Artificial Neural Network Non-Linear Transistor Behavioral Models: Structure and Parameter Determination Process yellowBased on the Cardiff Model

Mengyue Tian, James J. Bell, *Member, IEEE,* Roberto Quaglia, *Member, IEEE,* Ehsan M.Azad, *Member, IEEE,* and Paul J. Tasker, *Fellow, IEEE*

***Abstract*—This article introduces a novel Artificial Neural Network (ANN) structure magentadetermination process, based on the Cardiff Model (CM), to determine ANN-based transistor non-linear behavioral models. By relating the CM formulation and coefficients to the Taylor series expansion of the ANN model, a novel approach for determining the required values of a Fully Connected Cascaded (FCC) ANN structure has been formulated. The proposed method provides the chance to escape from the possible time-consuming ANN magentadetermination process. Experiments proved that the proposed ANN models using the magentadetermination method can provide accurate prediction for the behavior acquired from load-pull characterizations of a Wolfspeed 10 W packaged gallium nitride (GaN) High Electron Mobility Transistor (HEMT) simulation at 3.5 GHz, and a dense load-pull measurement of WIN NP12 4x75 um GaN HEMT at 20 GHz, with Normalized Mean Square Error (NMSE) levels lower than - 40 dB.**

***Index Terms*—Artificial Neural Network (ANN), behavioral model, yellowLoad-Pull measurement, yellowgallium nitride (GaN), power amplifier.**

I. Introduction

VARIOUS large-signal models for HEMT transistors have been reported in the literature. Those commonly used in CAD design are state function, current-voltage (I-V) and charge-voltage (Q-V), time-domain based compact models. While physically motivated from an equivalent circuit topology perspective, they generally use user-defined empirical closed-form mathematical expressions to fit the measured I-V and Q-V response [1], [2]. CAD implementations have also used look-up tables [3], [4], or a pure ANN based model [5]. Recently, especially for GaN HEMTs, state function formulations that are more physics-based have emerged, such as Advanced Spice Model (ASM) HEMT [6] and MIT Virtual Source GaN-HEMT (MVSG) [7] models, where the closed-form mathematical expressions are derived from either surface-potential or charge based analysis and so provide more insight of how the device operates and ideally better scalability vs. device geometry.

Manuscript received April 23, 2024. This work was supported in part by the Chinese Scholarship Council (CSC).*(Corresponding Author: Mengyue Tian)*

M. Tian, J. J. Bell, R. Quaglia, P. J. Tasker are with the Center for High Frequency Engineering (CHFE), School of Engineering, Cardiff University, Cardiff, CF24 3AA (Email: mengyuet@cardiff.ac.uk).

E. Azad is with CSA Catapult Innovation Centre, Celtic Way, Imperial Park, Newport, NP10 8BE

The state function modeling approach, while providing a robust general purposed CAD model unfortunately involves gathering extensive DC and S-parameter data, coupled with complex parameter extraction procedures, which can be very time-consuming. Behavioral models, typically formulated in the frequency domain, provide an alternative less general but more time-efficient approach. For example, the Poly-Harmonic Distortion (PHD) Model [8], the X-parameter Model [9], and the Cardiff Model [10], require no prior knowledge of the internal device structure but provide good model accuracy [11], [12] since they effectively enable the direct use of measurement data in non-linear CAD design.

Considering that the Artificial Neural Network (ANN) technique has been one of the research hotspots for transistor model extraction in recent decades. This work focuses on exploring the possibility of further enhancing behavioral model performance through the application of the ANN technique. The ANN technique has provided RF designers with flexibility in the model structure to account for more complexity, and the possibility of a faster optimization procedure to create models for fitting different measurement datasets [13]–[17]. More importantly, it potentially allows a better extrapolation prediction ability [12], [18], and flexibility on model parameter variation for a robust selection on the model's accuracy, regardless of the dataset for model extraction (explored later in this work). The design of RF power amplifiers is a good application area, for the ANN technique, to provide the required transistor non-linear behavioral models. However, the challenge is determining the appropriate ANN structure.

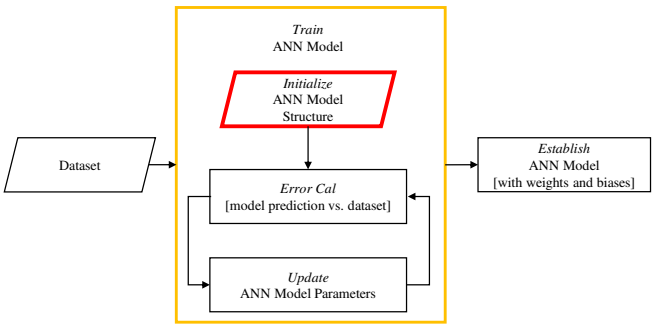


Fig. 1. Brief general flow diagram for the ANN training process.

Under different scenarios, the ANN structure is not fixed for systems with different non-linearity [19]–[21]. Therefore, different structures might be required for better prediction of behavior. The existing research recognizes that the Multilayer Perceptron (MLP) feedforward neural network structure [22], [23], and the Levenberg–Marquardt (LM) backpropagation algorithm [20], [24], [25], which provide efficient ANN-based model with lower complexity and reduced computational requirement training process, can be chosen for utilizing for solving non-linear RF device behavioral modeling problems [26]. However, as briefly shown in Fig.1, the ANN model fitting process generally requires an initial setup for the values of the hidden neuron numbers, weights, and biases. Different initialization points used when training ANNs yield different results with different associated errors [23], [27]–[29]. Overdetermining the values may lead to the phenomenon of overlearning [22], [30]. The optimization methods for finding the optimal configuration of an ANN behavioral model may rely on the necessary experience from the operator. In all cases, a successful ANN model training session may require multiple attempts.

Non-linear transistor behavioral models used in RF design, the look-up table based, curve-fitted mathematical formulations are typically black-box models. Consider, however, the Cardiff Model (CM), which provides a physics based mathematically formulated non-linear model [10], [31], [32]. It works by transferring the load-pull measurement data of the RF transistors to a set of mixing theory based coefficients. But it is typically extracted over a limited load-pull impedance range, due to both measurement system limitations and transistor operation constraints, which may lead to the non-convergence or non-physical solution when used in CAD optimization tools [26]. Knowing the system level non-linearity determined polynomial order, hence the required number of CM coefficients, it is able to provide reliable accurate models using tailored datasets [33]. If the CM could be used to indicate the structure of the ANN and influence parameter magentadetermination (values of the weights and biases), it would remove the requirement for random ANN structure selection and parameter initialization. Simultaneously, as an ANN model structure, the possibility to overcome the converging issues with reasonable extrapolation predictions [26] can be improved.

The article is organized as follows: In yellowSection II, a novel method that does allow the ANN model structure and magenta~~initial~~ parameter values to be determined from the CM is presented. In yellowSection III, the verification of the proposed method will be undertaken, using firstly, a set of load-pull simulation data from ADS on a Wolfspeed GaN device. Secondly, a dense set of data acquired from the load-pull measurement done on the WIN GaN on-wafer device is used to further prove the method's robustness in yellowSection IV. The CM and ANN models extracted from the WIN device measured data will be used to highlight the extrapolating advantages of using the ANN solution. Finally, the discussion and conclusion are given in yellow Sections V and VI.

II. Method Proposing Procedure

The overview of the proposed method is shown as a flow diagram in Fig. 2.

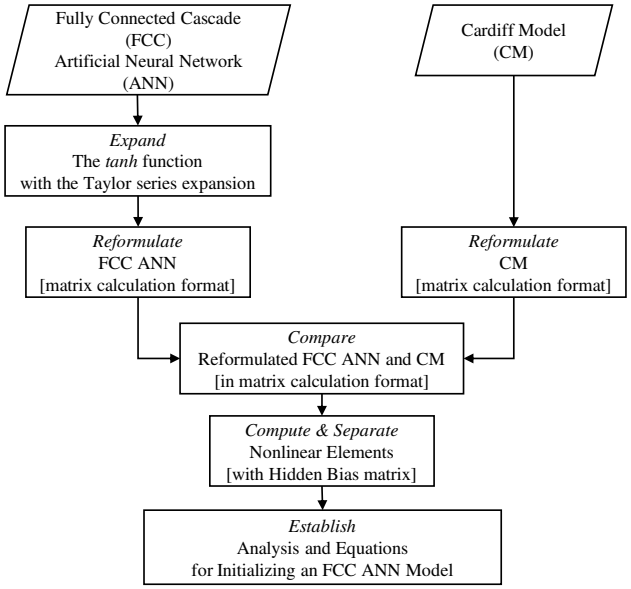


Fig. 2. Flow diagram summarizing the procedure used to establish the equivalence between ANN and CM coefficients.

It follows these steps in detail:

1) Since ANNs require complex numbers to be processed with the real and imaginary parts separated, the CM (structure diagram shown in Fig. 3) needs to be reformulated to accommodate the separation of the complex variables (following the details explained in Step 1 with Appendix A).
2) A single hidden layer structure ANN model [13] (conventional structure in Fig. 4) is reformulated, with the Taylor series expansion being applied to replace the *tanh* activation function. This provides a better understanding of how an ANN's structure operates on non-linear segments from layer to layer with the *tanh* activation function.
3) Following the details explained in Step 2 with Appendix B, it is important to note that the *tanh* function generates both linear and constant terms as well as higher order terms. To account for the linear and constant terms, a Fully Connected Cascaded (FCC) ANN structure is selected to ensure accurate prediction (interpolation) of measured results. It is the higher order terms that contribute to an improved extrapolation ability [12], [18].
4) After the ANN and CM models have been reformulated, their elements are matched to establish a new set of equations for identifying the required FCC ANN model configuration.
5) Finally, the proposed method for magentadetermining an FCC ANN model parameters is derived.

A. Step 1: CM Equations Reformulation

The original CM mathematically relates the input and output power waves of the device through a set of model coefficients

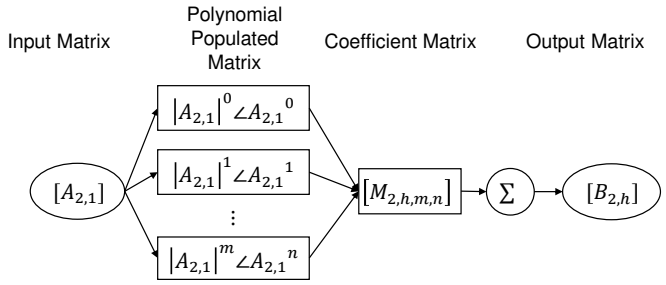


Fig. 3. Conventional CM model diagram.

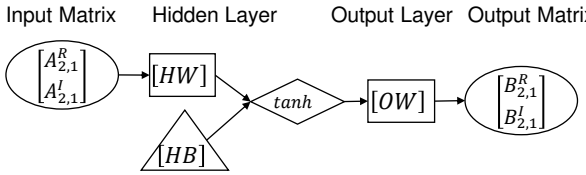


Fig. 4. Conventional ANN model diagram.

