Prior Knowledge Accelerated Transfer Learning (PKI-TL) for Machine Learning Assisted Uncertainty Quantification of MLGNR Interconnect Networks

Asha Kumari Jakhar, Surila Guglani, Avirup Dasgupta, Sourajeet Roy

Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, India
Email: sourajeet.roy@ece.iitr.ac.in

***Abstract* — In this paper, an algorithm to combine the distinct advantages of knowledge-based training and transfer learning has been developed for the fast artificial neural network (ANN) assisted uncertainty quantification of on-chip multi-layered graphene nanoribbon (MLGNR) interconnect networks. In particular, the proposed algorithm enables different modes of information such as the values of the weights and bias terms and the predicted responses from a pre-trained secondary ANN to guide the highly data-efficient training of the primary ANN. The goal of the primary ANN is to develop a parametric model of the MLGNR interconnect responses that can be used in a Monte Carlo framework for fast uncertainty quantification.**

***Keywords* — Artificial neural networks (ANNs), high-speed interconnects, multi-layered graphene nanoribbons (MLGNRs), signal integrity, transfer learning, uncertainty quantification.**

I. Introduction

Graphene-based interconnects such as multi-walled carbon nanotubes (MWCNTs) and multi-layer graphene nanoribbons (MLGNRs) are emerging as possible substitutes for conventional on-chip copper interconnects below 10-nm technology node [1]. Therefore, there is an urgent need to quantify the effects of fabrication process variations and manufacturing tolerances on the electrical performance of MWCNT/MLGNR interconnect networks in the early design cycles. Traditionally, the brute-force Monte Carlo method based on SPICE simulations of MWCNT/MLGNR interconnect networks was used for uncertainty quantification (UQ) [2]. However, the poor convergence of Monte Carlo coupled with the high computation time cost of a SPICE simulation of MWCNT/MLGNR interconnects make this method infeasible [3].

Recently, machine learning techniques such as artificial neural networks (ANNs) have been used to emulate the responses of interconnect networks as analytic functions of the geometrical, material, and physical parameters of the network [4]. Once trained, these ANNs act as surrogate models that can be probed far more efficiently than repeated SPICE simulations of the network for UQ. Despite this basic computational advantage of ANNs, they require a massive amount of training data to reliably emulate the responses of interconnect networks. This training data is obtained from repeated SPICE simulations of the network – a task that will naturally incur massively high time costs when applied for MWCNT/MLGNR interconnects.

In order to mitigate the high training time cost of conventional ANNs for MWCNT/MLGNR interconnects, various knowledge based ANNs (KBANNs) have been reported in the literature [5]. These KBANNs crosscut the numerical efficiency of an approximate equivalent single conductor (ESC) model of MWCNT/MLGNR interconnects with the accuracy of a rigorous multi-conductor circuit (MCC) model of the same to enable significantly faster training of ANNs. More recently, a naïve transfer learning (TL) approach was also presented for MWCNT/MLGNR interconnects [6]. In the naïve TL approach, the values of the weights and bias terms obtained from training a cheap secondary ANN were transferred to the primary ANN to expedite its training where the objective of the primary ANN was to emulate the responses of the accurate MCC model of MWCNT/MLGNR interconnects.

In this paper, the naïve TL approach of [6] is further improved by combining it with a known KBANN, specifically the prior knowledge input ANN (PKI-ANN) [5], [7]. In the proposed approach, when training the primary ANN, instead of simply transferring the values of the weights and bias terms from the secondary ANN to the primary ANN, the outputs of the secondary ANN are also taken in as new inputs to the primary ANN. The idea is to utilize the responses of the ESC model representation of the MLGNR interconnect networks predicted by the secondary ANN to further guide the training of the primary ANN above and beyond the benefits of naïve TL approach of [6]. Importantly, this work has developed an approach to initialize the values of the synaptic weights arising from the new inputs of the primary ANN (i.e., the outputs of the secondary ANN) such that these weights do not impede the training of the primary ANN but rather accelerate it. The proposed approach is referred to as PKI accelerated transfer learning (PKI-TL) approach and it is able to outperform the conventional PKI method and the naïve TL method of [6].

