Efficient energy-conservative dispersive transistor modelling using discrete-time convolution and artificial neural networks

Justin B. King

Department of Electronic and Electrical Engineering, Trinity College Dublin, Dublin, Ireland

Correspondence
Justin King, Department of Electronic and Electrical Engineering, Trinity College Dublin, Dublin, Ireland.
Email: justin.king@tcd.ie

Present Address
Justin King, Arás an Phiarsaigh, Trinity College Dublin, the University of Dublin, Dublin 2, Ireland.

Abstract
This paper describes an efficient simulation technique for field-effect transistor devices exhibiting significant frequency dispersive behavior. The device model is formulated such that a single energy function completely describes the channel displacement current, thus imposing charge and energy conservation principles from the outset. It is shown that adherence to energy conservation necessarily implies a reciprocal intrinsic displacement network, which is typically too rigid to follow measured device data. This problem is overcome by considering a frequency dispersive time-delay parameter, which has the effect of allowing nonreciprocal susceptance behavior to be captured by the overall device model. This dispersive time-delay parameter is included within the nonlinear device simulation via an efficient discrete-time convolution.

KEYWORDS
conservation, device model, dispersion, energy model, simulation

1 | INTRODUCTION

As we enter the next phase of the telecommunications era, the evidence suggests that the transformative change to the electronic engineering discipline, and to our societies, brought about by the recent generations of mobile communications network technologies is set to continue for the foreseeable future. A key enabler of this change is the rapidly increasing bitrates made possible with each new network generation. The Shannon-Hartley channel capacity theorem makes clear the necessary physical conditions required to meet this increasing demand, namely, increased bandwidth. The linear relationship between channel capacity (in bits per second) and the available bandwidth (in hertz) has led to a shift to the largely untapped mm-wave spectrum for the current fifth generation (5G) networks. Given the sublinear relationship between channel capacity and signal power, it is likely that the bitrate demands required of future generations of the telecommunications network infrastructure will be met through further excursions into the abundant natural resource available in the mm-wave and terahertz (THz) spectrum.

Continuation of this trend will be dependent on significant advances in physical layer technologies. Even today, there exists a wide range of advanced semiconductor processes capable of meeting this fundamental challenge, with reports of highly efficient silicon germanium (SiGe) transistors achieving amplification at close to THz frequencies, while reports of other, more exotic, compound devices, such as indium phosphide (InP) having surpassed this...
milestone, have been published some time ago. Nevertheless, as we have seen with the silicon versus gallium nitride (GaN) debate, the winning technology often depends on other factors, such as ease of fabrication, and the challenge becomes attaining maximum performance from the most cost-effective technologies.

As an example of the trend, consider the following. The Global System for Mobile Communications (GSM) network, still widely in use across Europe, operates with constant-amplitude signals via a modulation technique known as Gaussian Minimum Shift Keying (GMSK). This simple second generation (2G) network signaling very significantly alleviates the challenge on the RF physical layer; for example, power amplifiers may be operated close to saturation where they are most efficient without risk of clipping, while device models have relaxed gain accuracy requirements. Comparing this to the modern 5G network, operating with highly complex digitally modulated signals, such as orthogonal frequency division multiplexing (OFDM), at wide bandwidths, through highly efficient harmonically tuned power amplifiers, it is clear how the demands on the hardware have dramatically increased.

In particular, increasing frequencies and bandwidths place significant demands on the transistor device models. Modern design approaches have made the equivalent circuit modelling approach the natural choice; while behavioral models have their advantages in high fidelity and short development time, the equivalent circuit approach facilitates access to the intrinsic device, which is required for power amplifier designs based on waveform engineering. However, the standard equivalent circuit model topology suffers from many significant and well-known drawbacks.

Specifically, the traditional field effect transistor (FET) or high-electron mobility transistor (HEMT) equivalent circuit is based entirely on lumped circuit elements. There are two reasons for this: namely, (i) it is a simple approach that tends to work reasonably well at low-to-mid frequencies, (ii) distributed elements are more difficult to simulate in commercial simulators, particularly under transient conditions where short lines can cause significant numerical difficulties. Another issue with the standard equivalent circuit topology follows from the independent nature of the circuit element constitutive relations. In many models, independent capacitance functions are fit to the gate-source and gate-drain capacitances, which can lead to charge conservation issues. While this is addressed in some models through the use of a single capacitance function and the energy conservation issue is identified in Reference 12, no formal solution to ensure energy conservation within an equivalent circuit model while allowing an accurate fit to nonreciprocal capacitance measurement data has been presented.