'$M_{p,h,m,n}$' [33], [34]. The formulation is phased normalized to $A_{1,1}$ for fundamental load-pull measurement datasets is as (1):

$$B_{p,h} = \sum_r \sum_n M_{p,h,m,n} \left|A_{2,1}\right|^m (\angle A_{2,1})^n \tag{1}$$

where $\angle A_{2,1}$ represents the complex exponential of the phase of $A_{2,1}$, the 'p' and 'h' subscripts denote the respective port and harmonics.

The exponents choice is driven by mixing theory, which makes the model physically meaningful. 'm' and 'n' denote the coefficient related power wave's magnitude and complex exponential of the phase respectively. The 'm' and 'n' are related as '$m = |n| + 2r$', where 'r' is the magnitude indexing term is limited to integer values from 0 to 1 because of extrapolation concerns [31], [33], [34]. The simple CM diagram can be shown as in Fig. 3.

Having the '$A_{2,1}$' stimulus incident waves separated into real and imaginary parts the CM can be reformulated as (2).

$$\begin{aligned}
\begin{bmatrix} B_{p,h}^R \\ B_{p,h}^I \end{bmatrix} &= [M_0] \\
&+ [M_1] \begin{bmatrix} A_{2,1}^R \\ A_{2,1}^I \end{bmatrix} \\
&+ [M_2] \begin{bmatrix} \left(A_{2,1}^R\right)^2 \\ A_{2,1}^R A_{2,1}^I \\ \left(A_{2,1}^I\right)^2 \end{bmatrix} \\
&+ [M_3] \begin{bmatrix} \left(A_{2,1}^R\right)^3 \\ \left(A_{2,1}^R\right)^2 A_{2,1}^I \\ A_{2,1}^R \left(A_{2,1}^I\right)^2 \\ \left(A_{2,1}^I\right)^3 \end{bmatrix} \\
&\vdots \\
&+ [M_o] \begin{bmatrix} \cdot \\ \left(A_{2,1}^R\right)^{m-k} \left(A_{2,1}^I\right)^k \\ \cdot \end{bmatrix}
\end{aligned} \tag{2}$$

where 'R' and 'I' represent the real and imaginary parts of the complex data segments, the $[M_o]$ matrix is computed from sets of model coefficient '$M_{p,h,m,n}$', exponent pairs (m, n), associated with the respective mixing order [10], and 'k' ranges from 0 to 'o', where 'o' represents the expanded polynomial order. More details can be found in Appendix A.

In this case, the reformulated CM equation (2) can also be presented by the structure diagram shown in Fig. 5. When observing the structure of both models, as in Fig. 4 and 5, the similarities between the two model structures can be seen. The coefficient matrix columns in Fig. 5 can fulfill a similar process as the hidden layer and output layer in Fig. 4.

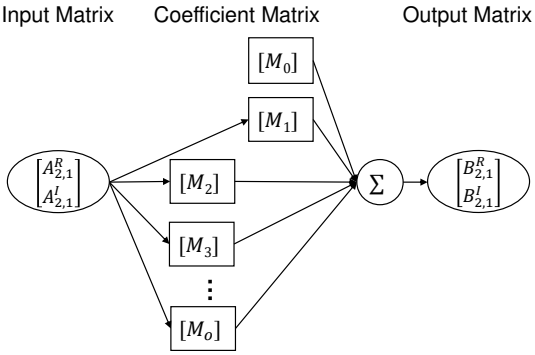


Fig. 5. Diagram of the CM model reformulated for Real and Imaginary parts.

B. Step 2: General ANN Equations Reformulation

Before moving on to the FCC ANN model structure, the general ANN model structure shown in Fig. 4, is first presented. Since the ANN-based model necessitates having any input and output matrices split from the complex form into separate real and imaginary matrices, it executes the following mathematical formulation:

$$\begin{bmatrix} B_{p,h}^R \\ B_{p,h}^I \end{bmatrix} = [OW] \tanh\left([HW] \begin{bmatrix} A_{2,1}^R \\ A_{2,1}^I \end{bmatrix} + [HB]\right) \tag{3}$$

where '*R*' and '*I*' are the abbreviations of real and imaginary parts of the complex data segments, while parameters '[*HW*]', '[*HB*]' and '[*OW*]' represent the number of hidden weights, hidden biases, and output weights respectively.

Inside the ANN structure, the non-linear transfer characteristic function *tanh*, also called the activation function [35], is the key to creating non-linearity within any ANN system. By formulating the *tanh* activation function with its Taylor series expansion, the non-linearity terms are clearly separated (the expanded version is shown in Appendix B) and so provide a formulation that can be compared with the reformulated CM structure.

It must be noted that the constant and linear elements inside the Taylor series reformulated ANN model will not be independent of the higher-order non-linearities. Hence, analysis done on the Taylor Series expansion starts from the higher order elements, respectively, because the parameters ('[*HW*]', '[*HB*]' and '[*OW*]') are shared between linear and non-linear elements, so lower order elements will be affected by the higher order elements due to the calculation operation order. Therefore, extra weight '[*CW*]' (Cascaded Weights) and bias matrices '[*OB*]' (Output Bias) will be required for correcting the constant and linear elements provided by the Taylor series expansion. Also, knowing from [18] that the Knowledge-based Neural Network structure is proven to be more robust than the multilayer perceptron (MLP) structure, especially when the dataset is insufficient. The general ANN structure, in Fig. 4, will then be transformed into the FCC ANN structure [22], which is a basic KBNN, shown in Fig. 6.

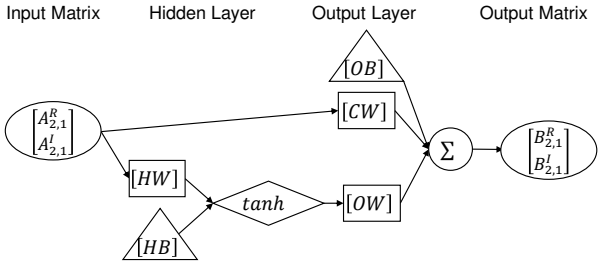


Fig. 6. FCC ANN model diagram.

C. Step 3: FCC ANN Equations Reformulation

The Taylor series formulation of this FCC ANN model can be written as:

$$\begin{aligned}\begin{bmatrix}B_{p,h}^{R}\\ B_{p,h}^{I}\end{bmatrix} &= ([OB]+[OW][\alpha])\\ &+([CW]+[OW]\Delta[\alpha][HW])\begin{bmatrix}A_{2,1}^{R}\\ A_{2,1}^{I}\end{bmatrix}\\ &+[OW]\Delta^{2}[\alpha]\left[HW^{2}\right]\begin{bmatrix}\left(A_{2,1}^{R}\right)^{2}\\ A_{2,1}^{R}A_{2,1}^{I}\\ \left(A_{2,1}^{I}\right)^{2}\end{bmatrix}\\ &+[OW]\Delta^{3}[\alpha]\left[HW^{3}\right]\begin{bmatrix}\left(A_{2,1}^{R}\right)^{3}\\ \left(A_{2,1}^{R}\right)^{2}A_{2,1}^{I}\\ A_{2,1}^{R}\left(A_{2,1}^{I}\right)^{2}\\ \left(A_{2,1}^{I}\right)^{3}\end{bmatrix}\\ &\vdots\\ &+[OW]\Delta^{o}[\alpha]\left[HW^{o}\right]\begin{bmatrix}.\\ (A_{2,1}^{R})^{o-k}(A_{2,1}^{I})^{k}\\ .\end{bmatrix}\end{aligned} \tag{4}$$

'*NH*' here represents the number of hidden neurons used in the ANN structure. And this is used to compute the Taylor series terms:

$$[\alpha]=\tanh[HB] \tag{5}$$

$$\Delta^{o}[\alpha]=\frac{1}{o!}\frac{\partial^{o}\tanh([HB])}{\partial([HB])^{o}}. \tag{6}$$

where $[HB]$ is given by:

$$[HB]=\begin{bmatrix}HB_{1,1}\\ \vdots\\ HB_{NH,1}\end{bmatrix} \tag{7}$$

and the $[HW]$ is given by:

$$[HW]=\begin{bmatrix}HW_{1,1} & HW_{1,2}\\ & \vdots\\ HW_{NH,1} & HW_{NH,2}\end{bmatrix}. \tag{8}$$

Since the defined FCC ANN model deals with datasets that consist of real and imaginary pairs, the two elements in (8) associated with each hidden neuron in the $[HW]$ matrix, will also represent the real and imaginary parts of the same complex number. Hence, $[HW]$ can be re-written as follows:

$$[HW]=\begin{bmatrix}\rho_{1}\cos\theta_{1} & \rho_{1}\sin\theta_{1}\\ & \vdots\\ \rho_{NH}\cos\theta_{NH} & \rho_{NH}\sin\theta_{NH}\end{bmatrix} \tag{9}$$

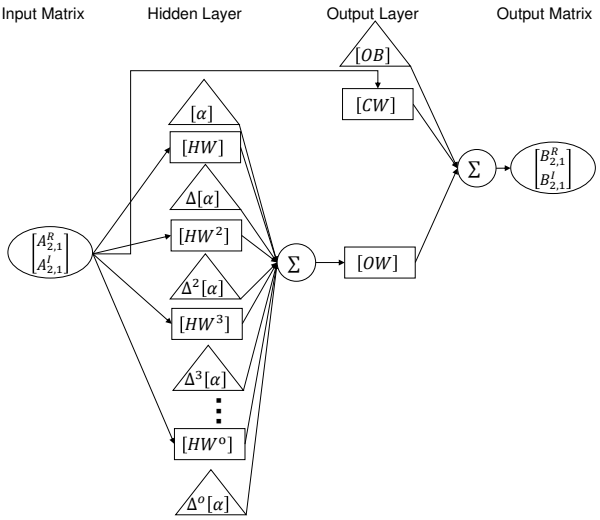


Fig. 7. Reformulated FCC ANN model diagram.

Defining the $[HW]$ matrix in the polar form provides an easier way to explore possible suitable values. This format is then used to determine corresponding $[HW^o]$ matrices:

$$[HW^o] = \begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \rho_{NH} \end{bmatrix} \begin{bmatrix} \cdot & \binom{o}{k}(\cos\theta_1)^{o-k}(\sin\theta_1)^k & \cdot \\ & \vdots & \\ \cdot & \binom{o}{k}(\cos\theta_{NH})^{o-k}(\sin\theta_{NH})^k & \cdot \end{bmatrix} \tag{10}$$

Note, the expansion of $[HW^o]$ matrix follows the binomial theorem. 'k' ranges from 0 to o, and 'o' refers to the polynomial order number. The positive integer $\binom{o}{k} = \frac{o!}{k!(o-k)!}$ is known as the binomial coefficient. More details about the reformulation process can be found in Appendix B.

By using the matrices formulation, shown in (4), the FCC ANN structure can be re-drawn as shown in Fig. 7. It is now clear how the levels of non-linearity in the system are being processed by different weights and biases inside the FCC ANN model. A direct comparison can now be done between the reformulated CM (Fig. 5) and the FCC ANN model (Fig. 7).

