no uncertainty was considered in the FinFETs.

(4) Finally, in this paper, the accuracy and the numerical efficiency of the proposed composite metamodels have been validated across multiple technology nodes in contrast to the work of [12] which was validated for only the 14 nm technology node. Furthermore, in the examples of Section III of this paper, a much larger number of input parameters is considered than what was taken in [12] for a more thorough validation of the scalability of the proposed composite metamodels.

## II. DEVELOPMENT OF PROPOSED COMPOSITE METAMODELS

### A. BASIC PROBLEM STATEMENT

Consider a multiconductor on-chip hybrid copper-graphene interconnect network as shown in Fig. 1(a) where the cross-section of the heterogeneous conductors is displayed in Fig. 1(b). Repeaters are usually inserted into each conductor of the network. These repeaters take the form of inverters [5], [6]. By tuning the number of repeaters and the size of the FinFET devices making up the repeaters, it is possible to optimize specific signal integrity (SI) quantities of interest such as 50% delay, power consumed, bandwidth, chip area, and power delay product (PDP) of the network [5], [6], [7], [8], [9], [10], [11], [12]. Now, let the fabrication process variations and layout uncertainty in the network be present in $n$ parameters of the interconnect structure $\mathbf{P_i} = [p_1, \cdots, p_n]$ and $m$ parameters of the FinFETs $\mathbf{P_d} = [p_{n+1}, \cdots, p_{n+m}]$. These parameters can be mathematically expressed as

$$p_i = p_{i0} + p_{i0}\, s_i \lambda_i; \quad 1 \le i \le n+m \tag{1}$$

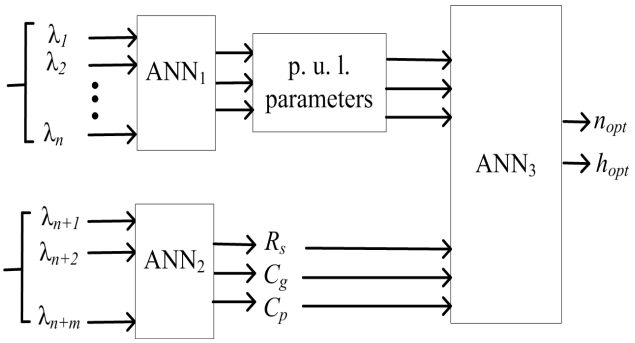



**FIGURE 2.** Block structure of the proposed composite metamodel 1 showing how the different ANNs are interconnected.

where $p_{i0}$ is the mean value, $s_i$ is the relative standard deviation, and $\lambda_i$ is the marginal random variable representing the uncertainty of the $i$-th parameter. Thus, the random variables corresponding to the parametric uncertainty in the interconnect structures is given by $\lambda_i = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ and that corresponding to the parametric uncertainty in the FinFETs is given by $\lambda_d = [\lambda_{n+1}, \lambda_{n+2}, \ldots, \lambda_{n+m}]$ where $\lambda = [\lambda_i, \lambda_d]$. Now, because of the presence of the random variables $\lambda$, the optimal number ($n_{opt}$) and size ($h_{opt}$) of the repeaters required to minimize the PDP of the network cannot be precisely known. In other words, the solution to the repeater optimization problem will be a function of the random variables $\lambda$, expressed as $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$. The task of uncertainty quantification (UQ) is to predict the statistical information such as the mean, variance, and probability density functions (PDFs) of $n_{opt}(\lambda)$ and $h_{opt}(\lambda)$. In order to perform UQ, in this paper two different multi-ANN composite metamodels are developed to emulate the unknowns $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ as described next.

### B. BASIC BLOCK STRUCTURE OF COMPOSITE METAMODEL 1

The block structure of the composite metamodel 1 is displayed in Fig. 2. It is observed from Fig. 2 that the first ANN (ANN$_1$) is used to emulate the p. u. l. parameters of

the copper-graphene interconnects as analytic functions of the geometrical, physical, and material parameters of the heterogeneous structure of Fig. 1(b). This ANN is trained on results extracted from the computationally costly full-wave EM solvers [12]. The second ANN ($ANN_2$) pertains to the repeaters used. It is pointed out that the repeaters are basically inverters comprising of PMOS and NMOS FinFET devices. In repeater optimization problems, these repeaters are modeled as equivalent RC circuits described by the switching resistance ($R_s$), gate capacitance ($C_g$), and the parasitic capacitance ($C_p$) of the inverter as shown in Fig. 3 [12], [19]. Now, the goal of the second ANN is to emulate the values of $\{R_s, C_g, C_p\}$ as functions of the geometrical, material, and physical parameters of the FinFETs. This ANN is trained on data generated from multiple SPICE simulations of the repeaters [12], [18]. Finally, based on the knowledge of the p. u. l. parameters of the interconnects and the RC elements $\{R_s, C_g, C_p\}$, the PSO algorithm can quantify the optimal number and size of the repeaters required $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ [11]. However, repeatedly evaluating the PSO algorithm for different values of the random variables $\lambda$ is a time-intensive process. To address this issue, a third ANN ($ANN_3$) is developed. This third ANN accepts as inputs the values of the p. u. l. parameters of the interconnects, and the RC elements $\{R_s, C_g, C_p\}$ and emulates the optimal number and size of the repeaters as functions of these inputs. Hence, this ANN is trained on data extracted from repeated evaluations of the PSO algorithm. In the following subsections, the training and testing details of each ANN is described.

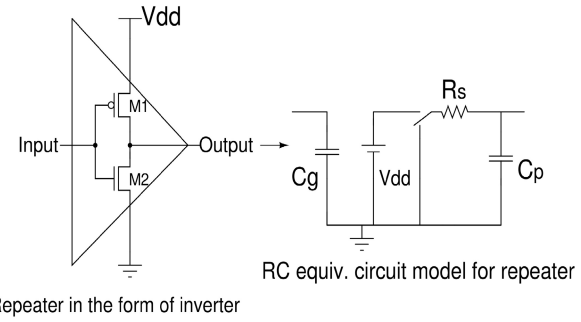



**FIGURE 3.** Equivalent RC circuit representation of repeaters fashioned as inverters.

### C. TRAINING OF $ANN_1$ OF COMPOSITE METAMODEL 1

In the work of [12], the training of $ANN_1$ proceeds in a conventional manner by directly using the dataset extracted from the full-wave EM solvers. In particular, the inputs to $ANN_1$ are the random variables $\lambda_i$ corresponding to the interconnect parameters $\mathbf{P}_i$. The outputs of $ANN_1$ consists of the p. u. l. resistance of any general $j$-th conductor $R_{jj}$, the self-capacitance of the $j$-th conductor to ground $C_{j0}$, the coupling capacitance between the $j$-th and $k$-th conductors $C_{jk}$, the self-inductance of the $j$-th conductor $L_{jj}$, and the mutual inductance between the $j$-th and $k$-th conductors $M_{jk}$. The main computational challenge of the conventional training approach seen in [12] is that usually hundreds of training samples, and consequently, hundreds of full-wave EM simulations of the interconnect structure of Fig. 1(b) are required to train $ANN_1$. To address the massive time costs associated with so many full-wave EM simulations, the authors have developed two prior knowledge techniques to expedite the training of $ANN_1$.

#### 1) SOURCE DIFFERENCE (SD) APPROACH

In the SD approach, $ANN_1$ attempts to emulate the error between the true p. u. l. parameters of the interconnects extracted from the full-wave EM simulations (i.e., $\mathbf{y}(\lambda_i)$) and the p. u. l. parameters estimated from the low-fidelity analytic model of [16], say $\mathbf{y}_{LF}(\lambda_i)$. In other words, the training dataset for $ANN_1$ using the SD approach is expressed as $\mathrm{D}_{SD} = \{(\lambda_{i,l}, \mathbf{E}(\lambda_{i,l}))\}_{l=1}^{L}$ where

$$\mathbf{E}(\lambda_{i,l}) = \mathbf{y}(\lambda_{i,l}) - \mathbf{y}_{LF}(\lambda_{i,l}) \tag{2}$$

In (2), $\lambda_{i,l} = [\lambda_1^{(l)}, \ldots, \lambda_n^{(l)}]$ refers to the $l$-th training sample and $L$ is the total number of training points required. Now, assuming $ANN_1$ has a single hidden layer, the predicted outputs of $ANN_1$ take the form

$$\tilde{\mathbf{E}}(\lambda_i) = [e_1(\lambda_i), \cdots, e_K(\lambda_i)];$$
$$e_k(\lambda_i) = \sigma_{k,3}\left(b_{k,3} + \sum_{i=1}^{N_h} w_{2,3}^{i,k}\,\sigma_{i,2}\left(b_{i,2} + \sum_{j=1}^{n} w_{1,2}^{j,i}\,\lambda_j\right)\right), \quad 1 \le k \le K \tag{3}$$