II. Proposed Prior Knowledge Accelerated Transfer Learning Approach

A. Problem Statement

Consider an M conductor MLGNR interconnect network where each conductor consists of N_l graphene nanoribbons

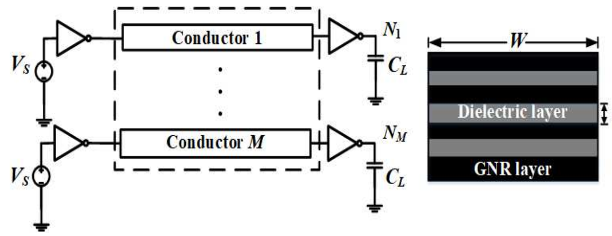

Fig. 1: Circuit schematic of the MLGNR interconnect network with cross-sectional view of MLGNR conductors.

stacked vertically and separated by dielectric layers as shown in Fig. 1. This network is driven and loaded by inverters consisting of fin-shaped field effect transistors (FinFETs). The variability present in the parameters of the interconnect structure and the FinFET devices are mapped to N mutually uncorrelated random variables $\boldsymbol{\lambda} = [\lambda_1, \lambda_2,\ldots, \lambda_N]$ located within the support Ω. Now, an ANN needs to be developed that can emulate the SI quantities of interest of the network of Fig. 1 as functions of the random variables $\boldsymbol{\lambda}$ as $\mathbf{y}(\boldsymbol{\lambda}) = [y_1(\boldsymbol{\lambda}), y_2(\boldsymbol{\lambda}),\ldots, y_P(\boldsymbol{\lambda})]$.

To train the ANN, a training dataset has to be extracted using repeated SPICE MCC model simulations of the MLGNR network of Fig. 1. Let this training dataset be expressed as $\{\boldsymbol{\lambda}^{(k)}, \mathbf{y}_{MCC}(\boldsymbol{\lambda}^{(k)})\}_{k=1}^{N_t}$ where $\mathbf{y}_{MCC}(\boldsymbol{\lambda}^{(k)})$ is the SI quantities of interest evaluated from a SPICE MCC model simulation of the network at the k-th training sample $\boldsymbol{\lambda}^{(k)} = [\lambda_1^{(k)}, \lambda_2^{(k)}, \ldots, \lambda_N^{(k)}]$. It is well known that developing a reliable ANN requires a massive number of training samples, N_t, where the time cost of even a solitary SPICE MCC simulation scale as $O(N_l^{\alpha})$ and $3 \le \alpha \le 4$ [3]. Thus, for MLGNR interconnects with even a few nanoribbons (i.e., a small N_l), the time cost to obtain the training dataset will be intractable. To address this problem, a PKI accelerated TL approach is proposed next.

B. Proposed PKI-TL Approach: Training the Secondary ANN

The first step of the proposed PKI-TL approach is to train an ANN that will emulate the target SI quantities of interest of the MLGNR network of Fig. 1 as a function of the random variables $\boldsymbol{\lambda}$ assuming an approximate ESC model representation of the network. This is referred to as the secondary ANN. The training dataset of this secondary ANN is $\{\boldsymbol{\lambda}^{(k)}, \mathbf{y}_{ESC}(\boldsymbol{\lambda}^{(k)})\}_{k=1}^{N_1}$ where $\mathbf{y}_{ESC}(\boldsymbol{\lambda}^{(k)})$ is the SI quantities of interest obtained from the SPICE ESC model simulation at the training point $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k)}$. Let the SI quantities of interest predicted by the secondary ANN be $\mathbf{z}(\boldsymbol{\lambda}) = [z_1(\boldsymbol{\lambda}), z_2(\boldsymbol{\lambda}),\ldots, z_P(\boldsymbol{\lambda})]$ where for a standard three-layer multi-perceptron architecture

$$z_i(\boldsymbol{\lambda}^{(k)}) = \sigma_{i,3}\left(b_{i,3} + \sum_{j=1}^{N_h} w_{2,3}^{j,i}\,\sigma_{j,2}\left(b_{j,2} + \sum_{p=1}^{N} w_{1,2}^{p,j}\,\lambda_p^{(k)}\right)\right) \quad (1)$$

In (1), $\sigma_{q,p}$ refers to the q-th nonlinear activation function used in the neurons of the p-th layer, $b_{q,p}$ is the bias value entering the βth neuron of the p-th layer, and $w_{q,p}^{\alpha,\beta}$ is the synaptic weight linking the α-th neuron of the q-th layer to the β-th neuron of the p-th layer. Next, all the weights and bias terms of (1) are optimized to minimize the error loss function

$$f_{Loss}(\mathbf{w},\mathbf{b}) = \frac{1}{N_t}\sum_{k=1}^{N_1}\left\|\mathbf{y}_{ESC}(\boldsymbol{\lambda}^{(k)}) - \mathbf{z}(\boldsymbol{\lambda}^{(k)})\right\|_2^2 \quad (2)$$

where $\mathbf{w}$ and $\mathbf{b}$ are weights and bias matrices in the secondary ANN. Note that the ESC model by virtue of compressing all N_l nanoribbons in a MLGNR conductor into a single nanoribbon ensures very fast SPICE simulation to generate the training dataset. Thus, the secondary ANN can be trained at very cheap time costs.

