IEEE Journal of Microwaves

Received 2 September 2020; revised 14 October 2020; accepted 19 October 2020.
Date of current version 7 January 2021.

Digital Object Identifier 10.1109/JMW.2020.3033780

Simulation and Automated Modeling of Microwave Circuits: State-of-the-Art and Emerging Trends

QI-JUN ZHANG [1] (Fellow, IEEE), EMAD GAD [2] (Senior Member, IEEE), BEHZAD NOURI [1] (Member, IEEE), WEICONG NA [3] (Member, IEEE), AND MICHEL NAKHLA [1] (Life Fellow, IEEE)

(Invited Paper)

[1]Department of Electronics, Carleton University, Ottawa, ON K1S 5B6, Canada
[2]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[3]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

CORRESPONDING AUTHOR: Qi-Jun Zhang (e-mail: qjz@ doe.carleton.ca).

This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grants RGPIN-2017-06420 and RGPIN 2017-06368.

ABSTRACT Microwave modeling and simulation are essential to designing microwave circuits and systems. Although fundamental concepts and approaches for modeling and simulation are mature, the drive to higher frequencies, tighter design margins, and more functionality/complexity of circuits continue to defy the capabilities of existing modeling and simulation methods. Newer algorithms are being developed to address the speed, accuracy and robustness of design algorithms. Coupled with the advent of more powerful computers and algorithms, microwave design automations are solving much more complex problems in much shorter time than previously imaginable. This paper describes the advances and state-of-the-art in automated modeling and simulation. Automated data-driven modeling approaches covering data sampling/generation, model structure adaptation, and model training/validation are described. Simulation of nonlinear microwave circuits is described covering formulations of simulation equations and advanced solution algorithms addressing problem size, convergence speed and solution accuracy. The descriptions highlight fundamental concepts, advanced methodologies, and future trends of development.

INDEX TERMS Microwave modeling, simulation, model order reduction, neural networks, harmonic balance.

I. INTRODUCTION

Microwave modeling and simulation are two of the most fundamental steps in designing today's and tomorrow's wireless components and systems for new generations of communication networks, autonomous systems and IoT. Built on the foundation for microwave modeling and simulation established throughout the past century, microwave designers have for many decades used models and simulations to estimate the microwave circuit behavior before the circuit is built, and to complete design efficiently. However, the drive to higher frequencies, more functionality, higher performance, and tighter design margins of microwave circuits/systems have resulted in significantly increased design complexity and computational cost, defying the capabilities of existing modeling and simulation methods. Newer algorithms constantly evolve to address the increasing demand on speed, accuracy and robustness of design algorithms. At the same time, computers are becoming faster and more powerful. The combination of newer computational algorithms and modern generation of computers is enabling microwave designers to solve much more complicated and challenging modeling and simulation problems not imaginable decades ago.

With ongoing progress in microwave design automation, an increasing scope of microwave design aspects are being simulated by computational algorithms. Various modeling and simulation tasks which previously required an often expensive trial-and-error process are being formulated into computational algorithms, so that problems can be solved

systematically and in a more automated way. Designers, in the pursuit of a quick closure to their design process, rely on the computer-aided design tools to perform a simulation of the design at the nominal values for the key design parameters. Besides the simulation of the circuit, the designer will often seek to optimize its performance through improving its key figure-of-merit or by making it fit within a pre-specified set of design constraints.

Simulation of microwave circuits requires accurate models of all devices in the circuit, where the outputs from device models are used in the overall circuit simulation equations. Detailed models for devices lead to good accuracy and are useful for accurate simulation. This accuracy requirement is typically motivated by the need to take into account the accurate behaviour of the devices under all possible environmental conditions. For microwave circuit/system design, on the other hand, simulations may need to be performed repeatedly. In this context, fully detailed device models may become too cumbersome. Instead, models that are fast to evaluate but with equally accurate characteristics are more desirable.

The purpose of this article is to outline the current state-of-the-art in modeling and simulations of microwave and RF circuits and devices. Automated modeling algorithms are described highlighting data sampling/data generation, model structural adaptation, and model training/validation. Advanced simulation algorithms are described with an emphasis on the tasks more often used in the context of nonlinear microwave and RF circuits, e.g., distortion analysis, intermodulation analysis, gain compression, noise analysis, etc. Those tasks typically require finding the steady-state response to a periodical or quasi-periodical stimulus in the frequency-domain using the harmonic balance (HB) approach.

The paper is organized as follows. Section II presents the state-of-the-art approaches used in automated modeling. Section III discusses applications of automated modeling for passive and active microwave devices and circuits. Section IV presents the mathematical formulation of the HB problem and Section V describes the various fundamental and advanced approaches used in the HB analysis. Section VI discusses some emerging trends with a significant potential to impact the current approaches to modeling and simulations.

II. AUTOMATED MODELING TECHNIQUES

Modeling is fundamental for microwave design automation. Models are mathematical representations of the microwave devices or subsystems, and are building blocks needed for simulation, optimization and design. Traditional approaches to develop microwave models are either based on physics laws, analytical/semi-analytical equations, equivalent circuits or empirical approximations. Model development has traditionally been based on physical insight of the problem and is a human intensive process. While the physical insight for a device model is valuable for design especially at device level design, faster models representing device behavior have gained increased use for higher level designs such as circuit and system level simulation and design. With the tremendous progress in design automation in the recent decades, computer-based algorithms for behavior modeling have evolved which makes the model development faster, more systematic and efficient. Automated modeling based on computational algorithms is nowadays one of the significant directions in design automation.

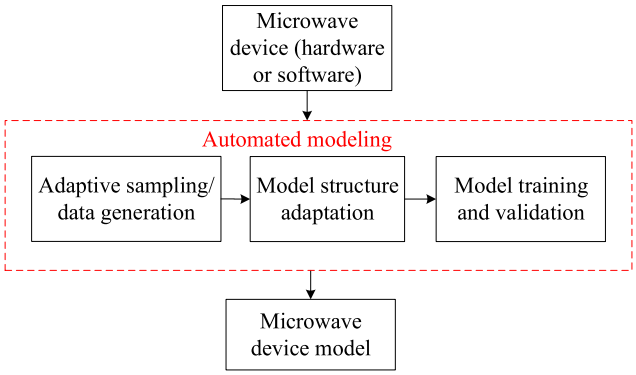


FIGURE 1. Flowchart of automated modeling process.

Automated modeling proceeds in a stepwise manner. Examples of the steps are: automated data sampling/data generation, automated model structure adaptation, automated training and model validation. Fig. 1 shows the flowchart of automated modeling process. For a given microwave device, automated sampling/data generation is firstly performed to generate training and validation data for model development. Then the model structure is formulated and refined through automated model structure adaptation. The model is trained and validated using the data samples obtained during automated sampling/data generation process. Finally, after model validation, a microwave device model with user-specified accuracy is produced.

One of the factors that stimulated automated modeling is the development of artificial neural networks for microwave modeling and design [1]–[3]. Significant speed-up of computer-aided-design by using neural network models in place of CPU-intensive electromagnetic (EM)/physics models resulted in a drive to develop advanced neural modeling techniques, including the automated microwave modeling techniques [4]–[6]. These techniques automate the neural network model development process by integrating all subtasks like data sampling/data generation, neural network model structure adaptation, training, and validation into a unified framework. By using automated modeling algorithm, microwave designers can produce models more efficiently and systematically, helping to shorten the cycle of microwave design.

In the following subsections, key aspects of the automated modeling techniques are described.

A. AUTOMATED DATA SAMPLING

In microwave modeling, data sampling is the starting point of modeling information, and its quality affects the efficiency and accuracy of the data-driven models [7]–[10], such as neural network, knowledge-based neural network, support

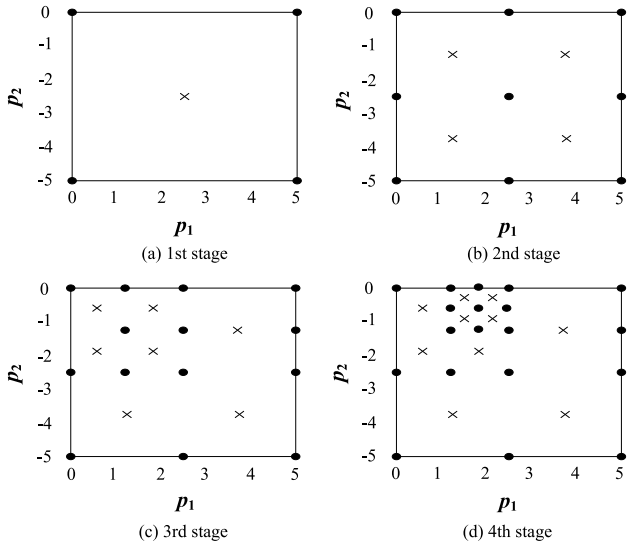


FIGURE 2. Intertwined automated distribution of training samples (•) and validation samples (×) obtained by automated sampling algorithm for a two-input modeling example [4]. The horizontal and vertical axises represent two inputs of the model. The algorithm identifies nonlinear subregions of the input space and automatically generates more samples in such regions [4].

vector machine (SVM) and Gaussian process (GP). While too many samples are expensive (e.g., measurement data and three-dimensional EM simulations), too few samples lead to over-learning of models. Therefore, automated data sampling algorithms [4], [5] are important for systematically determining the number of samples needed for developing a model with desired accuracy and their distribution in the model input space.