In (3), $\sigma_{p,q}$ refers to the nonlinear activation function used in the $p$-th neuron of the $q$-th layer, $b_{p,q}$ is the bias value entering the $p$-th neuron of the $q$-th layer, $w_{p,q}^{\alpha,\beta}$ is the synaptic weight linking the $\alpha$-th neuron of the $p$-th layer to the $\beta$-th neuron of the $q$-th layer, and $N_h$ is the number of hidden neurons. Next, the values of all the weight and bias terms in (3) are tuned to minimize the error loss function [15]

$$f_{ANN_1}(\mathbf{w}, \mathbf{b}) = \underset{\mathbf{w},\mathbf{b}\in\Re}{\arg\min}\left(\frac{1}{L}\sum_{l=1}^{L}\left\|\mathbf{E}(\lambda_{i,l}) - \tilde{\mathbf{E}}(\lambda_{i,l})\right\|_2^2\right) \tag{4}$$

where $(\mathbf{w}, \mathbf{b})$ refer to the real valued set of weights and bias terms in (3). This process of tuning the weights and bias terms is referred to as training of the ANN metamodel. Now, because the low-fidelity results $\mathbf{y}_{LF}(\lambda_i)$ are reasonably well-correlated, even if not exactly equal to the high-fidelity full-wave EM simulation results $\mathbf{y}(\lambda_i)$, the variance of the error function $\mathbf{E}(\lambda_i)$ of (2) will be significantly smaller than the variance of $\mathbf{y}(\lambda_i)$ (i.e., $var(\mathbf{E}(\lambda_i)) \ll var(\mathbf{y}(\lambda_i))$) [14], [15]. Therefore, a much smaller number of training samples will be required to train $ANN_1$ to emulate the error function $\mathbf{E}(\lambda_i)$ than to emulate the p. u. l. parameters of the interconnects $\mathbf{y}(\lambda_i)$ [14], [15]. Once $ANN_1$ has been sufficiently well-trained, the p. u. l. parameters of the interconnects are easily recovered from (2) and (3) as the sum of the outputs predicted by $ANN_1$ and the low-fidelity model as

$$\tilde{\mathbf{y}}(\lambda_i) = \tilde{\mathbf{E}}(\lambda_i) + \mathbf{y}_{LF}(\lambda_i) \tag{5}$$

In effect, in the SD approach, the low-fidelity results $\mathbf{y}_{LF}(\lambda_i)$ serves as prior and cheaply available information to expedite the training of $\text{ANN}_1$. Interested readers are directed to [14], [15] for further details of the SD approach.

#### 2) PRIOR KNOWLEDGE INPUT WITH SOURCE DIFFERENCE (PKID) APPROACH

In the PKID approach, the prior knowledge of the interconnects in the form of the low-fidelity results $\mathrm{y}_{LF}(\lambda_i)$ is leveraged in two distinct ways. First, the low-fidelity results $\mathrm{y}_{LF}(\lambda_i)$ are used as additional inputs to $\text{ANN}_1$. These additional inputs help guide $\text{ANN}_1$ to learn the functional dependence of the p. u. l. parameters of the hybrid interconnects on $\lambda$ in a more data-efficient manner than what is conventionally possible. Second, as in the SD approach, $\text{ANN}_1$ attempts to reduce the variance of its target output by emulating the error between the high-fidelity results extracted from the full-wave EM simulations (i.e., $\mathrm{y}(\lambda_i)$) and the low-fidelity results $\mathrm{y}_{LF}(\lambda_i)$. Thus, the training dataset for the PKID approach can be expressed as $\mathrm{D}_{PKID} = \{(\bar{\lambda}_{i,l}, \mathbf{E}(\lambda_{i,l}))\}_{l=1}^{L}$ where the new input feature space of $\text{ANN}_1$is

$$\bar{\lambda}_{i,l} = \left[\lambda_{i,l},\ \mathbf{y}_{LF}(\lambda_{i,l})\right] \tag{6}$$

As before, assuming $\text{ANN}_1$ has a single hidden layer, the outputs of $\text{ANN}_1$ take the form

$$\tilde{\mathbf{E}}\left(\lambda_i\right) = \left[e_1\left(\bar{\lambda}_i\right), \cdots, e_K\left(\bar{\lambda}_i\right)\right];\quad \bar{\lambda}_i = [\lambda_i,\ \mathbf{y}_{LF}(\lambda_i)]$$
$$e_k(\lambda_i) = \sigma_{k,3}\left(b_{k,3} + \sum_{i=1}^{N_h} w_{2,3}^{i,k}\,\sigma_{i,2}\left(b_{i,2} + \sum_{j=1}^{n+K} w_{1,2}^{j,i}\,\bar{\lambda}_j\right)\right),\quad 1 \le k \le K \tag{7}$$

Note that the key difference in (7) with respect to (3) is that the input feature space has changed from $\lambda_i$ to the expanded feature space $\bar{\lambda}_i = \left[\bar{\lambda}_1, \cdots, \bar{\lambda}_{n+K}\right] = [\lambda_i, y_{LF}(\lambda_i)]$ of (6). When training $\text{ANN}_1$using the PKID approach, the error loss function to be minimized remains the same as in (4). In effect, the PKID approach harnesses the advantages of both the SD approach of [14] and the prior knowledge input (PKI) approach of [15] and is, hence, expected to outperform both these approaches. This means that the number of training samples required by the PKID approach is expected to be even smaller than that required by the SD and PKI approaches individually to train $\text{ANN}_1$ to the same accuracy level.

At this juncture, the authors would like to emphasize that in their previous work of [18], the superior learning rate of PKID approach comapred to the SD and PKI approaches has already been validated. However, in that paper, the focus was on training an ANN to emualte the signal integrity quantities (e.g., delay, peak crosstalk, eye diagram characterstics) of multi-walled carbon nanotue interconencts using data extracted from SPICE simulations. Currently, no work has explored if the theoretical benefits of the PKID approach can be extended to other problems – for example, to emualte the p. u. l. parameters of complex interconenct structures (e.g., heterogeneous copper-graphene structures) using data extracted from the more time-intensive full-wave EM solvers.

### D. TRAINING OF $\text{ANN}_2$ OF COMPOSITE METAMODEL 1

The objective of $\text{ANN}_2$is to map the impact of the parametric uncertainty present in the FinFETs making up the repeaters on the optimal number and size of the repeaters. To that end, the repeaters are modeled as inverters where each inverter is, in turn, modeled using a resistance-capacitance (RC) network as shown in Fig. 3. Thus, the objective of $\text{ANN}_2$ is to emulate the values of the switching resistance ($R_s$), gate capacitance ($C_g$), and the parasitic capacitance ($C_p$) of the inverters as functions of the random variables associated with the FinFET parameters, $\lambda_d = [\lambda_{n+1}, \lambda_2, \ldots, \lambda_{n+m}]$. This objective can be met using the methodology outlined in [19]. However, that methodology is considered to be computationally prohibitive because it requires 6 discrete SPICE simulations to calculate the values of $\{R_s, C_g, C_p\}$ for each solitary set of FinFET parameter values. Hence, in this work, an ANN (i.e., $\text{ANN}_2$) is proposed as a more efficient alternative to analytically emulate the dependency of the values of $\{R_s, C_g, C_p\}$ on the FinFET parameters. The training dataset for $\text{ANN}_2$ is $D_R = \{(\lambda_{d,r}, \boldsymbol{\theta}(\lambda_{d,r}))\}_{r=1}^{R}$ where

$$\boldsymbol{\theta}\left(\lambda_{d,r}\right) = \left[R_s\left(\lambda_{d,r}\right), C_g\left(\lambda_{d,r}\right), C_p\left(\lambda_{d,r}\right)\right] \tag{8}$$

are the values of the RC circuit elements used to model the repeaters in Fig. 3 extracted from the SPICE simualtions using the methodology outlined in [19] at the r-th training sample $\lambda_{d,r} = [\lambda_{n+1}^{(r)}, \ldots, \lambda_{n+m}^{(r)}]$. Assuming a single hidden layer in $\text{ANN}_2$, the values of the RC circuit elements predicted by $\text{ANN}_2$take the form

$$\tilde{\boldsymbol{\theta}}\left(\lambda_d\right) = \left[\tilde{\theta}_1\left(\lambda_d\right), \tilde{\theta}_2\left(\lambda_d\right), \tilde{\theta}_3\left(\lambda_d\right)\right];$$
$$\tilde{\theta}_k(\lambda_d) = \sigma_{k,3}\left(b_{k,3} + \sum_{i=1}^{N_h} w_{2,3}^{i,k}\,\sigma_{i,2}\left(b_{i,2} + \sum_{j=1}^{m} w_{1,2}^{j,i}\,\lambda_{n+j}\right)\right),\quad 1 \le k \le 3 \tag{9}$$

In course of training $\text{ANN}_2$, the values of all the weight and bias terms in (11) are tuned to minimize the error loss function [17]

$$f_{ANN_2}(\mathbf{w}, \mathbf{b}) = \arg\min_{\mathbf{w},\mathbf{b}\in\Re}\left(\frac{1}{R}\sum_{r=1}^{R}\left\|\theta\left(\lambda_{d,r}\right) - \tilde{\boldsymbol{\theta}}\left(\lambda_{d,r}\right)\right\|_2^2\right) \tag{10}$$

Once $\text{ANN}_2$ has been sufficiently well-trained, the outputs of $\text{ANN}_2$ will automatically account for the parametric uncertainty present in the FinFETs making up the repeaters. Note that this ability of the proposed composite model 1 to account for the parametric uncertainty in the FinFETs when quantifying the statistics of the optimal number and size of the repeaters has not been seen in the existing literature.