TABLE I
NORMALLLY DISTRIBUTED NETWORK PARAMETERS

No.	Parameter	Mean	Relative SD
1	Conductor width (W)	10 nm	
2	Elevation of conductor above GND	18 nm	
3	Conductor spacing	9 nm	±15%
4	Dielectric constant (HfO_2)	25	
5	Dielectric constant (SiO_2)	3.9	
6	Fermi velocity (v_F)	8×10^5 m/s	
7	Dielectric thickness (δ)	0.34 nm	±10%
8	FinFET gate length	14 nm	
9	FinFET oxide layer thickness	1.2 nm	
10	Fin pitch	28 nm	
11	Fin height	21 nm	
12	Low field mobility of N-type FinFET ($\mu_{0,n}$)	0.0568 m^2/V-s	±10%
13	Low field mobility of P-type FinFET ($\mu_{0,p}$)	0.0376 m^2/V-s	
14	Fin width	8 nm	

C. Proposed PKI-TL Approach: Transfer of Knowledge

The next step is to now train a primary ANN from the knowledge of the secondary ANN to emulate the target SI quantities of interest of the MLGNR network of Fig. 1 as a function of $\boldsymbol{\lambda}$ assuming the rigorous MCC model representation of the network. To that end, the following two conditions are proposed.

(i) The primary ANN will possess the same architecture as the secondary ANN where the initial guess of the weights and bias terms will be inherited from the trained secondary ANN. This condition is true for the naïve TL approach of [6].

(ii) The primary ANN will take as additional inputs the outputs of the secondary ANN, $\mathbf{z}(\boldsymbol{\lambda})$. This is the new PKI condition proposed to ensure the faster convergence of the primary ANN than what is possible using only condition (i).

A key challenge in imposing the PKI condition of (ii) is that the initial value of the weights of the synapses emanating from the new input neurons $\mathbf{z}(\boldsymbol{\lambda})$ of the primary ANN are unknown. These weights cannot be randomly initialized because then the initial prediction of the primary ANN may be vastly different from the prediction of the trained secondary ANN, thus

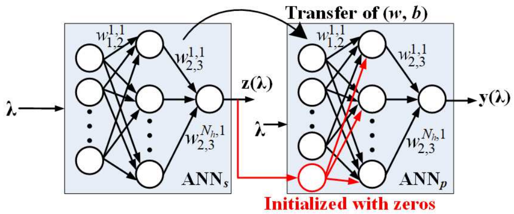

Fig. 2: Block diagram of the proposed PKI-TL approach.

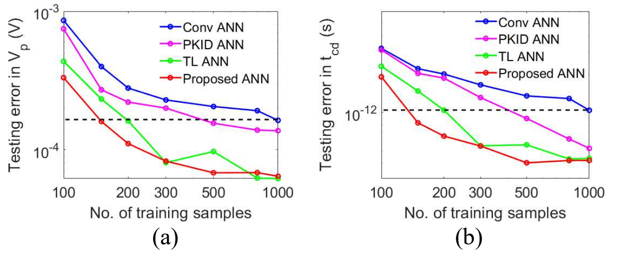

Fig. 3: Decay of the RMS testing error with the increasing number of training points for different ANNs. (a) Peak crosstalk at node N_2 and (b) crosstalk delay at node N_2.

preventing an intelligent starting point to the training of the primary ANN. To avoid this issue, in this work the synapses emanating from the new inputs $\mathbf{z}(\boldsymbol{\lambda})$ are initially set to zero as shown in Fig. 2. This ensures that the initial SI quantities predicted by the primary ANN is same as that predicted by the trained secondary ANN as in the naïve TL approach [6]. Hence, the advantage of the naïve TL approach is preserved. Moreover, after every training epoch, as the synaptic weights get updated, the impact of $\mathbf{z}(\boldsymbol{\lambda})$ will start to enter the hidden layer of the primary ANN and further accelerate its training. Thus, the proposed PKI-TL approach reaps the benefits of both the conventional PKI-ANN [5], [7] and the naïve TL approach [6].