D. Step 4: Equating Formulations and ANN Model Structure Identification

Comparing equations (2) and (4), and also Fig. 5 and 7, the non-linearities inside the two systems can be equated as follows: where the first line of the matrices' calculation in (4) will match with the constant $[M_0]$ matrix in (2); the second line of the matrices' calculation in (4) is generating the fundamental part of the response; and the 2^{nd} and 3^{rd} order responding output non-linearities are generated by the third and fourth row of the matrices' multiplication in (4), respectively. Hence, a set of equations (11) - (15) can be extracted by equating the linear and non-linear terms inside the calculation processes:

$$[M_0] = [OB] + [OW]\,[\alpha] \tag{11}$$

$$[M_1] = [CW] + [OW]\,\Delta\,[\alpha]\,[HW] \tag{12}$$

$$[M_2] = [OW]\,\Delta^2\,[\alpha]\,[HW^2] \tag{13}$$

$$[M_3] = [OW]\,\Delta^3\,[\alpha]\,[HW^3] \tag{14}$$

$$\vdots$$

$$[M_o] = [OW]\,\Delta^o\,[\alpha]\,[HW^o] \tag{15}$$

These equations provide a method for linking CM complexity and coefficients to FCC ANN based model structure and parameters.

By knowing the number and dimensions of the $[M_o]$ matrices required from the identified CM, the associated ANN matrices dimensions, $[OW]$, $[HB]$ and $[HW]$, necessary to ensure that the matrix equations (11) - (15) are self-consistent, can now be determined. The ANN structure with the required hidden neuron number can now be directly identified, by analyzing load-pull data complexity using the CM. Consider the case where the accurate modeling of the load-pull data is found to require a 3^{rd} order non-linear CM [33]. In this case, we need an ANN structure that can satisfy (11) - (14). Following the contribution of the non-linearities inside the expanded ANN structure, the higher order non-linearities segment will be analysed first, by solving (13) - (14), then the linear segment can be calculated, by solving (11) - (12).

The size of the $[M_2]$ matrix, given by the CM, in this case, is $[2 \times 3]$. Hence, the matrices dimensions on the right hand side of (13), $[OW]\,\Delta^2\,[\alpha]\,[HW^2]$, must also result in a $[2 \times 3]$ matrix, hence implying an ANN structure with 3 hidden neurons. There is the same scenario for solving (14). The size of the $[M_3]$ matrix, given by the CM, is $[2 \times 4]$. Hence the $[OW]\,\Delta^3\,[\alpha]\,[HW^3]$ product must also result in a $[2 \times 4]$ matrix, implying an ANN structure with 4 hidden neurons. However, in a given FCC ANN model the number of hidden neurons is a single value, hence different sizes for the required weights ($[OW]$ and $[HW]$), and bias ($[HB]$) matrices is not possible. This can be addressed by using the sum of the two different matrix sizes which, in this case, leads to an FCC ANN structure with 7 hidden neurons.

In summary, it has been identified that an FCC ANN model with 7 hidden neurons is the maximum complexity necessary to model a 3^{rd} order non-linear system. ~~Note, once this ANN model has been determined (weights and biases) through backpropagation training, the corresponding CM coefficients '$M_{p,h,m,n}$'/'$[M_o]$' matrices, can be computed from (11) - (14).~~

E. Step 5: FCC ANN Model Parameter magentadetermination

Now, the ANN model structure identification has been achieved. There is still a potential situation where the initial values of the ANN model parameters lead to a non-converging backpropagation training process. Therefore, having proper magentadetermined values of the ANN model parameters $[OW]$, $[HB]$ and $[HW]$ can be helpful for the training process. However, it is not possible to compute the ANN model parameters, the three matrices $[OW]$, $[HB]$ and $[HW]$, directly from the CM by simply reversing equations (11) - (14). Two of the ANN model matrices have to be predetermined. Following

the feed forward process of the ANN structure, the first set of matrices that should be defined are matrix $[HB]$, and $[HW]$. Once determined, $[OW]$ can then be computed directly using equations (16) and/or (17).

$$[OW] = [M_2]\left[\Delta^2[\alpha]\left[HW^2\right]\right]^{-1} \tag{16}$$

$$[OW] = [M_3]\left[\Delta^3[\alpha]\left[HW^3\right]\right]^{-1} \tag{17}$$

The selection of $[HB]$ is critical to ensuring that equations (16) and (17) give the same value of $[OW]$. It can be seen from the reformulated CM equation (2) that the $[M_2]$ matrix deals with the 2^{nd} order non-linearity and the $[M_3]$ matrix ideals with the 3^{rd} order non-linearity. To magentadetermine an FCC ANN using (13) and (14), according to the calculation through the *tanh* function, the $[HB]$ matrix can now be used to enable hidden neurons to target specific orders of non-linearity. The values of $\Delta^o[\alpha]$ matrices will be the key in this step.

The 3^{rd} order non-linear systems require an ANN model with 7 hidden neurons. Hence, $[HB]$ and $[HW]$ are as follows:

$$[HB] = \begin{bmatrix} HB_{1,1} \\ HB_{2,1} \\ HB_{3,1} \\ HB_{4,1} \\ HB_{5,1} \\ HB_{6,1} \\ HB_{7,1} \end{bmatrix} \tag{18}$$

$$[HW] = \begin{bmatrix} HW_{1,1} & HW_{1,2} \\ HW_{2,1} & HW_{2,2} \\ HW_{3,1} & HW_{3,2} \\ HW_{4,1} & HW_{4,2} \\ HW_{5,1} & HW_{5,2} \\ HW_{6,1} & HW_{6,2} \\ HW_{7,1} & HW_{7,2} \end{bmatrix} \tag{19}$$

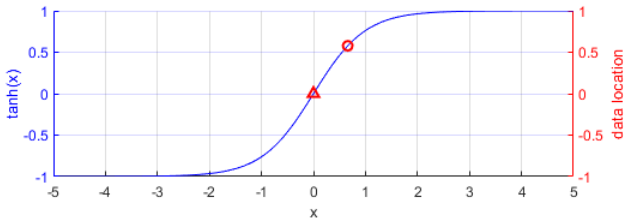


Fig. 8. Selected bias point (red markers) locations on the $tanh(x)$ function. Location without a 3^{rd} order derivative (triangle) and location without a 2^{nd} order derivative (circle).

Referring to the Taylor series expansion of the *tanh* function, see Appendix B, if an element of matrix $[HB]$ is set to 0.6585 (bias point Fig. 8-triangle marker) then the 3^{rd} order element will be filtered out, because the corresponding element of matrix $\Delta^3[\alpha]$ is 0; alternatively when an element of matrix $[HB]$ equals 0 (bias point Fig. 8-circle marker), then the 2^{nd} order element will be filtered out because the corresponding element of matrix $\Delta^2[\alpha]$ is 0.

Hence, using the required matrix sizes, mentioned in step 4, set the values of $[HB]$ matrix elements as follows:

$$[HB] = \begin{bmatrix} HB_{1,1} \\ HB_{2,1} \\ HB_{3,1} \\ HB_{4,1} \\ HB_{5,1} \\ HB_{6,1} \\ HB_{7,1} \end{bmatrix} = \begin{bmatrix} 0.6585 \\ 0.6585 \\ 0.6585 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

hence

$$\left[\Delta^2[\alpha]\right] = \begin{bmatrix} 0.38 \\ 0.38 \\ 0.38 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{21}$$

and

$$\left[\Delta^3[\alpha]\right] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -0.33 \\ -0.33 \\ -0.33 \\ -0.33 \end{bmatrix} \tag{22}$$

The FCC ANN structure for 2^{nd} and 3^{rd} order non-linear contributions has now be separated and assigned to hidden neurons $[1-3]$ and $[4-7]$ respectively. There is now no requirement to solve (16) and (17) simultaneously. They can be reformulated as follows:

$$[OW_2] = 0.38\,[M_2]\left[HW_2^2\right]^{-1} \tag{23}$$

$$[OW_3] = -0.33\,[M_3]\left[HW_3^3\right]^{-1} \tag{24}$$

where the $[HW]$ and $[OW]$ have been separated into two parts based on the non-linear order, as follows:

$$[HW] = \begin{bmatrix} HW_2 \\ HW_3 \end{bmatrix} \tag{25}$$

and

$$[OW] = \begin{bmatrix} OW_2 & OW_3 \end{bmatrix} \tag{26}$$

1) Segment Analysis for the 2^{nd} Order: Analytical solutions for the 2^{nd} order segment is achievable if (23) can be solved. This is possible provided the $\left[HW_2^2\right]$ matrix can be inverted. Since this matrix refers to 3 hidden neurons [1-3] then $\left[HW_2^2\right]$, shown below, is a square matrix. A square matrix can be inverted provided it is non-singular.

$$\left[HW_2^2\right] = \begin{bmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} \cos^2\theta_1 & 2\cos\theta_1\sin\theta_1 & \sin^2\theta_1 \\ \cos^2\theta_2 & 2\cos\theta_2\sin\theta_2 & \sin^2\theta_2 \\ \cos^2\theta_3 & 2\cos\theta_3\sin\theta_3 & \sin^2\theta_3 \end{bmatrix} \tag{27}$$

To make sure that (27) is a non-singular square matrix, the values the 3 phase angles have to be selected to ensure the following conditions:

$$\begin{cases} \theta_1 \neq \theta_2 \pm p\pi \\ \theta_2 \neq \theta_3 \pm q\pi \end{cases} \tag{28}$$

where p and q are any positive integer. For example, in this paper, $\theta_1 = 45°$, $\theta_2 = 90°$ and $\theta_3 = 135°$ have been used.

Then the elements of the $[HW_2^2]$ matrix used in (23) can be calculated by selecting the value of ρ. The example in the next section will discuss in detail the criteria for the selection of ρ.

2) Segment Analysis for the 3^{rd} Order : Similarly, an analytical solution the 3^{rd} order segment, is achievable if (24), can be solved. This is possible provided the $[HW_3^3]$ matrix can be inverted. Since this matrix refers to 4 hidden neurons $[4-7]$, then $[HW_3^3]$, shown below, is a square matrix which, again, can be inverted provided it is non-singular.

$$[HW_3^3] = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} \cos^3\theta_4 & 3\cos^2\theta_4\sin\theta_4 & 3\cos\theta_4\sin^2\theta_4 & \sin^3\theta_4 \\ \cos^3\theta_5 & 3\cos^2\theta_5\sin\theta_5 & 3\cos\theta_5\sin^2\theta_5 & \sin^3\theta_5 \\ \cos^3\theta_6 & 3\cos^2\theta_6\sin\theta_6 & 3\cos\theta_6\sin^2\theta_6 & \sin^3\theta_6 \\ \cos^3\theta_7 & 3\cos^2\theta_7\sin\theta_7 & 3\cos\theta_7\sin^2\theta_7 & \sin^3\theta_7 \end{bmatrix} \tag{29}$$

To make sure that (29) is a non-singular matrix, also follows the conditions shown with (28) for all the four angles, for example, $\theta_4 = 36°$, $\theta_5 = 72°$, $\theta_6 = 108°$ and $\theta_7 = 144°$ are used in this paper. Then the value of the $[HW_3^3]$ matrix used to solve (24) can be selected by sweeping the value of ρ, relating to the specific dataset.