Let $\mathbf{p}$ and $\mathbf{y}$ be vectors of model inputs and outputs, where $\mathbf{p}$ represents a vector containing $\mathbf{N}_p$ physical parameters of a microwave component and $\mathbf{y}$ represents a vector containing the outputs of the component model under consideration. The data sampling process proceeds stage by stage. To begin with, the data sampling algorithm regards the original bounded $\mathbf{N}_p$-dimensional inputs space ($\mathbf{p}$-space) of interest as one region, denoted as $\mathbf{R}_0$. Let $\mathbf{D}_T^k$ and $\mathbf{D}_V^k$ be defined as the training data set and validation data set for the k^{th} stage during data sampling process. Let $\mathbf{S}^k$ represent the model developed using the data at stage k. For the first stage (i.e., $k = 1$), training data ($\mathbf{D}_T^1$) and validation data ($\mathbf{D}_V^1$) are systematically generated in $\mathbf{R}_0$ in a predefined way as illustrated in Fig. 2(a). A first stage model $\mathbf{S}^1$ with a simple model structure is trained with data in $\mathbf{D}_T^1$. The resulting model is validated (tested) with data in $\mathbf{D}_V^1$. Let $\mathbf{E}_v$ be defined as the validation error which represents the error between model and all validation data at the present stage. Let $\mathbf{E}_d$ represent the corresponding accuracy threshold required from a user. The algorithm stops if the validation error of the current model satisfies the user-desired threshold, i.e., $\mathbf{E}_v \leq \mathbf{E}_d$. Otherwise, based on the model error criteria, over-learning of the model is detected and the algorithm moves to next stage by adding new data. New training data are added by decomposing the region into several subregions. New validation data are also added in each subregion. For the k^{th} stage, the validation sample $\mathbf{p}^* \in \mathbf{D}_V^k$ where the model has the largest error is identified by

$$\mathbf{p}^* = \arg\max_{\mathbf{p} \in \mathbf{D}_V^k} \mathbf{E}(\mathbf{p}) \tag{1}$$

where $\mathbf{E}(\mathbf{p})$ is defined as the error between the model at the present stage and a specific sample of validation data where $\mathbf{p} \in \mathbf{D}_V^k$. The subregion with the largest validation error is defined as the worst subregion $\mathbf{R}^*$, to which $\mathbf{p}^*$ belongs. Then in the next stage, $\mathbf{R}^*$ is further divided (split) into multiple new subregions. With these new subregions, $\mathbf{D}_T^k$ and $\mathbf{D}_V^k$ are updated, i.e.,

$$\mathbf{D}_T^{k+1} = \mathbf{D}_T^k \cup \mathbf{D}_T^{\text{new}} \tag{2}$$

$$\mathbf{D}_V^{k+1} = \mathbf{D}_V^k \cup \mathbf{D}_V^{\text{new}} \tag{3}$$

where $\mathbf{D}_T^{\text{new}}$ and $\mathbf{D}_V^{\text{new}}$ represent the sets of new data samples to be generated. The new training samples in $\mathbf{D}_T^{\text{new}}$ are located at the vertices of the new subregions. The new validation samples in $\mathbf{D}_V^{\text{new}}$ are located at the center of each new subregion. Fig. 2 shows an example of automated sampling process in a 2-dimensional input space after 4 stages. The intermediate model $\mathbf{S}^k$ is trained and validated using data in $\mathbf{D}_T^k$ and $\mathbf{D}_V^k$ in each stage until $\mathbf{E}_v \leq \mathbf{E}_d$, when the automated sampling process stops. As an example, such data sampling process has been used for neural network model development [4]. The automated sampling algorithm determines the numbers of training and validation samples and the sample distribution. Using such an algorithm, nonsmooth subregions will automatically end up with dense training data, while smooth subregions will automatically end up with sparse training data. In this way, data are generated only where they are most needed, minimizing cost of data generation while maintaining sufficient modeling information from data.

Interpolation techniques can be incorporated into data sampling algorithm to enhance efficiency [5]. Simple and local interpolation for each subregion can be used to replace intermediate models $\mathbf{S}^k$ and to evaluate the validation errors of various subregions to assess the adequacy of data samples. In this way, this advanced automated sampling method [5] takes advantage of availability of training data to produce localized interpolation, avoiding the training of intermediate model during data sampling, and speeding up model development.

B. AUTOMATED MODEL STRUCTURE ADAPTATION

For the development of data-driven models (such as neural network, knowledge-based neural network, SVM and GP), the model structure adaptation will help to achieve accurate, smooth and compact models avoiding over/under-learning issues. Several techniques are available, e.g., [4]–[6] to make the process of model structural adaptation systematic and efficient.

For example, the methods for neural network structure adaptation can be training with regularization or the use of over/under-learning criterion. In [4], automated neural model

structure adaptation algorithm was firstly applied to develop multilayer perceptrons (MLPs). Starting with a small-size MLP, the automated structure adaptation algorithm proceeds stage by stage. When the algorithm detects under-learning (i.e., training error is large) during neural network training process, it dynamically adds more hidden neurons in next stage to provide increased freedom in the model to better learn the nonlinearities in training data. The model structure adaptation process stops when $\mathbf{E}_v \leq \mathbf{E}_d$. The algorithm can also reduce the number of hidden neurons if good learning is detected while the validation error can continue remain within accuracy requirement [5]. This procedure allows the guess of the initial number of hidden neurons of MLP to be more flexible, and make the final model more compact.

Knowledge-based neural network modeling techniques [11]–[15] combine existing knowledge of microwave models (such as semi-analytical or empirical models) with neural network structures. In such models, the neural network learns the mapping between the spaces of knowledge models and the training data, i.e., space mapping [16] [17]. Such mapping can be at the input side of the knowledge model (input mapping), output side of the knowledge model (output mapping) or in parallel with the knowledge model (difference mapping). The mapping can also be applied to the frequency variable (frequency mapping). Any of these mappings can be either linear or nonlinear. The selection of mapping structure of a knowledge-based neural network model depends on several factors, such as the complexity of the specific modeling problem, the quality of the empirical model, and the modeling range of input parameters. Since the mapping structure is problem dependent, the development of automated model structure adaptation algorithm for knowledge-based neural network is very important. A genetic algorithm can be used to determine the topology of neural network mapping structure and the empirical model [18].

A further advance is an automated model structure adaptation algorithm for knowledge-based models [6]. A hybrid knowledge-based model structure is formulated to encompass input mapping, output mapping and frequency mapping with both linear and nonlinear mapping capabilities in all the mapping modules, as shown in Fig. 3. The error function for training includes error of the model versus training data and regularization terms with neural network weights. A two-stage training method with such formulation of error functions using l_1 optimization determines all the mappings in the knowledge-based model. The l_1 optimization has the distinctive property for feature selection within the training process. At the end of l_1 optimization, some weights in the mapping neural networks are zeros while others remain nonzeros. Zero weights mean that the corresponding parts of the mapping can be ignored and deleted. Using this property, l_1 optimization solutions are formulated to indicate whether a mapping is linear or nonlinear, and whether a mapping should be input mapping, frequency mapping, or output mapping. Compared to traditional knowledge-based models with a fixed mapping structure, this method uses a computational approach to adjust the mapping structure to achieve an accurate model with the most suitable and compact structure. It is a systematic technique and helps further speeding up the process of developing a knowledge-based neural network model.

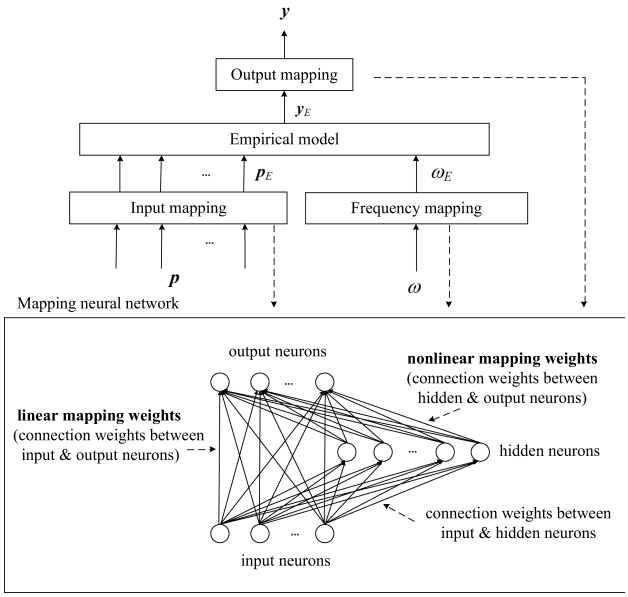


FIGURE 3. A hybrid knowledge-based model structure [6]. The overall model includes the empirical model (representing microwave knowledge of the component), the input mapping, output mapping and frequency mapping. Each of the three mappings is formulated as a neural network represented by the bottom half of the figure.

C. AUTOMATED TRAINING AND VALIDATION

During the automated modeling process for microwave applications, model training and verification are performed either as part of the data sampling stage [4], and/or in subsequent stages after data sampling has completed [5], [6]. Each training/validation stage will produce a training error and a validation error. For neural network and knowledge-based neural model, the training and validation are performed through an optimization process such that neural network internal synaptic weight parameters are adjusted to minimize training/validation errors. For SVM and GP models, the training and validation are performed to adjust the weights in SVM and hyper parameters in GP such that the training/validation errors are minimized to satisfy user-desired model accuracy. The automated modeling process terminates when the validation error satisfies accuracy requirement. The final model is valid in the entire modeling region represented by validation data.

D. ADDITIONAL METHODS FOR MODEL GENERATION

Well-established methods for microwave modeling are physics-based modeling, analytical/semi-analytical modeling, equivalent circuit/empirical modeling, table-lookup, among others. For specific applications, equivalent circuit models can be derived from electromagnetic analysis.

The automated modeling approaches described in this paper are focused on behavioral modeling for microwave simulations which are normally performed in the frequency domain. The emphasis is on data-driven behavioral model generation with multidimensional model inputs. These developments complement the synthesis approach for modeling, where equivalent circuit models are derived from circuit responses such as S-parameters [19]–[24]. Such synthesis approaches are widely used in commercial software presently. One of the motivations for the synthesis approach is the need of the model for time-domain simulations such as signal integrity and EMC analysis for high-speed IC packages and interconnects. Another motivation is to develop equivalent circuit models to provide physical insights of the components. An example of the synthesis process is to first express the frequency domain response in rational function format (such as pole/residue format), and then convert the pole/residue information into lumped equivalent circuit, e.g., [25]–[28]. Another example of the synthesis process is a fast method to synthesize each port-to-port connection as a branch model selected systematically out of all possible combinations of resistors, inductors and capacitors up to a certain level of complexity, subsequently forming a multiport lumped equivalent circuit model which is the best of a huge number (e.g., hundreds of millions) of possible models [29]. The realizability of the resulting equivalent circuit model in simulators has been an issue of investigation, e.g., [27]. The advances in SPICE-type simulators including the ability to handle negative elements, and to directly accept frequency-domain transfer functions, have helped in solving some of the important realizability issues. The synthesis process often involves the enforcement of passivity and causality of the model, which are crucial for time-domain simulations [20], [27], [30]. However, passivity and causality play an insignificant role in typical microwave simulations because such simulations are normally in the frequency domain. Therefore, detailed discussions on passivity-related investigations are not the focus of the present paper.

The focus of the present paper is on microwave-oriented modeling and simulation which are typically performed in the frequency domain. In such a situation, the S-parameters for passive components can be directly supplied to microwave simulators, without having to be converted into lumped equivalent circuits. For the passive modeling case, our emphasis is on addressing the challenges of evaluating the EM behavior for repetitive changes in geometrical design variables. Subsequently we focus our description on automated process for multidimensional parameterized modeling where multiple geometrical variables are the model inputs, and S-parameter outputs from the model can be directly supplied to simulators such as harmonic balance simulators.

The equivalent circuit models obtained from the synthesis methods can be utilized in the model generation methods described in this paper to relate the device behavior to multidimensional design variables (such as geometrical design variables). In this case, our aim is for the resulting models to provide fast answers of component behaviors for repetitive changes in component geometrical variables. In this sense, the equivalent circuit models especially with physical insights of the components [31] can be used as microwave knowledge such that mapping relationships between multidimensional geometrical variables and equivalent circuit parameters can be established.