III. Numerical Results and Discussions

In this section, a five conductor MLGNR interconnect network as shown in Fig. 1 is considered. The parametric variability in the network is listed in Table I. Lines 1, 3 and 5 are the active lines while lines 2 and 4 are quiet. The SI quantities of interest are the peak crosstalk (V_p) and the crosstalk delay (t_{cd}) for the victim line 2. The statistics of these SI quantities are evaluated using Monte Carlo analysis with 30,000 samples. The Monte Carlo analysis is performed using the following methods – the direct method using SPICE MCC model simulations, an ANN trained in the conventional sense, an ANN trained using the prior knowledge with source difference (PKID) approach because it is the fastest KBANN [5], an ANN trained using the naïve TL approach [6], and an ANN trained using the proposed PKI-TL approach. All ML techniques utilize the same training dataset of N_t = {100, 150, 200, 300, 500, 800, 1000} Latin hypercube sampling points and a common testing dataset of 1000 points. All ANN models use a single hidden layer, hyperbolic tangent activation function, and the Levenberg-Marquardt optimizer. In Fig. 3, the decay of the testing error with the increasing number of training points

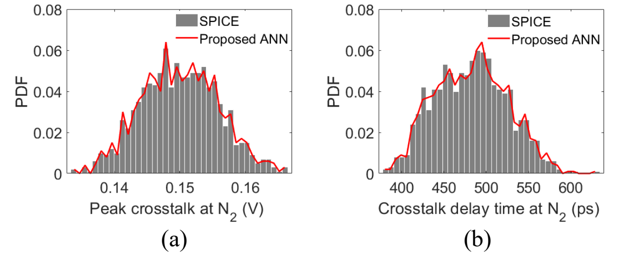

Fig. 4: PDF of (a) peak crosstalk at node N_2 and (b) crosstalk delay at node N_2 for 30,000 Monte Carlo samples.

for all the above ANNs is displayed. From Fig. 3, it is clear that the proposed PKI-TL approach requires the smallest number of training samples (150 MCC and 1000 ESC samples) compared to even the PKID and naïve TL ANNs (500 and 200 MCC samples respectively, each with a 1000 ESC samples to reach the same error threshold denoted by the black broken line). This corresponds to the proposed PKI-TL approach exhibiting the best speedup of roughly 6x over the conventional ANN, 1.3x over the naïve TL approach, and 3.3x over the PKID ANN during training. The accuracy of the proposed PKI-TL approach in predicting the full probability density function of the SI quantities of interest is illustrated in Fig. 4.

IV. Conclusion

In this work, a modified transfer learning approach that is accelerated with the prior knowledge input (PKI) formulation is proposed. This new PKI-TL approach is found to be far more data-efficient, and consequently, more time-efficient than conventional ANNs, the conventional PKID ANN, and the very recent naïve TL approach when training ANNs for the UQ of on-chip MLGNR interconnects.

References

[1] A. Naeemi and J. D. Meindl, "Performance benchmarking for graphene nanoribbon, carbon nanotube and Cu interconnects," *Proc. Int. Interconnect Technol. Conf.*, Burlingame, CA, USA, 2008, pp. 183-185

[2] A. Nieuwoudt and Y.Massoud, "On the optimal design, performance, and reliability of future carbon nanotube-based interconnect solutions," *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 2097–2010, Aug. 2008

[3] M. S. Sarto and A. Tamburrano, "Comparative analysis of TL models for multilayer graphene nanoribbon and multiwall carbon nanotube interconnects," *Proc. IEEE Int. Symp. on Electromagn. Compat.*, Fort Lauderdale, FL, USA, 2010, pp. 212-217

[4] Q.-J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Norwood, Massachusetts: Artech House 2000

[5] K. Dimple et al., "Exploring the impact of parametric variability on eye diagram of on-chip multi-walled carbon nanotube interconnects using fast machine learning techniques," *Proc. IEEE 72nd Electronic Comp. and Tech. Conf. (ECTC)*, San Diego, CA, 2022, pp. 981-986

[6] S. Guglani, K. Dimple, A. Dasgupta, R. Sharma, B. K. Kaushik and S. Roy, "A transfer learning approach to expedite training of artificial neural networks for variability-aware signal integrity analysis of MWCNT interconnects," *Proc. 31st IEEE Conf. Electrical Perform. Electronic Packag. and Systs.*, San Jose, CA, USA, 2022, pp. 1-3

[7] P. M. Watson, K. C. Gupta, and R. L. Mahajan, "Applications of knowledge-based artificial neural network modeling to microwave components," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 9, pp. 254–260, 1999