### E. TRAINING OF ANN$_3$ OF COMPODITE METAMODEL 1

The objective of ANN$_3$ of the proposed composite metamodel 1 of Fig. 2 is to predict the optimal number and size of the repeaters required to minimize the PDP of a given hybrid copper-graphene interconnect network – in other words, for any given values of the random variables $\lambda = [\lambda_i, \lambda_d]$. For this purpose, ANN$_3$ takes as inputs the outputs from ANN$_1$ and ANN$_2$ while the outputs are the optimal number and size of the repeaters $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$. The training dataset for ANN$_3$ is expressed as $D_{PSO} = \{(\boldsymbol{\alpha}_q, \mathbf{z}_q)\}_{q=1}^{Q}$ where

$$\boldsymbol{\alpha}_q = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{K+3} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{y}}\left(\lambda_{i,q}\right) \\ \tilde{\theta}\left(\lambda_{d,q}\right) \end{bmatrix}; \ \mathbf{z}_q = \begin{bmatrix} n_{opt}\left(\lambda_{i,q}, \lambda_{d,q}\right) \\ h_{opt}\left(\lambda_{i,q}, \lambda_{d,q}\right) \end{bmatrix} \tag{11}$$

In (11), the solution of the repeater optimization problem at each $q$-th training sample, $\mathbf{z}_q$, is obtained from the evaluation of the PSO algorithm [11]. Again, assuming a single hidden layer in ANN$_3$, the outputs predicted by ANN$_3$take the form

$$\tilde{\mathbf{z}}(\boldsymbol{\alpha}) = \begin{bmatrix} \tilde{z}_1(\boldsymbol{\alpha}) \\ \tilde{z}_2(\boldsymbol{\alpha}) \end{bmatrix} = \begin{bmatrix} \tilde{n}_{opt}(\boldsymbol{\alpha}) \\ \tilde{h}_{opt}(\boldsymbol{\alpha}) \end{bmatrix};$$
$$\tilde{z}_k(\boldsymbol{\alpha}) = \sigma_{k,3}\left(b_{k,3} + \sum_{i=1}^{N_h} w_{2,3}^{i,k}\, \sigma_{i,2}\left(b_{i,2} + \sum_{j=1}^{K+3} w_{1,2}^{j,i}\, \alpha_j\right)\right), \quad 1 \le k \le 2 \tag{12}$$

In course of training ANN$_3$, the values of all the weight and bias terms in (12) are tuned to minimize the error loss function [17]

$$f_{ANN_3}(\mathbf{w}, \mathbf{b}) = \underset{\mathbf{w},\mathbf{b}\in\Re}{\arg\min}\left(\frac{1}{Q}\sum_{q=1}^{Q}\left\|\mathbf{z}_q - \tilde{\mathbf{z}}\left(\alpha_q\right)\right\|_2^2\right) \tag{13}$$

Once ANN$_3$ has been sufficiently well-trained, it is clear from Fig. 2 that for any given values of the random variables $\lambda = [\lambda_i, \lambda_d]$, the corresponding optimal number and size of the repeaters $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ can be analytically obtained from the composite metamodel 1 consisting of ANN$_1$-ANN$_3$. Importantly, when utilizing the trained composite metamodel 1, there is no need for full-wave EM simulations to predict the p. u. l. parameters of the passive interconnects, nor is there any need for multiple SPICE simulations to predict the values of the RC elements making up the equivalent circuit model of the repeaters. Thus, the composite metamodel 1 can be repeated probed for different values of $\lambda$ in a Monte Carlo framework to infer the statistics of $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ in an analytic, and consequently, numerically efficient manner. Crucially, the number of full-wave EM simulations required to train the composite metamodel 1 will be significantly curtailed by using the SD and the new PKID approaches of Section II-C as opposed to a conventional training approach used in [12]. Additionally, the composite metamodel 1 will be able to include the parametric uncertainty in the FinFETs – something that was neglected in [12]. This makes the proposed composite metamodel 1 computationally more efficient as well as more reliable than the earlier work of [12].

### F. DEVELOPMENT OF COMPOSITE METAMODEL 2 USING PRIOR KNOWLEDGE INPUT WITH SPARSE POSTPROCESSING

In this subsection, the details of the proposed composite metamodel 2 are described. The block diagram of the proposed composite metamodel 2 is provided in Fig. 4. When comparing the composite metamodels of Figs. 2 and 4, three clear differences are noticed.

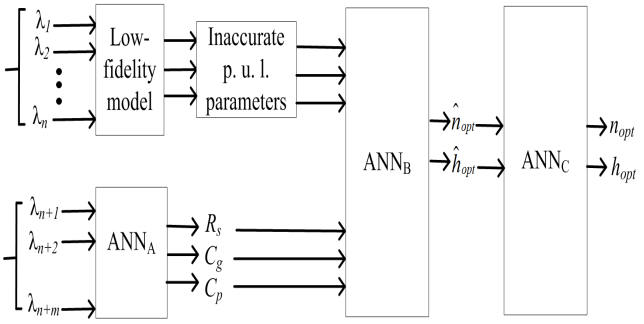



**FIGURE 4.** Block structure of the proposed composite metamodel 2 showing how the different ANNs are interconnected.

(i) First, the composite metamodel of 2 avoids the use of ANN$_1$. This is because this metamodel directly utilizes the low-fidelity results $\mathbf{y}_{LF}(\lambda_i)$ to represent the p. u. l. parameters of the hybrid copper-graphene interconnects. In so doing, it is acknowledged that these p. u. l. parameters of the interconnects will not be as accurate as those extracted from EM solvers.

(ii) Second, the ANN$_B$ of the composite metamodel 2 is very similar to the ANN$_3$ of the composite metamodel 1. The only difference between the two is that the inputs of ANN$_B$ include the inaccurate p. u. l. parameters of the interconnects obtained from the low-fidelity model $\mathbf{y}_{LF}(\lambda_i)$ instead of the accurate p. u. l. parameters predicted by the trained ANN$_1$ (see Fig. 4). Consequently, the outputs of ANN$_B$ will be the inaccurate optimal number and size of the repeaters $\left\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\right\}$instead of the accurate optimal number and size of the repeaters $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$.

(iii) Given the above differences between the two composite metamodels, it is evident that another neural postprocessing block is required to map the inaccurate optimal number and size of the repeaters $\left\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\right\}$ to the accurate optimal number and size of the repeaters $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$. This postprocessing block forms the ANN$_C$ of the proposed composite metamodel 2 as shown in Fig. 4.

In this work, it is argued that because the results of the low-fidelity model $\mathbf{y}_{LF}(\lambda_i)$ are reasonably well-correlated, even if not exactly equal to the results from the full-wave EM simulations $\mathbf{y}(\lambda_i)$, the inputs and outputs of ANN$_C$ in Fig. 4

will also be reasonably well-correlated. Thus, the training dataset for $\text{ANN}_\text{C}$ is $D_C = \left\{\hat{\xi}\left(\lambda_s\right), \xi\left(\lambda_s\right)\right\}_{s=1}^{S}$ where

$$\hat{\xi}\left(\lambda_s\right) = \begin{bmatrix} \hat{\xi}_1(\lambda_s) \\ \hat{\xi}_2(\lambda_s) \end{bmatrix} = \begin{bmatrix} \hat{n}_{opt}(\lambda_s) \\ \hat{h}_{opt}(\lambda_s) \end{bmatrix}; \xi\left(\lambda_s\right) = \begin{bmatrix} n_{opt}(\lambda_s) \\ h_{opt}(\lambda_s) \end{bmatrix} \tag{14}$$

In (14), the vector $\hat{\xi}\left(\lambda_s\right)$ contains the optimal number and size of the repeaters evaluated using the PSO algorithm where the interconnect p. u. l. parameters are the low-fidelity model results $\mathbf{y}_{LF}(\lambda_{i,s})$ and the values of the circuit elements $\{R_s, C_g, C_p\}$ are predicted by $\text{ANN}_2$ (i.e., $\tilde{\theta}(\lambda_{d,s})$) for the $s$-th training sample $\lambda_s = [\,\lambda_{i,s}, \lambda_{d,s}]$. Similarly, the vector $\xi\left(\lambda_s\right)$ contains the optimal number and size of the repeaters evaluated using the PSO algorithm where the interconnect p. u. l. parameters are extracted from full-wave EM simulations (i..e, $\mathbf{y}(\lambda_{i,s})$) and the values of $\{R_s, C_g, C_p\}$ are still predicted by $\text{ANN}_2$ (i.e., $\tilde{\theta}(\lambda_{d,s})$) for the same training sample $\lambda_s$. Next, assuming $\text{ANN}_\text{C}$ contains a single hidden layer, the outputs predicted by $\text{ANN}_\text{C}$ take the form

$$\tilde{\xi}\left(\lambda\right) = \begin{bmatrix} \tilde{\xi}_1(\lambda) \\ \tilde{\xi}_2(\lambda) \end{bmatrix} = \begin{bmatrix} \tilde{n}_{opt}(\lambda) \\ \tilde{h}_{opt}(\lambda) \end{bmatrix};$$
$$\tilde{\xi}_k(\lambda) = \sigma_{k,3}\left(b_{k,3} + \sum_{i=1}^{N_h} w_{2,3}^{i,k}\,\sigma_{i,2}\left(b_{i,2} + \sum_{j=1}^{2} w_{1,2}^{j,i}\,\hat{\xi}_j\right)\right); \quad 1 \le k \le 2 \tag{15}$$

where the inputs $\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\}$ will be available from the outputs of $\text{ANN}_\text{B}$ as shown in Fig. 4. In training $\text{ANN}_\text{C}$, the values of all the weight and bias terms in (17) are tuned to minimize the error loss function [17]

$$f_{ANN_C}(\mathbf{w}, \mathbf{b}) = \underset{\mathbf{w},\mathbf{b}\in\Re}{\arg\min}\left(\frac{1}{S}\sum_{s=1}^{S}\left\|\xi\left(\lambda_s\right) - \tilde{\xi}\left(\lambda_s\right)\right\|_2^2\right) \tag{16}$$

Similar to the composite metamodel 1, once $\text{ANN}_\text{C}$ has been sufficiently well-trained, it is clear from Fig. 4 that for any given values of the random variables $\lambda = [\,\lambda_i, \lambda_d]$, the corresponding optimal number and size of the repeaters $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ can be analytically obtained from the composite metamodel 2 consisting of $\text{ANN}_\text{A}$-$\text{ANN}_\text{C}$.