This paper focuses on addressing these issues through a unified approach. Energy and charge conservation requirements are automatically met through the prescription of a single energy function that describes completely the device displacement current. The resulting reciprocal model is augmented with a frequency-dispersive transconductance time-delay, introducing an asymmetry into the imaginary part of the model Y-parameters that accounts for the nonreciprocal susceptance typically seen in real device data. This avoids the requirement for any dissipative capacitive networks and eliminates the need for the troublesome transcapacitance elements. The frequency-dispersion in the transconductance time-delay is implemented using a highly efficient discrete-time technique. Finally, the use of artificial neural networks (ANNs) ensures a generalized semi-automated modelling approach is followed, suitable for any device technology.

2 | ENERGY-BASED DISPLACEMENT MODEL

Figure 1A shows an extrinsic parasitic network surrounding an intrinsic device model. The gate and drain shunt RC circuits account for the lossy silicon substrate inherent in certain device types, such as the GaN-on-Si device used for this...
work. All parasitic elements can be identified from measured data and removed according to standard modelling techniques.\textsuperscript{15,16} This allows the modelling task to focus on the intrinsic device, a commonly used topology for which is shown in Figure 1B, valid under small-signal conditions. In the intrinsic model, the current through each branch is separated into a conductive component due to the resistive elements, and a displacement component due to the capacitive elements.

The conductive gate leakage current is accounted for by $G_{gs}$ and $G_{gd}$, while the conductive drain current is due to $g_m$ and $g_{ds}$. Note that the transconductance $g_m$ exhibits a time-delay $\tau_m$, modelling the transit time delay of the channel, that is, the applied gate-source voltage gives rise to a current in the channel with an associated time delay, due to finite charge redistribution time and the finite travel time within the distributed channel. A similar delay is sometimes associated with the output conductance $g_{ds}$, particularly in GaN devices.\textsuperscript{17,18} The gate displacement current is accounted for by $C_{gs}$ and $C_{gd}$, while the channel displacement current is accounted for by $C_{ds}$.

This separation of conductance and displacement current contributions in the model of Figure 1B can be made explicit through the use of small-signal conductance and capacitance two-port networks, as shown in Figure 2. In Figure 2A, the conductance $G_{11}$ accounts for the gate conduction current due to the local port-1 voltage $v_g$, while the transconductance $G_{12}$ accounts for the gate conduction current due to the remote port-2 voltage $v_d$. A similar explanation for the drain conduction current, in terms of $G_{21}$ and $G_{22}$, applies at port-2. The situation for the device displacement current is entirely analogous, using capacitances instead of conductances, as shown in Figure 2B.

Mathematically, we can represent the total admittance $Y$-parameters of the parallel combination (in a two-port sense) of the two networks in Figure 2, as

$$Y(v_g, v_d, \omega) = G(v_g, v_d) + j\omega C(v_g, v_d),$$

(1)

where

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} G_{gs} + G_{gd} & -G_{gd} \\ g_m - G_{gd} & g_{ds} + G_{gd} \end{bmatrix},$$

(2)

and

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} C_{gs} + C_{gd} & -C_{gd} \\ -C_{gd} & C_{ds} + C_{gd} \end{bmatrix},$$

(3)

where the relationship between the two-ports of Figure 2 and the elements of small-signal intrinsic model of Figure 1B have been explicitly identified. In principle, it is possible to extract the elements of $G$ and $C$ across a range of biases, and develop large-signal two-port conductance and displacement networks with bias-dependent elements, as indicated by the explicit bias dependence in Equation (1). However, implementing the displacement current using arbitrary bias-dependent capacitances can lead to violation of terminal charge conservation, as well as numerical issues related to derivative approximation in circuit simulators.\textsuperscript{8,10} For these reasons, it is generally preferred to model the underlying gate and drain charge functions, as shown in Figure 3A. The capacitance matrix $C$ is related to these gate and drain charge functions via the gradient with respect to the port voltages $v_g$ and $v_d$, that is,
Note that the existence of the charge functions $Q_g$ and $Q_d$ implies that the capacitance functions obey the charge conservation condition, sometimes written as Reference 11

$$\frac{\partial C_{11}}{\partial v_d} = \frac{\partial C_{12}}{\partial v_g} \tag{5}$$

$$\frac{\partial C_{21}}{\partial v_d} = \frac{\partial C_{22}}{\partial v_g} \tag{6}$$