In summary, it has been shown that the proper selection of the hidden node bias values allows for the allocation of hidden nodes to a specific non-linear order, thus enabling magenta~~ini-tial~~ values of the FCC ANN model parameters to be directly computed from the corresponding CM coefficients '$M_{p,h,m,n}$'.

III. Method Verification With Simulations

In this section, a set of load-pull simulation data, with a complete 3^{rd} order data complexity, will be used as the first step for verifying the method using data that has a very low noise floor. The way of using the proposed method is summarized in the flow diagram, shown in Fig.9.

A. Simulation yellowData Acquisition

The load-pull simulation is set up in ADS for a Wolfspeed 10 W device (CG2H40010F), biased at V_{gs} = - 2.2 V, V_{ds} = 28 V. In Fig. 10-11, the data is collected with a constant input drive corresponding to 1 dB compression at the optimum load. According to the analysis in [33], 10 coefficients should be extracted here for an accurate CM with the -60 dB error threshold (red line in Fig. 11-right). Knowing that the ANN training process require data normalization before feeding the data in [26], the CM is extracted after normalizing the data based on the *tanh* function range, [-1 1]. The magnitude and phase restricted terms exponents and the extracted coefficients are determined, calculated, and listed in Table I. The performance of the extracted CM will be discussed later in the next step.

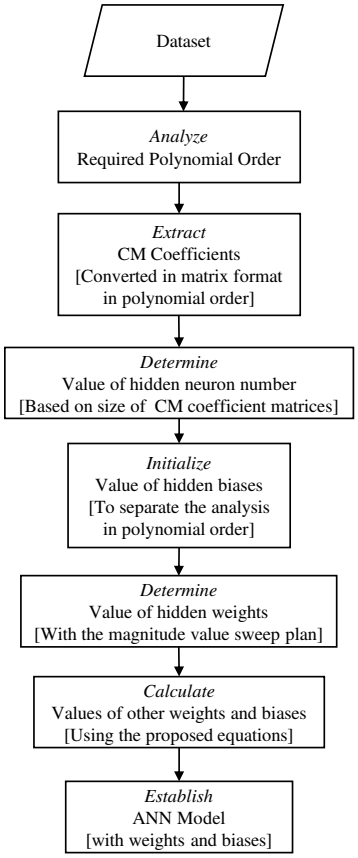
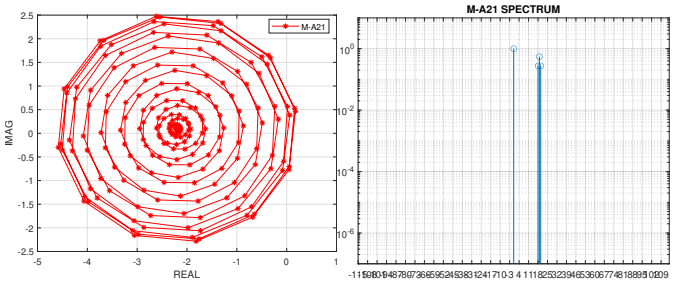


Fig. 9. Flow diagram summarizing the proposed ANN magentadetermining method.



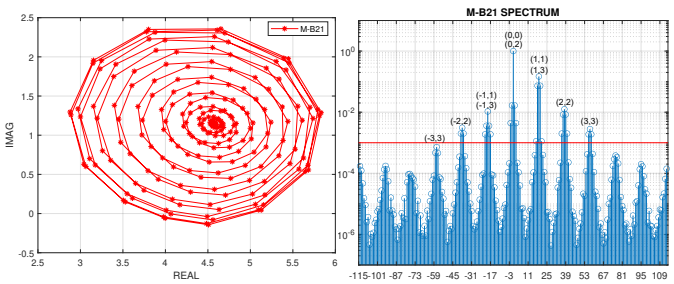
(a) (b)

Fig. 10. $A_{2,1}$ sample (a) and its Fourier Transformed spectrum (b) on a trajectory computed using 229 simulated points with the phase modulation rate of Sp = 19.

B. Method Created Model Analysis with ρ Sweep

According to the identified CM polynomial terms shown in Fig.11 using the analysis explained in [33], it is now clear that $[M_o]$ matrices with the value of m up to 3 are all that are required to model this dataset. As defined in section II-D, a dataset that requires a 3^{rd} order non-linear CM will determine a complete FCC ANN model structure as 1 hidden layer with 7 hidden neurons.

Following the analysis process in section II-E, with the selected phase angles $\theta_1 = 45°$, $\theta_2 = 90°$ and $\theta_3 = 135°$ for $[HW_2]$ and $\theta_4 = 36°$, $\theta_5 = 72°$, $\theta_6 = 108°$ and $\theta_7 = 144°$ for $[HW_3]$, the magentadetermined value of $[HW]$ and $[OW]$ can



(a) (b)

Fig. 11. $B_{2,1}$ sample (b) and its Fourier Transformed spectrum (a) with error bar (red) on the trajectory computation, and the expected exponents' $|A_{2,1}|^m(\angle A_{2,1})^n$ values of the CM for extracting coefficients in (n,m).

TABLE I
THE EXTRACTED CARDIFF MODEL COEFFICIENTS FOR $B_{2,1}$

				$M_{2,1,m,n}$	
index	*r*	*n*	*m*	*Real*	*Imag.*
1	0	0	0	0.7597	0.1
2	0	1	1	-0.5608	-0.0748
3	0	-1	1	-0.0502	0.0390
4	0	2	2	0.0741	-0.1155
5	1	0	2	-0.2547	-0.1019
6	0	-2	2	0.0222	0.0247
7	1	3	3	-0.0638	-0.0333
8	1	1	3	0.0108	0.1137
9	0	-1	3	0.1177	-0.1419
10	0	-3	3	-0.0162	-0.0070

be explored with the CM coefficients (listed in Table I) and a magnitude element ρ sweep plan for this specific dataset.

The effect brought by the different values of ρ is shown in Fig.12 and 13. It can be seen from the figures that a bigger value of ρ results in a bigger coverage on the *tanh* function within the acquired dataset range (Fig.12(a)-(b)), which then generates a $B_{2,1}$ model with a more complex Fourier Transformed spectrum, hence an effective "higher noise floor" comparing to the collected dataset (Fig.12(c)) and the CM predictions (Fig.12(d)). Also, as shown in Fig.13(a), the performance of the determined ANN model and the CM gets identical when the value of ρ get close to 0. Eventually, as ρ is swept from 0 up to 3, as shown in Fig.13(b), the accuracy of the model decreases.

Theoretically, the value of ρ can be selected from any point where the ANN model is performing an NMSE level below - 40 dB [32] for an accurate model. However, knowing from Fig.12 and 13 that the accuracy of determined ANN model decreases when ρ gets bigger, simultaneously, performs identically to the CM with ρ approaching 0. In this specific case, $\rho = 0.9$ is selected to determine an accurate ANN model, but also different from the CM. With the defined $[HB]$, $[HW_2]$ and $[HW_3]$. (23) and (24) can now be solved to determine the associated $[OW_2]$ and $[OW_3]$ matrices.

Then, the value of $[CW]$ and $[OB]$ can all be calculated and listed as follows:

TABLE II
THE EXTRACTED CARDIFF MODEL COEFFICIENTS FOR $B_{2,0}$

				$M_{2,0,m,n}$	
index	*r*	*n*	*m*	*Real*	*Imag.*
1	0	0	0	0.0126	0
2	0	1	1	0.0115	0.0234
3	0	-1	1	0.0115	-0.0234
4	0	2	2	0.0104	-0.0127
5	1	0	2	-0.0487	0
6	0	-2	2	0.0104	0.0127
7	1	3	3	-0.0065	-0.0029
8	1	1	3	0.0226	0.0224
9	0	-1	3	0.0226	-0.0224
10	0	-3	3	-0.0065	0.0029

$$[HW] = \begin{bmatrix} HW_2 \\ HW_3 \end{bmatrix}$$

$$[!ht] = \begin{bmatrix} HW_{1,1} & HW_{1,2} \\ HW_{2,1} & HW_{2,2} \\ HW_{3,1} & HW_{3,2} \\ HW_{4,1} & HW_{4,2} \\ HW_{5,1} & HW_{5,2} \\ HW_{6,1} & HW_{6,2} \\ HW_{7,1} & HW_{7,2} \end{bmatrix} = \begin{bmatrix} 0.6364 & 0.6364 \\ 0 & 0.9 \\ -0.6364 & 0.6364 \\ 0.7281 & 0.5290 \\ 0.2781 & 0.8560 \\ -0.2781 & 0.8560 \\ -0.7281 & 0.5290 \end{bmatrix} \quad (30)$$

$$[OW] = [OW_2 \quad OW_3]$$
$$= \begin{bmatrix} OW_{1,1} & OW_{1,2} & OW_{1,3} & OW_{1,4} & OW_{1,5} & OW_{1,6} & OW_{1,7} \\ OW_{2,1} & OW_{2,2} & OW_{2,3} & OW_{2,4} & OW_{2,5} & OW_{2,6} & OW_{2,7} \end{bmatrix} \quad (31)$$
$$= \begin{bmatrix} 0.0581 & 0.6176 & 0.9579 & 0.0191 & -0.1231 & 1.3936 & 0.3113 \\ 0.4517 & -0.5826 & 0.7844 & 0.7317 & -0.4842 & 0.5626 & 0.1400 \end{bmatrix}$$

$$\begin{bmatrix} CW_{1,1} & CW_{1,2} \\ CW_{2,1} & CW_{2,2} \end{bmatrix} = \begin{bmatrix} 0.4053 & -1.9500 \\ -0.0343 & -1.2136 \end{bmatrix} \quad (32)$$

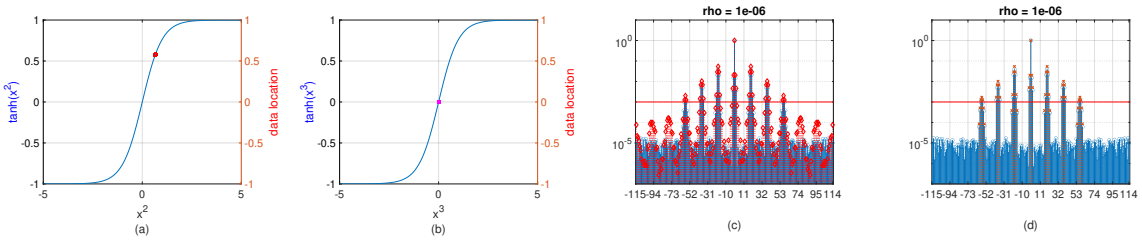
$$\begin{bmatrix} OB_{1,1} \\ OB_{2,1} \end{bmatrix} = \begin{bmatrix} -0.1835 \\ -0.2773 \end{bmatrix} \quad (33)$$

Performing a feedforward process on the generated FCC ANN, with the weights and biases listed above, results in a set of $B_{2,1}$ data that can be compared with the $B_{2,1}$ data that is collected from the simulation, see Fig. 14.