It is concluded from Fig. 4 that the majority of the time cost in training the composite metamodel 2 is incurred while training $\text{ANN}_\text{C}$. This is because extracting the training dataset for $\text{ANN}_\text{C}$ in (14) requires performing $S$ time-consuming full-wave EM simulations. However, it is pointed out that for $\text{ANN}_\text{C}$, the inputs $\left\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\right\}$ are well-correlated to the outputs $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$. This means that the map between the inputs of $\text{ANN}_\text{C}$ to the outputs will be a relatively simple one, and hence, $\text{ANN}_\text{C}$ can be trained using a very sparse set of training samples (i.e., $S$ will be very small). In effect, the composite metamodel 2 possesses the prior knowledge of the optimal number and size of the repeaters $\left\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\right\}$ required to minimize the PDP of the interconencts with the inaccurate p. u. l. parameters $\mathbf{y}_{LF}(\lambda_i)$ obtained from $\text{ANN}_\text{B}$. This prior knowledge is then used to enable the fast training of $\text{ANN}_\text{C}$ using a sparse training dataset. Hence, this approach is referred to as prior knowledge input with sparse postprocessing (PKISP). This PKISP approach is the reason why the composite metamodel 2 is trained in a far more numerically efficient manner than the conventional approach used in [12]. Hence, this metamodel too, like the proposed composite metamodel 1, is more efficient as well as more reliable than that of [12].

## III. NUMERICAL EXAMPLES

In this section, two numerical examples are presented to demonstrate the benefits of the proposed composite metamodels 1 and 2 for the UQ of the optimal number and size of repeaters required to minimize the PDP of hybrid copper-graphene interconnects. All ANNs used in these examples exhibit the multilayer perceptron architecture with a single hidden layer and the hyperbolic tangent activation function. These ANNs are trained and tested using the Statistics and Machine Learning Toolbox in MATLAB 2022a where the Levenberg-Marquardt optimizer with back-propagation is used to minimize the different error loss functions. For extraction of all p. u. l. parameters, the commercial ANSYS Q3D Extractor tool is used [20]. This tool employs a quasi-static 2D EM solver based on the finite element method (FEM) to extract the p. u. l. capacitance parameters of the interconnects. The same tool employs a quasi-static 3D EM solver using the method of moments (MoM) technique accelerated by the fast multipole method to extract the p. u. l. inductance parameters of the interconnects. For extracting the values of the RC circuit elements $\{R_s, C_g, C_p\}$ of the repeaters for a given set of values of the FinFET parameters, 6 SPICE simulations are required as per the methodology outlined in [19]. All these simulations are performed on a workstation with Intel Xeon Silver processor with maximum clock speed 2.1 GHz and 32 GB DDR4 RAM.

*Example 1:* In this example, a coupled on-chip hybrid copper-graphene interconnect network at 18nm technology node consisting of $N_c = 3$ conductors as shown in Fig. 1 is considered. All the interconnects of this network are driven and loaded with 20nm FinFET based inverters. The repeaters in the network also take the form of 20nm FinFET based inverters. All the uncertain parameters of the network are listed in Table 1. For this example, $n = 8$ parameters are considered in the passive hybrid interconnect structures and $m = 4$ parameters are considered in the active FinFET devices.

For the proposed composite metamodel 1, $\text{ANN}_1$ is trained using three different approaches – the conventional approach without any prior knowledge used in [12], the SD approach (described in Section II-C), and the PKID approach (described in Section II-C). All the three approaches use the same training dataset where the number of points in the

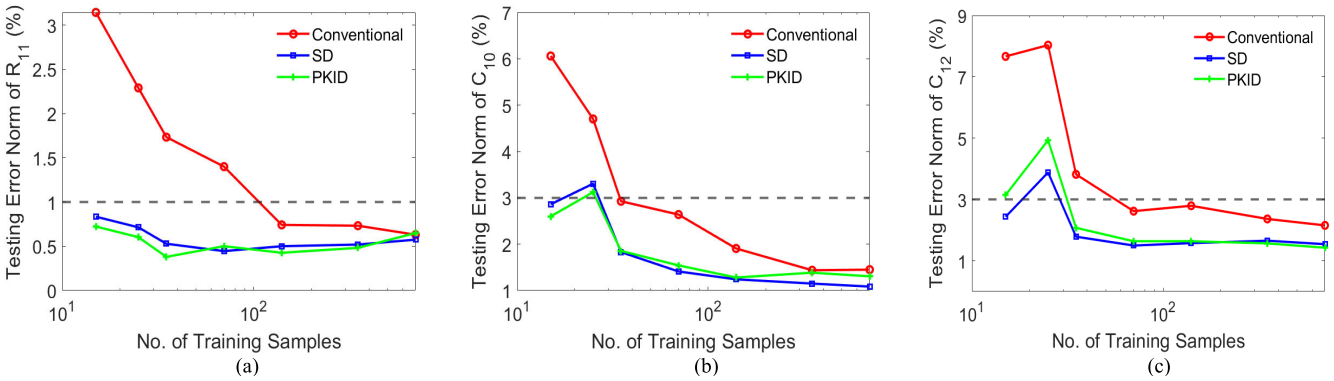



**FIGURE 5.** Scaling of the testing errors of $ANN_1$ with increasing number of training points using the conventional training approach of [12], the SD approach, and the PKID approach for Example 1. (a) Scaling of the testing error of the p. u. l. resistance of conductor 1. (b) Scaling of the testing error of the p. u. l. self-capacitance of conductor 1. (c) Scaling of the testing error of the coupling capacitance between conductors 1 and 2.

**TABLE 1.** Uncertain parameters of the interconnect network for Example 1.

| Passive Interconnects | | |
|---|---|---|
| Uncertain Parameters | Mean Values | Relative SD |
| Width (W) | 18 nm | +/- 15 % (Normal distribution) |
| Thickness (T) | 37.8 nm | |
| Height of interconnect from GND layer (h1) | 37 nm | |
| Total height of dielectric (H) | 113.8 nm | |
| Spacing between line 1 and line 2 (s1) | 13 nm | |
| Spacing between line 2 and line 3 (s2) | 17 nm | |
| Barrier layer thickness ($T_{gr}$) | 1 nm | |
| Dielectric constant ($\epsilon_r$) | 3.9 | |
| FinFET Device | | |
| Channel length | 24 nm | +/- 10 % (Normal distribution) |
| Channel doping concentration | $5x10^{23}$ $m^{-3}$ | |
| Gate oxide thickness | 1.4 nm | |
| Pitch of fin | 60 nm | |

dataset is progressively increased as $L = \{15, 25, 35, 70, 140, 350, 700\}$. The training points are selected using a Latin hypercube sampling (LHS) scheme over the $[-3\sigma\ +3\sigma]$ range of the interconnect parameters of Table 1. For testing, all the approaches use a single testing dataset comprising of 500 points uniformly distributed over the same $[-3\sigma\ +3\sigma]$ range of the interconnect parameters of Table 1. The decay of the testing error of $ANN_1$ with the increasing number of training points for all the above three approaches and different p. u. l. parameters of the interconnects is displayed in Fig. 5(a)-(c). From Fig. 5, it is clear that the learning ability of the SD and PKID approaches are significantly enhanced than that of the conventional training approach of [12]. Specifically, for this example, the minimum acceptable testing error threshold for all resistive, capacitive, and inductive p. u. l. parameters are set to 1%, 3%, and 3% respectively. It is observed from Fig. 5 that the conventional approach requires 140 training points, 7000 epochs, and 12 hidden neurons to reach this threshold while both the SD and PKID approaches do so in only a mere 15 training points (approximately) and 6000 number of epochs. The SD approach requires 5 hidden neurons and the PKID approach requires 15 hidden neurons. This means that the SD and PKID approaches uses roughly 9-10 times fewer full-wave EM simulations than the conventional training approach of [12] when training $ANN_1$. This learning efficiency is a direct outcome of exploiting the prior knowledge of the interconnects in the form of the p. u. l. parameters obtained from the low-fidelity analytic model of [16]. The faster learning ability of the SD and PKID approaches is further illustrated using the scatter plots in Fig. 6 where the number of training points is fixed at 15 for all three approaches.