The automated modeling approaches described in this paper provide systematic solutions for modeling new microwave devices when the existing methods became constrained in modeling accuracy, development time/cost, or flexibility in permitting repetitive changes in values of geometrical variables. Data driven modeling approaches used in the automated model generation process help address the gaps between existing device models and device data containing variations of device behavior with respect to multidimensional model input variables. Apart from the approaches described, another important approach is vector fitting based approach to parameterized macromodeling, where polytopic descriptor form [32] or state space form functions [33] are developed from given training data. The method when used to model passive microwave components, can preserve passivity. The order of the model is a user-controllable parameter that can be selected to help achieving required accuracy needs efficiently [32]–[37].

III. APPLICATIONS OF AUTOMATED MODELING TECHNIQUES

Two major categories of applications for automated modeling techniques are parametric EM modeling for passive components/circuits, e.g., [4]–[6], and nonlinear modeling for active components/circuits, e.g., [38]–[41].

A. AUTOMATED MODELING FOR PARAMETRIC EM MODELING

For parametric EM modeling, automated sampling algorithm drives EM simulators with varying values of geometrical parameters to generate training/validation data. The values of geometrical parameters for the EM structure are sampled according to data sampling algorithm. Typical outputs of the model, i.e., $\mathbf{y}$, can be scattering parameters which are used by circuit and system level simulations. Types of data-driven models can be, such as neural network, knowledge-based neural network, SVM [42] and GP [43]. Different types of models are suitable for different modeling cases. SVM and GP models have good generalization capability when the training data are limited [44], [45], while the neural network model is well suited to the case when the amount of training data is relatively large. Neural network is also used in complicated microwave modeling problems with a big database, such as inverse modeling with multi-valued solutions[46]. Deep neural networks are used to solve high-dimensional microwave modeling problems [47]. In addition, the extrapolation properties of the models [48]–[52] can also be considered with the model structure for a specific modeling problem. After training and validation, fast, accurate and compact models representing

EM behavior for varying values of geometrical parameters are produced. In addition to parametric EM modeling, application of data driven modeling is also exploited for multi-physics oriented modeling [53]–[55].

B. AUTOMATED MODELING FOR NONLINEAR DEVICES

For nonlinear device modeling, data samples for model training and validation are usually measurement data. For example, DC, small-signal, and large-signal data of the active devices can be measured. The measurements can be performed under different quiescent biases, pulsed biases, temperatures, frequencies, input powers and load impedances. After data generation, the model structures are formulated and refined with structure adaptation. Dynamic neural network, recurrent neural network, time-delay neural network and knowledge-based neural model are commonly used neural model structures for data-driven nonlinear device modeling. The outputs of the model, i.e., $\mathbf{y}$, typically are the terminal currents and charges of the device needed by the circuit level simulations.

In [38], nonlinear modeling for GaN transistors using measurement data from Nonlinear Vector Network Analyzer (NVNA) is presented. To include trapping and self-heating effects, temperature and two new dynamical input variables (i.e., gate trapping and drain trapping) are added to act as the knowledge part of the entire neural model. Using the measurement data and selected model structure, the model is trained and extensively validated on an advanced millimeter GaN FET. The final model can predict the broad-band, linear and high power nonlinear behavior of the GaN FET, including harmonic and intermodulation distortion and load-pull performance over the full range of device operation. This neural network based approach is currently one of the most accurate and systematic modeling approaches for solving the complicated task of GaN device modeling.

By taking advantage of the vast set of empirical/equivalent circuit models already available for microwave devices, knowledge-based approach can train neural networks to learn the mappings between the empirical/equivalent models and the device data. This provides a systematic way using computer-based algorithms to modify the behavior of existing models to match that demanded from new device data. Such an approach, called Neuro-Space Mapping (Neuro-SM), has been used for device modeling as well as statistical modeling [39], [40]. A further advance to Neuro-SM is the use of decomposed mapping for GaN HEMTs with trapping effects [41]. By moving the mapping space into the internal branches of the knowledge model (equivalent circuit), separate mappings for individual branches can be developed based on their vastly different behaviors. Therefore, special behaviors, such as trapping effects and frequency dispersion in GaN HEMTs, can be mapped separately and effectively. The trainings are applied to multiple mapping modules, achieving good accuracy of the overall model under DC, pulse I-V, small-signal, and large-signal conditions.

IV. HARMONIC BALANCE SIMULATION

The course of designing microwave and RF circuits, such as power amplifiers, oscillators and mixer circuits, often requires rigorous analysis to compute fundamental design metrics. Examples of those metrics, to name a few, include gain, gain compression, intermodulation products and noise figure. Computing those metrics requires a frequency-domain analysis. With the devices in the circuits being essentially modelled using nonlinear equations, the harmonic balance (HB) approach has emerged as one of the most efficient approaches to compute those metrics. Development of the HB was originally intended for frequency-domain analysis of microwave circuits. It has become one of the major research topics in the area of computer-aided design tools at large. This section describes the basis for the mathematical formulation of the HB problem, while the solution approaches are provided in the next section.

A. FORMULATION OF GENERAL CIRCUIT EQUATIONS

A general nonlinear circuit is typically formulated in the time domain using the Modified Nodal Analysis (MNA) [56] as follows

$$\frac{\mathrm{d}\mathbf{q}(\mathbf{x}(t))}{\mathrm{d}t} + \mathbf{i}(\mathbf{x}(t)) = \mathbf{b}(t) \tag{4}$$

where $\mathbf{x}(t) \in \mathbb{R}^N$ is a vector of node voltages, appended by currents in inductors and independent voltage sources, $\mathbf{i}(\mathbf{x}(t)), \mathbf{q}(\mathbf{x}(t)) \in \mathbb{R}^N$ are vectors representing, respectively, the currents and charges/fluxes in (linear or nonlinear) memoryless and memory elements, $\mathbf{b}(t) \in \mathbb{R}^N$ is a vector representing the independent stimuli, and N is the size of the MNA formulation.

One should note there that $\mathbf{i}(\mathbf{x}(t))$ and $\mathbf{q}(\mathbf{x}(t))$ are obtained based on models of the devices that could be constructed through physics-based models, or they could be based on the automated modeling techniques described in Section II.

There are two main classes of circuits that are encountered within the context of microwave or RFIC:

1) The input stimuli is a set of periodical sources having a set of W independent tones $\omega_i, (i = 1, \ldots, W)$ [1] and the objective is to compute the circuit response in its periodic or quasi periodic steady-state [57] without the transient phase. The circuits under this class are typically labelled *non-autonomous* circuits.
2) The independent stimuli are DC sources, but the circuit is designed such that its response settles into a periodic steady-state with some unknown angular frequency ω_0. Those types of circuits are usually called *autonomous* circuits.

We consider the problem formulation for each class independently.

[1] $\omega_i, (i = 1, \ldots, W)$ is said to be independent if there are no d_i's such that $\sum_{i=1}^{W} d_i \omega_i = 0$.

B. HB FOR NON-AUTONOMOUS CIRCUITS

In non-autonomous circuits, the HB approach to compute the quasi-periodic response is to expand all t-dependent quantities in (4) as a generalized Fourier series. For example,

$$\mathbf{x}(t) = \mathbf{X}_0 + \sum_{\boldsymbol{\kappa}\in\Lambda_G} \mathbf{X}_{\boldsymbol{\kappa}}^{[C]} \cos\left(\boldsymbol{\kappa}^\top \boldsymbol{\omega} t\right) + \mathbf{X}_{\boldsymbol{\kappa}}^{[S]} \sin\left(\boldsymbol{\kappa}^\top \boldsymbol{\omega} t\right) \quad (5)$$

where $\mathbf{X}_0, \mathbf{X}_{\boldsymbol{\kappa}}^{[C]}, \mathbf{X}_{\boldsymbol{\kappa}}^{[S]}$ are the Fourier coefficients and $\boldsymbol{\kappa} \in \mathbb{Z}^W$ is a multi-index[2], while $\boldsymbol{\omega}$ is given by $\boldsymbol{\omega} \in \mathbb{R}^W = [\omega_1, \omega_2, \ldots, \omega_W]^\top$, (superscript $\top$ represents the matrix transpose), which is a column-wise vector of the simulation tones, Λ_G is a set of multi-indices in $\mathbb{Z}^W$ with G elements. Typically Λ_G is created by including all multi-indices $\boldsymbol{\kappa} \in \mathbb{Z}^W$ in a box- or diamond-shaped regions in $\mathbb{Z}^W$ [58].

Expansions for $\mathbf{i}(\mathbf{x}(t))$, $\mathbf{q}(\mathbf{x}(t))$, and $\mathbf{b}(t)$ similar to (5) can be written with coefficients $\mathbf{I}_0, \mathbf{I}_{\boldsymbol{\kappa}}^{[C]}, \mathbf{I}_{\boldsymbol{\kappa}}^{[S]}, \mathbf{Q}_0, \mathbf{Q}_{\boldsymbol{\kappa}}^{[C]}, \mathbf{Q}_{\boldsymbol{\kappa}}^{[S]}$, and $\mathbf{B}_0, \mathbf{B}_{\boldsymbol{\kappa}}^{[C]}, \mathbf{B}_{\boldsymbol{\kappa}}^{[S]}$, respectively. The HB problem is formulated using the Galerkin projection which is carried out by substituting all the Fourier series expansions in (4) and requiring that the residual be orthogonal to all the basis functions used in the expansion. This process produces a system of nonlinear equations in $\bar{\mathbf{X}}$ defined by

$$\boldsymbol{\Phi}\left(\bar{\mathbf{X}}\right) \triangleq \bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Gamma}}\bar{\mathbf{Q}}\left(\bar{\mathbf{X}}\right) + \bar{\boldsymbol{\Gamma}}\bar{\mathbf{I}}\left(\bar{\mathbf{X}}\right) - \bar{\mathbf{B}} \quad (6)$$

where $\bar{\mathbf{X}}$, and $\bar{\mathbf{B}} \in \mathbb{R}^{NH}$ are vectors containing the Fourier coefficients $\mathbf{X}_0, \mathbf{X}_{\boldsymbol{\kappa}}^{[C]}, \mathbf{X}_{\boldsymbol{\kappa}}^{[S]}$ and $\mathbf{B}_0, \mathbf{B}_{\boldsymbol{\kappa}}^{[C]}, \mathbf{B}_{\boldsymbol{\kappa}}^{[S]}$, respectively, ordered by grouping all the Fourier coefficients of a single MNA variable in a contiguous subvector[3] of size H, where $H = 2G + 1$. $\bar{\mathbf{I}}(\bar{\mathbf{X}})$, $\bar{\mathbf{Q}}(\bar{\mathbf{X}})$, and $\bar{\boldsymbol{\Omega}}$ can all be described succinctly with the help of the Kronecker operator $\otimes$. For example, denoting the identity matrix of size M, by $\mathbf{1}_M$ and its k-th column by $\mathbf{e}_{k,M}$, $\bar{\mathbf{I}}(\bar{\mathbf{X}})$, $\bar{\mathbf{Q}}(\bar{\mathbf{X}})$ can be compactly expressed as

$$\left.\begin{matrix}\bar{\mathbf{I}}\\ \bar{\mathbf{Q}}\end{matrix}\right\}\left(\bar{\mathbf{X}}\right) = \sum_{u=1}^{H} \left.\begin{matrix}\mathbf{i}\\ \mathbf{q}\end{matrix}\right\}\left(\mathbf{1}_N \otimes \mathbf{e}_{u,H}^\top\, \bar{\boldsymbol{\Gamma}}^{-1}\bar{\mathbf{X}}\right) \otimes \mathbf{e}_{u,H}, \quad (7)$$