To ensure such “potential” functions exist, this condition can be imposed during model extraction.\(^{10}\) Note that so far we have made no assumptions on the symmetry of $C$, that is, in general, there is no requirement that $C_{12} = C_{21}$.

In order to include energy in this formulation, we note that charge $Q$ and energy $U$ are related by the definition of electric potential, that is,

$$U = \int Q dv. \tag{7}$$

For energy to be conserved in a two-port charge-based displacement current model, as implemented in Figure 3A, with the charge functions $Q_g$ and $Q_d$ derived via a 2D integration (with respect to the port voltages) of the capacitance data obeying Equation (4), we require

$$\frac{\partial Q_g}{\partial v_d} = \frac{\partial Q_d}{\partial v_g} \tag{8}$$

or, equivalently, that an energy function $U(v_g,v_d)$ must exist for the two-port, such that

$$\begin{bmatrix} Q_g \\ Q_d \end{bmatrix} = \nabla U(v_g,v_d) = \begin{bmatrix} \frac{\partial U}{\partial v_g} \\ \frac{\partial U}{\partial v_d} \end{bmatrix}. \tag{9}$$

The corresponding circuit implementation is shown in Figure 3B. Note how the full two-port is described by a single energy function $U(\cdot)$, from which follow the two charges [via Equation (9)], and the four capacitances [via Equation (9) then Equation (4)]. The price paid for ensuring energy conservation is the rather rigid structure imposed on the
capacitance matrix by Equation (8), which stipulates that \( \mathbf{C} \) is symmetric and hence that the two-port network of Figure 3B must be reciprocal. It is important to realize that \( C_{12} = C_{21} \) is a particularly strong condition to impose. Moreover, a symmetric \( 3m|\mathbf{Y}| \) generally not supported by experimental measurements on real transistor devices,\(^{13} \) which leads us to an apparent contradiction.

Following Reference 19, we consider that any proper displacement current component should obey the principle of energy conservation, and hence should indeed be implemented in model form using an energy-conservative network in the form of Figure 3. To account for the asymmetry seen experimentally in \( 3m|\mathbf{Y}| \), we again examine the conduction current network shown in Figure 2A. For simplicity, we can consider the device to be reverse biased at the gate, or to have an insulated gate, giving

\[
\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ g_m \exp[-j\tau_m(\omega)] & g_m \end{bmatrix}
\]

where \( g_m \) is the transconductance (sometimes referred to as mutual conductance, hence the \( m \) subscript), and \( \tau_m(\omega) \) is the transconductance time-delay parameter. Note that, in general, \( \tau_m(\omega) \) depends on frequency in order to account for frequency-dispersive behavior, while also showing bias dependence. By expanding \( G_{21} \), we obtain

\[
G_{21} = g_m \cos[(\omega \tau(\omega))] - jg_m \sin[(\omega \tau(\omega))],
\]

which, we see, has an imaginary component. This component exists in \( G_{21} \) only (i.e., it does not appear in \( G_{12} \)), and it therefore introduces an asymmetry into \( 3m|\mathbf{Y}| \) as

\[
3m|\mathbf{Y}| = \begin{bmatrix} \omega C_{11} \\ \omega C_{12} - g_m \sin[(\omega \tau(\omega))] \\ \omega C_{22} \end{bmatrix},
\]

which we can use to account for the asymmetry in the measured admittance parameter data. The permitted frequency dispersion in \( \tau_m \) helps considerably in this regard, albeit at the cost of a more complex model. While such frequency-dependent parameters may be handled in a straightforward manner using frequency domain simulations methods (such as harmonic balance), they have been traditionally difficult to incorporate into time domain transient simulations. An efficient solution to this problem will be described in Section 4.