Next, the training of $ANN_2$ of the proposed composite metamodel 1 is performed where the number of points in the training dataset is progressively increased as $L = \{10, 20, 30, 50, 80, 100, 200, 300, 400, 500\}$. The training points are selected using a Latin hypercube sampling (LHS) scheme over the $[-3\sigma\ +3\sigma]$ range of the FinFET parameters of Table 1. For testing, a single testing dataset comprising of 500 points uniformly distributed over the same $[-3\sigma\ +3\sigma]$ range of the FinFET parameters of Table 1 is used. For this example, the minimum acceptable RMS testing error threshold for $\{R_s, C_g, C_p\}$ is set to 0.03, 0.006, and 0.006 respectively. $ANN_2$ is able to reach this error threshold using 300 training points (1800 SPICE simulations), 400 epochs, and 6 hidden neurons as shown in Fig. 7.

The training of $ANN_3$ of the proposed composite metamodel 1 is performed as explained in Section II-D where the number of points in the training dataset is progressively increased as $L = \{10, 20, 30, 50, 80, 100, 200, 300, 400, 500\}$. The training points are selected using a Latin hypercube sampling (LHS) scheme over the $[-3\sigma\ +3\sigma]$ range of the full parameter set of Table 1. For testing, a single testing dataset comprising of 500 points uniformly distributed over the same $[-3\sigma\ +3\sigma]$ range of the full parameter set of Table 1 is used. For this example, the minimum acceptable RMS testing error threshold for $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ is set to 0.3 and 0.13,

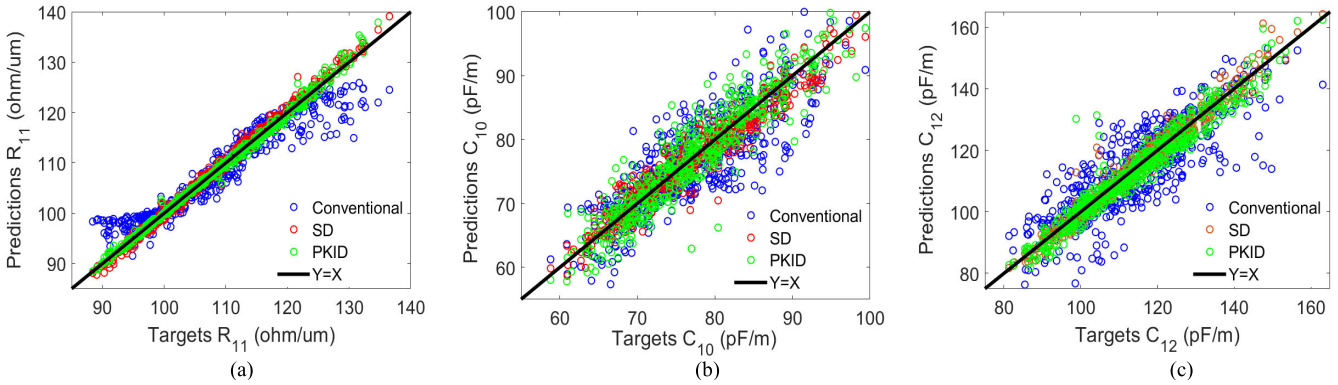



**FIGURE 6.** Scatter plots illustrating the better predictive accuracy of the SD and PKID approaches compared to the conventional training approach of [12] for Example 1 where the number of training points are fixed to 15. (a) Scatter plot of the p. u. l. resistance of conductor 1. (b) Scatter plot of the p. u. l. self-capacitance of conductor 1. (c) Scatter plot of the coupling capacitance between conductors 1 and 2.

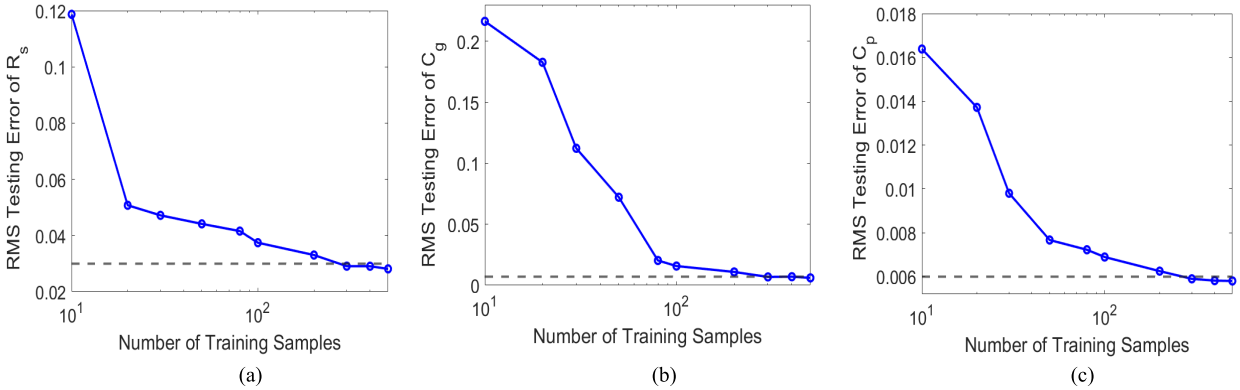



**FIGURE 7.** Scaling of the testing errors of $ANN_2$ with increasing number of training points for Example 1. (a) Scaling of the testing error of the switching resistance ($R_s$). (b) Scaling of the testing error of the gate capacitance ($C_g$). (c) Scaling of the testing error of the parasitic capacitance ($C_p$).

respectively. $ANN_3$ is able to reach this accuracy limit using 300 training points (300 PSO evaluations), 100 epochs, and 4 hidden neurons as shown in Fig. 8.

Now, coming to the proposed composite metamodel 2, $ANN_A$ is exactly the same as $ANN_2$ of the composite metamodel 1. For training $ANN_B$, the same training and testing datasets as used to train $ANN_3$ is used with the only difference being that the low-fidelity analytic model of [16] is used to predict the p. u. l. parameters of the interconnects. For this example, the minimum acceptable RMS testing error threshold for the inaccurate $\left\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\right\}$ is set to 0.2 and 0.1, respectively. As a result, 300 training points (i.e., 300 PSO evaluations), 100 epochs, and 6 hidden neurons is required by $ANN_B$ to reach this threshold. Thereafter, the mapping of the inaccurate $\left\{\hat{n}_{opt}(\lambda), \hat{h}_{opt}(\lambda)\right\}$ to the accurate $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ is performed in $ANN_C$. In $ANN_C$, the minimum acceptable RMS testing error threshold for $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ of this example is set to 0.73 and 0.21, respectively. For this example, 50 training points (i.e., 50 full-wave EM simulations plus 50 PSO evaluations), 300 epochs, and 17 hidden neurons is required by $ANN_C$ to reach this threshold as shown in Fig. 9. It is recalled that the conventional training approach of [12] where no prior knowledge was available requires 140 (i.e., roughly 3 times more) full-wave EM simulations. Next, for UQ, the mean, standard deviation (SD), and the PDF of the optimal number and size of the repeaters required $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ is determined using 5000 Monte Carlo sample points. For each Monte Carlo sample point, four methods are employed to quantify the optimal number and size of the repeaters required $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ – the proposed composite metamodels 1 and 2, the work of [12] where the parametric uncertainty in the FinFETs is neglected, and the brute-force method. In particular, the brute-force method extracts the p u. l. parameters of the interconnects using the commercial ANSYS Q3D Extractor tool [20], the values of the RC circuit elements $\{R_s, C_g, C_p\}$ modeling the repeaters using multiple SPICE simulations via the methodology of [19], and the corresponding optimal number and size of the repeaters using the PSO algorithm. The mean and standard deviation results obtained using all the four methods of above are listed in Table 2 while the PDF results are compared in Fig. 10. The results of Table 2 and Fig. 10 clearly underscore the good agreement between the proposed composite

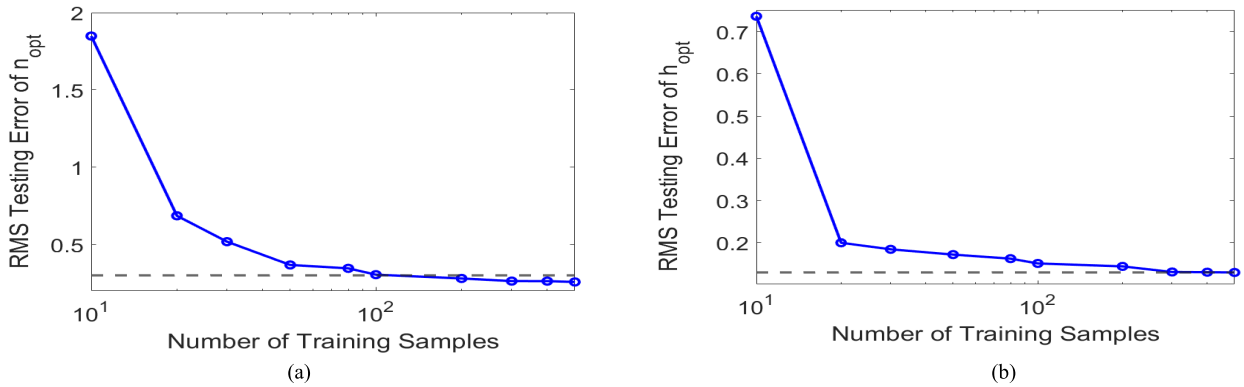



**FIGURE 8.** Scaling of the testing errors of $ANN_3$ with increasing number of training points for Example 1. (a) Scaling of the testing error of optimal number of repeaters ($n_{opt}$). (b) Scaling of the optimal size of repeaters ($h_{opt}$).