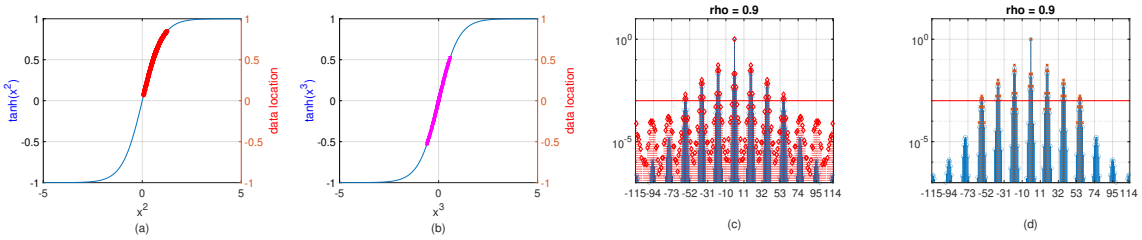
C. Repeated Analysis for DC yellowComponent

The $B_{2,0}$, which is the drain current collected from ADS simulation, is shown in Fig. 15, also has a spectrum complexity consistent with a model order of 3. Therefore, 10 CM coefficients are extracted for the ANN magentadetermination calculation, listed in Table II. The FCC ANN model can be magentadetermined for $B_{2,0}$ with the same 7 hidden neuron structure, using the same process as in the previous sections.

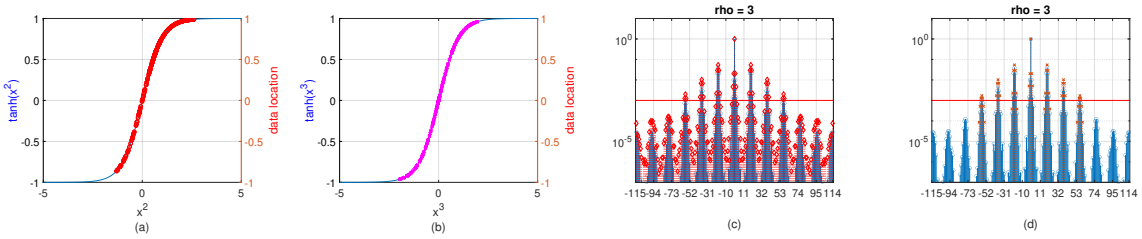
Here, $\rho = 0.9$ is still the selected case following the ρ sweep. Therefore, the value of $[HB]$ and $[HW]$ will stay the same as previously listed with the phase angles determined in section II-E. The calculated values of weights and biases (using (34)-(36)) are listed below. Performing a feedforward process on the determined ANN model, the predicted set of



(a) $\rho = 0$

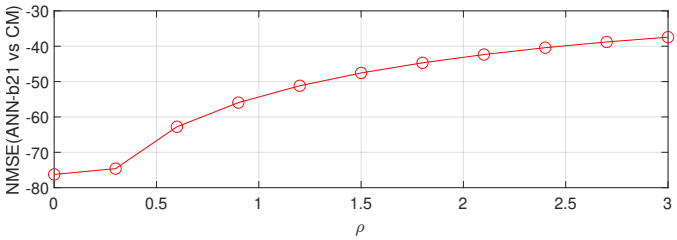


(b) $\rho = 0.9$

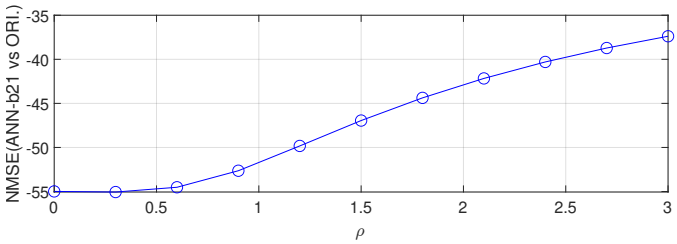


(c) $\rho = 3$

Fig. 12. The dataset coverage on the $tanh(x)$ function range without the $3^{rd}/2^{nd}$ order segment (two leftmost plots) location after magentadetermined weights and biases. The Fourier Transformed spectrum comparison plot (two rightmost plots) of the modeled $B_{2,1}$(blue) and the $B_{2,1}$ samples from the simulation (red) and the calculated from the CM (orange) with error bar (redline) on the trajectory computation.



(a) NMSE from ANN prediction vs. Cardiff Model prediction



(b) NMSE from ANN prediction vs. Simulation data

Fig. 13. The FCC ANN performance for $B_{2,1}$ prediction changing trend with ρ sweep from 0 to 3.

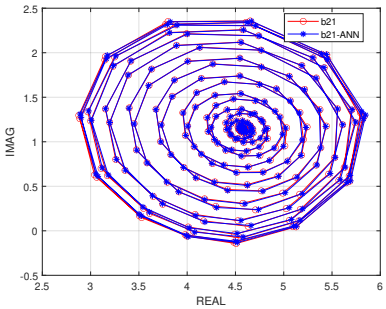


Fig. 14. The $B_{2,1}$ prediction from the magentadetermined FCC ANN (blue stars) and the $B_{2,1}$ acquired from the simulation with complete data complexity (red circles).

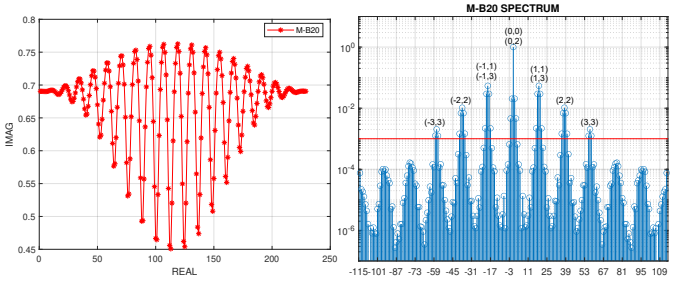
$B_{2,0}$ data can be compared with the $B_{2,0}$ data that is collected from the simulation, see Fig. 16.

$$[OW] = [OW_2 \quad OW_3]$$
$$= \begin{bmatrix} OW_{1,1} & OW_{1,2} & OW_{1,3} & OW_{1,4} & OW_{1,5} & OW_{1,6} & OW_{1,7} \\ OW_{2,1} & OW_{2,2} & OW_{2,3} & OW_{2,4} & OW_{2,5} & OW_{2,6} & OW_{2,7} \end{bmatrix} \tag{34}$$
$$= \begin{bmatrix} 0.0084 & 0.1328 & 0.1711 & -0.0985 & 0.0303 & 0.2010 & 0.1425 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} CW_{1,1} & CW_{1,2} \\ CW_{2,1} & CW_{2,2} \end{bmatrix} = \begin{bmatrix} 0.3148 & -0.4240 \\ 0 & 0 \end{bmatrix} \tag{35}$$

$$\begin{bmatrix} OB_{1,1} \\ OB_{2,1} \end{bmatrix} = \begin{bmatrix} -0.1677 \\ 0 \end{bmatrix} \tag{36}$$

When considering the coefficients in Table II, a clear pattern can be seen when the coefficients are extracted as conjugated



(a) (b)

Fig. 15. $B_{2,0}$ sample (a) and its Fourier Transformed spectrum (b) with error bar (red) on the trajectory computation, and the expected exponents' $|A_{2,1}|^m(\angle A_{2,1})^n$ values of the CM for extracting coefficients in (n,m).

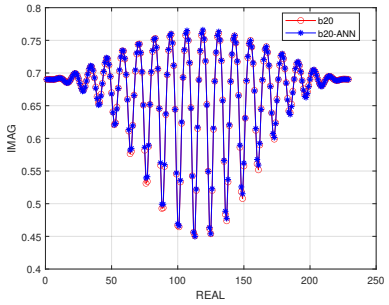


Fig. 16. The $B_{2,0}$ prediction from the magentadetermined FCC ANN (blue stars) and the $B_{2,0}$ acquired from the simulation with complete data complexity (red circles).

pairs. This will result in having the $[M_i]$ matrices simplified in dimension because the sum of the imaginary parts is 0. In this case, the imaginary part from the output of the FCC ANN does not have to be hard-wired to 0 as when using the conventional ANN structure. The imaginary part will be 0 directly after the calculation.

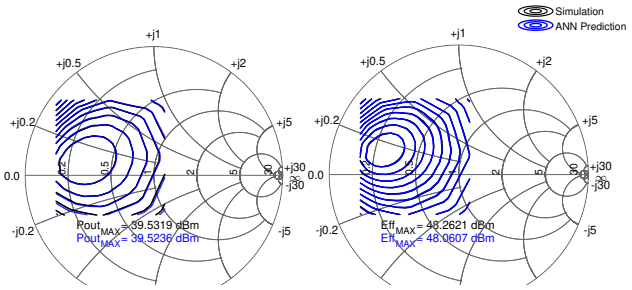
D. Power and Efficiency yellowContour Plots

With both $B_{2,1}$ and $B_{2,0}$, the prediction results for the output power and efficiency contours are shown in Fig. 17. The NMSE calculated between the proposed method (magentadetermined FCC ANN model) and the simulated dataset demonstrates an NMSE error level lower than - 50 dB [32] for both of the output power and efficiency contours. In this case, the proposed method for magentadetermining the FCC ANN behavioral model structure using the CM coefficients is proven accurate.

IV. Method Verification with Measurement

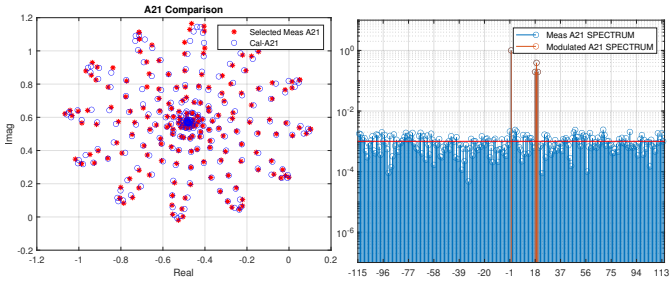
The proposed method will now be verified with a dataset acquired from load-pull measurements.

A high-density load-pull measurement was set up for a WIN NP12 4×75 GaN on-wafer device at 20 GHz, biased with $V_{DS} = 15\,V$, $I_{DS} = 30\,mA$ (setup as in [36]). In Fig. 18-19, the data is collected with a constant input drive corresponding to 3 dB compression at the optimum load for acquiring a dataset with higher complexity. To follow the analysis in [33], a dataset equivalent to a modulated stimulus signal $A_{2,1}$ is chosen from the measurement data points, as shown in Fig.



(a) (b)

Fig. 17. The predicted and simulated output power (a) and efficiency contours (b) comparison of the Wolfspeed device.



(a) (b)

Fig. 18. Selected $A_{2,1}$ samples (a) from the calculation (blue circles)and picked from measurement (red dots) and its Fourier Transformed spectrum comparison (b, calculated shown in red and measured shown in blue) on a trajectory computed using 229 simulated points.

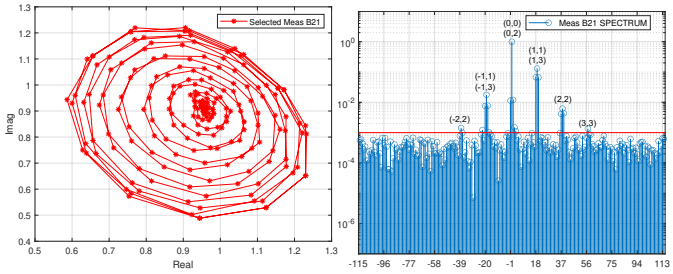
TABLE III
The Extracted Cardiff Model Coefficients For $B_{2,1}$

				$M_{2,1,m,n}$	
index	*r*	*n*	*m*	*Real*	*Imag.*
1	0	0	0	0.7589	0.7178
2	0	1	1	0.0010	-0.5840
3	0	-1	1	0.0599	-0.0622
4	0	2	2	0.0374	-0.0629
5	1	0	2	-0.1485	-0.1334
6	0	-2	2	-0.0098	0.0125
7	1	3	3	0.0021	-0.0319
8	1	1	3	0.0534	0.0152
9	0	-1	3	0.0130	0.1072
10	0	-3	3	0	0

19. To get a clean spectrum plot with a noise floor lower than -60 dB, the selected measurement $A_{2,1}$ dataset, needs to be a good match, at least within - 35 dB error level, to the ideal modulated $A_{2,1}$ dataset.

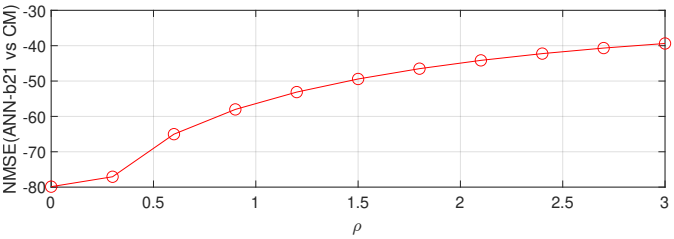
A. cyanModel Determination

According to the response Fourier Transformed spectrum in Fig. 19, 9 pairs of magnitude and phase restricted terms exponents can be defined. The correlated extracted coefficients are calculated and listed in Table III. Based on the CM reformulation (2), all the exponents pairs related coefficient components are required for a 3^{rd} polynomial order case, because of the $[M_o]$ matrices. When any of them do not exist on the spectrum plot, the coefficient values will be set to 0.

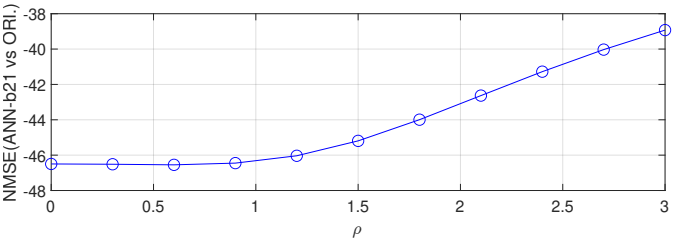


(a) (b)

Fig. 19. $B_{2,1}$ sample (a) and its Fourier Transformed spectrum (b) with - 60 dB error bar (red) on the trajectory computation, and the expected exponents' $|A_{2,1}|^m(\angle A_{2,1})^n$ values of the CM for extracting coefficients in (n,m).



(a) NMSE from ANN prediction vs. Cardiff Model prediction



(b) NMSE from ANN prediction vs. Simulation data

Fig. 20. The FCC ANN performance for $B_{2,1}$ prediction changing trend with ρ sweep from 0 to 3.

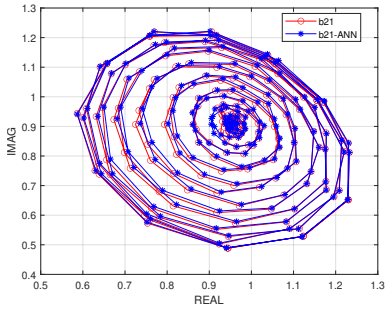
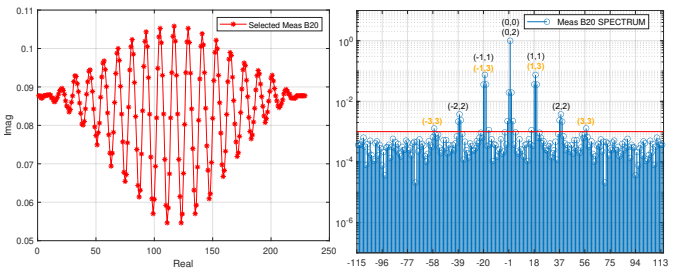


Fig. 21. The $B_{2,1}$ prediction from the magentadetermined FCC ANN (blue stars) and the $B_{2,1}$ acquired from the measurement with the analysed data complexity (red circles).

The required analysis with the ρ sweep plan and the calculation is done. According to the results shown in Fig. 20, $\rho = 1.2$ is selected in this case for a good model performance. The prediction results for the $B_{2,1}$ are shown in Fig. 21.

Repeating the same procedure for the DC component, $B_{2,0}$, which is the drain current. As shown in the spectrum plot from Fig. 22, there are 3^{rd} order terms located close to the - 60 dB noise floor, highlighted in yellow, which could be potentially



(a) (b)

Fig. 22. $B_{2,0}$ sample (a) and its Fourier Transformed spectrum (b) with - 60 dB error bar (red) on the trajectory computation, and the expected exponents' $|A_{2,1}|^m(\angle A_{2,1})^n$ values of the CM for extracting coefficients in (n,m).

TABLE IV
THE EXTRACTED CARDIFF MODEL COEFFICIENTS FOR $B_{2,0}$

				$M_{2,0,m,n}$	
index	*r*	*n*	*m*	*Real*	*Imag.*
1	0	0	0	0.0036	0
2	0	1	1	0.0096	0.0192
3	0	-1	1	0.0096	-0.0192
4	0	2	2	-0.0009	-0.0025
5	1	0	2	-0.0213	0
6	0	-2	2	-0.0009	0.0025

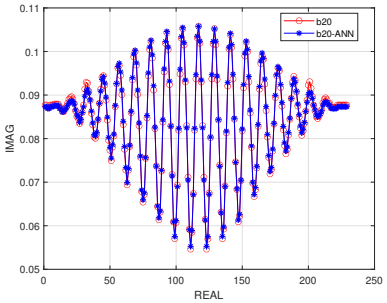


Fig. 23. The $B_{2,0}$ prediction from the magentadetermined FCC ANN (blue stars) and the $B_{2,0}$ acquired from the measurement with the analysed data complexity (red circles).

picked. However, the experiment demonstrates that the FCC ANN models magentadetermined without the 3^{rd} order terms are accurate enough, performing an NMSE level lower than - 40 dB. Here, with the 6 extracted CM coefficients listed in Table IV, the FCC ANN model magentadetermined using the proposed method with 3 hidden neurons in the hidden layer is providing $B_{2,0}$ predictions as in Fig. 23.

With both $B_{2,0}$ and $B_{2,1}$, the prediction results for the output power and efficiency contours for the WIN device are shown in Fig. 24. The NMSE calculated between the proposed method (magentadetermined FCC ANN model) and the measured dataset are both below - 40 dB [32] for the output power and efficiency contours.

B. Model Extrapolation Ability

While the range of $A_{2,1}$ values is limited during measurement because of transistor operation and measurement system limitations, much larger values of $A_{2,1}$ corresponding to loads covering the whole Smith Chart are possible during CAD

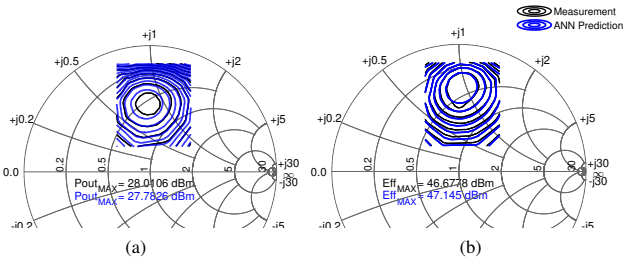


Fig. 24. The predicted and measured output power (a) and efficiency contours (b) comparison of the WIN device.

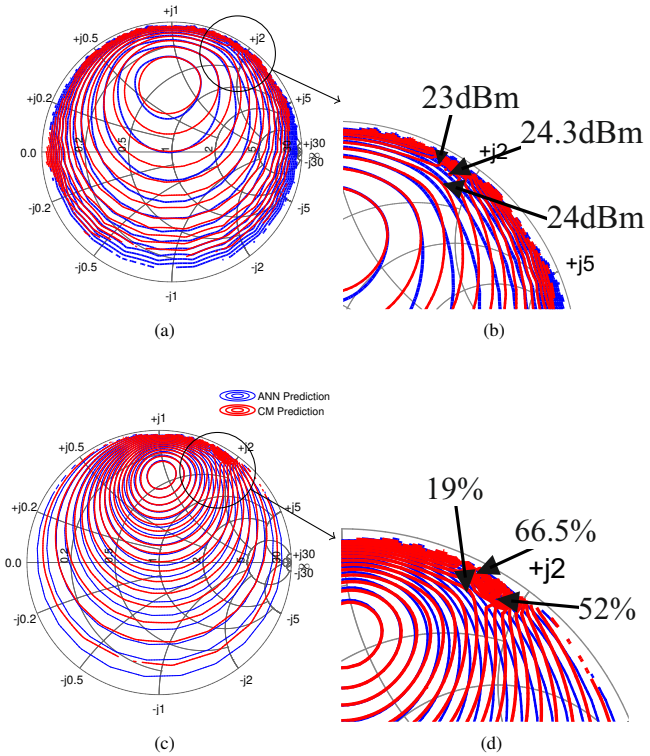


Fig. 25. Comparison of CM (red) and ANN (blue) modeling of the output power (a,b) and efficiency (c,d).

optimization. The extrapolation capability of both models therefore needs to be tested using an appropriately expanded $A_{2,1}$ range.

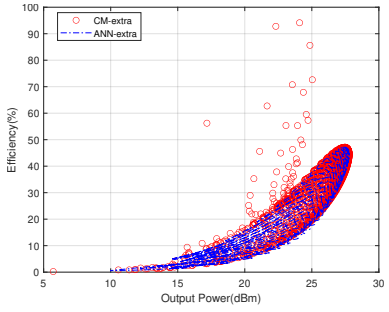


Fig. 26. Modeled efficiency vs. output power for the WIN device models, including the extrapolation region. CM (red circles) and ANN (blue line). The ANN does not show any of the CM unrealistic response.

The corresponding output power and efficiency contours are shown in Fig. 25. Zooming in a problem region identifies that the CM has some erroneous behavior. This can result in erroneous performance predictions, see Fig. 26. These features and erroneous predictions are not presented in the results offered by the proposed method determined ANN model.

V. Discussion

The method developed has proven to be robust enough to determine both an appropriate FCC ANN model structure and magenta~~provide an initial~~ determined set of ANN model parameters, weights and bias values, using both simulated and measured transistor load-pull data-sets.

A number of significant, additional, observations, need to be discussed. The phase angles selection, mentioned in Section II-E, highlights there are multiple valid, non-singular, solutions that can magentabe used for determining values for the $[HW]$ matrix, hence multiple magentadetermination options for the ANN model parameters. All these solutions provide similar model performance. Also, there are multiple (repeating) solutions of the $[HW]$ matrix, that provide an invalid, singular matrix condition.

Consider now the selection of ρ, which defines *tanh* function operational range of the ANN hidden nodes and the accuracy of the magentadetermined ANN model, ~~could also influence the ANN backpropogation process. Therefore, further investigation will need to be undertaken to determine if there is an optimal selection for ρ.~~ can be further investigated for a method for optimal selection. Besides, although there is no ANN training process presented in this paper, there is a potential that varying the value of ρ can have an impact on the training if required.

Alternatively, a pseudo inverse can also be applied to the calculation process. In this case, a square matrix is no longer required for the matrix calculation, hence neuron numbers could potentially be reduced without compromising the model accuracy.

VI. Conclusion

In this paper a process, based on the Cardiff Behavioral Model (CM), that determines the structure of a Fully Connected Cascaded (FCC) Artificial Neural Network (ANN) model, which is appropriate for non-linear transistor modeling, has been developed. In addition, a method for determining the values of the ANN model parameters, from the coefficients of the CM has been formulated.

With the dataset collected from the Wolfspeed GaN device load-pull simulation in ADS, the method has produced an accurate ANN model with an error level lower than -50 dB. For the WIN GaN device measurement data, the new proposed method also produces an accurate ANN model with an error level lower than -40 dB. In addition, the extrapolation ability of ANN models determined with the proposed methods are proven to be more reasonable than the CM.

Acknowledgments

The authors would like to thank WIN Semiconductors for the provision of the transistor samples that were used in this paper.

APPENDIX A
CARDIFF MODEL REFORMULATION

The original CM formulation (1) can be rearranged as follows:

$$\begin{aligned} B_{p,h} &= M_{p,h,0,0} \\ &+ M_{p,h,1,1}A_{2,1} + M_{p,h,1,-1}A_{2,1}^{*} \\ &+ M_{p,h,2,2}\left(A_{2,1}\right)^{2} + 2M_{p,h,2,0}A_{2,1}A_{2,1}^{*} \\ &+ M_{p,h,2,-2}\left(A_{2,1}^{*}\right)^{2} \\ &+ M_{p,h,3,3}\left(A_{2,1}\right)^{3} + 3M_{p,h,3,1}\left(A_{2,1}\right)^{2}\left(A_{2,1}^{*}\right) \\ &+ 3M_{p,h,3,-1}A_{2,1}\left(A_{2,1}^{*}\right)^{2} + M_{p,h,3,-3}\left(A_{2,1}^{*}\right)^{3} \\ &+ \dots \\ &+ \begin{bmatrix} \cdot & M_{p,h,m,n} & \cdot \end{bmatrix} \cdot \begin{bmatrix} \cdot \\ \binom{m}{k}\left(A_{2,1}\right)^{m-k}\left(A_{2,1}^{*}\right)^{k} \\ \cdot \end{bmatrix} \end{aligned} \tag{37}$$

where the value of n can now be calculated with $n = m - 2k$, the value of m equals to the polynomial order number [10] and k ranges from 0 to m. The positive integer $\binom{m}{k} = \frac{m!}{k!(m-k)!}$ is known as the binomial coefficient.

Rearranging (37) into matrix format with real and imaginary parts separated, then the $\begin{bmatrix} B_{p,h}^{R} \\ B_{p,h}^{I} \end{bmatrix}$ matrix can be calculated following the expansion of the dot product from matrices $\begin{bmatrix} \cdot & \left(M_{p,h,m,n}^{R} + jM_{p,h,m,n}^{I}\right) & \cdot \end{bmatrix}$ and $\begin{bmatrix} \cdot \\ \binom{m}{k}\left(A_{2,1}^{R}+jA_{2,1}^{I}\right)^{m-k}\left(A_{2,1}^{R}-jA_{2,1}^{I}\right)^{k} \\ \cdot \end{bmatrix}$. Hence, the formulation can be written with the binomial coefficients in the $A_{2,1}$ matrix expansion folded into the CM coefficients matrices, as (2).

In (2), the $[M_o]$ matrices with o as the polynomial order equals up to 3 are shown in the following as examples:

$$[M_0] = \begin{bmatrix} M_{p,h,0,0}^{R} \\ M_{p,h,0,0}^{I} \end{bmatrix} \tag{38}$$

$$[M_1] = \begin{bmatrix} \left(M_{p,h,1,1}^{R} + M_{p,h,1,-1}^{R}\right) & \left(-M_{p,h,1,1}^{I} + M_{p,h,1,-1}^{I}\right) \\ \left(M_{p,h,1,1}^{I} + M_{p,h,1,-1}^{I}\right) & \left(M_{p,h,1,1}^{R} - M_{p,h,1,-1}^{R}\right) \end{bmatrix} \tag{39}$$

$$\begin{aligned} [M_2] = &\begin{bmatrix} \left(M_{p,h,2,2}^{R} + M_{p,h,2,0}^{R} + M_{p,h,2,-2}^{R}\right) & \dots \\ \left(M_{p,h,2,2}^{I} + M_{p,h,2,0}^{I} + M_{p,h,2,-2}^{I}\right) & \dots \end{bmatrix} \\ &\begin{bmatrix} \dots & 2.\left(-M_{p,h,2,2}^{I} + M_{p,h,2,-2}^{I}\right) & \left(-M_{p,h,2,2}^{R} + M_{p,h,2,0}^{R} - M_{p,h,2,-2}^{R}\right) \\ \dots & 2.\left(-M_{p,h,2,2}^{R} - M_{p,h,2,-2}^{R}\right) & \left(-M_{p,h,2,2}^{I} + M_{p,h,2,0}^{I} - M_{p,h,2,-2}^{I}\right) \end{bmatrix} \end{aligned} \tag{40}$$

$$\begin{aligned} [M_3] = &\begin{bmatrix} \left(M_{p,h,3,3}^{R} + M_{p,h,3,1}^{R} + M_{p,h,3,-1}^{R} + M_{p,h,3,-3}^{R}\right) & \dots \\ \left(M_{p,h,3,3}^{I} + M_{p,h,3,1}^{I} + M_{p,h,3,-1}^{I} + M_{p,h,3,-3}^{I}\right) & \dots \end{bmatrix} \\ &\begin{bmatrix} \dots & \left(-3M_{p,h,3,3}^{I} - M_{p,h,3,1}^{I} + M_{p,h,3,-1}^{I} + 3M_{p,h,3,-3}^{I}\right) & \dots \\ \dots & \left(3M_{p,h,3,3}^{R} + M_{p,h,3,1}^{R} - M_{p,h,3,-1}^{R} - 3M_{p,h,3,-3}^{R}\right) & \dots \end{bmatrix} \\ &\begin{bmatrix} \dots & \left(-3M_{p,h,3,3}^{R} + M_{p,h,3,1}^{R} + M_{p,h,3,-1}^{R} - 3M_{p,h,3,-3}^{R}\right) & \dots \\ \dots & \left(-3M_{p,h,3,3}^{I} + M_{p,h,3,1}^{I} + M_{p,h,3,-1}^{I} - M_{p,h,3,-3}^{I}\right) & \dots \end{bmatrix} \\ &\begin{bmatrix} \dots & \left(M_{p,h,3,3}^{I} - M_{p,h,3,1}^{I} + M_{p,h,3,-1}^{I} - M_{p,h,3,-3}^{I}\right) \\ \dots & \left(-M_{p,h,3,3}^{R} + M_{p,h,3,1}^{R} - M_{p,h,3,-1}^{R} + M_{p,h,3,-3}^{R}\right) \end{bmatrix} \end{aligned} \tag{41}$$

The format (2) allows for the CM equation format to be compared to the ANN equation format.

APPENDIX B
ARTIFICIAL NEURAL NETWORK MODEL REFORMULATION

Following the original Artificial Neural Network (ANN) model structure, the calculation process can be presented in the matrix format, as (3). The *tanh* function is defined as:

$$\tanh\left(\underline{x}\right) = \frac{2}{1 + e^{-2(\underline{x})}} - 1 \tag{42}$$

The Taylor series expansion theory:

$$\begin{aligned} &f(x) \cong f(x_0) + \frac{f^{'}(x_0)}{1!}(x - x_0) + \\ &\frac{f^{''}(x_0)}{2!}(x - x_0)^2 + \frac{f^{'''}(x_0)}{3!}(x - x_0)^3 + ... \end{aligned} \tag{43}$$

The *tanh* function part in (3) can be reformulated as:

$$\begin{aligned} &\tanh\left([HW]\begin{bmatrix} A_{2,1}^{R} \\ A_{2,1}^{I} \end{bmatrix} + [HB]\right) \\ &= [\alpha] \\ &+ \Delta\,[\alpha]\,[HW]\begin{bmatrix} A_{2,1}^{R} \\ A_{2,1}^{I} \end{bmatrix} \\ &+ \Delta^2\,[\alpha]\,\left[HW^2\right]\begin{bmatrix} \left(A_{2,1}^{R}\right)^2 \\ A_{2,1}^{R}A_{2,1}^{I} \\ \left(A_{2,1}^{I}\right)^2 \end{bmatrix} \\ &+ \Delta^3\,[\alpha]\,\left[HW^3\right]\begin{bmatrix} \left(A_{2,1}^{R}\right)^3 \\ \left(A_{2,1}^{R}\right)^2 A_{2,1}^{I} \\ A_{2,1}^{R}\left(A_{2,1}^{I}\right)^2 \\ \left(A_{2,1}^{I}\right)^3 \end{bmatrix} \\ &\vdots \\ &+ \Delta^o\,[\alpha]\,[HW^o]\begin{bmatrix} \cdot \\ \left(A_{2,1}^{R}\right)^{o-k}\left(A_{2,1}^{I}\right)^{k} \\ \cdot \end{bmatrix} \end{aligned} \tag{44}$$

where according to (5) and (6), the $[HB]$ matrices can be calculated and shown in detail as:

$$\Delta\,[\alpha] = \frac{\partial \tanh\left([HB]\right)}{\partial\left([HB]\right)} = sech^2\left([HB]\right) \tag{45}$$

$$\Delta^2[\alpha] = \frac{1}{2}\frac{\partial^2 \tanh([HB])}{\partial([HB])^2} = -\tanh([HB]) sech^2([HB]) \tag{46}$$

$$\Delta^3[\alpha] = \frac{1}{6}\frac{\partial^3 \tanh([HB])}{\partial([HB])^3} = \frac{sech^2([HB])}{3}\{\tanh([HB])^2 - sech^2([HB])\} \tag{47}$$

Therefore, the original ANN model equation shown in (3) can now be reformulated as follows:

$$\begin{aligned}\begin{bmatrix} B^R_{p,h} \\ B^I_{p,h} \end{bmatrix} &= [OW][\alpha] \\ &+ [OW]\Delta[\alpha][HW]\begin{bmatrix} A^R_{2,1} \\ A^I_{2,1} \end{bmatrix} \\ &+ [OW]\Delta^2[\alpha]\left[HW^2\right]\begin{bmatrix} \left(A^R_{2,1}\right)^2 \\ A^R_{2,1}A^I_{2,1} \\ \left(A^I_{2,1}\right)^2 \end{bmatrix} \\ &+ [OW]\Delta^3[\alpha]\left[HW^3\right]\begin{bmatrix} \left(A^R_{2,1}\right)^3 \\ \left(A^R_{2,1}\right)^2 A^I_{2,1} \\ A^R_{2,1}\left(A^I_{2,1}\right)^2 \\ \left(A^I_{2,1}\right)^3 \end{bmatrix} \\ &\vdots \\ &+ [OW]\Delta^o[\alpha][HW^o]\begin{bmatrix} . \\ (A^R_{2,1})^{o-k}(A^I_{2,1})^k \\ . \end{bmatrix}\end{aligned} \tag{48}$$

where as mentioned, the two elements in (8) given by one hidden neuron in $[HW]$ matrix represent the real and imaginary parts of a complex number, because the defined FCC ANN model deals with datasets in real and imaginary parts respectively. Under this scenario, the $[HW]$ can written in the complex polar form as shown in (9).

Then $[HW^o]$ can then be simply computed (49).

$$[HW^o] = \begin{bmatrix} [\rho_1(\cos\theta_1 + \sin\theta_1)]^o \\ \vdots \\ [\rho_{NH}(\cos\theta_{NH} + \sin\theta_{NH})]^o \end{bmatrix} \tag{49}$$

since ρ_{NH} is set to limit the spread range of when the data is biased on the *tanh* activation function without exceeding a proper coverage range according to the non-linearity, it will be defined respectively unrelated to the phase angles when analysing the values of the $[HW]$ matrix. Therefore, the definition of the $[HW]$ matrices follows (8)-(10).

Bringing in the Fully Connected Cascaded (FCC) ANN structure for the dependent constant and linear elements inside the reformulated expanded equation in (48). Adding extra weights and biases in for correcting the constant and linear element generated by analyzing the 2^{nd} and 3^{rd} order respectively, where the original format of an FCC ANN equation (50) will then be reformulated as (4).

$$\begin{aligned}\begin{bmatrix} B^R_{p,h} \\ B^I_{p,h} \end{bmatrix} &= [OB] + [CW]\begin{bmatrix} A^R_{2,1} \\ A^I_{2,1} \end{bmatrix} \\ &+ [OW]\tanh\left([HW]\begin{bmatrix} A^R_{2,1} \\ A^I_{2,1} \end{bmatrix} + [HB]\right)\end{aligned} \tag{50}$$

REFERENCES

[1] I. Angelov, H. Zirath, and N. Rosman, "A new empirical nonlinear model for HEMT and MESFET devices," *IEEE Trans. Microw. Theory Techn.*, vol. 40, no. 12, pp. 2258–2266, Dec. 1992.

[2] I. Angelov, M. Thorsell, K. Andersson, A. Inoue, K. Yamanaka, and H. Noto, "On the large signal evaluation and modeling of GaN FET," *IEICE Transactions on Electronics*, vol. E93.C, no. 8, pp. 1225–1233, Aug. 2010.

[3] D. Root, "Technology-independent large-signal FET models: A measurement-based approach to active device modeling," in *15 ARMMS Conf., Bath*, Sept. 1991.

[4] D. E. Root, J. Xu, J. Horn, and M. Iwamoto, *The large-signal model: theoretical foundations, practical considerations, and recent trends*, ser. The Cambridge RF and Microwave Engineering Series. Cambridge: Cambridge University Press, 2011, p. 123–170.

[5] 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 International Microwave Symposium (IMS2014)*, Tampa, FL, USA, Jun. 2014, pp. 1–3.

[6] S. A. Albahrani, D. Mahajan, J. Hodges, Y. S. Chauhan, and S. Khandelwal, "ASM GaN: Industry standard model for GaN RF and power devices—part-ii: Modeling of charge trapping," *IEEE Trans. Electron Devices*, vol. 66, no. 1, pp. 87–94, Jan. 2019.

[7] U. Radhakrishna, T. Imada, T. Palacios, and D. Antoniadis, "MIT virtual source GaNFET-high voltage (MVSG-HV) model: A physics based compact model for HV-GaN HEMTs," *Phys. Status Solidi c*, vol. 11, no. 3-4, pp. 848–852, Mar. 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pssc.201300392

[8] Verspecht and D. Root, "Polyharmonic distortion modeling," *IEEE Microw. Mag.*, vol. 7, no. 3, pp. 44–57, Jun. 2006.

[9] D. E. Root, J. Verspecht, J. Horn, and M. Marcu, *X-parameters : characterization, modeling, and design of nonlinear RF and microwave components*, ser. The Cambridge RF and microwave engineering series. Cambridge: Cambridge University Press, Oct. 2013.

[10] P. J. Tasker and J. Benedikt, "Waveform inspired models and the harmonic balance emulator," *IEEE Microw. Mag.*, vol. 12, no. 2, pp. 38–54, Feb. 2011.

[11] M. Golio, L. Dunleavy, and T. Gneiting, "History and state-of-the-art in large signal modeling for RF/microwave power amplifier development," in *2015 IEEE MTT-S International Microwave Symposium*, Phoenix, AZ, USA, May 2015, pp. 1–4.

[12] J. Pedro and S. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

[13] V. Devabhaktuni, C. Xi, F. Wang, and Q.-J. Zhang, "Robust training of microwave neural models," in *IEEE MTT-S International Microwave Symposium Digest*, vol. 1, Anaheim, CA, USA, Jun. 1999, pp. 145–148 vol.1.

[14] J. Wood and D. Root, *Fundamentals of Nonlinear Behavioral Modeling for RF and Microwave Design*. Boston, Mass.: London: Artech House, 05 2005.

[15] J. Cai, J. Wang, C. Yu, H. Lu, J. Liu, and L. Sun, "An artificial neural network based nonlinear behavioral model for RF power transistors," in *2017 IEEE Asia Pacific Microwave Conference (APMC)*, Kuala Lumpur, Malaysia, Nov. 2017, pp. 600–603.

[16] J. Louro, C. Belchior, D. R. Barros, F. M. Barradas, L. C. Nunes, P. M. Cabral, and J. C. Pedro, "New transistor behavioral model formulation suitable for Doherty PA design," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 4, pp. 2138–2147, Apr. 2021.

[17] F. Feng, W. Na, J. Jin, J. Zhang, W. Zhang, and Q.-J. Zhang, "Artificial neural networks for microwave computer-aided design: The state of the art," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4597–4619, Nov. 2022.

[18] 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.

[19] J. de Villiers and E. Barnard, "Backpropagation neural nets with one and two hidden layers," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 136–141, Jan. 1993.
[20] T. Y. Kwok and D. Y. Yeung, "Constructive algorithms for structure learning in feedforward neural networks for regression problems," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 630–645, Mar. 1997.
[21] D. Hunter, H. Yu, M. S. Pukish, III, J. Kolbusz, and B. M. Wilamowski, "Selection of proper neural network sizes and architectures—a comparative study," *IEEE Trans. on Industrial Informatics*, vol. 8, no. 2, pp. 228–240, Feb. 2012.
[22] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design.* Boston, MS, USA: Artech House, 2000.
[23] A. Jarndal, S. Husain, and M. Hashmi, "Genetic algorithm initialized artificial neural network based temperature dependent small-signal modeling technique for GaN high electron mobility transistors," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 31, no. 3, p. e22542, Mar. 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.22542
[24] M. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.
[25] J. S. Smith, B. Wu, and B. M. Wilamowski, "Neural network training with Levenberg–Marquardt and adaptable weight compression," *IEEE Trans. Neural Netw.*, vol. 30, no. 2, pp. 580–587, Feb. 2019.
[26] M. Tian, J. J. Bell, R. Quaglia, and P. J. Tasker, "An extraction method based on artificial neural network techniques for novel Cardiff model with reasonable extrapolation behavior," *IEEE Microw. Wireless Technol. Lett.*, vol. 34, no. 1, pp. 5–8, Jan. 2024.
[27] R. J. Jesus, M. L. Antunes, S. N. D. R. A. da Costa, J. F. F. Mendes, and R. L. Aguiar, "Effect of initial configuration of weights on training and function of artificial neural networks," *Mathematics*, vol. 9, no. 18, p. 2246, Sept. 2021.
[28] G. Thimm and E. Fiesler, "High-order and multilayer perceptron initialization," *IEEE Trans. Neural Netw.*, vol. 8, no. 2, pp. 349–359, Feb. 1997.
[29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, vol. 9, pp. 249–256, May 2010.
[30] A. Khusro, S. Husain, M. S. Hashmi, and A. Q. Ansari, "Small signal behavioral modeling technique of gan high electron mobility transistor using artificial neural network: An accurate, fast, and reliable approach," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 30, no. 4, p. e22112, Apr. 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.22112
[31] H. Qi, J. Benedikt, and P. J. Tasker, "Nonlinear data utilization: From direct data lookup to behavioral modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 6, pp. 1425–1432, Jun. 2009.
[32] M. Rocio Moure, M. Casbon, M. Fernandez-Barciela, and P. J. Tasker, "A systematic investigation of behavioural model complexity requirements," in *13th European Microwave Integrated Circuits Conference (EuMIC)*, Madrid, Spain, Sept. 2018, pp. 77–80.
[33] P. J. Tasker, "Robust extraction of Cardiff model parameters from appropriately tailored measured load-pull data," in *2020 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS)*, Monterey, CA, USA, Nov. 2020, pp. 1–5.
[34] T. Husseini, A. Al-Rawachy, J. Benedikt, J. Bell, and P. Tasker, "Global behavioural model generation using coefficients interpolation," in *IEEE MTT-S International Microwave Symposium (IMS)*, Boston, MA, USA, Jun. 2019, pp. 200–203.
[35] B. Kalman and S. Kwasny, "Why tanh: choosing a sigmoidal function," in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, vol. 4, Baltimore, MD, USA, Jun. 1992, pp. 578–581 vol.4.
[36] E. M. Azad, J. J. Bell, R. Quaglia, J. J. Moreno Rubio, and P. J. Tasker, "Gate bias incorporation into Cardiff behavioural modelling formulation," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, Los Angeles, CA, USA, Aug. 2020, pp. 420–423.