$\bar{\boldsymbol{\Omega}}$ and $\bar{\boldsymbol{\Gamma}}$ are also represented using the $\otimes$ operator as $\bar{\boldsymbol{\Omega}} = \mathbf{1}_N \otimes \boldsymbol{\Omega}$, and $\bar{\boldsymbol{\Gamma}} = \mathbf{1}_N \otimes \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is the Discrete Fourier Transform (DFT) operator, and $\boldsymbol{\Omega}$ is given by

$$\boldsymbol{\Omega} = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}_L\end{bmatrix}, \text{ with } \boldsymbol{\Omega}_{\mathrm{L}} = \mathbf{T}_{\Lambda_{\mathrm{G}}} \otimes \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix} \quad (8)$$

with $\mathbf{T}_{\Lambda_G} \in \mathbb{R}^{G\times G}$ being a diagonal matrix having the elements of $\boldsymbol{\kappa}^\top\boldsymbol{\omega}$, $\forall\boldsymbol{\kappa} \in \Lambda_G$, on the diagonal.

The system of equations, $\boldsymbol{\Phi}(\bar{\mathbf{X}}) = \mathbf{0}$, is solved using the Newton method which, starting from an initial guess $\bar{\mathbf{X}}^{(0)}$, iterates through the update $\bar{\mathbf{X}}^{(i+1)} = \bar{\mathbf{X}}^{(i)} - \Delta\bar{\mathbf{X}}$, where $\Delta\bar{\mathbf{X}}$ is the solution of the linear system,

$$\left(\bar{\mathbf{J}}_{\mathrm{HB}}\left(\bar{\mathbf{X}}^{(i)}\right)\right)\Delta\bar{\mathbf{X}} = \boldsymbol{\Phi}\left(\bar{\mathbf{X}}^{(i)}\right), \quad (9)$$

which requires forming the Jacobian matrix $\bar{\mathbf{J}}_{\mathrm{HB}}(\bar{\mathbf{X}})$ expressed as

$$\bar{\mathbf{J}}_{\mathrm{HB}}\left(\bar{\mathbf{X}}\right) \triangleq \bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Gamma}}\bar{\mathbf{J}}_Q\left(\bar{\mathbf{X}}\right)\bar{\boldsymbol{\Gamma}}^{-1} + \bar{\boldsymbol{\Gamma}}\bar{\mathbf{J}}_I\left(\bar{\mathbf{X}}\right)\bar{\boldsymbol{\Gamma}}^{-1} \quad (10)$$

where

$$\bar{\mathbf{J}}_{\left\{\begin{matrix}Q\\ I\end{matrix}\right\}}\left(\bar{\mathbf{X}}\right) = \sum_{u=1}^{H} \mathbf{J}_{\left\{\begin{matrix}Q\\ I\end{matrix}\right\}}\left(\mathbf{1}_N \otimes \mathbf{e}_{u,H}^\top\bar{\boldsymbol{\Gamma}}^{-1}\bar{\mathbf{X}}\right) \otimes \mathbf{e}_{u,H}\mathbf{e}_{u,H}^\top \quad (11)$$

and $\mathbf{J}_I(\mathbf{x}) \in \mathbb{R}^{N\times N}$ ($\mathbf{J}_Q(\mathbf{x})$) is the $N \times N$ Jacobian matrix given by $\frac{\partial \mathbf{i}(\mathbf{x})}{\partial \mathbf{x}}$ ($\frac{\partial \mathbf{q}(\mathbf{x})}{\partial \mathbf{x}}$).

C. HB FOR AUTONOMOUS CIRCUITS

In autonomous circuits, the HB is developed similar to the above process for non-autonomous circuits, but with key modifications. Firstly, the response is expanded in a single tone, $\boldsymbol{\omega} = \omega_0$ ($W = 1$), where ω_0 is treated as an additional unknown in the HB problem. Secondly, the number of equations is balanced with the number of unknowns by using the fact that the phase of the circuit waveforms is arbitrary. This fact is utilized by setting the phase on arbitrary circuit variable in $\mathbf{x}(t)$ to an arbitrary value. Thus, for example, the additional equation needed to balance the system is obtained from $\mathbf{e}_{i,NH}^\top\bar{\mathbf{X}} = 0$, where i is chosen to select an arbitrary component in $\mathbf{X}_{\boldsymbol{\kappa}}^{[C]}$ or $\mathbf{X}_{\boldsymbol{\kappa}}^{[S]}$, thereby setting the phase angle of the selected component to 0 or 90°, respectively. It should be noted, however, that such a formulation for the HB encounters significant difficulties in convergence. More particularly, the reason for those difficulties arise from the fact that the oscillatory periodic solution exists alongside the quiescent DC operating point of the system. The method of auxiliary voltage or current probe has been introduced and is widely used to address this difficulty in convergence [59].

V. SOLUTION OF THE HB PROBLEM

Over the past two decades, extensive research efforts have been thrust on finding efficient ways to factorize the Jacobian matrix $\bar{\mathbf{J}}_{\mathrm{HB}}$. There is a two-fold reason for this fact. Firstly, this matrix does not exhibit the sparsity pattern that characterizes the matrices arising from typical circuits, where the number of non-zero entries is almost on par with the circuit size. It should be noted that $\bar{\mathbf{J}}_{\mathrm{HB}}$ inherits the sparsity pattern of the underlying circuit due to being constructed from the circuit Jacobians $\mathbf{J}_I$ and $\mathbf{J}_Q$. However, the DFT operator matrices, which are full, create dense size-$H \times H$ blocks in place of nonzero entries in $\mathbf{J}_I$ and $\mathbf{J}_Q$. In addition, the number of Fourier coefficients H grows exponentially with the number of tones, W, in the stimulus, making the simulation intractable in large circuits.

In the context of HB analysis, two main issues are typically addressed simultaneously, namely, the construction and factorization of $\bar{\mathbf{J}}_{\mathrm{HB}}$. The computational complexity of constructing $\bar{\mathbf{J}}_{\mathrm{HB}}$ approaches $\mathcal{O}(L_{\mathrm{nl}}H^2\log_2(H))$ as H approaches powers of 2, (L_{nl} being the number of nonlinear elements) if the fast Fourier transform (FFT) algorithm is used in multiplying a vector by the DFT $\bar{\boldsymbol{\Gamma}}$ or the inverse DFT ($\bar{\boldsymbol{\Gamma}}^{-1}$) [60] [Ch. 6].

[2]A multi-index is a vector whose components are restricted to the positive/negative integers domain, $\mathbb{Z}$.

[3]This mode of ordering is known as node-major or frequency-minor. It is possible to write the HB formulation in frequency-major by multiplying from the left by a permutation matrix $\mathbf{P}$ and doing a change of variables $\bar{\mathbf{X}} \rightarrow \mathbf{P}\bar{\mathbf{X}}$, where $\mathbf{P} = \sum_{u=1}^{H} \mathbf{e}_{u,H}{}^\top \otimes \mathbf{1}_N \otimes \mathbf{e}_{u,H}$.

In RFIC, $L_{\rm nl}$ is typically comparable to the circuit size N, due to the design placing more emphasis on using Silicon devices rather than passive off-chip components. Although the construction part places significant demands on the computational and memory resources of the used platform, the factorization part presents the true bottleneck in the HB analysis.

The research directions instigated by this challenge emerged in the late 1990's. It sought to forego the explicit construction and direct factorization of $\bar{\mathbf{J}}_{\rm HB}$ in the pursuit of solving (9) and instead use the iterative methods [61], the review of which is in order.

A. ITERATIVE METHODS IN HB ANALYSIS

In solving the linear system in (9), iterative techniques rely on multiplying $\bar{\mathbf{J}}_{\rm HB}$ by a sequence of vectors $\mathbf{V}_m$, $m = 0, 1, 2, \cdots$, where a vector $\mathbf{V}_m$ is obtained by first computing $\bar{\mathbf{J}}_{\rm HB}\mathbf{V}_{m-1}$, and then orthogonalizing the result with respect to all previous vectors $\mathbf{V}_j$, $j = 0, 1, 2, \ldots, m-1$. The solution is approximated from the subspace spanned by the vectors $\mathbf{V}_m$, known as the Krylov subspace, and the method is considered converged if the resulting residual error is below a user-specified threshold. The generalized minimal residual (GMRES) is a variant of the iterative techniques most widely used within the context of the HB. References [62] and [63] were among the first to report on the use of those methods in HB. Using a matrix-vector multiplication at the core of GMRES implies that $\bar{\mathbf{J}}_{\rm HB}$ need not be explicitly computed and stored; only its effect on a given vector is required: a process that can be done with complexity of $\mathcal{O}(NH(1+\log(H)))$. However, the orthogonalization process incurs a complexity of $\mathcal{O}(mNH)$, which is acceptable if the convergence occurs after few iterations, i.e., $m \ll NH$. However, convergence at small m is only possible if the eigenvalues of $\bar{\mathbf{J}}_{\rm HB}$ are clustered away from the origin of the complex plane, a condition that is not often satisfied for general circuits.

Accelerating the convergence of the iterative techniques has been suggested through the use of the preconditioner matrix, which is an approximate inverse of $\bar{\mathbf{J}}_{\rm HB}$ constructed at low computational cost. Although general purpose preconditioners are available, the preconditioner matrix is more effective in reducing the number of iterations to convergence, m, when it is derived based on prior knowledge of the specific domain in which the matrix arises. In fact, most of the research activities may be largely viewed through the prism of the approach chosen to construct the preconditioner. Those activities are briefly outlined next.

1) AVERAGED PRECONDITIONER

The earliest form of preconditioner $\tilde{\mathbf{J}}_{\rm ave}$, is constructed through using $\tilde{\mathbf{J}}_{\rm ave} = \bar{\mathbf{\Omega}}\bar{\mathbf{\Gamma}}\tilde{\mathbf{J}}_Q\bar{\mathbf{\Gamma}}^{-1} + \bar{\mathbf{\Gamma}}\tilde{\mathbf{J}}_Q\bar{\mathbf{\Gamma}}^{-1}$, where

$$\tilde{\mathbf{J}}_{\begin{Bmatrix} Q \\ I \end{Bmatrix}} \triangleq \sum_{u=1}^{H} \mathbf{J}^{[\rm ave]}_{\begin{Bmatrix} Q \\ I \end{Bmatrix}} \otimes \mathbf{e}_{u,H}\mathbf{e}_{u,H}^{\top} \tag{12}$$

and $\mathbf{J}^{[\rm ave]}_{I,Q} = \frac{1}{H}\sum_{u=1}^{H} \mathbf{J}_{I,Q}(\mathbf{1}_N \otimes \mathbf{e}_{u,H}^{\top}\bar{\mathbf{\Gamma}}^{-1}\bar{\mathbf{X}})$. It is easy to show that $\tilde{\mathbf{J}}_{\rm ave}$ can be permuted to a block diagonal matrix (with H sparse diagonal blocks of size N) and therefore is easy to invert (by inverting each block individually) which is an indispensable requirement in a preconditioner.

2) SCHUR COMPLEMENT

Another preconditioner approach, proposed in [64], is derived from the observation that $\bar{\mathbf{J}}_{\rm HB}$ can be viewed as a sparse 2D $N \times N$ array of blocks with size $H \times H$. It assumes that many of those blocks result from linear, or almost linear, elements and therefore are diagonal, with only few (c) that are full blocks, $c \ll N$. The preconditioner is constructed by permuting those c columns to the right of the matrix, using a (rectangular) block diagonal sparse solver to factorize the $N - c$ left columns and updating the last c columns with the L and U factors of the left part of the matrix.

3) TIME-DOMAIN-BASED PRECONDITIONER

Using the change of variables $\bar{\mathbf{X}} \rightarrow \bar{\mathbf{\Gamma}}\bar{\mathbf{X}}_{\rm TD}$, and multiplying (6) from the left by $\bar{\mathbf{\Gamma}}^{-1}$ converts the HB problem into a time-domain formulation. The Jacobian matrix in this case can be approximated using a low-rank displacement of the block-circulant matrix $\bar{\mathbf{\Gamma}}^{-1}\bar{\mathbf{\Omega}}\bar{\mathbf{\Gamma}}$, whose inverse is expressed as a short series of block-circulant matrices. This feature makes the multiplication of this (approximate) inverse by a vector a process implementable with complexity $\mathcal{O}(\alpha NH\log(NH))$, where α is the rank of the displacement structure [65]. This approximate inverse is then deployed as the required preconditioner matrix.

4) ONE-STEP CORRECTION PRECONDITIONER

The one-step correction preconditioner was proposed in [66] to improve the averaged preconditioner $\tilde{\mathbf{J}}_{\rm ave}$ presented earlier. The method views the full HB Jacobian matrix $\bar{\mathbf{J}}_{\rm HB}$ as being the summation of two matrices: the first part is the averaged matrix $\tilde{\mathbf{J}}_{\rm ave}$ and the second one is a correction matrix $\bar{\mathbf{J}}_{\Delta}$, i.e., $\bar{\mathbf{J}}_{\rm HB} \equiv \tilde{\mathbf{J}}_{\rm ave} + \bar{\mathbf{J}}_{\Delta}$. Using the first two terms in a Taylor series approximation,

$$\bar{\mathbf{J}}_{\rm HB}^{-1} \approx \left(\mathbf{1}_{NH} - \tilde{\mathbf{J}}_{\rm ave}^{-1}\bar{\mathbf{J}}_{\Delta}\right)\tilde{\mathbf{J}}_{\rm ave}^{-1} \tag{13}$$

The above matrix, being a first-order approximation to the inverse of the $\bar{\mathbf{J}}_{\rm HB}$, is then used as the required preconditioner in [66]. Note here that multiplications of this preconditioner with a vector is done at low computational cost, since it only requires the LU factorization of the block diagonal, $\tilde{\mathbf{J}}_{\rm ave}$, and multiplications by $\bar{\mathbf{J}}_{\Delta}$ which can be done in an implicit way using FFT. This approach to preconditioning is leveraged further through a hierarchical approach which was developed later for parallel platforms in [67].

5) PRECONDITIONER BASED ON GRAPH SPARSIFICATION

Results from graph theory [68] were utilized in [69] to create a preconditioner that can be factorized directly at a much lower

cost than direct factorization of the original HB Jacobian matrix $\bar{\mathbf{J}}_{\mathrm{HB}}$. The main idea used in this work starts from the graph representing $N \times N$ MNA matrix $\mathbf{J}_I + \mathbf{J}_Q$ and approximate it, using the Laplacian measure, with a tree-like "support graph" that has much fewer edges but maintains the same number of vertices. The circuit corresponding to the support graph, referred to as the support circuit, is then treated in the HB framework, where the resulting $NH \times NH$ Jacobian, $\tilde{\mathbf{J}}_{\mathrm{SCP}}$, is considered as an approximation to $\bar{\mathbf{J}}_{\mathrm{HB}}$, with its inverse used as the preconditioner matrix. Here, $\tilde{\mathbf{J}}_{\mathrm{SCP}}^{-1}$ is computed through direct block LU factorization similar to [70], where the main computational advantage, compared to direct factorization of $\bar{\mathbf{J}}_{\mathrm{HB}}$, is that the growth of (block) fillins is significantly controlled.

B. DIRECT LU FACTORIZATION

The success of iterative techniques remains largely dependent on the nature of the circuit. For example, increasing the input power limits the effectiveness of the preconditioner matrix in accelerating the convergence of the Krylov iterative methods. In addition, iterative techniques do not achieve the extremely accurate solution obtained in solving the linear problem using the direct factorization. These observations made the direct factorization receive renewed interest. In fact, direct factorization becomes more attractive if the entire HB Jacobian, as well as the fillins that arise during the factorization, can fit on the system memory. With the technology developments that enabled workstations with vast memory resources, direct factorization became again a viable alternative to solving the linear problem (9). Direct factorization was developed in [70] using a block form of the classical LU factorization used in transient circuit simulations, where results showed superior performance to iterative techniques in circuits with high power excitations. It was also shown in [71], [72] that using platforms with graphical processing units (GPU) significantly leverages the performance of the direct factorization.

C. ENVELOP TECHNIQUES

The origin for the envelope approaches in circuit simulations may be traced back to the area of numerical solution of differential equations in the time-domain. It was mainly instigated by the observation that waveforms appear to be evolving with two different time scales. The wide separation of the observed time scales prompted the work in [73]. This work, which was extended to time-domain circuit simulations in [74]–[76] is mainly premised on the idea that the time scale (T_C) at which the waveform moves faster is known *apriori*. The problem is subsequently formulated as tracking the waveform along the slower time scale by taking time steps equivalent to (large) multiples of T_C. In this line of work, the slower time scale is considered fictitious, in the sense that there is no system of differential equations that prescribes its evolution.

This concept was then employed in the domain of microwave circuits by incorporating the different time scales explicitly in the form of the solution as in [77] or to the form of the differential equations that model the circuit, as in [78], [79]. The idea was put more on firm theoretical grounds in [80], by showing the link between the partial differential equations and multi-rate circuit response. The various approaches that relied on the envelop methodology may be collectively described by rewriting the response of the circuit from (5) in the following form,

$$\hat{\mathbf{x}}(\mathbf{t}_e, \mathbf{t}_c) = \hat{\mathbf{X}}_0(\mathbf{t}_e) + \sum_{\boldsymbol{\kappa} \in \Lambda_G} \hat{\mathbf{X}}_{\boldsymbol{\kappa}}^{[C]}(\mathbf{t}_e) \cos\left(\boldsymbol{\kappa}^\top (\boldsymbol{\omega} \odot \mathbf{t}_c)\right) + \hat{\mathbf{X}}_{\boldsymbol{\kappa}}^{[S]}(\mathbf{t}_e) \sin\left(\boldsymbol{\kappa}^\top (\boldsymbol{\omega} \odot \mathbf{t}_c)\right) \quad (14)$$

where $\mathbf{t}_c \in \mathbb{R}^{W_F}$, $\mathbf{t}_e \in \mathbb{R}^{W_S}$ are vectors representing W_F fast, and W_S slow time scales in the circuit, and the operator $\odot$ denotes the Hadamard (entrywise) product. The system of differential equations to which (14) is a solution is given by

$$\sum_{i=1}^{W_S} \frac{\partial \mathbf{q}(\hat{\mathbf{x}}(\mathbf{t}_e, \mathbf{t}_c))}{\partial t_{i,e}} + \sum_{i=1}^{W_F} \frac{\partial \mathbf{q}(\hat{\mathbf{x}}(\mathbf{t}_e, \mathbf{t}_c))}{\partial t_{i,c}} + \mathbf{i}(\hat{\mathbf{x}}(\mathbf{t}_e, \mathbf{t}_c)) = \hat{\mathbf{b}}(\mathbf{t}_e, \mathbf{t}_c) \quad (15)$$

where $t_{i,e}$ and $t_{i,c}$ are assumed to be the individual components in $\mathbf{t}_e$ and $\mathbf{t}_c$ vectors, respectively. The relation between the solution $\hat{\mathbf{x}}(\mathbf{t}_e, \mathbf{t}_c)$ of (15) and the solution $\mathbf{x}(t)$ of (4) typically depends on the relation between $\hat{\mathbf{b}}(\mathbf{t}_e, \mathbf{t}_c)$ and $\mathbf{b}(t)$ when $t_{i,e} = t_{j,c} = t$, $i = 1, \ldots, W_S$, $j = 1, \ldots, W_F$ [80]. Several methods can be constructed by the particular choices of $\mathbf{t}_e$ and $\mathbf{t}_c$, with different stability criteria [81]. This method was applied in system-level steady-state including a propagation link [82], and also in different applications, e.g. [83]. It was extended for handling autonomous circuits in [84], [85].

D. HB WITH NON-SINUSOIDAL BASIS

Although the periodicity of the sinusoidal basis functions made it the ideal vessel to express the periodic steady-state response of nonlinear circuits, its infinite support over the time-domain was well recognized as the root cause for the dense structure of the resulting HB Jacobian matrix. This fact represented the main motivation in exploring other approaches that rely on different basis functions. Notable among those approaches, is the one that used the Wavelets basis functions in steady-state analysis, [86], [87], where it was shown that wavelet-based HB results in a much sparser Jacobian matrix compared with its sinusoidal counterpart.

E. HB PROBLEM SIZE REDUCTION

The HB problem was also tackled from the perspective of reducing the number of unknowns. One of the earlier works to consider this approach was the one presented in [88], where concepts of model-order reduction (MOR) were utilized to project a homotopy-like formulation of the HB equations onto a space of much smaller dimension than the dimension of the full HB problem (NH). The reduced equations were used to track the solution versus the homotopy parameter, thereby requiring the factorization of much smaller Jacobian matrix.

Another approach that relies on the principle of reduction was proposed more recently in [89] where the reduction is

approached on two steps. In the first step, the Galerkin projection is done by projecting the N residuals onto a smaller set of $2M + 1$ basis functions, $2M + 1 \ll H$. In the second step, the Galerkin projection is done by projecting only M of the residuals on the entire set of basis functions H. The first step produces of a problem with $N(2M + 1)$ unknowns and the second step produces a problem in MH unknowns. Both problems are solved separately using the iterative techniques such as the GMRES method, and the method would alternate between the two steps for few iterations until an acceptable error level has been achieved for the unreduced problem (NH unknowns). However, achieving a high accuracy in solving the full unreduced problem may require a large number of such two-tiered iterations. This is a drawback, that is mitigated in this framework by using the low accuracy solution achieved from the two-step projection as a starting vector for a final phase of the method where the GMRES is run on the original unreduced problem, thereby reducing significantly the number of iterations needed to reach convergence. The main computational gain from this method is that during the two-step iteration, the GMRES is run on reduced spaces, thereby reducing the cost of orthogonalization. On the other hand, the final phase requires the GMRES to be applied on the original problem but for a small number of iterations, which limits the growth of the orthogonalization process.

Another approach that eyed the growth in the problem size, especially when the goal is to run circuit-level simulation in the context of simulating an entire front-end communication system, is based on the concept of domain decomposition [90]. In this approach the system is divided into subsystems that operate at similar parts of the spectrum. Those subsystems are considered 1st tier system decomposition. A 2nd tier decomposition is carried out on those subsystems further dividing them into blocks connected through ports. Each block in the 2nd tier is further divided into two subblocks housing the nonlinear and linear networks of the block. The reduction in the problem size and computational complexity is enabled on two levels. On the first level, reduction is achieved by limiting the problem unknowns to the spectra (harmonics) of the waveforms at the ports of the nonlinear subnetworks [91] and the ports interconnecting the different blocks. On the second level, computational complexity is further reduced by boosting the sparsity of the Jacobian matrix due to decoupling the spectral components among the 1st tier subsystems. This approach leads to a Jacobian matrix with bordered structure, as it is customary in domain decomposition techniques.

VI. FUTURE TRENDS

The trend for further progress in speed, accuracy and efficiency of modeling and simulation algorithms will continue. Increased development of multi-physics oriented automated modeling covering electromagnetic, thermal, mechanical stress and other domains will be an important direction. Algorithms that handle the gap between computationally intensive multi-physics numerical solvers versus design needs for faster and accurate behavioral modeling will be essential. Incorporation of multi-physics domain knowledges into data driven models will play an important role. The increasing complexity of future generation wireless systems will drive the trend to systematically model more complex structures from components to packages and systems. Mathematical concepts such as decomposition and hierarchical analysis need to be used in combination with domain specific knowledges of the package and systems for systematic modeling. The increased capabilities from deep neural networks, machine learning, and other computational intelligence techniques will spur new advances in data driven modeling algorithms to handle higher dimensional design spaces, and to address increased complexity of nonlinear relationships in models. Together with the expanding computing resources and powers, advances in automated modeling would lead to new generations of systematic computational algorithms to address future design needs to model highly complex structures and relationships accurately and efficiently.

Simulation approaches will advance to handle larger circuits and systems, perhaps spanning multiple energy (electrical, mechanical, acoustic) domains. Advanced partitioning techniques, such as those based on domain decomposition or node tearing, may need to be developed to address the increasing complexity of those systems. Simulation approaches need to factor in the high variability of some of the physical design parameters and its effect on the performance of the circuit. The uncertainty arising from such variability will need to be accurately and efficiently quantified.

Recent developments in the mathematical sciences are creating new roads that are worth traveling. One of the new areas that may hold a significant potential in the steady-state analysis is related to the notion of compressed tensor train decomposition (TTD), which was presented in [92]. This approach was employed in [93] to create an algorithm to perform the Fourier transform. It works by structuring the vector, which needs to be Fourier transformed, into d-dimensional tensor assuming that the vector length is 2^d. Complexity analysis of this approach showed a Fourier transform that scales with $\mathcal{O}(d^2L^3)$ where L is the maximal rank of the tensors. This is to be compared with $\mathcal{O}(2^d d)$ of the regular FFT, which for low rank tensor representation provides an attractive competitive edge over the century-old FFT. On the front of using a preconditioner matrix to accelerate the convergence of Krylov subspace iterative methods, the development of preconditioner, for domain decomposition in the area of computational electromagnetic, appears to be largely untapped for HB problem.

Another class of techniques worthy of note for its promise in accelerating the simulation process is the one based on the paradigm of model-order reduction (MOR). MOR was first initiated for modeling large linear time-invariant (LTI) circuits using a system with much smaller order. Rather than starting with the input/output data generated through simulation or measurements, which represents the modeling approaches presented earlier, MOR constructs the model using the system of equations representing the system that needs to be

modelled. Those equations are mostly the Kirchoff current/voltage laws which are extracted in an automated manner. The model is constructed by projecting the system of equations onto a subspace with much smaller dimension, and the methodology is widely known as projection-based MOR [94], [95]. The success of projection-based MOR in compactly and accurately modeling large LTI circuits has prompted researchers to investigate its applicability in nonlinear circuits. The first idea in this direction was proposed in [96] for modeling weakly nonlinear circuits, and went through several phases improving its feasibility for circuits with strong nonlinear behaviour [97]–[99]. This area has recently witnessed a significant uptick in activity with the introduction of some novel ideas such as those based on the Lowener matrix [100], interpolation projection approach [105], and the Hankel norm balancing approach [101]. These approaches, although mostly developed and applied in domains remotely related to nonlinear microwave devices and RF circuits, hold significant potentials for advancing the simulation and modeling for those circuits. For example, the recent work in [104] shows the relation of some of those approaches to the Volterra series of nonlinear systems which are known to model distortion effects of nonlinear circuits.

Parametrized model-order reduction (PMOR) is a concept that was born out of MOR and was used in the framework of linear circuit analysis to accelerate tasks such as uncertainty quantification and variability analysis [106]. PMOR was recently extended to nonlinear circuits in the context of DC analysis [102]. PMOR shows a great potential in accelerating tasks such as variability analysis in the context of HB simulation.

VII. CONCLUSION

This paper has presented the current state-of-the-art in automated modeling and simulations of microwave and RF circuits and devices. Automated modeling algorithms have been described covering data sampling, model structure adaptation and training/validation. State of the art harmonic balance simulation algorithms have been described conquering historically difficult issues in solving nonlinear microwave circuits and addressing problem size, convergence speed and solution accuracy. The advanced simulation algorithms facilitate efficient simulation of steady-state response to periodical or quasi-periodical stimulus in the frequency-domain. The algorithms permit efficient simulation of nonlinear microwave and RF circuits, e.g., distortion, intermodulation analysis, gain compression, noise analysis, etc. Future trends in automated modeling and simulations have been highlighted. As modeling and simulation algorithms continue to expand to handle future design needs, exciting opportunities lie ahead in developing new breakthrough algorithms to handle larger and more complex systems than possible today.

REFERENCES

[1] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Norwood, MA, USA: Artech House, 2000.

[2] Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial neural networks for RF and microwave design-from theory to practice," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 4, pp. 1339–1350, Apr. 2003.

[3] J. E. Rayas-Snchez, "EM-based optimization of microwave circuits using artificial neural networks: The state-of-the-art," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 420–435, Jan. 2004.

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

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

[6] W. Na, F. Feng, C. Zhang, and Q. J. Zhang, "A unified automated parametric modeling algorithm using knowledge-based neural network and l_1 optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 3, pp. 729–745, Mar. 2017.

[7] J. D. Geest, T. Dhaene, N. Fache, and D. D. Zutter, "Adaptive sampling algorithm for accurate modeling of general interconnection structures," in *Proc. Eur. Microw. Conf.*, Munich, Germany, Oct. 1999, pp. 178–181.

[8] R. Lehmensiek and P. Meyer, "An efficient adaptive frequency sampling algorithm for model-based parameter estimation as applied to aggressive space mapping," *Microw. Opt. Technol. Lett.*, vol. 24, pp. 71–78, 2000.

[9] Z. Zhang, Q. S. Cheng, H. Chen, and F. Jiang, "An efficient hybrid sampling method for neural network-based microwave component modeling and optimization," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 7, pp. 625–628, Jul. 2020.

[10] L. Xiao, W. Shao, F. Jin, B. Wang, W. T. Joines, and Q. H. Liu, "Semisupervised radial basis function neural network with an effective sampling strategy," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 4, pp. 1260–1269, Apr. 2020.

[11] F. Wang and Q. J. Zhang, "Knowledge based neural models for microwave design," *IEEE Trans. Microw. Theory Techn.*, vol. 45, no. 12, pp. 2333–2343, Dec. 1997.

[12] J. W. Bandler, M. A. Ismail, J. E. Rayas-Snchez, and Q. J. Zhang, "Neuromodeling of microwave circuits exploiting space-mapping technology," *IEEE Trans. Microw. Theory Techn.*, vol. 47, no. 12, pp. 2417–2427, Dec. 1999.

[13] M. H. Bakr, J. W. Bandler, M. A. Ismail, J. E. Rayas-Snchez, and Q. J. Zhang, "Neural space-mapping optimization for EM-based design," *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 12, pp. 2307–2315, Dec. 2000.

[14] V. Gutirrez-Ayala and J. E. Rayas-Snchez, "Neural input space mapping optimization based on nonlinear two-layer perceptrons with optimized nonlinearity," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 20, no. 5, pp. 512–526, Sep. 2010.

[15] J. E. Rayas-Snchez and V. Gutirrez-Ayala, "EM-based Monte Carlo analysis and yield prediction of microwave circuits using linear-input neural-output space mapping," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4528–4537, Dec. 2006.

[16] J. W. Bandler *et al.*, "Space mapping: The state of the art," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 337–361, Jan. 2004.

[17] S. Koziel, Q. S. Cheng, and J. W. Bandler, "Space mapping," *IEEE Microw. Mag.*, vol. 9, no. 6, pp. 105–122, Dec. 2008.

[18] D. Gorissen, L. Zhang, Q. J. Zhang, and T. Dhaene, "Evolutionary neuro-space mapping technique for modeling of nonlinear microwave devices," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 2, pp. 213–229, Feb. 2011.

[19] J. C. Rautio, "Synthesis of lumped models from N-port scattering parameter data," *IEEE Trans. Microw. Theory Techn.*, vol. 42, no. 3, pp. 535–537, Mar. 1994.

[20] T. Mangold and P. Russer, "Full-wave modeling and automatic equivalent-circuit generation of millimeter-wave planar and multilayer structures," *IEEE Trans. Microw. Theory Techn.*, vol. 47, no. 6, pp. 851–858, Jun. 1999.

[21] I. Timmins and K. L. Wu, "An efficient systematic approach to model extraction for pasive microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 9, pp. 1565–1573, Sep. 2000.

[22] A. C. Watson, D. Melendy, P. Francis, K. Hwang, and A. Weisshaar, "A comprehensive compact-modeling methodology for spiral inductors in silicon-based RFICs," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 3, pp. 849–857, Mar. 2004.

[23] R. Neumayer, A. Stelzer, F. Haslinger, and R. Weigel, "On the synthesis of equivalent-circuit models for multiports characterized by frequency-dependent parameters," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 12, pp. 2789–2796, Dec. 2002.

[24] I. R. Pordanjani, H. E. Mazin, and W. Xu, "A novel genetic programming approach for frequency-dependent modeling," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 353–367, Jun. 2013.

[25] D. Romano, G. Antonini, U. Grossner, and I. Kovacevic-Badstubner, "Circuit synthesis techniques of rational models of electromagnetic systems: A tutorial paper," *Int. J. Numer. Model.*, vol. 32, no. 5, Sep./Oct. 2019.

[26] G. Antonini, "SPICE equivalent circuits of frequency-domain responses," *IEEE Trans. Electromagn. Compat.*, vol. 45, no. 3, pp. 502–512, Aug. 2003.

[27] K. L. Choi and M. Swaminathan, "Development of model libraries for embedded passives using network synthesis," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 47, no. 4, pp. 249–260, Apr. 2000.

[28] P. Zhao and K. L. Wu, "Circuit model extraction of parallel-connected dual passband coupled-resonator filters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 2, pp. 822–830, Feb.. 2018.

[29] J. C. Rautio, "Synthesis of compact lumped models from electromagnetic analysis results," *IEEE Trans. Microw. Theory Techn.*, vol. 55, no. 12, pp. 2548–2554, Dec. 2007.

[30] D. Saraswat, R. Achar, and M. Nakhla, "A fast algorithm and practical considerations for passive macromodeling of measured/simulated data," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 27, no. 1, pp. 57–70, Feb. 2004.

[31] J. C. Rautio, M. R. LeRoy, and B. J. Rautio, "Synthesis of perfectly casual parameterized compact models for planar transmission lines," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 12, pp. 2938–2947, Dec. 2009.

[32] P. Triverio, S. Grivet-Talocia, and M. S. Nakhla, "A parameterized macromodeling strategy with uniform stability test," *IEEE Trans. Adv. Packag.*, vol. 32, no. 1, pp. 205–215, Feb. 2009.

[33] F. Ferranti, L. Knockaert, and T. Dhaene, "Passivity-preserving parametric macromodeling by means of scaled and shifted state-space systems," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 10, pp. 2394–2403, Oct. 2011.

[34] D. Deschrijver, T. Dhaene, and D. De Zutter, "Robust parametric macromodeling using multivariate orthonormal vector fitting," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 7, pp. 1661–1667, Jul. 2008.

[35] F. Ferranti, L. Knockaert, and T. Dhaene, "Passivity-preserving interpolation-based parameterized macromodeling of scattered S-data," *IEEE Microw. Wireless Compon. Lett.*, vol. 20, no. 3, pp. 133–135, Jan. 2010.

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

[37] E.R. Samuel, L. Knockaert, F. Ferranti and R. Dhaene, "Guaranteed passive parameterized macromodeling by using sylvester state-space realizations," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 4, pp. 1444–1454, Apr. 2013.

[38] J. Xu, R. Jones, S. A. Harris, T. Nielsen, and D. E. Root, "Dynamic FET model - DynaFET - for GaN transistors from NVNA active source injection measurements," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2014, pp. 1–3.

[39] L. Zhang, J. J. Xu, M. Yagoub, R. Ding, and Q. J. Zhang, "Efficient analytical formulation and sensitivity analysis of neuro-space mapping for nonlinear microwave device modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 53, pp. 2752–2767, Sep. 2005.

[40] L. Zhang, Q. J. Zhang, and J. Wood, "Statistical neuro-space mapping technique for large-signal modeling of nonlinear devices," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 11, pp. 2453–2467, Nov. 2008.

[41] Z. Zhao, L. Zhang, F. Feng, W. Zhang, and Q. J. Zhang, "Space mapping technique using decomposed mappings for GaN HEMT modeling with trapping effects," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 8, pp. 3318–3341. Aug. 2020.

[42] P. H. Chen, R. E. Fan, and C. J. Lin, "A study on SMO-type decomposition methods for support vector machines," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 893–908, Jul. 2006.

[43] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[44] J. L. Chvez-Hurtado and J. E. Rayas-Snchez, "Polynomial-based surrogate modeling of RF and microwave circuits in frequency domain exploiting the multinomial theorem," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4371–4381, Dec. 2016.

[45] D. Gorissen, L. De Tommasi, K. Crombecq, and T. Dhaene, "Sequential modeling of a low noise amplifier with neural networks and active learning," *Neural Comput. Appl.*, vol. 18, no. 5, pp. 485–494, Jun. 2009.

[46] C. Zhang, J. Jin, W. Na, Q. J. Zhang, and M. Yu, "Multialued neural network inverse modeling and applications to microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 8, pp. 3781–3797, Aug. 2018.

[47] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q. J. Zhang, "Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4140–4155, Oct. 2019.

[48] C. Brezinski and M. R. Zaglia, *Extrapolation Methods: Theory and Practice*. Amsterdam, The Netherlands: Elsevier, 1991.

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

[50] L. C. Kiong, M. Rajeswari, and M. V. C. Rao, "Extrapolation detection and novelty-based node insertion for sequential growing multi-experts network," *Appl. Soft Comput.*, vol. 3, no. 2, pp. 159–175, Sep. 2003.

[51] J. Xu, M. C. E. Yagoub, R. Ding, and Q. J. Zhang, "Robust neural based microwave modelling and design using advanced model extrapolation," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Fort Worth, TX, USA, Jun. 2004, pp. 1549–1552.

[52] L. Zhang and Q. J. Zhang, "Simple and effective extrapolation technique for neural-based microwave modeling," *IEEE Microw. Wireless Compon. Lett.*, vol. 20, no. 6, pp. 301–303, Jun. 2010.

[53] W. Zhang *et al.*, "Space mapping approach to electromagnetic centric multiphysics parametric modeling of microwave components," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3169–3185, Jul. 2018.

[54] M. Iwamoto, J. Xu, W. Zhou, and D. E. Root, "Knowledge-based neural network (KBNN) modeling of HBT junction temperature and thermal resistance from electrical measurements," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Honolulu, HI, USA, Jun. 2017, pp. 1069–1071.

[55] J. L. Chvez-Hurtado, J. E. Rayas-Snchez, and Z. Brito-Brito, "Multiphysics polynomial-based surrogate modeling of microwave structures in frequency domain," in *Proc. IEEE MTT-S Latin Amer. Microw. Conf.*, Puerto Vallarta, Mexico, Dec. 2016, pp. 1–3.

[56] C.-W. Ho, A. Ruehli, and P. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. Circuits Syst.*, vol. 22, no. 6, pp. 504–509, Jun. 1975.

[57] K. S. Kundert, J. K. White, and A. Sangiovanni-Vincentelli, *Steady-State Methods for Simulating Analog and Microwave Circuits*. Norwell, MA, USA: Kluwer Academic, 1990.

[58] K. Kundert and A. Sangiovanni-Vincentelli, "Simulation of nonlinear circuits in the frequency domain," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 5, no. 4, pp. 521–535, Oct. 1986.

[59] E. Ngoya, A. Suárez, R. Sommet, and R. Quéré, "Steady state analysis of free or forced oscillators by harmonic balance and stability investigation of periodic and quasi-periodic regimes," *Int. J. Microw. Millimeter-Wave Comput.-Aided Eng.*, vol. 5, no. 3, pp. 210–223, 1995.

[60] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.

[61] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Boston, MA, USA: PWD, 1996.

[62] R. G. Melville, P. Feldmann, and J. Roychowdhury, "Efficient multi-tone distortion analysis of analog integrated circuits," in *Proc. IEEE Custom Integr. Circuits Conf.*, 1995, pp. 241–244.

[63] H. G. Brachtendorf, G. Welsch, and R. Laur, "A simulation tool for the analysis and verification of the steady state of circuit designs," *Int. J. Circuit Theory Appl.*, vol. 23, no. 4, pp. 311–323, Aug. 1995.

[64] D. Long, R. Melville, K. Ashby, and B. Horton, "Full-chip harmonic balance," in *Proc. IEEE Custom Integr. Circuits Conf.*, 1997, pp. 379–382.

[65] I. Gohberg and V. Olshevsky, "Complexity of multiplication with vectors for structured matrices," *Linear Algebra Appl.*, vol. 202, pp. 163–192, Apr. 1994.

[66] W. Dong and P. Li, "Hierarchical harmonic-balance methods for frequency-domain analog-circuit analysis," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 12, pp. 2089–2101, Dec. 2007.

[67] W. Dong and P. Li, "A parallel harmonic-balance approach to steady-state and envelope-following simulation of driven and autonomous circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 4, pp. 490–501, Apr. 2009.

[68] A. Kolla, Y. Makarychev, A. Saberi, and S.-H. Teng, "Subgraph sparsification and nearly optimal ultrasparsifiers," in *Proc. 42th ACM Symp. Theory Comput.*, 2010, pp. 57–66.

[69] L. Han, X. Zhao, and Z. Feng, "An adaptive graph sparsification approach to scalable harmonic balance analysis of strongly nonlinear post-layout RF circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 34, no. 2, pp. 173–185, Feb. 2015.

[70] A. Mehrotra and A. Somani, "A robust and efficient harmonic balance (HB) using direct solution of HB Jacobian," in *Proc. 46th Annu. Design Autom. Conf.*, 2009, pp. 370–375.

[71] B. Bandali, E. Gad, and M. Bolic, "Accelerated harmonic-balance analysis using a graphical processing unit platform," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 7, pp. 1017–1030, Jul. 2014.

[72] W. Lee, R. Achar, and M. S. Nakhla, "Dynamic GPU parallel sparse LU factorization for fast circuit simulation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 26, no. 11, pp. 2518–2529, Nov. 2018.

[73] L. R. Petzold, "An efficient numerical method for highly oscillatory ordinary differential equations," *SIAM J. Numer. Anal.*, vol. 18, no. 3, pp. 455–479, Jun. 1981.

[74] A. Brambilla and P. Maffezzoni, "Envelope following method for the transient analysis of electrical circuits," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 47, no. 7, pp. 999–1008, Jul. 2000.

[75] M. A. Farhan, E. Gad, M. S. Nakhla, and R. Achar, "High order and a-stable envelope following method for transient simulations of oscillatory circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 12, pp. 3309–3317, Dec. 2014.