**TABLE 2.** Statistics of the optimal number and size of the repeaters required in Example 1 calculated using the proposed composite metamodels 1 and 2, the work of [12], and the brute-force method.

| | Brute-force Method | | Composite Metamodel 1 | | Composite Metamodel 2 | | Work of [12] | |
|---|---|---|---|---|---|---|---|---|
| | $n_{opt}$ | $h_{opt}$ | $n_{opt}$ | $h_{opt}$ | $n_{opt}$ | $h_{opt}$ | $n_{opt}$ | $h_{opt}$ |
| Mean | 25 | 10.12 | 25 | 10.11 | 24 | 10.96 | 22 | 9.10 |
| SD | 7 | 0.63 | 7 | 0.59 | 7 | 0.60 | 6 | 0.53 |

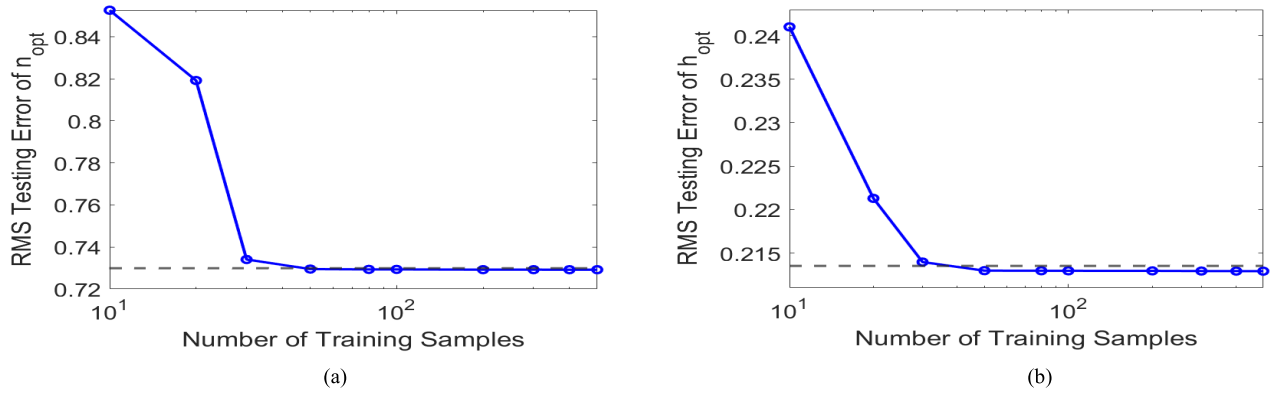



**FIGURE 9.** Scaling of the testing errors of $ANN_C$ with increasing number of training points for Example 1. (a) Scaling of the testing error of optimal number of repeaters ($n_{opt}$). (b) Scaling of the optimal size of repeaters ($h_{opt}$).

metamodels 1 and 2 and the brute-force method. However, the methodology of [12] clearly fails to provide accurate results with respect to the brute-force method. This is evidenced by the error of 12% in the mean of $n_{opt}(\lambda)$, 10% in mean of $h_{opt}(\lambda)$ and 15.87% in the SD of $h_{opt}(\lambda)$.

In Table 3, the computational time costs incurred by all the above four methods during the Monte Carlo analysis is recorded. In particular, the training, optimization, and execution time costs of all ANN metamodels involved is included for thoroughness of comparison. From the results of Table 3, it is observed that the proposed composite metamodels 1 and 2 achieve nearly two orders of magnitude in speedup over the brute-force method. Indeed, the proposed composite metamodels 1 and 2 are also more efficient than the work of [12]. Interestingly, the speedups achieved by the composite metamodels is possible due to the proposed three very distinct machine learning formulations – the SD, PKID, and PKISP formulations Moreover, it is noted that for this example, the composite metamodel 1 is much more efficient than the composite metamodel 2. This is because of two reasons. First, the composite metamodel 1 requires fewer full-wave EM simulations to train $ANN_1$ than the composite metamodel 2 to train $ANN_C$. Second, in training the composite metamodel 2, more evaluations of the PSO algorithm are required to train $ANN_B$ and $ANN_C$ unlike training the composite metamodel 1 where evaluations of the PSO algorithm is required only to train $ANN_3$.

*Example 2:* In this example, a coupled hybrid copper-graphene interconnect network at 14nm technology node consisting of three conductors is considered. Here, all

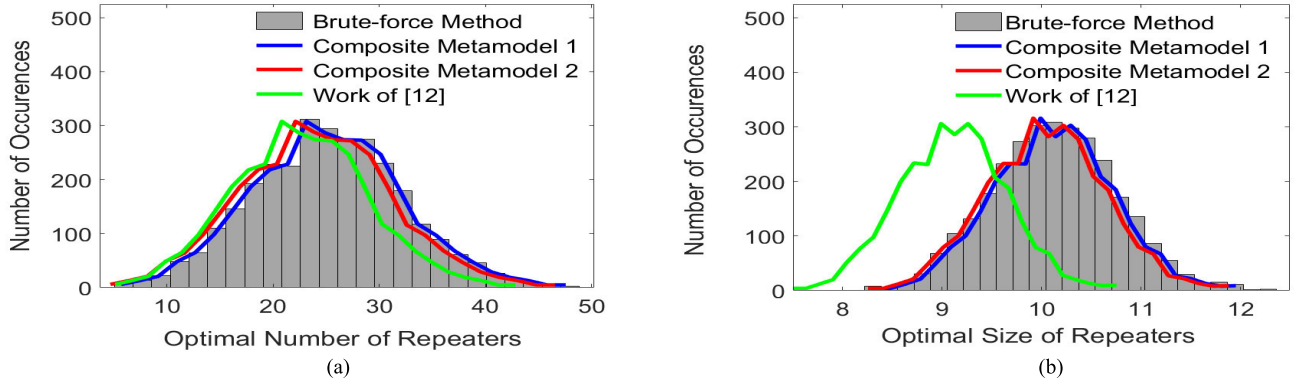



**FIGURE 10.** Probability density functions (PDFs) of the optimal number and size of the repeaters of Example 1 quantified using the brute-force method, the proposed conventional metamodels 1 and 2, and the work of [12]. (a) PDF of the optimal number of repeaters ($n_{opt}$). (b) PDF of the optimal size of repeaters ($h_{opt}$).

**TABLE 3.** Computational time cost of the brute-force method, proposed composite metamodels 1 and 2, and the work of [12] to perform uq for Example 1.

| | Training of ANNs | | | | Monte Carlo Analysis | | | Speedup |
|---|---|---|---|---|---|---|---|---|
| | Time Cost of EM Simulations (mins) | Time Cost of SPICE Simulations (mins) | Time Cost of PSO Evaluations (mins) | Time Cost for Optimization (mins) | Time Cost of ANSYS Simulations (mins) | Time Cost of SPICE Simulations (mins) | Time Cost of PSO Evaluations (mins) | |
| Brute-force Method | - | - | - | - | 15000 | 250.2 | 666.6 | - |
| Work of [12] | 420 | 0.05 | 40 | 1 | - | - | - | 34.53 |
| Composite Metamodel 1 | 45 | 15 | 40 | 1.53 | - | - | - | 156.76 |
| Composite Metamodel 2 | 150 | 15 | 46.67 | 1.46 | - | - | - | 74.67 |

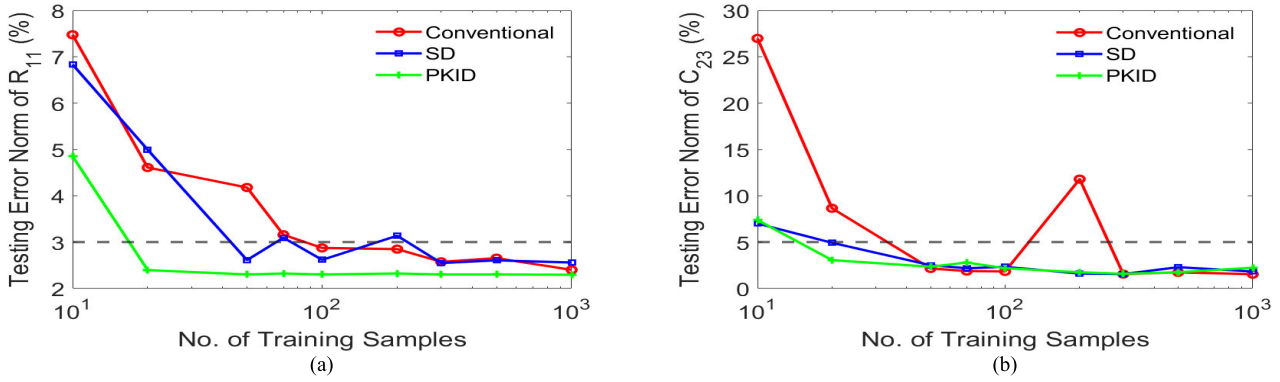



**FIGURE 11.** Scaling of the testing errors of $ANN_1$ with increasing number of training points using the conventional training approach of [12], the SD approach, and the PKID approach for Example 2. (a) Scaling of the testing error of the p. u. l. resistance of conductor 1. (b) Scaling of the testing error of the p. u. l. coupling capacitance between conductors 2 and 3.

the FinFETs are also of the 14nm technology node. All the uncertain parameters of the network are listed in Table 4. For this example, $n = 8$ parameters are considered in the passive hybrid interconnect structures and $m = 7$ parameters are considered in the active FinFET devices.