In a more general treatment, the output conductance also could be considered to exhibit some time-delay,\(^ {17,18} \) which would contribute a term to \( 3mY_{22} \), although this is less significant since in the present formulation there is no requirement for \( C_{11} = C_{22} \) that is, \( C_{11} \) and \( C_{22} \) may be fit independently in any case.

### 3 | ENERGY MODELLING WITH ARTIFICIAL NEURAL NETWORKS

The two-port energy function \( U(v_g, v_d) \) pertaining to a particular device cannot be measured directly. From Equations (7) and (9), we see that an integral relationship exists between energy and charge, and from Equation (4), we see a similar integral relationship exists between charge and capacitance, giving the following relationship between the capacitance matrix and the energy function

\[
\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix} = \mathbf{H}_U(v_g, v_d) = \begin{bmatrix} \frac{\partial^2 U}{\partial v_g^2} & \frac{\partial^2 U}{\partial v_g v_d} \\ \frac{\partial^2 U}{\partial v_d v_g} & \frac{\partial^2 U}{\partial v_d^2} \end{bmatrix},
\]

that is, the capacitance matrix is determined from the Hessian of the scalar energy function \( U \). This is useful: while we cannot measure \( U \), we can extract capacitance data from measured small-signal S-parameters. We could then attempt
to integrate these capacitance data, twice, and fit an analytically defined energy function to the resulting energy data. However, there are at least two problems with the above approach:

1. Integrating numerically, twice, will introduce significant errors into the computed final energy data \( \hat{U} \);
2. If there remains even a small degree of asymmetry in \( C \), the numerical integration will be path dependent.

With path-dependent integration, the computed final energy data \( \hat{U} \) depend on which integration path has been chosen to obtain it, and it is difficult to know which path leads to the least error. Another problem is that it is often not easy to find an analytical function that will easily fit the energy data \( \hat{U} \).

Almost all the above issues can be alleviated by representing the two-port energy function \( U \) by an ANN of the form shown in Figure 4. This ANN has two inputs that is, the two port voltages \( v_g \) and \( v_d \), has one hidden layer which may comprise of any number of neurons in general, and one output, corresponding to the modelled two-dimensional two-port energy function \( \bar{U}(v_g,v_d) \). Using matrix notation, we can express the output \( \bar{U} \) as

\[
\bar{U} = AW^{(2)},
\]

where

\[
A = f(vW^{(1)}).
\]

It is seen that \( \bar{U}(v_g,v_d) \) is a linear combination \( AW^{(2)} \) of the \( 1 \times m \) activation matrix \( A \) from the hidden layer neurons scaled by the \( m \times 1 \) second layer weight matrix \( W^{(2)} \). The activation \( A \) consists of a function \( f \) applied element-wise to the \( 1 \times m \) matrix \( vW^{(1)} \) that is, to the linear combination of the input voltages \( w_1v_g + w_2v_d \) in each column \( i \). The function \( f \) is usually of sigmoidal type for example, \( \tanh(\cdot) \), the \( 2 \times m \) matrix \( W^{(1)} \), and the \( m \times 1 \) matrix \( W^{(2)} \) are the layer-1 and layer-2 weights, respectively, and \( v = [v_g, v_d] \).

For this work, the hidden layer consists of 15 neurons. This implies \( 2 \times 15 + 15 \times 1 = 45 \) weight parameters. As with all ANNs, it is useful to scale the input to a limited range for example, \([-1, 1]\), using an affine transformation at the input. Similarly, the output should be trained to lie also in a limited range, with a second affine function scaling this back to the appropriate size for the target variable (\( \bar{U}(v_g,v_d) \) in this case). This aids the optimization process, ensuring the weights lie within a restricted range. In this work, the weights are initialized to a random value in the range \([-1, 1]\) at the beginning of the optimization procedure.