[76] M. A. Farhan, M. Nakhla, E. Gad, and R. Achar, "Parallel high-order envelope-following method for fast transient analysis of highly oscillatory circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 1, pp. 261–270, Jan. 2017.

[77] E. Ngoya and R. Larcheveque, "Envelop transient analysis: A new method for the transient and steady state analysis of microwave communication circuits and systems," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 1996, vol. 3, pp. 1365–1368.

[78] H. G. Brachtendorf, G. Welsch, R. Laur, and A. Bunse-Gerstner, "Numerical steady state analysis of electronic circuits driven by multi-tone signals," *Elect. Eng.*, vol. 79, no. 2, pp. 103–112, Apr. 1996.

[79] H. G. Brachtendorf, G. Welsch, and R. Laur, "A novel time-frequency method for the simulation of the steady state of circuits driven by multi-tone signals," in *Proc. Int. Symp. Circuits Syst.*, Hong Kong, Jun. 1997, pp. 1508–1511.

[80] J. Roychowdhury, "Analyzing circuits with widely separated time scales using numerical PDE methods," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 5, pp. 578–594, May 2001.

[81] T. Mei, J. Roychowdhury, T. S. Coffey, S. A. Hutchinson, and D. M. Day, "Robust, stable time-domain methods for solving MPDEs of fast/slow systems," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 2, pp. 226–239, Feb. 2005.

[82] N. Carvalho, J. Pedro, W. Jang, and M. Steer, "Nonlinear RF circuits and systems simulation when driven by several modulated signals," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 2, pp. 572–579, Feb. 2006.

[83] A. Pels, J. Gyselinck, R. V. Sabariego, and S. Schöps, "Solving nonlinear circuits with pulsed excitation by multirate partial differential equations," *IEEE Trans. Magn.*, vol. 54, no. 3, pp. 1–4, Mar. 2018.

[84] O. Narayan and J. Roychowdhury, "Analyzing oscillators using multitime PDEs," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 7, pp. 894–903, Jul. 2003.

[85] T. Mei and J. Roychowdhury, "A time-domain oscillator envelope tracking algorithm employing dual phase conditions," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 1, pp. 59–69, Jan. 2008.

[86] N. Soveiko and M. S. Nakhla, "Steady-state analysis of multitone nonlinear circuits in wavelet domain," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 3, pp. 785–797, Mar. 2004.

[87] N. Soveiko, E. Gad, and M. Nakhla, "A wavelet-based approach for steady-state analysis of nonlinear circuits with widely separated time scales," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 6, pp. 451–453, Jun. 2007.

[88] E. Gad, R. Khazaka, M. Nakhla, and R. Griffith, "A circuit reduction technique for finding the steady-state solution of nonlinear circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 12, pp. 2389–2396, Dec. 2000.

[89] F. Bizzarri, A. Brambilla, and L. Codecasa, "Harmonic balance based on two-step galerkin method," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 9, pp. 1476–1486, Sep. 2016.

[90] V. Rizzoli, D. Masotti, F. Mastri, and E. Montanari, "System-oriented harmonic-balance algorithms for circuit-level simulation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 2, pp. 256–269, Feb. 2011.

[91] V. Rizzoli, D. Masotti, F. Mastri, and C. Cecchetti, "Generalized distributed nonlinear device modeling for Krylov-subspace based microwave circuit analysis," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2000, vol. 1, pp. 453–456 .

[92] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, 2011.

[93] S. Dolgov, B. Khoromskij, and D. Savostyanov, "Superfast Fourier transform using QTT approximation," *J. Fourier Anal. Appl.*, vol. 18, no. 5, pp. 915–953, May 2012.

[94] W. H. Schilders, H. A. Van der Vorst, and J. Rommes, *Model Order Reduction: Theory, Research Aspects and Applications*. Berlin, Germany: Springer, 2008, vol. 13.

[95] P. Benner, M. Hinze, and E. J. W. Ter Maten, *Model Reduction for Circuit Simulation*. Berlin, Germany: Springer, 2011.

[96] J. R. Phillips, "Projection-based approaches for model reduction of weakly nonlinear, time-varying systems," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 22, no. 2, pp. 171–187, Feb. 2003.

[97] P. Li and L. T. Pileggi, "Compact reduced-order modeling of weakly nonlinear analog and RF circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 2, pp. 184–203, Feb. 2005.

[98] C. Gu, "QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 9, pp. 1307–1320, Sep. 2011.

[99] H. Liu, L. Daniel, and N. Wong, "Model reduction and simulation of nonlinear circuits via tensor decomposition," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 34, no. 7, pp. 1059–1069, Jul. 2015.

[100] I. V. Gosea and A. C. Antoulas, "Data-driven model order reduction of quadratic-bilinear systems," *Numer. Linear Algebra Appl.*, vol. 25, no. 6, 2018, Paper e2200.

[101] M. I. Ahmad, P. Benner, and L. Feng, "A new two-sided projection technique for model reduction of quadratic-bilinear descriptor systems," *Int. J. Comput. Math.*, vol. 96, no. 10, pp. 1899–1909, 2019.

[102] B. Nouri, E. Gad, A. Nouri, M. Nakhla, and Y. Tao, "DC-centric parameterized reduced-order model via moment-based interpolation projection (MIP) algorithm," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 10, no. 8, pp. 1348–1357, Aug. 2020.

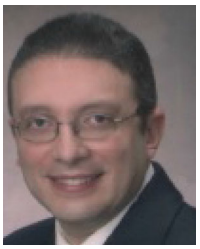
[103] P. Benner and T. Breiten, "Interpolation-based $\mathcal{H}_2$-model reduction of bilinear control systems," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 859–885, 2012.

[104] X. Cao, P. Benner, I. P. Duff, and W. Schilders, "Model order reduction for bilinear control systems with inhomogeneous initial conditions," *Int. J. Control*, to be published, doi: 10.1080/00207179.2020.1740945.

[105] S. Chaturantabut and D. C. Sorensen, "Nonlinear model reduction via discrete empirical interpolation," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2737–2764, 2010.

[106] Y. Tao, B. Nouri, M. S. Nakhla, M. A. Farhan, and R. Achar, "Variability analysis via parameterized model order reduction and numerical inversion of Laplace transform," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 7, no. 5, pp. 678–686, May 2017.

QI-JUN ZHANG (Fellow, IEEE) received the B.Eng. degree from the Nanjing University of Science and Technology, Nanjing, China, in 1982, and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 1987. He was a Research Engineer with Optimization Systems Associates, Inc., Dundas, ON, Canada, from 1988 to 1990, developing advanced optimization software for microwave modeling and design. He joined the Department of Electronics, Carleton University, Ottawa, Canada, in 1990, where he is currently a Chancellor's Professor. His research interests are modeling, optimization, and neural networks for high-speed/high-frequency electronic design, and has more than 300 publications in the area. He is the author of the book *Neural Networks for RF and Microwave Design* (Boston, MA, USA: Artech House, 2000), a Co-Editor of *Modeling and Simulation of High-Speed VLSI Interconnects* (Boston, MA, USA: Kluwer, 1994), and a Co-Editor of *Simulation-Driven Design Optimization and Modeling for Microwave Engineering* (London, U.K.: Imperial College Press, 2013). He is a Fellow of the Canadian Academy of Engineering. He is an Associate Editor for IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, a Topic Editor for IEEE JOURNAL OF MICROWAVES, and the Chair of the Technical Committee on Design Automation (MTT-2) of the IEEE Microwave Theory and Techniques Society.



EMAD GAD (Senior Member, IEEE) received the Ph.D. degree from Carleton University, Ottawa, ON, Canada, in 2003. He is currently an Associate Professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada. His current research interests include numerical simulation and modeling approaches of high speed and RF circuits. He was the corecipient of the 2002 IEEE Microwave Prize for a significant contribution in a field of endeavor of the IEEE Microwave and Theory Techniques Society and also the recipient of several honorary awards, including the Governor General Gold Medal, the Carleton University Medal for the Outstanding Academic Performance at the graduate level, and the Ottawa Center for Research and Innovation 2003 as a Student Researcher of the year. He is a Practicing Professional Engineer in the Province of Ontario, Canada.

BEHZAD NOURI (Member, IEEE) received the B.Eng. degree in electronics engineering from the University of Tabriz, Tabriz, Iran, in 1988, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2008 and 2014, respectively.

He served in various capacities with research laboratories and telecommunication industries, including the Iran Telecommunication Research Center, Tehran, Iran, from 1988 to 1990; Iran Communication Industries, Inc., Tehran, Iran, from 1990 to 1999; and T-Telecom (Alcatel agent), Tehran, Iran, from 1999 to 2002. Since 2014, he has been a Postdoctoral Research Associate and a Research Adjunct Professor with the Department of Electronics Engineering, Carleton University. His current research interests include computer-aided design for circuits and systems simulation, signal/power integrity analysis, model order reduction, data-driven macromodeling, parameter-varying systems, uncertainty quantification, and numerical algorithms.

Dr. Nouri was the recipient of various scholarships during his graduate studies. He was the recipient of the University-Gold-Medal for his outstanding doctoral work in 2014. He was the recipient of the Best Transactions Paper Award from IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY in 2013 and the Best Paper Award from the IEEE Workshop on Signal and Power Integrity in 2011 and 2018.

WEICONG NA (Member, IEEE) received the B.Eng. degree from Tianjin University, Tianjin, China, in 2012, and the Ph.D. degree from the School of Microelectronics, Tianjin University, Tianjin, China, and the Department of Electronics, Carleton University, Ottawa, ON, Canada, in 2018.

She is currently a Lecturer with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. Her research interests include microwave circuit modeling and design, automated neural based model generation, EM field knowledge-based modeling and optimization, and deep neural network modeling for microwave applications.

MICHEL NAKHLA (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada. From 1976 to 1988, he was with Bell-Northern Research, Ottawa, as the Senior Manager of the Computer-Aided Engineering Group. In 1988, he joined Carleton University, Ottawa, Canada as a Professor and holder of the Senior Computer-Aided Engineering Industrial Chair established by Bell-Northern Research and the Natural Sciences and Engineering Research Council of Canada. He is the Founder of the High-Speed CAD Research Group, Carleton University, where he is currently a Chancellor's Professor of Electrical Engineering. He has authored more than 380 peer-reviewed research articles. He is currently an Associate Editor for IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY and served as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He is on various international committees, including the Standing Committee of the IEEE Signal and Power Workshop, the Technical Program Committee of the IEEE International Microwave Symposium and the CAD Committee of the IEEE Microwave Theory and Techniques Society. He also served on several Canadian and international government-sponsored research grants selection panels. His current research interests include modeling and simulation of high-speed circuits and interconnects, uncertainty quantification, model-order reduction, parallel processing, and nonlinear and microwave circuits. He is a Fellow of the Canadian Academy of Engineering.