In this example, the training of all ANNs present in the proposed composite metamodels 1 and 2 proceeds in the same

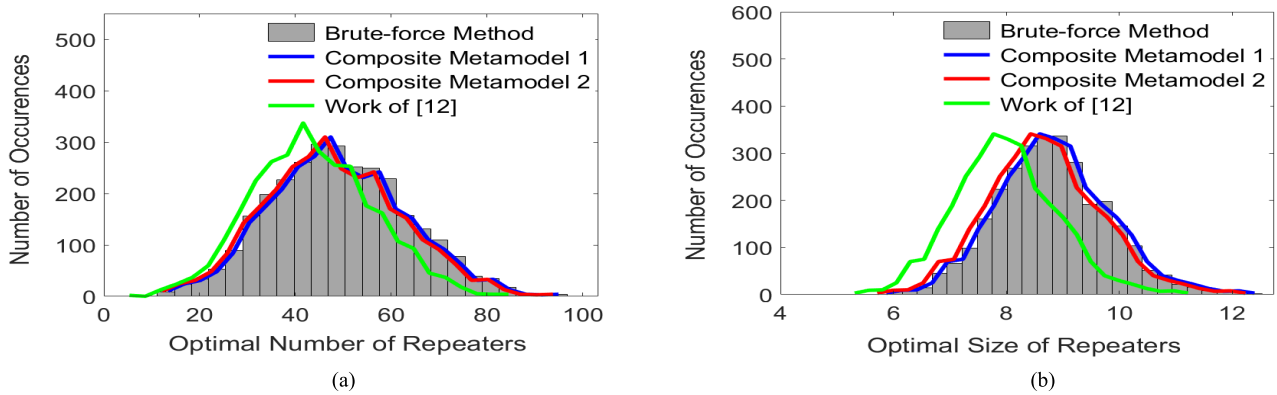



**FIGURE 12.** Probability density functions (PDFs) of the optimal number and size of the repeaters of Example 2 quantified using the brute-force method, the proposed conventional metamodels 1 and 2, and the work of [12]. (a) PDF of the optimal number of repeaters ($n_{opt}$). (b) PDF of the optimal size of repeaters ($h_{opt}$).

**TABLE 4.** Uncertain parameters of the interconnect network for Example 2.

| Passive Interconnects | | |
|---|---|---|
| Uncertain Parameters | Mean Values | Relative SD |
| Width (W) | 13 nm | +/- 25 % (Normal distribution) |
| Thickness (T) | 27.3 nm | |
| Height of interconnect from GND layer (h1) | 27 nm | |
| Total height of dielectric (H) | 82.9 nm | |
| Spacing between line 1 and 2 (s1) | 13 nm | |
| Spacing between line 2 and 3 (s2) | 13 nm | |
| Barrier layer thickness ($T_{gr}$) | 0.8 nm | |
| Dielectric constant ($\epsilon_r$) | 3.9 | |
| FinFET Device | | |
| Channel length | 24 nm | +/- 15 % (Normal distribution) |
| Channel doping concentration | $5x10^{23}$ $m^{-3}$ | |
| Gate oxide thickness | 1.4 nm | |
| Fin thickness | 10 nm | |
| Fin height | 23 nm | |
| Source/Drain concentration | $3x10^{26}$ $m^{-3}$ | |
| Height of raised source/drain on top of fin | 9 nm | |

manner as described in Example 1 with the only change being that in training $ANN_1$ the number of points in the training dataset is progressively increased as $L = \{10, 20, 50, 70, 100, 200, 300, 500, 1000\}$. The training, testing, and convergence properties of all the ANNs are listed in Table 5. To demonstrate the advantages of the SD and PKID approaches over the conventional ANN training approach without any prior knowledge as used in [12], the decay of the testing error of $ANN_1$ with increasing number of training samples for different p. u. l. parameters are compared in Fig. 11. For this example, the minimum acceptable testing error threshold for all resistive, capacitive, and inductive p. u. l. parameters are set to 3%, 5%, and 1% respectively. It is observed from Fig. 11 that the conventional approach requires 100 training points, 7000 epochs, and 10 hidden neurons to reach the prescribed error threshold for $ANN_1$. On the other hand, the SD approach can reach the same error threshold in 40 training points, 5000 epochs, and 16 hidden neurons. The PKID approach can reach the same error threshold in an even smaller 20 training samples, 5000 epochs, and 15 hidden neurons. Thus, the composite metamodel 1 needs to use 2.5-5 times fewer full-wave EM simulations than the work of [12] when training $ANN_1$ depending on whether the SD or PKID approach is used.

Now, for accuracy analysis, the mean, SD, and the PDF of the optimal number and size of the repeaters required $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ is determined using 5000 Monte Carlo sample points. The same four methods of Example 1 are again employed to quantify the optimal number and size of the repeaters required $\{n_{opt}(\lambda), h_{opt}(\lambda)\}$ at each Monte Carlo sample point. The mean and SD results obtained are listed in Table 6 while the PDF results are compared in Fig. 12. The results of Table 6 and Fig. 12 illustrate the good agreement between the results of the proposed composite metamodels 1 and 2 and the brute-force method. As expected, the approach of [12] fails to replicate the same accuracy. This is evidenced by the errors of roughly 13% in the mean of $n_{opt}(\lambda)$, 14.28% in the SD of $n_{opt}(\lambda)$, and 12.37% in the SD of $h_{opt}(\lambda)$.

In Table 7, the computational costs incurred by of all the above four methods during the Monte Carlo analysis is tabulated with the same level of detail as in Table 3. From the results of Table 7, it is clear that the proposed composite metamodels 1 and 2 can achieve substantial speedup over the brute-force method and the work of [12]. Thus, the improved performance of the proposed work, both in terms of predictive accuracy and computational time costs, over the competing work of [12] and the standard brute-force method is validated for different technology nodes and large number of network parameters.

**TABLE 5.** Training, testing, and convergence properties of all anns involved in the proposed composite metamodels 1 and 2 for Example 2.

| | $ANN_1$ | | | $ANN_2$ | | | $ANN_3$ | | $ANN_B$ | | $ANN_C$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | p. u. l. resistance | p. u. l. capacitance | p. u. l. inductance | $R_s$ | $C_g$ | $C_p$ | $n_{opt}$ | $h_{opt}$ | $\hat{n}_{opt}$ | $\hat{h}_{opt}$ | $n_{opt}$ | $h_{opt}$ |
| Error threshold (% or RMS) | 3% | 5% | 1% | 0.05 | 0.074 | 0.004 | 0.6 | 0.085 | 0.63 | 0.09 | 1.80 | 0.28 |
| # training points required | Conventional – 100<br>SD – 40, PKID - 20 | | | 200 | | | 300 | | 200 | | 30 | |
| # epochs required | Conventional – 7000<br>SD – 5000, PKID - 5000 | | | 400 | | | 100 | | 200 | | 500 | |
| # hidden neurons | Conventional – 10<br>SD – 16, PKID - 15 | | | 6 | | | 4 | | 6 | | 2 | |

**TABLE 6.** Statistics of the optimal number and size of the repeaters required in Example 2 calculated using the proposed composite metamodels 1 and 2, the work of [12], and the brute-force method.

| | Brute-force Method | | Composite Metamodel 1 | | Composite Metamodel 2 | | Work of [12] | |
|---|---|---|---|---|---|---|---|---|
| | $n_{opt}$ | $h_{opt}$ | $n_{opt}$ | $h_{opt}$ | $n_{opt}$ | $h_{opt}$ | $n_{opt}$ | $h_{opt}$ |
| Mean | 49 | 8.79 | 49 | 8.83 | 48 | 8.78 | 43 | 7.99 |
| SD | 14 | 0.97 | 14 | 1.00 | 14 | 1.01 | 12 | 0.85 |

**TABLE 7.** Computational time cost of the brute-force method, proposed composite metamodels 1 and 2, and the work of [12] to perform uq for Example 2.

| | Training of ANNs | | | | Monte Carlo Analysis | | | Speedup |
|---|---|---|---|---|---|---|---|---|
| | Time Cost of ANSYS Simulations (mins) | Time Cost of SPICE Simulations (mins) | Time Cost of PSO Evaluations (mins) | Time Cost for Optimization (mins) | Time Cost of ANSYS Simulations (mins) | Time Cost of SPICE Simulations (mins) | Time Cost of PSO Evaluations (mins) | |
| Brute-force Method | - | - | - | - | 15000 | 250.2 | 666.6 | - |
| Work of [12] | 300 | 0.05 | 40 | 1 | - | - | - | 46.67 |
| Composite Metamodel 1 (SD) | 120 | 10 | 40 | 1.53 | - | - | - | 92.25 |
| Composite Metamodel 1 (PKID) | 60 | | | 1.5 | - | - | - | 142.75 |
| Composite Metamodel 2 | 90 | 10 | 30.67 | 1.46 | - | - | - | 120.45 |

## IV. CONCLUSION

In this paper, two new multi-ANN composite metamodels are developed in order to quantify the statistics of the optimal number and size of repeaters required to minimize the PDP of hybrid on-chip copper-graphene interconnects in the presence of parametric uncertainty. The key feature of these composite metamodels is that they are able to harness the prior knowledge of the interconnects and exploit this knowledge via multiple different formulations – for example, the known SD formulation, the new PKID formulation, and the other new PKISP formulation. These formulations enable a far more efficient training of the composite metamodels than what is conventionally possible. Furthermore, the proposed composite metamodels are demonstrably more accurate and reliable than existing ANN based approaches on account of them being able to include the parametric uncertainty present in both the passive and active parts of the hybrid copper-graphene interconnect network.