Using this method to model the energy function allows the 4 s partial derivatives of \( \bar{U}(v_g,v_d) \) to be obtained analytically, which is a significant benefit. These partial derivatives correspond to the modelled approximations of the four entries of the extracted capacitance matrix, as indicated in Figure 4. The ANN can now be trained directly on these data according to the following optimization problem

\[
\min_{\mathbf{w}} \sum_{v_g} \sum_{v_d} \left( \sum_{i=1}^{2} \sum_{j=1}^{2} \left[ \hat{C}_{ij}(w,v_{gs},v_{ds}) - C_{ij}(v_{gs},v_{ds}) \right]^2 \right),
\]

where the two outer summations are over all gate-source and drain-source bias voltages, while the inner two summations are over all four elements of the capacitance matrix. The quantity \( \mathbf{w} \) is a vector of all ANN weights pertaining to both layers, which are adjusted during optimization to achieve the minimum modelled versus measured error.

Thus, no numerical integration is required; we fit extracted capacitance data over all bias points to the partially differentiated output of the ANN by adjusting the ANN weights. Ultimately, this provides us with an energy function \( \bar{U} \) that ensures the capacitances are close to our measured data, and which are also, by design, energy conservative. Any asymmetry error in \( C \) is evenly spread over the energy function. This approach also allows at least some degree of self-consistency reporting: if the measured data are not energy conservative, the error in the ANN fit will remain high. Finally, this energy function may be implemented directly in a circuit simulator to obtain the displacement current according to Figure 3B.
4 | INCORPORATING FREQUENCY DISPERSION USING A DISCRETE-TIME METHOD

4.1 | Adding a transmission line to the intrinsic equivalent circuit model

The discussion in Section 2 shows that in an energy-conserving model, which must necessarily have a reciprocal displacement current network, the transconductance time delay $\tau_m$ can be used to introduce the required asymmetry in the model susceptance matrix, $3m[Y]$. To explain how this is achieved, we return to element $3m[Y_{21}]$ from Equation (12). We require values for $C_{21}$, $g_m$, and $\tau_m$ such that the following optimization problem is satisfied

$$\min_{C_{21}, g_m, \tau_m} \left| Y_{21}^{\text{mod}}(C_{21}, g_m, \tau_m) - Y_{21}^{\text{meas}} \right|$$ (17)

where $Y_{21}^{\text{meas}}$ and $Y_{21}^{\text{mod}}$ represent the measured and modelled device $Y_{21}$ parameter, respectively, with $3m[Y_{21}^{\text{meas}}] = \omega C_{12} - g_m \sin[\omega r_m(\omega)]$. While almost all published compact models consider $\tau_m$ as a fixed value across frequency, in this work, it is proposed that a small but finite level of frequency dispersion is permitted that is, it is necessary for this time delay to be at least somewhat frequency dependent that is, $\tau_m = \tau_m(\omega)$. This improves the feasibility of Equation (17) and is justified by the inherent dispersion associated with the charge transit in the distributed channel under the gate. The resulting benefit of this approach is that a highly accurate $S$-parameter model, across a broad frequency range, is possible despite the symmetry imposed by the energy-based displacement current model.

The dispersion in $\tau_m$ can be represented by a transmission line, such as that shown in Figure 5A, with frequency-dependent propagation constant $\gamma(\omega)$ leading to a frequency-dependent phase constant $\beta(\omega) = \omega / \tau$, where $\tau$ is the line transit delay. The line models the true time delay that exists in any device exhibiting non-quasistatic behavior, that is, where there is an inherent delay in modulations of the intrinsic gate-source voltage $v_g$ effecting the steady-state charge distributions in the channel. The port-1 end of the transmission line is fed with the intrinsic gate voltage $v_1(t) = v_g$ and thus the port-2 voltage $v_2(t) = v_{gd}$ represents the same voltage after undergoing a frequency-dependent time delay. This delayed voltage is then used as the gate control voltage for the drain current source, as shown in Figure 5B. The line is matched at port-2 to ensure no reflections occur (for simplicity, the line $Z_0 = 1\Omega$).

Traditionally, due to simulator difficulties with variable time delays in full large-signal nonlinear transient simulations, the transconductance parameter is set to a fixed value, typically the average or median delay value within the device active region. In some simplistic models, the time delay is ignored altogether. Considering the discussion in Section 2, this has serious implications for the underlying physical basis of the model, and means that one has to accept that either the modelled $3m[Y]$ is symmetric, or the modelled intrinsic capacitances are adjusted to allow an asymmetric

FIGURE 4 ANN for the modelled energy function $\hat{U}$ with the four partial derivatives of this quantity giving the modelled approximation of the four entries of the extracted capacitance matrix $C$
Thus violating energy conservation. In the next section, it is shown how such a frequency-dependent time delay may be implemented within a transient simulation, using an efficient discrete time technique.

### 4.2 Time-domain S-parameters

The scattered wave $B_2$ on port-2 of the transmission line in Figure 5A may be related to the incident wave $A_1$ on port-1 through the line S-parameters as

$$B_2(3\omega) = S_{21}(3\omega)A_1(3\omega) + S_{22}(3\omega)A_2(3\omega),$$

where $A_2$ is the port-2 incident wave, which is zero due to the matched impedance at port-2. Similarly, there will be no scattered wave $B_1$ at port-1, hence we can rewrite Equation (18) in terms of the line port voltages as

$$V_2(3\omega) = S_{21}(3\omega)V_1(3\omega),$$

where $Z_0 = 1\ \Omega$ and $V_1(3\omega)$ and $V_2(3\omega)$ are the frequency domain port-1 and port-2 voltages, respectively. At a given frequency, $V_2$ is simply a time-delayed version of $V_1$. It is straightforward to determine the forward scattering parameter $S_{21}(3\omega)$ from the phase constant $\beta(\omega)$, as $S_{21} = e^{-3\beta(\omega)}$, where we have assumed a unit-length transmission line.

To be able to use the result of Equation (19) within a transient simulation, it is necessary to convert this equation to the time domain. If we consider $S_{21}(3\omega)$ to be periodic in the frequency domain, we could use the inverse discrete-time Fourier transform (DTFT) to obtain a discrete-time signal $v_2[n]$. If we also tabulate $S_{21}(3\omega)$ at discrete values of $\omega = \omega_k$, for $k \in \mathbb{Z}$, we can instead employ the discrete Fourier transform (DFT) as a practically implementable technique (either directly, or via the fast Fourier transform) to convert Equation (19) to the time-domain as the following discrete-time convolution

$$v_2[n] = s_{21}[n] * v_1[n].$$

Since transmission line frequency responses are inherently periodic across frequency (under normal conditions), this is usually a straightforward conversion. Non-periodic situations can be easily handled in any case. In addition, due to the smooth and bounded nature of the scattering parameters, the discrete-time scattering impulse response $s_{21}[n]$ is typically very compact, and may be effectively implemented as an finite impulse response filter, requiring a minimal number of taps, as indicated in Equation (21)

$$v_2[n] = \sum_{i=0}^{N} s_{21}[i] * v_1[n-i].$$
with \( N \) depending on the maximum frequency of interest but use no more than a few hundred taps. This discrete time approach is preferred to any continuous-time convolution implementation, where numerical integration would be required. In the discrete-time case, only sum-of-product type computations are required, which are very fast to implement on modern computers.

5 | RESULTS

To verify the proposed energy-based modelling method, small-signal models have been extracted for an 8 × 125 μm 10 W GaN-on-Si HEMT device, across a range of bias points. Figure 6 shows the fit obtained for the capacitance data from a single energy function \( U \) that is, \( C_{12} = C_{21} \). It can be seen that almost all capacitances are modelled with high fidelity, apart from a small inaccuracy in \( C_{22} \) in the high current region, where the conduction current tends to dominate the device response. This fit is achieved using just 15 neurons in the hidden layer.

Figure 7A shows the time delay parameter \( \tau_m \) (red squares) versus frequency, extracted from the (de-embedded) intrinsic device \( S \)-parameters, with dispersion clearly noticeable across the band. Also shown in Figure 7A is the modelled dispersion relation (blue line), which is used to synthesize the \( S \)-parameters \( S_{21}(j\omega) \) of the transmission line used within an equivalent circuit model, as per Figure 5. This frequency-dependent delay is implemented within the nonlinear large-signal transient simulation through use of a discrete-time convolution with the discrete-time impulse response \( s_{21}[n] \) (obtained via an inverse DFT of \( S_{21}(j\omega_k), k \in \mathbb{Z} \) ) shown in Figure 7B.

The measured and modelled \( S \)-parameters corresponding to two separate bias points may be seen in Figures 8 and 9. Also provided in Figure 8 are the \( S_{11} \) and \( S_{12} \) parameters (blue lines with circles) from an alternative small-signal model extracted in the traditional way without any attempt to ensure energy conservation. As can be seen, these modelled parameters show significant error compared with the measured data. The poor fit for the non-energy-conserving model is the result of nonphysical capacitance extraction due to the forcing of a fixed value of \( \tau_m \). The energy-based

![FIGURE 6](image_url) 3D plots of the extracted and modelled entries of the capacitance matrix \( C \) across the bias plane

![FIGURE 7](image_url) Shown in (A) is the extracted (●) and circuit-simulator-implemented (—) time delay parameter \( \tau_m \) across frequency for \( V_G = -2.00 \text{ V} \) and \( V_D = 0.2 \text{ V} \); while (B) shows the time-domain scattering impulse response \( s_{21}[n] \) which decays to zero beyond 150 ps, as shown
FIGURE 8  Modelled (---) and measured (○) S-parameters up to 10 GHz for the energy-conservative model. The device is biased at $V_G = -1$ V and $V_D = 5$ V. Also shown are the modelled $S_{11}$ and $S_{12}$ parameters from a corresponding non-energy conservative model (----) which shows a significant error compared to the energy-conserving model.

FIGURE 9  Modelled (---) and measured (○) S-parameters 10 GHz for the energy-conservative model at $V_G = -0.2$ V and $V_D = 2$ V.

FIGURE 10  Shown is a sinusoidal gate voltage $v_g$ (---) with unit amplitude at 9 at the input to the transmission line $[v_1(t)]$, along with the delayed version of this voltage $v_{gd}$ (○•) at the output of the line $[v_2(t)]$. This delayed voltage is determined using the proposed discrete time convolution method, with the initial ‘dead-zone’ visible at the start of the waveform. Also shown for comparison is the corresponding harmonic balance-derived delayed gate voltage (——) which is in perfect agreement with the convolution-based transient simulation in the steady state.
model permits a variable \( \tau_m(\omega) \) across frequency, allowing frequency dispersion in the model, thus enabling the extraction of accurate intrinsic capacitances on a model framework with a solid physical basis.

Finally, Figure 10 shows the results of a large-signal transient simulation using an input voltage waveform at 1 GHz, and the corresponding steady-state output waveform exhibiting the appropriate time delay. Also shown for validation purposes is the corresponding output from a frequency domain version of the model (where a time delay is trivial to implement) simulated using the harmonic balance technique.

6 | CONCLUSIONS

A method is demonstrated for associating asymmetry in the device capacitance matrix with the transconductance time-delay thus enabling formulation of an energy and charge conservative displacement current model. Results show that a single energy function can account for all device capacitances, and is well suited to circuit simulator implementation using an ANN. A discrete-time technique is used to facilitate inclusion of the frequency-dependent time delay into a large-signal transient simulation. Further work is required to handle voltage-dependent time delays, which cannot be implemented through the linear convolution proposed in this work. Validation of the large-signal model through measurement data should also be carried out. Finally, theory would suggest to ensure causality the transmission line dispersion and attenuation are related (via the Kramers-Kronig relations), further investigation is required to take this constraint into account correctly.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID
Justin B. King https://orcid.org/0000-0002-5144-1821

REFERENCES

1. Ericsson Mobility Report – June 2020. https://www.ericsson.com/en/mobility-report/reports/june-2020. Accessed September 30, 2020.
2. Chakraborty PS, Cardoso AS, Wier BR, et al. A 0.8 THz \( f_{\text{MAX}} \) SiGe HBT operating at 4.3 K. IEEE Electron Device Lett. 2014;35(2):151-153.
3. Mei X, Yoshida W, Lange M, et al. First demonstration of amplification at 1 THz using 25-nm InP high electron mobility transistor process. IEEE Electron Device Lett. 2015;36(4):327-329.
4. Spooner CM, Nicholls RB. Chapter 18 - Spectrum sensing based on spectral correlation. In: Fette BA, ed. Cognitive Radio Technology. 2nd ed. Oxford, England: Academic Press; 2009:593-634.
5. King JB, Brazil TJ. Nonlinear electrothermal GaN HEMT model applied to high-efficiency power amplifier design. IEEE Trans Microw Theory Tech. 2013;61(1):444-454.
6. Sharma T, Darrai J, Ghannouchi F. High efficiency continuous mode power amplifiers using waveform engineering. Paper presented at: Proceedings of 2014 Mediterranean Microwave Symposium (MMS2014); December 12-14, 2014; Marrakech, Morocco:1-4.
7. Mokht ZA, Tasker PJ, Lees J. Using waveform engineering to optimize Class-F power amplifier performance in an envelope tracking architecture. Paper presented at: 2014 9th European Microwave Integrated Circuit Conference; October 6-7, 2014; Rome, Italy:357-360.
8. King JB, Wilson C. Charge conservative FET modelling using ANNs. Paper presented at: 2017 12th European Microwave Integrated Circuits Conference (EuMIC); October 8-10, 2017; Nuremberg, Germany.
9. Root DE. Nonlinear charge modelling for FET large-signal simulation and its importance for IP3 and ACPR in communication circuits. Paper presented at: Proceedings of the 44th IEEE 2001 Midwest Symposium on Circuits and Systems. MWSCAS 2001 (Cat. No.01CH37257); August 14-17, 2001; Dayton, OH:768-772.
10. Wilson C, King JB. Ensuring charge conservation in GaN HEMT Large-Signal Models. Paper presented at: 2018 13th European Microwave Integrated Circuits Conference (EuMIC); September 23-25, 2018; Madrid, Spain.
11. Gibo GP, Santarelli A, Filicori F. Charge-conservative GaN HEMT nonlinear modeling from non-isodynamic multi-bias S-parameter measurements. Int J Microw Wireless Technol. 2019;11(5-6):431-440. https://doi.org/10.1017/S1759078719000059.
12. Schmidt-Szatowski M. Energy-based capacitance modeling for field-effect transistor stability analysis. Paper presented at: 2019 14th European Microwave Integrated Circuits Conference (EuMIC); September 2019:124-127.
13. Pedro JC, Root DE, Xu J, Nunes LC. Nonlinear Circuit Simulation and Modeling: Fundamentals for Microwave Design. The Cambridge RF and Microwave Engineering Series. Cambridge: Cambridge University Press; 2018.
14. Root DE, Hughes B. Principles of nonlinear active device modeling for circuit simulation. Paper presented at: 32nd ARFTG Conference Digest; December 1-2, 1988; Tempe, AZ:1-24.
15. Crupi G, Schreurs DMMP, Caddemi A. On the small signal modeling of advanced microwave FETs: a comparative study. *Int J RF Microw Comput-Aided Eng.* 2008;18(5):417-425. https://doi.org/10.1002/mmce.20300.

16. Crupi G, Xiao D, Schreurs DMMP, et al. Accurate multibias equivalent-circuit extraction for GaN HEMTs. *IEEE Trans. Microw. Theory Tech.* 2006;54(10):3616-3622.

17. Brady RG, Oxley CH, Brazil TJ. An improved small-signal extraction algorithm for GaN HEMT devices. *IEEE Trans. Microw. Theory Tech.* 2008;56(7):1535-1544.

18. Crupi G, Schreurs DMMP, Raffo A, Caddemi A, Vannini G. A new millimeter-wave small-signal model modeling approach for pHEMTs accounting for the output conductance time delay. *IEEE Trans Microw Theory Tech.* 2008;56(4):741-746.

19. Wilson C, Schmidt-Szalowski M, King JB. Energy and Charge Conservation for FET Models. Paper presented at: 2020 15th European Microwave Integrated Circuits Conference (EuMIC); January 10-15, 2021; Utrecht, Netherlands.

20. Brazil TJ. Nonlinear, transient simulation of distributed RF circuits using discrete-time convolution. Paper presented at: 2007 IEEE International Symposium on Circuits and Systems; May 27-30, 2007; New Orleans, LA:505-508.

How to cite this article: King JB. Efficient energy-conservative dispersive transistor modelling using discrete-time convolution and artificial neural networks. *Int J Numer Model.* 2021;1–12. https://doi.org/10.1002/jnm.2894