## REFERENCES


[1] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, "Interconnect limits on gigascale integration," *Proc. IEEE*, vol. 89, no. 3, pp. 305–324, Mar. 2001.

[2] P. S. Peercy, "The drive to miniaturization," *Nature*, vol. 406, no. 6799, pp. 1023–1026, Aug. 2000.

[3] W.-S. Zhao, D.-W. Wang, G. Wang, and W.-Y. Yin, "Electrical modeling of on-chip Cu-graphene heterogeneous interconnects," *IEEE Electron Device Lett.*, vol. 36, no. 1, pp. 74–76, Jan. 2015.

[4] R. Mehta, S. Chugh, and Z. Chen, "Enhanced electrical and thermal conduction in graphene-encapsulated copper nanowires," *Nano Lett.*, vol. 15, no. 3, pp. 2024–2030, Feb. 2015.

[5] Y. I. Ismail and E. G. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 2, pp. 195–206, Apr. 2000.

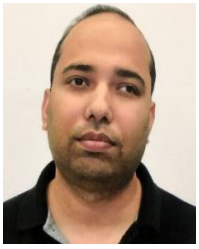
[6] K. Banerjee and A. Mehrotra, "Analysis of on-chip inductance effects for distributed RLC interconnects," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 21, no. 8, pp. 904–915, Aug. 2002.

[7] K. Banerjee and A. Mehrotra, "A power-optimal repeater insertion methodology for global interconnects in nanometer designs," *IEEE Trans. Electron Devices*, vol. 49, no. 11, pp. 2001–2007, Nov. 2002.

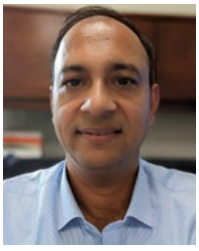
[8] F. Liang, G. Wang, and W. Ding, "Estimation of time delay and repeater insertion in multiwall carbon nanotube interconnects," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2712–2720, Aug. 2011.
[9] M. Rezaei Khezeli, M. H. Moaiyeri, and A. Jalali, "Analysis of crosstalk effects for multiwalled carbon nanotube bundle interconnects in ternary logic and comparison with Cu interconnects," *IEEE Trans. Nanotechnol.*, vol. 16, no. 1, pp. 107–117, Jan. 2017.
[10] S. Pathania, S. Kumar, and R. Sharma, "Analyzing crosstalk-induced effects in rough on-chip copper interconnects," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 9, no. 10, pp. 1984–1992, Oct. 2019.
[11] W.-S. Zhao, P.-W. Liu, H. Yu, Y. Hu, G. Wang, and M. Swaminathan, "Repeater insertion to reduce delay and power in copper and carbon nanotube-based nanointerconnects," *IEEE Access*, vol. 7, pp. 13622–13633, 2019.
[12] A. Sharif, S. Pathania, S. Kushwaha, S. Roy, R. Sharma, and B. K. Kaushik, "An artificial neural network surrogate model for repeater optimization in the presence of parametric uncertainty for hybrid copper-graphene interconnect networks," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Limoges, France, Jul. 2022, pp. 1–4, doi: 10.1109/NEMO51452.2022.10038956.
[13] M. A. El-Moursy and E. G. Friedman, "Optimum wire sizing of RLC interconnect with repeaters," *Integration*, vol. 38, no. 2, pp. 205–225, Dec. 2004.
[14] R. Kumar, S. S. L. Narayan, S. Kumar, S. Roy, B. K. Kaushik, R. Achar, and R. Sharma, "Knowledge-based neural networks for fast design space exploration of hybrid copper-graphene on-chip interconnect networks," *IEEE Trans. Electromagn. Compat.*, vol. 64, no. 1, pp. 182–195, Feb. 2022.
[15] S. Kushwaha, N. Soleimani, F. Treviso, R. Kumar, R. Trinchero, F. G. Canavero, S. Roy, and R. Sharma, "Comparative analysis of prior knowledge-based machine learning metamodels for modeling hybrid copper–graphene on-chip interconnects," *IEEE Trans. Electromagn. Compat.*, vol. 64, no. 6, pp. 2249–2260, Dec. 2022, doi: 10.1109/TEMC.2022.3205869.
[16] Z.-H. Cheng, W.-S. Zhao, D.-W. Wang, J. Wang, L. Dong, G. Wang, and W.-Y. Yin, "Analysis of Cu-graphene interconnects," *IEEE Access*, vol. 6, pp. 53499–53508, 2018.
[17] Q.-J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Norwood, MA, USA: Artech House, 2000.
[18] K. Dimple, S. Guglani, R. Kumar, S. Roy, B. K. Kaushik, S. Kushwaha, and R. Sharma, "Exploring the impact of parametric variability on eye diagram of on-chip multi-walled carbon nanotube interconnects using fast machine learning techniques," in *Proc. IEEE 72nd Electron. Compon. Technol. Conf. (ECTC)*, May 2022, pp. 1620–1625.
[19] N. H. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*. Chennai, India: Pearson, 2015.
[20] (2020). *ANSYS Q3D Extractor*. [Online]. Available: http://www.ansys.com/products/electronics/ansys-q3d-extractor

**SUYASH KUSHWAHA** (Graduate Student Member, IEEE) received the bachelor's degree in electronics and instrumentation engineering from Uttar Pradesh Technical University, Lucknow, India, in 2016, and the M.Tech. degree in photonics from the Indian Institute of Technology Roorkee, Roorkee, India, in 2019. He is currently pursuing the Ph.D. degree in electrical engineering with the Department of Electrical Engineering, Indian Institute of Technology Ropar, Rupnagar, India. His research interests include modeling and simulation of interconnects, signal integrity, and the application of machine learning in interconnect packaging.

**AVIRUP DASGUPTA** (Senior Member, IEEE) received the dual B.Tech.-M.Tech. and Ph.D. degrees from the Indian Institute of Technology Kanpur (IIT Kanpur), Kanpur, India, in 2014 and 2018, respectively. From 2018 to 2020, he was a Postdoctoral Scholar with the University of California at Berkeley, Berkeley, CA, USA. From 2019 to 2020, he was the Manager of the Berkeley Device Modeling Center (BDMC). Since 2021, he has been an Assistant Professor with the Department of Electronics and Communication Engineering, IIT Roorkee. He is currently an Assistant Professor with the Department of Electronics and Communication Engineering, IIT Roorkee, Roorkee, India. He leads the Device Research Laboratory (DiRac Lab), IIT Roorkee. He is also the co-developer of multiple industry standard compact models, including BSIM-BULK, BSIM-IMG, BSIM-CMG, BSIM-SOI, and ASM-HEMT. His research interest includes the analysis, modeling, and design of semiconductor devices and interconnects. He was a recipient of the prestigious IEEE EDS Early Career Award, in 2021.



**SOURAJEET ROY** (Senior Member, IEEE) received the B.Tech. degree in electrical engineering from Sikkim Manipal University, Gangtok, India, in 2006, and the M.E.Sc. and Ph.D. degrees in electrical engineering from the University of Western Ontario, London, ON, Canada, in 2009 and 2013, respectively. From 2013 to 2019, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA. Since 2019, he has been an Assistant Professor with the Department of Electrical and Communications Engineering, Indian Institute of Technology Roorkee, Roorkee, India, where he leads the Computational Modeling and Simulation Research Group. His current research interests include signal and power integrity analysis of integrated circuits, machine learning-based electronic design automation (EDA) of electronic packaging, and uncertainty quantification of microwave/RF circuits. He is a member of the Technical Program Committee for the IEEE Electrical Performance of Electronic Packaging and Integrated Systems Conference and the IEEE Workshop on Signal and Power Integrity. He is an Associate Editor of the IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY.



**ROHIT SHARMA** (Senior Member, IEEE) received the B.E. degree in electronics and telecommunication engineering from North Maharashtra University, India, in 2000, the M.Tech. degree in systems engineering from the Dayalbagh Educational Institute, India, in 2003, and the Ph.D. degree in electronics and communication engineering from the Jaypee University of Information Technology, Noida, Uttar Pradesh, India, in 2009.

He was a Postdoctoral Fellow with the Design Automation Laboratory, Seoul National University, Seoul, South Korea, from January 2010 to December 2010. He was a Postdoctoral Fellow with the Interconnect Focus Centre, Georgia Institute of Technology, Atlanta, GA, USA, from January 2011 to June 2012. He is currently an Associate Professor (on leave) with the Department of Electrical Engineering, Indian Institute of Technology Ropar, Rupnagar, Punjab, India. He is also a Research Professor with the School of Electrical Engineering and Computer Science, Penn State University, University Park, PA, USA. His current research interests include the design of high-speed chip–chip and on-chip interconnects, graphene-based nano interconnects, and signal and thermal integrity in high-speed interconnects. He is a Program Committee Member in all major IEEE packaging conferences, such as ECTC, EPEPS, SPI, and EDAPS. He is the Current Chair of the IEEE EPS Technical Committee on Electrical Design, Modeling, and Simulation. He is an Associate Editor of the IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY.